

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

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: A Survey of Stochastic Parameter Regression

Chapter Author: Barr Rosenberg

Chapter URL: <http://www.nber.org/chapters/c9933>

Chapter pages in book: (p. 381 - 397)

## A SURVEY OF STOCHASTIC PARAMETER REGRESSION

BY BARR ROSENBERG<sup>1</sup>

*Several important models of stochastic parameter variation are described: randomly dispersed parameters, sequentially varying parameters, stationary stochastic parameters, cross-section time-series models, and the shifting regressions approach. Theories and methods of stochastic parameter regression are surveyed.*

### I. STOCHASTIC PARAMETER REGRESSION

This article is a short survey of some of the more important issues in the theory of regression with stochastically varying parameters. The stochastic parameter regression problem arises when regression coefficients vary unsystematically in the familiar linear regression model:

$$(1) \quad y_{nt} = \sum_{i=1}^k b_{int} x_{int} + u_{nt} = \mathbf{x}'_{nt} \mathbf{b}_{nt} + u_{nt} \quad n = 1, \dots, N, \quad t = 1, \dots, T$$

where  $n$  connotes an individual within a cross section,  $t$  connotes a time period, and where one or the other of the subscripts will be suppressed when inappropriate. The regression parameters or coefficients  $b_1, \dots, b_k$ , one of which may be the intercept, are written with subscripts  $n$  and  $t$  to permit variation across individuals and time periods. Many articles in the literature on stochastic parameters have provided arguments for the nonconstancy of coefficients across observations, and it would be inappropriate to repeat these here. Suffice it to note that if the regression coefficients are to be regarded as the true partial derivatives of  $y_{nt}$  with respect to the regressors  $x_{int}$ , then it is improbable that these partial derivatives will be identical for two different observations.

Two types of parameter variation must be distinguished: *systematic* and *stochastic*. In systematic variation, the individual parameter vectors may be written as  $\mathbf{b}_{nt} = f(\lambda, \mathbf{z}_{nt})$  where the parameters  $\lambda$  specify a functional form determining  $\mathbf{b}_{nt}$  as a function of observable variables  $\mathbf{z}_{nt}$ . These observables may include the regressors  $\mathbf{x}_{nt}$  themselves, if the true regression model is nonlinear, as well as other characteristics of the individual. When parameter variation is systematic, the regression problem is a (possibly nonlinear) regression on  $\mathbf{x}_{nt}$  and  $\mathbf{z}_{nt}$  to estimate the parameters  $\lambda$ , and ordinary regression theory is applicable. For a discussion of these matters, see Belsley (1973).

The stochastic parameter problem arises when parameter variation includes a component which is a realization of some stochastic process in addition to whatever component is related to observable variables. Thus, stochastic parameter regression is a generalization of ordinary regression. Ideally, a model would be so well specified that no stochastic parameter variation would be present, and

<sup>1</sup> Assistant Professor of Business Administration, University of California, Berkeley. This research was supported by NSF grant GS 3306. This is, of necessity, a brief survey of the theory, and there must inevitably be omissions and perhaps errors as well, for which the author offers his apologies.

no generalization would be needed, but the world is less than ideal. The distinction between stochastic and systematic variation is particularly important, because in most models, stochastic parameter variation is assumed to be uncorrelated with the explanatory variables. Hence, these models are not intended to account for parameter variation resulting from misspecification in a nonlinear model, despite frequent assertions to this effect in the literature. Stochastic parameter regression should be employed as a supplement to analysis of systematic variation, rather than as an alternative.

All of the stochastic parameter processes that have been considered in economics may be written in the general form:

$$(2) \quad \mathbf{b}_{nt} = \mathbf{A}_{nt}(\boldsymbol{\theta})\mathbf{m} + \mathbf{B}_{nt}(\boldsymbol{\theta})\boldsymbol{\eta}, \quad n = 1, \dots, N, \quad t = 1, \dots, T$$

where  $\mathbf{m}$  is a  $(j \times 1)$  vector of unknown parameters,  $\boldsymbol{\eta}$  is a large vector of stochastic terms,  $(p \times 1)$  say, having a proper joint distribution with mean  $\bar{\boldsymbol{\eta}}(\boldsymbol{\theta})$  and variance matrix  $\sigma^2\boldsymbol{\Lambda}(\boldsymbol{\theta})$ , with  $\sigma^2$  a scale parameter, and where the notation  $(\boldsymbol{\theta})$  indicates that these terms may depend on a vector of *stochastic specification parameters*  $\boldsymbol{\theta}$ . When these relations are substituted into the linear regression (1) and all observations in the regression are combined, the result is a linear statistical model:

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{NT} \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_{11} & \mathbf{A}_{11}(\boldsymbol{\theta}) \\ \mathbf{x}'_{12} & \mathbf{A}_{12}(\boldsymbol{\theta}) \\ \vdots & \vdots \\ \mathbf{x}'_{NT} & \mathbf{A}_{NT}(\boldsymbol{\theta}) \end{pmatrix} \mathbf{m} + \begin{pmatrix} \mathbf{x}'_{11} & \mathbf{B}_{11}(\boldsymbol{\theta}) \\ \mathbf{x}'_{12} & \mathbf{B}_{12}(\boldsymbol{\theta}) \\ \vdots & \vdots \\ \mathbf{x}'_{NT} & \mathbf{B}_{NT}(\boldsymbol{\theta}) \end{pmatrix} \boldsymbol{\eta} + \begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{NT} \end{pmatrix}$$

or, in matrix form,

$$(3) \quad \mathbf{y} = \boldsymbol{\Xi}(\boldsymbol{\theta})\mathbf{m} + \boldsymbol{\Psi}(\boldsymbol{\theta})\boldsymbol{\eta} + \mathbf{u}.$$

To see that the model differs little from the traditional regression model, notice that it may be rewritten as:

$$(4) \quad \mathbf{y} = \boldsymbol{\Xi}(\boldsymbol{\theta})\mathbf{m} + (\boldsymbol{\Psi}(\boldsymbol{\theta}):\mathbf{I}) \begin{pmatrix} \boldsymbol{\eta} \\ \mathbf{u} \end{pmatrix}.$$

Thus, the model is a regression with an augmented vector of stochastic terms including both stochastic parameters and disturbances. In contrast to ordinary regression, where the disturbances are customarily eliminated as nuisance parameters at the first opportunity, in stochastic parameter regression the estimators for the stochastic terms are followed through carefully. However, as Theil has shown (1971, Ch. 5), there are implied estimators for the stochastic terms in the ordinary regression model, and Theil's BLUS procedure is an instance of the general procedures for optimal estimation of the stochastic terms in a linear statistical model such as (4).

The regression may be written in yet another way as:

$$(5) \quad \mathbf{y}^*(\boldsymbol{\theta}) \equiv \mathbf{y} - \boldsymbol{\Psi}(\boldsymbol{\theta})\bar{\boldsymbol{\eta}}(\boldsymbol{\theta}) = \boldsymbol{\Xi}(\boldsymbol{\theta})\mathbf{m} + \mathbf{v},$$

where

$$\mathbf{v} = (\boldsymbol{\Psi}(\boldsymbol{\theta}):\mathbf{I}) \begin{pmatrix} \boldsymbol{\eta} - \bar{\boldsymbol{\eta}}(\boldsymbol{\theta}) \\ \mathbf{u} \end{pmatrix}.$$

Assume that  $E(\mathbf{u}) = \mathbf{0}$ ,  $\text{var}(\mathbf{u}) = \sigma^2 \Sigma(\theta)$ ,  $\text{cov}(\mathbf{u}, \eta) = \mathbf{0}$ . Then,  $E(\mathbf{v}) = \mathbf{0}$ ,  $\text{var}(\mathbf{v}) = \sigma^2(\Psi(\theta)\Lambda(\theta)\Psi'(\theta) + \Sigma(\theta)) \equiv \sigma^2 \mathbf{V}(\theta)$ . Thus, the vector  $\mathbf{v}$  has the properties of a heteroscedastic disturbance vector, and the stochastic parameter regression model, insofar as it implies an estimator for  $\mathbf{m}$ , is seen to be no more than an ordinary regression with a complicated covariance among the "disturbances."

## II. STOCHASTIC PARAMETER MODELS

Before discussing the several stochastic parameter models that have been proposed in the literature, the concept of the Markovian Canonical Form for a model, which will be needed in the discussion, will be introduced.

### A. Markovian Canonical Form

A stochastic parameter regression model is in *Markovian Canonical Form* (MCF), when the (possibly transformed) set of parameter vectors,  $\mathbf{b}_1^*, \dots, \mathbf{b}_r^*$ , satisfy the following conditions:

- (i) The observations of the dependent variables appear in linear regressions of the form  $y_i^* = \Psi_i^*(\theta)\mathbf{b}_i^* + \mathbf{u}_i^*$ , such that for  $i \neq j$ ,  $\text{cov}(\mathbf{u}_i^*, \mathbf{u}_j^*) = \mathbf{0}$ .
- (ii) The joint distribution of the parameters may be represented by prior distributions for the parameter vectors, which are uncorrelated across different parameter vectors; and by a set of linear *transition relations*, each linking some pair of parameter vectors in the form  $\mathbf{b}_i^* = \Phi_{ij}(\theta)\mathbf{b}_j^* + \mathbf{d}_{ij}$ , where each stochastic *parameter shift* vector  $\mathbf{d}_{ij}$  has a specified mean and variance which may depend on  $\theta$ , and where each shift vector is uncorrelated with all other shift vectors and with all disturbances.

A model in MCF may be represented visually by a graph, in which the vertices correspond to the parameter vectors, and the links between vertices to the transition relations. For instance, in a time series regression in which disturbances are serially uncorrelated and parameters follow a first-order Markov process, so that successive parameter shifts are uncorrelated, the model is in MCF represented by the graph in Figure 1.



Figure 1

A model that is not immediately in MCF may often be transformed into this form by adjoining to the parameter vector *state variables* that transmit the correlation in the model. To clarify this, consider the familiar fixed-parameter model with first-order autoregressive disturbance process given by  $u_t = \rho u_{t-1} + e_t$ , where  $E(e_t) = 0$ ,  $\text{var}(e_t) = \sigma^2$ . By regarding the disturbance  $u_t$  as the "stochastic parameter," the model is represented in MCF, with graph as in Figure 1, as follows:

*Regression relations:*

$$y_t = (\mathbf{x}_t' : 1) \begin{pmatrix} \mathbf{b} \\ u_t \end{pmatrix} \quad t = 1, \dots, T$$

*Prior distribution:*

$$E(u_1) = 0, \quad \text{var}(u_1) = \sigma^2 / (1 - \rho^2)$$

Transition relations:

$$u_{t+1} = \rho u_t + e_t, \quad t = 1, \dots, T-1,$$
$$E[e_t] = 0, \quad \text{var}[e_t] = \sigma^2.$$

This illustration shows two interesting characteristics of stochastic parameter regression. For one, the distinction between disturbances and stochastic parameters is purely arbitrary. Secondly, there may, in fact, be no true disturbances at all; when  $u$  becomes a parameter, no additive disturbance remains.

The essential characteristic of the MCF is the absence of correlation across regression observations and transition relations. The regression observations and/or prior distribution for any parameter vector then provide an *atom* of information about that parameter vector, with random components uncorrelated with other stochastic terms in the model, which may be processed by the rules of ordinary fixed-parameter regression. Since the linear transition relations between parameters are uncorrelated with the atoms of information, the atoms may be combined by applying, at each transition, the rule for a given linear transformation with an uncorrelated additive stochastic shift. All information in the regression may be combined by recursive application of this procedure of *stepwise composition*, incorporating the information for each parameter vector by ordinary regression, and combining information by linear stochastic transformation or *extrapolation*. The process may be visualized as a tracing out of the links in the graph representing the MCF, with the step along each link being an extrapolation of information collected at the previous vertex, and with the action at each vertex being the combination of the atom of regression information for that parameter vector with the extrapolated prior information.

This approach to deriving estimators in the stochastic regression problem may be termed *atomistic* (Rosenberg, 1968b). It contrasts with the alternative *holistic* approach of direct derivation from the general linear form (3). Whatever the method by which the estimators are derived, the great significance of the MCF lies in the fact that the matrix algebra required for the atomistic regressions and for the extrapolations is never of greater dimension than the dimension of the parameter vectors  $b_t^*$ .<sup>2</sup> Hence, it follows that an MCF with parameter vectors of low dimension is a sufficient condition for estimation in the model to be computationally feasible. In addition, when the model cannot be exactly placed in MCF, an approximation in MCF often suggests an approximate computational procedure.

### B. Randomly Dispersed Parameter Models

The term "random," as opposed to "stochastic," is reserved for the specific model in which regression parameter vectors are random drawings from a common multivariate distribution, with mean vector  $m$  and variance matrix  $\sigma^2\Omega$ .

<sup>2</sup> For instance, the models proposed by Cooley and Prescott (1973c) and Sarris (1973) are easily transformed to have the MCF in Figure 1; with this insight, application of the formulae in Rosenberg ((1968a), (1973c, Sect. II)) allow the  $N$ -dimensional matrix operations proposed by these authors to be avoided entirely. There is a substantial savings in required computations when  $N$  is large (more than 4 orders of magnitude in the "Capital Market Application" mentioned by Cooley and Prescott).

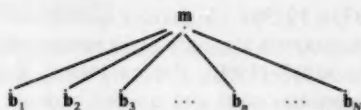


Figure 2

When only the constant term varies, the model becomes the analysis of covariance with random effects. If the parameters are represented as  $b_n = m + \eta_n$ ,  $n = 1, \dots, N$ , where  $m$  is the population mean vector, the process is in the general form (2). When disturbances are uncorrelated across the individual parameter vectors, the model falls immediately into the MCF in Figure 2.

The model is clearly applicable to a random sample of individuals in which only one observation is taken on each individual. When there are repeated observations on an individual, the model applies if successive parameter vectors for the individual are independently drawn, as assumed by Fisk (1967), or if the successive realizations for each individual are identical, as assumed by Rao (1965b).

Several early studies examined the case where only a single observation corresponded to each parameter vector. Rubin [1950] wrote down the likelihood function, Theil and Mennes (1959) and Hildreth and Houck (1968) studied estimation of  $m$  and  $\Omega$ , assuming  $\Omega$  to be diagonal, and Fisk (1967) studied this case with general  $\Omega$ . Rao (1965a, pp. 192-193; 1965b) and recently Efron and Morris (1972) studied the second case where multiple observations are generated by each individual parameter vector, but with the assumption that the regressor matrix is of rank  $k$  and identical for all individuals. Swamy (1970, 1971, and 1973) permitted differing  $X_n$  but retained the requirement of rank equal to  $k$ . Rosenberg (1973a) relaxed the rank  $k$  assumption. Applications of the model appear in Swamy (1971), Rosenberg and McKibben (1973), and Sheiner, Rosenberg, and Melmon (1972). In the last, the regressions are nonlinear, and an approximation suggested by the MCF is employed.

### C. Sequential or Markov Parameter Models

In these models, parameters follow a first-order Markov process, or more generally, an autoregressive or moving-average process of low order. The model is naturally applied to a time series, with the stochastic parameter process introducing random drift in the parameters. The process is represented as  $b_{t+1} = \Phi b_t + \eta_t$ ,  $t = 1, \dots, T$ . In order to place it in the general form (2), the parameter vector in the initial period may be chosen as the unknown vector  $m$ , and the parameters may be represented as  $b_t = \Phi^{t-1}m + \sum_{s=1}^{t-1} \Phi^{t-s-1}\eta_s$ ,  $t = 1, \dots, T$ . The MCF was given in Figure 1. This general model was first analyzed by Kalman (1960) and Kalman and Bucy (1961), who originated an extensive literature in control theory and the applied physical sciences, in which the optimal estimation methods are often referred to as the Wiener-Kalman or Kalman-Bucy filter.

A range of studies have introduced applications of this model into economics and statistics: Rosenberg (1967, 1968a), Teräsvirta (1970), Duncan and Horn (1972), Bowman and LaPorte (1972). Where only the constant or intercept term varies, the special case of the *adaptive regression model* arises: Cooley (1971) and

Cooley and Prescott (1973a, 1973b). The further special case of time series analysis, where a time varying constant is the only term present, has been studied extensively: Swerling (1959), Muth (1960), Parzen (1961), Box and Jenkins (1962, 1970). Another family of special cases was studied in depth by Brown (1963). See, also, the articles in this issue by Cooley and Prescott (1973c), Sarris (1973), and Rosenberg (1973c).

#### *D. Stationary Stochastic Parameter Models*

Burnett and Guthrie (1970) introduced the general model where parameter variation over time follows a stationary stochastic process. Although this model is of formal interest, it suffers from the defect that it cannot, in general, be transformed into Markovian Canonical Form with a parameter vector of low dimension, so that the computational problems associated with it are horrendous.

#### *E. Cross-Section Time-Series Models*

Statisticians have perceived the need for responsiveness to parameter variation most clearly in cross-section time-series analysis, not necessarily because it is intrinsically any more necessary here than in the analysis of a single cross section or a single time series, but rather because the greater number of observations allows more degrees of freedom to deal with the problem. The traditional method allows the intercept in the regression to vary randomly over time (time effects) and across the population (individual effects). An extensive literature on this subject, which is an interesting case of stochastic parameter regression, exists. Some important recent contributions are Tiao and Tan (1965, 1966), Chetty (1968), Wallace and Hussain (1969), Amemiya (1971), Maddala (1971), Swamy (1971, Ch. 3), Swamy and Arora (1972). Swamy has applied the random parameter model to a cross section of time series, assuming that individual parameters are fixed over time. He has also analyzed the case where a random time effect occurs each period; these parameters must be adjoined to the individual parameter vectors to bring the model into the MCF in Figure 2. Hsaio (1972, 1973) has proposed an extension, in which the parameter vectors are the sum of random individual parameter vectors and random time vectors, so that the traditional procedure applied to the intercept is generalized to the entire parameter vector (1972a). The MCF for this model cannot be fully simplified, and Hsaio's decomposition of the computational formulae requires a matrix inversion of rank  $Nk$  (1973, Eq. A.12), an infeasible computation in a large cross section. Therefore, an approximation, presumably based on approximate decomposition of the model into a randomly dispersed process over time, superimposed on another across the population, is required. Another article in this issue proposes a *convergent parameter model*, in which individual parameter vectors follow first-order Markov processes subordinated to a tendency to converge to the population norm. The MCF for this model, given in Rosenberg (1973c, Figure 1) in terms of parameter vectors of order  $Nk$ , can be accurately approximated by a simplified MCF of order  $k$ , thereby rendering the computations feasible.



### *F. Switching Parameter Models*

Quandt and Goldfeld (1973) have continued Quandt's earlier study (1958) of the model where a parameter vector switches stochastically between two values. The model is applicable, for instance, if supply and demand functions alternately dominate in a market. Any of the parameter processes introduced thus far could be formulated as a switching parameter model. However, the estimation methods generally considered for these processes are intended for parameter variation with a continuous distribution, and are less efficient in the switching parameter problem than the methods developed by Quandt and Goldfeld to exploit the binary distribution.

### III. THE IMPORTANCE OF RESPONSIVENESS TO STOCHASTIC PARAMETER VARIATION

In the general stochastic parameter model (2), the estimation problem can be viewed as consisting of three parts: (i) estimation of the unknown parameter vector  $\mathbf{m}$ , (ii) estimation of the stochastic parameter vector  $\boldsymbol{\eta}$ , and (iii) estimation of the stochastic specification,  $\sigma^2$  and  $\boldsymbol{\theta}$ . For instance, in the random parameter model, (i) corresponds to estimation of the population mean parameter vector, (ii) to estimation of the individual parameter vectors, and (iii) to estimation of the dispersion of the individual parameter vectors. In the sequential parameter model, (i) corresponds to estimating the initial value of the parameter vector (or, by a transformation, to estimation of the current value), (ii) corresponds to estimating the history of parameter realizations, and (iii) to estimating the Markov parameter process.

It is clear that (iii) can only be accomplished in the context of a random parameter model. Thus, when the process of parameter variation is of interest (as, for instance, if stochastic variation in corporate return on equity reflects competitive forces and the magnitude of these forces is to be determined), a random parameter model is essential.

With regard to estimating  $\mathbf{m}$ , it is clear from (5) that, so long as any nonzero mean in the stochastic parameters is adjusted for, an ordinary least squares (OLS) regression for  $\mathbf{m}$  will be defective only insofar as it ignores heteroscedasticity: thus, OLS will be unbiased but inefficient, and the advantage of a random parameter context consists in allowing the heteroscedasticity to be identified so that Aitken's Generalized Least Squares may be applied. This improvement in efficiency appears to be substantial in simulations (e.g., Cooley and Prescott (1973a), Rosenberg (1973c)). Equally important, in the presence of parameter variation, OLS sampling theory severely understates parameter estimation error variance. Thus, recognition of the correct specification removes a downward bias in estimated error variance, just as in the use of the familiar Cochrane-Orcutt transformation when disturbances are serially correlated.

Inappropriate use of a fixed-parameter model causes even more severe problems in estimating the individual parameter vectors, because stochastic parameter variation introduces random components in these vectors. Because the fixed-parameter model ignores these random components, the inefficiency and



invalidity of OLS are severe; in the simulations reported in Rosenberg (1973c), OLS error variance rises to five times the efficient variance, and OLS sampling theory underestimates OLS error variance by a factor of twenty or more. In the stochastic parameter model, minimum mean square estimation is achieved by attributing a proportion of the residuals  $y - \Xi(\theta)\hat{m}$  to the stochastic parameters: familiar econometric methods such as Theil's BLUS procedure and the estimation of the latest residual in the autoregressive disturbance model preparatory to forecasting, are special cases of this method.

If parameter variation is ignored in estimation, the estimate for  $m$  will tend to be an average of the realized individual parameter vectors. From this perspective, responsiveness to parameter variation is seen to be (i) valuable in achieving a more efficient estimator for this average, with more valid sampling theory; (ii) critically important in estimating the realized values of the individual parameters as distinct from the average, and (iii) essential in analyzing the stochastic parameter process.

As an illustration, consider the medical problem of estimating the physiological parameters describing an individual's response to a drug, with the purpose of recommending a correct dosage. These parameters will vary systematically across the patient population in relation to measurable patient characteristics such as weight, and will, in addition, vary stochastically about these systematic predictions in a manner that can be modeled, at a first approximation, by the random parameter model, assuming individual stochastic parameters to be fixed over time. Then one or more observations on the individual's response to the drug allow this random component to be estimated, thereby allowing dosages more appropriate to the individual to be prescribed. This approach has been applied with substantial success in a program under way at the University of California Medical Center, San Francisco (Sheiner, *et al.* (1972)). The method provides superior estimates of the population mean parameters and of the systematic parameter variation,<sup>3</sup> improved predictions of individual parameters, and medically useful descriptions of the extent of stochastic parameter variation.

#### IV. ESTIMATION OF THE STOCHASTIC SPECIFICATION

The estimation problem in a linear stochastic parameter regression breaks naturally into two stages: (i) estimation of the stochastic specification parameters,  $\theta$ ; and (ii) estimation of the scale parameter  $\sigma^2$ , of the unknown parameters  $m$ , and of the stochastic parameters  $\eta$ . The break occurs because the estimators for the latter parameters, conditional on  $\theta$ , can be expressed analytically, whereas most estimators for  $\theta$  must be computed by iterative procedures. Note that  $\sigma^2$  is treated as a scale parameter for all second moments (those of the parameter distribution as well as those of the disturbances), so that  $\theta$  specifies only the relative magnitudes of these moments. This allows the problem of estimating  $\theta$  to be reduced by one dimension and is therefore convenient. Estimation of  $\sigma^2$ , conditional on  $\theta$ , is always a straightforward matter, exactly analogous to the methods for an ordinary regression; moreover, this computation is always implicit in calculation of

<sup>3</sup> Inappropriate use of the fixed-parameter assumption in this case yields biased as well as inefficient estimators, because the regression models are nonlinear.

the sample likelihood or posterior distribution for  $\theta$ , so that it need not be treated separately. (It is a natural alternative to include  $\sigma^2$  in an expanded  $\theta$  vector, and the analysis could be carried out in this way if desired.)

In this section, the problem of estimating  $\theta$  will be discussed. Let  $R_\theta$  be the region of admissible values. Notice that there are two classes of criteria for optimal estimation of  $\theta$ : those which are specific to  $\theta$ , and those which refer to the properties of the estimators for  $m, \eta$  implied by the estimator of  $\theta$ .

When stochastic terms are normally distributed, two natural estimators to consider are the Maximum Likelihood estimator (MLE) and the Bayesian. In Rosenberg (1973b), the likelihood function and Bayesian posterior for  $\theta$  are derived for the general model (3). The formulae to compute these for each  $\theta$  also yield the MLE and Bayes posterior distributions for  $\sigma^2, m$ , and  $\eta$ , conditional on  $\theta$ . Maximum Likelihood estimation may be accomplished by a search over  $R_\theta$ , evaluating the likelihood function at each point. Bayes estimation is accomplished by numerical integration, with respect to the posterior distribution for  $\theta$ , again by an iteration over  $R_\theta$ . Computations at each step of the iteration are of the same order as required for estimation of  $m$ , conditional on  $\theta$ . Therefore, if the latter computation is feasible, as will be discussed in Section V, then the only computational problem arising is the need for repeated computation of the formulae for many  $\theta$  values in  $R_\theta$ . For an application of these methods to the sequential parameter problem, in which the MCF is exploited to simplify the computations for any  $\theta$ , see Rosenberg (1973c, Sections IIA, IIB). Cooley and Prescott have taken the same approach to the special case of the Adaptive Regression Problem (1973a, 1973b).<sup>4</sup> For the randomly dispersed parameter problem, see Fisk (1967, Sections, 5, 6), Swamy (1971, p. 111), Rosenberg (1973a).

The author's experience with these methods is quite favorable. In both the sequential and the random parameter models, the likelihood function has been well behaved, without extrema other than the global maximum, and convergence has been achieved without difficulty. The computations are quite feasible on third-generation computers; convergence in a nonlinear random parameter model with seven elements in  $\theta$  and roughly 500 observations is routinely accomplished in less than a minute on the CDC 7600, using a modified Fletcher-Powell algorithm. Moreover, the estimates for the stochastic specification parameters and for the regression parameters in several empirical applications have been consistent with a priori expectations, although the latter were not included in the estimation procedure in any way. This is perhaps the most robust test of any method.

The optimal properties of Bayesian methods, given that the prior distribution is appropriate, are well known. The asymptotic optimality of some Maximum Likelihood estimators is also well known, and Anderson (1970) extends these optimal properties to the case where the variance matrix  $V(\theta)$  is linear in  $\theta$ . He also notes that if a consistent initial estimator of  $\theta$  is available in this case, a single iteration of the Newton-Raphson procedure applied with these initial estimates

<sup>4</sup> However, in Cooley and Prescott (1973c) there appears to be a misunderstanding as to the existence of a general solution. CP appear to assert that their model is not a member of the family of models with sequential MCF as in Figure 1. As a matter of fact, the model is easily transformed to have the MCF in Figure 1, and the likelihood function for the model in this form is given in Rosenberg (1973c, Sect. II).

as a starting point will then provide asymptotically efficient estimators. Cooley (1971) and Cooley and Prescott (1972b) derive the asymptotic properties of the parameter shift variance estimator in the adaptive regression, which is a special case of Anderson's problem. Notice, however, that the transition matrices in the MCF enter nonlinearly into the matrix  $V(\theta)$ ; asymptotic efficiency when  $V$  is nonlinear in  $\theta$  has not been demonstrated, to the author's knowledge, although the demonstration should be relatively straightforward. See Kushner (1967) and Kashyap (1970) in this context. The asymptotic sampling properties of the Bayesian estimators are essentially equivalent to those of the Maximum Likelihood estimator (see, e.g., Johnson (1967) and Zellner (1971, p. 31)).

On the other hand, the small sample properties of Maximum Likelihood estimators are not necessarily optimal, nor are the Bayesian estimators' properties necessarily optimal, in a sampling theory sense, if the prior distribution is not believed. Efron and Morris (1972, 1973) approach this problem. For the randomly dispersed parameter model, with identical regressor matrices for all individuals, they are able to deduce the sampling properties of the regression parameter estimators implied by a class of estimators for the dispersion matrix  $\Omega(\theta)$ , and to deduce from this the implications of alternative prior distributions for elements of  $\theta$ . Although substantial generalizations will be needed before these results can be brought to bear on the complex regression problems usually encountered in econometrics, this approach is promising.

Within the class of estimators considered by EM are some estimators of  $\theta$  that are simple quadratic functions of the dependent variables, possibly truncated by setting to zero any negative estimated variances. These estimators fall within a general category of possibly truncated quadratic or iterative quadratic estimators of the variances in  $\theta$ , with the number of iterations, if there are any, being small. Within this general category, the simplest estimators are regressions of the form  $\hat{e} = \hat{X}\theta$ , where  $\hat{e}$  represents the squared values of residuals (or a transformation of these residuals) from (5), and  $\hat{X}$  represents the squares and cross products of the regressors in (5). Fisk (1967) pioneered this method and Hildreth and Houck (1968) developed it independently. Where only a single observation on each parameter vector arises in the random model with normally distributed stochastic terms, a single iteration of this method is asymptotically efficient, as Hildreth and Houck have shown, and as follows from Amemiya's general results on regression where the variance of the dependent variable is proportional to the square of its expectation, with the dependent variable following a gamma distribution (1973). In more complicated stochastic parameter regression models, however, a complicated heteroscedasticity in this "second moment regression" appears, and the method can no longer be made asymptotically efficient without arduous computations. Nevertheless, because of its computational simplicity, it may be proposed as a quick initial estimator for a subsequent Maximum Likelihood procedure, as in Rosenberg (1973c, Section IIc).

Analogous to these "second moment regression" estimators are a set of quadratic estimators for variances that have been conceived for various models, which have the virtue of unbiasedness but not necessarily any virtue of small-sample or even asymptotic optimality. Among these are Swamy's suggested estimator of  $\Omega$  for the randomly dispersed model with differing individual regres-

sor matrices of full rank (1970), and Anderson's suggested initial estimator for the general covariance matrix  $\mathbf{V}(\boldsymbol{\theta})$  that is linear in  $\boldsymbol{\theta}$  (1970, Eq. (2.16)). See also Arora (1973).

Rao (1970, 1971, 1972) has investigated quadratic estimators that are unbiased and optimal, either in the sense of minimum variance (MIVQUE) or in the weaker but computationally more accessible sense of minimum norm (MINQUE). For both classes, optimality is defined with respect to a norm depending on  $\boldsymbol{\theta}$ , so that construction of the optimal estimator for  $\boldsymbol{\theta}$  requires, in principle, that  $\boldsymbol{\theta}$  be known. If a guess at  $\boldsymbol{\theta}$  is used to define the estimator, the estimator will be unbiased but not efficient. If  $\hat{\boldsymbol{\theta}}$  yielded by the first iteration is used to define the optimal estimator for a second iteration, the estimator may become more efficient, but the property of unbiasedness is lost. Rao's approach is presumably best carried out as an iterative estimator. Rao (1972) has provided some general lemmas to aid in deriving the MINQUE estimators for a general linear model. MINQUE is very closely related to Anderson's iterative estimator.<sup>5</sup> Anderson's problem where  $\mathbf{V}(\boldsymbol{\theta})$  is linear in  $\boldsymbol{\theta}$  is equivalent to Rao's problem of variance and covariance components. Let  $\mathcal{L}(\boldsymbol{\theta})$  denote the classical likelihood function for this problem, and let  $\mathcal{L}^*(\boldsymbol{\theta})$  denote the Bayesian posterior distribution for  $\boldsymbol{\theta}$  conditional on the sample information, given diffuse prior densities for  $\sigma^2$  and  $\mathbf{m}$  (see Zellner (1971; Ch. 2)). Then  $\mathcal{L}^*(\boldsymbol{\theta}) \propto \mathcal{L}(\boldsymbol{\theta})|\mathbf{W}(\boldsymbol{\theta})|^{1/2}$ , where  $\mathbf{W}(\boldsymbol{\theta})$  is the variance matrix for the estimator of  $\mathbf{m}$ , given  $\boldsymbol{\theta}$ . It may be shown (Rosenberg (1973b)) that the MINQUE equations bear the same relationship to maximization of  $\mathcal{L}^*$  as Anderson's iterative procedure does to maximization of  $\mathcal{L}$ . Since  $\mathcal{L}$  and  $\mathcal{L}^*$  differ by a factor of order equal to the number of unknown parameters, MINQUE coincides with Anderson's procedure when no parameters are unknown, and for any vector  $\mathbf{m}$ , the two methods coincide asymptotically as the number of observations approaches infinity. The adjustment  $|\mathbf{W}(\boldsymbol{\theta})|^{1/2}$  serves to achieve unbiasedness in small-sample estimators of variance components.

Another desirable property of  $\mathcal{L}^*$ —and therefore of MINQUE—is that it is invariant with respect to a linear transformation of the parameters  $\mathbf{m}$  and  $\boldsymbol{\eta}$ , whereas  $\mathcal{L}$  is not. For example, MINQUE and Bayesian estimation are not affected, and Anderson's procedure and Maximum Likelihood estimation are affected, by Cooley and Prescott's decision to use the current parameter (which is a linear transformation of the initial parameter and of the stochastic parameter shifts in the sample history) as the unknown parameter in the adaptive regression model in place of the initial parameter.

Monte Carlo explorations of small-sample properties are necessary to compare these alternative estimators. A recent study by Froelich (1973), while a welcome step in this direction, is somewhat confusing in that it identifies "MINQUE" with one initial guess for  $\boldsymbol{\theta}$  (corresponding to homoscedastic disturbances), without noting that other MINQUE estimators exist. It is important to note that in all but the simplest stochastic parameter models, iterative procedures such as Anderson's and MINQUE are crucially dependent on the initial guess for  $\boldsymbol{\theta}$ . If the initial guess is good, one iteration may yield better small-sample properties

<sup>5</sup> I am indebted to Cheng Hsiao for several useful conjectures as to the general relationship between MINQUE and maximum likelihood estimation.

than continuation of the iterative procedure to the point of convergence, but if the initial guess is poor, the method may perform poorly.

There is a substantial applied physical sciences literature on "identification" of  $\theta$  in the sequential parameter model, e.g., Åström and Eykhoff (1971), Kashyap (1970), Mehra (1970, 1972), Sage and Melsa (1971). Much of the literature on the cross-section, time-series model with random time and individual effects cited in Section II.E, is also interesting in that it provides a detailed exploration of a special case.

All of the theory presented here is no more than a specialization of the theory for the general linear model, and the foundations for most of the results were originally derived on that more general level. An unfortunate limitation is that all the estimators are either optimal with respect to the normal distribution or else are quadratic functions of the stochastic terms, and in neither case are the methods robust against stochastic distributions with massive tails. This implies that, regardless of the method used, the experimenter would be wise to examine the outliers in the sample, to check the robustness of the estimated parameters against deletion of these outliers, and, quite possibly, to delete the outliers if the results are sensitive to their presence.

#### V. ESTIMATION OF THE UNKNOWN AND STOCHASTIC REGRESSION PARAMETERS

For the general linear stochastic parameter regression model, (3), with stochastic terms assumed normally distributed for Maximum Likelihood and Bayes purposes, Rosenberg (1968, 1973b) has derived Maximum Likelihood, Bayesian, and Minimum Mean Square Error Linear Unbiased Estimators (MMSLUE) conditional on  $\theta$ , and Duncan and Horn (1972) and Sarris (1973) have derived MMSLUE and Bayesian estimators, respectively, under the additional assumption that a proper prior distribution for the unknown vector  $\mathbf{m}$  exists. It turns out that, conditional on  $\theta$ , these three types of estimators (MLE, Bayesian, and MMSLUE) coincide, provided that careful attention is given to the concept of unbiasedness.

The problem with bias arises because there are several natural expectation operators to use in discussing the bias in an estimate of a stochastic parameter vector. Let  $\mathbf{b}_n = \mathbf{A}_n \mathbf{m} + \mathbf{B}_n \eta$ , where  $\mathbf{A}_n$  is nonsingular, as is usually the case. Then three possible expectation operators, with respect to which an estimator  $\hat{\mathbf{b}}_n$  might be defined as unbiased, are:

$$E_m(\cdot) \equiv E(\cdot | \mathbf{m})$$

$$E_n(\cdot) \equiv E(\cdot | \mathbf{b}_n) = E(E(\cdot | \mathbf{m}, \mathbf{b}_n) | \mathbf{b}_n)$$

$$E_{mn}(\cdot) \equiv E(\cdot | \mathbf{m}, \mathbf{b}_n).$$

The first "unconditional expectation," written as " $E_u$ " by Duncan and Horn, is the familiar expectation in a linear regression model of form (5). The second expectation is conditional on the realized value of the parameter vector  $\mathbf{b}_n$ , with the distribution for  $\mathbf{m}$  being taken as conditional on  $\mathbf{b}_n$ . The third expectation, conditioned on both vectors, is important in discussing the behavior of  $\hat{\mathbf{b}}_n$  when  $\mathbf{b}_n$  differs from  $\mathbf{m}$ . An estimator  $\hat{\mathbf{a}}_n$  will be called  $m - [n - , mn - ]$  unbiased if  $E_m[E_n, E_{mn}](\hat{\mathbf{a}}_n) = E_m[E_n, E_{mn}](\mathbf{a}_n)$ . For applications of these concepts, see Rosenberg (1972, 1973a).



It may be shown that the expectations  $E_m$  and  $E_n$  are essentially equivalent, in that any estimator that is  $m$ -unbiased is also  $n$ -unbiased. However, it turns out that the choice of expectation does affect the minimum variance linear unbiased estimator (MVLUE) for a stochastic parameter. Specifically, if a subscript denotes the expectation operator with respect to which the estimator is defined, then in estimating  $\mathbf{b}_n$ ,  $\text{MMSLUE}_n = \text{MMSLUE}_m = \text{MVLUE}_n$ , but these three estimators are not equal to  $\text{MVLUE}_m$ . In fact, it is easily seen that if  $\mathbf{A}_n = \mathbf{I}$ , then the  $\text{MVLUE}_m$  for  $\mathbf{b}_n$  is identical to the  $\text{MVLUE}_m$  for  $\mathbf{m}$ , regardless of the stochastic variation in  $\mathbf{b}_n$ . Since  $\text{MMSLUE}_n = \text{MMSLUE}_m$ , but  $\text{MVLUE}_n \neq \text{MVLUE}_m$ , it is preferable to speak of a minimum mean square error property for an estimator of a stochastic parameter rather than a minimum variance property.<sup>6</sup> It may now be asserted that the MLE and the Bayesian posterior means (with diffuse prior for  $\mathbf{m}$ ), for  $\mathbf{m}$  and  $\boldsymbol{\eta}$ , conditional on  $\boldsymbol{\theta}$ , coincide and are also MMSLUE when  $\boldsymbol{\theta}$  is known. Since these three estimators conditional on  $\boldsymbol{\theta}$  do coincide, they will be referred to hereafter as "the optimal estimators."

These estimators are *not*  $mn$ -unbiased. For instance, in the randomly dispersed parameter model, the population mean estimator  $\hat{\mathbf{m}}$  is slightly  $mn$ -biased toward  $\mathbf{b}_n$  and the individual parameter estimator  $\hat{\mathbf{b}}_n$  is  $mn$ -biased toward  $\mathbf{m}$ , if  $\mathbf{b}_n \neq \mathbf{m}$ .

Since the three optimal estimators coincide, it is a matter of indifference by which criterion the estimation formulae are derived for any special case, except that it is preferable to use an approach which also allows the likelihood function and posterior distribution for  $\boldsymbol{\theta}$  to be computed. The history of the derivations of the optimal estimators for the special models described in Section II will now be surveyed briefly.

For the randomly dispersed parameter model, Rao (1965) originally derived the optimal estimators for the population mean parameter vector by a holistic approach, and also the optimal estimators (predictors) for the individual parameter vectors, with the population mean assumed known. More recently, Efron and Morris (1972) and Rosenberg (1973a) derived the optimal estimators for the individual parameters with the population mean being unknown, the latter for the general case.

For the sequential parameter model, Kalman (1960) and Kalman and Bucy (1961) originally derived the optimal estimators for the current parameter vector (the Kalman-Bucy filter). The optimal estimators for the parameter vectors in the interior of the sample period (the smoothed estimators) and their variances were derived by Bryson and Frazier (1962), Lee (1964), Rauch (1963), Rauch, Tung, and Striebel (1965), and Meditch (1967). Other important early contributions were by Battin (1962), by Ho and Lee (1964) in expositing the Bayesian approach, and by Schweppe (1965) and Kushner (1964, 1967) in exhibiting the likelihood function. All of these works were atomistic in approach; the first holistic approach was by Fagin (1964). All of the above articles assumed that there was a proper prior

<sup>6</sup> Duncan and Horn avoid this difficulty in a rather confusing way by defining a minimum variance estimator as that estimator with minimum error variance, a violation of the tradition that the "variance" of an estimator is its own variance, not the variance of the estimation error. The theorems in Duncan and Horn hold only if "minimum variance" is defined in this unusual way; otherwise, the term "minimum mean square error" must be substituted.

distribution for the initial parameter vector, so that no unknown parameter vector was present. This "starting problem" or inability to deal with an unknown parameter vector was quite troublesome. As late as 1967, Aoki, in an otherwise excellent survey, proposed a solution to the starting problem that was erroneous, because it was based on false generalized matrix inversion "identities" (1967, p. 80). Duncan and Horn (1972) and Sarris (1973) in their recent treatments of the problem continue to assume a proper prior distribution for  $\mathbf{m}$ , independent of the data, in discussing estimation of the stochastic parameters. The first solutions to the starting problem were found independently by Fraser (1967) and Rosenberg (1967), and what is apparently the first general solution to this problem which permits computation of the sample likelihood is published in this issue (Rosenberg (1973c, Section IIB)). The formulae for the covariances between estimation errors in different periods, together with an analysis of the relationship between the atomistic and holistic approaches, appear in Rosenberg (1968a, 1968b).

For the stationary stochastic parameter model, Burnett and Guthrie (1970) derived the optimal estimators conditional on the mean parameter  $\mathbf{m}$  being known, and Rosenberg (1972) generalized these to the case where  $\mathbf{m}$  was unknown. With regard to cross-section time-series models and switching regressions, the reader is referred to the articles already cited in these contexts.

#### VI. CONCLUSION

On the whole, the theory of stochastic parameter regression is one of the most exciting areas of statistical investigation. Moreover, the theory seems to have already reached the point where it promises fruitful applications. The most productive applications in econometrics are likely to come in cross-section time-series analysis, where the wealth of data offers a real opportunity to identify the pattern of parameter variation; stochastic parameter methods are especially needed when estimation of the individual parameters, as distinct from the population mean alone, is of great importance, since fixed-parameter methods are relatively less efficient in estimating these parameters. Potential applications also arise whenever it is desired to analyze the process of parameter variation itself, an aspect of economic events that has been studied all too little. To cite just one example, of what strength are the competitive forces that cause corporate rates of return on equity to converge toward the norm for the economy, and of what magnitude are the stochastic shocks that allow corporations to achieve above-average returns. A third promising area of application is in short-term forecasting (Rosenberg (1968, Ch. 8), Cooley (1971)), where adaptation to sequential parameter shifts is essential.

*University of California, Berkeley*

#### BIBLIOGRAPHY

- Amemiya, T. (1971), "The Estimation of the Variances in a Variance-Components Model," *International Economic Review*, 12 (February), 1-13.
- . (1973), "Regression When the Variance of the Dependent Variable Is Proportional to the Square of Its Expectation," *Journal of the American Statistical Association* (December) (forthcoming).
- Anderson, T. W. (1969), "Statistical Inference for Covariance Matrices with Linear Structure," *Proceedings of the Second International Symposium on Multivariate Analysis*. Edited by P. R. Krishnaiah. New York: Academic Press.



- . (1970), "Estimation of Covariance Matrices Which Are Linear Combinations or Whose Inverses Are Linear Combinations of Given Matrices," *Essays in Probability and Statistics*. Edited by R. C. Bose, et al. Chapel Hill, N.C.: University of North Carolina Press.
- Aoki, Masano (1967), *Optimization of Stochastic Systems*. New York: Academic Press.
- Arora, Swarnjit S. (1973), "Error Components Regression Models and Their Applications," *Annals of Economic and Social Measurement*, 2 (October).
- Åström, K. J. and Eykhoff, P. (1971), "System Identification—A Survey," *Automatica*, 7 (March), 123–163.
- Battin, Richard H. (1962), "A Statistical Optimizing Navigation Procedure for Space Flight," *Journal of American Rocket Society*, 32 (November), 1681–1696.
- Belsley, David (1973), "On the Determination of Systematic Parameter Variation in the Linear Regression Model," *Annals of Economic and Social Measurement*, 2 (October).
- Bowman, H. W. and LaPorte, A. M. (1972), "Stochastic Optimization in Recursive Equation Systems with Random Parameters," *Annals of Economic and Social Measurement*, 1 (January), 419–436.
- Box, G. E. P. and Jenkins, G. M. (1962), "Some Statistical Aspects of Adaptive Optimization and Control," *Journal of the Royal Statistical Society, Series B*, Vol. 24, 297–343.
- . (1970), *Time Series Analysis, Forecasting, and Control*. San Francisco: Holden-Day, Inc.
- Brown, Robert Goodell (1963), *Smoothing, Forecasting, and Prediction of Discrete Time Series*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Bryson, A. E. and Frazier, M. (1963), "Smoothing for Linear and Nonlinear Dynamic Systems," *Proceedings of the Optimum Systems Synthesis Conference, September 1962*. Dayton, Ohio: Wright-Patterson Air Force Base, Aeronautical Systems Division, TOR-63-119, 353–364.
- Burnett, J. D. and Guthrie, D. (1970), "Estimation of Stationary Stochastic Regression Parameters," *Journal of the American Statistical Association*, 65 (December), 1547–1553.
- Chetty, V. K. (1968), "Pooling of Time Series and Cross Section Data," *Econometrica*, 36 (April), 279–290.
- Chew, V. (1970), "Covariance Matrix Estimation in Linear Structure," *Journal of the American Statistical Association*, 65 (March), 173–181.
- Cooley, T. F. (1971), "Estimation in the Presence of Sequential Parameter Variation," Unpublished Ph.D. dissertation, Department of Economics, University of Pennsylvania.
- Cooley, T. F. and Prescott, E. C. (1973a), "Tests of an Adaptive Regression Model," *Review of Economics and Statistics*, 55 (May), 248–256.
- . (1973b), "The Adaptive Regression Model," *International Economic Review*, 14 (June).
- . (1973c), "Varying Parameter Regression: A Theory and Some Applications," *Annals of Economic and Social Measurement*, 2 (October).
- Cox, Henry (1964), "On the Estimation of State Variables and Parameters for Noisy Dynamic Systems," *IEEE Transactions on Automatic Control*, AC-9 (January), 5–12.
- Duncan, D. B. and Horn, S. D. (1972), "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis," *Journal of the American Statistical Association*, 67 (December), 815–821.
- Efron, Bradley and Morris, C. (1972), "Empirical Bayes on Vector Observations: An Extension of Stein's Method," *Biometrika*, 59 (August).
- . (1973), "Stein's Estimation Rule and Its Competitors—An Empirical Bayes Approach," *Journal of the American Statistical Association*, 68 (March), 117–130.
- Fagin, Samuel L. (1964), "Recursive Linear Regression Theory, Optimal Filter Theory, and Error Analysis of Optimal Systems," *IEEE International Convention Record*, Part 1, 216–240.
- Fisk, P. R. (1967), "Models of the Second Kind in Regression Analysis," *Journal of the Royal Statistical Society, Series B*, 29, 266–281.
- Fraser, Donald (1967), "A New Method for the Optimal Smoothing of Data," Unpublished Sc.D. thesis, Department of Aeronautics, Massachusetts Institute of Technology.
- Froelich, B. R. (1973), "Some Estimators for a Random Coefficient Regression Model," *Journal of the American Statistical Association*, 68 (June), 329–335.
- Hartley, H. O. and Rao, J. N. K. (1967), "Maximum Likelihood Estimation for the Mixed Analysis of Variance Model," *Biometrika*, 54, Nos. 1 and 2, 89–108.
- Henderson, C. R., Jr. (1971), "Comment on the Use of Error Components Model in Combining Cross Section with Time Series Data," *Econometrica*, 39 (March), 397–401.
- Hildreth, C. and Houck, J. P. (1968), "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association*, 63, 584–595.
- Ho, Y. C. (1963), "On the Stochastic Approximation Method and Optimal Filtering Theory," *Journal of Math Analysis and Applications*, 6 (February), 152–155.
- 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 (October), 333–339.

- Hsiao, C. (1972a), "The Combined Use of Cross-Section and Time-Series Data in Econometric Analysis," Unpublished Ph.D. dissertation, Stanford University.
- . (1972b), "Some Estimation Methods for a Random Coefficients Model," Technical Report No. 77. Stanford, Calif.: Institute for Mathematical Studies in the Social Sciences, Stanford University.
- . (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.
- Johnson, R. A. (1967), "An Asymptotic Expansion for Posterior Distributions," *Annals of Mathematical Statistics*, 38, 1899-1906.
- Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," Transactions of ASME, Series D, *Journal of Basic Engineering*, 82, 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, 95-108.
- Kashyap, R. L. (1970), "Maximum Likelihood Identification of Stochastic Linear Systems," *IEEE Transactions on Automatic Control*, AC-15 (February), 25-34.
- Kushner, Harold J. (1964), "On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes, With Applications," *SIAM Journal on Control*, 2 (January), 106-119.
- . (1967), *Stochastic Stability and Control*, New York: Academic Press.
- . (1971), *Introduction to Stochastic Control*, New York: Holt, Rinehart & Winston.
- Lee, Robert C. K. (1964), *Optimal Estimation, Identification, and Control*, Cambridge, Mass.: MIT Press.
- Maddala, G. S. (1971), "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39 (March), 341-358.
- Meditch, James S. (1967), "Orthogonal Projection and Discrete Optimal Linear Smoothing," *SIAM Journal on Control*, 5 (February), 74-89.
- Mehra, R. K. (1970), "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Transactions on Automatic Control*, AC-15 (April), 175-184.
- . (1972), "Approaches to Adaptive Filtering," *IEEE Transactions on Automatic Control*, AC-17 (October), 693-698.
- Muth, J. F. (1960), "Optimal Properties of Exponentially Weighted Forecasts," *Journal of the American Statistical Association*, 55 (June), 299-306.
- Parks, Richard W. (1967), "Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated," *Journal of the American Statistical Association*, 62 (June), 500-509.
- Parzen, E. (1961), "An Approach to Time Series Analysis," *Annals of Mathematical Statistics*, 32 (December), 951-989.
- Quandt, R. E. (1958), "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53, 873-880.
- Quandt, Richard and Goldfeld, Steven (1973), "The Estimation of Structural Shifts by Switching Regressions," *Annals of Economic and Social Measurement*, 2 (October).
- Rao, C. R. (1965a), *Linear Statistical Inference and Its Applications*, New York: John Wiley and Sons.
- . (1965b), "The Theory of Least Squares When the Parameters Are Stochastic," *Biometrika*, 52, 447-458.
- . (1970), "Estimation of Heteroscedastic Variances in Linear Models," *Journal of the American Statistical Association*, 65 (March), 161-172.
- . (1971), "Minimum Variance Quadratic Unbiased Estimation of Variance Components," *Journal of Multivariate Analysis*, 1, 257-275.
- . (1972), "Estimation of Variance and Covariance Components in Linear Models," *Journal of the American Statistical Association*, 67 (March), 112-115.
- Rauch, H. E. (1963), "Solutions to the Linear Smoothing Problem," *IEEE Transactions on Automatic Control*, AC-8 (October), 371-372.
- Rauch, H. E., Tung, F., and Striebel, C. T. (1965), "Maximum Likelihood Estimates of Linear Dynamic Systems," *Journal of the American Institute of Aeronautics and Astronautics*, 3 (August), 1445-1450.
- Rosenberg, Barr (1967), "Regression in the Presence of Stochastically Varying Parameters," Paper presented at the Annual Meetings of the Econometric Society (December). Mimeographed.
- . (1968a), "Varying-Parameter Estimation," Unpublished Ph.D. dissertation, Department of Economics, Harvard University.
- . (1968b), "An Atomistic Approach to Linear Regression," Working Paper No. IP-137. Berkeley: Institute of Business and Economic Research, University of California.

- . (1972), "The Estimation of Stationary Stochastic Regression Parameters Re-examined," *Journal of the American Statistical Association*, 67 (September), 650-654.
- . (1973a), "Linear Regression With Randomly Dispersed Parameters," *Biometrika*, 60, 65-72.
- . (1973b), "Estimation in the General Linear Stochastic Parameter Model," Research Report. Cambridge, Mass. National Bureau of Economic Research, Computer Research Center.
- . (1973c), "The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, 2 (October).
- Rosenberg, Barr and McKibben, W. (1973), "The Prediction of Systematic and Specific Risk in Common Stocks," *Journal of Financial and Quantitative Analysis*, 8, 317-333.
- Rubin, H. (1950), "Note on Random Coefficients," *Statistical Inference in Dynamic Economic Models*. Edited by C. T. Koopmans. New York: John Wiley and Sons.
- Sage, A. P. and Melsa, J. L. (1971), *Estimation Theory With Applications to Communications and Control*, New York: McGraw-Hill Book Company, Inc.
- Sarris, Alexander (1973), "A Bayes Approach to Estimation of Time Varying Regression Coefficients," *Annals of Economic and Social Measurement*, 2 (October).
- Schweppe, Fred C. (1965), "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE Transactions on Information Theory*, 11 (January), 61-70.
- Searle, S. R. (1971), "Topics in Variance Components Estimation," *Biometrics*, 27 (March), 1-76.
- Sheiner, L. B., Rosenberg, Barr, and Melmon, K. L. (1972), "Modelling of Individual Pharmacokinetics for Computer-Aided Drug Dosage," *Computers in Biomedical Research*, 5, 441-459.
- Swamy, P. A. V. B. (1970), "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38, 311-323.
- . (1971), *Statistical Inference in Random Coefficient Regression Models*, Berlin-Heidelberg-New York: Springer-Verlag.
- . (1973), "Criteria, Constraints, and Multicollinearity in Random Coefficient Regression Models," *Annals of Economic and Social Measurement*, 2 (October).
- Swamy, P. A. V. B. and Arora, S. S. (1972), "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," *Econometrica*, 40 (March), 261-276.
- Swering, Peter (1959), "First Order Error Propagation in a Stagewise Smoothing Procedure for Satellite Observations," *Journal of the Astronautical Sciences*, 6 (Autumn), 46-52.
- Teräsvirta, Timo (1970), "On Stepwise Regression and Economic Forecasting," Finnish Economic Association, *Economic Studies XXXI*, Helsinki.
- Theil, Henri (1971), *Principles of Econometrics*, New York: John Wiley and Sons, Inc.
- Theil, Henri and Mennes, L. B. (1959), "Conception Stochastique de Coefficients Multiplicateurs dans l'Adjustment Lineaire des Series Temporelles," Publications de l'Institut de Statistique de l'Universite de Paris, 8, 211-227.
- Tiao, G. C. and Tan, W. Y. (1965), "Bayesian Analysis of Random-Effects Models in the Analysis of Variance, I. Posterior Distribution of Variance-Components," *Biometrika*, 52, 37-53.
- . (1966), "Bayesian Analysis of Random-Effects Models in the Analysis of Variance, II. Effect of Autocorrelated Errors," *Biometrika*, 53, 477-495.
- Wallace, T. D. and Hussain, Ashiq (1969), "The Use of Error Components Models in Combining Cross Section with Time Series Data," *Econometrica*, 3 (January), 55-72.
- Zellner, Arnold (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley & Sons, Inc.