UC Berkeley
CUDARE Working Papers

Title
Structural change, parameter variation, and forecasting

Permalink
https://escholarship.org/uc/item/094565f3

Authors
Mindlak, Yair
Rausser, Gordon C.

Publication Date
1976-12-01

eScholarship.org Powered by the California Digital Library
University of California

Department of Agricultural & Resource Economics, UCB
CUDARE Working Papers
(University of California, Berkeley)

Year 1976 *Paper* 76

Structural change, parameter variation, and forecasting

Yair Mindlak Gordon C. Rausser
University of California, Berkeley

This paper is posted at the eScholarship Repository, University of California.
http://repositories.cdlib.org/are_ucb/76
Copyright ©1976 by the authors.

CUDARE,

76

Division of Agricultural Sciences
UNIVERSITY OF CALIFORNIA

Working Paper No. 76

STRUCTURAL CHANGE, PARAMETER VARIATION, AND FORECASTING

by

Yair Mundlak and Gordon C. Rausser

GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS
LIBRARY
FEB 25 1980

GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS
LIBRARY
APR 12 1990

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
December 1979

STRUCTURAL CHANGE, PARAMETER VARIATION, AND FORECASTING*

Yair Mundlak† and Gordon C. Rausser††

1. Introduction

Attempts to capture useful relationships for forecasting purposes in non-experimental sciences such as economics have long been fraught with formidable obstacles. Many of these obstacles relate to *unobservables* and the lack of controlled effects. Important unobservables are generally associated with the evolution of taste, the formation of expectations and anticipations, and sociological phenomena. In the context of controlled effects, economists generally operate with some version of the classical linear statistical model; this model presumes that the effects remain constant over all sample observations.

For most economic systems, the assumption of parameter consistency is imposed in the face of noncontrolled effects and many important unobservable influences. Economists and other social scientists often neglect the problem of isolating an appropriate set of data for which it appears reasonable to assume parameters are "approximately constant." This problem is at the heart of the issues faced by the classical framework of experimental design. Economists, of course, have recognized that different data sets often result in noticeably different coefficient estimates. Perhaps the best example of this recognition is the typical treatment of pre- and postwar data on economic phenomena. To account for the difference in effects between pre- and postwar data, the general practice has been to introduce dummy variables to allow for possible significant shifts in intercept and slope parameters.

Although the dummy variable approach is indeed convenient, in may instances it will lead to grossly inaccurate forecasts. Neglecting issues of complexity, such specifications may be suboptimal. In a time series context,

it is appealing to view the data from, say, 1970 as more relevant to forecast for 1980 than data from the period of the early 1950s. Taste, expectation formation patterns, and sociological phenomena in 1970 are clearly far different than 1950; and it would appear that in this sense the year 1970 contains more relevant information than earlier periods in forecasting future phenomena.

In agricultural economic forecasting, the above observations appear particularly relevant. Models constructed for various commodity systems, especially feed and food grain commodity systems based on data prior to 1971, have revealed notorious inaccuracies in forecasting important economic variables for the years 1972 through 1976. These gross inaccuracies proved to be a bitter disappointment to the Cost of Living Council during the years 1972 and 1973 in their attempts to control inflation. In addition to the obvious need for respecification of the basic commodity models, other issues related to the conventional use of the constant parameter formulation have naturally arisen.

In other fields of economics, researchers have begun to question the validity of the constant parameter formulation. The potential advantages of utilizing the forthcoming information to update or revise estimates of the coefficients of econometric models has been clearly demonstrated. In part, this increased awareness has resulted from the growing body of evidence, both of a conceptual nature and from empirical observations, that parameters of econometric models generally change over time. For macroeconometric models, the empirical evidence is reported in Duffy (1969), Cooper (1972), and the experience of the econometric consulting community. Practitioners have noted the forecasting accuracy benefits that can be obtained from

mechanically "adjusting the constant term."[1] Evidence has also accumulated on the parameter variability associated with wage/price data, especially structural shifts in conventional Phillips Curve formulations. Moreover, numerous studies at the microlevel have revealed parameter instability [see, for example, Balestra and Nerlove (1966)].

It seems, therefore, that more attention should be devoted to modeling processes where the parameter effects themselves are subject to various sorts of perturbations. Although such modeling processes have appeared with increased frequency in statistics, as well as the quantitative economic literature, there is nevertheless much uncertainty about the value of such approaches in an empirical setting. Even though economists have recognized the possibility of parameter instability, the complexity of pinpointing the nature of such variations has caused the profession to gravitate to various constant parameter formulations. A view of many empirical researchers seem to be that much of the recent conceptual work on parameter variation in both statistics and economics represent new gimmicks which contribute little in the way of useful empirical information.

To be sure, there is some merit in considering certain parameter effects as relatively fixed in economic models. The principal merit is simplicity in providing insights about economic interrelationships, unclouded by the meticulous details of a changing real world. There is also, however, some merit in obtaining more accurate representations, forecast, and economic policy evaluations. Hence, from a research strategy standpoint, the principal issue faced in deciding whether to employ a constant parameter formulation or some varying parameter formulation depends critically upon the trade-off between inaccuracy and complexity. In the final section of this paper, the issue of complexity

will be discussed to provide a set of guidelines that researchers hopefully will find valuable in selecting an appropriate formulation.[2]

From the standpoint of potential model inaccuracies, many justifications can be offered for the parameter variation formulation. First, the "true" coefficients themselves can be viewed as generated by a nonstationary or time-varying random process. Numerous authors have dealt with a special case of this view, namely, the random, stationery parameter formulation. This formulation has been advanced principally for time series and cross-section observations [Rosenberg (1973b)]. In particular, a cross section of individuals are presumed to possess the same regression regime over time but whose individual behavior at a given point in time is viewed as a random sample from a population of coefficients with a constant mean. When the mean is allowed to change over time, there is the more general nonstationery formulation.

Even when the underlying "true" parameters are stable, situations arise in which the parameter variation approach will prove valuable. By their very nature, econometric models are abstractions; they involve simplifications imposed by available data, research time, and budgets as well as the desire to achieve tractable results. Such simplifications and abstractions often result in misspecifications which in turn influence the degree of accuracy of the model's forecasts. The effects of such misspecifications can be countered by an appropriate parameter variation structure. The most important types of misspecifications which arise in the construction of forecasting models include omitted variables, the use of proxy variables, the use of aggregate data, and nonlinearities.

The omission of important explanatory variables arise from inadequate theoretical frameworks, unavailable data, or the desire for simplicity. Such excluded variables often relate to structural changes resulting from taste evolution, technological developments, changes in institutional arrangements, and the like. The effects of such excluded variables are presumed to be random with a distribution which has a time invariant mean and variance. Such variables will have no effect upon the parameter effects of included variables, provided the omitted variables are independent of those that are included. However, time series for such omitted variables exhibit nonstationery behavior and are often correlated with the included variables. Under these circumstances, the estimated effects of the included variables can be expected to change with time. At a minimum, it seems reasonable to expect that excluded variables with nonzero effects will result in time variations in the intercept or constant effect.

Due to data limitations, proxy variables are often employed in the construction of econometric models. Such proxy variables are invariably introduced into dynamic representations which involve expectation formation patterns and measures of capital. Unfortunately, these and other proxy variables detect only partial changes in the level of economic stimuli they purport to measure. Furthermore, the relationship between the desired variable and its proxy can be expected to change over time. Under these circumstances, changes in the desired variable which measures the actual economic stimuli induces instability in the estimated parameters associated with the proxy variables.

In the context of aggregate data, the potential for parameter instability has been demonstrated on numerous occasions. Since aggregate data are measured by weighting the relative importance of the heterogeneous sets

of microunits, the parameters in the estimated aggregate equation will remain constant so long as these weights do not vary. In the context of time series data, the assumption of constant weights (i.e., relative importance of the individual components of the aggregate remains unchanged) is indeed unlikely to be satisfied. Hence, since shifts in the aggregation weights are the rule rather than the exception, parameter effects associated with the aggregate variables in the estimated model will vary across time [Zellner (1962)].

Another potential cause for parameter variation arises from the inappropriate specification of functional forms. For example, if under the pretext of a Taylor series expansion a linear relationship is estimated as an approximation to a nonlinear equation, the assumption of constant parameters for the simplified equation constitutes a reasonable approximation only if the observed explanatory variables remain within some narrow range. For variations beyond this range, it is a simple matter to demonstrate the nature of parameter variation for the simplified equation. Moreover, the secular evolution of many economic time series strongly suggest the rejection of any model that is based upon the assumption of narrow sample ranges. In general, the approximation of highly nonlinear "true" relationships by some simpler functional form along with observations outside a narrow sample range provides perhaps the strongest motivations for a varying parameter structure.

In addition to the misspecification rationale for varying parameter formulations, economic theory can also be advanced to justify their potential relevance. In many situations, the very nature of economic theory leads us to expect relationships that change over time. Lucas (1976), for example, has argued that the constant parameter formulation is inconsistent with economic theory. He notes that a change in policy, for example, will result in a change

in the environment facing economic units; and under the assumption that such units behave rationally, this will result in shifts of the equations representing the behavior of such units.

In the case of commodity systems, a number of illustrations of the points raised by Lucas are available. Perhaps one of the best examples occurred recently in the U. S. livestock sector. As the result of the United States Economic Stabilization Program over the period 1971-1974, price ceilings were imposed on red meats at the end of March, 1973. These ceilings, when combined with the biological nature of various red meat animals, led to a distortion and clouded price signals which in turn resulted in strategic errors on the part of various decision-makers. These signals, of course, led to instability in the expectation formation pattern of various decision-makers along the vertical commodity chains in beef, pork, and poultry. During this period, the cattle cycle poised for a sizable liquidation, was substantially altered by the distorted signals. Price ceilings in fact became the expected prices of producers for a short period of time. As a result, the liquidation phase was curtailed which in turn provided the basis for larger supplies, substantially lower prices, and significant negative margins. These dynamic effects resulted in an extended liquidation phase which exceeded all expectations. Hence, the effects of the price ceilings had the immediate effect of a substantial shift in price expectations which in turn resulted in rather drastic implications for dynamic supply responses and thus ultimate market realizations and cattle inventories.

Dynamic economic theory and the notion of rationale expectations does not in general provide us with well-defined behavioral equations whose

parameter effects can be treated as constant. In particular, Lucas notes that for the individual decision problem: "Find an optimal decision rule when certain parameters (future price say) follow an arbitrary path is simply not well formulated." He goes even further to suggest that, as expectations of future policy behavior change, economic theory predicts that this will result in shifts in the relevant estimatable behavioral equations. This source of parameter instability can only be avoided by reasonably accurate measurements of expectation formulation patterns and dynamic responses--a dubious hope at best.

Finally, in employing constructed models for various purposes, it is crucial that the models be tractable and interpretable. In using models for generating forecasts and/or policy analysis, a number of difficulties arise due to the model dimensions and problems of numerical accuracy. The accuracy is indeed an important issue when the structural model representation is non-linear. For such models which involve simultaneous representations of large systems, it is not possible to obtain a unique reduced form. In computing the necessary derivatives to obtain the reduced form for nonlinear models, issues of approximation and round-off problems naturally arise. More importantly, it is not possible to derive reliability statistics for highly nonlinear models. Thus, measures of forecast variance and risk associated with various policy actions are generally swept under the rug for such model representations. These problems can be largely avoided by specifying models as linear in the variable space but in essence nonlinear in the parameter space. By varying the parameter effects, any nonlinear representation can be appropriately approximated [Rausser (1978)]. This approach allows forecast probability distributions--unconditional or conditional on alternative policy actions--to

be generated for a particular point in the parameter space. Along similar lines, note that the approach also simplifies the validation of constructed models, especially the derivation of its dynamic properties.

The cumulative implications of the above observations is that, given the forces often neglected in economic models, it seems overly optimistic to presume that parameters will be identical over the complete sample record regardless of whether the model is linear or nonlinear. From an operational standpoint, the relevant issue is whether or not the explicit recognition of varying parameters will provide accuracy and implementation benefits which outweigh the additional construction complexities of such formulations. Can these formulations capture the enduring characteristics of the processes under examination? The purpose of this paper is to point us in the direction of answering this question in a definitive manner.

2. Parameter variation specification

In order to present the problem in the simplest possible form, a single equation with one explanatory variable will be presented, viz.,

$$y_t = x_t \beta_t + u_t, \tag{1}$$

where

$$u_t \sim \left(0, \sigma_u^2\right), \; E\,(u_t x_t) = E\,(u_t \beta_t) = 0, \; V_t.$$

Note that the parameter β_t is allowed to vary over time. The changes in β can be induced or be completely random without any structure imposed. Induced changes can be caused by the outside environment and be completely exogenous to the system or alternatively can be induced by variables in the

system itself. For instance, in demand equations it is possible that the demand coefficient depends on past consumption. A commodity can be habit forming or alternatively its consumption can be at a lead saturation point. In both cases the actual demand response depends on past experience. Furthermore, social change and its effect on the taste can lead to variation in the parameter structure. One instance of this behavior has been observed for the case of meat demand. During the early 1970s, forecasts of meat demand based on sample data through the year 1970 are consistently below actual levels of demand. One possible explanation for such forecasts is provided by the demand for convenience foods emanating from the women's liberation movement and the participation of females in the U. S. labor force. These and similar influences can be represented in a parameter variation equation by variables from outside the system (denoted by z) while factors associated with habit formation and saturation may be represented by variables appearing in the system denoted by the general function of L(x). These arguments can be summarized by

$$\beta_t = \beta_o + L(x_t) + z_t\alpha + e_t \qquad (2)$$

where

$$e_t \sim \left(o, \sigma_e^2\right), \quad \forall t$$

$$E\ (e_t z_t) = E\ [e_t\ L(x_t)] = 0.$$

As indicated above, the term z_t represents the outside effect on the coefficients whereas the term $L(x_t)$ represents the effect of the variables within the system. When these two effects are not present, then β_t is described as a random variable with mean β_0 and an error term e_t. In the most

special case when the variance of e is 0 and the effects z_t and $L(x_t)$ are not present, then β_t becomes constant coefficient which is the usual regression analysis formulation.

More generally, however, β_t is conceived as a random variable with a systematic component affected by the two forces described above. Combining equations (1) and (2), equation (3) is obtained which differs from equation (1) by the addition of two terms, viz.,

$$y_t = x_t \beta_o + L(x_t)\, x_t + x_t z_t \alpha + \varepsilon_t \tag{3}$$

where

$$\varepsilon_t = x_t e_t + u_t.$$

The third term in (3) is the product or interaction between x_t and z_t, and the second is the interaction between x_t and the general function $L(x_t)$.

Note that the variable z_t can be a qualitative variable (0, 1 variable) which introduces a switch in the regression coefficient. Of course, there can be more than one switch; and the introduction of more than one switch is a straightforward extension. The function $L(x_t)$ is a general function, but its main feature is that it is specified as a function of past experience in the x's. In order to detect an important and interesting feature of this function, it may be expressed as a geometric distributed lag function, i.e.,

$$\beta_t = \beta_o + \delta x_{t-1} + \delta^2 x_{t-2} + \ldots + \alpha z_t + e_t. \tag{4}$$

Multiplying equation (2.4) by δ and subtracting $\delta\beta_{t-1}$ from β_t, the following equation is obtained:

$$\beta_t = \delta\beta_{t-1} + (1 - \delta)\, \beta_o + \delta x_{t-1} + \alpha\, (z_t - \delta z_{t-1}) + (e_t - \delta e_{t-1}). \tag{5}$$

Note that equation (5) expresses β_t as proportional to β_{t-1} additional terms. It is, therefore, clear that expression (5) is a generalization and extenstion of the Markov chain formulation examined by Rosenberg (1973a) and Belsey (1973a,1973b); it is also widely used in the engineering literature. When δ is a fraction between 0 and 1, the term $\delta\beta_{t-1}$ represents a decay process. The convergence of this process is to β_o $(1 - \delta)$ plus the terms involving x_{t-1} and z_t. Hence, expression (5) generalizes the Markov process where the convergence is to a term β_o.

It should be noted that equation (3) has a heteroscedastic structure. Under the assumption of independence between u and e, for the variance of the error term ε_t the following expression is obtained:

$$\sigma^2_{\varepsilon_t} = x_t^2 \sigma_e^2 + \sigma_u^2. \tag{6}$$

2.1. Alternative specifications

The above framework admits a number of specifications which have appeared in the literature. The specification in (5) is obtained under the assumption that $L(x_t)$ has a specific distributed lag form. Other forms are possible and can be admitted by rewriting expression (5) as:

$$\beta_t = \delta_o\beta_o + \delta_1\beta_{t-1} + \delta_2 x_{t-1} + \alpha z_t^* + \xi_t \tag{7}$$

where the parameters associated with β_{t-1} and x_{t-1} differ. For (7) to be equivalent to (5), $\delta_1 = \delta_2 = \delta$; $\delta_o = 1 - \delta$; $z_t^* - \delta z_{t-1}$; and $\xi_t = e_t - \delta e_{t-1}$.

This general representation embeds as special cases all of the parameter variation specifications which have been introduced in the literature. More specifically,

a. Classical linear model: $\delta_o = 1$, $\delta_1 = 0$, $\delta_2 = 0$, $\alpha = 0$, $\xi_t = 0$. Hence,

$$\beta_t = \beta_o \text{ for all } t. \tag{7a}$$

b. Cooley and Prescott (1973a) adaptive regression model: $\delta_o = 0$, $\delta_1 = 1$, $\delta_2 = 0$, $\alpha = 0$, $u_t = 0$, $x_t = 1$, for all t. Hence,

$$\beta_t = \beta_{t-1} + \xi_t. \tag{7b}$$

c. Belsey (1973b) systematic parameter variation model: $\delta_o = 0$, $\delta_1 = 0$, $\delta_2 = 0$. Hence,

$$\beta_t = \alpha z_t + \xi_t. \tag{7c}$$

d. Swamy (1970) random coefficient model: $\delta_o = 1$, $\delta_1 = 0$, $\delta_2 = 0$, $\alpha = 0$. Hence,

$$\beta_t = \beta_o + \xi_t. \tag{7d}$$

e. Cooley and Prescott (1973b) time-varying parameter model: $\delta_o = 0$, $\delta_1 = 1$, $\delta_2 = 0$, $\alpha = 0$, $u_t = 0$, $\xi_t = \eta_t - \nu_{t-1} + \phi_t$. Thus,

$$\beta_t = \beta_t^* + \eta_t,$$

$$\beta_t^* = \beta_{t-1}^* + \phi_t,$$

and

$$\beta_t = \beta_{t-1} + \alpha_o w_t + \alpha_1 w_{t-1} \tag{7e}$$

where $w_t = [\phi_t, \eta_t]$; $\alpha_o = [1, 1]$; and $\alpha_1 = [0, -1]$.

f. Singh et al. (1976) mean response model: $\delta_o = 1$, $\delta_1 = 0$, $\delta_2 = 0$, $z_t^* = 1$, and $\alpha = \bar{\alpha} f(t)$. Thus,

$$\beta_t = \beta_o + \bar{\alpha} f(t) + \xi_t \tag{7f}$$

where $f(t)$ is some function of time.

g. Goldfeld and Quandt (1973) switching regression model: $\delta_o = \delta_1 = \delta_2 = \xi_t = 0$, $\alpha = 1$, and $z_t = \beta_1$ for $t \varepsilon I_1$ and $z_t = \beta_2$ for $t \varepsilon I_2$, where I_1 and I_2 represent the sets of indices for which two separate regression equations hold. Hence,

$$\begin{aligned} y_t &= y_t \beta_1 + u_t \qquad t \varepsilon I_1 \\ y_t &= x_t \beta_2 + u_t \qquad t \varepsilon I_2. \end{aligned} \tag{7g}$$

h. Spline regression model [Poirier (1976), Buse and Lim (1977)]. In this formulation linear, quadratic, cubic, and other special forms of splines can be specified. This simplest case is the linear spline where the intercept it assumed constant; and for the slope, $\delta_o = 1$, $\delta = \delta = \xi_t = 0$, and α and z are defined as the vectors, $\alpha' = (\alpha_1, \alpha_2)'$ and $z_t = (t, t - \bar{t}_1)$. Hence, for the slope coefficient

$$\beta_t = \beta_o + \alpha_1 t + \alpha_2 (t - \bar{t}), \tag{7h}$$

where $t - \bar{t}$ is restricted to be zero for $t < \bar{t}$. The parameter β_t is referred to as a linear spline across time with a known knot at $\bar{t}$.

The above special cases (a) through (h) admit at entire spectrum of possible parameter evolutions. The first case (a) is the conventional constant

parameter specification. The second, the adaptive regression model of Cooley and Prescott, treats the parameter evolution only in terms of the intercept or constant term. This parameter evolution evolves in accordance with a random walk model and clearly does not allow for turning points in behavior of the time-varying parameters. The third formulation is especially important when influences from outside the model motivate structural changes in the parameter effects. Such influences are often qualitative in nature and cannot generally be measured with any acceptable degree of accuracy. The fourth formulation or the random coefficient model has been widely applied especially in the context of time series and cross-section data [Mundlak (1978b)]. The varying parameter formulation of Cooley and Prescott (e) has been applied to money demand relationships [Rausser and Laumas (1976)] and supply response elasticity for wheat [Cooley and DeCanio (1973)]. This formulation allows systematic variation in both the intercept and slope coefficients which appear in multiple regression models. The two-component process on the unknown parameter effects allow similar interpretations to Friedman's permanent income hypothesis. Both transitory and permanent variations in the time-varying parameters are allowed by this formulation.

The sixth special case advanced by Singh et al. (1976) provides a mild generalization to the Belsey (c) and the random coefficient model of Swamy (d). A new feature offered by this formulation is the inclusion of a linear function of time which leads to a presumed continuous evolution of the parameter effects. The switching regression model generalizes the conventional Chow (1960) formulation. The latter formulation presumes that *a priori* information is available to classify various regimes whereas the Goldfeld and Quandt approach endogenizes the distribution of the regimes. When *a priori* information

is not readily available, the Goldfeld and Quandt approach is preferable, particularly when the parameters move by discrete jumps. Finally, the spline formulation offers some distinct advantages in structuring the nature of parameter variation. In the case of linear spline, the formulation is equivalent to a piecewise linear approximation.

2.2. *Desired estimates*

In operating with specifications which allow time-varying parameters, issues associated with the type of estimates desired naturally arise. The resolutions of these issues depends upon the amount of information available. An estimate for a particular point in time might depend not only on past and current information but also on future information. For example, if $\beta_{t/t+j}$ is to denote the best estimate of β_t based on information up to and including the time period $t + j$, three situations can be distinguished. The first pertains to smooth estimates of the parameter effects ($j > 0$), the second to filtering estimates on the parameter effects ($j = 0$), and the third to prediction estimates of the parameter effects ($j < 0$). Econometricians are often concerned with the best estimate based on the entire data sample, i.e., $\beta_{t/T}$; but this requires filter and prediction estimates of the parameter effects, i.e., $\beta_{t/t}$ and $\beta_{t/t+j}$ with $j < 0$. Furthermore, from the standpoint of forecasting and policy evaluation, there is special concern with the prediction estimates of the parameter effects ($j < 0$). The generation of these estimates are crucial in the updating and revision of empirical model representations.

3. Estimation procedure

In this section operational estimation procedures are developed for equation (3). The treatment begins with the consequences of specification errors

which arise with the application of conventional estimation methods followed by the suggested estimation procedure. This procedure involves a two-stage approach and places emphasis on the estimation of the variance components, e_t and u_t. Finally, the possibility of negative variance estimates is examined and the basic estimation method is modified to preclude this potential outcome.

3.1. Consequence of specification error

In order to motivate the estimation of the equation described above, the properties of the estimates which ignore the time-varying coefficients will be investigated. Within the framework of the simple regression, the estimate of the constant slope is given by

$$b_{y/x} = \frac{\Sigma y_t x_t}{\Sigma x_t^2} = \frac{\Sigma x_t^2 \beta_t}{\Sigma x_t^2} + \frac{\Sigma x_t u_t}{\Sigma x_t^2} . \tag{8}$$

The expected value of this estimator is given by

$$E(b_{y/x}) = \Sigma w_t \beta_t \tag{9}$$

where $w_t = x_t^2 / \Sigma x_t^2$; hence, $0 < w_t < 1$, $\Sigma w_t = 1$. It turns out that the expected value of the regression coefficient is a weighted average of the individual coefficients. The weights are given by the squared x's. The expression (9) in terms of the reduced form equation (3) may be written:

$$\begin{aligned} E(b_{y/x}) &= E\left[\beta_o + \Sigma L_t w_t + \alpha \Sigma z_t w_t + \Sigma e_t w_t\right] \\ &= \beta_o + r_{LW} TS_w S_L + T \bar{L} \bar{w} + \alpha T r_{zw} S_z S_w + \alpha T \bar{z} \bar{w} \end{aligned} \tag{10}$$

where r denotes correlation coefficients; S, the standard deviation; T, the number of sample observations; and the upper bar, the average mean value with all moments computed from the sample.

For the expected value to be equal to a parameter independent of the sample data, it is necessary that both the correlation coefficients between the weights and the variables L and z and the product of the averages of the corresponding variables be identically zero. The requirement on the zero product of the average is indeed very restrictive; therefore, it is likely that the expected value of the simple regression coefficient will always depend on the sample data.

Another consequence of the present formulation is the possible introduction of serial correlation into an equation which is used for the estimation of the constant parameter. In particular, note that equation (1) may be written as:

$$y_t = x_t\beta_o + [(\beta_t - \beta_o)\, x_t + u_t]. \tag{11}$$

The term in brackets is the combined error term. This error term may be serially correlated even though the u_t's are unrelated. Specifically,

$$E[(\beta_t - \beta_o)\, x_t + u_t]\,[(\beta_{t-1} - \beta_o)\, x_{t-1} + u_{t-1}] = x_t x_{t-1}\, E(\beta_t - \beta_o)(\beta_{t-1} - \beta_o). \tag{12}$$

Combining equation (12) with equation (2), it becomes obvious that, if there is a serial correlation in either L_t or z_t, the β_t is serially correlated. Consequently, (8) is not an efficient estimator of β_o. On the other hand, the reduced form expression (3) for y_t eliminates the serial correlation by the explicit inclusion of $L(x_t)$ and z_t and consequently allows a more efficient estimate of β_o.

For the varying parameter model, it is of particular interest to estimate the individual β_t. Equation (8) gives an estimator for the constant coefficient β_o but not the individual β_t. Of course, (8) is a biased estimator of the individual β_t. The square of the bias is given by:

$$B_t^2 = (\Sigma w_t \beta_t - \beta_t)^2. \tag{13}$$

The variance of the estimator conditional on x_t and β_t is given by the expression:

$$\text{var } (b_{y/x}) = \sigma_u^2 / \Sigma x_t^2. \tag{14}$$

Combining the bias square and the variance, the expressions for the mean square error (MSE) is obtained:

$$\text{MSE} = \sigma_u^2 / \Sigma x_t^2 + (\Sigma w_t \beta_t - \beta_t)^2. \tag{15}$$

Note that the mean square error can be reduced in estimating the varying parameter equation by incorporating equation (2).

3.2. Estimation method

The error term in (3) has a heteroscedastic structure; therefore, it is desirable to use a two-stage GLS estimator. The first stage involves estimation of (3) by OLS followed by the estimation of an appropriate covariance matrix. The second stage employs the estimated covariance matrix to obtain the GLS estimator. The computation of the covariance matrix is not immediate; this issue will be investigated in Section 3.4.

The procedure outlined above provides estimates of β_o and of α. The values of the parameters obtained in such a way can now be introduced in the

structural equation (2) from which an estimate of β_t may be obtained. Such an estimate may be represented by

$$\hat{\beta}_t = \hat{\beta}_o + \hat{L}(x_t) + \hat{\alpha} z_t. \tag{16}$$

The estimates of β_o and α are unbiased even for the first stage of the two-stage procedure.

It is interesting to note that the most recent data play an important role in the estimation of β_t. That appears specifically in the two terms $L(x_t)$ and z_t. Early values for x_t get very little weight in $L(x_t)$ whereas early values for z_t do not appear in (16). Thus, the early observations affect β_t only through the effects of β_o, α, and the estimates of $\hat{L}$. Clearly, under the present structure, there is no need to throw away early observations since they still provide information on the constant terms β_o and α.

3.3. The estimated equation

The major conceptual problems have been reviewed in the context of the simple regression framework. The model involving more than one explanatory variable can now be easily formulated. Let $\underline{x}_t$ be a k vector of explanatory variables and write for the tth observation a generalized version of (1):

$$y_t = \underline{x}_t' \underline{\beta}_t + u_t \tag{17}$$

$$\underline{y} = X^d \underline{\beta} + \underline{u} \tag{18}$$

where $X^d = \text{diag}\{\underline{x}_t'\}$ is a TxTk matrix; $\underline{\beta}' = [\underline{\beta}_1', \ldots, \underline{\beta}_T']$; β_t is kxl vector; and $\underline{y}$ and $\underline{u}$ are T component vectors. For this specification, a generalized version of (2) is given by:

$$\underline{\beta}_t = \underline{\beta}_0 + [I_k \otimes \underline{L}_t'] \underline{\Pi} + [I_k \otimes \underline{z}_t'] \underline{\alpha} + \underline{e}_t \tag{19}$$

or equivalently,

$$\underline{\beta} = [\underline{1}_T \otimes \beta_0] + L\underline{\Pi} + Z\underline{\alpha} + \underline{e} \tag{20}$$

where $\underline{\beta}_0$ is the mean value of $\underline{\beta}_t$ for zero values of $\underline{L}_t$ and $\underline{z}_t$; $\underline{1}_t$ is a T component vector of ones; $\underline{\Pi}' = [\underline{\Pi}_1', \ldots, \underline{\Pi}_k']$; $\underline{\Pi}_j$ is a vector of coefficients π_{jr}; j, r = 1, ..., k, expressing the effect of L_{rt} on β_{tj}; $\underline{L}_t' = [L_{1t}, \ldots, L_{kt}]$; L_{jt} is a function of past value of x_j; $\underline{z}_t' = [z_{1t}, \ldots, z_{mt}]$; and $\underline{e}' = [\underline{e}_1', \ldots, \underline{e}_T']$.

$$Z = \begin{bmatrix} I_k \otimes \underline{z}_1' \\ \\ I_k \otimes \underline{z}_T' \end{bmatrix} \qquad L = \begin{pmatrix} I_k \otimes L_1' \\ \\ I_k \otimes L_T' \end{pmatrix}. \tag{21}$$

Combining (18) and (20),

$$\underline{y} = X^d\underline{\beta}_0 + X^d\, L\, \underline{\Pi} + X^d\, Z\, \underline{\alpha} + [X^d\, \underline{e} + \underline{u}] \tag{22}$$

where $X \equiv \begin{pmatrix} \underline{x}_1' \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_T' \end{pmatrix}$.

Letting $\underline{\varepsilon} = X^d\, \underline{e} + \underline{u}$, it follows that

$$\underline{\varepsilon} \sim \left[0,\ X^d\, E\,(\underline{e}\ \underline{e}')X^{d'} + \sigma^2 I_T\right] \tag{23}$$

and

$$E[\underline{\varepsilon}\ \underline{x}_t'] = 0,\ E[\underline{\varepsilon}\ \underline{z}_t'] = 0,\ E(e\ e') = I_T \otimes \Delta$$

where the elements of the diagonal matrix Δ represent the variances of random coefficients β_k.

Note that in (22) there is a generalized form of L which appeared in equation (3). Each β may depend not only on the L function constructed on the x conjugate to this β but also on all other x's. This is a general specification, but it is not suggested that every β will necessarily depend on all the lagged x's. In empirical applications, it is likely that any particular β will depend on the x conjugate to it and perhaps a few other x's. Nevertheless, the general form will be maintained here to allow flexibility in adapting the model to various circumstances.

Since no observations are available on $L(x_t)$, several possibilities can be considered in estimating (22). It is possible to impose a structure on $L(x_t)$ such as a first-order distributed lag function. The weights of such a function are not known, but it is possible to iterate by assuming several alternative values for the weights and examine the associated likelihood functions. If the likelihood function is not particularly flat, then the choice will be easy. If the likelihood function is flat, then the choice is immaterial to a large extent; and any of the weighting schemes may give equally good results. In the special case where L and Π are known, the following transformation can be made:

$$\underline{y} - X^{d}L\underline{\Pi} = X\beta_{o} + X^{d}\,Z\underline{\alpha} + \underline{\varepsilon}. \tag{24}$$

Alternatively, instead of enforcing a structure on $L(x_t)$, a sequence of lag values for x_t may be specified. This becomes equivalent to introducing several terms; each one is a product of x_t and a lagged value of x. The advantage of this approach is that it does not assume a structure. However, the cost is rather high and perhaps excessively high. Too many lagged values for x will introduce multicollinearity in the system; and in many instances,

there will be sufficient information to sustain only one or two terms. This problem is particularly troublesome in the case of forecasting. The problem arises not only because of the $L(x_t)$ function but it is intrinsic to the basic framework. It is due to the dependence of the systematic components of the coefficients on the x and z. This issue will be discussed further in Section 5.

Equation (22) provides the form to be estimated. The error term of this equation has a heteroscedastic structure; therefore, a two-stage procedure must be developed. For that purpose, the covariance matrix of the error structure must be estimated. This issue is dealt with in the next subsection.

3.4. *Estimation of variance components*

The estimation of the variance components in varying coefficients model is discussed in Mundlak (1978a). The method developed by Mundlak (1978a) is a generalization of the standard method used in components of error ANOVA. Specifically, the method requires repeated observations on the sampling unit. As such, this method cannot be applied to the framework examined here since only one observation per year is available. Consequently, another method must be developed.

The method developed here utilizes the fact that the error term of (22) is a linear combination of the various e's and u, where the coefficients of the combinations are the x's themselves. Since the various error terms have different <u>known</u> "coefficients," it is possible to estimate the variances and covariances in question. Letting $X^* \equiv (X, X^d L, X^d Z)$ and $M(X^*) = I - X^* (X^{*\prime} X^*)^{-1} X^{*\prime}$, it follows that the vector of the computed residuals of (22) may be expressed as

$$\hat{\underline{\varepsilon}} = M(X^*)\ \underline{\varepsilon}. \tag{25}$$

Hence,

$$E[\hat{\underline{\varepsilon}}\ \hat{\underline{\varepsilon}}'] = \sigma^2 M(X^*) + M(X^*)X^d\ [I_T \otimes \Delta]X^{d'}\ M(X^*). \tag{26}$$

Under the assumption that X^* is of full rank, say, k^*, rank $\Pi = T - k^*$. Proceed by evaluating

$$E\left[\hat{\varepsilon}_t^2\right] = \sigma^2\ m_{tt} + \underline{m}_t'\left[X^d\ (I_t \otimes \Delta)X^{d'}\right]\underline{m}_t \tag{27}$$

where $\underline{m}_t$ is the tth column of $M(X^*)$. The second term on the right-hand side may be simplified by

$$X^d\ [I_T \otimes \Delta]X^{d'} = \text{diag}\ \{\underline{x}_t'\ \Delta\ \underline{x}_t\} \tag{28}$$

and

$$\underline{m}_t'\left[X^d(I_T \otimes \Delta)X^{d'}\right]\underline{m}_t = \text{tr}\{\text{diag}(\underline{x}_t'\ \Delta\ x_t)\underline{m}_t\underline{m}_t'\}$$

$$= \sum_{\tau=1}^{T} m_{t\tau}^2\ (\underline{x}_\tau'\ \Delta\ \underline{x}_\tau^*)$$

$$= \Delta_{11} \sum_{\tau=1}^{T} m_{t\tau}^2\ x_{t\tau}^2 + \ldots + \Delta_{kk}\ \Sigma m_{t\tau}^2\ x_{\tau k}^2 +$$

$$+ 2\Delta_{12}\ \Sigma\ m_{t\tau}^2\ x_{\tau 1}\ x_{\tau 2} + \ldots + 2\Delta_{kk-1}\ \Sigma m_{t\tau}^2\ x_{\tau k}\ x_{\tau k-1}.$$

Defining $w_{js',t} \equiv \sum_{\tau=1}^{T} m_{t\tau}^2\ x_{\tau j}\ x_{\tau s}$, $j, s = 1, \ldots, k$, expression (27) can be written as:

$$E\left[\hat{\varepsilon}_t^2\right] = \sigma^2 m_{tt} + \Delta_{11} w_{11,t} + \ldots + \Delta_{kk} w_{kk,t} + 2\Delta_{12} w_{12,t} + \ldots + 2\Delta_{kk-1} w_{kk-1,t}; \tag{29}$$

and the coefficients σ^2, Δ_{js} can be estimated from the regression of $\hat{\varepsilon}_t^2$ on m_{tt} and the w's, directly observable variables.

A necessary condition for the existence of such estimates is given by $T - k > \frac{k(k+1)}{2} + 1$ or $T > \frac{(k+1)(k+2)}{2}$. However, with large k, multicollinearity will preclude reliable estimates unless T is very large. Therefore, in time series studies where the number of observations is far from excessive, it would be desirable to keep the number of coefficients which are allowed to vary reasonably small so that their variance could be estimated from the data. This issue will be formally discussed in Section 5.

Note that only T of the elements of $(\hat{\underline{\varepsilon}}\,\hat{\underline{\varepsilon}}')$ are used in deriving the estimation. At first glance it might appear that some information is lost by ignoring the off-diagonal elements. However, this is not the case because the rank of $\hat{\underline{\varepsilon}}\,\hat{\underline{\varepsilon}}'$ is 1; and, therefore, knowledge of one column of this matrix facilitates the computation of the remaining columns. A given column is not used but rather the diagonal elements which, of course, can be obtained by elementary row and column operations on $\hat{\underline{\varepsilon}}\,\hat{\underline{\varepsilon}}'$.

For the special case of $\Delta \equiv 0$, note also that (26) becomes the standard formula for deriving an unbiased estimate of σ^2. As is well known, only $T - k^*$ elements of $\hat{\varepsilon}$ are independent in the sense that, knowing these elements and X^*, the remaining $\hat{\varepsilon}$'s can be derived from $X^{*\prime}\,\hat{\underline{\varepsilon}} = \underline{0}$. The "loss" of k^* independent observations (or degrees of freedom) is, of course, due to the estimation of the regression coefficients. In some cases, which are rather important in

statistical analysis, this is not a loss but rather a source of important information, specifically, when the estimated regression coefficients can be thought to be repeated drawings of a random variable with a given probability distribution. Indeed, this is the situation presumed by the standard components of error analysis of variance model. In its simplest version, the model can be written as:

$$y_{ij} = \alpha_i + u_{ij} \sim (0, \sigma^2 + \Delta) \tag{30}$$

where $\Delta = \text{var } \alpha$.

It is well known that Δ is estimated from

$$\hat{\Delta} = \frac{1}{I-1} \sum_{i=1}^{I} \hat{\alpha}_i^2 - \frac{\hat{\sigma}^2}{J} \tag{31}$$

where $\hat{\alpha}_i = y_{i\cdot} - y_{\cdot\cdot}$, $y_{i\cdot} = \frac{1}{J} \sum_{j}^{J} y_{ij}$, and $\hat{\sigma}^2 = \frac{1}{I(J-1)} \sum_{i} \sum_{j} (y_{ij} - y_{i\cdot})^2$.

The point made by this model is that $\hat{\alpha}_i$ are regression coefficients (of I dummy explanatory variables). Simultaneously, they are repeated observations on the random variable $\hat{\alpha}_i \sim 0, \frac{\sigma^2}{J} + \Delta$. Thus, the k degrees of freedom "lost" in estimating $\hat{\alpha}_i$ provide $I - 1$ independent observations for the estimation of Δ. This structure does not extend itself to the problem where only one observation on the vector of regression coefficients is available. Fortunately, the error term (25) has a structure that permits the decomposition of the variance to its components as demonstrated by (26)-(29).

In the ANOVA error components model, the estimator given above (obtained by subtraction) may actually be negative. The possibility of a negative estimate of some variance components is not precluded from the case as well

although it appears that this possibility is less likely. However, if a regression coefficient which serves as a variance estimate is negative and significantly different from zero, it may be desirable to repeat the estimation under constraint that none of the variances will be negative.

3.5. Nonnegative variance constraints

The possibility of negative variance estimates for (29) can be dealt with in a number of fashions. The most obvious is to apply the inequality estimator [Judge and Takayama (1966)] to (29). This simply involves introducing the restrictions

$$\sigma^2,\ \Delta_{kk} \geq 0, \qquad \text{for all } k. \tag{32}$$

The sampling properties of the resulting estimator recently have been derived by Judge and Yancey (1978) under a squared error loss measure. Operating with a general inequality estimator, they are able to demonstrate that the variance of this estimator is equal to or less than the variance of the maximum likelihood estimator. As in our case, they show that, if the direction of the inequality constraint information is known, the inequality restricted estimator is uniformly superior over the range of the parameter space to the conventional maximum likelihood estimator under a squared error loss measure. The sampling properties are derived by Judge and Yancey for a pretest estimator, and the relevant test statistic is distributed as a central t random variable with conventional degrees of freedom.

An alternative approach to the problem of negative variance estimates is possible using shrunken estimators.[3] For this approach, the negative variance obtains in the unrestricted maximum likelihood estimation is addressed by the incorporation of *a priori* information which forces the estimators toward zero. This information is less consistent with the true

a priori conditions on variances than the inequality restrictions. On the other hand, it is much easier to implement from a computational viewpoint. For example, the ridge regression estimator for which computational routines are commonly available is a stochastically shrunken estimator where the parameters are forced toward zero with a probability directly related to the value used to augment the diagonal of the sum of squares for the design matrix [Fomby and Johnson (197)]. The *a priori* information introduced may be interpreted in terms of zero variances with a given probability. Finally, it should be noted that ridge-type and other shrunken estimators dominate maximum likelihood estimators over certain regions in the parameter space.

4. Forecasting

The introduction of the varying parameter model is of particular importance for forecasting. Since the parameters change with time, it is important to capture the coefficients which pertain to the period of forecast. Thus, there is a joint prediction problem, i.e., β_t and y_t are predicted simultaneously. To place the problem of prediction within a useful framework, write the predicted value of y as:

$$\hat{y}^*_{T+1} = x^*_{T+1}\, b_{T+1} \tag{33}$$

where b_{T+1} in this case is an arbitrary estimator of β_{T+1}, and x^*_{T+1} is a known value. Such a formulation is directed at the prediction of y. The error of such a forecast is given by:

$$\begin{aligned} \hat{y}^*_{T+1} - y_{T+1} &= x^*_{T+1}\, b_{T+1} - x^*_{T+1}\, \beta_{T+1} - u_{T+1} \\ &= x^*_{T+1}\, (b_{T+1} - \beta_{T+1}) - u_{T+1} \end{aligned} \tag{34}$$

The expected value of the error is given by

$$E\ (\hat{y}^*_{T+1} - y_{T+1}) = E\ \{x^*_{T+1}\ (b_{T+1} - \beta_{T+1})\}\Big|_{T+1} = x^*_{T+1}\ B_{T+1} \tag{35}$$

where B_{T+1} represents the bias of predicting β_{T+1} by b_{T+1}. Of course, when b_{T+1} is an unbiased predictor, B vanishes. The MSE of the predictor y^*_{T+1} is given by:

$$\text{MSE}\ \hat{y}^*_{T+1} = x^{*2}_{T+1}\ \ B^2_{T+1} + \text{var}\ b_{T+1}\ \ + \sigma^2_u. \tag{36}$$

Within the Gauss-Markov framework, the bias is zero; and the variance is minimized by using a generalized least-square estimator. Here, this approach is infeasible since the variance of the error term is not known.

The properties of the predictor based on equation (8) provide some interesting insights. The determinants of bias which appears in (13) can best be evaluated by substituting equation (2) for β_t to obtain

$$\begin{aligned} B_{t+1} &= \Sigma w_t\ [\beta_o + L_t + \alpha z_t + e_t] - [\beta_o + L_{T+1} + \alpha z_{T+1} + e_{T+1}] \\ &= [\Sigma w_t L_t - L_{T+1}] + \alpha\ [\Sigma w_t z_t - z_{T+1}] + [\Sigma w_t e_t - e_{T+1}]. \end{aligned} \tag{37}$$

This expression underscores the scope and importance of taking into account the explicit formulation of time-varying coefficient. If the values of $L(x_t)$ and z_t in the period for which a forecast is obtained vary or differ considerably from the past values of these variables, the bracketed terms will contribute substantially to the bias. On the other hand, if values of $L(x_t)$ and z_t are well within the range of past observations, then these terms may be negligible; and the predictor based on ignoring the varying parameters may give

an unbiased prediction. Hence, it is particularly important to use varying parameter approach when recent values of the explanatory variables deviate from the average values for the sample. From this analysis, it appears that the present framework can detect turning points in cases which a constant parameter fails to do so.

The bias in the estimate of β_{T+1} can be avoided by utilizing equation (3) from which the following can be constructed:

$$\hat{\beta}_{T+1} = \hat{\beta}_o + \hat{L}_{T+1} + \hat{\alpha} z_{T+1}. \tag{38}$$

Note that x_T is the latest value of x that enters the function L_{T+1}, and this value is known at the time the forecast is constructed. Moreover, $\hat{\beta}_o$ and $\hat{\alpha}$ are unbiased estimates of the corresponding coefficients. Consequently, the expected value of b_{T+1} conditional on z_{T+1} is equal to β_{T+1}.

The problem with the above approach is that the resulting regression equation can be blessed with too many variables. Increasing the number of variables has a cost; it increases the sampling variance and thereby the forecast variance. There is, therefore, a trade-off between an increase in the sampling variance and a decline in the bias which are associated with the addition of variables to the regression. That creates a problem of a choice which is particularly important in the framework. This problem is spelled out explicitly in the next section.

4.1. Selection of explanatory variables for forecasting

The problem posed at the end of the previous section is of a general nature and can be treated as such. It is assumed that the true equation contains k explanatory variables. This set of variables is suggested by the

theory (or perhaps by a theory) underlying the equation. Using the sample data, the equation is estimated and the null (or other) hypotheses are tested. Having this information, it is now desirable to predict y for a set of values given by the row vector of values $\underline{x}^{*\prime}$ for the regressors: The Best Linear Unbiased Predictor (BLUP) is:

$$\hat{y}^* = \underline{x}^{*\prime}\ \underline{b}; \tag{39}$$

and the variance of the prediction error, conditional on X, $f = \hat{y}^* - y$, is

$$\sigma_f^2 = \sigma^2\ (1 + h) \tag{40}$$

where

$$h = \underline{x}^{*\prime}\ (X'X)^{-1}\ \underline{x}^*. \tag{41}$$

The question is what X's should be included in the predicting equation. As usual, the answer depends on the criterion. The problem is analyzed here within the framework of smallest conditional MSE.

To deal with this problem, X is partitioned into $X = (X_1, X_2)$ of order $(n \times k_1, n \times k_2)$, $k_1 + k_2 = k$; and $\underline{x}^{*\prime}$ is partitioned accordingly, $\underline{x}^{*\prime} = (\underline{x}_1^{*\prime}, \underline{x}_2^{*\prime})$. The question is whether to use the predictor in (39) or, alternatively, to use a predictor which is based only on X_1:

$$\tilde{y}^* = \underline{x}_1^{*\prime}\ \underline{a} \tag{42}$$

where $\underline{a}$ is the L. S. coefficient of $\underline{y}$ on X_1. The error of this predictor is $f_1 = \tilde{y}^* - y$ and its variance conditional on X_1,

$$\sigma_{f1}^2 = \sigma_1^2\ (1 + h_1) \tag{43}$$

$$h_1 = \underline{x}_1^{*\prime} (X_1'X_1)^{-1} \underline{x}^*. \tag{44}$$

The evaluation of σ_1^2 and the relationships between the two predictors require some further specification of the model. Basic in this specification is the relationships between X_2 and X_1 which are written as follows:

$$E(X_2|X_1) = X_1\Pi$$

and

$$X_2 = X_1\Pi + U_2. \tag{45}$$

The term $E(X_2|X_1)$ may reflect the design of the experiment if X is nonstochastic or the population relationships if the X's are random variables. Write for the unrestricted equation

$$\begin{aligned} \underline{y} &= X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{u} \\ \underline{u} &\sim (\underline{0}, \sigma^2 I_n). \end{aligned} \tag{46}$$

Combining (45) and (46), the restricted equation is written:

$$\underline{y} = \underline{X}_1\underline{\alpha} + \underline{u}_1 \tag{47}$$

where

$$\underline{\alpha} = \underline{\beta}_1 + \Pi\underline{\beta}_2, \quad \underline{u}_1 = \underline{u} + U_2\underline{\beta}_2.$$

Let u_{2i}' be the ith row of U_2 and assume

$$\underline{u}_{2i}|_{X_1} \sim (\underline{0}, \Sigma_2). \tag{48}$$

Hence,

$$\underline{u}_1|_{X_1} \sim \left(\underline{0}, \sigma_1^2 I_n\right), \tag{49}$$

$$\sigma_1^2 = \sigma^2 + \underline{\beta}_2'\Sigma_2\underline{\beta}_2. \tag{50}$$

Note that σ_1^2 is the conditional variance of y given X_1. At the same time it can be viewed as the MSE of $\underline{y}|_{X_1}$ given X_2.

The difference between σ_1^2 and σ^2 is the square of the bias generated by evaluating $E(y|X)$ from $E(y|X_1)$. The difference can be estimated by using $\hat{\sigma}^2 = 1/(n - k)\ \underline{y}'M\underline{y}$ and $\hat{\sigma}_1^2 = 1/(n - k_1)\ \underline{y}'M_1\underline{y}$ where $M = I - K$, $M_1 = I - K_1$, $K = X(X'X)^{-1}X'$ and $K_1 = X_1(X_1'X_1)^{-1}X_1'$.

Let F be the statistics used for testing the null hypothesis of $\underline{\beta}_2 = 0$. By a simple transformation, it can be shown that $\hat{\sigma}_1/\hat{\sigma}^2 = 1 + (F - 1)k_2/n - k_1$. Consequently, $\hat{\sigma}_1^2 \gtreqless \hat{\sigma}^2 \Leftrightarrow F \gtreqless 1$. Of course, the verdict with respect to the null hypothesis depends on the critical value of F (which is distributed with k_2 and $n - k_1$ degrees of freedom).

Deciding that $\sigma_1^2 - \sigma^2$ is significantly different from zero still does not justify the use of the unrestricted model for purpose of prediction. For that, σ_f^2 and σ_{f1}^2 needs to be evaluated which, in turn, requires an evaluation of h and h_1 as given by (41) and (44) above. Using the expression for an inverse of a partitioned matrix

$$(X'X)^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} P \\ -I \end{bmatrix} (V_2'V_2)^{-1} \begin{bmatrix} (P - I) \end{bmatrix} \tag{51}$$

where $P = (X_1'X_1)^{-1}X_1'X_2$ is the OLSE of Π, and V_2 is the residual of such regressions, $V_2 = M_1X_2$, with rank $V_2 = k_2 \leq n - k_1$.

Combining (41), (44), and (51):

$$h = h_1 + \underline{v}_2'(V_2'V_2)^{-1}\ \underline{v}_2 \tag{52}$$

where

$$\underline{v}_2 = \underline{x}_2^* - P'\underline{x}_1^* \tag{53}$$

is a vector of errors of forecasting $\underline{x}_2^*$ by $P'\underline{x}_1^*$.

There are several possible interpretations of (52). These interpretations are most easily developed by denoting $\underline{b}_1$ and $\underline{b}_2$ as the unrestricted OLSE of $\underline{\beta}_1$ and $\underline{\beta}_2$ in (46), respectively, and the variances of the two estimators by Σ_{b1} and Σ_{b2}. Then,

$$\sigma^2(V_2'V_2)^{-1} = \Sigma_{b2} \tag{54}$$

and, hence,

$$h - h_1 = \frac{1}{\sigma^2}\,\underline{v}_2'\,\Sigma_{b2}\,\underline{v}_2. \tag{55}$$

Clearly, the sampling component of the unrestricted forecast variance increases with the variance of $\underline{b}_2$. The less accurate the estimator of $\underline{\beta}_2$ is, the larger the sampling variance of this forecast. Perhaps it might be useful to recall the conditions leading to a large sampling variance with reference to the case on hand: small sample variance in X_2, high multicollinearity between X_2 and X_1, and a small sample size.[4]

The variance of the unrestricted forecast also increases with $|\underline{v}_2|$. This result is a generalization of a restatement of a known property--that the forecast variance increases with the square of the deviations of x*'s from their sample means. In the case under investigation, the marginal contribution of X_2 is evaluated. Consequently, the marginal contribution of x_2^* to the forecast variance is due only to that part of $\underline{x}_2^*$ that is not linearly accounted for by $\underline{x}_1^*$. Consequently, when $\underline{x}_2^*$ only deviates slightly from $p'\underline{x}_1^*$, this contribution is small.

When $\underline{x}_2$ is a random vector, the variance of $\underline{v}_2$ can be evaluated and expressed in terms of such variance. Write:

$$\underline{v}_2 = (\Pi - P)' \underline{x}_1^* + \underline{u}_2 \tag{56}$$

and

$$P - \Pi = (X_1'X_1)^{-1} X_1'U_2.$$

Hence,

$$\begin{aligned} \Sigma_{v2} &= E(\underline{u}_2\underline{u}_2') + E\left\{U_2' \left[X_1(X_1'X_1)^{-1} \underline{x}_1^* \underline{x}_1^{*\prime} (X_1'X_1)^{-1} X_1'\right] U_2\right\} \\ &= \Sigma_2 + E\{U_2' [\cdot] U_2\} \\ &= \Sigma_2 + E\{\cdot\cdot\} \\ &= \Sigma_2 (1 + h_1).^5 \end{aligned} \tag{57}$$

An unbiased estimator of Σ_2 is obtained from:

$$\hat{\Sigma}_2 = \frac{1}{n - k_1} V_2'V_2. \tag{58}$$

Utilizing (58), (52) can be rewritten as:

$$h - h_1 = \underline{v}_2' \hat{\Sigma}_2^{-1} \frac{\underline{v}_2}{n - k_1} \tag{59}$$

and utilizing (57),

$$h - h_1 = \underline{v}_2' \hat{\Sigma}_{v2}^{-1} \underline{v}_2 \frac{(1 + h_1)}{n - k_1}. \tag{60}$$

It is now possible to make the following comparisons:

$$\frac{\sigma^2_{f1}}{\sigma^2_f} \gtreqless 1 \iff \frac{\sigma^2_1}{\sigma^2} \gtreqless \frac{1+h}{1+h_1} = 1 + \frac{h - h_1}{1 + h_1}$$

$$\iff \frac{\underline{\beta}_2' \, \Sigma_2 \, \underline{\beta}_2}{\sigma^2} \gtreqless \frac{\underline{v}_2' \, \hat{\Sigma}^{-1}_{v2} \, \underline{v}_2}{n - k_1} \tag{61}$$

$$\iff \underline{\beta}_2' \, \Sigma_2 \, \underline{\beta}_2 \gtreqless \underline{v}_2' \, \Sigma_{b2} \, \underline{v}_2.$$

A large variance Σ_2 increases the LHS of the inequalities and tends to decrease the RHS and, hence, leads to preferring the unrestricted forecast (the actual values of the terms, of course, also depend on $\underline{\beta}_2$ and $\underline{v}_2$).

The RHS of the inequalities are random variables. A stronger statement with respect to their distribution can be made if an assumption of normal distribution is added:

$$\underline{u}_{2i} \big| X1 \sim N_{k_2} (\underline{0}, \Sigma_2)$$

$$E\left(\underline{u}_{2i}, \underline{u}'_{2i*} \big| X1\right) = 0, \qquad \text{for all } i \neq i^*. \tag{62}$$

It then immediately follows that[6]

$$\underline{v}_2' \, \Sigma^{-1}_{v2} \, \underline{v}_2 \, \frac{n}{n - k_1} \sim T^2 \tag{63}$$

where T^2 is Hotelling's generalized T^2 statistics. Utilizing the relationships between T^2 and F, (61) can be rewritten as a probability statement:[7]

$$P\left(\left.\frac{\sigma_{f1}^2}{\sigma_f^2}\right|_{X_1} > 1\right) = P\left[F\,(k_2,\ n - k_1 + 1) < \left(\frac{n - k_1 + 1}{k_2}\right) \underline{\beta}_2' \Sigma_2 \underline{\beta}_2 \equiv \tau\right] \tag{64}$$

$$= \int_0^{\tau} F\,(k_2,\ n - k + 1)\ dF \equiv P(\tau).$$

As $n \to \infty$, $P(\tau) \to 1$ so that the unrestricted model dominates $\left(\sigma_f^2 < \sigma_{f1}^2\right)$ with probability 1. For finite n (and not "excessively large"), it is possible to benefit from restricting the model by eliminating variables which contain no net information. The net information is directly related to the size of the variance of these variables or more accurately to the quadratic form in the covariance matrix. Thus, it is desirable to eliminate variables which vary a little and have low values for their coefficients.[8]

The condition in (64) depends implicitly on the number of variables which are already included in the restricted model (k_1). Other things being equal, the larger k_1 is, the smaller the variance Σ_2, that is, the variance of the unaccounted part of X_2. It seems that this property is overlooked in frameworks which are based on forecasting from as large a number of variables as possible.

It may seem strange that the relative size of the forecast variance is treated as a random variable. The reason is that the statement is conditional on X_1, in which case X_2 has a probability distribution and, consequently, h has a probability distribution. For any set of values $\underline{x}_1^*$ used for a forecast, there are many various values for $\underline{x}_2^*$ and, consequently, of $\underline{v}_2$. In some cases, $\underline{v}_2$ may contain large (absolute) values for some components. In this case, it is better to forecast from a restricted model. Such events are represented by large values for F which exceed any preassigned values for τ.

The foregoing discussion is pertinent for designing an operational approach to forecasting. In practice, a narrower question can be asked: Which forecast is more accurate, given X and $\underline{x}^*$, that is conditional on X_1, $\underline{x}_1^*$ and X_2, $\underline{x}_2^*$? In this case the distribution of $\underline{v}_2$ is now irrelevant as $\underline{v}_2$ is given. The answer again is provided by (61). It is very straightforward; the true forecast variances are given and their ranking is thus determined but unknown. It is possible to estimate $\hat{\sigma}^2$ and $\hat{\sigma}_1^2$ and accordingly to state:

$$\frac{\hat{\sigma}_{f1}^2}{\hat{\sigma}_f^2} \gtreqless 1 \iff \frac{\hat{\sigma}_1^2}{\hat{\sigma}^2} \gtreqless 1 + \underline{v}_2' \hat{\Sigma}_{v_2}^{-1} \frac{\underline{v}_2}{n - k_1}. \tag{65}$$

Recalling the discussion above following equation (50):

$$\frac{\hat{\sigma}_{f1}^2}{\hat{\sigma}_f^2} \gtreqless 1 \iff F \gtreqless \frac{1}{k_2} \underline{v}_2' \hat{\Sigma}_{v_2}^{-1} \underline{v}_2 + 1 \gtreqless 1 \tag{66}$$

where F is the statistic for testing the null hypothesis $\underline{\beta}_2 = 0$. Under the null hypothesis, $\underline{\beta}_2$ has a distribution of central F with k_2 and $n - k$ degrees of freedom. Consequently:

1. When $F < 1$, the restricted forecast is preferable.
2. Let F_α be the critical value for testing the null hypothesis; then the inequality $F \gtreqless F_\alpha$ is not directly related to that in (66) in the sense that the verdict on the null hypothesis is neither necessary nor sufficient for determining which forecast is more accurate. F can be larger than the critical value but smaller than the RHS of (66); consequently, the restricted forecast should be preferred. The converse is also possible.

The discussion can be concluded as follows: one has to differentiate between a choice of a forecasting method and an actual forecast. In the first case the relative accuracy of the restricted and unrestricted forecasts conditional on the restricted set of variables are evaluated. Under the assumption of multivariate normal distribution of the vector of the variables affected by the restriction ($\underset{\sim}{x}_2$), (64) summarizes the pertinent information.

When a particular forecast is evaluated, it can be done so conditional on all the variables, as they are all known.[9] In this case, the various forecast variances can be estimated. It is interesting to note that the test of the null hypothesis with respect to the coefficients of the restricted variables ($\underset{\sim}{\beta}_2 = 0$) is neither sufficient nor necessary for determining the relative precision of the forecasts in question.

The conditions leading to more precision of a restricted forecast are of a similar nature in both cases: a small sample, little net information embedded in the restricted variables used in the regression, and considerable information in the values of these variables used for forecasting.

4.2. *Utilization of principal components*

In view of the developments in the previous section, it is clear that, when the sample is of finite size, there is a limit to the number of variables that should be included in a regression equation forecasting purposes. This raises a common problem in econometric analysis, that of a choice of variables. *A priori*, the model may suggest more variables than the data can support. By eliminating some variables, the model is reduced to a manageable size. Such a reduction need not be done arbitrarily. The problem of too many variables is that of the sampling variance and, therefore, the choice of variables to be eliminated can be done in such a way as to minimize

the sampling variance for any predetermined number of explanatory variables. Thus, the framework of the previous section could be applied by adding variables as long as the MSE declines. Such an approach leads to a consideration of the principal components framework which is now briefly outlined.

An orthogonal matrix P ($P'P = I_k$) is selected such that

$$P'(X'X)P = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{pmatrix} \equiv D \tag{67}$$

where $\lambda_1 > \cdots > \lambda_k$ are the characteristic roots of $(X'X)$. Note that

$$I = PP' = \underline{p}_1\underline{p}_1' + \underline{p}_k\underline{p}_k' \tag{68}$$

$$X'X = PDP' = \lambda_1\underline{p}_1\underline{p}_1' + \ldots + \lambda\underline{p}_k\underline{p}_k'. \tag{69}$$

In this setting, as is well known, the basic equation to be estimated can be rewritten as:

$$\begin{aligned} \underline{y} &= X\underline{\beta} + \underline{u} \\ &= (XP)(P'\underline{\beta}) + u = S\underline{\delta} + \underline{u}. \end{aligned} \tag{70}$$

Then the OLSE of $\underline{\delta}$ is:

$$\underline{\hat{\delta}} = (P'X'XP)^{-1} P'X'\underline{y} = D^{-1} S'\underline{y} = \left\{ \frac{\underline{s}_j'\underline{y}}{\lambda_j},\ j=1, \ldots, k \right\} \tag{71}$$

and

$$\text{var } \underline{\hat{\delta}} = \sigma^2 D^{-1}. \tag{72}$$

From (67) and (72), it follows that the estimators of the various coefficients are uncorrelated.

The relationships to OLSE of $\underline{b}$ is given by:

$$\underline{\hat{\delta}} = P'\underline{b} \text{ and } \underline{b} = P\underline{\hat{\delta}}. \tag{73}$$

Combining (72) and (73) and recalling that the elements of D are ranked, it is evident that the first row of P', say, p_1', gives the a linear combination of b, which has the lowest variance in the class of all normalized linear combinations of $\underline{b}$. This result is due to Greenberg (1975). This interpretation is extended by Fomby, Hill, and Johnson (1978) by showing that, if $\underline{\beta}$ in (70) is to be estimated subject to some (say, $k_2 < k$) homogenous linear restrictions ($R\underline{\beta} = \underline{0}$), the trace of the covariance matrix of the estimator is bounded from below by the sum of the variances of the first $k_1 = k - k_2$ regression coefficients in $\underline{\hat{\delta}}$ $\left(\sigma^2 \sum_{i=1}^{k} \lambda^{-1}\right)$.

In applying this discussion to the problem of forecasting, divide S into submatrices,

$$S = (S_1, S_2) = X\,(P_1, P_2).$$

Consider the row vector $\underline{s}^* = (\underline{s}_1^*, \underline{s}_2^*) = \underline{x}^*\,(P_1, P_2)$ and rewrite (70):

$$y = \underline{s}_1^*\underline{\delta}_1 + \underline{s}_2^*\underline{\delta}_2 + u. \tag{74}$$

The unrestricted forecast of y is:

$$\hat{y}^* = \underline{s}^*\underline{\hat{\delta}}. \tag{75}$$

But noting that:

$$\underline{s}^*\underline{\hat{\delta}} = (x^*P)(P'\underline{b}) = \underline{x}^*\underline{b},$$

which is equivalent to (39), the unrestricted forecast obtained without the use of principal components. Thus, the interest in using principal components arises when some variables are eliminated.

Consider the forecast based on the first k_1 principal components:

$$\hat{y}_{p1} = \underline{s}_1^* \hat{\underline{\delta}}_1 \tag{76}$$

and the forecast error

$$e_{p1} = y - \hat{y}_{p1} = \underline{s}^* \underline{\delta} + u - \underline{s}_1^* \hat{\underline{\delta}}_1 = \underline{s}_1^* (\underline{\delta}_1 - \hat{\underline{\delta}}_1) + (\underline{s}_2^* \underline{\delta}_2 + u).$$

The variance of e_{p1} conditional on S_1 is:

$$\sigma_{fp}^2 = \sigma_{p1}^2 (1 + h_{p1}) \tag{77}$$

where

$$h_{p1} = \underline{s}_1^* (S_1'S_1)^{-1} \underline{s}_1^{*1} = \underline{s}_1^* D_1^{-1} \underline{s}_1^{*'} \tag{78}$$

By (67),

$$S'S = \begin{Bmatrix} S_1'S_1 & 0 \\ 0 & S_2'S_2 \end{Bmatrix} = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$

and

$$\sigma_{p'}^2 = \sigma^2 + \underline{\delta}_2' E (\underline{s}_2^{*'} \underline{s}_2^*) \underline{\delta}_2 = \sigma^2 + \underline{\delta}_2' \Delta_2 \underline{\delta}_2$$

$$\Delta^2 = E (\underline{s}_2^{*'} \underline{s}_2^*) = E \{P_2' \underline{x}^{*'} \underline{x}^* P_2\} | X. \tag{79}$$

If $\underline{x}^*$ is stochastic and randomly drawn, then $\underline{s}_2^*$ is also stochastic. Furthermore, if it is drawn from the same distribution as the rows of X,

then $S_2'S_2/n - 1 = D_2/n - 1$ is the estimator of Δ_2 which is the variance of $\underline{s}_2$. When $\underline{x}$ is nonstochastic, then Δ_2 is a parameter determined by the design of the experiment.

The restricted forecast has to be compared to the unrestricted forecast. Following (61),

$$\frac{\sigma_{fp}^2}{\sigma_f^2} \gtreqless 1 \iff \frac{\sigma_{pl}^2}{\sigma^2} \gtreqless \frac{1 + h}{1 + h_{pl}} = 1 + \frac{h - h_{pl}}{1 + h_{pl}} . \tag{80}$$

On the surface it is not quite clear that a restricted forecast with k_1 principal components is preferable to such a forecast with k_1 original variables. The procedure of selecting principal components according to the size of the characteristic roots does not provide such an optimality criterion. The optimal results call for entering the principal components in a decreasing order of their t ratios (in absolute values). Since the regression coefficients on the principal components are uncorrelated, the addition or omission of variables does not affect the values of the t ratios of those variables which are retained in the regression. This procedure maximizes the R^2 in the class of regression equations with k_1 explanatory variables; and as a consequence, it minimizes the expected value of the noncentrality parameter, namely, $\sigma_{pl}^2 - \sigma^2$ and, consequently, the ratio of σ_{pl}^2 / σ^2. At the same time it also minimizes in a probability sense, the value of h_{pl} in the same class. Consequently, for any forecast based on k_1 explanatory variables, the aforementioned procedure has the lowest forecast MSE. Once the forecast is optimized for any given k_1, the selection of the number of variables to be included in the regression follows the rules developed in Section 4.1.

4.3. Implications for parameter variation model

The detailed evaluation of the choice of variables for purposes of forecasting is of special interest to the model advanced in Section 3. The essence of the model calls for increasing the number of the variables and it turns that such a procedure is not priceless. It seems that a reasonable procedure would be to formulate the model according to *a priori* considerations. This stage of formulation should bring out the more important relationships that are expected to exist between the variables. The next step is that of the empirical analysis. It turns out that, for the purpose of forecasting, the model will have to be reduced in terms of the number of variables. It is suggested that the reduction will be made by shifting to principal component regressions where the variables are added to the regression in a declining order of the absolute value of their t ratios.

A possible objection to the use of principal component is that the variables do not appear in a natural form. For instance, one would want to have in a demand function the income and own price variables or, alternatively, in terms of the analysis, one may want to have the original variables, that is, the set of variables which appear in the model prior to the modification due to the variable coefficient framework. This preference can be accommodated in the analysis. However, first it is to be noted that it is possible to obtain an estimate of β from the regression on the first k_1 principal components. In terms of the coefficients of (70),

$$\underline{\hat{\beta}}_{p1} = P_1'\hat{\delta}_1$$

$$\text{var } (\hat{\beta}_{p1}) = \sigma^2 P_1' D_1^{-1} P_1 \qquad (81)$$

where $\hat{\underline{\beta}}_{p1}$ is the estimate of $\underline{\beta}$ in (70) based on a regression on the first k_1 principal components.

Next, to retain some variables in their original form in the regression, the design matrix can be partitioned into $X = (X^*, X^{**})$, and the basic equation becomes

$$\underline{y} = X^*\underline{\beta}^* + X^{**}\underline{\beta}^{**} + \underline{u}. \tag{82}$$

Let

$$V = M(X^*)X^{**}$$

and rewrite the basic equation as

$$\begin{aligned} \underline{y} &= X^*\underline{\beta}^* + V\underline{\beta}^{**} + \underline{u} \\ &= X^*\beta^* + S\underline{\delta} + \underline{u} \end{aligned} \tag{83}$$

where the relevant P matrix used to construct S is now the matrix of characteristic vectors of V'V and the rest follows along the previous discussion.

5. Simultaneous equations

In this section the foregoing formulation and approach are adapted to deal with the models of simultaneous equations. In part, such an adaptation is simply an extension of the size of the system resulting in an additional complexity of expressions and perhaps of computations. However, in part, the introduction of the time-varying parameters to simultaneous equations introduces conceptual considerations which have to be dealt with explicitly. This is particularly the case with the identifiability of the system. In any event, this section is devoted to the formulation of the problem in such a way as to allow the utilization of the foregoing discussion and thereby minimize repetition.

Consider a system of G simultaneous equations, with a row vector $\underline{y}_t$ of endogenous variables in time t, and a row vector $\underline{x}_t$ of exogenous variables:[10]

$$\underline{y}_t \, B_t + \underline{x}_t \, \Gamma_t = \underline{u}_t \tag{84}$$

where B_t is a G x G now singular matrix and Γ_t is a K x G matrix. Both matrices, B_t and Γ_t, are allowed to vary with time as the index t indicates. In addition, it is assumed that:

$$\underline{u}_t \sim (0, \Sigma_u) \text{ and } E\,(\underline{u}'_t \underline{u}^*_t) = 0 \qquad \forall\, t \neq t^* \tag{85}$$

$$E\,(u_{gt}, x_{g^*t^*}) = 0, \; \forall_{g,t,g^*,t^*} \tag{86}$$

$$E\,(u_{gt}, \beta^*_{gr,t^*}) = 0, \; \forall_{g,g^*,t,t^*,r} \tag{87}$$

$$E\,(u_{gt}, \gamma_{g^*,s,t^*}) = 0, \; \forall_{t,g^*,s,t,t^*} \tag{88}$$

$$\text{Plim } \frac{1}{T} X'X = M_x < \infty \tag{89}$$

where $X = \underline{x}_1 \; . \; . \; . \; \underline{x}_T$ is a T x K matrix.

As in the case of single equation, the coefficient matrices, B_t and Γ_t, can be decomposed into systematic and error components:

$$B_t = B^*_t + E_t \text{ and } \Gamma_t = \Gamma^*_t + H_t \tag{90}$$

where the starred matrices are the systematic components. Combining (84) and (90):

$$\underline{y}_t \, B^*_t + \underline{x}_t \, \Gamma^*_t = \underline{u}_t - \underline{y}_t E_t - \underline{x}_t \, H_t. \tag{91}$$

Each of the systematic components can be further decomposed into a constant, a function which depends on past values of internal variables (endogeneous and exogenous) and on variables external to the model. Such a decomposition is integrated into the following discussion. In so doing, the system is first rewritten for the T observations:

$$Y^d B + X^d \Gamma = U \tag{92}$$

where

$$Y^d = \text{diag } (\underline{y}_t),\ T \times GT$$

$$X^d = \text{diag } (\underline{x}_t),\ T \times KT$$

$$U = \{\underline{u}_t\},\ T \times G$$

$$B = \begin{pmatrix} \cdot \\ \cdot \\ B_t \\ \cdot \\ \cdot \\ \cdot \end{pmatrix},\ GT \times G$$

and

$$\Gamma = \begin{pmatrix} \cdot \\ \cdot \\ \Gamma_t \\ \cdot \\ \cdot \end{pmatrix},\ KT \times G.$$

It should be noted that the system in (92) is not a complete system; it artificially expands the number of endogenous variables from G to GT variables while at the same time the number of equations remain unchanged.

However, this apparent problem is handled by expressing the coefficients as functions of the various variables as was done in the case of the single equation. In so doing, it is not necessary at this point to distinguish between the various kinds of such variables. The systematic parts of the coefficients are, therefore, expressed as:

$$B_t^* = W_{Bt}A_B, \qquad \Gamma_t^* = W_{\Gamma t}A_\Gamma \tag{93}$$

where W_{Bt} and $W_{\Gamma t}$ are GxS_B and GXS_Γ matrices of variables and A_B and A_Γ are S_B x G and S_Γ x G matrices of their respective coefficients which are held constant over time. It is now possible to combine (93) and (84):

$$(\underline{y}_t\, W_{Bt})\, A_B + (\underline{x}_t W_{\Gamma t})\, A_\Gamma = \underline{u}_t - y_t\, E_t - \underline{x}_t\, H_t. \tag{94}$$

Combining (90), (92), (93), and (94), the system can now be written as:

$$\begin{pmatrix} \underline{y}_1 & & 0 \\ & \ddots & \\ 0 & & \underline{y}_T \end{pmatrix} \begin{pmatrix} W_{B1} \\ \vdots \\ W_{BT} \end{pmatrix} A_B + \begin{pmatrix} \underline{x}_1 & & 0 \\ & \ddots & \\ 0 & & \underline{x}_T \end{pmatrix} \begin{pmatrix} W_{\Gamma 1} \\ \vdots \\ W_{\Gamma T} \end{pmatrix} A_\Gamma =$$

$$\begin{pmatrix} \underline{u}_1 \\ \\ \underline{u}_t \end{pmatrix} - \begin{pmatrix} \underline{y}_1 & & 0 \\ & \ddots & \\ 0 & & \underline{y}_T \end{pmatrix} \begin{pmatrix} E_1 \\ \\ E_T \end{pmatrix} - \begin{pmatrix} \underline{x}_1 & & \\ & \ddots & \\ & & \underline{x}_t \end{pmatrix} \begin{pmatrix} H_1 \\ \\ H_T \end{pmatrix}$$

all in a compact form:

$$PA = E \tag{95}$$

where P represents the matrices of product variables:

$$P \equiv (P_B,\ P_\Gamma) = \left| \begin{pmatrix} \\ \vdots \\ \underline{y}_t W_{Bt} \\ \vdots \\ \\ \end{pmatrix}, \begin{pmatrix} \\ \vdots \\ \underline{x}_t W_{\Gamma t} \\ \vdots \\ \\ \end{pmatrix} \right|. \tag{96}$$

P_B and P_Γ are T x S_B and T x S_Γ, respectively; hence, P is a T x S matrix. The matrix $A \equiv \begin{pmatrix} A_B \\ A_\Gamma \end{pmatrix}$ is a S x G matrix, with $S = S_B + S_\Gamma$, and

$$\varepsilon = U - Y^d E - X^d H \tag{97}$$

$$E = \begin{pmatrix} \vdots \\ E_t \\ \vdots \end{pmatrix}, \quad H = \begin{pmatrix} \vdots \\ H_t \\ \vdots \end{pmatrix}.$$

In confronting this system with the data, a number of issues arise:

1. Estimation of structural parameters. Since these parameters vary with time, the best that can be accomplished is to estimate the systematic components B_t^* and Γ_t^*. Having obtained such estimates and making appropriate assumptions about the distribution of the error terms, inferences can be made with respect to the distribution of the coefficients themselves.
2. Predicting $\underline{y}$ for a given vector $\underline{x}$ on the basis of past observations of the y's, x's, and other variables which appear in the B* and Γ* functions.
3. Same as (2) above combined with policy constraints on some of the variables.

These issues are dealt with in the following two sections.

5.1. Identification and estimation

The estimation of A in (95) can be divided into two parts: First, obtaining a consistent estimator of A and, second, obtaining an efficient

estimator. This distinction requires emphasis in view of the complex variance structure of the error term in (95).

The problem of obtaining a consistent estimate for A becomes a standard problem in simultaneous equation estimation. However, it should be noted that the endogenous variables on the LHS are now replaced by the product terms with the W's which in general are exogenous or lagged endogenous variables. As such, the original endogenous variables are in the row space of the W's. Therefore, much of the simultaneous equation problem disappears; and the OLSE of (95) are likely to be consistent.

It may be useful to relate this discussion to the problem of identification in the original model.[11] If a particular equation is originally identified, it remains so under the present case. This is simply because, instead of using the original exogenous variables for estimation, their product with the W variables are now used; and, consequently, the number of exogenous variables are increased. For this reason, an equation that originally was not identified may now be identified. Basically, the identification is used through the variability of the coefficients, a factor which does not exist in the constant coefficient models. To further elaborate on this point, it might be helpful to view the system from a somewhat different point of view. The system basically consists of two blocks; the first is given by equation (84) and the second is given by equations (90) and (93), i.e.,

$$B_t = W_{Bt} A_B + E_t$$
$$\Gamma_t = W_{\Gamma t} A_P + H_t. \tag{98}$$

Since the W's are predetermined, it is possible to estimate (98) by OLS, and the only obstacle to this is the fact that the dependent variables are unobservables. It is for this reason that it is necessary to bring in (84) and

to use its structure so as to infer the B_t and Γ_t from the observed variables in (84). This, however, by itself does not destroy the identification if it exists and helps to identify the system when the identification does not exist. The latter is achieved by the introduction of the additional exogenous variables, W's. As it is well known, identification can be achieved through variations of equations caused by the variations of the exogenous variables or by the variations of the disturbances if those are known. The formulation utilizes both; it relates some of the variations in the disturbances to variations in the coefficients and those, in turn, are related to "new" exogenous variables. The reader may find it useful to illustrate this to himself using the simple framework of supply and demand.

Once a consistent estimate of the system is available, the residuals are computed and those, in turn, are used to estimate the various variance components using the same approach advanced in Section 3. These estimates can be used to obtain a final estimate using OLS procedure.

5.2. Forecasting

Once the system is estimated, it can be used for forecasting. Write the constrained reduced form equation for year t:

$$\underline{y}_t = \underline{x}_t \Pi_t + \underline{v}_t$$
$$-\Pi_t = \Gamma_t B_t^{-1}, \quad \underline{v}_t = \underline{u}_t B_t^{-1} \tag{99}$$

where Γ_t and B_t^{-1} contain the various restrictions of the system. For forecasting, estimates of the parameters in question replace the actual parameters. If policy affected the parameters in question, then this influence should be taken into account in the estimation.

The estimates of the parameters in the system are used for predicting the values of the parameters in B_t and Γ_t. It is possible to obtain direct reduced form estimation for Π_t from

$$\Pi_t = W_{\pi t} A_\pi + E_{\pi t} \tag{100}$$

where $W_{\pi t}$ consists of W_B and W_Γ. By substituting (100) into (99), i.e.,

$$\underline{y}_t = (\underline{x}_t W_{\pi t}) A_\pi + (\underline{x}_t E_{\pi t} + \underline{v}_t). \tag{101}$$

It is easy to show that OLS applied to (101) is asymptotically unbiased and very likely also unbiased in finite samples, depending on the relationships between the W's and the error terms in (84).

For direct structural equations predictions, estimates of A_B and A_Γ are first obtained; then, from (93), estimates of B_t^* and Γ_t^* which are also used as predictors are obtained. These values are, in turn, used to compute Π for the predicted period, and this estimate is used for predicting y.

With regard to W matrices, some explicit statements can be made. They are basically similar in nature to the variables employed in the case of regression. As for this earlier formulation, it is possible to construct lagged functions of the explanatory variables, but now they may be in terms of both endogenous and exogeneous variables. The data can be employed, in part, to determine the nature of the lag structure and the number of variables. In addition, external variables or qualitative variables can be incorporated. Principal component methods can, again, be used in order to reduce the number of variables in the system.

For prediction under policy constraints, basically the procedure is repeated; but the constraints are brought explicitly into the model. In general

the sampling error of the estimates declines when constraints are imposed. Does it, therefore, improve prediction? This, of course, depends a great deal not only on the statistical properties of the model but also on its actual performance--specifically, the gap between the constraint which was assured in the model and the level of this constraint in reality. The greater the discrepancy between the two, the larger is likely to be the net effect on the error term.

6. Concluding remarks

While there is some merit in treating parameter effects as relatively fixed, the most prevailing circumstances in economic analysis, especially for agricultural systems, call for the specification of parameter variation formulations. These circumstances relate to the nature of available secondary data and the abstractions that are imposed to construct econometric models. When these circumstances are added to the observations of Lucas on the theoretical justification for parameter variation formulation and their operational implementation in the context of model use, the overwhelming evidence is in favor of such formulations. The model-use argument is often neglected; but for those of us who have attempted to derive reliability measures, forecasts, or policy impact probability distributions, this argument is very important indeed. Finally, many circumstances arise in which the "true" coefficients themselves are generated by a nonstationary or time-varying process.

The general specification advanced in this paper has the unique feature of incorporating lag and current effects of included variables and the effects of excluded variables on the systematic movements of the parameters appearing in equations of interest. It has been shown that this general formulation admits as special cases the classical linear model; the Cooley-Prescott adaptive

regression model; the Belsey systematic variation model; Swamy random coefficient model; the Cooley-Prescott time-varying parameter model; the Singh et al. mean response model; the Goldfeld and Quandt switching regression model; and the spline regression model.

Due to the nonobservability of the varying parameters, the estimation takes place in the context of an estimable form which includes both additive and interaction effects. The estimator that was constructed for the general case conforms to the two-stage Aitken method for unknown variance; the first round involves application of the ordinary least squares to the estimable equation, the second round generates estimates of the induced heteroscedastic structure of the error terms in the estimable equations, and the third round reestimates the additive and interaction effects taking into account the heteroscedastic structure. As with all two-stage Aitken procedures of this sort, the nonnegativity property of the variances may be violated; hence, procedures requiring the application of inequality estimators or ridge regressions can be employed to force the conformity of the empirical results with this property.

It has been demonstrated that the general specification and estimation for the single equation model can be generalized to simultaneous equation systems. In this setting, one of the more interesting implications of the general specification is that the identifiability of the complete system is often enhanced. In contrast to conventional constant parameter simultaneous equation specifications, the formulation advanced in this paper replaces explanatory variables with the product of these variables with the systematic factors influencing equational parameters. In essence, this result augments the number of predetermined variables and exogenous distinguishing information to identify the

structural parameters. For this reason, an equation that originally was not identified may become identified. Hence, the identification is enhanced through the variability of the coefficients themselves--an influence which does not exist in constant coefficient models. In terms of the usual sources of identifiability, namely, variations and equations caused by variations of exogenous variables or by variations of the disturbances, the formulation relates some of the variations in the disturbances to variations in the coefficients and those, in turn, are related to "new" exogenous variables.

In the context of actual empirical applications, particularly those oriented toward model uses such as forecasting and policy analysis, the above story is incomplete. The estimable form resulting from the parameter variation formulation introduces problems of multicollinearity. Operational procedures for appropriately dealing with these multicollinearity problems have been examined in detail. Formal variable selection rules have been developed which recognize the trade-off between an increase in sampling variance and a decline in bias associated with the inclusion of additional variables. The conditions leading to a large sampling variance are small sample variance of potential explanatory variables, high multicollinearity between potential explanatory variables and included explanatory variables, and a small sample size. An operational test is developed to select explanatory variables for forecasting purposes. This test demonstrates that the conventional procedure of testing to determine whether or not the coefficients of potential explanatory variables are zero is neither necessary nor sufficient.

The above selection procedures in the context of the parameter variation formulation are made operational by the use of principal components. The multicollinearity problem arises primarily among the interaction terms and between the interaction terms and the additive terms included in the estimable equation. It is recommended that principal components be taken only on the

interaction terms to preserve the additive effects in terms of original variables. Here, too, conventional selection rules for the principal components are not appropriate. That is, the procedure of selecting principal components according to the size of the appropriate characteristic roots does not provide an optimal result; instead, optimality calls for entering the principal components in a decreasing order of their respective t ratios.

The integration of all of the results of this paper that properly balances the cost of the complexity and the value of simplicity leads to the following procedure: (1) form the estimable equation, derive principal components of the interaction terms according to the specified selection criteria, and apply ordinary least squares; (2) estimate the covariance matrix of the heteroscedastic structure of the estimable form; and (3) combine the estimated covariance matrix with the principal components of the interaction variables and the original additive variables to generate generalized least squares estimates. Of course, this approach can be easily augmented to deal with simultaneous equations by deriving appropriate instrument variables for the explanatory endogenous variables prior to implementing the suggested procedure.

Footnotes

*Giannini Foundation Paper No. (reprint identification only).

†Professor of Economics, Hebrew University, and Professor of Economics, University of Chicago.

††Professor and Chairman, Department of Agricultural and Resource Economics, University of California, Berkeley. Note that senior authorship is not assigned.

[1]These practitioners have also demonstrated that more variability is present in econometric models than can be captured by conventional autoregressive error specifications.

[2]In this paper, the issues of appropriate experimental designs in determining what data ought to be used shall not be examined explicitly in the construction of forecasting models. In nonexperimental areas such as econometrics and other social sciences, auxilliary conditions ought to be examined to determine whether a given set of data is appropriate for estimating parameter effects. Such auxilliary conditions play a crucial role in experimental fields. Procedures which incorporate the last piece of information minimize the importance of deciding whether or not the last observation satisfies experimental design conditions for estimating parameters, testing hypothesis, and making inferences. Focusing on updating or forecasting ignores other equally important problems in sample selection regarding what the first observation ought to be. In the experimental sciences, the first observation is easily identified; and updated estimates of the unknown effects are appropriate provided the same experiment is ongoing. Such clarity is unavailable in economic modeling. Specifically, it is not clear that the process generating

the data is continually under the influence of the same experiment. This, of course, is why the choice of sample data for econometric model construction contains many elements of art rather than science.

[3]The authors wish to thank S. R. Johnson for this suggestion.

[4]The discussion is of no interest for large samples where the sampling error is of no quantitative importance.

[5]The term $\Sigma_2 h_1$ is obtained from the evaluation of the last term on the RHS. For instance, to obtain the j, r element ($j, r = k_1 + 1, \ldots, k$), the jth row of U_2' and rth column of U_2 is taken:

$$E\{\underline{U}_{2j}'[\cdot]\underline{U}_{2r}\} = \text{tr } E\{\cdot\cdot\} = \sigma_{jr} \text{ tr } [\cdot] = \sigma_{jr} h_1.$$

[6]Under this assumption, $V_2'V_2$ has a Wishart distribution $W_{k_2}(n, \Sigma_2)$ and is independent of $\underline{v}_2$ (which does not belong to the sample) and has a k_2 - normal distribution with zero mean and variance given by (57). Consequently:

$$T^2 = \frac{n}{1 + h_1} \underline{v}_2' (V_2'V_2)^{-1} \underline{v}_2 = \frac{n}{n - k_1} \underline{v}_2' \left(\frac{V_2'V_2}{n - k_1} \right)^{-1} \underline{v}_2 \frac{1}{1 + h_1}$$

$$= \frac{n}{n - k_1} \underline{v}_2^1 \hat{\Sigma}_{v_2}^{-1} \underline{v}_2.$$

[7]Given the notations:

$$\frac{n - k_1 + 1}{k_2 n} T^2 \sim F(k_2, n - k_1 + 1)$$

and, hence,

$$\underline{v}_2' \hat{\Sigma}_{v_2}^{-1} \underline{v}_2 \sim k_2 F(\cdot).$$

[8]This statement may seem to depend on the scaling of the variables (units of β_2), but this is not so as the units of $||\underline{\beta}_2||^2$ and Σ_2 are reciprocals.

[9]It is this property that does-not generalize to allow a selection of a method.

[10]Note that B_t was previously used to denote a bias. The meaning of the symbol will be cleared from the content.

[11]For examination of the identification problem in context of simultaneous equation models with random parameters, see Kelejian (1974). Unfortunately, this treatment imposes the assumption that all random parameters are independent of one another and, thus, is of little value in the current context.

References

Balestra, P. and M. Nerlove, 1966, Pooling cross-section and time series data in the estimation of the dynamic model: The demand for natural gas, Econometrica 34, 585-612.

Belsey, David A., 1973a, On the determination of systematic parameter variation in the linear regression model, Annals of Economic and Social Measurement 2, 487-494.

Belsey, David A., 1973b, The applicability of the Cameron filter in the determination of systematic parameter variation, Annals of Economic and Social Measurement 2, 531-533.

Buse, A. and L. Lim, 1977, Cubic splines as a special case of restricted least squares, Journal of the American Statistical Association 72, 64-68.

Chow, Gregory C., 1960, Tests of equality between sets of coefficients in two linear regressions, Econometrica 28, 591-605.

Cooley, T. F. and S. J. DeCanio, 1973, Changing supply response in late nineteenth century agriculture, Working paper (Tufts University, Medford, Massachusetts).

Cooley, Thomas F. and Edward C. Prescott, 1973a, An adaptive regression model, International Economic Review 14, 364-371.

Cooley, Thomas F. and Edward C. Prescott, 1973b, Varying parameter regression: A theory and some applications, Annals of Economic and Social Measurement 2, 463-473.

Cooper, R. L., 1972, The predictive performance of quarterly econometric models of the U. S., Econometric Models of Cyclical Behavior, Studies in Income and Wealth, Vol. 36 (National Bureau of Economic Research, New York).

Duffy, W. J., 1969, Parameter variation in a postwar econometric model, unpublished Ph.D. dissertation, University of Minnesota.

Fomby, T. B., R. C. Hill, and S. R. Johnson, 1978, An optimal property of principal components in the context of restricted least squares, Journal of the American Statistical Association 73, 191-193.

Fomby, T. B. and S. R. Johnson, 1977, MSE evaluation of ridge estimates based on stochastic prior information, Communications in Statistics A6(13), 1245-1258.

Greenberg, Edward, 1975, Minimum variance properties of principal component regression, Journal of the American Statistical Association 70, 194-197.

Goldfeld, S. M. and R. E. Quandt, 1973, The estimation of structural shifts by switching regressions, Annals of Economic and Social Measurement 2, 475-485.

Judge, G. G. and T. Takayama, 1966, Inequality restrictions in regression analysis, Journal of the American Statistical Association 61, 166-181.

Judge, G. G. and T. A. Yancey, 1978, Inequality restricted estimation under squared error loss, working paper no. 78-020, College of Business Administration, University of Georgia.

Kelejian, H. H., 1974, Random parameters in a simultaneous equation framework: Identification and estimation, Econometrica 42, 517-527.

Lucas, R., 1976, Econometric policy evaluation: A critique, Journal of Monetary Economics I, Supplement 3, 19-46.

Mundlak, Yair, 1978a, On the pooling of time series and cross section data, Econometrica 46, 69-85.

Mundlak, Yair, 1978b, On variable coefficients models, Annals de L'Insee, France.

Poirier, D. H., 1976, The econometrics of structural change (North Holland Publishing Company, Amsterdam).

Rausser, G. C., 1978, Model specification and use: Nonlinear parameters versus nonlinear variables, unpublished mimeo. (Harvard University).

Rausser, G. C. and P. S. Laumas, 1976, The stability of demand for money in Canada, Journal of Monetary Economics 2, 367-380.

Rosenberg, B., 1973a, A survey of stochastic parameter regression, Annals of Economic and Social Measurement 2, 381-397.

Rosenberg, B., 1973b, The analysis of a cross section of time series by stochastically convergent parameter regression, Annals of Economic and Social Measurement 2, 399-428.

Singh, B., A. L. Nagar, N. K. Choudhry, and Raj. Baldev, 1976, On the estimation of structural change: A generalization of the random coefficients' regression model, International Economic Review 17, 340-361.

Swamy, P. A. V. B., 1970, Efficient inference in a random coefficient regression model, Econometrica 38, 311-323.

Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association 62, 348-368.