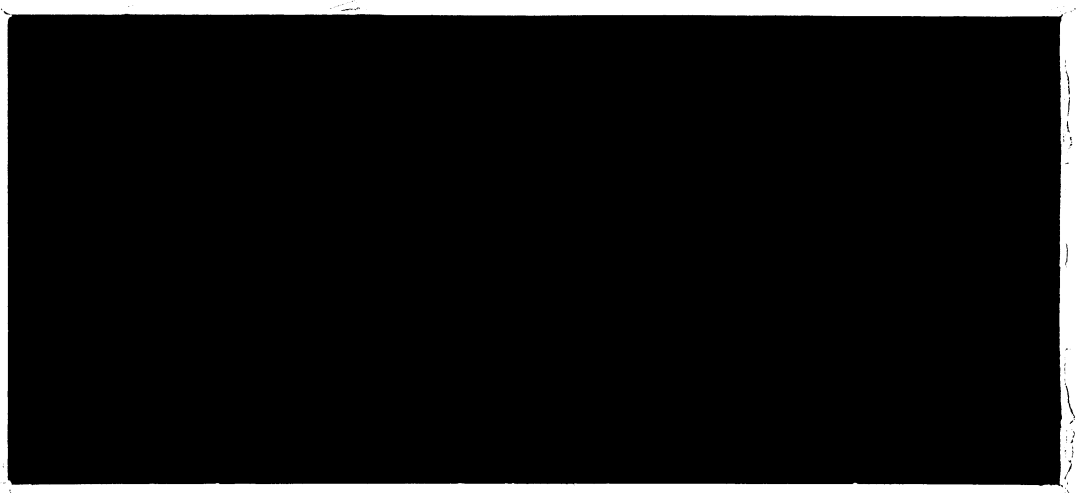
Working Paper Series



DEPARTMENT OF AGRICULTURAL AND
RESOURCE ECONOMICS

BERKELEY

CALIFORNIA AGRICULTURAL EXPERIMENT STATION

University of California

Division of Agricultural Sciences
UNIVERSITY OF CALIFORNIA

Working Paper No. 130

STRUCTURAL CHANGE, UPDATING, AND FORECASTING

Gordon C. Rausser, Yair Mundlak, and S. R. Johnson

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
April 1981

CHAPTER

STRUCTURAL CHANGE, UPDATING, AND FORECASTING*

Gordon C. Rausser, Yair Mundlak, and S. R. Johnson†

1. *Introduction*

Attempts to capture useful relationships for forecasting purposes in economics and other nonexperimental sciences 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 or environmental phenomena. In the context of controlled effects, economists generally operate with a version of the classical linear statistical model; this model presumes the parameters are constant for 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 data sets for which it is reasonable to assume parameters are "approximately constant." This problem is at the heart of the issues addressed 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 is the typical treatment of pre- and postwar data. To account for the difference in effects between pre- and postwar data, the general practice has been to introduce dummy variables to represent possible shifts in intercept and slope parameters.

---

†Gordon C. Rausser is Chairman and Professor of Agricultural and Resource Economics, University of California, Berkeley; Yair Mundlak is Professor of Economics, Hebrew University, Rehovot, Israel, and Professor of Economics, University of Chicago, Illinois; and S. R. Johnson is Professor of Economics and Agricultural Economics, University of Missouri, Columbia, and Visiting Professor of Agricultural and Resource Economics, University of California, Berkeley.

Although the dummy variable approach is indeed convenient, in many instances it can lead to grossly inaccurate forecasts. Neglecting issues of complexity, such specifications may be suboptimal. In time series regressions, it is appealing to view the data from, say, 1970 as more relevant to forecast for 1980 than data from the early 1950s. Relevant taste, expectation formation patterns, and sociological and environmental phenomena in 1970 were likely far different than 1950; and it would appear that in this sense the year 1970 contains more valuable information for forecasting.

In agricultural economic forecasting, the above observations are particularly relevant. Models constructed for various commodity systems, especially feed and food grain, and based on data prior to 1971, have revealed notorious inaccuracies in forecasting important economic variables for the years 1972 through 1976. These inaccuracies proved 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 constant parameter formulations have naturally arisen.

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

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 the 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. Many empirical researchers view much of the recent conceptual work on parameter variation in both statistics and economics as representing new gimmicks which contribute little in the way of useful empirical information.

To be sure, there is merit in considering certain parameter effects as fixed in economic models. Their principal advantage is simplicity in providing insights about economic interrelationships, unclouded by the meticulous details of a changing real world. There is also, however, merit in obtaining more accurate representations, forecasts, and economic policy evaluations. Hence, from a research strategy standpoint, the principal issue faced in deciding whether to employ a constant or 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, stationary parameter formulation. This formulation has been advanced principally for time series and cross-section observations (Rosenberg, 1973b). In particular, a cross section of individuals is presumed to possess the same regression regime over time, but their 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 hypothesized to change over time, a more general nonstationery formulation results.

1.1 *Misspecification and Parameter Variation*

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 involving 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 forecast accuracy. Effects of such misspecifications can be countered by introducing an appropriate parameter variation structure. Important types of misspecifications which arise in the construction of forecasting models include omitted variables, proxy variables, aggregate data, and nonlinearities.

The omission of important explanatory variables can 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 not alter the parameter effects of included variables, provided the omitted variables are orthogonal to those that are included. However, time series for such omitted variables exhibit nonstationery behavior, and they are often not orthogonal to 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 term.

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 expectations formation patterns and measures of capital. Unfortunately, these and other proxy variables detect only partially changes in the levels of economic stimuli they purport to measure. Furthermore, the relationship between the true variable and its proxy can be expected to change over time. Under these circumstances, changes in the true variables which measure the actual economic stimuli induce instability in the estimated parameters for the proxy variables.

For aggregate data, the possibility of parameter instability has been demonstrated widely. Since aggregate data are measured by weighting the

relative importance of the heterogenous sets of microunits, the parameters in the estimated aggregate equation will remain constant only so long as these weights do not vary. With 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).

Another potential cause for parameter variation is 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 is reasonable 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. The approximation of highly nonlinear "true" relationships by simpler functional forms, along with observations outside a narrow sample range, provides perhaps the strongest motivation for a varying parameter structure.

1.2 *Theory and Parameter Variation*

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, for example, has argued that the constant parameter formulation is inconsistent with economic theory. He notes that a change in policy will cause 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 their behavior.

In the case of commodity systems, a number of illustrations of the points raised by Lucas are available. One of the better 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 distorted 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 patterns 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 resulting in larger supplies, substantially lower prices, and significant negative margins. Hence, the price ceilings had the immediate effect of a substantial shift in price expectations which, in turn, had drastic implications for dynamic supply responses, ultimate market realizations, and cattle inventories.

Dynamic economic theory and the notion of rationale expectations does not in general provide for well-defined behavioral equations whose parameter effects can be treated as constants. 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 shifts in the relevant estimable behavioral equations. This source of parameter instability can only be avoided by reasonably accurate measurements of expectation formulation patterns and dynamic responses--a dubious prospect at best.

1.3 *Tractability and Parameter Variation*

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 model dimensions and problems of numerical accuracy. Accuracy is indeed an important issue when the structural model representation is nonlinear. For if such models are simultaneous representations of large systems, it is not in general possible to obtain unique reduced forms. In computing the necessary derivatives to obtain the reduced forms 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 risks 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 nonlinear in the parameter space. By varying the parameter effects, any nonlinear representation can be appropriately approximated (Rausser). This approach allows forecast probability distributions--unconditioned or conditioned on alternative policy actions--to be generated for a particular point in the parameter space. Along similar lines, the approach also simplifies the validation of constructed models, especially derivations of dynamic properties.

The cumulative implication of the above observations is that, given the forces often neglected in economic models, it is overly optimistic to presume that parameters will be identical over the complete sample 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 complexities of such formulations. Can these formulations capture the enduring characteristics of the processes under examination? The purpose of this chapter is to point us in the direction of answering this question in a definitive manner.

2. *Structural Change and Parameter Variation*

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

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

where

$$u_t \sim \left(0, \sigma_u^2\right), \; E(u_t x_t) = E(u_t \beta_t) = 0, \qquad t = 1, 2 \ldots .$$

Note that the parameter β_t is subscripted indicating that it can change for the different sample observations. The changes in β can be systematic or be completely random without a structure imposed. Systematic or structured

changes can be caused by the outside environment and be completely exogenous to the system or alternatively can be induced by variables within the system itself. For instance, in demand equations it is possible that the parameters depend on past consumption. A commodity can be habit forming or alternatively its consumption can be at a saturation point. In both cases the observed demand response depends on past experience. Furthermore, social change can affect taste and lead to variation in the parameter structure. One instance on this behavior has been observed for 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 the popularity of convenience foods emanating from the women's movement and increased participation of females in the U. S. labor force. These and similar influences can be sources for hypotheses of parameter variation conditioned by variables outside and within the system. Specifically, let

$$\beta_t = \beta_0 + L(x_t) + z_t\alpha + e_t \tag{2}$$

where

$$e_t \sim \left(0, \sigma_e^2\right), \quad E(e_t z_t) = E[e_t\, L(x_t)] = 0 \qquad t = 1, 2 \ldots .$$

The term z_t represents the outside effect on the coefficients, whereas $L(x_t)$ represents the effect of the variables within the system. When these two effects are not present, then β_t is random with mean β_0 and error e_t. In the most special case when the variance of e is zero and the effects of z_t and $L(x_t)$ are not present, β_t is constant as in the usual regression analysis formulation.

More generally, however, β_t is conceived as a random variable with a systematic component conditioned 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.,

$$y_t = x_t \beta_0 + L(x_t)\, x_t + x_t z_t \alpha + \varepsilon_t \tag{3}$$

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 an indicator variable (0, 1) which introduces a switch in the regression coefficient. Of course, there can be more than one switch; and the introduction of indicators for more than one switch is a straightforward extension. The function $L(x_t)$ is generally expressed but usually specified to reflect the impact of past values of the x's. An important and interesting feature of this function can be observed by expressing $L(x_t)$ as a geometric distributed lag, i.e.,

$$(4) \qquad \beta_t = \beta_0 + \delta x_{t-1} + \delta^2 x_{t-2} + \delta^2 x_{t-2} + \ldots + \alpha z_t + e_t,$$

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

$$(5) \qquad \beta_t = \delta\beta_{t-1} + (1 - \delta)\beta_0 + \delta x_{t-1} + \alpha(z_t - \delta z_{t-1}) + (e_t - \delta e_{t-1}).$$

Expression (5) is a generalization and extension of the Markov chain formulation examined by Rosenberg (1973a) and Belsey (1973a, 1973b). It is also widely used in the engineering literature. Since δ is between 0 and 1, the term $\delta\beta_{t-1}$ can be viewed as representing a decay process. The convergence of this process is to $\beta_0(1 - \delta)$ plus terms involving x_{t-1}, z_t, and z_{t-1}. Hence, expression (5) generalizes the Markov process where the convergence is to a constant conditioned by x_{t-1}, z_{t-1}, and z_t.

Finally, it should be observed that equation (3) has a heteroscedastic error structure. Under the assumption of independence between u and e, the variance of the error term ε_t is

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

2.1 *Alternative Specifications*

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

$$\beta_t = \delta_0\beta_0 + \delta_1\beta_{t-1} + \delta_2 x_{t-1} + \alpha z_t^* + \xi_t \tag{7}$$

where the parameters associated with β_{t-1} and x_{t-1} differ. For (7) to be equivalent to (5), $\delta_1 = \delta_2 = \delta$; $\delta_0 = 1 - \delta$; $z_t^* = (z_t - \delta z_{t-1})$; 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 (unless otherwise specified, ξ_t is assumed independent and identically distributed with zero mean and finite variance σ_ξ^2). More specifically,

a. Classical linear model: $\delta_0 = 1$, $\delta_1 = 0$, $\delta_2 = 0$, $\alpha = 0$, $\xi_t = 0$. Hence,

$$\beta_t = \beta_0 \quad \text{for all t.} \tag{7a}$$

b. Cooley and Prescott (1973a) adaptive regression model: $\delta_0 = 0$, $\delta_1 = 1$, $\delta_2 = 0$, $\alpha = 0$, $u_t = 0$, $x_t = 1$, for all t. Hence,

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

c. Belsey (1973b) systematic parameter variation model: $\delta_0 = 0$, $\delta_1 = 0$, $\delta_2 = 0$. Hence,

$$\beta_t = \beta_0 + \alpha z_t + \xi_t. \tag{7c}$$

d. Swamy random coefficient model: $\delta_0 = 1$, $\delta_1 = 0$, $\delta_2 = 0$, $\alpha = 0$. Hence,

(7d) $$\beta_t = \beta_0 + \xi_t.$$

e. Cooley and Prescott (1973b) time-varying parameter model: $\delta_0 = 0$, $\delta_1 = 1$, $\delta_2 = 0$, $\alpha = 0$, $u_t = 0$, $\xi_t = \eta_t - \nu_{t-1} + \phi_t$. Thus,

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

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

and

(7e) $$\beta_t = \beta_{t-1} + \alpha_0 w_t + \alpha_1 w_{t-1}$$

where $w_t = [\phi_t, \eta_t]$; $\alpha_0 = [1, 1]$; and $\alpha_1 = [0, -1]$.

f. Singh et al. mean response model: $\delta_0 = 1$, $\delta_1 = 0$, $\delta_2 = 0$, $z_t^* = 1$, and $\alpha = \bar{\alpha} f(t)$. Thus,

(7f) $$\beta_t = \beta_0 + \bar{\alpha} f(t) + \xi_t$$

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

g. Goldfeld and Quandt switching regression model: $\delta_0 = \delta_1 = \delta_2 = \xi_t = 0$, $\alpha = 1$, and $z_t = \beta_1$ for $t \epsilon I_1$ and $z_t = \beta_2$ for $t \epsilon I_2$, where I_1 and I_2 represent the sets of indices for which two separate regression equations for the two regimes hold. Hence,

(7g) $$y_t = x_t \beta_1 + u_t \qquad t \epsilon I_1$$

$$y_t = x_t \beta_2 + u_t \qquad t \epsilon I_2.$$

h. Spline regression model (Poirier; Buse and Lim). 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 is assumed constant; and for the slope, $\delta_0 = 1$, $\delta_1 = \delta_2 = \xi_t = 0$, and α and z are defined as the vectors, $\alpha' = (\alpha_1, \alpha_2)'$ and $z_t = (t, t - t_1)$. Hence, for the slope coefficient

$$\beta_t = \beta_0 + \alpha_1 t + \alpha_2 (t - \bar{t}), \tag{7h}$$

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 the index set t with a known knot at $\bar{t}$.

The above special cases (7a) through (7h) admit an entire spectrum of possible parameter evolutions. The first case (7a) is the conventional constant parameter specification. The second, the adaptive regression model of Cooley and Prescott, treats the parameter evolution only for the intercept or constant term. Under this hypothesis, the parameter 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 systematic changes in the parameters. Such influences are often qualitative in nature and cannot be generally measured with accuracy. The fourth formulation or 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 (7e) has been applied to money demand relationships (Rausser and Laumas) and supply response elasticity for wheat (Cooley and DeCanio). This formulation allows systematic variation in both the intercept and slope coefficients of multiple regression models. The two-component process on the unknown parameter effects allows interpretations similar to the permanent income hypothesis. Both transitory and permanent variations in the effects described by the parameters are allowed by this formulation.

The sixth special case advanced by Singh et al. provides a mild generalization to the Belsey (7c) and the random coefficient model of Swamy (7d). 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 parameters. The switching regression model generalizes the conventional

dummy variable formulation which presumes that *a priori* information is available to classify various regimes. The Goldfeld and Quandt approach endogenizes the distribution of the regimes. When *a priori* information on regimes is not readily available, the Goldfeld and Quandt approach is preferable, particularly when the parameters move by discrete jumps. Finally, the spline formulation offers advantages in structuring the nature of parameter variation. For example, with a linear spline, the formulation is equivalent to a piecewise linear approximation.

2.2 *Desired Estimates*

In operating with specifications allowing varying parameters, issues associated with the type of estimates desired naturally arise. Resolutions of these issues depend 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 $\beta_{t/t+j}$ denotes 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 smoothed 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 is crucial for 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 an analysis of the consequences for specification errors which arise with the application of conventional estimation methods. This is followed by a suggested estimation procedure. The procedure involves a two-stage approach and places emphasis on the estimation of the variance components associated with the random variables, 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 *Consequences 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. For the simple regression model, the estimate of the constant slope is

$$b = \frac{\Sigma y_t x_t}{\Sigma x_t^2}$$

(8)

$$= \frac{\Sigma x_t^2 \beta_t}{\Sigma x_t^2} + \frac{\Sigma x_t u_t}{\Sigma x_t^2} .$$

The expected value of this estimator is given by

$$E(b) = \Sigma w_t \beta_t \tag{9}$$

where $w_t = x_t^2 / \Sigma x_t^2$; hence, $0 < w_t < 1$, $\Sigma w_t = 1$. Thus, the expected value of the regression coefficient is a weighted average of the individual coefficients. The weights are given by the squared values for x_t as a proportion of the sum of the squared values, Σx_t^2. Expression (9) imposing the parameter structure hypothesis form equation (2) can be written

$$E(b) = E[\beta_o + \Sigma L(x_t)\, w_t + \alpha \Sigma z_t w_t + \Sigma e_t w_t]$$

(10)

$$= \beta_o + T r_{Lw}\, s_w s_L + T\, \bar{L}(x_t)\, \bar{w} + \alpha\, T\, r_{zw}\, s_z\, s_w + \alpha\, T\, \bar{z}\, \bar{w}$$

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 $L(x_t)$ and z_t 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 if a structure such as (2) is appropriate.

Another consequence of the misspecification of constant parameters is the possible introduction of serial correlation. 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 a composite error. This error may be serially correlated even though the u_t is uncorrelated. 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 $L(x_t)$ or z_t, the β_t will be serially correlated. Consequently, (8) will not be 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, albeit at the expense of also requiring estimates of the additional parameters defining (2).

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, as a result of the misspecification, (8) is a biased estimator of the individual β_t. The square of the bias is

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

The variance of the estimator conditional on x_t and hypothesis (2) for β_t is

(14) $$\text{var}\ (b) = \sigma_u^2 \Big/ \Sigma x_t^2.$$

Combining the bias square and the variance yields as the mean-square error (MSE)

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

Clearly, the MSE can be reduced in estimating the varying parameter equation by incorporating the true structure, equation (2).

3.2 *Estimation Method*

The error term in (3) has a heteroscedastic structure; therefore, it is desirable to use a two-stage generalized least squares (GLS) estimator. The first stage involves estimation of (3) by ordinary least squares (OLS), followed by use of the residuals in the estimation of the parameters for the appropriate covariance matrix. The second stage employs the estimated covariance to obtain a feasible GLS estimator. The approach for computation of the covariance matrix is not immediately clear. However, this issue is investigated in Section 3.4.

The procedure outlined above will provide estimates of β_0, α, and $\underline{L}$, the parameter vector associated with the structure $L(x_t)$. The values of the parameters so obtained can now be introduced into equation (2) to estimate β_t. This estimate is

(16) $$\hat{\beta}_t = \hat{\beta}_0 + \hat{L}(x_t) + \hat{\alpha} z_t.$$

The estimates of β_0, α, and $\underline{L}$ 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, specifically for $L(x_t)$ and z_t. Early values for x_t get very little weight in $L(x_t)$ as specified, for example, in equation (4), whereas early values for z_t do not appear in (16). Thus, the early observations affect β_t only through their effects on the

estimates of β_0, α, and the estimators $\hat{\underline{L}}$. Clearly, there is no need to discard early observations since, given the assumed structure, they still provide information on the constant terms β_0, α, and $\underline{L}$.

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 dimensional vector of explanatory variables and write for the tth observation for a generalized version of (1) as

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

or for a sample of size T,

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

where $X^d = \text{diag}\{\underline{x}_t'\}$ is a TxTk matrix; $\underline{\beta}' = [\underline{\beta}_1', \ldots, \underline{\beta}_t']$; $\underline{\beta}_t$ is kx1 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,

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

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 values of x_j; $\underline{z}_t' = [z_{1t}, \ldots, z_{mt}]$; and $\underline{e}' = [\underline{e}_1', \ldots, \underline{e}_T']$. Lastly, for equation (20),

$$Z = \begin{bmatrix} I_k \otimes \underline{z}_1' \\ \\ I_k \otimes \underline{z}_T' \end{bmatrix} \quad \text{and} \quad L = \begin{pmatrix} I_k \otimes \underline{L}_1' \\ \\ I_k \otimes \underline{L}_T' \end{pmatrix} . \tag{21}$$

Combining (18) and (20),

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

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

$$E[\underline{\varepsilon}\ x_t'] = 0,\ E[\underline{\varepsilon}\ z_t'] = 0,\ E(\underline{e}\ \underline{e}') = I_T \otimes \Delta$$

where the elements of the diagonal matrix Δ represent the variances of random coefficients $\underline{\beta}_k$.

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

Since no observations are available on $L(\underline{x}_t')$, several possibilities can be considered in estimating (22). It is possible to impose a structure for $L(\underline{x}_t')$ such as a first-order distributed lag. The weights of this function are not known, but it is possible to iterate by assuming several alternative values for the weights and examine the values of the likelihood function. If

the likelihood function is not particularly flat, then the choice is easy. If the likelihood function is flat, then the choice is immaterial given the information in the sample; and any of the weighting schemes examined will provide equally good results. In the special case where L and Π are known, the following transformation can be made:

$$\underline{y} - X^d L\underline{\Pi} = X\beta_0 + X^d Z\underline{\alpha} + \underline{\varepsilon}. \tag{24}$$

In fact, this is the form of the expression that would be typically employed in estimating the conditioned values of the likelihood function.

Alternatively, instead of enforcing a particular structure on $L(\underline{x}_t')$, a sequence of lag values for $\underline{x}_t'$ can be specified. This is equivalent to introducing several terms, each one a product of $\underline{x}_t'$ and a lagged value of $\underline{x}'$. The advantage of this approach is that it does not presuppose a distribution for the lag coefficients. However, the cost is high and perhaps excessively high. Too many lagged values for $\underline{x}'$ will introduce multicollinearity into the system, and in many instances there will be sufficient information to sustain only one or two terms. This problem is particularly troublesome for forecasting. The problem arises not only because of the $L(\underline{x}_t')$ 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 4.

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 directly applied for the present framework since only one observation per year or time frame is available.

The method developed here utilizes the fact that the error term of equation (22) is a linear combination of the various e's and u, where the coefficients of the combinations are the $\underline{x}$'s themselves. Since the various error terms have different known "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) can be expressed as

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

Hence,

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

Under the assumption that X^* is of full rank, say, k^*, implying that rank $M(X^*) = T - k^*$, proceed by evaluating

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

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

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

and

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

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

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

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

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

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

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 which are directly observable variables.

A necessary condition for the existence of these estimates is $t - k > \frac{k(k+1)}{2} + 1$ or $T > \frac{(k+1)(k+2)}{2}$. Also, with large k, multicollinearity will preclude reliable estimates unless T is very large. Therefore, in time series studies where the number of varying observations is limited, it is desirable to keep the number of varying coefficients reasonably small so that their variances can be reliably estimated from the sample data.

Note that only T of the elements of $(\underline{\hat{\varepsilon}}\ \underline{\hat{\varepsilon}}')$ 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 $\underline{\hat{\varepsilon}}\ \underline{\hat{\varepsilon}}'$ 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 $\underline{\hat{\varepsilon}}\ \underline{\hat{\varepsilon}}'$.

For the special case of $\Delta \equiv 0$, note 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 $X^{*\prime}\ \underline{\hat{\varepsilon}} = \underline{0}$. 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:

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

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

It is well known that Δ is estimated from

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

where $\hat{\alpha}_i = y_{i\cdot} - y_{\cdot\cdot}$, $y_{i\cdot} = 1/J \sum_j^J y_{ij}$, and $\hat{\sigma}^2 = \frac{1}{I(J-1)} \sum_i \sum_j (y_{ij} - y_i)^2$. The point made by this model is that $\hat{\alpha}_i$ is a regression coefficient (of I dummy explanatory variables). Simultaneously, they are repeated observations on the random variable $\hat{\alpha}_i \sim (0, \sigma^2/J + \Delta)$. 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 equations (26) through (29).

In the ANOVA error components model, the estimator given above (obtained by substraction) 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). This simply involves introducing the restrictions

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

The sampling properties of the resulting estimator recently have been derived by Judge and Yancey 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 unrestricted 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 with conventional degrees of freedom.

An alternative approach to the problem of negative variance estimates is possible using stochastic shrinkage factors. 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). The *a priori* information introduced may be interpreted in terms of zero variances with a given probability. Ridge-type and other shrunken estimators dominate the unrestricted 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 are hypothesized to 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., $\underline{\beta}_t$ and y_t are predicted simultaneously. To place the problem of prediction within a useful framework, write the predicted value of y from a simple model 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. The error of this forecast is:

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

The expected value of the error is

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

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 $\hat{y}^*_{T+1}$ is

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

Within the Gauss-Markov framework, the bias is zero and the variance is minimized by using a GLS estimator.

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

$$\begin{aligned} B_{T+1} &= \Sigma w_t\,[\beta_o + L(x_t) + \alpha z_t + e_t] - [\beta_o + L(x_{T+1}) + \alpha z_{T+1} + e_{T+1}] \\ &= [\Sigma w_t L(x_t) - L(x_{T+1})] + \alpha\,[\Sigma w_t z_t - z_{T+1}] + [\Sigma w_t e_t - e_{T+1}]. \end{aligned} \tag{37}$$

This expression underscores the scope and importance of taking into account the explicit formulation on varying coefficients. 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 without the varying parameter may be almost unbiased. Hence, it is 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:

$$\hat{\beta}_{T+1} = \hat{\beta}_0 + \hat{L}(x_{T+1}) + \hat{\alpha}z_{T+1}. \tag{38}$$

Note that x_T is the latest value of x that can enter the function $L(x_{t+1})$, and this value is known at the time the forecast is constructed. Moreover, $\hat{\beta}_0$ 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 associated with the addition of variables to the regression model. This creates a choice problem which is particularly important in this framework. This problem is analyzed in the Appendix to this chapter.

4.1 *Utilization of Principal Components*

In view of the results presented in the Appendix, there is a limit to the number of parameters that should be used for forecasting. This raises a common problem in econometric analysis, that of a choice of variables. *A priori*, the model may suggest inclusion of more variables than the data can support. By eliminating some variables, the model is reduced to a manageable size. Such reductions need not be arbitrary. The problem of too many parameters is that of the sampling variance and, therefore, the choice as to reduction can be made to minimize the sampling variance for any predetermined

number of explanatory variables. Thus, the framework advanced in the Appendix can be applied by adding variables as long as the MSE declines.

This approach leads to a consideration of the principal components framework now briefly outlined. To simplify the development of this framework, the heteroscedastic error structure in basic model of Section 3.3 is neglected. For X, a matrix of k explanatory variables with full rank,

$$P'(X'X)\,P = \begin{pmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_k \end{pmatrix} \equiv D \tag{39}$$

where $\lambda_1 > \cdot\ \cdot\ \cdot > \lambda_k$ are the characteristic roots of $(X'X)$ and P a matrix of corresponding ortho-normal characteristic vectors. Note that

$$I = PP' = P_1P_1' + \ldots + P_kP_k' \tag{40}$$

and

$$X'X = PDP' = \lambda_1\underline{p}_1\underline{p}_1' + \ldots + \lambda\underline{p}_k\underline{p}_k'. \tag{41}$$

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

$$\begin{aligned} \underline{y} &= X\underline{\beta} + \underline{u} \\ &= (XP)(P'\underline{\beta}) + u = R\underline{\delta} + \underline{u}. \end{aligned} \tag{42}$$

Then the OLS estimator of $\underline{\delta}$ is

$$\hat{\underline{\delta}} = (P'X'XP)^{-1}\,P'X'\underline{y} = D^{-1}\,R'\underline{y} = \left\{\frac{\underline{r}_j'\underline{y}}{\lambda_j},\ j = 1, \ldots, k\right\} \tag{43}$$

with

$$\text{var}\ \hat{\underline{\delta}} = \sigma^2 D^{-1}. \tag{44}$$

From (39) and (44), it follows that the estimators of the coefficients δ_j are uncorrelated.

The relationship to the OLS estimator of $\underline{b}$ is given by

(45) $$\hat{\underline{\delta}} = P'\underline{b} \quad \text{or} \quad \underline{b} = P\hat{\delta}.$$

Combining (44) and (45) and recalling that the elements of D are ranked by the relative values of the characteristic roots, it is evident that the first row of P', say, $\underline{p}_1'$, gives the a linear combination of $\underline{b}$, which has the lowest variance in the class of all normalized linear combinations of $\underline{b}$. This result is due to Greenberg. This interpretation is extended by Fomby, Hill, and Johnson by showing that, if $\underline{\beta}$ in (42) 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 $\underline{\delta}$, $\left(\sigma^2 \sum_{i=1}^{k_i} \lambda_i^{-1}\right)$.

In applying this discussion to the forecasting problem, divide R into two submatrices,

$$R = (R_1, R_2) = X(P_1, P_2).$$

Consider the row vector $\underline{r}^* = (\underline{r}_1^*, r_2^*) = \underline{x}^{*\prime}\ (P_1, P_2)$ and rewrite (42) as

(46) $$y^* = \underline{r}_1^*\underline{\delta}_1 + \underline{r}_2^*\underline{\delta}_2 + u^*.$$

The unrestricted forecast of y is:

(47) $$\hat{y}^* = \underline{r}^*\hat{\delta}.$$

Note that

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

the unrestricted forecast obtained without the use of principal components. It is clear that, under some circumstances, a reduction in the components

used may be desirable. The interest in using principal components arises when some variables are eliminated or, more properly, additional restrictions are placed on the parameter space.

Consider the forecast based on the first k_1 principal components

$$\hat{y}^*_{p1} = \underline{r}^*_1 \hat{\delta}_1 \tag{48}$$

and the forecast error

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

The variance of e_{p1} conditional on R_1 is

$$\sigma^2_{fp} = \sigma^2_{p1} (1 + h_{p1}) \tag{49}$$

where

$$h_{p1} = \underline{r}^*_1 (R_1' R_1)^{-1} \underline{r}^{*1}_1 = \underline{r}^*_1 D_1^{-1} \underline{r}^{*\prime}_1, \tag{50}$$

and u^* is from the same distribution as $\underline{u}$. By equation (39),

$$R'R = \begin{Bmatrix} R_1' R_1 & 0 \\ 0 & R_2' R_2 \end{Bmatrix} = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$

and

$$\sigma^2_{p1} = \sigma^2 + \underline{\delta}_2' E(\underline{r}^{*\prime}_2 \underline{r}^*_2) \underline{\delta}_2 = \sigma^2 + \underline{\delta}_2' \sigma^2_{r2} \underline{\delta}_2 \tag{51}$$

and

$$\sigma^2_{r2} = E(\underline{r}^{*\prime}_2 \underline{r}^*_2) = E\{P_2' \underline{x}^{*\prime} \underline{x}^* P_2\} | X.$$

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

then $R_2'R_2/(n-1) = D_2/(n-1)$ is the estimator of σ^2_{r2}, the variance of $\underline{r}_2$. When $\underline{x}^{*\prime}$ is nonstochastic, then σ^2_{r2} is a parameter determined by the design of the experiment.

The restricted forecast has to be compared to the unrestricted forecast. From the basic result of the Appendix,

$$(52) \qquad \frac{\sigma^2_{fp}}{\sigma^2_f} \gtreqless 1 \iff \frac{\sigma^2_{p1}}{\sigma^2} \gtreqless \frac{1+h}{1+h_{p1}} = 1 + \frac{h - h_{p1}}{1+h_{p1}} .$$

On the surface it is not 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, $\sigma^2_{p1} - \sigma^2$ and, consequently, the ratio σ^2_{p1}/σ^2. At the same time it also minimizes in a probability sense, the value of h_{p1} in the same class. Thus, for any forecast based on k_1 explanatory variables, the aforementioned procedure obtains the one with the lowest forecast MSE.

4.2 *Implications for Parameter Variation Model*

The detailed evaluation of the choice of variables for purposes of forecasting is of special interest for the model advanced in Section 3. This model calls for increasing the number of the explanatory variables. A reasonable procedure would be to formulate the model according to *a priori* considerations. This should bring out the more important relationships expected to exist between the variables. The next step is empirical. For the purpose of forecasting, the dimension of the model may have to be reduced. It is suggested that the reduction be made by shifting to principal

component regressions, where components 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 components 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 price variables. More generally, one may want to have the statistical results in terms of the original variables, that is, those which appeared in the model prior to the modification due to the variable coefficient framework. This preference can be accommodated. First note that it is possible to obtain an estimate of $\underline{\beta}$ from the regression on the first k_1 principal components. In terms of the notation for equation (42)

$$(53) \qquad \underline{\hat{\beta}}_{p1} = P_1' \underline{\hat{\delta}}_1 \quad \text{and} \quad \text{var}\,(\underline{\hat{\beta}}_{p1}) = \sigma^2 P_1' D_1^{-1} P_1$$

where $\underline{\hat{\beta}}_{p1}$ is the estimate of $\underline{\beta}$ in (42) based on a regression of 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

$$(54) \qquad \underline{y} = X^* \underline{\beta}^* + X^{**} \underline{\beta}^{**} + \underline{u}.$$

Let

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

and rewrite the basic equation as

$$\underline{y} = X^* \underline{\beta}^* + V \underline{\beta}^{**} + \underline{u}$$

$$(55) \qquad = X^* \underline{\beta}^* + R \underline{\delta} + \underline{u}$$

where the relevant P matrix used to construct R is now the matrix of characteristic vectors of V'V and the remainder of the expression parallels

the previous discussion. Due to potential multicollinearity problems, the actual matrix X** likely would be composed of the interaction terms appearing in the estimable form (22). That is,

$$X^{**} = [X^{d}_{L}\ X^{d}Z];\ \underline{\beta}^{**} = \begin{bmatrix} \underline{\Pi} \\ \underline{\alpha} \end{bmatrix}. \tag{56}$$

5. *Additional Sample Information and Updating*

To operationalize the results of Sections 3.3, 3.4, and 4 as a continuous process, methods of handling additional sample data must be recognized. As is well known, this is an updating problem which has dealt largely with Kalman filtering procedures or related methods (Belsey). It is a process of weighting the information contained in new sample data relative to old information reflected by current parameter estimates. This process can be conducted in accordance with formal experimental design methods (MacRae). The "information value" of the additional observations and their associated costs of collection and summarization can be utilized to evaluate the experimental updating decision. This requires an evaluation of the trade-off between the information gained by the experiment and the costs of the experiment.

In the above context, the sequential experimental design of additional observations is concerned with the quality of information, evaluated via some optimality criterion. For most formulations, the quality of information is reflected by the inverse of the estimator's variance matrix, and the problem is one of generating optimal families of parameter estimators. If sampling control is possible, a number of optimal experimental designs can be advanced including a D-optimal design (Covey-Crump and Silvey; and Silvey); an E-optimal design (Kiefer); sequential optimal designs (Wynn); and Bayesian designs (Guttman). Unfortunately, as noted in Section 1, controlled experimentation in econometric modeling is the exception rather than the rule. Economic data are often passively generated and commonly posted by governmental agencies and/or private economic institutions.

For the case of passively generated observations, it can be shown that updating of the model presented in Section 3.3 or its principal component representation (55) will, in general, significantly improve "information

value" even though the new sample observations grossly replicate the previous sample.

Updating parameter estimates without additional sample observations is trivial. From the estimation section, we have

$$\begin{pmatrix} \hat{\underline{\beta}}_0 \\ \hat{\underline{\Pi}} \\ \hat{\underline{\alpha}} \end{pmatrix} = [X^{*\prime}\ \hat{\Omega}^{-1}\ X^*]^{-1} X^*\ \hat{\Omega}^{-1}\ \underline{y} \tag{57}$$

where

$$\hat{\Omega} = X^d\ \hat{\Delta}\ X^{d\prime} + \hat{\sigma}^2\ I_T. \tag{58}$$

The elements of $\hat{\Delta}$ and $\hat{\sigma}^2$ are obtained from the estimation of equation (29). Given (57), an updated estimate of $\underline{\beta}_{T+1}$ can be derived directly from (19). This latter estimate in conditioned upon α_{T+1} and z_{t+1}. Once it enters (17), given X_{t+1}, we obtain the updated value y_{T+1}. The covariance matrix of the updated estimate $\hat{\beta}_{T+1}$ is correspondingly

$$\begin{aligned} V_{T+1} = E\Big\{&[(\hat{\underline{\beta}}_0 - \underline{\beta}_0) + (I_k \otimes L_{T+1})\ (\hat{\underline{\Pi}} - \underline{\Pi}) + (I_k \otimes Z'_{T+1})\ (\hat{\underline{\alpha}} - \underline{\alpha})]\ [\hat{\underline{\beta}}_0 - \underline{\beta}_0) \\ &+ (I_k \otimes L_{T+1})\ (\hat{\underline{\Pi}} - \underline{\Pi})\ (I_k \otimes Z'_{T+1})\ (\hat{\underline{\alpha}} - \underline{\alpha})]'\Big\} + \Delta. \end{aligned} \tag{59}$$

The above results can be extended to the more operational circumstances in which additional sample observations are employed to improve the precision of the estimated parameters. To motivate this formulation, consider the conventional constant parameter framework, viz., $\underline{\beta}_t = \underline{\beta}$, $\forall_t$ and $e_t = 0$, $\forall_t$. For this framework, the standard Kalman filter result may be obtained by applying mixed estimation (Theil and Goldberger) to the combined sample data, i.e.,

$$\begin{pmatrix} \underline{y} \\ y_{T+1} \end{pmatrix} = \begin{pmatrix} X \\ \underline{x}'_{T+1} \end{pmatrix} \underline{\beta} + \begin{pmatrix} \underline{u} \\ u_{T+1} \end{pmatrix}. \tag{60}$$

This combined system (60) can be viewed as containing two separate stochastic restrictions on $\underline{\beta}$. In the updating context, one set of restrictions, viz.,

$\underline{y} = X\underline{\beta} + \underline{u}$, may be regarded as *a priori* information from which $\underline{\beta}_T$ is estimated.

Application of OLS to (60) leads to the estimator

$$\underline{\hat{\beta}}_{T+1} = (X'X + \underline{x}_{T+1}\,\underline{x}'_{T+1})^{-1}\,(X'\,\underline{y} + \underline{x}_{T+1}\,y_{T+1}), \tag{61}$$

where the covariances for $\underline{u}$ and u_{T+1} have been normalized. Applying the standard "matrix inversion lemma," the estimator (61) can be restated as[3]

$$\underline{\hat{\beta}}_{T+1} = \underline{\hat{\beta}}_T + \underline{K}_{T+1}\,(y_{T+1} - \underline{x}'_{T+1}\,\underline{\hat{\beta}}_T) \tag{62}$$

where

$$\underline{K}_{T+1} = (X'X)^{-1}\,\underline{x}_{T+1}\left[I + \underline{x}_{T+1}\,(X'X)^{-1}\,\underline{x}'_{T+1}\right]^{-1}. \tag{63}$$

This linear function (62) updates the previous estimate $\underline{\hat{\beta}}_T$ by applying the matrix of proportionality or filter matrix K_{T+1} to the prediction error resulting from using the "un-updated" estimator $\underline{\hat{\beta}}_T$. The matrix K_{T+1} filters this prediction error and plays the role of processing the predictor discrepancy.

The covariance matrix associated with $\underline{\hat{\beta}}_{T+1}$ is

$$V_{T+1} = (X'X + \underline{x}_{T+1}\,\underline{x}'_{T+1})^{-1}. \tag{64}$$

Again applying the "matrix inversion lemma," the expression for the covariance matrix can be restated as

$$V_{T+1} = V_T - A\,\underline{K}'_{T+1}\,\underline{K}_{T+1}, \tag{65}$$

where

$$A = I + x_{T+1}\,(X'X)^{-1}\,x'_{T+1}. \tag{66}$$

Note that $A\underline{K}'_{T+1}\underline{K}_{T+1}$ is at least a positive semidefinite matrix, implying that $\hat{\underline{\beta}}_{T+1}$ has at least as small a variance as $\hat{\underline{\beta}}_T$.

The above results are equivalent to the optimal filter derived by Kalman; and for the classical linear model, $\hat{\underline{\beta}}_{T+1}$ is minimum variance, unbiased, and linear. They can be easily generalized to any number of updating periods; i.e., if outcome for period T + s is of interest, then

$$\hat{\underline{\beta}}_{T+s} = \left(X'X + \sum_{j=1}^{s} \underline{x}_j\underline{x}'_j\right)^{-1}\left(X'\underline{y} + \sum_{i=1}^{s} \underline{x}_j y_j\right) \tag{67}$$

and

$$V_{T+1} = \left(X'X + \sum_{j=1}^{s} \underline{x}_j\underline{x}'_j\right)^{-1}. \tag{68}$$

Updating in the context of random parameter variation presumed by the specifications in equations (17), (19), and the estimable form (22) is considerably more complicated. Not only must the prior information on parameter evolution (17) be taken into account but it must be accommodated as in the GLS framework due to nonscaler variance matrix Ω. The actual updating takes place in terms of the estimable or reparameterized form (22); and since Ω is unknown, only asymptotic properties of the resulting updated estimator can be derived.

To proceed in the context of generalized updating, rewrite the estimable equation (22) as

$$\underline{y} = X\,\underline{\beta}_{0T} + W\underline{\gamma}_T + [X^d\underline{e} + \underline{u}] \tag{69}$$

where $W = [X^d L, X^d Z]$ and $\underline{\gamma}' = [\underline{\Pi}', \underline{\alpha}']$; W is T x k(k + m); and $\underline{\gamma}$ is k(k + m) x 1. This compact representation simplifies the subsequent derivations and provides a tractable formulation for developing implications. Note that (69) could also refer to the principal component formulation where $X = X^*$, $\underline{\beta}_0 = \underline{\beta}^*$, $W = V$, and $\underline{\gamma} = \underline{\beta}^{**}$.

The additional sample data, $\underline{y}_{T+1}$, X_{T+1}, W_{T+1} can be combined with (69) to form the extended, condensed model

(70) $$\underset{\sim}{y}_{\star} = D_{\star}\ \underset{\sim}{\delta}_{\star} + \underset{\sim}{\xi}_{\star}$$

where

$$\underset{\sim}{y}_{\star} = \begin{pmatrix} \underset{\sim}{y}_T \\ \underset{\sim}{y}_{T+1} \end{pmatrix} \quad D_{\star} = \begin{bmatrix} X_T & W_T \\ X_{T+1} & W_{T+1} \end{bmatrix} = \begin{bmatrix} D_T \\ D_{T+1} \end{bmatrix}; \ \underset{\sim}{\delta}_{\star} = \begin{bmatrix} \underset{\sim}{\beta}_0 \\ \underset{\sim}{\gamma} \end{bmatrix} \equiv \underset{\sim}{\delta}_{T+1}$$

$$\underset{\sim}{\xi}_{\star} = \begin{bmatrix} \underset{\sim}{\xi} \\ \underset{\sim}{\xi}_{T+1} \end{bmatrix} = \begin{bmatrix} X_T^d\ \underset{\sim}{e}_T + \underset{\sim}{u}_T \\ X_{T+1}^d\ \underset{\sim}{e}_{T+1} + \underset{\sim}{u}_{T+1} \end{bmatrix}; \ \underset{\sim}{\xi}_{\star} \sim (0, \Omega_{\star}) \quad \text{and } \Omega_{\star} = \begin{bmatrix} \Omega_T & \Omega_{T,T+1} \\ \Omega'_{T+1,T} & \Omega_{T+1} \end{bmatrix}$$

i.e., $\Omega_{\star}$ is symmetric and its partitions refer to the variance matrix of $\underset{\sim}{\xi}$, Ω; the variance matrix for $\underset{\sim}{\xi}_{T+1}$; and the covariance matrix between $\underset{\sim}{\xi}$ and $\underset{\sim}{\xi}_{T+1}$. Applying generalized least squares to above formulations, we obtain

(71) $$\hat{\underset{\sim}{\delta}}_{\star} = (D'_{\star}\ \Omega_{\star}^{-1}\ D_{\star})^{-1}\ D'_{\star}\ \Omega_{\star}^{-1}\ \underset{\sim}{y}_{\star}.$$

The estimator $\hat{\underset{\sim}{\delta}}_{\star}$ can be restated as

(72) $$\hat{\underset{\sim}{\delta}}_{T+1} = \hat{\underset{\sim}{\delta}}_T + K(D_{\star}, \Omega_{\star})\ (\underset{\sim}{y}^{+}_{T+1} - \hat{\underset{\sim}{y}}^{+}_{T+1/T})$$

where $K(\cdot)$ is the appropriate form for the Kalman filter, here conditioned on $D_{\star}$ and $\Omega_{\star}$, and letting $\hat{y}_{T+1/T} = \underset{\sim}{D}_{T+1}\ \hat{\underset{\sim}{\delta}}_T$,

(73a) $$\hat{y}^{+}_{T+1/T} = \Omega'_{T+1,T}\ \Omega^{-1}\ \hat{y}_{T/T} - \hat{y}_{T+1/T}.$$

Thus, the last term in equation (72) can be written,

(73b)
$$y^{+}_{T+1} - \hat{y}^{+}_{T+1/T} = \Omega'_{T+1,T}\ \Omega_T^{-1}\ y_T - \hat{y}_{T+1} - \left(\Omega'_{T+1,T}\ \Omega_T^{-1}\ \hat{y}_{T/T} - \hat{y}_{T+1/T}\right)$$
$$= \Omega'_{T+1,T}\ \Omega^{-1}\ (y_T - \hat{y}_{T/T}) - (y_{T+1} - \hat{y}_{T+1/T}).$$

Clearly, if the covariance between $\underset{\sim}{\xi}$ and $\underset{\sim}{\xi}_{T+1}$ is zero, the matrix $\Omega'_{T+1,T}$ vanishes and expression (73b) reduces to the conventional Kalman error $y_{T+1} - \hat{y}_{T+1/T}$. If the matrix $\Omega_{T+1,T}$ fails to vanish, then the filtering

process must take into account past errors $(y_T - \hat{y}_{T/T})$, appropriately weighted (by $\Omega'_{T+1,T}\ \Omega_T^{-1}$) as well as current errors $(y_{T+1} - \hat{y}_{T+1/T})$. This simply reflects the correlation that might exist between ξ and ξ_{T+1}.

The modified Kalman filter, $K(D_*, \Omega_*)$ is a complex expression, viz.,

$$K(D_*, \Omega_*) = M_D\ D^+{}'\ [Q^{-1} - Q^{-1}\ D^+\ (D_*'\ \Omega_*^{-1}\ D_*)^{-1}\ D^+\ Q^{-1}], \tag{74}$$

$$= M_D\ D^+{}'\ G_D^{-1}$$

where

$$M_D = \left(D_T'\ \Omega_T^{-1}\ D_T\right)^{-1}, \tag{75}$$

$$D^+ = \Omega'_{T+1,T}\ \Omega_T^{-1}\ D_T - D_{T+1} \tag{76}$$

$$Q = \Omega_{T+1} - \Omega'_{T+1,T}\ \Omega_T^{-1}\ \Omega_{T+1,T}, \tag{77}$$

and

$$G_D^{-1} = Q^{-1} - Q^{-1}\ D^+\ (D_*'\ \Omega_*^{-1}\ D_*)^{-1}\ D^+\ Q^{-1}. \tag{78}$$

Expression (74) can be decomposed and compared to the constant parameters updating case by noting that the noncanonical (nonidentity covariance matrix Ω) Kalman filter under the exclusion of W can be represented as

$$K(X_*, \Omega_*) = M_X X^+{}'\ [Q^{-1} - Q^{-1}\ X^+\ (X_*'\ \Omega_*^{-1}\ X_*)^{-1}\ X^+\ Q^{-1}] = M_X\ X^+{}'\ G_X^{-1} \tag{79}$$

where

$$M_X = (X_T'\ \Omega_T^{-1}\ X_T)^{-1}\ , \tag{80}$$

$$X^+ = \Omega'_{T+1,T}\ \Omega_T^{-1}\ X_T - X_{T+1}, \tag{81}$$

$$X_* = \begin{bmatrix} X_T \\ X_{T+1} \end{bmatrix} \tag{82}$$

$$G_X^{-1} = Q^{-1} - Q^{-1}\ X^+\ (X_X'\ \Omega_*^{-1}\ X_*)^{-1}\ X^+\ Q^{-1}. \tag{83}$$

To compare (79) and (74) and the resulting implications for the updated parameter vector (72), we must decompose (74). Treating first M_D, after some simplication, the partitioned inverse in (75) can be represented as

$$M_D = \begin{bmatrix} (X_T' \Omega_T^{-1} X_T)^{-1} + \tilde{\psi}_T \tilde{F}_T^{-1} \tilde{\psi}_T' & -\tilde{\psi}_T \tilde{F}_T^{-1} \\ -\tilde{F}_T^{-1} \tilde{\psi}_T' & \tilde{F}_T^{-1} \end{bmatrix} \tag{84}$$

where

$$\tilde{\psi}_T = (X_T' \Omega_T^{-1} X_T)^{-1} X_T' \Omega_T^{-1} W_T \tag{85}$$

$$\tilde{F}_T = W_T' P_T' \tilde{M}_T P_T W_T = \tilde{W}_T' \tilde{M}_T \tilde{W}_T, \tag{86}$$

and

$$\tilde{M}_T = I - P_T X_T (X_T' \Omega_T^{-1} X_T)^{-1} X_T' P_T'. \tag{87}$$

Note that P_T is such that $\Omega_T^{-1} = P_T' P_T$. The expressions (85) through (87) can each be given a specific interpretation; $\tilde{\psi}_T$ is the generalized least squares estimator for the auxiliary regression of W_T on X_T; $\tilde{F}_T$ is the residual sum of squares for the auxiliary regression of W_T on X_T; and $\tilde{M}_T$ is the projection on the space ortho complementary to the space span by the columns of $\tilde{X}_1 = P_1 X_1$.

Given the expression for D* and its partition in terms of X and W, the product matrix $M_D D^{+\prime}$ can be written after substantial additional simplification as

$$M_D D^{+\prime} = \begin{bmatrix} M_X X^{+\prime} - \tilde{\psi}_T \tilde{F}_T^{-1} (W^+ - \hat{W})^{+\prime} \\ \tilde{F}_T^{-1} (W^+ - \hat{W})^{+\prime} \end{bmatrix} \tag{88}$$

where

$$W^+ - \hat{W}^+ = \left[\Omega_{T+1,T} \Omega_T^{-1} W_T - W_{T+1}\right] - \left[\Omega_{T+1,T}' \Omega_T^{-1} \hat{W}_T - \hat{W}_{T+1}\right] \tag{89a}$$

and

(89b) $$\hat{W}_T = \tilde{\psi}_T X_T \text{ and } \hat{W}_{T+1} = \tilde{\psi}_T X_{T+1}.$$

The balance of the expression for $K(D_*, \Omega_*)$ in (74), viz., G_D^{-1}, can be stated after much simplification as

(90) $$G_D^{-1} = G_X^{-1} - Q^{-1} (W^+ - \hat{W}_*^+) \tilde{F}_*^{-1} (W^+ - \hat{W}_*^+)' Q^{-1}$$

where

(91) $$\hat{W}_* = \begin{bmatrix} \hat{W}_{*T} \\ \hat{W}_{*T+1} \end{bmatrix} = \begin{bmatrix} \tilde{\psi}_* X_t \\ \tilde{\psi}_* X_{T+1} \end{bmatrix}.$$

Armed with expressions (88) and (90), the modified Kalman filter $K(D_*, \Omega_*)$ can be stated in terms of $K(X_X, \Omega_*)$, i.e.,

(92) $$K(D_*, \Omega_*) = \begin{bmatrix} K(X_*, \Omega_*) - \tilde{\psi}_T H(D_*, \Omega_*) \\ \\ H(D_*, \Omega_*) \end{bmatrix}$$

where

(93) $$H(D_*, \Omega_*) = \tilde{F}_T^{-1} (W^+ - \hat{W}^+)' [G_X^{-1} - Q^{-1}(W^+ - \hat{W}_*^+) \tilde{F}_*^{-1} (W^+ - \hat{W}_*^+)' Q^{-1}].$$

Returning to expression (72), the first component of (92) updates $\underline{\beta}_{oT}$ and the second updates $\underline{\gamma}_T$. Hence, the inclusion of the interaction terms represented in W requires an adjustment in the conventional noncanonical Kalman filter, $K(X_*, \Omega_*)$ which is reflected in the term $\tilde{\psi}_T H(D_*, \Omega_*)$. Note that if $\tilde{\psi}_T = 0$, i.e., W_T and X_T and W_{T+1} and X_{T+1} are orthogonal, the second term in the first component of equation (92) vanishes, and filter $K(D_*, \Omega_*)$ reduces to the conventional noncanonical Kalman filter in updating $\underline{\beta}_{oT}$.

The above results have implications for special cases of the generalized parameter variation structure. Referring to equation (19), observe that the

first component of (92) adjusts the constant term $\underline{\beta}_0$. This is clear since the specification of the model for $\underline{\alpha}$ and $\underline{\Pi}$ equal zero results in the Kalman filter of the form (72). The second component of (92) adjusts $\underline{\gamma}$ or $\underline{\Pi}$ and $\underline{\alpha}$. Both feed to update $\underline{\beta}_t$ by expression (19).

First consider the parameters $\underline{\alpha}$ corresponding to variables which themselves are updated in the period T + 1, the extraneous variables, $\underline{z}'_{t+1}$. In particular, suppose that the variance for $\underline{\alpha}$ goes to zero. Then the parameter structure $\underline{\beta}_t$ is deterministically updated, conditioned on $\underline{z}_{T+1}$. In this situation, with or without the additional sample data and the implications for $\underline{\gamma}$ and $\underline{\beta}_0$, the parameter vector $\underline{\delta}^*_T$ would differ from $\underline{\delta}^*_{T+1}$. Updating would occur based on the sample estimations from the data through T but the values of the parameter conditioning variables through T + 1. With $\underline{\alpha}$ known or var $\hat{\underline{\alpha}}$ approximating zero, the additional sample information could only affect $\underline{\beta}_t$ by impacting on $\underline{\beta}_0$ and $\underline{\Pi}$. The structure in this case is

$$\hat{\beta}_{T+1} - [I_K \otimes \underline{z}'_{T+1}]\,\underline{\alpha} = \hat{\beta}_{0T+1} + [I_K \otimes \underline{L}'_{T+1}]\,\hat{\underline{\Pi}}_{T+1}. \tag{94}$$

Thus, the updated versions of the parameters $\underline{\beta}_{0T+1}$ and $\underline{\Pi}_{T+1}$ would not be conditioned upon variables extraneous to the model. The parameter vector β_{T+1}, of course, may or may not change depending upon the extraneous values entering the right-hand side of equation (94).

If the parameter vector $\underline{\Pi}$ is also known with probability one, the structure for the varying parameters can be used to transform the model. The most trivial situation occurs when $\underline{\Pi} = \underline{\alpha} = \underline{0}$. In any event, with known $\underline{\alpha}$ and $\underline{\Pi}$, the Kalman filter, except for the random component of the parameter variation hypothesis, reduces to a conventional form. Abstracting from the random coefficients, the only additional information for updating the statistical model is contained in the new sample data.

Finally, the Kalman filter (92) can be specialized for the situation in which Ω is diagonal. This occurs when the *a priori* structure on the disturbances does not admit an intertemporal relationship, viz., the disturbances are assumed independent. With this assumption, $y^+_{T+1} = y_{T+1}$ and the development associated with equation (73b) can be omitted. Intuitively, the simplification can be made when the error structure for the

model contains no information about future values of y_t.[4] Again, these implications, as those for the parameter vectors $\underline{\alpha}$ and $\underline{\Pi}$, can be simplified if the coefficients are deterministically generated, i.e., $\underline{e}_t = \underline{0}$.

Thus, the various special cases illustrate that the generalizations to randomly varying coefficients and to disturbance structures involving intertemporal information can be handled by applying a filtering approach. The general representation has been addressed here because of its useful implications for hypothesis testing. The extensions admit structures which are nested within the general representation. Thus, likelihood ratio and commonly available MSE norm tests can be employed to examine the updated varying parameter structures and error assumptions for consistency with the sample data.

Finally, it should be noted that the results for the special cases show how extraneous information, say, in the form of commodity specialist judgments, can be used within the general framework to update the parameters. For example, external calibration of the forecasts by commodity specialists can be incorporated by modifying the results in the case for which $\underline{\alpha}$ was taken as known with probability one. In this situation the commodity specialists would have different information bases, represented by $\underline{z}_t$, and modify the parameters accordingly. If this information about $\underline{\alpha}$ were uncertain, if would be, of course, more efficient and appropriate to incorporate it in a mixed estimation framework.

A second and more likely possibility for calibrating the forecasts is the use of an "and factor" or "constant adjustment." The rationalization in this case is associated with the disturbance structure. In effect, such calibrations amount to assuming that $\Omega_{T+1,T}$ is known so that a constant relationship exists between $\hat{y}^{+}_{T+1/T}$ and $\hat{y}_{T+1/T}$. Hence, the representation developed provides a structure for justifying this common practice. As observed for the modifications associated with the varying parameters, if there is uncertainty about the nature of adjustments, the general structure can be integrated into the model and the parameters efficiently estimated using available extraneous and sample information.

6. *Simultaneous Equations*

In this section the basic formulation is adapted to deal with the simultaneous equation models. In part, this adaptation is simply an

extension of the size of the system resulting in an additional complexity of expressions and perhaps of computations. However, the introduction of the time-varying parameters to simultaneous equations introduces several conceptual problems which should be dealt with explicitly. This is particularly the case for identifiability of the system and the choice of instruments.

6.1 *Simultaneous Specification*

Consider the system of G simultaneous equations represented by

$$Y\Gamma + X\Theta = U \tag{95}$$

where Y is a T x G matrix of T observations on the G endogenous variables; Γ is a conformable G x G parameter matrix; X is the T x K matrix of sample values for the predetermined variables; Θ is the corresponding K x G parameter matrix; and, finally, U is a matrix formed by the T disturbance vectors for the G equations. Conventionally, it is assumed that the row vectors, corresponding to the values for the G disturbances associated with a particular observation, have expectation $\underline{0}$ and are only contemporaneously correlated. That is,

$$\left[E(\underline{u}_t') = \underline{0}' \text{ and } E(\underline{u}_t \, \underline{u}_t') = \begin{cases} \Sigma; & t = t' \\ 0; & t \neq t' \end{cases} \right]$$

where $\underline{u}_t'$ is the tth row vector from U. As well, assume that X is of rank K with probability one and Γ is nonsingular.

To conveniently develop the correspondence with the variational parameter specifications obtained in the previous sections, an alternative representation of the system is introduced. Let the prior information on the structural parameters be confined to exclusion restrictions. If these exclusion restrictions are imposed, the gth equation in the system can be expressed as

$$\underline{y}_g = Y_g \underline{\gamma}_g + X_g \underline{\theta}_g + \underline{u}_g \tag{96}$$

where $\underline{y}_g$ is the vector of T observations on the endogenous variable selected for normalization, Y_g and X_g are matrices of T observations on the included endogenous (less the one selected for normalization) and predetermined variables, $\underline{\gamma}_g$ and $\underline{\theta}_{1-g}$ are the appropriate parameter vectors, and $\underline{u}_g$ is the T x 1 vector of structural disturbances. Additionally, define the matrix

$$S_g = (Y_g, X_g)$$

and the vector

$$\underline{\beta}_g' = (\gamma_g, \theta_g).$$

Then the gth equation can be equivalently rewritten as

(97) $$\underline{y}_g = S_g \underline{\beta}_g + \underline{u}_g.$$

Based on expression (97), the G equations in the system are alternatively expressed (see Schmidt) in the form

(98) $$\underline{y} = S\underline{\beta} + \underline{u}$$

where the following stacking and partitioning conventions have been employed

$$\underline{y}' = (\underline{y}_1', \underline{y}_2', \ldots, \underline{y}_g', \ldots, \underline{y}_G')$$

$$\underline{u}' = (\underline{u}_1', \underline{u}_2', \ldots, \underline{u}_g', \ldots, \underline{u}_G')$$

$$\underline{\beta} = (\underline{\beta}_1', \underline{\beta}_2', \ldots, \underline{\beta}_g', \ldots, \underline{\beta}_G')$$

$$S = \begin{pmatrix} S_1 & 0 & 0, \ldots, 0 \\ 0 & S_2 & 0, \ldots, 0 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 0, \ldots, 0 & S_G \end{pmatrix}.$$

The previous assumptions on the structural disturbances imply

$$E(\underline{u}) = \underline{0} \text{ and } E(\underline{u}\ u') = \Sigma \otimes I.$$

Armed with the simultaneous equation system expressed in equation (97) and (98), the generalization to the variational parameter formulation is straightforward. In fact, except for identification, the choice of instruments and the more complex error structure, the results of Section 3.3 are directly applicable. To establish this claim, first simply subscript the vector $\underline{\beta}$. Corresponding to equation (17), the variational parameter specification for the gth equation is

$$y_{gt} = \underline{S}'_{gt}\ \underline{\beta}_{gt} + u_{gt}. \tag{99}$$

Similarly, for the sample and using the conventions for the parameters in equation (18), we have

$$\underline{y}_g = S_g^d\ \underline{\beta}_g^* + \underline{u}_g \tag{100}$$

where S_g^d is defined correspondingly to X^d and $\underline{\beta}_g^*$ is the T x (Txg*) vector of parameter values implied by the variational structure on the g* included, right-hand side endogenous and predetermined variables.

Explicitly imposing the parameter structure from equations (19) and (20), the expression for the expanded representation of the structural equations is

$$\underline{y}_g = S_g\ \underline{\beta}_{go} + S_g^d\ L_g\ \underline{\Pi}_g + S_g^d\ Z_g\ \underline{d}_g + [S_g^d\ \underline{e}_g + \underline{u}_g]. \tag{101}$$

Note that the operator, L; variables external to the system, Z; and the parameter vectors identifying the systematic variation in $\underline{\beta}_{gt}$ have been defined specific to the gth structural equation. Equation (101) can be expressed similarly to equation (97). In this form the questions regarding identification and instruments can be addressed. In particular, let

$$S_g^+ = \left[S_g,\ S_g^d\ L_g,\ S_g^d\ Z_g\right]$$

and

$$\underline{\beta}_g^{+\prime} = (\underline{\beta}_{go}^{\prime}, \underline{\Pi}_g^{\prime}, \underline{\alpha}_g^{\prime})$$

so that the gth equation with the structure for the variational parameters imposed is

$$\underline{y}_g = S_g^{+} \underline{\beta}_g^{+} + \underline{u}_g^{+} \tag{102}$$

where $\underline{u}_g^{+}$ is just the bracketed composit disturbance vector from equation (101). The dimension of the parameter vector is now $g^* + r_g + j_g$, with r_g determined by the lengths of the various lags imposed by L_g for each of the included endogenous and predetermined variables and j_g by the number of external or extraneous variables. Clearly, for $\underline{\Pi}_g = 0$, $\underline{\alpha}_g = 0$ and $\underline{e}_g = 0$, the system is identical to (97), the constant parameter case.

6.2 *Identification*

The nature of the identification problem introduced by the varying parameter structure is suggested by the expanded number of variables in S_g^{+}.[5] The variational parameters have the potential for completely changing the identification status of the gth equation. To illustrate, let g_γ denote the number of right-hand side endogenous variables included in constant parameters version (97) of equation g and g_θ denote the number of included exogenous variables, i.e., $g^* = g_\gamma + g_\theta$. The lag operator L_g can only expand the number of predetermined variables in the expression. Thus, according to the specification (102), there are r_g additional predetermined variables in the system.

The source of identifiability complication is the extraneous variables Z_g and the parameters that must be estimated on the "new" endogenous variables that they create. These variables are, of course, products of the extraneous variables and the endogenous variables. If we assume S_g^d is partitioned as $\left[Y_g^d, X_g^d\right]$, then the equations identifying these new endogenous variables are

$$Y_g^d Z_g = Y_g^* \tag{103}$$

where S_g^d and Z_g have been defined so that the first j_{g1} variables corresponding to the parameter vector $\underline{\alpha}_g$ are products of the extraneous variables with the included endogenous variables and the remaining j_{g2} are products with the included exogenous variables. Then the condition for the gth reparameterized equation to be identifiable is

$$g_\theta + r_g + j_{g2} \geq g_\gamma + j_{g1}. \tag{104}$$

This, of course, assumes that the rank of the expanded set of predetermined variables is equal to $g_\theta + r_g + j_{g2}$ and that the nonlinear identities associated with the j_{g2} equations (103) or defining Y_g^* have been added to the simultaneous system.

These j_{g2} equations can be added to the system without destroying the basic linearity by using a first-order Taylor's series expansion of expression (103). The expansion should be taken at the means or a reference value used for forecasting and only the linear term retained. Specifically, the nonlinear identities can be approximated by the j_{g2} behavioral equations

$$Y_g^* = \theta_{oj_{g1}} + Z_g \theta_{1j_{g2}} + \eta \tag{105}$$

where $\theta_{oj_{g1}}$ and $\theta_{1j_{g2}}$ are appropriately dimensioned diagonal parameter matrices and η is the matrix of disturbances created by the approximation. Thus, the identification problem is simple in principle. If the parameter structures are restricted so additional endogenous variables or nonlinear identities are not created, then equations which were identifiable without the information on the variational parameter structure will remain identifiable. Moreover, under these conditions, equations not identifiable in the absence of the information on variational parameters may become identifiable. These observations hold since, instead of using the original predetermined variables for estimation, their products with the variables conditioning the parameters can now be used. Thus, the number of predetermined variables is increased, in this case by $r_g + j_{g2}$. This is the reason an equation not originally identifiable may become identifiable. Basically, identifiability occurs due to the prior information on the variability of the coefficients, a factor which, of course, was not present for the constant coefficient specification.

For varying parameter structures which create additional endogenous variables through products with the existing endogenous variables, it is best to expand the simultaneous equations model. That is, treat the products with existing endogenous variables as "new" endogenous variables and provide the commensurate linear approximations for the identities (105) as the equations required to complete the expanded system.

6.3 *Instruments and Estimation*

The full system with the exclusion restrictions imposed has

$$\sum_{g=1}^{G} (g^* + r_g + j_g) = M_1 \tag{106}$$

parameters associated with the equations plus the parameters from the auxiliary equations used to approximate the induced nonlinearities. Since the auxiliary equations include a constant term plus the appropriate variables from Z_g, they each have two parameters. As shown in equation (105), there are in fact

$$\sum_{g=1}^{G} 2 \cdot j_{g1} = M_2$$

such parameters to estimate from these j_{g1} auxiliary equations.

Since the additional j_{g1} equations are in reduced form, they can be estimated by OLS to produce suitable instruments for the associated and endogenous variables, Y_g^*. These j_{g1} variables are then introduced directly into the structural system in the form (102). The resulting equations are linear in the parameters and include a number of endogenous variables equal to G. The system treated in this way can be directly solved for the unrestricted reduced form parameters. OLS applied to these reduced form equations provides a set of variables which satisfy the conditions appropriate for instruments to be applied in a two-stage estimation process. In fact, the estimators may be characterized as modified two-stage least squares. The modification in this case involves the treatment of the auxiliary equations.

If the suggested approach proves impractical (perhaps because of the approximations), a variation on iterated instrumental variables can be employed (Brundy and Jorgenson). This method would estimate the structure by OLS directly. The resulting parameter estimates would be used to solve for the reduced form. This time the auxiliary equations could be added in linear or nonlinear form. In the latter case, they would be substituted directly into the structural equations to calculate the instruments. Regardless of how the auxiliary relationships are treated, the first-round solutions using the OLS estimators (for the structural parameters to obtain the restricted reduced form) can be applied to generate instruments which assure consistent estimators for the structural parameters.

The choice between the two suggested approaches depends upon small sample efficiency, a matter about which we have little information; especially, in such potentially complex systems. From a pragmatic viewpoint, the iterated procedure is recommended when the linearized auxiliary equations do not fit well with the sample data.

Given the instruments for the structural equations, limited information or three-stage-least-square type estimators for the parameters in the system can be easily obtained. The heteroscedastic structure, due to the random components for the varying parameter structure, is estimated as indicated in Section 3.3 but with the instruments substituted into expression (102). The resulting efficient estimators for the parameters of the G structural equations can then be used to calculate the residuals employed to estimate Σ. With the estimate of Σ, the full and reparameterized system of G equations can be simultaneously estimated.

6.4 *Updating and New Sample Information*

The results from Section 5 can be extended to simultaneous specification of the present section. This should be clear from comparisons of equations (101) and (69). The simultaneous system as represented in stacked form shows that, aside from complexities associated with the contemporaneous across-equation error structure (which is of consequence for full-information estimation, identification, and choice of instruments), the extension is straightforward. The only aspect of the extension not explicitly covered in Section 5 involves the updated instruments. But by the arguments of Section 6.3, these updated instruments can be calculated using parameters

estimated from the existing sample data. New sample data implications for the instruments would likely be minimal but could be incorporated by iterating the updating process using "corrected" instruments calculated using both new and existing sample data.

7. *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 rationality and the theoretical justification for parameter variation formulation, the overwhelming evidence is in favor of such formulations. The model-use argument is often neglected; but for those 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 extraneous variables on the systematic movements of the parameters for the 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.

Estimation takes place in the context of a form which includes additive, interactive, and random effects. The estimator, as constructed for the general case, conforms to the two-stage Aitken method for unknown variance. The first round involves application of the OLS to the estimable equation. Residuals estimated using the OLS estimators for the structural parameters are then used to estimate the induced heteroscedastic structure of the error terms in the estimable equations. The second round reestimates the additive and interaction effects taking into account the heteroscedastic structure.

As with all two-stage Aitken or feasible GLS procedures of this sort, the nonnegativity property of the variances may be violated. Procedures involving the application of inequality estimators or ridge regressions can be employed to force the conformity of the empirical results with this property.

The specification and estimation results for the single-equation model have been generalized to simultaneous equation systems. In this setting, an interesting implication of the general specification is that the identifiability of the complete system is often enhanced. The formulation advanced can call for augmenting the explanatory variables with the products of these variables and those determining systematic components for the varying parameters. In essence, the *a priori* information for the structure is expanded by the hypotheses on parameter variation, making possible the identifiability of equations for which the original prior information was not sufficient. Hence, identification can be enhanced through prior information on the variability of the coefficients themselves. 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 for 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 rules for reducing the dimension of the estimation problem have been developed which recognize the tradeoff between increases in sampling variance and declines 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 rule is developed for selecting explanatory variables for forecasting purposes. This rule 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 involve 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 applied as a way of retaining as much of the limited sample information as possible while reducing the dimension of the estimation problem. Conventional selection rules for the principal components are not appropriate. That is, the procedure of selecting principal components according to the size of their 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.

All of the results for varying parameters can be as well viewed in an updating context. Accordingly, general expressions have been developed for updating the varying parameters models based on additional sample information. The results extend the Kalman filtering technique to the varying parameter models in single and simultaneous equation contexts. Moreover, the filter is generalized to accommodate the situation in which there is a persistence hypothesis for the structural disturbances. This general form for the updating problem is necessary for using empirically varying parameter structures since typically more parameters are introduced. In such specifications, sample information is of high potential value, especially when extending standard hypothesis testing to evaluate alternative parameter structures for conformity to the sample data.

The integration of the results obtained imply and approach to model development which properly balances the cost of the complexity and the value of simplicity. The steps of the procedures for applying the integrated approach are: (1) form the estimable equation, derive principal components of the interaction terms according to the specified selection criteria, and apply OLS; (2) estimate the covariance matrix of the heteroscedastic structure for the model as specified in estimable form; (3) combine the estimated covariance matrix to generate GLS estimates; and (4) update the estimations using a filtering process which recognizes the varying parameter structure and the possibility of persistence in errors. As shown, the approach can be generalized to deal with simultaneous equations by deriving appropriate instrumental variables for the explanatory endogenous variables and verifying the identifiability of the system prior to implementing the suggested procedure.

Appendix

Selection of Explanatory Variables for Forecasting

The selection of explanatory variables for forecasting purposes is a general problem and can be treated as such. It is assumed that the true equation contains k explanatory variables. This set of variables is suggested by a theory underlying the equation. Using the sample data, the equation is estimated and the null (or other) hypotheses 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 is

$$\hat{y}^* = \underline{x}^{*\prime}\,\hat{\underline{\beta}}; \tag{A.1}$$

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

$$\sigma_f^2 = \sigma^2\,(1 + h) \tag{A.2}$$

where

$$h = \underline{x}^{*\prime}\,(X'X)^{-1}\,\underline{x}^*. \tag{A.3}$$

The question is: What X's should be included in the predicting equation? As usual, the answer depends on the prediction criterion. The problem is to determine within this framework the smallest conditional MSE.

For reasons indicated in Section 4.1, it will be advantageous to consider this problem in the transformed parameter space $\underline{\delta} = P'\underline{\beta}$. Thus, the problem of specific interest is

$$\hat{\underline{y}}^* = \underline{r}^{*\prime}\,\hat{\underline{\delta}} = \underline{x}^{*\prime}\,P'P\hat{\underline{\beta}} \tag{A.4}$$

with forecast variance

$$e_f = e_{fp} = \sigma_p^2\,(1 + h_p) \tag{A.5}$$

where e_{fp} and h_p are defined analogous to equations (49) and (50). Producing the forecast results for this case permits taking advantage of the orthogonality properties of the principal components and the variance-minimizing result previously described.

To deal with this problem, R (= XP) is partitioned into R = (R_1, R_2) of order (n x k_1, n x k_2), $k_1 + k_2 = k$; and $\underline{r}^{*\prime}$ is partitioned correspondingly, $\underline{r}^{*\prime} = (\underline{r}_1^{*\prime}, \underline{r}_2^{*\prime})$. The question is whether to use the predictor in (A.1) [or identically (A.4)] or, alternatively, to use a predictor which is based only on R_1, i.e.,

(A.6) $$\tilde{y}^* = \underline{r}_1^{*\prime}\, \hat{\underline{\delta}}_1.$$

The error for the predictor (A.6) is f_1 (= $\tilde{y}^* - y$). The variance of the predictor $\tilde{y}^*$, conditional on R_1 is,

(A.7) $$\sigma^2_{fp1} = \sigma^2_{p1}(1 + h_{p1})$$

where

(A.8) $$h_{p1} = \underline{r}_1^* D_1^{-1} \underline{r}_1^{*\prime}.$$

The variance $\sigma^2_{p_1}$ is equal to $\sigma^2 + \underline{\delta}_2^1 \sigma^2_{r_2} \underline{\delta}_2$ as indicated in equation (51). Note that $\sigma^2_{p_1}$ is the conditional variance of y given R_1. At the same time, it can be viewed as the MSE of $\underline{y}|R_1$ given R.

The difference between $\sigma^2_{p_1}$ and σ^2 is the square of the bias generated by evaluating $E(y \mid R)$ from $E(y \mid R_1)$. The difference can be estimated by using $\hat{\sigma}^2 = 1/(n - k)\ \underline{y}'M\underline{y}$ and $\hat{\sigma}^2_{p1} = 1/(n - k_1)\ \underline{y}'M_1\underline{y}$ where $M = I - K$, $M_1 = I - K_1$, $K = X'(X'X)^{-1} X' = R(R'R)^{-1} R'$ and $K_1 = R_1(R_1' R_1)^{-1} R_1$.

Let F be the statistic used for testing the null hypothesis of $\underline{\delta}_2 = 0$. By a simple transformation, it can be shown that $\hat{\sigma}_{p1}/\hat{\sigma}^2 = 1 + (F - 1) \cdot$

$k_2/(n - k_1)$. Consequently, $\sigma_{p1} \gtreqless \hat{\sigma}^2 \iff F \gtreqless 1$. Of course, the verdict with respect to the null hypothesis depends on the critical value of F (distributed with k_2 and $n - k_1$ degrees of freedom).

Deciding that $\sigma^2_{p_1} - \sigma^2$ is significantly different from zero still does not justify the use of the unrestricted model for prediction. For that, σ_f^2 and σ_{f1}^2 need to be evaluated which, in turn, requires an evaluation of h and h_{p_1} as given by equation (50). From equation (50), it is clear that

$$h - h_{p_1} = \underline{r}_2^* \, D_2^{-1} \, \underline{r}_2^{*\prime} \tag{A.9}$$

or the portion of the variance associated with the omitted components. This result follows directly from the orthogonality of the partitions in the P matrix, (P_1 and P_2).

Clearly, the sampling component of the unrestricted forecast variance increases with the variance of the estimator for $\underline{\delta}_2$. The less accurate the estimator of $\underline{\delta}_2$, the larger the sampling variance of the forecast. It is useful to recall the conditions for a large sampling variance for the case on hand.[6] This component of the sampling variance will be large when the components forming $\underline{R}_2$ account for only a small proportion in the observed variation in X'X. This is the situation when there is multicollinearity.

Unbiased estimators for all of the components of the forecast problem are available, as indicated in Section 4.1. In fact, they can be obtained from an OLS regression on all the principal components. Thus, computationally the problem of applying these results is quite tractable. With these estimates or results, the following comparisons can be made:

$$\frac{\sigma^2_{f_{p_1}}}{\sigma_f^2} \gtreqless 1 \iff \frac{\sigma^2_{p_1}}{\sigma^2} \gtreqless \frac{1 + h}{1 + h_{p_1}} = 1 + \frac{h - h_{p_1}}{1 + h_{p_1}}$$

$$\iff \underline{\delta}_2' \, \sigma^2_{r2} \, \underline{\delta}_2 \lesseqgtr \underline{r}_2^* \, D_2^{-1} \, \underline{r}_2^{*\prime}. \tag{A.10}$$

A large variance, σ^2_{r2}, increases the left-hand side of the inequalities and tends to decrease the right-hand side and, hence, leads to preferring the

unrestricted forecast (the actual values of the terms, of course, also depends on $\underline{\delta}_2$ and $\underline{r}_2^*$.

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 R_1, $\underline{r}_1^*$ or R and $\underline{r}^*$? Under the null hypothesis, $\underline{\delta}_2 = 0$, the ratio

$$\frac{\left(\sigma^2_{p_1} - \sigma^2\right)/2}{\sigma^2/(N-k)}$$

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 (A.9) 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 right-hand side of (A.9); consequently, the restricted forecast would be preferred. The converse is also possible.

The discussion is 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 components is evaluated. It is interesting to note that the test of the null hypothesis, with respect to the coefficients of the restricted variables ($\underline{\delta}_2 = \underline{0}$), is neither sufficient nor necessary for determining the relative precision of the forecast in question. The conditions leading to more precision of a restricted forecast are of a similar nature in both cases: For a small sample, little net information embedded in the restricted variables not used in the regression and considerable information in the values of those variables used for estimating the parameters used for the forecast.

Footnotes

*Giannini Foundation Paper No. 601 (reprint identification only).

[1]These practitioners have also demonstrated that more variability is present in econometric models that 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, auxiliary conditions ought to be examined to determine whether a given set of data is appropriate for estimating parameter effects. Such auxiliary 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]Namely,

$$(X'X + x'_{T+1}\, x_{T+1})^{-1} = (X'X)^{-1} - (X'X)^{-1}\, x_{T+1}$$

$$[I + x_{T+1}\, (X'X)^{-1}\, x'_{T+1}]^{-1}\, x_{T+1}\, (X'X)^{-1}$$

or, in general, if $s = (M - 1 + AR^{-1}\, B)^{-1}$, then $S = M - MA\, [R + BMA]^{-1}$ BM. This lemma is formally proved in a number of places; see, for example, Sage and Melsa.

[4]The simultaneous occurrence of a specification involving varying parameters and a disturbance structure for which $\Omega_{T+1/T} \neq 0$ raises an interesting question. For example, if persistence hypothesis is incorporated in the varying parameter structure, why would additional persistence hypotheses be necessary for the disturbances? Generally, a choice should be made as to how to represent this aspect of the problem, as a fixed or random effect. In any case the specification for the model should be parsimonious, as numbers of parameters relative to data are likely to be a problem.

[5]For examination of the identification problem in context of simultaneous equation models with random parameters, see Kelejian. 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.

[6]The discussion is not relevant for large samples where the sampling error is of no quantitative importance.

References

Balestra, P., and M. Nerlove. "Pooling Cross-Section and Time Series Data in the Estimation of the Dynamic Model: The Demand for Natural Gas." Econometrica 34(1966):585-612.

Belsey, David A. "On the Determination of Systematic Parameter Variation in the Linear Regression Model." Annals of Econ. and Social Measurement 2(1973a):487-494.

_____. "The Applicability of the Kalman Filter in the Determination of Systematic Parameter Variation." Annals of Econ. and Social Measurement 2(1973b):531-533.

Brundy, J. M., and Dale W. Jorgenson. "Consistent and Efficient Estimation of Systems of Simultaneous Equations by Means of Instrumental Variables." Frontiers of Econometrics, ed. P. Zarembka. New York: Academic Press, 1974.

Buse, A., and L. Lim. "Cubic Splines as a Special Case of Restricted Least Squares." J. of the Amer. Statist. Assoc. 72(1977):64-68.

Cooley, T. F., and S. J. DeCanio. "Changing Supply Response in Late Nineteenth Century Agriculture." Working Paper. Tufts University, Medford, Massachusetts, 1973.

Cooley, Thomas F., and Edward C. Prescott. "An Adaptive Regression Model." Inter. Econ. Rev. 14(1973a):364-371.

_____. "Varying Parameter Regression: A Theory and Some Applications." Annals of Econ. and Social Measurement 2(1973b):463-473.

Cooper, R. L. "The Predictive Performance of Quarterly Econometric Models of the U. S." Econometric Models of Cyclical Behavior: Studies in Income and Wealth, Vol. 36. New York: National Bureau of Economic Research, 1972.

Covey-Crump, P. A. K., and S. D. Silvey. "Optimal Regression Designs with Previous Observations." Biometrika (1970):551-556

Duffy, W. J. "Parameter Variation in a Postwar Econometric Model." Unpublished Ph.D. dissertation, University of Minnesota, 1969.

Fomby, T. B., R. C. Hill, and S. R. Johnson. "An Optimal Property of Principal Components in the Context of Restricted Least Squares." J. of the Amer. Statist. Assoc. 73(1978):191-193.

Fomby, T. B., and S. R. Johnson. "MSE Evaluation of Ridge Estimates Based on Stochastic Prior Information." Communications in Statist. A6(13) (1977):1245-1258.

Greenberg, Edward. "Minimum Variance Properties of Principal Component Regression." J. of the Amer. Statist. Assoc. 70(1975):194-197.

Goldfeld, S. M., and R. E. Quandt. "The Estimation of Structural Shifts by Switching Regressions." Annals of Econ. and Social Measurement 2(1973):475-485.

Guttman, I. "A Remark on the Optimal Regression Designs with Previous Observations of Covey-Crump and Silvey." Biometrica 64(1971):633-684.

Judge, G. G., and T. Takayama, "Inequality Restrictions in Regression Analysis." J. of the Amer. Statist. Assoc. 61(1966):166-181.

Judge, G. G., and T. A. Yancey. "Inequality Restricted Estimation Under Squared Error Loss." Working Paper No. 78-020, College of Business Administration, University of Georgia, 1978.

Kelejian, H. H. "Random Parameters in a Simultaneous Equation Framework: Identification and Estimation." Econometrica 42(1974):517-527.

Kiefer, J. "On the Nonrandomized Optimality and Randomized Nonoptimality of Symmetrical Designs." Annals of Mathematical Statist. 29(1958):675-699.

Lucas, R. "Econometric Policy Evaluation: A Critique." J. of Monetary Econ. I Supplement 3(1976):19-46.

MacRae, Elizabeth Chase. "Optimal Experimental Design for Dynamic Econometric Models." Annals of Econ. and Social Measurement 6(1977):399-405.

Mundlak, Yair. "On the Pooling of Time Series and Cross Section Data." Econometrica 46(1978a):69-85.

_____. "On Variable Coefficients Models." Annals de L'Insee, France, 1978b.

Poirier, D. H. The Econometrics of Structural Change. Amsterdam: North Holland Publishing Company, 1976.

Rausser, G. C. "Model Specification and Use: Nonlinear Parameters Versus Nonlinear Variables." Unpublished mimeograph. Harvard University, 1978.

Rausser, G. C., and P. S. Laumas. "The Stability of Demand for Money in Canada." J. of Monetary Econ. 2(1976):367-380.

Rosenberg, B. "A Survey of Stochastic Parameter Regression." Annals of Econ. and Social Measurement 2(1973a):381-397.

_____. "The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression." Annals of Econ. and Social Measurement 2(1973b):399-428.

Sage, Andrew P., and James L. Melsa. Estimation Theory with Applications to Communications and Control. New York: McGraw-Hill Book Company, 1971.

Schmidt, Peter. Econometrics. New York: Marcel Dekker, Inc., 1976.

Silvey, S. D. "Multicollinearity and Imprecise Estimation." J. of the Royal Statist. Society, Series B 31(1969):539-552.

Singh, B., A. L. Nagar, N. K. Choudhry, and Raj. Baldev. "On the Estimation of Structural Change: A Generalization of the Random Coefficients' Regression Model." Inter. Econ. Rev. 17(1976):340-361.

Swamy, P. A. V. B. "Efficient Inference in a Random Coefficient Regression Model." Econometrica 38(1970):311-323.

Theil, H., and A. S. Goldberger. "On Pure and Mixed Statistical Estimation in Econometrics." Int. Econ. Rev. 2(1961):65-78.

Wynn, H. P. "The Sequential Generation of D-Optimal Experimental Designs." Annals of Mathematical Statist. 41(1970):1655-1664.

Zellner, Arnold. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." J. of the Amer. Statist. Assoc. 62(1962):348-368.