The Stochastic Coefficients Approach to Econometric Modeling, Part III: Estimation, Stability Testing, and Prediction

P.A.V.B. Swamy, Roger K. Conway, and Michael R. LeBlanc

Abstract. *In this final article of our three-part series, we demonstrate why stochastic coefficients models are well suited to predict future variables We analyze the forecasting problem and consider various criteria of prediction If a forecaster must choose one from among several coherent predictors, then the choice should be the one with the best track record Decomposing the forecast error shows that stochastic coefficients models can cover more possible sources of prediction error and correct for them The empirical record shows that stochastic coefficients models can substantially reduce out-of-sample forecast errors more than fixed coefficients models Our assessment of coefficient stability tests is they are contradictory, misleading, and without empirical value*

Keywords. *Stochastic coefficients, fixed coefficients, conditional expectation, Bayesian inference, coherence, estimation, prediction, stability tests*

Editor's note Part I A Critique of Fixed Coefficients Models appeared in Vol 40, No 2, Spring 1988 *Part II Description and Motivation* appeared in Vol 40, No 3, Summer 1988

Stochastic coefficients models are ideally suited to the problem of predicting future values of variables We say ideally because such models cover almost all possible sources of prediction errors and introduce suitable corrections for each error We also show that either parameter estimation or testing of hypotheses about parameters is a "halfway house" on the road to predicting future observations

If the objective of estimation is forecast accuracy, then one should attempt to find an estimation procedure that yields predictions as close to actual realizations as possible One should select the predictor that has the highest probability of taking values close to actual realizations [1] It is impossible to derive predictors based on this general criterion A necessary condition, however, for a predictor to take values close to actual realizations with the highest probability is that the mean square error (that is, the predictor's expected squared deviation from the actual realization) is a minimum

Predictors with uniformly minimum mean square error typically do not exist, a difficulty that can be avoided by replacing the criterion of minimum mean square error with the criterion of minimum average mean square error

The latter criterion selects a predictor if its expected squared deviation from a variable is a minimum Minimum average mean square error predictors take the form of conditional expectations, which can be evaluated exactly if their true functional forms are known and if they do not depend on unknown parameters Surfacing are problems in which the functional forms assumed for conditional expectations may not coincide with their true functional forms, and the errors of the estimates of the unknown parameters appearing in the assumed functional forms substantially affect the accuracy with which the desired values can be predicted

Our approach, then, is to use stochastic coefficients models to specify conditional expectations The motive for introducing stochastic coefficients models is the hope that such models can approximate true models better than fixed coefficients models This hope is not without a methodological basis If a functional form assumed for a conditional expectation is true, then it is appropriate to adopt the criterion of minimum variance unbiasedness for parameter estimation This criterion can satisfy a necessary condition for maximizing the probability that a predictor can generate predictions that are close to the true conditional expectation However, there still remains the problem of recognizing an operational unbiased predictor with minimum average mean square error

Swamy is a senior economist with the Board of Governors, Federal Reserve System, and adjunct professor of economics at The George Washington University (Washington, DC), and Conway and LeBlanc are agricultural economists with the Resources and Technology Division, ERS The authors received valuable comments and help from James Barth, Charlie Hallahan, Arthur Havenner, Tom Lutton, Ron Mittelhammer, Peter von zur Muehlen, Nadine Loften, and Douglas McManus

[1] Here, we use the term "predictors" to refer to a random variable's real valued measurable functions that are used to predict the future values of the random variable The term "predictions" refers to the values taken by these predictors We use the term "estimators" to refer to the real-valued measurable functions of random variables which are used to guess the unknown true value of a fixed parameter

(51)[2] Consequently, we may follow de Finetti's suggestion that the condition of coherence is the only minimum required condition one should impose on predictors

The derivation of coherent predictions is no easy task For example, Lane and Sudderth's method of deriving coherent predictions is difficult to use because we must specify a finitely additive probability distribution *(32)* We may find a Bayesian procedure which gives coherent predictions, however, to approximate our opinions via a probability density function (pdf) Even this procedure may be complicated because the specification of a consistent set of prior probabilities of models and the prior pdf's for their parameters require a demanding exercise in self-interrogation A Bayes procedure based on prior probabilities and prior pdf's can produce better forecasts than a non-Bayesian procedure in some cases But, expressing our opinions about models and their parameters in the form of prior probabilities and pdf's, respectively, and checking their consistency, are very difficult tasks The difficulty of checking the logical validity of models is, of course, common to both Bayesian and non-Bayesian methods

We can apply exact or approximate Bayesian and non-Bayesian methods to generate predictions under some assumptions about the data-generating process By dividing available data into fitting and forecast samples, we can use part of the data to generate predictions for the rest of the data, comparing these predictions with the realized values The result of these comparisons can guide the choice of models in other situations that share common features with the environment, resulting in the data used in the comparisons Our experience with such comparisons suggests that time-varying and stochastic modeling of regression slopes may contribute to improved forecasts These forecasts become useful in analyzing sources of so-called coefficient instability, predicting uncertainty that may arise in conventional models

The Forecasting Problem[3]

We are concerned with the following prediction problem we want to predict the value y_{T+s} that would be taken by a variable y_t^* in some future period s after T, where T is the terminal period of the currently available sample observations on y_t^* We will make this prediction having T+s observations up to time T+s on a vector x_t^* of K variables that are related to another variable y_t^* and also having T observations on y_t^* up to time T[4]

Formalizing this point prediction problem involves a random variable y_t^* taking on values in a sample space, according to a distribution which is assumed to belong to a family The currently available sample observations $y = (y_1, y_2, \ldots, y_T)'$ on $y^* = (y_1^*, y_2^*, \ldots, y_T^*)'$ constitute the data We may also have observations on variables with the symbol $x_t^* = (x_{1t}^*, x_{2t}^*, \ldots, x_{Kt}^*)'$ The observations x_t, t = 1,2, ..., T+s, on these variables are also part of the data when the distribution of y_t^* is determined by its own past values and by current and past values of x_t^*

The problem is the determination of a suitable predictor, that is, a real-valued function $\hat{y}_{T+s}$ defined over the sample space, of which it is hoped that $\hat{y}_{T+s}(y^{*\prime}, x^{*\prime})$ will tend to be close to the unknown value y_{T+s} The value $\hat{y}_{T+s}(y', x')$ taken on by $\hat{y}_{T+s}(y^{*\prime}, x^{*\prime})$ for the observed value $(y', x')'$ of $(y^{*\prime}, x^{*\prime})'$ is then the forecast of y_{T+s}, which will be our educated guess for the unknown value We say that a predictor is operational if it does not involve any unobservable quantities

A best predictor $\hat{y}_{T+s}$ should be sufficiently close to the actual realization y_{T+s} and because $\hat{y}_{T+s}$ is a random variable, the value y_{T+s} is covered by an interval of values which $\hat{y}_{T+s}$ takes with a high probability To make this requirement precise, we specify four measures of closeness of (or distance from) a predictor to y_{T+s}

Criteria of Prediction

Swamy and Schinasi extend well-known criteria used in point estimation to the problem of point prediction as follows *(51)*

(a) *Criterion of highest concentration* A predictor $\hat{y}_{T+s}$ of the actual value y_{T+s} is better than any other predictor $\tilde{y}_{T+s}$ when

$$\text{pr}(y_{T+s} - \lambda_1 < \hat{y}_{T+s} < y_{T+s} + \lambda_2) \geq \text{pr}(y_{T+s} - \lambda_1 < \tilde{y}_{T+s} < y_{T+s} + \lambda_2), \quad (1)$$

for all possible values of λ_1 and λ_2 in a chosen interval $(0, \lambda)$ and for all possible realizations y_{T+s} Here, pr denotes "probability"

(b) *Minimum mean square error* Swamy and Schinasi show that a necessary condition to satisfy criterion 1 for all λ and for all y_{T+s} is

$$E(\hat{y}_{T+s} - y_{T+s})^2 \leq E(\tilde{y}_{T+s} - y_{T+s})^2, \quad (2)$$

with the inequality being strict on a set of a positive Lebesgue measure, that is, the mean square error of $\hat{y}_{T+s}$ about the actual realization y_{T+s} is a minimum

(c) *Pitman's nearness (PN)* A predictor $\hat{y}_{T+s}$ is nearer to the value y_{T+s} than another predictor $\tilde{y}_{T+s}$ when

[2]Italicized numbers in parentheses cite sources listed in the References at the end of this article

[3]Several sections in this article are based on *(51)*

[4]We distinguish a random variable from its value by an asterisk For example, y_t is the value taken by the random variable y_t^* in period t

$$\text{pr}[L(\hat{y}_{T+s}, y_{T+s}) < L(\tilde{y}_{T+s}, y_{T+s})] > 1/2, \quad (3)$$

where the loss function, $L(\hat{y}_{T+s}, y_{T+s})$, represents the consequences of predicting y_{T+s} by a value of $\hat{y}_{T+s}$

Swamy and Schinasi state two sets of sufficient conditions for a predictor to be nearer to y_{T+s} than another predictor in the PN sense

(d) *Minimum average mean square error* A predictor $\hat{y}_{T+s}$ is called the minimum "average mean square error" predictor if it minimizes

$$E(\hat{y}_{T+s} - y^*_{T+s})^2 \quad (4)$$

If we wish to predict y_{T+s} from a Borel measurable function f(y*), say, of y*, then among all such functions of y* with finite second moment, that which minimizes the average mean square error 4 with $\hat{y}_{T+s} = f(y^*)$ is the conditional expectation of y^*_{T+s}, given $y^* = y$, denoted by [5]

$$f(y) = E(y^*_{T+s} \mid y) \quad (5)$$

When $\hat{y}_{T+s} = E(y^*_{T+s} \mid y)$, the average mean square error 4 reduces to the average conditional variance of y^*_{T+s}, given y

Criteria (a), (b), and (c) are based on distances from predictors to actual realizations, while criterion (d) is based on distances from predictors to y^*_{T+s}, which is a random variable In the definition of criteria (b) and (d), attention has been implicitly restricted to predictors with finite variance, because otherwise the problem of minimizing the (average) mean square error does not arise Predictors with infinite variance violate the necessary condition 2 and so they cannot satisfy the criterion of highest concentration (criterion 1) In fact, predictors satisfying either the criterion of highest concentration or the criterion of minimum mean square error (criterion 2) do not generally exist For this reason, the minimum average mean square error criterion 4 is the one used extensively in the econometric literature However, minimum average mean square error predictors sometimes have infinite mean square error For example, the forecast y_{T+s-1} is a minimum average mean square error forecast if y^*_t follows a random walk, $y^*_t = y^*_{t-1} + a^*_t$, $t = 0, \pm 1, \pm 2, \ldots$, where a^*_t is a white noise error term with zero mean For this process, $E(y^*_{T+s} \mid y_{T+s-1}) = y_{T+s-1}$, and $E(y_{T+s-1} - y^*_{T+s})^2$ is finite when viewed as the average conditional variance of y^*_{T+s}, given y_{T+s-1}, and is infinite when viewed as an unconditional mean square error of y^*_{T+s} about y_{T+s-1} The distinction between criteria (b) and (d) is not clearly explained in the econometric literature

[5] See (*36*, p 264) for a proof of this statement

Criterion (c) is different from criteria (b) and (d) in that attention is also given to predictors with infinite variance Keating and Mason give examples of predictors that are good in the PN sense (*27, 28*)

The result expressed in equation 5 is theoretically important but has little practical use unless one knows the true functional form of the conditional expectation of y^*_{T+s}, given y A conditional expectation that does not exist, however, could not have generated our data Conditions for the existence of various forms of conditional expectations are different (*52*) One usually cannot verify the truth of these conditions, and the best one can do is to argue (from coherent economic theories, for example) that, in many cases, one would expect $(y^*_{T+s}, y^{*\prime})'$ to follow a distribution which implies the existence of the conditional expectation (equation 5) of particular form (*49*) [6] The first step in any statistical method of generating predictions is to formulate a statistical model about the data-generating process The distribution implied by this model is the one assumed for $(y^*_{T+s}, y^{*\prime})'$

If the vector $(y^*_{T+s}, y^{*\prime})'$ is jointly normal, then the conditional expectation in equation 5 can be expressed as

$$E(y^*_{T+s} \mid y) = Ey^*_{T+s} + \text{cov}(y^*_{T+s}, y^{*\prime})\,[\text{var}(y^*)]^- (y - Ey^*), \quad (6)$$

where $\text{cov}(y^*_{T+s}, y^{*\prime}) = E(y^*_{T+s} - Ey^*_{T+s})(y^* - Ey^*)'$, and $[\text{var}(y^*)]^-$ is any generalized inverse of the covariance matrix of y^* (*36*, p 522) These variances and covariances may be time dependent if the variable $(y^*_{T+s}, y^{*\prime})'$ is nonstationary

Conditions other than normality may also be used to derive the predictor, equation 6, with the minimum average mean square error property Specifically, Chipman (*6*, pp 603-5) proved that the predictor (equation 6) has the minimum average mean square error within the class of linear (affine) predictors of y_{T+s} whenever $(y^*_{T+s}, y^{*\prime})'$ is not normal but possesses finite second-order moments So, one may be tempted to conjecture that only normal distributions give linear predictors with the minimum average mean square error property in linear regression While this conjecture is not true without further conditions, it is true for most practical purposes, as rigorously proved by Goel and DeGroot (*19*, p 899) and Rao (*38*)

[6] Unlike incoherent theories, coherent theories whose premises are not contradictory can be true The intuitively appealing concept of evidence stating that under no hypothesis, H, shall there be a high probability of outcomes being interpreted as strong evidence against H is useless unless our hypotheses are grounded in coherent economic theories

Where the mean vector, $E(y^*_{T+s}, y^{*\prime})'$, is unknown (which seems usual) and the covariance matrix of $(y^*_{T+s}, y^{*\prime})'$ is known (which seems unusual), Goldberger, Swamy and Mehta, and Harville have minimized criterion 4, subject to the restriction that the predictor $\hat{y}_{T+s}$ is equal to the homogeneous linear function, $c'y^*$, where c is a T×1 vector of constants, and to the unbiasedness restriction that $E\hat{y}_{T+s}$ is equal to the mean assumed for y^*_{T+s} *(20, 23, 47)*[7] The predictor that comes out of this constrained minimization procedure is called the minimum variance linear (homogeneous) "unbiased" predictor, and it is the same as the predictor expressed in equation 6 with Ey^*_{T+s} and Ey^* replaced by their respective minimum variance linear (homogeneous) unbiased estimators The minimum variance linear "unbiased" predictor of y_{T+s} will coincide with the minimum variance "unbiased" predictor of y_{T+s} in the normal case but not with the conditional expectation in equation 5, even in this case

Swamy and Schinasi show that the criterion of minimum variance unbiasedness satisfies a neces sary condition for maximizing the probability that a predictor generates predictions close to its expected value *(51)* No guarantee provides that actual realizations will be close to this expected value The expected value of the minimum variance linear "unbiased" predictor of y_{T+s} will not coincide with the true expected value of y^*_{T+s} if the unbiasedness restriction is erroneous[8] Imposing erroneous unbiasedness restrictions may have the undesirable consequence of yielding highly inaccurate forecasts

Conditions 1 and 2 logically lead to the criterion of minimum mean square error in satisfying a necessary condition foi maximizing the probability with which a predictor takes values close to actual realizations If condition 2 is true, then it follows that for at least one value of y^*_{T+s} the inequality 1 is true but not necessarily for all possible values of y^*_{T+s} *(36,* p 96) This result shows that a predictor, $\hat{y}_{T+s}$, which minimizes the mean square error, $E(\hat{y}_{T+s} - y_{T+s})^2$, for all values of y^*_{T+s} is useful if it satisfies the inequality 1 for those values of y^*_{T+s} which actually occur Unfortunately, such a predictor does not exist, as shown in the statistics literature *(34,* p 5) Nevertheless, comparing 2 with 4 shows that the conditional expectation of y^*_{T+s}, given a realization of y^*, nearly satisfies condition 2 if the true conditional distribution of y^*_{T+s}, given $y^* = y$, is sufficiently tight around its mean value Therefore, a necessary condition for obtaining accurate forecasts is that we specify and evaluate accurately the true conditional expectation of y^*_{T+s}, given $y^* = y$ Perhaps, we can better satisfy this necessary condition if we work with stochastic coefficients models rather than fixed coefficients models Again, any rigorous derivation of an econometric model using probability calculus naturally leads to a stochastic coefficients model unless severe restrictions are imposed on derivation True models are better approximated by stochastic coefficients models than by fixed coefficients models, particularly when the premises of the latter are contradictory

If we are interested in satisfying condition 2, why do we need condition 3? Our interest in the criterion of PN is justified by the following observations

- PN is an intrinsic measure of acceptability *(27)*
- Sufficient conditions can be found for satisfying the criterion of PN, whereas only a necessary condition can be found for satisfying the criterion of highest concentration 1
- Keating and Mason's results demonstrate that neither mean square error nor PN should reign exclusively in the comparison of estimators *(27)*

Fixed and Time-Varying Coefficients Approaches

The true functional form of equation 5 is unknown, so a functional form for equation 5 must be assumed The usual practice among econometricians is to presume that for every t, y^*_t follows the reduced-form model

$$y^*_t = x'_t\pi + \epsilon^*_t \quad (t = 0, \pm 1, \pm 2, \quad), \tag{7}$$

with fixed coefficients so that the minimum average mean square error linear predictor of y_{T+s} is

$$x'_{T+s}\pi + w'\frac{V^{-1}}{\sigma^2}(y - X\pi), \tag{8}$$

where $Ey^*_{T+s} = x'_{T+s}\pi$, $Ey^* = X\pi$, $\text{cov}(y^*_{T+s}, y^{*\prime}_t) = w'$ and $\text{var}(y^*) = \sigma^2 V$ are implied by the assumptions underlying model (7) with fixed x_t This predictor has the minimum variance within the class of linear "unbiased" predictors of y_{T+s}, if π in both the terms of the predictor 8 is replaced by $\hat{\pi} = (X'V^{-1}X)^{-1}\ X'V^{-1}y^*$ Several forms of w and V are given in *(25,* chaps 8 and 11) For suitable definitions of x'_{T+s} and X, the predictor

[7]This unbiasedness restriction ensures that both the distribution of $\hat{y}_{T+s}$ and the distribution assumed for y^*_{T+s} are located at the same value so that their variances are comparable It differs from the unbiasedness definition $E_\theta(\hat{\theta}) = \theta$ for all $\theta \epsilon \Theta$, where $\hat{\theta}$ is an estimator of the fixed parameter θ and Θ is the parameter space

[8]The unbiasedness restriction, $E\hat{y}_{T+s} = Ey^*_{T+s}$, is erroneous if the assumed functional form for the mean of y^*_{T+s} is different from the true form

8 also represents the minimum average mean square error linear predictor of an element of a vector variable following a vector autoregressive (VAR) model If equation 7 represents a univariate autoregressive model, then X consists of lagged y's, and w can be equal to 0 The vector w can be zero if equation 7 represents a regression model with a serially uncorrelated error term

The predictor 8 will not give accurate forecasts in the case where the slopes of the function 7 change over time The following model, developed in *(48)* may be appropriate

$$y_t^* = x_t'\Pi z_t + x_t'J\xi_t^* \quad (t = 0, \pm 1, \pm 2, \quad) \tag{9}$$

(For an explanation of these symbols, see (*54*, part II))

When this model is appropriate, the minimum average mean square error linear predictor of y_{T+s} is

$$x_{T+s}'(z_{T+s}' \otimes I)\,\mathrm{vec}(\Pi) + x_{T+s}'J\Phi^s\Sigma_{\xi T}'(I \otimes J')D_x'\Sigma_y^{-}(y - D_xZ_e\,\mathrm{vec}(\Pi)), \tag{10}$$

where the first term equals Ey_{T+s}^*, $Ey^* = D_xZ_e\,\mathrm{vec}(\Pi)$, $\mathrm{cov}(y_{T+s}^*, y^{*'}) = x_{T+s}'J\Phi^s\Sigma_{\xi T}'(I \otimes J')D_x'$, and $\mathrm{var}(y^*) = \Sigma_y$ are implied by the assumptions underlying model 9 with "fixed" x_t and z_t (*54*, p 27) This predictor becomes the minimum variance linear "unbiased" predictor of y_{T+s} if $\mathrm{vec}(\Pi)$ in both the terms of the predictor 10 is replaced by $\mathrm{vec}(\hat{\Pi}) = (Z_e'D_x'\Sigma_y^{-1}D_xZ_e)^{-1}\ Z_e'D_x'\Sigma_y^{-1}y^*$ whenever Σ_y is nonsingular

The model in 9 provides a useful approach for the decomposition of forecast error sources Partition several of the vectors of 9 as follows

$$x_t' = (1, x_{2t}'),\ z_t' = (1, z_{2t}'),\ J = (J_1, J_2)',\ \Pi = (\pi_1, \Pi_2) \tag{11}$$

Model 9 may then be expressed as the sum of terms similar to model 7 and additional terms involving x_{2t} and z_{2t}

$$y_t^* = x_t'(\pi_1, \Pi_2)\begin{pmatrix}1\\ z_{2t}\end{pmatrix} + (1, x_{2t}')\begin{pmatrix}J_1'\xi_t^*\\ J_2'\xi_t^*\end{pmatrix}$$

$$= x_t'\pi_1 + x_t'\Pi_2z_{2t} + J_1'\xi_t^* + x_{2t}'J_2'\xi_t^* \tag{12}$$

An estimated version of the fixed coefficients model 7 implies a forecast of y in some future period s after T, given by $\hat{y}_{T+s}$,

$$\hat{y}_{T+s} = \hat{x}_{T+s}'\hat{\pi} + \hat{\epsilon}_{T+s}, \tag{13}$$

where $\hat{x}_{T+s}'\hat{\pi}$ and $\hat{\epsilon}_{T+s}$ are some estimators or predictors of the first and second terms on the right-hand side of the predictor 8

The forecast error (the difference between the predictor, $\hat{y}_{T+s}$, and the future realization, y_{T+s}) that arises from using a fixed coefficients model when model 12 is true, may be decomposed as

$$\hat{y}_{T+s} - y_{T+s} = \hat{x}_{T+s}'(\hat{\pi} - \pi_1) + (\hat{x}_{T+s}' - x_{T+s}')\pi_1 + (\hat{\epsilon}_{T+s} - J_1'\xi_{T+s}) - x_{T+s}'\Pi_2z_{2T+s} - x_{2T+s}'J_2'\xi_{T+s}, \tag{14}$$

which, in order of appearance, is the sum of (1) a linear combination of the sampling errors of the coefficient estimates, (2) a linear combination of the errors in predicting future values of the independent variables, (3) the error in predicting stochastic shifts in the intercept, (4) the failure to predict deterministic shifts in regression intercept and slopes, and (5) the failure to predict stochastic shifts in regression slopes Except for (2), all these forecast error sources are accounted for when equation 9 is used Observe that an accounting of forecast error sources based on an estimate of equation 7 is limited to (1) and (3) The remaining error sources cannot be diagnosed using fixed coefficients models The error resulting from (2) is, of course, beyond the reach of any of the equations 7 and 9, because it originates from errors in forecasting exogenous events and/or comes from observation, sampling, and measurement deficiencies

One persistent problem in applied economic forecasting has been the recurrence of forecast drifts causing selected model equations to drift away from later historical realizations The conventional add-factoring of intercepts has not always proved satisfactory, especially in cases of suspected nonstationary regression slopes We have shown that one role of the z_2 variables in equation 12 is to account for sources of coefficient nonstationarities Equation 12 accounts for movements in coefficients that are caused by movements in certain observable variables suggested by theoretical considerations but neglected in equation 7 In diagnostic terms, if the z_2 variables are eliminated, then forecast error interpretations are limited because forecast errors cannot be based on errors in predicting deterministic shifts in regression slopes (see equation 14) Equation 12 is useful for distinguishing between errors arising from intercept instability, amenable to add-factor solutions, and errors arising from other sources

The list of potential sources of errors in equation 14 is exhaustive Although add-factoring (the judgmental adjustment of intercepts to realign errant equations to fit current data) has been useful, the exclusive focus on intercept instability, that is $(\hat{\epsilon}_{T+s} - J_1'\xi_{T+s})$, by add-factoring may mean that important sources of forecast error remain unaccounted for An econometric methodology with built-in features for measuring all coefficients variation, as in equation 9, however, could feasibly lend itself to being used as a diagnostic

tool for ascertaining all sources of equation instability, $x'_{T+s}(\Pi_2 z_{2T+s} + J\xi_{T+s})$ An estimate of equation 9 would yield an allocation of the total uncertainty over all components of an equation, as shown in equation 12, permitting quick investigation of the likely sources of future equation volatility

If alternative policy regimes have parametric implications for the behavior of the economy, as suggested by the so-called Lucas critique, a time-varying stochastic coefficients approach may provide a means for anticipating consequences $x'_{T+s}\Pi_2 z_{2T+s}$ of alternative conjectured policy assumptions, not available with conventional fixed coefficients techniques, whenever the z_{2t} elements include observable policy variables

Without a doubt, equation 7 is simpler to work with than model 9 because the second and fourth terms appearing on the right-hand side of equation 12 are ignored in equation 7 Even though including these terms complicates our models and possibly makes our parameter estimates imprecise and nonunique, we have no choice except to include them if model 7 does not give useful forecasts No logical principle warrants excluding these terms because no one knows for sure that these terms are absent from the true model We later show why the prediction principle advocated by Zellner (*61*, p 32) and others cannot conclusively reject equation 9 in favor of equation 7

Estimation Procedures

The predictors 8 and 10 are not operational because they involve unknown parameters To obtain computable forecasts, we need the estimates of these parameters The vector π or vec(Π), if fixed, can be estimated by one or more of the following procedures

- The least squares procedure,
- The generalized least squares procedure based on an estimated error covariance matrix,[9]
- A fully or partially restricted reduced-form procedure that fully or partially accounts for the connection between π and the coefficients of a structural model,
- A Bayes procedure,
- Shrinkage estimators, and
- Robust procedures

[9]Swamy and Tinsley's estimate of the error covariance matrix for model 9 may be singular (*48*), in which case Paige's numerically stable and efficient algorithm based on matrix decompositions should be used for estimating model 9 (*31*)

The corresponding methods of estimating w, V, σ^2, and the variances and covariances in equation 10 are also available Several methods of estimating w, V, and σ^2 are summarized in (*25*) and a method of estimating variances and covariances in equation 10 appears in (*48*) [10]

Swamy and Schinasi show that, if all the unknown parameters in equations 8 and 10 are replaced by their respective sample estimates, then we cannot in practice recognize an operational "unbiased" predictor with minimum variance in small samples (*51*) They also show that a universally preferred choice among different estimation procedures for equations 7 and 9 is not possible based on either the exact finite sample distribution theory or the asymptotic theory

Akaike's Information Criterion

Akaike has derived from information-theoretical considerations a probability density function (pdf) which may be expected to approximate the true pdf for a variable (*1*, *2*) The criterion he has used to find this approximation is

$$B(p, g) = -\int \frac{p(y)}{g(y)} \log\left[\frac{p(y)}{g(y)}\right] g(y)dy, \qquad (15)$$

where p(y) is the true pdf for a variable y^*, g(y) is an approximation to p(y), and the integration is over the entire range of y^* Clearly, this criterion can be written as

$$B(p, g) = \text{Elogg}(y) - \text{Elogp}(y) \leq 0, \qquad (16)$$

where the expectation is with respect to the true distribution of y^*

Because the quantity on the right-hand side of equation 16 is nonpositive, when $\int_{-\infty}^{\infty}[p(y) - g(y)]dy \geq 0$, as shown by Rao (*36*, p 59), the greater the value of Elogg(y) is, the closer the pdf g(y) is to the true pdf p(y) in the sense of B(p,g) However, the statement that the unknown true pdf, p(y), can be well approximated by g(y) if and only if g(y) maximizes Elogg(y) is useless as it stands Deciding whether the condition is or is not satisfied or taking the expectation of logg(y) with respect to p(y) is impossible without knowing the family of pdf's which covers the true pdf for y^* as a special case The maximum likelihood method is applied to a family of pdf's for this reason, which

[10]Swamy and Tinsley's (*48*) method of estimating variances and covariances extends Swamy's earlier work (*45*, *46*), which does considerably more than Chow's (*7*, p 340, *8*, p 1,237) perfunctory description of it as a "survey " Reinsel (1982, 1984) also presents estimators and predictors (*39*, *40*) He, however, basically repeats the results recorded earlier in the above papers (*47*)

presumably covers the true pdf as a special case The strict inequality in equation 16 is an important step in proving the consistency of maximum likelihood estimators (*29,* p 891) The only explicit statements about the interpretation of a pdf like p(y) in criterion 15 that we have found are in the applications of criterion 15, where p(y) is thought of as the pdf of the unknown true distribution Does this mean that distributions which do not possess pdf's cannot be true? A singular normal distribution does not possess a pdf except on a subspace Prior distributions satisfying Shiller's smoothness restrictions do not possess pdf's on the entire parameter space (*26, 57*) [11] Even though stable distributions have pdf's, these pdf's are generally expressed only as infinite series, which are not easy to work with Any of these distributions can be true We should not say, then, that p(y) in criterion 15 is the pdf of the unknown true distribution If p(y) is restricted *a priori* to belong to a particular family of pdf's, then criterion 15 may have the same defects as the maximum likelihood criterion (*53*, p 8) For example, if we assume that p(y) belongs to the family of pdf's implied by a mixed autoregressive, moving average model of finite but unknown order, then criterion 15 does not lead to consistent estimates of the order unless it is modified, as in (*22*) (See *(43)*)

Swamy and von zur Muehlen have developed some sufficient conditions for the existence of different families of distributions *(52)* Logic permits us to say only that these families are true if their sufficient conditions are true But, no one can determine the truth of these sufficient conditions, assuming they are coherent Our beliefs about these sufficient conditions may be expressed as subjective probabilities, which may then be transformed consistently into subjective probabilities on individual distributions (*52*) The defect of criterion 15 is that it is unable to take into account such probabilities

A justification of criterion 15 rests on the belief that the entropy of a distribution is a good measure of uncertainty Copas shows this belief is not correct in nonnormal cases by way of an example where a company is operating under much greater uncertainty in one of the two cases, though the entropies of distributions in the two cases are exactly the same (*15*) Copas wrote that this result arises as a direct consequence of the fact that the entropy of a distribution depends only on the distribution of the different heights of its pdf, paying no attention to the values of the variable at which these various heights are attained Entropy can be, therefore, a very imperfect measure of statistical uncertainty B(p, g) should not be used as a measure of the distance between g(y) and p(y) for this reason, regardless of any knowledge of the context and restrictions it puts on the shape of distribution one finds attractive

A possible alternative reaction is to note that, when some conditions are satisfied, equation 15 provides useful forecasts A set of such conditions is provided by Shibata (*44*) He has proved that if p(y) is determined by an autoregressive process of infinite order and if g(y) is determined by an autoregressive process of finite order K(<T)(AR(K)), where the order K is selected so as to maximize Elogg(y), or some other modification of Elogg(y), then an asymptotic lower bound is attained in the limit for the average mean square error of an estimated conditional mean of AR(K) This result is a large sample analog of the exact finite sample result that the conditional mean of AR(K) is a minimum average mean square error predictor if AR(K) is the true model The key assumption used by Shibata is that the order of the autoregression determining the true pdf is infinite Statisticians who believe in the principle of parsimony or simplicity assign to such an assumption the zero probability of being true (See *(43)*)

It is difficult to determine whether or not Shibata's demonstration constitutes an argument against the information criterion 15, or against the principle of parsimony or simplicity, or against de Finetti's (*16*) condition of finite additivity In any case, autoregressive models of finite or infinite order clearly ignore sources (4) and (5) of forecast errors described in equation 14 One cannot be sure that these sources are absent in any forecasting situation Garcia-Ferrei, Highfield, Palm, and Zellner's results showed that autoregressive models of order 3 for annual real output growth rates of nine countries did not generally result in lower root mean square forecast errors relative to naive models, so relying solely on Shibata's theoretical result is difficult (*17*)

Importance of Comparing Different Predictors

Oakes has proved that no universal algorithm guarantees accurate forecasts forever, so any attempt to prescribe a single forecasting procedure, applicable to all empirical situations, must be unsatisfactory (*35*) No agreement of the values taken by a predictor (based solely on the data known up to the current period) with the actuals for a finite past time period could possibly imply that the values of the predictor would agree with the actuals in the future Past success does not guarantee future success If we knew only that a predictor had produced accurate forecasts in a past period, we could not guarantee that any future

[11]There are applications of equation 15, where there is no mention of these points

forecasts generated by the predictor would be sufficiently accurate, because some predictors exist for which the initial values do not control the future values

For this reason, de Finetti set up minimal criteria that forecasts should be coherent based on data currently available (*16*) One predictor is as valid as any other, if they all satisfy the ,requirements for coherence based on what knowledge is available A predictor that conforms to probability calculus or does not violate any of the probability laws is coherent This means only that de Finetti's concept of coherence prohibits the use of any contradictory restrictions or premise that is inconsistent relative to the axioms of probability theory For example, if the premises of equation 7 are contradictory, then we cannot obtain coherent forecasts by using that model

If a forecaster must choose one predictor from among several coherent ones, a likely choice would be the one with the best track record The forecast can represent, at most, a measure of the confidence with which one expects that predictor to forecast an event based on currently available evidence, and not based on information yet to be observed Obtaining useful codification of statistics that yields a satisfactory predictor selector for all people in all settings is impossible Each experimenter must choose among various coherent predictors by comparing their past forecasting performance

A Coherent Approach to Prediction

We consider a Bayesian solution to the problem of finding the entire predictive distribution Jeffreys' book (*24*) is mainly responsible for the following Bayesian approach in Zellner's (*60*, pp 306-17) and Geisel's (*18*) seminal work on comparing models Given our beliefs in the form of a finite set of exhaustive and mutually exclusive models, $M_1, M_2, \ldots, M_n$, about the process that has generated the values of the variable y^*, we can compute the marginal probability density function (pdf) for y^* implied by the ith model by

$$p(y \mid M_i) = \int_{R_{\theta_i}} p(y \mid \theta_i, M_i) p(\theta_i \mid M_i) d\theta_i, \qquad (17)$$

where θ_i is the vector of parameters appearing in M_i, $p(y \mid \theta_i, M_i)$ is the conditional pdf for y^* given θ_i and M_i, $p(\theta_i \mid M_i)$ is the prior pdf for θ_i, and R_{θ_i} is the range of θ_i Let $pr(M_i)$ denote the prior probability of M_i being true When a particular value of the random variable y^*, say y, is observed, we may employ Bayes' theorem to revise the prior probability $pr(M_i)$ to become the posterior probability, that is

$$pr(M_i \mid y) = \frac{pr(M_i) p(y \mid M_i)}{\sum_{i=1}^{n} pr(M_i) p(y \mid M_i)},$$

$$= \frac{pr(M_i) p(y \mid M_i)}{p(y)} \quad (i = 1, 2, \ldots, n) \qquad (18)$$

We may derive the predictive pdf by

$$p(y_{T+s} \mid y) = \sum_{i=1}^{n} pr(M_i \mid y) p(y_{T+s} \mid M_i, y), \qquad (19)$$

where

$$p(y_{T+s} \mid M_i, y) = \int_{R_{\theta_i}} p(y_{T+s} \mid \theta_i, M_i) p(\theta_i \mid M_i, y) d\theta_i$$

The denominator of the ratio on the right-hand side of equation 18 is not equal to the unconditional pdf for y^* unless $M_1, M_2, \ldots, M_n$ are mutually exclusive and exhaustive We did not violate any probability laws in deriving equation 19 In this sense, the predictive pdf, equation 19, is coherent More important, equation 19 gives a coherent method of pooling the predictive pdf's given by different competing models of the same data-generating process, as long as the premises of any of these models are not contradictory

If we use only one model, say M_1, and do not use all other models to generate the predictive pdf for y^*_{T+s}, then we set $pr(M_1) = 1$ and $pr(M_i) = 0$ for $i \neq 1$ For these values of $pr(M_i)$, it is obvious from 19 that $p(y_{T+s} \mid y) = p(y_{T+s} \mid M_1, y)$ Formulas 8 and 10 are based on the assumptions that pr(model 7) =1 and pr(model 9) = 1, respectively These assumptions are false if we view models 7 and 9 as approximations to the true model because any approximately true model is neither absolutely true nor absolutely false Because we do not know of any models that are literally true, down to the last decimal point, some analysts feel that all models are false Boland says this opinion is a self-contradiction (*5*, p 179) We do not believe that self-contradiction is consistent with Bayesian *coherent* behavior If we truly believe that all the models $M_1, M_2, \ldots, M_n$ considered in equation 18 are indeed false, then as coherent Bayesians, we should be saying that $pr(M_i) = 0$ for $i = 1, 2, \ldots, n$ Otherwise, we would be contradicting ourselves If $pr(M_i) = 0$ for $i = 1, 2, \ldots, n$, then formula 18 is indeterminate Our models can be true if we satisfy the necessary condition of logical validity, although we cannot establish their truth status

Swamy and von zur Muehlen discussed probabilistic logic as a valid tool for scientific analysis and interpretation of causal relationships (*52*) This logic can be used to merely bound (rather than specify) pr(M_i), if we have some beliefs about the sufficient conditions under which M_i is true Thus, scientific beliefs are useful in quantifying pr(M_i) The prior pdf's for θ_i must also be consistent with these beliefs In this sense, pr(M_i) is related to $p(\theta_i \mid M_i)$ Because we do not know of any model that is literally true, we should assign positive probabilities to more than one logically valid model This assignment is warranted by the frequent disagreement among economists as to which model is superior to address a given issue If any consensus that ignores all but one of the opinions expressed is not satisfactory, then it is reasonable to have more than one model with a positive probability of being true The problem with the predictive pdf 19 is that coming up with an exhaustive and mutually exclusive set of models is difficult The prior pdf's $p(\theta_i \mid M_i)$, $i = 1,2,\ \ ,n$, which were selected based on considerations of mathematical convenience, may not be consistent with the values assigned to pr(M_i), $i = 1,2,\ \ ,n$, and may not represent anybody's beliefs If we prefer model 7 to model 9 because the Bayesian analysis of model 7 is simpler than the Bayesian analysis of model 9, then our inferences are incoherent if the premises of model 7 are contradictory

Linkage

Suppose that we have two different econometric models giving two different predictions of an unknown value y_{T+s} We do not know which one to choose because we do not know which one of these two predictions will be closer to the actual value y_{T+s} This is not unusual in economics We will have more than two models giving us more than two predictions about the same value If these models are not mutually exclusive and exhaustive, then we cannot use the previously discussed Bayesian approach However, we can use the following non-Bayesian approach under certain conditions

Let $\hat{y}_{1,T+s}, \ \ ,\hat{y}_{m,T+s}$ be the "unbiased" predictors of y_{T+s} given by m different econometric models, and assume that we have reason to believe that the expected squared deviation of $\hat{y}_{1,T+s}$ from y^*_{T+s} is smaller than that of any $\hat{y}_{j\,T+s}$ for $j = 2,\ \ ,m$ Let $\tilde{y}_{T+s} = (\hat{y}_{2,T+s}, \hat{y}_{3\,T+s}, \ \ ,\hat{y}_{m\,T+s})'$ and let $\iota = (1,1,\ \ ,1)'$ be an $(m-1)\times 1$ vector of unit elements Suppose that $\hat{y}_{1\,T+s}$ is correlated with the predictor $(\tilde{y}_{T+s} - \iota\hat{y}_{1,T+s})$ with zero expectation Then there exists an $(m+1)$th predictor whose expected squared deviation from y^*_{T+s} is smaller than that of $\hat{y}_{1\,T+s}$

This $(m+1)$th predictor is

$$\hat{y}_{m+1\,T+s} = \hat{y}_{1,T+s} - \text{cov}(\hat{y}_{1,T+s}, (\tilde{y}_{T+s} - \iota\hat{y}_{1\,T+s})') \{\text{var}(\tilde{y}_{T+s} - \iota\hat{y}_{1,T+s})\}^{-}(\tilde{y}_{T+s} - \iota\hat{y}_{1,T+s}), \quad (20)$$

where all the variances and covariances are about y^*_{T+s} and for any matrix A, A^- denotes a generalized inverse of A The predictor 20 results from making covariance adjustment in $\hat{y}_{1,T+s}$ with respect to the concomitant variable $(\tilde{y}_{T+s} - \iota\hat{y}_{1\,T+s})$ with zero expectation (*37*, p 359)

However, it is doubtful that the predictors $\hat{y}_{1,T+s}$ and $\tilde{y}_{T+s}$ given by different models suffering from different types of specification errors will be "unbiased" If $E\hat{y}_{1,T+s} \neq E\hat{y}_{j,T+s}$ for $j = 2,\ \ ,m$, which seems likely, then the expected squared deviation of (20) from y^*_{T+s} will not be smaller than that of $\hat{y}_{1,T+s}$ because the predictor $(\tilde{y}_{T+s} - \iota\hat{y}_{1,T+s})$ with nonzero expectation is not a concomitant variable suitable for making covariance adjustment in $\hat{y}_{1\,T+s}$, even when $\hat{y}_{1\,T+s}$ is "unbiased" If we use sample estimates of the variances and covariances in place of their known values used in equation 20, then we are no longer in a position to claim that the expected squared deviation of 20 from y^*_{T+s} is always smaller than that of any $\hat{y}_{j\,T+s}$ (*37*, p 360) The predictor 20, based on estimated variances and covariances will be incoherent if the estimates violate any of the assumptions under which the constituent predictors, $\hat{y}_{j\,T+s}$'s, are derived

The difficulties presented by equation 20 are not encountered if we use equation 9 alone Because equation 9 is a general model covering various fixed coefficients models as special cases (*54*) we can justify using this general model for predicting and abandoning the method of pooling the predictions of different fixed coefficients models, particularly when the premises of the fixed coefficients models are contradictory

Stability Tests

Some econometricians would like to see some evidence against the stability of the coefficients of equation 7 before they admit that a version of equation 9 deserves their consideration Stability tests are supposed to give such evidence A brief description of these tests appears in (*30*, pp 575-8) Based on our discussion of Birnbaum's confidence concept in Part I (*4*, *53*), a full disclosure of statistical evidence takes the form d_1^* = (reject H_0 in favor of H_1, α_I, β_{II}) or d_2^* = (reject H_1 in favor of H_0, α_I, β_{II}), where H_0 = a null hypothesis, H_1 = an alternative hypothesis, α_I = the probability of type I error, and β_{II} = the probability of type II error Can we come up with such disclosures about coefficients' stability?

We divide the available time series of length T on variables in equation 7 into G mutually exclusive subperiods, with m_1 observations in the first subperiod, and m_2 observations in the second subperiod, so that $\min(m_1, m_2, \quad, m_G) > K$ Note that $\Sigma_{i=1}^{G} m_i = T$ Assuming that the coefficient vector π varies between subperiods but not within each subperiod, we can depict the observations as

$$\begin{bmatrix} y_1^* \\ y_2^* \\ \\ \\ y_G^* \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ & & \\ & & \\ 0 & 0 & X_G \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \\ \\ \pi_G \end{bmatrix} + \begin{bmatrix} \epsilon_1^* \\ \epsilon_2^* \\ \\ \\ \epsilon_G^* \end{bmatrix}, \tag{21}$$

or more compactly as

$$y^* = X\pi + \epsilon^*, \tag{22}$$

where for $i = 1,2, \quad, G$, y_i^* is a $m_i \times 1$ vector of observations on the dependent variable, X_i is a $m_i \times K$ matrix of rank K of observations on K independent nonstochastic variables, π_i is a $K \times 1$ vector of regression coefficients, ϵ_i^* is a $m_i \times 1$ vector of stochastic disturbances, $y^* \equiv [y_1^{*\prime}, y_2^{*\prime}, \quad, y_G^{*\prime}]'$, $\pi \equiv [\pi_1', \pi_2', \quad, \pi_G']'$, $\epsilon^* \equiv [\epsilon_1^{*\prime}, \epsilon_2^{*\prime}, \quad, \epsilon_G^{*\prime}]'$, and X represents the block-diagonal matrix on the right-hand side of equation 21 The vector ϵ^* is assumed to have a normal distribution with mean zero and the covariance matrix Σ

The null hypothesis of coefficient stability can be stated as

$$H_0 \quad \pi_1 = \pi_2 \quad = \pi_G, \tag{23}$$

which can be expressed as

$$R\pi = \begin{bmatrix} I & -I & 0 & 0 & 0 & 0 \\ 0 & I & -I & 0 & 0 & 0 \\ & & & & & \\ & & & & & \\ 0 & 0 & 0 & I & -I & 0 \\ 0 & 0 & 0 & 0 & I & -I \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \\ \\ \pi_{G-1} \\ \pi_G \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \\ \\ 0 \\ 0 \end{bmatrix}. \tag{24}$$

The statistics literature states that under the null hypothesis (23), the statistic, (T−GK)/(G−1)K times

$$\frac{y^{*\prime}\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}R'[R(X'\Sigma^{-1}X)^{-1}R']^{-1}R(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y^*}{y^{*\prime}\Sigma^{-1}y^* - y^{*\prime}\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y^*}, \tag{25}$$

is distributed as F with (G − 1)K and T − GK degrees of freedom (*59*)

Suppose that we use statistic 25 to test the hypothesis 23 against the alternative hypothesis,

$$H_1 \quad R\pi \neq 0, \tag{26}$$

and come up with the decision,

$$(\text{reject } H_0 \text{ in favor of } H_1, \alpha_I, \beta_{II}) \tag{27}$$

If the values of α_I and β_{II} in the statement 27 are sufficiently small, then the statement 27 provides strong but inconclusive evidence against coefficients' stability The values of α_I and β_{II} in the statement 27, however, depend on the values of R and Σ used If we use an incorrect value of either R or Σ, then the values of α_I and β_{II} will be incorrect and the evidence of the statement 27 will be misleading Even if our assumptions about the forms of R and Σ are correct, and if we use a sample estimate of Σ in place of its known value used in the statistic 25, then we may not know the exact distribution of the statistic 25 If we use an asymptotic distribution of the statistic 25 to evaluate α_I and β_{II}, then due to the approximate nature of the values of α_I and β_{II}, the evidence in the statement 27 may be misleading It is also possible that for some sample estimates (or *a priori* values) of Σ, $\alpha_I \geq 0.5$ and $\beta_{II} \geq 0.5$ In that case, the evidence in the statement 27 is worthless

Other difficulties arise because 26 is not the only alternative hypothesis of interest There is no guarantee that the coefficients of model 7 do not change in periods other than those specified by the hypothesis 26 Model 9 is appropriate if the alternative hypothesis that π changes at any or every t is true Under this realistic alternative, the statistic 25 is not defined Little reason exists to use the test statistic 25 if we want to test the null hypothesis 23 against this realistic alternative hypothesis To divide a time series into event-conditioned subperiods as in equation 21, econometricians must have considerable knowledge of their data One cannot be content with casual inspection of a few stereotyped measures, such as F values, as is common practice in much applied econometric work

An alternative to the statistic 25 is the CUSUM or CUSUM-square statistic of Brown, Durbin, and Evans (*30,* pp 576-8) These alternative statistics are based on recursive residuals which are not unique We can get T − K nonzero and K zero recursive residuals for model 7 (where T is the number of observations and K is the number of independent variables) We get different T − K recursive residuals depending on which K of the T residuals are set equal to zero Computing these residuals usually means Σ is arbitrarily set equal to $\sigma^2 I$ Therefore, the values of

α_I and β_{II} for the test of the null hypothesis in 23 against the alternative hypothesis that the coefficient vector of equation 7 changes at some unknown periods based on the CUSUM (or CUSUM square) statistic depend on the value of Σ employed and also on which K of the T residuals are set equal to zero For this reason, two different econometricians working with two different recursive residuals for the same model and data can come up with two different pairs of values of (α_I, β_{II}) for the CUSUM (or CUSUM-square) test These pairs of values may give contradictory conclusions It is also clear that the CUSUM (or CUSUM-square) test cannot detect shifts in coefficients in any period if we set the recursive residuals of that period equal to zero This discussion and the discussion in the previous paragraphs show that the stability tests are not informative and can be misleading By contrast, we can conclude that the coefficients of equation 7 are unstable if using equation 9 produces a noticeable and important improvement in forecasting performance relative to that of equation 7 [12]

Even in large samples, the CUSUM (or CUSUM-square) test does not give correct conclusions because, under the alternative hypothesis that π in 7 changes at some unknown periods, in some unknown manner, the power, $(1 - \beta_{II})$, of this test does not tend to 1 as the sample size tends to ∞ The basic difficulty is that the time-varying coefficients of equation 9, $\Pi z_t + J\xi$, are not consistently estimable (See the uncertainty principle formulated by Swamy and Tinsley (*48*, p 117))The seductive danger of stability tests is that they pretend to a kind of relevance which their logical machinery cannot justify

Some Applications

The authors have employed stochastic coefficients models before to forecast several economic variables for several time periods The stochastic coefficients model they employed can be represented by

$$y_t^* = x_t'\pi_1 + x_t'\epsilon_t^*, \tag{28}$$

where the $K \times 1$ vector ϵ_t^* is assumed to follow a first-order autoregressive vector process Equation 28 is a restricted version of equation 9 obtained by zeroing the vector z_{2t} and setting $J\xi_t^*$ equal to ϵ_t^*

[12]It is fair to point out that, like recursive residuals, forecasts based on asymptotically efficient estimators of the parameters of equations 7 and 9 may also be arbitrary when the asymptotic efficiency is defined as in Lehmann (*34*, p 415) For example, when the solution of the likelihood equations for the parameters of equation 7 or 9 is not unique, asymptotically efficient likelihood estimators such as the one step estimator suggested by Lehmann (*34*, p 435) depend on a somewhat arbitrary initial estimator and need no longer agree with the maximum likelihood estimator even for large samples However, the arbitrariness of asymptotically efficient estimators is of a very different nature from those of recursive residuals

The reasons for considering equation 28 include

- Employing model 28, though restrictive, is more general than model 7
- Estimating model 28 does not need as much (magnetic) core on the computer as estimating model 9 needs
- Comparing the forecasts of models 7, 28, and 9 shows whether proceeding in order of increasing complexity increases the accuracy of forecasts
- Working with model 28 has computational benefits which produces fewer unknown parameters than model 9

A natural approach to investigating the advantages and disadvantages of equation 28 is to apply Zellner's prediction principle (*61*, p 32) Split the available time series into two nonoverlapping parts The period of the first part is called the estimation (or fitting) period and the period of the second, the forecast period Let $t = 1,2, \ldots, T$ be the fitting period and let $t = T+1, T+2, \ldots, T+n$ be the forecast period Estimate model 28 and its fixed coefficients counterpart by using the first part and then use these estimated models along with the values of the independent variables for the forecast period to predict the values of the dependent variable for the forecast period without revising the parameter estimates We call such forecasts nonsequential or *multi-step-ahead* [13]

The authors have used Swamy and Tinsley's (*48*) method described in Part II of this article (*54*) to estimate equation 28 The maximum likelihood estimates of the parameters of equation 28 may not exist, and Swamy and Tinsley's iterative method of estimating these parameters is not guaranteed to converge, so we do not iterate on Swamy and Tinsley's method until convergence (*48*) Because equation 28 fits the sample values perfectly whenever an estimate of $x_t'\pi$ is added to the corresponding prediction of $x_t'\epsilon_t^*$,

[13]It is obvious that this procedure is not operational if the values of the independent variables for the period T+s are not available at the time of forecasting y_{T+s} Since our purpose is to obtain separate estimates of the terms on the right hand side of equation 14, we have to use these values of independent variables Without separating the second of these terms from the rest, it is not possible to evaluate forecast errors arising from coefficients' instability Forecasters are also interested in knowing the magnitudes of each term on the right hand side of equation 14 Thus, we are solving here a problem which is more important than a practical forecasting problem

If we estimate sequentially the fixed parameters using all past data prior to each of the forecast periods, $T+1, T+2, \ldots, T+n$, then we call the corresponding forecasts sequential or one step ahead The primary purpose of Swamy and Schinasi's article (*51*) is to demonstrate that the one step ahead forecasts will not necessarily be closer to the realized values of the forecasted variable than the multi step-ahead forecasts There is no non-Bayesian theory which mandates prediction with sequential estimation

measures of within-sample fits are useless to discriminate among the estimates obtained at different iterations of Swamy and Tinsley's procedure To avoid overfitting, we choose estimates which minimize the root mean square forecast error

$$[\frac{1}{n}\sum_{s=1}^{n}(\hat{y}_{T+s} - y_{T+s})^2]^{1/2},$$

where $\hat{y}_{T+s}$ is a forecast of y_{T+s} in some period s after the terminal period T of the fitting period The root mean square forecast error is a generally good substitute for an averaged within-sample residual sum of squares From L iterations of the Swamy and Tinsley procedure, we obtain L different estimates of the unknown parameters in model 28 Inserting these estimates into formula 10 furnishes L different predictions for each forecast period We show these predictions by $\hat{y}_{T+s,l}$, s = 1,2, ,n, l = 1,2, ,L These predictions give L different values for the root mean square forecast error $[\frac{1}{n}\sum_{s=1}^{n}(\hat{y}_{T+s,l} - y_{T+s})^2]^{1/2}$, l =1,2, ,L We select the estimates of the unknown parameters by minimizing these L values [14] We use these minimum root mean square forecast error estimates to forecast the values of y_t^* beyond the period T+n The sample beyond the period T+n is used to compare the forecasting performance of an estimated stochastic coefficients model with those of other models

To estimate the fixed coefficients counterpart of equation 28, Swamy and his co-authors considered both the classical least squares procedure and an approximate generalized least squares procedure based on a sample estimate of the error covariance matrix They also applied approximate Bayes and ridge-type shrinkage estimators to equation 7 with and without serial correlation in the error term These estimates allow us to evaluate the corresponding minimum average mean square error predictors for the forecast period, T+1, T+2, ,T+n, and the root mean square forecast errors of these predictions are computed We obtain one root mean square forecast error for each fixed coefficients estimator We choose the estimates of the fixed coefficients corresponding to the smallest of these root mean square forecast errors

Table 1 shows the results of the authors' computations Use of time-varying, stochastic coefficients modeling may substantially reduce out-of-sample forecast errors, similar to reductions obtained in several earlier empirical applications of the stochastic coefficients models listed in table 1

The results in table 1 generally turned out favorably to the stochastic coefficients models because the minimum average mean square error predictors corresponding to these models are evaluated at their respective minimum root mean square forecast error estimates of unknown parameters These results cannot be reproduced using any arbitrary *a priori* values of parameters This statement not only elaborates upon footnote 14 but also explains why Alexander and Thomas (*3*) and Wolff (*58*) find that the forecasts of exchange rates generated by the Kalman filter with *a priori* values of parameters are poor relative to the forecasts of random walk models [15]

We used equation 28 to estimate an agricultural investment model (see *10* for a complete discussion of the model) Investment is assumed to be generated by a linear version of the flexible accelerator where

$$K_t - K_{t-1} = b_t + b_{wt}W_t + b_{ut}U_t + B_tK_{t-1}, \qquad (29)$$

and W is the ratio of input to output price, K is the capital stock, and U is the implicit rental rate of capital

We compared the usefulness of the stochastic coefficients specification's forecast accuracy with six other models in five-period out-of-sample tests Although the six alternatives do not exhaust all possible models, they help evaluate the predictive capability of the stochastic coefficients investment model

One of the six models is the fixed coefficient analogue of the stochastic coefficients investment model Net investment is regressed on a constant, an input/output price ratio, a rental rate, and lagged capital stock Two other models are variants of the fixed coefficients model One model includes net farm income (income) as a regressor, the other includes a time trend (time) The fourth model is a fixed coefficient, nonlinear flexible accelerator, where the adjustment parameter is a function of the ratio of input to output prices and the rental rate The final two models are atheoretical Investment is assumed to be a stochastic process following both a first-order autoregressive AR(1) process and a second-order autoregressive AR(2) process

Table 2 shows that the stochastic coefficients model is the superior predictor However, an unambiguous indicator of forecast accuracy does not exist Each indicator has its own risk function For example, a mean absolute error criterion is based on an absolute deviation loss function, while a mean square error criterion is based on a quadratic loss function Therefore, different analysts may prefer different models, depending on their assumed loss function Considering a wide variety

[14]The arbitrariness of these estimates is less harmful in terms of the accuracy of forecasts they lead to than the arbitrariness of prior distributions

[15]For examples of the use of stochastic coefficients models in a policy simulation framework, see (*9, 11, 14*)

Table 1—Out-of-sample root mean square forecast errors of stochastic coefficients and fixed coefficients estimators[1]

Source	Dependent variable	Fitting period[2]	Forecast period	Stochastic coefficients	Fixed coefficients[3]	Random walk[4]	Improvement over best alternative
							Percent
Conway, Hallahan, Stillman, and Prentice (1987)	Beef retail price (U S)	1968 Q1–1979 Q4	1980 Q1–1983 Q4	7 29	14 00	—	48
	Pork retail price (U S)	1968 Q1–1979 Q4	1980 Q1–1983 Q4	5 83	2 91	—	−100
	Broiler retail price (U S)	1968 Q3 1979 Q4	1980 Q1–1983 Q4	3 78	6 02	—	37
Conway and Gill (1987)	Fixed weight GNP inflation rate (U S)	1960 Q1 1980 Q4	1981 Q1 1984 Q4	200	395	—	49
LeBlanc, Kitchen, and Conway (1988)	Exchange rate U S $/Canada $	1975 9–1985 2	1985 3–1985 7	036	048	0 049	25
Swamy, Kennickell, and von zur Muehlen (1986)	M1 aggregate (U S)	1960 Q1–1982 Q2	1982 Q3–1985 Q2	4 644	19 804	—	77
Swamy and Tavlas (1989)	Monetary base (Australia)	1967 Q1–1984 Q3	1984 Q4–1985 Q4	562[5]	534	1 767	−5
Swamy and Tavlas (1989)	Monetary base (Australia)	1967 Q1–1984 Q3	1986 Q1–1987 Q2	883[5]	960	978	8
Swamy and Tavlas (1989)	M1 aggregate (Australia)	1967 Q1–1984 Q3	1984 Q4–1985 Q4	1 922[5]	2 109	2 716	9
Swamy and Tavlas (1989)	M1 aggregate (Australia)	1967 Q1–1984 Q3	1986 Q1–1987 Q2	939[5]	1 159	1 651	19
Swamy and Tavlas (1989)	M3 aggregate (Australia)	1967 Q1–1984 Q3	1984 Q4–1985 Q4	659[5]	1 173	2 706	44
Swamy and Tavlas (1989)	M3 aggregate (Australia)	1967 Q1–1984 Q3	1986 Q1–1987 Q2	774[5]	1 555	1 137	32
Swamy, Kolluri, and Singamsetti (1988)	Treasury bill rate (U S)	1960 Q1–1983 Q4	1984 Q1–1986 Q4	411	585	658	30
Swamy and Schinasi (1986)	Stock prices	1900–73	1974–83	680	877[6]	1 801	22
Schinasi and Swamy (1987)	Exchange rate dollar/pound	1973 3–1980 3	1980 4–1981 6	2 170	3 540	3 030	28
	dollar/yen	1973 3–1980 3	1980 4–1981 6	3 270	4 030	3 960	17
	dollar/deutschmark	1973 3–1980 3	1980 4–1981 6	2 170	2 560	3 690	15
Schinasi and Swamy (1988)	G-10 weighted average dollar	1975 1–1982 12	1983 1–1984 12	2 009	2 181	2 056	2
Schinasi and Swamy (1988)	G-10 weighted average dollar	1976 1–1983 12	1984 1–1985 12	2 618	2 715	2 873	4
Schinasi and Swamy (1988)	G-10 weighted average dollar	1973 1–1984 12	1985 1–1986 12	2 311	2 315	2 712	0

[1]Square root of an average of sum of squared deviations multiplied by 100
[2]Numbers shown after years signify either quarters (Q) or specific months
[3]Forecasts based on the best predicting estimates
[4]$y_t = y_{t-1}$ + white noise
[5]Forecasts of the same variable for these two different forecast periods are based on identical parameter estimates
[6]Sequential forecasts

Table 2—Out-of-sample net investment forecast, 1981-85

Year	Actual	Stochastic coefficients	Fixed coefficients	Income	Time	Flexible accelerator	AR1	AR2
			Million dollars (1972)					
1981	−993	−385	354	623	602	308	610	566
1982	−2,017	−1,169	321	−359	987	120	543	500
1983	−1,962	−1,369	343	−818	1,349	135	502	472
1984	−1,815	−1,359	449	614	1,764	157	478	460
1985	−2,104	−1,845	33	198	1,580	169	463	460

of forecast and other criteria, including goodness of fit and tracking measures, is preferable

Table 2 presents each model's forecasts for 1981-85 The forecast statistics, based on years with dramatic declines in agricultural investment, provide an excellent test of forecast accuracy The absolute error shows that the stochastic coefficients model dominates the fixed coefficients models each year After missing the actual value by a relatively wide margin in 1982 ($849 million), the stochastic coefficients' forecast improves through 1985, where the absolute error is $258 million The evaluation statistics for each model are mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and Theil's U2 coefficient Table 3 shows that the stochastic coefficients model is the most accurate out-of-sample forecaster The mean absolute error statistic (MAE) is representative of the stochastic coefficients' dominance over its competitors The nearest competitor, the flexible accelerator model, is more than three times greater in MAE The stochastic coefficients model outperforms the other six models for nearly any sensible risk function

Conclusions

By accepting that the aim of inference is to generate predictions for future observables, we can see that the problem of comparing alternative model specifications is resolved by comparing the accuracy of predictions the models generate and choosing the model that predicts best Experience with such comparisons suggests that allowing all coefficients in an economic relationship to vary over time may contribute to improved forecasts The economics literature has long recognized that slopes of economic relationships may not be constant through time because of aggregation effects and policy changes Therefore, the assumption of time-varying coefficients cannot be so easily dismissed on the grounds that, when coefficients vary, "the concept of seasonal adjustment can become rather confused" (*21*, p 1,014), or that increasing the complexity of the models used to generate predictions does not necessarily lead to better predictions

Table 3—Forecast evaluation statistics[1]

Model	MAE	MAPE	RMSE	Theil's U2
Stochastic coefficients	533	34	1 89	0 89
Fixed coefficients	2,297	133	2 17	1 03
Fixed coefficients with income	2,269	131	2 17	1 02
Fixed coefficients with time	2,078	119	2 15	1 02
Flexible accelerator	1,829	109	2 13	1 00
AR1	3,034	170	2 24	1 06
AR2	1,956	112	2 14	1 01

[1]Mean value for net investment during 1981-85 is $1 78 billion

Swamy and Tinsley's minimum root mean square forecast error estimates will also be useful in assessing Bayesian prior distributions (*48*) If these estimates imply a distribution for the coefficients, which is more general than a Bayesian prior distribution implied by our prior beliefs, then such prior distributions are incapable of producing accurate forecasts

References

1 Akaike, H "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models," *Biometrika* Vol 60, 1973, pp 255-65

2 Akaike, H "A New Look at the Stochastic Model Identification," *I E E E Transaction and Automatic Control* A-C-19, No 6, 1974, pp 716-22

3 Alexander, D , and L R Thomas "Monetary/Asset Models of Exchange Rate Determination How Well Have They Performed in the 1980's?" *International Journal of Forecasting* Vol 3, 1987, pp 53-64

4 Birnbaum, A "The Neyman-Pearson Theory as Decision Theory, and as Inference Theory, with a Criticism of the Lindley-Savage Argument for Bayesian Theory," *Synthese* Vol 36, 1977, pp 19-49

5 Boland, L A *The Foundation of Economic Method* London George Allen and Unwin, 1982

6 Chipman, J S "Estimation and Aggregation in Econometrics An Application of the Theory of Generalized Inverses," *Generalized Inverses and Applications* (ed M Zuhair Nashed) New York Academic Press, 1976

7 Chow, G C *Econometrics* New York McGraw-Hill Book Co , 1983

8 Chow, G C "Random and Changing Coefficient Models," *Handbook of Econometrics* (eds Z Griliches and M D Intriligator) Vol 2 Amsterdam North-Holland, 1984, pp 1,213-46

9 Conway, R K "An Examination of the 'Schuh Controversy' Have Agricultural Exports Become Elastic," *Applied Economics* Vol 19, 1987, pp 853-73

10 Conway, R K , J Hrubovcak, and M LeBlanc *A Forecast Evaluation of Capital Investment in Agriculture* U S Dept Agr , Econ Res Serv TB-1732, Aug 1987

11 Conway, R K , J Hrubovcak, and M LeBlanc "The Structure of Agricultural Investment Comparing a Flexible Accelerator with Stochastic Coefficients," *Journal of Business & Economic Statistics* Vol 6, 1988, pp 231-40

12 Conway, R K , C B Hallahan, R P Stillman, and P T Prentice *Forecasting Livestock Prices Fixed and Stochastic Coefficients Estimation* U S Dept Agr , Econ Res Serv TB-1725, May 1987

13 Conway, R K , and G Gill *Is the Phillips Curve Stable? A Time Varying Parameter Approach* U S Dept Agr , Econ Res Serv Staff Report No AGES861209, May 1987 and forthcoming *Journal of Policy Modeling*

14 Conway, R K , R Durst, J Hrubovcak, and M LeBlanc *Economic Consequences of Tax Reform on Agricultural Investment* U S Dept Agr , Econ Res Serv TB-1741, Feb 1988

15 Copas, J B "Discussion of Professor Bernardo's Paper, Reference Posterior Distributions for Bayesian Inference," *Journal of the Royal Statistical Society Series B* Vol 41, No 2, 1979, pp 128-30

16 de Finetti, B *The Theory of Probability* Vol 1 New York John Wiley & Sons, 1974

17 Garcia-Ferrer, A , R A Highfield, F Palm, and A Zellner "Macroeconomic Forecasting Using Pooled International Data," *Journal of Business & Economic Statistics* Vol 5, No 1, Jan 1987, pp 53-67

18 Geisel, M S "Bayesian Comparison of Simple Macroeconomic Models," *Studies in Bayesian Econometrics and Statistics* (eds S E Fienberg and A Zellner) Amsterdam North-Holland Publishing Company, 1975, p 227 56

19 Goel, P M , and M H DeGroot "Only Normal Distributions Have Linear Posterior Expectations in Linear Regression," *Journal of the American Statistical Association* Vol 75, No 372, Dec 1980, pp 895-900

20 Goldberger, A S "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association* Vol 57, No 298, June 1962, pp 369-75

21 Granger, C W J , and N W Watson "Time Series and Spectral Methods in Econometrics," *Handbook of Econometrics* (eds Z Griliches and M D Intriligator) Vol 2 Amsterdam North-Holland, 1984, pp 979-1,022

22 Hannan, E J "The Estimation of the Order of An ARMA Process," *The Annals of Statistics* Vol 8, No 5, 1980, pp 1,071-81

23 Harville, D "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects," *The Annals of Statistics* Vol 4, No 2, 1976, pp 384-95

24 Jeffreys, H *Theory of Probability* 3rd edition London Oxford University Press, 1967

25 Judge, G G , W E Griffiths, R Carter Hill, H Lutkepohl, and T C Lee *The Theory and Practice of Econometrics* 2nd edition New York John Wiley & Sons, 1985

26 Kashyap, A K , P A V B Swamy, J S Mehta, and R D Porter "Further Results on Estimating Linear Regression Models With Partial Prior Information," *Economic Modelling* Vol 5, No 1, Jan 1988, pp 49-57

27 Keating, J P , and R L Mason "Practical Relevance of an Alternative Criterion in Estimation," *The American Statistician* Vol 39, No 3, Aug 1985, pp 203-4

28 Keating, J P , and R L Mason "James-Stein Estimation From an Alternative Perspective," *The American Statistician* Vol 42, No 2, May 1988, pp 160-4

29 Kiefer, J , and J Wolfowitz "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics* Vol 27, 1956, pp 884-906

30 Kmenta, J *Elements of Econometrics* 2nd edition New York Macmillan Publishing Co , 1986

31 Kourouklis, S , and C C Paige "A Constrained Least Squares Approach to the General Gauss-Markov Linear Model," *Journal of the American Statistical Association* Vol 76, No 375, Sept 1981, pp 620-5

32 Lane, D A , and W D Sudderth "Coherent Predictions Are Strategic," *The Annals of Statistics* Vol 13, No 3, 1985, pp 1,244-48

33 LeBlanc, M , J Kitchen, and R K Conway "A Stochastic Coefficents Interpretation of Exchange Rate Models " Paper presented at Economic Dynamics and Control Conference, Arizona State University, Tempe, Mar 1988

34 Lehmann, E L *Theory of Point Estimation* New York John Wiley & Sons, 1983

35 Oakes, D "Self-Calibrating Priors Do Not Exist," *Journal of the American Statistical Association* Vol 80, No 390, June 1985, p 339

36 Rao, C R *Linear Statistical Inference and Its Applications* 2nd edition New York John Wiley & Sons, 1973

37 Rao, C R "Least Squares Theory Using An Estimated Dispersion Matrix and Its Application to Measurement of Signals," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* Vol 1 Berkeley Univ of California Press, 1967, pp 355-72

38 Rao, C R "Characterization of Prior Distributions and Solutions to a Compound Decision Problem," *The Annals of Statistics* Vol 4, No 5, 1976, pp 823-35

39 Reinsel, G "Multivariate Repeated-Measurement or Growth Curve Models With Multivariate Random-Effects Covariance Structure," *Journal of the American Statistical Association* Vol 77, No 377, Mar 1982, pp 190-95

40 Reinsel, G "Estimation and Prediction in a Multivariate Random Effects Generalized Linear Model," *Journal of the American Statistical Association* Vol 79, No 386, June 1984, pp 406-14

41 Schinasi, G J , and P A V B Swamy "The Out-of-Sample Forecasting Performance of Exchange Rate Models When Coefficients Are Allowed to Change," International Finance Discussion Paper # 301 Fed Res Board, 1987

42 Schinasi, G J , and P A V.B Swamy "Forecasting the G-10 Weighted Average Dollar," Memorandum to Exchange Rate Forecasting Group Fed Res Board, 1988

43 Shibata, R "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion," *Biometrika* Vol 63, No 1, 1976, pp 117-26

44 Shibata, R "Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process," *The Annals of Statistics* Vol 8, No 1, 1980, pp 147-64

45 Swamy, P A V B *Statistical Inference in Random Coefficient Regression Models* New York Springer-Verlag, 1971

46 Swamy, P A V B "Linear Models with Random Coefficients," *Frontiers in Econometrics* (ed P Zarembka) New York Academic Press, 1974, pp 143-68

47 Swamy, P A V B , and J S Mehta "Bayesian and Non-Bayesian Analysis of Switching Regressions and of Random Coefficient Regression Models," *Journal of the American Statistical Association* Vol 70, No 351, Part 1, Sept 1975, pp 593-602

48 Swamy, P A V B , and P A Tinsley "Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients," *Journal of Econometrics* Vol 12, Feb 1980, pp 103-42

49 Swamy, P A V B , R K Conway, and P von zur Muehlen "The Foundations of Econometrics—Are There Any?" *Econometric Reviews* Vol 4, No 1, 1985, pp 1-120

50 Swamy, P A V B , A B Kennickell, and P von zur Muehlen "Forecasting Money Demand With Econometric Models " Special Studies Paper No 196, Fed Res Board, 1986

51 Swamy, P A V B , and G J Schinasi "Should Fixed Coefficients be Reestimated Every Period for Extrapolation?" Special Studies Paper No 213, Fed. Res Board, 1986, and forthcoming *Journal of Forecasting*

52 Swamy, P A V B , and P von zur Muehlen "Further Thoughts on Testing for Causality with Econometric Models," *Journal of Econometrics* Vol 39, Nos 1 and 2, Sept /Oct. 1988, pp 105-47

53 Swamy, P A V B., R K Conway, and M R LeBlanc "The Stochastic Coefficients Approach to Econometric Modeling, Part I A Critique of Fixed Coefficients Models," *The Journal of Agricultural Economics Research* Vol 40, No 2, Spring 1988, pp 2-10

54 Swamy, P A V B , R K Conway, and M R LeBlanc "The Stochastic Coefficients Approach to Econometric Modeling, Part II Description and Motivation," *The Journal of Agricultural Economics Research* Vol 40, No 3, Summer 1988, pp 21-30.

55 Swamy, P A V B , and G S Tavlas "Financial Deregulation, The Demand for Money, and Monetary Policy in Australia," *IMF Staff Papers* Vol 36, No 1, Mar 1989

56 Swamy, P A V B , B R Kolluri, and R N Singamsetti "What Do Regressions of Interest Rates on Deficits Imply?" Finance and Economics Discussion Series No 3, Fed Res Board, 1988

57 Thurman, S S , P A V B Swamy, and J S Mehta "An Examination of Distributed Lag Model Coefficients Estimated With Smoothness Priors," *Communications in Statistics — Theory and Methods* Vol 15, No 6, June 1986, pp 1723-50

58 Wolff, C C P "Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models," *Journal of Business and Economic Statistics* Vol 5, No 1, Jan 1987, pp 87-98

59 Zellner, A "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association* Vol 57, No 298, June 1962, pp 348-68

60 Zellner, A *An Introduction to Bayesian Inference in Econometrics* New York John Wiley & Sons, 1971

61 Zellner, A "Bayesian Analysis in Econometrics," *Journal of Econometrics* Vol 37, No 1, Jan 1988, pp 27-50