The Stochastic Coefficients Approach to Econometric Modeling
Part I: A Critique of Fixed Coefficients Models

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

Abstract. Stochastic coefficients models can provide accurate agricultural sector forecasts and useful policy analysis. Coefficient variation may occur for many reasons including aggregating over micro units, omitting variables, using an incorrect functional form, and allowing for a dynamic economic theory of optimizing behavior. In the first article in this series of three, we address the logical problems with fixed coefficients models. A number of auxiliary and possibly contradictory assumptions are imposed on econometric models to make them empirically manageable. In the second and third articles, we will show how stochastic coefficients models eliminate the logical problems associated with fixed coefficients models.

Keywords: Stochastic coefficients, fixed coefficients, aggregation, classical logic, probabilistic logic, evidential interpretation.

"The coefficients arrived at are apparently assumed to be constant for 10 years or for a larger period. Yet, surely we know that they are not constant. There is no reason at all why they should not be different every year."

John Maynard Keynes, 1938

The economics profession recognizes increasingly that the classical regression assumption of constant slopes is dubious. Indeed, if there were not this recognition, why adjust the constant terms or add factors to improve forecasting and policy simulations? By tampering with their models, econometricians implicitly acknowledge more variability in their models than they can capture by classical autoregressive errors. Furthermore, the necessity to apply such first aid to classically estimated models with or without deterministic shifts in coefficients appears to have increased over the past decade.

Many econometricians today remember the sixties as halcyon years when policymakers believed they could "fine-tune" the economy and determine an optimal policy mix, first, by simulating their models of conventional type under a variety of policy assumptions and, then, by reviewing the influence of their policy assumptions on the pertinent endogenous variables. The experience of the seventies irrevocably altered the sanguine attitude of those policymakers. Severe and dramatic structural shocks and shifts such as crop failures, the Soviet grain deal, the Soviet grain embargo, the poor Peruvian anchovy harvest in 1973, the movement from fixed to floating exchange rates, wage and price controls, the two energy price shocks, the payment-in-kind (PK) program, and institutional changes in the financial sector wreaked havoc on the vain attempts by econometricians to provide reliable and consistent policy recommendations from the results of their models with fixed or deterministically changing slopes. Stochastic coefficients models have been developed to address these and other problems.

This article is the first of three articles devoted to the evaluation of stochastic coefficients models. Interest in stochastic coefficients modeling is no longer confined to the general economic literature. In the past few years a number of agricultural economists have discussed and applied these new methods in agricultural settings. We believe many benefits for agricultural decisionmakers and the agricultural community will flow from these efforts. Our series will highlight the variety of models available, their usefulness, their assumptions, their limitations, and their possible contradictions. We first demonstrate that stochastic coefficients modeling rests on firmer philosophical and logical foundations as an econometric methodology than does conventional constant coefficients modeling. We list the reasons why coefficients may be stochastic and discuss the inherent

---

Swamy is a senior economist with the Board of Governors, Federal Reserve System, and Adjunct Professor of Economics at The George Washington University. Conway is an agricultural economist with the Resources and Technology Division, ERS, and LeBlanc is an agricultural economist with the Resources and Technology Division, ERS. The authors received valuable comments and help from James Barth, Arthur Havenner, Nadine Loften, Tom Luten, Ron Mattelhammer, and Peter von zur Muehlen.

1 Keynes' comment on a proof copy of Jan Tinbergen's Business Cycles in the United States of America, as quoted in (23, p. 286).

2 Italized numbers in parentheses refer to items in the References at the end of this article.

See (5, 7, 8, 9, 10, 25, 32, 33).
weaknesses in the conventional econometric approach to modeling that lead to logical inconsistencies. In the second article, we will show how a methodology based on stochastic coefficients in a time-varying mode avoids this logical quagmire. However, the generality of stochastic coefficients models leads to certain practical problems which we will also discuss. These problems become less serious if we recognize that the real aim of statistical inference is usually to generate a prediction about the value of some future observables, as shown in the second article.

Why Should Coefficient Variation Occur?

There are a number of theoretical and empirical justifications for specifying a stochastic coefficients model. First, there is no a priori reason why the "true" coefficients themselves cannot be generated by a non-stationary or time-varying random process, as the opening quotation from Keynes shows. Second, omitted variables that exhibit nonstationary behavior and that are not orthogonal to the included variables can induce variability in the coefficients of included variables. Third, it is a conventional econometric practice to use proxy variables in place of unobservable explanatory variables. As we know, proxy variables will only imperfectly capture changes in the economic behavior of the true variable, and the relationship between the "true" variable and its proxy may change in time. Fourth, aggregation over micro units can induce variation. It is highly restrictive to assume the aggregation weights of microeconomic units will not change over time (see 36, 37 for a discussion of this point). There are surely few observed events that are not already the outcome of some aggregation. Therefore, since the topic is important, let us consider the aggregation issue in detail.

A problem with constant coefficients arises naturally from aggregation in economics. Neoclassical economists begin with the assumption that economic agents optimize. From there, microeconomic models of an individual consumer demand (or factor input demand schedules for a firm) can be derived from a consumer utility function (or firm output or profit function) (see 31).

Because micro data are typically not available, researchers are forced to aggregate up to a macroeconomic model. Here one assumes a homogeneous structure when constant coefficients are employed. However, it is not reasonable to assume the same utility function for all individuals or that every firm in a given industry faces the same objective function. Instead, it is more reasonable to assume coefficients will change across individuals and firms. Furthermore, it is more reason-able to assume coefficients will vary over time as a result of changes in taste and technology, among other reasons.

To explore the aggregation issue, we first state a general micro equation as

\[ y_{it} = \sum_{j=1}^{k} \beta_{jt} x_{jt} + \epsilon_{it}, \quad i = 1, 2, \ldots, n_t, \; t = 1, 2, \ldots, T \] (1)

where \( y_{it} \) and the \( x_{jt} \) represent micro units' dependent and independent variables, \( n_t \) is the number of micro units, \( i \) indexes a cross-section unit, and \( t \) indexes time. To allow for a combined intercept and additive error term, we set \( x_{it} \equiv 1 \), that is, \( \beta_{jt} \) represents the sum of the intercept and an error term. Since micro data are frequently not available, we aggregate equation 1 over \( i \) to obtain a macro function. This aggregation procedure requires some extra assumptions.

One way to aggregate the data described by Swamy, Barth, and Tinsley (28) is as follows. Subtracting and adding the function \( \sum_{j=1}^{k} \beta_{jt} x_{jt} \) with time-dependent coefficients on the right side of equation 1 gives an equation that, after summing over \( i \) and dividing through by \( n_t \), shows the aggregate function to be

\[ \frac{1}{n_t} \sum_{i=1}^{n_t} y_{it} = \sum_{j=1}^{k} \beta_{jt} \frac{1}{n_t} \sum_{i=1}^{n_t} x_{jt} + \frac{1}{n_t} \sum_{i=1}^{n_t} \left( \beta_{jt} - \beta_{jt} \right) x_{jt} \] (2)

This equation can be written as

\[ y_t = \sum_{j=1}^{k} \beta_{jt} x_{jt} + \xi_t \] (3)

where

\[ y_t = (1/n_t) \sum_{i=1}^{n_t} y_{it}, \]

\[ x_{jt} = (1/n_t) \sum_{i=1}^{n_t} x_{jt}, \text{ and} \]

\[ \xi_t = \frac{1}{n_t} \sum_{i=1}^{n_t} \sum_{j=1}^{k} (\beta_{jt} - \beta_{jt}) x_{jt} \] (4)
Equation 3 with $\xi_t$ suppressed is theoretically plausible if $\xi_t$ converges in probability to zero for every $t$. This convergence holds if and only if both equation 1 and the condition that $E(|\xi_t|^r/(1 + |\xi_t|^r)) \to 0$ as $n_t \to \infty$ for some $r > 0$ and every $t$ hold (see 15, p 69). An alternative set of conditions for the convergence of $\xi_t$ to 0 in probability appears in (24). Serfling (26) states five theorems that also give alternative sets of conditions for the stochastic convergence of $\xi_t$ to zero.

In the case where all the elements of the individual-invariant coefficients $\beta_{it}$ ($j = 2, \ldots, k$) corresponding to the slopes of equation 1 are also time-invariant, equation 3 reduces to

$$y_t = \beta_{it} + \sum_{j=2}^{k} \beta_j x_{jt} + \xi^*_t$$

(5)

where

$$\xi^*_t = \frac{1}{n_t} \sum_{i=1}^{n_t} (\beta_{it} - \beta_i) + \frac{1}{n_t} \sum_{i=1}^{n_t} \sum_{j=2}^{k} (\beta_{jt} - \beta_j) x_{jt}$$

If this $\xi^*_t$ converges in probability to zero, then equation 5 reduces to

$$y_t = \beta_{it} + \sum_{j=2}^{k} \beta_j x_{jt}$$

(6)

Equation 6 is the same as an aggregate equation of the conventional type because, in the traditional fixed-coefficients models, the intercept may be interpreted as a random coefficient. We may consider the quantity $\beta_{it} = \beta_i + u_t$ as a random variable with mean $\beta_i$, whereas the slope coefficients $\beta_j (j = 2, \ldots, k)$ are fixed.

As is well known in statistics, we cannot achieve stochastic convergence without centering and scaling the variables. There are different ways of centering and scaling (see 15, pp 68, 280). The definitions of $\xi_t$ and $\xi_t^*$ are different because in these definitions the micro coefficients are centered differently. This difference is the only reason why conditions for the stochastic convergence of $\xi_t^*$ to zero are more stringent than those for the stochastic convergence of $\xi_t$ to zero. The point here is that the term $\xi_t$ is more likely to converge in probability to zero and will do so more quickly, if the micro coefficients are centered by subtracting time-varying coefficients as in equation 4 than if they are centered by subtracting fixed coefficients as in equation 5. This advantage by centering with time-varying coefficients means that the aggregate equation 3 with $\xi_t$ suppressed and with variable slopes can exist under weaker conditions than the aggregate equation 6 with fixed slopes and random intercept. For any practical work, the existence conditions are important because a model that does not exist could not have generated our data and should not be used for empirical analysis. Thus, if one wants to assume weaker conditions to show the existence of an aggregate model and to avoid aggregation bias, one should use aggregate models with time-varying slopes.

A fifth justification for specifying stochastic coefficients is that coefficient variation may occur as a result of an incorrect functional form. Because the true functional forms are usually unknown, equation 1 is preferable to a linear equation with fixed slopes because the former can trace any nonlinear path depending on the time profiles of $\beta_{it}$'s. As Rausser, Mundlak, and Johnson observe, “The approximation of highly nonlinear 'true' relationships by simpler functional forms, along with observations outside the narrow sample range, provides perhaps the strongest motivations for a varying parameter structure” (25). Goldberger (14, pp 108–11) also makes the same point while commenting on the Rotterdam school demand models. Finally, conventional econometric models may be in contradiction to the dynamic economic theory of optimizing behavior. A change in economic or policy variables will result in a new environment that will, in turn, lead to new optimal decisions and a new micro and macroeconomic structure. Thus, formalization of Keynes' intuitive insight is the contribution of Lucas (20). As Lucas and Sargent note, this “[e]quilibrium theorizing...readily leads to a model of how process nonstationarity and Bayesian learning applied by agents to exogenous variables lead to time-dependent coefficients in an economic agent's decision rule” (21). This dynamic optimizing behavior further justifies our use of micro equations of the form of equation 1 with all its coefficients varying across individuals both at a point in time and through time. As we have shown, using an aggregate equation with all its coefficients varying over time represents an attempt to deduce an aggregate equation from these plausible micro equations under weaker conditions. Equation 3 with $\xi_t$ suppressed can be called a stochastic coefficients model if its slopes are stochastic. We will see that another advantage of the variable coefficients models over the fixed-coefficients models is that the former are free from the contradictions that may exist in the fixed-coefficients approach.

**The Logical Cracks in Conventional Econometric Models**

Although fixed coefficients models appear simple, there are serious logical problems with this mode of modeling...
that further support using variable coefficients models
To illustrate, let us review the traditional models and
their assumptions.

Traditional econometric models posit two sets of as-
sumptions. First, there is a set of behavioral assump-
tions relating to the economic agents under study. Sec-
ond, there is a set of additional assumptions deter-
mined by the form of simplifications imposed by empir-
cal necessity on these behavioral assumptions.

The familiar matrix formulation of a linear system of
behavioral equations is

\[ Y = X\Gamma + B = U \]  

where \( Y \) is the \( T \times G \) matrix of observations on \( G \)
endogenous variables, \( \Gamma \) is the \( G \times G \) matrix of constant
coefficients of the endogenous variables, \( X \) is the \( T \times K \)
matrix of observations on the \( K \) exogenous variables, \( B \)
is the \( K \times G \) matrix of constant coefficients of the
exogenous variables, and \( U \) is the \( T \times G \) matrix of
disturbances. Assume also that \( U \) is mean independent
of \( X \), that is, \( E(U|X) = E(U) = 0 \) and the \( E(U|Y) = E(U) \) in general. The assumption of mean independence
is stronger than the assumption of uncorrelatedness and
weaker than the assumption of independence (see 6).

If \( \Gamma \) is nonsingular, we can obtain the reduced form as

\[ Y = X\Pi + V \]

where \( \Pi = -\Gamma^{-1} \) and \( V = U\Gamma^{-1} \)

How should one evaluate such a model or econometric
models in general? Boland suggests “the only objective
and nonarbitrary test to be applied to theories or
models is that of logical consistency and validity” (4, p.
24) According to Boland, although logically valid mod-
els can be true, logically invalid models cannot be true
(Further discussion of this topic appears in 29.) We,
therefore, document the eight explicit or implicit as-
sumptions an econometrician must make to estimate
models like equation 7 including some assumptions that
are not well known.

First, the conditions of Kagan, Linuk, and Rao’s
(KLR’s) lemma in equation (6) hold for a given set of
endogenous and exogenous variables.

This lemma provides conditions for the existence of a
set of linear regression equations, or linear population
regression functions of the form of equation 8 between
each of a set of endogenous variables and a set of exogenous variables in every time period \( t \). This result
suggests that a dependent variable in time \( t \) has an
expectation linear in, and a variance independent of, the
conditioning vector of independent variables. We have
already pointed out the importance of existence condi-
tions for empirical work. When a condition of KLR’s
lemma holds, described in (6, p 6), the disturbances in
equation 8 are mean independent of the independent
variables \( X \) and our assumption about \( X \) in equation 7 is
correct.

Second, the rank of \( X \) is \( K \) if \( K \leq T \) and the rows of \( \Pi \)
corresponding to the linearly dependent columns of \( X \)
are null if \( T < K \). Alternatively, the matrix \( \Pi \) satisfies
some exclusion restrictions even when the rank of \( X \) is
\( K \).

Third, for \( j = 1, 2, \ldots, G \), the \( G - G_j - 1 \) elements of the \( j \)th
column of \( \Gamma \) and \( K - K_j \) elements of the \( j \)th column of \( B \)
are zero. One element of each column of \( \Gamma \) is unity, and
\( \Gamma \) stays nonsingular even after these restrictions are
imposed.

Fourth, the rank of a specific \((K - K_j) \times (G_j + 1)\) subma-
trix of \( \Pi \), denoted by \( \pi_j \), is \( G_j \) for \( j = 1, 2, \ldots, G \).

Fifth, the matrix \( U \) is stationary or nonstationary, and
the distribution of \( U \) is normal or nonnormal.

Sixth, prior distributions of the unknown elements of \( \Gamma \),
\( B \), and the covariance matrix of \( U \) are of a particular
form.

Seventh, either the system in equation 7 is interdepen-
dent or triangular-recursive in the sense of Wold (see 1).

Eighth, parameters \( \theta \) and \( \eta \) indexing the conditional and
marginal distributions in the equation \( F(Y|X, \theta) = F(Y|X, \theta)F(X|\eta) \) are unrelated in the sense of Basu (2, pp. 364-66).

Econometricians typically ignore the marginal distribu-
tion of \( X \) and base their inferences mainly on the
conditional distribution of \( Y \), given \( X \). It is difficult to
say whether the eighth assumption is satisfied if the
marginal distribution of \( X \) is ignored. In any case, the
Bayesian definition of parameter unrelatedness is clear.
In a frequentist (classical) context, the elements of a
vector of unknown constants like \( \theta \) can be said to be
unrelated to the vector of unknown constants like \( \eta \) if it
is possible to meaningfully isolate all the relevant infor-
mation about each element of \( \theta \) contained in the data.

Frequentists’ definitions of this concept of parameter
unrelatedness could not pass Basu’s (2) careful scrutiny,
although the Bayesian definition did.

With an abuse of notation we are using here the same symbol to
denote a random variable and the value taken by the random variable.
Econometricians may prefer model 7 to model 8 because the former, unlike the latter, may represent an economic law, in which case it has a natural causal interpretation. This interpretation follows from Feigl's definition of causality as predictability according to a law or set of laws (see 35).

By means of classical logic, we may establish sufficient conditions for the truth of model 7. Using the standard truth tables, given, for example, in McCawley (22), we may evaluate the truth of the eight assumptions and model 7 if each of the constituents of the statements (assumptions one through eight, model 7) is either true or false (but not both). The set (if assumptions one through eight, then model 7, given assumptions one through eight) semantically entails model 7 if, in all states of affairs in which assumptions one through eight are all true, model 7 is also true.

It is unfortunate one cannot guarantee that all eight assumptions are either true or false (but not both). Indeed, a serious flaw with fixed-coefficient models depicted in models 7 and 8 is their lack of logical clarity, thereby disallowing any claims about the logical validity of these models. More precisely, the set of behavioral assumptions implied by model 7 and the eight auxiliary assumptions required for estimation may conflict with each other. For example, Swamy (27) has shown the restrictions of assumption two imposed on reduced-form coefficients may contradict the identifying restrictions of assumption three imposed on structural coefficients. Furthermore, consider the case when an econometrician does not use all the exogenous variables in the system to compute an instrumental variable estimator. Use of a subset of exogenous variables in a system implies certain zero restrictions on the reduced-form coefficients that might contradict structural identifying restrictions. Therefore, restrictions on the reduced-form coefficients may be implicitly or explicitly imposed even when the matrix of observations on the exogenous variables, \( X \), has full column rank.

Model 7 exists if assumption one is true and \((K - K_g)\) equations, \( [\gamma^*_j, \Pi_j] [\gamma_0, - \gamma_j] = 0 \), in \( G + 1 \) unknowns, \( [\gamma_0, - \gamma_j] \), have a unique-up-to-a-fixed-factor-of-proportionality solution for \( j = 1, 2, \ldots, G \) (see 22, p 315). If the equations \( \Pi_j^{*} [\gamma^{*}_0, \gamma_j] = \gamma^*_j \) are inconsistent, which can happen when \( K - K_g \geq G \), and assumption one constacts assumption two or three, then model 7 does not exist under the eight assumptions.

Although, in principle, instrumental variables techniques can be used to estimate model 7 under the eight assumptions, these techniques rest on the hypothesis that certain observed variables used as instruments are truly exogenous and yet have an important influence on the endogenous right-hand variables with nonzero coefficients. These two requirements are contradictory if assumptions one and two contradict assumption three. Instrumental variables estimation in this case would be based on an uneasy compromise where the exogeneity of the instruments is uncertain. Thus, the use of certain observed variables as instruments may imply an unwanted imposition of a set of contradictory model restrictions. The possibility of contradictory model restrictions puts into question the entire class of proofs that have been used to establish the statistical consistency of instrumental variables estimation.

These examples do not end the discussion of possible contradictions among the eight assumptions. The nonstationarity condition of assumption five may imply that some of the errors in model 7 are heteroscedastic. In this case, not only is a restriction of assumption one false but so is assumption eight because those elements of \( \theta \) that represent the variances of heteroscedastic errors are related to exogenous variables. Thus, assumptions one and eight may contradict the nonstationarity or nonnormality implied by assumption five.

Under the rules of inference adopted by classical logicians, conclusions based on contradictory premises are arbitrary in the sense that both the truth and falsity of a structural model can be inferred from its contradictory premises (see 22, pp 29–30).

The above discussion does not capture all legitimate forms of reasoning. There is another system of logic known as probabilistic logic. The set (the eight assumptions) is logically inconsistent if and only if both model 7 and not-model 7 can be inferred from those assumptions. As we have shown, according to classical logic both model 7 and not-model 7 can be inferred from the eight assumptions if these assumptions contradict each other. However, the concept of logical consistency is relative to a given system of rules of inference (or set of general principles) that specifies what conclusions may be inferred from what premises (see 22, p 41). Thus, a set of propositions that is inconsistent with respect to the classical system of rules of inference could very well be consistent with respect to the system of statistical rules of inference. Therefore, let us check the consistency of the set (the eight assumptions) with respect to the latter system.
Suppose the eight assumptions impose restrictions on \( \Gamma, B \) and \( II \), which make the equation, \( F(Y|X;\theta) = F_r(Y|X,\theta)F_s(X|\eta) \), fail, then they are inconsistent relative to probability axioms, since consistency requires that both \( F_r(Y|X,\theta) \) and \( F_s(X|\eta) \) correspond to some joint distribution of \( Y \) and \( X \). Lane and Sudderth \((18)\) have constructed conditions for the existence of a joint distribution with given conditional distributions, showing that Bayesian inferences based on a conditional distribution, denoted by \( F_3(\theta|Y, X) \), are consistent with some non-Bayesian inferences based on the conditional distribution \( F_r(Y|X;\theta) \), if and only if both correspond to some joint distribution, denoted by \( F_4(Y, \theta|X) \). These are difficult conditions to check, particularly if a given structural form is nonlinear. In that case the reduced form may not exist and, even if it exists, it may not be unique. Because Lane and Sudderth's consistency conditions may not be satisfied in the nonlinear case, specifying a logically consistent, identifiable nonlinear structural model may be an intractable problem.

Thus, when the linearity assumption in model 7 is relaxed and we assume the structural relationships are nonlinear, an explicit form for the corresponding reduced form may be impossible. One can try, as Gallant and Holly \((12, p 699)\) do, to simply assume the existence of a unique reduced form and add this auxiliary assumption to the ever-expanding laundry list. It is unfortunate that our \textit{a priori} information may be insufficient to conclusively eliminate any possible contradiction between the identifying restrictions on the structure and the conditions for the existence of the nonlinear reduced form or the population nonlinear regression functions of the form \( E(Y_i|X_{1i}, \ldots, X_{Gi}) = g(X_{1i}, \ldots, X_{Gi}) \). In this example, as with the others, any virtues of the structural equation approach may founder beneath the waves of logical contradiction.

Finally, some of the eight assumptions require modifications if a subset of the right-side variables in model 7 represents unobserved expectations, as in rational expectations models \((34)\). In these models, subjective probabilities are equated to the probabilities implied by structural models. Such an equation is unreal because the conditions under which a frequency (or subjective) interpretation of probability is correct are neither necessary nor sufficient for the subjective (or frequency) interpretation of probability to be valid \((28, 29)\). Thus, the conjunction of the conditions under which frequency and subjective interpretations are correct is unrealistic and the rational expectations models violate Aristotle's axiom of identity: different statements in an argument should not use different definitions of the same words. Moreover, one can estimate the structural models like model 7 involving the rational expectations variables using the conditional expectations that are generated by stationary time series models. Swamy and von zur Muehlen \((30)\) list the conditions under which the stationary time series models of various forms exist. These conditions may contradict the eight assumptions that are required for the existence of linear structural models. As a result, the premises of a conjunction between a stationary time series model and a structural model can be contradictory.

According to classical logic, models with contradictory premises or models that violate Aristotle's axioms of logic are not admissible into logical arguments. There is no guarantee that the premises of a fixed coefficients structural model are not contradictory. By means of probabilistic logic, which is different from classical logic, coherent inferences are not possible if the specifying assumptions underlying a model violate the probability laws. There is also no guarantee that the assumptions underlying a fixed coefficients structural model do not violate the probability laws.

The Lane Question: "When Will We Stop Taking Parameters So Seriously?"

Besides the risk of logical inconsistencies and contradictions, there are deeper philosophical issues to raise regarding the fixed coefficients paradigm. Hypothesis testing traditionally plays a fundamental role in motivating the estimation of fixed coefficients. However, there are serious problems related to the meaning, adequacy, and relevance of the fixed parameter paradigm as typically employed in econometrics.

One may interpret an econometric model by assuming there is a joint probability distribution of the current endogenous variables conditional on the values of the exogenous variables. This distribution can then be written as:

\[
P_\theta = F(y_i|x_i,\theta)
\]

where \( \theta \) is a fixed parameter vector taking values in a parameter space \( \Theta \), and where \( S \) is a sample space in which \( y_i \) takes on its values.

The appropriateness of inference procedures that can be applied to equation 9 depends on the interpretation of \( \Theta \). Lane \((17)\) defines at least three possible interpretations of the elements of \( \Theta \):

1. \( \theta \) is the distribution equation 9;
2. \( \Theta \) is an abstract set, and \( \theta \) simply indexes the distribution equation 9.

\footnote{Lane meant that only experiments whose sampling distributions are identical share "the same \( \theta \)." That is, Lane's interpretation 1 means that the elements of two parameter spaces \( \Theta_1 \) and \( \Theta_2 \) are not the same parameter unless the distributions \( F(y_i|x_i,\theta_1) \) and \( F(y_i|x_i,\theta_2) \) are identical.}
3 $\theta$ is a possible value for some "real" physical parameter, and the distribution equation 9 is to be regarded as the distribution of the random vector $y_t$, given $x_t$, should $\theta$ be the true value of that parameter.

The choice Lane poses is important since both classical inferences and decision theory as well as the likelihood principle (LP) require that interpretation 3 holds (see 19, p 1, and 17). The statistical notions of consistency and efficiency are without meaning unless the true value of $\theta$ exists. If interpretation 3 does not hold, then the true value of $\theta$ may not exist. Furthermore, consistent estimation and testing of $\theta$ is impossible unless $\theta$ is identifiable because identification is a necessary condition for statistical consistency (19, p 335). However, interpretation 3 also elicits formidable philosophical questions, principally, when—and in what sense—do "real" physical parameters exist? It rather strains credibility to believe there are model-free physical quantities underlying each model parameter without some guidance as to what constitutes "reality" and how "reality" is linked to the mathematics congealed in specific models. It is unfortunate that it is not possible to give such guidance because, for reasons already mentioned, econometricians cannot prove the factual truth of their models (see also 29). In the absence of such a proof we could be wrong if we interpret the elements of $\theta$ as "real" physical quantities.

Interpretation 1 allows no scope for the mixture principle only experiments whose sampling distributions are identically share the "same $\theta$". As such, the LP is devoid of interesting consequences under this interpretation (17). Interpretation 2, in contrast, allows tremendous scope for mixing. Any two experiments with the same index set can be mixed. In this case, the LP is wrong, as shown by Lane (17). We will discuss how stochastic coefficients models relate to interpretation 2 in the second article.

Although econometricians may use economic models to guide the formulation of inference, the inferences have value to us only if they yield useful statements about the real world. Hypotheses stated in terms of the values of $\theta$ refer to the real world if $\theta$ has a physical interpretation or interpretation 3 is correct. Any evidence against a hypothesis about the world is useful if that evidence has a small probability of occurring when that hypothesis is true. Birnbaum's theory of evidential interpretation is most pertinent here. Evidential interpretations are in the form $d_1^* = (\text{reject } H_0 \text{ in favor of } H_1, \alpha_1, \beta_{11})$ and $d_2^* = (\text{reject } H_1 \text{ in favor of } H_0, \alpha_1, \beta_{11})$, where $\alpha_1$ is the probability of type I error and $\beta_{11}$ is the probability of type II error.

This intuitively appealing concept of evidence stating that under no $\theta$ shall there be a high probability of outcomes being interpreted as "strong evidence against $\theta$" has been articulated by Birnbaum (3) in the following terms:

The Confidence Concept. A concept of statistical evidence is not plausible unless it finds "strong evidence for $H_1$ against $H_0$" with small probability $\alpha_1$ when $H_0$ is true and with much larger probability $(1 - \beta_{11})$ when $H_1$ is true.

Birnbaum's confidence concept makes sense because no satisfactory justification of the choice of a test statistic exists except in terms of the alternatives of interest. Econometricians have all faced situations in which the propriety of considering large or small values of a test statistic as significant rests on the alternatives of interest. There may be little reason to use a given test statistic except in light of certain alternatives.

Birnbaum (3) interprets the decision $d_1^* = (\text{reject } H_0 \text{ in favor of } H_1, 0.01, 0.2)$ as very strong (but inconclusive) evidence for $H_1$ against $H_0$, and he interprets the decision $d_2^* = (\text{reject } H_0 \text{ in favor of } H_1, 0.5, 0.5)$ as worthless evidence. The latter evidence is no more useful than the result of a toss of a fair coin since the error probabilities $(0.5, 0.5)$ also represent the experiment of tossing a fair coin, with one side labeled "reject $H_0$" and the other "reject $H_1". Consequently, the precise values of $\alpha_1$ and $\beta_{11}$ go a long way toward interpreting statistical evidence, provided interpretation 3 is correct. Anything less than reporting the values of both $\alpha_1$ and $\beta_{11}$ would not achieve the full disclosure required to make a researcher's study convincing to a critical reader and palatable to a casual reader.

The validity and usefulness of the statistical conclusions based on fixed coefficients models appear to depend on two premises about the nature of inferences in the $(S, \theta, P_\theta)$ paradigm. First, the purpose of statistical inference is to make some statement about the "true" value of an unobservable parameter $\theta$ on the basis of the observed behavior of certain random variables. Second, $\theta$ exists independently of the specified model that presumably produced the given data, and information about $\theta$ can be separated into two components, one deriving just from the model and "other information" presumably preexisting the model. With all the aggregation and logical inconsistency problems we have pointed out, these two premises are rarely true in the practical situations to which statistical inference is applied, especially in econometrics. In the second article, we discuss the stochastic coefficients approach and show how it addresses these logical problems and philosophical issues.
Conclusions

In this article we have shown that the conventional fixed coefficients modeling approach must employ several and possibly contradictory auxiliary assumptions to be operational. Underlying the traditional technique is a philosophic stance on the nature of parameters that is hard for a social scientist to swallow (and, according to Lane (16), a mouthful for the physical scientists as well). In the next article, we will show that stochastic coefficients models ease the number of assumptions and allow a more reasonable foundation to interpret parameters. Thus, we offer a reasonable alternative to the fixed-coefficients approach to counter the pessimistic prospect for econometric modeling suggested in this article.

References


16 Lane, D A “Comment on ‘Some Subjective Considerations in the Selection of Models’ by Bruce Hill,” Econometric Review, Vol 4, 1986, pp 253-58


19 Lucas, Robert E “After Keynesian Macroeconomics,” After the Philips...

22 McCawley, J D. Everything That Linguists Have Always Wanted to Know About Logic (but were ashamed to ask) Chicago The Univ of Chicago Press, 1981


