USGS
science for a changing world

# A Theory for Modeling Ground-Water Flow in Heterogeneous Media

Professional Paper 1679

U.S. Department of the Interior
U.S. Geological Survey

# A Theory for Modeling Ground-Water Flow in Heterogeneous Media

By Richard L. Cooley

**U.S. Department of the Interior**
Gale A. Norton, Secretary

**U.S. Geological Survey**
Charles G. Groat, Director

U.S. Geological Survey, Reston, Virginia: 2004

For sale by U.S. Geological Survey, Information Services
Box 25286, Denver Federal Center
Denver, CO 80225

For more information about the USGS and its products:
Telephone: 1-888-ASK-USGS
World Wide Web: http://www.usgs.gov/

# CONTENTS

## FIGURES

## TABLES

# Principal Notation

| Variable | Page Where Defined | Variable | Page Where Defined | Variable | Page Where Defined |
|---|---|---|---|---|---|
| $\tilde{\mathbf{A}}$ | 196 | $\mathbf{f}_o(\cdots)$ | 72 | $q$ | 40, 89 |
| $\tilde{\mathbf{A}}_a$ | 202 | $F(p,n-p)$ | 46 | $\mathbf{q}$ | 28, 130 |
| $a$ | 48 | $F_a(p,n-p)$ | 46 | $\tilde{\mathbf{q}}$ | 144, 150 |
| $\hat{B}_L$ | 86 | $\tilde{\mathbf{F}}_i$ | 78, 196 | $\tilde{\mathbf{q}}_a$ | 156 |
| $\hat{B}_U$ | 86 | $\tilde{\mathbf{F}}_{ai}$ | 201 | $\mathbf{R}$ | 28 |
| $b$ | 47 | $g(\cdots)$ | 33 | $\hat{\mathbf{R}}$ | 49 |
| $\mathbf{C}_i$ | 50, 193 | $g_0(\cdots)$ | 76 | $\mathbf{R}_a$ | 62, 63, 157 |
| $\tilde{\mathbf{C}}_i$ | 78, 195 | $h$ | 5, 89 | $S(\cdots)$ | 27 |
| $\mathbf{C}_{ai}$ | 198 | $h(\cdots)$ | 155 | $S_a(\cdots)$ | 60, 61 |
| $\tilde{\mathbf{C}}_{ai}$ | 201 | $\mathbf{I}$ | 19 | $s^2$ | 51 |
| $\mathbf{C}_p$ | 60 | $\mathbf{I}_a$ | 62 | $T$ | 13, 89 |
| $c$ | 49 | $\mathbf{J}$ | 135 | $t(n-p)$ | 53 |
| $c_c$ | 54 | $\mathbf{J}_\beta$ | 138 | $t_{\alpha/2}(n-p)$ | 55 |
| $c_p$ | 61 | $L(\theta,\lambda)$ | 51, 55, 56 | $tr(\cdots)$ | 18 |
| $c_r$ | 48 | $L(\theta,\upsilon,\lambda)$ | 65, 68 | $\mathbf{U}$ | 27 |
| $\mathbf{Df}$ | 11, 29 | $\mathbf{l}$ | 28, 130 | $\mathbf{U}_*$ | 28 |
| $\mathbf{D}\hat{\mathbf{f}}$ | 85, 129 | $\mathbf{l}_*$ | 160 | $U_p^*$ | 59 |
| $\mathbf{D}^2\mathbf{f}$ | 11 | $\tilde{\mathbf{l}}$ | 144,146,147 | $\mathbf{U}_{*a}$ | 62, 155 |
| $\mathbf{D}g$ | 11 | $\tilde{\mathbf{l}}_*$ | 161 | $V_p$ | 60 |
| $\mathbf{D}^2 g$ | 11 | $\tilde{\mathbf{l}}_a$ | 155 | $V_{mx}$ | 55 |
| $\mathbf{D}_\beta\mathbf{f}$ | 11, 17 | $\tilde{\mathbf{l}}_{*a}$ | 155 | $V_{mxa}$ | 68 |
| $\mathbf{D}_\beta^2\mathbf{f}$ | 11, 17 | $\ell(\theta)$ | 41 | $\mathbf{V}_\beta$ | 13, 18, 19 |
| $\mathbf{D}_\beta g$ | 11, 33 | $\ell_a(\theta,\upsilon)$ | 67 | $\mathbf{V}_e$ | 15 |
| $\mathbf{D}_\beta^2 g$ | 11, 33 | $\hat{M}_{\min}$ | 86 | $\mathbf{V}_*$ | 21 |
| $\mathbf{D}_a\mathbf{f}_a$ | 157 | $m$ | 12 | $\mathbf{V}_{*a}$ | 62 |
| $\mathbf{D}_a^2\mathbf{f}_a$ | 156 | $m_i$ | 13 | $Var(\cdots)$ | 18 |
| $\mathbf{D}_a h$ | 156 | $\hat{N}$ | 87 | $W$ | 13, 89 |
| $\mathbf{D}_a^2 h$ | 156 | $N_{\min}$ | 75 | $W_p$ | 60 |
| $\mathbf{d}$ | 178 | $\hat{N}_{\min}$ | 85 | $\mathbf{W}_a$ | 60, 155 |
| $E(\cdots)$ | 18 | $n$ | 15 | $\mathbf{W}_p$ | 60 |
| $\mathbf{e}$ | 17 | $n_k$ | 40 | $w_{GK}$ | 40, 41 |
| $\mathbf{e}_*$ | 17 | $p$ | 13 | $Y_p$ | 45, 59 |
| $f(\cdots)$ | 12 | $\mathbf{Q}$ | 34 | $\mathbf{Y}$ | 15 |
| $\mathbf{f}(\cdots)$ | 15, 17, 28 | $\hat{\mathbf{Q}}$ | 54 | $\mathbf{Y}_a$ | 155 |
| $\mathbf{f}_a(\cdots)$ | 155 | $\mathbf{Q}_a$ | 62, 63, 157 | $\mathbf{Z}$ | 28, 163 |

| Variable | Page Where Defined | Variable | Page Where Defined | Variable | Page Where Defined |
|---|---|---|---|---|---|
| $\mathbf{Z}_*$ | 160 | $\boldsymbol{\gamma}$ | 12 | $\sigma_\varepsilon^2$ | 15 |
| $\widetilde{\mathbf{Z}}$ | 147 | $\boldsymbol{\Delta}$ | 180 | $\upsilon$ | 60 |
| $\mathbf{Z}_{*a}$ | 174 | $\varepsilon_p$ | 59 | $\upsilon_*$ | 59 |
| $\widetilde{\mathbf{Z}}_a$ | 157 | $\boldsymbol{\varepsilon}$ | 15 | $\widetilde{\upsilon}$ | 67 |
| $\mathbf{1}$ | 49 | $\theta_p$ | 61 | $\boldsymbol{\phi}(\boldsymbol{\theta})$ | 35 |
| $\mathbf{1}_i$ | 13 | $\hat{\theta}_p$ | 61 | $\chi^2(\cdots)$ | 162 |
| $\alpha$ | 44, 73 | $\widetilde{\theta}_p$ | 61 | $\boldsymbol{\Omega}$ | 47 |
| $\widetilde{\alpha}$ | 78 | $\boldsymbol{\theta}$ | 16, 27 | $\boldsymbol{\Omega}_a$ | 62 |
| $\boldsymbol{\alpha}(\boldsymbol{\beta})$ | 35 | $\overline{\boldsymbol{\theta}}$ | 13 | $\omega_p$ | 63 |
| $\beta$ | 73 | $\boldsymbol{\theta}_*$ | 16 | $\hat{\omega}_p$ | 62 |
| $\boldsymbol{\beta}$ | 12 | $\hat{\boldsymbol{\theta}}$ | 28 | $\omega_{Gk}$ | 40 |
| $\gamma_I$ | 47, 53, 73, 77, 185, 187 | $\widetilde{\boldsymbol{\theta}}$ | 52 | $\boldsymbol{\omega}$ | 27 |
| $\hat{\gamma}_I$ | 46, 73, 185 | $\hat{\boldsymbol{\theta}}_0$ | 74 | $\boldsymbol{\omega}_i$ | 28 |
| $\gamma_w$ | 41, 53, 185, 187 | $\boldsymbol{\theta}_a$ | 155 | $\boldsymbol{\omega}_j$ | 28 |
| $\hat{\gamma}_w$ | 46, 185 | $\lambda$ | 51 | $\hat{\boldsymbol{\omega}}$ | 48 |
| $\gamma_{Ia}$ | 62, 82 | $\boldsymbol{\lambda}$ | 91 | $\boldsymbol{\omega}_G$ | 55 |
| $\hat{\gamma}_{Ia}$ | 61, 82 | $\xi$ | 86 | $\boldsymbol{\omega}_a$ | 63 |
| $\gamma_{wa}$ | 61 | $\xi_a$ | 113 | | |
| $\hat{\gamma}_{wa}$ | 61 | $\sigma_\beta^2$ | 13 | | |

# Abstract

Construction of a ground-water model for a field area is not a straightforward process. Data are virtually never complete or detailed enough to allow substitution into the model equations and direct computation of the results of interest. Formal model calibration through optimization, statistical, and geostatistical methods is being applied to an increasing extent to deal with this problem and provide for quantitative evaluation and uncertainty analysis of the model. However, these approaches are hampered by two pervasive problems: 1) nonlinearity of the solution of the model equations with respect to some of the model (or hydrogeologic) input variables (termed in this report system characteristics) and 2) detailed and generally unknown spatial variability (heterogeneity) of some of the system characteristics such as log hydraulic conductivity, specific storage, recharge and discharge, and boundary conditions. A theory is developed in this report to address these problems. The theory allows construction and analysis of a ground-water model of flow (and, by extension, transport) in heterogeneous media using a small number of lumped or smoothed system characteristics (termed parameters). The theory fully addresses both nonlinearity and heterogeneity in such a way that the parameters are not assumed to be effective values.

The ground-water flow system is assumed to be adequately characterized by a set of spatially and temporally distributed discrete values, $\boldsymbol{\beta}$, of the system characteristics. This set contains both small-scale variability that cannot be described in a model and large-scale variability that can. The spatial and temporal variability in $\boldsymbol{\beta}$ are accounted for by imagining $\boldsymbol{\beta}$ to be generated by a stochastic process wherein $\boldsymbol{\beta}$ is normally distributed, although normality is not essential. Because $\boldsymbol{\beta}$ has too large a dimension to be estimated using the data normally available, for modeling purposes $\boldsymbol{\beta}$ is replaced by a smoothed or lumped approximation $\boldsymbol{\gamma\theta}_*$ (where $\boldsymbol{\gamma}$ is a spatial and temporal interpolation matrix). Set $\boldsymbol{\gamma\theta}_*$ has the same form as the expected value of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}\bar{\boldsymbol{\theta}}$, where $\bar{\boldsymbol{\theta}}$ is the set of drift parameters of the stochastic process; $\boldsymbol{\theta}_*$ is a best-fit vector to $\boldsymbol{\beta}$. A model function $f(\boldsymbol{\beta})$, such as a computed hydraulic head or flux, is assumed to accurately represent an actual field quantity, but the same function written using $\boldsymbol{\gamma\theta}_*$, $f(\boldsymbol{\gamma\theta}_*)$, contains error from lumping or smoothing of $\boldsymbol{\beta}$ using $\boldsymbol{\gamma\theta}_*$. Thus, the replacement of $\boldsymbol{\beta}$ by $\boldsymbol{\gamma\theta}_*$ yields nonzero mean model errors of the form $E(f(\boldsymbol{\beta}) - f(\boldsymbol{\gamma\theta}_*))$ throughout the model and covariances between model errors at points throughout the model. These nonzero means and covariances are evaluated through third- and fifth-order accuracy, respectively, using Taylor series expansions. They can have a significant effect on construction and interpretation of a model that is calibrated by estimating $\boldsymbol{\theta}_*$.

Vector $\boldsymbol{\theta}_*$ is estimated as $\hat{\boldsymbol{\theta}}$ using weighted nonlinear least squares techniques to fit a set of model functions $\mathbf{f}(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})$ to a corresponding set of observations of $\mathbf{f}(\boldsymbol{\beta})$, $\mathbf{Y}$. These observations are assumed to be corrupted by zero-mean, normally distributed observation errors, although, as for $\boldsymbol{\beta}$, normality is not essential. An analytical approximation of the nonlinear least squares solution is obtained using Taylor series expansions and perturbation techniques that assume model

and observation errors to be small. This solution is used to evaluate biases and other results to second-order accuracy in the errors. The correct weight matrix to use in the analysis is shown to be the inverse of the second-moment matrix $E(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'$, but the weight matrix is assumed to be arbitrary in most developments. The best diagonal approximation is the inverse of the matrix of diagonal elements of $E(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'$, and a method of estimating this diagonal matrix when it is unknown is developed using a special objective function to compute $\hat{\theta}$.

When considered to be an estimate of $\mathbf{f}(\gamma\theta_*)$, the estimate $\mathbf{f}(\gamma\hat{\theta})$ is biased because of nonlinearity in $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta)$ ($\theta = \theta_*$ or $\hat{\theta}$), but when considered to be an estimate of $\mathbf{f}(\beta)$, $\mathbf{f}(\gamma\hat{\theta})$ is biased only because of components of nonlinearity in $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta)$ known as the intrinsic nonlinearity. (Intrinsic nonlinearity in either $\mathbf{f}(\beta)$ or $\mathbf{f}(\gamma\theta)$ is a component of the total nonlinearity in either function that cannot be eliminated by a unique transformation of either $\beta$ or $\theta$, respectively.) Because both types of intrinsic nonlinearity can be small, $\mathbf{f}(\gamma\hat{\theta})$ can be nearly unbiased as an estimate of $\mathbf{f}(\beta)$. Analogous results hold for a prediction $g(\gamma\hat{\theta})$ (where $g$ is some function of parameters of interest to the investigator), except that in this case the intrinsic nonlinearity is for the combination of either $g(\beta)$ and $\mathbf{f}(\beta)$ or $g(\gamma\theta)$ and $\mathbf{f}(\gamma\theta)$, termed the combined intrinsic nonlinearity. The biases are evaluated to second-order accuracy using Taylor series expansions and the analytical least squares solution, but an investigator would probably be more interested in estimates of $\mathbf{f}(\beta)$ and $g(\beta)$ than in estimates of fictitious variables $\mathbf{f}(\gamma\theta_*)$ and $g(\gamma\theta_*)$. Predictive accuracy of a model is thus strongly tied to the degree of intrinsic nonlinearity of the models $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta_*)$ together with the combined intrinsic nonlinearity of the models and the predictions to be made with them.

Uncertainties in the estimates of $\theta_*$, $\mathbf{f}(\gamma\theta_*)$, $\mathbf{f}(\beta)$, $g(\gamma\theta_*)$, and $g(\beta)$ (or some future measurement of $g(\beta)$) are addressed through nonlinear confidence regions, confidence intervals, and prediction intervals. If $\beta$ and the observation errors are normally distributed, statistical distributions of functions of the weighted sums of squared errors in the estimates necessary to define the regions and intervals approximate $F$ distributions that are modified with correction factors. These functions are very similar to the standard ones developed for linear models except that the parameter set $\theta_*$ is stochastic rather than fixed. The correction factors correct the distributions to account for intrinsic nonlinearity of $\mathbf{f}(\gamma\theta)$ and deviation of the weight matrix from the correct one. The correction factors are derived using the Taylor series and perturbation method used for the least squares; the generality of the factors and concepts leading to them are verified by an independent method that does not rely on Taylor series and perturbations. Because of the effects of spatial correlation, confidence regions and confidence intervals would generally be too small without using components of the correction factors needed to correct for using an incorrect weight matrix unless the correct one is used; prediction intervals may often be nearly correct. Approximate bounds for the correction factors are developed for use when the information on nonlinearity and heterogeneity necessary to calculate them is not available. Measures of total model nonlinearity of $\mathbf{f}(\gamma\theta)$, intrinsic nonlinearity of $\mathbf{f}(\gamma\theta)$, and combined

intrinsic nonlinearity of $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$ help an investigator decide when the components of the correction factors accounting for the types of intrinsic nonlinearity are not important.

Two examples are analyzed to test the validity and robustness of the theory when the model error is large. Example 1 is for one-dimensional, steady-state flow in an aquifer having log transmissivity (ln*T*) that varies stochastically at small scale and recharge (*W*) that is constant. Example 2 is for two-dimensional, steady-state flow in a zoned aquifer where ln*T* and *W* vary spatially at both large and small scales, the small-scale variations being stochastic. Hydraulic head data $\mathbf{Y}$ were generated as $\mathbf{f}(\gamma\theta_*)$ plus the sum of model errors $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ and small, zero-mean, independent, normal observation errors, $\mathbf{Y}-\mathbf{f}(\beta)$. The most important results are as follows: 1) The total nonlinearity in $\mathbf{f}(\gamma\theta)$ is large for both examples, but the intrinsic nonlinearity in $\mathbf{f}(\gamma\theta)$ and the combined intrinsic nonlinearity of $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$ for $g(\gamma\theta)$ equal to ln*T*, *W*, and a predicted hydraulic head, are all small for both examples. As the theory predicts, the corresponding biases also were found to be small. 2) The sum of the model and observation errors, $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$, has a nonzero mean and is not normally distributed for either example. 3) Spatial correlations between elements of $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ are often large, and large values are usually positive for both examples. 4) Residual set $\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})$ does not differ significantly from the zero-mean, normally distributed set predicted by the theory for either example. 5) For example 2, use of the correct, full weight matrix produced accurate confidence and prediction intervals, as determined using a Monte Carlo procedure. (Because of severe ill conditioning, example 1 produced open-ended intervals that could not be formally analyzed.) 6) Use of the best diagonal matrix for example 2 produced accurate confidence intervals only when the correction factors were used. Otherwise, the intervals are too small as determined by the Monte Carlo procedure. The prediction intervals did not have to be corrected. 7) Use of the estimated weight matrix for example 2 produced confidence and prediction intervals that had to be corrected, but the minimum containment probability of 0.92 for the intervals after correction remains slightly too small compared to the nominal probability of 0.95 because the correction factors are more approximate for this case than when the best diagonal weight matrix is used.

# 1. Introduction

## Background

Ground-water models simulate the processes involved in ground-water flow and transport. They are among the most powerful tools available for use in water-resources studies. They are used to 1) analyze the effects of possible hydrologic, geologic, and man-made processes and features on synthetic flow systems, 2) analyze past and present actual flow systems to better understand rates, directions, and causes of water movement and (or) transport within them, and 3) predict responses of actual flow systems to future changes or conditions such as ground-water or surface-water development, changes in recharge or discharge rates, contaminant spills, and so forth (Cooley, 1985; Anderson and Woesner, 1992, p. 4). Most modern ground-water models represent the processes involved mathematically, and so in this report the term ground-water model refers to this representation. Also, in this report concern is explicitly on flow, but as will become obvious, the theory developed here is general enough to be applied to transport as well.

Even when only flow is considered, construction of a ground-water model for a field system using the types of data commonly available is not necessarily straightforward. Data are virtually never complete and detailed enough to allow substitution into the model equations and direct computation of the results of interest. A model can, of course, be fitted to observed data by manual, trial and error calibration, but, though simple in concept, even this procedure is not necessarily straightforward. (See, for example, the advice given by Konikow (1978).) In addition, as argued by Carrera and Neuman (1986, p. 199-200), Cooley and Naff (1990, Chapter 1), and Hill (1992, p. 3-4), for example, the trial and error procedure is often highly subjective and fails to provide a basis for both critical, quantitative evaluation and uncertainty analysis. For these reasons, the attention of a number of investigators has turned to more formal optimization, statistical, and geostatistical procedures. (See reviews by Yeh (1986), Carrera (1988), and Ginn and Cushman (1990).)

Application of optimization, statistical, and geostatistical procedures to model calibration and uncertainty analysis is hampered by two pervasive problems: 1) nonlinearity of the solution of the model equations with respect to some of the model (or hydrogeologic) input variables (termed in this report system characteristics), and 2) detailed and generally unknown spatial variability (referred to as heterogeneity) of some of the system characteristics. These problems are highly interrelated, as will be shown.

***Nonlinearity***. The model equations for ground-water flow include the flow equation(s), boundary conditions, and initial conditions. The fundamental sources of nonlinearity in the solution of these equations for hydraulic head are well known. They result from multiplicative relations between system characteristics such as hydraulic conductivity and specific storage, and hydraulic head or its spatial and temporal derivatives in the model equations. (See, for example, Hill (1992, p. 69-70) for a discussion of nonlinearity from Darcy's law.) These sources of nonlinearity are present even when the differential equations composing the model equations are classified as being linear; the solution is still a nonlinear function of the multiplicative input variables. For example, consider the standard linear differential equation for transient, three-dimensional, ground-water flow in an isotropic, heterogeneous porous formation, which is

$$\frac{\partial}{\partial x}(K(\mathbf{x})\frac{\partial h}{\partial x})+\frac{\partial}{\partial y}(K(\mathbf{x})\frac{\partial h}{\partial y})+\frac{\partial}{\partial z}(K(\mathbf{x})\frac{\partial h}{\partial z})=S_s(\mathbf{x})\frac{\partial h}{\partial t}-w(\mathbf{x},t) \tag{1-1}$$

where $\mathbf{x}$ = the cartesian spatial coordinates $x, y$, and $z$; $t$ = time; $h = h(\mathbf{x},t)$ = hydraulic head as a function of $\mathbf{x}$ and $t$; $K(\mathbf{x})$ = hydraulic conductivity as a function of $\mathbf{x}$; $S_s(\mathbf{x})$ = specific storage as a function of $\mathbf{x}$; and $w(\mathbf{x},t)$ = source (negative for a sink) as a function of $\mathbf{x}$ and $t$. The solution of (1-1) (and attendant boundary and initial conditions) is $h(\mathbf{x},t)$. The solution is a nonlinear function of $K(\mathbf{x})$ and $S_s(\mathbf{x})$, but is a linear function of $w(\mathbf{x},t)$. Additional sources of nonlinearity result when the model equations are classified as being nonlinear, as for unconfined flow problems (which have an unknown free surface) or variably saturated flow problems (where hydraulic conductivity and storativity are functions of pressure) or when a nonlinear hydraulic-head or flux process is present. In this case the two sources of nonlinearity compound in their effects on the solution of the flow equations.

Nonlinearity has a significant influence on methods of calibration and uncertainty analysis. Virtually all optimization, statistical, and geostatistical methods known to me, except Monte Carlo methods, are designed primarily to treat linear problems. Nonlinear problems are either linearized or solved through some kind of iterative or sequential solution procedure wherein each step solves a linear problem. For example, the Gauss-Newton method for fitting a nonlinear function (such as the solution of the model equations) to data using least squares uses a linearization of the function obtained from a truncated Taylor series at each iteration (Seber and Wild, 1989, p. 25-26). Similarly, the modification proposed by Yeh and others (1996) to include nonlinearity in the classical cokriging method (a linear geostatistical method) for simultaneously estimating log-transmissivity and hydraulic head fields involves an iterative solution wherein a linear cokriging problem is solved at each iteration. Finally, the statistical distributions used for classical uncertainty analyses (or inference) are generally based on linear functions. (See Seber and Wild (1989, Chapter 5).) All of these methods for incorporating nonlinearity are more complicated, and can be less numerically stable, than their linear counterparts. Moreover, they often only approximately account for nonlinearity (for example, the Yeh and others (1996) method).

One might conclude from the discussion in the previous paragraph that problems too nonlinear for linearization to provide good approximate results should be analyzed using other methods, such as Monte Carlo methods. Monte Carlo methods appear to be straightforward because they involve repeated sampling of a statistical distribution of system characteristics and use of the sets of variables in the model equations to generate a statistical distribution of a model function (such as the solution for hydraulic head) from which desired quantities such as means, variances, and percentiles can be computed. However, Monte Carlo methods are no panacea. In particular, when the underlying statistical distribution is unknown or involves unknown distributional parameters, as it does for application to ground-water models, then the Monte Carlo analysis method reduces to a bootstrap method (Efron, 1982), which may not be accurate unless the dimension of the observed-data set is large and (or) the degree of nonlinearity of the model function is small (Cooley, 1997, p. 871-872). Again, nonlinearity seems to be a significant factor. In addition, computational requirements may be large (Peck and others, 1988, p. 129-130; Cooley, 1997), and, if realizations of the system characteristics of interest are derived using least squares, then data censoring in generating sample data for ill-conditioned or highly nonlinear problems can become a problem because of nonconvergence of the least squares method (Cooley, 1997).

***Heterogeneity***. Characterization of both large- and small-scale spatial variability in hydraulic conductivity (or transmissivity) fields and incorporation of this spatial variability into ground-water models have been the subjects of extensive analysis spanning many years. (See, for example, discussions and reviews in Freeze (1975), Dagan (1986), Gelhar (1986), Peck and others (1988), and McLaughlin and Townley (1996).) However, other spatially defined system characteristics such as recharge (or discharge), leakance (for two-dimensional models), specific storage, and boundary conditions also vary spatially, at both large and small scales. Characterization of spatial variability of these variables and incorporation of the spatial variation into ground-water models have been the subjects of much less research. Freeze (1975) considered spatial variability in the storage coefficient in addition to spatial variability of hydraulic conductivity; Gomez-Hernandez and Gorelick (1989) examined spatial variability in hydraulic conductivity, leakance, and recharge; and Graham and Tankersley (1994a, 1994b) derived and used a geostatistical parameter estimation model for spatially variable transmissivity and recharge. In addition, Neuman and Orr (1993) and Tartakovsky and Neuman (1998) considered recharge to be stochastic and spatially variable in their derivations of conditional moment, flow equations. All of the cited studies considered hydraulic conductivity (or transmissivity) to be statistically independent of all other variables. (Gomez-Hernandez and Gorelick (1989) varied hydraulic conductivity, leakance, and recharge separately.) However, because of the strong relation of recharge, storativity, boundary conditions, and perhaps even leakance to the same rock properties governing hydraulic conductivites of rocks involved in at least shallow ground-water flow, these variables probably would not often be independent. Realistic geostatistical characterization of all of the system characteristics for a ground-water model does not seem to have been published.

Consideration of scales of variation is of major importance in characterizing heterogeneity and incorporating heterogeneity into a model (Di Federico and others, 1999). It is convenient to separate the scales into two classes: those that are too small to be explicitly identified and represented in a ground-water model and those that are large enough to be explicitly identified and represented. For example, a large, mappable, sudden facies change in a rock unit could be represented as a hydraulic-conductivity zone boundary in a model; whereas, a gradual facies change could be represented as a possible trend in hydraulic conductivity. Smaller scale variations might not be mappable or otherwise explicitly identifiable. From a purely operational standpoint, the smallest scale at which variations can be represented explicitly in a standard numerical model is at the grid-block scale. Smaller-scale variations must be lumped or smoothed. Even so, if these grid-block and sub grid-block scale variations are not explicitly identifiable, means must be devised for incorporating their influence into a model. Methods using geostatistical representation such as those of RamaRao and others (1995), Kitanidis (1995), and Yeh and others (1996) explicitly use the small scales of variation by relating small-scale variations in a system characteristic (such as log transmissivity) to small-scale variations in the model solution (such as hydraulic head). Methods that use lumping or smoothing of the system characteristics, such as trend fitting (for example, Yoon and Yeh, 1976, and Hill and others, 1998), zonation (for example, Cooley and others, 1986, and D'Agnese and others, 1999) and others such as classification of sediment types (Kuiper, 1994) do not represent small-scale variations explicitly. (See also reviews by Yeh (1986), Carrera (1988), and McLaughlin and Townley (1996).)

Methods of lumping or smoothing the system characteristics are convenient for handling heterogeneity because these methods result in a small number of variables to be estimated. However, questions arise as to what a solution (for example, hydraulic head) to a model equation using the smoothed variables actually represents. I suspect that most individuals constructing models of field areas would like the solutions and computed fluxes obtained using smoothed or lumped system characteristics to approximate spatial and temporal running averages of the solutions and computed fluxes that would be obtained using variables varying at small scale. These ideas also were discussed by McLaughlin and Townley (1996, p. 1134-1135) and immediately bring to mind the concept of "effective values," which can be loosely defined as lumped or smoothed system characteristics that yield the average quantities just mentioned. Note that this definition is determinstic in the sense that local values of the variables are considered to be fixed, not realizations from a stochastic process. A similar, but more precise, definition using spatial moments was used by Kitanidis (1990) to find the effective hydraulic conductivity for gradually varying flow in a periodic medium.

Another definition of effective values is based on concepts of stochastic flow theory. For this definition the running averages mentioned in the previous paragraph are replaced by ensemble averages. Neuman and Orr (1993, p. 144) make a clear distinction between the stochastic and other definitions of effective hydraulic conductivity, give a precise and useful stochastic definition, and give a number of references for studies involving alternative definitions. An important result of their study, which was for steady-state flow, is that in general an effective hydraulic conducitivity field as defined by them does not exist. A follow-up study by Tartakovsky and Neuman (1998) for transient flow came to a similar conclusion. Neuman and Orr (1993) and Tartakovsky and Neuman (1998) argued that effective hydraulic conductivity formulations obtained in previous studies only apply to special conditions. It seems quite probable that these results would generalize to apply for alternative definitions of effective hydraulic conductivity.

The difficulty in defining an effective hydraulic conductivity value is the result of nonlinearity of the model solution with respect to hydraulic conductivity. This can be deduced from the facts that 1) the term causing the difficulty (labeled $r_k(\mathbf{x})$ by Neuman and Orr, 1993, p. 342) involves the product of hydraulic conductivity fluctuations and hydraulic gradient fluctuations, which is the source of nonlinearity mentioned previously, and 2) effective values of system characteristics that appear linearly in the model solution are readily defined. For example, in forms of the stochastic flow equations derived by Neuman and Orr (1993) and Tartakovsky and Neuman (1998) in which the random variability of hydraulic conductivity is set to zero, the ensemble mean source term and boundary conditions are effective values because they yield the ensemble mean hydraulic head and computed flux fields.

Because the model solution is nonlinear when considered to be a function of all system characteristics, it appears that there is little hope of defining effective values of these variables for general ground-water models that are based on standard flow equations such as (1-1). One might conclude from this that lumped or smoothed system characteristics should not be used with these models. Such a conclusion seems implicit in the conclusions of Neuman and Orr (1993, p. 355-356) and Tartakovsky and Neuman (1998, p. 6). However, reasonable estimates for lumped or smoothed system characteristics have been obtained using nonlinear regression to calibrate models resulting from field studies (for example, Cooley, 1979; Cooley and others, 1986; Yager, 1996; Christensen and Cooley, 1999a, 1999b; and D'Agnese and others, 1999), and any biases in these estimates did not seem to be significant. (Note: other studies obtaining similarly reasonable results undoubtedly exist.) These results indicate the possible existence of a theoretical basis for using estimates of lumped or smoothed system characteristics in ground-water models.

A theory involving estimation of lumped or smoothed system characteristics for ground-water models must address the question of exactly what is being estimated. Yeh and others (1996, p. 85) question the identity of the hydraulic head and lumped or smoothed transmissivity fields estimated by minimizing an objective function of the differences between observed and computed data. They also make the point that, because these fields are "often undefined", "the uncertainty associated with the output can not be addressed." Although the work of Christensen and Cooley (1999b) casts considerable doubt on the latter assertion, the question concerning identity is valid. A nonlinear regression model is generally stated in terms of a true (or correct) parameter set, which for a ground-water model would ideally represent a set of effective values. Moreover, uncertainty in the estimates and predictions to be made with the model is with reference to the true set and quantities computed using it. However, if clearly definable effective values do not exist, then what is the true parameter set?

One possible approach to answering the question cited above is to consider the model to be empirical. A model written in terms of parameters that do not have concrete physical definitions fits the definition of an empirical model given by Jones (1983, p. 68-69). He defined the parameter set being estimated by assuming that predictions to be made with the model will be of the same types of data as included in the calibration-data set (Jones, 1983, p. 69). Then the true parameter set (termed the "best parameter" set by Jones, 1983, p. 69) is the set that minimizes the limit of the objective function as the data set gets infinitely large. Presumably, the infinitely large data set would encompass all potential predictions to be made with the model. A major problem with this conceptualization as applied to a ground-water model is that the model often is used to make predictions of types of quantities not potentially includable in the original data set (for example, a computed flux distribution or the results of future development), which again leaves the "best parameter" set undefined.

The above background discussion indicates the need for a new theory for modeling ground-water flow using lumped and (or) smoothed system characteristics. This theory must recognize and effectively deal with model nonlinearity that in general prevents lumping or smoothing of the heterogeneous system characteristics to obtain effective values and that causes the other difficulties mentioned. To address this need a viewpoint that differs from the viewpoint used in previous studies is adopted in this report. The model parameters are defined physically as the lumped or smoothed system characteristics, and estimates and uncertainty measures are derived based on the physical definitions. Properties of the estimates and uncertainty measures (for example, biases) are investigated using extensions of methods used to investigate classical nonlinear regression models to determine when the lumped or smoothed properties would produce accurate approximations.

## Purpose and Scope

The purpose of this report is to describe a new theory for modeling ground-water flow in heterogeneous media. The report

1. provides a sound theoretical framework and sound theoretical guidance for modeling ground-water motion in heterogeneous media using lumped and (or) smoothed system characteristics (termed model parameters, or simply parameters), and
2. provides a sound theoretical framework for estimating the parameters and assessing the uncertainty of the estimates, model functions computed using the estimates, and predictions to be made with the model. (The distinction between the last two quantities will become apparent.)

The theory developed in this report seeks to explain some results observed from the field studies cited earlier (Cooley, 1979; Cooley and others, 1986; Christensen and Cooley, 1999a, 1999b; Yager, 1996). That is:

1. Estimates of parameters, model functions, and predicted quantities are often physically realistic, or close to what was expected, even though effective values of the system characteristics probably do not exist.
2. Differences between the observed data and data computed using the parameter estimates are normally distributed. That is, these differences behave as if the model were linear and as if the differences between the observed data and data computed using the true parameters were normally distributed.

3. Measures of uncertainty (confidence intervals) for some parameters and computed model functions appear to exclude reasonable values (or to be too small); whereas, the measures for others do not.

These points are revisited in section 8 of this report after the theory has been completely developed.

The approach taken is derived from work by Beale (1960), Johansen (1983), and Hamilton and Wiens (1987). It uses a combination of Taylor series expansions and perturbation theory to derive approximations of parameter estimates and statistical distributions necessary to characterize model behavior and uncertainty when the model solution is a nonlinear function of the parameters and when error resulting from lumping and smoothing heterogeneous fields of system characteristics is significant. An approach that does not use the Taylor series and perturbation approximations is used to verify the principal results of the approximate analyses and extend the analyses to system characteristics having larger variances than those assumed for the approximate analyses. The viewpoint is classical, based on sampling of data from specified statistical distributions. A Bayesian viewpoint also could have been used, which would have lead to developments that are parallel to those given here. The classical viewpoint is used because the developments appeared to be more straightforward using it as compared to the Bayesian viewpoint, and because I suspect that most practicing hydrologists are more familiar with classical statistics than Bayesian statistics.

This report is designed to be read at three possible levels of understanding. If only the introduction, the summaries at the end of each section, and the final summary and conclusions section are read, the reader can obtain a quick overview of the principal results without having to delve into the mathematics. In addition to a good knowledge of ground-water hydrology, only a basic understanding of stochastic flow theory and statistics are required to understand these sections. If all of the main text, exclusive of the appendices, is read, then the reader can understand the main results without having to follow the derivations. A good understanding of engineering mathematics, stochastic flow theory, and statistics is necessary to derive full benefit of this reading. Finally, the appendices contain detailed derivations of all the results. Some of these require some specialized knowledge of theoretical statistics. The derivations are presented in detail to guide the interested reader unambiguously.

## Acknowledgments

Aldo “Skip” Vecchia, U.S. Geological Survey, provided valuable collaboration and put the author on the right track in developing theory and methodology for computing nonlinear confidence and prediction intervals. Brent Troutman, U.S. Geological Survey, tirelessly discussed numerous points involving the theory developed in the present report with the author. Steen Christensen, University of Aarhus, Aarhus, Denmark, provided field data and collaborated in testing the theory from its inception. Without the support of these individuals, the present work could not have been completed.

# 2. Notation for Matrix Derivatives

In this report operator notation (Johansen, 1983, p. 174-175) is used to denote matrices of partial derivatives. With this notation first and second partial derivatives of some vector **f** of order $n$ with respect to another vector $\boldsymbol{\theta}$ of order $p$ are given by the matrices

$$\mathbf{Df} = \left[\frac{\partial f_i}{\partial \theta_j}\right];\ i = 1,2,\ldots,n;\ j = 1,2,\ldots,p \tag{2-1}$$

and

$$\mathbf{D}^2\mathbf{f} = \left[\frac{\partial^2 f_i}{\partial \theta_j \partial \theta_k}\right];\ i = 1,2,\ldots,n;\ j = 1,2,\ldots,p;\ k = 1,2,\ldots,p \tag{2-2}$$

Note that (2-1) is a standard matrix of order $n \times p$, whereas (2-2) is a three-dimensional matrix of order $n \times p \times p$. Algebraic operations involving three-dimensional matrices can be confusing, so all algebraic operations involving a three-dimensional matrix such as $\mathbf{D}^2\mathbf{f}$ will be explicit on $i$ and will use the two-dimensional slice $\mathbf{D}^2 f_i$, which is a standard matrix of order $p \times p$. For example, a product of some vector $\mathbf{Z}$ of order $n$ with $\mathbf{D}^2\mathbf{f}$ would be given as $\sum_i Z_i \mathbf{D}^2 f_i$ or $\sum_i \mathbf{D}^2 f_i Z_i$ instead of $\mathbf{Z}'\mathbf{D}^2\mathbf{f}$ or $\mathbf{D}^2\mathbf{fZ}$, where the prime indicates transpose. Occasionally, the row vector $\mathbf{D}f_i$ or column vector $\mathbf{D}f_i'$ of order $p$ will need to be used in calculations involving $\mathbf{Df}$ and $\mathbf{D}^2\mathbf{f}$.

Partial derivatives with respect to vectors other than $\boldsymbol{\theta}$ are denoted by an appropriately subscripted operator. Thus, the partial derivatives of **f** with respect to a vector $\boldsymbol{\beta}$ of order $m$ are given by

$$\mathbf{D}_\beta\mathbf{f} = \left[\frac{\partial f_i}{\partial \beta_j}\right];\ i = 1,2,\ldots,n;\ j = 1,2,\ldots,m \tag{2-3}$$

and

$$\mathbf{D}^2_\beta\mathbf{f} = \left[\frac{\partial^2 f_i}{\partial \beta_j \partial \beta_k}\right];\ i = 1,2,\ldots,n;\ j = 1,2,\ldots,m;\ k = 1,2,\ldots,m \tag{2-4}$$

Equations (2-3) and (2-4) are matrices of order $n \times m$ and $n \times m \times m$, respectively; operations on the three-dimensional matrix are explicit on $i$ as before.

Other alternatives exist for algebraic operations using three-dimensional matrices such as (2-2) and (2-4). One example is the notation used by Johansen (1983, p. 181-184), and another is the use of Vetter calculus (for example, Dettinger and Wilson, 1981). However, the explicit notation adopted here has the advantage of being both simple and straightforward.

# 3. Basic Theory

## System Properties and Model

Initially, assume the ground-water flow system to be in steady state. The unsteady case is considered later in this section. Further, assume the flow system to be adequately characterized by three categories of hydrogeologic variables or system characteristics: 1) variables such as hydraulic conductivity (or a transformation of the variables such as log hydraulic conductivity), recharge from precipitation, and discharge from evapotranspiration that can vary spatially throughout the model; 2) variables such as hydraulic heads and fluxes that can vary spatially along internal and external boundaries of the model; and 3) variables such as spring and well discharges that occur locally, at points. The variables in the first two categories can be conceptualized as being continuously variable spatially as was done by Neuman and Orr (1993) and McLaughlin and Townley (1996) or as being discretely variable spatially as was done by Kitanidis (1995) and McLaughlin and Townley (1996). Discrete variation is often associated with the discretization of a model region into a grid for numerical simulation (for example, RamaRao and others, 1995; Kitanidis, 1995). However, discrete variation can be at as small a scale as desired, thus potentially making it virtually the same as continuous variation. Because all scales can be included in discrete variation and because it is straightforward to work with, the discrete viewpoint is adopted in this report.

All of the system characteristics can be assembled into a vector $\beta$ of order $m$. Each element in this vector is the value of a system characteristic in a particular volume element for category 1, a boundary segment for category 2, or a point for category 3. (An example is given later in this section.) Because $\beta$ includes all scales of variation necessary to produce an accurate model, any model function of $\beta$, $f(\beta)$, is almost free of model error, assuming, of course, that the model accurately represents the physical processes. The model function could be a computed hydraulic head at some point, a computed flux at some point, or any other physically relevant function of $\beta$.

The system characteristics in categories 1 and 2 contain scales of variability that are explicitly contained in a ground-water flow model and smaller scales that are not. For example, a model may be zoned for a particular characteristic so that none of the spatial variability within each zone is explicitly contained in the model; the zone simply represents an average for the characteristic. Models are constructed in this way because the order, $m$, of $\beta$ is generally so large that it is impossible to measure or otherwise estimate all of the elements in it. To include the influence of this unknown variability (at a smaller scale than represented in the model), it is common to imagine a stochastic process for $\beta$ and use the stochastic properties of $\beta$ in modeling; a vast literature has emerged based on this concept. (See reviews by Gelhar, 1986, and McLaughlin and Townley, 1996.)

Following Kitanidis (1995), assume the expected value of the stochastic vector $\beta$ to have the form $\gamma\bar{\theta}$, where $\gamma$ is an $m \times p$ interpolation or spatial averaging matrix to be examined later and $\bar{\theta}$ is a vector of drift parameters of order $p$. Also, assume vector $\beta$ to be normally distributed,

$$\beta \sim N\left(\gamma\bar{\theta},\ \mathbf{V}_\beta \sigma_\beta^2\right) \tag{3–1}$$

where $\mathbf{V}_\beta \sigma_\beta^2$ is an $m \times m$ covariance matrix that gives the spatial covariances among all of the elements of $\beta$; the covariance $V_{\beta ii}\sigma_\beta^2$ is simply the variance of $\beta_i$. The normality assumption is commonly made (for example Kitanidis, 1995; McLaughlin and Townley, 1996) and, for the elements of $\beta$ representing hydraulic conductivity, is known to be a good approximation if $\beta$ is written in terms of log hydraulic conductivity. This assumption is not essential to developing the present theory, but some aspects of the theory would be difficult to express analytically if the assumption were abandoned. As will be seen, the assumption can be indirectly tested.

A simple example for $\gamma$ is obtained if the hydrogeology of the region is such that the drift for categories 1 and 2 can be approximated by zones of constant value. In this case $\gamma$ assumes the form

$$\gamma = \begin{bmatrix} \mathbf{1}_1 & \mathbf{0}_1 & \mathbf{0}_1 & \cdots & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{1}_2 & \mathbf{0}_2 & \cdots & \mathbf{0}_2 \\ & & \cdots & & \\ \mathbf{0}_p & \mathbf{0}_p & \mathbf{0}_p & \cdots & \mathbf{1}_p \end{bmatrix} \tag{3-2}$$

where vectors $\mathbf{1}_i$ and $\mathbf{0}_i$ of order $m_i$ are given by

$$\mathbf{1}_i = \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix} \qquad \mathbf{0}_i = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} \tag{3-3}$$

and $\sum_{i=1}^{p} m_i = m$. Thus parameter $i$ lies in a zone having $m_i$ discrete values of $\beta_j$ in it. By performing the product $\gamma\,\bar{\theta}$ it can be seen that the mean value of $\beta_j$ at each point in a zone is given by $\bar{\theta}_i$. Note that if $m_i = 1$, a category 3 variable is obtained. Other forms for $\gamma$ result from using interpolation (Yeh, 1986, p. 98-99). An example of the use of finite element interpolation is given in Hill and others (1998).

An example involving zonation is shown in figure 1-1 and table 1-1 for two zones of transmissivity $T$, two zones of recharge rate $W$, two boundary segments for specified head $H_B$, and one pumping well $Q$. In this case, if $\ln T$, $W$, $H_B$, and $Q$ are normally distributed,

$$\bar{\theta} = \begin{bmatrix} \ln \bar{T}_1 \\ \ln \bar{T}_2 \\ \bar{W}_1 \\ \bar{W}_2 \\ \bar{H}_{B1} \\ \bar{H}_{B2} \\ \bar{Q} \end{bmatrix} \tag{3-4}$$

where the overbars signify drift values and the form for $\gamma\bar{\theta}$ is given in table 1-1. Note that, although all boundaries in the example are orthogonal and rectilinear for simplicity, they could be nonorthogonal and curvilinear, as could occur for a real system. This would not change the form of $\beta$, $\gamma$, or $\bar{\theta}$ for the example.

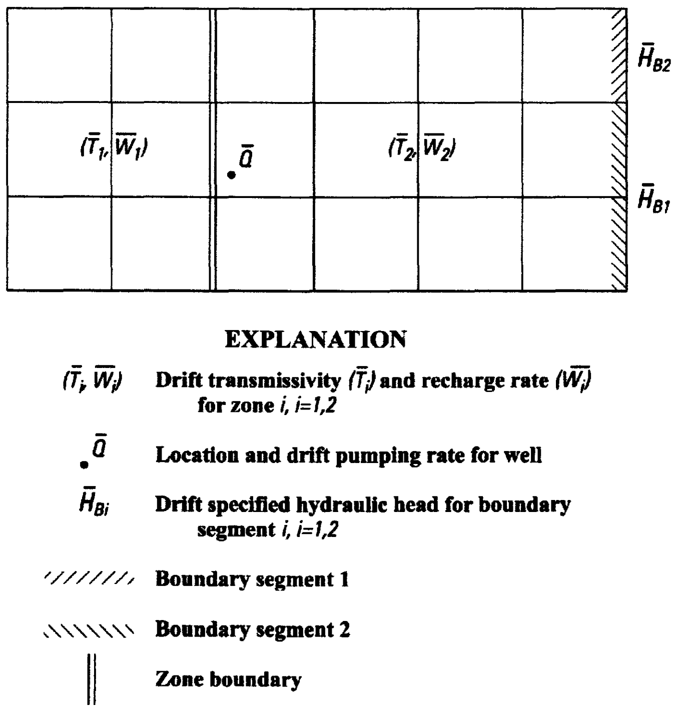


**Figure 1-1.** Example involving zonation in conjunction with small-scale variability. Square elements designate discrete elements of internal small-scale variability contained in $\beta$.

**Table 1-1.** Zonal information for the example.

[Vectors $\mathbf{1}_1, \mathbf{1}_2, \ldots, \mathbf{1}_7$ are defined in (3-2) and (3-3); $\overline{T}_i$ is geometric mean (drift) transmissivity in zone *i*; $\overline{W}_i$ is mean (drift) recharge rate in zone *i*; $\overline{H}_{Bi}$ is mean (drift) specified head in boundary segment *i*; $\overline{Q}$ is the mean (drift) pumping rate from the well; $\overline{\theta}_i$ is an element of $\overline{\boldsymbol{\theta}}$ defined in (3-4); and $m_i$ is the number of values of $\beta_j$ defined for parameter *i*.]

| Zone Definition Vector | $\overline{\theta}_i$ | $m_i$ |
|---|---|---|
| $\mathbf{1}_1$ | $\ln\overline{T}_1$ | 6 |
| $\mathbf{1}_2$ | $\ln\overline{T}_2$ | 12 |
| $\mathbf{1}_3$ | $\overline{W}_1$ | 6 |
| $\mathbf{1}_4$ | $\overline{W}_2$ | 12 |
| $\mathbf{1}_5$ | $\overline{H}_{B1}$ | 2 |
| $\mathbf{1}_6$ | $\overline{H}_{B2}$ | 1 |
| $\mathbf{1}_7$ | $\overline{Q}$ | 1 |
| | $m$ = | 40 |

Assume that a set of *n* observations corresponding to *n* values of the model function $f(\boldsymbol{\beta})$ can be expressed in the form

$$\mathbf{Y} = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon} \tag{3-5}$$

where $\mathbf{Y}$ is the vector of observations of order *n* corresponding to the vector of model function values $\mathbf{f}(\boldsymbol{\beta})$, and $\boldsymbol{\varepsilon}$ is a vector of observation errors of order *n*. Also, assume the errors to have zero mean and to have the normal distribution

$$\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \mathbf{V}_\varepsilon \sigma_\varepsilon^2\right) \tag{3-6}$$

where $\mathbf{V}_\varepsilon \sigma_\varepsilon^2$ is the observation-error covariance matrix. Matrix $\mathbf{V}_\varepsilon$ may often be block diagonal corresponding to different types of model function and corresponding data in $\mathbf{Y}$. As for $\boldsymbol{\beta}$, normality is not essential for the theory developed here, but some aspects of the theory are difficult to state analytically if normality is not assumed. The normality assumption can be indirectly tested. Finally, assume $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ to be statistically independent.

## A Spatial Average for the Vector of System Characteristics, $\boldsymbol{\beta}$

To construct a ground-water model, estimates of the system characteristics are needed. However, the dimension of $\boldsymbol{\beta}$ is too large to permit a unique estimate of $\boldsymbol{\beta}$ to be obtained. Furthermore, as discussed in section 1, reduction of the dimension of $\boldsymbol{\beta}$ by substituting a vector of "effective values" of much smaller dimension may not be possible because effective values may often not exist. That is, if a vector of smaller dimension is used, it may not be able to produce a model that reproduces both average fluxes and the average hydraulic head distribution. As stated in section 1, the solution to this problem followed in this report is to estimate a vector of reduced dimension that has a unique physical definition, but is not necessarily a vector of effective values, then investigate the properties of this vector, its estimate, and the resulting model and predictions to be made with it. If the properties are found to be favorable, then the model can be accepted and used.

A possible candidate for the reduced vector might seem to be the vector of drift parameters, $\bar{\theta}$. However, even though this vector has a unique physical definition, it is in fact fictitious because the stochastic process is fictitious. Often, estimates of parameters of the real physical system, which is imagined to be a realization of the stochastic process, are desired. Thus, a vector of drift parameters is not an ideal candidate for the reduced vector, and one pertaining to the real physical system should be selected. A vector containing the reduced vector is defined in this report as a spatial average of $\beta$ that has the same form as the drift. It is derived as the best-fit vector $\gamma\theta_*$ obtained by minimizing $(\beta-\gamma\theta)'(\beta-\gamma\theta)$ with respect to general parameter set $\theta$ to obtain

$$\theta_* = (\gamma'\gamma)^{-1}\gamma'\beta \tag{3-7}$$

Because $\beta$ has an expected value of $\gamma\bar{\theta}$, $\theta_*$ has an expected value of $\bar{\theta}$.

If zonation is used for the drift, the indicated products in (3-7) may be performed using (3-2) to give

$$(\gamma'\gamma)^{-1} = \begin{bmatrix} m_1^{-1} & 0 & \cdots & 0 \\ 0 & m_2^{-1} & \cdots & 0 \\ & \cdots & & \\ 0 & 0 & \cdots & m_p^{-1} \end{bmatrix} \tag{3-8}$$

$$(\gamma'\gamma)^{-1}\gamma' = \begin{bmatrix} m_1^{-1}\mathbf{1}_1' & \mathbf{0}_2' & \cdots & \mathbf{0}_p' \\ \mathbf{0}_1' & m_2^{-1}\mathbf{1}_2' & \cdots & \mathbf{0}_p' \\ & \cdots & & \\ \mathbf{0}_1' & \mathbf{0}_2' & \cdots & m_p^{-1}\mathbf{1}_p' \end{bmatrix} \tag{3-9}$$

and

$$(\gamma'\gamma)^{-1}\gamma'\beta = \begin{bmatrix} m_1^{-1}(\mathbf{1}_1',\mathbf{0}_2',\cdots,\mathbf{0}_p')\beta \\ m_2^{-1}(\mathbf{0}_1',\mathbf{1}_2',\cdots,\mathbf{0}_p')\beta \\ \cdots \\ m_p^{-1}(\mathbf{0}_1',\mathbf{0}_2',\cdots,\mathbf{1}_p')\beta \end{bmatrix} \tag{3-10}$$

Thus, $(\gamma'\gamma)^{-1}\gamma'\beta$ yields the average of values in $\beta$ for each parameter in each zone. That is, for parameter *i*

$$m_i^{-1}(\mathbf{0}'_1,\mathbf{0}'_2,\cdots,\mathbf{1}'_i,\cdots,\mathbf{0}'_p)\beta = \frac{1}{m_i}\sum_{j(i)}\beta_j \tag{3-11}$$

where $j(i)$ indicates summation over all values of $j$ for parameter $i$.

## Properties of the Vector of Model Function Differences, $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$

Corresponding to model function vector $\mathbf{f}(\beta)$ is the model function vector $\mathbf{f}(\gamma\theta_*)$, representing values of the same model function written using the spatial average instead of $\beta$. Systematic discrepancies between $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta_*)$ are indicated by the expected value and variance of the difference $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$. An expression for this difference is given by (3-18) using the following development. First, the errors $\mathbf{e}$ and $\mathbf{e}_*$ are defined as

$$\mathbf{e} = \beta - \gamma\bar{\theta} \tag{3-12}$$

$$\mathbf{e}_* = \beta - \gamma\theta_* \tag{3-13}$$

These two errors are related using (3-7) as follows.

$$\begin{aligned}
\mathbf{e}_* &= \beta - \gamma(\gamma'\gamma)^{-1}\gamma'\beta \\
&= (\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')(\mathbf{e}+\gamma\bar{\theta}) \\
&= (\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}
\end{aligned} \tag{3-14}$$

Second, expansion of $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta_*)$ to second order around $\mathbf{f}(\gamma\bar{\theta})$ using truncated Taylor series yields

$$f_i(\beta) = f_i(\gamma\bar{\theta}) + \mathbf{D}_\beta f_i(\beta-\gamma\bar{\theta}) + \frac{1}{2}(\beta-\gamma\bar{\theta})'\mathbf{D}_\beta^2 f_i\,(\beta-\gamma\bar{\theta});\ i=1,2,\ldots,n \tag{3-15}$$

$$f_i(\gamma\theta_*) = f_i(\gamma\bar{\theta}) + \mathbf{D}_\beta f_i\gamma(\theta_*-\bar{\theta}) + \frac{1}{2}(\theta_*-\bar{\theta})'\gamma'\mathbf{D}_\beta^2 f_i\,\gamma(\theta_*-\bar{\theta});\ i=1,2,\ldots,n \tag{3-16}$$

where $\mathbf{D}_\beta f_i$ and $\mathbf{D}_\beta^2 f_i$ are row-vector and matrix components of $\mathbf{D}_\beta\mathbf{f}$ and $\mathbf{D}_\beta^2\mathbf{f}$ as defined by (2-3) and (2-4). They are evaluated at $\beta=\gamma\bar{\theta}$. Third, $\theta_*-\bar{\theta}$ is expressed in terms of $\mathbf{e}$ as

$$\begin{aligned}
\theta_*-\bar{\theta} &= (\gamma'\gamma)^{-1}\gamma'\beta-\bar{\theta} \\
&= (\gamma'\gamma)^{-1}\gamma'(\mathbf{e}+\gamma\bar{\theta})-\bar{\theta} \\
&= (\gamma'\gamma)^{-1}\gamma'\mathbf{e}
\end{aligned} \tag{3-17}$$

Fourth, expansion of $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ results in

$$f_i(\beta)-f_i(\gamma\theta_*)= f_i(\beta)-f_i(\gamma\bar{\theta})+f_i(\gamma\bar{\theta})-f_i(\gamma\theta_*)$$
$$= \mathbf{D}_\beta f_i \mathbf{e}_* + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 f_i\, \mathbf{e} - \frac{1}{2}(\theta_* - \bar{\theta})'\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\theta_* - \bar{\theta})$$
$$= \mathbf{D}_\beta f_i(\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_i \; - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e};\, i=1,2,\ldots,n \qquad (3\text{-}18)$$

Equation (3-18) is used in theorem 4.6.1 in Graybill (1976, p. 139-140) to evaluate the expected value and variance of $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$. The final result is obtained by using the facts that $E(\mathbf{e})=\mathbf{0}$ and $Var(\mathbf{e})=\mathbf{V}_\beta\sigma_\beta^2$, where $E(\cdots)$ and $Var(\cdots)$ stand for the ensemble expected value (mean) and variance, respectively, and is

$$E(f_i(\beta)-f_i(\gamma\theta_*)) = \frac{1}{2}tr\;((\mathbf{D}_\beta^2 f_i \; - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\sigma_\beta{}^2;\, i=1,2,\ldots,n \qquad (3\text{-}19)$$

where $tr(\cdots)$ stands for matrix trace. Because $\beta$ has a symmetric distribution, (3-19) is third-order accurate. Note that if the model is linear so that $\mathbf{D}_\beta^2 f_i$ is zero, the expected value is zero.

The variance of $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ is the matrix

$$Var(\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)) = [E(f_i(\beta)-f_i(\gamma\theta_*))(f_j(\beta)-f_j(\gamma\theta_*))$$
$$-E(f_i(\beta)-f_i(\gamma\theta_*))E(f_j(\beta)-f_j(\gamma\theta_*))] \qquad (3\text{-}20)$$

Evaluation of the first expected value using (3-18) and the fact that any triple product of a zero-mean symmetrically distributed variable is zero yields

$$E(f_i(\beta)-f_i(\gamma\theta_*))(f_j(\beta)-f_j(\gamma\theta_*))$$
$$= E(\mathbf{D}_\beta f_i(\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_i \; - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})(\mathbf{D}_\beta f_j(\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}$$
$$+\frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_j \; - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})$$
$$= E(\mathbf{D}_\beta f_i(\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}\mathbf{e}'(\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{D}_\beta f_j')$$
$$+\frac{1}{4}E(\mathbf{e}'(\mathbf{D}_\beta^2 f_i \; - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})(\mathbf{e}'(\mathbf{D}_\beta^2 f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}) \qquad (3\text{-}21)$$

Because $\mathbf{e}$ is symmetrically distributed, (3-21) is fifth-order accurate. Evaluation of the second expected value uses the result from appendix A that for symmetric matrices $\mathbf{A}_i$ and $\mathbf{A}_j$

$$E(\mathbf{x}'\mathbf{A}_i\mathbf{x})(\mathbf{x}'\mathbf{A}_j\mathbf{x}) = tr(\mathbf{A}_i)tr(\mathbf{A}_j)\sigma^4 + 2tr(\mathbf{A}_i\mathbf{A}_j)\sigma^4 \qquad (3\text{-}22)$$

where $\mathbf{x} \sim N(\mathbf{0},\mathbf{I}\sigma^2)$. Let $\mathbf{x}=\mathbf{V}_\beta^{-1/2}\mathbf{e}$, $\sigma^2=\sigma_\beta^2$, and

$\mathbf{A}_i = \mathbf{V}_\beta^{1/2}(\mathbf{D}_\beta^2 f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta^{1/2}$. Then, from the definition of **e**, $E(\mathbf{ee}') = \mathbf{V}_\beta \sigma_\beta^2$, so that

$$
\begin{aligned}
&E(f_i(\beta) - f_i(\gamma\theta_*))(f_j(\beta) - f_j(\gamma\theta_*)) \\
&= \mathbf{D}_\beta f_i(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{D}_\beta f_j'\sigma_\beta^2 \\
&+ \frac{1}{4} tr((\mathbf{D}_\beta^2 f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta) tr((\mathbf{D}_\beta^2 f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\sigma_\beta^4 \\
&+ \frac{1}{2} tr((\mathbf{D}_\beta^2 f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{D}_\beta^2 f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\sigma_\beta^4 \qquad (3\text{-}23)
\end{aligned}
$$

where **I** is the identity matrix. Finally, substitution of (3-19) and (3-23) yields the variance as

$$
\begin{aligned}
&Var(\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*)) = \mathbf{D}_\beta \mathbf{f}(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{D}_\beta \mathbf{f}'\sigma_\beta^2 \\
&+ \frac{1}{2}\left[tr(\mathbf{D}_\beta^2 f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{D}_\beta^2 f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\, \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\right]\sigma_\beta^4 \qquad (3\text{-}24)
\end{aligned}
$$

Spatial covariance of $\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*)$ exists whether or not the model is nonlinear. The magnitudes of the sensitivities $\mathbf{D}_\beta \mathbf{f}$ strongly influence the magnitudes of the covariances.

The product $\gamma(\gamma'\gamma)^{-1}\gamma'$ can be evaluated for the zonation example and then applied to $\mathbf{D}_\beta^2 f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma'$ to illustrate the meaning of this term. From (3-2) and (3-9)

$$
\gamma(\gamma'\gamma)^{-1}\gamma' = \begin{bmatrix} m_1^{-1}\mathbf{1}_1\mathbf{1}_1' & \mathbf{0}_{12} & \cdots & \mathbf{0}_{1p} \\ \mathbf{0}_{21} & m_2^{-1}\mathbf{1}_2\mathbf{1}_2' & \cdots & \mathbf{0}_{2p} \\ & \cdots & & \\ \mathbf{0}_{p1} & \mathbf{0}_{p2} & \cdots & m_p^{-1}\mathbf{1}_p\mathbf{1}_p' \end{bmatrix} \qquad (3\text{-}25)
$$

where $\mathbf{0}_{kl}$ is an $m_k \times m_l$ submatrix of zeros. Thus, a submatrix of $\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma'$ corresponding to parameters *k* and *l* is

$$
(m_k m_l)^{-1}\mathbf{1}_k \sum_{q(k)}\sum_{r(l)} D_{\beta qr}^2 f_i \mathbf{1}_l' = \frac{1}{m_k m_l}\sum_{q(k)}\sum_{r(l)} \frac{\partial^2 f_i}{\partial\beta_q \partial\beta_r}\mathbf{1}_k\mathbf{1}_l' \qquad (3\text{-}26)
$$

which is a $m_k \times m_l$ submatrix for which each element is the average second derivative with respect to values of $\beta_j$ pertaining to parameters *k* and *l*. From this it is apparent that $\mathbf{D}_\beta^2 f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\, \gamma(\gamma'\gamma)^{-1}\gamma'$ is a matrix of deviations of $\partial^2 f_i / \partial\beta_j \partial\beta_k$ values from their averages as defined in (3-26). If these deviations are small, then the second term in (3-24) may be small, even if the magnitude of $\mathbf{D}_\beta^2\mathbf{f}$ were large. A similar observation may be made for (3-19).

For purposes of comparison, the expected value and variance of the difference $\mathbf{f}(\beta) - \mathbf{f}(\gamma\bar{\theta})$ also can be computed. From (3-12) and (3-15)

$$f_i(\beta)-f_i(\gamma\bar{\theta})=\mathbf{D}_\beta f_i\mathbf{e}+\frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 f_i\mathbf{e}\,;\,i=1,2,\ldots,n \tag{3-27}$$

Hence, using the same procedures as before,

$$E(f_i(\beta)-f_i(\gamma\bar{\theta}))=\frac{1}{2}tr(\mathbf{D}_\beta^2 f_i\mathbf{V}_\beta)\sigma_\beta^2\,;\,i=1,2,\ldots,n \tag{3-28}$$

and

$$Var\left(\mathbf{f}(\beta)-\mathbf{f}(\gamma\bar{\theta})\right)=\mathbf{D}_\beta\mathbf{f}\mathbf{V}_\beta\mathbf{D}_\beta\mathbf{f}'\sigma_\beta^2+\frac{1}{2}\left[tr\left(\mathbf{D}_\beta^2 f_i\mathbf{V}_\beta\mathbf{D}_\beta^2 f_j\mathbf{V}_\beta\right)\right]\sigma_\beta^4 \tag{3-29}$$

Because $\gamma\theta_*$ is a best-fit vector to $\beta$, $f_i(\gamma\theta_*)$ would be expected to be closer to $f_i(\beta)$ than $f_i(\gamma\bar{\theta})$ would. In this case $E(\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*))$ should be smaller in magnitude than $E(\mathbf{f}(\beta)-\mathbf{f}(\gamma\bar{\theta}))$, and $Var(\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*))$ should be smaller in magnitude than $Var(\mathbf{f}(\beta)-\mathbf{f}(\gamma\bar{\theta}))$.

Neuman and Orr (1993) obtained a result almost analogous to (3-28) using a much different method. They allowed for conditioning of their results on possible hydraulic conductivity data. They then showed that the conditional (and unconditional) ensemble mean hydraulic head and flux distribution (analogous to $E(\mathbf{f}(\beta))$ are not obtained from a solution of the standard ground-water flow equation written in terms of the conditional (or unconditional) ensemble mean hydraulic conductivity distribution (analogous to $\mathbf{f}(\gamma\bar{\theta})$). They also derived a correction term analogous in effect to the trace term in (3-28). The main conceptual differences are that (3-28) applies to all types of system characteristics but involves only the unconditional mean $\bar{\theta}$.

## Properties of the Error Vector, $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$

The above discussion focused on errors in the model function resulting from smoothing the small-scale variability inherent in $\beta$ by replacing $\beta$ with $\gamma\theta_*$. These errors are model errors. (See also discussion by Hill, 1992, p. 42-43.) By adding the observation-error vector $\mathbf{Y}-\mathbf{f}(\beta)$ from (3-5) to the model-error vector $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ from (3-18), the total-error vector $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ is obtained as

$$\begin{aligned}Y_i-f_i(\gamma\theta_*)&=Y_i-f_i(\beta)+f_i(\beta)-f_i(\gamma\theta_*)\\&=\varepsilon_i+\mathbf{D}_\beta f_i(\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}+\frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_i-\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}\,;\,i=1,2,\ldots,n\end{aligned} \tag{3-30}$$

The expected value and variance of the total error are obtained from (3-30). From (3-6) it can be seen that $E(\varepsilon)=\mathbf{0}$. Thus, the expected value of $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ is the same as the expected value of $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$, or

$$E(Y_i - f_i(\gamma\theta_*)) = \frac{1}{2} tr((\mathbf{D}^2_\beta f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_i \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\sigma^2_\beta \,;\, i = 1,2,\ldots,n \tag{3-31}$$

Also, because $\varepsilon$ and $\mathbf{e}$ are assumed to be statistically independent, the variance of $\mathbf{Y} - \mathbf{f}(\gamma\theta_*)$ is

$$\begin{aligned} &Var(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) = Var(\varepsilon) + Var(\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*)) \\ &= \mathbf{V}_\varepsilon\sigma^2_\varepsilon + \mathbf{D}_\beta\mathbf{f}(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{D}_\beta\mathbf{f}'\sigma^2_\beta \\ &+ \frac{1}{2}\left[tr((\mathbf{D}^2_\beta f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_i \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{D}^2_\beta f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_j \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\right]\sigma^4_\beta \end{aligned} \tag{3-32}$$

For purposes that will become apparent in sections 4 and 5, the linear-model component of the variance (which involves $\mathbf{D}_\beta\mathbf{f}$ but not $\mathbf{D}^2_\beta\mathbf{f}$ ) is defined as

$$\mathbf{V}_*\sigma^2_\varepsilon = \mathbf{V}_\varepsilon\sigma^2_\varepsilon + \mathbf{D}_\beta\mathbf{f}(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{D}_\beta\mathbf{f}'\sigma^2_\beta \tag{3-33}$$

so that

$$\begin{aligned} &Var(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) = \mathbf{V}_*\sigma^2_\varepsilon \\ &+ \frac{1}{2}\left[tr((\mathbf{D}^2_\beta f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_i \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{D}^2_\beta f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_j \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\right]\sigma^4_\beta \end{aligned} \tag{3-34}$$

For comparison $\mathbf{Y} - \mathbf{f}(\gamma\bar{\theta})$ can be expanded like (3-30) using (3-12) and (3-15) to obtain

$$Y_i - f_i(\gamma\bar{\theta}) = \varepsilon_i + \mathbf{D}_\beta f_i\mathbf{e} + \frac{1}{2}\mathbf{e}'\mathbf{D}^2_\beta f_i\mathbf{e} \,;\, i = 1,2,\ldots,n \tag{3-35}$$

from which

$$E(Y_i - f_i(\gamma\bar{\theta})) = \frac{1}{2} tr(\mathbf{D}^2_\beta f_i\mathbf{V}_\beta)\sigma^2_\beta;\; i = 1,2,\ldots,n \tag{3-36}$$

and

$$Var(\mathbf{Y} - \mathbf{f}(\gamma\bar{\theta})) = \mathbf{V}_\varepsilon\sigma^2_\varepsilon + \mathbf{D}_\beta\mathbf{f}\,\mathbf{V}_\beta\mathbf{D}_\beta\,\mathbf{f}'\sigma^2_\beta + \frac{1}{2}\left[tr(\mathbf{D}^2_\beta f_i\mathbf{V}_\beta\mathbf{D}^2_\beta f_j\mathbf{V}_\beta)\right]\sigma^4_\beta \tag{3-37}$$

Note that because of the quadratic terms involving $\mathbf{e}$ in both (3-30) and (3-35), neither set of total errors is normally distributed.

## Reducing Model Error

It is important to consider how the model error terms in the variance (3-32) might be reduced. Some ideas are obtained by examining $\mathbf{D}_\beta f_i(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{D}_\beta f'_j$ .

First of all, if $p = m$, then $\gamma(\gamma'\gamma)^{-1}\gamma' = \mathbf{I}$, and the term is zero. Because $m$ can be very large, generally much larger than $n$, letting $p = m$ is not generally possible. Temporarily,

$$\mathbf{V} = (\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma') \tag{3-38}$$

for convenience, so that $\mathbf{D}_\beta f_i \mathbf{V} \mathbf{D}_\beta f_j'$ may be examined. In (C-14), appendix C, $\mathbf{D}_\beta \mathbf{f}\gamma$ is shown to be independent of $m$, so that $\mathbf{D}_\beta \mathbf{f}$ has elements of order $m^{-1}$, termed $O(m^{-1})$, in magnitude. Therefore, the magnitude of the model error term may be small if $m$ is large and $\mathbf{V}$ is diagonal, because for $\mathbf{V}$ diagonal

$$\mathbf{D}_\beta f_i \mathbf{V} \mathbf{D}_\beta f_j' = \sum_{k=1}^{m} \frac{\partial f_i}{\partial \beta_k} V_{kk} \frac{\partial f_j}{\partial \beta_k} = \sum_{k=1}^{m} O(m^{-1}) V_{kk} O(m^{-1}) = O(m^{-1}) \tag{3-39}$$

Matrix $\mathbf{V}$ can approach diagonal if $\mathbf{V}_\beta$ approaches diagonal and $m$ is large. This is because as $m \to \infty$, $(\gamma'\gamma)^{-1} \to \mathbf{0}$, so that $\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma' \to \mathbf{I}$. (See, for example, (3-8) and (3-25).) Matrix $\mathbf{V}_\beta$ approaches diagonal as the correlation lengths implied by its covariances get shorter, but is full with increasingly larger covariances as correlation lengths get longer.

To some extent the stochastic process generating $\beta$ can be designed at the descretion of the investigator constructing a model. The above analysis suggests that all trends and features of large enough scale to be identified be removed from the stochastic process and incorporated into $\gamma\bar{\theta}$. This should be done even if geostatistical analysis indicates that $\beta$ (or some subset of it) could be represented differently, for example as a stationary random process having a long enough correlation length that an identified trend could be interpreted as a random fluctuation. It is better to assign the trend to the drift and reinterpret the stochastic process to reduce both the magnitude of $\mathbf{V}_\beta$ and the correlation length. Ideally, $\mathbf{V}_\beta$ should represent only random noise, with no nonzero covariances. If only short correlation lengths are implied by $\mathbf{V}_\beta$, then $\mathbf{D}_\beta f_i \mathbf{V} \mathbf{D}_\beta f_j'$ should be close to $O(m^{-1})$. For large $m$, then, the term $\mathbf{D}_\beta f_i \mathbf{V} \mathbf{D}_\beta f_j'$ could be small. However, if significant features are not identified and incorporated into $\gamma\bar{\theta}$, elements in $\mathbf{V}_\beta$ could be large and correlation lengths could be long simply as a result of the unidentified features. Because $\mathbf{V}_\beta$ would not be diagonal, a large value of $m$ would not necessarily make the magnitude of the term small. Thorough hydrogeologic field work, resulting in identification and incorporation of all significant hydrogeologic features, is very important.

## Modifications for Unsteady Flow

For general unsteady flow, $\beta$ is distributed in both space and time. The drift $\gamma\bar{\theta}$ also extends over both space and time, which can be accomplished by making $\gamma$ an interpolation matrix over both space and time and allowing elements of $\bar{\theta}$ to be different at different points in time. Vector $\beta$ is distributed as a correlated random function around the drift, but probably would often be continuously variable temporally rather than discretely variable. However, if discrete time elements are small enough, then discrete time variation approximates continuous variation. This viewpoint has the advantage of not requiring a separate formulation from the one adopted for spatial variability.

As an example of time-dependent variability, let the vector of drift parameters at any time $t$ between times $t_r$ and $t_{r+1}$ be well described by a linear function of time. Then

$$\bar{\theta}(t) = \sigma_r \bar{\theta}_r + \sigma_{r+1} \bar{\theta}_{r+1} \tag{3-40}$$

where $\bar{\theta}_r$ is the set of drift parameters at time $t_r$, $\bar{\theta}_{r+1}$ is the set of drift parameters at time $t_{r+1}$, and

$$\sigma_r = \frac{t_{r+1} - t}{t_{r+1} - t_r} \qquad \sigma_{r+1} = \frac{t - t_r}{t_{r+1} - t_r} \tag{3-41}$$

If the spatial variability of the drift can be approximated by zones of constant value, then $\gamma$ of (3-2) is replaced for $r$=1, for example, by

$$\gamma = \begin{bmatrix} \sigma_1^1 \mathbf{1}_1 & \mathbf{0}_1 & \mathbf{0}_1 & \cdots & \mathbf{0}_1 & \sigma_2^1 \mathbf{1}_1 & \mathbf{0}_1 & \mathbf{0}_1 & \cdots & \mathbf{0}_1 \\ \mathbf{0}_2 & \sigma_1^1 \mathbf{1}_2 & \mathbf{0}_2 & \cdots & \mathbf{0}_2 & \mathbf{0}_2 & \sigma_2^1 \mathbf{1}_2 & \mathbf{0}_2 & \cdots & \mathbf{0}_2 \\ & & & & \cdots & & & & & \\ \mathbf{0}_p & \mathbf{0}_p & \mathbf{0}_p & \cdots & \sigma_1^1 \mathbf{1}_p & \mathbf{0}_p & \mathbf{0}_p & \mathbf{0}_p & \cdots & \sigma_2^1 \mathbf{1}_p \\ \sigma_1^2 \mathbf{1}_1 & \mathbf{0}_1 & \mathbf{0}_1 & \cdots & \mathbf{0}_1 & \sigma_2^2 \mathbf{1}_1 & \mathbf{0}_1 & \mathbf{0}_1 & \cdots & \mathbf{0}_1 \\ \mathbf{0}_2 & \sigma_1^2 \mathbf{1}_2 & \mathbf{0}_2 & \cdots & \mathbf{0}_2 & \mathbf{0}_2 & \sigma_2^2 \mathbf{1}_2 & \mathbf{0}_2 & \cdots & \mathbf{0}_2 \\ & & & & \cdots & & & & & \\ \mathbf{0}_p & \mathbf{0}_p & \mathbf{0}_p & \cdots & \sigma_1^2 \mathbf{1}_p & \mathbf{0}_p & \mathbf{0}_p & \mathbf{0}_p & \cdots & \sigma_2^2 \mathbf{1}_p \\ & & & & \cdots & & & & & \end{bmatrix} \tag{3-42}$$

where $\mathbf{1}_i$ and $\mathbf{0}_i$ are given by (3-3) and superscripts of the form $k$ on the $\sigma$ functions designate discrete time elements of $\beta$ centered at $t^k$ between $t_1$ and $t_2$. Any number of these elements may be accommodated. Vector $\bar{\theta}$ is redefined to correspond with $\gamma$ as

$$\bar{\theta} = \begin{bmatrix} \bar{\theta}_1 \\ \bar{\theta}_2 \end{bmatrix} \tag{3-43}$$

This redefined vector is not time dependent. Finally, note that a more complicated time dependent drift can be accommodated by using more than one linear function contiguously in time so that $r > 1$. The added functions add rows and columns in block diagonal form to (3-42). A more complicated function of time than linear might often reduce the number $r$ and thus the number of parameters in $\bar{\theta}$.

Some variables such as hydraulic conductivity might be constant with time, but others such as recharge might vary, even somewhat erratically, with time. Vector $\beta$ and drift $\gamma\bar{\theta}$ must reflect these types of variability. The type of drift variability is easily specified using parameter subsets $\bar{\boldsymbol{\theta}}_r$, but specifying the type of time variability in $\boldsymbol{\beta}$ also requires the use of $\mathbf{V}_\beta\sigma_\beta^2$. Let some variable at a point in space be given at two different points in time as $\beta_i$ and $\beta_j$, and let corresponding rows of $\boldsymbol{\gamma}$ be $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$. Then, if the variable is constant in time, $\boldsymbol{\gamma}_i\bar{\boldsymbol{\theta}} = \boldsymbol{\gamma}_j\bar{\boldsymbol{\theta}}$ and $\beta_i = \beta_j$ so that for all $k$, $Cov(\beta_i, \beta_k) = Cov(\beta_j, \beta_k)$ (and symmetric relations). Thus, $V_{\beta ik} = V_{\beta jk} = V_{\beta ki} = V_{\beta kj}$. These relations further induce the relations $V_{\beta ii} = V_{\beta jj} = V_{\beta ij} = V_{\beta ji}$, which specify that the correlation between $\beta_i$ and $\beta_j$ is unity. Note that, although a correlation of unity makes $\mathbf{V}_\beta$ singular, matrices such as $\mathbf{D}_\beta\mathbf{f}(\mathbf{I}-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{V}_\beta(\mathbf{I}-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{D}_\beta\mathbf{f}'\sigma_\beta^2$ probably would not be singular as shown below; $\mathbf{V}_*\sigma_\varepsilon^2$ almost certainly would not be singular.

As an example, let the spatial order of $\boldsymbol{\beta}$ be 3 and the temporal order be 2, so that each of the pairs $(\beta_1, \beta_4)$, $(\beta_2, \beta_5)$, and $(\beta_3, \beta_6)$ is a variable at a fixed spatial location for two different time elements. Then the full covariance matrix $\mathbf{V}_\beta\sigma_\beta^2$ is

$$\mathbf{V}_\beta\sigma_\beta^2 = \begin{bmatrix} V_{\beta 11} & V_{\beta 12} & V_{\beta 13} & V_{\beta 14} & V_{\beta 15} & V_{\beta 16} \\ V_{\beta 21} & V_{\beta 22} & V_{\beta 23} & V_{\beta 24} & V_{\beta 25} & V_{\beta 26} \\ V_{\beta 31} & V_{\beta 32} & V_{\beta 33} & V_{\beta 34} & V_{\beta 35} & V_{\beta 36} \\ V_{\beta 41} & V_{\beta 42} & V_{\beta 43} & V_{\beta 44} & V_{\beta 45} & V_{\beta 46} \\ V_{\beta 51} & V_{\beta 52} & V_{\beta 53} & V_{\beta 54} & V_{\beta 55} & V_{\beta 56} \\ V_{\beta 61} & V_{\beta 62} & V_{\beta 63} & V_{\beta 64} & V_{\beta 65} & V_{\beta 66} \end{bmatrix}\sigma_\beta^2 \tag{3-44}$$

Now let $\beta_1 = \beta_4$, $\beta_2 = \beta_5$, and $\beta_3 = \beta_6$ so that all three variables are constant in time. Then $V_{\beta 1k} = V_{\beta 4k} = V_{\beta k1} = V_{\beta k4}$, $V_{\beta 2k} = V_{\beta 5k} = V_{\beta k2} = V_{\beta k5}$, and $V_{\beta 3k} = V_{\beta 6k} = V_{\beta k3} = V_{\beta k6}$. Application of the three sets of equalities yields

$$\mathbf{V}_\beta\sigma_\beta^2 = \begin{bmatrix} V_{\beta 11} & V_{\beta 12} & V_{\beta 13} & V_{\beta 11} & V_{\beta 12} & V_{\beta 13} \\ V_{\beta 21} & V_{\beta 22} & V_{\beta 23} & V_{\beta 21} & V_{\beta 22} & V_{\beta 23} \\ V_{\beta 31} & V_{\beta 32} & V_{\beta 33} & V_{\beta 31} & V_{\beta 32} & V_{\beta 33} \\ V_{\beta 11} & V_{\beta 12} & V_{\beta 13} & V_{\beta 11} & V_{\beta 12} & V_{\beta 13} \\ V_{\beta 21} & V_{\beta 22} & V_{\beta 23} & V_{\beta 21} & V_{\beta 22} & V_{\beta 23} \\ V_{\beta 31} & V_{\beta 32} & V_{\beta 33} & V_{\beta 31} & V_{\beta 32} & V_{\beta 33} \end{bmatrix}\sigma_\beta^2 \tag{3-45}$$

which is a four-fold repetition of the underlying $3 \times 3$ spatial covariance matrix. Thus, if the rank of the spatial covariance matrix is 3, the rank of the full covariance matrix is also 3 because each row and column is repeated three times. (The determinant of a matrix is zero if any row or column is repeated.) This argument generalizes inductively, so that, in general, the rank of $\mathbf{V}_\beta\sigma_\beta^2$ is no less than the rank of the spatial covariance matrix contained within it.

Finally, elements of matrices such as $\mathbf{D}_\beta\mathbf{f}$ and $\mathbf{D}_\beta^2\mathbf{f}$ are computed at their respective spatial and temporal points. Thus, even if all variables specified in $\boldsymbol{\beta}$ are constant in time, covariance matrices such as $\mathbf{D}_\beta\mathbf{f}(\mathbf{I}-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{V}_\beta(\mathbf{I}-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{D}_\beta\mathbf{f}'\sigma_\beta^2$ or $\mathbf{D}_\beta\mathbf{f}\mathbf{V}_\beta\mathbf{D}_\beta\mathbf{f}'\sigma_\beta^2$

reflect the temporal variation of $f(\beta)$ inherent in unsteady flow. Consider the example for which $\mathbf{V}_\beta \sigma_\beta^2$ was computed using (3-45) for simplicity. Then an element of the matrix is

$$\mathbf{D}_\beta f_i \mathbf{V}_\beta \mathbf{D}_\beta f_j' = \sum_{k=1}^{6}\sum_{\ell=1}^{6} \frac{\partial f_i}{\partial \beta_k} V_{\beta k\ell} \frac{\partial f_j}{\partial \beta_\ell} = \sum_{k=1}^{3}\sum_{\ell=1}^{6} \frac{\partial f_i}{\partial \beta_k} V_{\beta k\ell} \frac{\partial f_j}{\partial \beta_\ell} + \sum_{k=1}^{3}\sum_{\ell=1}^{6} \frac{\partial f_i}{\partial \beta_{k+3}} V_{\beta k\ell} \frac{\partial f_j}{\partial \beta_\ell}$$

$$= \sum_{k=1}^{3}\sum_{\ell=1}^{3} \frac{\partial f_i}{\partial \beta_k} V_{\beta k\ell} \frac{\partial f_j}{\partial \beta_\ell} + \sum_{k=1}^{3}\sum_{\ell=1}^{3} \frac{\partial f_i}{\partial \beta_k} V_{\beta k\ell} \frac{\partial f_j}{\partial \beta_{\ell+3}}$$

$$+ \sum_{k=1}^{3}\sum_{\ell=1}^{3} \frac{\partial f_i}{\partial \beta_{k+3}} V_{\beta k\ell} \frac{\partial f_j}{\partial \beta_\ell} + \sum_{k=1}^{3}\sum_{\ell=1}^{3} \frac{\partial f_i}{\partial \beta_{k+3}} V_{\beta k\ell} \frac{\partial f_j}{\partial \beta_{\ell+3}}$$

$$= \sum_{j=1}^{3}\sum_{k=1}^{3} \left(\frac{\partial f_i}{\partial \beta_k} + \frac{\partial f_i}{\partial \beta_{k+3}}\right) V_{\beta k\ell} \left(\frac{\partial f_j}{\partial \beta_\ell} + \frac{\partial f_j}{\partial \beta_{\ell+3}}\right) \tag{3-46}$$

Now, the sensitivities are not truly additive because $\beta_k$ (or $\beta_\ell$) applies for one time element and $\beta_{k+3}$ (or $\beta_{\ell+3}$) applies for the next contiguous time element. Thus, the sums represent the time varying sensitivities. If flow were steady, then the sensitivities would be constant in time, so that the full covariance matrix would be just the steady-state covariance matrix.

The above analysis shows that all results obtained for steady flow can be applied for unsteady flow, if $\gamma$, $\bar{\theta}$, and $\beta$ incorporate the time variant properties. Thus, no distinction between the two types of flow is made in subsequent developments.

## Summary of Principal Results

The ground-water flow system is assumed to be adequately characterized by a set of $m$ system characteristics, termed $\beta$, that fall into three categories: 1) variables such as hydraulic conductivity, recharge, and discharge that can vary spatially throughout the model; 2) variables such as hydraulic heads and fluxes that can vary spatially along internal and external boundaries of the model; and 3) variables such as spring and well discharges that occur locally, at points. Initially, flow also is assumed to be steady state so that $\beta$ varies only spatially. Vector $\beta$ thus contains discrete values of the system characteristics in volume elements for category 1, boundary segments for category 2, or points for category 3. The discretization is assumed to be fine enough that any model function of $\beta$, $f(\beta)$, is almost free of model error.

Vector $\beta$ represents small-scale variability that cannot be explicitly described in a model and larger-scale variability that can. The influence of the small-scale variability is included in a model by imagining a stochastic process for $\beta$ and using the stochastic properties of $\beta$ in modeling. Specifically, in (3-1) $\beta$ is assumed to be normally distributed with a mean given by an interpolation $\gamma\bar{\theta}$ of a set of $p$ drift parameters $\bar{\theta}$ and covariance given by a spatial covariance matrix $\mathbf{V}_\beta \sigma_\beta^2$.

A set of $n$ observations $\mathbf{Y}$ differs from a set of $n$ corresponding values $\mathbf{f}(\beta)$ of the model function $f(\beta)$ by an observation-error vector $\varepsilon$, which is considered to be normally distributed with a mean vector of zero and covariance matrix $\mathbf{V}_\varepsilon \sigma_\varepsilon^2$. The model and distribution are stated by (3-5) and (3-6), respectively. The normality assumed for $\beta$ and $\varepsilon$ is not essential for the theory developed in this report, and the assumption can be indirectly tested.

The vector $\beta$ has too large a dimension to be estimated. Hence, a vector of reduced dimension that has the same form as the drift is to be estimated. The vector to be estimated is obtained as the vector $\gamma\theta_*$ that is the best fit to $\beta$. It is derived by minimizing the criterion $(\beta-\gamma\theta)'(\beta-\gamma\theta)$ with respect general set of parameters $\theta$ to obtain $\theta_* = (\gamma'\gamma)^{-1}\gamma'\beta$, which is (3-7). The estimate of $\theta_*$ is derived in section 4.

The model function $\mathbf{f}(\beta)$ and model function $\mathbf{f}(\gamma\theta_*)$ do not in general have the same expected value (ensemble average) if the model is nonlinear in $\beta$ or $\gamma\theta_*$. A third-order correct result for the expected value of the difference, or model error, $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ is given by (3-19). The covariance matrix for this difference is given by (3-24) and indicates that model error resulting from replacing $\beta$ with $\gamma\theta_*$ can be highly correlated throughout the model.

Error vector $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ is obtained by adding the observation error $\mathbf{Y}-\mathbf{f}(\beta)$ to the model error $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$. The expected value of this error is the same as the expected value of $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ because the expected value of the observation error is zero. The covariance matrix for the error vector, (3-32), is obtained by adding $\mathbf{V}_\varepsilon\sigma_\varepsilon^2$ to the variance of $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ because $\beta$ and $\varepsilon$ are assumed to be statistically independent. If $\mathbf{V}_\varepsilon\sigma_\varepsilon^2$ is diagonal or nearly so, the correlations among the errors $Y_i - f_i(\gamma\theta_*)$, $i = 1,2,\ldots,n$, are reduced over the correlations among the differences $f_i(\beta)-f_i(\gamma\theta_*)$, $i = 1,2,\ldots,n$.

Model error can be reduced by selecting the stochastic process so that correlation lengths in $\mathbf{V}_\beta$ are as short as possible. That is, trends and significant hydrogeologic features should be represented in $\gamma\bar{\theta}$ so that $\mathbf{V}_\beta\sigma_\beta^2$ represents mostly short-correlation length variability. This deduction results from an analysis of the covariance matrix for $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ given by (3-39) that showed that the magnitudes of the covariance terms can be small if $\mathbf{V}_\beta$ is diagonal and $m$ is large.

For general unsteady flow $\beta$ is distributed in both space and time, and the drift $\gamma\bar{\theta}$ varies in both space and time. Time variation of the drift is accomplished in the theory developed in this report by making $\gamma$ an interpolation matrix in both space and time and by allowing elements of $\bar{\theta}$ to be different at different points in time. Time variation of $\beta$ is approximated using the same discrete viewpoint adopted for spatial variation because time elements can be made small enough to approximate continuous variation. All results obtained for steady flow can be applied for unsteady flow, if $\gamma$, $\bar{\theta}$, and $\beta$ incorporate the time variant properties. The vectors and matrices are simply augmented to account for any number of time elements and a time variant drift. Thus, further developments will not distinguish between the two types of flow.

# 4. Estimation and Prediction

## Estimation of the Vector of Spatial Average System Characteristics, $\theta_*$

Vector $\theta_*$ must be estimated because, being a linear combination involving $\beta$, it is unknown. Vector $\theta_*$ and the procedure used to estimate it must both be constructed so that $\theta_*$ has a unique estimate. Weighted least squares estimation is shown here to lead to desirable properties and uncertainty estimates for parameters and predictions. For this method the following objective function is minimized.

$$S(\theta) = (\mathbf{Y} - \mathbf{f}(\gamma\theta))'\omega(\mathbf{Y} - \mathbf{f}(\gamma\theta)) \tag{4-1}$$

where $\theta$ is an arbitrary vector of parameters of order $p$ and $\omega$ is an arbitrary, positive definite $n \times n$ weight matrix, possible forms for which are to be developed. Note that, strictly speaking, weighted least squares is the term often applied when $\omega$ is diagonal (Draper and Smith, 1998, p. 223). However, Seber and Wild (1989, p. 27) use the term as a synonym for generalized least squares, which is expressed for the theory developed in this report as $\omega^{-1} \propto E(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'$ as shown in section 5. For nomenclatural convenience, the term weighted least squares is generalized further in this report so that $\omega$ can be arbitrary and nondiagonal, but is positive definite. The term generalized least squares (also called Gauss-Markov estimation) is applied when $\omega^{-1} \propto E(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'$. Another objective function is introduced when $\omega$ is unknown. Because weights are approximated, this case can be termed approximate weighted least squares.

Note that model error is included in the matrix $E(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'$, which is not standard statistical usage (for example, Seber and Wild, 1989, p. 28). This is not without precedent. For example, Tasker and Stedinger (1989) employed a similar idea to derive a generalized least squares model for regional regression analysis of floods.

In modeling studies, $S(\theta)$ is minimized using standard techniques of nonlinear regression such as adaptive least squares (Cooley and Hill, 1992). However, to develop the theory that is used to analyze the estimates and predictions to be made with them, and to develop the theory underlying the uncertainty analysis methods, an approximate analytical solution of the minimization problem is needed. This solution is obtained using extensions of methods given by Johansen (1983). First, the linear-model component of the error vector $\mathbf{Y} - \mathbf{f}(\gamma\bar{\theta})$ is defined as

$$\mathbf{U} = \varepsilon + \mathbf{D}_\beta \mathbf{fe} \tag{4-2}$$

which from (3-1) and (3-6) has the normal distribution

$$\mathbf{U} \sim N(\mathbf{0}, \mathbf{V}_\varepsilon \sigma_\varepsilon^2 + \mathbf{D}_\beta \mathbf{fV}_\beta \mathbf{D}_\beta \mathbf{f}'\sigma_\beta^2) \tag{4-3}$$

Next the estimate $\hat{\theta}$ is expressed as $\bar{\theta}$, plus a term $\mathbf{l}$ that is first order in $\mathbf{U}$, plus a term $\mathbf{q}$ that is second order in $\mathbf{U}$, $\mathbf{e}$, and their product, or

$$\hat{\theta} = \bar{\theta} + \mathbf{l} + \mathbf{q} \tag{4-4}$$

Vectors $\mathbf{l}$ and $\mathbf{q}$ are obtained in appendix B by a combination Taylor series expansion and perturbation technique that formally assumes $Var(\mathbf{U})$ to be small. The solutions are

$$\mathbf{l} = (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{U} \tag{4-5}$$

and

$$\mathbf{q} = (\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_i \mathbf{D}^2 f_i \mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z} + \tfrac{1}{2}\mathbf{Df}'\sum_j \omega_j\,(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l})) \tag{4-6}$$

where $\omega_i^{1/2}$ and $\omega_j$ stand for row $i$ and column $j$ of $\omega^{1/2}$ and $\omega$, respectively, and

$$\mathbf{Z} = (\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}\mathbf{U} \tag{4-7}$$

in which

$$\mathbf{R} = \omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}} \tag{4-8}$$

The robustness of results obtained using (4-4) is explored in section 7.

## Bias in the Estimate $\hat{\theta}$ of $\theta_*$

Bias in the estimate $\hat{\theta}$ of $\theta_*$ results because the model functions $\mathbf{f}(\beta)$, $\mathbf{f}(\gamma\theta_*)$, and $\mathbf{f}(\gamma\hat{\theta})$ are nonlinear in $\beta$, $\theta_*$, and $\hat{\theta}$, respectively. The bias is derived here in the same manner as used in standard nonlinear regression (Seber and Wild, 1989, p. 182) except that in the present instance there are additional influences from model error. The following development leads to the main result, (4-15).

First the linear-model component of the error vector $\mathbf{Y} - \mathbf{f}(\gamma\theta_*)$ is defined as

$$\mathbf{U}_* = \varepsilon + \mathbf{D}_\beta \mathbf{f}\mathbf{e}_* = \varepsilon + \mathbf{D}_\beta \mathbf{f}(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} \tag{4-9}$$

which has the normal distribution

$$\begin{aligned}\mathbf{U}_* &\sim N(\mathbf{0}, \mathbf{V}_\varepsilon\sigma_\varepsilon^2 + \mathbf{D}_\beta\mathbf{f}(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{D}_\beta\,\mathbf{f}'\sigma_\beta^2) \\ &= N(\mathbf{0}, \mathbf{V}_*\sigma_\varepsilon^2)\end{aligned} \tag{4-10}$$

where $\mathbf{V}_*\sigma_\varepsilon^2 = Var(\mathbf{U}_*)$ as defined by (3-33).

The difference $\hat{\theta}-\theta_*$ can be written using (3-17), (4-5), (4-6), and (4-9) as

$$\begin{aligned}
\hat{\theta}-\theta_* &= \hat{\theta}-\bar{\theta} - (\theta_* - \bar{\theta}) \\
&= \mathbf{l}+\mathbf{q}-(\gamma'\gamma)^{-1}\gamma'\mathbf{e} \\
&= (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{U} + (\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_i \mathbf{D}^2 f_i \mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z} + \frac{1}{2}\mathbf{Df}'\sum_j \omega_j(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})) \\
&\quad - (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{Df}(\gamma'\gamma)^{-1}\gamma'\mathbf{e} \\
&= (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{U}_* + (\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_i \mathbf{D}^2 f_i \mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z} + \frac{1}{2}\mathbf{Df}'\sum_j \omega_j(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))
\end{aligned} \tag{4-11}$$

where the result $\mathbf{Df} = \mathbf{D}_\beta\mathbf{f}\gamma$ ((C-14), appendix C) was used. Hence, the bias is

$$\begin{aligned}
E(\hat{\theta}-\theta_*) &= E(\mathbf{q}) \\
&= (\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_i E(\mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z}) + \frac{1}{2}\mathbf{Df}'\sum_j \omega_j E(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))
\end{aligned} \tag{4-12}$$

For reasons that are apparent later, $E(\mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z})$ is not evaluated. The second expected value is

$$E(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}) = tr(\mathbf{D}^2_\beta f_j\mathbf{V}_\beta\sigma^2_\beta - \mathbf{D}^2 f_j Var(\mathbf{l})) \tag{4-13}$$

in which (4-5) is used to yield

$$Var(\mathbf{l}) = (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega(\mathbf{V}_\varepsilon\sigma^2_\varepsilon + \mathbf{D}_\beta\mathbf{f}\mathbf{V}_\beta\mathbf{D}_\beta\mathbf{f}'\sigma^2_\beta)\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1} \tag{4-14}$$

With (4-13), the bias is written

$$E(\hat{\theta}-\theta_*) = (\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_i E(\mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z}) + \frac{1}{2}\mathbf{Df}'\sum_j \omega_j\, tr(\mathbf{D}^2_\beta f_j\mathbf{V}_\beta\sigma^2_\beta - \mathbf{D}^2 f_j Var(\mathbf{l}))) \tag{4-15}$$

Because $E(\theta_*) = \bar{\theta}$, the bias in $\hat{\theta}$ as an estimate of $\theta_*$ is the same as the bias in $\hat{\theta}$ as an estimate of $\bar{\theta}$.

## Bias in Estimates of the Model Function Vectors $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta_*)$

Expression of the difference $f_i(\gamma\hat{\theta}) - f_i(\beta)$ using (B-11), appendix B, yields

$$\begin{aligned}
f_i(\gamma\hat{\theta}) - f_i(\beta) &= -(f_i(\beta) - f_i(\gamma\bar{\theta})) + f_i(\gamma\hat{\theta}) - f_i(\gamma\bar{\theta}) \\
&\approx -\mathbf{D}_\beta f_i\,\mathbf{e} - \frac{1}{2}\mathbf{e}'\mathbf{D}^2_\beta f_i\mathbf{e} + \omega_i^{-\frac{1}{2}}\mathbf{R}(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}) + \frac{1}{2}\omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}
\end{aligned}$$

$$+\mathbf{D}f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}$$

$$=-\omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{D}_\beta f_j\mathbf{e}+\frac{1}{2}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))+\omega_i^{-\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\varepsilon}$$

$$+\mathbf{D}f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z} \qquad (4\text{-}16)$$

so that the bias may be written by using (4-13) to obtain

$$E(f_i(\gamma\hat{\theta})-f_i(\beta))\approx-\frac{1}{2}\omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}tr(\mathbf{D}_\beta^2 f_j\mathbf{V}_\beta\sigma_\beta^2-\mathbf{D}^2 f_j Var(\mathbf{l}))$$

$$+\mathbf{D}f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k E(\mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}) \qquad (4\text{-}17)$$

Again, model nonlinearity can cause bias. Similarly, writing the difference $f_i(\gamma\hat{\theta})-f_i(\gamma\theta_*)$ using (3-18) and (4-16) results in

$$f_i(\gamma\hat{\theta})-f_i(\gamma\theta_*)=f_i(\beta)-f_i(\gamma\theta_*)+f_i(\gamma\hat{\theta})-f_i(\beta)$$

$$\approx\mathbf{D}_\beta f_i(\mathbf{I}-\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}+\frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_i-\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\,\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}$$

$$-\omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{D}_\beta f_j\mathbf{e}+\frac{1}{2}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))+\omega_i^{-\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\varepsilon}$$

$$+\mathbf{D}f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z} \qquad (4\text{-}18)$$

so that the bias is

$$E(f_i(\gamma\hat{\theta})-f_i(\gamma\theta_*))\approx\frac{1}{2}tr((\mathbf{D}_\beta^2 f_i-\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\,\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\sigma_\beta^2$$

$$-\frac{1}{2}\omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}tr(\mathbf{D}_\beta^2 f_j\mathbf{V}_\beta\sigma_\beta^2-\mathbf{D}^2 f_j Var(\mathbf{l}))+\mathbf{D}f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k E(\mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}) \qquad (4\text{-}19)$$

The bias in $\mathbf{f}(\gamma\hat{\theta})$ as an estimate of $\mathbf{f}(\gamma\theta_*)$ is the difference between the bias in $\mathbf{f}(\gamma\hat{\theta})$ as an estimate of $\mathbf{f}(\beta)$ and the bias in $\mathbf{f}(\gamma\theta_*)$ as an estimate of $\mathbf{f}(\beta)$.

## Bias and Other Properties of the Residuals, $\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})$

***Bias.*** Residuals $\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})$ are the estimates of the errors $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$. Second-order approximations are computed in appendix B and are

$$Y_i - f_i(\gamma\hat{\theta}) \approx \omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l})) - \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}\,;$$
$$i = 1,2,\ldots,n \qquad (4\text{-}20)$$

Residuals have the same form when written in terms of $\mathbf{U}_*$ rather than $\mathbf{U}$. This important result is obtained by using the identity $\omega_i^{-1/2}(\mathbf{I}-\mathbf{R})\omega^{1/2}\mathbf{Df}(\gamma'\gamma)^{-1}\gamma'\mathbf{e} = \mathbf{0}$, which results because $(\mathbf{I}-\mathbf{R})\omega^{1/2}\mathbf{Df} = \mathbf{0}$:

$$Y_i - f_i(\gamma\hat{\theta}) \approx \omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l})) - \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}\,;$$
$$i = 1,2,\ldots,n \qquad (4\text{-}21)$$

Hence, either $\mathbf{U}$ or $\mathbf{U}_*$ can be used to get the expected values of the residuals as

$$E(Y_i - f_i(\gamma\hat{\theta})) \approx \frac{1}{2}\omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}} E(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}) - \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k E(\mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z})$$
$$= \frac{1}{2}\omega_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}} tr(\mathbf{D}_\beta^2 f_j \mathbf{V}_\beta \sigma_\beta^2 - \mathbf{D}^2 f_j Var(\mathbf{l})) - \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k E(\mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z})\,;$$
$$i = 1,2,\ldots,n \qquad (4\text{-}22)$$

Model nonlinearity can cause the expected values of the residuals to be biased as estimates of expected values of the errors $Y_i - f_i(\gamma\theta_*)$ given by (3-31). Model nonlinearity also can cause the residuals to have a non-normal distribution. (Note the quadratic terms involving $\mathbf{e}$ and $\mathbf{l}$ in (4-21).)

***Effects of nonlinearity on measures of non randomness from model error.*** Residuals from a modeling problem are commonly analyzed for indications of non randomness resulting from model error (Draper and Smith, 1998, p. 59-61; Cooley and Naff, 1990, p. 167-171; Hill, 1998, p. 20-24). A sum of all residuals of nearly zero is taken to indicate a good overall fit of a ground-water model to the data, and a nearly horizontal band of data for a plot of weighted residuals $\omega_i^{1/2}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}))$ in relation to weighted estimated model function values $\omega_i^{1/2}\mathbf{f}(\gamma\hat{\theta}))$ is taken to indicate a lack of model error. These measures also can be affected by nonlinearity. First the sum is examined. A weighted residual is written as

$$\omega_i^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}))$$
$$\approx (\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l})) - \omega_i^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z} \qquad (4\text{-}23)$$

where $(\mathbf{I}-\mathbf{R})_i$ is row $i$ of $\mathbf{I}-\mathbf{R}$. For a linear model the sum of weighted residuals should not be significant because, from $E(\mathbf{U}_*) = \mathbf{0}$, its expected value is zero. However, for a nonlinear model the expected value is

$$E(\sum_i \omega_i^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))) \approx \sum_i (\mathbf{I}-\mathbf{R})_i \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}} tr(\mathbf{D}_\beta^2 f_j \mathbf{V}_\beta \sigma_\beta^2 - \mathbf{D}^2 f_j Var(\mathbf{l}))$$
$$-\sum_i \omega_i^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k E(\mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}) \qquad (4\text{-}24)$$

It is possible for model nonlinearity to substantially increase the magnitude of the sum of residuals.

The extent to which a plot of weighted residuals in relation to estimated weighted model function values deviates from a horizontal band can be evaluated by computing the slope of a line through the data, which is given from the following development as (4-27). (This development can be skipped, if desired.) The standard equation for the slope in linear regression (Draper and Smith, 1998, p. 25) indicates that the slope is proportional to $\sum_i \omega_i^{1/2}\mathbf{f}(\gamma\hat{\theta})\omega_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})) - \sum_i \omega_i^{1/2}\mathbf{f}(\gamma\hat{\theta})\sum_i \omega_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))/n$. Evaluation of the first term in this expression through second order terms in $\mathbf{U}$, $\mathbf{e}$, and their product using (4-23) and (B-11), appendix B, yields

$$\sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\hat{\theta})\omega_i^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$$
$$\approx \sum_i (\omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\bar{\theta}) + \mathbf{R}_i(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}) + \frac{1}{2}(\mathbf{I}-\mathbf{R})_i\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}$$
$$+\omega_i^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z})((\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{l}'\mathbf{D}_\beta^2 f_j\mathbf{l}))$$
$$-\omega_i^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z})$$
$$\approx \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\bar{\theta})((\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))$$
$$-\omega_i^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}) \qquad (4\text{-}25)$$

where $\mathbf{R}_i$ is row $i$ of $\mathbf{R}$ and, because $\mathbf{R}$ is symmetric, idempotent (Cooley and Naff, 1990, p. 165), $\mathbf{R}(\mathbf{I}-\mathbf{R}) = \mathbf{R}'(\mathbf{I}-\mathbf{R}) = \mathbf{0}$. Similarly, evaluation of the second term results in

$$\sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\hat{\theta})\sum_i \omega_i^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))/n$$
$$\approx \sum_i (\omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\bar{\theta}) + \mathbf{R}_i(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}) + \frac{1}{2}(\mathbf{I}-\mathbf{R})_i\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}$$
$$+\omega_i^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z})\sum_i((\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{l}'\mathbf{D}_\beta^2 f_j\mathbf{l}))$$
$$-\omega_i^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z})/n$$

$$\approx \sum_i \omega_i^{\frac{1}{2}} \mathbf{f}(\gamma\bar{\theta}) \sum_i ((\mathbf{I}-\mathbf{R})_i (\omega^{\frac{1}{2}} \mathbf{U}_* + \frac{1}{2} \sum_j \omega_j^{\frac{1}{2}} (\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}_\beta^2 f_j \mathbf{l}))$$

$$- \omega_i^{\frac{1}{2}} \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1} \sum_k \mathbf{D}^2 f_k \mathbf{l} \omega_k^{\frac{1}{2}} \mathbf{Z})/n \qquad (4\text{-}26)$$

Combination of (4-25) and (4-26) gives

$$\sum_i \omega_i^{\frac{1}{2}} \mathbf{f}(\gamma\hat{\theta}) \omega_i^{\frac{1}{2}} (\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta})) - \sum_i \omega_i^{\frac{1}{2}} \mathbf{f}(\gamma\hat{\theta}) \sum_i \omega_i^{\frac{1}{2}} \mathbf{f}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}))/n$$

$$\approx \sum_i \omega_i^{\frac{1}{2}} \mathbf{f}(\gamma\bar{\theta}) ((\mathbf{I}-\mathbf{R})_i (\omega^{\frac{1}{2}} \mathbf{U}_* + \frac{1}{2} \sum_j \omega_j^{\frac{1}{2}} (\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))$$

$$- \omega_i^{\frac{1}{2}} \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1} \sum_k \mathbf{D}^2 f_k \mathbf{l} \omega_k^{\frac{1}{2}} \mathbf{Z}) - \sum_i \omega_i^{\frac{1}{2}} \mathbf{f}(\gamma\bar{\theta}) \sum_i ((\mathbf{I}-\mathbf{R})_i (\omega^{\frac{1}{2}} \mathbf{U}_* + \frac{1}{2} \sum_j \omega_j^{\frac{1}{2}} (\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}_\beta^2 f_j \mathbf{l}))$$

$$- \omega_i^{\frac{1}{2}} \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1} \sum_k \mathbf{D}^2 f_k \mathbf{l} \omega_k^{\frac{1}{2}} \mathbf{Z})/n \qquad (4\text{-}27)$$

The slope of a line through the plot is altered from what it would be for a linear model. For a linear model the slope should not be significant unless an intercept is needed, but for a nonlinear model it could be significant just because of model nonlinearity. Note that the slope has an expected value of zero for a linear model, but this is not necessarily true for a nonlinear model. Seber and Wild (1989, p. 179) cite the same behavior for the classical nonlinear model. The slope and its implications are analyzed further later in this section.

## Bias in Predictions, $g(\gamma\hat{\theta})$

Predictions to be made with the model also are affected by model nonlinearity. A prediction, defined as any function of $\gamma\hat{\theta}$ of interest that is not contained in $\mathbf{f}(\gamma\hat{\theta})$, is termed $g(\gamma\hat{\theta})$. That is, $g(\gamma\hat{\theta})$ can be the same type of function as any element $f_i(\gamma\hat{\theta})$ (such as a hydraulic head or flux), but was not observed as $Y_i$. Variables $g(\beta)$ and $g(\gamma\theta_*)$ are predicted using $g(\gamma\hat{\theta})$.

Development of the bias $E(g(\gamma\hat{\theta}) - g(\beta))$ starts with

$$g(\beta) = g(\gamma\bar{\theta}) + \mathbf{D}_\beta g(\beta - \gamma\bar{\theta}) + \frac{1}{2}(\beta - \gamma\bar{\theta})'\mathbf{D}_\beta^2 g(\beta - \gamma\bar{\theta})$$

$$= g(\gamma\bar{\theta}) + \mathbf{D}_\beta g \mathbf{e} + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 g \mathbf{e} \qquad (4\text{-}28)$$

where $\mathbf{D}_\beta g$ is the row vector $[\partial g / \partial \beta_j]$ evaluated at $\beta = \gamma\bar{\theta}$ and $\mathbf{D}_\beta^2 g$ is the matrix $[\partial^2 g / \partial \beta_i \partial \beta_j]$ evaluated at $\beta = \gamma\bar{\theta}$. Then, the difference $g(\gamma\hat{\theta}) - g(\beta)$ is written using (4-28) and (B-13), appendix B, as

$$g(\gamma\hat{\theta}) - g(\beta) = -(g(\beta) - g(\gamma\bar{\theta})) + g(\gamma\hat{\theta}) - g(\gamma\bar{\theta})$$
$$\approx -\mathbf{D}_\beta g\mathbf{e} - \frac{1}{2}\mathbf{e}'\mathbf{D}^2_\beta g\mathbf{e} + \mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}(\mathbf{l}'\mathbf{D}^2 g\mathbf{l} - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})$$
$$+\frac{1}{2}\mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z}$$
$$= -\mathbf{D}_\beta g\mathbf{e} + \mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U} - \frac{1}{2}(\mathbf{e}'\mathbf{D}^2_\beta g\mathbf{e} - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}) + \frac{1}{2}(\mathbf{l}'\mathbf{D}^2 g\mathbf{l} - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})$$
$$+\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z} \tag{4-29}$$

where

$$\mathbf{Q} = \boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g' \tag{4-30}$$

The bias is

$$E(g(\gamma\hat{\theta}) - g(\beta)) \approx -\frac{1}{2}(tr(\mathbf{D}^2_\beta g\mathbf{V}_\beta) - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}} tr(\mathbf{D}^2_\beta f_j\mathbf{V}_\beta))\sigma^2_\beta$$
$$+\frac{1}{2}(tr(\mathbf{D}^2 g Var(\mathbf{l})) - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}} tr(\mathbf{D}^2 f_j Var(\mathbf{l}))) + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i E(\mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z}) \tag{4-31}$$

Note that the form of (4-31) is similar to (4-17).

Development of the bias $E(g(\gamma\hat{\theta}) - g(\gamma\theta_*))$ proceeds from

$$g(\gamma\theta_*) = g(\gamma\bar{\theta}) + \mathbf{D}g(\theta_* - \bar{\theta}) + \frac{1}{2}(\theta_* - \bar{\theta})'\mathbf{D}^2 g(\theta_* - \bar{\theta})$$
$$= g(\gamma\bar{\theta}) + \mathbf{D}g(\gamma'\gamma)^{-1}\gamma'\mathbf{e} + \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e} \tag{4-32}$$

Then, (4-28), (4-29), and (4-32) are combined, to get

$$g(\gamma\hat{\theta}) - g(\gamma\theta_*) = g(\beta) - g(\gamma\theta_*) + g(\gamma\hat{\theta}) - g(\beta)$$
$$\approx \mathbf{D}_\beta g(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} + \frac{1}{2}\mathbf{e}'(\mathbf{D}^2_\beta g - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta g\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}$$
$$-\mathbf{D}_\beta g\mathbf{e} + \mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U} - \frac{1}{2}(\mathbf{e}'\mathbf{D}^2_\beta g\mathbf{e} - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}) + \frac{1}{2}(\mathbf{l}'\mathbf{D}^2 g\mathbf{l} - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})$$
$$+\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z} \tag{4-33}$$

Again, the bias is

$$E(g(\gamma\hat{\theta}) - g(\gamma\theta_*)) \approx \frac{1}{2} tr((\mathbf{D}_\beta^2 g - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 g\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta)\sigma_\beta^2$$

$$-\frac{1}{2}(tr(\mathbf{D}_\beta^2 g\mathbf{V}_\beta) - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}} tr(\mathbf{D}_\beta^2 f_j \mathbf{V}_\beta))\sigma_\beta^2 + \frac{1}{2}(tr(\mathbf{D}^2 g Var(\mathbf{l})) - \mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}} tr(\mathbf{D}^2 f_j Var(\mathbf{l})))$$

$$+\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i E(\mathbf{D}^2 f_i \mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z}) \qquad (4\text{-}34)$$

Equation (4-34) is analogous to (4-19). The bias in $g(\gamma\hat{\theta})$ as an estimate of $g(\gamma\theta_*)$ is the difference between the bias in $g(\gamma\hat{\theta})$ as an estimate of $g(\beta)$ and the bias in $g(\gamma\theta_*)$ as an estimate of $g(\beta)$.

## The Concept of Intrinsic Nonlinearity

Model nonlinearity can cause potentially significant bias in estimates of $\theta_*$, estimates of model functions $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta_*)$, and predictions of $g(\beta)$ and $g(\gamma\theta_*)$. If these biases were truly significant, then use of a ground-water model could be severely compromised, unless, of course, the biases could be estimated. It turns out that many of the bias terms are small if unique transformations $\phi(\theta)$ and $\alpha(\beta)$, of vectors $\theta$ and $\beta$, nearly linearize the model so that transformed second-derivative matrices of the form $\mathbf{D}_\phi^2\mathbf{f}$ and $\mathbf{D}_\alpha^2\mathbf{f}$ are small. (Note that this symbolic notation for the matrices is slightly improper in that **f** is not the same function of $\phi$ or $\alpha$ as it is of $\theta$ or $\beta$. The notation is used to keep the number of variable names at a minimum, and simply implies substitution of $\theta(\phi)$ for $\theta$ or $\beta(\alpha)$ for $\beta$ in **f** when evaluating the derivatives.) Only the existence of the transformations is needed because they never need be used. If the transformations substantially reduce, or even eliminate, nonlinearity, then certain terms in the biases can be small, even though $\theta$ and $\beta$ actually used in the model functions are not necessarily the sets that produce minimum degrees of nonlinearity. This is because, as will be shown, the terms involve the second-derivative matrices and can be invariant under transformations such as $\phi(\theta)$ and $\alpha(\beta)$.

Beale (1960, p. 57) introduced a quantitative measure of the degree to which a model can be linearized by transformation of $\theta$ to $\phi(\theta)$, which he termed the intrinsic nonlinearity. (He did not consider the $\mathbf{f}(\beta)$ model.) Bates and Watts (1980) expanded on this concept and introduced another measure, which they termed the intrinsic curvature. Seber and Wild (1989, Chapter 4) give a good discussion of the ideas, and the interrelations among the ideas, of Beale (1960), Bates and Watts (1980), and others. For the theory developed in this report curvature measures such as introduced by Bates and Watts (1980) are not nearly as useful as extensions of Beale's (1960) results. The extensions are obtained using Johansen's (1983) methods of analysis, which are a simplification of Beale's (1960) methods. In addition, in this report the term intrinsic nonlinearity is not just applied to a single measure, but instead is applied to the model as a whole. Thus, the term low (degree of) intrinsic nonlinearity is applied to a model that can nearly be linearized, and the term high (degree of) intrinsic nonlinearity applied to a model that cannot nearly be linearized.

There are two types of intrinsic nonlinearity, one for **f** as a function of $\boldsymbol{\theta}$, $\mathbf{f}(\boldsymbol{\gamma\theta})$, and one for **f** as a function of $\boldsymbol{\beta}$, $\mathbf{f}(\boldsymbol{\beta})$. The former involves second derivative matrix $\mathbf{D}^2\mathbf{f}$, and the latter involves second derivative matrix $\mathbf{D}^2_\beta\mathbf{f}$. Both of these types could be referred to as model intrinsic nonlinearity. However, the form of the terms that express the intrinsic nonlinearity for $\mathbf{f}(\boldsymbol{\beta})$ involve both $\mathbf{D}^2_\beta\mathbf{f}$ and $\mathbf{I}-\mathbf{R}$ or $\mathbf{Q}$ in such a way that, to be small, the terms must satisfy some special requirements. This type of intrinsic nonlinearity is referred to in this report as system intrinsic nonlinearity, and unless otherwise indicated, the term model intrinsic nonlinearity in this report refers only to the type for $\mathbf{f}(\boldsymbol{\gamma\theta})$. Some measures of intrinsic nonlinearity are indicated as the ideas are developed further.

In appendix C the method given in Seber and Wild (1989, p. 692-694) is used to show that terms of the form $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\,\mathbf{x}'\mathbf{D}^2 f_j\mathbf{y}$ and $\mathbf{Q}'\sum_j \omega_j^{1/2}\,\mathbf{x}'\mathbf{D}^2 f_j\mathbf{y}-\mathbf{x}'\mathbf{D}^2 g\mathbf{y}$ reflecting types of model intrinsic nonlinearity are invariant under transformation of $\boldsymbol{\theta}$. In these terms **x** and **y** are given vectors of order *p*. Extensions of Seber and Wild's (1989) methods are used to show that terms of the form of $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\,\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}$ and $\mathbf{Q}'\sum_j \omega_j^{1/2}\,\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}-\mathbf{e}'\mathbf{D}^2_\beta\, g\mathbf{e}$ reflecting types of system intrinsic nonlinearity also can be approximately invariant under transformation of $\boldsymbol{\beta}$. However, as shown in appendix C, the nonlinear component of the transformation from $\mathbf{D}^2_\beta\mathbf{f}$ to $\mathbf{D}^2_\alpha\mathbf{f}$ must behave similarly to the nonlinear component of the transformation from $\mathbf{D}^2\mathbf{f}$ to $\mathbf{D}^2_\phi\mathbf{f}$ in order to allow for approximate invariance of the terms. The invariance, and approximate invariance for the terms involving **e**, causes the terms to take on values dictated by the smallest values of the matrices $\mathbf{D}^2_\phi\mathbf{f}$, $\mathbf{D}^2_\alpha\mathbf{f}$, $\mathbf{D}^2_\phi g$, and $\mathbf{D}^2_\alpha g$, which are transformations of $\mathbf{D}^2\mathbf{f}$, $\mathbf{D}^2_\beta\mathbf{f}$, $\mathbf{D}^2 g$, and $\mathbf{D}^2_\beta g$, respectively. The same transformation of $\boldsymbol{\theta}$ or $\boldsymbol{\beta}$ must make both $\mathbf{D}^2_\phi\mathbf{f}$ and $\mathbf{D}^2_\phi g$ or $\mathbf{D}^2_\alpha\mathbf{f}$ and $\mathbf{D}^2_\alpha g$ small for terms involving both $\mathbf{D}^2\mathbf{f}$ and $\mathbf{D}^2 g$ or $\mathbf{D}^2_\beta\mathbf{f}$ and $\mathbf{D}^2_\beta g$, so these terms may often be larger than terms involving only $\mathbf{D}^2\mathbf{f}$ or $\mathbf{D}^2_\beta\mathbf{f}$. Because the former terms involve both the model function **f** and the prediction function *g*, the intrinsic nonlinearity indicated by these terms is termed in this report the combined intrinsic nonlinearity. As before, there are two types of combined intrinsic nonlinearity. The type for $\mathbf{f}(\boldsymbol{\gamma\theta})$ and $g(\boldsymbol{\gamma\theta})$ is termed the model combined intrinsic nonlinearity, and the type for $\mathbf{f}(\boldsymbol{\beta})$ and $g(\boldsymbol{\beta})$ is termed the system combined intrinsic nonlinearity.

## Effect of Intrinsic Nonlinearity on Estimates of Model Functions and Residuals

***Effect on bias.*** All of the bias terms involve one or more of the terms $\mathbf{D}f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k E(\mathbf{D}^2 f_k\mathbf{l}\omega_k^{1/2}\mathbf{Z})$, $\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k E(\mathbf{D}^2 f_k\mathbf{l}\omega_k^{1/2}\mathbf{Z})$, $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\,E(\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})$, $(\mathbf{I}-\mathbf{R})\sum_j\omega_j^{1/2}\,E(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e})$, $E(\mathbf{l}'\mathbf{D}^2 g\mathbf{l})-\mathbf{Q}'\sum_j\omega_j^{1/2}E(\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})$, or $E(\mathbf{e}'\mathbf{D}^2_\beta g\mathbf{e})-\mathbf{Q}'\sum_j\omega_j^{1/2}E(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e})$. In appendix C the first three terms are shown to be small when the model intrinsic nonlinearity is small; the fourth is shown to be small when the system intrinsic nonlinearity is small; the fifth is shown to be small when the model combined intrinsic nonlinearity is small; and the sixth is shown to be small when the system combined intrinsic nonlinearity is small. It is worthwhile to explore the effects of intrinsic nonlinearity on the bias terms.

The bias in $\hat{\boldsymbol{\theta}}$ as an estimate of $\boldsymbol{\theta}_*$ is given by (4-12) or (4-15). With $g(\gamma\boldsymbol{\theta}) = \theta_i$, $(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})_i^{-1}\sum_k E(\mathbf{D}^2 f_k \mathbf{l}\omega_k^{1/2}\mathbf{Z})$, where $(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})_i^{-1}$ is row $i$ of $(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}$, is of the form $\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k E(\mathbf{D}^2 f_k \mathbf{l}\omega_k^{1/2}\mathbf{Z})$. Therefore, if the model intrinsic nonlinearity is small, the bias becomes

$$E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*) = \frac{1}{2}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\sum_j \omega_j\, tr(\mathbf{D}_\beta^2 f_j \mathbf{V}_\beta \sigma_\beta^2 - \mathbf{D}^2 f_j Var(\mathbf{l})) \tag{4-35}$$

Note that the trace term is the difference between the contribution of nonlinearity to $E(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\bar{\boldsymbol{\theta}}))$ (which can be seen from (3-28)) and the contribution of nonlinearity to $E(\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{f}(\gamma\bar{\boldsymbol{\theta}}))$, where $\hat{\boldsymbol{\theta}}_0 = \bar{\boldsymbol{\theta}} + \mathbf{l}$. It is tempting to speculate that the similarity of the two terms could often cause their difference to be small. This possibility has not been investigated in this report.

For small model and system types of intrinsic nonlinearity, the bias in $\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$ as an estimate of $\mathbf{f}(\boldsymbol{\beta})$ is obtained from (4-17) as

$$E(f_i(\gamma\hat{\boldsymbol{\theta}}) - f_i(\boldsymbol{\beta})) \approx 0\,;\ i = 1,2,\ldots,n \tag{4-36}$$

In other words, estimates of $\mathbf{f}(\boldsymbol{\beta})$ are nearly unbiased if both types of intrinsic nonlinearity are small. However, under these same circumstances the bias in $\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$ as an estimate of $\mathbf{f}(\gamma\boldsymbol{\theta}_*)$ is given from (4-19) as

$$E(f_i(\gamma\hat{\boldsymbol{\theta}}) - f_i(\gamma\boldsymbol{\theta}_*)) \approx \frac{1}{2}tr((\mathbf{D}_\beta^2 f_i - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_i\, \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{V}_\beta)\sigma_\beta^2\,;\ i = 1,2,\ldots,n \tag{4-37}$$

which is the bias in $\mathbf{f}(\gamma\boldsymbol{\theta}_*)$ as an estimate of $\mathbf{f}(\boldsymbol{\beta})$. Because interest is generally in replicating $\mathbf{f}(\boldsymbol{\beta})$, not $\mathbf{f}(\gamma\boldsymbol{\theta}_*)$, the bias given by (4-37) would not seem to be too important.

***Effects on residuals.*** Properties of residuals change materially when both model and system types of intrinsic nonlinearity are small. In this case the residuals given by (4-21) can be written

$$Y_i - f_i(\gamma\hat{\boldsymbol{\theta}})) \approx \omega_i^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*\,;\ i = 1,2,\ldots,n \tag{4-38}$$

so that

$$E(Y_i - f_i(\gamma\hat{\boldsymbol{\theta}})) \approx 0\,;\ i = 1,2,\ldots,n \tag{4-39}$$

Even though the errors $Y_i - f_i(\gamma\boldsymbol{\theta}_*)$ may have nonzero expected values, the residuals have expected values of nearly zero. The variance also is simplified. From (4-38)

$$Var(\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})) \approx \boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}Var(\mathbf{U}_*)\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{-\frac{1}{2}}$$
$$= \boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{-\frac{1}{2}}\sigma_\varepsilon^2 \qquad (4\text{-}40)$$

Another form of the variance is useful in section 5 for examining model uncertainty. Let $\mathbf{T}_*$ be the nonlinear terms in (3-21). Then from the results of appendix C the product $(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{T}_*\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R})$ is exactly of the form that is small when system intrinsic nonlinearity is small. Therefore in this instance

$$Var(\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})) \approx \boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{V}_\varepsilon\sigma_\varepsilon^2 + E(\mathbf{f}(\boldsymbol{\beta})-\mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{f}(\boldsymbol{\beta})-\mathbf{f}(\gamma\boldsymbol{\theta}_*))')\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{-\frac{1}{2}}$$
$$= \boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}E(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{-\frac{1}{2}} \qquad (4\text{-}41)$$

Finally, because $\mathbf{U}_*$ is assumed to be normally distributed

$$\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) \sim N(\mathbf{0}, \boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{-\frac{1}{2}}\sigma_\varepsilon^2)$$
$$\approx N(\mathbf{0}, \boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}E(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{-\frac{1}{2}}) \qquad (4\text{-}42)$$

The residuals can be normally distributed even when the errors $\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*)$ are not. In addition, the residuals behave as if the model were nearly linear.

***Detection of intrinsic nonlinearity***. An analysis of the residuals from a specific modeling problem indicates the possible importance of intrinsic nonlinearity. If model and system types of intrinsic nonlinearity are both small, then from (4-23) the sum of weighted residuals is

$$\sum_i \omega_i^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})) \approx \sum_i (\mathbf{I}-\mathbf{R})_i\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* \qquad (4\text{-}43)$$

and from (4-27) the slope of the weighted residual plot is proportional to

$$\sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})\omega_i^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})) - \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})\sum_i \omega_i^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))/n$$
$$\approx \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\bar{\boldsymbol{\theta}})(\mathbf{I}-\mathbf{R})_i\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* - \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\bar{\boldsymbol{\theta}})\sum_i(\mathbf{I}-\mathbf{R})_i\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*/n \qquad (4\text{-}44)$$

That is, if both types of intrinsic nonlinearity are small, both the sum of weighted residuals and the slope of the plot of weighted residuals in relation to $\omega_i^{1/2}\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$ are the values expected for a linear model. The slope should not be significant unless an intercept is needed.

From (4-42), a sample distribution of weighted residuals that is not significantly different from normal adds evidence that both types of intrinsic nonlinearity are small and suggests in addition that $\mathbf{U}_*$ is normally distributed. If $\mathbf{U}_*$ is normal, then the deviation of the distribution of $\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma\theta}_*)$ from normality results from model nonlinearity.

A check for model intrinsic nonlinearity only is to premultiply the weighted residual vector by **R**. From (4-23) and the fact that $\mathbf{R}(\mathbf{I}-\mathbf{R})=\mathbf{0}$, the result should be a vector of nearly zero values if model intrinsic nonlinearity is small. In theory **Df** used in **R** is computed at $\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}$, which is unknown. However, note that $\mathbf{DfJ}=\mathbf{D}_\phi\mathbf{f}$, where **J** is the Jacobian defined by (C-1), appendix C, and $\mathbf{D}_\phi\mathbf{f}$ is nearly constant when model intrinsic nonlinearity is small. Hence, because **R** is invariant under transformation of parameters, it is nearly constant when $\boldsymbol{\theta}$ is varied if model intrinsic nonlinearity is small. In this case any set $\boldsymbol{\theta}$ not too remote from $\bar{\boldsymbol{\theta}}$ can be used to compute **Df**, and thus **R**, for the check except $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$. If $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$, the product of **R** and the weighted residual vector is always zero.

## Effect of Intrinsic Nonlinearity and Combined Intrinsic Nonlinearity on Predictions

The effect of intrinsic nonlinearity and combined intrinsic nonlinearity on biases in predictions is similar to the effect of intrinsic nonlinearity on biases in estimates. That is, if the model intrinsic nonlinearity is small, then (4-31) becomes

$$E(g(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})-g(\boldsymbol{\beta}))$$
$$\approx -\frac{1}{2}(tr(\mathbf{D}^2_\beta g\mathbf{V}_\beta)-\mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}tr(\mathbf{D}^2_\beta f_j\mathbf{V}_\beta))\sigma^2_\beta+\frac{1}{2}(tr(\mathbf{D}^2 gVar(\mathbf{l}))-\mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}tr(\mathbf{D}^2 f_j Var(\mathbf{l}))) \quad (4\text{-}45)$$

and, if both the model and system types of combined intrinsic nonlinearity are small, then

$$E(g(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})-g(\boldsymbol{\beta}))\approx 0 \quad (4\text{-}46)$$

Because (4-46) requires that **f** and g both be nearly linearized by the same transformations of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, (4-46) may be harder to satisfy than (4-45). For example, if $g(\boldsymbol{\beta})=\beta_i$ and $g(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})=\boldsymbol{\gamma}_i\hat{\boldsymbol{\theta}}$, where $\boldsymbol{\gamma}_i$ is row *i* of $\boldsymbol{\gamma}$, then (4-45) is equivalent to (4-35). (There is no additional bias from estimating $\beta_i$ with $\boldsymbol{\gamma}_i\boldsymbol{\theta}_*$.) However, if $g(\boldsymbol{\beta})$ were, for example, a hydraulic head, and most of the data in **Y** were hydraulic head data, then small model and system types of intrinsic nonlinearity would probably imply small model and system types of combined intrinsic nonlinearity as well. Evaluation of bias $E(g(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})-g(\boldsymbol{\gamma\theta}_*))$ adds the term $\frac{1}{2}tr((\mathbf{D}^2_\beta g-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}^2_\beta g\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{V}_\beta)\sigma^2_\beta$ to both (4-45) and (4-46). (See (4-34).) As for (4-37), interest generally is in predicting $g(\boldsymbol{\beta})$ not $g(\boldsymbol{\gamma\theta}_*)$, so the extra component of bias may not be too important.

Residuals that pertain to predictions can be derived using constrained regression as discussed in section 5 and appendix E. These residuals can be analyzed with methods analogous to those used for the standard residuals to detect both types of combined intrinsic nonlinearity. Discussion of this is deferred to section 5 after discussion of the basic concepts of the constrained regression.

## Estimation and Prediction When the Weights are Unknown

The theory developed in this report thus far is valid for any weight matrix, $\boldsymbol{\omega}$. For Gauss-Markov estimation $\boldsymbol{\omega}$ is defined using $\boldsymbol{\omega}^{-1} \propto E(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'$, which, from (3-32) is generally full (that is, not diagonal). This definition (the second-moment matrix) is used instead of $Var(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))$ (the covariance matrix) used in classical regression because of the nonzero vector $E(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))$. As shown in (B-18)-(B-22), any squared linear combination of the form $E(\boldsymbol{l}'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_*))^2$ is minimized through third-order terms by using this definition. Obenchain (1975, p. 378) considered the classical linear model with correlated errors and suggested letting $\boldsymbol{\omega}^{-1}$ be diagonal, with diagonal elements given by the variances of the errors. He cited several benefits from this definition, including the fact that the model would fit the data (with the residuals having a mixture of positive and negative signs) instead of being systematically offset from the data (with the residuals tending to have one sign) as can happen in Gauss-Markov estimation. For the theory developed in this report the variances of the errors would be replaced by $E(Y_i - f_i(\gamma\boldsymbol{\theta}_*))^2$. However, if $E(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'$ is unknown, its diagonal elements $E(Y_i - f_i(\gamma\boldsymbol{\theta}_*))^2$ might also be unknown.

Diagonal elements $E(Y_i - f_i(\gamma\boldsymbol{\theta}_*))^2$ can be estimated from field evidence for large-scale heterogeneity (Christensen and Cooley, 1999b), estimates of observation-error variances (Hill, 1998, p. 45-49), and analysis of residuals (Cooley and Naff, 1990, Chapter 5). These methods generally result in grouping the diagonal elements into $q$ groups, each of which appears to have nearly uniform values. In this case the objective function $S(\boldsymbol{\theta})$ is written

$$S(\boldsymbol{\theta}) = \sum_{k=1}^{q} \omega_{Gk} \sum_{i(k)} (Y_i - f_i(\gamma\boldsymbol{\theta}))^2 \tag{4-47}$$

where $\omega_{Gk}$ is the weight for group $k$ and $i(k)$ indicates summation over the observations in group $k$.

An analysis of residuals yields estimated weights $w_{Gk}$ that give apparently uniform average variability of weighted residuals for all groups (Cooley and Naff, 1990, p. 168-171), that is, so that

$$\frac{1}{n_1}\sum_{i(1)}(Y_i - f_i(\gamma\hat{\boldsymbol{\theta}}))^2 w_{G1} \approx \frac{1}{n_2}\sum_{i(2)}(Y_i - f_i(\gamma\hat{\boldsymbol{\theta}}))^2 w_{G2} \approx \cdots \approx \frac{1}{n_q}\sum_{i(q)}(Y_i - f_i(\gamma\hat{\boldsymbol{\theta}}))^2 w_{Gq} \tag{4-48}$$

where $n_k$ is the number of observations in group k. From (4-48) the estimated weights approximately satisfy

$$w_{Gk}^{-1} \propto \frac{1}{n_k} \sum_{i(k)} (Y_i - f_i(\gamma\hat{\theta}))^2 \tag{4-49}$$

An objective function that, when minimized with respect to $\theta$, gives normal equations having weights exactly satisfying (4-49) is

$$\ell(\theta) = \frac{1}{2} \sum_{k=1}^{q} n_k \ln\left( \sum_{i(k)} (Y_i - f_i(\gamma\theta))^2 \right) \tag{4-50}$$

Appendix D shows that, through third order in $Y_i - f_i(\gamma\theta_*)$, $\ell(\theta)$ and $S(\theta) + constants$ are proportional when $\omega_{Gk}$ is defined as

$$\omega_{Gk} = \frac{\sigma_\varepsilon^2 n_k}{\sum_{i(k)} E(Y_i - f_i(\gamma\theta_*))^2}; \; k = 1,2,\ldots,q \tag{4-51}$$

This definition is approximately equivalent to the definition proposed by Obenchain (1975) discussed earlier, the approximation resulting from grouping the errors. Appendix D also shows that the normal equations obtained from $\ell(\theta)$ and $S(\theta)$ are equivalent through second order in $Y_i - f_i(\gamma\theta_*)$ under the same conditions. Barlebo and others (1998, p. 154) used (4-49) to compute the weights in the normal equations, but they did not formally justify the procedure.

The above analysis leads to the conclusion that the results for estimation and prediction, including effects of the various types of intrinsic nonlinearity, may be approximately applied when the weights are unknown.

## Summary of Principal Results

Vector $\theta_*$ is estimated using nonlinear least squares based on the objective function given by (4-1), $S(\theta) = (\mathbf{Y} - \mathbf{f}(\gamma\theta))'\omega(\mathbf{Y} - \mathbf{f}(\gamma\theta))$. In this function $\omega$ is a positive definite weight matrix that is proportional to the inverse of $E(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'$ for generalized least squares (Gauss-Markov estimation) but is arbitrary for most developments. The inverse of the second-moment matrix $E(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'$ is shown to be the correct weight matrix to use for Gauss-Markov-type estimation instead of the inverse of the matrix $Var(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))$ used in classical regression because of the nonzero vector $E(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))$. As a basis for the theory needed to analyze the estimate $\hat{\theta}$ of $\theta_*$ and functions of the estimate, an analytical solution for the estimate that is second-order correct in $\varepsilon$, $\mathbf{e} = \beta - \gamma\bar{\theta}$, and their products is obtained using Taylor series and perturbation expansions; the solution is given by (4-4)-(4-6).

The second-order-correct approximation for $\hat{\theta}$ is used to develop approximate expressions for the biases $E(\hat{\theta} - \theta_*)$, $E(\mathbf{f}(\gamma\hat{\theta}) - \mathbf{f}(\beta))$, and $E(\mathbf{f}(\gamma\hat{\theta}) - \mathbf{f}(\gamma\theta_*))$ given by (4-15), (4-17), and (4-19), respectively. These biases can all be nonzero because of model nonlinearity with respect to $\theta$ and $\beta$. In spite of the presence of model error, the biases are zero for a linear model.

Model residuals, defined as $\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})$, are estimates of the errors $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$. An approximate expression (4-22) shows that the estimates can be biased as estimates of $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ because of model nonlinearity. Measures used to gage the quality of fit of a specific model to field data such as the sum of weighted residuals $\Sigma\omega_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ (where $\omega_i^{1/2}$ is row $i$ of $\omega^{1/2}$) and the plot of weighted residuals $\omega_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ in relation to weighted model function values $\omega_i^{1/2}\mathbf{f}(\gamma\hat{\theta})$ also can be affected by model nonlinearity. For a linear model both the sum of weighted residuals and the slope of the plot should not be significant if the model is correct. However, for a general model both the sum and the slope can be significant because of model nonlinearity, which can be seen from (4-23) and (4-27), respectively.

A prediction to be made with the model is defined as any function of $\gamma\hat{\theta}$ of interest that is not contained in $\mathbf{f}(\gamma\hat{\theta})$. It is termed $g(\gamma\hat{\theta})$ and is used to predict $g(\beta)$ or $g(\gamma\theta_*)$. Note that function $g(\gamma\hat{\theta})$ can be the same type of function as any element $f_i(\gamma\hat{\theta})$ (for example hydraulic head or flux), but it was not observed as $Y_i$. Biases $E(g(\gamma\hat{\theta})-g(\beta))$ and $E(g(\gamma\hat{\theta})-g(\gamma\theta_*))$ are given by (4-31) and (4-34), respectively. As for the model function biases, the prediction biases can be nonzero because of model nonlinearity.

If the biases in $\mathbf{f}(\gamma\hat{\theta})$, $\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})$, and $g(\gamma\hat{\theta})$ were large, they could severely compromise use of a ground-water model, unless they could be adequately estimated. However, many of the bias terms can be small if unique transformations $\phi(\theta)$ and $\alpha(\beta)$ nearly linearize the models $\mathbf{f}(\gamma\theta)$ and $\mathbf{f}(\beta)$. The transformations do not have to be known. If such a transformation for $\theta$ exists, then the model intrinsic nonlinearity is said to be small. If such a transformation for $\beta$ exists, and if in addition certain approximations applied to second derivatives of $\mathbf{f}(\beta)$ and explained in appendix C are accurate, then the system intrinsic nonlinearity is said to be small. Other bias terms pertaining to the predictions are small if the transformations nearly linearize $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$, and $\mathbf{f}(\beta)$ and $g(\beta)$, simultaneously. If these transformations exist, then model and system types of combined intrinsic nonlinearity are both said to be small, assuming for the latter accurate approximations for the second derivatives of $\mathbf{f}(\beta)$ and $g(\beta)$.

If the model intrinsic nonlinearity is small, then as shown by (4-35) some of the bias in $\hat{\theta}$ is eliminated, and as shown by (4-36), if model and system types of intrinsic nonlinearity are small, $\mathbf{f}(\gamma\hat{\theta})$ is nearly unbiased as an estimate of $\mathbf{f}(\beta)$. The generally nonzero component $E(\mathbf{f}(\gamma\theta_*)-\mathbf{f}(\beta))$ remains as bias in the estimate of $\mathbf{f}(\gamma\theta_*)$, as shown in (4-37). However, a ground-water study generally is concerned with estimating $\mathbf{f}(\beta)$, not $\mathbf{f}(\gamma\theta_*)$, so this bias may not be important. Properties of the residuals also change when intrinsic nonlinearity is small. That is, when model and system types of intrinsic nonlinearity are small, the residuals become nearly unbiased ($E(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))\approx\mathbf{0}$ as given by (4-39)), their covariance matrix is nearly like the covariance matrix for a linear model (as shown in (4-40) and (4-41)), and, if $\mathbf{e}$ and $\varepsilon$ are normally distributed, they can be normally distributed as given by (4-42).

The presence of significant model and system types of intrinsic nonlinearity can be tested for by examining the slope of the plot of weighted residuals in relation to weighted estimated function values and the product of the $\mathbf{R}$ matrix (defined by (4-8)) with the vector of weighted residuals. Although $\mathbf{R}$ is defined using quantities computed at $\theta = \bar{\theta}$, any $\theta$ not too remote from $\bar{\theta}$ except $\hat{\theta}$ may be used to compute $\mathbf{R}$ for the test. From (4-23), a nearly zero vector of the product of $\mathbf{R}$ and the vector of weighted residuals indicates that the model intrinsic nonlinearity is small, and from (4-44) the absence of a significant slope suggests that both model and system types of intrinsic nonlinearity are small. A sample distribution of weighted residuals that is nearly normal adds evidence that both types of intrinsic nonlinearity are small. Similar tests for combined intrinsic nonlinearity are developed in section 5.

The second-moment matrix necessary for Gauss-Markov estimation often would be unknown. Obenchain (1975) suggested using the diagonal elements of this matrix and indicated several benefits from this definition, including a better fit of the model to the data than often results from Gauss-Markov estimation. Although it is likely that the diagonal elements also would be unknown, they might be estimated from field evidence for large-scale heterogeneity, estimates of observation-error variances, and analysis of residuals. These methods generally result in grouping the diagonal elements into $q$ groups, each of which appears to have nearly uniform values. Objective function $S(\theta)$ is then written in the form of (4-47) to incorporate these groups. A formulation that automatically weights residuals in each of the groups according to the apparent variance for the group results from minimizing the objective function (4-50), $\ell(\theta) = \frac{1}{2}\sum_{k=1}^{q} n_k \ln(\sum_{i(k)} (Y_i - f_i(\gamma\theta))^2)$, where $n_k$ is the number of observations in group $k$. Through second order in $\mathbf{Y} - \mathbf{f}(\gamma\theta_*)$, the normal equations resulting from using $\ell(\theta)$ and $S(\theta)$ are shown to be the same. Hence, the bias and other analyses developed for the case when the weight matrix is known may be used as an approximation when the weight matrix is unknown.

# 5. Uncertainty Analyses

## Confidence Regions, Confidence Intervals, and Prediction Intervals

Section 4 dealt with estimation of $\theta_*$ and functions $\mathbf{f}(\gamma\theta_*)$, $\mathbf{f}(\beta)$, $g(\gamma\theta_*)$, and $g(\beta)$. Principal concerns were with possible biases in the estimates $\hat{\theta}$, $\mathbf{f}(\gamma\hat{\theta})$, and $g(\gamma\hat{\theta})$. Expressions were derived to indicate how accurate the estimates are, on the average. These expressions do not indicate either the precision of the estimates, or how close specific estimates might be to the values of interest (values of $\theta_*$, $f_i(\gamma\theta_*)$, $f_i(\beta)$, $g(\gamma\theta_*)$, or $g(\beta)$). Uncertainty in prediction of specific future observations, which contain additional future measurement error, also was not addressed by the average-accuracy expressions. However, estimates of uncertainty are necessary in order to express the confidence that an investigator has in the results (flow system analysis and predictions) of a model study. In this report uncertainty in estimates is expressed through confidence regions and confidence intervals. Uncertainty in predictions of future observations is expressed through prediction intervals.

A joint confidence region for all parameters (simply referred to in this report as a confidence region) is defined as a usually closed but possibly open region that has a specified probability $1-\alpha$ of containing the true (as opposed to estimated) parameter set $\theta_*$. It is a random region that always encloses $\hat{\theta}$. An interpretation is that if many realizations of $\mathbf{e}$ and $\varepsilon$ were used for an equal number of regressions to find values of $\hat{\theta}$, then the fraction of associated confidence regions containing $\theta_*$ would be approximately $1-\alpha$. A major difference between the confidence region defined here and a confidence region defined for a classical linear or nonlinear model is that the true parameter set for a classical model is considered to be fixed; whereas, the true parameter set $\theta_*$ used here is stochastic. The theories pertaining to the two types of regions turn out to be analogous, however. Cooley and Naff (1990, p. 172-175) give a thorough discussion of confidence regions, including a generalization for parameter subsets, for a linear approximation of the classical nonlinear model. Graybill (1976, p. 183-192) gives a complete theoretical foundation for confidence regions for a classical linear model, and Seber and Wild (1989, Chapter 5) discuss approximate confidence regions for the classical nonlinear model.

A type of confidence interval derived directly from a confidence region is called a Scheffé interval (Graybill, 1976, p. 199; Seber and Wild, 1989, p. 194). For the confidence region used in this report it is an interval for some function of parameters $\theta_*$, say $g(\gamma\theta_*)$, computed as the maximum and minimum values of $g(\gamma\theta)$ taken over all parameter sets $\theta$ lying within the confidence region. It is a simultaneous interval so that $g(\gamma\theta_*)$ lies within its Scheffé interval with probability $1-\alpha$ while all other linearizable functions of $\theta_*$ lie within their Scheffé intervals with the same probability. That is, the probability is only $\alpha$ that any one of the virtually infinite number of possible Scheffé intervals will not contain its respective function of $\theta_*$. The term "linearizable" was used to eliminate pathologic functions of $\theta$ that cannot be linearized using Taylor series, as is required for the analysis used in this report. Scheffé intervals for nonlinear models are discussed within a classical context by Vecchia and Cooley

(1987), and are further discussed and applied to a field problem by Christensen and Cooley (1999a). Approximate Scheffé intervals are briefly discussed for the classical nonlinear model by Seber and Wild (1989, p. 194), and a thorough discussion of the theory as applied to the classical linear model is given by Graybill (1976, p. 195-200).

An individual confidence interval for $g(\gamma\theta_*)$ is defined as a usually closed but possibly open interval around $g(\gamma\hat{\theta})$ that contains $g(\gamma\theta_*)$ with specified probability $1-\alpha$. As for a confidence region and Scheffé interval, an individual confidence interval is random, but it is not simultaneous. It applies only to the selected function so that, in contrast to the Scheffé interval, a fraction $\alpha$ of all individual confidence intervals for linearizable functions of $\theta_*$ will not contain their respective functions of $\theta_*$. Also, a confidence interval as defined in this report differs from the classical one in that the parameter set $\theta_*$ is stochastic in the present case instead of being fixed as in the classical case. As noted by Cooley (2000, p. 1161) this gives the confidence interval as defined here some of the properties of a prediction interval. An individual confidence interval could also be obtained for $g(\beta)$, which would in general be larger than an individual confidence interval for $g(\gamma\theta_*)$ because it receives an extra component of variance from the difference $g(\beta)-g(\gamma\theta_*)$. This interval has more in common with prediction intervals than with confidence intervals, so is included with them. Individual confidence intervals for classical nonlinear models are discussed by Seber and Wild (1989, Chapter 5), and a thorough discussion of their theoretical basis as one type in a family of joint intervals for a classical linear model is given by Graybill (1976, p. 201-204).

Finally, an individual prediction interval for some future observation $Y_p$ of $g(\beta)$ is defined as a usually closed but possibly open interval around $g(\gamma\hat{\theta})$ that contains $Y_p$ with specified probability $1-\alpha$. Because $Y_p$ and $g(\beta)$ are both stochastic, the only difference between a prediction interval and a confidence interval for $g(\beta)$ is the extra component of variance from the difference $Y_p - g(\beta)$, which is the variance from a future measurement error. A classical prediction interval also is defined for the stochastic variable $Y_p$, so it does not differ in definition from the one defined in this report. Approximate prediction intervals for a classical linearized model are discussed in Seber and Wild (1989, p. 193-194), and prediction intervals for a classical linear model are derived and discussed by Graybill (1976, p. 267-270). Christensen and Cooley (1999b) discuss and apply prediction intervals to two field sites using a nonlinear model in a classical context where model error was assumed to add variance analogous in form to measurement error variance. They showed that both new and old data are contained in their prediction intervals with apparently correct or slightly conservative probability.

In the following sections and associated appendices, confidence regions, confidence intervals, and prediction intervals are derived using the combination Taylor series and perturbation methods used for estimation. These methods were used by Johansen (1983) and Hamilton and Wiens (1987) for similar analyses applied to the classical nonlinear model. In the last section of appendix F many of the results are extended to apply approximately when the small-variance conditions for the perturbation analysis are violated.

## Development of Confidence Regions and Scheffé Intervals

***Statistical distribution and confidence region when the weight matrix is known.*** For the formal derivation of the statistical distribution necessary to define a confidence region, it is assumed that $Var(\mathbf{U})$, $Var(\mathbf{U}_*)$, and $Var(\mathbf{D}_\beta\,\mathbf{fe})$ are all small, and that $Var(\mathbf{D}_\beta\,\mathbf{fe})$ is much smaller than $Var(\mathbf{U}_*)$. These assumptions were used to derive distribution (F-57) in appendix F, which is repeated here as

$$\frac{(S(\boldsymbol{\theta}_*)-S(\hat{\boldsymbol{\theta}}))/p}{S(\hat{\boldsymbol{\theta}})/(n-p)} \sim c_r F(p,n-p) \tag{5-1}$$

where $F(p,n-p)$ signifies an $F$ random variable with $p$ and $n-p$ degrees of freedom and $c_r$ is a correction factor defined by (5-3). Equation (5-1) implies the $(1-\alpha)\times 100$ percent confidence region

$$S(\boldsymbol{\theta}_*)-S(\hat{\boldsymbol{\theta}}) \le \frac{p}{n-p} S(\hat{\boldsymbol{\theta}}) c_r F_\alpha(p,n-p) \tag{5-2}$$

where $F_\alpha(p,n-p)$ is the upper $\alpha$ point of the $F(p,n-p)$ random variable. That is, $(1-\alpha)\times 100$ percent of the continuum of possible $F(p,n-p)$ values are less than $F_\alpha(p,n-p)$, which yields the inequality in (5-2). As discussed earlier, (5-2) defines a region that has a $(1-\alpha)\times 100$ percent chance of containing $S(\boldsymbol{\theta}_*)$. If equality is used, then $S(\boldsymbol{\theta}_*)$ is replaced with a bounding surface of $S(\boldsymbol{\theta})$ values that forms the limit of the confidence region.

***Correction factor.*** The correction factor is defined as

$$c_r = \frac{\sigma_\varepsilon^2 + (\gamma_w \sigma_\beta^2 + \gamma_I \sigma_\varepsilon^4)/p}{\sigma_\varepsilon^2 + (\hat{\gamma}_w \sigma_\beta^2 + \hat{\gamma}_I \sigma_\varepsilon^4)/(n-p)} \tag{5-3}$$

where, from the definitions following (F-59),

$$\hat{\gamma}_w \sigma_\beta^2 = (tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}) - n + p)\sigma_\varepsilon^2 \tag{5-4}$$

and

$$\hat{\gamma}_I \sigma_\varepsilon^4 = E(S(\hat{\boldsymbol{\theta}})) - tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2 \tag{5-5}$$

and, from the definitions following (F-63),

$$\gamma_w \sigma_\beta^2 = (tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}) - p)\sigma_\varepsilon^2 \tag{5-6}$$

and

$$\gamma_I \sigma_\varepsilon^4 = E(S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}})) - tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2 \tag{5-7}$$

More detailed definitions of the terms (termed component correction factors) $\hat{\gamma}_w \sigma_\beta^2$, $\hat{\gamma}_I \sigma_\varepsilon^4$, $\gamma_w \sigma_\beta^2$, and $\gamma_I \sigma_\varepsilon^4$ from the perturbation analysis are given by (F-60)-(F-62), (F-64), and (F-65). Factors $\hat{\gamma}_w \sigma_\beta^2$ and $\gamma_w \sigma_\beta^2$ correct for the possibility that $\boldsymbol{\omega}^{-1} \neq \mathbf{V}_*$, and factors $\hat{\gamma}_I \sigma_\varepsilon^4$ and $\gamma_I \sigma_\varepsilon^4$ correct for model intrinsic nonlinearity. The latter two factors are zero when there is neither model nor system types of intrinsic nonlinearity. Finally, note that $\boldsymbol{\omega}^{-1}$ cannot be entirely arbitrary; it should be composed of $\mathbf{V}_\varepsilon$ plus a matrix $\mathbf{V}_\omega \sigma_\beta^2 / \sigma_\varepsilon^2$, where $\mathbf{V}_\omega$ is dependent only on model error ((F-61), appendix F).

In the last section of appendix F the mean and variance of an expression that is analogous to, but more general than, (5-1) are analyzed using a method that does not rely on Taylor series expansions or perturbations. The analysis shows that the small-variance assumptions can be relaxed if the correction factors are written in terms of $\boldsymbol{\Omega}$ instead of $\mathbf{V}_*$, where

$$\boldsymbol{\Omega} = E(\mathbf{Y} - \mathbf{f}(\boldsymbol{\gamma\theta}_*))(\mathbf{Y} - \mathbf{f}(\boldsymbol{\gamma\theta}_*))' / \sigma_\varepsilon^2 \tag{5-8}$$

Distribution (5-1) written using the redefined correction factors, (F-133)-(F-135), is most accurate if $(V_{\varepsilon ii})^{1/2}\sigma_\varepsilon$ is twice or more $(E(f_i(\boldsymbol{\beta}) - f_i(\boldsymbol{\gamma\theta}_*))^2)^{1/2}$, $i = 1,2,\ldots,n$. The redefinitions give a more accurate indication of errors resulting from the form of $\boldsymbol{\omega}$ than the original factors because $\boldsymbol{\Omega}$ is the correct form for $\boldsymbol{\omega}^{-1}$ for Gauss-Markov estimation, not $\mathbf{V}_*$. However, as indicated by (G-7) and associated text in appendix G, the difference between $\boldsymbol{\Omega}$ and $\mathbf{V}_*$ is of order larger than the order of terms dropped for the perturbation analysis, so that any differences resulting from the two different variances may often be small. Hence, in this section $\boldsymbol{\Omega}$ and $\mathbf{V}_*$ are often used interchangeably. Finally, note from the forms of (5-3)-(5-7) that the value of $c_r$ is unchanged by the redefinition.

A useful form for the correction factor is given by (5-18) and can be developed as follows. When model intrinsic nonlinearity is small the expected value of $S(\hat{\boldsymbol{\theta}})$ can be obtained from (F-99) as

$$E(S(\hat{\boldsymbol{\theta}})) \approx tr((\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2 \tag{5-9}$$

Now let

$$tr((\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2 = b(n - ap)\sigma_\varepsilon^2 \tag{5-10}$$

where

$$b = tr(\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{\omega}^{\frac{1}{2}}) / n \tag{5-11}$$

Then to evaluate $a$, (5-10) is used with (5-11) to yield

$$bap = bn - tr(\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}}) + tr(\mathbf{R}\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}})$$
$$= tr(\mathbf{R}\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}}) \tag{5-12}$$

from which

$$a = tr(\mathbf{R}(\omega / b)^{\frac{1}{2}}\Omega(\omega / b)^{\frac{1}{2}})/p \tag{5-13}$$

Also

$$E(S(\theta_*)) = tr(\omega\Omega)\sigma_\varepsilon^2 = tr(\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}})\sigma_\varepsilon^2 = bn\sigma_\varepsilon^2 \tag{5-14}$$

so that, for small model intrinsic nonlinearity,

$$E(S(\theta_*) - S(\hat{\theta})) \approx bn\sigma_\varepsilon^2 - b(n-ap)\sigma_\varepsilon^2 = bap\sigma_\varepsilon^2 \tag{5-15}$$

From (5-4) and (5-10) using redefined correction factors

$$(n-p)\sigma_\varepsilon^2 + \hat{\gamma}_w\sigma_\beta^2 + \hat{\gamma}_I\sigma_\varepsilon^4 = b(n-ap)\sigma_\varepsilon^2 + \hat{\gamma}_I\sigma_\varepsilon^4 \tag{5-16}$$

Similarly, from (5-6) and (5-12)

$$p\sigma_\varepsilon^2 + \gamma_w\sigma_\beta^2 + \gamma_I\sigma_\varepsilon^4 = bap\sigma_\varepsilon^2 + \gamma_I\sigma_\varepsilon^4 \tag{5-17}$$

so that the correction factor can be written as

$$c_r = \frac{\sigma_\varepsilon^2 + (\gamma_w\sigma_\beta^2 + \gamma_I\sigma_\varepsilon^4)/p}{\sigma_\varepsilon^2 + (\hat{\gamma}_w\sigma_\beta^2 + \hat{\gamma}_I\sigma_\varepsilon^4)/(n-p)} = \frac{(n-p)(ap + \gamma_I\sigma_\varepsilon^2/b)}{p(n - ap + \hat{\gamma}_I\sigma_\varepsilon^2/b)} \tag{5-18}$$

If the model intrinsic nonlinearity is negligible, then (5-18) becomes

$$c_r \approx \frac{\sigma_\varepsilon^2 + \gamma_w\sigma_\beta^2/p}{\sigma_\varepsilon^2 + \hat{\gamma}_w\sigma_\beta^2/(n-p)} = \frac{(n-p)a}{n-ap} \tag{5-19}$$

***Relation between the correction factor and spatial correlation.*** An approximation given by (5-20) can be used to illustrate the relation between the correction factor and spatial correlation. The approximation is based on replacing arbitrary positive definite matrix $b\omega^{-1}$ with $b\hat{\omega}^{-1}$, defined as a diagonal matrix composed of the diagonal elements of $\mathbf{\Omega}$. Then $(\hat{\omega}/b)^{1/2}\Omega(\hat{\omega}/b)^{1/2}$ is similar to a correlation matrix. This matrix is approximated with the form

$$(\hat{\omega}/b)^{\frac{1}{2}}\Omega(\hat{\omega}/b)^{\frac{1}{2}} \approx \begin{bmatrix} 1 & c & c & \cdots & c \\ c & 1 & c & \cdots & c \\ & & \cdots & & \\ c & c & c & \cdots & 1 \end{bmatrix} = (1-c)\mathbf{I} + c\mathbf{1} \tag{5-20}$$

where $0 \le c \le 1$ so that significant spatial correlation from model error is assumed to be positive, and $\mathbf{1}$ is a matrix of ones. Equation (5-20) has the correct limits when $\Omega$ is diagonal and when all correlations implied in $\Omega$ are unity. If $\sum_i (\mathbf{I}-\mathbf{R})_i$ is small (as is often the case and is easily checked for any particular model), then the sum of any row or column of $\mathbf{R}$ is approximately unity. This assumption yields the two useful results

$$\hat{\mathbf{R}}(\hat{\omega}/b)^{\frac{1}{2}}\Omega(\hat{\omega}/b)^{\frac{1}{2}} \approx (1-c)\hat{\mathbf{R}} + c\mathbf{1} \tag{5-21}$$

and

$$(\mathbf{I}-\hat{\mathbf{R}})(\hat{\omega}/b)^{\frac{1}{2}}\Omega(\hat{\omega}/b)^{\frac{1}{2}} \approx (1-c)(\mathbf{I}-\hat{\mathbf{R}}) \tag{5-22}$$

in which

$$\hat{\mathbf{R}} = \hat{\omega}^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\hat{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\hat{\omega}^{\frac{1}{2}} \tag{5-23}$$

Use of the approximation to investigate spatial correlation effects is based on replacement of $\omega$ with $\hat{\omega}$ and use of (5-21) so that (5-13) becomes

$$a = tr(\hat{\mathbf{R}}(\hat{\omega}/b)^{\frac{1}{2}}\Omega(\hat{\omega}/b)^{\frac{1}{2}})/p \approx 1 - c + cn/p \tag{5-24}$$

which gives

$$n - ap \approx (1-c)(n-p) \tag{5-25}$$

and

$$ap \approx (1-c)p + cn \tag{5-26}$$

Then expression of (5-18) and (5-19) in terms of $c$ gives

$$c_r = \frac{\sigma_\varepsilon^2 + (\gamma_w\sigma_\beta^2 + \gamma_I\sigma_\varepsilon^4)/p}{\sigma_\varepsilon^2 + (\hat{\gamma}_w\sigma_\beta^2 + \hat{\gamma}_I\sigma_\varepsilon^4)/(n-p)} \approx \frac{(n-p)((1-c)p + cn + \gamma_I\sigma_\varepsilon^2/b)}{p((1-c)(n-p) + \hat{\gamma}_I\sigma_\varepsilon^2/b)} \tag{5-27}$$

and

$$c_r \approx \frac{\sigma_\varepsilon^2 + \gamma_w \sigma_\beta^2 / p}{\sigma_\varepsilon^2 + \hat{\gamma}_w \sigma_\beta^2 / (n-p)} \approx \frac{(1-c)p + cn}{(1-c)p} \tag{5-28}$$

Equations (5-27) and (5-28) show the dependence of the ratio $((S(\theta_*) - S(\hat{\theta}))/p) / (S(\hat{\theta})/(n-p))$ on the spatial correlation from model error when a diagonal weight matrix is used. When positive correlation is large as indicated by a large value of $c$, the ratio becomes much larger than given by $F(p, n-p)$. The large spatial correlation causes $S(\hat{\theta})$ to be too small compared to $S(\theta_*)$ as indicated by a large value of $a$. (See (5-24).) This type of behavior is verified by synthetic examples, some of which are given in section 7. In the last section of appendix F it is argued that use of $\hat{\omega}$ should cause the ratio to be more nearly $c_r F(p, n-p)$ distributed than use of arbitrary choices for $\omega$. If $\Omega$ and $\hat{\omega}$ are unknown, then $c_r$ cannot be exactly evaluated for practical work, and a different technique given in the final part of this section can be used.

***Approximate evaluation of the correction factor.*** The component correction factor $\hat{\gamma}_I \sigma_\varepsilon^4$ is given by (G-1), appendix G, and (G-1) is evaluated in (G-2)-(G-7). Examination of (G-1)-(G-7) shows that $\hat{\gamma}_I \sigma_\varepsilon^4$ can be written in terms of $(\omega / b)^{1/2} \Omega (\omega / b)^{1/2}$ and $b^2 \sigma_\varepsilon^4$. As for $c_r$, $\hat{\gamma}_I$ cannot be evaluated exactly unless $\Omega$ and $\omega$ are known. In the simplest case where $b\omega^{-1}$ is set equal to $\Omega$ (for example, when spatial correlation is small enough to be ignored so that $\Omega \approx b\hat{\omega}^{-1}$ or when using Gauss-Markov estimation), $(\omega / b)^{1/2} \Omega (\omega / b)^{1/2}$ can be set equal to **I**, so that

$$\hat{\gamma}_I \sigma_\varepsilon^4 \approx \frac{1}{4} \sum_i (tr^2(\mathbf{C}_i) - 2tr(\mathbf{C}_i^2)) b^2 \sigma_\varepsilon^4 \tag{5-29}$$

where

$$\mathbf{C}_i = (\mathbf{I} - \mathbf{R})_i \sum_j \omega_j^{\frac{1}{2}} \omega^{\frac{1}{2}} \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1} \mathbf{D}^2 f_j (\mathbf{Df}'\omega\mathbf{Df})^{-1} \mathbf{Df}'\omega^{\frac{1}{2}} \tag{5-30}$$

In all cases, from (F-65), $\gamma_I \sigma_\varepsilon^4$ equals $-\hat{\gamma}_I \sigma_\varepsilon^4$. In section 6, $\hat{\gamma}_I \sigma_\varepsilon^4$ is derived and investigated using general concepts of intrinsic nonlinearity. Finally, examination of (F-57) using (F-59) and (F-63) or (F-66) shows that substitution of $\omega$ for $\omega / b$ does not change the confidence region or any of the confidence intervals.

***An estimate of*** $b\sigma_\varepsilon^2$. Substitution of (5-29) into (5-27) gives a correction for model intrinsic nonlinearity in terms of $b\sigma_\varepsilon^2$, which is unknown and so must be estimated. If model intrinsic nonlinearity is small, $b\sigma_\varepsilon^2$ may be written using the combination of (5-9) and (5-10) as

$$b\sigma_\varepsilon^2 \approx \frac{E(S(\hat{\theta}))}{n - ap} \tag{5-31}$$

which has as an estimate

$$s^2 = \frac{S(\hat{\theta})}{n - ap} \tag{5-32}$$

Model intrinsic nonlinearity is rarely large (Seber and Wild, 1989, p. 136), so $\hat{\gamma}_I \sigma_\varepsilon^4$ will often not be needed.

***Computation of a Scheffé interval when the weight matrix is known.*** As discussed earlier in this section, a Scheffé interval for $g(\gamma\theta_*)$ is found from the maximum and minimum values of $g(\gamma\theta)$ over the confidence region. Vecchia and Cooley (1987, p. 1240-1241) argued that, if there are no maxima or minima of $g(\gamma\theta)$ within the confidence region that are more extreme than those on the boundary, the interval can be obtained by finding extreme values of $g(\gamma\theta)$ on the boundary of the confidence region. Christensen and Cooley (1999a, p. 816) gave a graphical proof that the interval can be obtained using the method of Lagrange multipliers by finding extreme values of

$$L(\theta,\lambda) = g(\gamma\theta) + \lambda\left(\frac{p}{n-p} S(\hat{\theta}) c_r F_\alpha(p, n-p) - S(\theta) + S(\hat{\theta})\right) \tag{5-33}$$

where $\lambda$ is the Lagrange multiplier. Cooley (1999, p. 118) argued that the assumption about existence of alternative maxima and minima is almost nonrestrictive, and Christensen and Cooley (1999a, p. 812) showed that the assumption held for a field case. Numerical methodology for solution of (5-33) is given in Vecchia and Cooley (1987).

***Confidence region and Scheffé interval when the weight matrix is unknown.*** When the weight matrix is unknown, $\ell(\theta)$ given by (4-50) should be the objective function instead of $S(\theta)$. The analysis developed for the $\hat{\omega}$ -known case can be used as an approximation when the weight matrix is unknown. This is shown as follows. For convenience let

$$\boldsymbol{\eta}_* = \mathbf{Y} - \mathbf{f}(\gamma\theta_*) \tag{5-34}$$

and

$$\hat{\boldsymbol{\eta}} = \mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}) \tag{5-35}$$

Then, use of (D-9), appendix D, yields, through third order in $\boldsymbol{\eta}_*$ and $\hat{\boldsymbol{\eta}}$,

$$\ell(\theta_*) - \ell(\hat{\theta}) \approx \frac{1}{\sigma_\varepsilon^2}(S(\theta_*) - S(\hat{\theta})) \tag{5-36}$$

where $S(\theta)$ is defined using (4-47). From (F-55), appendix F, the right-hand side is approximately proportional to a chi squared random variable. However, to help compensate for the approximate nature of (5-36), $\sigma_\varepsilon^2$ is approximated with $S(\hat{\theta})/(b(n-ap))$ so that an $F$ random variable may be used. This idea is developed as follows. Modification of (5-36) by using the approximation and the idea that $b \approx 1$ if the error groups adequately reflect the error structure results in:

$$\ell(\theta_*) - \ell(\hat{\theta}) \approx (n - ap)\frac{S(\theta_*) - S(\hat{\theta})}{S(\hat{\theta})} \tag{5-37}$$

Then (5-1) gives

$$\frac{(S(\theta_*) - S(\hat{\theta}))/p}{S(\hat{\theta})/(n-p)} \approx \frac{n-p}{p(n-ap)}(\ell(\theta_*) - \ell(\hat{\theta})) \sim c_r F(p, n-p)$$

or

$$\ell(\theta_*) - \ell(\hat{\theta}) \sim \frac{p(n-ap)}{n-p} c_r F(p, n-p) \tag{5-38}$$

If model intrinsic nonlinearity is small so that (5-19) is approximately valid, then

$$\ell(\theta_*) - \ell(\hat{\theta}) \sim apF(p, n-p) \tag{5-39}$$

so that an approximate confidence region is given by

$$\ell(\theta_*) - \ell(\hat{\theta}) \le apF_\alpha(p, n-p) \tag{5-40}$$

Because $n - ap = 0$ if $a = n/p$, $a$ is bounded above by $n/p$. In this case a very conservative confidence region is given by

$$\ell(\theta_*) - \ell(\hat{\theta}) \le nF_\alpha(p, n-p) \tag{5-41}$$

A Scheffé interval based on (5-40) is computed by finding extreme values of

$$\mathrm{L}(\theta, \lambda) = \mathrm{g}(\gamma\theta) + \lambda(apF_\alpha(p, n-p) - \ell(\theta) + \ell(\hat{\theta})) \tag{5-42}$$

Numerical methodology for computing a Scheffé interval from (5-42) is given in Cooley (1999).

## Development of Individual Confidence Intervals

***Statistical distribution and confidence interval when the weight matrix is known.*** Consider the distribution of $(S(\tilde{\theta}) - S(\hat{\theta}))/(S(\hat{\theta})/(n-p))$, where $\tilde{\theta}$ is a regression estimate that is constrained so that $g(\gamma\tilde{\theta}) = g(\gamma\theta_*)$. This regression estimate is derived in appendix E using the same methods as used to derive the unconstrained regression estimate $\hat{\boldsymbol{\theta}}$. In appendix F, $\tilde{\boldsymbol{\theta}}$ is used to derive the stated distribution using the same methods and assumptions as used for (5-1). The result is (F-57), which is repeated here in the form

$$\frac{S(\tilde{\boldsymbol{\theta}})-S(\hat{\boldsymbol{\theta}})}{S(\hat{\boldsymbol{\theta}})/(n-p)} \sim c_c F(1,n-p) \tag{5-43}$$

where $c_c$ is a correction factor defined by (5-45). From (5-43) note that, by analogy with (5-2),

$$S(\tilde{\boldsymbol{\theta}})-S(\hat{\boldsymbol{\theta}}) \le \frac{S(\hat{\boldsymbol{\theta}})}{n-p} c_c F_\alpha(1,n-p) \tag{5-44}$$

The maximum and minimum values of $g(\boldsymbol{\gamma\theta})$ over region (5-44) (termed a likelihood region) are the maximum and minimum values of $g(\boldsymbol{\gamma\theta})$ that could equal $g(\boldsymbol{\gamma\theta}_*)$ at probability level $\alpha$. That these limits form a $(1-\alpha)\times 100$ percent individual confidence interval is shown graphically by Christensen and Cooley (1999b, p. 2637-2638). For further developments in this section note that $F(1,n-p)=t^2(n-p)$, where $t(n-p)$ is the Student $t$ random variable with $n-p$ degrees of freedom.

***Correction factor***. The correction factor is defined analogously to $c_r$ as

$$c_c = \frac{\sigma_\varepsilon^2+\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4}{\sigma_\varepsilon^2+(\hat{\gamma}_w\sigma_\beta^2+\hat{\gamma}_I\sigma_\varepsilon^4)/(n-p)} \tag{5-45}$$

where $\hat{\gamma}_w\sigma_\beta^2$ and $\hat{\gamma}_I\sigma_\varepsilon^4$ are defined by (5-4) and (5-5), respectively. From the definitions following (F-66), $\gamma_w\sigma_\beta^2$ and $\gamma_I\sigma_\varepsilon^4$ are defined for (5-45) as

$$\gamma_w\sigma_\beta^2 = \left(\frac{1}{\mathbf{Q'Q}}\mathbf{Q'}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Q}-1\right)\sigma_\varepsilon^2 \tag{5-46}$$

and

$$\gamma_I\sigma_\varepsilon^4 = E(S(\tilde{\boldsymbol{\theta}})-S(\hat{\boldsymbol{\theta}}))-\frac{1}{\mathbf{Q'Q}}\mathbf{Q'}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Q}\sigma_\varepsilon^2 \tag{5-47}$$

More detailed definitions based on the perturbation analysis are given by (F-67) and (F-68). Factor $\gamma_w\sigma_\beta^2$ corrects for the possibility that $\boldsymbol{\omega}^{-1}\ne\mathbf{V}_*$, and factor $\gamma_I\sigma_\varepsilon^4$ corrects for model intrinsic nonlinearity and model combined intrinsic nonlinearity.

As before, an expression analogous to, but more general than, (5-43) is analyzed in the last section of appendix F. The same conclusions as reached before, regarding use of $\boldsymbol{\Omega}$ instead of $\mathbf{V}_*$ to improve the accuracy of the correction factors, are again reached.

***Approximate evaluation of the correction factor.*** To evaluate $c_c$ in terms of $\boldsymbol{\Omega}$ and $\boldsymbol{\omega}$, first note from (5-46) that

$$\sigma_\varepsilon^2 + \gamma_w \sigma_\beta^2 + \gamma_I \sigma_\varepsilon^4 = \frac{1}{\mathbf{Q}'\mathbf{Q}} \mathbf{Q}' \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{Q} \sigma_\varepsilon^2 + \gamma_I \sigma_\varepsilon^4 \tag{5-48}$$

As indicated previously, up to the order of the perturbation approximations used, $\boldsymbol{\Omega}$ and $\mathbf{V}_*$ may be used interchangeably. However, by using the definitions of $\mathbf{V}_*$ and $\boldsymbol{\Omega}$ in (3-21) it can be seen that $\mathbf{Q}'\boldsymbol{\omega}^{1/2}(\boldsymbol{\Omega} - \mathbf{V}_*)\boldsymbol{\omega}^{1/2}\mathbf{Q} \geq 0$ so that $\mathbf{Q}'\boldsymbol{\omega}^{1/2}\boldsymbol{\Omega}\boldsymbol{\omega}^{1/2}\mathbf{Q} \geq \mathbf{Q}'\boldsymbol{\omega}^{1/2}\mathbf{V}_*\boldsymbol{\omega}^{1/2}\mathbf{Q}$. Thus, a larger value for (5-48) is

$$\sigma_\varepsilon^2 + \gamma_w \sigma_\beta^2 + \gamma_I \sigma_\varepsilon^4 \approx \frac{1}{\mathbf{Q}'\mathbf{Q}} \mathbf{Q}' \boldsymbol{\omega}^{\frac{1}{2}} \boldsymbol{\Omega} \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{Q} \sigma_\varepsilon^2 + \gamma_I \sigma_\varepsilon^4 \tag{5-49}$$

Correction factor $c_c$ is obtained by using (5-49) and the definition for $\hat{\gamma}_w \sigma_\beta^2$ in (5-45) as

$$c_c = \frac{(n-p)(\mathbf{Q}'(\boldsymbol{\omega}/b)^{\frac{1}{2}} \boldsymbol{\Omega} (\boldsymbol{\omega}/b)^{\frac{1}{2}} \mathbf{Q} + \mathbf{Q}'\mathbf{Q} \gamma_I \sigma_\varepsilon^2 / b)}{\mathbf{Q}'\mathbf{Q}(n - ap + \hat{\gamma}_I \sigma_\varepsilon^2 / b)} \tag{5-50}$$

When spatial correlation is small or Gauss-Markov estimation is used, $\hat{\gamma}_I \sigma_\varepsilon^2 / b$ may be evaluated using (5-29). Computation of $\gamma_I \sigma_\varepsilon^2 / b$ is discussed in section 6.

***Relation between the correction factor and spatial correlation.*** Equation (5-50) shows possible strong dependence of the ratio $(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}))/(S(\hat{\boldsymbol{\theta}})/(n-p))$ on spatial correlation when $\boldsymbol{\omega} = \hat{\boldsymbol{\omega}} \neq b\boldsymbol{\Omega}^{-1}$. Expansion the form $\sum_i(\hat{Q}_i - \sum_i \hat{Q}_i / n)^2$ leads to the relation $0 \leq \hat{\mathbf{Q}}'\mathbf{1}\hat{\mathbf{Q}} \leq n\hat{\mathbf{Q}}'\hat{\mathbf{Q}}$, where $\hat{\mathbf{Q}}$ is $\mathbf{Q}$ using $\boldsymbol{\omega} = \hat{\boldsymbol{\omega}}$. Therefore, if spatial correlation were positive and large ($(\hat{\boldsymbol{\omega}}/b)^{1/2}\boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}/b)^{1/2}$ near $\mathbf{1}$) and if most entries in $\hat{\mathbf{Q}}$ had the same sign, the numerator of $c_c$ could be large if $n$ were large, as is the numerator of $c_r$ from (5-27). However, unlike $c_r$, $c_c$ could be small, even if $n$ and correlation were large, if $\hat{\mathbf{Q}}'(\hat{\boldsymbol{\omega}}/b)^{1/2}\boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}/b)^{1/2}\hat{\mathbf{Q}}$ were small compared with $\hat{\mathbf{Q}}'\hat{\mathbf{Q}}$ as could occur if entries in $\hat{\mathbf{Q}}$ were fairly uniform and equally divided by sign. Component correction factor $\gamma_I \sigma_\varepsilon^2 / b$ defined for $c_c$ is for model combined intrinsic nonlinearity and so is less likely to be near zero than is $\gamma_I \sigma_\varepsilon^2 / b$ defined for $c_r$. The former component correction factor is given as (G-8), appendix G, and is evaluated in (G-9), (G-11), (G-14), (G-15), and (G-17)-(G-19). The final result is lengthy, but straightforward, and so is not given here. In section 6, $\gamma_I \sigma_\varepsilon^4$ is investigated using general concepts of combined intrinsic nonlinearity.

***Computation of an individual confidence interval when the weight matrix is known.*** As discussed after (5-44), an individual confidence interval is found from the maximum and minimum values of $g(\boldsymbol{\gamma\theta})$ over the likelihood region, (5-44). If there are no maxima or minima of $g(\boldsymbol{\gamma\theta})$ more extreme than those on the boundary of (5-44), then an individual confidence interval is calculated in the same manner as for the Scheffé interval. That is, an individual confidence interval is calculated from extreme values of

$$L(\boldsymbol{\theta},\lambda) = g(\boldsymbol{\gamma\theta}) + \lambda\left(\frac{S(\hat{\boldsymbol{\theta}})}{n-p} c_c t^2_{\alpha/2}(n-p) - S(\boldsymbol{\theta}) + S(\hat{\boldsymbol{\theta}})\right) \tag{5-51}$$

where $t_{\alpha/2}(n-p)$ is the $(1-\alpha/2)\times 100$ percentile of the $t$ distribution, and $t^2_{\alpha/2}(n-p) = F_\alpha(1, n-p)$. Numerical methodology is the same as used for a Scheffé interval.

***Individual confidence interval when the weight matrix is unknown.*** As for confidence regions and Scheffé intervals, when the weights are unknown $\ell(\boldsymbol{\theta})$ defined by (4-50) should be used in place of $S(\boldsymbol{\theta})$. An analysis like the one given by (5-34)-(5-42) applies for the present case if $\tilde{\boldsymbol{\eta}} = \mathbf{Y} - \mathbf{f}(\boldsymbol{\gamma}\tilde{\boldsymbol{\theta}})$ replaces $\boldsymbol{\eta}_* = \mathbf{Y} - \mathbf{f}(\boldsymbol{\gamma\theta}_*)$. Thus,

$$\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}) \approx (n-ap)\frac{S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}})}{S(\hat{\boldsymbol{\theta}})} \tag{5-52}$$

so that, from (5-43), as an approximation

$$\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}) \sim \frac{n-ap}{n-p} c_c t^2(n-p) \tag{5-53}$$

If model intrinsic nonlinearity is small, then $c_c$ is given by (5-50) so that (5-53) becomes

$$\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}) \sim \left(\frac{\mathbf{Q}'(\boldsymbol{\omega}_G/b)^{\frac{1}{2}}\boldsymbol{\Omega}(\boldsymbol{\omega}_G/b)^{\frac{1}{2}}\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}} + \gamma_I\sigma^2_\varepsilon/b\right)t^2(n-p) \tag{5-54}$$

where $\boldsymbol{\omega}$ is replaced by the diagonal matrix of $\omega_{Gk}$ values, $\boldsymbol{\omega}_G$.

Because $\boldsymbol{\omega}_G$ and $\boldsymbol{\Omega}$ are assumed to be unknown, (5-54) cannot be directly used. A useful approximate bound for $\mathbf{Q}'(\boldsymbol{\omega}_G/b)^{1/2}\boldsymbol{\Omega}(\boldsymbol{\omega}_G/b)^{1/2}\mathbf{Q}$ is obtained as follows. For convenience, temporarily let $(\boldsymbol{\omega}_G/b)^{1/2}\boldsymbol{\Omega}(\boldsymbol{\omega}_G/b)^{1/2} = \mathbf{C}$. Then, if $\boldsymbol{\omega}_G$ adequately approximates $\hat{\boldsymbol{\omega}}$, diagonal entries of $\mathbf{C}$ are all approximately 1's and off diagonal entries are all less than or equal to approximately 1 in magnitude. Assuming that all significant off-diagonal entries in $\mathbf{C}$ are positive, the approximate maximum magnitude of an entry of the row vector $\mathbf{Q}'\mathbf{C}$ is given either by the sum of all positive values of $Q_i$ or the negative of the sum of all negative values of $Q_i$. Let either sum be $\sum_{i(s)} Q_i$, where *i(s)* indicates the sum over all entries having the same sign. Then the approximate maximum value that $\mathbf{Q}'\mathbf{C}\mathbf{Q}$ could have, $V_{mx}$, is obtained as

$$V_{mx} = \max_{S} \sum_{j(s)}\left(\sum_{i(s)} Q_i \times Q_j\right) = \max_{S}\left(\sum_{i(s)} Q_i\right)^2 \tag{5-55}$$

where the maximum over $s$ ($\max\limits_{s}(\cdots)$) indicates the maximum for either the sum over positive signs or the sum over negative signs. Finally, as an approximate bound

$$\frac{\mathbf{Q}'(\omega_G/b)^{\frac{1}{2}}\,\Omega(\omega_G/b)^{\frac{1}{2}}\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}} \le \frac{V_{mx}}{\mathbf{Q}'\mathbf{Q}} \tag{5-56}$$

Hamilton and Wiens (1987) and Cooley (1997) found that for the cases they studied $\gamma_I\sigma_\varepsilon^2/b$ did not increase the size of a confidence interval more than about 6 percent, although Hamilton and Wiens (1987) found that $\gamma_I\sigma_\varepsilon^2/b$ could decrease the size of a confidence interval as much as about 35 percent. Thus, an approximate, perhaps conservative, confidence interval can be computed by using (5-56) in (5-54), neglecting $\gamma_I\sigma_\varepsilon^2/b$, and finding extreme values of

$$L(\theta,\lambda) = g(\gamma\theta) + \lambda\left(\frac{V_{mx}}{\mathbf{Q}'\mathbf{Q}}t_{\alpha/2}^2(n-p) - \ell(\theta) + \ell(\hat{\theta})\right) \tag{5-57}$$

## Detection of Combined Intrinsic Nonlinearity

Analysis of weighted residuals to detect possible significant model and system types of intrinsic nonlinearity was discussed in section 4, where it was found that significant model and system types of intrinsic nonlinearity can be indicated by a slope of the plot of weighted residuals in relation to weighted function values that is significantly different than zero. It also was found that premultiplication of the weighted residual vector by $\mathbf{R}$ should yield a vector of nearly zero values if model intrinsic nonlinearity is small. Similar measures for model and system types of combined intrinsic nonlinearity are developed using a weighted constrained residual vector $(\mathbf{I}-\mathbf{QQ}'/\mathbf{Q}'\mathbf{Q})\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))$ that equals the standard residual vector when model intrinsic nonlinearity and model combined intrinsic nonlinearity are both small. This is shown as follows. Use of (E-29), appendix E, yields

$$\begin{aligned}
&(\mathbf{I}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))\\
&\approx(\mathbf{I}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}\mathbf{U}_*+\frac{1}{2}\sum_j\omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))\\
&-(\mathbf{I}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_k\mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}}-\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)\\
&+\frac{1}{2}(\mathbf{I}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}}-\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})\\
&=(\mathbf{I}-\mathbf{R})(\omega^{\frac{1}{2}}\mathbf{U}_*+\frac{1}{2}\sum_j\omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))
\end{aligned}$$

$$-(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df'}\omega\mathbf{Df})^{-1}(\sum_k \mathbf{D}^2 f_k \tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}}-\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q'Q}}\mathbf{Q'}\omega^{\frac{1}{2}}\mathbf{U}_*) \qquad (5\text{-}58)$$

Next (C-4) and (C-11), appendix C, show that the nonlinear model terms are small in magnitude if the model intrinsic nonlinearity and the model combined intrinsic nonlinearity are small. Finally, comparison of (5-58) with (4-23) shows that, in this instance, weighted constrained and standard weighted residuals are both approximately given by $(\mathbf{I}-\mathbf{R})(\omega^{1/2}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{1/2}\mathbf{e'D}_\beta^2 f_j \mathbf{e})$.

If model intrinsic nonlinearity and model combined intrinsic nonlinearity are both small, the sum of weighted constrained residuals should be nearly the same as the sum of weighted residuals, or

$$\sum_i(\mathbf{I}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})_i\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta})) \approx \sum_i(\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}\mathbf{e'D}_\beta^2 f_j \mathbf{e}) \approx \sum_i \omega_i^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})) \qquad (5\text{-}59)$$

where $(\mathbf{I}-\mathbf{QQ'}/\mathbf{Q'Q})_i$ is row $i$ of $\mathbf{I}-\mathbf{QQ'}/\mathbf{Q'Q}$. Additionally, if the weighted constrained residual vector is premultiplied by **R**, the result should be a vector of nearly zero values if the model combined intrinsic nonlinearity is small. Matrix **R** and vector **Q** can be computed for the test using the same set of parameters as used to compute **R** for the test applied to residuals. (See discussion in the paragraph following (4-44).)

The slope of the plot of weighted constrained residuals in relation to weighted function values $\omega_i^{1/2}\mathbf{f}(\gamma\tilde{\theta})$ should not be significant if model and system types of intrinsic nonlinearity and model and system types of combined intrinsic nonlinearity are all small, as shown by the following development of (5-62). (This development can be skipped if desired.) The slope is proportional to $\sum_i \omega_i^{1/2}\mathbf{f}(\gamma\tilde{\theta})(\mathbf{I}-\mathbf{QQ'}/\mathbf{Q'Q})_i\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))$ $-\sum_i \omega_i^{1/2}\mathbf{f}(\gamma\tilde{\theta})\sum_i(\mathbf{I}-\mathbf{QQ'}/\mathbf{Q'Q})_i\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))/n$. Evaluation of the first term to second order using (E-5), (E-14), (E-28), and (5-58) results in

$$\sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\tilde{\theta})(\mathbf{I}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})_i\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))$$

$$\approx \sum_i(\omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\theta_*) + (\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e'D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))$$

$$+(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})_i\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df'}\omega\mathbf{Df})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}}-\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q'Q}}\mathbf{Q'}\omega^{\frac{1}{2}}\mathbf{U}_*)$$

$$-\frac{1}{2}\frac{1}{\mathbf{Q'Q}}Q_i(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}}-\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) + \frac{1}{2}\sum_j \omega_{ij}^{\frac{1}{2}}(\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}-\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 f_j(\gamma'\gamma)^{-1}\gamma'\mathbf{e}))$$

$$\bullet((\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e'D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))-(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})_i\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df'}\omega\mathbf{Df})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}}$$

$$-\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q'Q}}\mathbf{Q'}\omega^{\frac{1}{2}}\mathbf{U}_*))$$

$$\approx \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\theta_*)((\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})) - (\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}$$
$$\bullet(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}} - \mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)) \tag{5-60}$$

Similarly evaluation of the second term yields

$$\sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\tilde{\theta})\sum_i(\mathbf{I}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))/n$$
$$\approx \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\theta_*)\sum_i((\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))$$
$$-(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}} - \mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*))/n \tag{5-61}$$

Combination of (5-60) and (5-61) results in

$$\sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\tilde{\theta})(\mathbf{I}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta})) - \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\tilde{\theta})\sum_i(\mathbf{I}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))/n$$
$$\approx \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\theta_*)((\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})) - (\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}$$
$$\bullet(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}} - \mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)) - \sum_i \omega_i^{\frac{1}{2}}\mathbf{f}(\gamma\theta_*)\sum_i((\mathbf{I}-\mathbf{R})_i(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))$$
$$-(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}} - \mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*))/n \tag{5-62}$$

When the model and system types of intrinsic nonlinearity and the model and system types of combined intrinsic nonlinearity are negligible, the slope is the slope for a linear model, which should not be significant. Note that in this case the expected value of the slope is zero.

The constrained regression estimate $\tilde{\theta}$, which yields the constrained weighted residuals, is obtained with the constraint $g(\gamma\tilde{\theta}) = g(\gamma\theta_*)$. However, $g(\gamma\theta_*)$ is unknown. Therefore, the confidence limits of $g(\gamma\theta_*)$ must be used instead of $g(\gamma\theta_*)$, and the necessary values of $\tilde{\theta}$ are found by solving the appropriate extreme value problem, (5-51) or (5-57), depending on whether or not $\omega$ is known. Note that this gives two sets of constrained weighted residuals to analyze for each confidence interval, one set at each confidence limit.

## Development of Individual Prediction Intervals

***Forms for predicted variables, covariances, and sum of squared errors.*** A predicted observation is given by

$$Y_p = g(\beta) + \varepsilon_p \tag{5-63}$$

where $\varepsilon_p$ is a predicted observation error that is assumed to have the marginal normal distribution

$$\varepsilon_p \sim N(0, V_{\varepsilon p}\sigma_\varepsilon^2) \tag{5-64}$$

and $V_{\varepsilon p}\sigma_\varepsilon^2 = Var(\varepsilon_p)$. Equation (5-63) can be written in terms of a combined model and observation error, $Y_p - g(\gamma\theta_*)$, by using (4-28) and (4-32) as follows.

$$\begin{aligned}
Y_p &= g(\gamma\theta_*) + g(\beta) - g(\gamma\bar{\theta}) + g(\gamma\bar{\theta}) - g(\gamma\theta_*) + \varepsilon_p \\
&= g(\gamma\theta_*) + \mathbf{D}_\beta g\mathbf{e} + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 g\mathbf{e} - \mathbf{D}g\,(\gamma'\gamma)^{-1}\gamma'\mathbf{e} - \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma\gamma')^{-1}\gamma'\mathbf{e} + \varepsilon_p \\
&= g(\gamma\theta_*) + U_p^* + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 g - \gamma(\gamma\gamma')^{-1}\mathbf{D}^2 g(\gamma\gamma')^{-1}\gamma')\mathbf{e} \\
&= g(\gamma\theta_*) + \upsilon_*
\end{aligned} \tag{5-65}$$

where

$$\upsilon_* = U_p^* + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 g - \gamma(\gamma\gamma')^{-1}\mathbf{D}^2 g(\gamma\gamma')^{-1}\gamma')\mathbf{e} \tag{5-66}$$

and

$$U_p^* = \mathbf{D}_\beta\, g(\mathbf{I} - \gamma(\gamma\gamma')^{-1}\gamma')\mathbf{e} + \varepsilon_p \tag{5-67}$$

The form for $\upsilon_*$ implicitly assumes that model errors **e** remain the same for predictions as they were for the original model. This should be true for errors in framework properties such as hydraulic conductivity, and is correct for all types of error if the prediction interval is used to bound a present value of $g(\beta)$ by setting $V_{\varepsilon p}$ to zero. The form is an approximation for future values of hydrologic variables such as recharge and discharge. In fact, if the model errors for recharge and discharge were controlled mainly by transient processes, then there might be little to tie the original model errors for recharge and discharge to the predicted model errors for recharge and discharge. In this instance the two sets of errors could have the same mean and spatial covariance, but could be nearly independent of each other. This case is addressed by setting the appropriate covariances to zero as explained below.

Assume $\varepsilon_p$ and **e** to be independent, as are $\varepsilon$ and **e**. However, $\varepsilon_p$ and $\varepsilon$ may be correlated, so let

$$Cov(\varepsilon, \varepsilon_p) = \mathbf{C}_{\varepsilon p}\sigma_\varepsilon^2 \tag{5-68}$$

Also, let

$$Var(U_p^*) = \mathbf{D}_\beta g(\mathbf{I} - \gamma(\gamma\acute{\gamma})^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma\acute{\gamma})^{-1}\gamma')\mathbf{D}_\beta g'\sigma_\beta^2 + V_{\varepsilon p}\sigma_\varepsilon^2 = V_p\sigma_\varepsilon^2 \tag{5-69}$$

and

$$\begin{aligned} Cov(\mathbf{U}_*, U_p^*) &= E(\mathbf{D}_\beta \mathbf{f}(\mathbf{I} - \gamma(\gamma\acute{\gamma})^{-1}\gamma')\,\mathbf{e} + \boldsymbol{\varepsilon})(\mathbf{e}'(\mathbf{I} - \gamma(\gamma\acute{\gamma})^{-1}\gamma')\,\mathbf{D}_\beta g' + \varepsilon_p) \\ &= \mathbf{D}_\beta \mathbf{f}(\mathbf{I} - \gamma(\gamma\acute{\gamma})^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma\acute{\gamma})^{-1}\gamma')\,\mathbf{D}_\beta g'\sigma_\beta^2 + \mathbf{C}_{\varepsilon p}\sigma_\varepsilon^2 \\ &= \mathbf{C}_{ep}\sigma_\beta^2 + \mathbf{C}_{\varepsilon p}\sigma_\varepsilon^2 = \mathbf{C}_p\sigma_\varepsilon^2 \end{aligned} \tag{5-70}$$

where $\mathbf{C}_{ep}\sigma_\beta^2 = \mathbf{D}_\beta \mathbf{f}(\mathbf{I} - \gamma(\gamma\acute{\gamma})^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma\acute{\gamma})^{-1}\gamma')\,\mathbf{D}_\beta g'\sigma_\beta^2$ and the covariance between $\mathbf{U}_*$ and $U_p^*$ results because of (5-68) and because $\mathbf{U}_*$ and $U_p^*$ involve the same model errors, **e**. If some of the model errors for the predictions were envisioned to be independent of the original model errors as discussed above, then these covariances could be set equal to zero.

Addition of the predicted observation to the set of observations implies a corresponding augmented sum of squares function. For a Gauss-Markov type of estimation this would be of the following form (which neglects nonlinear terms in the covariance matrix because only the form of the matrix is of interest here).

$$\begin{aligned} &\begin{bmatrix} \mathbf{Y} - \mathbf{f}(\gamma\theta) \\ \upsilon \end{bmatrix}' \begin{bmatrix} \mathbf{V}_* & \mathbf{C}_p \\ \mathbf{C}_p' & V_p \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y} - \mathbf{f}(\gamma\theta) \\ \upsilon \end{bmatrix} \\ &= (\mathbf{Y} - \mathbf{f}(\gamma\theta))'(\mathbf{V}_*^{-1} + \mathbf{V}_*^{-1}\mathbf{C}_p\mathbf{C}_p'\mathbf{V}_*^{-1}/C)(\mathbf{Y} - \mathbf{f}(\gamma\theta)) - 2(\mathbf{Y} - \mathbf{f}(\gamma\theta))'\mathbf{V}_*^{-1}\mathbf{C}_p\upsilon/C \\ &+ \upsilon^2/C \end{aligned} \tag{5-71}$$

where $\upsilon = Y_p - g(\gamma\theta)$, $C = V_p - \mathbf{C}_p'\mathbf{V}_*^{-1}\mathbf{C}_p$, and the augmented matrix was inverted by partitioning (Hohn, 1964, p. 108-109). In terms of a general weight matrix, the augmented sum of squares has the same form and is defined as

$$S_a(\theta, \upsilon) = (\mathbf{Y} - \mathbf{f}(\gamma\theta))'\mathbf{W}(\mathbf{Y} - \mathbf{f}(\gamma\theta)) + 2(\mathbf{Y} - \mathbf{f}(\gamma\theta))'\mathbf{W}_p\upsilon + W_p\upsilon^2 \tag{5-72}$$

where $\mathbf{W}$, $\mathbf{W}_p$, and $W_p$ compose a general, augmented weight matrix, $\mathbf{W}_a$, defined by

$$\mathbf{W}_a = \begin{bmatrix} \mathbf{W} & \mathbf{W}_p \\ \mathbf{W}_p' & W_p \end{bmatrix} \tag{5-73}$$

***Statistical distribution and prediction interval when the weight matrix is known.*** As for the statistical distribution used to define the individual confidence interval, derivation of the appropriate distribution to define a prediction interval involves minimizing the sum of squares subject to a constraint. In the present case the function to minimize is $S_a(\theta, \upsilon)$ subject to the constraint $g(\gamma\theta) + \upsilon = Y_p$. The constrained parameter vector is therefore $(\theta, \upsilon)$. It is convenient to replace $\upsilon$ with $Y_p - \theta_p$, so that $\theta_p$ is the new parameter. With this definition the augmented sum of squares becomes

$$S_a(\theta,\theta_p) = (\mathbf{Y}-\mathbf{f}(\gamma\theta))'\mathbf{W}(\mathbf{Y}-\mathbf{f}(\gamma\theta)) + 2(\mathbf{Y}-\mathbf{f}(\gamma\theta))'\mathbf{W}_p(Y_p-\theta_p)$$
$$+W_p(Y_p-\theta_p)^2 \tag{5-74}$$

Note from the relations $\upsilon = Y_p - g(\gamma\theta) = Y_p - \theta_p$ that

$$\theta_p = g(\gamma\theta) \tag{5-75}$$

Equation (5-75) becomes the new constraint for the constrained regression. Special cases of (5-75) are $\bar{\theta}_p = g(\gamma\bar{\theta})$, $\theta_p^* = g(\gamma\theta_*)$, and $\tilde{\theta}_p = g(\gamma\tilde{\theta})$. The constrained regression estimate $(\tilde{\theta},\tilde{\theta}_p)$ is derived to second-order accuracy in appendix E.

In appendix F the appropriate distribution to define a prediction interval is derived as (F-89), which is repeated here in the form

$$\frac{S_a(\tilde{\theta},\tilde{\theta}_p) - S_a(\hat{\theta},\hat{\theta}_p)}{S_a(\hat{\theta},\hat{\theta}_p)/(n-p)} \sim c_p t^2(n-p) \tag{5-76}$$

where $(\hat{\theta},\hat{\theta}_p)$ are unconstrained regression estimates obtained by minimizing $S_a(\theta,\theta_p)$ and $c_p$ is a correction factor to be defined by (5-78). Equation (5-76) implies the likelihood region

$$S_a(\tilde{\theta},\tilde{\theta}_p) - S_a(\hat{\theta},\hat{\theta}_p) \le \frac{S_a(\hat{\theta},\hat{\theta}_p)}{n-p} c_p t_{\alpha/2}^2(n-p) \tag{5-77}$$

Maximum and minimum values of $g(\gamma\theta)+\upsilon$ over the likelihood region are the maximum and minimum values that $Y_p$ could have at probability level $\alpha$, which defines the prediction interval.

***Correction factor.*** The correction factor $c_p$ is defined as

$$c_p = \frac{\sigma_\varepsilon^2 + \gamma_{wa}\sigma_\beta^2 + \gamma_{Ia}\sigma_\varepsilon^4}{\sigma_\varepsilon^2 + (\hat{\gamma}_{wa}\sigma_\beta^2 + \hat{\gamma}_{Ia}\sigma_\varepsilon^4)/(n-p)} \tag{5-78}$$

where, from (F-91)-(F-95), appendix F, perturbation-based forms for the component correction factors are

$$\hat{\gamma}_{wa}\sigma_\beta^2 = (tr((\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}) - n + p)\sigma_\varepsilon^2 \tag{5-79}$$

$$\hat{\gamma}_{Ia}\sigma_\varepsilon^4 = E(S_a(\hat{\theta},\hat{\theta}_p)) - tr((\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^2 \tag{5-80}$$

$$\gamma_{wa}\sigma_\beta^2 = (\frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{Q}_a'\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\mathbf{Q}_a - 1)\sigma_\varepsilon^2 \tag{5-81}$$

and

$$\gamma_{Ia}\sigma_\varepsilon^4 = E(S_a(\tilde{\theta},\tilde{\theta}_p) - S_a(\hat{\theta},\hat{\theta}_p)) - \frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{Q}_a'\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\mathbf{Q}_a\sigma_\varepsilon^2 \tag{5-82}$$

In (5-79)-(5-82), $\mathbf{I}_a$ is the identity matrix of order $n+1$,

$$\mathbf{R}_a = \mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a^{\frac{1}{2}} \tag{5-83}$$

$$\mathbf{Q}_a = \mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h' \tag{5-84}$$

and

$$\mathbf{V}_{*a}\sigma_\varepsilon^2 = Var(\mathbf{U}_{*a}) \tag{5-85}$$

in which $\mathbf{D}_a\mathbf{f}_a$ and $\mathbf{D}_a h$ are augmented variables defined by (E-46) and (E-48), and

$$\mathbf{U}_{*a} = \begin{bmatrix} \mathbf{U}_* \\ U_p^* \end{bmatrix} \tag{5-86}$$

As before, factors $\hat{\gamma}_{wa}\sigma_\beta^2$ and $\gamma_{wa}\sigma_\beta^2$ correct for the possibility that $\mathbf{W}_a^{\frac{1}{2}} \neq \mathbf{V}_{*a}$, and factors $\hat{\gamma}_{Ia}\sigma_\varepsilon^4$ and $\gamma_{Ia}\sigma_\varepsilon^4$ correct for model intrinsic nonlinearity and model combined intrinsic nonlinearity.

An analysis of an expression analogous to, but more general than, (5-76) was performed using the same methods as used for individual confidence intervals. This analysis is omitted here because it is almost identical to the analysis for individual confidence intervals and reaches nearly identical conclusions. The only difference in conclusions is that the generalization of distribution (5-76) behaves even more like a $(\textit{correction factor}) \times t^2(n-p)$ distribution than does the generalization of (5-43).

***Approximate evaluation of the correction factor.*** First the general augmented form of $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}_a$, is defined by

$$\boldsymbol{\Omega}_a = \begin{bmatrix} \boldsymbol{\Omega} & E(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))(Y_p - g(\gamma\boldsymbol{\theta}_*))/\sigma_\varepsilon^2 \\ E(Y_p - g(\gamma\boldsymbol{\theta}_*))(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'/\sigma_\varepsilon^2 & b\hat{\omega}_p^{-1} \end{bmatrix} \tag{5-87}$$

in which, using (5-65)-(5-67) and (5-69),

$$b\hat{\omega}_p^{-1} = E(Y_p - g(\gamma\boldsymbol{\theta}_*))^2/\sigma_\varepsilon^2 = E(\upsilon_*^2)/\sigma_\varepsilon^2$$

$$= V_p + \frac{1}{4} tr^2((\mathbf{D}_\beta^2 g - \gamma(\gamma\gamma)^{-1}\gamma' \mathbf{D}_\beta^2 g\gamma(\gamma\gamma)^{-1}\gamma')\mathbf{V}_\beta)\sigma_\beta^4 / \sigma_\varepsilon^2$$

$$+ \frac{1}{2} tr(((\mathbf{D}_\beta^2 g - \gamma(\gamma\gamma)^{-1}\gamma' \mathbf{D}_\beta^2 g\gamma(\gamma\gamma)^{-1}\gamma')\mathbf{V}_\beta)^2)\sigma_\beta^4 / \sigma_\varepsilon^2 \qquad (5\text{-}88)$$

Second, the general weight matrix $\mathbf{W}_a$ is replaced by a block diagonal weight matrix created by ignoring terms corresponding to the off-diagonal covariances between $\mathbf{Y} - \mathbf{f}(\gamma\theta_*)$ and $Y_p - g(\gamma\theta_*)$. The general weight matrix $\omega$ replaces $\mathbf{W}$ and the general weight $\omega_p$ replaces $W_p$. Thus, the weight matrix is given by

$$\omega_a = \begin{bmatrix} \omega & \mathbf{0} \\ \mathbf{0}' & \omega_p \end{bmatrix} \qquad (5\text{-}89)$$

Third, by analogy with (5-48) and (5-49), the numerator of $c_p$ becomes

$$\sigma_\varepsilon^2 + \gamma_{wa}\sigma_\beta^2 + \gamma_{Ia}\sigma_\varepsilon^4 = \frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{Q}_a'\omega_a^{\frac{1}{2}}\mathbf{V}_{*a}\omega_a^{\frac{1}{2}}\mathbf{Q}_a\sigma_\varepsilon^2 + \gamma_{Ia}\sigma_\varepsilon^4$$

$$\leq \frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{Q}_a'\omega_a^{\frac{1}{2}}\Omega_a\omega_a^{\frac{1}{2}}\mathbf{Q}_a\sigma_\varepsilon^2 + \gamma_{Ia}\sigma_\varepsilon^4 \qquad (5\text{-}90)$$

where now

$$\mathbf{Q}_a = \omega_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\omega_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h' \qquad (5\text{-}91)$$

Fourth, evaluation of $\mathbf{Q}_a$ using (5-89), (E-48), and (E-56), appendix E, yields

$$\mathbf{Q}_a = \begin{bmatrix} \mathbf{Q} \\ -\omega_p^{-\frac{1}{2}} \end{bmatrix} \qquad (5\text{-}92)$$

Fifth, (5-90) is expressed using (5-87), (5-89), and (5-92) as

$$\sigma_\varepsilon^2 + \gamma_{wa}\sigma_\beta^2 + \gamma_{Ia}\sigma_\varepsilon^4 \approx \frac{(\mathbf{Q}'(\omega/b)^{\frac{1}{2}}\Omega(\omega/b)^{\frac{1}{2}}\mathbf{Q} - 2\omega_p^{-\frac{1}{2}}(\omega_p/b)^{\frac{1}{2}}\mathbf{C}'(\omega/b)^{\frac{1}{2}}\mathbf{Q} + \hat{\omega}_p^{-1})b\sigma_\varepsilon^2}{\mathbf{Q}'\mathbf{Q} + \omega_p^{-1}}$$

$$+\gamma_{Id}\sigma_\varepsilon^4 \qquad (5\text{-}93)$$

where $\mathbf{C} = E(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))(Y_p - g(\gamma\theta_*))/\sigma_\varepsilon^2$.

Evaluation of the denominator of $c_p$ uses the definition of $\mathbf{R}_a$ as $\mathbf{R}_a = \omega_a^{1/2}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\omega_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\omega_a^{1/2}$. Then, from (5-89) and (E-56), appendix E,

$$\mathbf{R}_a = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0}' & 1 \end{bmatrix} \tag{5-94}$$

Evaluation of $\hat{\gamma}_{wa}\sigma_\beta^2$ shows that

$$\begin{aligned}\hat{\gamma}_{wa}\sigma_\beta^2 &= (tr((\mathbf{I}_a - \mathbf{R}_a)\omega_a^{\frac{1}{2}}\mathbf{V}_{*a}\omega_a^{\frac{1}{2}}) - n + p)\sigma_\varepsilon^2 \\ &\le (tr((\mathbf{I}_a - \mathbf{R}_a)\omega_a^{\frac{1}{2}}\Omega_a\omega_a^{\frac{1}{2}}) - n + p)\sigma_\varepsilon^2 \\ &= (tr((\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}}) - n + p)\sigma_\varepsilon^2 \\ &= \hat{\gamma}_w\sigma_\beta^2\end{aligned} \tag{5-95}$$

The component correction factor and, therefore, the form of the denominator for $c_p$ are the same as for $c_r$ and $c_c$. Finally, $c_p$ can be expressed as

$$c_p \approx \frac{(n-p)(\mathbf{Q}'(\omega/b)^{\frac{1}{2}}\Omega(\omega/b)^{\frac{1}{2}}\mathbf{Q} - 2\omega_p^{-\frac{1}{2}}(\omega_p/b)^{\frac{1}{2}}\mathbf{C}'(\omega/b)^{\frac{1}{2}}\mathbf{Q} + \hat{\omega}_p^{-1} + (\mathbf{Q}'\mathbf{Q} + \omega_p^{-1})\gamma_{Ia}\sigma_\varepsilon^2/b)}{(\mathbf{Q}'\mathbf{Q} + \omega_p^{-1})(n - ap + \hat{\gamma}_I\sigma_\varepsilon^2/b)} \tag{5-96}$$

where, as shown in appendix G, $\hat{\gamma}_{Ia}\sigma_\varepsilon^2 = \hat{\gamma}_I\sigma_\varepsilon^2$. Component correction factor $\gamma_{Ia}\sigma_\varepsilon^4$ is evaluated in (G-31)-(G-42), appendix G. The result is lengthy and is not given here. In section 6, $\gamma_{Ia}\sigma_\varepsilon^4$ is investigated using general concepts of combined intrinsic nonlinearity. When spatial correlation is small or Gauss-Markov estimation is used, $\hat{\gamma}_I\sigma_\varepsilon^2$ may be evaluated using (5-29).

***The relation between the correction factor and spatial correlation.*** Although (5-96) shows dependence of the ratio $(S_a(\tilde{\theta},\tilde{\theta}_p) - S_a(\hat{\theta},\hat{\theta}_p))/(S_a(\hat{\theta},\hat{\theta}_p)/(n-p))$ on spatial correlation, the dependence is less than that displayed by (5-50) because of terms involving $\omega_p^{-1}$. In fact, if $\omega_p = \hat{\omega}_p$ and $\omega_p^{-1} >> \mathbf{Q}'\mathbf{Q}$, then from (G-46), appendix G, $\gamma_{Ia}\sigma_\varepsilon^2 \approx 0$, and, letting $n - ap = (1-c)(n-p)$,

$$c_p \approx \frac{1}{1 - c + \hat{\gamma}_I\sigma_\varepsilon^2/(b(n-p))} \tag{5-97}$$

Furthermore, because $\hat{\gamma}_I\sigma_\varepsilon^2/b$ is usually small,

$$c_p \approx \frac{1}{1-c} \tag{5-98}$$

which, unlike $c_c$, has no dependence on $n$. Behavior of prediction intervals when $\omega_p^{-1} >> \mathbf{Q}'\mathbf{Q}$ is analyzed further later in this section.

***Computation of a prediction interval when the weight matrix is known and is block diagonal.*** When $\mathbf{W}_a = \omega_a$, (5-74), written in terms of estimates $\hat{\theta}$ and $\hat{\theta}_p$, reduces to

$$S_a(\hat{\theta},\hat{\theta}_p) = S(\hat{\theta}) + \omega_p(Y_p - \hat{\theta}_p)^2 \qquad (5\text{-}99)$$

so that $\hat{\theta}$ is the standard regression estimate obtained by minimizing $S(\theta)$, and $\hat{\theta}_p = Y_p$. Therefore, when $\mathbf{W}_a = \omega_a$, $S_a(\hat{\theta},\hat{\theta}_p) = S(\hat{\theta})$. Similarly, when $\mathbf{W}_a = \omega_a$, $S_a(\theta,\upsilon)$ from (5-72) reduces to

$$S_a(\theta,\upsilon) = S(\theta) + \omega_p\upsilon^2 \qquad (5\text{-}100)$$

which is more convenient than $S_a(\theta,\theta_p)$ for calculating prediction intervals. Finally, because the prediction limits are the maximum and minimum limits (extreme values) of $g(\gamma\theta)+\upsilon$ over the likelihood region (5-77), a prediction interval can be computed under the same assumptions as used for (5-51) from the extreme values of

$$L(\theta,\upsilon,\lambda) = g(\gamma\theta) + \upsilon + \lambda\left(\frac{S(\hat{\theta})}{n-p}c_p t^2_{\alpha/2}(n-p) - S(\theta) - \omega_p\upsilon^2 + S(\hat{\theta})\right) \qquad (5\text{-}101)$$

Numerical methodology for finding the extreme values is given by Vecchia and Cooley (1987).

Often the second moment of $Y_p - g(\gamma\theta_*)$ will be known rather than the weight $\hat{\omega}_p$. This second moment is given by $\hat{\omega}_p^{-1}b\sigma_\varepsilon^2$, so (5-100) would be written incorporating it as

$$S_a(\theta,\upsilon) = S(\theta) + \left(\frac{\hat{\omega}_p}{b\sigma_\varepsilon^2}\right)b\sigma_\varepsilon^2\upsilon^2 \qquad (5\text{-}102)$$

Variance $b\sigma_\varepsilon^2$ is unknown. Christensen and Cooley (1999b, p. 2629) argued that replacement of $b\sigma_\varepsilon^2$ with its estimate $s^2$ (given by (5-32)) should lead to a slightly conservative prediction interval. Thus, when $\hat{\omega}_p^{-1}b\sigma_\varepsilon^2$ is known, the prediction interval should be computed from

$$L(\theta,\upsilon,\lambda) = g(\gamma\theta) + \upsilon + \lambda\left(\frac{S(\hat{\theta})}{n-p}c_p t^2_{\alpha/2}(n-p) - S(\theta) - \left(\frac{\hat{\omega}_p}{b\sigma_\varepsilon^2}\right)s^2\upsilon^2 + S(\hat{\theta})\right) \qquad (5\text{-}103)$$

The estimated weight $(\hat{\omega}_p/(b\sigma_\varepsilon^2))s^2$ simply replaces the known, general weight $\omega_p$ in (5-101) when making the calculations.

***Approximate prediction interval when the predicted error predominates.*** With the above background, the approximate form for a prediction interval for the fairly common case where $\omega_p = \hat{\omega}_p$ and $\omega_p^{-1} >> \mathbf{Q'Q}$ can now be developed. The development utilizes linearized models for $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$, but the interval also will be shown to apply for nonlinear models. The linearized models are obtained as Taylor series expansions about $\gamma\bar{\theta}$ that retain only the first-order terms rather than both the first- and second-order terms as previously done. This yields

$$\mathbf{f}(\gamma\theta) \approx \mathbf{f}(\gamma\bar{\theta}) + \mathbf{Df}(\theta - \bar{\theta}) \qquad (5\text{-}104)$$

and

$$g(\gamma\boldsymbol{\theta}) \approx g(\gamma\bar{\boldsymbol{\theta}}) + \mathbf{D}g(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) \qquad (5\text{-}105)$$

As shown in appendix H, insertion of (5-104) and (5-105) into (5-101) followed by solution of the extreme value problem gives the limits of $g(\gamma\tilde{\boldsymbol{\theta}}) + \tilde{\upsilon} = \tilde{Y}_p$ as

$$\tilde{Y}_p = g(\gamma\hat{\boldsymbol{\theta}}) \pm \left(\frac{S(\hat{\boldsymbol{\theta}})}{n-p} c_p t^2_{\alpha/2}(n-p)(\mathbf{Q}'\mathbf{Q} + \omega_p^{-1})\right)^{\frac{1}{2}} \qquad (5\text{-}106)$$

which has the form of a standard linear prediction interval (Seber and Wild, 1989, p. 193). When $\omega_p = \hat{\omega}_p$ and $\hat{\omega}_p^{-1} >> \mathbf{Q}'\mathbf{Q}$, (5-32), (5-98), and (5-106) can be combined to obtain

$$\tilde{Y}_p = g(\gamma\hat{\boldsymbol{\theta}}) \pm t_{\alpha/2}(n-p)(\hat{\omega}_p^{-1}s^2)^{\frac{1}{2}} \qquad (5\text{-}107)$$

If the variance $\hat{\omega}_p^{-1}b\sigma_\varepsilon^2$ is known, then $(\hat{\omega}_p/(b\sigma_\varepsilon^2))s^2$ should replace weight $\hat{\omega}_p$ so that $\hat{\omega}_p^{-1}b\sigma_\varepsilon^2$ replaces $\hat{\omega}_p^{-1}s^2$.

Equation (5-106) also applies for nonlinear models when values of $\hat{\gamma}_I\sigma_\varepsilon^2$ and $\gamma_{Ia}\sigma_\varepsilon^2$ are both small in magnitude. This is because model intrinsic nonlinearity and model combined intrinsic nonlinearity are both small so that an equation of the form of (5-106) could have been derived using parameters $\boldsymbol{\phi}$ for which the models are nearly linear. Since $\mathbf{Q}'\mathbf{Q}$ and $g(\gamma\hat{\boldsymbol{\theta}})$ are both invariant under transformation of parameters, (5-106) is valid for nonlinear models $g(\gamma\hat{\boldsymbol{\theta}})$ and $\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$. This analysis also applies to (5-107) when $\omega_p^{-1} >> \mathbf{Q}'\mathbf{Q}$. Equation (5-107) implies that the variance of $g(\gamma\boldsymbol{\theta}_*) - g(\gamma\hat{\boldsymbol{\theta}})$ is small compared to the variance of $Y_p - g(\gamma\boldsymbol{\theta}_*)$ because $\hat{\omega}_p^{-1}s^2$ is the estimated variance of $Y_p - g(\gamma\boldsymbol{\theta}_*)$, but (5-107) implies that $(Y_p - g(\gamma\hat{\boldsymbol{\theta}}))/(\hat{\omega}_p^{-1}s^2)^{1/2}$ has an approximate $t(n-p)$ distribution.

An important conclusion to be drawn from (5-107) is that, if $\omega_p = \hat{\omega}_p$ and $\omega_p^{-1} >> \mathbf{Q}'\mathbf{Q}$, the effects of spatial correlation are contained almost wholly in the estimates $g(\gamma\hat{\boldsymbol{\theta}})$ and $s^2$ when $\hat{\omega}_p$ is known and in the estimate $g(\gamma\hat{\boldsymbol{\theta}})$ when $\hat{\omega}_p^{-1}b\sigma_\varepsilon^2$ is known. In the latter case spatial correlation has negligible effect on width of the prediction interval.

***Testing prediction intervals for accuracy.*** Christensen and Cooley (1999b) showed that general prediction intervals can be tested for overall accuracy using a cross-validation procedure whereby $Y$ values are withdrawn from the data set $\mathbf{Y}$ and predicted using prediction intervals one (or a few) at a time. The percentage of $Y$ values that should be contained in their prediction intervals at probability $\alpha$ can be determined and compared with the actual number of $Y$ values that are contained in their prediction intervals. This procedure should test primarily for the possibility that one or more correction factors $c_p$ should be larger than the values used. In the two field cases that Christensen and Cooley (1999b) studied, values of $c_p$ were implicitly assumed to be unity, and no evidence was found to indicate that the prediction intervals were too small. Christensen and Cooley (1999b) also used new data to test the prediction intervals and again found no evidence to indicate that they were too small.

***Prediction interval when the weight matrix is unknown.*** When the weight matrix is unknown, an extension of $\ell(\theta)$ can be used instead of $S_a(\theta,\theta_p)$. Two separate formulations are used depending on whether $Y_p$ is contained in one of the $q$ error groups or not. Both formulations lead to approximations of (5-76). If $Y_p$ is contained in one of the error groups, for example group $j$, then the augmented form $\ell_a(\theta,\upsilon)$ is defined by

$$\ell_a(\theta,\upsilon)=\frac{1}{2}\sum_{k=1}^{q}(n_k+\delta_{kj})\ln\Big(\sum_{i(k)}(Y_i-f_i(\gamma\theta))^2+\delta_{kj}\upsilon^2\Big) \tag{5-108}$$

where $\delta_{kj}$ is the Kronecker delta defined by (A-3), appendix A. If $Y_p$ is not in an error group, then an equation like (5-108) would involve a separate error group and term $\ln(\upsilon^2)$, which has a value of negative infinity when $\upsilon$ is zero (such as, for example, when $\upsilon=\hat{\upsilon}$). This problem is eliminated by defining $\ell_a(\theta,\upsilon)$ when $\upsilon$ is not in an error group as

$$\ell_a(\theta,\upsilon)=\ell(\theta)+\frac{\hat{\omega}_p}{b\sigma_\varepsilon^2}\upsilon^2 \tag{5-109}$$

where $\hat{\omega}_p^{-1}b\sigma_\varepsilon^2$ is presumed to be known.

If $Y_p$ is contained in one of the error groups, then expansion of (5-108) using the same ideas as employed in appendix D yields

$$\ell_a(\tilde{\theta},\tilde{\upsilon})-\ell_a(\hat{\theta},\hat{\upsilon})\approx\frac{1}{\sigma_\varepsilon^2}(S(\tilde{\theta})+\omega_{Gj}\tilde{\upsilon}^2-S(\hat{\theta})) \tag{5-110}$$

where $\hat{\upsilon}=0$ and $\omega_{Gj}$ replaces $\hat{\omega}_p$. Next, substitution of $S(\hat{\theta})/(n-ap)$ for $\sigma_\varepsilon^2$ as was done to obtain (5-37) results in

$$\ell_a(\tilde{\theta},\tilde{\upsilon})-\ell_a(\hat{\theta},\hat{\upsilon})\approx(n-ap)\frac{S(\tilde{\theta})+\omega_{Gj}\tilde{\upsilon}^2-S(\hat{\theta})}{S(\hat{\theta})} \tag{5-111}$$

If $Y_p$ is not in an error group, then estimation of $\hat{\omega}_p/(b\sigma_\varepsilon^2)$ with $(n-ap)\hat{\omega}_p/S(\hat{\theta})$ in (5-109) and use of (5-52) shows that

$$\ell_a(\tilde{\theta},\tilde{\upsilon})-\ell_a(\hat{\theta},\hat{\upsilon})\approx(n-ap)\frac{S(\tilde{\theta})+\hat{\omega}_p\tilde{\upsilon}^2-S(\hat{\theta})}{S(\hat{\theta})} \tag{5-112}$$

Use of (5-111) or (5-112) with (5-76) results in the approximate distribution

$$\ell_a(\tilde{\theta},\tilde{\upsilon})-\ell_a(\hat{\theta},\hat{\upsilon})\sim\frac{n-ap}{n-p}c_p t^2(n-p) \tag{5-113}$$

If model intrinsic nonlinearity is small, then $c_p$ can be simplified so that (5-113) can be written

$$\ell_a(\tilde{\theta},\tilde{\upsilon}) - \ell_a(\hat{\theta},\hat{\upsilon}) \sim \Big(\frac{\mathbf{Q}'(\omega_G/b)^{\frac{1}{2}}\Omega(\omega_G/b)^{\frac{1}{2}}\mathbf{Q} - 2\omega_p^{-\frac{1}{2}}(\omega_p/b)^{\frac{1}{2}}\mathbf{C}'(\omega_G/b)^{\frac{1}{2}}\mathbf{Q} + \hat{\omega}_p^{-1}}{\mathbf{Q}'\mathbf{Q} + \omega_p^{-1}}$$
$$+\gamma_{Ia}\sigma_\varepsilon^2/b\Big)t^2(n-p) \tag{5-114}$$

where, as in (5-54), $\omega_G$ replaces $\hat{\omega}$, and, if $Y_p$ is in an error group, $\omega_p^{-1} = \omega_{Gj}^{-1}$ and, if $Y_p$ is not in an error group, $\omega_p^{-1} = \hat{\omega}_p^{-1}$.

Matrices $\Omega$ and $\omega_G$ are assumed to be unknown, so the numerator of the correction factor must be approximately bounded with a form like (5-56). The same argument as used to obtain (5-55) may be used for the numerator written in the form $\mathbf{Q}'_a(\omega_{Ga}/b)^{\frac{1}{2}}\Omega_a(\omega_{Ga}/b)^{\frac{1}{2}}\mathbf{Q}_a$ (where $\omega_{Ga}$ is $\omega_G$ augmented with $\omega_p$) to obtain

$$V_{mxa} = \max_S(\sum_{i(s)} Q_{ai})^2 \tag{5-115}$$

Use of (5-92) to evaluate (5-115) yields

$$\left.\begin{aligned} V_{mxa} &= V_{mx} \text{ for a sum of positive values} \\ V_{mxa} &= V_{mx} + 2\omega_p^{-\frac{1}{2}}V_{mx}^{\frac{1}{2}} + \omega_p^{-1} \text{ for a sum of negative values} \end{aligned}\right\} \tag{5-116}$$

As an approximate bound

$$\frac{\mathbf{Q}'_a(\omega_{Ga}/b)^{\frac{1}{2}}\Omega_a(\omega_{Ga}/b)^{\frac{1}{2}}\mathbf{Q}_a}{\mathbf{Q}'\mathbf{Q} + \omega_p^{-1}} \le \frac{V_{mxa}}{\mathbf{Q}'\mathbf{Q} + \omega_p^{-1}} \tag{5-117}$$

As for (5-57), $\gamma_{Ia}\sigma_\varepsilon^2/b$ probably will not increase the width of a prediction interval significantly. Therefore (5-114) and (5-117) can be used to compute an approximate prediction interval by finding extreme values of

$$L(\theta,\upsilon,\lambda) = g(\gamma\theta) + \upsilon + \lambda\Big(\frac{V_{mxa}}{\mathbf{Q}'\mathbf{Q} + \omega_p^{-1}}t_{\alpha/2}^2(n-p) - \ell_a(\theta,\upsilon) + \ell_a(\hat{\theta},\hat{\upsilon})\Big) \tag{5-118}$$

where either (5-108) or (5-109) is used to compute $\ell_a(\theta,\upsilon)$ and $\ell_a(\hat{\theta},\hat{\upsilon})$. Note that because of the logarithmic form of $\ell_a(\theta,\upsilon)$, solution of (5-118) is always a nonlinear problem, even if $g(\gamma\theta)$ and $\mathbf{f}(\gamma\theta)$ are both linear. This also applies to (5-57).

## Summary of Principal Results

Results of section 4 indicate how accurate an average estimate of $\boldsymbol{\theta}_*$, $\mathbf{f}(\boldsymbol{\gamma\theta}_*)$, $\mathbf{f}(\boldsymbol{\beta})$, $g(\boldsymbol{\gamma\theta}_*)$, or $g(\boldsymbol{\beta})$ (or some future measurement of $g(\boldsymbol{\beta})$) might be, but do not indicate either precision of the estimates or how close a specific estimate might be to the value it estimates. These uncertainties are addressed in this section through confidence regions, confidence intervals, and prediction intervals. A joint confidence region for all parameters (referred to here simply as a confidence region) is a usually closed but possibly open region around $\hat{\boldsymbol{\theta}}$ that has a specified probability $1-\alpha$ of containing the true (as opposed to estimated) parameter set $\boldsymbol{\theta}_*$. As used here, it differs from a classical confidence region in that $\boldsymbol{\theta}_*$ is stochastic rather than fixed. A Scheffé-type confidence interval for $g(\boldsymbol{\gamma\theta}_*)$ is derived from a confidence region as the maximum and minimum limits of $g(\boldsymbol{\gamma\theta})$ over the confidence region. This interval is simultaneous in that $g(\boldsymbol{\gamma\theta}_*)$ lies within its Scheffé interval with probability $1-\alpha$ while all other linearizable functions of $\boldsymbol{\theta}_*$ lie within their Scheffé intervals with the same probability. (The limitation to linearizable functions eliminates pathologic functions of $\boldsymbol{\theta}_*$ to which the theory developed here does not apply.) An individual confidence interval for $g(\boldsymbol{\gamma\theta}_*)$ is a usually closed but possibly open interval around $g(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})$ that contains $g(\boldsymbol{\gamma\theta}_*)$ with specified probability $1-\alpha$. It is not simultaneous; it applies only to the selected function so that a fraction $\alpha$ of all individual confidence intervals for linearizable functions of $\boldsymbol{\theta}_*$ will not contain their respective functions of $\boldsymbol{\theta}_*$. Again, as used here, the individual confidence interval differs from the classical one in that $\boldsymbol{\theta}_*$ is stochastic rather than fixed. Finally, an individual prediction interval for some future observation $Y_p$ of $g(\boldsymbol{\beta})$ is a usually closed but possibly open interval around $g(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})$ that contains $Y_p$ with specified probability $1-\alpha$.

The approximate confidence region given by (5-2) is derived by a combination second-order Taylor series and perturbation method that formally assumes model and observation error variances to be small, with the model-error variance $Var(\mathbf{D}_\beta\mathbf{fe})$ being much smaller than the observation-error variance $Var(\boldsymbol{\varepsilon})$. However, a different method is used to show that approximate validity also holds when the variances are not small. The confidence region is defined using the standard upper $\alpha$ point of the $F$ random variable for the sum of squares ratio $((S(\boldsymbol{\theta}_*)-S(\hat{\boldsymbol{\theta}}))/p)/(S(\hat{\boldsymbol{\theta}})/(n-p))$ and applies for nonlinear models $\mathbf{f}(\boldsymbol{\gamma\theta})$ and $\mathbf{f}(\boldsymbol{\beta})$ and an arbitrary weight matrix $\boldsymbol{\omega}$. Equation (5-2) contains a correction factor $c_r$ that is defined by (5-3) and corrects for both $\boldsymbol{\omega}^{-1}$ not being proportional to $E(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma\theta}_*))(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma\theta}_*))'$ and nonzero model intrinsic nonlinearity. Matrix $E(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma\theta}_*))(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma\theta}_*))'$ would have to be known to exactly calculate $c_r$. An approximation given by (5-20) is used to illustrate the effect of correlations implied by the matrix. The resulting approximation for $c_r$ is given by (5-27) and uses a diagonal weight matrix $\hat{\boldsymbol{\omega}} \propto [E(Y_i - f_i(\boldsymbol{\gamma\theta}_*))^2]^{-1}$ and a constant effective spatial correlation $c$. The approximate form for $c_r$ shows that when positive spatial correlation from model error is significant, a confidence region and Scheffé interval would be too small without using $c_r$. This results primarily because $S(\hat{\boldsymbol{\theta}})$ is too small compared to $S(\boldsymbol{\theta}_*)$. Sum of squares $S(\hat{\boldsymbol{\theta}})$ does not measure the systematic variation of $\mathbf{Y}$ from $\mathbf{f}(\boldsymbol{\gamma\theta}_*)$ that is caused by the spatial correlation.

A Scheffé interval for $g(\gamma\theta_*)$ is computed by finding the limits of $g(\gamma\theta)$ over the confidence region. If there are no maxima or minima of $g(\gamma\theta)$ within the confidence region that are more extreme than those on the boundary, then the computation can be made using a Lagrange multiplier formulation given by (5-33) that yields limits of $g(\gamma\theta)$ on the boundary of the confidence region.

When the weight matrix is unknown, $\ell(\theta)$ should be used instead of $S(\theta)$ to describe the confidence region. An analysis shows that an approximate confidence region and Scheffé interval based on $\ell(\theta)$ can be constructed using the theory developed for the case where the weight matrix is known. The confidence region is given by (5-41), and the Lagrange multiplier formulation for a Scheffé interval is given by (5-42).

An individual confidence interval is computed using methods analogous to those used to compute a Scheffé interval. In this report a region defined by (5-44), analogous to a confidence region, is constructed from the ratio $(S(\tilde{\theta}) - S(\hat{\theta}))/(S(\hat{\theta})/(n-p))$, where $\tilde{\theta}$ is a regression estimate that is constrained so that $g(\gamma\tilde{\theta}) = g(\gamma\theta_*)$. The region is defined using the square of the standard upper $\alpha/2$ point of the Student $t$ random variable and a correction factor $c_c$ that is analogous to $c_r$. Maximum and minimum limits of $g(\gamma\theta)$ over the region, termed a likelihood region, form the confidence interval. Correction factor $c_c$, which is defined by (5-45), corrects for $\omega^{-1}$ not being proportional to $E(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'$, model intrinsic nonlinearity, and model combined intrinsic nonlinearity. Spatial correlation can cause $c_c$ to be large. However, $c_c$ also can be near unity (no correction) even if spatial correlation is large, as shown by the analysis following (5-50).

When the weight matrix is known, the Lagrange multiplier formulation for finding the limits of $g(\gamma\theta)$ is given by (5-51). When the weight matrix is unknown, the theory developed for when the weight matrix is known is again used to obtain an approximate likelihood region and Lagrange multiplier formulation based on $\ell(\theta)$ instead of $S(\theta)$. The Lagrange multiplier formulation is given by (5-57), which is written in terms of an approximate bound developed for the correction factor that is needed when the weight matrix is unknown.

Model and system types of combined intrinsic nonlinearity might be detected by analyzing the weighted residuals obtained from the constrained regression. The analysis is analogous to the analysis of standard weighted residuals used to detect model and system types of intrinsic nonlinearity. The components of the product of $\mathbf{R}$ and the weighted constrained residual vector $(\mathbf{I} - \mathbf{QQ}'/\mathbf{Q}'\mathbf{Q})\omega^{1/2}(\mathbf{Y} - \mathbf{f}(\gamma\tilde{\theta}))$ defined by (5-58) will not necessarily be approximately zero if the model combined intrinsic nonlinearity is significant, and the plot of the components of the weighted constrained residual vector in relation to components of the weighted function vector $\omega^{1/2}\mathbf{f}(\gamma\tilde{\theta})$ may exhibit a significantly nonzero slope if model and system types of intrinsic nonlinearity and model and system types of combined intrinsic nonlinearity are significant.

A prediction interval is derived using a regression that is constrained so that $g(\gamma\tilde{\theta}) + \tilde{\upsilon} = Y_p$, where $\tilde{\upsilon}$ is the predicted error and $(\tilde{\theta}, \tilde{\upsilon})$ is the constrained parameter vector. The constrained regression estimate is obtained by constrained minimization of a sum of squares function that is augmented by including terms resulting from the predicted error $\tilde{\upsilon}$. The augmented problem is completely analogous to the problem for individual confidence intervals except in the augmented case there is one more parameter and one more observation. Correction factor $c_p$, which is analogous to $c_c$, corrects for the possibility that $\mathbf{W}_a^{-1}$, the inverse of the augmented weight matrix, is not proportional to $E(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'$ (augmented to include products involving the predicted error) and for model intrinsic, and model combined intrinsic, nonlinearity. An approximation for $c_p$ is derived by defining $\boldsymbol{\omega}_a^{-1}$ as a block diagonal matrix with $\boldsymbol{\omega}$ and a general prediction weight $\omega_p$ forming the diagonal blocks. Analysis of $c_p$ as given by (5-96) shows that spatial correlation can cause $c_p$ to be large, but that, in general, $c_p$ should be less dependent on spatial correlation than $c_c$. A cross-validation method developed by Christensen and Cooley (1999b) can be used to test the possibility that one or more values of $c_p$ are too small.

When the weight matrix is known, the usual Lagrange multiplier formulation, augmented to include the predicted error, is used to compute limits of the prediction interval. This is given by (5-101). If the model intrinsic and model combined intrinsic types of nonlinearity are small, then (5-106), which has the form of a standard linear prediction interval, may be used instead of (5-101). When the weight matrix is unknown, augmented forms, $\ell_a(\theta,\upsilon)$, of $\ell(\theta)$ are used for the prediction interval. The form of $\ell_a(\theta,\upsilon)$ depends on whether the prediction is contained in one of the $q$ error groups or not. If it is, $\ell_a(\theta,\upsilon)$ is given by (5-108) and if it is not $\ell_a(\theta,\upsilon)$ is given by (5-109). Approximate theory based on the weight-matrix-known case is used to obtain the Lagrange multiplier formulation given by (5-118), which applies to both forms of $\ell_a(\theta,\upsilon)$.

# 6. Further Analysis of Intrinsic Nonlinearity and Combined Intrinsic Nonlinearity

As shown in Section 5, the $F_\alpha$ value for confidence regions, and the $t_{\alpha/2}$ value for confidence and prediction intervals, have to be adjusted with correction factors because the $F$ and $t$ distributions do not approximate the actual distributions of the pertinent variables well when $\omega^{-1} \neq \Omega$ and when model and system types of intrinsic nonlinearity and model and system types of combined intrinsic nonlinearity are significant. The correction factors are derived in appendix F using the combined Taylor series and perturbation method and are evaluated in appendix G. The correction factors for $\omega^{-1} \neq \Omega$ are further investigated without using the Taylor series and perturbation method in the last section of appendix F where the original assumptions for the method are shown to be overly restrictive if the correction factors are redefined in terms of $\Omega$ rather than $\mathbf{V}_*$. In the present section the analysis of the correction factors for model and system types of intrinsic nonlinearity and model and system types of combined intrinsic nonlinearity is continued.

## Analysis of Intrinsic Nonlinearity

***General forms for component correction factors.*** To begin the analysis, the residual vector is expanded as the sum of a vector of the form of the linear-model residual vector and two components that will be shown to be functions of model intrinsic nonlinearity:

$$\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}) = \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)) + \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}(\gamma\hat{\theta}))$$
$$+\,\omega^{-\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})) \qquad (6\text{-}1)$$

Use of linear model (5-104) labeled $\mathbf{f}_0(\gamma\theta)$, where $\theta = \theta_*$ and $\theta = \hat{\theta}$, facilitates showing that the second term on the right-hand side is a function of model intrinsic nonlinearity. With these relations and the relation $(\mathbf{I}-\mathbf{R})\omega^{1/2}\mathbf{Df} = \mathbf{0}$, (6-1) can be written

$$\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}) = \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)) + \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}(\gamma\hat{\theta})+\mathbf{f}_0(\gamma\hat{\theta}))$$
$$+\,\omega^{-\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})) \qquad (6\text{-}2)$$

In appendix I the last two terms on the right-hand side of (6-1) or (6-2) are shown to be zero if the model intrinsic nonlinearity is zero. The effects of system intrinsic nonlinearity are contained solely in the error vector $\omega^{-1/2}(\mathbf{I}-\mathbf{R})\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))$ in which $\mathbf{Y}-\mathbf{f}(\gamma\theta_*) = \mathbf{U}_* + \mathbf{d}$, where $\mathbf{d} = \mathbf{f}(\beta) - \mathbf{f}_0(\beta) - \mathbf{f}(\gamma\theta_*) + \mathbf{f}_0(\gamma\theta_*)$ ((F-103), appendix F). The intrinsic system nonlinearity is small if $\omega^{-1/2}(\mathbf{I}-\mathbf{R})\omega^{1/2}\mathbf{d}$ is small, which is shown in (F-104)-(F-108). These analyses are accomplished without using perturbation approximations. Equation (6-2) also is shown in appendix I to correspond to the perturbation form given by (B-12), appendix B.

Component correction factors $\hat{\gamma}_I\sigma_\varepsilon^4$ and $\gamma_I\sigma_\varepsilon^4$ for $c_r$ can be expressed in terms of $\mathbf{\Omega}$ rather than $\mathbf{V}_*$ by using (6-2) in (F-134) and (F-136), appendix F:

$$\begin{aligned}\hat{\gamma}_I\sigma_\varepsilon^4 &= E(S(\hat{\boldsymbol{\theta}}) - (\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*)))\\ &= 2E(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))\\ &+ E(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}))\\ &+ E(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))\end{aligned} \tag{6-3}$$

$$\begin{aligned}\gamma_I\sigma_\varepsilon^4 &= E(S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}}) - (\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*)))\\ &= -E(S(\hat{\boldsymbol{\theta}}) - (\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*)))\\ &= -\hat{\gamma}_I\sigma_\varepsilon^4\end{aligned} \tag{6-4}$$

Equations (6-3) and (6-4) substantiate that the component correction factors are a direct function of the degree of model intrinsic nonlinearity, but do not correct for system intrinsic nonlinearity.

***Approximations and approximate bounds for component correction factors.*** Analysis of the terms on the right-hand side of (6-3) would be very difficult in general. An approximate analysis can be made for Gauss-Markov estimation using the assumptions adopted for the perturbation analysis together with the additional assumption that $\boldsymbol{\omega}^{-1} = \mathbf{V}_* \approx \mathbf{\Omega}$. The approximation for $\hat{\gamma}_I\sigma_\varepsilon^4$ is given by (6-9) using the following development. In appendix I the three terms on the right-hand side of (6-3) are evaluated using the above assumptions as (6-5)-(6-7):

$$\begin{aligned}&2E(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))\\ &\approx -2\sum_i tr(\mathbf{C}_i^2)\sigma_\varepsilon^4 = -2\beta\sigma_\varepsilon^4\end{aligned} \tag{6-5}$$

$$\begin{aligned}&E(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}))\\ &\approx \frac{1}{4}\sum_i tr^2(\mathbf{C}_i)\sigma_\varepsilon^4 + \frac{1}{2}\sum_i tr(\mathbf{C}_i^2)\sigma_\varepsilon^4 = \frac{1}{4}\alpha\sigma_\varepsilon^4\end{aligned} \tag{6-6}$$

$$\begin{aligned}&E(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))\\ &\approx \sum_i tr(\mathbf{C}_i^2)\sigma_\varepsilon^4 = \beta\sigma_\varepsilon^4\end{aligned} \tag{6-7}$$

Variables $\alpha$ and $\beta$ are as defined by Johansen (1983, p. 183-184), and $\mathbf{C}_i$ is defined by (G-3), which is

$$\mathbf{C}_i = (\mathbf{I}-\mathbf{R})_i \sum_j \boldsymbol{\omega}_j^{\frac{1}{2}} \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}^2 f_j (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}} \tag{6-8}$$

From (6-3) and (6-5)-(6-7)

$$\begin{aligned}\hat{\gamma}_I\, \sigma_\varepsilon^4 &\approx \frac{1}{4}\Big(\sum_i tr^2(\mathbf{C}_i) - 2\sum_i tr(\mathbf{C}_i^2)\Big)\sigma_\varepsilon^4 \\ &= \frac{1}{4}(\alpha - 4\beta)\sigma_\varepsilon^4\end{aligned} \tag{6-9}$$

Following Johansen (1983, p. 184), bounds for $\hat{\gamma}_I$ can be developed from $\alpha$ and $\beta$. From (6-6) and (6-7) $\alpha = 2\beta + tr^2(\mathbf{C}_i)$ so that $\alpha \geq 2\beta$ or $\beta/\alpha \leq \frac{1}{2}$. Also, because $tr(\mathbf{C}_i^2) \geq 0$, $\beta/\alpha \geq 0$, so that $0 \leq \beta/\alpha \leq \frac{1}{2}$. Finally, because $\hat{\gamma}_I$ can be written as

$$\hat{\gamma}_I = \frac{\alpha}{4}\Big(1 - 4\frac{\beta}{\alpha}\Big) \tag{6-10}$$

it follows that

$$-\frac{\alpha}{4} \leq \hat{\gamma}_I \leq \frac{\alpha}{4} \tag{6-11}$$

Equations (6-11) and (I-12), appendix I, indicate that $\hat{\gamma}_I \sigma_\varepsilon^4$ is bounded by

$$\begin{aligned}&\pm\frac{1}{4}E\Big(\sum_i \mathbf{l}'_*\mathbf{D}^2 f_i \mathbf{l}_* \boldsymbol{\omega}_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}\mathbf{l}'_*\mathbf{D}^2 f_j \mathbf{l}_*\Big) \\ &\approx \pm E(\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}_0))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}_0))\end{aligned} \tag{6-12}$$

where $\hat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_* + \mathbf{l}_*$ and $\mathbf{l}_*$ is defined by (F-5). The ideas of (I-1)-(I-5) also allow the bounds to be written as

$$\begin{aligned}&\pm E(\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}_0))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}_0)) \\ &= \pm E(\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{Df}\hat{\boldsymbol{\psi}})'\boldsymbol{\omega}(\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{Df}\hat{\boldsymbol{\psi}})\end{aligned} \tag{6-13}$$

where

$$\hat{\boldsymbol{\psi}} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}(\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}_0) - \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}_0)) \tag{6-14}$$

The analysis of bounds and the development leading to (6-12)-(6-14) are very similar to the original development of the concepts of model intrinsic nonlinearity given by Beale (1960). In fact, Johansen (1983, p. 178) scaled (6-12) to define a measure of model intrinsic nonlinearity from Beale (1960, p. 59) as (in the notation used in this report)

$$N_{\min} = \frac{1}{4(p+2)\sigma_\varepsilon^2} E(\sum_i \mathbf{l}'_* \mathbf{D}^2 f_i \mathbf{l}_* \omega_i^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \sum_j \omega_j^{\frac{1}{2}} \mathbf{l}'_* \mathbf{D}^2 f_j \mathbf{l}_*) \tag{6-15}$$

where

$$(p+2)\sigma_\varepsilon^2 = E(\mathbf{l}'_* \mathbf{Df}'\omega\mathbf{Df}\mathbf{l}_*)^2 / p\sigma_\varepsilon^2 \tag{6-16}$$

This measure, which is directly proportional to the bounds, is a scaled measure of the sum of squared weighted discrepancies between linear $(\mathbf{f}_0(\gamma\theta(\hat{\phi}_0)))$ and nonlinear $(\mathbf{f}(\gamma\theta(\hat{\phi}_0)))$ model results when the models are written in terms of best-transformation parameters $\phi$. The reasons for the scaling $((p+2)\sigma_\varepsilon^2)$ are explained by Beale (1960, p. 54-59).

***Evaluation of approximate bounds.*** If $\hat{\theta}$ is a good estimate of $\theta_*$, then $\left|\mathbf{f}(\gamma\theta_*) - \mathbf{f}_0(\gamma\theta_*) - \mathbf{f}(\gamma\hat{\theta}) + \mathbf{f}_0(\gamma\hat{\theta})\right| < \left|\mathbf{f}(\gamma\hat{\theta}) - \mathbf{f}_0(\gamma\hat{\theta})\right|$ so that $\hat{\gamma}_I\sigma_\varepsilon^4$ as defined by (6-9) could be greater in magnitude than $\hat{\gamma}_I\sigma_\varepsilon^4$ as defined without the perturbation approximations by (6-3), assuming $\hat{\theta}$ is the same for both. In this case, the bounds for (6-9) given by (6-12) or (6-13) could bound (6-3) as well. If $\Omega$ and $\sigma_\varepsilon^2$ were known, then one possibility for computing (6-12) would be to evaluate it directly from (6-6). However, computer codes are not available for computing the required second derivative matrices for most ground-water models. Another possibility is to use Monte Carlo simulations of the right-hand side of (6-12) in which as an approximation $\mathbf{l}_* \sim N(\mathbf{0}, (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{1/2}(\omega/b)^{1/2}\Omega(\omega/b)^{1/2}\omega^{1/2}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}b\sigma_\varepsilon^2)$. In most cases $b\sigma_\varepsilon^2$ is unknown, so that its estimate $s^2$ must be substituted in the calculations. Similarly, $\theta_*$ also is unknown, so that the estimate $\hat{\theta}$ must be substituted for $\theta_*$, and $\hat{\theta}_0$ must now be computed as $\hat{\theta} + \mathbf{l}_*$, where $\hat{\theta}$ is fixed during the Monte Carlo simulations. In this case, the Monte Carlo method is a type of percentile bootstrap method (Efron, 1982, p. 78-84). If $\Omega$ is unknown, then (6-12) cannot be approximated using the correct distribution for $\hat{\theta}_0$. In this case the importance of intrinsic nonlinearity is gaged by an empirical measure described later in this section.

## Analysis of Combined Intrinsic Nonlinearity for Confidence Intervals

***General forms for component correction factors.*** As before, expansion of the constrained residual vector yields an identity, the right-hand side of which is the sum of a vector having the form of the linear-model constrained residual vector and two components that will be shown to be functions of model combined intrinsic nonlinearity:

$$\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}) = \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)) + \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}(\gamma\tilde{\theta}))$$
$$+\omega^{-\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta})) \tag{6-17}$$

The second term on the right-hand side is a function of model combined intrinsic nonlinearity. The equation needed to show this is obtained by expanding (6-17) using the idea used to obtain (6-2), along with the constraint $g(\gamma\theta_*) = g(\gamma\tilde{\theta})$ and the relation

$$\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}_0(\gamma\tilde{\theta})) = \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{Df}(\theta_*-\tilde{\theta})$$
$$= \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g(\theta_*-\tilde{\theta}) = \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g_0(\gamma\theta_*)-g_0(\gamma\tilde{\theta})) \tag{6-18}$$

where $g_0(\gamma\theta)$ is linear model (5-105). The equation needed is

$$\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}) = \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)) + \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)$$
$$-\mathbf{f}(\gamma\tilde{\theta})+\mathbf{f}_0(\gamma\tilde{\theta})) + \omega^{-\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta})) + \omega^{-\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}(\gamma\tilde{\theta}))$$
$$-\omega^{-\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}_0(\gamma\tilde{\theta})) + \omega^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g_0(\gamma\theta_*)-g_0(\gamma\tilde{\theta}))$$
$$-\omega^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\theta_*)-g(\gamma\tilde{\theta}))$$
$$= \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)) + \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}(\gamma\tilde{\theta})+\mathbf{f}_0(\gamma\tilde{\theta}))$$
$$-\omega^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\theta_*)-g_0(\gamma\theta_*)-g(\gamma\tilde{\theta})+g_0(\gamma\tilde{\theta})) + \omega^{-\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta})) \tag{6-19}$$

In appendix I all terms on the right-hand side of (6-17) or (6-19) except the first term are shown to be zero if the model combined intrinsic nonlinearity is zero. This is accomplished using an extension of the argument used for (6-2). As before, the system intrinsic nonlinearity is contained in the error vector $\omega^{-1/2}(\mathbf{I}-\mathbf{R})\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))$. If $\omega^{-1/2}(\mathbf{I}-\mathbf{R})\omega^{1/2}\mathbf{d}$ is small, the system intrinsic nonlinearity is small. Equation (6-19) also is shown in appendix I to correspond to the perturbation form (E-29), appendix E.

The component correction factor $\gamma_I\sigma_\varepsilon^4$ pertaining to an individual confidence interval can be expressed in terms of $\mathbf{\Omega}$ by using (6-19) in (F-147) to obtain

$$
\begin{aligned}
\gamma_I\sigma_\varepsilon^4 &= E(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}) - (\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))) \\
&= E(S(\tilde{\boldsymbol{\theta}}) - (\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))) - \hat{\gamma}_I\sigma_\varepsilon^4 \\
&= 2E((\mathbf{f}(\gamma\boldsymbol{\theta}_*)-\mathbf{f}_0(\gamma\boldsymbol{\theta}_*)-\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}})+\mathbf{f}_0(\gamma\tilde{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}} - (g(\gamma\boldsymbol{\theta}_*)-g_0(\gamma\boldsymbol{\theta}_*)-g(\gamma\tilde{\boldsymbol{\theta}}) \\
&\quad + g_0(\gamma\tilde{\boldsymbol{\theta}}))\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*)) + E((\mathbf{f}(\gamma\boldsymbol{\theta}_*)-\mathbf{f}_0(\gamma\boldsymbol{\theta}_*)-\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}}) \\
&\quad + \mathbf{f}_0(\gamma\tilde{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}} - (g(\gamma\boldsymbol{\theta}_*)-g_0(\gamma\boldsymbol{\theta}_*)-g(\gamma\tilde{\boldsymbol{\theta}})+g_0(\gamma\tilde{\boldsymbol{\theta}}))\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}) \\
&\quad \bullet(\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\boldsymbol{\theta}_*)-\mathbf{f}_0(\gamma\boldsymbol{\theta}_*)-\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}})+\mathbf{f}_0(\gamma\tilde{\boldsymbol{\theta}})) - \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\boldsymbol{\theta}_*)-g_0(\gamma\boldsymbol{\theta}_*)-g(\gamma\tilde{\boldsymbol{\theta}}) \\
&\quad + g_0(\gamma\tilde{\boldsymbol{\theta}}))) + E(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}})) - \hat{\gamma}_I\sigma_\varepsilon^4
\end{aligned}
\tag{6-20}
$$

Equation (6-20) substantiates that the component correction factor is a direct function of the degree of model combined intrinsic nonlinearity, but does not correct for system combined intrinsic nonlinearity.

***Approximations and approximate bounds for component correction factors.*** As before, an approximate analysis for Gauss-Markov estimation uses the assumptions from perturbation analysis and the additional assumption that $\boldsymbol{\omega}^{-1} = \mathbf{V}_* \approx \boldsymbol{\Omega}$. The approximation for $\gamma_I\sigma_\varepsilon^4$ is given by (6-27) using the following development. In appendix I the three terms on the right-hand side of (6-20) are evaluated as (6-21)-(6-23):

$$
\begin{aligned}
&2E((\mathbf{f}(\gamma\boldsymbol{\theta}_*)-\mathbf{f}_0(\gamma\boldsymbol{\theta}_*)-\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}})+\mathbf{f}_0(\gamma\tilde{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}} - (g(\gamma\boldsymbol{\theta}_*)-g_0(\gamma\boldsymbol{\theta}_*)-g(\gamma\tilde{\boldsymbol{\theta}}) \\
&+ g_0(\gamma\tilde{\boldsymbol{\theta}}))\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*)) \\
&\approx -2\sum_i tr(\tilde{\mathbf{C}}_i^2)\sigma_\varepsilon^4 + 2E(\sum_i \boldsymbol{\omega}_i^{\frac{1}{2}}\tilde{\mathbf{Z}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Df} \\
&\bullet(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}_*\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{U}_*)
\end{aligned}
\tag{6-21}
$$

$$
\begin{aligned}
&E((\mathbf{f}(\gamma\boldsymbol{\theta}_*)-\mathbf{f}_0(\gamma\boldsymbol{\theta}_*)-\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}})+\mathbf{f}_0(\gamma\tilde{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}} - (g(\gamma\boldsymbol{\theta}_*)-g_0(\gamma\boldsymbol{\theta}_*)-g(\gamma\tilde{\boldsymbol{\theta}})+g_0(\gamma\tilde{\boldsymbol{\theta}}))\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}-\mathbf{R} \\
&+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})(\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\boldsymbol{\theta}_*)-\mathbf{f}_0(\gamma\boldsymbol{\theta}_*)-\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}})+\mathbf{f}_0(\gamma\tilde{\boldsymbol{\theta}})) - \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\boldsymbol{\theta}_*)-g_0(\gamma\boldsymbol{\theta}_*)-g(\gamma\tilde{\boldsymbol{\theta}})+g_0(\gamma\tilde{\boldsymbol{\theta}}))) \\
&\approx \frac{1}{4}\sum_i tr^2(\tilde{\mathbf{C}}_i)\sigma_\varepsilon^4 + \frac{1}{2}\sum_i tr(\tilde{\mathbf{C}}_i^2)\sigma_\varepsilon^4 - \frac{1}{2}\sum_i tr(\tilde{\mathbf{C}}_i)tr(\tilde{\mathbf{F}}_i)\sigma_\varepsilon^4 - \sum_i tr(\tilde{\mathbf{C}}_i\tilde{\mathbf{F}}_i)\sigma_\varepsilon^4
\end{aligned}
$$

$$+\frac{1}{4}\sum_i tr^2(\widetilde{\mathbf{F}}_i)\sigma_\varepsilon^4+\frac{1}{2}\sum_i tr(\widetilde{\mathbf{F}}_i^2)\sigma_\varepsilon^4=\frac{1}{4}\sum_i tr^2(\widetilde{\mathbf{C}}_i-\widetilde{\mathbf{F}}_i)\sigma_\varepsilon^4+\frac{1}{2}\sum_i tr((\widetilde{\mathbf{C}}_i-\widetilde{\mathbf{F}}_i)^2)\sigma_\varepsilon^4 \tag{6-22}$$

$$\begin{aligned}&E(\mathbf{Y}-\mathbf{f}(\gamma\widetilde{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\widetilde{\boldsymbol{\theta}}))\\&\approx\sum_i tr(\widetilde{\mathbf{C}}_i^2)\sigma_\varepsilon^4+\sum_i tr(\widetilde{\mathbf{F}}_i^2)\sigma_\varepsilon^4\\&-2E(\sum_i\boldsymbol{\omega}_i^{\frac{1}{2}}\widetilde{\mathbf{Z}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i\,(\mathbf{Df'\omega\,Df})^{-1}\mathbf{Df'}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Df}\,(\mathbf{Df'\omega\,Df})^{-1}\mathbf{D}^2 g\,\widetilde{\mathbf{l}}_*\frac{1}{\mathbf{Q'Q}}\mathbf{Q'}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*)\end{aligned} \tag{6-23}$$

where $\widetilde{\mathbf{C}}_i$, $\widetilde{\mathbf{F}}_i$, and $\widetilde{\mathbf{A}}$ are defined by (G-10), (G-12), and (G-13), appendix G, respectively, which are (6-24)-(6-26):

$$\widetilde{\mathbf{C}}_i=(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})_i\sum_j\boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Df}\,(\mathbf{Df'\omega\,Df})^{-1}\mathbf{D}^2 f_j\,(\mathbf{Df'\omega\,Df})^{-1}\mathbf{Df'}\,\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \tag{6-24}$$

$$\widetilde{\mathbf{F}}_i=\frac{Q_i}{\mathbf{Q'Q}}\widetilde{\mathbf{A}} \tag{6-25}$$

$$\widetilde{\mathbf{A}}=(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Df}\,(\mathbf{Df'\omega\,Df})^{-1}\mathbf{D}^2 g(\mathbf{Df'\omega\,Df})^{-1}\mathbf{Df'}\,\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \tag{6-26}$$

Substitution of (6-21)-(6-23) into (6-20) yields

$$\begin{aligned}&\gamma_I\sigma_\varepsilon^4\approx\frac{1}{4}\sum_i tr^2(\widetilde{\mathbf{C}}_i-\widetilde{\mathbf{F}}_i)\sigma_\varepsilon^4+\frac{1}{2}\sum_i tr((\widetilde{\mathbf{C}}_i-\widetilde{\mathbf{F}}_i)^2)\sigma_\varepsilon^4\\&-\sum_i tr(\widetilde{\mathbf{C}}_i^2)\sigma_\varepsilon^4+\sum_i tr(\widetilde{\mathbf{F}}_i^2)\sigma_\varepsilon^4-\hat{\gamma}_I\sigma_\varepsilon^4\end{aligned} \tag{6-27}$$

To form general bounds for $(\gamma_I+\hat{\gamma}_I)\sigma_\varepsilon^4$, let

$$\widetilde{\alpha}=\sum_i tr^2(\widetilde{\mathbf{C}}_i-\widetilde{\mathbf{F}}_i)+2\sum_i tr((\widetilde{\mathbf{C}}_i-\widetilde{\mathbf{F}}_i)^2) \tag{6-28}$$

Then

$$\gamma_I+\hat{\gamma}_I\le\frac{\widetilde{\alpha}}{4}+\frac{1}{2}\sum_i tr^2(\widetilde{\mathbf{F}}_i)+\sum_i tr(\widetilde{\mathbf{F}}_i^2) \tag{6-29}$$

and

$$\gamma_I+\hat{\gamma}_I\ge\frac{\widetilde{\alpha}}{4}-\frac{1}{2}\sum_i tr^2(\widetilde{\mathbf{C}}_i)-\sum_i tr(\widetilde{\mathbf{C}}_i^2) \tag{6-30}$$

When $g(\gamma\theta)$ is linear, as, for example, when the confidence interval is for a parameter, $\widetilde{\mathbf{A}} = \mathbf{0}$ and (6-29) and (6-30) form the bounds $\pm\widetilde{\alpha}/4$. These are of the same form as the bounds for (6-9). When the model combined intrinsic nonlinearity is very small so that $\widetilde{\mathbf{C}}_i \approx \widetilde{\mathbf{F}}_i$, the bounds given by (6-29) and (6-30) may be very wide.

Equation (6-29) indicates that the upper bound of $(\gamma_I + \hat{\gamma}_I)\sigma_\varepsilon^4$ is

$$(\gamma_I + \hat{\gamma}_I)\sigma_\varepsilon^4 \le \frac{1}{4}E(\sum_i \widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \widetilde{\mathbf{l}}_* \,\omega_i^{\frac{1}{2}} - \widetilde{\mathbf{l}}_*'\mathbf{D}^2 g \widetilde{\mathbf{l}}_* \frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I} - \mathbf{R} + \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\sum_j \omega_j^{\frac{1}{2}} \widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \widetilde{\mathbf{l}}_* - \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g \widetilde{\mathbf{l}}_*)$$
$$+\frac{1}{2}\frac{1}{\mathbf{Q}'\mathbf{Q}}E(\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g \widetilde{\mathbf{l}}_*)^2$$
$$\approx E((\mathbf{f}(\gamma\widetilde{\theta}_0) - \mathbf{f}_0(\gamma\widetilde{\theta}_0))'\omega^{\frac{1}{2}} - (g(\gamma\widetilde{\theta}_0) - g_0(\gamma\widetilde{\theta}_0))\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I} - \mathbf{R} + \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\widetilde{\theta}_0)$$
$$-\mathbf{f}_0(\gamma\widetilde{\theta}_0)) - \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\widetilde{\theta}_0) - g_0(\gamma\widetilde{\theta}_0))) + \frac{2}{\mathbf{Q}'\mathbf{Q}}E(g(\gamma\widetilde{\theta}_0) - g_0(\gamma\widetilde{\theta}_0))^2 \tag{6-31}$$

which also can be written in the form

$$(\gamma_I + \hat{\gamma}_I)\sigma_\varepsilon^4 \le E(\mathbf{f}(\gamma\widetilde{\theta}_0) - \mathbf{f}_0(\gamma\widetilde{\theta}_0) - \mathbf{D}\mathbf{f}\widetilde{\psi})'\omega(\mathbf{f}(\gamma\widetilde{\theta}_0) - \mathbf{f}_0(\gamma\widetilde{\theta}_0) - \mathbf{D}\mathbf{f}\widetilde{\psi})$$
$$+\frac{2}{\mathbf{Q}'\mathbf{Q}}E(g(\gamma\widetilde{\theta}_0) - g_0(\gamma\widetilde{\theta}_0))^2 \tag{6-32}$$

where $\widetilde{\theta}_0 = \theta_* + \widetilde{\mathbf{l}}_*$, in which $\widetilde{\mathbf{l}}_*$ is defined by (F-7), and

$$\widetilde{\psi} = -\left(\frac{(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}\mathbf{D}g'\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}} - (\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}\mathbf{D}\mathbf{f}'\omega^{\frac{1}{2}}\right)\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\widetilde{\theta}_0) - \mathbf{f}_0(\gamma\widetilde{\theta}_0))$$
$$+\frac{(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}\mathbf{D}g'}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\widetilde{\theta}_0) - g_0(\gamma\widetilde{\theta}_0)) \tag{6-33}$$

as can be verified by substituting (6-33) into (6-32). Equation (6-30) indicates that the lower bound of $(\gamma_I + \hat{\gamma}_I)\sigma_\varepsilon^4$ is

$$(\gamma_I + \hat{\gamma}_I)\sigma_\varepsilon^4 \ge \frac{1}{4}E(\sum_i \widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \widetilde{\mathbf{l}}_* \,\omega_i^{\frac{1}{2}} - \widetilde{\mathbf{l}}_*'\mathbf{D}^2 g \widetilde{\mathbf{l}}_* \frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I} - \mathbf{R} + \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\sum_j \omega_j^{\frac{1}{2}} \widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \widetilde{\mathbf{l}}_* - \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g \widetilde{\mathbf{l}}_*)$$
$$-\frac{1}{2}E(\sum_i \widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \widetilde{\mathbf{l}}_* \,\omega_i^{\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}} \widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \widetilde{\mathbf{l}}_*)$$
$$\approx E((\mathbf{f}(\gamma\widetilde{\theta}_0) - \mathbf{f}_0(\gamma\widetilde{\theta}_0))'\omega^{\frac{1}{2}} - (g(\gamma\widetilde{\theta}_0) - g_0(\gamma\widetilde{\theta}_0))\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I} - \mathbf{R} + \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\widetilde{\theta}_0)$$
$$-\mathbf{f}_0(\gamma\widetilde{\theta}_0)) - \frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\widetilde{\theta}_0) - g_0(\gamma\widetilde{\theta}_0))) - 2E(\mathbf{f}(\gamma\widetilde{\theta}_0) - (\mathbf{f}_0(\gamma\widetilde{\theta}_0))'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}$$

$$\bullet(\mathbf{f}(\gamma\tilde{\theta}_0)-\mathbf{f}_0(\gamma\tilde{\theta}_0)) \tag{6-34}$$

or

$$\begin{aligned}(\gamma_I+\hat{\gamma}_I)\sigma_\varepsilon^4 \geq\ & E(\mathbf{f}(\gamma\tilde{\theta}_0)-\mathbf{f}_0(\gamma\tilde{\theta}_0)-\mathbf{D}\mathbf{f}\tilde{\psi})'\omega(\mathbf{f}(\gamma\tilde{\theta}_0)-\mathbf{f}_0(\gamma\tilde{\theta}_0)-\mathbf{D}\mathbf{f}\tilde{\psi})\\ &-2E(\mathbf{f}(\gamma\tilde{\theta}_0)-\mathbf{f}_0(\gamma\tilde{\theta}_0)-\mathbf{D}\mathbf{f}\tilde{\psi}_0)'\omega(\mathbf{f}(\gamma\tilde{\theta}_0)-\mathbf{f}_0(\gamma\tilde{\theta}_0)-\mathbf{D}\mathbf{f}\tilde{\psi}_0)\end{aligned} \tag{6-35}$$

where

$$\tilde{\psi}_0=-\left(\frac{(\mathbf{D}\mathbf{f}'\omega\,\mathbf{D}\mathbf{f})^{-1}\mathbf{D}g'\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}-(\mathbf{D}\mathbf{f}'\omega\,\mathbf{D}\mathbf{f})^{-1}\mathbf{D}\mathbf{f}'\omega^{\frac{1}{2}}\right)\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\tilde{\theta}_0)-\mathbf{f}_0(\gamma\tilde{\theta}_0)) \tag{6-36}$$

as can be verified by substituting (6-36) into (6-35). If $\Omega$ were known, the bounds could be obtained by Monte Carlo simulation similar to the method outlined for (6-12).

## Analysis of Combined Intrinsic Nonlinearity for Prediction Intervals

***General forms for component correction factors.*** Prediction intervals are analyzed in the same way as were confidence intervals. The final form for the augmented constrained residual vector is given by (6-41) as shown by the following development. First, expansion of the augmented residual vector analogously to (6-1) yields

$$\begin{aligned}\mathbf{Y}_a-\mathbf{f}_a(\gamma\hat{\theta},\hat{\theta}_p)=\ & \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a-\mathbf{f}_a(\gamma\theta_*,\theta_p^*))+\mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\theta_*,\theta_p^*)\\ &-\mathbf{f}_a(\gamma\hat{\theta},\hat{\theta}_p))+\mathbf{W}_a^{-\frac{1}{2}}\mathbf{R}_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a-\mathbf{f}_a(\gamma\hat{\theta},\hat{\theta}_p))\end{aligned} \tag{6-37}$$

or

$$\begin{aligned}\mathbf{Y}_a-\mathbf{f}_a(\gamma\hat{\theta},\hat{\theta}_p)=\ & \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a-\mathbf{f}_a(\gamma\theta_*,\theta_p^*))+\mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\theta_*,\theta_p^*)\\ &-\mathbf{f}_{0a}(\gamma\theta_*,\theta_p^*)-\mathbf{f}_a(\gamma\hat{\theta},\hat{\theta}_p)+\mathbf{f}_{0a}(\gamma\hat{\theta},\hat{\theta}_p))+\mathbf{W}_a^{-\frac{1}{2}}\mathbf{R}_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a-\mathbf{f}_a(\gamma\hat{\theta},\hat{\theta}_p))\end{aligned} \tag{6-38}$$

All terms on the right-hand side of (6-37) or (6-38) except the first are zero if model intrinsic nonlinearity is zero as can be shown using an analysis exactly analogous to the one used for (6-1) or (6-2). Also, the effects of system intrinsic nonlinearity are contained in the first term on the right-hand side of (6-38). Finally, the expansion given by (6-38) corresponds to the expansion obtained by perturbation analysis. All of these results are so similar to the ones obtained previously that they are not elaborated further here.

Next, expansion of the augmented, constrained residual vector analogously to (6-17) produces

$$\mathbf{Y}_a - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) = \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*)) + \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a$$
$$+ \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) + \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) \qquad (6\text{-}39)$$

Then, use of the relation

$$\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_{0a}(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) = \frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{D}_a h(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a(\boldsymbol{\theta}_{*a} - \tilde{\boldsymbol{\theta}}_a)$$
$$= \frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{D}_a h(\boldsymbol{\theta}_{*a} - \tilde{\boldsymbol{\theta}}_a) = \frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(\mathbf{D}g(\boldsymbol{\theta}_* - \tilde{\boldsymbol{\theta}}) - \theta_p^* + \tilde{\theta}_p)$$
$$= \frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(g_0(\gamma\boldsymbol{\theta}_*) - g_0(\gamma\tilde{\boldsymbol{\theta}}) - \theta_p^* + \tilde{\theta}_p) \qquad (6\text{-}40)$$

and the constraint $g(\gamma\boldsymbol{\theta}_*) + \upsilon_* = g(\gamma\tilde{\boldsymbol{\theta}}) + \tilde{\upsilon}$ in (6-39) shows that

$$\mathbf{Y}_a - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) = \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*)) + \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*)$$
$$- \mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) + \mathbf{f}_{0a}(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) + \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p))$$
$$+ \mathbf{W}_a^{-\frac{1}{2}}\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) - \mathbf{W}_a^{-\frac{1}{2}}\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*)$$
$$- \mathbf{f}_{0a}(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) + \mathbf{W}_a^{-\frac{1}{2}}\frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(g_0(\gamma\boldsymbol{\theta}_*) - g_0(\gamma\tilde{\boldsymbol{\theta}}) - \theta_p^* + \tilde{\theta}_p)$$
$$- \mathbf{W}_a^{-\frac{1}{2}}\frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(g(\gamma\boldsymbol{\theta}_*) + \upsilon_* - g(\gamma\tilde{\boldsymbol{\theta}}) - \tilde{\upsilon})$$
$$= \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*)) + \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*)$$
$$- \mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) + \mathbf{f}_{0a}(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) - \mathbf{W}_a^{-\frac{1}{2}}\frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(g(\gamma\boldsymbol{\theta}_*) - g_0(\gamma\boldsymbol{\theta}_*) - g(\gamma\tilde{\boldsymbol{\theta}})$$
$$+ g_0(\gamma\tilde{\boldsymbol{\theta}})) + \mathbf{W}_a^{-\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) \qquad (6\text{-}41)$$

where the relations $\upsilon_* = Y_p - \theta_p^*$ and $\tilde{\upsilon} = Y_p - \tilde{\theta}_p$ were used. As before, all terms on the right-hand side of (6-39) or (6-41) except the first are zero if the model combined intrinsic nonlinearity is zero. The method of showing this is the same as used for (6-17) or (6-19) and is not given in this report. The expansion of $\mathbf{Y}_a - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)$ given by (6-41) corresponds to the expansion obtained by the perturbation method.

The component correction factors $\hat{\gamma}_{1a}\sigma_\varepsilon^4$ and $\gamma_{1a}\sigma_\varepsilon^4$ can be expressed analogously to (6-3) and (6-20) as

$$\begin{aligned}
\hat{\gamma}_{1a}\sigma_\varepsilon^4 &= E(S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p) - (\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*))'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*))) \\
&= 2E(\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\hat{\boldsymbol{\theta}}, \hat{\theta}_p) + \mathbf{f}_{0a}(\gamma\hat{\boldsymbol{\theta}}, \hat{\theta}_p))'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a \\
&\quad - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*)) + E(\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\hat{\boldsymbol{\theta}}, \hat{\theta}_p) + \mathbf{f}_{0a}(\gamma\hat{\boldsymbol{\theta}}, \hat{\theta}_p))\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a \\
&\quad - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\hat{\boldsymbol{\theta}}, \hat{\theta}_p) + \mathbf{f}_{0a}(\gamma\hat{\boldsymbol{\theta}}, \hat{\theta}_p)) \\
&\quad + E(\mathbf{Y}_a - \mathbf{f}_a(\gamma\hat{\boldsymbol{\theta}}, \hat{\theta}_p))'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\hat{\boldsymbol{\theta}}, \hat{\theta}_p))
\end{aligned} \tag{6-42}$$

and

$$\begin{aligned}
\gamma_{1a}\sigma_\varepsilon^4 &= E(S_a(\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) - S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p) - (\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*))'\mathbf{W}_a^{\frac{1}{2}}\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*))) \\
&= E(S_a(\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) - (\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*))'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*))) \\
&\quad - \hat{\gamma}_{1a}\sigma_\varepsilon^4 \\
&= 2E((\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) + \mathbf{f}_{0a}(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p))'\mathbf{W}_a^{\frac{1}{2}} - (g(\gamma\boldsymbol{\theta}_*) - g_0(\gamma\boldsymbol{\theta}_*) \\
&\quad - g(\gamma\tilde{\boldsymbol{\theta}}) + g_0(\gamma\tilde{\boldsymbol{\theta}}))\frac{\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*)) + E((\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*) \\
&\quad - \mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) + \mathbf{f}_{0a}(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p))'\mathbf{W}_a^{\frac{1}{2}} - (g(\gamma\boldsymbol{\theta}_*) - g_0(\gamma\boldsymbol{\theta}_*) - g(\gamma\tilde{\boldsymbol{\theta}}) - g_0(\gamma\tilde{\boldsymbol{\theta}})) \\
&\quad \bullet \frac{\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})(\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_{0a}(\gamma\boldsymbol{\theta}_*, \theta_p^*) - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) + \mathbf{f}_{0a}(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) \\
&\quad - \frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(g(\gamma\boldsymbol{\theta}_*) - g_0(\gamma\boldsymbol{\theta}_*) - g(\gamma\tilde{\boldsymbol{\theta}}) - g_0(\gamma\tilde{\boldsymbol{\theta}}))) + E(\mathbf{Y}_a - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p))'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a}) \\
&\quad \bullet \mathbf{W}_a^{\frac{1}{2}}(\mathbf{Y}_a - \mathbf{f}_a(\gamma\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p)) - \hat{\gamma}_{1a}\sigma_\varepsilon^4
\end{aligned} \tag{6-43}$$

***Approximations and approximate bounds for component correction factors.*** The assumptions adopted for the perturbation analysis and the assumption that $\mathbf{W}_a^{-1} = \mathbf{V}_{*a} \approx \boldsymbol{\Omega}_a$ allow an idealized analysis of the same type as that leading to (6-13) to conclude that, for Gauss-Markov estimation, $\hat{\gamma}_{1a}\sigma_\varepsilon^4$ is bounded by

$$\pm E(\mathbf{f}_a(\gamma\hat{\theta}_0,\hat{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\hat{\theta}_0,\hat{\theta}_p^0))'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\hat{\theta}_0,\hat{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\hat{\theta}_0,\hat{\theta}_p^0))$$
$$=\pm E(\mathbf{f}_a(\gamma\hat{\theta}_0,\hat{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\hat{\theta}_0,\hat{\theta}_p^0)-\mathbf{D}_a\mathbf{f}_a\hat{\psi}_a)'\mathbf{W}_a(\mathbf{f}_a(\gamma\hat{\theta}_0,\hat{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\hat{\theta}_0,\hat{\theta}_p^0)-\mathbf{D}_a\mathbf{f}_a\hat{\psi}_a) \quad (6\text{-}44)$$

where $[\hat{\theta}_0'\ \hat{\theta}_p^0]=\hat{\theta}_{0a}'=\hat{\theta}_{*a}'+\mathbf{l}_{*a}'$ and

$$\hat{\psi}_a=(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a(\mathbf{f}_a(\gamma\hat{\theta}_0,\hat{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\hat{\theta}_0,\hat{\theta}_p^0)) \quad (6\text{-}45)$$

Similarly, the upper bound for $(\gamma_{Ia}+\hat{\gamma}_{Ia})\sigma_\varepsilon^4$ is analogous to (6-31) and (6-32):

$$(\gamma_{Ia}+\hat{\gamma}_{Ia})\sigma_\varepsilon^4\le E((\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0))'\mathbf{W}_a^{\frac{1}{2}}-(g(\gamma\tilde{\theta}_0)-g_0(\gamma\tilde{\theta}_0))\frac{\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})(\mathbf{I}_a-\mathbf{R}_a$$
$$+\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})(\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0))-\frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(g(\gamma\tilde{\theta}_0)-g_0(\gamma\tilde{\theta}_0)))$$
$$+\frac{2}{\mathbf{Q}_a'\mathbf{Q}_a}E(g(\gamma\tilde{\theta}_0)-g_0(\gamma\tilde{\theta}_0))^2$$
$$=E(\mathbf{f}_a(\gamma\tilde{\theta}_0,\hat{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{D}_a\mathbf{f}_a\tilde{\psi}_a)'\mathbf{W}_a(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{D}_a\mathbf{f}_a\tilde{\psi}_a)$$
$$+\frac{2}{\mathbf{Q}_a'\mathbf{Q}_a}E(g(\gamma\tilde{\theta}_0)-g_0(\gamma\tilde{\theta}_0))^2 \quad (6\text{-}46)$$

where $[\tilde{\theta}_0'\ \tilde{\theta}_p^0]=\tilde{\theta}_{0a}'=\theta_{*a}'+\tilde{\mathbf{l}}_{*a}'$ and

$$\tilde{\psi}_a=-\left(\frac{(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h'\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a}-(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a^{\frac{1}{2}}\right)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)$$
$$-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0))+\frac{(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h'}{\mathbf{Q}_a'\mathbf{Q}_a}(g(\gamma\tilde{\theta}_0)-g_0(\gamma\tilde{\theta}_0)) \quad (6\text{-}47)$$

Finally, the lower bound is analogous to (6-34) and (6-35):

$$(\gamma_{Ia}+\hat{\gamma}_{Ia})\sigma_\varepsilon^4\ge E((\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0))'\mathbf{W}_a^{\frac{1}{2}}-(g(\gamma\tilde{\theta}_0)-g_0(\gamma\tilde{\theta}_0))\frac{\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})(\mathbf{I}_a-\mathbf{R}_a$$
$$+\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})(\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0))-\frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(g(\gamma\tilde{\theta}_0)-g_0(\gamma\tilde{\theta}_0)))$$
$$-2E(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0))'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0))$$
$$=E(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{D}_a\mathbf{f}_a\tilde{\psi}_a)'\mathbf{W}_a(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{D}_a\mathbf{f}_a\tilde{\psi}_a)$$
$$-2E(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{D}_a\mathbf{f}_a\tilde{\psi}_{0a})'\mathbf{W}_a(\mathbf{f}_a(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)-\mathbf{f}_{0a}(\gamma\tilde{\theta}_0,\tilde{\theta}_p^0)$$

$$-\mathbf{D}_a\mathbf{f}_a\widetilde{\psi}_{0a}) \tag{6-48}$$

where

$$\widetilde{\psi}_{0a} = -\left(\frac{(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h'\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a} - (\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a^{\frac{1}{2}}\right)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{f}_a(\gamma\widetilde{\theta}_0, \widetilde{\theta}_p^0)$$
$$-\mathbf{f}_{0a}(\gamma\widetilde{\theta}_0, \widetilde{\theta}_p^0)) \tag{6-49}$$

If the elements corresponding to the covariances in $\Omega_a$ are neglected when forming the weight matrix, $\mathbf{W}_a = \boldsymbol{\omega}_a$ so that (6-44)-(6-49) can be expressed in terms of unaugmented variables using (E-38)-(E-40), (E-48), (E-56), appendix E, (5-89), (5-91), (5-92), and (5-94). The results are that (6-44) and (6-45) become (6-13) and (6-14), respectively, and that (6-46)-(6-49) become

$$(\gamma_{Ia}+\hat{\gamma}_{Ia})\sigma_\varepsilon^4 \le E((\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0)-\mathbf{Df}\widetilde{\psi})'\boldsymbol{\omega}(\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0)-\mathbf{Df}\widetilde{\psi})$$
$$+\widetilde{\psi}_p\omega_p\widetilde{\psi}_p)+\frac{2}{\mathbf{Q}'\mathbf{Q}+\omega_p^{-1}}E(g(\gamma\widetilde{\theta}_0)-g_0(\gamma\widetilde{\theta}_0))^2 \tag{6-50}$$

where

$$\widetilde{\psi} = -\left(\frac{(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}+\omega_p^{-1}} - (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}\right)\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0))$$
$$+\frac{(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'}{\mathbf{Q}'\mathbf{Q}+\omega_p^{-1}}(g(\gamma\widetilde{\theta}_0)-g_0(\gamma\widetilde{\theta}_0)) \tag{6-51}$$

$$\widetilde{\psi}_p = \frac{\omega_p^{-1}}{\mathbf{Q}'\mathbf{Q}+\omega_p^{-1}}(\mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0))-(g(\gamma\widetilde{\theta}_0)-g_0(\gamma\widetilde{\theta}_0))) \tag{6-52}$$

and

$$(\gamma_{Ia}+\hat{\gamma}_{Ia})\sigma_\varepsilon^4 \ge E((\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0)-\mathbf{Df}\widetilde{\psi})'\boldsymbol{\omega}(\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0)-\mathbf{Df}\widetilde{\psi})$$
$$+\widetilde{\psi}_p\omega_p\widetilde{\psi}_p)-2E((\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0)-\mathbf{Df}\widetilde{\psi}_0)'\boldsymbol{\omega}(\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0)-\mathbf{Df}\widetilde{\psi}_0)$$
$$+\widetilde{\psi}_{0p}\omega_p\widetilde{\psi}_{0p}) \tag{6-53}$$

where

$$\widetilde{\psi}_0 = -\left(\frac{(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}+\omega_p^{-1}} - (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}\right)\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\widetilde{\theta}_0)-\mathbf{f}_0(\gamma\widetilde{\theta}_0)) \tag{6-54}$$

$$\tilde{\psi}_{0p} = \frac{\omega_p^{-1}}{\mathbf{Q}'\mathbf{Q} + \omega_p^{-1}} \mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\tilde{\boldsymbol{\theta}}_0) - \mathbf{f}_0(\gamma\tilde{\boldsymbol{\theta}}_0)) \tag{6-55}$$

The result that (6-44) and (6-45) become (6-13) and (6-14) when $\mathbf{W}_a = \boldsymbol{\omega}_a$ confirms the result obtained with the perturbation method that, at least when $\mathbf{W}_a = \boldsymbol{\omega}_a$, $\hat{\gamma}_{Ia}\sigma_\varepsilon^4 = \hat{\gamma}_I\sigma_\varepsilon^4$. Also, when $\omega_p^{-1} >> \mathbf{Q}'\mathbf{Q}$, (6-50) and (6-53) become (6-13) because, from (H-12), appendix H, $\tilde{\boldsymbol{\theta}}_0 \to \hat{\boldsymbol{\theta}}_0$. This confirms the result of (G-46) that $\gamma_{Ia}\sigma_\varepsilon^4 \approx 0$ when $\omega_p^{-1} >> \mathbf{Q}'\mathbf{Q}$.

A Monte Carlo method to compute bounds for $\gamma_{Ia}\sigma_\varepsilon^4$ could be developed as an extension of the method used to compute bounds for $\gamma_I\sigma_\varepsilon^4$ discussed previously.

## Empirical Measures of Model Intrinsic Nonlinearity, Model Combined Intrinsic Nonlinearity, and Total Model Nonlinearity

***A measure of model intrinsic nonlinearity.*** If $\mathbf{V}_*$ (or $\boldsymbol{\Omega}$) is unknown, then (6-12) cannot be approximated using the correct distribution for $\hat{\boldsymbol{\theta}}_0$. In this case an empirical measure of model intrinsic nonlinearity similar to the one derived by Beale (1960, p. 57-59) and revised by Linssen (1975, p. 97-98) can be developed to indicate the importance of model intrinsic nonlinearity. A measure similar to the square of Linssen's measure combines (6-15) and (6-16) with the expected values replaced by sample averages computed using sets $\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}$ on the periphery of the linear confidence region $(\mathbf{Y} - \mathbf{f}_0(\gamma\boldsymbol{\theta}_0))'\boldsymbol{\omega}(\mathbf{Y} - \mathbf{f}_0(\gamma\boldsymbol{\theta}_0)) - S(\hat{\boldsymbol{\theta}})$ $= (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'\mathbf{D}\hat{\mathbf{f}}'\boldsymbol{\omega}\mathbf{D}\hat{\mathbf{f}}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) \le pc_r S(\hat{\boldsymbol{\theta}})F_\alpha(p, n-p)/(n-p)$, where $\mathbf{f}_0(\gamma\boldsymbol{\theta}_0) = \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{D}\hat{\mathbf{f}}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})$ and $\mathbf{D}\hat{\mathbf{f}}$ indicates evaluation at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. However, values of the earlier form given by Beale (1960, p. 58) (without the correction factor) presented by Guttman and Meeter (1965) are consistently larger than corresponding values of (6-15) also presented by them. A possibly better approximation would be to place the sets $\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}$ on the surface where $(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'\mathbf{D}\hat{\mathbf{f}}'\boldsymbol{\omega}\mathbf{D}\hat{\mathbf{f}}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})$ is equal to its estimated expected value, $aps^2$, obtained by substituting $s^2$ for $b\sigma_\varepsilon^2$ in (5-15).

Beale (1960) and Linssen (1975) both defined their empirical measures in terms of specific regression $\hat{\boldsymbol{\theta}}$. However, to conform to the theoretical measure (6-15), the empirical measure is defined here to be independent of any specific regression. For practical use, the measure must be evaluated at $\hat{\boldsymbol{\theta}}$, which makes it analogous to Beale's and Linssen's measures. The measure of model intrinsic nonlinearity based on the foregoing ideas is thus defined as

$$\hat{N}_{\min} = \frac{1}{apb\sigma_\varepsilon^2} \sum_{\ell=1}^{2p} (\mathbf{f}(\gamma\boldsymbol{\theta}_\ell) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell) - \mathbf{Df}\boldsymbol{\psi}_\ell)'\boldsymbol{\omega}(\mathbf{f}(\gamma\boldsymbol{\theta}_\ell) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell) - \mathbf{Df}\boldsymbol{\psi}_\ell)/(2p) \tag{6-56}$$

where

$$\boldsymbol{\psi}_\ell = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}(\mathbf{f}(\gamma\boldsymbol{\theta}_\ell) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell)) \tag{6-57}$$

and the $2p$ parameter sets $\boldsymbol{\theta}_\ell$ are obtained as $\boldsymbol{\theta}_\ell = (\boldsymbol{\theta}_\ell - \bar{\boldsymbol{\theta}}) + \bar{\boldsymbol{\theta}}$ for which the change $\boldsymbol{\theta}_\ell - \bar{\boldsymbol{\theta}}$ is computed using the Cooley and Naff (1990, p. 189) method, except that Cooley and Naff's $d^2$ is set equal to the expected value $apb\sigma_\varepsilon^2$. This definition uses a linear probability region $(\mathbf{Y} - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell))'\boldsymbol{\omega}(\mathbf{Y} - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell)) - S(\bar{\boldsymbol{\theta}}) = (\boldsymbol{\theta}_\ell - \bar{\boldsymbol{\theta}})'\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df}(\boldsymbol{\theta}_\ell - \bar{\boldsymbol{\theta}})$ centered on $\bar{\boldsymbol{\theta}}$ and having a diameter $apb\sigma_\varepsilon^2$ instead of the linear confidence region centered on $\hat{\boldsymbol{\theta}}$ used by Beale and Linssen. The measure is, therefore, defined to be independent of any specific regression. For practical calculation $apb\sigma_\varepsilon^2$ is replaced in (6-56) with $aps^2$ because $b\sigma_\varepsilon^2$ is unknown. Similarly, the sensitivity matrix $\mathbf{Df}$ and parameter set $\bar{\boldsymbol{\theta}}$ are replaced with $\hat{\mathbf{Df}}$ and $\hat{\boldsymbol{\theta}}$ because $\mathbf{Df}$ and $\bar{\boldsymbol{\theta}}$ is unknown. If $a$ is unknown, then its bound $n/p$ can be used.

***A measure of model combined intrinsic nonlinearity***. The measure of the model combined intrinsic nonlinearity is based on (6-32) and (6-35). It is independent of a specific regression and is defined in three parts. The first part is drawn from the first terms on the right-hand sides of (6-32) and (6-35) and is

$$\hat{M}_{\min} = \frac{1}{\xi\, b\sigma_\varepsilon^2}\sum_{\ell=1}^{2}(\mathbf{f}(\gamma\boldsymbol{\theta}_\ell) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell) - \mathbf{Df}\boldsymbol{\psi}_\ell)'\boldsymbol{\omega}(\mathbf{f}(\gamma\boldsymbol{\theta}_\ell) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell) - \mathbf{Df}\boldsymbol{\psi}_\ell)/2 \tag{6-58}$$

where

$$\boldsymbol{\psi}_\ell = \boldsymbol{\psi}_\ell^0 + \frac{(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\boldsymbol{\theta}_\ell) - g_0(\gamma\boldsymbol{\theta}_\ell)) \tag{6-59}$$

$$\boldsymbol{\psi}_\ell^0 = -\left(\frac{(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}} - (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}\right)\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\boldsymbol{\theta}_\ell) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell)) \tag{6-60}$$

$$\xi = \frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'(\boldsymbol{\omega}/b)^{\frac{1}{2}}\boldsymbol{\Omega}(\boldsymbol{\omega}/b)^{\frac{1}{2}}\mathbf{Q} \tag{6-61}$$

Parameter sets $\boldsymbol{\theta}_\ell = (\boldsymbol{\theta}_\ell - \bar{\boldsymbol{\theta}}) + \bar{\boldsymbol{\theta}}$ form limits of a probability interval of the same width as a linear confidence interval, but are centered on $\bar{\boldsymbol{\theta}}$ instead of $\hat{\boldsymbol{\theta}}$. They are computed as described later in this section. Two quantities in addition to $\hat{M}_{\min}$ are needed to measure the importance of the sum $\gamma_I + \hat{\gamma}_I$ related to bounds (6-32) and (6-35). In scaled form these are

$$\hat{B}_U = \frac{1}{\xi\, b\sigma_\varepsilon^2}\sum_{\ell=1}^{2}(g(\gamma\boldsymbol{\theta}_\ell) - g_0(\gamma\boldsymbol{\theta}_\ell))^2/(2\,\mathbf{Q}'\mathbf{Q}) \tag{6-62}$$

and

$$\hat{B}_L = \frac{1}{\xi\, b\sigma_\varepsilon^2}\sum_{\ell=1}^{2}(\mathbf{f}(\gamma\boldsymbol{\theta}_\ell) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell) - \mathbf{Df}\boldsymbol{\psi}_\ell^0)'\boldsymbol{\omega}(\mathbf{f}(\gamma\boldsymbol{\theta}_\ell) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_\ell) - \mathbf{Df}\boldsymbol{\psi}_\ell^0)/2 \tag{6-63}$$

where the importance of $(\gamma_I + \hat{\gamma}_I)\sigma_\varepsilon^4$ is measured by the larger in magnitude of $(\hat{M}_{\min} + 2\hat{B}_U)\xi\, b\sigma_\varepsilon^2$ and $(\hat{M}_{\min} - 2\hat{B}_L)\xi\, b\sigma_\varepsilon^2$. Note that $\hat{B}_U$ is $\hat{M}_{\min}$ as if $\mathbf{f}(\gamma\theta_\ell)$ were the linear model $\mathbf{f}_0(\gamma\theta_\ell)$ and $\hat{B}_L$ is $\hat{M}_{\min}$ as if $g(\gamma\theta_\ell)$ were the linear function $g_0(\gamma\theta_\ell)$.

Parameter sets $\theta_\ell$ are located at points on the surface where $(\theta_\ell - \bar{\theta})'\mathbf{Df}'\omega\mathbf{Df}(\theta_\ell - \bar{\theta})$ is equal to its expected value when $\theta_\ell - \bar{\theta}$ has the same distribution as $\tilde{\mathbf{l}}_*$. From (F-66), this expected value is given by $\sigma_\varepsilon^2 + \gamma_w\sigma_\beta^2$, where, from (5-49),

$$\sigma_\varepsilon^2 + \gamma_w\sigma_\beta^2 \approx \frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'(\omega/b)^{\frac{1}{2}}\Omega(\omega/b)^{\frac{1}{2}}\mathbf{Q}b\sigma_\varepsilon^2 = \xi\, b\sigma_\varepsilon^2 \tag{6-64}$$

If $\hat{\omega}$ or $\omega_G$ is used for $\omega$, (6-64) has the approximate bound given from (5-56) as $V_{mx}b\sigma_\varepsilon^2/\mathbf{Q}'\mathbf{Q}$, which can be used when $\xi$ is unknown. Points on the surface corresponding to the ends of the probability interval can be used to evaluate $\theta_\ell$, since nonlinearity effects in these areas would seem to be most important. This idea follows from an alternative to the use of expected values in $\hat{\gamma}_I\sigma_\varepsilon^4$ suggested by Spjötvoll (Johansen, 1983, p. 189). Values of $\theta_\ell$ at the ends of the linearized probability interval are computed using (H-12) with $\omega_p^{-1}$ set equal to zero, $d_a^2$ set equal to $\mathbf{Q}'(\omega/b)^{1/2}\Omega(\omega/b)^{1/2}\mathbf{Q}b\sigma_\varepsilon^2/\mathbf{Q}'\mathbf{Q}$, and $g(\gamma\hat{\theta})$ replaced with $g(\gamma\bar{\theta})$. The average using the two values is used as the sample average. For practical use $b\sigma_\varepsilon^2$ is replaced by $s^2$, $\bar{\theta}$ is replaced by $\hat{\theta}$, and derivatives are evaluated at $\hat{\theta}$.

***A measure of total model nonlinearity***. Finally, a measure of model nonlinearity for a model using parameters $\theta$ instead of $\phi$ can be defined. This type of nonlinearity is referred to as the total model nonlinearity because it is the sum of model intrinsic nonlinearity and parameter effects nonlinearity, the latter of which is the part of the total model nonlinearity that can be removed by the transformation $\phi(\theta)$ (Draper and Smith, 1998, p. 528-529). A modification of the measure of the total model nonlinearity as defined by Beale (1960, p. 54) is defined in this report as the average sum of weighted, squared discrepancies between nonlinear and linear model values scaled using the squared diameter of the expected value of the linear probability region, $apb\sigma_\varepsilon^2$ :

$$\hat{N} = \frac{1}{apb\sigma_\varepsilon^2}\sum_{\ell=1}^{2p}(\mathbf{f}(\gamma\theta_\ell) - \mathbf{f}_0(\gamma\theta_\ell))'\omega(\mathbf{f}(\gamma\theta_\ell) - \mathbf{f}_0(\gamma\theta_\ell))/(2p) \tag{6-65}$$

Parameter sets $\theta_\ell$ are defined in the same way as for $\hat{N}_{\min}$, and the measure is modified for practical evaluation analogously to $\hat{N}_{\min}$.

## Summary of Principal Results

Component correction factors $\hat{\gamma}_I\sigma_\varepsilon^4$, $\gamma_I\sigma_\varepsilon^4$, $\hat{\gamma}_{Ia}\sigma_\varepsilon^4$, and $\gamma_{Ia}\sigma_\varepsilon^4$, which are components of correction factors $c_r$, $c_c$, and $c_p$ developed in section 5, correct for model intrinsic nonlinearity and model combined intrinsic nonlinearity that can affect the confidence regions, and confidence and prediction intervals, also developed in section 5. These component factors are analyzed using a method that is free of the assumptions and restrictions inherent in the combined Taylor series and perturbation method originally used to derive them. The new method is based on expansion of the residual vector $\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$ as an identity that is the sum of three vectors, two of which are shown to be zero when model intrinsic nonlinearity is zero, and a third given as $\boldsymbol{\omega}^{-1/2}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))$, which has the form of the residual vector for a classical linear model. System intrinsic nonlinearity is contained in the third term. Factor $\hat{\gamma}_I\sigma_\varepsilon^4$ is given by (6-3), which is $\hat{\gamma}_I\sigma_\varepsilon^4 = E(\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})) - E(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))$ as obtained using the identity. Factor $\gamma_I\sigma_\varepsilon^4$ equals $-\hat{\gamma}_I\sigma_\varepsilon^4$ when used in $c_r$ and is given by (6-20) when used in $c_c$. Development of (6-20) is analogous to development of (6-3), but is more complex because of the constraints used in developing quantities for individual confidence intervals. The factors correspond exactly to the same factors derived using the Taylor series/perturbation method when the perturbation approximations are used. However, they apply more generally and establish the fact that the concepts of model intrinsic nonlinearity and model combined intrinsic nonlinearity are valid beyond the Taylor series/perturbation approximation.

An approximate analysis for Gauss-Markov estimation uses the Taylor series/ perturbation forms of the factors together with the assumption that $\boldsymbol{\omega}^{1/2}E(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{1/2} \approx \mathbf{I}$ to give approximate bounds (6-11) for $\hat{\gamma}_I\sigma_\varepsilon^4$ that are the same as bounds originally obtained for the classical nonlinear model by Johansen (1983). The bounds are written in nonperturbation form as (6-12) and (6-13). The analyses are extended to yield approximate bounds for $\gamma_I\sigma_\varepsilon^4$ as used in $c_c$. The bounds are given in terms of $\hat{\gamma}_I\sigma_\varepsilon^4$ by (6-31) (or (6-32)) and (6-34) (or (6-35)).

All of the methods extend readily to apply for $\hat{\gamma}_{Ia}\sigma_\varepsilon^4$ and $\gamma_{Ia}\sigma_\varepsilon^4$ as obtained for prediction intervals. Component correction factors $\hat{\gamma}_{Ia}\sigma_\varepsilon^4$ and $\gamma_{Ia}\sigma_\varepsilon^4$ are given by (6-42) and (6-43), respectively; approximate bounds for $\hat{\gamma}_{Ia}\sigma_\varepsilon^4$ for Gauss-Markov estimation are given by (6-44); and similar bounds for $\gamma_{Ia}\sigma_\varepsilon^4$ are given by (6-46) and (6-48). These are all of the same form as the bounds for $\hat{\gamma}_I\sigma_\varepsilon^4$ and $\gamma_I\sigma_\varepsilon^4$. When the augmented weight matrix is given by $\boldsymbol{\omega}_a$, the bounds assume the simplified forms given by $\hat{\gamma}_{Ia}\sigma_\varepsilon^4 = \hat{\gamma}_I\sigma_\varepsilon^4$, (6-50), and (6-53). These forms confirm the generality of results obtained using the Taylor series/perturbation method. In particular, when $E(Y_p - g(\gamma\boldsymbol{\theta}_*))^2 >> Var(\mathbf{D}g(\tilde{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_*))$ (where $\tilde{\boldsymbol{\theta}}_0$ is $\tilde{\boldsymbol{\theta}}$ obtained using a linear model approximation), $\gamma_I\sigma_\varepsilon^4 \approx 0$.

Measures of model intrinsic nonlinearity and model combined intrinsic nonlinearity are defined to indicate the possible importance of $\gamma_I\sigma_\varepsilon^4$ and $\hat{\gamma}_I\sigma_\varepsilon^4$ when estimates of bounds for these factors cannot be computed because the second moment matrix of $\mathbf{Y}-\mathbf{f}(\gamma\boldsymbol{\theta}_*)$ is unknown. These measures are given by (6-56)-(6-64). A quantity known as the total model nonlinearity measures the weighted sum of squared discrepancies between model functions $\mathbf{f}(\gamma\boldsymbol{\theta})$ and linearized approximations of them. This measure is given by (6-65) and is always greater than or equal to the measure of model intrinsic nonlinearity.

# 7. Experimental Results

The objective of this section is to explore consequences of the theory presented in sections 1-6 for two specific hypothetical examples. The investigation is intended to test, for the two examples, the validity and robustness of the theory when the model error is large. Thus, the investigation is not, and is not intended to be, exhaustive. Such an investigation is beyond the scope of this report. The first example is for one-dimensional, steady-state flow in an aquifer having spatially (one-dimensionally) varying transmissivity and constant recharge. Transmissivity varies stochastically and only at small scale. Although the example is numerically simple, it turns out to be rather ill conditioned. The second example is for two-dimensional, steady-state flow and is based on the example used by Cooley and Naff (1990, p. 79-81). Both transmissivity and recharge vary spatially (two dimensionally) at both large and small scales, with the small-scale variations being stochastic. This example is much more numerically complex than the first, but turns out to be well conditioned.

To provide for flexibility in interpretation of results, all variables in both examples are scaled with arbitrary length parameters $l_d$ and $l_q$ and an arbitrary time parameter $t_c$ to make them dimensionless. Thus, $h$ is (*hydraulic head*)/$l_d$; $x$ and $y$ are *distance*/$l_d$; $T$ is *transmissivity* $\times t_c/(l_d l_q)$; $W$ is (*recharge rate*) $\times t_c/l_q$; $q$ is *flux* $\times t_c/(l_d l_q)$; and $Q$ is (*pumping discharge*) $\times t_c/(l_d^2 l_q)$. Specific values of these variables such as $x = L$ are scaled consistently. For simplicity, the modifier "dimensionless" is omitted when discussing these variables.

## Example 1 – One-Dimensional, Steady-State Flow with Recharge

***Models and stochastic properties.*** Consider a general, one-dimensional, steady-state flow system for which the hydrogeology is depicted in figure 7-1. Hydraulic head is known at the lower $(x = 0)$ end of the system and the (Darcy) flux is known at the upper $(x = L)$ end. Recharge and transmissivity can vary from block $(\Delta x_i)$ to block in the system. The solution for hydraulic head $h(x)$ at any point $x$ that lies in block $j$ is given by

$$h(x) = -\frac{q_{j-1}}{T_j}(x - x_{j-1}) - \sum_{i=1}^{j-1}\frac{q_{i-1}\Delta x_i}{T_i} - \frac{1}{2}\left(\frac{W_j}{T_j}(x - x_{j-1})^2 + \sum_{i=1}^{j-1}\frac{W_i}{T_i}(\Delta x_i)^2\right) + h_0 \qquad (7\text{-}1)$$

where $T_i$ is transmissivity in block $i$, $W_i$ is recharge rate in block $i$, $h_0$ is the known hydraulic head at $x = 0$, and

$$q_{j-1} = q_L - \sum_{i=j}^{N} W_i \Delta x_i \qquad (7\text{-}2)$$

in which $q_L$ is the known flux at $x = L$ and $N$ is the number of blocks. This is the $f(\beta)$ model.

For simplicity let $W$ be invariant and constant over the flow region, $0 \le x \le L$. Also, let all $T_i$ be independently and identically log-normally distributed so that for this example (3-1) is

$$\left.\begin{aligned}\beta_i &= \ln T_i \sim N(\bar{\theta}_1,\sigma_\beta^2);\ i=1,2,\ldots,N\\ \beta_{i+N} &= W_i \sim N(\bar{\theta}_2,0);\ i=1,2,\ldots,N\end{aligned}\right\} \qquad (7\text{-}3)$$

where $\bar{\theta}_1$ is the $\ln T$ drift parameter and $\bar{\theta}_2$ is the $W$ drift parameter. Recharge $W$ was placed in β because it is to be estimated as a parameter. Giving all $\beta_{i+N}$ a variance of zero fixes them at the value of $\bar{\theta}_2$ in the stochastic process. For this system, γ is defined by (3-2), and $\mathbf{1}_1$ and $\mathbf{1}_2$ each have dimension $N$ ($m_1 = N$ and $m_2 = N$). Specific values for the quantities needed to specify example 1 are given in table 7-1.

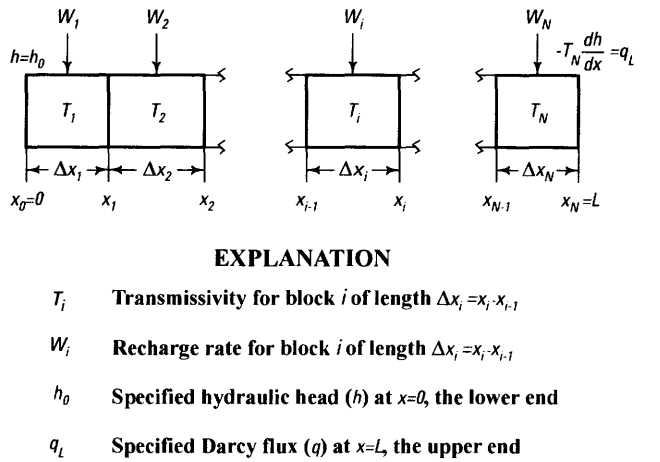


**Figure 7-1.** Assumed hydrogeology of the one-dimensional, ground-water flow system for example 1.

**Table 7-1.** Model specifications for example 1.

| | |
|---|---|
| Drift transmissivity: | $\exp(\bar{\theta}_1) = \bar{T} = 1{,}000$ |
| Drift recharge rate: | $\bar{\theta}_2 = \bar{W} = 0.003$ |
| Known hydraulic head at $x = 0$: | $h_0 = 10$ |
| Known flux at $x = L$: | $q_L = 20$ |
| Block size: | $\Delta x_i = 100,\ i = 1,2,\ldots,N$ |
| Standard deviation of the $\ln T$ process: | $\sigma_\beta = 0.5$ |
| Observation-error covariance matrix: | $\mathbf{V}_\varepsilon = \mathbf{I},\ \sigma_\varepsilon = 0.1$ |
| Number of blocks: | $N$=30 |
| Number of observations: | $n$=11, at $x$ = 50, 350, 650, 950, 1,250, 1,550, 1,850, 2,150, 2,450, 2,750, 2,950 |
| Number of parameters: | $p = 2$ |

A one-dimensional stochastic transmissivity process with no spatial correlation is not physically realistic (Bakr and others, 1978). However, this simple example is intended to provide an initial test of the validity and robustness of the theory presented here, and the uncorrelated, one-dimensional process is sufficient for this purpose. Some results for spatially correlated transmissivities are given later in this section.

From (3-7), (3-10), and (3-11), $\theta_{*j}$ is simply the arithmetic average, (3-11), of the $\beta_i$ values over all values of *i* pertaining to parameter *j*, so that the transmissivity corresponding to $\theta_{*1}$ is simply the geometric mean, $T_*$, of the $T_i$ values. By replacing all $T_i$ with $T_*$ and all $W_i$ with $\bar{\theta}_2 = \bar{W}$, (7-1) becomes $f(\gamma\theta_*)$, which can be rearranged to become

$$h_*(x) = -\frac{q_L}{T_*}x + \frac{\bar{W}}{T_*}x(L - \frac{1}{2}x) + h_0 \tag{7-4}$$

Examination of (7-1) reveals that it can be written as a linear model in terms of the 2*N* quantities $q_{i-1}/T_i$, $i = 1,2,\ldots,N$, and $W_i/T_i$, $i = 1,2,\ldots,N$. These constitute a one-to-one transformation of the 2*N* linearly independent system characteristics $\ln T_i$, $i = 1,2,\ldots,N$, and $W_i$, $i = 1,2,\ldots,N$, and thus form the transformation $\alpha(\beta)$. Similarly, (7-4) can be written as a linear model in terms of the two quantities $q_L/T$ and $W/T$. These constitute a one-to-one transformation of the two parameters $\ln T$ and *W* composing $\theta$ and thus form the transformation $\phi(\theta)$. Because both transformations linearize their respective models, both models ((7-1) and (7-4)) have no model intrinsic nonlinearity. The system intrinsic nonlinearity is defined in this report to be in terms of $\mathbf{I} - \mathbf{R}$, which would be computed using (7-4), and $\mathbf{D}^2_\beta \mathbf{f}$, which would be computed using (7-1). For the system intrinsic nonlinearity to be small, $\lambda$ defined as $\lambda = \mathbf{J}_\beta^{-1}\gamma\mathbf{J}$ by (C-24), appendix C, must be nearly constant. It is a straightforward task to compute $\mathbf{J}_\beta^{-1}$ and $\gamma\mathbf{J}$, then $\lambda$ to show that $\lambda$ is constant (not a function of $\beta$ or $\theta$). However, the approximation given by (C-27) also must be accurate. Evaluation of this expression revealed that it is not exact. Whether or not it is accurate enough to yield small system intrinsic nonlinearity must be determined by testing numerical results, which is done later in this section.

***Mean errors, covariances, and other population properties.*** The vector of mean errors $E(\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*))$, the matrix of second moments $E(\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*))(\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*))'$, and the covariance matrix $Var(\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*))$ were approximated for the 11 observation points given in table 7-1. Rather than develop the computer codes necessary to use (3-19), (3-23), and (3-24) to approximate these quantities, they were approximated using straightforward Monte Carlo simulations. For each realization a value of $\beta$ was generated from (7-3), $\theta_{*1}$ was computed, then these were used directly in the integrated finite difference solutions for $h(x)$ and $h_*(x)$ given by Cooley and Naff (1990, p. 81-83), which are exact for the boundary value problems leading to (7-1) and (7-4). Means, second moments, and covariances of the errors were computed as standard sample quantities. Values of sample skewness and kurtosis of the errors also were computed to check for deviations of the error distributions from normality. This would have been tedious using the Taylor series expansions, but was straightforward using the Monte Carlo method.

Results for a Monte Carlo sample size of 15,000 are given in figure 7-2 and tables 7-2 and 7-3. Mean errors $E(\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*))$ increase steadily from the known head boundary to the known flux boundary, where they are large, over 9. This suggests a large degree of total system nonlinearity (nonlinearity in $\mathbf{f}(\beta)$). However, although mean errors $E(\mathbf{f}(\gamma\theta_*) - \mathbf{f}(\gamma\bar{\theta}))$ show a similar increase, the increase is small because the average, $\theta_{*1}$, varies with the small standard

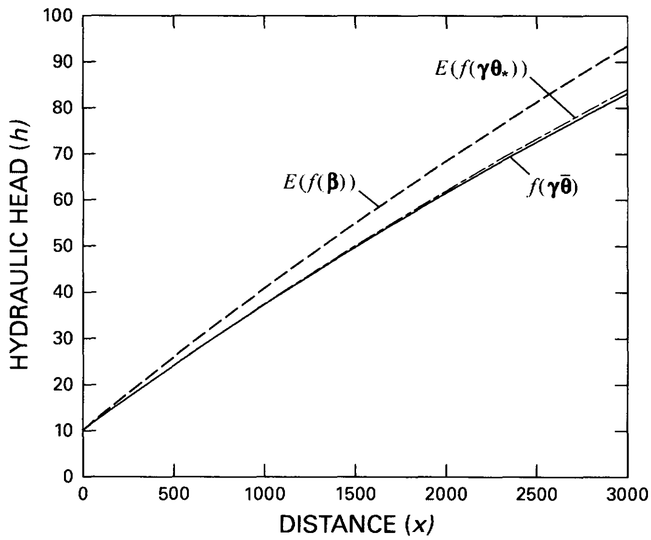


**Figure 7-2.** Curves containing mean model functions $E(\mathbf{f}(\boldsymbol{\beta}))$, $E(\mathbf{f}(\gamma\boldsymbol{\theta}_*))$, and $\mathbf{f}(\gamma\bar{\boldsymbol{\theta}})$ representing mean hydraulic heads for small-scale $\ln T$ variability, average $\ln T$ variability over $0 \le x \le L$, and the fixed $\ln T$ drift for example 1.

deviation of $0.5/\sqrt{30} = 0.091$ about $\bar{\theta}_1$. That produces a small value of $E(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_i\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e})/2$ in (3-18) for all $i$. Values of elements in the second moment matrix $E(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'$ increase in the same manner as do errors $E(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\boldsymbol{\theta}_*))$. Values of the standard deviations reach a maximum at about the halfway point in the system, but then decrease toward the known flux boundary. Correlations are large for most pairs of errors, but are especially large for pairs near the known flux boundary. Values of skewness and kurtosis indicate that the distribution of errors is not normal, being especially skewed and leptokurtic near the known head boundary. A similar increase in deviation of model errors from normality also was found to occur near a specified head boundary by Smith and Freeze (1979, p. 525). These results show that model errors have the potential of having a significant detrimental influence on regression modeling of the flow system.

**Table 7-2.** Values of mean, skewness, and kurtosis for the distribution of $f_i(\boldsymbol{\beta}) - f_i(\gamma\boldsymbol{\theta}_*)$ at observation points, $i$, for example 1.

| Obs. no. | Mean | Skewness* | Kurtosis+ |
|---|---|---|---|
| 1 | 0.187 | 1.55 | 7.21 |
| 2 | 1.30 | 0.848 | 4.29 |
| 3 | 2.35 | 0.657 | 3.81 |
| 4 | 3.37 | 0.515 | 3.45 |
| 5 | 4.39 | 0.455 | 3.37 |
| 6 | 5.36 | 0.430 | 3.36 |
| 7 | 6.28 | 0.417 | 3.40 |
| 8 | 7.17 | 0.396 | 3.40 |
| 9 | 8.03 | 0.441 | 3.48 |
| 10 | 8.85 | 0.567 | 3.63 |
| 11 | 9.38 | 0.752 | 3.94 |

*Compare with the theoretical value of 0 for a normal distribution of $f_i(\boldsymbol{\beta}) - f_i(\gamma\boldsymbol{\theta}_*)$.
+Compare with the theoretical value of 3 for a normal distribution of $f_i(\boldsymbol{\beta}) - f_i(\gamma\boldsymbol{\theta}_*)$.

**Table 7-3.** Second moment and correlation matrices for the distribution of $f_i(\beta) - f_i(\gamma\theta_*)$ at observation points, *i*, for example 1.

a. Second moment matrix

| Obs. no. | Second moment values with the diagonal in the first column | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.746 | 1.56 | 1.65 | 1.73 | 1.81 | 1.85 | 1.90 | 1.96 | 2.01 | 2.07 | 2.11 |
| 2 | 10.1 | 11.4 | 11.9 | 12.4 | 12.9 | 13.3 | 13.8 | 14.2 | 14.6 | 14.8 | |
| 3 | 19.1 | 21.3 | 22.2 | 23.1 | 23.9 | 24.6 | 25.4 | 26.1 | 26.6 | | |
| 4 | 29.6 | 31.6 | 32.8 | 33.9 | 35.0 | 36.1 | 37.1 | 37.8 | | | |
| 5 | 40.2 | 42.3 | 43.7 | 45.2 | 46.5 | 47.8 | 48.7 | | | | |
| 6 | 50.3 | 52.5 | 54.4 | 56.1 | 57.9 | 58.8 | | | | | |
| 7 | 60.5 | 63.1 | 65.1 | 67.0 | 68.2 | | | | | | |
| 8 | 70.8 | 73.5 | 75.6 | 77.1 | | | | | | | |
| 9 | 81.0 | 83.8 | 85.4 | | | | | | | | |
| 10 | 91.0 | 93.2 | | | | | | | | | |
| 11 | 97.7 | | | | | | | | | | |

b. Correlation matrix

| Obs. no. | Standard deviation in the first column and correlations in the remaining columns | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.843 | 0.536 | 0.378 | 0.304 | 0.254 | 0.215 | 0.186 | 0.162 | 0.143 | 0.133 | 0.130 |
| 2 | 2.90 | 0.759 | 0.602 | 0.507 | 0.440 | 0.388 | 0.347 | 0.316 | 0.296 | 0.291 | |
| 3 | 3.78 | 0.830 | 0.693 | 0.600 | 0.528 | 0.468 | 0.424 | 0.396 | 0.388 | | |
| 4 | 4.28 | 0.860 | 0.741 | 0.647 | 0.576 | 0.518 | 0.476 | 0.464 | | | |
| 5 | 4.57 | 0.883 | 0.769 | 0.682 | 0.609 | 0.550 | 0.529 | | | | |
| 6 | 4.65 | 0.887 | 0.783 | 0.695 | 0.622 | 0.591 | | | | | |
| 7 | 4.59 | 0.895 | 0.786 | 0.696 | 0.652 | | | | | | |
| 8 | 4.40 | 0.891 | 0.776 | 0.717 | | | | | | | |
| 9 | 4.07 | 0.881 | 0.798 | | | | | | | | |
| 10 | 3.57 | 0.915 | | | | | | | | | |
| 11 | 3.10 | | | | | | | | | | |

***Nonlinearity measures using*** $\omega = \hat{\omega}$. Values of the nonlinearity measures $\hat{N}$, $\hat{N}_{\min}$, $\hat{M}_{\min}$, $\hat{B}_L$ and $\hat{B}_U$ given by (6-56)-(6-65) and obtained using $\omega = \hat{\omega}$ are tabulated for $\ln T$, $W$, and head at $x = 3{,}000$ in table 7-4. (All three of these are specific functions that have been labeled $g(\gamma\theta)$ in general here. However, for clarity $\ln T$ and $W$ are used instead of $g(\gamma\theta)$. Similarly for clarity $h$ at $x = 3{,}000$ is labeled $h_p$. Weight matrix $\hat{\omega}$ used to compute the measures is obtained as explained later in this section.) Total model nonlinearity $\hat{N}$ is large; the sum of weighted, squared discrepancies is about 38 times the squared diameter of the region on the periphery of which the discrepancies are computed. However, as was expected, model intrinsic nonlinearity is, to within round-off error, zero. Model combined intrinsic nonlinearity is small for $\ln T$ but is somewhat larger for $W$. It is, to within round-off error, zero for $h_p$ as should be expected because the solution for head, (7-4), has no intrinsic nonlinearity. The bounds for $(\gamma_I + \hat{\gamma}_I)\sigma_\varepsilon^4$ would be large for $W$ and $h_p$ but the large bound for $h_p$ would be far too conservative, the reason for which is discussed after (6-30). Additional inquiry into model combined intrinsic nonlinearity for $W$ is presented after the following regression results.

**Table 7-4.** Values of the nonlinearity measures for log transmissivity, recharge rate, and hydraulic head at $x = 3{,}000$ for example 1.

[The value of $a$ needed was computed using (5-13) as 4.6056; the value of $b$ needed was computed using (5-11) as 1.0000; and the value of $\xi$ needed was computed using (6-62) as 0.97400.]

| | | |
|---|---|---|
| | $\hat{N} = 38.0$ | |
| | $\hat{N}_{\min} = 2.98\times10^{-8}$ | |
| For log transmissivity: | For recharge rate: | For head at 3,000: |
| $\hat{M}_{\min} = 0.0346$ | $\hat{M}_{\min} = 0.171$ | $\hat{M}_{\min} = 1.10\times10^{-7}$ |
| $\hat{B}_U = 7.75\times10^{-18}$ | $\hat{B}_U = 0$ | $\hat{B}_U = 0.253$ |
| $\hat{B}_L = 0.0346$ | $\hat{B}_L = 0.171$ | $\hat{B}_L = 0.253$ |

***Regression results and analysis of residuals using*** $\omega = \hat{\omega}$. A regression was performed using the method of Cooley and Hill (1992) on hydraulic-head data $\mathbf{Y}$ from a realization of the Monte Carlo process. A vector of zero-mean random normal deviates having a standard deviation of 0.1 (table 7-1) was added to $\mathbf{f}(\beta)$ to account for a small observation error. Note that the diagonal elements of the second moment matrix of table 7-3a are increased by only 0.01 by adding these deviates to the stochastic process; model error completely dominates the process. Hence, the theoretical conditions for validity of the theory are not satisfied, and robustness of the theory is being tested. Initially the weight matrix $\omega$ used was $\hat{\omega}$, the diagonal matrix, each element of which is the inverse of the sum of 0.01 and the diagonal value of the second moment matrix from table 7-3a. Partial results for $\omega = \Omega^{-1}$ and $\omega = \mathbf{I}$ are discussed later in this section.

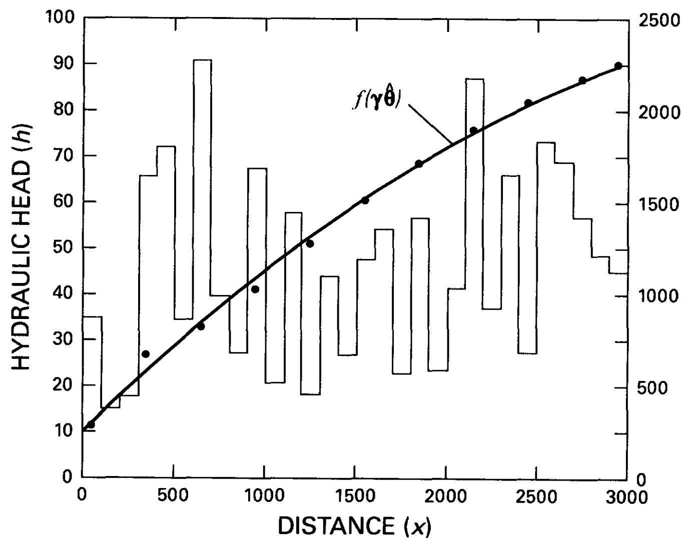


**Figure 7-3.** Hydraulic head data $Y_i$ at observation points $i$ (dots) along $x$, fitted model $f(\gamma\hat{\theta})$ using $\omega = \hat{\omega}$, and the $T$ distribution for example 1.

The data, regression curve, and spatial $T$ distribution are illustrated in figure 7-3. The systematic bias shown in figure 7-2 is completely absent from the fit of the regression curve to the data, which suggests small system intrinsic nonlinearity. (Several other regressions also were performed for comparison and produced the same results.) Note that the pattern of residuals appears to be both rather systematic and related to the local variations in $T$. However, the four groups of like signs has a greater than 10 percent chance of occurring by chance alone (Draper and Smith, 1998, p. 193-198). Thus, correlation from model error could not be conclusively identified from the pattern of residuals in this plot.

Weighted residuals $\hat{\omega}_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ are compared with measures of their theoretically correct values in figure 7-4. Measures of the theoretically correct values are 1) the sample means of ordered simulated, weighted residuals generated at the 11 observation points by Monte Carlo simulation from the weighted form of (4-42), and 2) the means plus and minus 2 times the sample standard deviations of these same ordered, simulated, weighted residuals. For comparison, in figure 7-5 the weighted residuals are plotted with the same theoretical measures, but are obtained from the incorrect distribution $N(\mathbf{0},(\mathbf{I}-\hat{\mathbf{R}})S(\hat{\theta})/(n-p))$, which, noting that $\hat{\mathbf{R}}=\hat{\omega}^{1/2}\mathbf{Df}(\mathbf{Df}'\hat{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\hat{\omega}^{1/2}$ $=\hat{\omega}^{1/2}\mathbf{D\hat{f}}(\mathbf{D\hat{f}}'\hat{\omega}\mathbf{D\hat{f}})^{-1}\mathbf{D\hat{f}}'\hat{\omega}^{1/2}$ in the present example, generally would be used for field studies. Plotting positions (approximations of the cumulative percents for the statistical distribution) $F_i=100(i-1/2)/n,\ i=1,2,\ldots,n,$ (Draper and Smith, 1998, p. 71) were used for both plots, and 10,000 Monte Carlo simulations were used to generate both sets of theoretical measures.

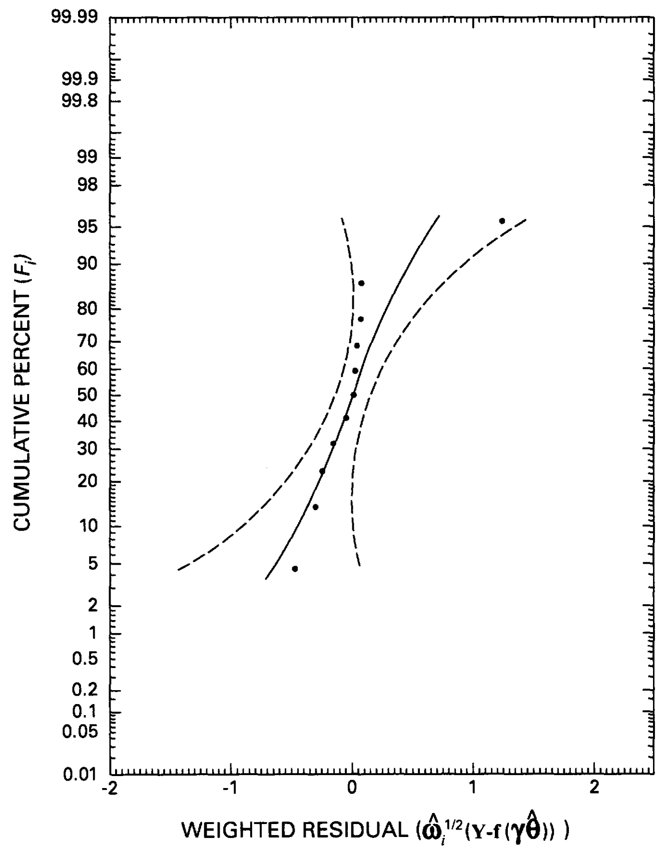


**Figure 7-4.** Probability plot of weighted residuals $\hat{\omega}_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ (dots), sample mean (solid line) of ordered, simulated, weighted residuals from the theoretically correct distribution, and plus and minus 2 standard deviation limits (dashed lines) of the ordered, simulated, weighted residuals for example 1.

All of the residuals are contained within the two-standard deviation limits when the simulated residuals are correct, but three are not when the simulated residuals are incorrect. Note that the means and limits when the simulated residuals are correct differ from the means and limits when the simulated residuals are incorrect. The mean has more of an S shape, the tails are more variable, and the center is less variable for

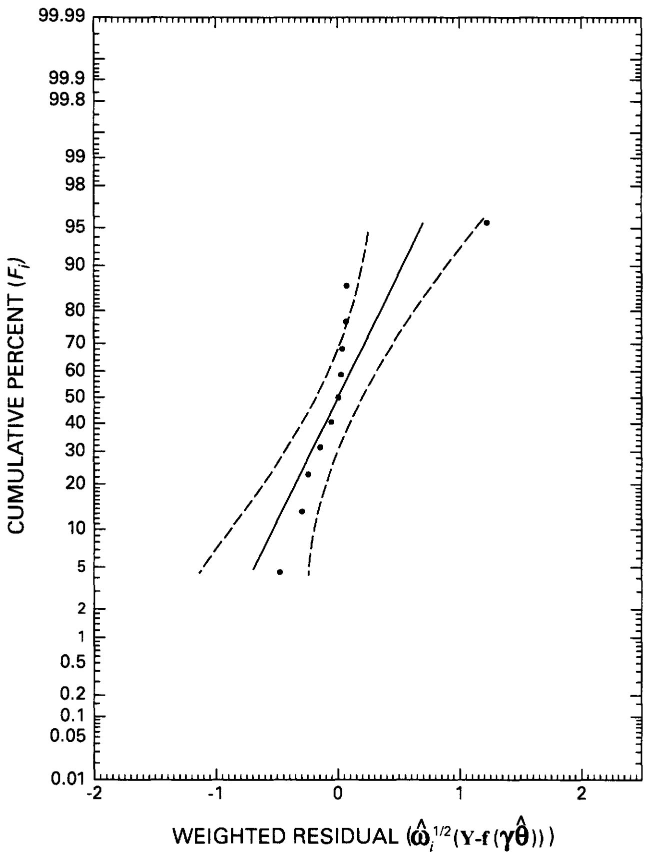


**Figure 7-5.** Probability plot of weighted residuals $\hat{\omega}_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ (dots), sample mean (solid line) of ordered, simulated, weighted residuals from the incorrect distribution $N(\mathbf{0},(\mathbf{I}-\hat{\mathbf{R}})S(\hat{\theta})/(n-p))$, and plus and minus 2 standard deviation limits (dashed lines) of the ordered, simulated, weighted residuals from the same distribution for example 1.

the correct residuals as compared to the incorrect residuals. The S shape appears to be reflected by the residuals.

The plot of weighted residuals versus weighted function values shown in figure 7-6 has a mean weighted residual of 0.0149 and a slope of –0.0921, neither of which are large. Visually, it appears to have a wave-like pattern, although the four groups of like signs could occur by chance greater than 10 percent of the time (Draper and Smith, 1998, p. 193-198). Thus, there is no quantitative evidence from the mean residual, the slope of the plot, or the sequence of signs that the residuals plot is abnormal. Any correlation of residuals resulting from the correlation of errors shown in table 7-3b is not conclusively shown in the plot.

A measure of model intrinsic nonlinearity that can be used in addition to (6-56) and (6-57) for a specific regression was indicated in section 4 to be the product of $\mathbf{R}$ and the weighted residual vector. The squared length of this product vector, given by $(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))'\omega^{1/2}\mathbf{R}\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$, can be used as a summary measure. Note that the expected value of this function appears in the component correction factor defined by (6-3). Similarly, a measure of model combined intrinsic nonlinearity was indicated in section 5 to be the product of $\mathbf{R}$ and the weighted constrained residual vector. The squared length of this product vector, given by $(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))'\omega^{1/2}(\mathbf{R}-\mathbf{QQ}'/\mathbf{Q}'\mathbf{Q})\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))$, can be used as a summary measure in addition to the general measures given by (6-58)-(6-63). The expected value of this function appears in the component correction factor defined by (6-20). Values of the functions were computed using $\omega=\hat{\omega}$ and found to be near zero. In particular, the value of the second function for $W$ is only slightly larger than values for $\ln T$ and $h_p$. Therefore, model combined intrinsic nonlinearity for $W$ does not appear to be significant, even though $\hat{M}_{\min}$ for it is not near zero.

***Confidence intervals using*** $\omega=\hat{\omega}$. Values of $\theta_{*i}$, their estimates $\hat{\theta}_i$, 95 percent linearized confidence intervals computed from $\hat{\theta}_i \pm t_{\alpha/2}(n-p)(c_c S(\hat{\theta})(\mathbf{D}\hat{\mathbf{f}}'\hat{\omega}\mathbf{D}\hat{\mathbf{f}})_{ii}^{-1}/(n-p))^{1/2}$

(where $(\mathbf{D\hat{f}}'\hat{\omega}\mathbf{D\hat{f}})_{ii}^{-1}$ is the $i$th diagonal element of $(\mathbf{D\hat{f}}'\hat{\omega}\mathbf{D\hat{f}})^{-1}$), and 95 percent nonlinear confidence intervals computed using (5-51) are shown together with their uncorrected ($c_c = 1$) counterparts in figure 7-7. Value $h_{p*} = g(\lambda\theta_*)$, its estimate $\hat{h}_p$, its 95 percent linearized confidence interval computed from $\hat{h}_p \pm t_{\alpha/2}(n-p)(c_c S(\hat{\theta})\mathbf{D}\hat{g}(\mathbf{D\hat{f}}'\hat{\omega}\mathbf{D\hat{f}})^{-1}\mathbf{D}\hat{g}'/(n-p))^{1/2}$, and its 95 percent npnlinear confidence interval computed using (5-51) also are shown in corrected and uncorrected forms in the figure.

The uncorrected confidence intervals are apparently too small; one out of the three linearized intervals and two out of the three nonlinear intervals do not contain their true values. Correction factors $c_c$ were computed using (5-50) with $\hat{\gamma}_I$ and $\gamma_I$ set to zero. Linearized intervals using the correction factors are large, true values falling well within the intervals. The nonlinear intervals show the effects of severe ill conditioning, which is indicated by a value of 0.997 for the linearized correlation (Cooley and Naff, 1990, p. 117) between the estimates $\ln\hat{T}$ and $\hat{W}$. Because of this ill conditioning, only ratios $W/\ln T$ are unique for the upper limits of the confidence intervals for $\ln T$ and $W$ so that the confidence intervals are open ended (unbounded). Also, the solution for the lower limit of the confidence interval for $h_p$ is unique only in terms of the ratio. The uncorrected nonlinear confidence interval for $h_p$ is the same as the linear one, and the lower limit of the corrected interval probably only differs from the lower limit for the corrected linear interval because of the influence of the nonuniqueness. This correspondence occurs because of the absence of model intrinsic and model combined intrinsic types of nonlinearity, as explained in the paragraph following (5-107). Corrected confidence intervals all contain their true values, and so may be accurate. However, because of the ill-conditioning problem, the actual containment probabilities could not be investigated by Monte Carlo analysis as was done for example 2 discussed later in this section.

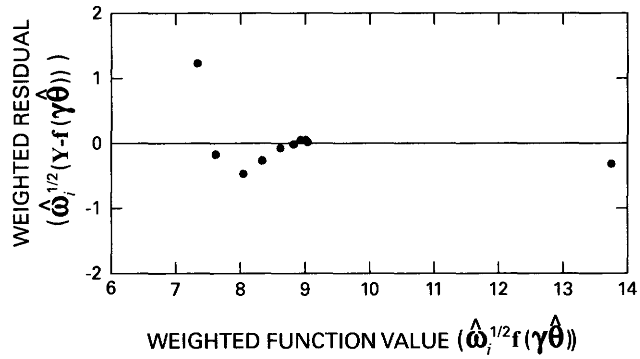


**Figure 7-6.** Plot of weighted residuals $\hat{\omega}_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ in relation to weighted function values $\hat{\omega}_i^{1/2}\mathbf{f}(\gamma\hat{\theta})$ for example 1.

***Results for alternative weight matrices.*** Regressions also were performed using $\omega = \Omega^{-1}$ (Gauss-Markov estimation) and $\omega = \mathbf{I}$ (ordinary least squares). Gauss-Markov estimation produced nearly the same parameter estimates and model fit to the data as obtained using $\omega = \hat{\omega}$, and ordinary least squares produced results only slightly different. Linearized confidence intervals for $\ln T_*$, $\overline{W}$, and $h_{p*}$ are shown in figure 7-8. Both corrected and uncorrected intervals are shown for ordinary least squares; no correction is required for Gauss-Markov estimation. Corrected intervals using both methods are nearly the same and both are smaller than comparable intervals obtained using $\omega = \hat{\omega}$ (figure 7-7). This may indicate that the latter intervals are somewhat conservative. Uncorrected intervals using ordinary least squares

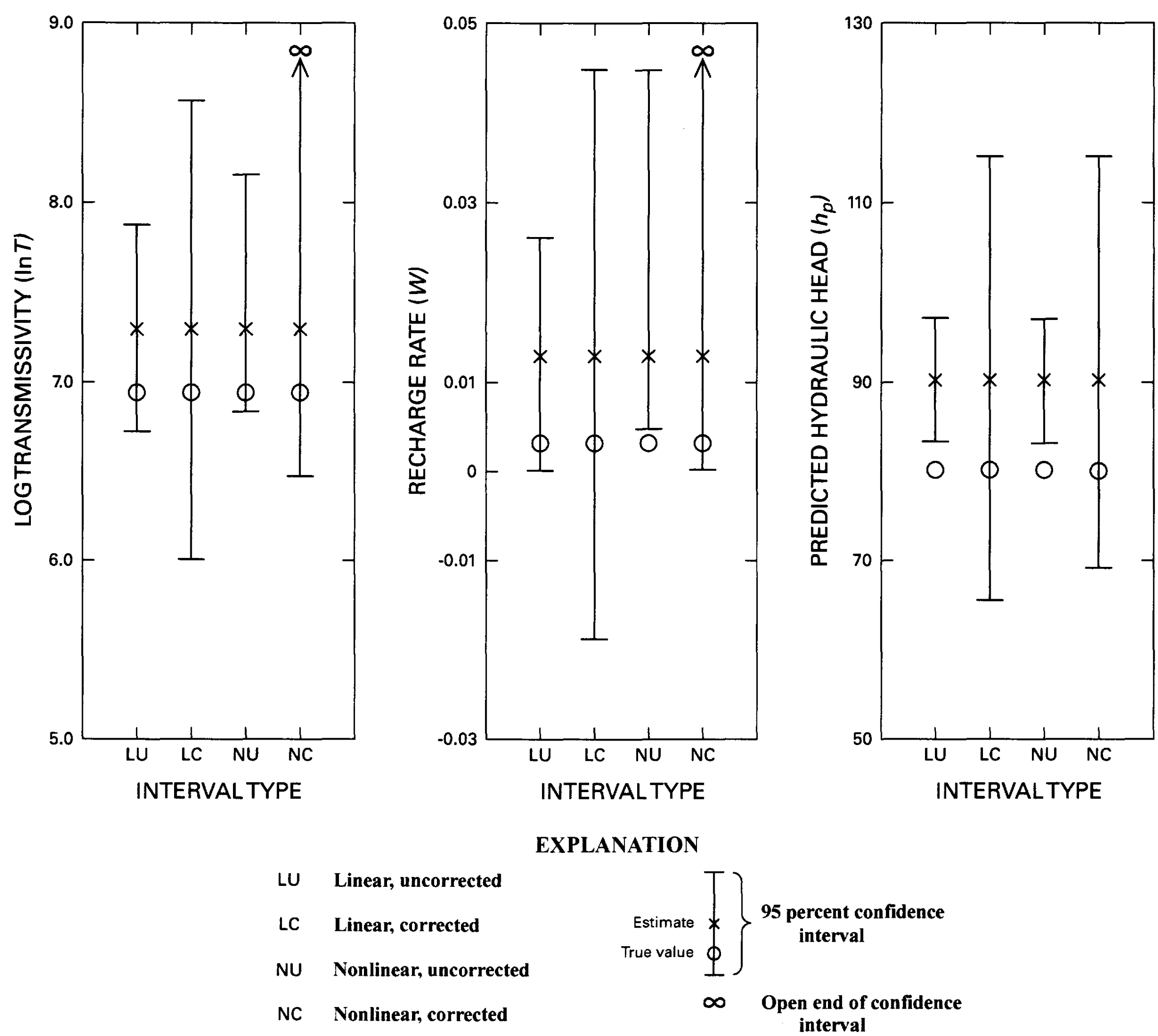


**Figure 7-7.** True values $\ln T_*$, $\overline{W}$, and $h_{p*}$, their estimates obtained using $\omega = \hat{\omega}$, and their 95 percent confidence intervals for example 1.

are smaller than comparable, uncorrected intervals obtained using $\omega = \hat{\omega}$. The former intervals for $\overline{W}$ and $h_{p*}$ do not contain their true values, and are likely to be more inaccurate than the latter intervals. However, all intervals except those using Gauss-Markov estimation require correction for model error. Finally, although not illustrated, neither set of residuals appears to differ from a set expected for the theoretical zero-mean normal distribution.

***Correction factors and bounds.*** Correction factors and bounds $V_{mx} / \mathbf{Q}'\mathbf{Q}$ for $\xi$ given by (5-56) (with $\omega_G$ replaced by both $\hat{\omega}$ and **I**) are given in table 7-5. Note that the values of $a$ suggest that, as predicted by theory, $S(\hat{\theta})/(n-p)$ considerably underestimates $b\sigma_\varepsilon^2$, the value of $(n-ap)/(n-p)$ being 0.199 when $\omega = \hat{\omega}$ and 0.0707 using ordinary least squares. Only a small part of the total variance, $b\sigma_\varepsilon^2$, is variance about the regression curves. Values of $c_r$ and $c_c$ are therefore large. Correction factors $c_r$ and $c_c$ obtained using ordinary least squares are

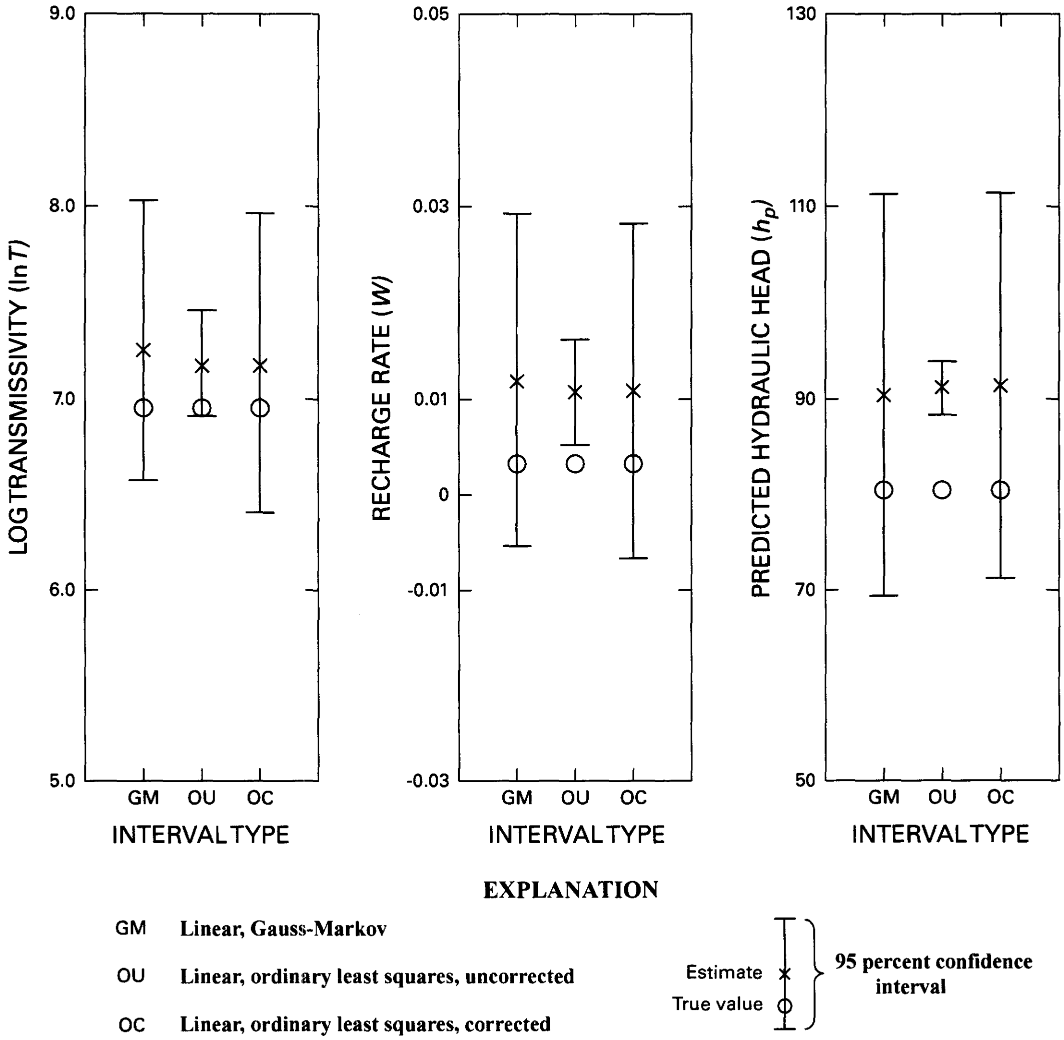


**Figure 7-8.** True values $\ln T_*$, $\overline{W}$, and $h_{p*}$, their estimates obtained using Gauss-Markov estimation and ordinary least squares, and their 95 percent linearized confidence intervals for example 1.

larger than correction factors $c_r$ and $c_c$ obtained using $\omega = \hat{\omega}$, reflecting the theoretically predicted result that $F$ and $t$ distributions are approximated better when $\omega = \hat{\omega}$ than when $\omega$ is arbitrary. The bounds $V_{mx} / \mathbf{Q}'\mathbf{Q}$ for $\xi$ are twice or more the value of $\xi$ except for the bound pertaining to $h_p$ obtained using ordinary least squares. The requirement for the bound to be more than approximate is that the error groups accurately reflect the error structure. Because **I** does not approximate $\hat{\omega}$ well, the bound is less than $\xi$. Correction factors are needed for all confidence intervals and both confidence regions. The bounds for $\xi$ generally would give much larger confidence intervals than using $\xi$.

**Table 7-5.** Correction factors for example 1.

[$a$ is computed using (5-13); $b$ is computed using (5-11); $c_r$ is computed using (5-19); $c_c$ is computed using (5-50) with $\gamma_I = 0$; $\xi$ is computed using (6-62); $V_{mx}/\mathbf{Q'Q}$ is computed using (5-56) with $\boldsymbol{\omega}_G$ replaced by $\hat{\boldsymbol{\omega}}$ or **I**.]

| Weight matrix | Variable | $a$ | $b$ | $c_r$ | $c_c$ | $\xi$ | $V_{mx}/\mathbf{Q'Q}$ |
|---|---|---|---|---|---|---|---|
| $\hat{\boldsymbol{\omega}}$ | $\ln T$ | 4.61 | 1.00 | 23.2 | 4.90 | 0.974 | 2.62 |
| | $W$ | 4.61 | 1.00 | 23.2 | 5.85 | 1.16 | 3.53 |
| | $h_p$ | 4.61 | 1.00 | 23.2 | 12.6 | 2.51 | 3.77 |
| **I** | $\ln T$ | 5.18 | 50.2 | 73.3 | 9.19 | 0.649 | 3.29 |
| | $W$ | 5.18 | 50.2 | 73.3 | 9.93 | 0.702 | 4.10 |
| | $h_p$ | 5.18 | 50.2 | 73.3 | 52.8 | 3.73 | 2.96 |

***Correlated errors***. Thus far, values of $\ln T_i$ have been uncorrelated. However, $\ln T$ values often are considered to be correlated (Bakr and others, 1978; Delhomme, 1979; Gelhar, 1986). As is suggested in section 3, when the correlation is manifested as a definite trend, the stochastic process used to apply the present theory should probably consider the trend to be the drift $\boldsymbol{\gamma}\bar{\boldsymbol{\theta}}$, even if the data showing the trend also could be considered to be a realization of a process having stationary drift. This idea is illustrated here for a one-dimensional, stationary, exponentially correlated, $\ln T$ process. To specify this process, (7-3) is replaced with a form in which for simplicity $W$ is not a parameter:

$$\boldsymbol{\beta} = \ln \mathbf{T} \sim N(\boldsymbol{\gamma}\bar{\boldsymbol{\theta}}, \mathbf{V}_\beta \sigma_\beta^2) \tag{7-5}$$

where $\boldsymbol{\beta}$ has order $m = N$, $\boldsymbol{\gamma} = \mathbf{1}$, $\bar{\theta} = \ln \bar{T}$ is the stationary drift, and

$$V_{\beta ij} = \exp(-|x_i - x_j|/\ell) \tag{7-6}$$

In (7-6) $\ell$ is the correlation length. Realizations of $\boldsymbol{\beta}$ can be generated using the simple method given by Kitanidis (1997, p. 191).

Figure 7-9 illustrates a realization of the correlated $\ln T$ process that shows a distinct linear trend. The drift $\bar{\theta} = \ln \bar{T}$ and the mean $\theta_* = \ln T_*$ also are shown. For this realization, $\sigma_\beta = 0.5$ and $\ell = 3{,}000$. It is apparent that estimating $\theta_*$ would yield a very crude model of the system; the trend needs to be included. One means of doing this is to revise the drift to be the linear trend, so that the revised drift is $\boldsymbol{\gamma}\bar{\boldsymbol{\theta}}$, where $\bar{\boldsymbol{\theta}} = [\bar{\theta}_1, \bar{\theta}_2]'$, $\bar{\theta}_1$ and $\bar{\theta}_2$ are drift values of $\ln T$ at $x = 0$ and $x = L$, $\gamma_{i1} = (L - \bar{x}_i)/L$, and $\gamma_{i2} = \bar{x}_i/L$, $i = 1,2,\ldots,m$, where $\bar{x}_i$ is $x$ at the midpoint in block $\Delta x_i$. Note that because of this revision, $\boldsymbol{\theta}_* = \bar{\boldsymbol{\theta}}$ for this realization. Revised residuals $\boldsymbol{\beta} - \boldsymbol{\gamma}\boldsymbol{\theta}_*$ from the fit of $\boldsymbol{\gamma}\boldsymbol{\theta}_*$ to $\boldsymbol{\beta}$ are given by

$$\boldsymbol{\beta} - \boldsymbol{\gamma}\boldsymbol{\theta}_* = \boldsymbol{\beta} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\boldsymbol{\beta} = (\mathbf{I} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\boldsymbol{\beta} \tag{7-7}$$

Then, revised values of $\boldsymbol{\beta}$, $\boldsymbol{\beta}_{rev}$, dispersed about the revised drift $\boldsymbol{\gamma}\bar{\boldsymbol{\theta}}$ are given by

$$\boldsymbol{\beta}_{rev} = \boldsymbol{\gamma}\bar{\boldsymbol{\theta}} + (\mathbf{I} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\boldsymbol{\beta} \tag{7-8}$$

which has revised covariance matrix $(\mathbf{I} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{V}_\beta(\mathbf{I} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\sigma_\beta^2$. Thus, because the rank of this covariance matrix is $m - p$,

$$\boldsymbol{\beta}_{rev} \sim N(\boldsymbol{\gamma}\bar{\boldsymbol{\theta}}, (\mathbf{I} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{V}_\beta(\mathbf{I} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\sigma_\beta^2) \tag{7-9}$$

is a singular normal distribution.

The change from distribution (7-5) to distribution (7-9) changes $Var(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\boldsymbol{\gamma\theta}_*))$. For example, the linearized version of this covariance matrix, which was originally $\mathbf{D}_\beta\mathbf{f}(\mathbf{I} - \mathbf{11}'/m)\mathbf{V}_\beta(\mathbf{I} - \mathbf{11}'/m)\mathbf{D}_\beta\mathbf{f}'\sigma_\beta^2$, is revised to be $\mathbf{D}_\beta\mathbf{f}(\mathbf{I} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{V}_\beta(\mathbf{I} - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{D}_\beta\mathbf{f}'\sigma_\beta^2$ where $\boldsymbol{\gamma}$ is for the linear drift. Note that, although $Var(\boldsymbol{\beta}_{rev})$ is singular, if $m - p \geq n$, which would generally be the case, $Var(\mathbf{f}(\boldsymbol{\beta}_{rev}) - \mathbf{f}(\boldsymbol{\gamma\theta}_*))$ would be full rank, *n*, if $\mathbf{D}_\beta\mathbf{f}$ has rank *n*. Original and revised, linearized correlation matrices are given in table 7-6. For the revised process, standard deviations of the model errors are greatly reduced, positive correlations are reduced, and negative correlations replace some of the original positive correlations.

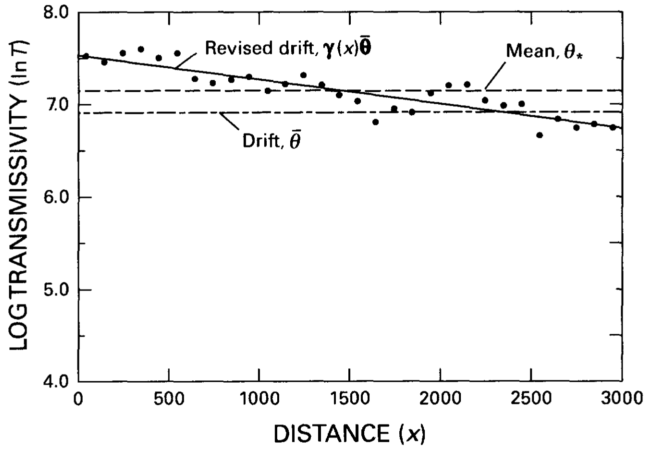


**Figure 7-9.** Realization (dots) from a stationary random process having an exponential covariance function for which $\sigma_\beta = 0.5$ and $\ell = 3{,}000$. The drift $\bar{\theta}$ of the process, the revised drift (the drift of the residuals), and the mean $\theta_*$ of the realization also are shown.

Correction factors *a*, $c_r$, $\xi$, and $c_c$ also were computed in the same way as in table 7-5 assuming $W = 0$, using the covariance matrices derived from table 7-6, and setting $\gamma_I$ and $\hat{\gamma}_I$ to zero. For the original process where $\ln \overline{T}$ is constant, the model used to calculate the correction factors is (7-4), where $\overline{W} = 0$. For the revised case a model with linearly varying $\ln T$ is needed. This is

$$h_*(x) = \frac{q_L L}{\ln(T_L/T_0)}\frac{1}{T_0}\left(\left(\frac{T_0}{T_L}\right)^{\frac{x}{L}} - 1\right) + h_0 \tag{7-10}$$

where $T_0 = e^{\theta_{*1}}$ is $T$ at $x = 0$ and $T_L = e^{\theta_{*2}}$ is $T$ at $x = L$. With use of (7-4) and table 7-6a, correction factors $a$ and $c_r$ are computed to be $a = 4.68$ and $c_r = 7.42$; for $\ln T_*$, $\xi = 4.68$ and $c_c = 7.42$. With use of (7-10) and table 7-6b, correction factors $a$ and $c_r$ are computed to be $a = 1.05$ and $c_r = 1.06$; for $\ln T_0$, $\xi = 0.601$ and $c_c = 0.608$; for $\ln T_L$, $\xi = 0.596$ and $c_c = 0.603$. Examination of the equations used to compute these correction factors shows that the reduction in magnitudes by the revision results primarily from the change in correlation structure. Thus, not only are the magnitudes of the model errors greatly reduced, but also, because of the change in correlation structure, the need for correction is virtually eliminated by the revision for this example. Not using the correction factors would give conservative individual confidence intervals.

**Table 7-6.** Standard deviations and correlations of model errors resulting from the original and revised, correlated log-transmissivity processes for example 1.

a. Correlation matrix, original process using (7-4)

| Obs. no. | Standard deviation in the first column and correlations in the remaining columns | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.484 | 0.950 | 0.883 | 0.818 | 0.756 | 0.697 | 0.644 | 0.601 | 0.575 | 0.599 | 0.699 |
| 2 | 3.01 | 0.966 | 0.908 | 0.845 | 0.784 | 0.728 | 0.680 | 0.651 | 0.676 | 0.782 | |
| 3 | 4.89 | 0.975 | 0.924 | 0.866 | 0.809 | 0.759 | 0.727 | 0.749 | 0.855 | | |
| 4 | 6.15 | 0.979 | 0.934 | 0.881 | 0.830 | 0.795 | 0.812 | 0.910 | | | |
| 5 | 6.83 | 0.981 | 0.940 | 0.893 | 0.856 | 0.865 | 0.948 | | | | |
| 6 | 6.96 | 0.983 | 0.945 | 0.908 | 0.908 | 0.970 | | | | | |
| 7 | 6.60 | 0.984 | 0.952 | 0.942 | 0.978 | | | | | | |
| 8 | 5.76 | 0.985 | 0.968 | 0.974 | | | | | | | |
| 9 | 4.50 | 0.988 | 0.967 | | | | | | | | |
| 10 | 2.87 | 0.976 | | | | | | | | | |
| 11 | 1.70 | | | | | | | | | | |

b. Correlation matrix, revised process using (7-10)

| Obs. no. | Standard deviation in the first column and correlations in the remaining columns | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.337 | 0.895 | 0.729 | 0.511 | 0.170 | -0.359 | -0.646 | -0.598 | -0.481 | -0.361 | -0.293 |
| 2 | 1.73 | 0.900 | 0.676 | 0.282 | -0.367 | -0.763 | -0.739 | -0.614 | -0.478 | -0.399 | |
| 3 | 2.17 | 0.875 | 0.488 | -0.267 | -0.820 | -0.863 | -0.751 | -0.606 | -0516 | | |
| 4 | 1.99 | 0.780 | 0.007 | -0.745 | -0.919 | -0.857 | -0.722 | -0.627 | | | |
| 5 | 1.51 | 0.532 | -0.377 | -0.776 | -0.840 | -0.761 | -0.677 | | | | |
| 6 | 1.16 | 0.482 | -0.102 | -0.375 | -0.447 | -0.426 | | | | | |
| 7 | 1.29 | 0.743 | 0.423 | 0.219 | 0.151 | | | | | | |
| 8 | 1.56 | 0.861 | 0.651 | 0.541 | | | | | | | |
| 9 | 1.54 | 0.890 | 0.774 | | | | | | | | |
| 10 | 0.985 | 0.941 | | | | | | | | | |
| 11 | 0.235 | | | | | | | | | | |

## Example 2 –Two-Dimensional, Steady-State Flow in a Zoned System

***Model and stochastic properties.*** This example is based on an example problem used by Cooley and Naff (1990, p. 79-81). The assumed hydrogeology is illustrated in figure 7-10. There are three zones of constant drift for $T$ and $W$; a river having a known streambed specific conductance, $R_2$; two wells pumping at known rates $Q_1$ and $Q_2$; a specified head boundary on which hydraulic head varies linearly between the values $h_{B1}$, $h_{B2}$, and $h_{B3}$; and a known flux boundary composed of five no-flow ($q_B = 0$) segments and two segments along the north boundary where $q_{B1}$ and $q_{B2}$ are known, nonzero, values.

The stochastic process involves both $W$ and $T$ and is defined only for small scale (grid-cell to grid-cell) variability. For simplicity, $\ln T_i$ and $W_i$ for grid-cells $i$ are all assumed to be statistically independent. Specifically, the assumed distributions are

$$\left.\begin{aligned} &\beta_i = \ln T_i \sim N(\bar{\theta}_k, V_{\beta ii}\sigma_\beta^2);\ i = i(k);\ k = 1,2,3 \\ &\beta_{i+N} = W_i \sim N(\bar{\theta}_k, V_{\beta i+N, i+N}\sigma_\beta^2);\ i = i(k);\ k = 4,5,6 \end{aligned}\right\} \qquad (7\text{-}11)$$

where $\sqrt{V_{\beta ii}} = 0.5$ and $\sqrt{V_{\beta i+N, i+N}} = 0.0001$. At river cells, $W$ does not appear in the flow equation, so $\beta_{i+N}$ is not defined by (7-11) and can be set to zero at these cells. Matrix $\gamma$ is defined by (3-2) with $\mathbf{1}_k$ having dimension $m_k$ equal to the number of grid cells occupied by parameter $k$. Three more parameters, $h_{B1}$, $h_{B2}$, and $h_{B3}$, are estimated in the regression. Distributions (7-11) could be augmented as for the first example to incorporate these parameters, but, because the distributions are never used, they are ignored. Specific values for the variables needed to specify example 2 are given in table 7-7.

As for example 1, a two-dimensional transmissivity process with no spatial correlation is not physically realistic (Bakr and others, 1978). This idea might be extended to the recharge process as well. However, as for example 1 the example is intended to provide a test of the validity and robustness of the theory, and the uncorrelated process is sufficient for this purpose.

***Mean errors, covariances, and other population properties.*** As for example 1, the vector of mean model functions (mean hydraulic heads) $E(\mathbf{f}(\boldsymbol{\beta}))$, the vector of mean errors $E(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\boldsymbol{\gamma\theta}_*))$, the matrix of second moments $E(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\boldsymbol{\gamma\theta}_*))(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\boldsymbol{\gamma\theta}_*))'$, the covariance matrix $Var(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\boldsymbol{\gamma\theta}_*))$, and the vectors of skewness and kurtosis were approximated for the 32 observation points shown in figure 7-10. Monte Carlo simulation using the integrated finite difference model in Cooley and Naff (1990, p. 81-83) was again used for the calculations. Partial results for a Monte Carlo sample size of 5,000 are given in table 7-8; off-diagonal second moments and correlations are not shown. The magnitudes of the mean errors range between 0 and 4.15 and correlate closely with the magnitudes of the mean hydraulic heads; the mean errors are not large but do indicate measurable total system nonlinearity in zone 3 and at pumping wells. All positive values of mean error occur in zone 1 under the influence of ground-water discharge and pumping from the two wells. The values that have smallest magnitude occur at and near the river. The largest second-moment values are associated with the largest magnitudes

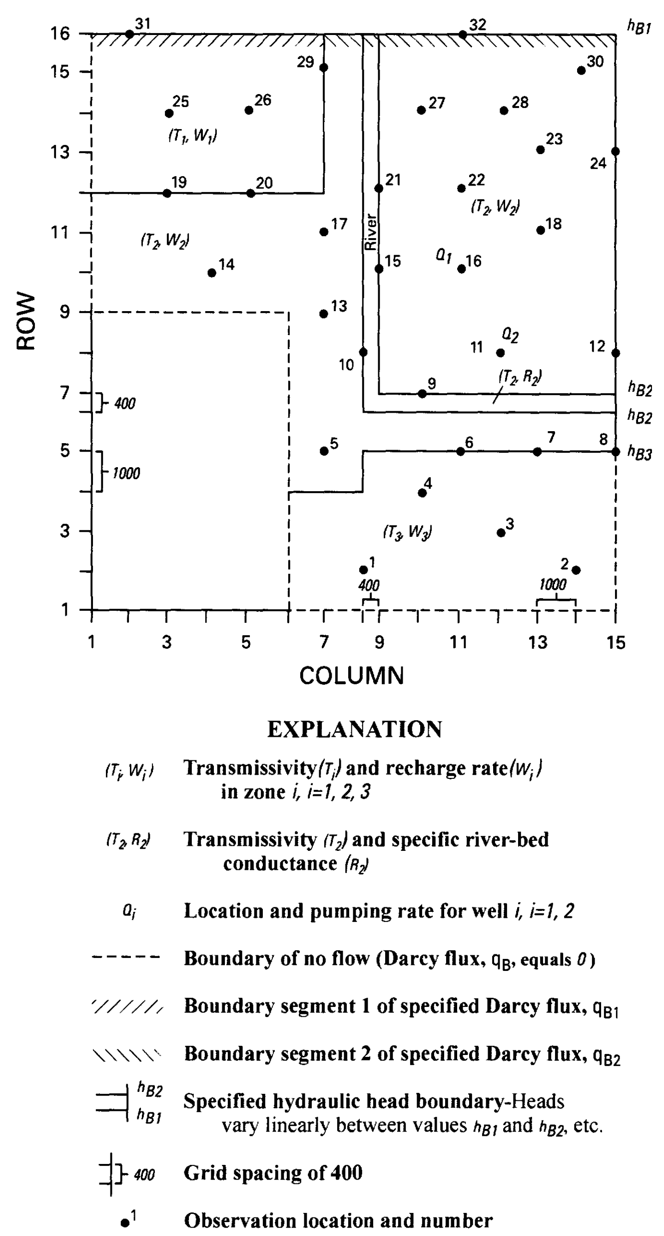


**Figure 7-10.** Assumed hydrogeology of the two-dimensional, ground-water flow system for example 2.

of mean error; they are variable and again are small near the river. Because of its permeable bed, the river nearly forms a specified head boundary of river elevation 4.5; hence, the smaller errors and variances of errors near it. Skewness ranges from $-0.661$ to $0.964$, with 13 negative and 19 positive values. Most negative values are associated with the positive errors. Kurtosis ranges from 2.95 to 4.87, with all but one, except along the specified head boundary, being greater than 3. Thus, even though the mean errors are not large, the error distribution is not normal. It is skewed and generally leptokurtic, as for example 1.

Most off-diagonal second moments and correlations are small. Out of the 496 distinct correlations, 44 lie between –0.2 and 1.0; 12 lie between 0.5 and 1.0; 3 lie between 0.8 and 1.0, being 0.826, 0.832, and 0.837 between observations 28 and 32, 14 and 20, and 14 and 19, respectively; 6 lie between –0.36 and –0.26, and are between observation 15 and each of observations 18, 21, 23, 28, and 32. Most large correlations involve observation 13 and higher numbers. As is shown later in this section, these small numbers of correlations are sufficient to yield significant model-error influences.

***Nonlinearity measures using*** $\omega = \hat{\omega}$. Values of the nonlinearity measures $\hat{N}$ and $\hat{N}_{\min}$ were computed using $\hat{\omega}$ obtained in the same way as for example 1. Values of $\hat{M}_{\min}$, $\hat{B}_U$, and $\hat{B}_L$ were computed using the same $\hat{\omega}$ for the six parameters of $\ln T$ and $W$, and for hydraulic head $h_p$ at grid point (13, 9), the point selected for prediction. Values of $\hat{N}$ and $\hat{N}_{\min}$ are $\hat{N} = 1.43$ and $\hat{N}_{\min} = 0.0157$. Values of $\hat{M}_{\min}$, $\hat{B}_U$, and $\hat{B}_L$ are all less than, or on the order of, $10^{-3}$, the largest being $\hat{M}_{\min} = 8.06 \times 10^{-3}$ for $W_1$. Thus, only total model nonlinearity is

**Table 7-7.** Model specifications for example 2.

| | |
|---|---|
| Drift transmissivities: | $\exp(\bar{\theta}_1) = \bar{T}_1 = 50$ |
| | $\exp(\bar{\theta}_2) = \bar{T}_2 = 500$ |
| | $\exp(\bar{\theta}_3) = \bar{T}_3 = 20$ |
| Drift recharge rates: | $\bar{\theta}_4 = \bar{W}_1 = 0.0003$ |
| | $\bar{\theta}_5 = \bar{W}_2 = -0.0001$ |
| | $\bar{\theta}_6 = \bar{W}_3 = 0.0002$ |
| Known streambed specific conductance: | $R_2 = 0.1$ |
| Known river elevation above an arbitrary datum: | 4.5 |
| Specified hydraulic heads: | $h_{B1} = 10$ |
| | $h_{B2} = 5$ |
| | $h_{B3} = 5.5$ |
| Known fluxes: | $q_{B1} = 0.5$ |
| | $q_{B2} = 0.28$ |
| Known pumping rates: | $Q_1 = -100{,}000$ |
| | $Q_2 = -50{,}000$ |
| Standard deviation of the $\ln T$ process: | $0.5\sigma_\beta$, with $\sigma_\beta = 1$ |
| Standard deviation of the $W$ process: | $0.0001\sigma_\beta$, with $\sigma_\beta = 1$ |
| Observation-error covariance matrix: | $\mathbf{V}_\varepsilon = \mathbf{I}, \sigma_\varepsilon = 0.1$ |
| Number of blocks: | $N = 162$ |
| Number of observations: | $n = 32$ |
| Number of parameters: | $p = 9$ |

significant, and both model intrinsic, and model combined intrinsic, types of nonlinearity are probably negligible. Values of $(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))'\hat{\boldsymbol{\omega}}^{1/2}\hat{\mathbf{R}}\hat{\boldsymbol{\omega}}^{1/2}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))$ and $(\mathbf{Y} - \mathbf{f}(\gamma\tilde{\boldsymbol{\theta}}))'\hat{\boldsymbol{\omega}}^{1/2}(\hat{\mathbf{R}} - \hat{\mathbf{Q}}\hat{\mathbf{Q}}'/\hat{\mathbf{Q}}'\hat{\mathbf{Q}})\hat{\boldsymbol{\omega}}^{1/2}(\mathbf{Y} - \mathbf{f}(\gamma\tilde{\boldsymbol{\theta}}))$ obtained using sensitivities **Df** and results discussed next were always near zero, which corresponds with this conclusion.

***Regression results and analysis of residuals using*** $\boldsymbol{\omega} = \hat{\boldsymbol{\omega}}$. Hydraulic head data **Y** used for the regression were obtained by adding a vector of zero-mean random normal deviates (small observation errors) having a standard deviation of 0.1 to a realization $\mathbf{f}(\boldsymbol{\beta})$ of the Monte Carlo process. As for example 1, this increases the diagonal elements of the second moment matrix (table 7-8) by only 0.01, so that again model error completely dominates the stochastic process. Initially, the weight matrix used was $\hat{\boldsymbol{\omega}}$. Partial results for $\boldsymbol{\omega} = \boldsymbol{\Omega}^{-1}$ and $\boldsymbol{\omega} = \mathbf{I}$ are considered later in this section.

The data **Y** and residuals $\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$ are illustrated in map form in figure 7-11. There seems to be no systematic pattern in the residuals. Weighted residuals $\hat{\boldsymbol{\omega}}^{1/2}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}))$ are plotted with their theoretically correct measures in the same way as for example 1 in figure 7-12. For comparison the weighted residuals are plotted in figure 7-13 with theoretical measures of the (incorrect) distribution $N(\mathbf{0}, (\mathbf{I} - \hat{\mathbf{R}})S(\hat{\boldsymbol{\theta}})/(n-p))$ that generally would be used for field studies.

**Table 7-8.** Values of mean model function, mean error, second moment, variance, skewness, and kurtosis for the distribution of $f_i(\beta) - f_i(\gamma\theta_*)$ at observation points, $i$, for example 2.

| Obs. no. | Mean model function | Mean error | Second moment | Variance | Skewness* | Kurtosis+ |
|---|---|---|---|---|---|---|
| 1 | 58.1 | -2.06 | 52.8 | 48.6 | 0.270 | 3.30 |
| 2 | 74.2 | -2.10 | 86.5 | 82.1 | 0.245 | 3.22 |
| 3 | 57.6 | -2.26 | 49.4 | 44.3 | 0.206 | 3.18 |
| 4 | 29.6 | -1.30 | 25.6 | 23.9 | 0.316 | 3.30 |
| 5 | 6.80 | -0.077 | 0.318 | 0.312 | 0.563 | 3.65 |
| 6 | 5.74 | -0.032 | 0.149 | 0.148 | 0.821 | 4.12 |
| 7 | 5.83 | -0.031 | 0.173 | 0.172 | 0.964 | 4.87 |
| 8 | 5.50 | 0 | 0 | 0 | 0 | 3 |
| 9 | 4.20 | 0.003 | 0.007 | 0.007 | -0.661 | 3.94 |
| 10 | 4.50 | -0.002 | <0.001 | <0.001 | -0.547 | 3.94 |
| 11 | -40.7 | 2.45 | 83.4 | 77.4 | -0.330 | 3.27 |
| 12 | 5.56 | 0 | 0 | 0 | 0 | 3 |
| 13 | 5.64 | -0.058 | 0.115 | 0.112 | 0.672 | 3.77 |
| 14 | 12.1 | -0.122 | 1.84 | 1.83 | 0.432 | 3.37 |
| 15 | 3.70 | 0.008 | 0.026 | 0.025 | -0.265 | 2.95 |
| 16 | -85.0 | 4.15 | 205.7 | 188.5 | -0.632 | 3.83 |
| 17 | 6.29 | -0.079 | 0.215 | 0.209 | 0.552 | 3.65 |
| 18 | -14.9 | 1.18 | 14.2 | 12.8 | -0.364 | 3.19 |
| 19 | 16.6 | -0.198 | 3.03 | 2.99 | 0.335 | 3.23 |
| 20 | 12.4 | -0.146 | 1.87 | 1.85 | 0.368 | 3.28 |
| 21 | 4.19 | 0.001 | 0.006 | 0.006 | -0.519 | 3.40 |
| 22 | -16.7 | 1.24 | 13.5 | 11.9 | -0.277 | 3.24 |
| 23 | -3.25 | 0.628 | 4.08 | 3.69 | -0.389 | 3.26 |
| 24 | 8.33 | 0 | 0 | 0 | 0 | 3 |
| 25 | 53.8 | -1.60 | 20.6 | 18.1 | 0.149 | 3.20 |
| 26 | 38.3 | -1.28 | 16.2 | 14.6 | 0.233 | 3.44 |
| 27 | 0.342 | 0.325 | 1.45 | 1.35 | -0.615 | 3.54 |
| 28 | -2.11 | 0.512 | 2.60 | 2.34 | -0.340 | 3.12 |
| 29 | 7.36 | -0.040 | 0.860 | 0.859 | 0.881 | 4.30 |
| 30 | 5.57 | 0.199 | 1.19 | 1.15 | -0.622 | 3.68 |
| 31 | 83.1 | -1.83 | 25.4 | 22.0 | 0.323 | 3.66 |
| 32 | 1.50 | 0.236 | 1.06 | 1.01 | -0.511 | 3.46 |

*Compare with the theoretical value of 0 for a normal distribution of $f_i(\beta) - f_i(\gamma\theta_*)$.

+Compare with the theoretical value of 3 for a normal distribution of $f_i(\beta) - f_i(\gamma\theta_*)$.

As in example 1, 10,000 Monte Carlo simulations were used to obtain both sets of theoretical measures. All but one of the weighted residuals are contained within the $\pm 2$ standard deviation limits for the theoretically correct distribution. The limits for the correct distribution allow for greater variability in tail values from set to set of weighted residuals than does the incorrect distribution. However, all but two of the weighted residuals are contained within the limits for the incorrect distribution, and the data are well within the limits at the tails. Both sets of measures display similar S shapes that appear to be reflected in the weighted residuals. Hence, the weighted residuals appear to follow the normal distribution (4-42) expected for a model having negligible model and system types of intrinsic nonlinearity. Furthermore, neither the map distribution of residuals nor the weighted residuals plot that would be used for field studies detect the influence of correlation of model errors.

The plot of weighted residuals in relation to weighted function values $\hat{\omega}_i^{1/2}\mathbf{f}(\gamma\hat{\theta})$ is shown in figure 7-14. Except for the three near-zero weighted residual values beyond the weighted function value of 50, the plot does not appear abnormal. The three cited values occur for specified head nodes 8, 12, and 24 for which the data values are only subject to the small observation errors and the weights are large. Therefore, the large weights make the values of $\hat{\omega}_i^{1/2}\mathbf{f}(\gamma\hat{\theta})$ large, and the linear segments of head distribution along the boundary are fitted closely to only three values of head on the boundary. Thus, this plot too does not show any obvious effects of model-error correlation. Finally, neither the mean weighted residual value of $-0.0428$ nor the slope of $-3.07\times10^{-3}$ are large in magnitude, which corresponds with the other indications of small model and system types of intrinsic nonlinearity.

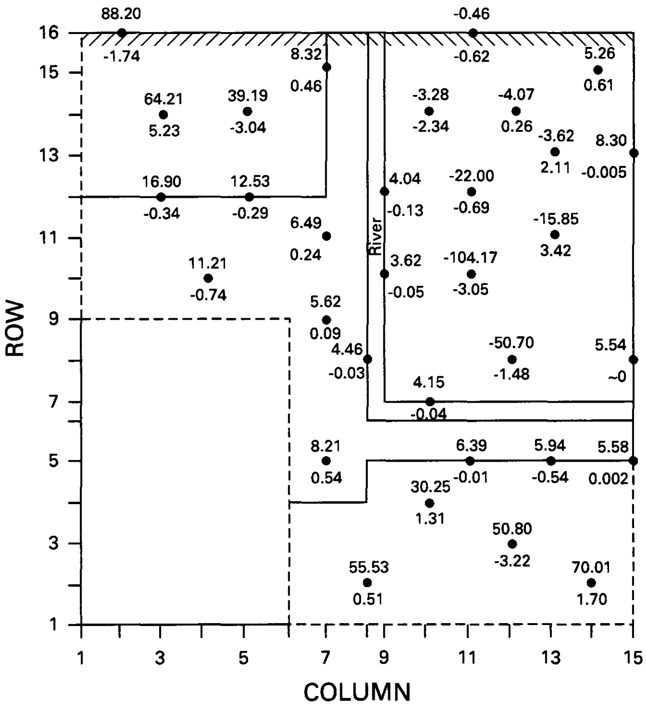


**Figure 7-11.** Hydraulic head data $Y_i$ (upper number) and residuals $Y_i - f_i(\gamma\hat{\theta})$ (lower number) at observation points $i$ for $\omega = \hat{\omega}$ for example 2.

***Confidence and prediction intervals using*** $\omega = \hat{\omega}$. Values of $\theta_{*i}$, $i = 1,2,\ldots,6$, and $h_{p*}$, their estimates $\hat{\theta}_i$ and $\hat{h}_p$, and both their corrected 95 percent linearized, and corrected 95 percent nonlinear confidence intervals are shown in figure 7-15. All are computed in the same manner as for example 1. Linear and nonlinear prediction intervals for $Y_p$, computed using $\hat{h}_p \pm t_{\alpha/2}(n-p)(c_p S(\hat{\theta})(\mathbf{D}\hat{g}(\mathbf{D}\hat{\mathbf{f}}'\hat{\omega}\mathbf{D}\hat{\mathbf{f}})^{-1}\mathbf{D}\hat{g}' + \hat{\omega}_p^{-1})/(n-p))^{1/2}$ and (5-101), also are shown on the figure. Correction factors $c_c$ and $c_p$ used in the calculations are listed in table 7-9. The effects of model nonlinearity are small; linear and nonlinear confidence intervals have nearly the same size. The effects of model error are variable. The value of $c_c$ is 4.22 for $h_p$, so an uncorrected confidence interval for $h_{p*}$ would be only about half the size it should be. However, an uncorrected confidence interval for $\ln T_{*1}$ or $W_{*2}$ would be about 81 percent of its corrected size. The value of $\ln T_{*3}$ is slightly outside of both its linear and its nonlinear confidence intervals.

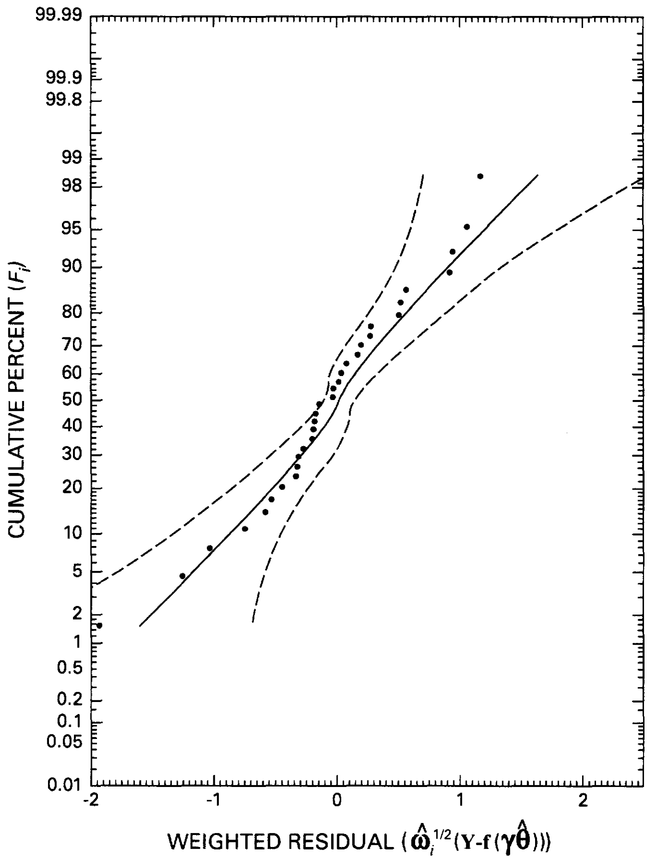


**Figure 7-12.** Probability plot of weighted residuals $\hat{\omega}_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ (dots), sample mean (solid line) of ordered, simulated, weighted residuals from the theoretically correct distribution, and plus and minus 2 standard deviation limits (dashed lines) of the ordered, simulated, weighted residuals for example 2.

Finally, the prediction interval for $Y_p$ contains its predicted value, which is considerably different than $h_{p*}$. This difference shows the large effect of heterogeneity in $T$, $W$, or both. Note the small value of $c_p$ compared with the corresponding value of $c_c$, which is why confidence and prediction intervals have similar sizes in spite of the large predicted error variance of approximately $\hat{\omega}_p^{-1}b\sigma_\varepsilon^2 = 19.6$.

***Monte Carlo accuracy checks using*** $\omega = \hat{\omega}$. A Monte Carlo analysis was performed to check the accuracy of the confidence and prediction intervals. For each realization a data set **Y** was generated as for the above example; then values of $\theta_*$, $h_{p*}$ or $Y_p$, $\hat{\theta}$, and $\hat{h}_p$ or $\tilde{Y}_p$ were computed; finally a nonlinear confidence or prediction interval was computed and checked to see if it contained a value $\ln T_{*i}, W_{*i}, i = 1,2,3, h_{p*}$, or $Y_p$. For comparison, the analysis also was performed for uncorrected confidence intervals. Results for 500 realizations are given in table 7-10. Corrected confidence and prediction intervals appear to be accurate. The average containment probability for the confidence intervals is 0.945, and the lowest is 0.934. For the prediction interval the containment probability of 0.952 is nearly exact. In contrast, uncorrected confidence intervals are too small, with the largest containment probability being 0.892 and the smallest 0.656. These results show the importance of correcting confidence intervals for model-error correlations. Because $c_p = 0.902$, uncorrected prediction intervals would be about 5 percent larger than corrected ones, and so would be accurate or slightly conservative. The theory presented in section 5, indicates that uncorrected prediction intervals should be more accurate than corresponding confidence intervals if the value of $\omega_p$ used is accurate.

The theory predicts that, if a model has small model and system types of intrinsic nonlinearity, then $E(\mathbf{f}(\gamma\hat{\theta}))$ approximately equals $E(\mathbf{f}(\beta))$ as given by (4-36) and additionally, if the model and system types of combined intrinsic nonlinearity are small, then $E(g(\gamma\hat{\theta}))$ approximately equals $E(g(\beta))$ as given in (4-46). These ideas were checked using the Monte

Carlo results. All values of $E(f_i(\beta))$ are predicted to within less than 1 percent by $E(f_i(\gamma\hat{\theta}))$ except at observations number $i = 23$, 27, and 28, where the values are 0.0783, 0.0588, and 0.0761 units greater than $E(f_i(\gamma\hat{\theta}))$. These errors represent –2.39, 18.8, and –3.60 percent of $E(f_i(\beta))$, respectively. For the six parameters, the percent changes of $E(\hat{\theta}_i)$ from the drift parameters $\bar{\theta}_i = E(\theta_{*i})$, $i = 1,2,\ldots,6$, are 1.18, 0.982, 2.29, 7.35, 6.74, and 4.41. In addition, the percent decrease of $E(\hat{h}_p)$ from $\bar{h}_p$ is –0.789 percent. Thus, the biases are small, which corresponds with the other indications of small model and system types of intrinsic nonlinearity.

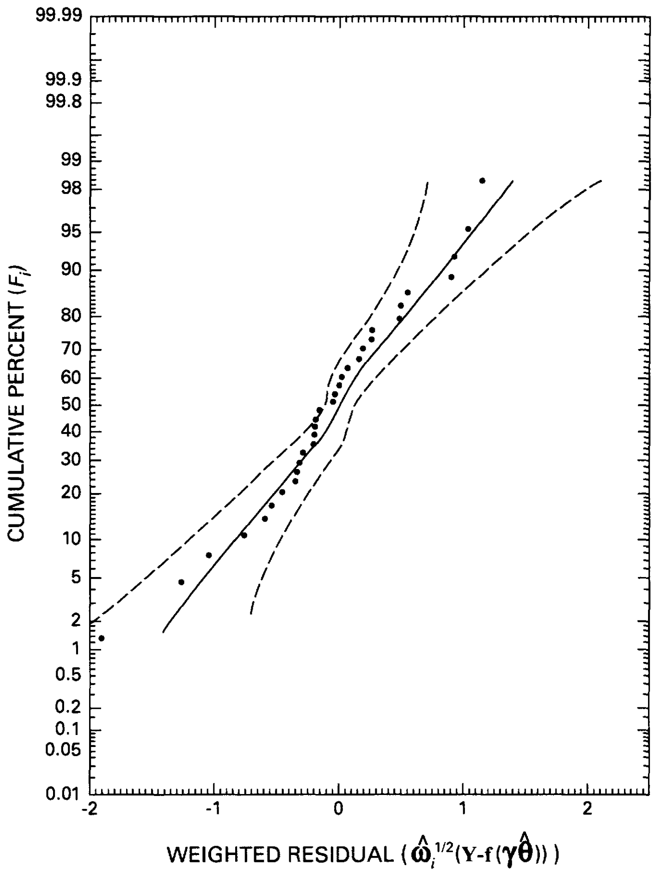


**Figure 7-13.** Probability plot of weighted residuals $\hat{\omega}_i^{1/2}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}))$ (dots), sample mean (solid line) of ordered, simulated, weighted residuals from the incorrect distribution $N(\mathbf{0}, (\mathbf{I} - \hat{\mathbf{R}})S(\hat{\theta})/(n-p))$, and plus and minus 2 standard deviation limits (dashed lines) of the ordered, simulated, weighted residuals from the same distribution for example 2.

***Results for alternative weight matrices.*** As for example 1, Gauss-Markov ($\omega = \Omega^{-1}$) and ordinary least squares ($\omega = \mathbf{I}$) regressions also were performed. In contrast to the results of example 1, figure 7-16 shows that ordinary least squares estimates are noticeably different from Gauss-Markov estimates for $\ln T_3$, $W_2$, $W_3$, and $h_p$, and corrected, linearized confidence intervals for ordinary least squares are different from (generally larger than) the same intervals for Gauss-Markov estimates. Comparison with figure 7-15 shows that Gauss-Markov estimates and confidence intervals are generally closer to corresponding estimates and intervals using $\omega = \hat{\omega}$. Thus, in contrast to example 1, Gauss-Markov estimation and regression using $\omega = \hat{\omega}$ produce very similar estimates and confidence intervals, if the confidence intervals using $\omega = \hat{\omega}$ are corrected for model-error correlation. Finally, as for example 1, neither set of residuals (not illustrated) appears to differ from one expected for the theoretical zero-mean normal distribution.

***Results for an unknown weight matrix.*** The entire analysis yielding regression estimates, confidence intervals, and prediction intervals was repeated for the case where $\Omega$ and $\hat{\omega}$ are unknown. For this case, the observations must be grouped using assumptions regarding similarities of model and observation errors. The three error groups illustrated in figure 7-17 were obtained according to the following criteria.

Group 1-- Upland areas of similar hydrogeology (apparent $T$ and $W$); large magnitudes of sensitivities and most residuals based on ordinary least squares.

Group 2-- Lowland area of nearly uniform hydrogeology (apparent $T$ and $W$); small magnitudes of sensitivities and most residuals based on ordinary least squares.

Group 3-- Lowland area of hydrogeology similar to that of group 2; large magnitudes of sensitivities and most residuals based on ordinary least squares.

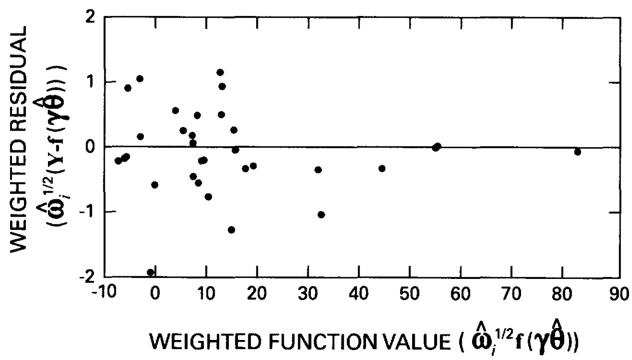


**Figure 7-14.** Plot of weighted residuals $\hat{\omega}_i^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ in relation to weighted function values $\hat{\omega}_i^{1/2}\mathbf{f}(\gamma\hat{\theta})$ for example 2.

The theoretical weight matrix $\omega_G$ (which would be unknown for field studies) was approximated by inverting the diagonal matrix of group averages of $\Omega_{ii}$ values. Elements $\omega_{G11,11}$ and $\omega_{G16,16}$ for observations 11 and 16 (the pumping wells) were additionally weighted by the ratio of the group average to the average of the two second moments for these observations. This ratio also is used as a multiplier occupying the position of a weight in $\ell(\theta)$. The ratio would be unknown and would have to be estimated, if needed, for field studies.

Values of the nonlinearity measures were computed in the same manner as computed previously, except weight matrix $\omega_G$ was used in the present case. The value of $\hat{N}$ is 1.58 and the value of $\hat{N}_{\min}$ is 0.0230, only slightly larger than those obtained using $\hat{\omega}$. The values of $\hat{M}_{\min}$, $\hat{B}_U$, and $\hat{B}_L$ for the six parameters of $\ln T$ and $W$, and for $h_p$, are all less than, or on the order of, $10^{-3}$, with the largest being $\hat{M}_{\min}=6.66\times10^{-3}$ for $W_1$. Thus, only total model nonlinearity is significant, and model intrinsic, and model combined intrinsic, types of nonlinearity are negligible. As before, values of $(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))'\omega_G^{1/2}\mathbf{R}\omega_G^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ and $(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))'\omega_G^{1/2}(\mathbf{R}-\mathbf{QQ}'/\mathbf{Q}'\mathbf{Q})\omega_G^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))$ (where $\mathbf{R}$ and $\mathbf{Q}$ are computed using sensitivities $\mathbf{Df}$ and $\mathbf{Dg}$ and weight matrix $\omega_G$) are near zero, which agrees with this conclusion.

The same hydraulic-head data set $\mathbf{Y}$ as used for the previous analyses also was used for the present one. The data and residuals $\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})$ are illustrated in map form in figure 7-18, and weighted residuals are plotted together with correct and incorrect theoretical distributional measures in figures 7-19 and 7-20. Weights used to compute the theoretical measures are the ones estimated by the regression, which are $w_{Gk}$, $k=1,2,3$, as given by (4-49) with equality replacing proportionality. Otherwise, the plots were constructed in the same manner as were the analogous ones in figures 7-11 - 7-13. Features of the plots in figures 7-18 - 7-20 are very similar to features in figures 7-11 - 7-13 discussed previously. Weighted residuals are plotted in

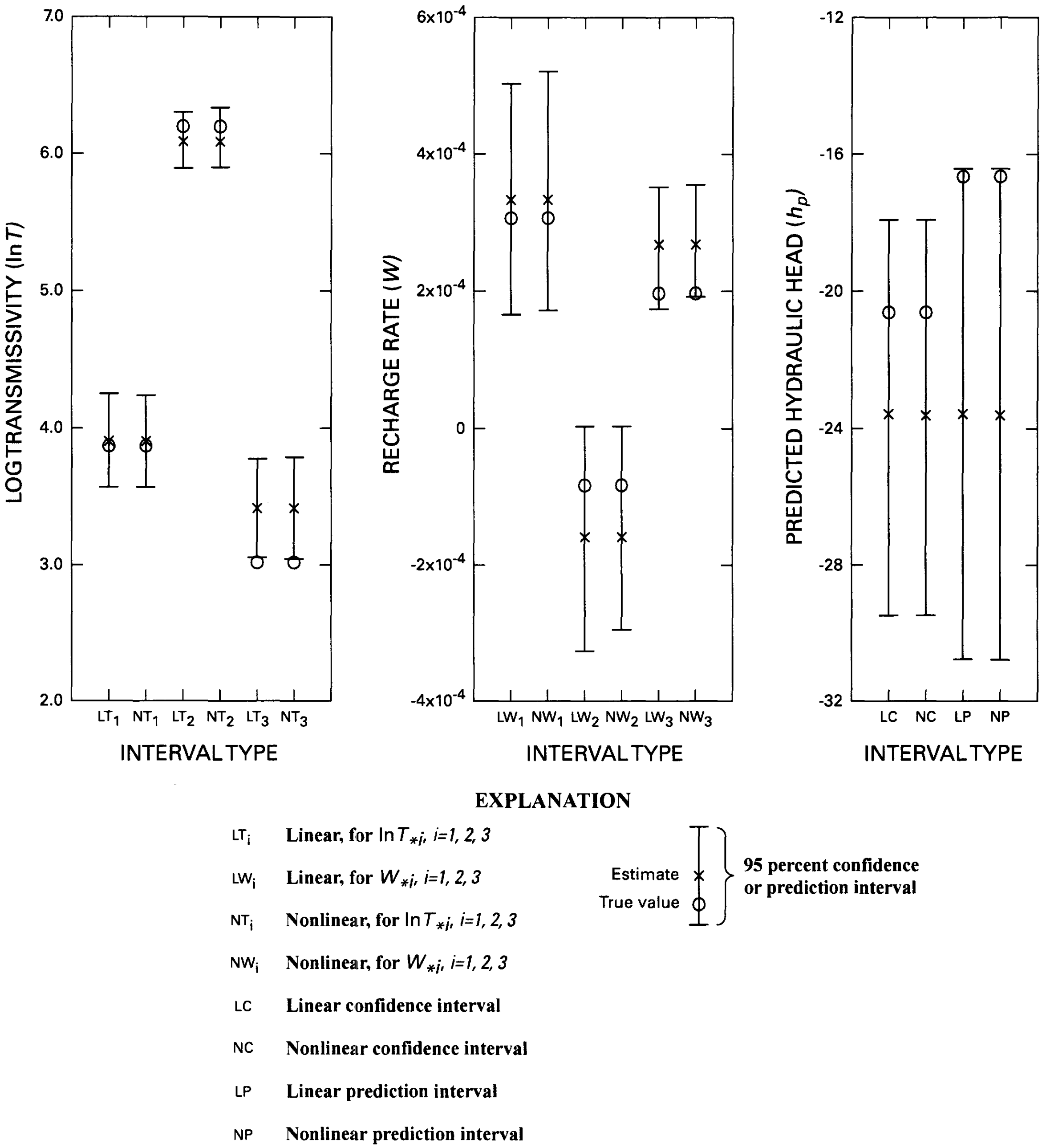


**Figure 7-15.** True zonal values $\ln T_{*i}$, $W_{*i}$ $(i = 1,2,3)$, and $h_{p*}$, their estimates obtained using $\omega = \hat{\omega}$, and their 95 percent confidence or prediction intervals for example 2.

relation to weighted function values in figure 7-21. Note that the three residuals for observations on the constant head boundary do not correspond to large weighted function values in the present case because group weights were used for them. No abnormalities are obvious, so that model-error correlation is not obvious from the plots. Also, neither the mean residual value of 0.0788 nor the slope of $-2.46 \times 10^{-3}$ are large in magnitude, which again agrees with the other indications of small model and system types of intrinsic nonlinearity.

**Table 7-9.** Correction factors for example 2.

[$a$ is computed using (5-13); $b$ is computed using (5-11); $c_r$ is computed using (5-19); $c_c$ is computed using (5-50) with $\gamma_I = 0$; $c_p$ is computed using (5-96) with $\gamma_I = 0$; $\xi$ is computed using (6-62); $V_{mx}/\mathbf{Q}'\mathbf{Q}$ is computed using (5-56); weight matrix used is $\hat{\boldsymbol{\omega}}$, $\mathbf{I}$, or $\boldsymbol{\omega}_G$ as appropriate; $\hat{\omega}_p^{-1} b \sigma_\varepsilon^2 = 19.624$; $\omega_{G3}^{-1}\sigma_\varepsilon^2 = 4.1496$; -, not computed]

| Weight matrix | Variable | $a$ | $b$ | $c_r$ | $c_c$ | $c_p$ | $\xi$ | $V_{mx}/\mathbf{Q}'\mathbf{Q}$ |
|---|---|---|---|---|---|---|---|---|
| $\hat{\boldsymbol{\omega}}$ | $\ln T_1$ | 1.70 | 1.00 | 2.34 | 1.51 | - | 1.10 | 5.47 |
| | $\ln T_2$ | 1.70 | 1.00 | 2.34 | 3.62 | - | 2.63 | 9.47 |
| | $\ln T_3$ | 1.70 | 1.00 | 2.34 | 1.66 | - | 1.20 | 7.19 |
| | $W_1$ | 1.70 | 1.00 | 2.34 | 1.55 | - | 1.13 | 7.08 |
| | $W_2$ | 1.70 | 1.00 | 2.34 | 1.51 | - | 1.10 | 4.87 |
| | $W_3$ | 1.70 | 1.00 | 2.34 | 1.55 | - | 1.13 | 8.38 |
| | $h_p$ or $Y_p$ | 1.70 | 1.00 | 2.34 | 4.22 | 0.902 | 3.07 | 9.78 |
| $\mathbf{I}$ | $\ln T_1$ | 2.27 | 19.0 | 4.51 | 1.88 | - | 0.948 | 3.42 |
| | $\ln T_2$ | 2.27 | 19.0 | 4.51 | 18.1 | - | 9.13 | 3.21 |
| | $\ln T_3$ | 2.27 | 19.0 | 4.51 | 2.19 | - | 1.10 | 5.69 |
| | $W_1$ | 2.27 | 19.0 | 4.51 | 1.89 | - | 0.951 | 3.58 |
| | $W_2$ | 2.27 | 19.0 | 4.51 | 1.33 | - | 0.672 | 6.16 |
| | $W_3$ | 2.27 | 19.0 | 4.51 | 2.11 | - | 1.06 | 5.82 |
| | $h_p$ or $Y_p$ | 2.27 | 19.0 | 4.51 | 8.44 | 1.95 | 4.25 | 8.06 |
| $\boldsymbol{\omega}_G$ | $\ln T_1$ | 1.90 | 0.914 | 2.93 | 1.83 | - | 1.19 | 5.10 |
| | $\ln T_2$ | 1.90 | 0.914 | 2.93 | 7.78 | - | 5.04 | 7.74 |
| | $\ln T_3$ | 1.90 | 0.914 | 2.93 | 2.75 | - | 1.78 | 6.70 |
| | $W_1$ | 1.90 | 0.914 | 2.93 | 1.98 | - | 1.28 | 6.89 |
| | $W_2$ | 1.90 | 0.914 | 2.93 | 1.99 | - | 1.29 | 4.72 |
| | $W_3$ | 1.90 | 0.914 | 2.93 | 2.52 | - | 1.64 | 7.55 |
| | $h_p$ or $Y_p$ | 1.90 | 0.914 | 2.93 | 9.22 | 3.75 | 5.97 | 7.48 |

True values $\theta_{*i}$, $i = 1,2,\ldots,6$, and $h_{p*}$, their estimates, their corrected 95 percent linearized and nonlinear confidence intervals, and the 95 percent linearized and nonlinear prediction interval are shown in figure 7-22. The linearized confidence intervals are computed as in figure 7-15 except weights $w_{Gk}$ replace weights $\hat{\omega}_{ii}$ used before. To correspond with this, group weight $w_{G3}$ replaces $\hat{\omega}_p$ for the prediction interval. Nonlinear confidence intervals are computed using (5-57), and the nonlinear prediction interval is computed using (5-118). To make the intervals as accurate as possible for testing purposes, known factor $\xi$ is used in place of bound $V_{mx}/\mathbf{Q}'\mathbf{Q}$ in (5-57) and the extension of $\xi$ applying for prediction intervals is used in place of $V_{mxa}/(\mathbf{Q}'\mathbf{Q} + \omega_p^{-1})$ in (5-118). This extension is defined as

$$\xi_a = \frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}\mathbf{Q}_a'(\mathbf{W}_a/b)^{\frac{1}{2}}\boldsymbol{\Omega}_a(\mathbf{W}_a/b)^{\frac{1}{2}}\mathbf{Q}_a \tag{7-12}$$

In the present example $\xi$ is computed using $\boldsymbol{\omega} = \boldsymbol{\omega}_G$ and $\xi_a$ is computed using $\mathbf{W}_a = \boldsymbol{\omega}_{Ga}$. Matrix $\boldsymbol{\omega}_{Ga}$ is $\boldsymbol{\omega}_G$ augmented with $\omega_{Gj}$ (in which $j = 3$ here) by analogy with (5-89). Correction factors $c_c$, $c_p$, and $\xi$ used are given in table 7-9; the value of $\xi_a$ is 2.43.

**Table 7-10**. Containment probabilities for 95 percent confidence and prediction intervals obtained using $\boldsymbol{\omega} = \hat{\boldsymbol{\omega}}$ for example 2.

a. Corrected intervals

| Variable | No. outside interval* | Containment probability |
|---|---|---|
| $\ln T_{*1}$ | 27 | 0.946 |
| $\ln T_{*2}$ | 28 | 0.944 |
| $\ln T_{*3}$ | 23 | 0.954 |
| $W_{*1}$ | 32 | 0.936 |
| $W_{*2}$ | 33 | 0.934 |
| $W_{*3}$ | 25 | 0.950 |
| $h_{p*}$ | 26 | 0.948 |
| $Y_p$ | 24 | 0.952 |

* Computed for 500 realizations

b. Uncorrected intervals

| Variable | No. outside Interval* | Containment probability |
|---|---|---|
| $\ln T_{*1}$ | 54 | 0.892 |
| $\ln T_{*2}$ | 156 | 0.688 |
| $\ln T_{*3}$ | 63 | 0.874 |
| $W_{*1}$ | 69 | 0.862 |
| $W_{*2}$ | 56 | 0.888 |
| $W_{*3}$ | 62 | 0.876 |
| $h_{p*}$ | 174 | 0.656 |

* Computed for 500 realizations

Linearized confidence intervals for parameters in figure 7-22 are only slightly larger than corresponding intervals obtained using $\boldsymbol{\omega} = \hat{\boldsymbol{\omega}}$ shown in figure 7-15. The linearized confidence interval for $h_p$ is much smaller than the one in figure 7-15. Nonlinear confidence intervals can be much larger than the corresponding linear intervals, probably because of the nonlinearity resulting from the logarithmic form of the objective function $\ell(\boldsymbol{\theta})$ rather than because of model nonlinearity. The large value of $\xi$ for $h_p$ (5.97 in table 7-9) has apparently combined with this process to yield a large nonlinear confidence interval for $h_{p*}$. Because of this, this confidence interval is much larger than the prediction interval for which $\xi_a$ is much smaller (2.43). When the two intervals were recalculated with no correction, the prediction interval was computed to be (–27.0, –19.0), and the confidence interval was computed to be (–26.0, –19.8), so that the expected larger prediction interval was obtained. As shown next, the confidence interval for $h_{p*}$ is extremely conservative because of the large value of $\xi$. Because the correction factors are computed using constant weight matrices $\boldsymbol{\omega}_G$ or $\boldsymbol{\omega}_{Ga}$, they are only approximate when applied to confidence and prediction intervals computed using $\ell(\boldsymbol{\theta})$ or $\ell_a(\boldsymbol{\theta})$. These intervals use new weights computed for each realization of $\mathbf{Y}$.

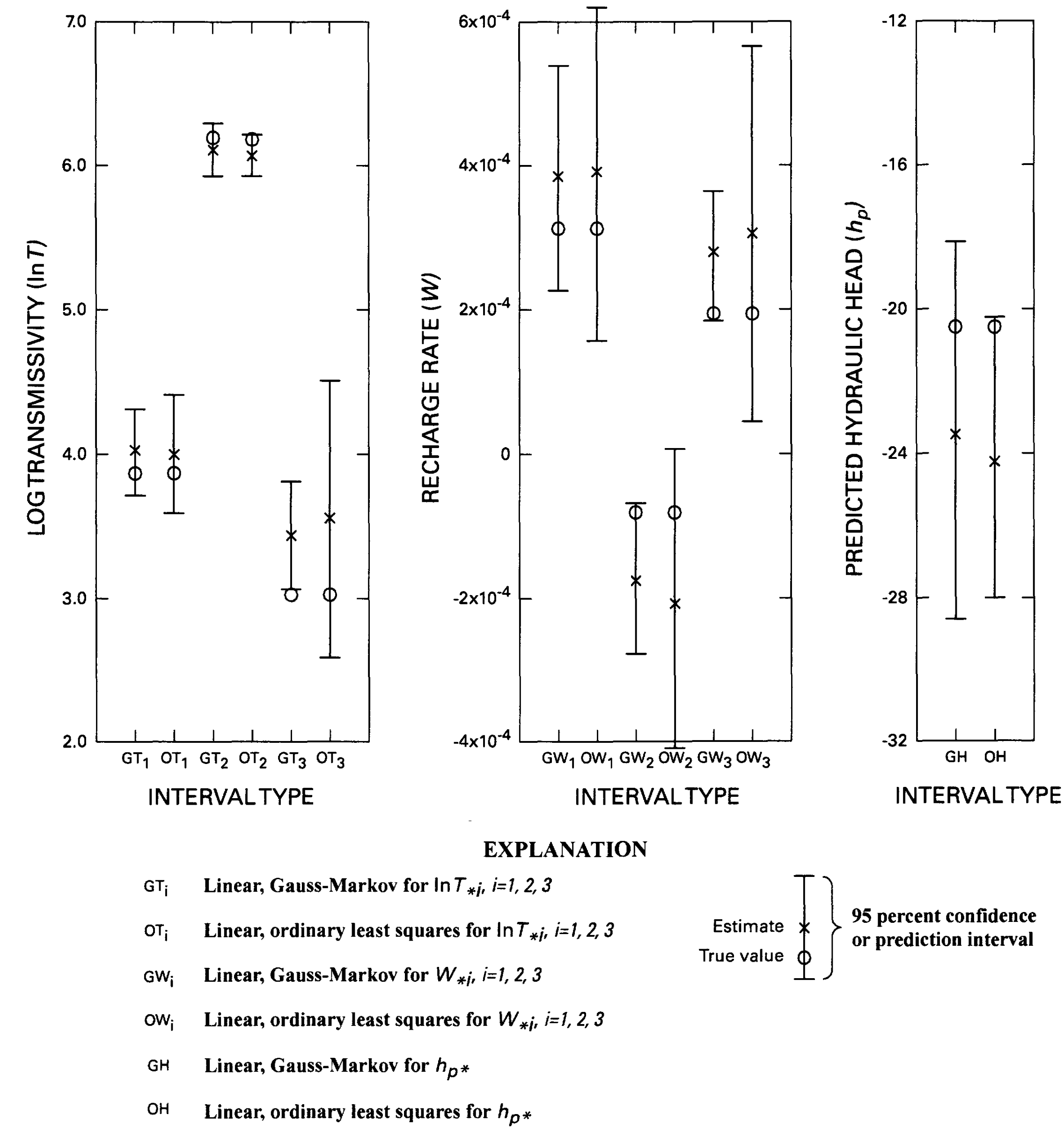


**Figure 7-16.** True zonal values $\ln T_{*i}$, $W_{*i}$ $(i=1,2,3)$, and $h_{p*}$, their estimates obtained using Gauss-Markov estimation or ordinary least squares, and their 95 percent linearized confidence intervals for example 2.

A Monte Carlo analysis was performed to check the containment probabilities by generating data sets **Y** and calculating corresponding nonlinear confidence and prediction intervals as was done using $\omega=\hat{\omega}$. The results are shown in table 7-11. Because of the greater degree of approximation used when $\hat{\omega}$ is unknown than when $\omega=\hat{\omega}$, results when $\hat{\omega}$ is unknown are not as accurate as when $\omega=\hat{\omega}$. The confidence intervals for $\ln T_{*2}$ and $h_{p*}$ are large compared to confidence intervals using $\omega=\hat{\omega}$ (figures 7-15 and 7-22), and are apparently very conservative because their containment probabilities are both 1.00. As discussed in the previous paragraph, this probably results because the correction factors are too large. For example, when the Monte Carlo analysis applying for $h_p$ was rerun using $\xi = 2.8$, a containment probability of 0.948 was obtained. The confidence intervals for quantities other than $\ln T_{*2}$ and $h_{p*}$ seem to be too small because the containment probabilities are all less than 0.95. However, the smallest probability is approximately 0.92, so that the intervals are not greatly in error. This bias probably is because the correction factors are too small. Finally, correction of all confidence intervals and the prediction interval is needed. When the intervals are uncorrected, the largest containment probability is 0.892 and the smallest is only 0.624 (table 7-11).

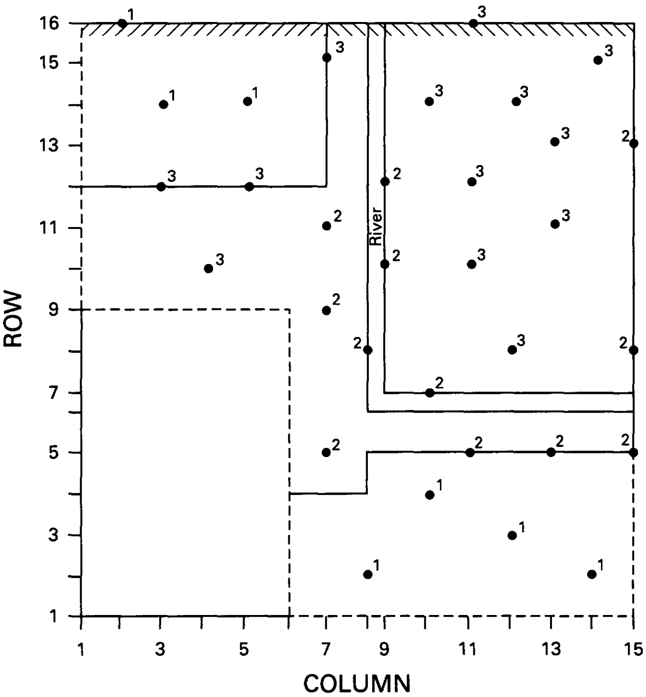


**Figure 7-17.** Error group numbers for observation points for example 2.

It appears that correction factors computed assuming $\omega_G$ is known and used when $\omega_G$ is unknown could be less accurate than correction factors computed and used when $\omega_G$ is known. To check this idea, the Monte Carlo analysis was repeated using $\omega = \omega_G$. The results for corrected confidence intervals are shown in table 7-11c. The results are better than those in table 7-11a in that the very large containment probabilities for $\ln T_{*2}$ and $h_{p*}$ have been decreased and are now accurate, and the remaining probabilities have been increased and are slightly more accurate, averaging approximately 0.934 now, whereas for $\omega_G$ unknown they average approximately 0.928. However, grouping the errors yields confidence intervals that are less accurate than those obtained by not grouping the errors (that is, using $\omega = \hat{\omega}$, table 7-10a). This corresponds with the theory developed in appendix F, which predicts that the likelihood region is most nearly $t^2$ distributed when $\Omega^{-1}$ is approximated as a weight matrix by $\hat{\omega}$. It is encouraging that confidence and prediction intervals using $\omega = \hat{\omega}$, $\omega = \omega_G$, and unknown weights all are of acceptable accuracy when the error structure is dominated by model errors.

***Correction factors and bounds.*** As has been shown in the previous discussions, it is generally necessary to use the correction factors to increase the accuracy of confidence and prediction intervals to an acceptable level. Values of $a$ indicate that $S(\hat{\theta})/(n-p)$ underestimates $b\sigma_\varepsilon^2$ (table 7-9), with the bias for ordinary least squares being the largest ($(n-ap)/(n-p) = 0.503$) as it is for example 1. The smallest bias is obtained using $\omega = \hat{\omega}$, for which $(n-ap)/(n-p) = 0.726$. If the matrix of group averages $\omega_G$ were used as a known weight matrix, then the underestimate would be $(n-ap)/(n-p) = 0.648$. Thus, while the biases

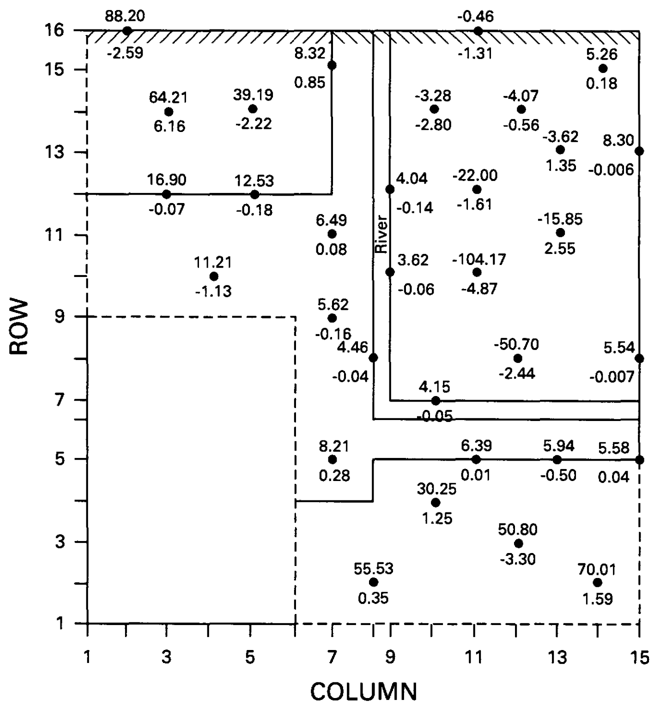


**Figure 7-18.** Hydraulic head data $Y_i$ (upper number) and residuals $Y_i - f_i(\gamma\hat{\theta})$ (lower number) at observation points $i$ for $\omega$ unknown for example 2.

are considerably smaller than those for example 1, they are still significant, and the rankings are predictable by the theory as explained in the previous paragraph. Values of $c_r$ and $c_c$ generally are largest using ordinary least squares and smallest using $\omega = \hat{\omega}$, although some small reverses in value occur. This again conforms with the theory. The approximate bounds for $\xi$ are all large except for the bounds pertaining to $\ln T_2$ using ordinary least squares. The failure of this bound is explained by the fact that $\mathbf{I}$ is not a good approximation of $\hat{\omega}$. All values of $\xi$ obtained using group averages are bounded by $V_{mx}/\mathbf{Q}'\mathbf{Q}$, which conforms with the idea that $\omega_G$ is an adequate approximation of $\hat{\omega}$ and that the error groups are adequate. Use of the bounds often would yield much larger confidence intervals than use of $\xi$.

## Summary of Principal Results

Two examples are analyzed to test the validity and robustness of the theory developed in this report when the model error is large. Example 1 is for one-dimensional, steady-state flow in an aquifer having transmissivity ($T$) that varies stochastically and one dimensionally at small scale and recharge ($W$) that is constant. Example 2 is for two-dimensional, steady-state flow in a zoned aquifer where transmissivity and recharge vary spatially at both large and small scales, the small-scale variations being stochastic.

An analytical solution (7-1) for example 1 allows for potential block-to-block variation in $\ln T$ and $W$ (that is, $\beta$) along the flow path. This solution is simplified in (7-4) to be in terms of constant average $\ln T$ over all blocks ($\underline{\theta}_{*1} = \ln T_*$) and the constant drift value of recharge ($\bar{\theta}_2 = \overline{W}$). The two solutions (one using block-to-block variations in $T$ and $W$ and the other using spatially constant $T$ and $W$) show that the intrinsic nonlinearity of both models is zero and that transformations $\alpha(\beta)$ and $\phi(\theta)$ linearize both models. Even so, system intrinsic nonlinearity may not be small unless the approximations discussed in section 4 are accurate.

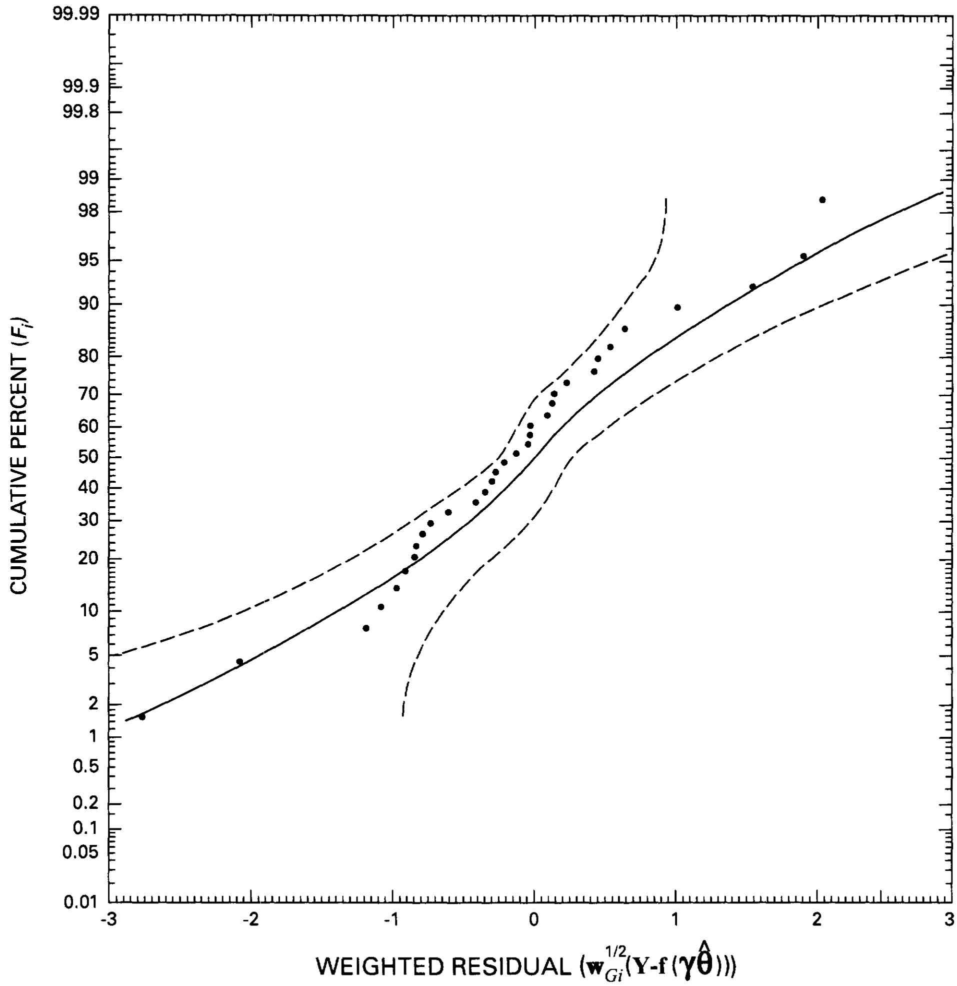


**Figure 7-19.** Probability plot of weighted residuals $\mathbf{w}_{Gi}^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ (dots), sample mean (solid line) of ordered, simulated, weighted residuals from the theoretically correct distribution, and plus and minus 2 standard deviation limits (dashed lines) of the ordered, simulated weighted residuals for example 2. Matrix $\mathbf{w}_G$ is the diagonal matrix $[w_{Gk}]$.

Values of $\beta_i = \ln T_i$, $i = 1,2,\ldots,N$ (where $N$ is the number of blocks), are assumed to be independently and identically normally distributed as $N(\bar{\theta}_1, \sigma_\beta^2)$, where $\sigma_\beta^2 = 0.25$. Values of $\beta_{i+N} = W_i$ are assumed to be constant at the drift value $\bar{\theta}_2$.

The vector of mean errors $E(\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*))$, the matrix of second moments $E(\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*))(\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*))'$, the covariance matrix $Var(\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*))$, and the vectors of skewness and kurtosis for the errors $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ were approximated for 11 observation points using a Monte Carlo method. Mean errors increase steadily from the known head boundary at the lower end of the system to the known flux boundary at the upper end, where they are large. Terms in the second moment matrix and correlations computed from the covariance matrix are large, especially near the known flux boundary. Values of skewness and kurtosis indicate that the distribution of $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ is not normal. These results all show that the system total

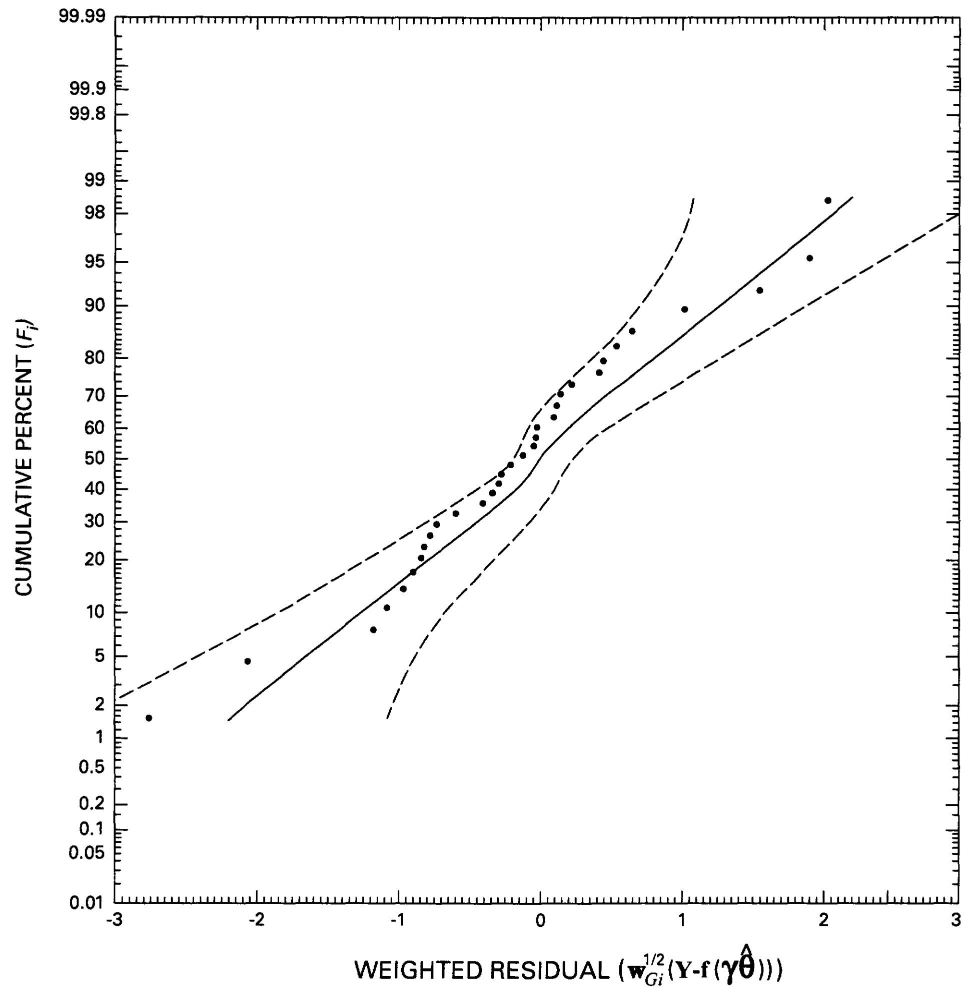


**Figure 7-20.** Probability plot of weighted residuals $\mathbf{w}_{Gi}^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$ (dots), sample mean (solid line) of ordered, simulated, weighted residuals from the incorrect distribution $N(\mathbf{0}, (\mathbf{I}-\mathbf{R})S(\hat{\theta})/(n-p))$, and plus and minus 2 standard deviation limits (dashed lines) of the ordered, simulated, weighted residuals from the same distribution for example 2. Matrix $\mathbf{w}_G$ is used to compute $\mathbf{R}$ and is the diagonal matrix $[w_{Gk}]$.

nonlinearity is large and that model error has the potential of having a large, possibly detrimental, effect on regression modeling of the flow system.

Measures of total model nonlinearity, model intrinsic nonlinearity, and model combined intrinsic nonlinearity defined by (6-56) - (6-65) were computed using weight matrix $\hat{\omega}$, calculated as described in the next paragraph. These values show that the model as written in terms of average values is highly nonlinear, but confirm that it has no model intrinsic nonlinearity. Model combined intrinsic nonlinearity also is zero for $g(\gamma\theta) = h_p$ (the predicted value of hydraulic head at the known flux boundary), but model combined intrinsic nonlinearity for parameters $\ln T$ and $W$ is larger. Subsequent analyses of regression results indicate that these model combined types of intrinsic nonlinearity are both negligible.

Hydraulic head data $\mathbf{Y}$ for regression analysis were obtained by adding a vector of zero-mean random normal deviates having a standard deviation of 0.1 to $\mathbf{f}(\boldsymbol{\beta})$ obtained as a realization from the Monte Carlo process. This simulates the additive influence of a small observation error on the data and adds only 0.01 to the diagonal elements of the second moment matrix; model error completely dominates the process. Thus, analysis using these data tests the robustness of the theory when the assumption of small model-error variances is not satisfied. Each diagonal element of $\hat{\omega}$ used for the regression was obtained by inverting the sum of 0.01 and a diagonal element of the second-moment matrix.

Regression using $\omega = \hat{\omega}$ produced the estimate $\hat{\theta}$. The vectors of residuals $\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta})$ and weighted residuals $\hat{\omega}^{1/2}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}))$ were analyzed for signs of spatial correlation inherited from model-error correlations and non-normality resulting from the non-normal distribution of $\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\theta_*)$. Probability plots were made of weighted residuals together with theoretical measures consisting of the means plus and minus two standard deviation limits of ordered, simulated residuals generated by Monte Carlo simulation from both the theoretically correct distribution of residuals and the (incorrect) distribution that would be used in field studies $(N(\mathbf{0}, (\mathbf{I} - \mathbf{R})S(\hat{\theta})/(n-p))$. The residuals lie within the limits of the correct distribution, but three points lie outside of the limits for the incorrect distribution. The plot of weighted residuals in relation to weighted function values $\hat{\omega}_i^{1/2}\mathbf{f}(\gamma\hat{\theta})$ shows a suspicious, but not abnormal, pattern. Both the mean weighted residual and the slope are negligible, which is predicted if both model and system types of intrinsic nonlinearity are small.

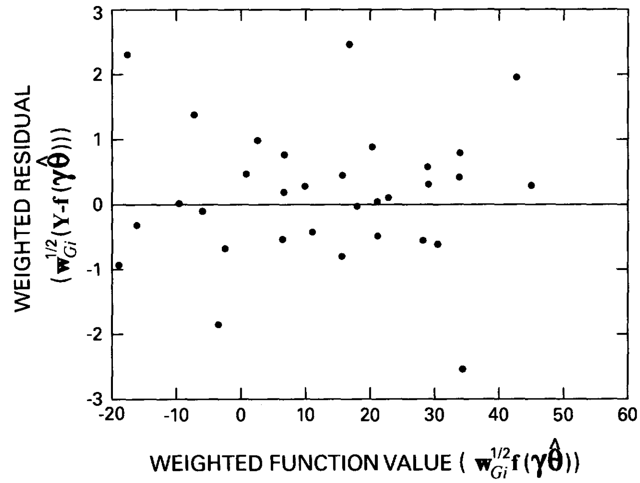


**Figure 7-21.** Plot of weighted residuals $\mathbf{w}_{Gi}^{1/2}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}))$ in relation to weighted function values $\mathbf{w}_{Gi}^{1/2}\mathbf{f}(\gamma\hat{\theta})$ for example 2. Matrix $\mathbf{w}_G$ is the diagonal matrix $[w_{Gk}]$.

Linearized and nonlinear 95 percent confidence intervals for $\ln T_*$, $\overline{W}$, and $g(\gamma\theta_*) = h_{p*}$ demonstrate that correction factors ($c_c$ values) must be used because several uncorrected confidence intervals do not contain their true values $\ln T_*$, $\overline{W}$, or $h_{p*}$, whereas their corrected counterparts do. The nonlinear confidence intervals also show the effect of a severe ill-conditioning problem involving a very large correlation between $\hat{\theta}_1$ ($\ln\hat{T}$) and $\hat{\theta}_2$ ($\hat{W}$). That is, one limit of each of the intervals is unique only for the ratio $W/\ln T$. Thus, the nonlinear confidence intervals for $\ln T_*$ and $\overline{W}$ are unbounded on one side.

Regressions also were conducted using Gauss-Markov estimation and ordinary least squares $(\omega = \mathbf{I})$. Only linearized confidence intervals were computed. When no correction factors are used, linearized confidence intervals obtained using Gauss-Markov estimation (which

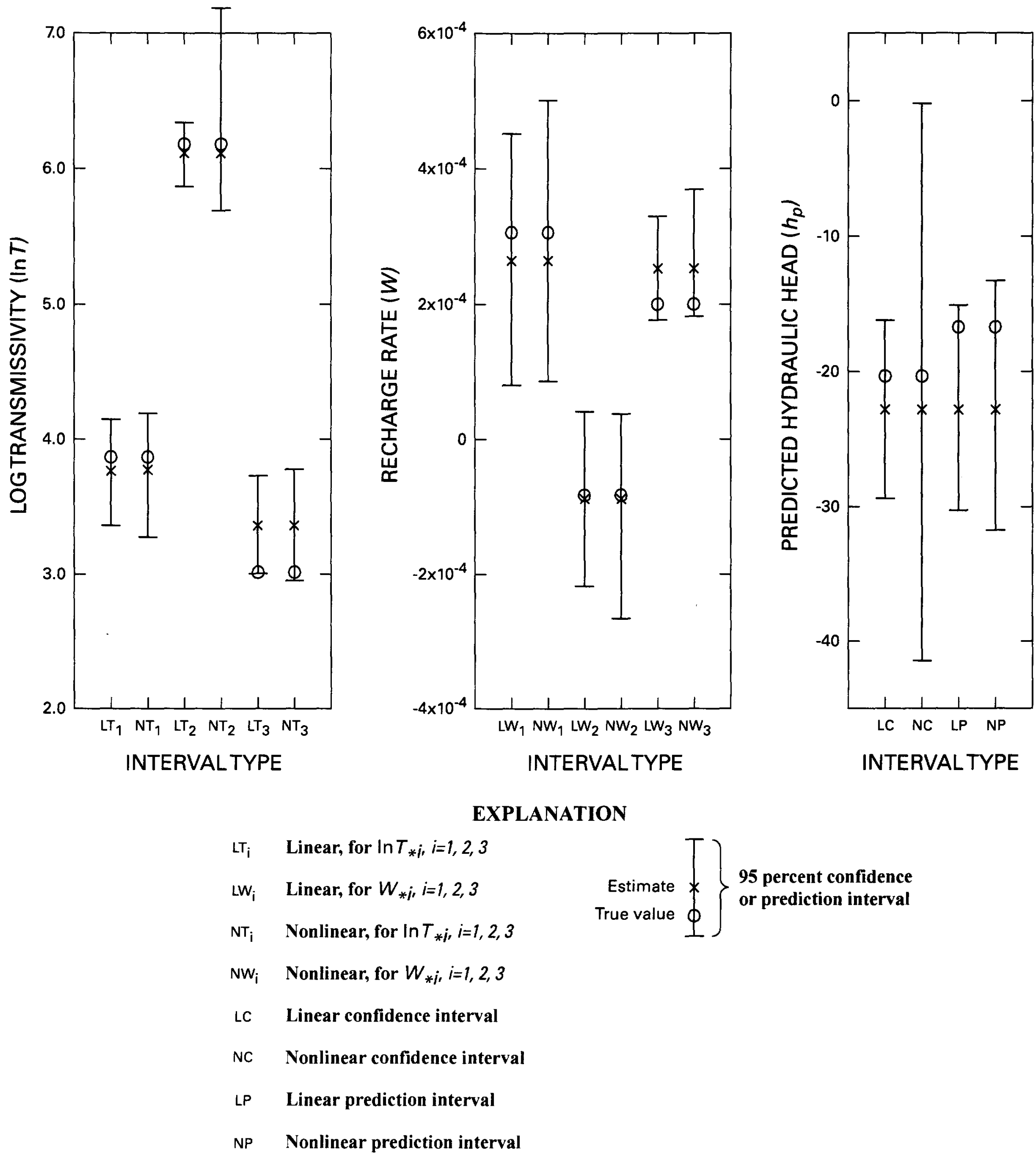


**Figure 7-22.** True zonal values $\ln T_{*i}, W_{*i}$ $(i = 1,2,3)$, and $h_{p*}$, their estimates obtained when $\omega$ is unknown, and their 95 percent confidence or prediction intervals for example 2.

require no correction) are largest and corresponding uncorrected intervals obtained using ordinary least squares are smallest, with uncorrected intervals obtained using $\omega = \hat{\omega}$ lying in between. These results primarily reflect the fact that unless $S(\hat{\theta})/(n-p)$ is divided by a factor that corrects for spatial correlation, this quantity considerably underestimates $E(S(\theta_*))/n$ when $\omega = \hat{\omega}$ (where the factor equals 0.199) and for ordinary least squares (where the factor equals 0.0707). Correction using $c_c$, which contains the factor, produces confidence intervals obtained

**Table 7-11.** Containment probabilities for 95 percent confidence and prediction intervals when $\hat{\omega}$ is unknown for example 2.

a. Corrected intervals

| Variable | No. outside interval* | Containment probability |
|---|---|---|
| $\ln T_{*1}$ | 37 | 0.926 |
| $\ln T_{*2}$ | 0 | 1.00 |
| $\ln T_{*3}$ | 23 | 0.954 |
| $W_{*1}$ | 42 | 0.916 |
| $W_{*2}$ | 38 | 0.924 |
| $W_{*3}$ | 41 | 0.918 |
| $h_{p*}$ | 0 | 1.00 |
| $Y_p$ | 13 | 0.974 |

* Computed for 500 realizations

b. Uncorrected intervals

| Variable | No. outside Interval* | Containment probability |
|---|---|---|
| $\ln T_{*1}$ | 54 | 0.892 |
| $\ln T_{*2}$ | 177 | 0.646 |
| $\ln T_{*3}$ | 77 | 0.846 |
| $W_{*1}$ | 62 | 0.876 |
| $W_{*2}$ | 63 | 0.874 |
| $W_{*3}$ | 85 | 0.830 |
| $h_{p*}$ | 188 | 0.624 |
| $Y_p$ | 153 | 0.694 |

* Computed for 500 realizations

c. Corrected intervals computed using $\omega = \omega_G$ .

| Variable | No. outside interval* | Containment probability |
|---|---|---|
| $\ln T_{*1}$ | 33 | 0.934 |
| $\ln T_{*2}$ | 27 | 0.946 |
| $\ln T_{*3}$ | 29 | 0.942 |
| $W_{*1}$ | 34 | 0.932 |
| $W_{*2}$ | 31 | 0.938 |
| $W_{*3}$ | 37 | 0.926 |
| $h_{p*}$ | 26 | 0.948 |

* Computed for 500 realizations

using $\omega = \hat{\omega}$ that are larger than corrected intervals obtained using ordinary least squares, which are in turn similar to the intervals obtained using Gauss-Markov estimation.

The case involving correlated values of $\ln T_i$ was examined to test the concept outlined in section 3 that when correlation is manifested as a trend, the trend should be considered to be the drift. It was found that model error and the effect of correlation could be reduced considerably by removing the trend from linearly trending $\ln T$ data generated from a stationary process having exponential covariance with a long correlation length. The need to use correction factors when computing confidence intervals was eliminated.

For example 2, the stochastic process involves both $T$ and $W$ varying grid cell by grid cell. Drift values are defined for three zones. All grid-cell values $\ln T_i$ and $W_i$, where $i$ is a grid-cell number and there are $N$ cells, are assumed to be statistically independent, so that $\beta_i = \ln T_i$ has the normal distribution $N(\bar{\theta}_k, 0.5^2)$, $k = 1,2,3$, and $\beta_{i+N} = W_i$ has the normal distribution $N(\bar{\theta}_k, 0.0001^2)$, where $k = 4,5,6$.

As for example 1, the vector of mean errors $E(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\boldsymbol{\theta}_*))$, the matrix of second moments $E(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\boldsymbol{\theta}_*))(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'$, the covariance matrix $Var(\mathbf{f}(\boldsymbol{\beta}) - \mathbf{f}(\gamma\boldsymbol{\theta}_*))$, and the vectors of skewness and kurtosis were approximated for 32 observation points using a Monte Carlo method. Model results were obtained with an integrated finite difference model. The magnitudes of mean model errors are generally small, but are measurable in one zone and at pumping wells. Diagonal elements of the second moment matrix correlate closely with magnitudes of mean hydraulic heads, $E(\mathbf{f}(\boldsymbol{\beta}))$. Correlations computed from the covariance matrix are often small; only 12 out of the possible 496 distinct correlations are over 0.5. Values of skewness and kurtosis again indicate that the statistical distribution of model errors is not normal. Thus, even in the absence of widespread large correlations, model errors may have a significant effect on model analyses.

Values of the same model nonlinearity measures used for example 1 indicate that the model for example 2 is not as nonlinear as the model for example 1. Model intrinsic nonlinearity is larger than for example 1, but is small. Values of the model combined intrinsic nonlinearity measures for all six parameters and a predicted head $(h_p)$ are all of the order of $10^{-3}$. Therefore, model intrinsic nonlinearity and model combined intrinsic nonlinearity should have little influence on model predictions and uncertainty analysis.

The philosophy of obtaining the hydraulic head data $\mathbf{Y}$ for example 2 is the same as used for example 1. Zero-mean random normal deviates having a standard deviation of 0.1 were added to $\mathbf{f}(\boldsymbol{\beta})$ obtained as a realization of the Monte Carlo process to yield an error structure dominated by model error. Weight matrix $\hat{\boldsymbol{\omega}}$ was obtained in the same way as for example 1. Regression produced estimates of the three zonal $\ln T$ parameters, the three zonal $W$ parameters, and three hydraulic heads defining the hydraulic heads along a specified head boundary.

Probability plots for the weighted residuals obtained as for example 1 suggest that the weighted residuals follow the theoretically expected normal distribution. However, in contrast to example 1, the weighted residuals also lie inside of the two-standard deviation limits for the incorrect distribution, which means that this plot, which would be used in field studies, would not detect the influence of model-error correlations. The plot of weighted residuals in relation to weighted function values has no visually apparent abnormalities; the slope and mean are both small in magnitude. These results suggest small model and system types of intrinsic nonlinearity.

Comparison of corrected linear and nonlinear confidence intervals for all six $\ln T$ and $W$ parameters and $h_p$ shows that the effect of model nonlinearity is small. Magnitudes of the correction factors show that all intervals need correction using the correction factors. The prediction interval contains its predicted value $Y_p$, which is considerably different than $h_{p*}$. This difference shows the large effect of heterogeneity in $T$, $W$, or both.

A Monte Carlo process consisting of repeated applications of the procedure used to calculate the nonlinear 95 percent confidence and prediction intervals was used to check the accuracy of the intervals. Both corrected and uncorrected intervals were computed. Corrected confidence intervals and the corrected prediction interval were found to be accurate, the average containment probability being nearly 0.95. In contrast, uncorrected confidence intervals were found to be too small, the largest and smallest containment probabilities being 0.892 and 0.656, respectively. The uncorrected prediction interval is accurate, however,

The Monte Carlo results also were used to check for bias in $E(\mathbf{f}(\gamma\hat{\theta}))$ and $E(g(\gamma\hat{\theta}))$, the latter applying for the six $\ln T$ and $W$ parameters and $h_p$. These means should be nearly unbiased for a model with small model and system intrinsic, and model and system combined intrinsic, types of nonlinearity. This was found to be true, with the biases in $E(\mathbf{f}(\gamma\hat{\theta}))$ being negligible and the largest percent change in $E(g(\gamma\hat{\theta}))$ from $E(g(\gamma\theta_*))$ being 7.35 percent for $g(\gamma\theta) = \theta_4 = W_1$.

Regression estimates and linearized confidence intervals also were computed based on Gauss-Markov estimation and ordinary least squares. In contrast to the results of example 1, Gauss-Markov estimates are different from estimates obtained using ordinary least squares, and linearized confidence intervals obtained using the former method are generally smaller than corresponding corrected intervals obtained using the latter. Also in contrast to example 1, estimates and confidence intervals obtained using Gauss-Markov estimation are very similar to corresponding estimates and corrected confidence intervals obtained using $\omega = \hat{\omega}$.

The entire analysis yielding regression estimates, confidence intervals, and prediction intervals was repeated for the case where the weight matrix is unknown. Errors were grouped into three groups according to similarities in hydrogeology, sensitivities, and residuals obtained using ordinary least squares. An additional weighting adjustment was made to observations at the two pumping wells to account for the large model errors at these points.

The nonlinearity measures and the results of analyzing the residuals from the regression are very similar to the corresponding measures and results obtained using $\omega = \hat{\omega}$. Regression estimates and linearized confidence intervals for parameters are similar to the corresponding estimates and intervals obtained using $\omega = \hat{\omega}$; the prediction interval is smaller than the one obtained before. Nonlinear confidence and prediction intervals can be much larger than those obtained using $\omega = \hat{\omega}$. This appears to result more from the logarithmic form of $\ell(\theta)$, which causes weights to be calculated as part of the procedure to obtain the intervals, than from model nonlinearity.

A Monte Carlo analysis of the accuracy of the confidence and prediction intervals confirms that the large intervals are very conservative. The other corrected intervals are not quite as accurate as the ones obtained using $\omega = \hat{\omega}$ because the containment probabilities are all less than 0.95, the smallest being approximately 0.92. The correction factors, which were derived assuming known weights, are less accurate when calculated using weights calculated during the regression; use of the bounds would produce much larger intervals. All intervals need to be corrected; the containment probabilities for uncorrected nonlinear confidence intervals range from 0.624 to 0.892.

# 8. Summary and Conclusions

Application of geostatistical and statistical optimization procedures to ground-water model calibration and uncertainty analysis is hampered by two pervasive problems: 1) nonlinearity of the solution of the model equations with respect to some of the hydrogeologic input variables (termed system characteristics), and 2) detailed and generally unknown spatial variability (heterogeneity) of some of the system characteristics. Because of the lack of detailed site-specific information on heterogeneity, heterogeneity is often described geostatistically. However, efficient inclusion of both nonlinearity and heterogeneity in geostatistical and statistical optimization formulations of ground-water models has remained elusive. This report describes a new theory and approach for efficient modeling of ground-water flow to include nonlinearity and a geostatistical description of heterogeneity using a small number of model parameters. The theory provides a sound framework for 1) lumping and smoothing the system characteristics to define the model parameters and 2) estimating the parameters and assessing the uncertainty of the estimates, model functions computed using the estimates, and predictions to be made with the model.

The following general conclusions can be drawn from this report. A brief summary of the results leading to each conclusion also is included.

1. The vector of system characteristics, $\boldsymbol{\beta}$, can be replaced with a lumped or smoothed approximation $\boldsymbol{\gamma\theta}_*$ (where $\boldsymbol{\gamma}$ is a spatial and temporal interpolation matrix) when constructing a ground-water model. This idea is used because $\boldsymbol{\beta}$, which contains both small and large scales of variability (or heterogeneity) in such properties as hydraulic conductivity, recharge, discharge, boundary conditions, pumping rates from wells, and other quantities that characterize the ground-water system, has too large a dimension to be estimated using the data normally available. The small-scale variability contained in $\boldsymbol{\beta}$ is accounted for by imagining $\boldsymbol{\beta}$ to be generated by a stochastic process. Vector $\boldsymbol{\gamma\theta}_*$ is a spatial and temporal average having the same form as the drift, $\boldsymbol{\gamma\bar{\theta}}$, of the stochastic process, but is a best-fit vector to $\boldsymbol{\beta}$. Vector $\boldsymbol{\theta}_*$ does not have to be a vector of effective values. A model function $f(\boldsymbol{\beta})$, such as a computed hydraulic head or flux, is assumed to accurately represent a field quantity because of the detailed nature of $\boldsymbol{\beta}$, but a model function $f(\boldsymbol{\gamma\theta}_*)$ contains error resulting from lumping or smoothing of $\boldsymbol{\beta}$ by $\boldsymbol{\gamma\theta}_*$. Thus, the replacement process yields mean model errors of the form $E(f(\boldsymbol{\beta})-f(\boldsymbol{\gamma\theta}_*))$ throughout the model and correlations between model errors at points throughout the model. This can be regarded as the penalty paid when replacing $\boldsymbol{\beta}$ with $\boldsymbol{\gamma\theta}_*$. The nonzero means and correlations can have a significant effect on construction and interpretation of a model that is calibrated by estimating $\boldsymbol{\theta}_*$.
2. Vector $\boldsymbol{\theta}_*$ can be estimated as $\hat{\boldsymbol{\theta}}$ using weighted nonlinear least squares techniques. The estimate $\mathbf{f}(\boldsymbol{\gamma\hat{\theta}})$ (where $\mathbf{f}$ is a vector of computed values of one or more model functions corresponding to data $\mathbf{Y}$ at observation points used for the least squares) is a biased estimate of $\mathbf{f}(\boldsymbol{\gamma\theta}_*)$ because of (total) nonlinearity in $\mathbf{f}(\boldsymbol{\beta})$ and intrinsic nonlinearity in both $\mathbf{f}(\boldsymbol{\beta})$ and $\mathbf{f}(\boldsymbol{\gamma\theta})$. (Intrinsic nonlinearity in $\mathbf{f}(\boldsymbol{\beta})$ or $\mathbf{f}(\boldsymbol{\gamma\theta})$ is the portion of total nonlinearity that could not in theory be eliminated by some unique transformation of $\boldsymbol{\beta}$ or $\boldsymbol{\theta}$, respectively, although to yield

only small bias, small intrinsic nonlinearity in $\mathbf{f}(\beta)$ must be combined with the requirement that approximations of second derivatives of $\mathbf{f}(\beta)$ explained in appendix C be accurate. Special terminology was applied for the various types of nonlinearity, and the reader is referred to section 4 for an explanation.) When considered to be an estimate of $\mathbf{f}(\beta)$, $\mathbf{f}(\gamma\hat{\theta})$ is biased only because of intrinsic nonlinearity in $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta)$. Because the intrinsic nonlinearity can be small and the approximations can be accurate, $\mathbf{f}(\gamma\hat{\theta})$ can be a nearly unbiased estimate of $\mathbf{f}(\beta)$, but, because total nonlinearity is generally much larger, $\mathbf{f}(\gamma\hat{\theta})$ can often be more biased as an estimate of $\mathbf{f}(\gamma\theta_*)$. Analogously, the prediction $g(\gamma\hat{\theta})$ (where $g$ is some function of parameters of interest to the investigator) is a biased estimate of $g(\gamma\theta_*)$ because of (total) nonlinearity in $g(\beta)$ and combined intrinsic nonlinearity in $g(\beta)$, $\mathbf{f}(\beta)$, $g(\gamma\theta)$, and $\mathbf{f}(\gamma\theta)$. (Combined intrinsic nonlinearity is the portion of total nonlinearity in the pairs $g(\beta)$ and $\mathbf{f}(\beta)$ or $g(\gamma\theta)$ and $\mathbf{f}(\gamma\theta)$ that could not in theory be eliminated by some unique transformation of $\beta$ or $\theta$, respectively. Again, special terminology was applied for the various types of nonlinearity, and the reader is referred to section 4 for an explanation.) When considered to be an estimate of $g(\beta)$, $g(\gamma\hat{\theta})$ is biased only because of combined intrinsic nonlinearity or inaccurate approximations of the second derivatives of $\mathbf{f}(\beta)$ and $g(\beta)$, as explained in appendix C. Combined intrinsic nonlinearity may often be larger than intrinsic nonlinearity, but should generally be much smaller than total nonlinearity, so $g(\gamma\hat{\theta})$ should generally be expected to be less biased as an estimate of $g(\beta)$ than as an estimate of $g(\gamma\theta_*)$. An investigator would probably be more interested in estimates of the real variables $\mathbf{f}(\beta)$ and $g(\beta)$ than the fictitious variables $\mathbf{f}(\gamma\theta_*)$ and $g(\gamma\theta_*)$, so the extra component of bias with respect to the functions of $\gamma\theta_*$ may not be too important. In any case, the predictive accuracy of a model is strongly tied to the degree of intrinsic nonlinearity and combined intrinsic nonlinearity of the models $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta)$, and predictions $g(\beta)$ and $g(\gamma\theta)$. As a final point of interest, the forms of the terms expressing bias from total nonlinearity contained in the biases $E(\mathbf{f}(\gamma\hat{\theta})-\mathbf{f}(\gamma\theta_*))$ and $E(g(\gamma\hat{\theta})-g(\gamma\theta_*))$ show that the terms express an interaction of heterogeneity and nonlinearity because the terms equal zero if there is no small-scale heterogeneity, a condition that is expressed as $\beta=\theta$ and $\gamma=\mathbf{I}$ (the identity matrix).

3. The correct weight matrix to use when estimating $\hat{\theta}$ and evaluating uncertainty in $\hat{\theta}$, $\mathbf{f}(\gamma\hat{\theta})$, and $g(\gamma\hat{\theta})$ is the inverse of the second moment matrix for the total error vector, $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)=\mathbf{Y}-\mathbf{f}(\beta)+\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$, where $\mathbf{Y}-\mathbf{f}(\beta)$ is an observation-error vector and $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ is the model-error vector. However, Obenchain (1975) argued that use of this matrix can produce a poor model fit to the data $\mathbf{Y}$. In cases where this proves to be true or where the data on $\beta$ are insufficient to compute this matrix, a diagonal estimate of it may be used. This diagonal weight matrix is ideally composed of inverses of the diagonal elements of the second moment matrix, but for practical computation the errors may be grouped based on similarities in factors believed to cause the errors. The theory implies that a weight matrix based only on observation errors would not be a good substitute unless model errors were small.

4. Distributions of functions of the sums of squared, weighted errors that are $F$ and $t^2$ distributed for a classical linear model form the basis for confidence regions for $\theta_*$, confidence intervals for $g(\gamma\theta_*)$, and prediction intervals for $g(\gamma\theta_*)+\upsilon_*$ (where $\upsilon_*$ is a predicted

combination model and observation error) used for model uncertainty analyses. A combination Taylor series and perturbation technique that assumes the variances of $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ and $\mathbf{f}(\beta)-\mathbf{f}(\gamma\theta_*)$ to be small, with the latter being much smaller than the former, was used to derive these distributions for a nonlinear model and an arbitrary weight matrix, the latter to allow for the choices for the weight matrix given earlier in this section. The functions of the sums of squared, weighted errors were found to have distributions that are multiples of the $F$ and $t^2$ distributions, where the multipliers are termed correction factors. The correction factors are functions of intrinsic nonlinearity of $\mathbf{f}(\gamma\theta)$, combined intrinsic nonlinearity of $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$, and the deviation of the weight matrix from the inverse of the second moment matrix for $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$. Corrections for nonlinearity in $\mathbf{f}(\beta)$ and $g(\beta)$ cannot be made readily.

5. Additional analyses (appendix F) that do not use Taylor series and perturbation expansions show that the corrected $F$ and $t^2$ distributions apply approximately even when the model and observation errors are large. However, the analyses also show that the approximations should get worse the further the distribution of $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ deviates from normal, and the correction factors do not directly include these non-normality effects. Two examples where the model error variance composes most of the total error variance, and both are large, yield results that are predicted accurately by the theory, even though $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ in both cases is definitely not normally distributed. In particular, approximate nonlinear confidence and prediction intervals have close to correct containment probability if the critical $t$ value is multiplied by the appropriate correction factor, but intervals can be much too small if the factor is omitted. More study is needed to determine the potential for inaccuracies from assuming $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ to be normally distributed when it is not.

6. Magnitudes of the correction factors are problem dependent. Normally the contribution from intrinsic nonlinearity of $\mathbf{f}(\gamma\theta)$ can be neglected; the contribution from combined intrinsic nonlinearity of $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$ was observed to be small for the two example problems, but this may not always be the case. The importance of both of these contributions can be tested. Contribution from model error can be minimized by using a structure $\gamma\theta$ that accounts for as much variability in the field set $\beta$ as possible, consistent with the necessity of designing a nonsingular problem. However, results from example 2 show that even localized large correlations resulting from model error may have a significant effect on the magnitudes of the correction factors and thus on the confidence and prediction intervals. The localized correlations would probably not be detected by an analysis of residuals. Therefore, prediction intervals should be tested for accurate containment probability by using techniques such as the cross-validation techniques proposed by Christensen and Cooley (1999b). Any new data also should be similarly tested to determine whether or not they are contained in their prediction intervals with nearly correct probability.

7. When the geostatistical data are insufficient to permit estimation of the correction factors or a diagonal weight matrix, then a method that assumes the weights to be unknown can be used to estimate $\theta_*$ and compute confidence and prediction intervals, with approximate bounds replacing the correction factors. To use this method, only the error grouping discussed in section 4 needs to be known. The method will probably produce conservative confidence and prediction

intervals if most of the variability in the diagonal second moments can be accounted for by grouping the errors. Prediction intervals often will be more accurate than confidence intervals if the variance of $\upsilon_*$ is accurately known.

8. The theory developed for modeling ground-water flow in heterogeneous media using regression methodology explains the observed field results listed in the introduction:

a) Estimates $\mathbf{f}(\gamma\hat{\theta})$ and $g(\gamma\hat{\theta})$ often are physically realistic or close to what should be expected, even when effective values of the form of $\theta$ to replace $\beta$ are known not to exist. Biases are tied to the magnitudes of intrinsic nonlinearity of $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta)$, and combined model intrinsic nonlinearity of $\mathbf{f}(\beta)$ and $g(\beta)$, and $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$, all which can be small. If the major hydrogeologic features are accurately contained in a model, the biases can be reduced further. The importance of the intrinsic and combined intrinsic types of nonlinearity for $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$ can be evaluated for any particular model by using measures derived in this report and in previous studies (Linssen, 1975; Johansen, 1983). Evaluation of the types of intrinsic nonlinearity for $\mathbf{f}(\beta)$ and $g(\beta)$ is more difficult, but can be carried out by analyzing residuals. The analyses carried out for the two examples in this report indicate that biases from intrinsic nonlinearity are insignificant for both them. However, development of better diagnostic methodology is needed, and this is beyond the scope of this report.

b) Residuals $\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})$ often behave as if the model were linear and as if errors $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ had a zero-mean normal distribution. If intrinsic nonlinearity in $\mathbf{f}(\beta)$ and $\mathbf{f}(\gamma\theta)$ is small and $\beta$ has a normal distribution, then the residuals have a zero-mean normal distribution as if the model were linear and as if $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ had a zero-mean normal distribution. Model nonlinearity and model error do not manifest themselves as an abnormality in the residuals in this case. Both examples have residuals that appear to have zero-mean, normal distributions. In a field situation, log hydraulic conductivity is known to often have nearly a normal distribution, which yields the required near-normal distribution for at least some elements of $\beta$.

c) Some confidence intervals appear to exclude reasonable values; whereas, others do not. This characteristic is explained by correction factors that can vary greatly from one confidence interval to another; uncorrected confidence intervals can range from accurate to highly inaccurate. For example, in example 2 the ordinary least squares confidence interval for the recharge in zone 2 is only about 15 percent too small without correction; whereas, the ordinary least squares confidence interval for log transmissivity for zone 2 is about 425 percent too small without correction.

9. Finally, although not emphasized earlier, it is worth noting that model function $\mathbf{f}(\beta)$ can be interpreted very broadly. For example, $\mathbf{f}(\beta)$ can be a model function for various types of models, including stochastic models such as proposed by Neuman and Orr (1993) and Tartakovsky and Neuman (1998). Function $\mathbf{f}(\beta)$ can even be interpreted as data directly on $\beta$, such as $f_i(\beta)=\beta_j$ (the *i*th error-free observation, which is the *j*th element of $\beta$). If this is the only type of data in $\mathbf{f}(\beta)$, then the entire theory would cover the use of direct observations of $\beta$ to construct a model. More commonly, the data could be mixtures of more than one type, as, for example, data on a model function such as hydraulic head and direct observations of $\beta$, and $\mathbf{f}(\beta)$ would be defined accordingly.

# Appendix A – Evaluation of $E(\mathbf{x}'\mathbf{A}_i\mathbf{x})(\mathbf{x}'\mathbf{A}_j\mathbf{x})$

Let $\mathbf{x} = [x_i]$ be a zero-mean normal random variable so that

$$\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}\sigma^2) \tag{A-1}$$

where $\mathbf{I}$ is the identity matrix and $\sigma^2$ is an arbitrary variance. Then, for arbitrary symmetric matrices $\mathbf{A}_i = [a_{ikl}]$ and $\mathbf{A}_j = [a_{jkl}]$,

$$\begin{aligned}
&E(\mathbf{x}'\mathbf{A}_i\mathbf{x})(\mathbf{x}'\mathbf{A}_j\mathbf{x}) \\
&= E(\sum_k\sum_l a_{ikl}x_k x_l)(\sum_q\sum_r a_{jqr}x_q x_r) \\
&= \sum_k\sum_l\sum_q\sum_r a_{ikl}a_{jqr}E(x_k x_l x_q x_r) \\
&= \sum_k\sum_l\sum_q\sum_r a_{ikl}a_{jqr}(\delta_{kr}\delta_{lq} + \delta_{kq}\delta_{lr} + \delta_{kl}\delta_{qr})\sigma^4 \\
&= (\sum_k\sum_l a_{ikl}a_{jlk} + \sum_k\sum_l a_{ikl}a_{jkl} + \sum_k\sum_l a_{ikk}a_{jll})\sigma^4 \\
&= (\sum_k\sum_l a_{ikk}a_{jll} + 2\sum_k\sum_l a_{ikl}a_{jlk})\sigma^4 \\
&= tr(\mathbf{A}_i)tr(\mathbf{A}_j)\sigma^4 + 2tr(\mathbf{A}_i\mathbf{A}_j)\sigma^4
\end{aligned} \tag{A-2}$$

where $\delta_{ij}$ is the Kronecker delta,

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{A-3}$$

and the standard result $E(x_k x_l x_q x_r) = (\delta_{kr}\delta_{lq} + \delta_{kq}\delta_{lr} + \delta_{kl}\delta_{qr})\sigma^4$ is obtained by using the characteristic function for a multivariate normal distribution (Anderson, 1958, p. 39).

# Appendix B – Second-Order Correct Parameter and Parameter Function Estimates; Third-Order Correct Sum of Squares Estimate

## Second-Order Correct Parameter Estimates

To develop $\mathbf{l}$ and $\mathbf{q}$ in $\hat{\theta} - \bar{\theta} = \mathbf{l} + \mathbf{q}$, methods used by Johansen (1983) are extended to include model error from heterogeneity. First, expansion of $\mathbf{f}(\gamma\theta)$ through second order around $\mathbf{f}(\gamma\bar{\theta})$ using a truncated Taylor series yields

$$f_i(\gamma\theta) = f_i(\gamma\bar{\theta}) + \mathbf{D}f_i(\theta - \bar{\theta}) + \frac{1}{2}(\theta - \bar{\theta})'\mathbf{D}^2 f_i\,(\theta - \bar{\theta});\ i = 1,2,\ldots,n \tag{B-1}$$

where $\mathbf{D}f_i$ and $\mathbf{D}^2 f_i$ are row-vector and matrix components of $\mathbf{Df}$ and $\mathbf{D}^2\mathbf{f}$ as defined by (2-1) and (2-2). They are evaluated at $\theta = \bar{\theta}$. Second, use of (B-1) and (4-4) gives, up through second order in $\mathbf{U}$ and $\mathbf{e}$,

$$\begin{aligned} f_i(\gamma\hat{\theta}) &= f_i(\gamma\bar{\theta}) + \mathbf{D}f_i(\hat{\theta} - \bar{\theta}) + \frac{1}{2}(\hat{\theta} - \bar{\theta})'\mathbf{D}^2 f_i\,(\hat{\theta} - \bar{\theta}) \\ &= f_i(\gamma\bar{\theta}) + \mathbf{D}f_i(\mathbf{l} + \mathbf{q}) + \frac{1}{2}(\mathbf{l} + \mathbf{q})'\mathbf{D}^2 f_i\,(\mathbf{l} + \mathbf{q}) \\ &\approx f_i(\gamma\bar{\theta}) + \mathbf{D}f_i\mathbf{l} + \mathbf{D}f_i\mathbf{q} + \frac{1}{2}\mathbf{l}'\mathbf{D}^2 f_i\mathbf{l} \end{aligned} \tag{B-2}$$

$$\begin{aligned} \mathbf{D}\hat{f}_i &= \mathbf{D}f_i + (\hat{\theta} - \bar{\theta})'\mathbf{D}^2 f_i \\ &\approx \mathbf{D}f_i + \mathbf{l}'\mathbf{D}^2 f_i \end{aligned} \tag{B-3}$$

where $\mathbf{D}\hat{f}_i$ indicates evaluation at $\theta = \hat{\theta}$. Third, $S(\theta)$ is minimized using (3-35), (4-2), (B-2), and (B-3), keeping terms up through second order in $\mathbf{U}$, $\mathbf{e}$, or their product. The result is

$$\begin{aligned} &\sum_i\sum_j \mathbf{D}\hat{f}_i'\omega_{ij}(Y_j - f_j(\gamma\hat{\theta})) \\ &\approx \sum_i\sum_j(\mathbf{D}f_i' + \mathbf{D}^2 f_i\mathbf{l})\omega_{ij}(Y_j - f_j(\gamma\bar{\theta}) - \mathbf{D}f_j\mathbf{l} - \mathbf{D}f_j\mathbf{q} - \frac{1}{2}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}) \\ &\approx \sum_i\sum_j \mathbf{D}f_i'\omega_{ij}(U_j + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{D}f_j\mathbf{l} - \mathbf{D}f_j\mathbf{q} - \frac{1}{2}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}) + \sum_i\sum_j \mathbf{D}^2 f_i\mathbf{l}\omega_{ij}(U_j - \mathbf{D}f_j\mathbf{l}) \\ &= \mathbf{0} \end{aligned} \tag{B-4}$$

Because $\mathbf{l}$ is the first-order solution, it must satisfy

$$\sum_i\sum_j \mathbf{D}f_i'\omega_{ij}(U_j - \mathbf{D}f_j\mathbf{l}) = \mathbf{0} \tag{B-5}$$

or, in matrix form,

$$\mathbf{l} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}\mathbf{U} \tag{B-6}$$

Putting (B-5) into (B-4) gives

$$\sum_i\sum_j \mathbf{D}f_i'\omega_{ij}(\frac{1}{2}\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} - \mathbf{D}f_j\,\mathbf{q} - \frac{1}{2}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}) + \sum_i\sum_j \mathbf{D}^2 f_i\,\mathbf{l}\,\omega_{ij}(U_j - \mathbf{D}f_j\mathbf{l})$$
$$= \mathbf{0} \tag{B-7}$$

Now, use of (B-6) in the last term in (B-7) yields

$$\sum_i\sum_j \mathbf{D}^2 f_i\,\mathbf{l}\,\omega_{ij}(U_j - \mathbf{D}f_j\mathbf{l})$$
$$= \sum_i \mathbf{D}^2 f_i\,\mathbf{l}\boldsymbol{\omega}_i(\mathbf{U} - \mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}\mathbf{U})$$
$$= \sum_i \mathbf{D}^2 f_i\,\mathbf{l}\boldsymbol{\omega}_i\boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I} - \boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}$$
$$= \sum_i \mathbf{D}^2 f_i\,\mathbf{l}\boldsymbol{\omega}_i\boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}$$
$$= \sum_i \mathbf{D}^2 f_i\,\mathbf{l}\boldsymbol{\omega}_i\boldsymbol{\omega}^{-1}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Z}$$
$$= \sum_i \mathbf{D}^2 f_i\,\mathbf{l}\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z} \tag{B-8}$$

where $\boldsymbol{\omega}_i$ is row $i$ of $\boldsymbol{\omega}$, $\boldsymbol{\omega}_i^{1/2}$ is row $i$ of $\boldsymbol{\omega}^{1/2}$, $\mathbf{Z} = (\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{U}$, and $\mathbf{R} = \boldsymbol{\omega}^{1/2}\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{1/2}$. Thus, (B-7) becomes

$$\frac{1}{2}\sum_j \mathbf{Df}'\boldsymbol{\omega}_j(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}) - \mathbf{Df}'\boldsymbol{\omega}\mathbf{Df}\mathbf{q} + \sum_i \mathbf{D}^2 f_i\mathbf{l}\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z} = \mathbf{0} \tag{B-9}$$

or

$$\mathbf{q} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}(\sum_i \mathbf{D}^2 f_i\mathbf{l}\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z} + \frac{1}{2}\mathbf{Df}'\sum_j \boldsymbol{\omega}_j(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})) \tag{B-10}$$

where $\boldsymbol{\omega}_j$ is column $j$ of $\boldsymbol{\omega}$.

## Second-Order Correct Parameter Function Estimates

An estimated model function value $f_i(\gamma\hat{\boldsymbol{\theta}})$, a residual $Y_i - f_i(\gamma\hat{\boldsymbol{\theta}})$, and a prediction $g(\gamma\hat{\boldsymbol{\theta}})$ are developed using $\mathbf{l}$ and $\mathbf{q}$. An estimated model function value is expressed to second order using (B-2), (B-6), and (B-10) as

$$f_i(\gamma\hat{\theta}) \approx f_i(\gamma\bar{\theta}) + \mathbf{D}f_i\mathbf{l} + \mathbf{D}f_i\mathbf{q} + \frac{1}{2}\mathbf{l}'\mathbf{D}^2 f_i\,\mathbf{l}$$

$$= f_i(\gamma\bar{\theta}) + \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{U} + \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_i \mathbf{D}^2 f_k\mathbf{l}\omega^{\frac{1}{2}}\mathbf{Z}$$

$$+\frac{1}{2}\mathbf{Df}'\sum_j \omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})) + \frac{1}{2}\mathbf{l}'\mathbf{D}^2 f_i\mathbf{l}$$

$$= f_i(\gamma\bar{\theta}) + \omega_i^{-\frac{1}{2}}\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\omega^{\frac{1}{2}}\omega^{-1}\sum_j \omega_j\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e})$$

$$+\frac{1}{2}\omega_i^{-\frac{1}{2}}(\mathbf{I} - \omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}})\omega^{\frac{1}{2}}\omega^{-1}\sum_j \omega_j\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l} + \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}$$

$$= f_i(\gamma\bar{\theta}) + \omega_i^{-\frac{1}{2}}\mathbf{R}(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}) + \frac{1}{2}\omega_i^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}$$

$$+ \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z} \qquad \text{(B-11)}$$

where $\omega_k^{1/2}$ is row $k$ of $\omega^{1/2}$.

A residual is defined as $Y_i - f_i(\gamma\hat{\theta})$. Then a second-order approximation is computed using (3-5), (3-27), and (B-11) as follows.

$$Y_i - f_i(\gamma\hat{\theta}) = Y_i - f_i(\beta) + f_i(\beta) - f_i(\gamma\bar{\theta}) - (f_i(\gamma\hat{\theta}) - f_i(\gamma\bar{\theta}))$$

$$\approx \varepsilon_i + \mathbf{D}_\beta f_i\,\mathbf{e} + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 f_i\mathbf{e} - \omega_i^{-\frac{1}{2}}\mathbf{R}(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}) - \frac{1}{2}\omega_i^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}$$

$$- \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}$$

$$= \omega_i^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R})(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})) - \mathbf{D}f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z} \qquad \text{(B-12)}$$

A prediction $g(\gamma\hat{\theta})$ of $g(\beta)$ or $g(\gamma\theta_*)$ can be computed using an equation analogous to (B-2) together with (B-6) and (B-10):

$$g(\gamma\hat{\theta}) = g(\gamma\bar{\theta}) + \mathbf{D}g(\hat{\theta} - \bar{\theta}) + \frac{1}{2}(\hat{\theta} - \bar{\theta})'\mathbf{D}^2 g(\hat{\theta} - \bar{\theta})$$

$$\approx g(\gamma\bar{\theta}) + \mathbf{D}g\mathbf{l} + \mathbf{D}g\mathbf{q} + \frac{1}{2}\mathbf{l}'\mathbf{D}^2 g\,\mathbf{l}$$

$$= g(\gamma\bar{\theta}) + \mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{U} + \mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum \mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z}$$

$$+\frac{1}{2}\mathbf{Df}'\sum_j \omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})) + \frac{1}{2}\mathbf{l}'\mathbf{D}^2 g\,\mathbf{l}$$

$$= g(\gamma\bar{\theta}) + \mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}(\mathbf{l}'\mathbf{D}^2 g\mathbf{l} - \mathbf{Q}'\omega^{\frac{1}{2}}\omega^{-1}\sum_j \omega_j\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})$$

$$+\frac{1}{2}\mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\omega}^{-1}\sum_j \boldsymbol{\omega}_j\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}+\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z}$$

$$=g(\gamma\overline{\boldsymbol{\theta}})+\mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}+\frac{1}{2}(\mathbf{l}'\mathbf{D}^2g\mathbf{l}-\mathbf{Q}'\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})+\frac{1}{2}\mathbf{Q}'\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}$$

$$+\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z} \qquad \text{(B-13)}$$

where $\mathbf{Q}=\boldsymbol{\omega}^{1/2}\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'$.

## Third-Order Correct Sum of Squares Estimate

To obtain all of the terms through fourth order in **U**, **e**, and their product resulting from the second-order expansion for $\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$, the term $\mathbf{l}'\mathbf{D}^2 f_i\mathbf{q}$ must be kept in (B-2). This does not result in true fourth-order accuracy in **U**, **e**, and their product for the sum of squares because, in order to obtain this, a third-order term of the Taylor series would have to be retained. This term is not retained because evaluation of third-order derivatives of **f** is not practical.

With the added term, (B-12) becomes

$$Y_i-f_i(\gamma\hat{\boldsymbol{\theta}})\approx\boldsymbol{\omega}_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}+\frac{1}{2}\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))-\mathbf{D}f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\boldsymbol{\omega}_k^{\frac{1}{2}}\mathbf{Z}$$

$$-\mathbf{l}'\mathbf{D}^2 f_i\mathbf{q} \qquad \text{(B-14)}$$

so that

$$S(\hat{\boldsymbol{\theta}})=\sum_i\sum_\ell(Y_i-f_i(\gamma\hat{\boldsymbol{\theta}}))\omega_{i\ell}(Y_\ell-f_\ell(\gamma\hat{\boldsymbol{\theta}}))$$

$$\approx\sum_i\sum_\ell(\boldsymbol{\omega}_i^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}+\frac{1}{2}\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))-\mathbf{D}f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\boldsymbol{\omega}_k^{\frac{1}{2}}\mathbf{Z}$$

$$-\mathbf{l}'\mathbf{D}^2 f_i\mathbf{q})\omega_{i\ell}(\boldsymbol{\omega}_\ell^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}+\frac{1}{2}\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))$$

$$-\mathbf{D}f_\ell(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\boldsymbol{\omega}_k^{\frac{1}{2}}\mathbf{Z}-\mathbf{l}'\mathbf{D}^2 f_\ell\mathbf{q})$$

$$\approx(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}+\frac{1}{2}\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))'(\mathbf{I}-\mathbf{R})\sum_i\sum_\ell(\boldsymbol{\omega}_i^{-\frac{1}{2}})'\omega_{i\ell}\boldsymbol{\omega}_\ell^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}$$

$$+\frac{1}{2}\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))-2(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}+\frac{1}{2}\sum_j \boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))'(\mathbf{I}-\mathbf{R})$$

$$\bullet\sum_i\sum_\ell(\boldsymbol{\omega}_i^{-\frac{1}{2}})'\omega_{i\ell}\mathbf{D}f_\ell(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\boldsymbol{\omega}_k^{\frac{1}{2}}\mathbf{Z}-2\mathbf{U}'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_i\sum_\ell(\boldsymbol{\omega}_i^{-\frac{1}{2}})'\omega_{i\ell}\mathbf{l}'\mathbf{D}^2 f_\ell\mathbf{q}$$

$$+\sum_k \boldsymbol{\omega}_k^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_k(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i\sum_\ell \mathbf{D}f_i'\omega_{i\ell}\mathbf{D}f_\ell(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k\mathbf{l}\boldsymbol{\omega}_k^{\frac{1}{2}}\mathbf{Z}$$

$$
\begin{aligned}
&= (\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))'(\mathbf{I}-\mathbf{R})(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))\\
&- 2(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))'(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}\\
&- 2\mathbf{U}'\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_i(\omega_i^{\frac{1}{2}})'\mathbf{l}'\mathbf{D}^2 f_i \mathbf{q} + \sum_k \omega_k^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_k(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}\\
&= (\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))'(\mathbf{I}-\mathbf{R})(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))\\
&- 2\mathbf{U}'\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_i(\omega_i^{\frac{1}{2}})'\mathbf{l}'\mathbf{D}^2 f_i \mathbf{q} + \sum_k \omega_k^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_k(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z} \qquad \text{(B-15)}
\end{aligned}
$$

where $(\mathbf{I}-\mathbf{R})\omega^{1/2}\mathbf{Df} = \mathbf{0}$. Now, use of (B-10) gives

$$
\begin{aligned}
&\mathbf{U}'\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_i(\omega_i^{\frac{1}{2}})'\mathbf{l}'\mathbf{D}^2 f_i \mathbf{q} = \mathbf{Z}'\sum_i(\omega_i^{\frac{1}{2}})'\mathbf{l}'\mathbf{D}^2 f_i \mathbf{q}\\
&= \mathbf{Z}'\sum_i(\omega_i^{\frac{1}{2}})'\mathbf{l}'\mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_j \mathbf{D}^2 f_j \mathbf{l}\omega_j^{\frac{1}{2}}\mathbf{Z} + \frac{1}{2}\mathbf{Df}'\sum_j \omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))\\
&= \sum_k \omega_k^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_k(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}\\
&+ \frac{1}{2}\sum_k \omega_k^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_k(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\sum_j \omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}) \qquad \text{(B-16)}
\end{aligned}
$$

Finally, substitution of (B-16) into (B-15) results in

$$
\begin{aligned}
S(\hat{\theta}) &\approx (\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))'(\mathbf{I}-\mathbf{R})(\omega^{\frac{1}{2}}\mathbf{U} + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}))\\
&- \sum_k \omega_k^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_k(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_k \mathbf{D}^2 f_k \mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}\\
&- \sum_k \omega_k^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_k(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\sum_j \omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j \mathbf{l}) \qquad \text{(B-17)}
\end{aligned}
$$

## Proof that Any Squared Linear Combination $E(l'(\hat{\theta}-\theta_*))^2$ is Minimized Through Third-Order Terms When $\omega^{-1} \propto E(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'$

This proof is an adaptation of the Gauss-Markov theorem as given by Beck and Arnold (1977, p. 232-234). Use of (3-30) and (4-11) allows the solution for $\hat{\theta}-\theta_*$ to be expressed in the form

$$
\hat{\theta}-\theta_* = (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)) + \Delta \qquad \text{(B-18)}
$$

where

$$\Delta = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}(\sum_i \mathbf{D}^2 f_i \mathbf{l}\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z} + \frac{1}{2}\mathbf{Df}'\sum_j \boldsymbol{\omega}_j(\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_j\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})) \tag{B-19}$$

Let $\Omega\sigma_\varepsilon^2 = E(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*))(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*))'$. Then, if $\boldsymbol{\omega} \propto \boldsymbol{\Omega}^{-1}$, the solution $\hat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_*$ is given by

$$\hat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_* = (\mathbf{Df}'\boldsymbol{\Omega}^{-1}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\Omega}^{-1}(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*)) + \Delta_v \tag{B-20}$$

where $\Delta_v$ is $\Delta$ for which $\boldsymbol{\omega} \propto \boldsymbol{\Omega}^{-1}$. Combination of (B-18) and (B-20) gives

$$\begin{aligned}
\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_* &= \hat{\boldsymbol{\theta}}_v - \boldsymbol{\theta}_* + \mathbf{C}'(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*)) + \Delta - \Delta_v \\
&= (\mathbf{Df}'\boldsymbol{\Omega}^{-1}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\Omega}^{-1}(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*)) + \mathbf{C}'(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*)) + \Delta
\end{aligned} \tag{B-21}$$

where $\mathbf{C}' = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega} - (\mathbf{Df}'\boldsymbol{\Omega}^{-1}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\Omega}^{-1}$. Now (B-21), the definition of $\boldsymbol{\Omega}$, the fact that $\mathbf{C}'\mathbf{Df} = \mathbf{0}$, and the fact that expected values of third order products involving $\mathbf{U}_*$ and $\mathbf{e}$ are zero are used to obtain, through third order in $\mathbf{U}_*$ and $\mathbf{e}$, the squared linear combination

$$\begin{aligned}
E(\boldsymbol{l}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*))^2 &= \boldsymbol{l}'E((\mathbf{Df}'\boldsymbol{\Omega}^{-1}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\Omega}^{-1}(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*)) + \mathbf{C}'(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*)) + \Delta) \\
&\bullet ((\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*))'\boldsymbol{\Omega}^{-1}\mathbf{Df}(\mathbf{Df}'\boldsymbol{\Omega}^{-1}\mathbf{Df})^{-1} + (\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*))\mathbf{C}' + \Delta)\boldsymbol{l} \\
&\approx \boldsymbol{l}'((\mathbf{Df}'\boldsymbol{\Omega}^{-1}\mathbf{Df})^{-1}\sigma_\varepsilon^2 + 2\mathbf{C}'E(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*))(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\boldsymbol{\theta}_*))'\boldsymbol{\Omega}^{-1}\mathbf{Df}(\mathbf{Df}'\boldsymbol{\Omega}^{-1}\mathbf{Df})^{-1} \\
&+ \mathbf{C}'\boldsymbol{\Omega}\mathbf{C}\sigma_\varepsilon^2)\boldsymbol{l} \\
&= \boldsymbol{l}'(\mathbf{Df}'\boldsymbol{\Omega}^{-1}\mathbf{Df})^{-1}\boldsymbol{l}\sigma_\varepsilon^2 + \boldsymbol{l}'\mathbf{C}'\boldsymbol{\Omega}^{-1}\mathbf{C}\boldsymbol{l}\sigma_\varepsilon^2
\end{aligned} \tag{B-22}$$

Thus, because $\mathbf{C}' = \mathbf{0}$ when $\boldsymbol{\omega} \propto \boldsymbol{\Omega}^{-1}$, $E(\boldsymbol{l}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*))^2$ is minimized through third-order terms by selecting $\boldsymbol{\omega}$ to be proportional to $\boldsymbol{\Omega}^{-1}$.

# Appendix C – Invariance of Terms Expressing Intrinsic Nonlinearity and Combined Intrinsic Nonlinearity

## Terms Expressing Intrinsic Nonlinearity with Respect to $\mathbf{f}(\gamma\theta)$

An expression of the form $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{x}'\mathbf{D}^2 f_j\mathbf{y}$ is shown to be invariant under a unique transformation of $\theta$ to $\phi(\theta)$. Under these circumstances the expression takes on a value conforming to the smallest magnitude that the transformation of the matrix $\mathbf{D}^2 f_j$ could have, and the magnitude of the expression can be small if the transformation nearly linearizes **f**. Note that the actual transformation that makes **f** most nearly linear does not have to be obtained to show the required invariance.

The Jacobian for the transformation evaluated at $\theta=\bar{\theta}$ is defined as the nonsingular $p\times p$ matrix

$$\mathbf{J}=\left[\frac{\partial\theta_i}{\partial\phi_j}\right];\ i=1,2,\ldots,p;\ j=1,2,\ldots,p \tag{C-1}$$

Then **Df** transforms to $\mathbf{D}_\phi\mathbf{f}$ and $\mathbf{D}^2\mathbf{f}$ transforms to $\mathbf{D}_\phi^2\mathbf{f}$, where

$$\mathbf{DfJ}=\mathbf{D}_\phi\mathbf{f}=\left[\frac{\partial f_i}{\partial\phi_j}\right];\ i=1,2,\ldots,n;\ j=1,2,\ldots,p \tag{C-2}$$

and

$$\mathbf{D}_\phi^2\mathbf{f}=\left[\frac{\partial^2 f_i}{\partial\phi_j\partial\phi_k}\right];\ i=1,2,\ldots,n;\ j=1,2,\ldots,p;\ k=1,2,\ldots,p \tag{C-3}$$

Both matrices are evaluated at $\bar{\phi}=\phi(\bar{\theta})$.

With the above relations, the subject expression transforms from $\theta$ to $\phi$ as follows.

$$(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{x}'\mathbf{D}^2 f_j\mathbf{y}$$

$$=(\mathbf{I}-\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}})\sum_j\omega_j^{\frac{1}{2}}\sum_{i=1}^{p}\sum_{k=1}^{p}\frac{\partial^2 f_j}{\partial\theta_i\partial\theta_k}x_i y_k$$

$$=(\mathbf{I}-\omega^{\frac{1}{2}}\mathbf{DfJ}(\mathbf{J}'\mathbf{Df}'\omega\mathbf{DfJ})^{-1}\mathbf{J}'\mathbf{Df}'\omega^{\frac{1}{2}})\sum_j\omega_j^{\frac{1}{2}}\sum_{i=1}^{p}\sum_{k=1}^{p}\left(\frac{\partial\phi'}{\partial\theta_i}\mathbf{D}_\phi^2 f_j\frac{\partial\phi}{\partial\theta_k}+\mathbf{D}_\phi f_j\frac{\partial^2\phi}{\partial\theta_i\partial\theta_k}\right)x_i y_k$$

$$=(\mathbf{I}-\omega^{\frac{1}{2}}\mathbf{D}_\phi\mathbf{f}(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}})\sum_j\omega_j^{\frac{1}{2}}\sum_{i=1}^{p}\sum_{k=1}^{p}\left(\frac{\partial\phi'}{\partial\theta_i}\mathbf{D}_\phi^2 f_j\frac{\partial\phi}{\partial\theta_k}+\mathbf{D}_\phi f_j\frac{\partial^2\phi}{\partial\theta_i\partial\theta_k}\right)x_i y_k$$

$$= (\mathbf{I} - \boldsymbol{\omega}^{\frac{1}{2}}\mathbf{D}_{\phi}\mathbf{f}(\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}\mathbf{D}_{\phi}\mathbf{f})^{-1}\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}^{\frac{1}{2}})\sum_{j}\omega_{j}^{\frac{1}{2}}\sum_{i=1}^{p}\sum_{k=1}^{p}\frac{\partial\boldsymbol{\phi}'}{\partial\theta_{i}}\mathbf{D}_{\phi}^{2}f_{j}\frac{\partial\boldsymbol{\phi}}{\partial\theta_{k}}x_{i}y_{k}$$

$$= (\mathbf{I} - \mathbf{R}_{\phi})\sum_{j}\omega_{j}^{\frac{1}{2}}(\mathbf{J}^{-1}\mathbf{x})'\mathbf{D}_{\phi}^{2}f_{j}(\mathbf{J}^{-1}\mathbf{y}) \tag{C-4}$$

where $(\mathbf{I} - \boldsymbol{\omega}^{1/2}\mathbf{D}_{\phi}\mathbf{f}(\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}\mathbf{D}_{\phi}\mathbf{f})^{-1}\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}^{1/2})\sum_{j}\omega_{j}^{1/2}\mathbf{D}_{\phi}f_{j} = \mathbf{0}$, and

$$\mathbf{R}_{\phi} = \boldsymbol{\omega}^{\frac{1}{2}}\mathbf{D}_{\phi}\mathbf{f}(\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}\mathbf{D}_{\phi}\mathbf{f})^{-1}\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}^{\frac{1}{2}} \tag{C-5}$$

Equation (C-4) shows that $(\mathbf{I} - \mathbf{R}_{\phi})\sum_{j}\omega_{j}^{1/2}(\mathbf{J}^{-1}\mathbf{x})'\mathbf{D}_{\phi}^{2}f_{j}(\mathbf{J}^{-1}\mathbf{y})$ has the same form as, and is equal to, $(\mathbf{I} - \mathbf{R})\sum_{j}\omega_{j}^{1/2}\mathbf{x}'\mathbf{D}^{2}f_{j}\mathbf{y}$ no matter what transformation of the form $\boldsymbol{\phi}(\boldsymbol{\theta})$ (including $\boldsymbol{\theta}$ itself) is used. Thus, it is invariant under transformation of $\boldsymbol{\theta}$ to $\boldsymbol{\phi}(\boldsymbol{\theta})$. Note that **R** also is invariant.

Forms for **x** (and **y**) encountered in this report are $\mathbf{x} = \mathbf{l}$, $\mathbf{x} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}f_i'$, $\mathbf{x} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'$, and $\mathbf{x} = \boldsymbol{\theta}_{*} - \bar{\boldsymbol{\theta}} = (\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}$. Then forms for $\mathbf{J}^{-1}\mathbf{x}$ are

$$\begin{aligned}\mathbf{J}^{-1}\mathbf{l} &= \mathbf{J}^{-1}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}\mathbf{U}\\ &= (\mathbf{J}'\mathbf{Df}'\boldsymbol{\omega}\mathbf{DfJ})^{-1}\mathbf{JDf}'\boldsymbol{\omega}\mathbf{U}\\ &= (\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}\mathbf{D}_{\phi}\mathbf{f})^{-1}\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}\mathbf{U}\end{aligned} \tag{C-6}$$

$$\begin{aligned}\mathbf{J}^{-1}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}f_i' &= (\mathbf{J}'\mathbf{Df}'\boldsymbol{\omega}\mathbf{DfJ})^{-1}\mathbf{J}'\mathbf{D}f_i'\\ &= (\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}\mathbf{D}_{\phi}\mathbf{f})^{-1}\mathbf{D}_{\phi}f_i'\end{aligned} \tag{C-7}$$

$$\begin{aligned}\mathbf{J}^{-1}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g' &= (\mathbf{J}'\mathbf{Df}'\boldsymbol{\omega}\mathbf{DfJ})^{-1}\mathbf{JD}g'\\ &= (\mathbf{D}_{\phi}\mathbf{f}'\boldsymbol{\omega}\mathbf{D}_{\phi}\mathbf{f})^{-1}\mathbf{D}_{\phi}g'\end{aligned} \tag{C-8}$$

$$\mathbf{J}^{-1}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e} = ((\boldsymbol{\gamma}\mathbf{J})'\boldsymbol{\gamma}\mathbf{J})^{-1}(\boldsymbol{\gamma}\mathbf{J})'\mathbf{e} \tag{C-9}$$

where $\mathbf{D}g\mathbf{J} = \mathbf{D}_{\phi}g$ and

$$\boldsymbol{\gamma}\mathbf{J} = \left[\sum_{k=1}^{p}\frac{\partial\beta_i}{\partial\theta_k}\,\frac{\partial\theta_k}{\partial\phi_j}\right] = \left[\frac{\partial\beta_i}{\partial\phi_j}\right];\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,p \tag{C-10}$$

which is evaluated at $\bar{\boldsymbol{\phi}}$. Equation (C-10) is the transformation of $\boldsymbol{\gamma}$, where $\boldsymbol{\beta}$ is written as a function of $\boldsymbol{\gamma\theta}$, and $\boldsymbol{\theta}$ is written as $\boldsymbol{\theta}(\boldsymbol{\phi})$ (the inverse transformation of $\boldsymbol{\phi}(\boldsymbol{\theta})$). Because **e** and **U** are random variables that are not functions of $\boldsymbol{\theta}$, they do not transform.

## Terms Expressing Combined Intrinsic Nonlinearity with Respect to $\mathbf{f}(\gamma\theta)$ and $g(\gamma\theta)$

Next, an expression of the form $\mathbf{Q}'\sum_j \omega_j^{1/2}\mathbf{x}'\mathbf{D}^2 f_j\mathbf{y} - \mathbf{x}'\mathbf{D}^2 g\mathbf{y}$ is shown to be invariant under transformation of $\theta$ to $\phi(\theta)$. Therefore, if the same transformation transforms matrices $\mathbf{D}^2\mathbf{f}$ and $\mathbf{D}^2 g$ to matrices that are small in magnitude, the expression will be small in magnitude. That is, if the same transformation nearly linearizes both **f** and $g$, the expression will be small in magnitude.

Transformation of the subject expression shows invariance in the same way as does (C-4) as follows.

$$\mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}\mathbf{x}'\mathbf{D}^2 f_j\mathbf{y} - \mathbf{x}'\mathbf{D}^2 g\mathbf{y}$$

$$= \mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^{p}\sum_{k=1}^{p}\frac{\partial^2 f_j}{\partial\theta_i\partial\theta_k}x_i y_k - \sum_{i=1}^{p}\sum_{k=1}^{p}\frac{\partial^2 g}{\partial\theta_i\partial\theta_k}x_i y_k$$

$$= \mathbf{D}g\mathbf{J}(\mathbf{J}'\mathbf{Df}'\omega\mathbf{DfJ})^{-1}\mathbf{J}'\mathbf{Df}'\omega^{\frac{1}{2}}\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^{p}\sum_{k=1}^{p}\left(\frac{\partial\phi'}{\partial\theta_i}\mathbf{D}_\phi^2 f_j\frac{\partial\phi}{\partial\theta_k} + \mathbf{D}_\phi f_i\frac{\partial^2\phi}{\partial\theta_i\partial\theta_k}\right)x_i y_k$$

$$-\sum_{i=1}^{p}\sum_{k=1}^{p}\left(\frac{\partial\phi'}{\partial\theta_i}\mathbf{D}_\phi^2 g\frac{\partial\phi}{\partial\theta_k} + \mathbf{D}_\phi g\frac{\partial^2\phi}{\partial\theta_i\partial\theta_k}\right)x_i y_k$$

$$= \mathbf{D}_\phi g(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}}\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^{p}\sum_{k=1}^{p}\frac{\partial\phi'}{\partial\theta_i}\mathbf{D}_\phi^2 f_j\frac{\partial\phi}{\partial\theta_k}x_i y_k - \sum_{i=1}^{p}\sum_{k=1}^{p}\frac{\partial\phi'}{\partial\theta_i}\mathbf{D}_\phi^2 g\frac{\partial\phi}{\partial\theta_k}x_i y_k$$

$$= \mathbf{Q}_\phi\sum_j \omega_j^{\frac{1}{2}}(\mathbf{J}^{-1}\mathbf{x})'\mathbf{D}_\phi^2 f_j(\mathbf{J}^{-1}\mathbf{y}) - (\mathbf{J}^{-1}\mathbf{x})'\mathbf{D}_\phi^2 g(\mathbf{J}^{-1}\mathbf{y}) \qquad \text{(C-11)}$$

where $\mathbf{D}_\phi g(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{1/2}\sum_j \omega_j^{1/2}\mathbf{D}_\phi f_j - \mathbf{D}_\phi g = \mathbf{0}$ and

$$\mathbf{Q}_\phi = \omega^{\frac{1}{2}}\mathbf{D}_\phi\mathbf{f}(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi g' \qquad \text{(C-12)}$$

## Terms Expressing Intrinsic Nonlinearity with Respect to $\mathbf{f}(\beta)$

Invariance of the form $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}$ can only be shown to be approximate. In this case unique transformations $\phi(\theta)$ and $\alpha(\beta)$ are used. First some preliminary relations are needed. Because of the equality $\beta = \gamma\bar{\theta} + \mathbf{e}$, where $\mathbf{e}$ is not functionally dependent on $\bar{\theta}$, the chain rule of calculus gives

$$\frac{\partial f_i}{\partial\bar{\theta}_j} = \sum_{k=1}^{m}\frac{\partial f_i}{\partial\beta_k}\frac{\partial\beta_k}{\partial\bar{\theta}_j} = \sum_{k=1}^{m}\frac{\partial f_i}{\partial\beta_k}\gamma_{kj} \qquad \text{(C-13)}$$

or, if evaluated at the set $\beta = \gamma\bar{\theta}$,

$$\mathbf{D}\mathbf{f} = \mathbf{D}_\beta \mathbf{f} \boldsymbol{\gamma} \tag{C-14}$$

The derivative of (C-13) is

$$\frac{\partial}{\partial \bar{\theta}_\ell}\left(\frac{\partial f_i}{\partial \bar{\theta}_j}\right) = \sum_{r=1}^{m} \frac{\partial}{\partial \beta_r}\left(\sum_{k=1}^{m} \frac{\partial f_i}{\partial \beta_k}\gamma_{kj}\right)\frac{\partial \beta_r}{\partial \bar{\theta}_\ell} = \sum_{k=1}^{m}\sum_{r=1}^{m} \frac{\partial^2 f_i}{\partial \beta_k \partial \beta_r}\gamma_{kj}\gamma_{r\ell} \tag{C-15}$$

or, if evaluated at the set $\boldsymbol{\beta} = \boldsymbol{\gamma}\bar{\boldsymbol{\theta}}$,

$$\mathbf{D}^2 f_i = \boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_i \boldsymbol{\gamma};\ i = 1,2,\ldots,n \tag{C-16}$$

Finally, the Jacobian of the transformation $\boldsymbol{\alpha}(\boldsymbol{\beta})$ evaluated at $\boldsymbol{\beta} = \boldsymbol{\gamma}\bar{\boldsymbol{\theta}}$ is defined as the $m \times m$ nonsingular matrix

$$\mathbf{J}_\beta = \left[\frac{\partial \beta_i}{\partial \alpha_j}\right];\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,m \tag{C-17}$$

so that

$$\mathbf{D}_\beta \mathbf{f} \mathbf{J}_\beta = \mathbf{D}_\alpha \mathbf{f} = \left[\frac{\partial f_i}{\partial \alpha_j}\right];\ i = 1,2,\ldots,n;\ j = 1,2,\ldots,m \tag{C-18}$$

Next, the approximate relation of $\bar{\boldsymbol{\alpha}}$ to $\bar{\boldsymbol{\phi}}$ is obtained. From the definition $\bar{\boldsymbol{\beta}} = \boldsymbol{\gamma}\bar{\boldsymbol{\theta}}$,

$$d\bar{\boldsymbol{\beta}} = \boldsymbol{\gamma}\, d\bar{\boldsymbol{\theta}} \tag{C-19}$$

Also

$$d\bar{\boldsymbol{\beta}} = \mathbf{J}_\beta\, d\bar{\boldsymbol{\alpha}} \tag{C-20}$$

and

$$d\bar{\boldsymbol{\theta}} = \mathbf{J}\, d\bar{\boldsymbol{\phi}} \tag{C-21}$$

Substitution of (C-20) and (C-21) into (C-19) and premultiplication of the result by $\mathbf{J}_\beta^{-1}$ yield

$$d\bar{\boldsymbol{\alpha}} = \mathbf{J}_\beta^{-1}\boldsymbol{\gamma}\mathbf{J}\, d\bar{\boldsymbol{\phi}} \tag{C-22}$$

Now, if $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ transform similarly because $\boldsymbol{\beta}$ is just at smaller scale than $\boldsymbol{\theta}$, then $\boldsymbol{\gamma}\mathbf{J}$ is an approximate interpolation of $\mathbf{J}_\beta$, so that at $\boldsymbol{\beta} = \boldsymbol{\gamma}\bar{\boldsymbol{\theta}}$, $\mathbf{J}_\beta^{-1}\boldsymbol{\gamma}\mathbf{J}$ is approximately constant. With this approximation, (C-22) can be integrated to give

$$\bar{\alpha} \approx \lambda\bar{\phi} + [\textit{integration constants}] \tag{C-23}$$

where

$$\lambda = \mathbf{J}_\beta^{-1}\gamma\mathbf{J} \tag{C-24}$$

With the above results, transformation of $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}$ in an analogous manner to transformation of $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{x}'\mathbf{D}^2 f_j\mathbf{y}$ results in

$$\begin{aligned}
&(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} \\
&= (\mathbf{I}-\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}})\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^m\sum_{k=1}^m \frac{\partial^2 f_j}{\partial\beta_i\partial\beta_k}e_i e_k \\
&= (\mathbf{I}-\omega^{\frac{1}{2}}\mathbf{D}_\phi\mathbf{f}(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}})\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^m\sum_{k=1}^m \left(\frac{\partial\alpha'}{\partial\beta_i}\mathbf{D}_\alpha^2 f_j\frac{\partial\alpha}{\partial\beta_k} + \mathbf{D}_\alpha f_j\frac{\partial^2\alpha}{\partial\beta_i\partial\beta_k}\right)e_i e_k
\end{aligned} \tag{C-25}$$

where the derivatives are evaluated at $\beta = \gamma\bar{\theta}$. The term involving $\mathbf{D}_\alpha\mathbf{f}$ must be written in terms of $\mathbf{D}_\phi\mathbf{f}$ in order to remove it as was done in (C-4). First (C-2), (C-14), (C-18), and (C-24) are used to obtain

$$\mathbf{D}_\phi\mathbf{f} = \mathbf{DfJ} = \mathbf{D}_\beta\mathbf{f}\gamma\mathbf{J} = \mathbf{D}_\alpha\mathbf{f}\mathbf{J}_\beta^{-1}\gamma\mathbf{J} = \mathbf{D}_\alpha\mathbf{f}\lambda \tag{C-26}$$

Next, approximation of $\partial^2\alpha/\partial\beta_i\partial\beta_k$ evaluated at $\beta = \gamma\bar{\theta}$ with its best-fit vector $\lambda\mathbf{a}_{ik}$ gives

$$\frac{\partial^2\alpha}{\partial\beta_i\partial\beta_k} \approx \lambda\mathbf{a}_{ik} \tag{C-27}$$

where $\mathbf{a}_{ik}$; $i = 1,2,\ldots,m$; $k = 1,2,\ldots,m$ is a set of vectors equal to the set $(\lambda'\mathbf{w}_{ik}\lambda)^{-1}\lambda'\mathbf{w}_{ik}\partial^2\alpha/\partial\beta_i\partial\beta_k$ and $\mathbf{w}_{ik}$ is a set of weight matrices to be explained. Again, the derivatives are evaluated at $\beta = \gamma\bar{\theta}$. Finally, substitution of (C-26) and (C-27) into (C-25) yields

$$\begin{aligned}
&(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} \\
&\approx (\mathbf{I}-\omega^{\frac{1}{2}}\mathbf{D}_\phi\mathbf{f}(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}})\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^m\sum_{k=1}^m \left(\frac{\partial\alpha'}{\partial\beta_i}\mathbf{D}_\alpha^2 f_j\frac{\partial\alpha}{\partial\beta_k} + \mathbf{D}_\phi f_j\mathbf{a}_{ik}\right)e_i e_k \\
&= (\mathbf{I}-\mathbf{R}_\phi)\sum_j \omega_j^{\frac{1}{2}}(\mathbf{J}_\beta^{-1}\mathbf{e})'\mathbf{D}_\alpha^2 f_j(\mathbf{J}_\beta^{-1}\mathbf{e})
\end{aligned} \tag{C-28}$$

If each approximation of $\mathbf{D}_\alpha f_j\partial^2\alpha/\partial\beta_i\partial\beta_k$ with $\mathbf{D}_\phi f_j\mathbf{a}_{ik}$ is accurate, then $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}$ is nearly invariant and the term is small if $\mathbf{D}_\alpha^2 f_j$ is small. If each approximation is not accurate, then the term may not be small even if $\mathbf{D}_\alpha^2 f_j$ is small. The weight matrices are conceptually designed to give the most nearly invariant final result. An approximation analogous to (C-27) is developed with a different perspective in paragraphs containing (F-104)-(F-108), appendix F.

Finally, (C-4), (C-16), and (C-28) are used to show that the expressions $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}$ and $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{e}'(\mathbf{D}_\beta^2 f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}$ are at least approximately invariant. First

$$(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}$$

$$= (\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 f_j (\gamma'\gamma)^{-1}\gamma'\mathbf{e} \tag{C-29}$$

so that, with $(\gamma'\gamma)^{-1}\gamma'\mathbf{e} = \mathbf{x} = \mathbf{y}$, (C-29) is of the form $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{x}'\mathbf{D}^2 f_j \mathbf{y}$, which from (C-4) is invariant. The difference between (C-28) and (C-29) gives the second expression, which, therefore, is approximately invariant.

## Terms Expressing Combined Intrinsic Nonlinearity with Respect to f(β) and $g(\beta)$

As for $(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{1/2}\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e}$, invariance of the form $\mathbf{Q}'\sum_j \omega_j^{1/2}\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{e}'\mathbf{D}_\beta^2 g\mathbf{e}$ can only be shown to be approximate. The same basis and approximations as used to get (C-28) are again used, but they are extended to involve $g(\boldsymbol{\beta})$ also. For this note that

$$\mathbf{D}_\phi g = \mathbf{D}g\mathbf{J} = \mathbf{D}_\beta g \gamma \mathbf{J} = \mathbf{D}_\alpha g \mathbf{J}_\beta^{-1}\gamma\mathbf{J} = \mathbf{D}_\alpha g \lambda \tag{C-30}$$

Then

$$\mathbf{Q}'\sum_j \omega_j^{\frac{1}{2}}\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \mathbf{e}'\mathbf{D}_\beta^2 g\mathbf{e}$$

$$= \mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^{m}\sum_{k=1}^{m}\frac{\partial^2 f_j}{\partial\beta_i\partial\beta_k}e_i e_k - \sum_{i=1}^{m}\sum_{k=1}^{m}\frac{\partial^2 g}{\partial\beta_i\partial\beta_k}e_i e_k$$

$$= \mathbf{D}_\phi g(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}}\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^{m}\sum_{k=1}^{m}\left(\frac{\partial\alpha'}{\partial\beta_i}\mathbf{D}_\alpha^2 f_j \frac{\partial\alpha}{\partial\beta_k} + \mathbf{D}_\alpha f_j \frac{\partial^2\alpha}{\partial\beta_i\partial\beta_k}\right)e_i e_k$$

$$-\sum_{i=1}^{m}\sum_{k=1}^{m}\left(\frac{\partial\alpha'}{\partial\beta_i}\mathbf{D}_\alpha^2 g\frac{\partial\alpha}{\partial\beta_k} + \mathbf{D}_\alpha g\frac{\partial^2\alpha}{\partial\beta_i\partial\beta_k}\right)e_i e_k$$

$$\approx \mathbf{D}_\phi g(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}}\sum_j \omega_j^{\frac{1}{2}}\sum_{i=1}^{m}\sum_{k=1}^{m}\left(\frac{\partial\alpha'}{\partial\beta_i}\mathbf{D}_\alpha^2 f_j \frac{\partial\alpha}{\partial\beta_k} + \mathbf{D}_\phi f_j \mathbf{a}_{ik}\right)e_i e_k$$

$$-\sum_{i=1}^{m}\sum_{k=1}^{m}\left(\frac{\partial\boldsymbol{\alpha}'}{\partial\beta_i}\mathbf{D}_\alpha^2 g\frac{\partial\boldsymbol{\alpha}}{\partial\beta_k}+\mathbf{D}_\phi g\mathbf{a}_{ik}\right)e_i e_k$$
$$=\mathbf{Q}'_\phi\sum_j\omega_j^{\frac{1}{2}}(\mathbf{J}_\beta^{-1}\mathbf{e})'\mathbf{D}_\alpha^2 f_j(\mathbf{J}_\beta^{-1}\mathbf{e})-(\mathbf{J}_\beta^{-1}\mathbf{e})'\mathbf{D}_\alpha^2 g(\mathbf{J}_\beta^{-1}\mathbf{e}) \qquad \text{(C-31)}$$

so that $\mathbf{Q}'\sum_j\omega_j^{1/2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{e}'\mathbf{D}_\beta g\mathbf{e}$ is approximately invariant.

Finally, use of (C-11), (C-16), and (C-31) shows that the expressions $\mathbf{Q}'\sum_j\omega_j^{1/2}\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_j\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}-\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 g\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}$ and $\mathbf{Q}'\sum_j\omega_j^{1/2}\mathbf{e}'(\mathbf{D}_\beta^2 f_j-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_j\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{e}-\mathbf{e}'(\mathbf{D}_\beta^2 g-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 g\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{e}$ also are approximately invariant. The development is analogous to the development in (C-29) and the discussion following (C-29).

# Appendix D – Third-Order Analysis of the Objective Function When the Weights are Unknown

When the weights are unknown, an objective function can be written as (4-50):

$$\ell(\theta) = \frac{1}{2}\sum_{k=1}^{q} n_k \ln(\sum_{i(k)} \eta_i^2) \tag{D-1}$$

where

$$\eta_i = Y_i - f_i(\gamma\theta) \tag{D-2}$$

Through third order in $\eta_i$, $\ell(\theta)$ is shown in this section to be proportional to (4-47) plus some constants, so that through second order in $\eta_i$ the normal equations obtained by minimizing $\ell(\theta)$ are equivalent to the normal equations obtained by minimizing (4-47).

First a set of weights $\omega_{Gk}$, $k = 1,2,...,q$, are defined from

$$\omega_{Gk}^{-1} = \frac{\sigma_\varepsilon^{-2}}{n_k}\sum_{i(k)} E(Y_i - f_i(\gamma\theta_*))^2;\ k = 1,2,...,q \tag{D-3}$$

Because the expected value and variance of $Y_i - f_i(\gamma\theta_*)$ are taken to be uniform in each group,

$$\sum_{i(k)} E(Y_i - f_i(\gamma\theta_*))^2 = n_k E(Y_i - f_i(\gamma\theta_*))^2 \tag{D-4}$$

so that

$$E(Y_i - f_i(\gamma\theta_*))^2 = \omega_{Gk}^{-1}\sigma_\varepsilon^2 \tag{D-5}$$

Now expansion of $\ell(\theta)$ through second order in $\eta_i^2 - E(Y_i - f_i(\gamma\theta_*))^2$ using a truncated Taylor series yields

$$\begin{aligned}\ell(\theta) \approx& \frac{1}{2}\sum_{k=1}^{q} n_k \ln(n_k\omega_{Gk}^{-1}\sigma_\varepsilon^2) + \frac{1}{2}\sum_{k=1}^{q} n_k \sum_{i(k)} \frac{\partial}{\partial(\eta_i^2)} \ln(\sum_{\ell(k)} \eta_\ell^2)(\eta_i^2 - \omega_{Gk}^{-1}\sigma_\varepsilon^2)\\ &+ \frac{1}{4}\sum_{k=1}^{q} n_k \sum_{i(k)}\sum_{j(k)} \frac{\partial^2}{\partial(\eta_i^2)\partial(\eta_j^2)} \ln(\sum_{\ell(k)} \eta_\ell^2)(\eta_i^2 - \omega_{Gk}^{-1}\sigma_\varepsilon^2)(\eta_j^2 - \omega_{Gk}^{-1}\sigma_\varepsilon^2)\end{aligned} \tag{D-6}$$

where the derivatives are evaluated at $\eta_i^2 = \eta_{*i}^2 = E(Y_i - f_i(\gamma\theta_*))^2$:

$$\frac{\partial}{\partial(\eta_i^2)} \ln(\sum_{\ell(k)} \eta_\ell^2) = \frac{1}{\sum_{\ell(k)} \eta_{*\ell}^2} = \frac{\omega_{Gk}}{\sigma_\varepsilon^2 n_k} \tag{D-7}$$

$$\frac{\partial^2}{\partial(\eta_i^2)\partial(\eta_j^2)}\ln(\sum_{\ell(k)}\eta_\ell^2) = \frac{\partial}{\partial(\eta_j^2)}\frac{1}{\sum_{\ell(k)}\eta_\ell^2} = -\frac{\omega_{Gk}^2}{\sigma_\varepsilon^4 n_k^2} \tag{D-8}$$

Substitution of (D-7) and (D-8) into (D-6) produces

$$\begin{aligned}
\ell(\theta) &\approx \frac{1}{2}\sum_{k=1}^{q} n_k \ln(n_k\omega_{Gk}^{-1}\sigma_\varepsilon^2) + \frac{1}{2\sigma_\varepsilon^2}\sum_{k=1}^{q}\omega_{Gk}\sum_{i(k)}(\eta_i^2 - \omega_{Gk}^{-1}\sigma_\varepsilon^2) \\
&- \frac{1}{4\sigma_\varepsilon^4}\sum_{k=1}^{q}\frac{\omega_{Gk}^2}{n_k}\sum_{i(k)}\sum_{j(k)}(\eta_i^2 - \omega_{Gk}^{-1}\sigma_\varepsilon^2)(\eta_j^2 - \omega_{Gk}^{-1}\sigma_\varepsilon^2) \\
&= \frac{1}{2}\sum_{k=1}^{q} n_k \ln(n_k\omega_{Gk}^{-1}\sigma_\varepsilon^2) + \frac{1}{2\sigma_\varepsilon^2}\sum_{k=1}^{q}\omega_{Gk}\sum_{i(k)}\eta_i^2 - \frac{n}{2} - \frac{1}{4\sigma_\varepsilon^4}\sum_{k=1}^{q}\frac{\omega_{Gk}^2}{n_k}\sum_{i(k)}\sum_{j(k)}\eta_i^2\eta_j^2 \\
&+ \frac{1}{4\sigma_\varepsilon^2}\sum_{k=1}^{q}\omega_{Gk}\sum_{i(k)}\eta_i^2 + \frac{1}{4\sigma_\varepsilon^2}\sum_{k=1}^{q}\omega_{Gk}\sum_{j(k)}\eta_j^2 - \frac{n}{4} \\
&= \frac{1}{\sigma_\varepsilon^2}S(\theta) - \frac{1}{4\sigma_\varepsilon^4}\sum_{k=1}^{q}\frac{\omega_{Gk}^2}{n_k}\sum_{i(k)}\sum_{j(k)}\eta_i^2\eta_j^2 + \frac{1}{2}\sum_{k=1}^{q} n_k\ln(n_k\omega_{Gk}^{-1}\sigma_\varepsilon^2) - \frac{3n}{4}
\end{aligned} \tag{D-9}$$

where $n = \sum_{k=1}^{q} n_k$. Thus, through third order in $\eta_i$, $\ell(\theta)$ and $S(\theta) + \mathit{constants}$ are proportional, the proportionality factor being $\sigma_\varepsilon^{-2}$.

Minimization of $\ell(\theta)$ with respect to $\theta$ using (D-1) yields

$$\sum_{k=1}^{q} w_{Gk}\sum_{i(k)}\mathbf{D}\hat{f}_i\,(Y_i - f_i(\gamma\hat{\theta})) = \mathbf{0} \tag{D-10}$$

where

$$w_{Gk}^{-1} \propto \frac{1}{n_k}\sum_{i(k)}(Y_i - f_i(\gamma\hat{\theta}))^2 \tag{D-11}$$

Also, minimization of $S(\theta)$ using (4-47) results in

$$\sum_{k=1}^{q}\omega_{Gk}\sum_{i(k)}\mathbf{D}\hat{f}_i\,(Y_i - f_i(\gamma\hat{\theta})) = \mathbf{0} \tag{D-12}$$

From (D-9) it can be seen that (D-10) and (D-12) differ by terms of third order in $\eta_i$. Thus, through second order in $\eta_i$, the two sets of normal equations are equivalent.

# Appendix E – Second-Order Correct Constrained Regression Estimates; Third-Order Correct Constrained Sum of Squares Estimates

## Second-Order Correct Constrained Parameter Estimates for Confidence Intervals

A Lagrange multiplier formulation is used to obtain a constrained regression estimate, $\tilde{\theta}$, of $\theta_*$. The approximation method and solution procedure are similar to the ones used in appendix B to obtain $\hat{\theta}$.

First let $\tilde{\theta}-\bar{\theta}=\tilde{\mathbf{l}}+\tilde{\mathbf{q}}$, where $\tilde{\mathbf{l}}$ is the first-order term and $\tilde{\mathbf{q}}$ is the second-order term, both of which are to be obtained. Then, a second-order Taylor series expansion gives, up through second-order in $\mathbf{U}_*$ and $\mathbf{e}$,

$$
\begin{aligned}
f_i(\gamma\tilde{\theta}) &= f_i(\gamma\bar{\theta})+\mathbf{D}f_i(\tilde{\theta}-\bar{\theta})+\frac{1}{2}(\tilde{\theta}-\bar{\theta})'\mathbf{D}^2 f_i(\tilde{\theta}-\bar{\theta}) \\
&= f_i(\gamma\bar{\theta})+\mathbf{D}f_i\tilde{\mathbf{l}}+\mathbf{D}f_i\tilde{\mathbf{q}}+\frac{1}{2}(\tilde{\mathbf{l}}+\tilde{\mathbf{q}})'\mathbf{D}^2 f_i(\tilde{\mathbf{l}}+\tilde{\mathbf{q}}) \\
&\approx f_i(\gamma\bar{\theta})+\mathbf{D}f_i\tilde{\mathbf{l}}+\mathbf{D}f_i\tilde{\mathbf{q}}+\frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_i\tilde{\mathbf{l}}
\end{aligned}
\tag{E-1}
$$

and

$$
\begin{aligned}
g(\gamma\tilde{\theta}) &= g(\gamma\bar{\theta})+\mathbf{D}g(\tilde{\theta}-\bar{\theta})+\frac{1}{2}(\tilde{\theta}-\bar{\theta})'\mathbf{D}^2 g(\tilde{\theta}-\bar{\theta}) \\
&\approx g(\gamma\bar{\theta})+\mathbf{D}g\tilde{\mathbf{l}}+\mathbf{D}g\tilde{\mathbf{q}}+\frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}}
\end{aligned}
\tag{E-2}
$$

Next using (3-17),

$$
\begin{aligned}
f_i(\gamma\theta_*) &= f_i(\gamma\bar{\theta})+\mathbf{D}f_i(\theta_*-\bar{\theta})+\frac{1}{2}(\theta_*-\bar{\theta})'\mathbf{D}^2 f_i(\theta_*-\bar{\theta}) \\
&= f_i(\gamma\bar{\theta})+\mathbf{D}f_i(\gamma'\gamma)^{-1}\gamma'\mathbf{e}+\frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 f_i(\gamma'\gamma)^{-1}\gamma'\mathbf{e}
\end{aligned}
\tag{E-3}
$$

and

$$
\begin{aligned}
g(\gamma\theta_*) &= g(\gamma\bar{\theta})+\mathbf{D}g(\theta_*-\bar{\theta})+\frac{1}{2}(\theta_*-\bar{\theta})'\mathbf{D}^2 g(\theta_*-\bar{\theta}) \\
&= g(\gamma\bar{\theta})+\mathbf{D}g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}+\frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}
\end{aligned}
\tag{E-4}
$$

Combination of (E-1) and (E-3), and (E-2) and (E-4), results in

$$f_i(\gamma\tilde{\theta}) = f_i(\gamma\theta_*) + \mathbf{D}f_i(\tilde{\theta} - \theta_*) + \frac{1}{2}(\tilde{\theta} - \bar{\theta})'\mathbf{D}^2 f_i(\tilde{\theta} - \bar{\theta}) - \frac{1}{2}(\theta_* - \bar{\theta})'\mathbf{D}^2 f_i(\theta_* - \bar{\theta})$$

$$\approx f_i(\gamma\theta_*) + \mathbf{D}f_i(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) + \mathbf{D}f_i\tilde{\mathbf{q}} + \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_i\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 f_i\,\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e} \qquad \text{(E-5)}$$

and

$$g(\gamma\tilde{\theta}) = g(\gamma\theta_*) + \mathbf{D}g(\tilde{\theta} - \theta_*) + \frac{1}{2}(\tilde{\theta} - \bar{\theta})'\mathbf{D}^2 g(\tilde{\theta} - \bar{\theta}) - \frac{1}{2}(\theta_* - \bar{\theta})'\mathbf{D}^2 g\,(\theta_* - \bar{\theta})$$

$$\approx g(\gamma\theta_*) + \mathbf{D}g(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) + \mathbf{D}g\tilde{\mathbf{q}} + \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e} \qquad \text{(E-6)}$$

The constrained regression estimate is obtained by minimizing $S(\theta)$ subject to the constraint that $g(\gamma\theta) = g(\gamma\theta_*)$. This can be formulated as the Lagrange multiplier problem (Boas, 1966, p. 145-150)

$$L(\theta, \lambda) = S(\theta) + 2\lambda(g(\gamma\theta_*) - g(\gamma\theta)) \qquad \text{(E-7)}$$

Minimization of (E-7) is accomplished using (3-30), (4-9), (C-16), (E-5), and (E-6), keeping terms up through second order in $\mathbf{U}_*$, $\mathbf{e}$, or their product. The result with respect to $\theta$ is

$$\frac{\partial L}{\partial \theta} = -2\sum_i\sum_j \mathbf{D}\tilde{f}_i'\omega_{ij}(Y_j - f_j(\gamma\tilde{\theta})) - 2\lambda\mathbf{D}\tilde{g}'$$

$$\approx -2\sum_i\sum_j (\mathbf{D}f_i' + \mathbf{D}^2 f_i\tilde{\mathbf{l}})\omega_{ij}(Y_j - f_j(\gamma\theta_*) - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\tilde{\mathbf{q}}$$

$$-\frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}} + \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 f_j\,(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - 2\lambda(\mathbf{D}g' + \mathbf{D}^2 g\tilde{\mathbf{l}})$$

$$= -2\sum_i\sum_j (\mathbf{D}f_i' + \mathbf{D}^2 f_i\tilde{\mathbf{l}})\omega_{ij}(U_{*j} + \frac{1}{2}\mathbf{e}'(\mathbf{D}^2_\beta f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} - \mathbf{D}f_j(\tilde{\mathbf{l}}$$

$$-(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\tilde{\mathbf{q}} - \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}} + \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 f_j\,(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - 2\lambda(\mathbf{D}g' + \mathbf{D}^2 g\tilde{\mathbf{l}})$$

$$\approx -2\sum_i\sum_j \mathbf{D}f_i'\omega_{ij}(U_{*j} + \frac{1}{2}\mathbf{e}'\mathbf{D}^2_\beta f_j\,\mathbf{e} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\tilde{\mathbf{q}} - \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$-2\sum_i\sum_j \mathbf{D}^2 f_i\tilde{\mathbf{l}}\,\omega_{ij}(U_{*j} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e})) - 2\lambda(\mathbf{D}g' + \mathbf{D}^2 g\tilde{\mathbf{l}})$$

$$= \mathbf{0}$$

or

$$\sum_i\sum_j \mathbf{D}f_i'\omega_{ij}(U_{*j} + \frac{1}{2}\,\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\tilde{\mathbf{q}} - \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$+\sum_i\sum_j \mathbf{D}^2 f_i \tilde{\mathbf{l}}\,\omega_{ij}(U_{*j} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e})) + \lambda(\mathbf{D}g' + \mathbf{D}^2 g\tilde{\mathbf{l}}) = \mathbf{0} \tag{E-8}$$

where $\mathbf{D}\tilde{f}_i$ and $\mathbf{D}\tilde{g}$ indicate evaluation at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$. The result with respect to $\lambda$ is

$$\begin{aligned}\frac{\partial L}{\partial \lambda} &= 2(g(\gamma\boldsymbol{\theta}_*) - g(\gamma\tilde{\boldsymbol{\theta}}))\\
&\approx -2(\mathbf{D}g(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) + \mathbf{D}g\tilde{\mathbf{q}} + \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})\\
&= 0\end{aligned}$$

or

$$\mathbf{D}g(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) + \mathbf{D}g\tilde{\mathbf{q}} + \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e} = 0 \tag{E-9}$$

Because $\tilde{\mathbf{l}}$ is the first-order solution, it must satisfy

$$\sum_i\sum_j \mathbf{D}f_i'\omega_{ij}(U_{*j} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e})) + \lambda\,\mathbf{D}g' = \mathbf{0} \tag{E-10}$$

or, in matrix form,

$$\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}\mathbf{U}_* + \lambda(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g' \tag{E-11}$$

and

$$\mathbf{D}g(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) = 0 \tag{E-12}$$

To solve for $\lambda$ (E-11) is premultiplied by $\mathbf{D}g$ and (E-12) is used to get

$$\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}\mathbf{U}_* + \lambda\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g' = 0$$

or

$$\lambda = -\frac{\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}\mathbf{U}_*}{\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'} \tag{E-13}$$

Then, substitution of (E-13) into (E-11) yields

$$\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}\mathbf{U}_* - \frac{\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}\mathbf{U}_*}{\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g'$$

$$= (\mathbf{Df}'\omega\mathbf{Df})^{-1}(\mathbf{I} - \frac{\mathbf{D}g'\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}}{\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'})\mathbf{Df}'\omega\mathbf{U}_*$$

$$= (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}$$

$$- \frac{\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}}{\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'})\omega^{\frac{1}{2}}\mathbf{U}_*$$

$$= (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{U}_* \tag{E-14}$$

To evaluate $\widetilde{\mathbf{q}}$ (E-14) is substituted into (E-8) to get

$$\sum_j \mathbf{Df}'\omega_j(U_{*j} + \frac{1}{2}\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e}) - \mathbf{Df}'\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{U}_*$$

$$- \mathbf{Df}'\omega\mathbf{Df}\widetilde{\mathbf{q}} - \frac{1}{2}\sum_j \mathbf{Df}'\omega_j\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j\widetilde{\mathbf{l}} + \sum_i \mathbf{D}^2 f_i\widetilde{\mathbf{l}}\omega_i(\mathbf{U}_* - \mathbf{Df}(\widetilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}))$$

$$+ \lambda(\mathbf{D}g' + \mathbf{D}^2 g\widetilde{\mathbf{l}})$$

$$= \frac{1}{2}\sum_j \mathbf{Df}'\omega_j\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} + \mathbf{Df}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{U}_* - \mathbf{Df}'\omega\mathbf{Df}\widetilde{\mathbf{q}} - \frac{1}{2}\sum_j \mathbf{Df}'\omega_j\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j\widetilde{\mathbf{l}}$$

$$+ \sum_i \mathbf{D}^2 f_i\widetilde{\mathbf{l}}\omega_i(\mathbf{U}_* - \mathbf{Df}(\widetilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e})) + \lambda(\mathbf{D}g' + \mathbf{D}^2 g\widetilde{\mathbf{l}}) = \mathbf{0} \tag{E-15}$$

where $\mathbf{Df}'\omega^{1/2}\mathbf{R}\omega^{1/2} = \mathbf{Df}'\omega$. Now

$$\mathbf{U}_* - \mathbf{Df}(\widetilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) = \mathbf{U}_* - \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{U}_*$$

$$= (\mathbf{I} - \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}})\mathbf{U}_*$$

$$= \omega^{-\frac{1}{2}}(\mathbf{I} - \omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}))\omega^{\frac{1}{2}}\mathbf{U}_*$$

$$= \omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{U}_* \tag{E-16}$$

where $\mathbf{RQQ}' = \mathbf{QQ}'$. Let $\widetilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{R} + \mathbf{QQ}'/\mathbf{Q}'\mathbf{Q})\omega^{1/2}\mathbf{U}_*$. Then (E-15) becomes

$$\frac{1}{2}\sum_j \mathbf{Df}'\omega_j\mathbf{e}'\mathbf{D}^2_\beta f_j\mathbf{e} + \mathbf{Df}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{U}_* - \mathbf{Df}'\omega\mathbf{Df}\widetilde{\mathbf{q}} - \frac{1}{2}\sum_j \mathbf{Df}'\omega_j\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j\widetilde{\mathbf{l}}$$

$$+ \sum_i \mathbf{D}^2 f_i\widetilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}} + \lambda(\mathbf{D}g' + \mathbf{D}^2 g\widetilde{\mathbf{l}}) = \mathbf{0} \tag{E-17}$$

or

$$\tilde{\mathbf{q}} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}(\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\boldsymbol{\omega}_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Df}'\boldsymbol{\omega}_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j \tilde{\mathbf{l}})$$
$$+\lambda(\mathbf{D}g' + \mathbf{D}^2 g\tilde{\mathbf{l}})) \tag{E-18}$$

Putting (E-14) and (E-18) into (E-9) permits evaluation of $\lambda$:

$$\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}(\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\boldsymbol{\omega}_i^{\frac{1}{2}}\tilde{\mathbf{Z}}$$
$$+\frac{1}{2}\sum_j \mathbf{Df}'\boldsymbol{\omega}_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}) + \lambda(\mathbf{D}g' + \mathbf{D}^2 g\tilde{\mathbf{l}})) + \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\mathbf{D}^2 g(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}$$
$$= \mathbf{Q}'(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}}\boldsymbol{\omega}_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Q}'\boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$
$$+\lambda(\mathbf{Q}'\mathbf{Q} + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}) + \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\mathbf{D}g(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}$$
$$= \mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}}\boldsymbol{\omega}_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Q}'\boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$
$$+\lambda(\mathbf{Q}'\mathbf{Q} + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}) + \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\mathbf{D}g(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}$$
$$= 0$$

where

$$\mathbf{Q}'(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}) = \mathbf{Q}'\mathbf{R} - \mathbf{Q}' = \mathbf{0} \tag{E-19}$$

Therefore,

$$\lambda = -\frac{1}{d}(\mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}}\boldsymbol{\omega}_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Q}'\boldsymbol{\omega}_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$
$$+\frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\mathbf{D}^2 g(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}) \tag{E-20}$$

where, through first order in the temporary variable $\varepsilon$ (which is all that is required),

$$d^{-1} = (\mathbf{Q}'\mathbf{Q} + \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}})^{-1} = (\mathbf{Q}'\mathbf{Q} + \varepsilon)^{-1}$$
$$\approx (\mathbf{Q}'\mathbf{Q})^{-1} - (\mathbf{Q}'\mathbf{Q} + \varepsilon)^{-2}\big|_{\varepsilon=0}\,\varepsilon$$
$$= (\mathbf{Q}'\mathbf{Q})^{-1} - (\mathbf{Q}'\mathbf{Q})^{-2}\mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}} \tag{E-21}$$

Putting (E-21) into (E-20) yields, through second order

$$\lambda \approx -\frac{1}{\mathbf{Q}'\mathbf{Q}}(\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_* + \mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Q}'\omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j \tilde{\mathbf{l}})$$

$$+\frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) + \frac{1}{(\mathbf{Q}'\mathbf{Q})^2}\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_* \qquad \text{(E-22)}$$

Next, substitution of (E-22) into (E-18) gives

$$\tilde{\mathbf{q}} \approx (\mathbf{Df}'\omega\mathbf{Df})^{-1}(\mathbf{Df}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{U}_* + \sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Df}'\omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$-(\frac{1}{\mathbf{Q}'\mathbf{Q}}(\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_* + \mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Q}'\omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$+\frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \frac{1}{(\mathbf{Q}'\mathbf{Q})^2}\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)(\mathbf{D}g' + \mathbf{D}^2 g\tilde{\mathbf{l}}))$$

$$\approx (\mathbf{Df}'\omega\mathbf{Df})^{-1}(\mathbf{Df}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{U}_* + \sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Df}'\omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$-\frac{1}{\mathbf{Q}'\mathbf{Q}}(\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_* + \mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}} + \frac{1}{2}\sum_j \mathbf{Q}'\omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$+\frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})\mathbf{D}g' + \frac{1}{(\mathbf{Q}'\mathbf{Q})^2}\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*\mathbf{D}g'$$

$$-\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*\mathbf{D}^2 g\tilde{\mathbf{l}}) \qquad \text{(E-23)}$$

Pairs of terms in (E-23) evaluate to become the following.

$$\mathbf{Df}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{U}_* - \frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*$$

$$= \mathbf{Df}'\omega^{\frac{1}{2}}\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{U}_* - \frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*$$

$$= \mathbf{0} \qquad \text{(E-24)}$$

$$\sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}} - \frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\,\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}}$$

$$= (\mathbf{I} - \frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1})\sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}}$$

$$= \mathbf{Df}'\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\mathbf{I} - \frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1})\mathbf{Df}'\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}}$$

$$= \mathbf{Df}'\omega^{\frac{1}{2}}(\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}} - \frac{1}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}})\omega^{\frac{1}{2}}$$

$$\bullet\ \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}}\,\omega_i^{\frac{1'}{2}}\tilde{\mathbf{Z}}$$

$$= \mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}}\,\omega_i^{\frac{1'}{2}}\tilde{\mathbf{Z}} \qquad \text{(E-25)}$$

$$\frac{1}{2}\sum_j \mathbf{Df}'\omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})-\frac{1}{2}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'\sum_j \mathbf{Q}'\omega_j^{\frac{1'}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$=\frac{1}{2}\mathbf{Df}'\omega\mathbf{Df}(\mathbf{Df}\omega\mathbf{Df})^{-1}\sum_j \mathbf{Df}'\omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$-\frac{1}{2}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Df}'\omega\mathbf{Df}(\mathbf{Df}\omega\mathbf{Df})^{-1}\mathbf{D}g'\sum_j \mathbf{Q}'\omega_j^{\frac{1'}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$=\frac{1}{2}\mathbf{Df}'\omega^{\frac{1}{2}}(\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\sum_j\omega_j-\frac{1}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'\mathbf{Q}'\sum_j\omega_j^{\frac{1}{2}})(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$

$$=\frac{1}{2}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j\omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}) \qquad \text{(E-26)}$$

$$\frac{1}{(\mathbf{Q}'\mathbf{Q})^2}\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*\mathbf{D}g'-\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*\mathbf{D}^2 g\tilde{\mathbf{l}}$$

$$=-(\mathbf{I}-\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'\mathbf{D}g(\mathbf{Df}'\omega\mathbf{Df})^{-1})\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*$$

$$=-\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_* \qquad \text{(E-27)}$$

Use of (E-24)-(E-27) in (E-23) yields

$$\tilde{\mathbf{q}}=(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}}\,\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}}+\frac{1}{2}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}$$

$$-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j\omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})-\frac{1}{2}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}}-\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})$$

$$-\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)$$

$$=(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}(\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}}\,\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}}-\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)$$

$$+\frac{1}{2}\sum_j\omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))-\frac{1}{2}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}}-\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})) \qquad \text{(E-28)}$$

## Third-Order Correct Constrained Sum of Squares Estimate for Confidence Intervals

First a second-order correct constrained residual is computed using (3-30), (E-5), (E-14), and (E-28) as follows.

$$
\begin{aligned}
&Y_i - f_i(\gamma\tilde{\theta}) = Y_i - f_i(\gamma\theta_*) + f_i(\gamma\theta_*) - f_i(\gamma\tilde{\theta}) \\
&\approx U_{*i} + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} - \mathbf{D}f_i(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_i\tilde{\mathbf{q}} \\
&-\frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_i \tilde{\mathbf{l}} + \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 f_i(\gamma'\gamma)^{-1}\gamma'\mathbf{e} \\
&= U_{*i} - \mathbf{D}f_i(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}\mathbf{D}\mathbf{f}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}(\mathbf{e}'\mathbf{D}_\beta^2 f_i\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_i\tilde{\mathbf{l}}) \\
&- \mathbf{D}f_i(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}(\mathbf{D}\mathbf{f}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}\mathbf{D}\mathbf{f}(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}} - \mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*) \\
&+\frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})) - \frac{1}{2}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})) \\
&= \omega_i^{-\frac{1}{2}}(\mathbf{I} - \omega^{\frac{1}{2}}\mathbf{D}\mathbf{f}(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}\mathbf{D}\mathbf{f}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}))(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})) \\
&- \mathbf{D}f_i(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}(\mathbf{D}\mathbf{f}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{D}\mathbf{f}(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}} - \mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*) \\
&- \frac{1}{2}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})) \\
&= \omega_i^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})) \\
&- \omega_i^{-\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{D}\mathbf{f}(\mathbf{D}\mathbf{f}'\omega\mathbf{D}\mathbf{f})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}} - \mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*) \\
&+ \frac{1}{2}\omega_i^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) \qquad \text{(E-29)}
\end{aligned}
$$

As for computation of (B-15), the term $\tilde{\mathbf{l}}'\mathbf{D}^2 f_i\tilde{\mathbf{q}}$ must be kept in the approximation (E-1) in order to obtain fourth-order accuracy in the approximation of the second-order Taylor series expansion of $\mathbf{f}(\gamma\tilde{\theta})$. Thus, to obtain the approximation for $S(\tilde{\theta})$, this term is subtracted from (E-29) and the result is used to get, through fourth order in $\mathbf{U}_*$, $\mathbf{e}$, and their products,

$$
\begin{aligned}
&S(\tilde{\theta}) = (\mathbf{Y} - \mathbf{f}(\gamma\tilde{\theta}))'\omega(\mathbf{Y} - \mathbf{f}(\gamma\tilde{\theta})) \\
&\approx (\omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))
\end{aligned}
$$

$$
\begin{aligned}
&-\omega^{-\frac{1}{2}}(R-\frac{QQ'}{Q'Q})\omega^{\frac{1}{2}}Df(Df'\omega Df)^{-1}(\sum_i D^2 f_i\tilde{l}\,\omega_i^{\frac{1}{2}}\tilde{Z}-D^2 g\tilde{l}\,\frac{1}{Q'Q}Q'\omega^{\frac{1}{2}}U_*)+\frac{1}{2}\omega^{-\frac{1}{2}}\frac{Q}{Q'Q}(\tilde{l}'D^2 g\tilde{l}\\
&-e'\gamma(\gamma'\gamma)^{-1}D^2 g(\gamma'\gamma)^{-1}\gamma'e)-\omega^{-\frac{1}{2}}\sum_j\omega_j^{\frac{1}{2}}\tilde{l}'D^2 f_j\tilde{q})'\omega(\omega^{-\frac{1}{2}}(I-R+\frac{QQ'}{Q'Q})(\omega^{\frac{1}{2}}U_*+\frac{1}{2}\sum_j\omega_j^{\frac{1}{2}}(e'D_\beta^2 f_j e\\
&-\tilde{l}'D^2 f_j\tilde{l}))-\omega^{-\frac{1}{2}}(R-\frac{QQ'}{Q'Q})\omega^{\frac{1}{2}}Df(Df'\omega Df)^{-1}(\sum_i D^2 f_i\tilde{l}\,\omega_i^{\frac{1}{2}}\tilde{Z}-D^2 g\tilde{l}\,\frac{1}{Q'Q}Q'\omega^{\frac{1}{2}}U_*)\\
&+\frac{1}{2}\omega^{-\frac{1}{2}}\frac{Q}{Q'Q}(\tilde{l}'D^2 g\tilde{l}-e'\gamma(\gamma'\gamma)^{-1}D^2 g(\gamma'\gamma)^{-1}\gamma'e)-\omega^{-\frac{1}{2}}\sum_j\omega_j^{\frac{1}{2}}\tilde{l}'D^2 f_j\tilde{q})\\
&\approx(\omega^{\frac{1}{2}}U_*+\frac{1}{2}\sum_j\omega_j^{\frac{1}{2}}(e'D_\beta^2 f_j e-\tilde{l}'D^2 f_j\tilde{l}))'(I-R+\frac{QQ'}{Q'Q})(\omega^{\frac{1}{2}}U_*+\frac{1}{2}\sum_j\omega_j^{\frac{1}{2}}(e'D_\beta^2 f_j e-\tilde{l}'D^2 f_j\tilde{l}))\\
&+(\omega^{\frac{1}{2}}U_*+\frac{1}{2}\sum_j\omega_j^{\frac{1}{2}}(e'D_\beta^2 f_j e-\tilde{l}'D^2 f_j\tilde{l}))'\frac{Q}{Q'Q}(\tilde{l}'D^2 g\tilde{l}-e'\gamma(\gamma'\gamma)^{-1}D^2 g(\gamma'\gamma)^{-1}\gamma'e)\\
&+\frac{1}{4}\frac{1}{Q'Q}(\tilde{l}'D^2 g\tilde{l}-e'\gamma(\gamma'\gamma)^{-1}D^2 g(\gamma'\gamma)^{-1}\gamma'e)^2+(\sum_i D^2 f_i\tilde{l}\,\omega_i^{\frac{1}{2}}\tilde{Z}-D^2 g\tilde{l}\,\frac{1}{Q'Q}Q'\omega^{\frac{1}{2}}U_*)'(Df'\omega Df)^{-1}\\
&\bullet Df'\omega^{\frac{1}{2}}(R-\frac{QQ'}{Q'Q})\omega^{\frac{1}{2}}Df(Df'\omega Df)^{-1}(\sum_i D^2 f_i\tilde{l}\,\omega_i^{\frac{1}{2}}\tilde{Z}-D^2 g\tilde{l}\,\frac{1}{Q'Q}Q'\omega^{\frac{1}{2}}U_*)\\
&-2U_*'\omega^{\frac{1}{2}}(I-R+\frac{QQ'}{Q'Q})\sum_j\omega_j^{\frac{1}{2}}\tilde{l}'D^2 f_j\tilde{q} \qquad \text{(E-30)}
\end{aligned}
$$

Evaluation of the combination of the last two terms, using (E-28) for $\tilde{q}$, results in

$$
\begin{aligned}
&(\sum_i D^2 f_i\tilde{l}\,\omega_i^{\frac{1}{2}}\tilde{Z}-D^2 g\tilde{l}\,\frac{1}{Q'Q}Q'\omega^{\frac{1}{2}}U_*)'(Df'\omega Df)^{-1}Df'\omega^{\frac{1}{2}}(R-\frac{QQ'}{Q'Q})\omega^{\frac{1}{2}}Df(Df'\omega Df)^{-1}(\sum_i D^2 f_i\tilde{l}\,\omega_i^{\frac{1}{2}}\tilde{Z}\\
&-D^2 g\tilde{l}\,\frac{1}{Q'Q}Q'\omega^{\frac{1}{2}}U_*)-2\tilde{Z}'\sum_j\omega_j^{\frac{1}{2}}\tilde{l}'D^2 f_j(Df'\omega Df)^{-1}(Df'\omega^{\frac{1}{2}}(R-\frac{QQ'}{Q'Q})(\omega^{\frac{1}{2}}Df(Df'\omega Df)^{-1}(\sum_i D^2 f_i\tilde{l}\,\omega_i^{\frac{1}{2}}\tilde{Z}\\
&-D^2 g\tilde{l}\,\frac{1}{Q'Q}Q'\omega^{\frac{1}{2}}U_*)+\frac{1}{2}\sum_j\omega_j^{\frac{1}{2}}(e'D_\beta^2 f_j e-\tilde{l}'D^2 f_j\tilde{l}))-\frac{1}{2}\frac{1}{Q'Q}Dg'(\tilde{l}'D^2 g\tilde{l}-e'\gamma(\gamma'\gamma)^{-1}D^2 g(\gamma'\gamma)^{-1}\gamma'e))\\
&=-\tilde{Z}'\sum_j\omega_j^{\frac{1}{2}}\tilde{l}'D^2 f_j(Df'\omega Df)^{-1}Df'\omega^{\frac{1}{2}}(R-\frac{QQ'}{Q'Q})\omega^{\frac{1}{2}}Df(Df'\omega Df)^{-1}\sum_i D^2 f_i\tilde{l}\,\omega_i^{\frac{1}{2}}\tilde{Z}\\
&+(\frac{1}{Q'Q}Q'\omega^{\frac{1}{2}}U_*)^2\tilde{l}'D^2 g(Df'\omega Df)^{-1}Df'\omega^{\frac{1}{2}}(R-\frac{QQ'}{Q'Q})\omega^{\frac{1}{2}}Df(Df'\omega Df)^{-1}D^2 g\tilde{l}\\
&-\tilde{Z}'\sum_j\omega_j^{\frac{1}{2}}\tilde{l}'D^2 f_j(Df'\omega Df)^{-1}Df'\omega^{\frac{1}{2}}(R-\frac{QQ'}{Q'Q})\sum_j\omega_j^{\frac{1}{2}}(e'D_\beta^2 f_j e-\tilde{l}'D^2 f_j\tilde{l})\\
&+\frac{1}{Q'Q}\tilde{Z}'\sum_j\omega_j^{\frac{1}{2}}\tilde{l}'D^2 f_j(Df'\omega Df)^{-1}Dg'(\tilde{l}'D^2 g\tilde{l}-e'\gamma(\gamma'\gamma)^{-1}D^2 g(\gamma'\gamma)^{-1}\gamma'e) \qquad \text{(E-31)}
\end{aligned}
$$

Substitution of (E-31) into (E-30) yields the final result,

$$S(\widetilde{\theta}) \approx (\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \widetilde{\mathbf{l}}'\mathbf{D}^2 f_j \widetilde{\mathbf{l}}))'(\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \widetilde{\mathbf{l}}'\mathbf{D}^2 f_j \widetilde{\mathbf{l}}))$$
$$+ (\omega^{\frac{1}{2}}\mathbf{U}_* + \frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \widetilde{\mathbf{l}}'\mathbf{D}^2 f_j \widetilde{\mathbf{l}}))'\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\widetilde{\mathbf{l}}'\mathbf{D}^2 g \widetilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})$$
$$+ \frac{1}{4}\frac{1}{\mathbf{Q}'\mathbf{Q}}(\widetilde{\mathbf{l}}'\mathbf{D}^2 g \widetilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})^2$$
$$- \widetilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \widetilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}}$$
$$+ (\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)^2\widetilde{\mathbf{l}}'\mathbf{D}^2 g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g \widetilde{\mathbf{l}}$$
$$- \widetilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \widetilde{\mathbf{l}}'\mathbf{D}^2 f_j \widetilde{\mathbf{l}})$$
$$+ \frac{1}{\mathbf{Q}'\mathbf{Q}}\widetilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'(\widetilde{\mathbf{l}}'\mathbf{D}^2 g \widetilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) \qquad \text{(E-32)}$$

## Second-Order Correct Constrained Parameter Estimates for Prediction Intervals

Again, a Lagrange multiplier formulation is used to obtain the required constrained regression estimates, this time $\widetilde{\theta}$ and $\widetilde{\theta}_p$, of $\theta_*$ and $\theta_p^*$. As will be shown, the solution can be put into the same form as the solution for $\widetilde{\theta}$ for confidence intervals. Similarly, the required functions of $\widetilde{\theta}$ and $\widetilde{\theta}_p$ have the same forms as the functions of $\widetilde{\theta}$ for confidence intervals.

The constrained regression estimates $\widetilde{\theta} = \overline{\theta} + \widetilde{\mathbf{l}} + \widetilde{\mathbf{q}}$ and $\widetilde{\theta}_p = \overline{\theta}_p + \widetilde{l}_p$ are obtained by minimizing $S_a(\theta, \theta_p)$ subject to the constraint that $g(\gamma\theta) + \upsilon = Y_p$, where $\upsilon = Y_p - \theta_p$. This can be formulated as before as the Lagrange multiplier problem

$$L(\theta, \theta_p, \lambda) = S_a(\theta, \theta_p) + 2\lambda(Y_p - g(\gamma\theta) - \upsilon)$$
$$= (\mathbf{Y} - \mathbf{f}(\gamma\theta))'\mathbf{W}(\mathbf{Y} - \mathbf{f}(\gamma\theta)) + 2(\mathbf{Y} - \mathbf{f}(\gamma\theta))'\mathbf{W}_p(Y_p - \theta_p) + W_p(Y_p - \theta_p)^2$$
$$+ 2\lambda(\theta_p - g(\gamma\theta)) \qquad \text{(E-33)}$$

To solve the problem (E-33) is minimized using (3-30), (4-9), (5-63)-(5-65), (5-73), (C-16), (E-5), and (E-6), keeping terms up through second order in $\mathbf{U}_*$, $U_p^*$, $\mathbf{e}$, $\theta_p^* - \overline{\theta}_p$, or their product. Definitions $\widetilde{l}_p = \widetilde{\theta}_p - \overline{\theta}_p$ and $e_p = \theta_p^* - \overline{\theta}_p$ allow the result with respect to $\theta$ to be expressed as

$$\frac{\partial L}{\partial \theta} = -2\sum_i\sum_j \mathbf{D}\widetilde{f}_i' W_{ij}(Y_j - f_j(\gamma\widetilde{\theta})) - 2\sum_i \mathbf{D}\widetilde{f}_i' W_{pi}(Y_p - \widetilde{\theta}_p) - 2\lambda\,\mathbf{D}\widetilde{g}'$$
$$\approx -2\sum_i\sum_j(\mathbf{D}f_i' + \mathbf{D}^2 f_i \widetilde{\mathbf{l}})W_{ij}(Y_j - f_j(\gamma\theta_*) - \mathbf{D}f_j(\widetilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\widetilde{\mathbf{q}}$$
$$- \frac{1}{2}\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j \widetilde{\mathbf{l}} + \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 f_j(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - 2\sum_i(\mathbf{D}f_i' + \mathbf{D}^2 f_i \widetilde{\mathbf{l}})W_{pi}(\upsilon_* - (\widetilde{\theta}_p - \overline{\theta}_p) + (\theta_p^* - \overline{\theta}_p))$$

$$-2\lambda(\mathbf{D}g' - \mathbf{D}^2 g\tilde{\mathbf{l}})$$

$$\approx -2\sum_i\sum_j \mathbf{D}f_i' W_{ij}(U_{*j} + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\tilde{\mathbf{q}} - \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$
$$-2\sum_i\sum_j \mathbf{D}^2 f_i\tilde{\mathbf{l}} W_{ij}(U_{*j} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e})) - 2\sum_i \mathbf{D}f_i' W_{pi}(U_p^* + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 g$$
$$-\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma')\mathbf{e} - \tilde{l}_p + e_p) - 2\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}} W_{pi}(U_p^* - \tilde{l}_p + e_p) - 2\lambda(\mathbf{D}g' - \mathbf{D}^2 g\tilde{\mathbf{l}})$$
$$= \mathbf{0}$$

or

$$\sum_i\sum_j \mathbf{D}f_i' W_{ij}(U_{*j} + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\tilde{\mathbf{q}} - \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$
$$+\sum_i\sum_j \mathbf{D}^2 f_i\tilde{\mathbf{l}} W_{ij}(U_{*j} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e})) + \sum_i \mathbf{D}f_i' W_{pi}(U_p^* + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 g$$
$$-\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma')\mathbf{e} - \tilde{l}_p + e_p) + \sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}} W_{pi}(U_p^* - \tilde{l}_p + e_p) + \lambda(\mathbf{D}g' - \mathbf{D}^2 g\tilde{\mathbf{l}})$$
$$= \mathbf{0} \tag{E-34}$$

The result with respect is $\theta_p$ is

$$\frac{\partial L}{\partial \theta_p} = -2\sum_j (Y_j - f_j(\gamma\tilde{\theta}))W_{pj} - 2W_p(Y_p - \tilde{\theta}_p) + 2\lambda$$
$$\approx -2\sum_j (U_{*j} + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\tilde{\mathbf{q}} - \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})W_{pj}$$
$$-2W_p(U_p^* + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 g - \gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma')\mathbf{e} - \tilde{l}_p + e_p) + 2\lambda$$
$$= 0$$

or

$$\sum_j (U_{*j} + \frac{1}{2}\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{D}f_j(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}f_j\tilde{\mathbf{q}} - \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})W_{pj}$$
$$+W_p(U_p^* + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 g - \gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma')\mathbf{e} - \tilde{l}_p + e_p) - \lambda$$
$$= 0 \tag{E-35}$$

Finally, the result with respect to $\lambda$ is

$$\frac{\partial L}{\partial \lambda} = 2(\tilde{\theta}_p - g(\gamma\tilde{\theta})) = 2(\tilde{\theta}_p - \theta_p^* + \theta_p^* - g(\gamma\theta_*) + g(\gamma\theta_*) - g(\gamma\tilde{\theta}))$$
$$\approx 2(\tilde{l}_p - e_p - \mathbf{D}g(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \mathbf{D}g\tilde{\mathbf{q}} - \frac{1}{2}\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} + \frac{1}{2}\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})$$

$= 0$

where $\theta_p^* - g(\gamma\boldsymbol{\theta}_*) = 0$, or

$$\mathbf{D}g(\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}) + \mathbf{D}g\tilde{\mathbf{q}} + \frac{1}{2}(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) - \tilde{l}_p + e_p = 0 \tag{E-36}$$

Equations (E-34), (E-35), and (E-36) can be written as two matrix equations having the form of (E-8) and (E-9). First, $\mathbf{D}_a$ and $\mathbf{D}_a^2$ are defined as derivative operators augmented to include $\theta_p$ in addition to $\boldsymbol{\theta}$ (that is, operators with respect to $\boldsymbol{\theta}_a$). Next, the following augmented matrices and vectors are defined:

$$\mathbf{Y}_a = \begin{bmatrix} \mathbf{Y} \\ Y_p \end{bmatrix} \tag{E-37}$$

$$\mathbf{f}_a(\gamma\boldsymbol{\theta}, \theta_p) = \begin{bmatrix} \mathbf{f}(\gamma\boldsymbol{\theta}) \\ \theta_p \end{bmatrix} \tag{E-38}$$

$$\boldsymbol{\theta}_a = \begin{bmatrix} \boldsymbol{\theta} \\ \theta_p \end{bmatrix} \tag{E-39}$$

$$h(\gamma\boldsymbol{\theta}, \theta_p) = g(\gamma\boldsymbol{\theta}) + \upsilon = g(\gamma\boldsymbol{\theta}) + Y_p - \theta_p \tag{E-40}$$

$$\tilde{\mathbf{l}}_a = \begin{bmatrix} \tilde{\mathbf{l}} \\ \tilde{l}_p \end{bmatrix} \tag{E-41}$$

$$\tilde{\mathbf{l}}_{*a} = \begin{bmatrix} \tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e} \\ \tilde{l}_p - e_p \end{bmatrix} \tag{E-42}$$

$$\tilde{\mathbf{q}}_a = \begin{bmatrix} \tilde{\mathbf{q}} \\ 0 \end{bmatrix} \tag{E-43}$$

$$\mathbf{U}_{*a} = \begin{bmatrix} \mathbf{U}_* \\ U_p^* \end{bmatrix} \tag{E-44}$$

$$\mathbf{W}_a = \begin{bmatrix} \mathbf{W} & \mathbf{W}_p \\ \mathbf{W}_p' & W_p \end{bmatrix} \tag{E-45}$$

$$\mathbf{D}_a f_{aj} = [\mathbf{D}f_j \quad 0] \text{ for } f_{aj} = f_j \qquad \mathbf{D}_a f_{aj} = [\mathbf{0} \quad 1] \text{ for } f_{aj} = \theta_p \tag{E-46}$$

$$\mathbf{D}_a^2 f_{aj} = \begin{bmatrix} \mathbf{D}^2 f_j & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} \text{ for } f_{aj} = f_j \qquad \mathbf{D}_a^2 f_{aj} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} \text{ for } f_{aj} = \theta_p \tag{E-47}$$

$$\mathbf{D}_a h' = \begin{bmatrix} \mathbf{D}g' \\ -1 \end{bmatrix} \tag{E-48}$$

$$\mathbf{D}_a^2 h = \begin{bmatrix} \mathbf{D}^2 g & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} \tag{E-49}$$

$$\mathbf{G}_a = [\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1} \quad \mathbf{0}] \tag{E-50}$$

$$E_{aj} = \mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e}\,; j = 1,2,\ldots,n \qquad E_{aj} = \mathbf{e}'(\mathbf{D}_\beta^2 g - \boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\mathbf{D}^2 g(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{e}\,; j = n+1 \tag{E-51}$$

Finally, use of (E-37)-(E-51), transforms (E-34)-(E-36) to the two equations

$$\sum_i \sum_j \mathbf{D}_a f'_{ai} W_{aij}(U_{*aj} + \frac{1}{2}E_{aj} - \mathbf{D}_a f_{aj}\tilde{\mathbf{l}}_{*a} - \mathbf{D}_a f_{aj}\tilde{\mathbf{q}}_a - \frac{1}{2}\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 f_{aj}\tilde{\mathbf{l}}_a)$$
$$+ \sum_i \sum_j \mathbf{D}_a^2 f_{ai}\tilde{\mathbf{l}}_a W_{aij}(U_{*aj} - \mathbf{D}_a f_{aj}\tilde{\mathbf{l}}_{*a}) + \lambda(\mathbf{D}_a h' + \mathbf{D}_a^2 h\tilde{\mathbf{l}}_a) = \mathbf{0} \tag{E-52}$$

where $i, j = 1,2,\ldots,n+1$, and

$$\mathbf{D}_a h\tilde{\mathbf{l}}_{*a} + \mathbf{D}_a h\tilde{\mathbf{q}}_a + \frac{1}{2}\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 h\tilde{\mathbf{l}}_a - \frac{1}{2}\mathbf{e}'\mathbf{G}_a\mathbf{D}_a^2 h\mathbf{G}'_a\mathbf{e} = 0 \tag{E-53}$$

which are of the form of (E-8) and (E-9), respectively.

Because (E-52) and (E-53) are of the form of (E-8) and (E-9), the solutions for $\tilde{\mathbf{l}}_{*a}$ and $\mathbf{q}_a$ are of the forms of (E-14) and (E-28), or

$$\tilde{\mathbf{l}}_{*a} = (\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a} \tag{E-54}$$

and

$$\tilde{\mathbf{q}}_a = (\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})(\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}(\sum_i \mathbf{D}_a^2 f_{ai}\tilde{\mathbf{l}}_a\mathbf{W}_{ai}^{\frac{1}{2}}\tilde{\mathbf{Z}}_a$$
$$- \mathbf{D}_a^2 h\tilde{\mathbf{l}}_a\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{Q}'_a\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a}) + \frac{1}{2}\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj} - \tilde{\mathbf{l}}'_a\mathbf{D}_a^2 f_{aj}\tilde{\mathbf{l}}_a)) - \frac{1}{2}\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{D}_a h'(\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 g_a\tilde{\mathbf{l}}_a$$
$$- \mathbf{e}'\mathbf{G}_a\mathbf{D}_a^2 h\mathbf{G}'_a\mathbf{e})) \tag{E-55}$$

where

$$\mathbf{D}_a\mathbf{f}_a = \begin{bmatrix} \mathbf{Df} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \tag{E-56}$$

$$\mathbf{R}_a = \mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a^{\frac{1}{2}} \tag{E-57}$$

$$\mathbf{Q}_a = \mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h' \tag{E-58}$$

$$\widetilde{\mathbf{Z}}_a = (\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a} \tag{E-59}$$

and $\mathbf{W}_{ai}^{1/2}$ and $\mathbf{W}_{aj}^{1/2}$ signify row $i$ and column $j$, respectively, of $\mathbf{W}_a^{1/2}$. Matrix $\mathbf{I}_a$ is the identity matrix of order $n+1$.

## Third-Order Correct Constrained Sum of Squares Estimate for Prediction Intervals

In conformance with (E-37) and (E-38) an augmented constrained residual can be defined as

$$\begin{bmatrix} \mathbf{Y}-\mathbf{f}(\gamma\widetilde{\theta}) \\ Y_p - \widetilde{\theta}_p \end{bmatrix} = \begin{bmatrix} \mathbf{Y}-\mathbf{f}(\gamma\theta_*) \\ Y_p - \theta_p^* \end{bmatrix} + \begin{bmatrix} \mathbf{f}(\gamma\theta_*)-\mathbf{f}(\gamma\widetilde{\theta}) \\ \theta_p^* - \widetilde{\theta}_p \end{bmatrix} \tag{E-60}$$

or

$$\mathbf{Y}_a - \mathbf{f}_a(\gamma\widetilde{\theta},\widetilde{\theta}_p) = \mathbf{Y}_a - \mathbf{f}_a(\gamma\theta_*,\theta_p^*) + \mathbf{f}_a(\gamma\theta_*,\theta_p^*) - \mathbf{f}_a(\gamma\widetilde{\theta},\widetilde{\theta}_p) \tag{E-61}$$

Because the forms of (E-14), (E-28), and $\mathbf{Y}-\mathbf{f}(\gamma\widetilde{\theta})$ are the same as the forms of (E-54), (E-55), and $\mathbf{Y}_a - \mathbf{f}_a(\gamma\widetilde{\theta},\widetilde{\theta}_p)$, respectively, the augmented sum of squares estimate can be written in the form of (E-32), or

$$\begin{aligned} S_a(\widetilde{\theta},\widetilde{\theta}_p) &= (\mathbf{Y}_a - \mathbf{f}_a(\gamma\widetilde{\theta},\widetilde{\theta}_p))'\mathbf{W}_a(\mathbf{Y}_a - \mathbf{f}_a(\gamma\widetilde{\theta},\widetilde{\theta}_p)) \\ &\approx (\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a} + \frac{1}{2}\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj} - \widetilde{\mathbf{l}}_a'\mathbf{D}_a^2 f_{aj}\widetilde{\mathbf{l}}_a))'(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})(\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a} + \frac{1}{2}\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj} \\ &- \widetilde{\mathbf{l}}_a'\mathbf{D}_a^2 f_{aj}\widetilde{\mathbf{l}}_a)) + (\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a} + \frac{1}{2}\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj} - \widetilde{\mathbf{l}}_a'\mathbf{D}_a^2 f_{aj}\widetilde{\mathbf{l}}_a))'\frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(\widetilde{\mathbf{l}}_a'\mathbf{D}_a^2 h\widetilde{\mathbf{l}}_a - \mathbf{e}'\mathbf{G}_a\mathbf{D}_a^2 h\mathbf{G}_a'\mathbf{e}) \\ &+ \frac{1}{4}\frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}(\widetilde{\mathbf{l}}_a'\mathbf{D}_a^2 h\widetilde{\mathbf{l}}_a - \mathbf{e}'\mathbf{G}_a\mathbf{D}_a^2 h\mathbf{G}_a'\mathbf{e})^2 \end{aligned}$$

$$
\begin{aligned}
&-\widetilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}} \widetilde{\mathbf{I}}'_a \mathbf{D}_a^2 f_{aj} (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a^{\frac{1}{2}} (\mathbf{R}_a - \frac{\mathbf{Q}_a \mathbf{Q}'_a}{\mathbf{Q}'_a \mathbf{Q}_a}) \mathbf{W}_a^{\frac{1}{2}} \mathbf{D}_a \mathbf{f}_a (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \sum_i \mathbf{D}_a^2 f_{ai} \widetilde{\mathbf{I}}_a \mathbf{W}_{ai}^{\frac{1}{2}} \widetilde{\mathbf{Z}}_a \\
&+ (\frac{1}{\mathbf{Q}'_a \mathbf{Q}_a} \mathbf{Q}'_a \mathbf{W}_a^{\frac{1}{2}} \mathbf{U}_{*a})^2 \widetilde{\mathbf{I}}'_a \mathbf{D}_a^2 h (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a^{\frac{1}{2}} (\mathbf{R}_a - \frac{\mathbf{Q}_a \mathbf{Q}'_a}{\mathbf{Q}'_a \mathbf{Q}_a}) \mathbf{W}_a^{\frac{1}{2}} \mathbf{D}_a \mathbf{f}_a (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a^2 h \widetilde{\mathbf{I}}_a \\
&-\widetilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}} \widetilde{\mathbf{I}}'_a \mathbf{D}_a^2 f_{aj} (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a^{\frac{1}{2}} (\mathbf{R}_a - \frac{\mathbf{Q}_a \mathbf{Q}'_a}{\mathbf{Q}'_a \mathbf{Q}_a}) \sum_j \mathbf{W}_{aj}^{\frac{1}{2}} (E_{aj} - \widetilde{\mathbf{I}}'_a \mathbf{D}_a^2 f_{aj} \widetilde{\mathbf{I}}_a) \\
&+ \frac{1}{\mathbf{Q}'_a \mathbf{Q}_a} \widetilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}} \widetilde{\mathbf{I}}'_a \mathbf{D}_a^2 f_{aj} (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a h' (\widetilde{\mathbf{I}}'_a \mathbf{D}_a^2 h \widetilde{\mathbf{I}}_a - \mathbf{e}' \mathbf{G}_a \mathbf{D}_a^2 h \, \mathbf{G}'_a \mathbf{e}) \qquad \text{(E-62)}
\end{aligned}
$$

# Appendix F – Derivation of Statistical Distributions Used to Define Confidence Regions, Confidence Intervals, and Prediction Intervals

## Distributions for Confidence Regions and Confidence Intervals

***Forms of sum of squares functions for perturbation analysis.*** Procedures generalized from Johansen (1983, p. 183-184) are used here to derive an approximate probability density function (pdf) for a ratio involving the sum of squared errors objective function $S(\theta)$. Let the ratio be a scalar multiple of $(S(\theta)-S(\hat{\theta}))/S(\hat{\theta})$, where $\theta$ is either the spatial average set $\theta_*$ or the set $\tilde{\theta}$ produced by the constrained regression. The functions, $S(\hat{\theta})$ and $S(\tilde{\theta})$ are given by (B-17) and (E-32), respectively. Equation (B-17) is repeated here in expanded form as

$$
\begin{aligned}
S(\hat{\theta}) \approx\ & \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}) \\
& +\frac{1}{4}\sum_i(\mathbf{e}'\mathbf{D}_\beta^2 f_i\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_i\mathbf{l})\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}) \\
& -\sum_i \omega_i^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z} \\
& -\sum_i \omega_i^{\frac{1}{2}}\mathbf{Z}\mathbf{l}'\mathbf{D}^2 f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\sum_j \omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})
\end{aligned}
\qquad \text{(F-1)}
$$

where the identity $(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{Df}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}=\mathbf{0}$ was used to allow the first two terms to be written in terms of $\mathbf{U}_*$ rather than $\mathbf{U}$. Equation (E-32) is given in expanded form as

$$
\begin{aligned}
S(\tilde{\theta}) \approx\ & \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}) \\
& +\frac{1}{4}\sum_i(\mathbf{e}'\mathbf{D}_\beta^2 f_i\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_i\tilde{\mathbf{l}})\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}) \\
& +(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*+\frac{1}{2}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}}))'\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}}-\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\mathbf{D}^2 g(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}) \\
& +\frac{1}{4}\frac{1}{\mathbf{Q}'\mathbf{Q}}(\tilde{\mathbf{l}}'\mathbf{D}^2 g\tilde{\mathbf{l}}-\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\mathbf{D}^2 g(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e})^2 \\
& -\tilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\tilde{\mathbf{l}}\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}} \\
& +(\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*)^2\tilde{\mathbf{l}}'\mathbf{D}^2 g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}} \\
& -\tilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}'\mathbf{D}^2 f_j(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})
\end{aligned}
$$

$$+\frac{1}{\mathbf{Q'Q}}\widetilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j(\mathbf{Df'\omega Df})^{-1}\mathbf{D}g'(\widetilde{\mathbf{l}}'\mathbf{D}^2 g\widetilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}) \tag{F-2}$$

The function $S(\theta_*)$ can be approximated using (3-30) and (4-9) as

$$S(\theta_*) = (\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'\omega(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) = \sum_i\sum_j (Y_i - f_i(\gamma\theta_*))\omega_{ij}(Y_j - f_j(\gamma\theta_*))$$

$$= \sum_i\sum_j (\varepsilon_i + \mathbf{D}_\beta f_i(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_i\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})\omega_{ij}(\varepsilon_j$$

$$+ \mathbf{D}_\beta f_j(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} + \frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})$$

$$= \mathbf{U}_*'\omega\mathbf{U}_* + \mathbf{U}_*'\omega^{\frac{1}{2}}\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'(\mathbf{D}_\beta^2 f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})$$

$$+\frac{1}{4}(\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'(\mathbf{D}_\beta^2 f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}))'\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'(\mathbf{D}_\beta^2 f_j$$

$$- \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}) \tag{F-3}$$

Equations (F-1)-(F-3) need to be put into forms that will allow perturbation analysis. This is accomplished as follows. First, errors $\mathbf{U}$ and $\mathbf{U}_*$ are assumed to be much larger in magnitude than errors $\mathbf{D}_\beta\mathbf{fe}$, all of which are small. This permits dropping terms of higher than third order in **e**, higher than fourth order in **U** or $\mathbf{U}_*$, and higher than a total of third order when products of **e** and **U** or $\mathbf{U}_*$ are involved. From (4-9)

$$\mathbf{U} = \mathbf{U}_* + \mathbf{Df}(\gamma'\gamma)^{-1}\gamma'\mathbf{e} \tag{F-4}$$

so that, from (4-5),

$$\mathbf{l} = (\mathbf{Df'\omega Df})^{-1}\mathbf{Df'\omega U}_* + (\gamma'\gamma)^{-1}\gamma'\mathbf{e}$$
$$= \mathbf{l}_* + (\gamma'\gamma)^{-1}\gamma'\mathbf{e} \tag{F-5}$$

where $\mathbf{l}_* = (\mathbf{Df'\omega Df})^{-1}\mathbf{Df'\omega U}_*$. Next (F-5) is used to substitute for $\mathbf{l}$ in (F-1), and all fourth-order terms involving **e** and products of **e** and $\mathbf{U}_*$ are dropped to give

$$S(\hat{\theta}) \approx \mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}\mathbf{U}_* + \mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})$$

$$+\frac{1}{4}\sum_i \mathbf{l}_*'\mathbf{D}^2 f_i\mathbf{l}_*\omega_i^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\sum_j \omega_j^{\frac{1}{2}}\mathbf{l}_*'\mathbf{D}^2 f_j\mathbf{l}_* - \sum_i \omega_i^{\frac{1}{2}}\mathbf{Z}_*\mathbf{l}_*'\mathbf{D}^2 f_i(\mathbf{Df'\omega Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}_*\omega_i^{\frac{1}{2}}\mathbf{Z}_*$$

$$+\sum_i \omega_i^{\frac{1}{2}}\mathbf{Z}_*\mathbf{l}_*'\mathbf{D}^2 f_i(\mathbf{Df'\omega Df})^{-1}\mathbf{Df}'\sum_j \omega_j\mathbf{l}_*'\mathbf{D}^2 f_j\mathbf{l}_* \tag{F-6}$$

where $\mathbf{Z}_* = (\mathbf{I} - \mathbf{R})\omega^{1/2}\mathbf{U}_*$. Similarly from (E-14)

$$\widetilde{\mathbf{l}} = (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{U}_* + (\gamma'\gamma)^{-1}\gamma'\mathbf{e}$$
$$= \widetilde{\mathbf{l}}_* + (\gamma'\gamma)^{-1}\gamma'\mathbf{e} \tag{F-7}$$

where $\widetilde{\mathbf{l}}_* = (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{1/2}(\mathbf{R} - \mathbf{QQ}'/\mathbf{Q}'\mathbf{Q})\omega^{1/2}\mathbf{U}_*$. Use of (F-7) to substitute for $\widetilde{\mathbf{l}}$ in (F-2) in which fourth order terms in **e**, and in products of **e** and $\mathbf{U}_*$, are dropped produces

$$S(\widetilde{\theta}) \approx \mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{U}_* + \mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_\beta^2 f_j \mathbf{e} - \widetilde{\mathbf{l}}'\mathbf{D}^2 f_j \widetilde{\mathbf{l}})$$
$$+\frac{1}{4}\sum_i \widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \widetilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \widetilde{\mathbf{l}}_* + \mathbf{U}_*'\omega^{\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\widetilde{\mathbf{l}}'\mathbf{D}^2 g\widetilde{\mathbf{l}} - \mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^2 g(\gamma'\gamma)^{-1}\gamma'\mathbf{e})$$
$$-\frac{1}{2}\sum_i \widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \widetilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g\widetilde{\mathbf{l}}_* + \frac{1}{4}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g\widetilde{\mathbf{l}}_*)^2$$
$$-\widetilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \widetilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}}$$
$$+(\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)^2\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\widetilde{\mathbf{l}}_*$$
$$+\widetilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \widetilde{\mathbf{l}}_*$$
$$+\frac{1}{\mathbf{Q}'\mathbf{Q}}\widetilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g\widetilde{\mathbf{l}}_* \tag{F-8}$$

***Approximate characteristic function for sum of squares ratio.*** The statistical distributions are derived using characteristic functions, which are Fourier transforms of pdf's (Papoulis, 1965, p. 153). They are used to simplify the derivations, and readers not familiar with their use should read Papoulis (1965, p. 153-162, 213-214, 244-245) for an excellent discussion.

The joint characteristic function for the distribution of the ratio $(S(\theta) - S(\hat{\theta}))/S(\hat{\theta})$ is

$$\psi(s,t) = E(\exp\{is(S(\theta) - S(\hat{\theta})) + itS(\hat{\theta})\}) \tag{F-9}$$

where $i = \sqrt{-1}$ and $s$ and $t$ are Fourier transform variables analogous to $\omega_1$ and $\omega_2$ of Papoulis (1965, p. 213). Equation (F-9) can be expanded for evaluation by writing $S(\theta) - S(\hat{\theta})$ and $S(\hat{\theta})$ in the form of chi-squared ($\chi^2$) distributed variables plus deviations, or

$$S(\theta) - S(\hat{\theta})) = Q_1(\mathbf{U}_*) + (S(\theta) - S(\hat{\theta}) - Q_1(\mathbf{U}_*))$$
$$= Q_1(\mathbf{U}_*) + D_1 \tag{F-10}$$

and

$$S(\hat{\theta}) = Q_2(\mathbf{U}_*) + (S(\hat{\theta}) - Q_2(\mathbf{U}_*))$$

$$= Q_2(\mathbf{U}_*) + D_2 \tag{F-11}$$

where

$$Q_1(\mathbf{U}_*) = \mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{M}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_* \tag{F-12}$$

$$Q_2(\mathbf{U}_*) = \mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}(\mathbf{I}-\mathbf{H})\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_* \tag{F-13}$$

$$\left.\begin{aligned} &\mathbf{M} = \mathbf{H} \text{ when } \theta = \theta_* \\ &\mathbf{M} = \frac{\mathbf{PP}'}{\mathbf{P}'\mathbf{P}} \text{ when } \theta = \widetilde{\theta} \end{aligned}\right\} \tag{F-14}$$

$$\mathbf{H} = \mathbf{V}_*^{-\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\mathbf{V}_*^{-1}\mathbf{Df})^{-1}\mathbf{Df}'\mathbf{V}_*^{-\frac{1}{2}} \tag{F-15}$$

$$\mathbf{P} = \mathbf{V}_*^{-\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\mathbf{V}_*^{-1}\mathbf{Df})^{-1}\mathbf{D}g' \tag{F-16}$$

$$D_1 = S(\theta) - S(\hat{\theta}) - Q_1(\mathbf{U}_*) \tag{F-17}$$

$$D_2 = S(\hat{\theta}) - Q_2(\mathbf{U}_*) \tag{F-18}$$

Note using (4-10) that $\mathbf{V}_*^{-1/2}\mathbf{U}_* \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$. Then, because $\mathbf{I}-\mathbf{H}$ is symmetric and idempotent (Cooley and Naff, 1990, p. 165) with a rank of $n-p$, $Q_2(\mathbf{U}_*)/\sigma_\varepsilon^2$ has a $\chi^2$ distribution with $n-p$ degrees of freedom (Theorem 4.4.1, Graybill, 1976, p. 134), or

$$Q_2(\mathbf{U}_*)/\sigma_\varepsilon^2 \sim \chi^2(n-p) \tag{F-19}$$

Similarly, because $\mathbf{M}$ is symmetric and idempotent with a rank of $p_1 = p$ when $\mathbf{M} = \mathbf{H}$ and with a rank of $p_1 = 1$ when $\mathbf{M} = \mathbf{PP}'/\mathbf{P}'\mathbf{P}$,

$$Q_1(\mathbf{U}_*)/\sigma_\varepsilon^2 \sim \chi^2(p_1) \tag{F-20}$$

Use of (F-10) and (F-11) in (F-9) allows an approximate characteristic function to be expressed as a product of the joint characteristic function for $Q_1(\mathbf{U}_*)$ and $Q_2(\mathbf{U}_*)$ and a correction factor. To start the evaluation, $\psi(s,t)$ is expanded and approximated to get

$$\begin{aligned} \psi(s,t) &= E(\exp\{is(Q_1(\mathbf{U}_*)+D_1)+it(Q_2(\mathbf{U}_*)+D_2)\}) \\ &= E(\exp\{is(Q_1(\mathbf{U}_*)+it(Q_2(\mathbf{U}_*)\}\exp\{isD_1+itD_2\}) \\ &\approx E(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}(1+isD_1+itD_2)) \end{aligned} \tag{F-21}$$

***Evaluation of approximate characteristic function.*** To evaluate (F-21), $\mathbf{V}_*^{-1/2}\mathbf{U}_*$ needs to be written in terms of statistically independent variables that also appear in $Q_1(\mathbf{U}_*)$ and $Q_2(\mathbf{U}_*)$. This is accomplished by writing $\mathbf{V}_*^{-1/2}\mathbf{U}_*$ as

$$\begin{aligned}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_* &= \mathbf{M}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_* + (\mathbf{H}-\mathbf{M})\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_* + (\mathbf{I}-\mathbf{H})\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_* \\ &= \mathbf{W}+\mathbf{T}+\mathbf{Z}\end{aligned} \tag{F-22}$$

where $\mathbf{W} = \mathbf{M}\mathbf{V}_*^{-1/2}\mathbf{U}_*$, $\mathbf{T} = (\mathbf{H}-\mathbf{M})\mathbf{V}_*^{-1/2}\mathbf{U}_*$, and, in the following development only, $\mathbf{Z} = (\mathbf{I}-\mathbf{H})\mathbf{V}_*^{-1/2}\mathbf{U}_*$. It can be verified readily that $\mathbf{HM} = \mathbf{MH} = \mathbf{M}$, from which it can be seen that $Cov(\mathbf{W},\mathbf{T}) = Cov(\mathbf{W},\mathbf{Z}) = Cov(\mathbf{T},\mathbf{Z}) = \mathbf{0}$. Hence, **W**, **T**, and **Z** are uncorrelated, so that, because they are also normally distributed, they are statistically independent (Theorem 3.5.1, Graybill, 1976, p. 105). Functions $Q_1(\mathbf{U}_*)$ and $Q_2(\mathbf{U}_*)$ can be expressed in terms of **W** and **Z** as

$$Q_1(\mathbf{U}_*) = \mathbf{W}'\mathbf{W} = |\mathbf{W}|^2 \tag{F-23}$$

$$Q_2(\mathbf{U}_*) = \mathbf{Z}'\mathbf{Z} = |\mathbf{Z}|^2 \tag{F-24}$$

Use of (F-6) in (F-18) shows that $D_2$ can be written as a sum of second- and fourth-order polynomial functions of $\mathbf{V}_*^{-1/2}\mathbf{U}_*$ and a third-order polynomial function of **e**, $\mathbf{U}_*$, and **U**. Also, from (F-22), the sum of second- and fourth-order polynomial functions of $\mathbf{V}_*^{-1/2}\mathbf{U}_*$ can be written as a sum of second- and fourth-order polynomial functions of **T**, **W**, and **Z**. Thus, the *k*th term in this sum can be expressed in the factored form

$$C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z}) = |\mathbf{T}|^{2\xi}|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon}C_{2k}(\mathbf{T}/|\mathbf{T}|,\mathbf{W}/|\mathbf{W}|,\mathbf{Z}/|\mathbf{Z}|) \tag{F-25}$$

where $2\xi$, $2\mu$, and $2\upsilon$ are powers to be determined, and $\xi$, $\mu$, and $\upsilon$ are integers. With use of (F-25) the *k*th term in (F-21) is

$$\begin{aligned}&E(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}itC_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) \\ &= itE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}|\mathbf{T}|^{2\xi}|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon}C_{2k}(\mathbf{T}/|\mathbf{T}|,\mathbf{W}/|\mathbf{W}|,\mathbf{Z}/|\mathbf{Z}|)) \\ &= itE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}|\mathbf{T}|^{2\xi}|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon})E(C_{2k}(\mathbf{T}/|\mathbf{T}|,\mathbf{W}/|\mathbf{W}|,\mathbf{Z}/|\mathbf{Z}|))\end{aligned} \tag{F-26}$$

where the fact was used that $|\mathbf{T}|$, $|\mathbf{W}|$, $|\mathbf{Z}|$, $\mathbf{T}/|\mathbf{T}|$, $\mathbf{W}/|\mathbf{W}|$, and $\mathbf{Z}/|\mathbf{Z}|$ are all mutually independent (Johansen, 1983, p. 183). Now, using (F-25),

$$E(C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) = E(|\mathbf{T}|^{2\xi}|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon})E(C_{2k}(\mathbf{T}/|\mathbf{T}|,\mathbf{W}/|\mathbf{W}|,\mathbf{Z}/|\mathbf{Z}|)) \tag{F-27}$$

Solution of (F-27) for $E(C_{2k}(\mathbf{T}/|\mathbf{T}|,\mathbf{W}/|\mathbf{W}|,\mathbf{Z}/|\mathbf{Z}|))$ and substitution of the result into (F-26) yields

$$itE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z}))$$
$$= itE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}|\mathbf{T}|^{2\xi}|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon})E(C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z}))/E(|\mathbf{T}|^{2\xi}|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon}) \qquad \text{(F-28)}$$

A similar development of each term $C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})$ in $D_1$ in (F-21) can be used to get

$$isE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z}))$$
$$= isE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}|\mathbf{T}|^{2\xi}|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon})E(C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z}))/E(|\mathbf{T}|^{2\xi}|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon}) \qquad \text{(F-29)}$$

The third-order term in $D_2$ can be expressed as $C_2(\mathbf{e},\boldsymbol{\varepsilon})$ because $\mathbf{U}_*$ and $\mathbf{U}$ are both linear functions of $\mathbf{e}$ and $\boldsymbol{\varepsilon}$. Then use of this term in (F-21) and expansion of the exponential in a Taylor series produces

$$itE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}C_2(\mathbf{e},\boldsymbol{\varepsilon}))$$
$$= itE(1+isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)+\frac{1}{2}(isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*))^2+\frac{1}{6}(isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*))^3$$
$$+\cdots)C_2(\mathbf{e},\boldsymbol{\varepsilon})) \qquad \text{(F-30)}$$

Each power of $isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)$ yields an overall even function of $\mathbf{e}$ and $\boldsymbol{\varepsilon}$ because each term of the series is a sum of terms involving forms $e_i^a\varepsilon_j^b$, where $a+b$ is even. Each term in $C_2(\mathbf{e},\boldsymbol{\varepsilon})$ involves the form $e_i^c\varepsilon_j^d$, where $c+d$=3. Therefore, the sums of powers for the products of terms in the series and $C_2(\mathbf{e},\boldsymbol{\varepsilon})$ is 0+3=3, 2+3=5, 4+3=7, 6+3=9, $\cdots$, which are all odd. This implies that either $e_i$ or $\varepsilon_i$ in each product of terms always has an odd power, so that its expected value is zero. Hence, the value of (F-30) is zero. An analogous analysis applies for the third-order terms in $D_1$ so that the third-order terms do not contribute to the final expression.

Characteristic functions for $Q_1(\mathbf{U}_*)$ and $Q_2(\mathbf{U}_*)$, which have statistical distributions given by (F-19) and (F-20), are (Johansen, 1983, p. 183)

$$\phi_{p_1}(s) = E(\exp\{isQ_1(\mathbf{U}_*)\}) = (1-2is\sigma_\varepsilon^2)^{-\frac{1}{2}p_1} \qquad \text{(F-31)}$$

and

$$\phi_{n-p}(t) = E(\exp\{itQ_2(\mathbf{U}_*)\}) = (1-2it\sigma_\varepsilon^2)^{-\frac{1}{2}(n-p)} \qquad \text{(F-32)}$$

Equations (F-31) and (F-32) are manipulated to give expressions used to simplify (F-28) and (F-29). Taking successive derivatives of $\phi_{p_1}(s)$ yields the general term

$$\frac{d^\mu}{ds^\mu}\phi_{p_1}(s) = \frac{d^\mu}{ds^\mu}E(\exp\{isQ_1(\mathbf{U}_*)\}) = E((i|\mathbf{W}|)^{2\mu}\exp\{is|\mathbf{W}|^2\})$$

$$= \frac{d^{\mu}}{ds^{\mu}}(1-2is\sigma_{\varepsilon}^{2})^{-\frac{1}{2}p_1} = p_1(p_1+2)\cdots(p_1+2(\mu-1))\sigma_{\varepsilon}^{2\mu}i^{\mu}(1-2is\sigma_{\varepsilon}^{2})^{-\frac{1}{2}(p_1+2\mu)}$$
$$= p_1(p_1+2)\cdots(p_1+2(\mu-1))\sigma_{\varepsilon}^{2\mu}i^{\mu}\phi_{p_1+2\mu}(s) \qquad \text{(F-33)}$$

At $s$=0, (F-33) becomes

$$E(i|\mathbf{W}|)^{2\mu} = p_1(p_1+2)\cdots(p_1+2(\mu-1))\sigma_{\varepsilon}^{2\mu}i^{\mu} \qquad \text{(F-34)}$$

Combination of (F-33) and (F-34), then cancellation of $i$, reveals that

$$E(|\mathbf{W}|^{2\mu}\exp\{is|\mathbf{W}|^{2}\}) = E(|\mathbf{W}|^{2\mu})\phi_{p_1+2\mu}(s) \qquad \text{(F-35)}$$

A similar analysis of $\phi_{n-p}(t)$ shows that

$$E(|\mathbf{Z}|^{2\upsilon}\exp\{it|\mathbf{Z}|^{2}\}) = E(|\mathbf{Z}|^{2\upsilon})\phi_{n-p+2\upsilon}(t) \qquad \text{(F-36)}$$

Next, combination of (F-35) and (F-36), in which $|\mathbf{W}|^2$ and $|\mathbf{Z}|^2$ are statistically independent, gives

$$E(|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon}\exp\{is|\mathbf{W}|^{2}+it|\mathbf{Z}|^{2}\})$$
$$= E(|\mathbf{W}|^{2\mu}\exp\{is|\mathbf{W}|^{2}\})E(|\mathbf{Z}|^{2\upsilon}\exp\{it|\mathbf{Z}|^{2}\})$$
$$= E(|\mathbf{W}|^{2\mu}|\mathbf{Z}|^{2\upsilon})\phi_{p_1+2\mu}(s)\phi_{n-p+2\upsilon}(t) \qquad \text{(F-37)}$$

Finally, substitution of (F-37) into (F-28) and (F-29) yields the expressions

$$itE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z}))$$
$$= it\phi_{p_1+2\mu}(s)\phi_{n-p+2\upsilon}(t)E(C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) \qquad \text{(F-38)}$$

and

$$isE(\exp\{isQ_1(\mathbf{U}_*)+itQ_2(\mathbf{U}_*)\}C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z}))$$
$$= is\phi_{p_1+2\mu}(s)\phi_{n-p+2\upsilon}(t)E(C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) \qquad \text{(F-39)}$$

By definition, $2\mu$ and $2\upsilon$ are the powers on $|\mathbf{W}|$ and $|\mathbf{Z}|$ that are factored out of $C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})$ and $C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})$. (See (F-25), for example.) Examination of (F-3) (omitting the fourth-order term in **e**), (F-6), and (F-8) shows that possible values of $2\mu$ and $2\upsilon$ for second-order terms are $(2\mu, 2\upsilon) = (0, 0), (2, 0), (0, 2)$, and possible values for fourth-order terms are $(2\mu, 2\upsilon) = (0, 0), (2, 0), (0, 2), (4, 0), (2, 2), (0, 4)$. Expected values are evaluated further on in this section and in appendix G where the second-order terms are shown to be functions of $\sigma_{\beta}^{2}$, whereas the fourth-order terms are functions of $\sigma_\varepsilon^4$. Thus, the sums over $C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})$ for each combination $(2\mu, 2\upsilon)$ are defined as

$$\left.\begin{array}{lll}
\gamma_1\sigma_\beta^2 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,0) \\
\gamma_2\sigma_\beta^2 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (2,0) \\
\gamma_3\sigma_\beta^2 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,2) \\
\gamma_4\sigma_\varepsilon^4 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,0) \\
\gamma_5\sigma_\varepsilon^4 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (2,0) \\
\gamma_6\sigma_\varepsilon^4 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,2) \\
\gamma_7\sigma_\varepsilon^4 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (4,0) \\
\gamma_8\sigma_\varepsilon^4 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (2,2) \\
\gamma_9\sigma_\varepsilon^4 = E(\sum_k C_{1k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,4)
\end{array}\right\} \tag{F-40}$$

where the sums over $k$ involve only the indicated powers. Similarly the sums over $C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})$ for each combination $(2\mu, 2\upsilon)$ are defined as

$$\left.\begin{array}{lll}
\hat{\gamma}_1\sigma_\beta^2 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,0) \\
\hat{\gamma}_2\sigma_\beta^2 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (2,0) \\
\hat{\gamma}_3\sigma_\beta^2 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,2) \\
\hat{\gamma}_4\sigma_\varepsilon^4 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,0) \\
\hat{\gamma}_5\sigma_\varepsilon^4 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (2,0) \\
\hat{\gamma}_6\sigma_\varepsilon^4 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,2) \\
\hat{\gamma}_7\sigma_\varepsilon^4 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (4,0) \\
\hat{\gamma}_8\sigma_\varepsilon^4 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (2,2) \\
\hat{\gamma}_9\sigma_\varepsilon^4 = E(\sum_k C_{2k}(\mathbf{T},\mathbf{W},\mathbf{Z})) & \text{for} & (0,4)
\end{array}\right\} \tag{F-41}$$

where, again, sums involve only the indicated powers. With (F-31), (F-32), and (F-38)-(F-41), (F-21) can be expressed as

$$\psi(s,t) = \phi_{p_1}(s)\phi_{n-p}(t) + is\phi_{p_1}(s)\phi_{n-p}(t)\gamma_1\sigma_\beta^2 + is\phi_{p_1+2}(s)\phi_{n-p}(t)\gamma_2\sigma_\beta^2$$
$$+ is\phi_{p_1}(s)\phi_{n-p+2}(t)\gamma_3\sigma_\beta^2 + is\phi_{p_1}(s)\phi_{n-p}(t)\gamma_4\sigma_\varepsilon^4 + is\phi_{p_1+2}(s)\phi_{n-p}(t)\gamma_5\sigma_\varepsilon^4$$

$$+ is\phi_{p_1}(s)\phi_{n-p+2}(t)\gamma_6\sigma_\varepsilon^4 + is\phi_{p_1+4}(s)\phi_{n-p}(t)\gamma_7\sigma_\varepsilon^4 + is\phi_{p_1+2}(s)\phi_{n-p+2}(t)\gamma_8\sigma_\varepsilon^4$$
$$+ is\phi_{p_1}(s)\phi_{n-p+4}(t)\gamma_9\sigma_\varepsilon^4 + it\phi_{p_1}(s)\phi_{n-p}(t)\hat{\gamma}_1\sigma_\beta^2 + it\phi_{p_1+2}(s)\phi_{n-p}(t)\hat{\gamma}_2\sigma_\beta^2$$
$$+ it\phi_{p_1}(s)\phi_{n-p+2}(t)\hat{\gamma}_3\sigma_\beta^2 + it\phi_{p_1}(s)\phi_{n-p}(t)\hat{\gamma}_4\sigma_\varepsilon^4 + it\phi_{p_1+2}(s)\phi_{n-p}(t)\hat{\gamma}_5\sigma_\varepsilon^4$$
$$+ it\phi_{p_1}(s)\phi_{n-p+2}(t)\hat{\gamma}_6\sigma_\varepsilon^4 + it\phi_{p_1+4}(s)\phi_{n-p}(t)\hat{\gamma}_7\sigma_\varepsilon^4 + it\phi_{p_1+2}(s)\phi_{n-p+2}(t)\hat{\gamma}_8\sigma_\varepsilon^4$$
$$+ it\phi_{p_1}(s)\phi_{n-p+4}(t)\hat{\gamma}_9\sigma_\varepsilon^4 \tag{F-42}$$

Equation (F-42) must now be put into an approximate form to derive the distributions of $S(\theta) - S(\hat{\theta})$ and $S(\hat{\theta})$ separately, which can be done only if $S(\theta) - S(\hat{\theta})$ and $S(\hat{\theta})$ are approximately statistically independent. Statistical independence is indicated if $\psi(s,t)$ can be written as the product of a characteristic function in terms of $s$ and a characteristic function in terms of $t$ (Papoulis, 1965, p. 213-214). Approximate statistical independence is shown and the forms of the two distributions are developed as follows. From (F-31) or (F-32) note that for $r = p_1$ or $r = n-p$ and $w = s$ or $w = t$

$$\phi_r(w) = (1 - 2iw\sigma_\varepsilon^2)\phi_{r+2}(w) \tag{F-43}$$

and, by approximating $(1 - 2iw\sigma_\varepsilon^2)^{-1}$ through first order (which is all that is required),

$$\phi_{r+2}(w) \approx (1 + 2iw\sigma_\varepsilon^2)\phi_r(w) \tag{F-44}$$

Now, substitution of (F-43) and (F-44) into (F-42) keeping terms through orders $\sigma_\beta^2\sigma_\varepsilon$ and $\sigma_\varepsilon^4$ yields

$$\psi(s,t) \approx \phi_{p_1}(s)\phi_{n-p}(t) + is\phi_{p_1+2}(s)\phi_{n-p}(t)((\gamma_1+\gamma_2+\gamma_3)\sigma_\beta^2 + (\gamma_4+\gamma_5+\gamma_6+\gamma_7+\gamma_8+\gamma_9)\sigma_\varepsilon^4)$$
$$+ it\phi_{p_1}(s)\phi_{n-p+2}(t)((\hat{\gamma}_1+\hat{\gamma}_2+\hat{\gamma}_3)\sigma_\beta^2 + (\hat{\gamma}_4+\hat{\gamma}_5+\hat{\gamma}_6+\hat{\gamma}_7+\hat{\gamma}_8+\hat{\gamma}_9)\sigma_\varepsilon^4)$$
$$\approx (\phi_{p_1}(s) + is\phi_{p_1+2}(s)(\gamma_w\sigma_\beta^2 + \gamma_I\sigma_\varepsilon^4))(\phi_{n-p}(t) + it\phi_{n-p+2}(t)(\hat{\gamma}_w\sigma_\beta^2 + \hat{\gamma}_I\sigma_\varepsilon^4)) \tag{F-45}$$

where

$$\gamma_w = \gamma_1 + \gamma_2 + \gamma_3 \tag{F-46}$$

$$\gamma_I = \gamma_4 + \gamma_5 + \gamma_6 + \gamma_7 + \gamma_8 + \gamma_9 \tag{F-47}$$

$$\hat{\gamma}_w = \hat{\gamma}_1 + \hat{\gamma}_2 + \hat{\gamma}_3 \tag{F-48}$$

$$\hat{\gamma}_I = \hat{\gamma}_4 + \hat{\gamma}_5 + \hat{\gamma}_6 + \hat{\gamma}_7 + \hat{\gamma}_8 + \hat{\gamma}_9 \tag{F-49}$$

Terms involving $\gamma_9$ and $\hat{\gamma}_7$ in (F-42) require shifting characteristic functions from $\phi_{n-p+4}(t)$ to $\phi_{n-p}(t)$ and from $\phi_{p_1+4}(s)$ to $\phi_{p_1}(s)$, respectively. Because of these large shifts, the forms of the terms dropped should be obtained to be sure that accuracy is not appreciably less than for the other terms. Use of (F-43) shows that

$$
\begin{aligned}
&is\phi_{p_1}(s)\phi_{n-p+4}(t)\gamma_9\sigma_\varepsilon^4 - is\phi_{p_1+2}(s)\phi_{n-p}(t)\gamma_9\sigma_\varepsilon^4 \\
&= is\phi_{p_1+2}(s)\phi_{n-p+4}(t)(1-2is\sigma_\varepsilon^2 - (1-2it\sigma_\varepsilon^2)^2)\gamma_9\sigma_\varepsilon^2
\end{aligned}
$$

where

$$
\begin{aligned}
&1-2is\sigma_\varepsilon^2 - (1-2it\sigma_\varepsilon^2)^2 \\
&= 1-2is\sigma_\varepsilon^2 - 1 + 4it\sigma_\varepsilon^2 - 4i^2t^2\sigma_\varepsilon^4 = 2i(2t-s)\sigma_\varepsilon^2 - 4i^2t^2\sigma_\varepsilon^4 \\
&\approx 2i(2t-s)\sigma_\varepsilon^2
\end{aligned}
$$

so that

$$is\phi_{p_1}(s)\phi_{n-p+4}(t)\gamma_9\sigma_\varepsilon^4 \approx is\phi_{p_1+2}(s)\phi_{n-p}(t)\gamma_9\sigma_\varepsilon^4 + 2i^2s(2t-s)\phi_{p_1+2}(s)\phi_{n-p+4}(t)\gamma_9\sigma_\varepsilon^6 \qquad \text{(F-50)}$$

Similarly

$$
\begin{aligned}
&it\phi_{p_1+4}(s)\phi_{n-p}(t)\hat{\gamma}_7\sigma_\varepsilon^4 - it\phi_{p_1}(s)\phi_{n-p+2}(t)\hat{\gamma}_7\sigma_\varepsilon^4 \\
&= it\phi_{p_1+4}(s)\phi_{n-p+2}(t)(1-2it\sigma_\varepsilon^2 - (1-2is\sigma_\varepsilon^2)^2)\hat{\gamma}_7\sigma_\varepsilon^4
\end{aligned}
$$

where

$$
\begin{aligned}
&1-2it\sigma_\varepsilon^2 - (1-2is\sigma_\varepsilon^2)^2 \\
&= 1-2it\sigma_\varepsilon^2 - 1 + 4is\sigma_\varepsilon^2 - 4i^2s^2\sigma_\varepsilon^4 = 2i(2s-t)\sigma_\varepsilon^2 - 4i^2s^2\sigma_\varepsilon^4 \\
&\approx 2i(2s-t)\sigma_\varepsilon^2
\end{aligned}
$$

so that

$$it\phi_{p_1+4}(s)\phi_{n-p}(t)\hat{\gamma}_7\sigma_\varepsilon^4 \approx it\phi_{p_1}(s)\phi_{n-p+2}(t)\hat{\gamma}_7\sigma_\varepsilon^4 + 2i^2t(2s-t)\phi_{p_1+4}(s)\phi_{n-p+2}(t)\hat{\gamma}_7\sigma_\varepsilon^6 \qquad \text{(F-51)}$$

Terms dropped from (F-50) and (F-51) are of accuracy similar to the other terms dropped to obtain (F-45).

The terms in (F-45) can be put into standard forms for $\chi^2$ distributions by noting the following.

$$
\begin{aligned}
&\phi_{p_1}(s(1+(\gamma_w\sigma_\beta^2/\sigma_\varepsilon^2+\gamma_I\sigma_\varepsilon^2)/p_1)) = (1-2is(1+(\gamma_w\sigma_\beta^2/\sigma_\varepsilon^2+\gamma_I\sigma_\varepsilon^2)/p_1)\sigma_\varepsilon^2)^{-\frac{1}{2}p_1} \\
&= (1-2is\sigma_\varepsilon^2 - 2is(\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4)/p_1)^{-\frac{1}{2}p_1} = (1-2is\sigma_\varepsilon^2-\varepsilon)^{-\frac{1}{2}p_1}
\end{aligned}
$$

$$\approx (1-2is\sigma_\varepsilon^2)^{-\frac{1}{2}p_1} + \frac{d}{d\varepsilon}(1-2is\sigma_\varepsilon^2-\varepsilon)_{\varepsilon=0}^{-\frac{1}{2}p_1}\,\varepsilon$$

$$= \phi_{p_1}(s) + is\phi_{p_1+2}(s)(\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4) \tag{F-52}$$

where $\varepsilon = 2is(\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4)/p_1$, and

$$\phi_{n-p}(t(1+(\hat{\gamma}_w\sigma_\beta^2/\sigma_\varepsilon^2+\hat{\gamma}_I\sigma_\varepsilon^2)/(n-p))) = (1-2it(1+(\hat{\gamma}_w\sigma_\beta^2/\sigma_\varepsilon^2+\hat{\gamma}_I\sigma_\varepsilon^2)/(n-p))\sigma_\varepsilon^2)^{-\frac{1}{2}(n-p)}$$

$$= (1-2it\sigma_\varepsilon^2-2it(\hat{\gamma}_w\sigma_\beta^2+\hat{\gamma}_I\sigma_\varepsilon^4)/(n-p))^{-\frac{1}{2}(n-p)} = (1-2it\sigma_\varepsilon^2-\varepsilon)^{-\frac{1}{2}(n-p)}$$

$$\approx (1-2it\sigma_\varepsilon^2)^{-\frac{1}{2}(n-p)} + \frac{d}{d\varepsilon}(1-2it\sigma_\varepsilon^2-\varepsilon)_{\varepsilon=0}^{-\frac{1}{2}(n-p)}\,\varepsilon$$

$$= \phi_{n-p}(t) + it\phi_{n-p+2}(t)(\hat{\gamma}_w\sigma_\beta^2+\hat{\gamma}_I\sigma_\varepsilon^4) \tag{F-53}$$

where $\varepsilon = 2it(\hat{\gamma}_w\sigma_\beta^2+\hat{\gamma}_I\sigma_\varepsilon^4)/(n-p)$. With (F-52) and (F-53), (F-45) becomes

$$\psi(s,t) \approx \phi_{p_1}(s(1+(\gamma_w\sigma_\beta^2/\sigma_\varepsilon^2+\gamma_I\sigma_\varepsilon^2)/p_1))\phi_{n-p}(t(1+(\hat{\gamma}_w\sigma_\beta^2/\sigma_\varepsilon^2+\hat{\gamma}_I\sigma_\varepsilon^2)/(n-p))) \tag{F-54}$$

***Approximate statistical distributions.*** By definition (Papoulis, 1965, p. 154) $\phi_r(cw) = (1-2iwc\sigma_\varepsilon^2)^{-r/2}$ is the characteristic function for the pdf for a $c\sigma_\varepsilon^2\chi^2(r)$ random variable, where $c$ is some constant. Therefore, from (F-54) and to the order of accuracy of approximations used, $S(\theta)-S(\hat{\theta})$ and $S(\hat{\theta})$ are independently distributed as

$$S(\theta)-S(\hat{\theta}) \sim \chi^2(p_1)(\sigma_\varepsilon^2+(\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4)/p_1) \tag{F-55}$$

$$S(\hat{\theta}) \sim \chi^2(n-p)(\sigma_\varepsilon^2+(\hat{\gamma}_w\sigma_\beta^2+\hat{\gamma}_I\sigma_\varepsilon^4)/(n-p)) \tag{F-56}$$

Also, by definition (Graybill, 1976, p. 66) $(\chi^2(p_1)/p_1)/(\chi^2(n-p)/(n-p))$ has an $F(p_1,n-p)$ distribution with $p_1$ and $n-p$ degrees of freedom, so that

$$\frac{(S(\theta)-S(\hat{\theta}))/p_1}{S(\hat{\theta})/(n-p)} \sim F(p_1,n-p)\frac{\sigma_\varepsilon^2+(\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4)/p_1}{\sigma_\varepsilon^2+(\hat{\gamma}_w\sigma_\beta^2+\hat{\gamma}_I\sigma_\varepsilon^4)/(n-p)} \tag{F-57}$$

If desired, the correction factor can be approximated to the order of accuracy used in the derivations as

$$\frac{\sigma_\varepsilon^2+(\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4)/p_1}{\sigma_\varepsilon^2+(\hat{\gamma}_w\sigma_\beta^2+\hat{\gamma}_I\sigma_\varepsilon^4)/(n-p)} \approx (1+(\gamma_w\sigma_\beta^2/\sigma_\varepsilon^2+\gamma_I\sigma_\varepsilon^4)/p_1)(1-(\hat{\gamma}_w\sigma_\beta^2/\sigma_\varepsilon^2+\hat{\gamma}_I\sigma_\varepsilon^2)/(n-p))$$

$$\approx 1+(\gamma_w\sigma_\beta^2/\sigma_\varepsilon^2+\gamma_I\sigma_\varepsilon^2)/p_1-(\hat{\gamma}_w\sigma_\beta^2/\sigma_\varepsilon^2+\hat{\gamma}_I\sigma_\varepsilon^2)/(n-p) \tag{F-58}$$

***Evaluation of correction factor.*** From (F-6), (F-10)-(F-18) and the definitions of the $\hat{\gamma}_i, i=1,9$, given by (F-41) note that $(n-p)$ times the denominator of the correction factor in (F-57) is

$$
\begin{aligned}
&(n-p)\sigma_\varepsilon^2+\hat{\gamma}_w\sigma_\beta^2+\hat{\gamma}_I\sigma_\varepsilon^4 \\
&=E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}(\mathbf{I}-\mathbf{H})\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)+E(\mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*-\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}(\mathbf{I}-\mathbf{H})\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*) \\
&+E(S(\hat{\boldsymbol{\theta}})-\mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*) \\
&=E(S(\hat{\boldsymbol{\theta}}))
\end{aligned}
\tag{F-59}
$$

where $E(\mathbf{U}_*'\mathbf{V}_*^{-1/2}(\mathbf{I}-\mathbf{H})\mathbf{V}_*^{-1/2}\mathbf{U}_*)=(n-p)\sigma_\varepsilon^2$,
$E(\mathbf{U}_*'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{U}_*-\mathbf{U}_*'\mathbf{V}_*^{-1/2}(\mathbf{I}-\mathbf{H})\mathbf{V}_*^{-1/2}\mathbf{U}_*)=\hat{\gamma}_w\sigma_\beta^2$, and
$E(S(\hat{\boldsymbol{\theta}})-\mathbf{U}_*'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{U}_*)=\hat{\gamma}_I\sigma_\varepsilon^4$. The component correction factor $\hat{\gamma}_w\sigma_\beta^2$ is given by

$$
\begin{aligned}
\hat{\gamma}_w\sigma_\beta^2&=E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}(\mathbf{V}_*^{\frac{1}{2}}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}-\mathbf{I}+\mathbf{H})\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*) \\
&=(tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}})-n+p)\sigma_\varepsilon^2 \\
&=tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}-(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\omega}^{-1}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2 \\
&=tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{V}_*-\boldsymbol{\omega}^{-1})\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2
\end{aligned}
\tag{F-60}
$$

To make (F-60) solely a function of model error, $\boldsymbol{\omega}^{-1}$ is expressed as the sum of $\mathbf{V}_\varepsilon$ and a matrix $\mathbf{V}_\omega\sigma_\beta^2/\sigma_\varepsilon^2$, where $\mathbf{V}_\omega$ depends only on model error. If $\mathbf{V}$ is defined by $\mathbf{V}=\mathbf{D}_\beta\mathbf{f}(\mathbf{I}-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{V}_\beta(\mathbf{I}-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{D}_\beta\mathbf{f}'$, then use of (3-33) in (F-60) gives

$$
\hat{\gamma}_w\sigma_\beta^2=tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{V}-\mathbf{V}_\omega)\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^2
\tag{F-61}
$$

The component correction factor $\hat{\gamma}_I\sigma_\varepsilon^4$ is expressed by using (F-6) to yield

$$
\begin{aligned}
\hat{\gamma}_I\sigma_\varepsilon^4\approx{}&\frac{1}{4}E(\sum_i\mathbf{l}_*'\mathbf{D}^2f_i\mathbf{l}_*\boldsymbol{\omega}_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j\boldsymbol{\omega}_j^{\frac{1}{2}}\mathbf{l}_*'\mathbf{D}^2f_j\mathbf{l}_*)-E(\sum_i\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z}_*\mathbf{l}_*'\mathbf{D}^2f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i\mathbf{D}^2f_i\mathbf{l}_*\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z}_*) \\
&+E(\sum_i\boldsymbol{\omega}_i^{\frac{1}{2}}\mathbf{Z}_*\mathbf{l}_*'\mathbf{D}^2f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\sum_j\boldsymbol{\omega}_j\mathbf{l}_*'\mathbf{D}^2f_j\mathbf{l}_*)
\end{aligned}
\tag{F-62}
$$

Evaluation of the expected values in (F-62) is deferred to appendix G.

The numerator of the correction factor times $p_1$ evaluates in the same way as the denominator. That is, for $\mathbf{M}=\mathbf{H}$ and $p_1=p$

$$
p\sigma_\varepsilon^2+\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4
$$

$$= E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{H}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*) + E(\mathbf{U}_*'\boldsymbol{\omega}\mathbf{U}_* - \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* - \mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{H}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)$$
$$+ E(S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}}) - \mathbf{U}_*'\boldsymbol{\omega}\mathbf{U}_* + \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*)$$
$$= E(S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}})) \qquad \text{(F-63)}$$

where $E(\mathbf{U}_*'\mathbf{V}_*^{-1/2}\mathbf{H}\mathbf{V}_*^{-1/2}\mathbf{U}_*) = p\sigma_\varepsilon^2$,
$E(\mathbf{U}_*'\boldsymbol{\omega}\mathbf{U}_* - \mathbf{U}_*'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{U}_* - \mathbf{U}_*'\mathbf{V}_*^{-1/2}\mathbf{H}\mathbf{V}_*^{-1/2}\mathbf{U}_*) = \gamma_w\sigma_\beta^2$,
and $E(S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}}) - \mathbf{U}_*'\boldsymbol{\omega}\mathbf{U}_* + \mathbf{U}_*'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{U}_*) = \gamma_I\sigma_\varepsilon^2$. Then

$$\gamma_w\sigma_\beta^2 = E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}(\mathbf{V}_*^{\frac{1}{2}}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}} - \mathbf{H})\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)$$
$$= (tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}) - p)\sigma_\varepsilon^2$$
$$= tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{V}_* - \boldsymbol{\omega}^{-1})\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2$$
$$= tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{V} - \mathbf{V}_\omega)\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^2 \qquad \text{(F-64)}$$

Finally, expression of $\gamma_I\sigma_\varepsilon^4$ using (F-3) (through third-order terms) and (F-6) shows that

$$\gamma_I\sigma_\varepsilon^4 = -\hat{\gamma}_I\sigma_\varepsilon^4 \qquad \text{(F-65)}$$

For $\mathbf{M} = \mathbf{P}\mathbf{P}'/\mathbf{P}'\mathbf{P}$ and $p_1 = 1$

$$\sigma_\varepsilon^2 + \gamma_w\sigma_\beta^2 + \gamma_I\sigma_\varepsilon^4$$
$$= E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\frac{\mathbf{P}\mathbf{P}'}{\mathbf{P}'\mathbf{P}}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*) + E(\mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* - \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* - \mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\frac{\mathbf{P}\mathbf{P}'}{\mathbf{P}'\mathbf{P}}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)$$
$$+ E(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}) - \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + \mathbf{U}_*'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*)$$
$$= E(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}})) \qquad \text{(F-66)}$$

where $E(\mathbf{U}_*'\mathbf{V}_*^{-1/2}\mathbf{P}\mathbf{P}'/\mathbf{P}'\mathbf{P}\mathbf{V}_*^{-1/2}\mathbf{U}_*) = \sigma_\varepsilon^2$,
$E(\mathbf{U}_*'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R}+\mathbf{Q}\mathbf{Q}'/\mathbf{Q}'\mathbf{Q})\boldsymbol{\omega}^{1/2}\mathbf{U}_* - \mathbf{U}_*'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{U}_* - \mathbf{U}_*'\mathbf{V}_*^{-1/2}\mathbf{P}\mathbf{P}'/\mathbf{P}'\mathbf{P}\mathbf{V}_*^{-1/2}\mathbf{U}_*)$
$= \gamma_w\sigma_\beta^2$, and $E(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}) - \mathbf{U}_*'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R}+\mathbf{Q}\mathbf{Q}'/\mathbf{Q}'\mathbf{Q})\boldsymbol{\omega}^{1/2}\mathbf{U}_*$
$+ \mathbf{U}_*'\boldsymbol{\omega}^{1/2}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{1/2}\mathbf{U}_*) = \gamma_I\sigma_\varepsilon^4$. As before

$$\gamma_w\sigma_\beta^2 = E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}(\mathbf{V}_*^{\frac{1}{2}}\boldsymbol{\omega}^{\frac{1}{2}}\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}} - \frac{\mathbf{P}\mathbf{P}'}{\mathbf{P}'\mathbf{P}})\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)$$
$$= (tr(\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}) - 1)\sigma_\varepsilon^2$$

$$= tr(\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{V}_* - \omega^{-1})\omega^{\frac{1}{2}})\sigma_\varepsilon^2$$

$$= \frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}(\mathbf{V} - \mathbf{V}_\omega)\omega^{\frac{1}{2}}\mathbf{Q}\sigma_\beta^2 \qquad \text{(F-67)}$$

and, from (F-8),

$$\gamma_I \sigma_\varepsilon^4 \approx \frac{1}{4}E(\sum_i \tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \tilde{\mathbf{l}}_* \omega_i^{\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \tilde{\mathbf{l}}_*)$$
$$-\frac{1}{2}E(\sum_i \tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \tilde{\mathbf{l}}_* \omega_i^{\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 g \tilde{\mathbf{l}}_*) + \frac{1}{4}\frac{1}{\mathbf{Q}'\mathbf{Q}}E(\tilde{\mathbf{l}}_*'\mathbf{D}^2 g \tilde{\mathbf{l}}_*)^2$$
$$- E(\tilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \tilde{\mathbf{l}}_* \omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}})$$
$$+ E((\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)^2\tilde{\mathbf{l}}_*'\mathbf{D}^2 g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g \tilde{\mathbf{l}}_*)$$
$$+ E(\tilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \tilde{\mathbf{l}}_*)$$
$$+ \frac{1}{\mathbf{Q}'\mathbf{Q}}E(\tilde{\mathbf{Z}}'\sum_j \omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'\tilde{\mathbf{l}}_*'\mathbf{D}^2 g \tilde{\mathbf{l}}_*) - \hat{\gamma}_I \sigma_\varepsilon^4 \qquad \text{(F-68)}$$

Evaluation of the expected values in (F-68) is deferred to appendix G.

## Distribution for Prediction Intervals

***Forms of sum of squares functions for perturbation analysis.*** The same procedures as used to derive the distribution for confidence intervals are used to derive the distribution for prediction intervals. The distribution, correction factors, and the variables composing these results will be shown to have the same forms as those for confidence intervals. The function $S_a(\tilde{\theta}, \tilde{\theta}_p)$ given by (E-62) is repeated here in expanded form as

$$S_a(\tilde{\theta}, \tilde{\theta}_p) \approx \mathbf{U}_{*a}'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a} + \mathbf{U}_{*a}'\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj} - \tilde{\mathbf{l}}_a'\mathbf{D}_a^2\mathbf{f}_{aj}\tilde{\mathbf{l}}_a)$$
$$+ \frac{1}{4}\sum_i (E_{ai} - \tilde{\mathbf{l}}_a'\mathbf{D}_a f_{ai}\tilde{\mathbf{l}}_a)\mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj} - \tilde{\mathbf{l}}_a'\mathbf{D}_a^2 f_{aj}\tilde{\mathbf{l}}_a)$$
$$+ (\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a} + \frac{1}{2}\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj} - \tilde{\mathbf{l}}_a'\mathbf{D}_a^2 f_{aj}\tilde{\mathbf{l}}_a))\frac{\mathbf{Q}_a}{\mathbf{Q}_a'\mathbf{Q}_a}(\tilde{\mathbf{l}}_a'\mathbf{D}_a^2 h \tilde{\mathbf{l}}_a - \mathbf{e}'\mathbf{G}_a\mathbf{D}_a^2 h \mathbf{G}_a'\mathbf{e})$$
$$+ \frac{1}{4}\frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}(\tilde{\mathbf{l}}_a'\mathbf{D}_a^2 h \tilde{\mathbf{l}}_a - \mathbf{e}'\mathbf{G}_a\mathbf{D}_a^2 h \mathbf{G}_a'\mathbf{e})^2$$

$$-\tilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\sum_i \mathbf{D}_a^2 f_{ai}\tilde{\mathbf{l}}_a\mathbf{W}_{ai}^{\frac{1}{2}}\tilde{\mathbf{Z}}_a$$
$$+(\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{Q}'_a\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a})^2\,\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 h(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\,\mathbf{D}_a^2 h\,\tilde{\mathbf{l}}_a$$
$$-\tilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj}-\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 f_{aj}\,\tilde{\mathbf{l}}_a)$$
$$+\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\tilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h'(\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 h\,\tilde{\mathbf{l}}_a-\mathbf{e}'\mathbf{G}_a\mathbf{D}_a^2 h\mathbf{G}'_a\mathbf{e}) \tag{F-69}$$

The perturbation analysis uses (E-41) and (E-42) in the form

$$\tilde{\mathbf{l}}_a=\mathbf{l}_{*a}+\begin{bmatrix}(\gamma'\gamma)^{-1}\gamma'\mathbf{e}\\ e_p\end{bmatrix} \tag{F-70}$$

then uses (F-70) for $\tilde{\mathbf{l}}_a$ in (F-69) and drops fourth-order terms involving $\mathbf{e}$ or $e_p$ to get

$$S_a(\tilde{\theta},\tilde{\theta}_p)\approx \mathbf{U}'_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a}+\mathbf{U}'_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj}-\tilde{\mathbf{l}}'_a\mathbf{D}_a^2\mathbf{f}_{aj}\,\tilde{\mathbf{l}}_a)$$
$$+\frac{1}{4}\sum_i \tilde{\mathbf{l}}'_{*a}\,\mathbf{D}_a^2 f_{ai}\tilde{\mathbf{l}}_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_{*a}\,\mathbf{D}_a^2 f_{aj}\tilde{\mathbf{l}}_{*a}+\mathbf{U}'_{*a}\mathbf{W}_a^{\frac{1}{2}}\frac{\mathbf{Q}_a}{\mathbf{Q}'_a\mathbf{Q}_a}(\tilde{\mathbf{l}}'_a\mathbf{D}_a^2 h\,\tilde{\mathbf{l}}_a-\mathbf{e}'\mathbf{G}_a\mathbf{D}_a^2 h\mathbf{G}'_a\mathbf{e})$$
$$-\frac{1}{2}\sum_i \tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{ai}\tilde{\mathbf{l}}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}\frac{\mathbf{Q}_a}{\mathbf{Q}'_a\mathbf{Q}_a}\tilde{\mathbf{l}}'_{*a}\,\mathbf{D}_a^2 h\tilde{\mathbf{l}}_{*a}+\frac{1}{4}\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}(\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 h\,\tilde{\mathbf{l}}_{*a})^2$$
$$-\tilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\sum_i \mathbf{D}_a^2 f_{ai}\tilde{\mathbf{l}}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}\tilde{\mathbf{Z}}_a$$
$$+(\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{Q}'_a\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{U}_{*a})^2\,\tilde{\mathbf{l}}'_{*a}\,\mathbf{D}_a^2 h(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\,\mathbf{D}_a^2 h\,\tilde{\mathbf{l}}_{*a}$$
$$+\tilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{aj}\tilde{\mathbf{l}}_{*a}$$
$$+\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\tilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h'\,\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 h\,\tilde{\mathbf{l}}_{*a} \tag{F-71}$$

A function analogous to $S(\hat{\theta})$, but applying for prediction intervals, also is needed. To obtain this function, $\lambda$ is set to zero in (E-33) so that minimization of the resulting formulation is simply minimization of $S_a(\theta,\theta_p)$. The result of this minimization is just (E-52) with $\lambda=0$, which is of the form of (B-4), the equation set from the regression solution. That is,

$$\sum_i\sum_j \mathbf{D}_a f'_{ai}W_{aij}(U_{*aj}+\frac{1}{2}E_{aj}-\mathbf{D}_a f_{aj}\mathbf{l}_{*a}-\mathbf{D}_a f_{aj}\mathbf{q}_a-\frac{1}{2}\mathbf{l}'_a\mathbf{D}_a^2 f_{aj}\mathbf{l}_a)$$

$$+\sum_i\sum_j \mathbf{D}_a^2 f_{ai}\mathbf{l}_a W_{aij}(U_{*aj} - \mathbf{D}_a f_{aj}\mathbf{l}_{*a}) = \mathbf{0} \tag{F-72}$$

where

$$\mathbf{l}_a = \mathbf{l}_{*a} + \begin{bmatrix} (\gamma'\gamma)^{-1}\gamma'\mathbf{e} \\ e_p \end{bmatrix} \tag{F-73}$$

After solving (F-72) for $\mathbf{l}_a$ and $\mathbf{q}_a$ in the same way as (B-4) was solved for $\mathbf{l}$ and $\mathbf{q}$, the results in appendix B can be used to obtain an augmented sum of squares of the form of (F-1). Finally, substitution of (F-73) for $\mathbf{l}_a$ in this augmented sum of squares yields a result of the form of (F-6):

$$\begin{aligned}
S_a(\hat{\theta}, \hat{\theta}_p) \approx\; & \mathbf{U}'_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a} + \mathbf{U}'_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}(E_{aj} - \mathbf{l}'_a\mathbf{D}_a^2\mathbf{f}_{aj}\mathbf{l}_a) \\
& + \frac{1}{4}\sum_i \mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}\mathbf{l}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{aj}\mathbf{l}_{*a} \\
& - \sum_i \mathbf{W}_{ai}^{\frac{1}{2}}\mathbf{Z}_{*a}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\sum_i \mathbf{D}_a^2 f_{ai}\mathbf{l}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}\mathbf{Z}_{*a} \\
& + \sum_i \mathbf{W}_{ai}^{\frac{1}{2}}\mathbf{Z}_{*a}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\sum_j \mathbf{W}_{aj}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{aj}\mathbf{l}_{*a}
\end{aligned} \tag{F-74}$$

where $\mathbf{Z}_{*a} = (\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{1/2}\mathbf{U}_{*a}$.

***Approximate characteristic function for sum of squares ratio.*** The joint characteristic function for $S_a(\tilde{\theta}, \tilde{\theta}_p) - S_a(\hat{\theta}, \hat{\theta}_p)$ and $S_a(\hat{\theta}, \hat{\theta}_p)$ is

$$\psi_a(s,t) = E(\exp\{is(S_a(\tilde{\theta}, \tilde{\theta}_p) - S_a(\hat{\theta}, \hat{\theta}_p)) + itS_a(\hat{\theta}, \hat{\theta}_p)\}) \tag{F-75}$$

Equation (F-75) can be written in terms of $\chi^2$ distributed variables and deviations from them as before so that

$$S_a(\tilde{\theta}, \tilde{\theta}_p) - S_a(\hat{\theta}, \hat{\theta}_p) = Q_{1a}(\mathbf{U}_{*a}) + D_{1a} \tag{F-76}$$

$$S_a(\hat{\theta}, \hat{\theta}_p) = Q_{2a}(\mathbf{U}_{*a}) + D_{2a} \tag{F-77}$$

$$Q_{1a}(\mathbf{U}_{*a}) = \mathbf{U}'_{*a}\mathbf{V}_{*a}^{-\frac{1}{2}}\frac{\mathbf{P}_a\mathbf{P}'_a}{\mathbf{P}'_a\mathbf{P}_a}\mathbf{V}_{*a}^{-\frac{1}{2}}\mathbf{U}_{*a} \tag{F-78}$$

$$Q_{2a}(\mathbf{U}_{*a}) = \mathbf{U}'_{*a}\mathbf{V}_{*a}^{-\frac{1}{2}}(\mathbf{I}_a - \mathbf{H}_a)\mathbf{V}_{*a}^{-\frac{1}{2}}\mathbf{U}_{*a} \tag{F-79}$$

where $\mathbf{V}_{*a} = Var(\mathbf{U}_{*a})$,

$$\mathbf{P}_a = \mathbf{V}_{*a}^{-\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\mathbf{V}_{*a}^{-1}\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h' \tag{F-80}$$

$$\mathbf{H}_a = \mathbf{V}_{*a}^{-\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\mathbf{V}_{*a}^{-1}\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{V}_{*a}^{-\frac{1}{2}} \tag{F-81}$$

$$D_{1a} = S_a(\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) - S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p) - Q_{1a}(\mathbf{U}_{*a}) \tag{F-82}$$

$$D_{2a} = S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p) - Q_{2a}(\mathbf{U}_{*a}) \tag{F-83}$$

Then, because $\mathbf{U}_{*a}$ has a normal distribution with variance matrix $\mathbf{V}_{*a}\sigma_\varepsilon^2$,

$$\mathbf{V}_{*a}^{-\frac{1}{2}}\mathbf{U}_{*a} \sim N(\mathbf{0}, \mathbf{I}_a\sigma_\varepsilon^2) \tag{F-84}$$

Finally, the properties of $\mathbf{P}_a$ and $\mathbf{H}_a$ allow concluding that

$$Q_{1a}(\mathbf{U}_{*a})/\sigma_\varepsilon^2 \sim \chi^2(1) \tag{F-85}$$

$$Q_{2a}(\mathbf{U}_{*a})/\sigma_\varepsilon^2 \sim \chi^2(n-p) \tag{F-86}$$

***Approximate statistical distributions***. The derivation to find the distributions of $S_a(\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) - S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p)$, $S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p)$, and their ratio is completely analogous to the derivation used to find $S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}})$, $S(\hat{\boldsymbol{\theta}})$, and their ratio. The results are

$$S_a(\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) - S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p) \sim \chi^2(1)(\sigma_\varepsilon^2 + \gamma_{wa}\sigma_\beta^2 + \gamma_{Ia}\sigma_\varepsilon^4) \tag{F-87}$$

$$S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p) \sim \chi^2(n-p)(\sigma_\varepsilon^2 + (\hat{\gamma}_{wa}\sigma_\beta^2 + \hat{\gamma}_{Ia}\sigma_\varepsilon^4)/(n-p)) \tag{F-88}$$

$$\frac{S_a(\tilde{\boldsymbol{\theta}}, \tilde{\theta}_p) - S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p)}{S_a(\hat{\boldsymbol{\theta}}, \hat{\theta}_p)/(n-p)} \sim F(1, n-p)\frac{\sigma_\varepsilon^2 + \gamma_{wa}\sigma_\beta^2 + \gamma_{Ia}\sigma_\varepsilon^4}{\sigma_\varepsilon^2 + (\hat{\gamma}_{wa}\sigma_\beta^2 + \hat{\gamma}_{Ia}\sigma_\varepsilon^4)/(n-p)} \tag{F-89}$$

where

$$\begin{aligned}&\frac{\sigma_\varepsilon^2 + \gamma_{wa}\sigma_\beta^2 + \gamma_{Ia}\sigma_\varepsilon^4}{\sigma_\varepsilon^2 + (\hat{\gamma}_{wa}\sigma_\beta^2 + \hat{\gamma}_{Ia}\sigma_\varepsilon^4)/(n-p)}\\ &\approx 1 + \gamma_{wa}\sigma_\beta^2 + \gamma_{Ia}\sigma_\varepsilon^4 - (\hat{\gamma}_{wa}\sigma_\beta^2 + \hat{\gamma}_{Ia}\sigma_\varepsilon^4)/(n-p)\end{aligned} \tag{F-90}$$

***Evaluation of correction factor.*** The factors $\gamma_{wa}$, $\gamma_{Ia}$, $\hat{\gamma}_{wa}$, and $\hat{\gamma}_{Ia}$ in the correction factor (F-90) are analogous to $\gamma_w$, $\gamma_I$, $\hat{\gamma}_w$, and $\hat{\gamma}_I$, and they evaluate in an analogous manner. Thus,

$$\hat{\gamma}_{wa}\sigma_\beta^2 = E(\mathbf{U}'_{*a}\mathbf{V}_{*a}^{-\frac{1}{2}}(\mathbf{V}_{*a}^{\frac{1}{2}}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}^{\frac{1}{2}} - \mathbf{I}_a + \mathbf{H}_a)\mathbf{V}_{*a}^{-\frac{1}{2}}\mathbf{U}_{*a})$$
$$= tr((\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{V}_{*a} - \mathbf{W}_a^{-1})\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^2$$
$$= tr((\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}(\mathbf{V}_a - \mathbf{V}_{\omega a})\mathbf{W}_a^{\frac{1}{2}})\sigma_\beta^2 \qquad \text{(F-91)}$$

where $\mathbf{V}_{\omega a}$ is analogous to $\mathbf{V}_\omega$,

$$\mathbf{V}_a = \begin{bmatrix} \mathbf{V} & \mathbf{C}_{ep} \\ \mathbf{C}'_{ep} & V_{ep} \end{bmatrix} \qquad \text{(F-92)}$$

and $V_{ep} = \mathbf{D}_\beta g(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{V}_\beta(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{D}_\beta g'$. Next, (F-74) is used to get

$$\hat{\gamma}_{Ia}\sigma_\varepsilon^4 = E(S_a(\hat{\theta}, \hat{\theta}_p) - \mathbf{U}'_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a})$$
$$\approx \frac{1}{4}E(\sum_i \mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}\mathbf{l}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a)\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{aj}\mathbf{l}_{*a})$$
$$- E(\sum_i \mathbf{W}_{ai}^{\frac{1}{2}}\mathbf{Z}_{*a}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\sum_i \mathbf{D}_a^2 f_{ai}\mathbf{l}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}\mathbf{Z}_{*a})$$
$$+ E(\sum_i \mathbf{W}_{ai}^{\frac{1}{2}}\mathbf{Z}_{*a}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\sum_j \mathbf{W}_{aj}\mathbf{l}'_*\mathbf{D}_a^2 f_{aj}\mathbf{l}_{*a}) \qquad \text{(F-93)}$$

Evaluation of $\gamma_{wa}\sigma_\beta^2$ gives

$$\gamma_{wa}\sigma_\beta^2 = E(\mathbf{U}'_{*a}\mathbf{V}_{*a}^{-\frac{1}{2}}(\mathbf{V}_{*a}^{\frac{1}{2}}\mathbf{W}_a^{\frac{1}{2}}\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}^{\frac{1}{2}} - \frac{\mathbf{P}_a\mathbf{P}'_a}{\mathbf{P}'_a\mathbf{P}_a})\mathbf{V}_{*a}^{-\frac{1}{2}}\mathbf{U}_{*a})$$
$$= tr(\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{V}_{*a} - \mathbf{W}_a^{-1})\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^2$$
$$= \frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{Q}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{V}_a - \mathbf{V}_{\omega a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{Q}_a)\sigma_\beta^2 \qquad \text{(F-94)}$$

and use of (F-71) yields

$$\gamma_{Ia}\sigma_\varepsilon^4 = E(S_a(\tilde{\theta}, \tilde{\theta}_p) - S_a(\hat{\theta}, \hat{\theta}_p) - \mathbf{U}'_{*a}\mathbf{W}_a^{\frac{1}{2}}\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{W}_a^{\frac{1}{2}}\mathbf{U}_{*a})$$
$$\approx \frac{1}{4}E(\sum_i \tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{ai}\tilde{\mathbf{l}}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{aj}\tilde{\mathbf{l}}_{*a})$$
$$-\frac{1}{2}E(\sum_i \tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{ai}\tilde{\mathbf{l}}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}\frac{\mathbf{Q}_a}{\mathbf{Q}'_a\mathbf{Q}_a}\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 h\tilde{\mathbf{l}}_{*a}) + \frac{1}{4}\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}E(\tilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 h\tilde{\mathbf{l}}_{*a})^2$$

$$
\begin{aligned}
&- E(\widetilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}} \widetilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 f_{aj} (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a^{\frac{1}{2}} (\mathbf{R}_a - \frac{\mathbf{Q}_a \mathbf{Q}'_a}{\mathbf{Q}'_a \mathbf{Q}_a}) \mathbf{W}_a^{\frac{1}{2}} \mathbf{D}_a \mathbf{f}_a (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \sum_i \mathbf{D}_a^2 f_{ai} \widetilde{\mathbf{l}}_{*a} \mathbf{W}_{ai}^{\frac{1}{2}} \widetilde{\mathbf{Z}}_a) \\
&+ E((\frac{1}{\mathbf{Q}'_a \mathbf{Q}_a} \mathbf{Q}'_a \mathbf{W}_a^{\frac{1}{2}} \mathbf{U}_{*a})^2 \widetilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 h (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a^{\frac{1}{2}} (\mathbf{R}_a - \frac{\mathbf{Q}_a \mathbf{Q}'_a}{\mathbf{Q}'_a \mathbf{Q}_a}) \mathbf{W}_a^{\frac{1}{2}} \mathbf{D}_a \mathbf{f}_a (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a^2 h \widetilde{\mathbf{l}}_{*a}) \\
&+ E(\widetilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj} \widetilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 f_{aj} (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a^{\frac{1}{2}} (\mathbf{R}_a - \frac{\mathbf{Q}_a \mathbf{Q}'_a}{\mathbf{Q}'_a \mathbf{Q}_a}) \sum_j \mathbf{W}_{aj}^{\frac{1}{2}} \widetilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 f_{aj} \widetilde{\mathbf{l}}_{*a}) \\
&+ \frac{1}{\mathbf{Q}'_a \mathbf{Q}_a} E(\widetilde{\mathbf{Z}}'_a \sum_j \mathbf{W}_{aj}^{\frac{1}{2}} \widetilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 f_{aj} (\mathbf{D}_a \mathbf{f}'_a \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a h' \widetilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 h \widetilde{\mathbf{l}}_{*a}) - \hat{\gamma}_{Ia} \sigma_\varepsilon^4
\end{aligned}
\qquad \text{(F-95)}
$$

Evaluation of the expected values in (F-93) and (F-95) is deferred to appendix G.

## Further Analysis of Distributions and Correction Factors

***Distribution of*** $S(\hat{\theta})$. The distributions of $S(\hat{\theta})$, $S(\theta) - S(\hat{\theta})$, and their ratio can be analyzed under somewhat less restrictive conditions than those required for the perturbation analyses. For $S(\hat{\theta})$ this is accomplished as follows. First, from (F-56) and (F-59), as an approximation,

$$\frac{S(\hat{\theta})}{E(S(\hat{\theta}))/(n-p)} \sim \chi^2(n-p) \qquad \text{(F-96)}$$

If (F-96) were a good approximation, then the mean and variance of the left-hand side of (F-96) would nearly equal the mean and variance expected for a $\chi^2(n-p)$ random variable, $n-p$ and $2(n-p)$, respectively. It is apparent that the means match exactly.

The variance of the variable in (F-96) is

$$Var\left(\frac{(n-p)S(\hat{\theta})}{E(S(\hat{\theta}))}\right) = \frac{(n-p)^2 Var(S(\hat{\theta}))}{(E(S(\hat{\theta})))^2} \qquad \text{(F-97)}$$

If model and system types of intrinsic nonlinearity are significant, it is difficult to compute the variance of $S(\hat{\theta})$ without using the Taylor series and perturbation expansions. Hence, these types of intrinsic nonlinearity are assumed to be negligible, which is the same as considering only the effect of $\omega^{-1} \neq \Omega$ on the variance. First, the residual vector $\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta})$ is expressed in terms of $\phi$, noting that the model in terms of $\phi$ is linear in the absence of model intrinsic nonlinearity. That is,

$$
\begin{aligned}
&\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}) = \mathbf{Y} - \mathbf{f}(\gamma\theta(\hat{\phi})) \\
&= \mathbf{Y} - \mathbf{f}(\gamma\theta(\phi_*)) - (\mathbf{f}(\gamma\theta(\hat{\phi})) - \mathbf{f}(\gamma\theta(\phi_*)))
\end{aligned}
$$

$$= \mathbf{Y} - \mathbf{f}(\gamma\theta(\phi_*)) - \mathbf{D}_\phi\mathbf{f}(\hat{\phi} - \phi_*)$$
$$= \mathbf{Y} - \mathbf{f}(\gamma\theta(\phi_*)) - \mathbf{D}_\phi\mathbf{f}(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega(\mathbf{Y} - \mathbf{f}(\gamma\theta(\phi_*)))$$
$$= \mathbf{Y} - \mathbf{f}(\gamma\theta(\phi_*)) - \mathbf{DfJ}(\mathbf{J}'\mathbf{Df}'\omega\mathbf{DfJ})^{-1}\mathbf{J}'\mathbf{Df}'\omega(\mathbf{Y} - \mathbf{f}(\gamma\theta(\phi_*)))$$
$$= \omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) \qquad \text{(F-98)}$$

where $\hat{\phi}$ and $\phi_*$ are equal to $\phi(\hat{\theta})$ and $\phi(\theta_*)$, respectively, and $\hat{\phi}$ was obtained by linear least squares using the linear model $\mathbf{f}(\gamma\theta(\phi))$. Thus, when model intrinsic nonlinearity is negligible

$$S(\hat{\theta}) \approx (\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) \qquad \text{(F-99)}$$

Next, error vector $\mathbf{Y} - \mathbf{f}(\gamma\theta_*)$ is written using (3-5) and (4-9) to get

$$\mathbf{Y} - \mathbf{f}(\gamma\theta_*) = \mathbf{U}_* + (\mathbf{Y} - \mathbf{U}_* - \mathbf{f}(\gamma\theta_*))$$
$$= \mathbf{U}_* + (\varepsilon + \mathbf{f}(\beta) - \varepsilon - \mathbf{D}_\beta\mathbf{f}(\mathbf{I} - \gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e} - \mathbf{f}(\gamma\theta_*))$$
$$= \mathbf{U}_* + \mathbf{f}(\beta) - \mathbf{f}_0(\beta) - \mathbf{f}(\gamma\theta_*) + \mathbf{f}_0(\gamma\theta_*)$$
$$= \mathbf{U}_* + \mathbf{d} \qquad \text{(F-100)}$$

where

$$\mathbf{f}_0(\beta) = \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\beta\mathbf{fe} \qquad \text{(F-101)}$$

$$\mathbf{f}_0(\gamma\theta_*) = \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\beta\mathbf{f}\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e} \qquad \text{(F-102)}$$

$$\mathbf{d} = \mathbf{f}(\beta) - \mathbf{f}_0(\beta) - \mathbf{f}(\gamma\theta_*) + \mathbf{f}_0(\gamma\theta_*)$$
$$\approx \frac{1}{2}\left[\mathbf{e}'(\mathbf{D}^2_\beta f_i - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_i\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}\right] \qquad \text{(F-103)}$$

If both model and system types of intrinsic nonlinearity are negligible, the variance of (F-99) can be evaluated in terms of $\mathbf{U}_*$ using (F-100). This was previously done using (4-38), which results from the Taylor series and perturbation expansions. It is done here without using these expansions by showing that the term $(\mathbf{I} - \mathbf{R})\omega^{1/2}\mathbf{d}$ directly reflects model and system types of intrinsic nonlinearity, as follows. The increment $\phi_* - \bar{\phi}$ in best transformation set $\phi$ can be written in terms of the $\theta$ set as $\mathbf{J}^{-1}(\theta_* - \bar{\theta} + \psi_*)$, where, given values of $\phi_*$, $\bar{\phi}$, $\theta_*$, $\bar{\theta}$, and $\mathbf{J}$, $\psi_*$ is uniquely defined. Next, a linear-model approximation $\mathbf{p}(\psi_*)$ of $\mathbf{f}(\gamma\theta_*)$ is written using $\phi$ as

$$\mathbf{p}(\psi_*) = \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\phi\mathbf{f}(\phi_* - \bar{\phi})$$
$$= \mathbf{f}(\gamma\bar{\theta}) + \mathbf{DfJ}(\phi_* - \bar{\phi})$$
$$= \mathbf{f}(\gamma\bar{\theta}) + \mathbf{Df}(\theta_* - \bar{\theta} + \psi_*)$$

$$= \mathbf{f}_0(\gamma\theta_*) + \mathbf{Df}\psi_* \tag{F-104}$$

and an analogous linear-model approximation of $\mathbf{f}(\beta)$ is written using $\alpha$ as

$$\begin{aligned}
\mathbf{p}_\alpha(\psi_\alpha) &= \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\alpha\mathbf{f}(\alpha - \bar{\alpha}) \\
&= \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\beta\mathbf{f}\mathbf{J}_\beta(\alpha - \bar{\alpha}) \\
&= \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\beta\mathbf{f}(\beta - \gamma\bar{\theta} + \zeta) \\
&\approx \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\beta\mathbf{f}(\beta - \gamma\bar{\theta} + \gamma\psi_\alpha) \\
&= \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\beta\mathbf{f}(\beta - \gamma\bar{\theta}) + \mathbf{Df}\psi_\alpha \\
&= \mathbf{f}_0(\beta) + \mathbf{Df}\psi_\alpha
\end{aligned} \tag{F-105}$$

where, given values of $\alpha$, $\bar{\alpha}$, $\beta$, $\gamma\bar{\theta}$, and $\mathbf{J}_\beta$, $\zeta$ is uniquely defined. The approximation $\zeta \approx \gamma\psi_\alpha$ is analogous to the approximation (C-27), appendix C. The best fit of $\mathbf{p}(\psi) = \mathbf{p}_\alpha(\psi_\alpha) - \mathbf{p}(\psi_*)$ to $\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*)$ is obtained by minimizing, with respect to $\psi = \psi_\alpha - \psi_*$, the objective function

$$S(\psi) = (\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*) - \mathbf{p}(\psi))'\omega(\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*) - \mathbf{p}(\psi)) \tag{F-106}$$

The result is

$$\psi = (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{d} \tag{F-107}$$

Thus,

$$\begin{aligned}
\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*) - \mathbf{p}(\psi) &= \mathbf{d} - \mathbf{Df}\psi \\
&= \omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}\mathbf{d}
\end{aligned} \tag{F-108}$$

If $\mathbf{p}(\psi)$ fits $\mathbf{f}(\beta) - \mathbf{f}(\gamma\theta_*)$ closely, then the linear model $\mathbf{p}(\psi) = \mathbf{f}_0(\beta) - \mathbf{f}_0(\gamma\theta_*) + \mathbf{Df}\psi$ is almost exact, indicating negligible intrinsic nonlinearity of $\mathbf{f}(\beta)$ (system intrinsic nonlinearity) and $\mathbf{f}(\gamma\theta_*)$ (model intrinsic nonlinearity). In this case (F-108) is nearly $\mathbf{0}$. From the preceding analysis, if model and system types of intrinsic nonlinearity are negligible, (F-99) written using (F-100) is approximated as

$$\begin{aligned}
S(\hat{\theta}) &\approx \mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}\mathbf{U}_* + 2\mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}\mathbf{d} + \mathbf{d}'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}\mathbf{d} \\
&\approx \mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\omega^{\frac{1}{2}}\mathbf{U}_*
\end{aligned} \tag{F-109}$$

which is the same relation found using the perturbation analysis.

Next, the variance of $S(\hat{\theta})$ as given by (F-109) is evaluated. The variance of the form $\mathbf{U}_*'\mathbf{A}\mathbf{U}_*$, where $\mathbf{A}$ is symmetric, is

$$Var(\mathbf{U}_*'\mathbf{A}\mathbf{U}_*) = E(\mathbf{U}_*'\mathbf{A}\mathbf{U}_*)^2 - (E(\mathbf{U}_*'\mathbf{A}\mathbf{U}_*))^2 \tag{F-110}$$

in which

$$E(\mathbf{U}_*'\mathbf{A}\mathbf{U}_*) = tr(\mathbf{A}\ Var(\mathbf{U}_*)) \tag{F-111}$$

$$E(\mathbf{U}_*'\mathbf{A}\mathbf{U}_*)^2 = tr^2(\mathbf{A}\ Var(\mathbf{U}_*)) + 2tr(\mathbf{A}\ Var(\mathbf{U}_*)\mathbf{A}\ Var(\mathbf{U}_*)) \tag{F-112}$$

where the result of appendix A was employed. Hence,

$$Var(\mathbf{U}_*'\mathbf{A}\mathbf{U}_*) = 2tr(\mathbf{A}\ Var(\mathbf{U}_*)\mathbf{A}\ Var(\mathbf{U}_*)) \tag{F-113}$$

Application of (F-113) to (F-109) using the definitions of $\mathbf{V}_*$ and $\boldsymbol{\Omega}$ implicit in (3-21) produces

$$\begin{aligned} Var(S(\hat{\theta})) &\approx 2tr(\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*)\sigma_\varepsilon^4 \\ &\leq 2tr(\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega})\sigma_\varepsilon^4 \end{aligned} \tag{F-114}$$

By definition

$$\Delta\sigma_\beta^2/\sigma_\varepsilon^2 = \boldsymbol{\Omega} - \boldsymbol{\omega}^{-1} \tag{F-115}$$

where $\Delta$ is analogous to $\mathbf{V}-\mathbf{V}_\omega$ in (F-61), and so depends only on model error. Then (F-114) becomes

$$\begin{aligned} Var(S(\hat{\theta})) &\approx 2tr(\mathbf{I}-\mathbf{R})\sigma_\varepsilon^4 + 4tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\Delta\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2 \\ &+ 2tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\Delta\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\Delta\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^4 \\ &= 2(n-p)\sigma_\varepsilon^4 + 4tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\Delta\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2 \\ &+ 2tr((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\Delta\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\Delta\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^4 \end{aligned} \tag{F-116}$$

Similarly,

$$(E(S(\hat{\theta})))^2 = tr^2((\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$= (n - p + tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2/\sigma_\varepsilon^2)^2\sigma_\varepsilon^4$$

$$= (n-p)^2\sigma_\varepsilon^4 + 2(n-p)tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2 + tr^2((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4 \quad \text{(F-117)}$$

so that

$$Var\left(\frac{(n-p)S(\hat{\theta})}{E(S(\hat{\theta}))}\right)$$

$$\approx \frac{(n-p)^2(2(n-p)\sigma_\varepsilon^4 + 4tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2 + 2tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4}{(n-p)^2\sigma_\varepsilon^4 + 2(n-p)tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2 + tr^2((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4}$$

$$= 2(n-p)\frac{(n-p)\sigma_\varepsilon^4 + 2tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2 + tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4}{(n-p)\sigma_\varepsilon^4 + 2tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2 + tr^2((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4/(n-p)} \quad \text{(F-118)}$$

Thus, through the second terms in the numerator and denominator, $Var((n-p)S(\hat{\theta})/E(S(\hat{\theta}))) = 2(n-p)$, which is the correct value for the variance.

It is instructive to evaluate the last terms in the numerator and denominator using approximation (5-22) (with $b$=1 for simplicity):

$$tr((\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}}(\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}})\sigma_\beta^4$$

$$= (tr((\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Omega\hat{\omega}^{\frac{1}{2}}(\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Omega\hat{\omega}^{\frac{1}{2}}) - 2tr((\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Omega\hat{\omega}^{\frac{1}{2}}) + tr(\mathbf{I}-\hat{\mathbf{R}}))\sigma_\varepsilon^4$$

$$\approx c^2(n-p)\sigma_\varepsilon^4 \quad \text{(F-119)}$$

and

$$tr^2((\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}})\sigma_\beta^4/(n-p)$$

$$= (tr((\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Omega\hat{\omega}^{\frac{1}{2}}) - tr(\mathbf{I}-\hat{\mathbf{R}}))^2\sigma_\varepsilon^4/(n-p)$$

$$\approx c^2(n-p)\sigma_\varepsilon^4 \quad \text{(F-120)}$$

Therefore, if the approximation is accurate and $\omega = \hat{\omega}$, the last two terms also are approximately equal.

To relate $c$ to relative sizes of model and observation errors, use is made of

$$\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}} = \left[\frac{(1-\delta_{ij})E(f_i(\beta)-f_i(\gamma\theta_*))(f_j(\beta)-f_j(\gamma\theta_*))}{(V_{\varepsilon ii}\sigma_\varepsilon^2 + E(f_i(\beta)-f_i(\gamma\theta_*))^2)^{\frac{1}{2}}(V_{\varepsilon jj}\sigma_\varepsilon^2 + E(f_j(\beta)-f_j(\gamma\theta_*))^2)^{\frac{1}{2}}}\right] \qquad \text{(F-121)}$$

Thus, if $(V_{\varepsilon ii})^{1/2}\sigma_\varepsilon$ is only twice $(E(f_i(\beta)-f_i(\gamma\theta_*))^2)^{1/2}$, then, even if correlations among the model errors are very large, an off-diagonal element of (F-121) is roughly only $1/((2^2+1)^{1/2}(2^2+1)^{1/2}) = 0.2$, so that $c^2 \approx (0.2)^2 = 0.04$. The last two terms in the numerator and denominator of (F-118) are not only approximately equal, in the present example their magnitudes are only about 4 percent of the magnitudes of the first terms.

The preceding analysis does not assume either model or observation errors to be small in magnitude. Thus, in general, the mean of $(n-p)S(\hat{\theta})/E(S(\hat{\theta}))$ equals the mean of a $\chi^2(n-p)$ random variable, and, if $\omega$ is set equal to $\hat{\omega}$, the variance of $(n-p)S(\hat{\theta})/E(S(\hat{\theta}))$ can be a good approximation of the variance of a $\chi^2(n-p)$ random variable if both model and system types of intrinsic nonlinearity are negligible.

***Distribution of*** $S(\theta_*)-S(\hat{\theta})$. From (F-63) and (F-55) in which $p_1 = p$ and $\theta = \theta_*$, as an approximation,

$$\frac{S(\theta_*)-S(\hat{\theta})}{E(S(\theta_*)-S(\hat{\theta}))/p} \sim \chi^2(p) \qquad \text{(F-122)}$$

As before, if (F-122) were a good approximation, then the mean and variance of the left-hand side would nearly equal the mean and variance, $p$ and $2p$, of a $\chi^2(p)$ random variable. It is apparent that the means are the same.

The variance of the variable in (F-122) is

$$Var\left(\frac{p(S(\theta_*)-S(\hat{\theta}))}{E(S(\theta_*)-S(\hat{\theta}))}\right) = \frac{p^2 Var(S(\theta_*)-S(\hat{\theta}))}{(E(S(\theta_*)-S(\hat{\theta})))^2} \qquad \text{(F-123)}$$

As before, both model and system types of intrinsic nonlinearity are assumed to be negligible, so that use of (F-99), (F-100), and (F-109) yields

$$\begin{aligned} S(\theta_*)-S(\hat{\theta}) &\approx (\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'\omega(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)) - (\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)) \\ &\approx (\mathbf{U}_*+\mathbf{d})'\omega(\mathbf{U}_*+\mathbf{d}) - \mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{U}_* \\ &= \mathbf{U}_*'\omega^{\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}\mathbf{U}_* + 2\mathbf{U}_*'\omega\mathbf{d} + \mathbf{d}'\omega\mathbf{d} \end{aligned} \qquad \text{(F-124)}$$

The last two terms are negligible only if total system nonlinearity (nonlinearity in $\mathbf{f}(\beta)$) is negligible. Hence, use of (F-103) produces

$$Var(S(\theta_*) - S(\hat{\theta})) \approx Var(\mathbf{U}'_*\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_* + 2\mathbf{U}'_*\boldsymbol{\omega}\mathbf{d} + \mathbf{d}'\boldsymbol{\omega}\mathbf{d})$$
$$\approx Var(\mathbf{U}'_*\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*) + 2Cov(\mathbf{U}'_*\boldsymbol{\omega}\mathbf{U}_*, \mathbf{d}'\boldsymbol{\omega}\mathbf{d})$$
$$+ 4Var(\mathbf{U}'_*\boldsymbol{\omega}\mathbf{d}) + Var(\mathbf{d}'\boldsymbol{\omega}\mathbf{d}) \qquad \text{(F-125)}$$

in which

$$Var(\mathbf{U}'_*\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{U}_*) = 2tr(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*)\sigma_\varepsilon^4$$
$$\leq 2tr(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega})\sigma_\varepsilon^4 \qquad \text{(F-126)}$$

Equations (F-125) and (F-126) combine to give

$$Var(S(\theta_*) - S(\hat{\theta})) \approx 2tr(\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega})\sigma_\varepsilon^4 + r \qquad \text{(F-127)}$$

where *r* signifies the terms left out, which from (G-7) have a leading value of order no lower than $\sigma_\varepsilon^2\sigma_\beta^4$. If model and system types of intrinsic nonlinearity are negligible, these terms result from total system nonlinearity, which would be expected to increase the variance, and from using $\boldsymbol{\Omega}$ in place of $\mathbf{V}_*$, which would decrease the variance. Remainder *r* represents the deviation of $\mathbf{Y} - \mathbf{f}(\gamma\theta_*)$ from a zero-mean, normally distributed random variable, because, if $\mathbf{Y} - \mathbf{f}(\gamma\theta_*)$ has a mean of zero and is normally distributed, use of the first line of (F-124) to compute the variance shows that $r = 0$.

Substitution of (F-115) into (F-127) yields

$$Var(S(\theta_*) - S(\hat{\theta})) \approx 2p\sigma_\varepsilon^4 + 4tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2$$
$$+ 2tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^4 + r \qquad \text{(F-128)}$$

Similarly, substitution of (F-115) into $(E(S(\theta_*) - S(\hat{\theta})))^2$ as evaluated using the first line of (F-124) produces

$$(E(S(\theta_*) - S(\hat{\theta})))^2 \approx tr^2(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Omega}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^4$$
$$= (p + tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^2/\sigma_\varepsilon^2)^2\sigma_\varepsilon^4$$
$$= p^2\sigma_\varepsilon^4 + 2ptr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2 + tr^2(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\beta^4 \qquad \text{(F-129)}$$

Thus,

$$Var\left(\frac{p(S(\theta_*)-S(\hat{\theta}))}{E(S(\theta_*)-S(\hat{\theta}))}\right)$$
$$\approx \frac{p^2(2p\sigma_\varepsilon^4+4tr(\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2+2tr(\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4+r)}{p^2\sigma_\varepsilon^4+2ptr(\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2+tr^2(\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4}$$
$$=2p\frac{p\sigma_\varepsilon^4+2tr(\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2+tr(\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4+r/2}{p\sigma_\varepsilon^4+2tr(\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^2\sigma_\varepsilon^2+tr^2(\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4/p} \quad \text{(F-130)}$$

Again, through the second terms in the numerator and denominator, $Var(p(S(\theta_*)-S(\hat{\theta}))/E(S(\theta_*)-S(\hat{\theta})))=2p$, which is the correct value for the variance.

The third terms in the numerator and denominator can be evaluated using approximation (5-21) to get

$$tr(\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}}\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}})\sigma_\beta^4$$
$$=(tr(\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Omega\hat{\omega}^{\frac{1}{2}}\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Omega\hat{\omega}^{\frac{1}{2}})-2tr(\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Omega\hat{\omega}^{\frac{1}{2}})+tr(\hat{\mathbf{R}}))\sigma_\varepsilon^4$$
$$\approx c^2((n-p)^2+n^2(p-1))\sigma_\varepsilon^4/p \quad \text{(F-131)}$$

and

$$tr^2(\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}})\sigma_\beta^4/p$$
$$=(tr(\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Omega\hat{\omega}^{\frac{1}{2}})-tr(\hat{\mathbf{R}}))^2\sigma_\varepsilon^4/p$$
$$\approx c^2(n-p)^2\sigma_\varepsilon^4/p \quad \text{(F-132)}$$

The term represented by (F-131) is equal to or larger than the term represented by (F-132), with equality occurring only for $p = 1$.

As before, the preceding analysis does not assume small model or observation errors. Thus, the mean of $p(S(\theta_*)-S(\hat{\theta}))/E(S(\theta_*)-S(\hat{\theta}))$ equals the mean of a $\chi^2(p)$ random variable. However, the variance of $p(S(\theta_*)-S(\hat{\theta}))/E(S(\theta_*)-S(\hat{\theta}))$ could be larger than the variance of a $\chi^2(p)$ random variable, even if $\omega$ is set equal to $\hat{\omega}$. The $r$ terms would probably increase in magnitude as $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ becomes progressively more non-normal.

***Redefinition of component correction factors.*** Based on the perturbation analysis, $E(S(\hat{\theta}))$ and $E(S(\theta_*)-S(\hat{\theta}))$ are expressed in terms of sums of correction factors in (F-59) and (F-63), respectively. The present analysis yields $S(\hat{\theta})$ and $S(\theta_*)-S(\hat{\theta})$ in the absence of model intrinsic nonlinearity as (F-99) and the first line of (F-124), respectively. These latter equations indicate that the component correction factors might be more accurate for practical use if they are redefined in terms of $\mathbf{\Omega}$ rather than $\mathbf{V}_*$ as follows.

$$\hat{\gamma}_w\sigma_\beta^2 = E(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*)) - (n-p)\sigma_\varepsilon^2$$

$$= (tr((\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{\Omega}\boldsymbol{\omega}^{\frac{1}{2}}) - n + p)\sigma_\varepsilon^2 \tag{F-133}$$

$$\hat{\gamma}_I\sigma_\varepsilon^4 = E(S(\hat{\boldsymbol{\theta}}) - (\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*)))$$

$$= E(S(\hat{\boldsymbol{\theta}})) - tr((\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{\Omega}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2 \tag{F-134}$$

$$\gamma_w\sigma_\beta^2 = E(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*)) - p\sigma_\varepsilon^2$$

$$= (tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{\Omega}\boldsymbol{\omega}^{\frac{1}{2}}) - p)\sigma_\varepsilon^2 \tag{F-135}$$

$$\gamma_I\sigma_\varepsilon^4 = E(S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}}) - (\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*))'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\boldsymbol{\theta}_*)))$$

$$= E(S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}})) - tr(\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{\Omega}\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^2 \tag{F-136}$$

Redefined factors $\hat{\gamma}_w\sigma_\beta^2$ and $\gamma_w\sigma_\beta^2$ both measure only the influence of $\boldsymbol{\omega}^{-1} \neq \mathbf{\Omega}$ on the distributions of $S(\hat{\boldsymbol{\theta}})$ and $S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}})$. In addition (F-99) and (F-124) show that both $\hat{\gamma}_I\sigma_\varepsilon^4$ and $\gamma_I\sigma_\varepsilon^4$ are zero in the absence of model intrinsic nonlinearity, so both measure model intrinsic nonlinearity. Because $E(S(\boldsymbol{\theta}_*)) = tr(\boldsymbol{\omega}^{1/2}\mathbf{\Omega}\boldsymbol{\omega}^{1/2})\sigma_\varepsilon^2$, $\gamma_I\sigma_\varepsilon^4 = -\hat{\gamma}_I\sigma_\varepsilon^4$, as also was found from the perturbation analysis. Finally, as before, $E(S(\hat{\boldsymbol{\theta}})) = (n-p)\sigma_\varepsilon^2 + \hat{\gamma}_w\sigma_\beta^2 + \hat{\gamma}_I\sigma_\varepsilon^4$ and $E(S(\boldsymbol{\theta}_*) - S(\hat{\boldsymbol{\theta}})) = p\sigma_\varepsilon^2 + \gamma_w\sigma_\beta^2 + \gamma_I\sigma_\varepsilon^4$. Thus, the redefinitions do not change any critical properties of the correction factors.

***Distribution of*** $S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}})$. The last approximate $\chi^2$ distribution to be examined is obtained using (F-55) and (F-66) with $p_1 = 1$ and $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$:

$$\frac{S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}})}{E(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}))} \sim \chi^2(1) \tag{F-137}$$

For (F-137) to be a good approximation, the mean and variance of the left-hand side should be approximately 1 and 2, respectively. The mean is correct.

The variance of the left-hand side of (F-137) is

$$Var\left(\frac{S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}})}{E(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}))}\right) = \frac{Var(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}))}{(E(S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}})))^2} \tag{F-138}$$

The model combined intrinsic nonlinearity is considered to be negligible so that $S(\tilde{\theta})$ can be obtained from the constrained regression using linear functions $\mathbf{f}(\gamma\theta(\phi))$ and $g(\gamma\theta(\phi))$ written as

$$\mathbf{f}(\gamma\theta(\phi)) = \mathbf{f}(\gamma\theta_*) + \mathbf{D}_\phi \mathbf{f}(\phi - \phi_*) \tag{F-139}$$

and

$$g(\gamma\theta(\phi)) = g(\gamma\theta_*) + \mathbf{D}_\phi g(\phi - \phi_*) \tag{F-140}$$

Solution of the Lagrange multiplier problem for the constrained regression estimates $\tilde{\phi}$ is obtained by substituting (F-139) and (F-140) into (E-7), then minimizing $L(\phi,\lambda)$ in the same manner as used to obtain $\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}$ in (E-10)-(E-14). The result is

$$\tilde{\phi} - \phi_* = (\mathbf{D}_\phi \mathbf{f}'\omega \mathbf{D}_\phi \mathbf{f})^{-1} \mathbf{D}_\phi \mathbf{f}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) \tag{F-141}$$

from which

$$\begin{aligned}
\mathbf{Y} - \mathbf{f}(\gamma\tilde{\theta}) &= \mathbf{Y} - \mathbf{f}(\gamma\theta(\tilde{\phi})) \\
&= \mathbf{Y} - \mathbf{f}(\gamma\theta(\phi_*)) - (\mathbf{f}(\gamma\theta(\tilde{\phi})) - \mathbf{f}(\gamma\theta(\phi_*))) \\
&= \mathbf{Y} - \mathbf{f}(\gamma\theta(\phi_*)) - \mathbf{D}_\phi \mathbf{f}(\tilde{\phi} - \phi_*) \\
&= \mathbf{Y} - \mathbf{f}(\gamma\theta_*) - \mathbf{D}_\phi \mathbf{f}(\mathbf{D}_\phi \mathbf{f}'\omega \mathbf{D}_\phi \mathbf{f})^{-1} \mathbf{D}_\phi \mathbf{f}'\omega^{\frac{1}{2}}(\mathbf{R} - \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) \\
&= \omega^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\theta_*))
\end{aligned} \tag{F-142}$$

Therefore,

$$S(\tilde{\theta}) \approx (\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'\omega^{\frac{1}{2}}(\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) \tag{F-143}$$

and

$$\begin{aligned}
S(\tilde{\theta}) - S(\hat{\theta}) &\approx (\mathbf{Y} - \mathbf{f}(\gamma\theta_*))'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{Y} - \mathbf{f}(\gamma\theta_*)) \\
&= \mathbf{U}_*'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{U}_* + 2\mathbf{U}_*'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{d} + \mathbf{d}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{d}
\end{aligned} \tag{F-144}$$

Terms expressing model and system types of intrinsic nonlinearity cancel in (F-144), but the two terms containing **d** do not express either model or system combined intrinsic nonlinearity. Thus, these terms cannot be neglected, even when model and system types of combined intrinsic nonlinearity are small.

Evaluation of the variance of (F-144) and subsequently (F-138) is completely analogous to the evaluation given by (F-125)-(F-130) if $\omega^{1/2}(\mathbf{QQ}'/\mathbf{Q}'\mathbf{Q})\omega^{1/2}$ is substituted for either $\omega^{1/2}\mathbf{R}\omega^{1/2}$ or $\omega$, as appropriate. The result is

$$Var\left(\frac{S(\tilde{\theta})-S(\hat{\theta})}{E(S(\tilde{\theta})-S(\hat{\theta}))}\right)$$

$$\approx 2\frac{\sigma_\varepsilon^4+2tr(\mathbf{QQ}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}/\mathbf{Q}'\mathbf{Q})\sigma_\beta^2\sigma_\varepsilon^2+tr(\mathbf{QQ}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\mathbf{QQ}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}/(\mathbf{Q}'\mathbf{Q})^2)\sigma_\beta^4+r/2}{\sigma_\varepsilon^4+2tr(\mathbf{QQ}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}/\mathbf{Q}'\mathbf{Q})\sigma_\beta^2\sigma_\varepsilon^2+tr^2(\mathbf{QQ}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}/\mathbf{Q}'\mathbf{Q})\sigma_\beta^4}$$

$$=2\frac{\sigma_\varepsilon^4+(2\mathbf{Q}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\mathbf{Q}/\mathbf{Q}'\mathbf{Q})\sigma_\beta^2\sigma_\varepsilon^2+(\mathbf{Q}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\mathbf{Q}/\mathbf{Q}'\mathbf{Q})^2\sigma_\beta^4+r/2}{\sigma_\varepsilon^4+(2\mathbf{Q}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\mathbf{Q}/\mathbf{Q}'\mathbf{Q})\sigma_\beta^2\sigma_\varepsilon^2+(\mathbf{Q}'\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\mathbf{Q}/\mathbf{Q}'\mathbf{Q})^2\sigma_\beta^4} \qquad \text{(F-145)}$$

This time, the variance is correct, that is, $Var((S(\tilde{\theta})-S(\hat{\theta}))/E(S(\tilde{\theta})-S(\hat{\theta})))=2$, through all terms except $r/2$. As before, the $r/2$ terms would be expected to increase in magnitude as $\mathbf{Y}-\mathbf{f}(\gamma\theta_*)$ becomes progressively more non-normal. Although the $r/2$ terms are not equal to the $r/2$ terms defined previously, they have the same lowest possible leading order of $\sigma_\varepsilon^2\sigma_\beta^4$.

***Redefinition of component correction factors.*** As before, the component correction factors for $S(\tilde{\theta})-S(\hat{\theta})$ are probably more accurate for practical use if they are redefined in terms of $\Omega$ rather than $\mathbf{V}_*$ as follows.

$$\gamma_w\sigma_\beta^2=E(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))-\sigma_\varepsilon^2$$

$$=(\mathbf{Q}'\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}}\mathbf{Q}/\mathbf{Q}'\mathbf{Q}-1)\sigma_\varepsilon^2 \qquad \text{(F-146)}$$

$$\gamma_I\sigma_\varepsilon^4=E(S(\tilde{\theta})-S(\hat{\theta})-(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)))$$

$$=E(S(\tilde{\theta})-S(\hat{\theta}))-\mathbf{Q}'\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}}\mathbf{Q}/\mathbf{Q}'\mathbf{Q}\sigma_\varepsilon^2 \qquad \text{(F-147)}$$

Factor $\gamma_w\sigma_\beta^2$ measures only the influence of $\omega^{-1}\neq\Omega$ on the distribution of $S(\tilde{\theta})-S(\hat{\theta})$, and factor $\gamma_I\sigma_\varepsilon^4$, which is zero for zero model combined intrinsic nonlinearity, measures the influence of model combined intrinsic nonlinearity on the distribution of $S(\tilde{\theta})-S(\hat{\theta})$. As can be seen, $E(S(\tilde{\theta})-S(\hat{\theta}))=\sigma_\beta^2+\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4$.

***Distribution of ratio***. The final analysis is of the ratio $((S(\theta)-S(\hat{\theta}))/p_1)/(S(\hat{\theta})/(n-p))$. This ratio has an $F(p_1, n-p)$ distribution if the two component random variables are $\chi^2$-distributed and independent. If $\chi^2(n-p)$ and $\chi^2(p_1)$ are independent, then by the chi-square summation theorem (Miller and Kahn, 1962, p. 463) the sum $\chi^2(n-p)+\chi^2(p_1)$ is a $\chi^2(n-p+p_1)$ random variable. Thus, the mean and variance of the sum of $(n-p)S(\hat{\theta})/E(S(\hat{\theta}))$ and $p_1(S(\theta)-S(\hat{\theta}))/E(S(\theta)-S(\hat{\theta}))$ may be checked to see if they have approximate values of $n-p+p_1$ and $2(n-p+p_1)$, as expected. The mean is

$$E\left(\frac{(n-p)S(\hat{\theta})}{E(S(\hat{\theta}))}+\frac{p_1(S(\theta)-S(\hat{\theta}))}{E(S(\theta)-S(\hat{\theta}))}\right)=n-p+p_1 \tag{F-148}$$

which is as expected.

The variance is

$$\begin{aligned}&Var\left(\frac{(n-p)S(\hat{\theta})}{E(S(\hat{\theta}))}+\frac{p_1(S(\theta)-S(\hat{\theta}))}{E(S(\theta)-S(\hat{\theta}))}\right)\\&=Var\left(\frac{(n-p)S(\hat{\theta})}{E(S(\hat{\theta}))}\right)+2Cov\left(\frac{(n-p)S(\hat{\theta})}{E(S(\hat{\theta}))},\ \frac{p_1(S(\theta)-S(\hat{\theta}))}{E(S(\theta)-S(\hat{\theta}))}\right)\\&+Var\left(\frac{p_1(S(\theta)-S(\hat{\theta}))}{E(S(\theta)-S(\hat{\theta}))}\right)\end{aligned} \tag{F-149}$$

If the covariance term is nearly zero, the variance is equal to $2(n-p+p_1)$ + *error terms*. The error terms already have been analyzed. The covariance term is analyzed in the same manner as used for the variance of $S(\hat{\theta})$ and $S(\theta)-S(\hat{\theta})$. First,

$$\begin{aligned}&Cov\left(\frac{(n-p)S(\hat{\theta})}{E(S(\hat{\theta}))},\ \frac{p_1(S(\theta)-S(\hat{\theta}))}{E(S(\theta)-S(\hat{\theta}))}\right)\\&=\frac{(n-p)p_1Cov(S(\hat{\theta}),\ S(\theta)-S(\hat{\theta}))}{E(S(\hat{\theta}))E(S(\theta)-S(\hat{\theta}))}\end{aligned} \tag{F-150}$$

Then, use of a derivation like the one used in (F-110)-(F-113) shows that for symmetric matrices **A** and **B**,

$$Cov(\mathbf{U}_*'\mathbf{A}\mathbf{U}_*,\ \mathbf{U}_*'\mathbf{B}\mathbf{U}_*)=2tr(\mathbf{A}Var(\mathbf{U}_*)\mathbf{B}Var(\mathbf{U}_*)) \tag{F-151}$$

For $\theta=\theta_*$ and negligible model and system types of intrinsic nonlinearity, (F-109), (F-115), (F-124), and (F-151) yield

$$Cov(S(\hat{\theta}),\ S(\theta_*)-S(\hat{\theta}))\approx Cov(\mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{U}_*, \mathbf{U}_*'\omega^{\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}\mathbf{U}_*+2\mathbf{U}_*'\omega\mathbf{d}+\mathbf{d}'\omega\mathbf{d})$$

$$= 2tr(\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}\mathbf{V}_*)\sigma_\varepsilon^4 + Cov(\mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{U}_*,\ \mathbf{d}'\omega\mathbf{d})$$
$$= 2tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}})\sigma_\varepsilon^4 + r$$
$$= 2tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4 + r \tag{F-152}$$

where $r$ signifies the remaining terms. Approximation (5-21) gives

$$tr((\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}}\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}})\sigma_\beta^4$$
$$\approx tr((1-c)(\mathbf{I}-\hat{\mathbf{R}})\hat{\mathbf{R}}\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}})\sigma_\beta^4$$
$$= 0 \tag{F-153}$$

For $\theta = \tilde{\theta}$, negligible model combined intrinsic nonlinearity, and negligible model and system types of intrinsic nonlinearity, a development similar to (F-152) yields

$$Cov(S(\hat{\theta}),\ S(\tilde{\theta}) - S(\hat{\theta})) \approx Cov(\mathbf{U}_*'\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{U}_*, \mathbf{U}_*'\omega^{\frac{1}{2}}\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{U}_*$$
$$+ 2\mathbf{U}_*'\omega^{\frac{1}{2}}\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{d} + \mathbf{d}'\omega^{\frac{1}{2}}\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\mathbf{d})$$
$$= 2tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}}\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\Omega\omega^{\frac{1}{2}})\sigma_\varepsilon^4 + r$$
$$= 2tr((\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}}\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}\Delta\omega^{\frac{1}{2}})\sigma_\beta^4 + r \tag{F-154}$$

Now, approximation (5-21) results in

$$tr((\mathbf{I}-\hat{\mathbf{R}})\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}}\frac{\hat{\mathbf{Q}}\hat{\mathbf{Q}}'}{\hat{\mathbf{Q}}'\hat{\mathbf{Q}}}\hat{\omega}^{\frac{1}{2}}\Delta\hat{\omega}^{\frac{1}{2}})\sigma_\beta^4 \approx 0 \tag{F-155}$$

Thus, from (5-21), (F-109), (F-150), and (F-155), when $\omega = \hat{\omega}$ the covariance term is approximately given by

$$\frac{(n-p)p_1 r}{E(S(\hat{\theta}))E(S(\theta) - S(\hat{\theta}))} \approx \frac{p_1 r}{(1-c)E(S(\theta) - S(\hat{\theta}))\sigma_\varepsilon^2} \tag{F-156}$$

where, for the expected value in the denominator, $E(S(\theta_*) - S(\hat{\theta})) = p\sigma_\varepsilon^2 + tr(\hat{\mathbf{R}}\hat{\omega}^{1/2}\Delta\hat{\omega}^{1/2})\sigma_\beta^2$ and $E(S(\tilde{\theta}) - S(\hat{\theta})) = \sigma_\varepsilon^2 + (\hat{\mathbf{Q}}'\hat{\omega}^{1/2}\Delta\hat{\omega}^{1/2}\hat{\mathbf{Q}}/\hat{\mathbf{Q}}'\hat{\mathbf{Q}})\sigma_\beta^2$. Unless $r$ is significant, the covariance term can be small, so that, if $\omega = \hat{\omega}$, the sum of $(n-p)S(\hat{\theta})/E(S(\hat{\theta}))$ and

$p_1(S(\theta)-S(\hat{\theta}))/E(S(\theta)-S(\hat{\theta}))$ can have a mean and variance approximately the same as the mean and variance of a $\chi^2(n-p+p_1)$ random variable.

The analysis for prediction intervals using augmented variables is analogous to the preceding analysis given by (F-96)-(F-121) and (F-137)-(F-156). However, as suggested by the discussion leading to (5-97), the approximations are more accurate than the approximations in the preceding analysis. The development is not repeated here.

***An alternative noncentral chi-square approximation.*** As a final note, the last two terms of both (F-124) and (F-144) are nonzero because **d** is nonzero. This also causes $E(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))$ to be nonzero, as can be seen from (F-103) and (3-31). If **d** were approximated as a nearly normal random variable, then for $\mathbf{V}=Var(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))$

$$(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'\mathbf{V}^{-\frac{1}{2}}\mathbf{H}\mathbf{V}^{-\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))\sim\chi^2(p,\rho) \tag{F-157}$$

where $\mathbf{H}$ is computed using $\mathbf{V}$ instead of $\mathbf{V}_*$ and $\chi^2(p,\rho)$ is a noncentral $\chi^2$ random variable having $p$ degrees of freedom and noncentrality parameter $\rho$ (Graybill, 1976, p. 125). Parameter $\rho$ is defined by

$$\rho=\frac{1}{2}E(\mathbf{d})'\mathbf{V}^{-\frac{1}{2}}\mathbf{H}\mathbf{V}^{-\frac{1}{2}}E(\mathbf{d}) \tag{F-158}$$

Now

$$\begin{aligned}&E(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'\mathbf{V}^{-\frac{1}{2}}\mathbf{H}\mathbf{V}^{-\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))\\&=(p+2\rho)\sigma_\varepsilon^2\end{aligned} \tag{F-159}$$

so that, as an approximation,

$$\frac{S(\theta_*)-S(\hat{\theta})}{E(S(\theta_*)-S(\hat{\theta}))/(p+2\rho)}\sim\chi^2(p,\rho) \tag{F-160}$$

for which, as in (F-63), correction factors are defined from

$$E(S(\theta_*)-S(\hat{\theta}))=(p+2\rho)\sigma_\varepsilon^2+\gamma_w\sigma_\beta^2+\gamma_I\sigma_\varepsilon^4 \tag{F-161}$$

as $\gamma_w\sigma_\beta^2=E(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'(\omega^{1/2}\mathbf{R}\omega^{1/2}-\mathbf{V}^{-1/2}\mathbf{H}\mathbf{V}^{-1/2})(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))$ and $\gamma_I\sigma_\varepsilon^4=E(S(\theta_*)-S(\hat{\theta})-(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))'\omega^{1/2}\mathbf{R}\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*)))$, the latter of which is the same as (6-14). A similar approach also could be used for $S(\tilde{\theta})-S(\hat{\theta})$. Both distributions could be analyzed using the procedures used to analyze (F-122) and (F-137). The ratios of the approximate noncentral $\chi^2$ random variables, such as (F-160), to the approximate central $\chi^2$ random variable given by (F-96) lead to noncentral $F$ distributions to define confidence regions and intervals (Graybill, 1976, p. 128). Experiments indicate that this approach does not improve accuracy compared to the approach followed for the two examples considered in section 7. Also, it is doubtful that the noncentrality parameters could be estimated using the information usually available. For these reasons the analyses are not pursued further in this report.

# Appendix G – Evaluation of Component Correction Factors for Model Intrinsic Nonlinearity

## Component Correction Factor Pertaining to Confidence Regions

Component correction factor $\hat{\gamma}_I\sigma_\varepsilon^4$ is given by (F-62), appendix F, as

$$\hat{\gamma}_I\sigma_\varepsilon^4 \approx \frac{1}{4}E(\sum_i \sum_j \mathbf{l}_*'\mathbf{D}^2 f_i \mathbf{l}_* \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}}\mathbf{l}_*'\mathbf{D}^2 f_j \mathbf{l}_*) - E(\sum_i \sum_k \omega_i^{\frac{1}{2}}\mathbf{Z}_*\mathbf{l}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_k \mathbf{l}_*\omega_k^{\frac{1}{2}}\mathbf{Z}_*)$$
$$+ E(\sum_i \sum_j \omega_i^{\frac{1}{2}}\mathbf{Z}_*\mathbf{l}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega_j \mathbf{l}_*'\mathbf{D}^2 f_j \mathbf{l}_*) \quad \text{(G-1)}$$

Equation (G-1) is evaluated term-by-term using the result of appendix A. For the first term

$$E(\sum_i \sum_j \mathbf{l}_*'\mathbf{D}^2 f_i \mathbf{l}_* \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}}\mathbf{l}_*'\mathbf{D}^2 f_j \mathbf{l}_*)$$
$$= \sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}} E(\mathbf{l}_*'\mathbf{D}^2 f_i \mathbf{l}_*)(\mathbf{l}_*'\mathbf{D}^2 f_j \mathbf{l}_*)$$
$$= \sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}} E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{V}_*^{\frac{1}{2}}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)$$
$$\bullet(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{V}_*^{\frac{1}{2}}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)$$
$$= \sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}} tr(\mathbf{V}_*^{\frac{1}{2}}\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{V}_*^{\frac{1}{2}})$$
$$\bullet tr(\mathbf{V}_*^{\frac{1}{2}}\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{V}_*^{\frac{1}{2}})\sigma_\varepsilon^4 + 2\sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}} tr(\mathbf{V}_*^{\frac{1}{2}}\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}$$
$$\bullet \mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{V}_*\omega\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{V}_*^{\frac{1}{2}})\sigma_\varepsilon^4$$
$$= \sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}} tr(\mathbf{A}_i)tr(\mathbf{A}_j)\sigma_\varepsilon^4 + 2\sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}} tr(\mathbf{A}_i\mathbf{A}_j)\sigma_\varepsilon^4$$
$$= (\sum_j (\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}} tr(\mathbf{A}_j))'\sum_j (\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}} tr(\mathbf{A}_j)\sigma_\varepsilon^4 + 2\sum_j \sum_\ell ((\mathbf{I}-\mathbf{R})\omega_j^{\frac{1}{2}})'(\mathbf{I}-\mathbf{R})\omega_\ell^{\frac{1}{2}} tr(\mathbf{A}_j\mathbf{A}_\ell)\sigma_\varepsilon^4$$
$$= (\sum_i \sum_j (\mathbf{I}-\mathbf{R})_i \omega_j^{\frac{1}{2}} tr(\mathbf{A}_j) \sum_j (\mathbf{I}-\mathbf{R})_i \omega_j^{\frac{1}{2}} tr(\mathbf{A}_j)\sigma_\varepsilon^4 + 2\sum_i \sum_j \sum_\ell tr((\mathbf{I}-\mathbf{R})_i \omega_j^{\frac{1}{2}}\mathbf{A}_j(\mathbf{I}-\mathbf{R})_i \omega_\ell^{\frac{1}{2}}\mathbf{A}_\ell)\sigma_\varepsilon^4$$
$$= \sum_i tr^2(\mathbf{C}_i\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4 + 2\sum_i tr(\mathbf{C}_i\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\mathbf{C}_i\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4 \quad \text{(G-2)}$$

where temporary variable $\mathbf{A}_i$ is defined implicitly, and

$$\mathbf{C}_i = (\mathbf{I}-\mathbf{R})_i \sum_j \omega_j^{\frac{1}{2}} \omega^{\frac{1}{2}} \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}} \tag{G-3}$$

For the second term

$$E(\sum_i \sum_k \omega_i^{\frac{1}{2}} \mathbf{Z}_* \mathbf{l}_*' \mathbf{D}^2 f_i (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_k \mathbf{l}_* \omega_k^{\frac{1}{2}} \mathbf{Z}_*)$$

$$= \sum_j \sum_k E(\mathbf{Z}_*' \omega_j^{\frac{1}{2}} \omega_k^{\frac{1}{2}} \mathbf{Z}_*)(\mathbf{l}_*' \mathbf{D}^2 f_j (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_k \mathbf{l}_*)$$

$$= \sum_j \sum_k E(\mathbf{U}_*' \mathbf{V}_*^{-\frac{1}{2}} \mathbf{V}_*^{\frac{1}{2}} \omega^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega_j^{\frac{1}{2}} \omega_k^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega^{\frac{1}{2}} \mathbf{V}_*^{\frac{1}{2}} \mathbf{V}_*^{-\frac{1}{2}} \mathbf{U}_*)$$

$$\bullet (\mathbf{U}_*' \mathbf{V}_*^{-\frac{1}{2}} \mathbf{V}_*^{\frac{1}{2}} \omega \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_k (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega \mathbf{V}_*^{\frac{1}{2}} \mathbf{V}_*^{-\frac{1}{2}} \mathbf{U}_*)$$

$$= \sum_j \sum_k tr(\mathbf{V}_*^{\frac{1}{2}} \omega^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega_j^{\frac{1}{2}} \omega_k^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega^{\frac{1}{2}} \mathbf{V}_*^{\frac{1}{2}}) tr(\mathbf{V}_*^{\frac{1}{2}} \omega \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_k (\mathbf{Df}'\omega\mathbf{Df})^{-1}$$

$$\bullet \mathbf{Df}'\omega \mathbf{V}_*^{\frac{1}{2}})\sigma_\varepsilon^4 + 2\sum_j \sum_k tr(\mathbf{V}_*^{\frac{1}{2}} \omega^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega_j^{\frac{1}{2}} \omega_k^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega^{\frac{1}{2}} \mathbf{V}_* \omega \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j (\mathbf{Df}'\omega\mathbf{Df})^{-1}$$

$$\bullet \mathbf{D}^2 f_k (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega \mathbf{V}_*^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$= \sum_j \sum_k \omega_k^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega_j^{\frac{1}{2}} tr(\mathbf{B}_j \mathbf{B}_k \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+ 2\sum_j \sum_k \omega_k^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}} \mathbf{B}_j \mathbf{B}_k \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega_j^{\frac{1}{2}} \sigma_\varepsilon^4 \tag{G-4}$$

where

$$\mathbf{B}_j = \omega^{\frac{1}{2}} \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}} \tag{G-5}$$

For the third term

$$E(\sum_i \sum_j \omega_i^{\frac{1}{2}} \mathbf{Z}_* \mathbf{l}_*' \mathbf{D}^2 f_i (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega_j \mathbf{l}_*' \mathbf{D}^2 f_j \mathbf{l}_*)$$

$$= \sum_i \sum_j E(\mathbf{l}_*' \mathbf{D}^2 f_i (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega_j \omega_i^{\frac{1}{2}} \mathbf{Z}_*)(\mathbf{l}_*' \mathbf{D}^2 f_j \mathbf{l}_*)$$

$$= \sum_i \sum_j E(\mathbf{U}_*' \mathbf{V}_*^{-\frac{1}{2}} \mathbf{V}_*^{\frac{1}{2}} \omega \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}} \omega_j^{\frac{1}{2}} \omega_i^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega^{\frac{1}{2}} \mathbf{V}_*^{\frac{1}{2}} \mathbf{V}_*^{-\frac{1}{2}} \mathbf{U}_*)$$

$$\bullet (\mathbf{U}_*' \mathbf{V}_*^{-\frac{1}{2}} \mathbf{V}_*^{\frac{1}{2}} \omega \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j (\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega \mathbf{V}_*^{\frac{1}{2}} \mathbf{V}_*^{-\frac{1}{2}} \mathbf{U}_*)$$

$$= \sum_i \sum_j tr(\mathbf{B}_i \omega_j^{\frac{1}{2}} \omega_i^{\frac{1}{2}} (\mathbf{I}-\mathbf{R}) \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}}) tr(\mathbf{B}_j \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+2\sum_i\sum_j tr(\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}\mathbf{B}_i\omega_j^{\frac{1}{2}}\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\mathbf{B}_j\omega^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$=\sum_i\sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\mathbf{B}_i\omega_j^{\frac{1}{2}}tr(\mathbf{B}_j\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+2\sum_i\sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\mathbf{B}_j\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\mathbf{B}_i\omega_j^{\frac{1}{2}}\sigma_\varepsilon^4 \qquad \text{(G-6)}$$

Matrix $\mathbf{V}_*$ can be replaced with $\Omega$, as can be deduced using (3-23), (3-33), (5-8) and the fact that

$$E(Y_i - f_i(\gamma\theta_*))(Y_j - f_j(\gamma\theta_*))$$
$$= E(Y_i - f_i(\beta) + f_i(\beta) - f_i(\gamma\theta_*))(Y_j - f_j(\beta) + f_j(\beta) - f_j(\gamma\theta_*))$$
$$= E(\varepsilon_i\varepsilon_j) + E(f_i(\beta) - f_i(\gamma\theta_*))(f_j(\beta) - f_j(\gamma\theta_*))$$

to yield

$$\Omega = \mathbf{V}_* + O(\sigma_\beta^4/\sigma_\varepsilon^2) \qquad \text{(G-7)}$$

where $O(\sigma_\beta^4/\sigma_\varepsilon^2)$ signifies terms of order $\sigma_\beta^4/\sigma_\varepsilon^2$. Use of (G-7) shows that replacement of $\omega^{1/2}\mathbf{V}_*\omega^{1/2}$ with $\omega^{1/2}\Omega\omega^{1/2}$ in (G-2), (G-4), and (G-6) involves dropping terms of order $\sigma_\beta^4\sigma_\varepsilon^2$, which are of higher order than kept in the perturbation analysis.

## Component Correction Factor Pertaining to Individual Confidence Intervals

Component correction factor $\gamma_I\sigma_\varepsilon^4$ is given by (F-68), appendix F, as

$$\gamma_I\sigma_\varepsilon^4 \approx \frac{1}{4}E(\sum_i\sum_j \tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i\tilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j\tilde{\mathbf{l}}_*)$$

$$-\frac{1}{2}E(\sum_i \tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i\tilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 g\tilde{\mathbf{l}}_*)+\frac{1}{4}\frac{1}{\mathbf{Q}'\mathbf{Q}}E(\tilde{\mathbf{l}}_*'\mathbf{D}^2 g\tilde{\mathbf{l}}_*)^2$$

$$-E(\sum_j\sum_i \tilde{\mathbf{Z}}'\omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i\tilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}\tilde{\mathbf{Z}})$$

$$+E((\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)^2\tilde{\mathbf{l}}_*'\mathbf{D}^2 g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\tilde{\mathbf{l}}_*)$$

$$+E(\sum_\ell\sum_j \tilde{\mathbf{Z}}'\omega_\ell^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_\ell(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j\tilde{\mathbf{l}}_*)$$

$$+\frac{1}{\mathbf{Q}'\mathbf{Q}}E(\sum_j \tilde{\mathbf{Z}}'\omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}g'\tilde{\mathbf{l}}_*'\mathbf{D}^2 g\tilde{\mathbf{l}}_*)-\hat{\gamma}_I\sigma_\varepsilon^4 \qquad \text{(G-8)}$$

Equation (G-8) is evaluated term-by-term using the result of appendix A as was done for (G-1). For the first term

$$E(\sum_i \sum_j \tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \tilde{\mathbf{l}}_* \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \tilde{\mathbf{l}}_*)$$

$$=\sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\, E(\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i \tilde{\mathbf{l}}_*)(\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_j \tilde{\mathbf{l}}_*)$$

$$=\sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\, E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\,\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}$$

$$\bullet(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\,\mathbf{V}_*^{\frac{1}{2}}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\,\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}$$

$$\bullet(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\,\mathbf{V}_*^{\frac{1}{2}}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)$$

$$=\sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\, tr(\mathbf{A}_i)tr(\mathbf{A}_j)\sigma_\varepsilon^4+2\sum_i \sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}} tr(\mathbf{A}_j\mathbf{A}_\ell)\sigma_\varepsilon^4$$

$$=(\sum_j(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\, tr(\mathbf{A}_j))'\sum_j(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\, tr(\mathbf{A}_j)\sigma_\varepsilon^4$$

$$+2\sum_j \sum_\ell((\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}})'(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega_\ell^{\frac{1}{2}}\, tr(\mathbf{A}_j\mathbf{A}_\ell)\sigma_\varepsilon^4$$

$$=\sum_i \sum_j(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega_j^{\frac{1}{2}}\, tr(\mathbf{A}_j)\sum_j(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega_j^{\frac{1}{2}}\, tr(\mathbf{A}_j)\sigma_\varepsilon^4$$

$$+2\sum_i \sum_j \sum_\ell tr\,((\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega_j^{\frac{1}{2}}\mathbf{A}_j(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega_\ell^{\frac{1}{2}}\,\mathbf{A}_\ell)\sigma_\varepsilon^4$$

$$=\sum_i tr^2(\tilde{\mathbf{C}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}})\sigma_\varepsilon^4+2\sum_i tr(\tilde{\mathbf{C}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}}\tilde{\mathbf{C}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4 \qquad \text{(G-9)}$$

where temporary variable $\mathbf{A}_i$ is defined implicitly, and

$$\tilde{\mathbf{C}}_i=(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\sum_j \omega_j^{\frac{1}{2}}\,(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\,\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j\,(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}) \qquad \text{(G-10)}$$

For the second term

$$E(\sum_i \tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i\tilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}\,\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\,\tilde{\mathbf{l}}_*'\mathbf{D}^2 g\,\tilde{\mathbf{l}}_*)$$

$$=\sum_i \omega_i^{\frac{1}{2}}\,\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}E(\tilde{\mathbf{l}}_*'\mathbf{D}^2 f_i\tilde{\mathbf{l}}_*)(\tilde{\mathbf{l}}_*'\mathbf{D}^2 g\tilde{\mathbf{l}}_*)$$

$$=\sum_i \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\,tr(\mathbf{A}_i)tr(\mathbf{A})\sigma_\varepsilon^4+2\sum_i \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\,tr(\mathbf{A}_i\,\mathbf{A})\sigma_\varepsilon^4$$

$$= \sum_j \frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\, tr(\mathbf{A}_j) tr(\mathbf{A})\sigma_\varepsilon^4 + 2\sum_j \frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\, tr(\mathbf{A}_j\,\mathbf{A})\sigma_\varepsilon^4$$

$$= \sum_i \sum_j (\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega_j^{\frac{1}{2}}\, tr(\mathbf{A}_j)\frac{Q_i}{\mathbf{Q}'\mathbf{Q}} tr(\mathbf{A})\sigma_\varepsilon^4 + 2\sum_i \sum_j (\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})_i\,\omega_j^{\frac{1}{2}}\frac{Q_i}{\mathbf{Q}'\mathbf{Q}}\, tr(\mathbf{A}_j\mathbf{A})\sigma_\varepsilon^4$$

$$= \sum_i tr(\widetilde{\mathbf{C}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}}) tr(\widetilde{\mathbf{F}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4 + 2\sum_i tr(\widetilde{\mathbf{C}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}}\widetilde{\mathbf{F}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4 \qquad \text{(G-11)}$$

where temporary variables $\mathbf{A}$ and $\mathbf{A}_i$ are defined implicitly,

$$\widetilde{\mathbf{F}}_i = \frac{Q_i}{\mathbf{Q}'\mathbf{Q}}\widetilde{\mathbf{A}} \qquad \text{(G-12)}$$

and

$$\widetilde{\mathbf{A}} = (\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\,\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}) \qquad \text{(G-13)}$$

For the third term

$$\frac{1}{\mathbf{Q}'\mathbf{Q}}E(\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g\widetilde{\mathbf{l}}_*)^2 = \sum_i \frac{Q_i}{\mathbf{Q}'\mathbf{Q}} tr(\mathbf{A})\frac{Q_i}{\mathbf{Q}'\mathbf{Q}} tr(\mathbf{A})\sigma_\varepsilon^4 + 2\sum_i tr(\frac{Q_i}{\mathbf{Q}'\mathbf{Q}}\mathbf{A}\frac{Q_i}{\mathbf{Q}'\mathbf{Q}}\mathbf{A})\sigma_\varepsilon^4$$

$$= \sum_i tr^2(\widetilde{\mathbf{F}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}})\sigma_\varepsilon^4 + 2\sum_i tr(\widetilde{\mathbf{F}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}}\widetilde{\mathbf{F}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4 \qquad \text{(G-14)}$$

For the fourth term

$$E(\sum_j \sum_i \widetilde{\mathbf{Z}}'\omega_j^{\frac{1}{2}}\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i\widetilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}})$$

$$= \sum_j \sum_i E(\widetilde{\mathbf{Z}}'\omega_j^{\frac{1}{2}}\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}})(\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i\widetilde{\mathbf{l}}_*)$$

$$= \sum_j \sum_i E(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)(\mathbf{U}_*'\mathbf{V}_*^{-\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})$$

$$\bullet\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_j(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}$$

$$\bullet\,(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}\mathbf{V}_*^{-\frac{1}{2}}\mathbf{U}_*)$$

$$= \sum_j \sum_i tr(\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}}) tr(\widetilde{\mathbf{B}}_j\widetilde{\mathbf{B}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+ 2\sum_j \sum_i tr(\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega_j^{\frac{1}{2}}\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\widetilde{\mathbf{B}}_j\widetilde{\mathbf{B}}_i\omega^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$=\sum_j\sum_i \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega_j^{\frac{1}{2}}\,tr(\widetilde{\mathbf{B}}_j\widetilde{\mathbf{B}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+2\sum_j\sum_i \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\widetilde{\mathbf{B}}_j\,\widetilde{\mathbf{B}}_i\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\sigma_\varepsilon^4 \qquad \text{(G-15)}$$

where

$$\widetilde{\mathbf{B}}_i=(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\,\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 f_i\,(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}) \qquad \text{(G-16)}$$

For the fifth term

$$E((\frac{1}{\mathbf{Q}'\mathbf{Q}}\,\mathbf{Q}'\omega^{\frac{1}{2}}\,\mathbf{U}_*)^2\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^2 g\,\widetilde{\mathbf{l}}_*)$$

$$=\frac{1}{(\mathbf{Q}'\mathbf{Q})^2}\,tr(\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}\mathbf{QQ}'\omega^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}})tr(\widetilde{\mathbf{A}}^2\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+\frac{2}{(\mathbf{Q}'\mathbf{Q})^2}\,tr(\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}\mathbf{QQ}'\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\widetilde{\mathbf{A}}^2\omega^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$=\frac{1}{(\mathbf{Q}'\mathbf{Q})^2}\,\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\mathbf{Q}\,tr(\widetilde{\mathbf{A}}^2\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4+\frac{2}{(\mathbf{Q}'\mathbf{Q})^2}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\widetilde{\mathbf{A}}^2\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\mathbf{Q}\sigma_\varepsilon^4 \qquad \text{(G-17)}$$

For the sixth term

$$E(\sum_\ell\sum_j \widetilde{\mathbf{Z}}'\omega_\ell^{\frac{1}{2}}\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_\ell(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j\,\widetilde{\mathbf{l}}_*)$$

$$=\sum_i\sum_j E(\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega_j^{\frac{1}{2}}\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}})(\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j\,\widetilde{\mathbf{l}}_*)$$

$$=\sum_i\sum_j tr(\widetilde{\mathbf{B}}_i\omega_j^{\frac{1}{2}}\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})tr(\widetilde{\mathbf{B}}_j\,\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+2\sum_i\sum_j tr(\mathbf{V}_*^{\frac{1}{2}}\omega^{\frac{1}{2}}\,\widetilde{\mathbf{B}}_i\omega_j^{\frac{1}{2}}\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\widetilde{\mathbf{B}}_j\omega^{\frac{1}{2}}\mathbf{V}_*^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$=\sum_i\sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\widetilde{\mathbf{B}}_i\omega_j^{\frac{1}{2}}\,tr(\widetilde{\mathbf{B}}_j\,\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+2\sum_i\sum_j \omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{V}_*\omega^{\frac{1}{2}}\widetilde{\mathbf{B}}_j\,\omega^{\frac{1}{2}}\mathbf{V}_*\,\omega^{\frac{1}{2}}\,\widetilde{\mathbf{B}}_i\omega_j^{\frac{1}{2}}\sigma_\varepsilon^4 \qquad \text{(G-18)}$$

For the seventh term

$$
\begin{aligned}
&\frac{1}{\mathbf{Q'Q}} E(\sum_j \widetilde{\mathbf{Z}}' \omega_j^{\frac{1}{2}} \widetilde{\mathbf{l}}_*' \mathbf{D}^2 f_j (\mathbf{Df'\omega Df})^{-1} \mathbf{D}g' \, \widetilde{\mathbf{l}}_*' \mathbf{D}^2 g \widetilde{\mathbf{l}}_*) \\
&= \frac{1}{\mathbf{Q'Q}} \sum_i E(\widetilde{\mathbf{l}}_*' \mathbf{D}^2 f_i (\mathbf{Df'\omega Df})^{-1} \mathbf{D}g' \omega_i^{\frac{1}{2}} \widetilde{\mathbf{Z}})(\widetilde{\mathbf{l}}_*' \mathbf{D}^2 g \widetilde{\mathbf{l}}_*) \\
&= \frac{1}{\mathbf{Q'Q}} \sum_i E(\mathbf{U}_*' \omega^{\frac{1}{2}} (\mathbf{R} - \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \mathbf{B}_i \mathbf{Q} \omega_i^{\frac{1}{2}} (\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \omega^{\frac{1}{2}} \mathbf{U}_*)(\mathbf{U}_*' \omega^{\frac{1}{2}} \widetilde{\mathbf{A}} \omega^{\frac{1}{2}} \mathbf{U}_*) \\
&= \frac{1}{\mathbf{Q'Q}} \sum_i tr\,((\mathbf{R} - \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \mathbf{B}_i \mathbf{Q} \omega_i^{\frac{1}{2}} (\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}}) tr(\widetilde{\mathbf{A}} \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}}) \sigma_\varepsilon^4 \\
&+ \frac{2}{\mathbf{Q'Q}} \sum_i tr\,((\mathbf{R} - \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \mathbf{B}_i \mathbf{Q} \omega_i^{\frac{1}{2}} (\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}} \widetilde{\mathbf{A}} \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}}) \sigma_\varepsilon^4 \\
&= \frac{1}{\mathbf{Q'Q}} \sum_i \omega_i^{\frac{1}{2}} (\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}} (\mathbf{R} - \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \mathbf{B}_i \mathbf{Q} tr(\widetilde{\mathbf{A}} \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}}) \sigma_\varepsilon^4 \\
&+ \frac{2}{\mathbf{Q'Q}} \sum_i \omega_i^{\frac{1}{2}} (\mathbf{I} - \mathbf{R} + \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}} \widetilde{\mathbf{A}} \omega^{\frac{1}{2}} \mathbf{V}_* \omega^{\frac{1}{2}} (\mathbf{R} - \frac{\mathbf{QQ'}}{\mathbf{Q'Q}}) \mathbf{B}_i \mathbf{Q} \sigma_\varepsilon^4 \qquad \text{(G-19)}
\end{aligned}
$$

## Component Correction Factors Pertaining to Individual Prediction Intervals

***Evaluations using a general weight matrix.*** Component correction factor $\hat{\gamma}_{la}\sigma_\varepsilon^4$ is given by (F-93), appendix F, as

$$
\begin{aligned}
&\hat{\gamma}_{la}\sigma_\varepsilon^4 \approx \frac{1}{4} E(\sum_i \sum_j \mathbf{l}_{*a}' \mathbf{D}_a^2 f_{ai} \mathbf{l}_{*a} \mathbf{W}_{ai}^{\frac{1}{2}} (\mathbf{I}_a - \mathbf{R}_a) \mathbf{W}_{aj}^{\frac{1}{2}} \mathbf{l}_{*a}' \mathbf{D}_a^2 f_{aj} \mathbf{l}_{*a}) \\
&- E(\sum_i \sum_k \mathbf{W}_{ai}^{\frac{1}{2}} \mathbf{Z}_{*a} \mathbf{l}_{*a}' \mathbf{D}_a^2 f_{ai} (\mathbf{D}_a \mathbf{f}_a' \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a^2 f_{ak} \mathbf{l}_{*a} \mathbf{W}_{ak}^{\frac{1}{2}} \mathbf{Z}_{*a}) \\
&+ E(\sum_i \sum_j \mathbf{W}_{ai}^{\frac{1}{2}} \mathbf{Z}_{*a} \mathbf{l}_{*a}' \mathbf{D}_a^2 f_{ai} (\mathbf{D}_a \mathbf{f}_a' \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}_a' \mathbf{W}_{aj} \mathbf{l}_*' \mathbf{D}_a^2 f_{aj} \mathbf{l}_{*a}) \qquad \text{(G-20)}
\end{aligned}
$$

Equation (G-20) is analogous to (G-1) and evaluates analogously. Hence, for the first term

$$
\begin{aligned}
&E(\sum_i \sum_j \mathbf{l}_{*a}' \mathbf{D}_a^2 f_{ai} \mathbf{l}_{*a} \mathbf{W}_{ai}^{\frac{1}{2}} (\mathbf{I}_a - \mathbf{R}_a) \mathbf{W}_{aj}^{\frac{1}{2}} \mathbf{l}_{*a}' \mathbf{D}_a^2 f_{aj} \mathbf{l}_{*a}) \\
&= \sum_i tr^2 (\mathbf{C}_{ai} \mathbf{W}_a^{\frac{1}{2}} \mathbf{V}_{*a} \mathbf{W}_a^{\frac{1}{2}}) \, \sigma_\varepsilon^4 + 2 \sum_i tr(\mathbf{C}_{ai} \mathbf{W}_a^{\frac{1}{2}} \mathbf{V}_{*a} \mathbf{W}_a^{\frac{1}{2}} \mathbf{C}_{ai} \mathbf{W}_a^{\frac{1}{2}} \mathbf{V}_{*a} \mathbf{W}_a^{\frac{1}{2}}) \sigma_\varepsilon^4 \qquad \text{(G-21)}
\end{aligned}
$$

where

$$
\mathbf{C}_{ai} = (\mathbf{I}_a - \mathbf{R}_a)_i \sum_j \mathbf{W}_{aj}^{\frac{1}{2}} \mathbf{W}_a^{\frac{1}{2}} \mathbf{D}_a \mathbf{f}_a (\mathbf{D}_a \mathbf{f}_a' \mathbf{W}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a^2 f_{aj} (\mathbf{D}_a \mathbf{f}_a' \mathbf{W}_a \mathbf{D}_a \mathbf{f})^{-1} \mathbf{D}_a \mathbf{f}_a' \mathbf{W}_a^{\frac{1}{2}} \qquad \text{(G-22)}
$$

For the second term

$$E(\sum_i\sum_k \mathbf{W}_{ai}^{\frac{1}{2}}\mathbf{Z}_{*a}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a^2 f_{ak}\mathbf{l}_{*a}\mathbf{W}_{ak}^{\frac{1}{2}}\mathbf{Z}_{*a})$$
$$=\sum_j\sum_k \mathbf{W}_{ak}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_{aj}^{\frac{1}{2}}tr(\mathbf{B}_{aj}\mathbf{B}_{ak}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4$$
$$+2\sum_j\sum_k \mathbf{W}_{ak}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\mathbf{B}_{aj}\mathbf{B}_{ak}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_{aj}^{\frac{1}{2}}\sigma_\varepsilon^4 \tag{G-23}$$

where

$$\mathbf{B}_{aj}=\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a^{\frac{1}{2}} \tag{G-24}$$

For the third term

$$E(\sum_i\sum_j \mathbf{W}_{ai}^{\frac{1}{2}}\mathbf{Z}_{*a}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_{aj}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{aj}\mathbf{l}_{*a})$$
$$=\sum_i\sum_j \mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\mathbf{B}_{ai}\mathbf{W}_{aj}^{\frac{1}{2}}tr(\mathbf{B}_{aj}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4$$
$$+2\sum_i\sum_j \mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\mathbf{B}_{aj}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\mathbf{B}_{ai}\mathbf{W}_{aj}^{\frac{1}{2}}\sigma_\varepsilon^4 \tag{G-25}$$

***Evaluations using a known, block diagonal weight matrix***. Equations (G-21)-(G-25) can be expressed in terms of $\boldsymbol{\omega}_a$ by using (5-89), (5-94), and definitions (E-46), (E-47), and (E-56), appendix E. When making the multiplications, the explicit sums on $i, j, k$, or $\ell$ that involve augmented vectors and matrices extend over $n+1$ terms and the $n+1$th slice is $\mathbf{D}_a^2 f_{an+1}=\mathbf{0}$, which follows from the definition given by (E-46). Also, $\mathbf{W}_{ai}^{1/2}=[\boldsymbol{\omega}_i^{1/2}, \mathbf{0}]$ for $i=1,2,\ldots,n$ and $\mathbf{W}_{ai}^{1/2}=[\mathbf{0}, \omega_p^{1/2}]$ for $i=n+1$. Equation (G-22) becomes

$$\mathbf{C}_{ai}=(\mathbf{I}_a-\mathbf{R}_a)_i\sum_j \boldsymbol{\omega}_{aj}^{\frac{1}{2}}\boldsymbol{\omega}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}_a'\boldsymbol{\omega}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}_a'\boldsymbol{\omega}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}_a'\boldsymbol{\omega}_a^{\frac{1}{2}}$$
$$=\begin{bmatrix}\mathbf{C}_i & \mathbf{0}\\ \mathbf{0} & 0\end{bmatrix} \tag{G-26}$$

so that for (G-21) written in terms of $\boldsymbol{\omega}_a$

$$E(\sum_i\sum_j \mathbf{l}'_{*a}\mathbf{D}_a^2 f_{ai}\mathbf{l}_{*a}\boldsymbol{\omega}_{ai}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a)\boldsymbol{\omega}_{aj}^{\frac{1}{2}}\mathbf{l}'_{*a}\mathbf{D}_a^2 f_{aj}\mathbf{l}_{*a})$$
$$=\sum_i tr^2(\mathbf{C}_i\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^4+2\sum_i tr(\mathbf{C}_i\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{C}_i\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{V}_*\boldsymbol{\omega}^{\frac{1}{2}})\sigma_\varepsilon^4 \tag{G-27}$$

which is the same as (G-2). Equation (G-24) becomes

$$\mathbf{B}_{aj} = \boldsymbol{\omega}_a^{\frac{1}{2}} \mathbf{D}_a \mathbf{f}_a (\mathbf{D}_a \mathbf{f}'_a \boldsymbol{\omega}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a^2 f_{aj} (\mathbf{D}_a \mathbf{f}'_a \boldsymbol{\omega}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}'_a \boldsymbol{\omega}_a^{\frac{1}{2}}$$

$$= \begin{bmatrix} \mathbf{B}_j & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \tag{G-28}$$

which gives, for (G-23) written using $\boldsymbol{\omega}_a$,

$$E(\sum_i \sum_k \boldsymbol{\omega}_{ai}^{\frac{1}{2}} \mathbf{Z}_{*a} \mathbf{l}'_{*a} \mathbf{D}_a^2 f_{ai} (\mathbf{D}_a \mathbf{f}'_a \boldsymbol{\omega}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a^2 f_{ak} \mathbf{l}_{*a} \boldsymbol{\omega}_{ak}^{\frac{1}{2}} \mathbf{Z}_{*a})$$

$$= \sum_j \sum_k \boldsymbol{\omega}_k^{\frac{1}{2}} (\mathbf{I} - \mathbf{R}) \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}} (\mathbf{I} - \mathbf{R}) \boldsymbol{\omega}_j^{\frac{1}{2}} tr(\mathbf{B}_j \mathbf{B}_k \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}}) \sigma_\varepsilon^4$$

$$+ 2 \sum_j \sum_k \boldsymbol{\omega}_k^{\frac{1}{2}} (\mathbf{I} - \mathbf{R}) \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{B}_j \mathbf{B}_k \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}} (\mathbf{I} - \mathbf{R}) \boldsymbol{\omega}_j^{\frac{1}{2}} \sigma_\varepsilon^4 \tag{G-29}$$

Equation (G-29) is the same as (G-4). Finally, use of (G-28) yields (G-25) written using $\boldsymbol{\omega}_a$ as

$$E(\sum_i \sum_j \boldsymbol{\omega}_{ai}^{\frac{1}{2}} \mathbf{Z}_{*a} \mathbf{l}'_{*a} \mathbf{D}_a^2 f_{ai} (\mathbf{D}_a \mathbf{f}'_a \boldsymbol{\omega}_a \mathbf{D}_a \mathbf{f}_a)^{-1} \mathbf{D}_a \mathbf{f}'_a \boldsymbol{\omega}_{aj} \mathbf{l}'_{*a} \mathbf{D}_a^2 f_{aj} \mathbf{l}_{*a})$$

$$= \sum_i \sum_j \boldsymbol{\omega}_i^{\frac{1}{2}} (\mathbf{I} - \mathbf{R}) \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{B}_i \boldsymbol{\omega}_j^{\frac{1}{2}} tr(\mathbf{B}_j \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}}) \sigma_\varepsilon^4$$

$$+ 2 \sum_i \sum_j \boldsymbol{\omega}_i^{\frac{1}{2}} (\mathbf{I} - \mathbf{R}) \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{B}_j \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{V}_* \boldsymbol{\omega}^{\frac{1}{2}} \mathbf{B}_i \boldsymbol{\omega}_j^{\frac{1}{2}} \sigma_\varepsilon^4 \tag{G-30}$$

which is (G-6).

Equations (G-26)-(G-30) lead to the conclusion that $\hat{\gamma}_{Ia}\sigma_\varepsilon^4 = \hat{\gamma}_I \sigma_\varepsilon^4$. That is, the component correction factor for model intrinsic nonlinearity for prediction intervals is the same as the component correction factor for model intrinsic nonlinearity for confidence regions and confidence intervals.

***Evaluations using a general weight matrix***. Component correction factor $\gamma_{Ia}\sigma_\varepsilon^4$ is given by (F-95), which is

$$\gamma_{Ia}\sigma_\varepsilon^4 \approx \frac{1}{4} E(\sum_i \sum_j \tilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 f_{ai} \tilde{\mathbf{l}}_* \mathbf{W}_{ai}^{\frac{1}{2}} (\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a \mathbf{Q}'_a}{\mathbf{Q}'_a \mathbf{Q}_a}) \mathbf{W}_{aj}^{\frac{1}{2}} \tilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 f_{aj} \tilde{\mathbf{l}}_{*a})$$

$$- \frac{1}{2} E(\sum_i \tilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 f_{ai} \tilde{\mathbf{l}}_{*a} \mathbf{W}_{ai}^{\frac{1}{2}} \frac{\mathbf{Q}_a}{\mathbf{Q}'_a \mathbf{Q}_a} \tilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 h \tilde{\mathbf{l}}_{*a}) + \frac{1}{4} \frac{1}{\mathbf{Q}'_a \mathbf{Q}_a} E(\tilde{\mathbf{l}}'_{*a} \mathbf{D}_a^2 h \tilde{\mathbf{l}}_{*a})^2$$

$$-E(\sum_j\sum_i \widetilde{\mathbf{Z}}'_a \mathbf{W}_{aj}^{\frac{1}{2}}\widetilde{\mathbf{I}}'_{*a}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a^2 f_{ai}\widetilde{\mathbf{I}}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}\widetilde{\mathbf{Z}}_a)$$
$$+E((\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{Q}'_a\mathbf{W}_a^{\frac{1}{2}}\ \mathbf{U}_{*a})^2\widetilde{\mathbf{I}}'_{*a}\ \mathbf{D}_a^2 h(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\ \mathbf{D}_a^2 h\widetilde{\mathbf{I}}_{*a}$$
$$+E(\sum_\ell\sum_j \widetilde{\mathbf{Z}}'_a \mathbf{W}_{a\ell}^{\frac{1}{2}}\widetilde{\mathbf{I}}'_{*a}\mathbf{D}_a^2 f_{a\ell}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_{aj}^{\frac{1}{2}}\widetilde{\mathbf{I}}'_{*a}\mathbf{D}_a^2 f_{aj}\widetilde{\mathbf{I}}_{*a})$$
$$+\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}E(\sum_j\widetilde{\mathbf{Z}}'_a\mathbf{W}_{aj}^{\frac{1}{2}}\widetilde{\mathbf{I}}'_{*a}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h'\widetilde{\mathbf{I}}'_{*a}\mathbf{D}_a^2 h\widetilde{\mathbf{I}}_{*a})-\hat{\gamma}_{Ia}\sigma_\varepsilon^4 \qquad \text{(G-31)}$$

Equation (G-31) is analogous to (G-8) and evaluates analogously. Thus, for the first term

$$E(\sum_i\sum_j \widetilde{\mathbf{I}}'_{*a}\ \mathbf{D}_a^2 f_{ai}\widetilde{\mathbf{I}}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_{aj}^{\frac{1}{2}}\widetilde{\mathbf{I}}'_{*a}\ \mathbf{D}_a^2 f_{aj}\widetilde{\mathbf{I}}_{*a})$$
$$=\sum_i tr^2(\widetilde{\mathbf{C}}_{ai}\ \mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4+2\sum_i tr(\widetilde{\mathbf{C}}_{ai}\ \mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\widetilde{\mathbf{C}}_{ai}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4 \qquad \text{(G-32)}$$

where

$$\widetilde{\mathbf{C}}_{ai}=(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})_i\sum_j \mathbf{W}_{aj}^{\frac{1}{2}}\ (\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}$$
$$\bullet\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a}) \qquad \text{(G-33)}$$

For second term

$$E(\sum_i \widetilde{\mathbf{I}}'_{*a}\ \mathbf{D}_a^2 f_{ai}\widetilde{\mathbf{I}}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}\frac{\mathbf{Q}_a}{\mathbf{Q}'_a\mathbf{Q}_a}\widetilde{\mathbf{I}}'_{*a}\mathbf{D}_a^2 h\widetilde{\mathbf{I}}_{*a})$$
$$=\sum_i tr(\widetilde{\mathbf{C}}_{ai}\ \mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})tr(\widetilde{\mathbf{F}}_{ai}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4+2\sum_i tr(\widetilde{\mathbf{C}}_{ai}\ \mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\widetilde{\mathbf{F}}_{ai}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4 \qquad \text{(G-34)}$$

where

$$\widetilde{\mathbf{F}}_{ai}=\frac{Q_{ai}}{\mathbf{Q}'_a\mathbf{Q}_a}\widetilde{\mathbf{A}}_a \qquad \text{(G-35)}$$

in which

$$\widetilde{\mathbf{A}}_a = (\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a^2 h(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})$$

For the third term

$$\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}E(\widetilde{\mathbf{l}}'_{*a}\,\mathbf{D}_a^2 h\widetilde{\mathbf{l}}_{*a})^2 = \sum_i tr^2(\widetilde{\mathbf{F}}_{ai}\,\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\,\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4 + 2\sum_i tr(\widetilde{\mathbf{F}}_{ai}\,\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\,\mathbf{W}_a^{\frac{1}{2}}\widetilde{\mathbf{F}}_{ai}\,\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\,\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4 \quad \text{(G-37)}$$

For the fourth term

$$E(\sum_j\sum_i \widetilde{\mathbf{Z}}'_a\mathbf{W}_{aj}^{\frac{1}{2}}\widetilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a^2 f_{ai}\widetilde{\mathbf{l}}_{*a}\mathbf{W}_{ai}^{\frac{1}{2}}\widetilde{\mathbf{Z}}_a)$$
$$= \sum_j\sum_i \mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_{aj}^{\frac{1}{2}}tr(\widetilde{\mathbf{B}}_{aj}\widetilde{\mathbf{B}}_{ai}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4$$
$$+ 2\sum_j\sum_i \mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\widetilde{\mathbf{B}}_{aj}\widetilde{\mathbf{B}}_{ai}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\,\mathbf{W}_{aj}^{\frac{1}{2}}\sigma_\varepsilon^4 \quad \text{(G-38)}$$

where

$$\widetilde{\mathbf{B}}_{ai} = (\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a^2 f_{ai}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}\,(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})$$
$$\text{(G-39)}$$

For the fifth term

$$E((\frac{1}{\mathbf{Q}'_a\mathbf{Q}_a}\mathbf{Q}'_a\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{U}_{*a})^2\widetilde{\mathbf{l}}'_{*a}\,\mathbf{D}_a^2 h(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{D}_a\mathbf{f}_a(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\,\mathbf{D}_a^2 h\,\widetilde{\mathbf{l}}_{*a})$$
$$= \frac{1}{(\mathbf{Q}'_a\mathbf{Q}_a)^2}\mathbf{Q}'_a\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{Q}_a tr(\widetilde{\mathbf{A}}_a^2\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4 + \frac{2}{(\mathbf{Q}'_a\mathbf{Q}_a)^2}\mathbf{Q}'_a\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\,\widetilde{\mathbf{A}}_a^2\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{Q}_a\sigma_\varepsilon^4$$
$$\text{(G-40)}$$

For the sixth term

$$E(\sum_\ell\sum_j \widetilde{\mathbf{Z}}'_a\mathbf{W}_{a\ell}^{\frac{1}{2}}\widetilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{a\ell}(\mathbf{D}_a\mathbf{f}'_a\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a\mathbf{f}'_a\,\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a - \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_{aj}^{\frac{1}{2}}\,\widetilde{\mathbf{l}}'_{*a}\mathbf{D}_a^2 f_{aj}\,\widetilde{\mathbf{l}}_{*a})$$
$$= \sum_i\sum_j \mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a - \mathbf{R}_a + \frac{\mathbf{Q}_a\mathbf{Q}'_a}{\mathbf{Q}'_a\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\,\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\,\widetilde{\mathbf{B}}_{ai}\mathbf{W}_{aj}^{\frac{1}{2}}tr(\widetilde{\mathbf{B}}_{aj}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4$$

$$+2\sum_i\sum_j \mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\widetilde{\mathbf{B}}_{aj}\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\widetilde{\mathbf{B}}_{ai}\mathbf{W}_{aj}^{\frac{1}{2}}\sigma_\varepsilon^4 \tag{G-41}$$

Finally, for the seventh term

$$\frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}E(\sum_j \widetilde{\mathbf{Z}}_a'\mathbf{W}_{aj}^{\frac{1}{2}}\widetilde{\mathbf{l}}_{*a}'\mathbf{D}_a^2 f_{aj}(\mathbf{D}_a\mathbf{f}_a'\mathbf{W}_a\mathbf{D}_a\mathbf{f}_a)^{-1}\mathbf{D}_a h'\widetilde{\mathbf{l}}_{*a}'\mathbf{D}_a^2 h\,\widetilde{\mathbf{l}}_{*a})$$
$$=\frac{1}{\mathbf{Q}_a'\mathbf{Q}_a}\sum_i \mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{B}_{ai}\mathbf{Q}_a tr(\widetilde{\mathbf{A}}_a\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}})\sigma_\varepsilon^4$$
$$+\frac{2}{\mathbf{Q}_a'\mathbf{Q}_a}\sum_i \mathbf{W}_{ai}^{\frac{1}{2}}(\mathbf{I}_a-\mathbf{R}_a+\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}\widetilde{\mathbf{A}}_a\mathbf{W}_a^{\frac{1}{2}}\mathbf{V}_{*a}\mathbf{W}_a^{\frac{1}{2}}(\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a})\mathbf{B}_{ai}\mathbf{Q}_a\sigma_\varepsilon^4 \tag{G-42}$$

***Evaluation when the prediction error predominates and the weight matrix is block diagonal.*** The following evaluation of the terms in (G-31) for the case where $\omega_p^{-1} >> \mathbf{Q'Q}$ yields an important result. First, from (5-92),

$$\mathbf{Q}_a\mathbf{Q}_a' = \begin{bmatrix}\mathbf{Q}\\ -\omega_p^{-\frac{1}{2}}\end{bmatrix}[\mathbf{Q}' \; -\omega_p^{-\frac{1}{2}}] = \begin{bmatrix}\mathbf{QQ}' & -\mathbf{Q}\omega_p^{-\frac{1}{2}}\\ -\omega_p^{-\frac{1}{2}}\mathbf{Q}' & \omega_p^{-1}\end{bmatrix} \tag{G-43}$$

and

$$\mathbf{Q}_a'\mathbf{Q}_a = \mathbf{Q'Q}+\omega_p^{-1} \tag{G-44}$$

Second, use of (5-94) and $\omega_p^{-1} >> \mathbf{Q'Q}$ produces

$$\mathbf{R}_a-\frac{\mathbf{Q}_a\mathbf{Q}_a'}{\mathbf{Q}_a'\mathbf{Q}_a} = \begin{bmatrix}\mathbf{R}-\dfrac{\mathbf{QQ}'}{\mathbf{Q'Q}+\omega_p^{-1}} & \dfrac{\mathbf{Q}\omega_p^{-\frac{1}{2}}}{\mathbf{Q'Q}+\omega_p^{-1}}\\ \dfrac{\omega_p^{-\frac{1}{2}}\mathbf{Q}'}{\mathbf{Q'Q}+\omega_p^{-1}} & 1-\dfrac{\omega_p^{-1}}{\mathbf{Q'Q}+\omega_p^{-1}}\end{bmatrix} \approx \begin{bmatrix}\mathbf{R} & \mathbf{0}\\ \mathbf{0} & 0\end{bmatrix} \tag{G-45}$$

Third, when making multiplications 1) the $n+i$th slice of $\mathbf{D}_a^2\mathbf{f}_a$ is a matrix of zeros and 2) $\mathbf{W}_{ai}^{1/2} = [\omega_i^{1/2}, \mathbf{0}]$ for $i = 1,2,\ldots,n$ and $\mathbf{W}_{ai}^{1/2} = [\mathbf{0}, \omega_p^{1/2}]$ for $i = n+1$. Finally, application of the above ideas to (G-32)-(G-42) shows from (G-31) that when $\omega_p^{-1} >> \mathbf{Q'Q}$

$$\gamma_{Ia}\sigma_\varepsilon^4 \approx \hat{\gamma}_I\sigma_\varepsilon^4 - \hat{\gamma}_I\sigma_\varepsilon^4$$
$$=0 \tag{G-46}$$

# Appendix H – Prediction Interval for Linearized Models

Extreme values of (5-101) using linear models (5-104) and (5-105) are obtained in this appendix. Specific forms of (5-104) and (5-105) needed are

$$\mathbf{f}(\gamma\tilde{\theta}) \approx \mathbf{f}(\gamma\bar{\theta}) + \mathbf{Df}(\tilde{\theta} - \bar{\theta}) \tag{H-1}$$

and

$$\mathbf{f}(\gamma\hat{\theta}) \approx \mathbf{f}(\gamma\bar{\theta}) + \mathbf{Df}(\hat{\theta} - \bar{\theta}) \tag{H-2}$$

so that, by subtraction,

$$\mathbf{f}(\gamma\tilde{\theta}) \approx \mathbf{f}(\gamma\hat{\theta}) + \mathbf{Df}(\tilde{\theta} - \hat{\theta}) \tag{H-3}$$

Similarly,

$$g(\gamma\tilde{\theta}) \approx g(\gamma\hat{\theta}) + \mathbf{D}g(\tilde{\theta} - \hat{\theta}) \tag{H-4}$$

Extreme values of (5-101) are obtained by taking derivatives with respect to $\theta$, $\upsilon$, and $\lambda$ to get: for $\theta$

$$\frac{\partial L}{\partial \theta} \approx \mathbf{D}g' + 2\lambda\,\mathbf{Df}'\omega(\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta}) - \mathbf{Df}(\theta - \hat{\theta})) = \mathbf{0}$$

or

$$\begin{aligned}\mathbf{Df}'\omega\mathbf{Df}(\tilde{\theta} - \hat{\theta}) &= \tilde{\lambda}\,\mathbf{D}g' + \mathbf{Df}'\omega(\mathbf{Y} - \mathbf{f}(\gamma\hat{\theta})) \\ &= \tilde{\lambda}\,\mathbf{D}g'\end{aligned} \tag{H-5}$$

where $\tilde{\lambda} = 1/(2\lambda)$ and the second term on the right-hand side is zero because it is the gradient for the least squares solution. For $\upsilon$

$$\frac{\partial L}{\partial \upsilon} \approx 1 - 2\lambda\,\omega_p \upsilon = 0$$

or

$$\omega_p \tilde{\upsilon} = \tilde{\lambda} \tag{H-6}$$

Finally, for $\lambda$

$$\frac{\partial L}{\partial \lambda} = d_\alpha^2 - S(\boldsymbol{\theta}) - \omega_p \upsilon^2 + S(\hat{\boldsymbol{\theta}}) = 0$$

or

$$d_\alpha^2 = S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}) + \omega_p \tilde{\upsilon}^2 \tag{H-7}$$

where

$$d_\alpha^2 = \frac{S(\hat{\boldsymbol{\theta}})}{n-p} c_p t_{\alpha/2}^2 (n-p) \tag{H-8}$$

Solution of (H-5)-(H-7) for $\tilde{\boldsymbol{\theta}}$ and $\tilde{\lambda}$ yields equations from which $\tilde{Y}_p$ is obtained. First, (H-3) and (H-5) are used to obtain

$$\begin{aligned}
S(\tilde{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}) &\approx (\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) - \mathbf{Df}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) - \mathbf{Df}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})) - S(\hat{\boldsymbol{\theta}}) \\
&= S(\hat{\boldsymbol{\theta}}) - 2(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})'\mathbf{Df}'\boldsymbol{\omega}(\mathbf{Y} - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}})) + (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})'\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) - S(\hat{\boldsymbol{\theta}}) \\
&= (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})'\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \\
&= \tilde{\lambda}^2 \mathbf{D}g(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g' \\
&= \tilde{\lambda}^2 \mathbf{Q}'\mathbf{Q}
\end{aligned} \tag{H-9}$$

Then substitution of (H-6) and (H-9) into (H-7) gives

$$\begin{aligned}
d_\alpha^2 &= \tilde{\lambda}^2 \mathbf{Q}'\mathbf{Q} + \omega_p(\tilde{\lambda}^2/\omega_p^2) \\
&= \tilde{\lambda}^2(\mathbf{Q}'\mathbf{Q} + \omega_p^{-1})
\end{aligned} \tag{H-10}$$

from which

$$\tilde{\lambda} = \pm\left(\frac{d_\alpha^2}{\mathbf{Q}'\mathbf{Q} + \omega_p^{-1}}\right)^{\frac{1}{2}} \tag{H-11}$$

Next, (H-11) in (H-5) are used to arrive at

$$\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}} = \pm\left(\frac{d_\alpha^2}{\mathbf{Q}'\mathbf{Q} + \omega_p^{-1}}\right)^{\frac{1}{2}} (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{D}g' \tag{H-12}$$

Finally, (H-12) is premultiplied by $\mathbf{D}g$ and $\tilde{\upsilon}$ is added to get

$$\mathbf{D}g(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \tilde{\upsilon} = g(\gamma\tilde{\boldsymbol{\theta}}) - g(\gamma\hat{\boldsymbol{\theta}}) + \tilde{\upsilon}$$

$$= \pm\left(\frac{d_\alpha^2}{\mathbf{Q}'\mathbf{Q}+\omega_p^{-1}}\right)^{\frac{1}{2}} \mathbf{D}g(\mathbf{Df}'\mathbf{\omega}\mathbf{Df})^{-1}\mathbf{D}g' \pm \left(\frac{d_\alpha^2}{\mathbf{Q}'\mathbf{Q}+\omega_p^{-1}}\right)^{\frac{1}{2}} \omega_p^{-1}$$

$$= \pm d_\alpha (\mathbf{Q}'\mathbf{Q}+\omega_p^{-1})^{\frac{1}{2}}$$

or

$$\widetilde{Y}_p = g(\gamma\hat{\mathbf{\theta}}) \pm d_\alpha (\mathbf{Q}'\mathbf{Q}+\omega_p^{-1})^{\frac{1}{2}} \qquad \text{(H-13)}$$

# Appendix I – Analysis of Equations (6-2) and (6-19)

Equations (6-2) and (6-19) are analyzed in this appendix to show 1) that certain terms in these equations are zero when model intrinsic nonlinearity and model combined intrinsic nonlinearity are zero and 2) that these equations correspond to perturbation equations derived in appendices B and E. Equations (6-2) and (6-19) also are used together with perturbation expansions and the assumption that $\boldsymbol{\omega}^{1/2}\mathbf{V}_*\boldsymbol{\omega}^{1/2} \approx \boldsymbol{\omega}^{1/2}\boldsymbol{\Omega}\boldsymbol{\omega}^{1/2} \approx \mathbf{I}$ to evaluate component correction factors $\hat{\gamma}_I\sigma_\varepsilon^4$ and $\gamma_I\sigma_\varepsilon^4$.

## Analysis of (6-2)

***Form for small model intrinsic nonlinearity.*** To show that the second term on the right-hand side of (6-2) is zero when model intrinsic nonlinearity is zero, an analysis analogous to the one used to obtain (F-108), appendix F, is used. First, the linear-model approximation of $\mathbf{f}(\gamma\boldsymbol{\theta}_*)$ in (F-104) is

$$\mathbf{p}(\boldsymbol{\psi}_*) = \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) + \mathbf{Df}\boldsymbol{\psi}_* \tag{I-1}$$

Next, an analogous linear-model approximation of $\mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$ is

$$\begin{aligned}\mathbf{p}(\hat{\boldsymbol{\psi}}) &= \mathbf{f}(\gamma\bar{\boldsymbol{\theta}}) + \mathbf{D}_\phi\mathbf{f}(\hat{\boldsymbol{\phi}} - \bar{\boldsymbol{\phi}}) \\ &= \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{Df}\hat{\boldsymbol{\psi}}\end{aligned} \tag{I-2}$$

Now, the best fit of $\mathbf{p}(\boldsymbol{\psi}) = \mathbf{p}(\boldsymbol{\psi}_*) - \mathbf{p}(\hat{\boldsymbol{\psi}})$ to $\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}})$ is obtained by minimizing with respect to $\boldsymbol{\psi} = \boldsymbol{\psi}_* - \hat{\boldsymbol{\psi}}$ the objective function

$$S(\boldsymbol{\psi}) = (\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) - \mathbf{p}(\boldsymbol{\psi}))'\boldsymbol{\omega}(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) - \mathbf{p}(\boldsymbol{\psi})) \tag{I-3}$$

to obtain

$$\boldsymbol{\psi} = (\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\boldsymbol{\omega}(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}})) \tag{I-4}$$

The residual vector for this problem is

$$\begin{aligned}\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) - \mathbf{p}(\boldsymbol{\psi}) &= \mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}) - \mathbf{Df}\boldsymbol{\psi} \\ &= \boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I} - \mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\gamma\boldsymbol{\theta}_*) - \mathbf{f}_0(\gamma\boldsymbol{\theta}_*) - \mathbf{f}(\gamma\hat{\boldsymbol{\theta}}) + \mathbf{f}_0(\gamma\hat{\boldsymbol{\theta}}))\end{aligned} \tag{I-5}$$

If $\mathbf{p}(\psi)$ fits $\mathbf{f}(\gamma\theta_*) - \mathbf{f}(\gamma\hat{\theta})$ exactly, then the linear model $\mathbf{p}(\psi) = \mathbf{f}_0(\gamma\theta_*) - \mathbf{f}_0(\gamma\hat{\theta}) + \mathbf{Df}\psi$ is exact, indicating no model intrinsic nonlinearity. In this case (I-5), which is the second term on the right-hand side of (6-2), equals **0**.

For the last term on the right-hand side of (6-2),

$$
\begin{aligned}
&\omega^{-\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})) = \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))\\
&= \mathbf{DfJ}(\mathbf{J}'\mathbf{Df}'\omega\mathbf{DfJ})^{-1}\mathbf{J}'\mathbf{Df}'\omega(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))\\
&= \mathbf{D}_\phi\mathbf{f}(\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega(\mathbf{Y}-\mathbf{f}(\gamma\theta(\hat{\phi}))) \qquad \text{(I-6)}
\end{aligned}
$$

If there is no model intrinsic nonlinearity, then $\mathbf{D}_\phi\mathbf{f}$ is not a function of $\phi$ so that $\mathbf{D}_\phi\mathbf{f}'\omega(\mathbf{Y}-\mathbf{f}(\gamma\theta(\hat{\phi})))$ is the gradient vector for the least squares solution for $\hat{\phi}$ (and, thus $\hat{\theta}$), which is **0**.

***Correspondance to perturbation form.*** The expansion of $\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})$ given by (6-2) corresponds exactly to the perturbation form given by (B-12), appendix B. Expansion of each term in (6-2) through second order in **e** and **U** shows this result:

$$
\begin{aligned}
&\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))\\
&\approx \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(U_{*j} + \frac{1}{2}\mathbf{e}'(\mathbf{D}^2_\beta f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}) \qquad \text{(I-7)}
\end{aligned}
$$

$$
\begin{aligned}
&\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}(\gamma\hat{\theta})+\mathbf{f}_0(\gamma\hat{\theta}))\\
&\approx \frac{1}{2}\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_j\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}) \qquad \text{(I-8)}
\end{aligned}
$$

$$
\begin{aligned}
&\omega^{-\frac{1}{2}}\mathbf{R}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta})) \approx \mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i(\mathbf{D}\hat{f}_i - \mathbf{D}^2 f_i(\hat{\theta}-\bar{\theta}))\omega_i(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))\\
&\approx -\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i(\mathbf{l}+\mathbf{q})\omega_i(\mathbf{Y}-\mathbf{f}(\gamma\bar{\theta})-\mathbf{Df}(\mathbf{l}+\mathbf{q}))\\
&\approx -\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\omega_i(\mathbf{U}-\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{Df}'\omega\mathbf{U})\\
&= -\mathbf{Df}(\mathbf{Df}'\omega\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z} \qquad \text{(I-9)}
\end{aligned}
$$

(Equation (I-9) also could have been obtained by simply substituting (B-12) into $\omega^{-1/2}\mathbf{R}\omega^{1/2}(\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}))$.) Then, substitution of (I-7)-(I-9) into (6-2) results in

$$
\begin{aligned}
&\mathbf{Y}-\mathbf{f}(\gamma\hat{\theta}) \approx \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(U_{*j} + \frac{1}{2}\mathbf{e}'(\mathbf{D}^2_\beta f_j - \gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})\\
&+\frac{1}{2}\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_j\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e} - \mathbf{l}'\mathbf{D}^2 f_j\mathbf{l})
\end{aligned}
$$

$$-\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i \mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z}$$

$$=\boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(U_{*j}+\frac{1}{2}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))-\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z}$$

$$=\boldsymbol{\omega}^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j \omega_j^{\frac{1}{2}}(U_j+\frac{1}{2}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))-\mathbf{Df}(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_i \mathbf{D}^2 f_i\mathbf{l}\omega_i^{\frac{1}{2}}\mathbf{Z} \qquad \text{(I-10)}$$

which is the same as (B-12).

***Approximate evaluation of terms in component correction factor.*** Equations (I-7)-(I-9), (G-2)-(G-6), appendix G, the assumption that $\boldsymbol{\omega}^{1/2}\mathbf{V}_*\boldsymbol{\omega}^{1/2}\approx\boldsymbol{\omega}^{1/2}\boldsymbol{\Omega}\boldsymbol{\omega}^{1/2}\approx\mathbf{I}$, and the definitions of pertinent variables are used to develop the three expected values on the right-hand side of (6-3) as

$$2E(\mathbf{f}(\boldsymbol{\gamma\theta}_*)-\mathbf{f}_0(\boldsymbol{\gamma\theta}_*)-\mathbf{f}(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})+\mathbf{f}_0(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma\theta}_*))$$

$$\approx E(\sum_i(\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_i\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_i\mathbf{l}-2\mathbf{l}'\mathbf{D}^2 f_i\mathbf{q})\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j\omega_j^{\frac{1}{2}}(U_{*j}$$

$$+\frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_j-\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_j\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}')\mathbf{e}))$$

$$\approx -2E(\sum_i\mathbf{l}'\mathbf{D}^2 f_i\mathbf{q}\omega_i^{\frac{1}{2}}\mathbf{Z})$$

$$\approx -2E(\sum_i\mathbf{l}'\mathbf{D}^2 f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}(\sum_k\mathbf{D}^2 f_k\mathbf{l}\omega_k^{\frac{1}{2}}\mathbf{Z}+\frac{1}{2}\mathbf{Df}'\sum_j\omega_j(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))\omega_i^{\frac{1}{2}}\mathbf{Z})$$

$$\approx -2E(\sum_i\omega_i^{\frac{1}{2}}\mathbf{Z}_*\mathbf{l}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\sum_k\mathbf{D}^2 f_k\mathbf{l}_*\omega_k^{\frac{1}{2}}\mathbf{Z}_*)$$

$$+E(\sum_i\omega_i^{\frac{1}{2}}\mathbf{Z}_*\mathbf{l}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\boldsymbol{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\sum_j\omega_j\mathbf{l}_*'\mathbf{D}^2 f_j\mathbf{l}_*)$$

$$=-2\sum_i tr(\mathbf{C}_i^2)\sigma_\varepsilon^4 \qquad \text{(I-11)}$$

$$E(\mathbf{f}(\boldsymbol{\gamma\theta}_*)-\mathbf{f}_0(\boldsymbol{\gamma\theta}_*)-\mathbf{f}(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})+\mathbf{f}_0(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{f}(\boldsymbol{\gamma\theta}_*)-\mathbf{f}_0(\boldsymbol{\gamma\theta}_*)-\mathbf{f}_0(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}})+\mathbf{f}_0(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}}))$$

$$\approx\frac{1}{4}E(\sum_i(\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_i\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}-\mathbf{l}'\mathbf{D}^2 f_i\mathbf{l})\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j\omega_j^{\frac{1}{2}}(\mathbf{e}'\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{D}_\beta^2 f_j\boldsymbol{\gamma}(\boldsymbol{\gamma}'\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}'\mathbf{e}$$

$$-\mathbf{l}'\mathbf{D}^2 f_j\mathbf{l}))$$

$$\approx\frac{1}{4}E(\sum_i\mathbf{l}_*'\mathbf{D}^2 f_i\mathbf{l}_*\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R})\sum_j\omega_j^{\frac{1}{2}}\mathbf{l}_*'\mathbf{D}^2 f_j\mathbf{l}_*)$$

$$=\frac{1}{4}\sum_i tr^2(\mathbf{C}_i)\sigma_\varepsilon^4+\frac{1}{2}\sum_i tr(\mathbf{C}_i^2)\sigma_\varepsilon^4 \qquad \text{(I-12)}$$

$$E(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}}))'\boldsymbol{\omega}^{\frac{1}{2}}\mathbf{R}\boldsymbol{\omega}^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\boldsymbol{\gamma}\hat{\boldsymbol{\theta}}))$$

$$\approx E(\sum_i \omega_i^{\frac{1}{2}} \mathbf{Z} \mathbf{l}' \mathbf{D}^2 f_i (\mathbf{Df}'\omega\mathbf{Df})^{-1} \mathbf{Df}'\omega\mathbf{Df} (\mathbf{Df}'\omega\mathbf{Df})^{-1} \sum_k \mathbf{D}^2 f_k \mathbf{l} \omega_k^{\frac{1}{2}} \mathbf{Z})$$
$$\approx E(\sum_i \omega_i^{\frac{1}{2}} \mathbf{Z}_* \mathbf{l}_*' \mathbf{D}^2 f_i (\mathbf{Df}'\omega\mathbf{Df})^{-1} \sum_k \mathbf{D}^2 f_k \mathbf{l}_* \omega_k^{\frac{1}{2}} \mathbf{Z}_*)$$
$$= \sum_i tr(\mathbf{C}_i^2)\sigma_\varepsilon^4 \qquad \text{(I-13)}$$

Matrix $\mathbf{C}_i$ is defined by (G-3).

## Analysis of (6-19)

***Form for small model combined intrinsic nonlinearity.*** To show that the second and third terms on the right-hand side of (6-19) sum to zero when the model combined intrinsic nonlinearity is zero, an analysis similar to the one used to obtain (I-5) is used. First, a linear-model approximation of $\mathbf{f}(\gamma\tilde{\theta})$ analogous to (I-1) is

$$\mathbf{p}(\tilde{\psi}) = \mathbf{f}(\gamma\bar{\theta}) + \mathbf{D}_\phi \mathbf{f}(\tilde{\phi} - \bar{\phi})$$
$$= \mathbf{f}_0(\gamma\tilde{\theta}) + \mathbf{Df}\tilde{\psi} \qquad \text{(I-14)}$$

Next, the constraint is expressed using the same ideas:

$$g(\gamma\theta_*) - g(\gamma\tilde{\theta}) = g(\gamma\bar{\theta}) + \mathbf{D}g(\theta_* - \bar{\theta} + \psi_*) - g(\gamma\bar{\theta}) - \mathbf{D}g(\tilde{\theta} - \bar{\theta} + \tilde{\psi})$$
$$= g_0(\gamma\theta_*) - g_0(\gamma\tilde{\theta}) + \mathbf{D}g\psi \qquad \text{(I-15)}$$

where $\psi = \psi_* - \tilde{\psi}$. Finally, a constrained best fit of $\mathbf{p}(\psi) = \mathbf{p}(\psi_*) - \mathbf{p}(\tilde{\psi})$ to $\mathbf{f}(\gamma\theta_*) - \mathbf{f}(\gamma\tilde{\theta})$ is obtained by minimizing with respect to $\psi$ the Lagrangian function

$$L(\psi, \lambda) = (\mathbf{f}(\gamma\theta_*) - \mathbf{f}(\gamma\tilde{\theta}) - \mathbf{p}(\psi))'\omega(\mathbf{f}(\gamma\theta_*) - \mathbf{f}(\gamma\tilde{\theta}) - \mathbf{p}(\psi))$$
$$- 2\lambda(g(\gamma\theta_*) - g_0(\gamma\theta_*) - g(\gamma\tilde{\theta}) + g_0(\gamma\tilde{\theta}) - \mathbf{D}g\psi) \qquad \text{(I-16)}$$

The same method as used to solve for $\tilde{\mathbf{l}} - (\gamma'\gamma)^{-1}\gamma'\mathbf{e}$ in (E-7)-(E-14), appendix E, produces

$$\psi = -\left( \frac{(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{D}g'\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}} - (\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}} \right)\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*) - \mathbf{f}_0(\gamma\theta_*) - \mathbf{f}(\gamma\tilde{\theta}) + \mathbf{f}_0(\gamma\tilde{\theta}))$$
$$+ \frac{(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{D}g'}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\theta_*) - g_0(\gamma\theta_*) - g(\gamma\tilde{\theta}) + g_0(\gamma\tilde{\theta})) \qquad \text{(I-17)}$$

The constrained residual vector is

$$\mathbf{f}(\gamma\theta_*) - \mathbf{f}(\gamma\tilde{\theta}) - \mathbf{p}(\psi) = \mathbf{f}(\gamma\theta_*) - \mathbf{f}(\gamma\tilde{\theta}) - \mathbf{f}_0(\gamma\theta_*) + \mathbf{f}_0(\gamma\tilde{\theta})$$

$$+\mathbf{Df}(\left(\frac{(\mathbf{Df}\acute{\omega}\mathbf{Df})^{-1}\mathbf{D}g'\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}}-(\mathbf{Df}\acute{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}\right)\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}(\gamma\tilde{\theta})+\mathbf{f}_0(\gamma\tilde{\theta}))$$
$$-\frac{(\mathbf{Df}\acute{\omega}\mathbf{Df})^{-1}\mathbf{D}g'}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\theta_*)-g_0(\gamma\theta_*)-g(\gamma\tilde{\theta})+g_0(\gamma\tilde{\theta})))$$
$$=\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}(\gamma\tilde{\theta})+\mathbf{f}_0(\gamma\tilde{\theta}))$$
$$-\omega^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\theta_*)-g_0(\gamma\theta_*)-g(\gamma\tilde{\theta})+g_0(\gamma\tilde{\theta})) \qquad \text{(I-18)}$$

If $\mathbf{p}(\psi)$ as constrained by (I-15) fits $\mathbf{f}(\gamma\theta_*)-\mathbf{f}(\gamma\tilde{\theta})$ exactly, then the combined linear models $\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}_0(\gamma\tilde{\theta})+\mathbf{Df}\psi$ and $g_0(\gamma\theta_*)-g_0(\gamma\tilde{\theta})+\mathbf{D}g\psi$ are exact, indicating no model combined intrinsic nonlinearity. In this case (I-18), which composes the second and third terms on the right-hand side of (6-19), equals $\mathbf{0}$.

The last term on the right-hand side of (6-19) also is zero when there is no model combined intrinsic nonlinearity, which is shown as follows. If there is no model combined intrinsic nonlinearity, then the Lagrange multiplier formulation to solve for the constrained regression estimate $\tilde{\phi}$ is given by

$$L(\phi,\lambda)=(\mathbf{Y}-\mathbf{f}(\gamma\theta(\phi_*))-\mathbf{D}_\phi\mathbf{f}(\phi-\phi_*))'\omega(\mathbf{Y}-\mathbf{f}(\gamma\theta(\phi_*))-\mathbf{D}_\phi\mathbf{f}(\phi-\phi_*))$$
$$+2\lambda(g(\gamma\theta(\phi_*))-g(\gamma\theta(\phi_*))-\mathbf{D}_\phi g(\phi-\phi_*)) \qquad \text{(I-19)}$$

The solution for $\tilde{\phi}$ takes the form

$$\tilde{\phi}-\phi_*=(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega(\mathbf{Y}-\mathbf{f}(\gamma\theta(\phi_*)))$$
$$-(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta(\phi_*))) \qquad \text{(I-20)}$$

from which, using the constraint $\mathbf{D}_\phi g(\tilde{\phi}-\phi_*)=0$, the gradient is obtained as

$$(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega(\mathbf{Y}-\mathbf{f}(\gamma\theta(\phi_*))-\mathbf{D}_\phi\mathbf{f}(\tilde{\phi}-\phi_*))$$
$$-(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta(\phi_*)))-(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}_\phi g(\tilde{\phi}-\phi_*)$$
$$=(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega(\mathbf{Y}-\mathbf{f}(\gamma\theta(\tilde{\phi})))-(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}}\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}}\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta(\phi_*)))$$
$$-(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega^{\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}_\phi g(\mathbf{D}_\phi\mathbf{f}\acute{\omega}\mathbf{D}_\phi\mathbf{f})^{-1}\mathbf{D}_\phi\mathbf{f}'\omega\mathbf{D}_\phi\mathbf{f}(\tilde{\phi}-\phi_*)$$
$$=\mathbf{J}^{-1}(\mathbf{Df}\acute{\omega}\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))=\mathbf{0} \qquad \text{(I-21)}$$

Equation (I-21) shows that the last term of (6-19) is zero when there is no model combined intrinsic nonlinearity.

***Correspondance to perturbation form.*** The expansion of $\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta})$ given by (6-19) corresponds to the perturbation expansion given by (E-29). Expansion of each term in (6-19) to second order shows this result:

$$\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_*))$$
$$\approx \omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}(U_{*j}+\frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_j-\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e}) \tag{I-22}$$

$$\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}(\gamma\tilde{\theta})+\mathbf{f}_0(\gamma\tilde{\theta}))-\omega^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\theta_*)-g_0(\gamma\theta_*)$$
$$-g(\gamma\tilde{\theta})+g_0(\gamma\tilde{\theta}))$$
$$\approx\frac{1}{2}\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\tilde{\mathbf{l}})$$
$$-\frac{1}{2}\omega^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 g\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 g\,\tilde{\mathbf{l}}) \tag{I-23}$$

$$\omega^{-\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta}))$$
$$\approx-\omega^{-\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\,\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}}-\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*) \tag{I-24}$$

where (E-29) was used for $\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta})$ to obtain (I-24). Then substitution of (I-22)-(I-24) into (6-19) yields

$$\mathbf{Y}-\mathbf{f}(\gamma\tilde{\theta})=\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}\,(U_{*j}+\frac{1}{2}\mathbf{e}'(\mathbf{D}_\beta^2 f_j-\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})$$
$$+\frac{1}{2}\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 f_j\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 f_j\,\tilde{\mathbf{l}})$$
$$-\frac{1}{2}\omega^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_\beta^2 g\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^2 g\,\tilde{\mathbf{l}})$$
$$-\omega^{-\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}(\sum_k \mathbf{D}^2 f_k\tilde{\mathbf{l}}\,\omega_k^{\frac{1}{2}}\tilde{\mathbf{Z}}-\mathbf{D}^2 g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)$$
$$=\omega^{-\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{QQ}'}{\mathbf{Q}'\mathbf{Q}})\sum_j \omega_j^{\frac{1}{2}}\,(U_{*j}+\frac{1}{2}(\mathbf{e}'\mathbf{D}_\beta^2 f_j\mathbf{e}-\tilde{\mathbf{l}}'\,\mathbf{D}_\beta^2 f_j\tilde{\mathbf{l}}\,))$$

$$-\frac{1}{2}\omega^{-\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_{\beta}^{2}g\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^{2}g\,\tilde{\mathbf{l}})$$

$$-\omega^{-\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}(\sum_{k}\mathbf{D}^{2}f_{k}\tilde{\mathbf{l}}\,\omega_{k}^{\frac{1}{2}}\tilde{\mathbf{Z}}-\mathbf{D}^{2}g\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_{*}) \qquad \text{(I-25)}$$

which is (E-29).

***Approximate evaluation of terms in component correction factor.*** Finally, use of (I-22)-(I-24), (G-9)-(G-19), the assumption that $\omega^{1/2}\mathbf{V}_{*}\omega^{1/2}\approx\omega^{1/2}\Omega\omega^{1/2}=\mathbf{I}$, and the definitions of pertinent variables yields the three expected values on the right-hand side of (6-20) as

$$2E((\mathbf{f}(\gamma\theta_{*})-\mathbf{f}_{0}(\gamma\theta_{*})-\mathbf{f}(\gamma\tilde{\theta})+\mathbf{f}_{0}(\gamma\tilde{\theta}))'\omega^{\frac{1}{2}}-(g(\gamma\theta_{*})-g_{0}(\gamma\theta_{*})-g(\gamma\tilde{\theta})$$
$$+g_{0}(\gamma\tilde{\theta}))\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}-\mathbf{R}+\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\theta_{*}))$$
$$\approx E(\sum_{i}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_{\beta}^{2}f_{i}\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^{2}f_{i}\tilde{\mathbf{l}}-2\,\tilde{\mathbf{l}}'\mathbf{D}^{2}f_{i}\tilde{\mathbf{q}})\,\omega_{i}^{\frac{1}{2}}-(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_{\beta}^{2}g\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}$$
$$-\tilde{\mathbf{l}}'\mathbf{D}^{2}g\,\tilde{\mathbf{l}})\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}-\mathbf{R}+\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\sum_{j}\omega_{j}(U_{*j}+\frac{1}{2}\mathbf{e}'(\mathbf{D}_{\beta}^{2}f_{j}-\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}_{\beta}^{2}f_{j}\gamma(\gamma'\gamma)^{-1}\gamma')\mathbf{e})$$
$$\approx-2E(\sum_{i}\tilde{\mathbf{l}}'\mathbf{D}^{2}f_{i}\tilde{\mathbf{q}}\omega_{i}^{\frac{1}{2}}\tilde{\mathbf{Z}})$$
$$\approx-2E(\sum_{i}\tilde{\mathbf{l}}'\mathbf{D}^{2}f_{i}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}(\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}(\sum_{k}\mathbf{D}^{2}f_{k}\tilde{\mathbf{l}}\,\omega_{k}^{\frac{1}{2}}\tilde{\mathbf{Z}}$$
$$-\mathbf{D}^{2}g\,\tilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_{*})+\frac{1}{2}\sum_{j}\omega_{j}^{\frac{1}{2}}(\mathbf{e}'\mathbf{D}_{\beta}^{2}f_{j}\mathbf{e}-\tilde{\mathbf{l}}'\mathbf{D}^{2}f_{i}\tilde{\mathbf{l}}))-\frac{1}{2}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{D}g'(\tilde{\mathbf{l}}'\mathbf{D}^{2}g\,\tilde{\mathbf{l}}$$
$$-\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\mathbf{D}^{2}g(\gamma'\gamma)^{-1}\gamma'\mathbf{e}))\omega_{i}^{\frac{1}{2}}\tilde{\mathbf{Z}})$$
$$\approx-2E(\sum_{i}\omega_{i}^{\frac{1}{2}}\tilde{\mathbf{Z}}\tilde{\mathbf{l}}_{*}'\mathbf{D}^{2}f_{i}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}\mathbf{Df}\acute{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}\sum_{k}\mathbf{D}^{2}f_{k}\tilde{\mathbf{l}}_{*}\omega_{k}^{\frac{1}{2}}\tilde{\mathbf{Z}})$$
$$+2E(\sum_{i}\omega_{i}^{\frac{1}{2}}\tilde{\mathbf{Z}}\tilde{\mathbf{l}}_{*}'\mathbf{D}^{2}f_{i}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}\mathbf{Df}\acute{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}\mathbf{D}^{2}g\tilde{\mathbf{l}}_{*}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_{*})$$
$$+E(\sum_{i}\omega_{i}^{\frac{1}{2}}\tilde{\mathbf{Z}}\tilde{\mathbf{l}}_{*}'\mathbf{D}^{2}f_{i}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}\mathbf{Df}\acute{\omega}^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\sum_{j}\omega_{j}\tilde{\mathbf{l}}_{*}'\mathbf{D}^{2}f_{j}\tilde{\mathbf{l}}_{*})$$
$$+\frac{1}{\mathbf{Q}'\mathbf{Q}}E(\sum_{i}\omega_{i}^{\frac{1}{2}}\tilde{\mathbf{Z}}\tilde{\mathbf{l}}_{*}'\mathbf{D}^{2}f_{i}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}\mathbf{D}g'\tilde{\mathbf{l}}_{*}'\mathbf{D}^{2}g\tilde{\mathbf{l}}_{*})$$
$$=-2\sum_{i}tr(\tilde{\mathbf{C}}_{i}^{2})\sigma_{\varepsilon}^{4}+2E(\sum_{i}\omega_{i}^{\frac{1}{2}}\tilde{\mathbf{Z}}\tilde{\mathbf{l}}_{*}'\mathbf{D}^{2}f_{i}(\mathbf{Df}\acute{\omega}\,\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}$$
$$\bullet(\mathbf{Df}'\omega\mathbf{Df})^{-1}\mathbf{D}^{2}g\tilde{\mathbf{l}}_{*}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega_{i}^{\frac{1}{2}}\mathbf{U}_{*}) \qquad \text{(I-26)}$$

$$E((\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}(\gamma\widetilde{\theta})+\mathbf{f}_0(\gamma\widetilde{\theta}))'\omega^{\frac{1}{2}}-(g(\gamma\theta_*)-g_0(\gamma\theta_*)-g(\gamma\widetilde{\theta})+g_0(\gamma\widetilde{\theta}))\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}-\mathbf{R}$$
$$+\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\omega^{\frac{1}{2}}(\mathbf{f}(\gamma\theta_*)-\mathbf{f}_0(\gamma\theta_*)-\mathbf{f}(\gamma\widetilde{\theta})+\mathbf{f}_0(\gamma\widetilde{\theta}))-\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(g(\gamma\theta_*)-g_0(\gamma\theta_*)-g(\gamma\widetilde{\theta})+g_0(\gamma\widetilde{\theta})))$$
$$\approx\frac{1}{4}E(\sum_i(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_i\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\widetilde{\mathbf{l}}'\mathbf{D}^2 f_i\widetilde{\mathbf{l}})\omega_i^{\frac{1}{2}}-(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta g\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\widetilde{\mathbf{l}}'\mathbf{D}^2 g\widetilde{\mathbf{l}})\frac{\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\mathbf{I}$$
$$-\mathbf{R}+\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})(\sum_j\omega_j^{\frac{1}{2}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta f_j\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\widetilde{\mathbf{l}}'\mathbf{D}^2 f_j\widetilde{\mathbf{l}})-\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}(\mathbf{e}'\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{D}^2_\beta g\gamma(\gamma'\gamma)^{-1}\gamma'\mathbf{e}-\widetilde{\mathbf{l}}'\mathbf{D}^2 g\widetilde{\mathbf{l}}))$$
$$\approx\frac{1}{4}E(\sum_i\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i\widetilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}(\mathbf{I}-\mathbf{R}+\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\sum_j\omega_j^{\frac{1}{2}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_j\widetilde{\mathbf{l}}_*)-\frac{1}{2}E(\sum_i\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i\widetilde{\mathbf{l}}_*\omega_i^{\frac{1}{2}}\frac{\mathbf{Q}}{\mathbf{Q}'\mathbf{Q}}\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g\widetilde{\mathbf{l}}_*)$$
$$+\frac{1}{4}\frac{1}{\mathbf{Q}'\mathbf{Q}}E(\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g\widetilde{\mathbf{l}}_*)^2$$
$$=\frac{1}{4}\sum_i tr^2(\widetilde{\mathbf{C}}_i)\sigma_\varepsilon^4+\frac{1}{2}\sum_i tr(\widetilde{\mathbf{C}}_i^2)\sigma_\varepsilon^4-\frac{1}{2}\sum_i tr(\widetilde{\mathbf{C}}_i)tr(\widetilde{\mathbf{F}}_i)\sigma_\varepsilon^4-\sum_i tr(\widetilde{\mathbf{C}}_i\widetilde{\mathbf{F}}_i)\sigma_\varepsilon^4$$
$$+\frac{1}{4}\sum_i tr^2(\widetilde{\mathbf{F}}_i)\sigma_\varepsilon^4+\frac{1}{2}\sum_i tr(\widetilde{\mathbf{F}}_i^2)\sigma_\varepsilon^4=\frac{1}{4}\sum_i tr^2(\widetilde{\mathbf{C}}_i-\widetilde{\mathbf{F}}_i)\sigma_\varepsilon^4+\frac{1}{2}\sum_i tr((\widetilde{\mathbf{C}}_i-\widetilde{\mathbf{F}}_i)^2)\sigma_\varepsilon^4 \qquad \text{(I-27)}$$

$$E(\mathbf{Y}-\mathbf{f}(\gamma\widetilde{\theta}))'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}(\mathbf{Y}-\mathbf{f}(\gamma\widetilde{\theta}))$$
$$=E(\sum_i\mathbf{D}^2 f_i\widetilde{\mathbf{l}}\,\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}}-\mathbf{D}^2 g\widetilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)'(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\,\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}$$
$$\bullet(\sum_k\mathbf{D}^2 f_k\widetilde{\mathbf{l}}\,\omega_k^{\frac{1}{2}}\widetilde{\mathbf{Z}}-\mathbf{D}^2 g\widetilde{\mathbf{l}}\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)$$
$$\approx E(\sum_i\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}}\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\sum_k\mathbf{D}^2 f_k\widetilde{\mathbf{l}}_*\omega_k\widetilde{\mathbf{Z}})$$
$$-2E(\sum_i\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}}\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{D}^2 g\,\widetilde{\mathbf{l}}_*\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)$$
$$+E((\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)^2\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 g(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{D}^2 g\widetilde{\mathbf{l}}_*)$$
$$=\sum_i tr(\widetilde{\mathbf{C}}_i^2)\sigma_\varepsilon^4-2E(\sum_i\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}}\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{D}^2 g\,\widetilde{\mathbf{l}}_*$$
$$\bullet\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*)+\frac{1}{\mathbf{Q}'\mathbf{Q}}tr(\widetilde{\mathbf{A}}^2)\sigma_\varepsilon^4$$
$$=\sum_i tr(\widetilde{\mathbf{C}}_i^2)\sigma_\varepsilon^4+\sum_i tr(\widetilde{\mathbf{F}}_i^2)\sigma_\varepsilon^4$$
$$-2E(\sum_i\omega_i^{\frac{1}{2}}\widetilde{\mathbf{Z}}\,\widetilde{\mathbf{l}}_*'\mathbf{D}^2 f_i(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{Df}'\omega^{\frac{1}{2}}(\mathbf{R}-\frac{\mathbf{Q}\mathbf{Q}'}{\mathbf{Q}'\mathbf{Q}})\omega^{\frac{1}{2}}\mathbf{Df}(\mathbf{Df}'\omega\,\mathbf{Df})^{-1}\mathbf{D}^2 g\widetilde{\mathbf{l}}_*\frac{1}{\mathbf{Q}'\mathbf{Q}}\mathbf{Q}'\omega^{\frac{1}{2}}\mathbf{U}_*) \qquad \text{(I-28)}$$

where $\tilde{\mathbf{C}}_i$, $\tilde{\mathbf{F}}_i$, and $\tilde{\mathbf{A}}$ are defined by (G-10), (G-12), and (G-13), respectively.

# REFERENCES

Anderson, M.P., and Woesner, W.W., 1992, Applied groundwater modeling: San Diego, California, Academic, 381 p.

Anderson, T.W., 1958, An introduction to multivariate statistical analysis: New York, John Wiley, 374 p.

Bakr, A.A., Gelhar, L.W., Gutjahr, A.L., and MacMillan, J.R., 1978, Stochastic analysis of spatial variability in subsurface flows. 1. Comparison of one- and three-dimensional flows: Water Resources Research, v. 14, no. 2, p. 263-271.

Barlebo, H.C., Hill, M.C., Rosbjerg, Dan, and Jensen, K.H., 1998, Concentration data and dimensionality in groundwater models – evaluation using inverse modeling: Nordic Hydrology, v. 29, p. 149-178.

Bates, D.M., and Watts, D.G., 1980, Relative curvature measures of nonlinearity: Journal of the Royal Statistical Society, Series B, v. 42, no. 1, p. 1-25.

Beale, E.M.L., 1960, Confidence regions in non-linear estimation: Journal of the Royal Statistical Society, Series B, v. 22, no. 1, p. 41-76.

Beck, J.V., and Arnold, K.J., 1977, Parameter estimation in engineering and science: New York, John Wiley, 501 p.

Boas, M.L., 1966, Mathematical methods in the physical sciences: New York, John Wiley, 778 p.

Carrera, Jesus, 1988, State of the art of the inverse problem applied to the flow and solute transport equations, in Custudio, E., Gurgui, A., Lobo Ferreira, J.P., eds., Groundwater flow and quality modelling: Hingham, Massachusetts, D. Reidel, p. 549-583.

Carrera, Jesus, and Neuman, S.P., 1986, Estimation of aquifer parameters under transient and steady-state conditions. 1. Maximum likelihood method incorporating prior information: Water Resources Research, v. 22, no. 2, p. 199-210.

Christensen, Steen, and Cooley, R.L., 1999a, Evaluation of confidence intervals for a steady-state leaky aquifer model: Advances in Water Resources, v. 22, no. 8, p. 807-819.

_____ 1999b, Evaluation of prediction intervals for expressing uncertainties in groundwater flow model predictions: Water Resources Research, v. 35, no. 9, p. 2627-2639.

Cooley, R.L., 1979, A method of estimating parameters and assessing reliability for models of steady-state groundwater flow. 2. Application of statistical analysis: Water Resources Research, v. 15, no. 3, p. 603-617.

_____ 1985, Conceptual framework for regression modeling of ground-water flow, *in* Torno, H.C., ed., Computer applications in water resources – Proceedings of the Specialty Conference on Computer Applications in Water Resources, Buffalo, New York, June 10-12, 1985: New York, American Society of Civil Engineers, p. 891-896.

_____ 1997, Confidence intervals for ground-water models using linearization, likelihood, and bootstrap methods: Ground Water, v. 35, no. 5, p. 869-880.

_____ 1999, Practical Scheffé-type credibility intervals for variables of a groundwater model: Water Resources Research, v. 35, no. 1, p. 113-126.

_____ 2000, An analysis of the pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: Water Resources Research, v. 36, no. 4, p. 1159-1163.

Cooley, R.L., and Hill, M.C., 1992, A comparison of three Newton-like nonlinear least-squares methods for estimating parameters of ground-water flow models, *in* Russell, T.F., Ewing, R.E., Brebbia, C.A., Gray, W.G., and Pinder, G.F., eds., Computational methods in water resources IX, v. 1, Numerical methods in water resources – Proceedings of the International Conference on Computational Methods in Water Resources, 9th, Denver, Colorado, 1992: Oxford, UK, Computational Mechanics Publications, p. 379-386.

Cooley, R.L., Konikow, L.F., and Naff, R.L., 1986, Nonlinear-regression modeling of a deep regional aquifer system: Water Resources Research, v. 22, no. 13, p. 1759-1778.

Cooley, R.L., and Naff, R.L., 1990, Regression modeling of ground-water flow: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. B4, 232p.

Dagan, Gedeon, 1986, Statistical theory of groundwater flow and transport – Pore to laboratory, laboratory to formation, formation to regional scale: Water Resources Research, v. 22, no. 9, p. 1205-1345.

D'Agnese, F.A., Faunt, C.C., Hill, M.C., and Turner, A.K., 1999, Death Valley regional ground-water flow model calibration using optimal parameter estimation methods and geoscientific information systems: Advances in Water Resources, v. 22, no. 8, p. 777-790.

Delhomme, J.P., 1979, Spatial variability and uncertainty in ground-water flow parameters – a geostatistical approach: Water Resources Research, v. 15, no. 2, p. 269-280.

Dettinger, M.D., and Wilson, J.L., 1981, First-order analysis of uncertainty in numerical models of groundwater flow. 1. Mathematical development: Water Resources Research, v. 17, no. 1, p. 149-161.

Di Federico, Vittorio, Neuman, S.P., and Tartakovsky, D.M., 1999, Anisotropy, lacunarity, and upscaled conductivity and its autocovariance in multiscale random fields with truncated power variograms: Water Resources Research, v. 35, no. 10, p. 2891-2908.

Draper, N.R., and Smith, Harry, 1998, Applied regression analysis (3rd ed.): New York, John Wiley, 706 p.

Efron, Bradley, 1982, The jackknife, the bootstrap, and other resampling plans: Society for Industrial and Applied Mathematics, Regional Conference Series in Applied Mathematics, v. 38, 92 p.

Freeze, R.A., 1975, A statistical-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media: Water Resources Research, v. 11, no. 5, p. 725-741.

Gelhar, L.W., 1986, Stochastic subsurface hydrology from theory to applications: Water Resources Research, v. 22, no. 9, p. 135S-145S.

Ginn, T.R., and Cushman, J.H., 1990, Inverse methods for subsurface flow – A critical review of stochastic techniques: Stochastic Hydrology and Hydraulics, v. 4, p. 1-26.

Gomez-Hernandez, J.J., and Gorelick, S.M., 1989, Effective groundwater model values – influence of spatial variability of hydraulic conductivity, leakance, and recharge: Water Resources Research, v. 25, no. 3, p. 405-419.

Graham, W.D. and Tankersley, C.D., 1994a, Optimal estimation of spatially variable recharge and transmissivity fields under steady-state groundwater flow. Part 1. Theory: Journal of Hydrology, v. 157, p. 247-266.

_____1994b, Optimal estimation of spatially variable recharge and transmissivity fields under steady-state groundwater flow. Part 2. Case study: Journal of Hydrology, v. 157, p. 267-285.

Graybill, F.A., 1976, Theory and application of the linear model: Pacific Grove, Calif., Wadsworth and Brooks/Cole, 704 p.

Guttman, Irwin, and Meeter, D.A., 1965, On Beale's measures of non-linearity: Technometrics, v. 7, no. 4, p. 623-637.

Hamilton, David, and Wiens, Douglas, 1987, Correction factors for *F* ratios in nonlinear regression: Biometrika, v. 74, no. 2, p. 423-425.

Hill, M.C., 1992, A computer program (MODFLOWP) for estimating parameters of a transient, three-dimensional, ground-water flow model using nonlinear regression: U.S. Geological Survey Open-File Report 92-484, 358 p.

_____ 1998, Methods and guidelines for effective model calibration: U.S. Geological Survey Water-Resources Investigations Report 98-4005, 90 p.

Hill, M.C., Cooley, R.L., and Pollock, D.W., 1998, A controlled experiment in ground water flow model calibration: Ground Water, v. 36, no. 3, p. 520-535.

Hohn, F.E., 1964, Elementary matrix algebra (2nd ed.): New York, Macmillan, 395 p.

Johansen, Søren, 1983, Some topics in regression: Scandinavian Journal of Statistics, v. 10, p. 161-194.

Jones, D.A., 1983, Statistical analysis of empirical models fitted by optimization: Biometrika, v. 70, no. 1, p. 67-88.

Kitanidis, P.K, 1990, Effective hydraulic conductivity for gradually varying flow: Water Resources Research, v. 26, no. 6, p. 1197-1208.

_____ 1995, Quasi-linear geostatistical theory for inversing: Water Resources Research, v. 31, no. 10, p. 2411-2419.

_____ 1997, Introduction to geostatistics: New York, Cambridge University, 249 p.

Konikow, L.F, 1978, Calibration of ground-water models, *in* Verification of mathematical and physical models in hydraulic engineering – Proceedings of the 26th Annual Hydraulics Division Specialty Conference, College Park, Maryland, August 9-11, 1978: New York, American Society of Civil Engineers, p. 87-93.

Kuiper, Logan, 1994, Nonlinear-regression flow model of the Gulf Coast aquifer systems in the south-central United States: U.S. Geological Survey Water Resources Investigations Report 93-4020, 171 p.

Linssen, H.N., 1975, Nonlinearity measures – a case study: Statistica Neerlandica, v. 29, p. 93-99.

McLaughlin, Dennis, and Townley, L.R., 1996, A reassessment of the groundwater inverse problem: Water Resources Research, v. 32, no. 5, p. 1131-1161.

Miller, R.L., and Kahn, J.S., 1962, Statistical analysis in the geological sciences: New York, John Wiley, 483 p.

Neuman, S.P., and Orr, Shlomo, 1993, Prediction of steady state flow in nonuniform geologic media by conditional moments – exact formalism, effective conductivities, and weak approximations: Water Resources Research, v. 29, no. 2, p. 341-364.

Obenchain, R.L., 1975, Residual optimality – ordinary vs. weighted vs. biased least squares: Journal of the American Statistical Association, v. 70, no. 350, p. 375-379.

Papoulis, Athanasios, 1965, Probability, random variables, and stochastic processes: New York, McGraw-Hill, 583 p.

Peck, Adrian, Gorelick, Steven, de Marsily, Ghislain, Foster, Stephen, and Kovalevsky, Vladimir, 1988, Consequences of spatial variability in aquifer properties and data limitations for groundwater modelling practice: International Association of Hydrological Sciences Publication No. 175, 272 p.

RamaRao, B.S., LaVenue, A.M., de Marsily, Ghislain, and Marietta, M.G., 1995, Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 1. Theory and computational experiments: Water Resources Research, v. 31, no. 3, p. 475-493.

Seber, G.A.F., and Wild, C.J., 1989, Nonlinear regression: New York, John Wiley, 768 p.

Smith, Leslie, and Freeze, R.A., 1979, Stochastic analysis of steady state groundwater flow in a bounded domain. 1. One-dimensional simulations: Water Resources Research, v. 15, no. 3, p. 521-528.

Tartakovsky, D.M., and Neuman, S.P., 1998, Transient flow in bounded randomly heterogeneous domains. 1. Exact conditional moment equations and recursive approximations: Water Resources Research, v. 34, no. 1, p. 1-12.

Tasker, G.D., and Stedinger, J.R., 1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 111, p. 361-375.

Vecchia, A.V., and Cooley, R.L., 1987, Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model: Water Resources Research, v. 23, no. 7, p. 1237-1250.

Yager, R.M., 1996, Simulated three-dimensional ground-water flow in the Lockport Group, a fractured-dolomite aquifer near Niagara Falls, New York: U.S. Geological Survey Water Supply Paper 2487, 41 p.

Yeh, J.T-C., Jin, Minghui, and Hanna, Samuel, 1996, An iterative stochastic inverse method – conditional effective transmissivity and hydraulic head fields: Water Resources Research, v. 32, no. 1, p. 85-92.

Yeh, W. W-G., 1986, Review of parameter identification procedures in groundwater hydrology – the inverse problem: Water Resources Research, v. 22, no. 2, p. 95-108.

Yoon, Y.S., and Yeh, W. W-G., 1976, Parameter identification in an inhomogeneous medium with the finite element method: Society of Petroleum Engineers Journal, August, 1976, p. 217-226.