AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

The Stochastic Coefficients Approach to Econometric Modeling, Part II: Description and Motivation

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

Abstract. *A general stochastic coefficients model developed by Swamy and Tinsley serves as a reference point for discussion in this second of a series of three articles Other well-known specifications are related to the model. The authors weigh the advantages and disadvantages of stochastic coefficients and suggest procedures to address the identification and estimation problem with weaker and noncontradictory assumptions They argue that the real aim of inference is prediction and that "imprecise" parameter estimates of a coherent model are acceptable if they forecast well.*

Keywords. *Stochastic coefficients, fixed coefficients, time series analysis, Bayesian inference, identification, coherence, estimation.*

Editor's note *Part I A Critique of Fixed Coefficients Models* appeared in Vol 40, No 2, Spring 1988 *Part III Stability Tests, Estimation, and Prediction* will be published in Vol 41, No 1, Winter 1989

Although classical logic and probabilistic logic provide different rules of inference, both types of logic are useful in econometrics, as Swamy, Conway, and von zur Muehlen (*27*) and Swamy and von zur Muehlen (*31*) have shown [1] The purpose of these rules of inference is to indicate what conclusions may be inferred from what premises If we do not adhere to either type of logic, the conclusions we draw from given premises may be invalid For example, after estimating a fixed-coefficients model, the econometrician who adjusts the estimate of a constant term or the estimates of some other parameters like the autoregressive coefficient of the error process, while retaining the original estimates of the remaining parameters to obtain good forecasts, may violate one or more of the probability laws If any probability law is violated in the process of drawing an econometric inference, the resulting inference will be incoherent There is no logic associated with such an inference

Our economic inferences may also be invalid if our premises are contradictory Surely any inference based on a model is valid if the assumptions underlying the model, including those used for estimation and forecasting, do not contradict each other Extraneous restrictions necessitated by a fixed-coefficients approach need not be free from contradictions, as we have shown in Part I of this article (*26*) (See also Swamy and von zur Muehlen (*31*)) One alternative is to consider models with coefficients that are not fixed and thereby remove the necessity for extraneous restrictions that introduce contradictions In Part I, we showed that fixed-coefficients models may be inappropriate for other reasons as well These reasons include aggregation effects, changes in tastes, technology, institutions, and even policy

Many research papers have dealt with the estimation of a regression model in which some or all of the slopes are both time-dependent and stochastic Our primary purpose here is to evaluate the estimation methods suggested for these models To do so, we describe a general model, developed by Swamy and Tinsley (*30*), to serve as a touchstone We also compare other specifications with the general model to handle the identification and estimation problems with weaker and possibly noncontradictory assumptions

Introducing a Model with Stochastic Slopes

A regression model whose coefficient vector is time-dependent asserts that a scalar dependent variable y_t is time generated in accordance with

$$y_t = x_t'\beta_t \qquad (t=1, 2, \ldots, T) \qquad (1)$$

where x_t' is a $1\times K$ vector of observations on K independent variables, and β_t is a $K\times 1$ vector of coefficients

It is assumed that

$$\beta_t = \Pi z_t + J\xi_t \qquad (t=1, 2, \ldots, T) \qquad (2)$$

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, and Nadine Loften

[1] Italicized numbers in parentheses refer to items in the References at the end of this article

where Π is a $K \times m$ matrix of fixed coefficients, z_t is an $m \times 1$ vector of observable variables, J is a $K \times n$ matrix of fixed elements, and ξ_t is an $n \times 1$ vector of unobservable variables [2]

Regarding ξ_t, it is assumed that

$$\xi_t = \Phi \xi_{t-1} + e_t \qquad (3)$$

where Φ is an $n \times n$ matrix, e_t is an $n \times 1$ random vector with $E(e_t | \xi_{t-1}, z_t, x_t) = E(e_t) = 0$ for every t, and $E(e_t e_s' | \xi_{t-1}, z_t, x_t) = \Delta_e$ if $t = s$ and 0 if $t \neq s$

The first element of each x_t and z_t may be identically equal to 1 for all t, with the coefficient corresponding to these unit elements representing a time-varying intercept and a constant vector, respectively Since the usual additive disturbance term that appears in conventional econometric models with fixed coefficients cannot be distinguished from a time-dependent stochastic intercept term without the imposition of severe restrictions, both specifications are combined into a single term That is, the coefficient corresponding to the unit element of x_t represents the sum of the additive disturbance term and a time-varying intercept Thus, it is not correct to say that the usual disturbance term is omitted in equation 1

Note that the vectors x_t and z_t may not be completely distinct All those elements of x_t that are believed to be correlated with β_t can be included in z_t We will clarify the reasons for defining z_t this way

Note also that equation 3 is less restrictive than it may seem The equation represents a first-order autoregressive process only when $J = I_K$ and $n = K$ If J is a $K \times (p+q)K$ matrix having the columns of I_K as its first K columns and zeros elsewhere and if $n = (p+q)K$, then equation 3 represents the following mixed, autoregressive, moving-average process

$$\varepsilon_t = \sum_{i=1}^{p} \Phi_i \varepsilon_{t-i} + \sum_{j=1}^{q} \Theta_j a_{t-j} + a_t \qquad (4)$$

where ε_t, ε_{t-i} 's, a_t and a_{t-j}'s are Kxl vectors, Φ_i's and Θ_j's are $K \times K$ matrices, and $\{a_t\}$ is a sequence of uncorrelated $K \times 1$ random vectors, each with mean vector zero and constant covariance matrix $\sigma^2 \Delta_a$ *(30)*

[2]Kmenta (*13*, p 574) comments that "A troublesome aspect of Swamy's model is the assumption that the 'average' coefficients are constant over time He is apparently unaware of Swamy and Tinsley's (*30*) equation 2 above, which permits the "average" coefficients to vary over time However, Kmenta's criticism does apply to the other stochastic coefficients models he surveyed

The time profile of $\beta = (\beta_1', \beta_2', \ldots, \beta_T')'$, which is smoother than the profile implied by equations 2 and 3, is obtained if these equations are replaced by the following equation

$$R\beta = v \qquad (5)$$

where $R\beta$ represents the d_j+1 th order difference of the values of the jth coefficient in β and where v is a random variable with mean vector zero and a diagonal covariance matrix It follows from Shiller (*25*) that equation 5 implies some smoothness restrictions on β

The First Two Moments of Variables in Equation 1

First, consider equations 1-3 Then, inserting equation 2 into equation 1 gives

$$y_t = x_t' \Pi z_t + x_t' J \xi_t \qquad (6)$$

Equation 3 implies that, given x_t and z_t, the mean of y_t is $x_t' \Pi z_t$ It also implies that

$$E(\xi) = 0,\ E(\xi_t \xi_t') = \Gamma_0 = \Phi \Gamma_0 \Phi' + \sigma^2 \Delta_e,$$
$$E(\xi_t \xi_{t-s}') = \Phi^s \Gamma_0 \text{ and } E(\xi_t \xi_{t+s}') = \Gamma_0 \Phi'^s \qquad (7)$$

Therefore, defining $u_t = x_t' J \xi_t$ gives

$$E(u_t) = 0 \text{ and } E(u_t u_{t-s}) = x_t' J \Phi^s \Gamma_0 J' x_{t-s} \qquad (8)$$

From equation 8 it follows that the covariance matrix of $u = (u_1, u_2, \ldots, u_T)'$ is

$$\Sigma_y = [x_i' J \Phi^{i-j} \Gamma_0 J' x_j \text{ if } i \geq j \text{ and } x_i' J \Gamma_0 \Phi'^{j-i} J' x_j \text{ if } i < j] \qquad (9)$$

Σ_y is also the covariance matrix of $y = (y_1, y_2, \ldots, y_T)'$

Next, if equation 5 is assumed instead of equations 2 and 3, then the first two moments of y and β may be derived as in Kashyap and others (*12*)

Special Cases of Equations 1-3

The following 10 models, which are frequently considered by econometricians, are special cases of equations 1-3

Box-Jenkins' type model (*1*) In the case where all the elements of Π, Φ_i's, Θ_j's, and Δ_a other than the leading diagonal elements are zero, model 1 reduces to a univariate autoregressive-moving-average (ARMA) model In turn, this univariate ARMA model reduces to the autoregressive-integrated-moving-average (ARIMA) variant if some of the roots of the following autoregressive polynomial equation are unity

$$1-\phi_{111}\zeta-\phi_{211}\zeta^2-, \quad , -\phi_{p11}\zeta^p = 0 \qquad (10)$$

where ϕ_{111}, ϕ_{211}, , ϕ_{p11} are the leading diagonal elements of the matrices, Φ_1, Φ_2, , Φ_p, respectively, and ζ is a complex number

A fixed-coefficients model with an ARMA or ARIMA error term. The conventional fixed-coefficients model with an ARMA or ARIMA error term is obtained if all the elements of Φ_i's, Θ_j's, and Δ_a other than the leading diagonal elements are zero This fixed-coefficients model will not be a reduced-form equation if some of the elements of x_t are endogenous It also reduces to a model having dummy or deterministic variables as some of its explanatory variables if these dummy or deterministic variables are included in z_t

A fixed-coefficients model with a heteroscedastic error term. Equations 1-3 are equivalent to a fixed-coefficients model with a heteroscedastic error term if $\Phi_i = 0$ (i=1, 2, , p), $\Theta_j = 0$ (j=1, 2, , q), and all the elements of Δ_a except one diagonal element corresponding to a nonconstant element of x_t are zero Such a fixed-coefficients model will not be a reduced-form equation if some of the elements of x_t are endogenous

A model with time-dependent deterministic slopes. The last set of K−1 equations in equation 4 reduces to a set of deterministic difference equations if all the elements of Δ_a and Θ_j's except the leading diagonal elements are zero In this case, the slope coefficients of equation 1 are time-dependent deterministic parameters even when all the columns of Π except the first are null

Autoregressive conditional heteroscedastic (ARCH) models. The ARCH model proposed by Engle (*6*) can be considered as a special case of equations 1–3, obtained by replacing y_t in equation 1 by $y_t^* = (y_t - x_t'\alpha)$ with fixed α, replacing x_t in equation 1 by $(1, y_{t-1}^*, \quad , y_{t-K+1}^*)$, settng Π in equation 2 equal to a null matrix, zeroing the Φ_i's and Θ_j's in equation 4, and setting Δ_a equal to a diagonal matrix

An equation embedded in a VAR model. Equation 1 will have the form of an equation embedded in a vector autoregressive (VAR) model if the vector x_t consists of lagged y's and current and lagged values of variables other than y, if all the columns of Π except the first are null, if all the matrices Φ_i's and Θ_j's are null, and if all the elements of Δ_a other than the leading diagonal element are zero

Hildreth and Houck's model. Equation 1 reduces to Hildreth and Houck's (*10*) model if all the columns of Π except the first are null, if all the matrices Φ_i's and Θ_j's are null, if Δ_a is diagonal, and if all the elements of x_t' are fixed

Fisk's model. The model consisting of equations 1, 2, and 4 is the same as that of Fisk (*7*) if all the elements of x_t' are fixed, if all the columns of Π other than the first are null, and if all the matrices, Φ_i's and Θ_j's, are null

A disequilibrium model. Equation 1 is either a demand or a supply function for a market in disequilibrium if two conditions hold First, if the first element of β_t can be separated from an additive disturbance term so that β_t is independent of the disturbance term Second, if β_t follows a discrete distribution such that the conditional probability density function (pdf) of y_t, given x_t', is equal to the ratio of two probability values, each of which is less than unity, times the pdf of the normal distribution with mean $x_t'\beta$ and constant variance σ^2 These probability values are time de pendent (*18*)

A restricted version of the Kalman model. Equations 1-3 give a version of Kalman's model if an additive error term can be separated from the first element of β_t, if the additive error term is serially uncorrelated and also uncorrelated with β_t in every period, if Π is null, if $J = I_K$, if Φ and Δ_e are known, if the mean and the variance of the additive error term are known, and if the conditional mean and the conditional covariance matrix of β_t, given y_t for t = 1, are known (*5*) In some Kalman filter applications, equations 2 and 4 with the restrictions that $\Pi = 0$, $\Phi_1 = I_K$, $\Phi_i = 0$ (i = 2, , p), and $\Theta_j = 0$ (j = 1, 2, , q) are used In this case the constant mean of β_t is indeterminate and equation 1 cannot be written in the form of equation 6

The Cooley-Prescott and Rosenberg Models

Cooley and Prescott (*2*) and Rosenberg (*24*) also consider equation 1, but make assumptions that differ from our equations 2, 3, and 5 The maximum likelihood estimators for all the unknown parameters of these models do not exist and there are no other operational methods of estimating these models

A Priori Restrictions

Several of the models considered in the econometric literature are the restrictive forms of equations 1–3 Building models so as to have each equation satisfy one or another set of these restrictions requires a priori considerations about the forms of economic laws that are within the purview of coherent economic theories [3] Any set of contradictory restrictions or restrictions violating the conditions under which empirically interpretable models exist should be rejected outright The a priori restrictions needed to deduce conventional fixed-coefficients models from equations 1–3 may be contradictory and may violate the conditions under which these models exist, as shown in Part I of our series of articles (*26*), see also (*31*) Restricted models could be justified only if one could find empirically that models so restricted were still coherent (or free from contradictions) and performed better in prediction than models not so restricted

This is not to say that equations 1–3 (or 1 and 5) should be used without restrictions The above argument only calls for caution in imposing any restrictions on equations 1–3 (or 1 and 5) If we want to analyze equations 1–3 (or 1 and 5) without imposing any restrictions because we are afraid that any restrictions on these equations might introduce contradictions, then we may have to use arbitrary values for the parameters of the equations These arbitrary values might lead to unreasonable results or poor forecasts Coherent zero restrictions on Π, $J\Phi$, and Δ_a may exist We may find these restrictions by comparing the out-of-sample forecasting performance of different restrictions We present examples of such comparisons in Part III Thus, models consisting of equations of the type 1–3 (or 1 and 5) with unknown Π, Φ, and Δ_a (or moments of β implied by equation 5) have advantages as well as disadvantages over conventional fixed-coefficients models

Advantages of Stochastic Coefficients Models

In stating equations 1–3, we have not violated any probability laws Indeed, all three assumptions represented by equations 1–3 are consistent with a formal axiomatic foundation of probability theory Therefore, the procedure for verifying the logical consistency of these equations is relatively straightforward The additive error term that appears in nearly every conventional fixed-coefficients model can be added to its fixed intercept Thus, conventional fixed-coefficients models can be viewed as models with random intercept and fixed slopes The assumption that an intercept is random and slopes are fixed is just as arbitrary as the assumption that all coefficients are random Because any assumption about the unobservable β_t is necessarily arbitrary, equations 1 and 2 prudently impose a *minimal* set of assumptions that avoid contradictory restrictions Of course, no one can *prove* that assumptions 1–3 are true, but at least the logical requirement of coherency is satisfied Use of equations 1 and 5 may involve some contradictions if we are not careful The necessary precautions we should take to avoid contradictions are explicitly stated in Thurman and others (*33*) and Kashyap and others (*12*)

Conventional approaches yield a set of problems that require some major modifications to capture important higher order nonlinearities and nonstationarities One offered solution is to find stationarity-inducing transformations (for example, Box-Jenkins) yielding forms that can be subjected to conventional techniques suitable to stationary processes Because in any given case the appropriateness of such transformations is always uncertain, it would be desirable to find a methodology for which this doubt is not present One such approach is given by equations 1–3 where it is shown that problems of first- and second-moment nonlinearities and nonstationarities, including those caused by heteroscedasticity, can be dealt with in natural ways that do not require the imposition of unverifiable and possibly contradictory assumptions

Equations 1–3 will coincide with a stochastic law defined by nature's behavior if and only if x_t and z_t are uncorrelated with e_t, as shown by Pratt and Schlaifer (*22*) for a model that is simpler than equation 1 Assumptions 2 and 3 state that ξ_t is mean independent of x_t and z_t It follows from Pratt and Schlaifer's argument (*22*) that this mean independence condition is satisfied unless any nonconstant elements x_{it} of x_t and z_{it} of z_t are directly or indirectly affected by any element β_{jt} of β_t or by any nonconstant variable not included in x_t and z_t that either affects or is affected by β_{jt}

We have shown in Part I (*26*) that an instrumental variables method of estimating an equation with fixed coefficients can introduce contradictions In contrast, one can handle "simultaneous equations" complications within the framework of equations 1-3 without using any instruments Suppose that equation 2 is part of a larger model In this case, regressors may be correlated with the contemporaneous errors Then, elements of x_t are correlated with those of β_t, which means they also appear in z_t on the right side of equation 2 The vector $J\xi_t$ is that part of β_t not correlated with z_t if equations 1-3 define a stochastic law Under this condition, we may assume that $J\xi_t$ is mean independent of z_t Thus, to estimate equations 1-3, we do not need any instruments excluded from equation 1, even when some or all of the elements of x_t are endogenous The possible correlations between β_t and

[3]Zellner (*34*) has shown that Feigl's definition of causality, namely "predictability according to a law or set of laws," applies to complete econometric models regardless of whether the models satisfy these a priori restrictions or not

x_t and between an additive error term and x_t are ignored by Pagan (*21*)

Another feature of equations 1-3 is that they are identifiable even when the number of exogenous variables excluded from equation 1 is zero or smaller than the number of endogenous variables included in equation 1 minus one Even if all the right-hand-side variables in equation 1 are endogenous, they can all be included on the right-hand side of equation 2, and equation 6 is identifiable in this case This feature is a clear advantage when we do not know the truth of exclusion restrictions and the exogeneity of variables Equations 1-3 give a coherent method of identifying and estimating coefficients that change over time without the need for instruments, restrictions, or ad hoc adjustments that might introduce contradictions

Equation 6, obtained by inserting equation 2 into 1, reveals that the representation of a process in 1 and 2 is equivalent to a fixed-coefficients nonlinear model with serially correlated and heteroscedastic errors of a very general form If regression models with heteroscedastic or serially correlated error terms covered in econometrics textbooks represent economic laws, then so does equation 6 To obtain a textbook-type model, we should impose certain zero restrictions on Φ_i's, Θ_j's, and Δ_a that could contradict each other Our intent in working with equation 6 is not to ask "woolly" questions and receive "woolly" answers (in Maddala's sense (*17*, p 403)) or to unnecessarily complicate the analysis, but to avoid introducing contradictory restrictions It seems that some researchers would rather use a fixed-coefficients model because such a model supposedly answers "nonwoolly" questions than use the coherent set of equations 1-3, even if the former is incoherent Models with contradictory premises cannot be true Therefore, we prefer to risk so-called "woolly" answers if the likely alternative is incoherence In any case, the temptation to think that equations 1 3 lack the explanatory power of a conventional fixed-coefficients model is unwarranted

Kmenta makes two criticisms of stochastic coefficients models (1) "the models are not justified by theory" and (2) the "use of varying coefficients models implies that we have given up trying to find the real causes of [coefficient] variation" (*13*, p 578) Since these criticisms are representative of general comments made by others, we think it is appropriate to respond to Kmenta here

A widespread practice among econometricians is to add a stochastic error term to a mathematical model somewhat arbitrarily to represent *unidentified* factors and to make the meaningless or false assumption that at least some of the included variables are uncorrelated with those unidentified factors, as rightly pointed out by Pratt and Schlaifer (*22*, p 11) Kmenta follows this practice and criticizes stochastic coefficients models that depart from it Just as the mathematical calculus is used by economists to rigorously derive mathematical models of economics, so the probability calculus should be used to rigorously derive stochastic models of economics The derivation of stochastic coefficients models, unlike the derivation of fixed-coefficients models, does not violate the probability laws, as Swamy and von zur Muehlen show (*31*) Therefore, it is not true that the stochastic coefficients models are not justified by theory Furthermore, the use of stochastic coefficients models represents an attempt to acknowledge as well as to model explicitly the coefficient variation but not to give up trying to find the real causes of coefficient variation, as suggested by Kmenta Of course, no one can prove that a model of coefficient variation or the convenient assumption of fixed slopes is true Any assumption about the purely unobservable coefficients is largely arbitrary The reason is that the tests of the constancy of regression slopes against a general alternative have low power and hence are not informative In any case, stochastic coefficients models have the advantage of being able to predict future values of observable variables at least as well as their fixed-coefficients counterparts, as we will show in Part III

In view of Swamy and von zur Muehlen's (*31*) demonstration that it is impossible to be sure of the true causes of even the observable effects, it may be impossible to follow Kmenta's suggestion that we can find the real causes of coefficient variation The probability theory teaches us how to be coherent, but it does not tell us how to find the real causes of coefficient variation

The Kalman model, which separates an additive error term from the first element of β_t, does not have the advantages of equations 1-3 because it cannot take into account the possible correlations between an additive error term and x_t' Besides, how can any econometrician know the values of Π, Φ_i's, Θ_j's, and Δ_a to implement the Kalman filter formula empirically? Meinhold and Singpurwalla's stereotyped Bayesian interpretations (*20*) of a Kalman filter do not apply to equations 1-3, if in equation 1, an additive error term is correlated with β_t and, hence, cannot be separated from the first element of β_t Furthermore, the convenient prior distributions employed by Meinhold and Singpurwalla (*20*) and by Doan, Litterman, and Sims (*4*) in their applications of Kalman's filter are arbitrary

When the derivation of a subjective probability distribution from the Bayesian assumptions of coherent behavior is not possible, then an arbitrary and convenient distribution is used in place of a subjective prior distribution Using Pratt and Schlaifer's argument (*22*, p 21), we can show that, if Πz_t is the effect of x_t on y_t and if α is specified incorrectly as the coefficient vector of x_t in the regression of y_t on $(x_t', z_t')'$, a Bayesian analysis that is based on a prior distribution of α alone and that ignores the difference between Πz_t and α will be as inconsistent as the usual methods Any full Bayesian analysis will be inconsistent unless one keeps in mind the possible reasons for differences between Πz_t and α given in this article when assessing a distribution of $\Pi z_t - \alpha$ or of Πz_t, given α One should not use arbitrary prior distributions regardless of the accuracy of the out-of-sample forecasts they produce

To see clearly another advantage of equations 1-3 over fixed-coefficients models, consider the case where x_t itself is not observed but the observations on x_t contain measurement errors It is known in the econometric literature that, when both the left- and right-side variables in a regression equation are measured with error, the regression equation between the observables is not identified unless the ratios of these error variances are known No econometrician can ever possess this type of prior information By contrast, such prior information is not needed to estimate consistently equations 1-3 To see why, suppose that the vector x_t' in equation 1 is not observable and the observations on x_t contain measurement errors In this case, if we replace x_t' in equation 1 by its observable counterpart, say $x_t^{*\prime}$, then $x_t^{*\prime}$ and its coefficient vector, say β_t^*, will be correlated If, in equation 2, we replace β_t by β_t^* and if x_t^* is a subvector of z_t, then Πz_t represents that part of β_t^* that is correlated with x_t^* and the remaining subvector of z_t, and $J\xi_t$ represents that part of β_t^* that is uncorrelated with x_t^* and the remaining subvector of z_t Thus, it is correct to treat the coefficients in the error-in-the-variables models as stochastic and the analysis of such models can proceed even when the ratios of the variances of measurement errors in y_t and x_t are not known, provided x_t^* is a subvector of z_t, and β_t^* is not considered as fixed

Now we should interpret β_t In applications considered in nonexperimental sciences such as economics, the model is estimated either from the data that are already available or perhaps from a subjective view of what the data would be like if they were available In such cases, there is no way to separate what the data say about β_t from "prior" information about β_t Indeed, β_t cannot be said to exist prior to the formulation of a model, even though there may be much prior information about which data might be observed In these situations it is reasonable to assume that the interpretation of β_t is defined in terms of the assumed model and may not refer to the physical reality that the model is intended to represent We owe this view to Lane (*14*) As a result, we prefer to adopt Lane's interpretation 2 in Part I, that the coefficient vector β_t in equation 1 taking values in an abstract set merely indexes that distribution of y_t As Lane (*14*) observes, any two experiments with the same index set can be mixed

Disadvantages of Stochastic Coefficients Models

Like the fixed-coefficients models, equations 1-3 may not represent a real physical process Yet, the assumption that equations 1-3 represent a real physical process is needed for the validity of the argument here A convenient algebraic expression for this assumption is that the distribution of $y = (y_1, y_2, \quad , y_T)'$, given x_t' and z_t' for $t=1, 2, \quad , T$, implied by equations 1-3, is indexed by θ and belongs to the following known class

$$\mathcal{P} = \{ P_\theta, \ \theta \epsilon \Theta \} \tag{11}$$

where each of the parameters of equations 2 and 3 is an element of θ and Θ is the parameter space Here θ is a possible value for some real physical parameter, and the distribution P_{θ_0} belonging to $\mathcal{P}$ is to be regarded as the distribution that actually generated the data when θ_0 was the true value of that parameter

Makelainen, Schmidt, and Styan have shown that the maximum likelihood estimate of θ exists and is unique if a twice continuously differentiable likelihood function is constant on the boundary of the parameter space Θ and if the Hessian matrix of second partial derivatives of the likelihood function is negative definite at the points where the gradient vector of the function vanishes (*19*) They have also shown that the condition of constancy on the boundary cannot be completely removed when there is more than one unknown parameter Asymptotic theory ensures, for a sufficiently regular family of distributions, that a consistent sequence of solutions to the likelihood equations will be unique from some sample size onwards However, it is important to find out, as a partial check on the applicability of asymptotic maximum likelihood theory or, more generally, as a step in inspecting the likelihood function, whether the likelihood equations admit a unique solution and whether such a solution actually maximizes the likelihood This partial check is particularly important in the case of equations 1-3, where the unknown parameters are abundant and the assumption that equations 1-3 represent a real physical process is questionable Furthermore, if the solution of the likelihood equations is not unique, the usual regularity conditions do not establish the existence of an efficient estimator of θ (*16*, p 435)

Applying Makelainen, Schmidt, and Styan's argument (*19*, Section 4 3) to equations 1-3 shows that when Je_t is normal and when Π, $J\Phi$, and $J\Delta_e J'$ are unknown, the likelihood function for equations 1-3 does not necessarily tend to zero as the diagonal elements of $J\Delta_e J'$ tend to 0 or ∞ This result means that the likelihood function is not constant on the boundary of the parameter space and, hence, this boundary is not necessarily the region of "minimal likelihood" Therefore, there is a basis to assume that the likelihood equations do not admit a unique solution for Π, $J\Phi$, and $J\Delta_e J'$ and that any such solution is only a local maximum either *inside* the parameter space or on the *boundary* The occurrence of several maxima of about the same magnitude would mean that the likelihood-based confidence regions are formed from disjoint regions and summarization of data by means of a maximum likelihood estimate and its asymptotic variance could be quite misleading, as is pointed out in the statistics literature Furthermore, when the sizes of the unknown parameter matrices, Π, $J\Phi$, and $J\Delta_e J'$, are big, the estimates of these parameter matrices obtained by numerically maximizing the likelihood function may be quite unsatisfactory because of overfitting These difficulties with the maximum likelihood procedure are not appreciated by Rosenberg (*24*), Cooley and Prescott (*2*), Pagan (*21*), Harvey and Phillips (*8*), and Judge, Griffiths, Hill, Lutkepohl, and Lee (*11*, pp 809-14), among others If the maximum likelihood estimate of θ does not exist, then Pagan's conditions (*21*), unlike Swamy and Tinsley's conditions (*30*), for the identification of θ are irrelevant [4] Pagan also mechanically reproduces Crowder's consistency conditions without verifying them The nonexistence of maximum likelihood estimates or the nonuniqueness of the solutions of the likelihood equations is not a difficulty that arises exclusively in the context of equations 1 3 Swamy and Mehta (*28, 29*) give instances of disequilibrium and simultaneous equations models where the maximum likelihood estimates of fixed coefficients do not exist

We do not think that anyone seriously believes that he or she can know exactly the values of Π, $J\Phi$, $J\Delta_e J'$ and the conditional mean and conditional covariance matrix of β_t, given y_t, for $t = 1$ appearing in the Kalman filter We also doubt that, for a Bayesian analysis of equations 1-3, one can find reasonable prior distributions of the parameters Π, $J\Phi$, and $J\Delta_e J'$ with known hyperparameters There may be no virtue in using arbitrary prior distributions Therefore, we should have some data-based estimates of these parameters to compare the consequences of using arbitrary a priori values with those of using data-based estimates Swamy and Tinsley (*30*) developed a technique that provides data-based estimates of Π, $J\Phi$, and $J\Delta_e J'$ [5]

It follows from the derivation of Swamy and Tinsley (*30*), Swamy and Mehta (*28*, p 596), and Harville (*9*) that, if equations 1-3 are true, then the predictor of a value of y in an out-of-sample period T+s with the smallest variance within the class of linear unbiased predictors is

$$\tilde{y}_{T+s} = x'_{T+s}(z'_{T+s} \otimes I_K)\mathrm{vec}[\tilde{\Pi}] + x'_{T+s} J\Phi^s \Sigma'_{\xi T}(I_T \otimes J')D'_x \Sigma_y^{-1} \cdot (y - D_x Z_e \mathrm{vec}[\tilde{\Pi}]) \tag{12}$$

where T is the terminal period of the sample, $\mathrm{vec}[\tilde{\Pi}] = (Z'_e D'_x \Sigma_y^{-1} D_x Z_e)^{-1} Z'_e D'_x \Sigma_y^{-1} y$ is the generalized least squares estimator of vec $[\Pi]$, which is the column stack of Π, Φ is as defined in equation 3, $\Sigma_{\xi T}$ is the matrix made up of the last (p+q)K columns of the covariance matrix of ξ_t, $t=1,2, \ldots, T$, $D_x = \mathrm{diag}[x'_1, x'_2, \ldots, x'_T]$, $y = (y_1, y_2, \ldots, y_T)'$, Σ_y is as defined in equation 9, and $Z'_e = [z_1 \otimes I_K, \ldots, z_T \otimes I_K]$

Clearly, the optimal predictor equation 12 is not operational if the parameter matrices $J\Phi$ and $J\Delta_e J'$ are unknown, as they usually are Swamy and Tinsley (*30*) develop the following estimating equations

$$\mathrm{vec}[\hat{\Pi}] = (Z'_e D'_x \hat{\Sigma}_y^{-1} D_x Z_e)^{-1} Z'_e D'_x \hat{\Sigma}_y^{-1} y \tag{13}$$

$$\hat{u} = [I_T - D_x Z_e (Z'_e D'_x \hat{\Sigma}_y^{-1} D_x Z_e)^{-1} Z'_e D'_x \hat{\Sigma}_y^{-1}] y \tag{14}$$

$$\hat{\xi}_t = \hat{\Sigma}'_{\xi t} (I_T \otimes J') D'_x \hat{\Sigma}_y^{-1} \hat{u} \quad (t = 1, 2, \ldots, T) \tag{15}$$

$$\begin{bmatrix} x'_2 J\hat{\xi}_2 \\ x'_3 J\hat{\xi}_3 \\ \cdot \\ \cdot \\ \cdot \\ x'_T J\hat{\xi}_T \end{bmatrix} = \begin{bmatrix} (\hat{\xi}'_1 \otimes x'_2) \\ (\hat{\xi}'_2 \otimes x'_3) \\ \cdot \\ \cdot \\ \cdot \\ (\hat{\xi}'_{T-1} \otimes x'_T) \end{bmatrix} \mathrm{vec}[J\hat{\Phi}] + \mathrm{error} \tag{16}$$

$$\begin{bmatrix} x'_2 \otimes x'_2 \\ x'_3 \otimes x'_3 \\ \cdot \\ \cdot \\ x'_T \otimes x'_T \end{bmatrix} \mathrm{vec}[\hat{\Delta}_a] + \mathrm{error} = \begin{bmatrix} x'_2 J(\hat{\xi}_2 - \hat{\Phi}\hat{\xi}_1)(\hat{\xi}_2 - \hat{\Phi}\hat{\xi}_1)' J' x_2 \\ x'_3 J(\hat{\xi}_3 - \hat{\Phi}\hat{\xi}_2)(\hat{\xi}_3 - \hat{\Phi}\hat{\xi}_2)' J' x_3 \\ \cdot \\ \cdot \\ x'_T J(\hat{\xi}_T - \hat{\Phi}\hat{\xi}_{T-1})(\hat{\xi}_T - \hat{\Phi}\hat{\xi}_{T-1})' J' x_T \end{bmatrix} (1/\hat{\sigma}^2) \tag{17}$$

$$\hat{\sigma}^2 = (y - D_x Z_e \mathrm{vec}[\hat{\Pi}])' \hat{\Sigma}_y^{-1} (y - D_x Z_e \mathrm{vec}[\hat{\Pi}])/T \tag{18}$$

$$\hat{y}_{T+s} = x'_{T+s}(z'_{T+s} \otimes I_K)\mathrm{vec}[\hat{\Pi}] + c x'_{T+s} J\hat{\Phi}^s \hat{\Sigma}'_{\xi T}(I_T \otimes J') \cdot D'_x \hat{\Sigma}_y^{-1}(y - D_x Z_e \mathrm{vec}[\hat{\Pi}]) \tag{19}$$

where $c \in [0, 1]$ is a constant

[4] Since Swamy and Tinsley's conditions (*30*) for the identification of θ are related to an estimation method that always works, their conditions are always relevant

[5] In the National Bureau of Economic Research National Science Foundation Seminar on Bayesian Inference in Econometrics held at the University of Michigan, Ann Arbor MI, on Nov 3-4, 1978, Haitovsky explicitly questioned whether or not Swamy's work had led the profession in the wrong direction This article is written partly to let the readers judge whether or not Swamy and his associates' work has misled the profession

Equations 16 and 17 reduce to the usual estimating equations given in econometrics textbooks if all the elements of Φ and Δ_e other than the leading diagonal elements are zero To solve equations 13-19, we follow an iterative procedure in which $J\Phi$ and Δ_a are initially arbitrarily chosen, but, through iteration, the dependence of estimators on these arbitrary values is eliminated However, convergence of this iterative procedure may not be achieved unless the conditions of Szatrowski's theorem 5 (*32*) are satisfied These conditions may not be satisfied if $\Phi \neq 0$ and if Δ_a is not diagonal The reason for following any iterative procedure that converges is to find the maximum likelihood or nonlinear least squares estimates If such estimates do not exist, then no iterative procedure converges to those estimates We have already pointed out that, in cases where $\Phi \neq 0$ and Δ_a is not diagonal, the sufficient conditions for the existence of the maximum likelihood estimate of θ are not satisfied Therefore, choosing among the estimates obtained in different iterations of Swamy and Tinsley's procedure (*30*) is a problem One solution to this problem is to choose estimates that give a (local) minimum value for the root mean square error of forecasts of y_{T+s} for s = 1, 2, , S This procedure avoids overfitting We should emphasize that the estimates of $J\Phi$ and Δ_a obtained in any iteration may be quite imprecise However, it is possible that Π is more precisely estimated than either $J\Phi$ or Δ_a, and so the accuracy of the forecast equation 19 might improve if c is set equal to a value less than 1, since the second term on the right side of equation 19 is more heavily influenced by the estimates of $J\Phi$ and Δ_a than is the first term Rao (*23*) gives an optimal value of c for a model that is simpler than equation 1, and this value is less than 1

The results based on equations 13-19 are highly nonrobust in the sense that a small change in an observation can make a substantial difference in the parameter estimates This result occurs because the number of observations per unknown parameter is quite low, unless $J\Phi$ and Δ_a are severely restricted For this reason, the values of $p>1$ and $q>0$ in equation 4 are not recommended

A Faustian Bargain? Trading Dilemmas

Perhaps econometricians generally prefer models with fixed slopes because of the disadvantages of assuming that all coefficients in a regression are varying But fixed-coefficients models also give rise to difficulties, as is shown in Part I Here then is a dilemma The robustness of results given by equations 1-3 is quite low and cannot be increased unless we put a sufficient number of restrictions on the parameters of these equations, in which case the equations may reduce to a fixed-coefficients model However, once restricted, equations 1-3 may have no advantages over a fixed-coefficients model and may suffer from contradictions

After all, every econometric procedure is based on some often quite special assumptions about underlying distributions and about the relation between the mathematical parameters of those distributions and the "true state" of the world That these assumptions may only be subjective and may not be factually true is argued by several Bayesians including de Finetti (*3*) and Lane (*14, 15*) From a subjective viewpoint, the assumption of fixed coefficients, implying that the distribution of each regression slope is degenerate at a point, is more stringent than de Finetti's notion of prevision and his requirements of coherence (*3*) Statisticians and econometricians in the past have appealed to the two contrary principles of parsimony and profligacy to justify ARIMA models of finite order and VAR models of finite order, respectively Swamy and von zur Muehlen (*31*) demonstrate that the premises of these models can be contradictory If the principles of parsimony and profligacy clash with the principle of coherence, the former principles should be rejected in favor of the latter principle The principle of coherence is preeminent, and equations 1-3 may help us empirically implement that principle The coherence approach prohibits the use of models with contradictory premises, but does not prohibit the use of imprecise parameter estimates, provided those estimates give successful forecasts of future observable values and plausible explanations of past experience

The nonrobustness of results given by equations 1-3 is of no concern if these equations do not represent a real physical process To avoid the problem of justifying the unjustifiable physical interpretation of parameters, we follow Lane (*14, 15*) and argue that the real aim of inference is usually to generate a prediction about the value of some future observables This goal is particularly appropriate when the model parameters do not represent "real" physical quantities In this case, the true values of parameters do not exist, and the precision of parameter estimates is not defined Parameter estimation may then be viewed as a "half-way house" on the road to predicting some relevant future observation Stochastic coefficients models are ideally suited to the problem of predicting future variables, as we shall see in the next article in this series of three articles

Conclusions

We have shown here that it is possible to develop an operational set of estimators for all the parameters appearing in a general stochastic coefficients model, but the precision of those estimators may be quite low The only way to improve this precision is to impose a large number of zero restrictions on the parameters of the model However, a stochastic coefficients model so restricted may reduce to a fixed-coefficients model of the conventional type and may suffer from contradictions We cannot accept contradictory restrictions Furthermore, even if certain restrictions do not contradict each other, the increases in the precision of the estimators resulting from these restrictions may be spurious More important, the low precision of an estimator of a parameter is a real cause for concern if the true value of the parameter exists We cannot be sure that the true value of a parameter exists unless we are sure that the model in which the parameter appears is true A model with contradictory premises is false, and the true values of its parameters do not exist

Since the premises of a fixed-coefficients model can be contradictory, we cannot be happy with the robust results that a fixed-coefficients model may give Econometric logic permits us to say only that, if a model is coherent (or free from contradictions), then it can be true We cannot establish the truth of a coherent model We prefer a stochastic coefficients model to its fixed-coefficients counterpart if we can establish only the coherence of the former but not of the latter Since the real aim of inference is prediction and not parameter estimation, we should not be overly concerned about the imprecision of parameter estimators given by a coherent stochastic coefficients model Therefore, any parameter estimates, however imprecise, are acceptable if they give successful forecasts of future observations and provide plausible explanations of past experience

References

1 Box, G E P , and G M Jenkins *Time-Series Analysis Forecasting and Control* 2nd ed San Francisco Holden-Day, 1976

2 Cooley, T , and E Prescott "Estimation in the Presence of Stochastic Paramater Variation," *Econometrica*, Vol 44, No 1, Jan 1976, pp 167-84

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

4 Doan, T , R Litterman, and C Sims "Forecasting and Conditional Projection Using Realistic Prior Distributions," (with discussion), *Econometric Reviews*, Vol 3, No 1, 1984, pp 1-144

5 Duncan, D B , and S D Horn, "Linear Dynamic Recursive Estimation From the Viewpoint of Regression Analysis," *Journal of the American Statistical Association*, Vol 67, No 340, Dec 1972, pp 815-21

6 Engle, R F "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, Vol 50, No 4, July 1982, pp 987-1008

7 Fisk, P R "Models of the Second Kind in Regression Analysis," *Journal of the Royal Statistical Society*, Series B, Vol 29, No 2, 1967, pp 266-81

8 Harvey, A , and G Phillips "Estimation of Regression Models with Time Varying Parameters," *Games, Economic Dynamics and Time Series Analysis* (eds M Deistler, E Furst, and G Schwodiauer) Vienna-Wurzburg Physica-Verlag, 1982, pp 306-21

9 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

10 Hildreth, C , and J P Houck "Some Estimators for a Linear Model With Random Coefficients," *Journal of the American Statistical Association*, Vol 63, No 322, June 1968, pp 584-95

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

12 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

13 Kmenta, J *Elements of Econometrics* 2nd ed New York Macmillan, 1986

14 Lane, D A "Discussion of the Likelihood Principle by J O Berger and R C Wolpert," *Lecture Notes-Monograph Series* Vol 6 Hayward, CA Institute of Mathematical Statistics, 1984

15 ________ "Comment on 'Some Subjective Considerations in the Selection of Models' by Bruce M Hill," *Econometric Reviews*, Vol 4, No 2, 1986, pp 253-58

16 Lehmann, E L *Theory of Point Estimation* New York John Wiley and Sons, 1983

17 Maddala, G S *Econometrics* New York McGraw-Hill, 1977

18 ———, and F D Nelson, "Maximum Likelihood Methods for Models of Markets in Disequilibrium," *Econometrica*, Vol 42, No 6, Nov 1974, pp 1013-30

19 Makelainen, T, K Schmidt, and G P H Styan "On the Existence and Uniqueness of the Maximum Likelihood Estimate of a Vector-Valued Parameter in Fixed-Sized Samples," *The Annals of Statistics*, Vol 9, No 4, 1981, pp 758-67

20 Meinhold, R J, and N J Singpurwalla "Understanding the Kalman Filter," *The American Statistician*, Vol 37, No 2, May 1983, pp 123-27

21 Pagan, A "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients," *Journal of Econometrics*, Vol 13, No 3, Aug 1980, pp 341-63

22 Pratt, J W, and R Schlaifer "On the Nature and Discovery of Structure" (with discussion), *Journal of the American Statistical Association*, Vol 79, No 385, Mar 1984, pp 9-33

23 Rao, C R "Simultaneous Estimation of Parameters in Different Linear Models and Applications to Biometric Problems," *Biometrics*, Vol 31, June 1975, pp 545-54

24 Rosenberg, B "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, Vol 2, No 4, Oct 1973, pp 399-428

25 Shiller, R J "A Distributed Lag Estimator Derived From Smoothness Priors," *Econometrica*, Vol 41, No 4, July 1973, pp 775-88

26 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 1, Spring 1988, pp 2-10

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

28 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, Sept 1975, pp 593-602

29 ——— "Further Results on Zellner's Minimum Expected Loss and Full Information Maximum Likelihood Estimators for Undersized Samples," *Journal of Business & Economic Statistics*, Vol 1, No 2, Apr 1983, pp 154-62

30 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, No 2, Feb 1980, pp 103-42

31 Swamy, P A V B, and P von zur Muehlen "Further Thoughts on Testing for Causality with Econometric Models" Special Studies Paper No 211 Federal Reserve Board, Washington, DC, 1988, and forthcoming issue of *Journal of Econometrics*

32 Szatrowski, T H "Necessary and Sufficient Conditions for Explicit Solutions in the Multivariate Normal Estimation Problem for Patterned Means and Covariances," *The Annals of Statistics*, Vol 8, No 4, 1980, pp 802-10

33 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, 1986, pp 1723-50

34 Zellner, A "Causality and Econometrics" (with discussion), *Carnegie-Rochester Conference Series on Public Policy, Vol. 10* (eds K Brunner and A H Meltzer) Supplement to *Journal of Monetary Economics* Amsterdam North-Holland Publishing Co, 1979, pp 9-54, 97-111