Volume Title: Annals of Economic and Social Measurement, Volume 2, number 4

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm73-4

Publication Date: October 1973

Chapter Title: The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression

Chapter Author: Barr Rosenberg

Chapter URL: http://www.nber.org/chapters/c9934

Chapter pages in book: (p. 399 - 428)

Annals of Economic and Social Measurement, 2/4, 1973

RANDOM COEFFICIENTS MODELS

THE ANALYSIS OF A CROSS SECTION OF TIME SERIES BY STOCHASTICALLY CONVERGENT PARAMETER REGRESSION[1]

BY BARR ROSENBERG

This paper develops a "convergent-parameter" regression model for a cross section of time series. Cross-sectional diversity in the regression parameters results from sequential random increments to the individual parameters. These random walks are subordinated to a continual tendency for individual parameters to converge to the population norm. The model implies stationary cross-sectional parameter dispersion, with nonconstant but serially correlated individual parameters. Maximum likelihood and Bayesian estimation methods are derived for the model. An approximation that makes the computations feasible is evaluated and found to be satisfactorily efficient. The estimators are compared with ordinary least squares.

I. THE "CONVERGENT PARAMETER" MODEL

A. Consider the familiar cross-section, time-series regression problem, where an endogenous variable y and exogenous variables $x_1, \ldots, x_k$ are observed for each of N individuals, $n = 1, \ldots, N$ in each of T time periods, $t = 1, \ldots, T$. The regression parameters $b_1, \ldots, b_k$ are the partial derivatives of the endogenous with respect to the exogenous variables. The parameter vector $\mathbf{b}_{nt} = (b_{1nt} : \ldots : b_{knt})'$ specific to individual n in period t is determined by the behavior and environment of that individual at that date. In most economic applications, it is unreasonable to expect these parameters to be the same for all individuals in all periods.

A variety of cross section, time series regression models have previously introduced stochastic variation in individual parameters. The most widely known methods are extensions of the analysis of covariance: shifts in the intercept term are associated with each individual ("individual effects") and with each time period ("time effects"). Sometimes these shifts in the intercept are introduced as dummy variables, or equivalently, as stochastic terms with diffuse prior distributions (Hildreth (1949, 1950), Hoch (1962), Wilks (1943: 195–200)). In other applications, these shifts are treated as stochastic terms with proper prior distributions, or "error components" (Wallace and Hussain (1969)). Serial correlation in individual disturbances may be superimposed upon these models (Parks (1967)). However, this class of models has the deficiency of postulating that regression parameters other than the intercept are identical for all individuals in all periods.

Where regression parameters do vary, an estimator assuming constant parameters has two important defects. First, the estimator is inefficient and the associated sampling theory is invalid, usually leading to downward-biased estimates of error variance. Second, when the pattern of parameter variation is of interest in

[1] The bulk of this research, reported in "Varying Parameter Regression in the Analysis of a Cross Section of Time Series," IBER Working Paper No. IP 165, 1969 (revised 1973), was completed under NSF Grant GS 2102, aided by subsidized funds of the Computer Center, University of California, Berkeley. The research was completed under NSF Grant GS 3306. The resourceful assistance of Daryl Carlson, and the indomitable work of Mrs. Ellen McGibbon in preparing various stages of the manuscript, are gratefully acknowledged.

its own right, a constant parameter model is totally incapable of shedding light on this aspect of the economic process.

Two models have introduced more general parameter variation. In Swamy's work, individual parameters are randomly dispersed across the population, but are constant over time (1970, 1971). In Hsiao's recent paper (1973), regression parameters are the sums of "individual effects" and "time effects," so that the model extends to the regression parameters the methods previously applied to the intercept term alone. These two approaches are appealing. However, they do not allow the individual parameters to vary independently of the rest of the population. If individual parameters do vary stochastically, these methods cannot track the individual parameter vectors nor model the stochastic variations.

B. What pattern of parameter variation can be expected in a cross section of economic decision units? There are certainly tendencies for different individuals' parameters to be alike. Social interaction within a population tends to preserve similarity among individuals playing the same role. When conformity is highly valued, or when the role of a deviate is, for any reason, difficult, individuals will tend to converge in behavior and in environment toward group norms, or toward subgroup norms if a deviant subgroup coalesces. Under competition, individuals will strive for profitable differentiation from the population, but as soon as such differentiation is achieved, competitive responses by others will tend to offset it. Uniformity may be enforced by institutional devices, such as trade organizations, or may result from interdependent individual responses to similar environments, as, for example, in loosely organized groups such as consumers.

On the other hand, within a group of individuals, each being somewhat different in innate characteristics and in environment, freedom of action will facilitate continual developments which are in opposition to, or at least independent of, the converging trends. These independent events will be a source of diversity which, when balanced against the conforming forces, may preserve a relatively stable degree of differentiation in the population. Individual characteristics will be different, but will not remain constant over time. The differences may behave as if subjected to sequential random increments and as if continually converging toward zero from the position randomly arrived at. Individual differences will then be serially correlated but nonconstant.

To fix ideas, it may be helpful to consider an example. In analyzing the returns to stockholders, it is useful to write for each stock in a universe of N stocks and for each holding period within a sequence of T holding periods:

$$r_{nt} = b_{0nt} + b_{1nt}r_{Mt} + b_{2nt}f_{2t} + \dots + b_{k-1,nt}f_{k-1,t} + u_{nt},$$
$$n = 1, \dots, N, \quad t = 1, \dots, T$$

where r_{nt} is the (excess) return on stock n over holding period t, r_{Mt} is the (excess) return on a stock market index in period t, and the f_{it}, $i = 2, \dots, k-1$, are other major economic or social factors which influence the returns on securities. The coefficient b_1, widely known in finance as the stock's "beta," is a partial derivative with respect to return on the index. The "beta" and the other coefficients are important in the theory and practice of investment management, since they determine the risk of a diversified portfolio (see, for example, Sharpe (1970)). The "beta," in particular, has been widely studied empirically. It has been shown that "beta," for any security, is serially correlated but nonconstant. A possible stochastic model for "beta" is:

$$b_{nt} = (1 - \phi)\bar{b}_n + \phi b_{n,t-1} + \varepsilon_{nt}.$$

The autoregressive parameter ϕ induces serial correlation, the term $(1 - \phi)$ implements a tendency to converge toward a normal value $\bar{b}_n$, and the serially independent random increments ε introduce stochastic variation over time. The characteristics of this process have been studied by Rosenberg and Ohlson (1973). The results support the model, and, in particular, show significant nonconstancy in beta and confirm the tendency of beta to converge toward a normal value $\bar{b}_n$.

This paper is concerned with the case where the normal value is a population norm common to several individuals. Every individual parameter vector is regarded as the sum of a population mean parameter vector and an individual difference, with the latter tending to converge toward zero.

Each individual difference is assumed to converge at the same rate and to be subject to random shocks of the same variance. This is clearly an oversimplification as a model of many economic processes. For example, in a study of competition in the computer industry, one would suspect that the tendency of IBM to converge toward the group norm would differ from other firms. Also, in many populations, individuals fall naturally into subgroups, so that a two-level hierarchy, in which individuals converge toward subgroup norms and subgroups may or may not converge toward the population norm, may be more appropriate. Nevertheless, the simple convergence structure is used here for several reasons.

One reason is heuristic: although the computational difficulty of the estimation problem does not increase as the convergence patterns become more complex, the notation becomes more painful. A second reason is one of operational usefulness. When the stochastic parameter process is known a priori, as it may be when the process determining behavioral modifications is well understood, it is quite possible to operate in the fully general framework. However, when the parameter process is to be estimated from the data, a simple structure must be postulated. The simplification that all individual parameters have convergence and stochastic-shift characteristics which are identical and unchanging over time is analogous to the traditional regression assumption that all parameters are identical, in that it asserts a similarity across the population which is necessary to develop an operationally feasible method. However, while the assumption of fixed parameters was originally thought to be needed before computations could be carried out at all, here the simplifying assumption is imposed, not by computational necessity, but by the experimenter's ignorance as to the exact nature of the parameter process.

There may also be events which induce simultaneous shifts in all of the individual parameters. It will be assumed that the effects of these constitute a series of serially independent communal increments occurring in all parameter vectors.

The individual parameter vector may contain both parameters which vary across the population ("cross-varying parameters") and parameters which are the same for all individuals in any time period ("cross-fixed parameters"). Accordingly, each k-element individual parameter vector is partitioned as $\mathbf{b}_{nt} = (\mathbf{c}_t' : \mathbf{a}_{nt}')'$, where $\mathbf{c}_t$ is a (possibly empty) κ-element subvector of cross-fixed parameters and $\mathbf{a}_{nt}$ is a

λ-element subvector of cross-varying parameters, with $k = \kappa + \lambda$. The explanatory variables $x_1, \ldots, x_k$ are partitioned correspondingly, with $w_1, \ldots, w_\kappa$, the explanatory variables having cross-fixed coefficients, and $z_1, \ldots, z_\lambda$, the explanatory variables having cross-varying coefficients. Let $\bar{\mathbf{b}}_t \equiv \begin{pmatrix} \mathbf{c}_t \\ \bar{\mathbf{a}}_t \end{pmatrix} \equiv \sum_{n=1}^{N} \mathbf{b}_{nt}/N$ be the population mean parameter vector.

The convergent parameter regression structure then takes the form:

$$(1) \qquad y_{nt} = \sum_{i=1}^{\kappa} w_{int}c_{it} + \sum_{j=1}^{\lambda} z_{jnt}a_{jnt} + u_{nt} \qquad t = 1, \ldots, T, n = 1, \ldots, N$$

$$E(u_{nt}) = 0 \qquad E(u_{ms}u_{nt}) = \delta_{st}\sigma^2(\delta_{mn}R_n + R_G)$$

or in vector notation,

$$y_{nt} = (\mathbf{w}'_{nt} : \mathbf{z}'_{nt})\begin{pmatrix} \mathbf{c}_t \\ \mathbf{a}_{nt} \end{pmatrix} + u_{nt} = \mathbf{x}'_{nt}\mathbf{b}_{nt} + u_{nt}.$$

Parameter Transition Relations:

$$(2) \qquad \mathbf{c}_{t+1} = \mathbf{c}_t + \boldsymbol{\gamma}_t \qquad t = 1, \ldots, T-1$$

and

$$(3) \qquad \mathbf{a}_{n,t+1} = \bar{\mathbf{a}}_t + \boldsymbol{\Delta}_\phi(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t) + \boldsymbol{\eta}_{nt} \qquad t = 1, \ldots, T-1, n = 1, \ldots, N$$

where $\quad E(\boldsymbol{\gamma}_t) = 0 \qquad E(\boldsymbol{\gamma}_s\boldsymbol{\gamma}'_t) = \delta_{st}\sigma^2\mathbf{Q}_c$

$$E(\boldsymbol{\eta}_{nt}) = 0 \qquad E(\boldsymbol{\eta}_{ms}\boldsymbol{\eta}'_{nt}) = \delta_{st}\sigma^2(\delta_{mn}\mathbf{Q}_a + \mathbf{Q}_G)$$

and $\quad E(u_{ms}\boldsymbol{\gamma}'_t) = 0 \qquad E(u_{ms}\boldsymbol{\eta}'_{nt}) = 0 \qquad E(\boldsymbol{\gamma}_s\boldsymbol{\eta}'_{nt}) = \delta_{st}\sigma^2\mathbf{Q}_{ca}.$

Here, δ_{ij} is the Kronecker delta equal to 1 if $i = j$, equal to zero otherwise. The disturbances are assumed to be serially uncorrelated, and to be composed of a communal disturbance with variance $\sigma^2 R_G \geq 0$, and uncorrelated individual terms with possibly heteroscedastic variances $\sigma^2 R_n$, $n = 1, \ldots, N$, with $R_n > 0$ for all n. The cross-fixed parameter vector is subject to serially uncorrelated increments having mean zero and variance matrix $\sigma^2\mathbf{Q}_c$. The convergence matrix $\boldsymbol{\Delta}_\phi$ is diagonal with diagonal entries ϕ_i, $0 \leq \phi_i < 1$, for $i = 1, \ldots, \lambda$. These diagonal entries are "convergence rates," in that ϕ_j is the proportion of the individual divergence $a_{jnt} - \bar{a}_{jt}$ which survives to period $t + 1$. The cross-varying parameter vectors are subject to serially uncorrelated individual parameter shifts. Each shift is the sum of a communal component with zero mean and variance matrix $\sigma^2\mathbf{Q}_G$ and an individual component with zero mean and variance matrix $\sigma^2\mathbf{Q}_a$. The disturbances are uncorrelated with the parameter process. The contemporaneous covariance between the cross-fixed parameter shift vector and any individual cross-varying parameter shift vector, or, equivalently, the covariance between the cross-fixed parameter shift and the communal component of the cross-varying parameter shifts, is $\sigma^2\mathbf{Q}_{ca}$. The variance matrices of parameter shifts may be positive semi-definite, permitting some parameters to remain fixed over time. All stochastic terms are assumed to be independent of the exogenous variables.

C. It is important for some purposes to view all individual parameter vectors as components of a single "grand parameter vector" $\boldsymbol{\beta}_t = (\mathbf{c}' : \mathbf{a}'_1 : \ldots : \mathbf{a}'_N)'_t$, with dimension $K = N\lambda + \kappa$. All the individual regressions in each period make up a single regression for the grand parameter vector

$$
(4) \qquad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}_t = \begin{pmatrix} \mathbf{w}_1' & \mathbf{z}_1' & & \\ \mathbf{w}_2' & & \mathbf{z}_2' & \mathbf{0} \\ \vdots & & \mathbf{0} & \ddots \\ \mathbf{w}_N' & & & \mathbf{z}_N' \end{pmatrix}_t \begin{pmatrix} \mathbf{c} \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_N \end{pmatrix}_t + \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}_t \qquad t = 1, \ldots, T
$$

or

$$
\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta}_t + \mathbf{u}_t, \qquad E[\mathbf{u}_t \mathbf{u}_t'] = \sigma^2 \mathbf{R},
$$

$$
\mathbf{R} = \begin{pmatrix} R_1 + R_G & R_G & \cdots & R_G \\ R_G & R_2 + R_G & \cdots & R_G \\ \vdots & \vdots & & \vdots \\ R_G & \cdots & R_G & \cdots\; R_N + R_G \end{pmatrix} = \begin{pmatrix} R_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & R_N \end{pmatrix} + R_G \boldsymbol{\iota}\boldsymbol{\iota}',
$$

where $\boldsymbol{\iota}$ denotes a vector of units. The parameter transition relations coalesce similarly into a single transition relation

$$
(5) \qquad \begin{pmatrix} \mathbf{c} \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_N \end{pmatrix}_{t+1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_\phi + \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} & \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} & \cdots & \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} \\ \mathbf{0} & \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} & \boldsymbol{\Delta}_\phi + \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} & \cdots & \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} & \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} & \cdots & \boldsymbol{\Delta}_\phi + \dfrac{(\mathbf{I} - \boldsymbol{\Delta}_\phi)}{N} \end{pmatrix} \times \begin{pmatrix} \mathbf{c} \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_N \end{pmatrix}_t + \begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_N \end{pmatrix}_t
$$

or

$$
\boldsymbol{\beta}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\beta}_t + \mathbf{d}_t, \qquad E[\mathbf{d}_t \mathbf{d}_t'] = \sigma^2 \mathbf{Q}
$$

where

$$
\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_c & \mathbf{Q}_{ca} & \mathbf{Q}_{ca} & \cdots & \mathbf{Q}_{ca} \\ \mathbf{Q}_{ca}' & \mathbf{Q}_a + \mathbf{Q}_G & \mathbf{Q}_G & \cdots & \mathbf{Q}_G \\ \mathbf{Q}_{ca}' & \mathbf{Q}_G & \mathbf{Q}_a + \mathbf{Q}_G & \cdots & \mathbf{Q}_G \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}_{ca}' & \mathbf{Q}_G & \mathbf{Q}_G & & \mathbf{Q}_a + \mathbf{Q}_G \end{pmatrix}
$$

D. One important property of the convergent parameter model is the stationary cross-sectional parameter dispersion which it generates. If the cross section of the individual parameter vectors is examined in any one period, for any individual n,

$$\mathbf{a}_{n,t+1} - \bar{\mathbf{a}}_{t+1} = \Delta_\phi(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t) + \boldsymbol{\eta}_{nt} - \bar{\boldsymbol{\eta}}_t. \tag{6}$$

Since parameter shifts between periods t and $t + 1$ are uncorrelated with the parameters in period t,

$$E[(\mathbf{a}_{n,t+1} - \bar{\mathbf{a}}_{t+1})(\mathbf{a}_{n,t+1} - \bar{\mathbf{a}}_{t+1})'] = \Delta_\phi E[(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t)(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t)']\Delta'_\phi + \frac{N-1}{N}(\sigma^2\mathbf{Q}_a) \tag{7}$$

and for $m \neq n$

$$E[(a_{m,t+1} - \bar{\mathbf{a}}_{t+1})(\mathbf{a}_{n,t+1} - \bar{\mathbf{a}}_{t+1})'] = \Delta_\phi E[(\mathbf{a}_{mt} - \bar{\mathbf{a}}_t)(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t)']\Delta'_\phi - \frac{1}{N}(\sigma^2\mathbf{Q}_a). \tag{8}$$

Since Δ_ϕ is diagonal, the stationary solutions to these difference equations are easily found to be:

$$\{E[(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t)(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t)']\}_{ij} = \frac{N-1}{N}\,\frac{\sigma^2\{\mathbf{Q}_a\}_{ij}}{1 - \phi_i\phi_j} \tag{9}$$

$$\{E[(\mathbf{a}_{mt} - \bar{\mathbf{a}}_t)(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t)']\}_{ij} = \frac{-1}{N}\,\frac{\sigma^2\{\mathbf{Q}_a\}_{ij}}{1 - \phi_i\phi_j} \quad \text{for } m \neq n \tag{10}$$

where $\{\mathbf{A}\}_{ij}$ denotes element (i, j) in the matrix $\mathbf{A}$. Since the eigenvalues of Δ_ϕ are smaller than one, this is, indeed, the stationary joint distribution of the cross-varying parameter vectors about their sample mean in any single time period. Notice that the dispersion about the sample mean is identical to that in a sample of vectors drawn independently from a multivariate population with variance matrix $\sigma^2\boldsymbol{\Omega}$ given by

$$\omega_{ij} \equiv \{\boldsymbol{\Omega}\}_{ij} = \frac{\{\mathbf{Q}_a\}_{ij}}{1 - \phi_i\phi_j}. \tag{11}$$

Thus, in any single cross section, the individual cross-varying parameter vectors in a convergent-parameter structure are distributed as if randomly drawn from a population with dispersion matrix $\sigma^2\boldsymbol{\Omega}$. Cross-sectional regressions of this kind, often called random or randomly dispersed parameter regressions, have been studied previously (Rao (1965), Swamy (1970), Rosenberg (1973a)).

The parameter interrelationships in the convergent-parameter model are diagrammed in two ways in Figure 1. In both diagrams, a link between vectors denotes a transition relation. Figure 1a exhibits the interrelationships among individual parameter vectors. At the top of the diagram is a representation of the stationary joint distribution of the individual parameter vectors in the initial period. The vector $\mathbf{b}_0$ is brought in as the mean of the hypothetical multivariate population from which the initial parameter vectors are drawn.

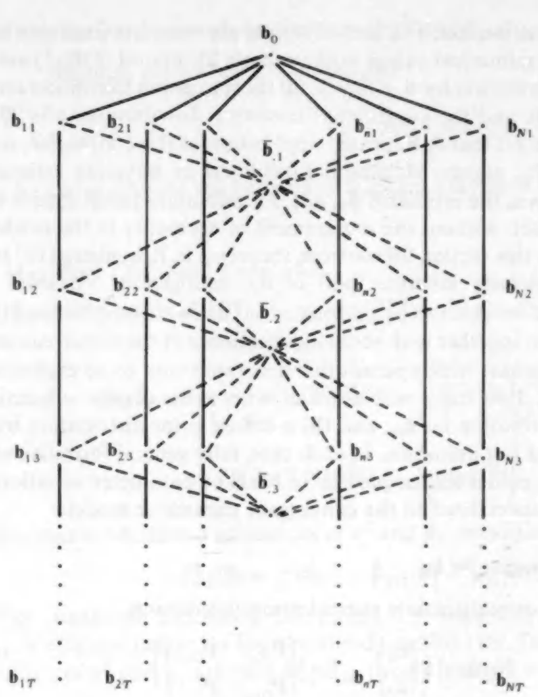

Figure 1a

Figure 1b

In the transitions between successive periods in Figure 1a, the solid lines denote the contributions of the individual parameter vectors to their own subsequent values, and the broken lines denote the contribution of the sample mean to the subsequent values of the individual vectors.

Figure 1b shows the elementary structure of the serially independent transitions between successive grand parameter vectors. The grand regression is a Markovian or sequential parameter regression problem in that the grand parameter vector obeys a first order Markov process.

II. Estimation in the Convergent Parameter Model

Let $\boldsymbol{\theta}$ denote the vector of parameters in the stochastic specification, including the second moments of the stochastic terms $R_1, \ldots, R_N, R_G$, $\mathbf{Q}_c, \mathbf{Q}_{ca}, \mathbf{Q}_a, \mathbf{Q}_G$ and the convergence rates $\phi_1, \ldots, \phi_\lambda$, but excluding the scale parameter σ^2. Let R_θ denote the admissible region of parameter values, which may be constrained by a priori information as well as nonnegativity and symmetry conditions on the second moments. Let $\mathbf{y}^s = (\mathbf{y}_1' : \ldots : \mathbf{y}_s')'$ denote the vector of all observations through period s.

In this section, Maximum Likelihood and Bayesian methods for estimating $\boldsymbol{\theta}$, σ^2, and $\boldsymbol{\beta}_T$ are developed under the assumption that all stochastic terms follow a multivariate normal distribution. The central results are recursive formulae which yield: (i) for any $\boldsymbol{\theta}$, the numerical values of the sample likelihood $\mathscr{L}(\boldsymbol{\theta}|\mathbf{y}^T)$ and the marginal posterior distribution for $\boldsymbol{\theta}$, $p''(\boldsymbol{\theta}|\mathbf{y}^T)$; (ii) the maximum likelihood estimators $\hat{\boldsymbol{\beta}}_{T|T}(\boldsymbol{\theta})$ and $\hat{\sigma}^2_{ML}(\boldsymbol{\theta})$, and the conditional posterior distributions $p''(\sigma^2|\boldsymbol{\theta}, y^T)$, $p''(\boldsymbol{\beta}_T|\boldsymbol{\theta}, \mathbf{y}^T)$, conditional on that $\boldsymbol{\theta}$. Repeated application of these formulae, over a range of $\boldsymbol{\theta}$ values in R_θ, allows Maximum Likelihood or Bayesian estimation. Moreover, if $\boldsymbol{\theta}$ be known, the estimator $\hat{\boldsymbol{\beta}}_{T|T}(\boldsymbol{\theta})$ is a minimum mean square error linear unbiased estimator, without the requirement of normality in the stochastic terms. The formulae in this section follow from theorems in Rosenberg (1973b).

The probability density function (pdf) of the endogenous variables may always be decomposed as $p(\mathbf{y}^T) = \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{y}^{t-1})$. The Markov process for the grand parameter vector, together with serial independence in the disturbances, are key simplifying assumptions which permit this decomposition to be exploited by a recursive procedure. Two cases will be dealt with in successive subsections: (A) a proper prior distribution for $\mathbf{b}_0$; and (B) a diffuse prior distribution for $\mathbf{b}_0$, or equivalently, $\mathbf{b}_0$ fixed but unknown. In each case, fully general formulae which hold for any regression model with sequential or Markov parameter variation will be exhibited and then specialized to the convergent parameter model.

A. Proper Prior Distribution for $\mathbf{b}_0$

Let $\mathbf{b}_0$ have a proper multivariate normal prior distribution

$$(12)\qquad \mathbf{b}_0 \sim \text{Normal}\left(\begin{pmatrix}\bar{\mathbf{c}}_0\\ \bar{\mathbf{a}}_0\end{pmatrix}, \quad \sigma^2\begin{pmatrix}\mathbf{P}_{0,c} & \mathbf{P}_{0,ca}\\ \mathbf{P}'_{0,ca} & \mathbf{P}_{0,a}\end{pmatrix}\right)$$

independently of all other stochastic terms. Then all regression parameters and endogenous variables follow a joint proper multivariate normal pdf, and it is easily shown that

$$(13)\quad p(\mathbf{y}^T|\sigma, \boldsymbol{\theta}) = \prod_{t=1}^{T} (2\pi\sigma^2)^{-(N/2)}|\mathbf{F}_t(\boldsymbol{\theta})|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta})\|_{\mathbf{F}_t(\boldsymbol{\theta})^{-1}}\right\},$$

where

$$\sigma^2\mathbf{F}_t(\boldsymbol{\theta}) \equiv \text{var}\,[\mathbf{y}_t|\sigma, \boldsymbol{\theta}, \mathbf{y}^{t-1}] = \sigma^2(\mathbf{X}_t\mathbf{M}_{t|t-1}(\boldsymbol{\theta})\mathbf{X}'_t + \mathbf{R}),$$

and where, in general,

$$\boldsymbol{\mu}_{r|s}(\boldsymbol{\theta}) \equiv E[\boldsymbol{\beta}_r|\boldsymbol{\theta}, \mathbf{y}^s], \qquad \sigma^2\mathbf{M}_{r|s}(\boldsymbol{\theta}) \equiv \text{var}\,[\boldsymbol{\beta}_r|\sigma, \boldsymbol{\theta}, \mathbf{y}^s].$$

The notation $\|\mathbf{e}\|_{\mathbf{A}}$ denotes the norm $\mathbf{e}'\mathbf{A}\mathbf{e}$. The subscript $r|s$ denotes an estimator or distribution for an item in period r, conditional on regression information up to and including period s.

Therefore, when $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\mathbf{F}(\boldsymbol{\theta})$ are computed by the recursive formulae provided below, the sample likelihood is

$$(14)\qquad \mathscr{L}(\sigma, \boldsymbol{\theta}\,|\mathbf{y}^T) = (2\pi\sigma^2)^{-TN/2}\left(\prod_{t=1}^{T}\zeta_t(\boldsymbol{\theta})\right)^{-1/2} \exp\left\{-\frac{TNs^2(\boldsymbol{\theta})}{2\sigma^2}\right\},$$

where

$$\zeta_t(\boldsymbol{\theta}) = |F_t(\boldsymbol{\theta})|, \qquad v_t(\boldsymbol{\theta}) = \|\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta})\|_{\mathbf{F}_t^{-1}(\boldsymbol{\theta})}, \qquad s^2(\boldsymbol{\theta}) = \frac{\sum_{t=1}^{T} v_t(\boldsymbol{\theta})}{TN}.$$

Also, from the joint normal distribution of $\mathbf{y}^T$ and $\boldsymbol{\beta}_T$,

$$(15) \qquad p(\boldsymbol{\beta}_T|\sigma, \boldsymbol{\theta}, y^T) = (2\pi\sigma^2)^{-(K/2)}|\mathbf{M}_{T|T}(\boldsymbol{\theta})|^{-1/2} \times \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{\beta}_T - \boldsymbol{\mu}(\boldsymbol{\theta})_{T|T}\|_{\mathbf{M}_{T|T}(\boldsymbol{\theta})^{-1}}\right\}.$$

These formulae provide the basis for Maximum Likelihood and Bayesian estimation.

A.1. *Maximum Likelihood Estimation*

The maximum value of the natural log of the likelihood function (14), for any $\boldsymbol{\theta}$, is

$$(16) \qquad l(\boldsymbol{\theta}|\mathbf{y}^T) \equiv \max_{\sigma^2} \ln \mathcal{L}(\sigma, \boldsymbol{\theta}|\mathbf{y}^T) = -\frac{1}{2}\left(TN\left(\ln\left(\frac{2\pi}{TN}\right) + 1\right) + TN\ln(TNs^2(\boldsymbol{\theta})) + \sum_{t=1}^{T} \ln \zeta_t(\boldsymbol{\theta})\right).$$

The maximum likelihood estimators of σ^2 and $\boldsymbol{\beta}_T$, conditional on $\boldsymbol{\theta}$, are

$$(17) \qquad \hat{\sigma}^2_{ML}(\boldsymbol{\theta}) = s^2(\boldsymbol{\theta}), \qquad \hat{\boldsymbol{\beta}}_{T|T}(\boldsymbol{\theta}) = \boldsymbol{\mu}_{T|T}(\boldsymbol{\theta}).$$

For maximum likelihood estimation, it is necessary to search R_θ for that $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{ML}$, which maximizes the log likelihood function (16). The maximum likelihood estimators of σ^2 and $\boldsymbol{\beta}_T$ are then $\hat{\sigma}^2_{ML}(\hat{\boldsymbol{\theta}}_{ML})$ and $\hat{\boldsymbol{\beta}}_{T|T}(\hat{\boldsymbol{\theta}}_{ML})$.

A.2. *Bayesian Estimation*

Let $p'(\boldsymbol{\theta})$ be a possibly diffuse prior pdf for $\boldsymbol{\theta}$, and let $p'(\sigma) \propto 1/\sigma$ be a diffuse prior for σ, following Zellner (1971: Ch. 2). Then the posterior pdf for $\boldsymbol{\beta}_T, \sigma, \boldsymbol{\theta}$, is

$$(18) \qquad p''(\boldsymbol{\beta}_T, \sigma, \boldsymbol{\theta}) = p(\boldsymbol{\beta}_T|\sigma, \boldsymbol{\theta}, \mathbf{y}^T)p''(\sigma, \boldsymbol{\theta}),$$

where the conditional pdf for $\boldsymbol{\beta}_T$ is given in (15), and the marginal posterior pdf for σ and $\boldsymbol{\theta}$ is, from (14),

$$(19) \qquad p''(\sigma, \boldsymbol{\theta}) \propto \mathcal{L}(\sigma, \boldsymbol{\theta}|\mathbf{y}^T)p'(\sigma)p'(\boldsymbol{\theta}) \propto \sigma^{-(TN+1)}p'(\boldsymbol{\theta})\left(\prod_{t=1}^{T} \zeta_t(\boldsymbol{\theta})\right)^{-1/2} \exp\left\{-\frac{TNs^2(\boldsymbol{\theta})}{2\sigma^2}\right\}.$$

This may be decomposed into the marginal posterior pdf for $\boldsymbol{\theta}$,

$$(20) \qquad p''(\boldsymbol{\theta}) = \int_{R_\sigma} p''(\sigma, \boldsymbol{\theta})\, d\sigma \propto p'(\boldsymbol{\theta})\left(\prod_{t=1}^{T} \zeta_t(\boldsymbol{\theta})\right)^{-1/2} (s^2(\boldsymbol{\theta}))^{-TN/2},$$

and the conditional posterior pdf for σ,

$$(21) \qquad p''(\sigma|\boldsymbol{\theta}) \propto \sigma^{-(NT+1)} \exp\left\{-\frac{TNs^2(\boldsymbol{\theta})}{2\sigma^2}\right\}.$$

Let $\overline{\sigma^2}(\boldsymbol{\theta})$ be the conditional posterior mean of σ^2, $\overline{\sigma^2}(\boldsymbol{\theta}) = TNs^2(\boldsymbol{\theta})/(TN - 2)$.

The conditional posterior pdf for $\boldsymbol{\beta}_T$ is multivariate Student t:

(22) $p''(\boldsymbol{\beta}_T|\boldsymbol{\theta}) \propto |s^2(\boldsymbol{\theta})\mathbf{M}_{T|T}(\boldsymbol{\theta})|^{-1/2}(TN + \|\boldsymbol{\beta}_T - \boldsymbol{\mu}(\boldsymbol{\theta})_{T|T}\|_{(s^2(\boldsymbol{\theta})\mathbf{M}_{T|T}(\boldsymbol{\theta}))^{-1}})^{-(NT+K)/2}$,

Hence, the moments of the marginal posterior pdf are

$$(23)\quad \begin{cases} \breve{\boldsymbol{\beta}}_{T|T} \equiv E[\boldsymbol{\beta}_T|\mathbf{y}^T] = \int_{R_\theta} \boldsymbol{\mu}_{T|T}(\boldsymbol{\theta}))p''(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \\ \breve{\mathbf{M}}_{T|T} \equiv \operatorname{var}[\boldsymbol{\beta}_T|\mathbf{y}^T] \\ \quad = \int_{R_\theta} (\overline{\sigma^2}(\boldsymbol{\theta})\mathbf{M}_{T|T}(\boldsymbol{\theta}) + (\boldsymbol{\mu}_{T|T}(\boldsymbol{\theta}) - \breve{\boldsymbol{\beta}}_{T|T})(\boldsymbol{\mu}'_{T|T}(\boldsymbol{\theta}) - \breve{\boldsymbol{\beta}}_{T|T}{}')p''(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \end{cases}$$

Thus, the posterior pdfs for $\boldsymbol{\beta}_T$ and σ, conditional on $\boldsymbol{\theta}$, are available in analytical form, so that Bayesian estimation may be carried out by numerical integration, with respect to $p''(\boldsymbol{\theta})$, over R_θ.

A.3. The Recursive Formulae

The required recursive formulae are well known in the applied physical sciences, and are often referred to as the Kalman-Bucy filter. (See, for example, Aoki (1967), Ho and Lee (1964), Kalman (1960), and Kalman and Bucy (1961).) For the special case of the convergent parameter regression model, the predictive pdf for the grand parameter vector in the initial period follows from the prior pdf(12) for $\mathbf{b}_0$ and the stationary dispersion of the individual parameter vectors (11):

$$(24a)\qquad \boldsymbol{\mu}_{1|0} = \begin{pmatrix} \bar{\mathbf{c}}_0 \\ \bar{\mathbf{a}}_0 \\ \vdots \\ \bar{\mathbf{a}}_0 \end{pmatrix}, \quad \mathbf{M}_{1|0}(\boldsymbol{\theta}) =$$

$$\begin{pmatrix} \mathbf{P}_{0,c} & \mathbf{P}_{0,ca} & \mathbf{P}_{0,ca} & \cdots & \mathbf{P}_{0,ca} \\ \mathbf{P}'_{0,ca} & \mathbf{P}_{0,a} + \boldsymbol{\Omega}(\boldsymbol{\theta}) & \mathbf{P}_{0,a} & \cdots & \mathbf{P}_{0,a} \\ \mathbf{P}'_{0,ca} & \mathbf{P}_{0,a} & \mathbf{P}_{0,a} + \boldsymbol{\Omega}(\boldsymbol{\theta}) & \cdots & \mathbf{P}_{0,a} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{P}'_{0,ca} & \mathbf{P}_{0,a} & \mathbf{P}_{0,a} & \cdots & \mathbf{P}_{0,a} + \boldsymbol{\Omega}(\boldsymbol{\theta}) \end{pmatrix}.$$

In a later period t, suppose that the regression information through period $t-1$ has been exploited to yield the posterior moments $\boldsymbol{\mu}_{t-1|t-1}(\boldsymbol{\theta})$ and $\sigma^2\mathbf{M}_{t-1|t-1}(\boldsymbol{\theta})$. Then the conditional predictive pdf for the parameters in period t, has moments given by the

Parameter Extrapolation Formulae:

$$(24b)\qquad \boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}) = \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{\mu}_{t-1|t-1}(\boldsymbol{\theta})$$

$$(24c)\qquad \mathbf{M}_{t|t-1}(\boldsymbol{\theta}) = \boldsymbol{\Phi}(\boldsymbol{\theta})\mathbf{M}_{t-1|t-1}(\boldsymbol{\theta})\boldsymbol{\Phi}'(\boldsymbol{\theta}) + \mathbf{Q}(\boldsymbol{\theta})$$

The predictive pdf for the parameters (24a) or (24b,c) implies a predictive pdf for the endogenous variables in that period.

Forecasting Formulae

(24e) $$\mathbf{e}_t(\boldsymbol{\theta}) \equiv \mathbf{y}_t - E(\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{y}^{t-1}) = \mathbf{y}_t - \mathbf{X}_t\boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta})$$

(24f) $$\mathbf{F}_t(\boldsymbol{\theta}) \equiv \frac{1}{\sigma^2}\operatorname{var}(\mathbf{e}_t|\sigma, \theta, \mathbf{y}^{t-1}) = \mathbf{X}_t\mathbf{M}_{t|t-1}(\boldsymbol{\theta})\mathbf{X}_t' + \mathbf{R}(\boldsymbol{\theta})$$

(24g) $$\mathbf{L}_t(\boldsymbol{\theta}) \equiv \frac{1}{\sigma^2}\operatorname{cov}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}), \mathbf{e}_t(\boldsymbol{\theta})|\sigma, \boldsymbol{\theta}, \mathbf{y}^{t-1}) = \mathbf{M}_{t|t-1}(\boldsymbol{\theta})\mathbf{X}_t'$$

(24h) $$v_t(\boldsymbol{\theta}) = \mathbf{e}_t'(\boldsymbol{\theta})\mathbf{F}_t^{-1}(\boldsymbol{\theta})\mathbf{e}_t(\boldsymbol{\theta})$$

(24i) $$\zeta_t(\boldsymbol{\theta}) = |\mathbf{F}_t(\boldsymbol{\theta})|.$$

Finally, the observations on the endogenous variables in period t are incorporated into a revised conditional pdf, given by the

Revision Formulae

(24m) $$\mathbf{K}_t(\boldsymbol{\theta}) = \mathbf{L}_t(\boldsymbol{\theta})\mathbf{F}_t^{-1}(\boldsymbol{\theta})$$

(24n) $$\boldsymbol{\mu}_{t|t}(\boldsymbol{\theta}) = \boldsymbol{\mu}_{t|t-1}(\boldsymbol{\theta}) + \mathbf{K}_t(\boldsymbol{\theta})\mathbf{e}_t(\boldsymbol{\theta})$$

(24o) $$\mathbf{M}_{t|t}(\boldsymbol{\theta}) = \mathbf{M}_{t|t-1}(\boldsymbol{\theta}) - \mathbf{L}_t(\boldsymbol{\theta})\mathbf{F}_t^{-1}(\boldsymbol{\theta})\mathbf{L}_t'(\boldsymbol{\theta}) = (\mathbf{I} - \mathbf{K}_t(\boldsymbol{\theta})\mathbf{X}_t)\mathbf{M}_{t|t-1}(\boldsymbol{\theta}).$$

B. No Prior Distribution for $\mathbf{b}_0$

Where no prior distribution for $\mathbf{b}_0$ exists (or, equivalently, where $\mathbf{b}_0$ is a fixed but unknown vector from the classical viewpoint), a "starting problem" exists. This problem proved to be quite troublesome. Indeed, the solution proposed in Aoki (1967) was erroneous, because it was based on false "identities" for generalized matrix inverses (p. 80). Fortunately, there is a straightforward solution to the problem. It may be shown (Rosenberg (1973b)) that the pdf for $\boldsymbol{\beta}_s$, conditional on $\mathbf{y}^t, \boldsymbol{\theta}, \sigma^2$, and $\mathbf{b}_0$, is of the form

(25) $$p(\boldsymbol{\beta}_s|\mathbf{b}_0, \sigma, \boldsymbol{\theta}, \mathbf{y}^t) = \text{Normal}\,(\boldsymbol{\xi}_{s|t}(\mathbf{b}_0, \boldsymbol{\theta}), \sigma^2\mathbf{M}_{s|t}^*(\boldsymbol{\theta})),$$

where the mean value is linear in $\mathbf{b}_0$,

$$\boldsymbol{\xi}_{s|t}(\mathbf{b}_0, \boldsymbol{\theta}) \equiv E[\mathbf{b}_s|\mathbf{b}_0, \boldsymbol{\theta}, \mathbf{y}^t] \equiv \boldsymbol{\mu}_{s|t}^*(\boldsymbol{\theta}) + \boldsymbol{\Xi}_{s|t}(\boldsymbol{\theta})\mathbf{b}_0.$$

It follows that

(26) $$p(\mathbf{y}^T|\mathbf{b}_0, \sigma, \boldsymbol{\theta}) = \prod_{t=1}^{T}(2\pi\sigma^2)^{-N/2}|\mathbf{F}_t^*(\boldsymbol{\theta})|^{-1/2}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\mu}_{t|t-1}^*(\boldsymbol{\theta}) - \mathbf{X}_t\boldsymbol{\Xi}_{t|t-1}(\boldsymbol{\theta})\mathbf{b}_0\|_{\mathbf{F}_t^*(\boldsymbol{\theta})^{-1}}\right\}.$$

where

$$\mathbf{F}_t^*(\boldsymbol{\theta}) \equiv \frac{1}{\sigma^2}\operatorname{var}[\mathbf{y}_t|\mathbf{b}_0, \sigma, \boldsymbol{\theta}, \mathbf{y}^{t-1}] = \mathbf{X}_t\mathbf{M}_{t|t-1}^*(\boldsymbol{\theta})\mathbf{X}_t' + \mathbf{R}(\boldsymbol{\theta}).$$

This is formally equivalent to the pdf in a regression with regressands $\mathbf{e}_t^*(\boldsymbol{\theta})$, regressor matrices $\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})$, and with $\mathbf{b}_0$ the unknown parameter vector, where

$$\mathbf{e}_t^*(\boldsymbol{\theta}) = \mathbf{y}_t - \mathbf{X}_t\boldsymbol{\mu}_{t|t-1}^*(\boldsymbol{\theta}), \qquad \boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \mathbf{X}_t\boldsymbol{\Xi}_{t|t-1}(\boldsymbol{\theta}).$$

In analogy with the familiar linear regression, it may be shown that

$$(27)\qquad \mathscr{L}(\mathbf{b}_0, \sigma, \boldsymbol{\theta}|\mathbf{y}^T) = (2\pi\sigma^2)^{-TN/2}\left(\prod_{t=1}^{T} \zeta_t^*(\boldsymbol{\theta})\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}((TN-k)s^2(\boldsymbol{\theta}) + \|\mathbf{b}_0 - \hat{\mathbf{b}}_0(\boldsymbol{\theta})\|_{\mathbf{W}_0(\boldsymbol{\theta})})\right\},$$

where

$$(28)\qquad \hat{\mathbf{b}}_0(\boldsymbol{\theta}) = \left(\sum_{t=1}^{T} \mathbf{H}_t^*(\boldsymbol{\theta})\right)^{-1} \sum_{t=1}^{T} \mathbf{h}_t^*(\boldsymbol{\theta}), \qquad \mathbf{W}_0(\boldsymbol{\theta}) = \left(\sum_{t=1}^{T} \mathbf{H}_t^*(\boldsymbol{\theta})\right)^{-1},$$

$$s^2(\boldsymbol{\theta}) = \frac{\sum_{t=1}^{T} \|\mathbf{e}_t^*(\boldsymbol{\theta}) - \boldsymbol{\Upsilon}_t(\boldsymbol{\theta})\hat{\mathbf{b}}_0(\boldsymbol{\theta})\|_{\mathbf{F}_t^*(\boldsymbol{\theta})^{-1}}}{TN-k} = \frac{\sum_{t=1}^{T} v_t^*(\boldsymbol{\theta}) - \hat{\mathbf{b}}_0' \sum_{t=1}^{T} \mathbf{h}_t^*(\boldsymbol{\theta})}{TN-k},$$

and where, for each t,

$$v_t^*(\boldsymbol{\theta}) = \mathbf{e}_t^*(\boldsymbol{\theta})'\mathbf{F}_t^{*-1}(\boldsymbol{\theta})\mathbf{e}_t^*(\boldsymbol{\theta}), \qquad \mathbf{H}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Upsilon}_t'(\boldsymbol{\theta})\mathbf{F}_t^{*-1}(\boldsymbol{\theta})\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})$$

$$\zeta_t^*(\boldsymbol{\theta}) = |\mathbf{F}_t^*(\boldsymbol{\theta})|, \qquad \mathbf{h}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Upsilon}_t'(\boldsymbol{\theta})\mathbf{F}_t^{*-1}(\boldsymbol{\theta})\mathbf{e}_t^*(\boldsymbol{\theta}).$$

B.1. Maximum Likelihood Estimation

From (27), the maximum value of the natural log of the likelihood function, for any $\boldsymbol{\theta}$, is

$$(29)\qquad l(\boldsymbol{\theta}|\mathbf{y}^T) \equiv \max_{\sigma,\mathbf{b}_0} \ln \mathscr{L}(\sigma, \boldsymbol{\theta}, \mathbf{b}_0|\mathbf{y}^T) = -\frac{1}{2}\left(TN\left(\ln\left(\frac{2\pi}{TN}\right) + 1\right) + TN\ln((TN-k)s^2(\boldsymbol{\theta})) + \sum_{t=1}^{T} \ln \zeta_t^*(\boldsymbol{\theta})\right).$$

The Maximum Likelihood estimator of $\mathbf{b}_0$, conditional on $\boldsymbol{\theta}$, is $\hat{\mathbf{b}}_0(\boldsymbol{\theta})$ given in (28). The Maximum Likelihood estimators for σ^2 and $\boldsymbol{\beta}_T$, conditional on $\boldsymbol{\theta}$, are

$$(30)\qquad \hat{\sigma}^2_{ML}(\boldsymbol{\theta}) = \frac{(TN-k)}{TN}s^2(\boldsymbol{\theta}), \qquad \hat{\boldsymbol{\beta}}_{T|T}(\boldsymbol{\theta}) = \boldsymbol{\mu}^*_{T|T}(\boldsymbol{\theta}) + \boldsymbol{\Xi}_{T|T}(\boldsymbol{\theta})\hat{\mathbf{b}}_0(\boldsymbol{\theta}).$$

As in A.1 above, the unconditional Maximum Likelihood estimators are $\hat{\boldsymbol{\theta}}_{ML}$, $\hat{\mathbf{b}}_0(\hat{\boldsymbol{\theta}}_{ML})$, $\hat{\sigma}^2_{ML}(\hat{\boldsymbol{\theta}}_{ML})$, and $\hat{\boldsymbol{\beta}}_{T|T}(\hat{\boldsymbol{\theta}}_{ML})$, where $\hat{\boldsymbol{\theta}}_{ML}$ maximizes (29) over R_θ.

B.2. Bayesian Estimation

Assume the same prior densities for $\boldsymbol{\theta}$ and σ as in A.2, above. The posterior pdf for all parameters is

$$(31)\qquad p''(\boldsymbol{\beta}_T, \mathbf{b}_0, \sigma, \boldsymbol{\theta}) = p(\boldsymbol{\beta}_T|\mathbf{b}_0, \sigma, \boldsymbol{\theta}, \mathbf{y}^T)p''(\mathbf{b}_0, \sigma, \boldsymbol{\theta}).$$

The conditional pdf for $\boldsymbol{\beta}_T$ is given in (25). The marginal posterior pdf for the other parameters is

$$(32)\qquad p''(\mathbf{b}_0, \sigma, \boldsymbol{\theta}) = p(\mathbf{b}_0, \sigma, \boldsymbol{\theta}|\mathbf{y}^T)p'(\sigma)p'(\boldsymbol{\theta})$$

$$\propto \sigma^{-(TN+1)}p'(\boldsymbol{\theta})\left(\prod_{t=1}^{T} \zeta_t^*(\boldsymbol{\theta})\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}((TN-k)s^2(\boldsymbol{\theta}) + \|\mathbf{b}_0 - \hat{\mathbf{b}}_0(\boldsymbol{\theta})\|_{\mathbf{W}_0(\boldsymbol{\theta})})\right\}.$$

Integrating with respect to $\mathbf{b}_0$ and σ, the marginal posterior pdf for $\boldsymbol{\theta}$ is found to be

$$(33)\qquad p''(\boldsymbol{\theta}) \propto p'(\boldsymbol{\theta})\left(\prod_{t=1}^{T} \zeta_t^*(\boldsymbol{\theta})\right)^{-1/2} |\mathbf{W}_0(\boldsymbol{\theta})|^{1/2}(s^2(\boldsymbol{\theta}))^{-(TN-k)/2}.$$

The conditional posterior pdf for σ is

$$(34)\qquad p''(\sigma|\boldsymbol{\theta}) \propto \sigma^{-(NT+1-k)} \exp\left\{-\frac{(TN-k)s^2(\boldsymbol{\theta})}{2\sigma^2}\right\}.$$

The mean is $\overline{\sigma^2}(\boldsymbol{\theta}) = [(TN-k)s^2(\boldsymbol{\theta})/(TN-k-2)]$. The conditional posterior pdf for $\boldsymbol{\beta}_T$ is again multivariate Student t:

$$(35)\qquad p''(\boldsymbol{\beta}_T|\boldsymbol{\theta}) \propto |s^2(\boldsymbol{\theta})\mathbf{M}_{T|T}(\boldsymbol{\theta})|^{-1/2} \times (TN-k+\|\boldsymbol{\beta}_T-\boldsymbol{\mu}_{T|T}(\boldsymbol{\theta})\|_{(s^2(\boldsymbol{\theta})\mathbf{M}_{T|T}(\boldsymbol{\theta}))^{-1}})^{-(TN-k+K)/2},$$

where

$$(36)\qquad \boldsymbol{\mu}_{T|T}(\boldsymbol{\theta}) = \boldsymbol{\mu}^*_{T|T}(\boldsymbol{\theta}) + \boldsymbol{\Xi}_{T|T}(\boldsymbol{\theta})\hat{\mathbf{b}}_0(\boldsymbol{\theta}),$$
$$\mathbf{M}_{T|T}(\boldsymbol{\theta}) = \mathbf{M}^*_{T|T}(\boldsymbol{\theta}) + \boldsymbol{\Xi}_{T|T}(\boldsymbol{\theta})\mathbf{W}_0(\boldsymbol{\theta})\boldsymbol{\Xi}'_{T|T}(\boldsymbol{\theta}).$$

The moments of the marginal posterior pdf of $\boldsymbol{\beta}_T$ are again given by formula (23).

B.3. The Recursive Formulae

The recursive formulae are closely related to those in the previous case. The initial conditions are somewhat changed.

Initial Conditions:

$$(37a)\qquad \boldsymbol{\mu}^*_{1|0}(\boldsymbol{\theta}) = \mathbf{0},\qquad \boldsymbol{\Xi}_{1|0}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \\ \vdots & \vdots \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$

$$\mathbf{M}_{1|0}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}(\boldsymbol{\theta}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Omega}(\boldsymbol{\theta}) & \cdots & \mathbf{0} \\ \vdots & \vdots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Omega}(\boldsymbol{\theta}) \end{pmatrix}.$$

All other formulae in the previous list (24b, ..., 24o) carry over to the present case, with the variables $\boldsymbol{\mu}, \mathbf{M}, \mathbf{e}, \mathbf{F}, \upsilon, \zeta, \mathbf{L}, \mathbf{K}$ having a superscript *. In addition, the following formulae are inserted in the list in alphabetical order:

Parameter Extrapolation:

$$(37d)\qquad \boldsymbol{\Xi}_{t|t-1}(\boldsymbol{\theta}) = \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{\Xi}_{t-1|t-1}(\boldsymbol{\theta})$$

Forecasting:

(37j) $$\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = \mathbf{X}_t\boldsymbol{\Xi}_{t|t-1}(\boldsymbol{\theta})$$

(37k) $$\mathbf{h}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Upsilon}_t'(\boldsymbol{\theta})\mathbf{F}_t^{*-1}(\boldsymbol{\theta})\mathbf{e}_t^*(\boldsymbol{\theta})$$

(37l) $$\mathbf{H}_t^*(\boldsymbol{\theta}) = \boldsymbol{\Upsilon}_t'(\boldsymbol{\theta})\mathbf{F}_t^{*-1}(\boldsymbol{\theta})\boldsymbol{\Upsilon}_t(\boldsymbol{\theta})$$

Revision:

(37p) $$\boldsymbol{\Xi}_{t|t}(\boldsymbol{\theta}) = \boldsymbol{\Xi}_{t|t-1}(\boldsymbol{\theta}) - \mathbf{K}_t^*(\boldsymbol{\theta})\boldsymbol{\Upsilon}_t(\boldsymbol{\theta}) = (\mathbf{I} - \mathbf{K}_t^*(\boldsymbol{\theta})\mathbf{X}_t)\boldsymbol{\Xi}_{t|t-1}(\boldsymbol{\theta}).$$

C. Both Maximum Likelihood and Bayesian estimation require an efficient means of searching R_θ. It is sometimes convenient to transform the parameters to a vector $\boldsymbol{\theta}^*$ such that the admissible region for the transformed parameters, R_{θ^*}, coincides with Euclidean space. For instance, the variance matrices are required to be positive semi-definite symmetric. This constraint may be imposed by expressing each matrix as the product of a lower-triangular matrix with its transpose, for instance, $\mathbf{Q}_b = \mathbf{TT}'$. Searching the space of unconstrained lower-triangular matrices $\mathbf{T}$ is equivalent to searching the space of positive semi-definite symmetric matrices $\mathbf{Q}_b$, and the constraints are removed from the transformed problem. Similarly, for the convergence rates ϕ_i, a convenient transformation is $\phi_i = d_i^2/(1 + d_i^2)$, since the admissable range $0 \leq \phi_i < 1$ is equivalent to the range $-\infty < d_i < \infty$. However, note that in both cases $-\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^*$ yield identical values for $\boldsymbol{\theta}$, and also that $\partial\boldsymbol{\theta}/\partial\boldsymbol{\theta}^*|_{\boldsymbol{\theta}^*=0} = \mathbf{0}$, so that attention must be given to avoiding the spurious local extremum at $\boldsymbol{\theta}^* = \mathbf{0}$.

A good initial estimate of the stochastic specification is also helpful. The following algorithm provides an initial estimate when the sample size is large:

(i) First, under the temporary simplifying assumption that parameters are not dispersed across the population, estimates of the mean parameters in every period, $\hat{\bar{\mathbf{b}}}_1, \ldots, \hat{\bar{\mathbf{b}}}_T$, are generated. If the population mean is assumed to be essentially unchanging over time, ($\mathbf{Q}_c = \mathbf{Q}_{ca} = \mathbf{Q}_G = \mathbf{0}$), this is done by ordinary least squares. Otherwise, the population mean changes sequentially over time according to a Markov process with incremental variance

$$\sigma^2\begin{pmatrix} \mathbf{Q}_c : & \mathbf{Q}_{ca} \\ \mathbf{Q}_{ca}' : & \mathbf{Q}_G + \dfrac{\mathbf{Q}_a}{N} \end{pmatrix}.$$

This variance, together with the realized values of the population mean parameters, may be estimated by an application of the previous formulae to this simpler sequential model. The communal disturbance variance $\sigma^2 R_G$ may also be estimated at this stage.

(ii) If the sample size is large, the residuals about these sample mean parameter estimates will approximate the contributions of the parameter dispersion and the disturbances,

$$\hat{e}_{nt} = y_{nt} - \mathbf{x}_{nt}'\hat{\bar{\mathbf{b}}}_t \simeq y_{nt} - \mathbf{x}_{nt}'\bar{\mathbf{b}}_t = \mathbf{z}_{nt}'(\mathbf{a}_{nt} - \bar{\mathbf{a}}_t) + u_{nt}.$$

Therefore,

(38) $$E[\hat{e}_{nt}^2] \simeq g_0 + \sum_{i=1}^{\lambda}\sum_{j=i}^{\lambda} g_{ij}z_{int}z_{jnt} \qquad n = 1, \ldots, N \quad t = 1, \ldots, T$$

where

$$g_0 = \sigma^2(1 + R_G), \qquad g_{ii} = \sigma^2\omega_{ii}, \quad \text{and} \quad g_{ij} = 2\sigma^2\omega_{ij} \quad \text{for } j > i.$$

(Note that, for simplicity, R_n is assumed here to equal unity for all n.) Also, for any time lag τ,

$$(39)\quad E[\hat{e}_{nt}\hat{e}_{n,t-\tau}] \simeq \sum_{i=1}^{\lambda}\sum_{j=1}^{\lambda} g_{\tau ij} z_{int} z_{jn,t-\tau} \qquad n = 1,\ldots,N \quad t = \tau + 1,\ldots,T$$

where

$$g_{\tau ij} = \sigma^2\omega_{ij}\phi_i^{\tau}.$$

If (38) is treated as a regression equation, with the squared residuals regressed on the cross products of the explanatory variables, then estimates of $g_0, g_1, \ldots, g_{1\lambda}, \ldots, g_{\lambda\lambda}$ and, hence, of σ^2 and $\mathbf{\Omega}$ are obtained. Similarly, for each time lag τ, a regression of the lagged products of the residuals on the lagged products of the explanatory variables of form (39) provides estimates of $\sigma^2\mathbf{\Delta}_\phi^\tau\mathbf{\Omega}$.

The various g's are nonlinear functions of the underlying parameters $\mathbf{\Delta}_\phi$ and $\mathbf{Q}_a$. The estimates $\hat{g}_{\tau ij}$ may be examined for their implications about the pattern of parameter variation, and initial estimates of the underlying parameters may be obtained by inspection or, if necessary, by nonlinear regression of the various $\hat{g}_{\tau ij}$ onto $\mathbf{\Delta}_\phi$ and $\mathbf{Q}_a$.

D. Minimum Mean Square Error Linear Estimation

Suppose that $\boldsymbol{\theta}$ is known. Let a minimum mean square linear unbiased estimator be defined as follows:

(i) An estimator $\boldsymbol{\beta}_{T|T}$ is linear unbiased iff it is a linear function of $\mathbf{y}^T$ such that $E[\boldsymbol{\beta}_{T|T}|\boldsymbol{\theta}] = E[\boldsymbol{\beta}_T|\boldsymbol{\theta}]$.

(ii) The minimum mean square error linear unbiased estimator $\hat{\boldsymbol{\beta}}_{T|T}$ is defined by the condition that for every linear combination of the parameters, $\boldsymbol{\alpha}'\boldsymbol{\beta}_T$, and for every linear unbiased estimator $\boldsymbol{\beta}_{T|T}$, $E[(\boldsymbol{\alpha}'\hat{\boldsymbol{\beta}}_{T|T} - \boldsymbol{\alpha}'\boldsymbol{\beta}_T)^2|\boldsymbol{\theta}] \leq E[(\boldsymbol{\alpha}'\boldsymbol{\beta}_{T|T} - \boldsymbol{\alpha}'\boldsymbol{\beta}_T)^2|\boldsymbol{\theta}]$.

Then it may be shown (Rosenberg (1973b)) that the estimators $\hat{\boldsymbol{\beta}}_{T|T}(\boldsymbol{\theta})$ derived in Sections II.A.2. and II.B.2. are minimum mean square error linear unbiased estimators, with mean square error matrices $\sigma^2\mathbf{M}_{T|T}(\boldsymbol{\theta})$. Also, $s^2(\boldsymbol{\theta})$ is an unbiased estimator of σ^2. These properties do not require that the stochastic terms be normally distributed.

III. Approximate Formulae

The number of arithmetic operations in the recursive formulae increases as $N^3\lambda^2$, and the number of entries in $\mathbf{M}$ increases as $N^2\lambda^2$. Consequently, the exact method requires excessive computer time and storage when N is large. Fortunately, a natural simplifying approximation eliminates these problems.

The parameter covariance matrix $\sigma^2\mathbf{M}$ may be partitioned as

$$\sigma^2\mathbf{M}_{s|t} \equiv E\left[\begin{pmatrix} \hat{\mathbf{c}}_{s|t} - \mathbf{c}_s \\ \hat{\mathbf{a}}_{1s|t} - \mathbf{a}_{1s} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mathbf{a}}_{Ns|t} - \mathbf{a}_{Ns} \end{pmatrix}\begin{pmatrix} \hat{\mathbf{c}}_{s|t} - \mathbf{c}_s \\ \hat{\mathbf{a}}_{1s|t} - \mathbf{a}_{1s} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mathbf{a}}_{Ns|t} - \mathbf{a}_{Ns} \end{pmatrix}'\right] \equiv \sigma^2 \begin{bmatrix} \mathbf{C} & \mathbf{D}_1 & \mathbf{D}_2 & \ldots & \mathbf{D}_N \\ \mathbf{D}_1' & \mathbf{A}_{11} & \mathbf{A}_{12} & \ldots & \mathbf{A}_{1N} \\ \mathbf{D}_2' & \mathbf{A}_{21} & \mathbf{A}_{22} & \ldots & \mathbf{A}_{2N} \\ \vdots & & & & \vdots \\ \mathbf{D}_N' & \mathbf{A}_{N1} & \mathbf{A}_{N2} & \ldots & \mathbf{A}_{NN} \end{bmatrix}_{s|t}$$

Throughout the recursive procedure, the largest part of the covariance between the parameters of different individuals arises from the common influence of the population mean. As a consequence, the matrices $\sigma^2\mathbf{A}_{mn}$, $m \neq n$, giving the covariance between the mth and nth individual parameter vectors, are similar for all pairs of individuals, as are the matrices $\mathbf{D}_n$ for all individuals. Accordingly, the following approximation suggests itself:

(40)
$$\tilde{\mathbf{M}} = \begin{pmatrix} \tilde{\mathbf{C}} & \tilde{\mathbf{D}} & \tilde{\mathbf{D}} & \ldots & \tilde{\mathbf{D}} \\ \tilde{\mathbf{D}}' & \tilde{\mathbf{A}}_G + \tilde{\mathbf{A}}_1 & \tilde{\mathbf{A}}_G & \ldots & \tilde{\mathbf{A}}_G \\ \tilde{\mathbf{D}}' & \tilde{\mathbf{A}}_G & \tilde{\mathbf{A}}_G + \tilde{\mathbf{A}}_2 & \ldots & \tilde{\mathbf{A}}_G \\ \vdots & \vdots & & \vdots & \\ \tilde{\mathbf{D}}' & \tilde{\mathbf{A}}_G & \tilde{\mathbf{A}}_G & \ldots & \tilde{\mathbf{A}}_G + \tilde{\mathbf{A}}_N \end{pmatrix}.$$

Here

$$\sigma^2\tilde{\mathbf{A}}_G = \sigma^2\left(\frac{\sum_{\substack{m,n=1 \\ m\neq n}}^{N} \mathbf{A}_{mn}}{N(N-1)}\right)$$

is the average interindividual covariance,

$$\sigma^2\tilde{\mathbf{D}} = \sigma^2\left(\frac{\sum_{n=1}^{N} \mathbf{D}_n}{N}\right)$$

is the average covariance between cross-fixed parameters and individual cross-varying parameters, and the matrices $\sigma^2\tilde{\mathbf{A}}_n = \sigma^2(\mathbf{A}_{nn} - \tilde{\mathbf{A}}_G)$, $n = 1, \ldots, N$ are the excess of intra-individual over average interindividual covariance. The superscript tilde denotes an approximation to a statistic.

The simplifying approximation reduces the number of distinct entries in $\mathbf{M}$ to order $\lambda^2 N$ and the number of arithmetic operations to order $k^2 N$. Estimation for a given $\boldsymbol{\theta}$ then requires the same order of magnitude of storage and computations as would be required by ordinary regressions for all individuals in the population, in which similarities across individuals would be in no way exploited.

In this section, the recursive formulae resulting from this approximation are given in terms of the individual parameters. These formulae, the exact recursive formulae, and the formulae for another approximation were derived in detail in (Rosenberg (1973c)), but only the approximation that was found to be preferable will be reported here. To simplify the presentation, the notation $(\boldsymbol{\theta})$ and the subscript t on the variables $\mathbf{y}$, $\mathbf{e}$, $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}$, $\mathbf{F}$, $\mathbf{K}$, $\mathbf{L}$, $\boldsymbol{\Upsilon}$ will be omitted where no confusion can result.

A. Approximate Recursive Formulce

The initial conditions (24a) and (37a) both satisfy the approximation exactly, and may be used in the forms already given.

Parameter Extrapolation

Suppose that for some t, $\tilde{\mathbf{M}}_{t-1|t-1}$ satisfies the approximation (40). Then the parameter extrapolation formulae are

$$
\text{b}\begin{cases}\tilde{\mathbf{c}}_{t|t-1} = \tilde{\mathbf{c}}_{t-1|t-1} \\ \tilde{\mathbf{a}}_{n,t|t-1} = \boldsymbol{\Delta}_{\phi}\tilde{\mathbf{a}}_{n,t-1|t-1} + (\mathbf{I}-\boldsymbol{\Delta}_{\phi})\bar{\tilde{\mathbf{a}}}_{t-1|t-1} \qquad n = 1,\ldots,N\end{cases}
$$

$$
\text{c}\begin{cases}\tilde{\mathbf{C}}_{t|t-1} = \tilde{\mathbf{C}}_{t-1|t-1} + \mathbf{Q}_c \\ \tilde{\mathbf{D}}_{t|t-1} = \tilde{\mathbf{D}}_{t-1|t-1} + \mathbf{Q}_{ca} \\ \tilde{\mathbf{A}}_{G,t|t-1} = \tilde{\mathbf{A}}_{G,t-1|t-1} + \mathbf{Q}_G + \dfrac{\bar{\bar{\mathbf{A}}}_{t-1|t-1} - \boldsymbol{\Delta}_{\phi}\bar{\bar{\mathbf{A}}}_{t-1|t-1}\boldsymbol{\Delta}_{\phi}}{N} \\ \tilde{\mathbf{A}}_{n,t|t-1} = \boldsymbol{\Delta}_{\phi}\tilde{\mathbf{A}}_{n,t-1|t-1}\boldsymbol{\Delta}_{\phi} + \mathbf{Q}_a \\ \quad + \dfrac{\boldsymbol{\Delta}_{\phi}(\tilde{\mathbf{A}}_{n,t-1|t-1} - \bar{\bar{\mathbf{A}}}_{t-1|t-1})(\mathbf{I}-\boldsymbol{\Delta}_{\phi}) + (\mathbf{I}-\boldsymbol{\Delta}_{\phi})(\tilde{\mathbf{A}}_{n,t-1|t-1} - \bar{\bar{\mathbf{A}}}_{t-1|t-1})\boldsymbol{\Delta}_{\phi}}{N} \\ \hfill n = 1,\ldots,N\end{cases}
$$

$$
\text{d}\begin{cases}\tilde{\boldsymbol{\Xi}}_{c,t|t-1} = \tilde{\boldsymbol{\Xi}}_{c,t-1|t-1} \\ \tilde{\boldsymbol{\Xi}}_{n,t|t-1} = \boldsymbol{\Delta}_{\phi}\tilde{\boldsymbol{\Xi}}_{n,t-1|t-1} + (\mathbf{I}-\boldsymbol{\Delta}_{\phi})\bar{\tilde{\boldsymbol{\Xi}}}_{t-1|t-1} \qquad n = 1,\ldots,N\end{cases}
$$

where $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Xi}}$ have been partitioned as

$$
\tilde{\boldsymbol{\mu}} = \begin{pmatrix}\tilde{\mathbf{c}} \\ \tilde{\mathbf{a}}_1 \\ \vdots \\ \tilde{\mathbf{a}}_N\end{pmatrix} \qquad \tilde{\boldsymbol{\Xi}} = \begin{pmatrix}\tilde{\boldsymbol{\Xi}}_c \\ \tilde{\boldsymbol{\Xi}}_1 \\ \vdots \\ \tilde{\boldsymbol{\Xi}}_N\end{pmatrix}
$$

and where the bar denotes an average over $n = 1,\ldots,N$, e.g.,

$$
\bar{\tilde{\mathbf{a}}}_{t-1|t-1} = \frac{\sum_{m=1}^{N}\tilde{\mathbf{a}}_{m,t-1|t-1}}{N}.
$$

Note that if $\mathbf{M}_{t-1|t-1}$ satisfies the approximation, then $\mathbf{M}_{t|t-1}$ also exactly satisfies it.

Forecasting

The forecast error vector is

$$
\text{e}\left\{\mathbf{e} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}_{t|t-1} = \begin{pmatrix} y_1 - (\mathbf{w}_1' : \mathbf{z}_1')\begin{pmatrix}\tilde{\mathbf{c}} \\ \tilde{\mathbf{a}}_1\end{pmatrix}_{t|t-1} \\ \vdots \\ y_N - (\mathbf{w}_N' : \mathbf{z}_N')\begin{pmatrix}\tilde{\mathbf{c}} \\ \tilde{\mathbf{a}}_N\end{pmatrix}_{t|t-1}\end{pmatrix} \equiv \begin{pmatrix}e_1 \\ \vdots \\ e_N\end{pmatrix}.\right.
$$

When $\mathbf{M}_{t-1|t-1}$ satisfies the approximation (40), $\mathbf{F}$ simplifies to

$$\mathbf{F} = \begin{pmatrix} \mathbf{w}_1' & \mathbf{z}_1' \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \mathbf{w}_N' & \mathbf{z}_N' \end{pmatrix} \begin{pmatrix} \bar{\mathbf{C}} & \bar{\mathbf{D}} \\ \bar{\mathbf{D}}' & \bar{\mathbf{A}}_G \end{pmatrix}_{t|t-1} \begin{pmatrix} \mathbf{w}_1 & \dots & \mathbf{w}_N \\ \mathbf{z}_1 & \dots & \mathbf{z}_N \end{pmatrix} + R_G \boldsymbol{\iota}\boldsymbol{\iota}' + \begin{pmatrix} f_1 & & 0 \\ & \ddots & \\ 0 & & f_N \end{pmatrix},$$

where

$$f_n = \mathbf{z}_n' \mathbf{A}_{n,t|t-1} \mathbf{z}_n + R_n, \qquad n = 1, \dots, N,$$

and where $\boldsymbol{\iota}$ is again a vector of units.

When the communal disturbance variance R_G is zero, the middle term vanishes. Otherwise, it may be adjoined to the first term:

$$f\left\{ \mathbf{F} = \begin{pmatrix} \mathbf{w}_1' & \mathbf{z}_1'[1] \\ \vdots & \vdots \\ \mathbf{w}_N' & \mathbf{z}_N'[1] \end{pmatrix} \begin{pmatrix} \mathbf{C} & \mathbf{D} & [0] \\ \mathbf{D}' & \mathbf{A}_G & [0] \\ [0] & [0] & [R_G] \end{pmatrix}_{t|t-1} \begin{pmatrix} \mathbf{w}_1 & \dots & \mathbf{w}_N \\ \mathbf{z}_1 & \dots & \mathbf{z}_N \\ [1] & \dots & [1] \end{pmatrix} + \begin{pmatrix} f_1 & & 0 \\ & \ddots & \\ 0 & & f_N \end{pmatrix} \right.$$

$$\equiv \boldsymbol{\Psi}\mathbf{P}_{t|t-1}\boldsymbol{\Psi}' + \boldsymbol{\Delta}_f.$$

Let $\boldsymbol{\psi}_n = (\mathbf{w}_n' : \mathbf{z}_n' : [1])'$ denote the nth column of $\boldsymbol{\Psi}'$. Here the communal disturbance changes status from a component of the disturbances with variance $\sigma^2 R_G$ to a cross-fixed parameter, with a coefficient vector of units, having forecast value of zero and forecast error variance of $\sigma^2 R_G$. Square brackets enclose terms which appear only when this artifice is in use. k [or $k+1$] dimensional matrices such as $\mathbf{P}$ will be partitioned in the self-explanatory notation:

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_c \\ \mathbf{P}_a \\ [\mathbf{P}_u] \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{cc} & \mathbf{P}_{ca} & [P_{cu}] \\ \mathbf{P}_{ac} & \mathbf{P}_{aa} & [P_{au}] \\ [P_{uc} & [P_{ua}] & [P_{uu}] \end{pmatrix}.$$

When $\mathbf{M}_{t-1|t-1}$ satisfies the approximation, $\mathbf{L}$ simplifies to

$$g\left\{ \mathbf{L} = \begin{pmatrix} \bar{\mathbf{C}} & \bar{\mathbf{D}} \\ \bar{\mathbf{D}}' & \bar{\mathbf{A}}_G \\ \vdots & \vdots \\ \bar{\mathbf{D}}' & \bar{\mathbf{A}}_G \end{pmatrix}_{t|t-1} \begin{pmatrix} \mathbf{w}_1 & \dots & \mathbf{w}_N \\ \mathbf{z}_1 & \dots & \mathbf{z}_N \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \dots & \mathbf{0} \\ \bar{\mathbf{A}}_{1,t|t-1}\mathbf{z}_1 & & \\ \mathbf{0} & \ddots & \mathbf{0} \\ & & \bar{\mathbf{A}}_{N,t|t-1}\mathbf{z}_N \end{pmatrix} \right.$$

$$= \begin{pmatrix} \mathbf{P}_c \\ \mathbf{P}_a \\ \vdots \\ \mathbf{P}_a \end{pmatrix} \boldsymbol{\Psi}' + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \boldsymbol{\lambda}_1 & \mathbf{0} & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \boldsymbol{\lambda}_N \end{pmatrix}$$

where $\boldsymbol{\lambda}_n = \bar{\mathbf{A}}_{n,t|t-1}\mathbf{z}_n$.

The inversion of $\mathbf{F}$ can be simplified by the matrix inversion identity

$$\text{(41)} \qquad \mathbf{F}^{-1} = \boldsymbol{\Delta}_f^{-1} - \boldsymbol{\Delta}_f^{-1}\boldsymbol{\Psi}(\boldsymbol{\Psi}'\boldsymbol{\Delta}_f^{-1}\boldsymbol{\Psi} + \mathbf{P}^{-1})^{-1}\boldsymbol{\Psi}'\boldsymbol{\Delta}_f^{-1}.$$

The matrix $\boldsymbol{\Psi}'\boldsymbol{\Delta}_f^{-1}\boldsymbol{\Psi} = \sum_{n=1}^{N} (\boldsymbol{\psi}_n\boldsymbol{\psi}_n'/f_n)$, which has the form of a precision matrix, will be denoted by $\mathbf{H}$. Let $\mathbf{S} = (\mathbf{\Psi}'\mathbf{\Delta}_f^{-1}\mathbf{\Psi} + \mathbf{P}^{-1})^{-1} = (\mathbf{H} + \mathbf{P}^{-1})^{-1}$. Also, let $\mathbf{h} = \mathbf{\Psi}'\mathbf{\Delta}_f^{-1}\mathbf{e} = \sum_{n=1}^{N} (\boldsymbol{\psi}_n e_n/f_n)$. Then the residual sum of squares is

$$(h)\left\{\quad v_t = \mathbf{e}'\mathbf{F}^{-1}\mathbf{e} = \mathbf{e}'\mathbf{\Delta}_f^{-1}\mathbf{e} - \mathbf{e}'\mathbf{\Delta}_f^{-1}\mathbf{\Psi}\mathbf{S}\mathbf{\Psi}'\mathbf{\Delta}_f^{-1}\mathbf{e} = \sum_{n=1}^{N} \frac{e_n^2}{f_n} - \mathbf{h}'\mathbf{S}\mathbf{h}.\right.$$

The determinant of $\mathbf{F}$ is given by the determinantal identity

$$|\mathbf{F}| = |\mathbf{\Delta}_f + \mathbf{\Psi}\mathbf{P}\mathbf{\Psi}'| = |\mathbf{\Delta}_f| \cdot |\mathbf{P}| \cdot |\mathbf{P}^{-1} + \mathbf{\Psi}'\mathbf{\Delta}_f^{-1}\mathbf{\Psi}| \tag{42}$$

which yields

$$(i)\left\{\quad \zeta_t = \left(\prod_{n=1}^{N} f_n\right) \cdot |\mathbf{P}| \cdot |\mathbf{S}^{-1}|.\right.$$

Whether or not $\mathbf{M}_{t-1|t-1}$ satisfies the approximation, $\mathbf{\Upsilon}$ is given by

$$(j)\left\{\quad \mathbf{\Upsilon} = \mathbf{X}_t\mathbf{\Xi}_{t|t-1} = \begin{pmatrix} \mathbf{w}_1'\mathbf{\Xi}_{c,t|t-1} & + & \mathbf{z}_1'\mathbf{\Xi}_{1,t|t-1} \\ & \vdots & \\ \mathbf{w}_N'\mathbf{\Xi}_{c,t|t-1} & + & \mathbf{z}_N'\mathbf{\Xi}_{N,t|t-1} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{\Upsilon}_1' \\ \vdots \\ \mathbf{\Upsilon}_N' \end{pmatrix}.\right.$$

Therefore,

$$(k)\left\{\quad \mathbf{h}_t^* = \mathbf{\Upsilon}'\mathbf{F}^{-1}\mathbf{e} = \sum_{n=1}^{N} \mathbf{\Upsilon}_n\left(\frac{e_n - \boldsymbol{\psi}_n'\mathbf{S}\mathbf{h}}{f_n}\right)\right.$$

$$(l)\left\{\quad \mathbf{H}_t^* = \mathbf{\Upsilon}'\mathbf{F}^{-1}\mathbf{\Upsilon} = \sum_{n=1}^{N} \frac{\mathbf{\Upsilon}_n\mathbf{\Upsilon}_n'}{f_n} - \left(\sum_{n=1}^{N} \frac{\mathbf{\Upsilon}_n\boldsymbol{\psi}_n'}{f_n}\right)\mathbf{S}\left(\sum_{n=1}^{N} \frac{\mathbf{\Upsilon}_n\boldsymbol{\psi}_n'}{f_n}\right)'.\right.$$

Revision

The first term of $\mathbf{L}$, when post-multiplied by $\mathbf{F}^{-1}$, assumes the simple form

$$\begin{pmatrix} \mathbf{P}_c \\ \mathbf{P}_a \end{pmatrix}\mathbf{\Psi}'(\mathbf{\Delta}_f^{-1} - \mathbf{\Delta}_f^{-1}\mathbf{\Psi}(\mathbf{H} + \mathbf{P}^{-1})^{-1}\mathbf{\Psi}'\mathbf{\Delta}_f^{-1}) = \begin{pmatrix} \mathbf{P}_c \\ \mathbf{P}_a \end{pmatrix}(\mathbf{I} - \mathbf{H}(\mathbf{H} + \mathbf{P}^{-1})^{-1})\mathbf{\Psi}'\mathbf{\Delta}_f^{-1}$$

$$= \begin{pmatrix} \mathbf{P}_c \\ \mathbf{P}_a \end{pmatrix}(\mathbf{P}^{-1}(\mathbf{H} + \mathbf{P}^{-1})^{-1})\mathbf{\Psi}'\mathbf{\Delta}_f^{-1} = (\mathbf{I} : [0])\mathbf{S}\mathbf{\Psi}'\mathbf{\Delta}_f^{-1}.$$

The revision matrix $\mathbf{K}$ may therefore be written:

$$m\left\{\quad \mathbf{K} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & [0] \\ \mathbf{0} & \mathbf{I} & [0] \\ & \vdots & \\ \mathbf{0} & \mathbf{I} & [0] \end{pmatrix}\mathbf{S}\mathbf{\Psi}'\mathbf{\Delta}_f^{-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \boldsymbol{\lambda}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \vdots & & & \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\lambda}_N \end{pmatrix}(\mathbf{\Delta}_f^{-1} - \mathbf{\Delta}_f^{-1}\mathbf{\Psi}\mathbf{S}\mathbf{\Psi}'\mathbf{\Delta}_f^{-1}).\right.$$

Row-by-row evaluation of the revision equation for $\boldsymbol{\beta}$ yields

$$n\left\{\begin{aligned} \tilde{\mathbf{c}}_{t|t} &= \tilde{\mathbf{c}}_{t|t-1} + \mathbf{S}_c\mathbf{h} \\ \tilde{\mathbf{a}}_{n,t|t} &= \tilde{\mathbf{a}}_{n,t|t-1} + \mathbf{S}_a\mathbf{h} + \boldsymbol{\lambda}_n\left(\frac{e_n - \boldsymbol{\psi}_n'\mathbf{S}\mathbf{h}}{f_n}\right) \qquad n = 1, \ldots, N. \end{aligned}\right.$$

Thus, a communal revision equal to **Sh** is made. Each cross-varying parameter estimate vector is further incremented by a multiple of the corresponding vector $\boldsymbol{\lambda}$.[1]

For revision of **M**, it is necessary to evaluate the term $-\mathbf{L}\mathbf{F}^{-1}\mathbf{L}'$. After substitution of the expressions for $\mathbf{F}^{-1}$ and **L**, use of the equality $\mathbf{P} - \mathbf{SHP} = (\mathbf{I} - \mathbf{SH})\mathbf{P} = \mathbf{SP}^{-1}\mathbf{P} = \mathbf{S}$ in partitioned form yields the revision formulae for the various components of the matrix:

$$o \left\{ \quad \bar{\mathbf{C}}_{t|t} = \mathbf{S}_{cc} \right.$$

$$(43) \quad \mathbf{D}_{n,t|t} = \mathbf{S}_{ca} - \mathbf{S}_c \frac{\boldsymbol{\psi}_n \boldsymbol{\lambda}'_n}{f_n}, \qquad n = 1, \ldots, N$$

$$(44) \quad \mathbf{A}_{mn,t|t} = \mathbf{S}_{aa} + \delta_{mn}\left(\bar{\mathbf{A}}_{m,t|t-1} - \frac{\boldsymbol{\lambda}_m \boldsymbol{\lambda}'_m}{f_m}\right) + \frac{\boldsymbol{\lambda}_m \boldsymbol{\psi}'_m \mathbf{S} \boldsymbol{\psi}_n \boldsymbol{\lambda}'_n}{f_m^2} - \frac{\boldsymbol{\lambda}_m \boldsymbol{\psi}'_m \mathbf{S}'_a}{f_m}$$
$$- \frac{\mathbf{S}_a \boldsymbol{\psi}_n \boldsymbol{\lambda}'_n}{f_n}, \qquad m = 1, \ldots, N \quad n = 1, \ldots, N.$$

From these formulae, it is apparent that **S** gives the variance in an individual estimate stemming from the communal sources of error *after* the new regression information has been incorporated.

The revised interindividual covariances (43), (44) are not identical unless $\mathbf{x}_{nt}, t = 1, \ldots T$ and R_n are the same for all individuals. Hence, if the regressors and disturbance variances are identical for all n, so that **M** satisfies (40) without adjustment, the "approximate" formulae in this section coincide with the true recursive formulae. When this is not the case, in order to preserve the simplifying conditions of the approximation, it is natural to force the interindividual covariances to equal their averages. This arbitrary adjustment is the sole cause of inefficiency in the approximation. The average values are:

$$\bar{\mathbf{D}}_{t|t} = \frac{\sum_{n=1}^{N} \mathbf{D}_{n,t|t}}{N} = \mathbf{S}_{ca} - \mathbf{S}_c\left(\frac{\sum_{n=1}^{N} (\boldsymbol{\psi}_n \boldsymbol{\lambda}'_n / f_n)}{N}\right)$$

$$o \quad \bar{\mathbf{A}}_{G,t|t} = \frac{\sum_{\substack{m,n=1 \\ m \neq n}}^{N} \mathbf{A}_{mn,t|t}}{N(N-1)} = \mathbf{S}_{aa} - \mathbf{S}_a\left(\frac{\sum_{n=1}^{N} (\boldsymbol{\psi}_n \boldsymbol{\lambda}'_n / f_n)}{N}\right)$$
$$-\left(\mathbf{S}_a\left(\frac{\sum_{n=1}^{N} (\boldsymbol{\psi}_n \boldsymbol{\lambda}'_n / f_n)}{N}\right)\right)' + \frac{\sum_{\substack{m,n=1 \\ m \neq n}}^{N} (\boldsymbol{\lambda}_m \boldsymbol{\psi}'_m \mathbf{S} \boldsymbol{\psi}_n \boldsymbol{\lambda}'_n / f_m f_n)}{N(N-1)}.$$

[1] Note that the factor multiplying $\boldsymbol{\lambda}$ is equal to that part of the forecast error not explained by the communal parameter revision, divided by the individual increment to forecast error variance. Thus, one part of the forecast error, $\boldsymbol{\psi}'_n\mathbf{Sh}$, is attributed to an error in estimating the population mean parameter vector; a proportion of the communally unexplained forecast error, equal to $\mathbf{z}'_n\boldsymbol{\lambda}_n/f_n$, is attributed to an error in estimating the individual cross-varying parameters; and the balance of the communally unexplained error, the proportion $1 - (\mathbf{z}'_n\boldsymbol{\lambda}_n/f_n) = R_n/f_n$, remains as a residual after revision of the parameter estimates. The communally unexplained forecast error is therefore divided between error in forecasting individual parameters and the individual disturbance in proportion to the contributions of these sources of error to prediction error variance.

For computational purposes, the last term can be simplified:

$$\frac{\sum\limits_{\substack{m,n=1\\ m\neq n}}^{N} (\boldsymbol{\lambda}_m \boldsymbol{\psi}_m' \mathbf{S} \boldsymbol{\psi}_n \boldsymbol{\lambda}_n' / f_m f_n)}{N(N-1)} = \frac{N}{N-1}\left(\frac{\sum_{n=1}^{N} (\boldsymbol{\lambda}_n \boldsymbol{\psi}_n' / f_n)}{N}\right)\mathbf{S}\left(\frac{\sum_{n=1}^{N} (\boldsymbol{\lambda}_n \boldsymbol{\psi}_n' / f_n)}{N}\right)' - \frac{\sum_{n=1}^{N} (\boldsymbol{\lambda}_n \boldsymbol{\psi}_n' \mathbf{S} \boldsymbol{\psi}_n \boldsymbol{\lambda}_n' / f_n^2)}{N(N-1)}.$$

The intra-individual variance increments are then set to be exact.

o{

$$\tilde{\mathbf{A}}_{n,t|t} = \mathbf{A}_{nn,t|t} - \tilde{\mathbf{A}}_{G,t|t} \qquad n = 1, \ldots, N.$$

When N is small, an "increment" $\tilde{\mathbf{A}}_{n,t|t}$ may occasionally fail to be positive definite in the first few periods of the sample, because the previous data for that individual have provided more information about an individual parameter than all sample data have provided about the sample mean. In this event, the approximated matrix $\tilde{\mathbf{M}}$ is not positive definite, and the method can break down. During the recursive algorithm, difficulties arise only when f_n is nonpositive in the following period $t + 1$, in which case the negative eigenvalues of $\tilde{\mathbf{A}}_{n,t|t}$ can be retrospectively adjusted to equal 0. After completion of the algorithm, the individual increments $\tilde{\mathbf{A}}_{n,T|T}$ can be checked for nonpositive eigenvalues, but this check is probably unnecessary, since nonpositive eigenvalues were never encountered in more than 150 simulations with $N = 10, 20$, or 40 at times $T = 10, 15$, or 20.

B. An Approximation to the Distribution of $\boldsymbol{\beta}_T$

In Maximum Likelihood estimation, the asymptotic approximate distribution for $\hat{\boldsymbol{\beta}}_{T|T}(\boldsymbol{\theta}_{ML})$ is normal $(\boldsymbol{\beta}_T, \hat{\sigma}^2_{ML}(\boldsymbol{\theta}_{ML})\mathbf{M}_{T|T}(\boldsymbol{\theta}_{ML}))$. In order for this distribution to be tractable, $\mathbf{M}_{T|T}$ may be approximated by $\tilde{\mathbf{M}}_{T|T}$, so that the variance matrix for $\boldsymbol{\beta}$ will satisfy (40). In Bayesian estimation, where $\boldsymbol{\beta}_{T|T}$ has the second moment given in (23), the numerical integration is facilitated by the use of $\tilde{\mathbf{M}}_{T|T}$ and by the further approximation:

$$\boldsymbol{\mu}_{T|T}(\boldsymbol{\theta}) - \bar{\boldsymbol{\beta}}_{T|T} \simeq \begin{pmatrix} \hat{\mathbf{c}}_{T|T}(\boldsymbol{\theta}) - \bar{\mathbf{c}}_{T|T} \\ \hat{\mathbf{a}}_{T|T}(\boldsymbol{\theta}) - \bar{\mathbf{a}}_{T|T} \\ \vdots \\ \hat{\mathbf{a}}_{T|T}(\boldsymbol{\theta}) - \bar{\mathbf{a}}_{T|T} \end{pmatrix} \quad \text{for all } \boldsymbol{\theta}. \tag{45}$$

After this simplification, the integrand satisfies (40), and hence $\tilde{\mathbf{M}}_{T|T}$ will satisfy (40) as well.

Statistical inference in the presence of a distribution with variance matrix $\tilde{\mathbf{M}}$ satisfying (40) requires evaluation of $|\tilde{\mathbf{M}}|$, and of the term

$$q = \begin{pmatrix} \tilde{\mathbf{c}} - \mathbf{c}^0 \\ \tilde{\mathbf{a}}_1 - \mathbf{a}_1^0 \\ \tilde{\mathbf{a}}_2 - \mathbf{a}_2^0 \\ \vdots \\ \tilde{\mathbf{a}}_N - \mathbf{a}_N^0 \end{pmatrix}' \tilde{\mathbf{M}}^{-1} \begin{pmatrix} \tilde{\mathbf{c}} - c^0 \\ \tilde{\mathbf{a}}_1 - \mathbf{a}_1^0 \\ \tilde{\mathbf{a}}_2 - \mathbf{a}_2^0 \\ \vdots \\ \tilde{\mathbf{a}}_N - \mathbf{a}_N^0 \end{pmatrix}.$$

By an application of the determinantal identity (42),

$$|\tilde{\mathbf{M}}| = \left| \begin{pmatrix} \tilde{\mathbf{C}} & & & & \\ & \mathbf{A}_1 & & \mathbf{0} & \\ & & \mathbf{A}_2 & & \\ & \mathbf{0} & & \ddots & \\ & & & & \tilde{\mathbf{A}}_N \end{pmatrix} + \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \\ \vdots & \vdots \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \tilde{\mathbf{D}} \\ \tilde{\mathbf{D}}' & \tilde{\mathbf{A}}_G \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \cdot \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{I} \cdot \mathbf{I} \end{pmatrix} \right| \tag{46}$$

$$= \begin{vmatrix} \tilde{\mathbf{C}} & \tilde{\mathbf{D}} \\ \tilde{\mathbf{D}}' & \tilde{\mathbf{A}}_G + \left(\sum_{n=1}^{N} \tilde{\mathbf{A}}_n^{-1} \right)^{-1} \end{vmatrix} \cdot \left| \sum_{n=1}^{N} \tilde{\mathbf{A}}_n^{-1} \right| \cdot \prod_{n=1}^{N} \left| \tilde{\mathbf{A}}_n \right|.$$

An application of the matrix inversion identity (41) yields a rank k formula for $\tilde{\mathbf{M}}^{-1}$. After some matrix manipulations, an expression for the statistic q may be derived in terms of the matrix

$$\mathbf{\Lambda} = \left(\sum_{n=1}^{N} \tilde{\mathbf{A}}_n^{-1} + (\tilde{\mathbf{A}}_G - \tilde{\mathbf{D}}'\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{D}})^{-1} \right):$$

$$q = \|\tilde{\mathbf{c}} - \mathbf{c}^0\|_{[\mathbf{C}^{-1}]} + \sum_{n=1}^{N} \|\tilde{\mathbf{a}}_n - \mathbf{a}_n^0 - \tilde{\mathbf{D}}'\tilde{\mathbf{C}}^{-1}(\tilde{\mathbf{c}} - \mathbf{c}^0)\|_{[\mathbf{A}_n^{-1}]} \tag{47}$$

$$- \left\| \sum_{n=1}^{N} \tilde{\mathbf{A}}_n^{-1}(\tilde{\mathbf{a}}_n - \mathbf{a}_n^0 - \tilde{\mathbf{D}}'\tilde{\mathbf{C}}^{-1}(\tilde{\mathbf{c}} - \mathbf{c}^0)) \right\|_{[\mathbf{\Lambda}^{-1}]}.$$

IV. The Statistical Efficiency and Validity of the Approximation

In this section, the properties of the approximation (hereafter referred to as A.I), conditional on $\boldsymbol{\theta}$ being correctly specified, will be analyzed. Upon examination of the recursive formulae that make up A.I, it may be seen to yield a linear unbiased estimator that is inefficient as a result of the simplifications in step (o). Recursive formulae for the *true* mean square error matrix of $\tilde{\boldsymbol{\beta}}_{T|T}$, as opposed to the approximation $\tilde{\mathbf{M}}$, may be derived. Then, for any $\boldsymbol{\theta}$, and for any set of explanatory variables $\mathbf{X}$, the exact properties of A.I may be computed, and two questions may be answered:

(i) How much larger is the mean square error of A.I than that of the exact, fully efficient method?

(ii) How valid is the approximated mean square error matrix $\tilde{\mathbf{M}}_{T|T}$ as an estimate of the true mean square error matrix for the approximate estimator, and how accurate is the approximated likelihood?

In addition, the properties of A.I may be compared with those of Ordinary Least Squares (OLS). These calculations, for a variety of convergent parameter regression structures $(\boldsymbol{\theta}, \mathbf{X})$, are reported in detail in Rosenberg (1973c, Sec. 5). The broad outlines will be summarized here.

A. Convergent Parameter Structures To Be Analyzed

Under the simplifying assumption that the cross-fixed parameters are constant over time and that the individual disturbance variances $R_n = R$ are identical for all n, a convergent parameter structure is specified by:

(i) the explanatory variables, $\mathbf{X}$

(ii) the communal disturbance variance, $\sigma^2 R_G$

(iii) the communal parameter shift variance, $\sigma^2 \mathbf{Q}_G$

(iv) the individual parameter shift variance, $\sigma^2 \mathbf{Q}_a$

(v) the convergence rates for parameters, $\phi_1, \ldots, \phi_\lambda$.

In selecting a set of representative structures among the infinite variety of options, the first problem is to construct the explanatory variables. The performance of the approximation is easily seen to be invariant to a linear transformation on the explanatory variables and a simultaneous inverse transformation on the parameter process. Accordingly, the explanatory variables can be normalized to have mean zero and variance unity, with inclusion of a constant being optional, provided that effects of changing scale are introduced through the parameter process. The correlation structure of the explanatory variables may be specified by four parameters, $\mathbf{X}(\rho_T, \rho_0, \rho_V, \rho_I)$, as follows:

$$\operatorname{corr}(x_{int}, x_{jmt}) = \begin{Bmatrix} \rho_0 \\ \rho_0 + \rho_V \\ \rho_0 + \rho_I \end{Bmatrix} \quad \text{for} \quad \begin{cases} i \neq j, & m \neq n \\ i = j, & m \neq n \\ i \neq j, & m = n \end{cases}$$

$$\operatorname{corr}(x_{int}, x_{jm,t-s}) = \rho_T^s \operatorname{corr}(x_{int}, x_{jmt}).$$

Thus, ρ_0 is the correlation between different variables for different individuals in the same period, ρ_V is the increment to this when the same variable is observed for different individuals, ρ_I is the increment when two different variables are observed for the same individual, and ρ_T is the attenuating factor for serial correlation. A set of pseudo-random, normally distributed explanatory variables obeying this correlation is easily constructed. In specifying $\boldsymbol{\theta}$, the covariances between parameter shifts for different parameters can be assumed to be zero, since variations in correlation are introduced in $\mathbf{X}$.

For each specification of $\mathbf{X}$, any combination of the remaining options—R_G, $\mathbf{Q}_G$, $\mathbf{Q}_a$, and $\boldsymbol{\Delta}_\phi$—may be selected. The stochastic specification can be summarized by two statistics: the average convergence rate, $\bar{\phi} = \sum_{i=1}^{\lambda} \phi_i/\lambda$, and the approximate proportion of variance due to parameter dispersion, $\bar{f} = \boldsymbol{\iota}'\boldsymbol{\Omega}\boldsymbol{\iota}/(\boldsymbol{\iota}'\boldsymbol{\Omega}\boldsymbol{\iota} + R + R_G)$. The first statistic captures the degree of serial memory in the parameter dispersion, and the second expresses the importance of parameter dispersion as a source of noise in the system.

In Rosenberg (1973c), efficiency and validity measures were computed for 166 structures. In all of these, κ and λ were set to 3. Cross-section sizes of $N = 10, 20, 40$ were tried, with 40 being the largest feasible cross section because efficiency evaluation requires calculations increasing as N^3. The performance of the approximation was evaluated after each five time periods through to a maximum of thirty time periods, and it was found to stabilize within fifteen periods. Accordingly, all results are based on evaluations after fifteen or more periods.

Fifty-one widely varied structures were tried first in an effort to discover which of the parameters in the specification most influenced efficiency. Then fifty-one additional structures were studied to analyze the effects of extreme values in the more influential parameters. Finally, a study of sixty-four structures was carried out to compare A.I with OLS, again for extreme values of the influential parameters. In these last structures, communal parameter shift variance $\sigma^2\mathbf{Q}_G$ was set to zero, so that the inefficiency observed in OLS would be due solely to nonresponsiveness to parameter dispersion.

The most important conclusions based on results in all the structures are summarized below. Also, detailed results for the last 64 structures are reported by grouping the results according to the presence or absence of serial correlation in $\mathbf{X}$, and by eight pairs of values for the two summary statistics $\bar{\phi}$ and $\bar{f}$. In this way, the 64 structures are segregated into 16 groups, and the results will be summarized by the worst value for each group. This simplification hides the systematic effects of variations other than serial correlation that were made in $\mathbf{X}$, but since these effects are small relative to the effect of serial correlation, the summary tables do give an accurate representation of the performance of the approximation.

B. The Statistical Efficiency of the Approximation

Each measure of efficiency will be reported as a percentage inefficiency, i.e., as $100(z_a/z_e - 1)$, where z_a is a mean square error measure for the method under analysis, and z_e is the same measure for the exact method. Perhaps the most interesting single measure of efficiency is the kth root of the determinant of the mean square error matrix (the "generalized mean square error") for the population mean parameter vector. The pattern of inefficiency is summarized in Table 1.

The inefficiency of A.I is far less than the inefficiency of OLS, but inefficiency does increase as serial correlation in $\mathbf{X}$ increases. Detailed analysis of mean square estimation errors for the separate parameters shows that almost all inefficiency in A.I arises in estimating the cross-fixed parameters. The maximal inefficiency of A.I. for a cross-fixed parameter is 95 percent, whereas the maximal inefficiency for a cross-varying parameter is only 2.5 percent. (OLS reaches 258 percent inefficiency for a cross-varying parameter.) In a large sample, the mean square error in cross-fixed parameters, even when inflated by substantial inefficiency, is very small relative to the mean square error in cross-varying parameters. For this

TABLE 1

MAXIMUM PERCENTAGE INEFFICIENCY IN GENERALIZED MEAN SQUARE ERROR FOR THE POPULATION MEAN PARAMETER VECTOR

(16 Groupings from 64 Different Specifications, with $N = 20$, $T = 15$)

Serial Correlation in X		Stochastic Specifications							
		$\bar{\phi} = 0.600$ $\bar{f} = 0.957$	0.833 0.977	0.600 0.938	0.833 0.972	0.800 0.963	0.517 0.938	0.800 0.971	0.517 0.980
$\rho_T = 0$	A.I	07	17	06	16	13	06	14	12
	OLS	232	392	269	378	338	234	368	536
$\rho_T = 0.6$ or $\rho_T = 0.9$	A.I	10	26	09	25	36	20	38	34
	OLS	317	546	269	510	363	267	369	644

reason, if the criterion of performance is taken as the arithmetic average of the eigenvalues of the mean square error matrix (rather than the geometric average implied by the generalized variance), A.I performs extremely well, with a maximum inefficiency of less than 5 percent versus over 200 percent for OLS.

The following influences of the parameters in the stochastic specification emerge:

(a) As N increases, the inefficiency of A.I tends to decrease.

(b) As ϕ increases for any parameter, inefficiency increases for that parameter, and as $\bar{\phi}$ increases for a regression, inefficiency increases for that regression.

(c) As $\bar{f}$ increases for a regression, inefficiency increases.

(d) As the communal parameter shift variance $\mathbf{Q}_G$ increases, inefficiency decreases.

(e) The variance of the communal disturbance, R_G, has little effect.

(f) With regard to the structure of the explanatory variables, the presence of a constant has little effect, the presence of serial correlation increases inefficiency, the presence of correlation across variables for the same individual has little effect, and correlation of the variables across individuals reduces inefficiency. The last is to be expected, since if the correlation rises to one, the approximation becomes exact and hence perfectly efficient.

Comparison of forecasting efficiency provides another important test of the approximation. Consider forecast errors for single dependent variables ($e_{nT} = y_{nT} - \mathbf{x}'_{nT}\hat{\mathbf{b}}_{n,T|T-1}$, $n = 1, \ldots, N$) and for the population aggregate ($e_{\cdot T} = \sum_n y_{nT} - \sum_n \mathbf{x}'_{nT}\hat{\mathbf{b}}_{n,T|T-1}$). The sources of error are the unpredictable disturbances and parameter shifts in period T, and the estimation error for the parameters in period $T-1$. Differences across methods in mean square estimation error in period $T-1$ therefore determine differences in the mean square forecast error. Moreover, since the explanatory variables are generated by a stationary stochastic process, the mean square forecast error weighs the efficiency of estimating various dimensions of the parameter vectors by the expected magnitude of the components of the explanatory variables corresponding to these dimensions.

For A.I, two possible forecasting procedures are available: to forecast each individual by the estimated parameters for that individual (Method I), or to forecast all individuals by the population mean parameter estimate (Method M). Method M should be less efficient, since it discards the disaggregated parameter estimates. For OLS with fixed parameters, these two methods coincide.

The criterion of forecasting performance for the single dependent variables is the sum of the mean square errors in the individual forecasts:

$$S_I = \sum_{n=1}^{N} E[(y_{nT} - \mathbf{x}'_{nT}\hat{\mathbf{b}}_{n,T|T-1})^2], \qquad S_M = \sum_{n=1}^{N} E[(y_{nT} - \mathbf{x}'_{nT}\hat{\mathbf{b}}_{T|T-1})^2],$$

where the subscripts indicate the use of individual or population mean parameter estimates. For the aggregate forecast, the criterion is the mean square error:

$$A_I = E\left[\left(\sum_{n=1}^{N} y_{nT} - \sum_{n=1}^{N} \mathbf{x}'_{nT}\hat{\mathbf{b}}_{nT|T-1}\right)^2\right],$$

$$A_M = E\left[\left(\sum_{n=1}^{N} y_{nT} - \left(\sum_{n=1}^{N} \mathbf{x}'_{nT}\right)\hat{\mathbf{b}}_{T|T-1}\right)^2\right].$$

TABLE 2

Maximum Percent Inefficiency in Sum of Mean Square Errors in Individual Forecasts (16 Groupings from 64 Different Specifications, with $N = 20$, $T = 15$)

			Using the Individual Parameter Forecasts (S_I)	Using the Forecast Population Mean Parameters (S_M)	
Specification			A.I	A.I	OLS
	$\bar\phi = 0.600$	0.957	1	39	53
	0.833	0.977	1	138	131
	0.600	0.938	0	29	43
	0.833	0.972	0	156	149
$\rho_T = 0$	0.800	0.963	1	97	117
	0.517	0.938	1	45	53
	0.800	0.971	0	78	118
	0.517	0.980	1	189	212
	$\bar\phi = 0.600$	$\bar f = 0.957$	1	76	71
	0.833	0.977	0	155	184
$\rho_T = 0.6$	0.600	0.938	0	56	55
or	0.833	0.972	0	166	199
$\rho_T = 0.9$	0.800	0.963	0	115	131
	0.517	0.938	0	76	62
	0.800	0.971	0	141	130
	0.517	0.980	0	338	233

TABLE 3

Maximum Percent Inefficiency in Mean Square Error in Forecasting the Aggregate (16 Groupings from 64 Different Specifications, with $N = 20$, $T = 15$)

			Using the Individual Parameter Forecasts (A_I)	Using the Forecast Population Mean Parameters (A_M)	
Specification			A.I	A.I	OLS
	$\bar\phi = 0.600$	$\bar f = 0.957$	1	85	148
	0.833	0.977	3	28	170
	0.600	0.938	1	51	153
	0.833	0.972	3	8	183
$\rho_T = 0$	0.800	0.963	4	15	255
	0.517	0.938	4	44	278
	0.800	0.971	2	56	234
	0.517	0.980	14	228	326
	$\bar\phi = 0.600$	$\bar f = 0.957$	1	50	112
	0.833	0.977	3	60	182
$\rho_T = 0.6$	0.600	0.938	1	29	105
or	0.833	0.972	2	71	205
$\rho_T = 0.9$	0.800	0.963	1	139	148
	0.517	0.938	1	5	79
	0.800	0.971	1	10	131
	0.517	0.980	0	12	210

In Tables 2 and 3, maximal percentage inefficiences of A.I and OLS are compared. A.I is almost perfectly efficient in forecasting the individual dependent variables but suffers a percentage inefficiency of up to 14 percent in forecasting the aggregate, due to relatively greater inefficiency in estimating the cross-fixed parameters. OLS has a percentage inefficiency of more than 200 percent in many cases. Notice that the results are dependent upon the $(\mathbf{X}, \boldsymbol{\theta})$ specifications chosen, but that for each specification, the results are the exact theoretical values, not the output of some sampling experiment.

C. Validity of Approximated Mean Square Error and Goodness of Fit

Let $\tilde{\sigma}^2$ denote $s^2(\boldsymbol{\theta})$ from A.I or from OLS, and let $\hat{\sigma}^2$ denote $s^2(\boldsymbol{\theta})$ from the Exact Method. Let $\tilde{l}$ and l denote the approximate and exact log likelihoods of the true structure, and let l_{FP} denote the exact log likelihood of the fixed-parameter structure.

In order to validate the approximated mean square error yielded by A.I or OLS, the statistics

$$V_n = \sqrt[k]{\frac{|\text{approximated mean square error matrix for } \bar{\tilde{\mathbf{b}}}_n|}{|\text{true mean square error matrix for } \bar{\tilde{\mathbf{b}}}_n|}} \quad \text{for } n = 1, \ldots, N$$

and

$$V_{\cdot} = \sqrt[k]{\frac{|\text{approximated mean square error matrix for } \bar{\tilde{\mathbf{b}}}|}{|\text{true mean square error matrix for } \bar{\tilde{\mathbf{b}}}|}}$$

are computed. The generalized mean square error ratios V_n are relatively constant across the population, so their value is summarized by the arithmetic mean $\bar{V} = \sum_{n=1}^{N} V_n/N$. The effect of estimation error in σ^2, which is omitted in these ratios, is introduced by computation of the additional ratios $(\tilde{\sigma}^2/\hat{\sigma}^2)V_{\cdot}$ and $((\tilde{\sigma}^2/\hat{\sigma}^2)\bar{V})$. The ratio $(\tilde{\sigma}^2/\hat{\sigma}^2)$ and the difference in log likelihoods are also computed. If A.I were exact, all ratios would be equal to their ideal value of unity, and the difference in log likelihood would be zero.

The results show a clear pattern. The validity of the approximation increases with N in more than 95 percent of the cases, an extremely encouraging property since sample sizes will be much larger in applications. Moreover, as the sample size doubles from $N = 20$ to $N = 40$, the difference $l - \tilde{l}$ declines in almost all cases, although the magnitude of l typically doubles. Thus, the proportional error in l declines more rapidly than $1/N$. If these results persist in large samples, the approximated log likelihood should be virtually perfect.

The values of the statistics that deviated most from the ideal values are given in Table 4 for the sixty-four structures already reported. The approximation is everywhere more valid than OLS. Moreover, the error in the approximated log likelihood is nowhere more than one-twentieth of the difference between the approximated log likelihood for the convergent-parameter structure and the log likelihood of the fixed-parameter structure. Hence, the approximated log likelihood reliably rejects the fixed-parameter model despite the small sample size.

TABLE 4

APPROXIMATED VERSUS TRUE PROPERTIES OF A.I AND OLS: MOST DEVIANT CASES

$N = 20$ $T = 15$[a]

		$\bar{V}$	$V.$	$\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\bar{V}$	$\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}V.$	$\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}$	$l - l$ or $l_{FP} - l$
Ideal Values		1.0	1.0	1.0	1.0	1.0	0.0
$\rho_T = 0$	A.I	1.025	1.028	1.043	1.039	1.032	8.5
	OLS	0.002	0.007	0.069	0.239	64.66	−369.4
$\rho_T = 0.6$ or $\rho_T = 0.9$	A.I	0.493	0.446	0.499	0.451	1.027	8.9
	OLS	0.001	0.003	0.048	0.117	67.91	−269.6

[a] For OLS, the values computed under the erroneous assumption of fixed parameters are compared to the true properties of OLS. The difference in log likelihoods is an exception: the figure is the (exact) log likelihood of fixed parameters minus the exact log likelihood of the true structure.

Throughout the results, A.I appears to be entirely valid when the explanatory variables are serially independent, but to understate the estimation error variance when the explanatory variables are serially correlated. In the most severe case, one with serial correlation of 0.9, the approximated mean square error falls to 45 percent of the true value. This is a serious defect, in view of the prevalence of serial correlation in economic variables. It will have to be taken into account in applications. Fortunately, the degree of understatement decreases with N and, in large samples, the downward bias may be small. It is interesting to note that the approximated sampling properties of OLS are far worse. In fact, the estimated generalized mean square error of OLS falls below one-twentieth of the true value for individual parameters and below one-ninth of the true value for the population mean parameters. These deficiencies highlight the dangers of using the fixed-parameter assumption where it is inappropriate.

In summary, the approximation is highly efficient in estimating the cross-varying parameters and satisfactorily efficient in estimating the cross-fixed parameters, and the approximated likelihood can apparently be used with confidence. The only defect of the approximation that must be taken into account is understatement of the mean square error in the case of serially correlated explanatory variables. Subject to this caution, the approximation may be substituted into the recursive formulae of Section II. The results also imply that the method is sharply superior to ordinary least squares—in terms of efficiency and in terms of validity of sampling theory—when parameter dispersion is present. These results are overly favorable to the method, since $\boldsymbol{\theta}$ is presumed known, whereas, in fact, it must be estimated. However, the very large difference in sample log likelihood between the true structure and the fixed-parameter structure suggests that, if $\boldsymbol{\theta}$ were estimated by maximum approximated likelihood, then the estimated structure would be relatively close to the true structure. Hence, much of the gain in efficiency due to recognition of parameter variation would be achieved. Moreover, the very large sample sizes in many cross-section, time-series applications promise excellent estimates of $\boldsymbol{\theta}$, and therefore full exploitation of the potential efficiency of the method—provided, of course, that the model permits an appropriate description of the true parameter process.

Finally, notice that the computations involved in the method are feasible: the calculations required to evaluate a single stochastic specification with $N = 40$ were equivalent to repeating the approximation more than 500 times, enough iterations for Maximum Likelihood estimation or Bayesian estimation with $\boldsymbol{\theta}$ of reasonable dimension.

V. Conclusion

There are numerous extensions of the method that need not be added to an already lengthy paper. "Smoothed" estimates of parameter vectors $\boldsymbol{\beta}_t$ for $t < T$ may be computed by modifications of the recursive formulae derived here (see, e.g., Rosenberg (1973b)). A more complex model, where individual parameters converge to subgroup norms, which in turn may converge to the population norm, is relatively easy to implement. An underlying population mean, which serves as the norm for convergence in place of the sample mean in every period, may be added to the model if variations in the sample mean are not desired to affect the convergence pattern. Nonconstant variances or convergence rates, which differ across individuals or over time as functions of known characteristics of the individual or time period may be easily introduced, and the parameters specifying these functions may be adjoined to $\boldsymbol{\theta}$ without changing the estimation approach.

To summarize, a model of parameter variation in a cross section of time series was presented, in which individual parameters obey random walks subordinated to a tendency to converge toward the population norm. The model involves an intuitively plausible dynamic model of the determinants of individual diversity, and it is consistent with the empirical observation that, in some cross sections of time series, individual parameters vary relative to one another as if subjected to sequential random increments, but that cross-sectional parameter dispersion nevertheless remains roughly constant. Next, a computationally feasible method for Maximum Likelihood or Bayesian estimation of the parameters specifying the stochastic structure, as well as of the individual regression parameters themselves, was derived. The approximation involved in these computations was validated, subject to the one defect of understating mean square error when explanatory variables are serially correlated. The method was shown to be superior to Ordinary Least Squares in the presence of stochastic parameter variation of the type conjectured.

University of California, Berkeley

Bibliography

Aoki, Masanao (1967), *Optimization of Stochastic Systems*, New York: Academic Press.

Balestra, P. and Nerlove, M. (1966), "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica*, 34 (July), 585–612.

Chetty, V. K. (1968), "Pooling of Time Series and Cross Section Data," *Econometrica*, 36 (April), 279–290.

Hildreth, Clifford (1949), "Preliminary Considerations Regarding Time Series and/or Cross-Section Studies," Cowles Commission Discussion Paper No. 333.

———. (1950), "Combining Cross Section and Time Series Data," Cowles Commission Discussion Paper No. 347.

Ho, Y. C. and Lee, R. C. K. (1964), "A Bayesian Approach to Problems in Stochastic Estimation and Control," *IEEE Transactions on Automatic Control*. AC-9, pp. 333–339.
Hoch, Irving (1962), "Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data," *Econometrica*. 30 (January), 34–53.
Hsiao, Cheng (1973), "Statistical Inference for a Model with Both Random Cross-Sectional and Time Effects," Technical Report No. 83, Stanford, Calif.: Institute for Mathematical Studies in the Social Sciences, Stanford University.
Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," Transactions of ASME, Series D, *Journal of Basic Engineering*, 82, pp. 35–45.
Kalman, R. E. and Bucy, R. S. (1961), "New Results in Linear Filtering and Prediction Theory," Transactions of ASME, Series D, *Journal of Basic Engineering*, 83, pp. 95–108.
Kuh, E. (1959), "The Validity of Cross-Sectionally Estimated Behavioral Equations in Time Series Applications," *Econometrica*, 27 (April), 197–214.
Kuh, E. and Meyer, J. R. (1957), "How Extraneous Are Extraneous Estimates?" *Review of Economics and Statistics*, 39 (November), 380–393.
Mundlak, Yair (1963), "Estimation of Production and Behavioral Functions from a Combination of Cross-Section and Time-Series Data," *Measurement in Economics*. Edited by Carl F. Christ, *et al.* Stanford, Calif.: Stanford University Press.
Nerlove, Marc (1965), *Estimation and Identification of Cobb-Douglas Production Functions*, Chicago: Rand McNally & Co.
———. (1968), "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections," Cowles Commission Discussion Paper No. 257.
Parks, Richard W. (1967), "Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated," *JASA*, 62 (June), 500–509.
Rao, C. R. (1965), "The Theory of Least Squares When the Parameters Are Stochastic," *Biometrika*, 52, pp. 447–458.
Rosenberg, Barr (1973a), "Linear Regression with Randomly Dispersed Parameters," *Biometrika*, 60, pp. 65–72.
———. (1973b), "Estimation in the General Linear Stochastic Parameter Model," Research Report. Cambridge, Mass: National Bureau of Economic Research, Computer Research Center.
———. (1973c), "Varying Parameter Regression in the Analysis of a Cross Section of Time Series," Working Paper No. IP 165. Berkeley: Institute of Business and Economic Research, University of California.
Sharpe, William F. (1970), *Portfolio Theory and Capital Markets*, New York: McGraw Hill Book Company.
Swamy, P. A. V. B. (1970), "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38, pp. 311–323.
———. (1971), *Statistical Inference in Random Coefficient Regression Models*, Berlin-Heidelberg-New York: Springer-Verlag.
Telser, Lester (1964), "Iterative Estimation of a Set of Linear Regression Equations," *JASA*, 59 (September), 845–862.
Wallace, T. D. and Hussain, Ashiq (1969), "The Use of Error Components Models in Combining Cross Section with Time Series Data," *Econometrica*, 3 (January), 37.1, 55–72.
Wilks, S. S. (1943), *Mathematical Statistics*, Princeton, N.J.: Princeton University Press.
Zellner, Arnold (1962), "An Efficient Method of Estimating Seemingly Unrelated Relations and Tests for Aggregation Bias," *JASA*, 57 (June), 348–367.
———. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley & Sons.