



Estimation in the Presence of Stochastic Parameter Variation

Thomas F. Cooley; Edward C. Prescott

Econometrica, Vol. 44, No. 1. (Jan., 1976), pp. 167-184.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28197601%2944%3A1%3C167%3AEITPOS%3E2.0.CO%3B2-S>

Econometrica is currently published by The Econometric Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/econosoc.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

ESTIMATION IN THE PRESENCE OF STOCHASTIC PARAMETER VARIATION¹

BY THOMAS F. COOLEY AND EDWARD C. PRESCOTT¹

In this paper we consider the estimation of a model with time varying structure. The parameters of the model are assumed to be subject to permanent and transitory changes over time. Estimation methods are developed, and the asymptotic properties of the estimates are derived.

1. INTRODUCTION

ECONOMETRICS IS TYPICALLY CONCERNED with the estimation of relationships that are characterized by:

$$(1.1) \quad y_t = F(y_{t-1}, x_t, \theta, \varepsilon_t).$$

The form of the function F and the value of the parameter θ are determined from the behavior of the individual agents in the economy. The assumption that underlies most estimation theory is that F and θ will be stable over time. In this paper, we argue that the assumption of a stable F and θ is frequently untenable and that it would often be more reasonable to assume in estimation that relationships vary over time.

The problem of structural instability in econometric relationships has been recognized by econometricians (see, for example, Dusenberry and Klein [8]), but the underlying reasons for the instability have not been well explored. One, perhaps obvious, source of instability in econometric relationships is misspecification. If the relationship has been misspecified it implies that the true form (the true F and θ of (1.1)) either has not been discovered or cannot be estimated. Cooley [4] and Rosenberg [20] explore a number of ways in which misspecification may lead to parameter variation over time.

One need not appeal to ignorance, however, to justify the assumption that econometric relationships vary over time. In many instances economic theory suggests that relationships will change over time. Lucas [14], for example, has shown that econometric models, as they are now structured, are inappropriate tools for long-term policy evaluation precisely because they assume a stable structure. The structure of an econometric model represents the optimal decision rules of economic agents. From dynamic economic theory we know that optimal decision rules vary systematically with changes in the structure of series relevant to the decision makers. It follows that changes in policy will systematically alter the structure of the series being forecasted by decision makers, and, therefore, the behavioral relationships as well. It also follows that other exogenous events, such as changes in technology, will alter the structure of individual decision rules.

For all but the simplest kinds of exogenous changes, the derivation of the changes in the optimal decision rules is analytically intractable. Thus, while theory suggests that it would be appropriate to view behavioral relationships as varying over time, it does not, in general, suggest how to capture the variation structurally.

¹ The authors would like to acknowledge the helpful comments of an anonymous referee.

We do know, however, that most policy or other exogenous changes will cause changes in the decision rules that are permanent and thus permanent changes in the behavioral relationships as well [14, pp. 15-17].

Some attempts have been made to deal with problems of parameter variation as is evidenced by the work on the random coefficient model [12 and 22] and the problem of testing for structural change [13 and 17]. Until the work of Rosenberg [20], however, little attention was given to the fact that the parameters in the econometric relationships are likely to vary systematically over time. In this paper we develop an approach to varying parameter regression that is an extension of the work in Cooley and Prescott [5, 6]. In the next section, we develop the general model of parameter variation to be considered. In developing a set of assumptions to be used in estimation, we focus on a particular type of parameter process although the estimation technique can be extended to others. In Section 3, we present a transformation which makes estimation no more capital intensive than many commonly used nonlinear estimation techniques. Section 4 develops the asymptotic properties of the estimates and presents their limiting distribution. The final section presents the conclusions.

2. A THEORY OF VARYING PARAMETER REGRESSION

The regression structure with which we shall be concerned has the following form:

$$(2.1) \quad y_t = x_t' \beta_t \quad (t = 1, 2, \dots, T),$$

where x_t is a k component vector of explanatory variables, β_t is a k component vector of parameters subject to stochastic variation, and y_t is the t th observation of the dependent variable. If there is an intercept, then

$$(2.2) \quad x_{t1} = 1 \quad (t = 1, 2, \dots, T),$$

and β_1 represents the intercept. Since parameter changes are likely to come from a variety of sources, it is reasonable to assume that some of them may persist while others may not. Thus, we assume the parameters to be adaptive in nature, subject to permanent and transitory changes. The hypothesized pattern of variation is:

$$(2.3) \quad \begin{aligned} \beta_t &= \beta_t^p + u_t \\ \beta_t^p &= \beta_{t-1}^p + v_t, \end{aligned} \quad (t = 1, \dots, T),$$

where the superscript p denotes the permanent component of the parameters.

The u_t and v_t are identically and independently distributed normal variates with mean vectors 0 and covariance structures known up to different scale factors. A particularly convenient parameterization of this is as follows:

$$(2.4) \quad \text{cov}(u_t) = (1 - \gamma)\sigma^2 \Sigma_u \quad \text{and} \quad \text{cov}(v_t) = \gamma\sigma^2 \Sigma_v$$

where Σ_u and Σ_v are known up to scale factors. This assumption implies one of the

elements of both Σ_u and Σ_v can be normalized to 1. When an intercept is present and is subject to both permanent and transitory changes, setting $\sigma_{11}^u = \sigma_{11}^v = 1$ is a convenient normalization. The transitory change in the intercept then corresponds to the additive disturbance term in the conventional regression model. Subsequently, for expository purposes, we assume β_{1t} is the intercept and that the above normalization has been made. The unknown parameters are the β_t , and the unchanging elements σ^2 and γ which specify the covariance structure. The objective of the estimation techniques is to estimate σ^2 and γ and the permanent components of the β_t .

This particular specification of permanent and transitory change appeals to us as useful for many econometric applications. As we noted in the introduction, most exogenous changes that affect microeconomic decision rules are likely to cause permanent changes in behavioral relationships. If knowledge of exogenous events is imperfect, theory suggests that it is appropriate to view behavioral relationships as changing slowly over time. Many of the kinds of misspecification that are analyzed in [3 and 20] imply changes in parameters that are permanent over time. The structure we are assuming in the subsequent analyses implies that the permanent components of the parameters will drift systematically over time away from their initial value with no inherent tendency to return to a mean value. The methodology developed below, however, can be easily adapted to different prior specifications of the parameter process if it is felt in a particular application that the parameters might adapt differently.

The process generating the parameters is non-stationary, and it is impossible to specify the likelihood function. For the purpose of estimation, however, we are interested in specific realizations of the parameter process. The likelihood function conditional on the value of the parameter process at some point in time is well defined so we can treat specific realizations of the parameter process as random parameters to be estimated. The most convenient procedure for forecasting is to focus on the value of the parameter process one period past the sample. In this case, it follows that:

$$(2.5) \quad \beta_{T+1}^p = \beta_t^p + v_T = \beta_t^p + \sum_{s=t+1}^{T+1} v_s,$$

$$(2.6) \quad \beta_t = \beta_{T+1}^p - \sum_{s=t+1}^{T+1} v_s + u_t,$$

and (2.1) can be rewritten as:

$$(2.7) \quad y_t = x_t' \beta + \mu_t,$$

where

$$(2.8) \quad \beta = \beta_{T+1}^p$$

and

$$(2.9) \quad \mu_t = x_t' u_t - x_t' \sum_{s=t+1}^{T+1} v_s.$$

It is easily verified that μ is distributed normally with mean zero and covariance matrix:

$$(2.10) \quad \text{cov}(\mu) = \sigma^2[(1 - \gamma)R + \gamma Q] \equiv \sigma^2\Omega(\gamma)$$

where R is a diagonal matrix with

$$(2.11) \quad r_{ii} = (x_i' \Sigma_u x_i),$$

and Q is a matrix such that

$$(2.12) \quad q_{ij} = \min(T - i + 1, T - j + 1)x_i' \Sigma_v x_j.$$

If one is concerned with the value of the permanent part of the parameter vector in period t , that is β_t^p , the appropriate formulae for the q_{ij} are

$$(2.13) \quad q_{ij} = \min\{|t - i|, |t - j|\}x_i' \Sigma_v x_j,$$

if both i and j exceed or are less than t . Otherwise, $q_{ij} = 0$. This generalization is useful in situations where one is not forecasting future values of the dependent variable y_t but rather attempting to draw inference about the path of the coefficients. This is of interest because economic theory sometimes suggests movements in the coefficients and such information is needed to test the validity of the theory. Alternatively, systematic drifts in the coefficients may suggest that the model is subject to specification errors of a particular kind and the information contained in the parameter changes may be useful in modifying the theory.

The full model can be rewritten as:

$$(2.14) \quad Y = X\beta + \mu,$$

where β is the k component vector

$$(2.15) \quad \beta = \begin{bmatrix} \beta_{1,T+1}^p \\ \beta_{2,T+1}^p \\ \vdots \\ \beta_{k,T+1}^p \end{bmatrix};$$

X is the $T \times k$ matrix:

$$(2.16) \quad \begin{bmatrix} x_{11} \dots x_{k1} \\ x_{12} \dots x_{k2} \\ \vdots \\ x_{1T} \dots x_{kT} \end{bmatrix};$$

and Y is the T component vector of the y_t . From (2.10) it follows that Y is distributed as:

$$(2.17) \quad Y \sim N[X\beta, \sigma^2\Omega(\gamma)].$$

If γ were known, then the estimation would be a trivial application of Aitkens generalized least squares (GLS) analysis because R and Q are functions of the observed exogenous variables. The parameter γ , however, plays a crucial role in the analysis and is unlikely to be known in most econometric applications. The parameter γ tells us how fast the β 's are adapting to structural change. If γ is large (close to 1), then the permanent changes are large relative to the transitory changes.

Using (2.10), we can write the log likelihood function of the observations as:

$$(2.18) \quad L(Y; \beta, \sigma^2, \gamma, X) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \ln |\Omega(\gamma)| \\ - \frac{1}{2\sigma^2} (\gamma - X\beta)' \Omega(\gamma)^{-1} (\gamma - X\beta).$$

We can maximize (2.18) partially with respect to β and σ^2 to obtain the estimators conditional on γ :

$$(2.19) \quad B(\gamma) = [X' \Omega(\gamma)^{-1} X]^{-1} X' \Omega(\gamma)^{-1} Y,$$

$$(2.20) \quad s^2(\gamma) = \frac{1}{T} [(Y - XB(\gamma))' \Omega(\gamma)^{-1} (Y - XB(\gamma))].$$

These are substituted in (2.18) to determine the concentrated likelihood function as:

$$(2.21) \quad L_c(Y; \gamma) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln s^2(\gamma) - \frac{1}{2} \ln |\Omega(\gamma)| - \frac{T}{2} \\ = -\frac{T}{2} (\ln 2\pi + 1) - \frac{T}{2} \ln s^2(\gamma) - \frac{1}{2} \ln |\Omega(\gamma)|.$$

Thus, globally maximizing the log likelihood function (2.18) is equivalent to maximizing this concentrated likelihood function. Note that γ , because it is the fraction of parameter variation due to permanent changes, is restricted to fall within the range

$$(2.22) \quad 0 \leq \gamma \leq 1.$$

The strategy of estimation then is to divide the range for γ into a number of points

$$(2.23) \quad \{\gamma_i; i = 1, 2, \dots, n\}.$$

For every γ_i evaluate (2.21) and choose as the estimator of γ , say g , the value such that:

$$(2.24) \quad L_c(Y; g, X) \geq L_c(Y; \gamma_i, X), \quad \text{all } i.$$

The estimates of β and σ^2 are determined from (2.19) and (2.20) above as $B(g)$ and $s^2(g)$ respectively.²

² Usually $T/(T-k)s^2(g)$ would be a better estimate of σ^2 than the maximum likelihood estimator because it is unbiased if $g = \gamma$.

To apply the estimation technique Σ_u and Σ_v , which along with γ and σ^2 specify the covariances of the permanent and transitory changes, must be known up to scale factors. In many applications, it is likely that only one parameter will be subject to variation over time.³ In such cases, the appropriate elements of Σ_u and Σ_v should be assumed equal to unity and the estimates of σ^2 and γ will completely specify the variance component of interest. In other cases, such as speed of adjustment and adaptive expectations models, the relationship between parameters is known and thus the appropriate elements of Σ_u and Σ_v can be specified a priori.⁴ Thus, there is a wide variety of well studied economic problems in which the estimation technique proposed here would be appropriate.

In certain applications, other techniques would be more appropriate. For example, if $\Sigma_u = \Sigma_v$ and $\gamma = 0$, then the model reduces to the random coefficient model and the MINQUE procedure proposed by Rao [19] would be more appropriate. In principle, it is also possible to estimate the elements Σ_u and Σ_v by direct maximum likelihood techniques (see [1]). Even with well behaved data, however, the number of observations required to achieve successful identification of several variance components is very large. Given that most economic time series exhibit substantial collinearity, the identification of variance components from the data would be even more difficult and the number of observations necessary correspondingly larger. Consequently, direct maximum likelihood techniques are unlikely to be feasible in any applications.⁵ The estimation technique proposed here, however, is well suited to the data commonly available in econometric research, and the transformation presented in the next section greatly simplifies the estimation.⁶

3. TRANSFORMATION OF THE MODEL

A possible drawback of the estimation scheme presented in the previous section is that it requires the inversion of the $T \times T$ matrix $(1 - \gamma)R + \gamma Q$ for each value of γ_i . In this section, we present a transformation which greatly reduces the number of computations required to obtain the estimators. The strategy of the transformation is to make the matrices R and Q diagonal so that inversion of the covariance matrix is a trivial computation. The elements of R and Q are known since they depend on the exogenous variables and the matrices Σ_u and Σ_v . To eliminate the matrix R , the model can be transformed as follows:

$$(3.1) \quad \begin{aligned} y_t^* &= y_t \sqrt{r_{tt}} & (t = 1, 2, \dots, T), \\ x_{it}^* &= x_{it} \sqrt{r_{tt}} & (i = 1, 2, \dots, k), \end{aligned}$$

³ This is the case, for example, in the aggregate consumption function discussed by Lucas [14], and in the efficient markets model discussed by Fama [9].

⁴ Dynamic adjustment models of the type introduced by Nerlove [16] are an example with which we have experimented.

⁵ An additional problem is that the properties of the estimates have not been developed when many components must be identified.

⁶ Extensive tests of the robustness of these estimations are reported in [4] and they show surprisingly small losses in estimation efficiency even for sizeable errors in a priori specification of the covariance structure.

where r_{tt} is the t th diagonal element of the matrix R . This yields a transformed model where Y^* is distributed as

$$(3.2) \quad Y^* \sim N\{X^*\beta, \sigma^2[(1 - \gamma)I + \gamma Q^*]\}$$

where

$$(3.3) \quad q_{ij}^* = q_{ij}/\sqrt{r_{ii}r_{jj}}.$$

Now, there exists an orthogonal matrix P whose columns are a set of orthonormal eigenvectors of the matrix Q^* so that

$$(3.4) \quad P'P = I,$$

and (see [11, p. 255])

$$(3.5) \quad P'Q^*P = D.$$

D is a diagonal matrix whose elements are the eigenvalues of Q^* . Now let

$$(3.6) \quad \bar{Y} = P'Y^*, \quad \bar{X} = P'X^*, \quad \bar{\mu} = P'\mu^*.$$

Observe that \bar{Y} is now distributed as:

$$(3.7) \quad \begin{aligned} \bar{Y} &\sim N[\bar{X}\beta, \sigma^2(P'P + \gamma P'Q^*P - \gamma P'P)] \\ &\sim N[\bar{X}\beta, \sigma^2[I + \gamma(D - I)]]. \end{aligned}$$

The matrix Q^* is known so that its eigenvalues need only be computed once. After this is done, estimation is trivial for each γ_i that is searched.

The computation of the eigenvalues and eigenvectors of Q^* is a well-studied problem. It is clear that every root of the characteristic equation must be obtained in order to have the matrix D completely specified. We found the use of Householder's tri-diagonalization followed by the QR method [18] quite accurate and fast. This calculation need only be done once and the transformation reduces the total number of computations significantly making these estimators less capital intensive than many commonly used non linear estimation techniques.⁷

4. ASYMPTOTIC PROPERTIES

Thus far, we have developed a class of models in which the parameters are subject to stochastic variation over time. The nature of the process, however, prohibits any simple application of the usual asymptotic results because no consistent estimator exists for the parameter set $(\beta, \gamma, \sigma^2)$. The variance of the β 's are bounded away from zero because these parameters are subject to random changes in every period.⁸ In this section we prove that g is a consistent estimator of γ which implies that the maximum likelihood estimator $B(g)$ is asymptotically efficient and yields asymptotically optimal predictions.

⁷ If the number of observations available is very large, Kalman filtering techniques such as those proposed in [20] might be more appropriate.

⁸ An additional problem is that generally verifiable conditions for consistency of maximum likelihood estimators for dependent processes have not been developed as has been done by Wald [24] for independent observations. Silvey [21] has developed conditions which are almost impossible to verify.

Following standard analysis (see [7, pp. 124–129]) it is easily seen that $B(\gamma)$ satisfies the conventional definition of efficiency since it is unbiased and its covariance satisfies the Cramer-Rao minimum variance bound. Thus, if γ is known, $B(\gamma)$ has minimum variance in the class of unbiased estimators. It can also be established that, conditional on future values of the explanatory variables X_τ , the forecaster $X_\tau B(\gamma)$ of future Y_τ has minimum quadratic loss in the class of unbiased forecasters. It is now shown that the maximum likelihood estimator of γ is consistent so that all of these optimality results hold asymptotically.

The maximum likelihood estimators of σ^2 and β conditional on γ for a given sample of size T will be denoted by $s_T^2(\gamma)$ and $B_T(\gamma)$ respectively and the true values of (γ, σ^2) by (γ_0, σ_0^2) . In the subsequent discussion, we are working with the transformed model with diagonalized covariance matrix $D(\gamma)$. The t th diagonal element of $D(\gamma)$ is

$$(4.1) \quad d_t(\gamma) = (1 - \gamma) + \gamma d_t$$

where d_t is the t th eigenvalue of the matrix Q^* defined in equation (3.3). In the proofs of the following lemmas we shall make use of the result that is established in the Appendix that the d_t exceed δ^* where $0 < \delta^* \leq 1$. The bars and subscript will be dropped from functions and variables for convenience, and all summations are implicitly from 1 to T . We also assume that x_{it} are bounded by $|x_{it}| \leq \bar{x}$.

Letting $S(\gamma)$ be the generalized sum of squared residuals conditional on γ , the concentrated log likelihood function, (2.21), divided by $T/2$ is, save for a constant,

$$(4.2) \quad L(\gamma; T) = -\ln(S(\gamma)/T) - T^{-1} \ln |D(\gamma)|.$$

The function $f(\gamma; T)$ is defined to be

$$(4.3) \quad f(\gamma; T) = -\ln \left[\sigma_0^2 \sum \frac{d_t(\gamma_0)}{d_t(\gamma)} \right] - T^{-1} \ln |D(\gamma)|.$$

Before presenting the consistency proof it is perhaps worthwhile to provide the following guide to the reader: The functions $L(\gamma; T) - f(\gamma; T)$ are shown to converge pointwise in probability to zero. Pointwise convergence, however, does not imply uniform convergence, which is convergence in the supremum norm. Consequently we next establish a continuity condition, which along with pointwise convergence implies uniform convergence over any interval $[\gamma^*, 1]$ where $\gamma^* > 0$. The functions $L(\gamma; T)$ and $f(\gamma; T)$ have in probability the same maxima (actually the difference in maxima converge to zero) over such sets.

Lemma C is the identification condition. The functions $f(\gamma; T)$ are shown to be strictly less than $f(\gamma_0; T)$ over any compact set not containing γ_0 . This along with the uniform convergence of $L(\gamma; T) - f(\gamma; T)$ to zero insures that the m.l.e. of γ over an interval $[\gamma^*, 1]$ where $0 < \gamma^* < \gamma_0$ must converge in probability to γ_0 .

The final part of the proof is to show that when γ_0 exceeds zero, there is a $\gamma^* > 0$ such that the probability that the m.l.e. over the entire interval is less than γ^* goes to zero. We first show that there is a constant K which does not depend upon

γ^* such that $L(\gamma; T) \leq L(\gamma^*; T) + K$ for all $\gamma \leq \gamma^*$. Then we show that $f(\gamma_0; T) - f(\gamma^*; T)$ exceeds any positive number for γ^* sufficiently small and all T sufficiently large. Thus if $|L(\gamma; T) - f(\gamma; T)|$ is small for both γ^* and γ_0 , $L(\gamma_0; T)$ will exceed $L(\gamma; T)$ for all $0 \leq \gamma \leq \gamma^*$ and consequently, the m.l.e. will exceed γ^* .

LEMMA A: For $\gamma \in (0, 1]$, $\text{plim} [L(\gamma; T) - f(\gamma; T)] = 0$.

PROOF: Assuming the variables have been transformed via transformation P , the generalized sum of squared residuals is

$$(4.4) \quad S(\gamma) = w'[I - M(\gamma)]w = w'w - w'M(\gamma)w$$

where $w \sim N[0, \sigma_0^2 D(\gamma)^{-1} D(\gamma_0)]$ and $M(\gamma) = D(\gamma)^{-\frac{1}{2}} X(X'D(\gamma)^{-1} X)^{-1} X'D(\gamma)^{-\frac{1}{2}}$. Since $M(\gamma)$ is idempotent of rank k , $[w'M(\gamma)w]/T$ is non-negative with expectation zero in the limit. By Tchebyshev's inequality its probability limit is 0. Therefore to establish the lemma it must be shown that

$$(4.5) \quad \text{plim } w'w/T = \frac{\sigma_0^2}{T} \sum \frac{d_i(\gamma_0)}{d_i(\gamma)}$$

If $\gamma > \gamma_0$, the $d_i(\gamma_0)/d_i(\gamma)$ are increasing in d_i and are therefore bounded by $1/\gamma$. For $\gamma \leq \gamma_0$, they are monotonically decreasing and therefore bounded by $1/\delta^*$; both results follow from (4.1). This implies

$$(4.6) \quad 0 < d_i(\gamma_0)/d_i(\gamma) \leq 1/\delta^* + 1/\gamma$$

Since this ratio is uniformly bounded in T , the variances of the w_i^2 are uniformly bounded which implies

$$(4.7) \quad \lim_{T \rightarrow \infty} \text{var} (w'w/T) = 0$$

This along with the fact that

$$E(w'w/T) = \frac{\sigma_0^2}{T} \sum \frac{d_i(\gamma_0)}{d_i(\gamma)}$$

proves the lemma.

LEMMA B: The convergence of $L(\gamma; T)$ to $f(\gamma; T)$ is uniform in probability over any such set Γ ; that is $\text{plim} \{ \sup_{\gamma^* \leq \gamma \leq 1} |L(\gamma; T) - f(\gamma; T)| \} = 0$.

PROOF: The extension of pointwise convergence to uniform convergence in probability for a finite set is straightforward. One selects a sufficiently large T^* such that for all $T \geq T^*$ and all $i = 1, \dots, N$,

$$(4.8) \quad \text{pr} \{ |L(\gamma_i; T) - f(\gamma_i; T)| < \varepsilon/2 \} \leq 1 - \delta/N$$

Then

$$(4.9) \quad \text{pr} \{ \sup_i |L(\gamma_i; T) - f(\gamma_i; T)| < \varepsilon/2 \} \geq 1 - \delta$$

We next prove that with probability one there is a constant K which bounds the slope of the function $L(\gamma; T) - f(\gamma; T)$ everywhere on the set Γ . Differentiating this function yields

$$(4.10) \quad \frac{\sum \frac{w_t^2}{d_t(\gamma)} \frac{d_t - 1}{d_t(\gamma)}}{\sum \frac{w_t^2}{d_t(\gamma)}} - \frac{\sum \frac{d_t(\gamma_0)}{d_t(\gamma)} \frac{d_t - 1}{d_t(\gamma)}}{\sum \frac{d_t(\gamma_0)}{d_t(\gamma)}}$$

Both of the above expressions are weighted averages of the $(d_t - 1)/d_t(\gamma)$. As these are bounded from below by $-1/\delta^*$ and from above by $1/\gamma^*$ for $\gamma \in \Gamma$, the above expression is bounded in absolute value by

$$(4.11) \quad 2\{\sup_{t, \gamma \in \Gamma} |(d_t - 1)/d_t(\gamma)|\} \leq \frac{2}{\gamma^*} + \frac{2}{\delta^*} \equiv K.$$

With this result the condition that $|L(\gamma_i; T) - f(\gamma_i; T)| < \varepsilon/2$ implies

$$(4.12) \quad \sup_{|\gamma - \gamma_i| \leq \varepsilon/2K} |L(\gamma; T) - f(\gamma; T)| < \varepsilon.$$

The finite set of points can be selected such that, for all $\gamma \in \Gamma$, $|\gamma - \gamma_i| \leq \varepsilon/2K$ for some i . Then for all $T \geq T^*$

$$(4.13) \quad \text{pr} \left\{ \sup_{\gamma \in \Gamma} |L(\gamma; T) - f(\gamma; T)| < \varepsilon \right\} \geq 1 - \delta,$$

proving uniform convergence over Γ in probability.

LEMMA C: *There is a $K(\gamma; T) \geq > 0$ such that*

$$(4.14) \quad f'(\gamma; T) = K(\gamma; T)(\gamma_0 - \gamma)$$

for $\gamma \in (0, 1]$.

PROOF: Differentiating $f(\gamma; T)$ yields

$$f'(\gamma; T) = \left\{ \frac{1}{T} \sum (d_t - 1) \frac{d_t(\gamma_0)}{d_t(\gamma)^2} \right\} / \left\{ \frac{1}{T} \sum \frac{d_t(\gamma_0)}{d_t(\gamma)} \right\} \frac{1}{T} \sum \frac{d_t - 1}{d_t(\gamma)}.$$

Let $c_t = 1 - d_t$ which implies $d_t(\gamma) = 1 + \gamma c_t$ and

$$K_1(\gamma; T) = T \left[\sum \frac{d_t(\gamma_0)}{d_t(\gamma)} \right]^{-1}.$$

The derivative can be rewritten as

$$f'(\gamma; T) = K_1(\gamma; T) \frac{1}{T^2} \left[\sum_{i,j} \frac{c_i d_i(\gamma_0)}{d_i(\gamma)^2} - \sum_{i,j} \frac{c_i d_j(\gamma_0)}{d_i(\gamma) d_j(\gamma)} \right].$$

Combining the summations, this simplifies to

$$\begin{aligned}
 f'(\gamma; T) &= K_1(\gamma; T) \frac{1}{T^2} \sum_{i,j} \frac{c_i[d_i(\gamma)d_j(\gamma) - d_j(\gamma_0)d_i(\gamma)] d_j(\gamma)}{d_i(\gamma)^2 d_j(\gamma)^2} \\
 &= K_1(\gamma; T) \frac{\gamma_0 - \gamma}{T^2} \sum_{i,j} \frac{c_i^2 c_j \gamma - c_i c_j^2 \gamma + c_i^2 - c_j c_i}{d_i(\gamma)^2 d_j(\gamma)^2} \\
 &= K_1(\gamma; T) \frac{\gamma_0 - \gamma}{T} \sum_{i,j} \frac{c_i^2 - c_i c_j}{d_i(\gamma)^2 d_j(\gamma)^2} \\
 &= K_1(\gamma; T) \frac{\gamma_0 - \gamma}{2T\gamma} \sum_{i,j} \frac{\gamma(c_i - c_j)^2}{d_i(\gamma)^2 d_j(\gamma)^2} \\
 &= K_1(\gamma; T) \frac{\gamma_0 - \gamma}{2T\gamma} \sum_{i,j} \frac{[d_i(\gamma) - d_j(\gamma)]^2}{d_i(\gamma)^2 d_j(\gamma)^2}.
 \end{aligned}$$

In the Appendix it is shown that $d_i(\gamma) > 1 - \gamma + \gamma\delta^*$, which implies the leading term exceeds γ/δ^* . Thus, with some additional algebra,

$$f'(\gamma; T) = K_2(\gamma; T) [T^{-1} \sum d_i(\gamma)^{-2} - (T^{-1} \sum d_i(\gamma)^{-1})^2 (\gamma_0 - \gamma)]$$

with $K_2(\gamma; T) > \delta^*/2$. In the Appendix (see Result A1), a positive lower bound is established for the average squared deviation of the $d_i(\gamma)^{-1}$ from their average which holds for all T sufficiently large. This completes the proof of Lemma C.

Lemma C implies the functions $f(\gamma; T)$ have unique maxima at γ_0 . Furthermore, if $|\gamma - \gamma_0| > \varepsilon$, then $|f(\gamma_0; T) - f(\gamma; T)| \geq K\varepsilon^2/2$. By Lemma B the probability that $L(\gamma; T)$ will be arbitrarily close to $f(\gamma; T)$ uniform over any set of the form $0 < \gamma^* \leq \gamma \leq 1$ approaches 1. If $\gamma_0 > \gamma^*$, then the maximum likelihood estimator of γ over the set $[\gamma^*, 1]$ will converge in probability to γ_0 ; otherwise if $\gamma_0 \leq \gamma^*$ it will converge in probability to γ^* .

We now will show that when $\gamma_0 > 0$ for a sufficiently small γ^* , the probability that g , the m.l.e. over the entire unit interval is smaller than γ^* converges to zero; this implies g is consistent. The following results established in the Appendix will be used in the argument:

RESULT A2: For any constant $C > 0$ and $\gamma_0 > 0$, there are a T^* and $\gamma^* > 0$ such that

$$(4.15) \quad \ln \left[\frac{\sigma_0^2}{T} \sum \frac{d_i(\gamma_0)}{d_i(\gamma)} \right] > C$$

for all $T \geq T^*$ and $0 \leq \gamma \leq \gamma^*$.

RESULT A3: The function $T^{-1} \ln |D(\gamma)|$ is bounded from above and below uniformly in γ and T ; that is

$$(4.16) \quad \sup_{\gamma, T} \frac{1}{T} \ln |D(\gamma)| - \inf_{\gamma, T} \frac{1}{T} \ln |D(\gamma)| = K^* < \infty.$$

Using Result A2, a $\gamma^* < \gamma_0$ and T exist such that for any $\varepsilon > 0$

$$(4.17) \quad \text{pr} [L(\gamma^*; T) \leq f(\gamma_0; T) - K^* - 1/\delta^* - 2\varepsilon] > 1 - \delta/2.$$

In Lemma B, a lower bound of $-1/\delta^*$ was established for the derivative of $-\ln [S(\gamma)/T]$. Using this fact along with Result A3, we conclude that

$$(4.18) \quad L(\gamma; T) - K^* - 1/\delta^* \leq L(\gamma^*; T) \quad \text{for} \quad 0 \leq \gamma \leq \gamma^*.$$

These two relationships imply

$$(4.19) \quad \text{pr} [\sup_{\gamma \leq \gamma^*} L(\gamma; T) \leq f(\gamma_0; T) - 2\varepsilon] \geq 1 - \delta/2$$

for $T \geq T_1$. Pick T_2 sufficiently large that

$$(4.20) \quad \text{pr} [|L(\gamma_0; T) - f(\gamma_0; T)| < \varepsilon] > 1 - \delta/2$$

for $T \geq T_2$. Given (4.19) and (4.20),

$$(4.21) \quad \text{pr} [\sup_{\gamma \leq \gamma^*} L(\gamma; T) < L(\gamma_0; T)] > 1 - \delta$$

for $T \geq \max [T_1, T_2]$. This proves that

$$\lim_{T \rightarrow \infty} \text{pr} [g_T \leq \gamma^*] = 0.$$

We have established that g converges in probability to γ_0 if γ_0 is positive and now establish this result for $\gamma_0 = 0$. In this case, by Lemmas B and C, the m.l.e. over any interval $[\varepsilon, 1]$, $0 < \varepsilon < 1$, converges in probability to ε . Denoting this restricted m.l.e. by g_ε ,

$$\lim \text{pr} [g_\varepsilon > 2\varepsilon] = 0.$$

But

$$\text{pr} [g > 2\varepsilon] \leq \text{pr} [g_\varepsilon > 2\varepsilon],$$

so we can conclude that

$$\lim \text{pr} [g > 2\varepsilon] = 0.$$

Since ε can be any positive number less than one, g converges in probability to 0 if $\gamma_0 = 0$.

The above discussion can be summarized by the following theorem:

THEOREM: *The maximum likelihood estimator g of γ converges in probability to γ_0 , the true parameter value.*

COROLLARY: *The maximum likelihood estimator, $s^2(g)$, of σ^2 converges in probability to σ_0^2 .*

PROOF: Since $s^2(g)$ is just $S(g)/T$, and by Lemma A, $\text{plim} S(\gamma_0)/T = \sigma_0^2$, the consistency of g implies this result.

Following the usual analysis but using the log likelihood function concentrated on β only,⁹ the asymptotic distribution of $\hat{\theta}$, the m.l.e. for $\theta' = (\gamma, \sigma^2)$, will be

$$\sqrt{T}(\hat{\theta} - \theta_0) \sim N[0, I(\theta_0)^{-1}],$$

where the information matrix $I(\theta_0)$ is

$$I(\theta_0) = -\frac{1}{T}E\left(\frac{\partial^2 L}{\partial \theta^2}\right) = \begin{bmatrix} \frac{1}{2T} \sum \frac{(d_i - 1)^2}{d_i(\gamma_0)} & \frac{-\sigma_0^2}{2T} \sum \frac{(d_i - 1)}{d_i(\gamma_0)} \\ \frac{-\sigma_0^2}{2T} \sum \frac{(d_i - 1)}{d_i(\gamma_0)} & \frac{\sigma_0^4}{2} \end{bmatrix}$$

Inverting $I(\theta_0)$, the variance of the asymptotic distribution of g is obtained, namely

$$\text{var} [\sqrt{T}(g - \gamma_0)] = 2 / \left[\frac{1}{T} \sum \frac{(d_i - 1)^2}{d_i(\gamma_0)^2} - \left(\frac{1}{T} \sum \frac{(d_i - 1)^2}{d_i(\gamma_0)} \right)^2 \right]$$

This may be used to test for the significance of the estimator g .

In practice one does not search the entire interval but searches over a finite number of points as we proposed in Section 2. This estimator will converge in probability to either the γ_i immediately to the right of γ_0 or the one immediately to the left. If every point in the unit interval is within ε of some point in the search set, then the value to which g converges is within 2ε of the true value.

5. CONCLUDING REMARKS

It may frequently occur in econometric research that the parameters of an equation are subject to stochastic variation over time. In this paper we have developed models and their estimators which explicitly incorporate this possibility. Because parameter variation may arise from a variety of sources, the variation process is assumed to be a general one, incorporating both permanent and transitory changes. The model proposed has a number of important special cases which have different implications about the form of parameter variation.

The maximum likelihood estimation techniques developed provide estimates of the permanent components of the parameters, the fraction of variation due to permanent changes in the parameters and the variance. The transformation developed makes estimation of the models quite feasible. The computations required are no more capital intensive than many commonly used nonlinear estimation techniques.¹⁰ The estimates of the permanent component of the β 's will not be consistent because of the assumed probability structure. A subset of the parameters (γ and σ^2) can be estimated consistently, however, which implies

⁹ Analysis of this type is presented in detail in [7].

¹⁰ The authors have developed a Fortran program that efficiently performs this estimation. It is available on the NBER TROLL system.

the estimate of the permanent component of the β 's will be asymptotically efficient and yield optimal predictions.

*Tufts University and National Bureau of Economic Research
and*

Carnegie-Mellon University

Manuscript received August, 1972; last revision received November, 1974.

APPENDIX

In the proof of consistency of the maximum likelihood estimators of γ and σ^2 certain propositions about eigenvalues were relied upon. The transformed model is¹¹

$$(A1) \quad Y^* = X^*\beta + \mu^*.$$

The vector μ^* is distributed as

$$(A2) \quad \mu^* \sim N[0, \sigma^2(1 - \gamma)I + \gamma Q^*],$$

where Q^* is defined as

$$(A3) \quad q_{ij}^* = \min [T - i + 1, T - j + 1] x_i^* \Sigma_v x_j^* \quad \text{and} \quad x_n^* = x_n \sqrt{x_n' \Sigma_u x_n}.$$

It is the properties of the eigenvalues Q^* that are of concern. The proof of consistency requires that the eigenvalues be uniformly bounded away from zero and that they possess finite variability.

For the special case of the intercept variation model the necessary properties of the eigenvalues are easily shown. For that case the matrix H corresponding to Q^* has ij th element

$$(A4) \quad h_{ij} = \min [T - i + 1, T - j + 1].$$

The j th largest eigenvalue of H is¹²

$$(A5) \quad \lambda_j(H) = 2 + 2 \cos [2\pi(T - j + 1)/(2T + 1)]$$

which is bounded away from zero by $\frac{1}{4}$. Also, since for this case $d_i = \lambda_i(H)$, the expression

$$(A6) \quad \frac{1}{T} \sum \left(\frac{1}{d_i(\gamma)} - \frac{1}{T} \sum \frac{1}{d_i(\gamma)} \right)^2$$

has a positive lower bound for any $\gamma > 0$. This is required for Lemma C to hold.

The following lemmas generalize the above result for the general model. Let X_i^* be the diagonal matrix with i th diagonal element equal to x_i^* and assume Σ_v is diagonal.¹³ The matrix Q^* can be written as

$$(A7) \quad Q^* = \sum_{i=1}^k \sigma_{ii}^v X_i^* H X_i^*,$$

where σ_{ii}^v is the appropriate element of Σ_v .

In the subsequent discussion all matrices are symmetric and the notation $A \geq B$ means matrix $A - B$ is positive semi-definite.

LEMMA A1: *If B is positive semi-definite, then $\lambda_j(A + B) \geq \lambda_j$, where λ_j denotes the j th largest eigenvalue of a matrix.*

¹¹ The * notation on the transformed variables which was dropped for convenience is re-employed here. It is necessary because it is important to make explicit the fact that the X 's and Y 's have been transformed in a particular way.

¹² In [3, p. 66] the eigenvalues for the inverse of H (numbering of rows and columns reversed) is developed, implying (A5).

¹³ This assumption can be made without loss of generality because, alternatively, an appropriate triangular factorization of Σ_v could be made.

PROOF: See Bellman [3, p. 115].

LEMMA A2: If A and B are diagonal, H positive definite, and for all i $|a_{ii}| \geq |b_{ii}|$, then $\lambda_j(AHA) \geq \lambda_j(BHB)$.

PROOF: See Anderson and Das Gupta [2].

LEMMA A3: The j th largest eigenvalue of Q^* satisfies the following inequalities:

$$(A8) \quad k_1 \lambda_j(H) \leq \lambda_j(Q^*) \leq k_2 \lambda_j(H),$$

where

$$(A9) \quad k_1 = \max \sigma_{ii}^v \bar{x}_i^2 \quad \text{and} \quad k_2 = \max \{2, \sum \sigma_{ii}^v \bar{x}_i^2\}.$$

PROOF: By Lemma A1 and equation (A.7)

$$(A10) \quad \sigma_{ii}^v X_i^* H X_i^* \leq Q^*.$$

Next using Lemma A2,

$$(A11) \quad \lambda_j(\bar{x}_i^2 H) \leq \lambda_j(X_i^* H X_i^*)$$

which implies the left-hand inequality of (A12).

Since¹⁴

$$(A12) \quad \lambda_j(Q^*) = \underset{y_1, \dots, y_{j-1}}{\text{minimum}} \underset{z', y_n = 0; n=1, \dots, j-1; \|z\| \leq 1}{\text{maximum}} \{z' \sum \sigma_{ii}^v X_i^* H X_i^* z\},$$

it follows that

$$(A13) \quad \begin{aligned} \lambda_j(Q^*) &\leq \underset{y_1, \dots, y_{j-1}}{\text{minimum}} \{ \sum \sigma_{ii}^v \underset{z', y_n = 0; n=1, \dots, j-1; \|z\| \leq 1}{\text{maximum}} [z' X_i H X_i^* z] \} \\ &\leq \underset{y_1, \dots, y_{j-1}}{\text{minimum}} \{ \sum \sigma_{ii}^v \bar{x}_i^2 \underset{z', y_n = 0; n=1, \dots, j-1; \|z\| \leq 1}{\text{maximum}} [z' X_i z] \} \\ &= \sum \sigma_{ii}^v \bar{x}_i^2 \lambda_i(H). \end{aligned}$$

This completes the proof.

RESULT A1: The $1/d_j(\gamma)$ have finite variability in the limit in the sense that the limit inferior of

$$(A14) \quad \frac{1}{T} \sum \left[\frac{1}{d_j(\gamma)} - \frac{1}{T} \sum \frac{1}{d_j(\gamma)} \right]^2$$

is positive.

PROOF: The nature of the proof is to show that the limit inferior of the fraction of $d_j(\gamma)$ exceeding $2k_2$ is positive as is the limit inferior of the fraction less than k_2 . Clearly the limit inferior of expression (A14) must then be positive.

By Lemma A3 the limit inferior of the fraction of the $d_j(\gamma)$ exceeding $2k_2$ exceeds the limit of the number of $\lambda_j(H)$ exceeding $2k_2/k_1$. Since the latter limit is positive, the limit inferior of the fraction of the $d_j(\gamma)$ exceeding $2k_2$ is positive. Also by Lemma A3, the limit inferior of the number of $d_j(\gamma)$ that are less than k_2 exceeds the limit of the number of $\lambda_j(H)$ that are less than 1. The latter limit is positive and the result is established.

RESULT A2: For any $\gamma_0 > 0$ and $K > 0$, there exist a T^* and $\gamma^* > 0$ such that

$$(A15) \quad h(\gamma, \gamma_0) \equiv \frac{1}{T} \sum \frac{d_j(\gamma_0)}{d_j(\gamma)} > K$$

for all $T \geq T^*$ and $0 \leq \gamma \leq \gamma^*$.

¹⁴ This representation of the eigenvalues derives from the Courant-Fischer min-max theorem. See Bellman [3, pp. 113-115].

PROOF: Since $d_t(\gamma) = (1 - \gamma) + \gamma\lambda_t(Q^*)$, Lemma A3 implies that

$$(A16) \quad d_t(\gamma_0) \geq \gamma_0 k_1 \lambda_t(H) \geq \gamma_0 k_1 / 4$$

and that

$$(A17) \quad d_t(\gamma) \leq k_2(1 - \gamma) + \gamma k_2 \lambda_t(H).$$

Therefore

$$(A18) \quad h(\gamma, \gamma_0) \geq \frac{\gamma_0 k_1}{4k_2} \frac{1}{T} \sum \frac{\lambda_t(H)}{(1 - \gamma) + \gamma \lambda_t(H)}.$$

Using (A5) to substitute for $\lambda_t(H)$ yields

$$(A19) \quad \begin{aligned} h(\gamma, \gamma_0) &\geq \frac{\gamma_0 k_1}{4k_2} \frac{1}{T} \sum \frac{1}{(1 - \gamma) \left[2 + 2 \cos \left(\frac{2\pi t}{2T + 1} \right) \right] + \gamma} \\ &= \frac{\gamma_0 k_1}{4k_2} \frac{1}{T} \sum \frac{1}{(1 - \gamma) \left[2 - 2 \cos \left(\frac{2\pi t - \pi}{2t + 1} \right) \right] + \gamma} \\ &\geq \frac{\gamma_0 k_1}{4k_2} \int_{1/(2T+1)}^{(2T-3)/(2T+1)} \frac{1}{(1 - \gamma)(2 - 2 \cos \pi x) + \gamma} dx. \end{aligned}$$

The last result follows from numerical integration theory for monotonically decreasing functions. By substituting the smaller function $1 - \pi x$ for $\cos(\pi x)$ and reducing the range of the integration of this positive function, we conclude

$$(A20) \quad h(\gamma, \gamma_0) > \frac{\gamma_0 k_1}{4k_2} \int_{1/(2T+1)}^{\frac{1}{2}} \frac{1}{(1 - \gamma)\pi x + \gamma} dx.$$

There are a T^* and positive γ^* such that the right-hand side of (A20) exceeds any positive constant K . Since that expression is increasing in T and decreasing in γ , $h(\gamma, \gamma_0) > K$ for all $\gamma < \gamma^*$ and $T > T^*$.

RESULT A3: The determinant of the matrix $\Omega^*(\gamma) = (1 - \gamma)I + \gamma Q^*$ satisfies the following bounds:

$$(A21) \quad \ln \{ \min(1, k_1/4) \} \leq \frac{1}{T} \ln |\Omega^*(\gamma)| \leq \ln(k_2) + 4/k_1.$$

PROOF: The eigenvalues of $\Omega^*(\gamma)$ are $(1 - \gamma) + \gamma\lambda_j(Q^*)$. Since the determinant of a symmetric matrix is the product of the eigenvalues

$$(A22) \quad \frac{1}{T} \ln |\Omega^*(\gamma)| \leq \ln \{ \inf \lambda_j(\Omega^*) \}$$

by equation (A5) and Lemma A3

$$(A23) \quad \lambda_j(Q^*) \leq k_1/4,$$

so

$$(A24) \quad \inf \lambda_j(\Omega^*) \leq \min(1, k_1/4)$$

which proves the left hand side of (A21).

To prove the right-hand inequality we observe that

$$(A25) \quad \frac{1}{T} \ln |\Omega^*(\gamma)| = \frac{1}{T} \sum \ln \{ (1 - \gamma) + \gamma\lambda_j(Q^*) \}.$$

Differentiating twice yields a negative function which implies that the above function is convex. For $\gamma = 1$, we find

$$(A26) \quad \frac{1}{T} \ln |\Omega^*(1)| = \frac{1}{T} \sum \ln \lambda_j(Q^*).$$

From Lemma A3 we know that $\lambda_j(Q^*) \leq k_2 \lambda_j(H)$, which implies

$$(A27) \quad \frac{1}{T} \ln \Omega^*(1) \leq \frac{1}{T} \sum (\ln k_2 + \ln \lambda_j(H)).$$

The determinant of H is 1 so

$$(A28) \quad \frac{1}{T} \ln |\Omega^*(1)| \leq \ln k_2.$$

Differentiating we find

$$(A29) \quad \begin{aligned} \frac{d}{d\gamma} \left[\frac{1}{T} \ln |\Omega^*(\gamma)| \right]_{\gamma=1} &= \frac{1}{T} \sum \left[\frac{\lambda_j(\Omega^*) - 1}{\lambda_j(Q^*)} \right] \\ &\geq -\frac{1}{T} \sum \frac{1}{\lambda_j(Q^*)} \\ &\geq -4/k_1. \end{aligned}$$

Given that the function is convex, has an upper bound at $\lambda = 1$, and a lower bound for its derivative, it follows that¹⁵

$$(A30) \quad \frac{1}{T} \ln |\Omega^*(\gamma)| \leq \ln k_2 + \frac{4}{k_1}$$

for $0 \leq \gamma \leq 1$.

REFERENCES

- [1] ANDERSON, T. W.: "Estimation of Covariance Matrices Which Are Linear Combinations or Whose Inverses Are Linear Combinations of Given Matrices," in *Essays in Probability and Statistics*, Chapel Hill, N.C.: University of North Carolina Press, 1970.
- [2] ANDERSON, T. W., AND S. DAS GUPTA: "A Monotonicity Property of the Power Functions of Some Tests of the Equality of Two Covariances Matrices," *Annals of Mathematics and Statistics*, 35 (1964), 1059-1063.
- [3] BELLMAN, R.: *Introduction to Matrix Analysis*. New York: McGraw-Hill, 1960.
- [4] COOLEY, T.: "Estimation In the Presence of Sequential Parameter Variation," unpublished Ph.D. Thesis, University of Pennsylvania, 1971.
- [5] COOLEY, T., AND E. PRESCOTT: "An Adaptive Regression Model," *International Economic Review*, 14 (1973), 364-371.
- [6] ———: "Tests of An Adaptive Regression Model," *Review of Economics and Statistics*, 55 (1973), 248-256.
- [7] DHRYMES, P.: *Econometrics*. New York: Harper and Row, 1970.
- [8] DUSENBERRY ET AL: *The Brookings Quarterly Econometric Model of The United States*. Chicago: Rand McNally, 1965.
- [9] FAMA, E. T.: "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25 (1970), 383-417.
- [10] FELLER, W.: *An Introduction To Probability Theory and Its Application*. New York: Wiley, 1966.
- [11] HADLEY, G.: *Linear Algebra*. Reading, Mass.: Addison-Wesley, 1961.
- [12] HILDRETH, C., AND J. P. HOUCK: "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association*, 63 (1968), 584-595.
- [13] HINICH, M. J., AND J. V. FARLEY: "Theory and Application of an Estimation Model for Time Series with Nonstationary Means," *Management Science*, 12 (1966), 648-658.
- [14] LUCAS, R.: "Econometric Policy Evaluation: A Critique," Carnegie-Mellon Working Paper, 1973.
- [15] MALINVAUD, E.: *Statistical Methods of Econometrics*. Amsterdam: North-Holland, 1966.
- [16] NERLOVE, M.: *The Dynamics of Supply: Estimation of Farmers Response To Price*. Baltimore: The Johns Hopkins Press, 1958.

¹⁵ We used the fact that if $g(x)$ is convex, then $g(x) \leq g(x_0) + g'(x_0)(x - x_0)$.

- [17] QUANDT, R.: "Estimation of Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of The American Statistical Association*, 52 (1958), 873-880.
- [18] RALSTON, A., AND H. F. WILF, EDS.: *Mathematical Methods for Digital Computers*, Vol. 2. New York: Wiley, 1967.
- [19] RAO, C. R.: "Estimation of Heteroscedastic Variances in Linear Models," *Journal of The American Statistical Association*, 65 (1970), 161-172.
- [20] ROSENBERG, B.: "Varying Parameter Estimation," unpublished Ph.D. thesis, Harvard University, 1968.
- [21] SILVEY, S. D.: "A Note on Maximum Likelihood in the Case of Dependent Random Variables," *Journal of the Royal Statistical Society, Series B*, 23 (1961), 444-457.
- [22] SWAMY, P. A. V. B.: "Efficient Inference In a Random Coefficients Regression Model," *Econometrica*, 28 (1970), 311-323.
- [23] THEIL, H.: *Economic Forecasts and Policy*. Amsterdam: North-Holland, 1965.
- [24] WALD, A.: "A Note on the Consistency of Maximum Likelihood Estimates," *The Annals of Mathematical Statistics*, 20 (1949), 595-601.