

STOCHASTIC COEFFICIENTS ESTIMATION PROGRAM (SCEP): A USER'S GUIDE

I-Lok Chang*

American University, Washington DC 20016

P.A.V.B. Swamy

U.S. Bureau of Labor Statistics, Washington DC 20212-0001

July 22, 2008

Contents

1	Introduction	2
2	Statement of the Problem	2
2.1	Single-Group Random Coefficient Model	3
2.2	Multiple-Group Random Coefficient Model	9
3	Setting Up Data Files for SCEP	11
4	How to Execute SCEP?	31
5	Five Levels of Output	32
6	Computational Steps Coded in SCEP	38
7	An Example	58
	Appendices	63
	A: Estimation of Direct Effects—A Step-by-Step Argument	63
	C: A Numerically Stable Method of Computing Generalized Least Squares Estimators and their Standard Errors	91
	D: A Stochastic Coefficients Approach to Forecasting	94

*Supported by a grant from the Office of the Comptroller of the Currency.

1 Introduction

This manual will introduce you to SCEP which is a computer software system for econometric estimation of stochastic coefficients models.¹ This program searches for a stable solution and uses an algorithm that is both numerically efficient and stable.² A set of questions which you can answer using this program is given in Appendix A to this manual.

Section 2 explains what a stochastic coefficients model is and how it differs from a regression model. Section 3 explains how to set up SCEP files. Section 4 describes the command you need to issue to execute SCEP. Section 5 describes the different levels of output that can be produced by constructing a stochastic coefficients model in SCEP. Section 6 presents the flow diagram for SCEP. An example is given in Section 7.

2 Statement of the Problem

Given a set of observations, a random coefficient model can be estimated to fit this set of observation. By one observation we shall mean a set of numerical values of the form $\{y_t, x_{1t}, \dots, x_{Kt}, z_{1t}, \dots, z_{pt}\}$, where t is an index sequencing the observations. The first value y_t is a value the dependent variable \tilde{y} of the model, the next sequence of x -values are data for the independent variables $\{\tilde{x}_j, j = 1, 2, \dots, K\}$ of the model, and the final z -values are the for the concomitant variables $\{\tilde{z}_l, l = 1, 2, \dots, p\}$ of the model.

Two methods for constructing a random coefficient model are available in SCEP. If the number of observations does not exceed 250, a model can be calculated directly and efficiently from the entire set of observations. We call such a model a Single-

¹This model was developed by Swamy, P.A.V.B. and P.A. Tinsley (1980), "Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients," *Journal of Econometrics*, 12, 103-142.

²The basic set of algebraic formulas that are coded in SCEP for a single group are given in Chang, I-Lok, C. Hallahan and P.A.V.B. Swamy (1992), "Efficient Computation of Stochastic Coefficients Models," in: H.M. Amman, D.A. Belsley and L.F. Pau, eds., *Computational Economics and Econometrics*, Kluwer Academic Publishers, Boston, 43-53.

Group Random Coefficient Model (**SGRCM**). If the data set is large, with the number of observations exceeding 250, sometimes as many as 15,000, then the data set must be partitioned into groups so that computation will be possible when using a desktop computer with no more than 64 megabytes of RAM. The size of each group of data is to be kept within the upper limit of 250 observations. SCEP executes two passes of calculation through this partitioned data set. In the first pass, a local random coefficient model is estimated for each group. The second pass then follows by combining the local models to obtain a global model for the entire data set. We call this model a Multiple-Group Random Coefficient Model (**MGRCM**). Computer hardware is not necessarily the only factor in choosing the method of MGRCM. A data set sometimes can yield more statistical information if it is partitioned into groups and MGRCM is used to give insight into local and global statistics.

2.1 Single-Group Random Coefficient Model

In this model, SCEP permits you to estimate the following equation:

$$\begin{matrix} \tilde{y}_t & = & \sum_{j=1}^K x_{jt} \tilde{\beta}_{jt} & = & \mathbf{x}'_t \tilde{\boldsymbol{\beta}}_t, & t = 1, 2, \dots, T, \\ (1 \times 1) & & & & (1 \times K)(K \times 1) & \end{matrix} \quad (1)$$

where \tilde{y}_t is a dependent variable, $\mathbf{x}'_t = (x_{1t}, x_{2t}, \dots, x_{Kt})$ is a vector of independent variables, and $\tilde{\boldsymbol{\beta}}_t = (\tilde{\beta}_{1t}, \tilde{\beta}_{2t}, \dots, \tilde{\beta}_{Kt})'$ is a vector of coefficients. Here and in the sequel the symbols with a curl over them represent random variables and the same symbols without a curl represent the values of those random variables. For example, \tilde{y}_t is a random variable and y_t is its value.

We call the equation $\tilde{y}_t = \sum_{j=1}^K x_{jt} \beta_j + \tilde{u}_t$ a fixed- $\boldsymbol{\beta}$ regression. Equation (1) differs from this regression in that all the coefficients are assumed to be stochastic. Note that in the fixed- $\boldsymbol{\beta}$ regression, if you set $x_{1t} = 1$ for all t , then β_1 is the intercept which can be distinguished from the disturbance \tilde{u}_t , provided the latter has mean zero. This is not the case with equation (1). If you set $x_{1t} = 1$ for all t in equation (1), then you cannot identify \tilde{u}_t and a stochastic intercept term separately

and should combine both specifications into a single term, $\tilde{\beta}_{1t}$.

Quite possibly, each coefficient of equation (1) is the sum of two parts, a direct effect of an explanatory variable on \tilde{y}_t , and an indirect or proxy effect (or omitted variables bias) due to the fact that the explanatory variable affects or proxies for omitted variables and the omitted variables, in turn, affect \tilde{y}_t .³ You can separate these two effects by assuming that $\tilde{\beta}_t$ satisfies the equation

$$\begin{matrix} \tilde{\beta}_t & = & \Pi \mathbf{z}_t & + & L \tilde{\epsilon}_t, \\ (K \times 1) & & (K \times p)(p \times 1) & & (K \times m)(m \times 1) \end{matrix} \quad (2)$$

where the elements of $\mathbf{z}_t = (z_{1t}, z_{2t}, \dots, z_{pt})'$ with $z_{1t} \equiv 1$ are called concomitants and L is a known matrix. The conditions under which the concomitants separate proxy effects from direct effects are given in Swamy et al.⁴

The matrix L is completely in your control and you can use it to impose a priori restrictions on the covariance matrix of $\tilde{\epsilon}_t$. For example, if you set the j th row of L equal to a null vector, then the j th element of $\tilde{\beta}_t$ becomes an exact linear deterministic function of \mathbf{z}_t . Alternatively, if you set $L = I_K$, an identity matrix of order K , then every coefficient of equation (1) is partly deterministic and partly random. **If you tell SCEP to use the default value for L , then SCEP sets $L = I_K$.**

The elements of Π , L , and $\tilde{\epsilon}_t$ may be depicted as

$$\begin{aligned} \Pi &= \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1p} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2p} \\ \vdots & \vdots & & \vdots \\ \pi_{K1} & \pi_{K2} & \cdots & \pi_{Kp} \end{bmatrix}, \\ L &= \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & & \vdots \\ l_{K1} & l_{K2} & \cdots & l_{Km} \end{bmatrix}, \end{aligned} \quad (3)$$

³For additional discussion of these effects, see Appendix A to this manual and Swamy, P.A.V.B., J.S. Mehta, R.N. Singamsetti (1992), "Circumstances in Which Different Criteria of Estimation Can be Applied to Estimate Policy Effects," Discussion Paper, Federal Reserve Board, Washington, DC 20551.

⁴Swamy, Mehta and Singamsetti, *op. cit.*

and

$$\tilde{\epsilon}_t = \begin{bmatrix} \tilde{\epsilon}_{1t} \\ \tilde{\epsilon}_{2t} \\ \vdots \\ \tilde{\epsilon}_{mt} \end{bmatrix}.$$

Equation (1) may be neither too broad nor too narrow if you assume that $\tilde{\epsilon}_t$ satisfies the equation

$$\begin{matrix} \tilde{\epsilon}_t \\ (m \times 1) \end{matrix} = \begin{matrix} \Phi \tilde{\epsilon}_{t-1} \\ (m \times m)(m \times 1) \end{matrix} + \begin{matrix} \tilde{\mathbf{a}}_t, \\ (m \times 1) \end{matrix} \quad (4)$$

where $E\tilde{\mathbf{a}}_t = 0$ for all t , $E\tilde{\mathbf{a}}_t\tilde{\mathbf{a}}'_s = \begin{cases} \sigma_a^2\Delta_a & \text{if } t = s \\ (m \times m) & \\ 0 & \text{if } t \neq s, \end{cases}$

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1m} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2m} \\ \vdots & \vdots & & \vdots \\ \phi_{m1} & \phi_{m2} & \cdots & \phi_{mm} \end{bmatrix}, \quad \tilde{\mathbf{a}}_t = \begin{bmatrix} \tilde{a}_{1t} \\ \tilde{a}_{2t} \\ \vdots \\ \tilde{a}_{mt} \end{bmatrix},$$

and

$$\Delta_a = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1m} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2m} \\ \vdots & \vdots & & \vdots \\ \delta_{m1} & \delta_{m2} & \cdots & \delta_{mm} \end{bmatrix}.$$

The eigenvalues of Φ which are either real or complex are restricted to be less than 1 in absolute value. A rationale for this restriction is given in Swamy et al.⁵ and an important generalization of assumption (3) is given in Swamy and Tinsley.⁶

If you think that equations (2) and (3) are too broad relative to your data set, then you can restrict either Φ or Δ_a or both to be diagonal.⁷ Or you can impose

⁵Swamy, Mehta and Singamsetti, *op. cit.*

⁶Swamy and Tinsley, *op. cit.*, p. 106.

⁷We will explain later how to impose these restrictions.

the following linear restrictions other than the diagonality restrictions:

$$\begin{aligned}
\frac{R_{\Pi}^e \text{vec}(\Pi)}{(\ell_{\Pi}^e \times Kp)(Kp \times 1)} &= \frac{\mathbf{r}_{\Pi}^e}{(\ell_{\Pi}^e \times 1)} \\
\frac{R_{\Pi}^i \text{vec}(\Pi)}{(\ell_{\Pi}^i \times Kp)(Kp \times 1)} &\geq \frac{\mathbf{r}_{\Pi}^i}{(\ell_{\Pi}^i \times 1)} \\
\frac{R_{\Phi} \text{vec}(\Phi)}{(\ell_{\Phi} \times m^2)(m^2 \times 1)} &= \frac{\mathbf{r}_{\Phi}}{(\ell_{\Phi} \times 1)} \\
\frac{R_{\Delta_a} \text{vec}(\Delta_a)}{(\ell_{\Delta_a} \times \bar{m})(\bar{m} \times 1)} &= \frac{\mathbf{r}_{\Delta_a}}{(\ell_{\Delta_a} \times 1)}
\end{aligned} \tag{5}$$

where for any matrix P , $\text{vec}(P)$ denotes its column stack, Δ_a is a triangular matrix consisting of only the diagonal and below diagonal elements of Δ_a , and $\bar{m} = m(m+1)/2$. For example, $\text{vec}(\Pi) = [(\pi_{11}, \pi_{21}, \dots, \pi_{K1}), (\pi_{12}, \pi_{22}, \dots, \pi_{K2})', \dots, (\pi_{1p}, \pi_{2p}, \dots, \pi_{Kp})']'$ and $\text{vec}(\Delta_a) = [(\delta_{11}, \delta_{21}, \dots, \delta_{m1}), (\delta_{22}, \delta_{32}, \dots, \delta_{m2}), \dots, (\delta_{mm})']'$.

Thus, you can impose three types of restrictions: (i) some elements of $\tilde{\beta}_t$ are the exact linear functions of \mathbf{z}_t , (ii) $\Phi = 0$ or either Φ or Δ_a or both are diagonal, and (iii) some linear combinations of the elements of Π , Φ , and Δ_a are equal to known values.

Examples of (4): The matrix Π is symmetric; some columns of Φ and Δ_a are null.

If $\mathbf{x}_t = \mathbf{z}_t$, then it is correct to constrain Π to be a symmetric matrix. To show this, substitute the right-hand side of equation (2) for $\tilde{\beta}_t$ in equation (1). This gives

$$\tilde{y}_t = \mathbf{x}_t' \Pi \mathbf{z}_t + \mathbf{x}_t' L \tilde{\epsilon}_t. \tag{6}$$

You can impose the restrictions that all the elements of $\tilde{\epsilon}_t$ except the first element are degenerate at zero using equation (4). You can also impose the constraints that $L = I_K$ and $\mathbf{x}_t = \mathbf{z}_t$. When $\mathbf{x}_t = \mathbf{z}_t$, the restriction that the matrix Π be symmetric is not a serious one, as is shown in the literature. An example of an equation which satisfies all these restrictions is a translog cost function. However, translog and other flexible functional forms may give better approximations to the true functions

if all the elements of $\tilde{\epsilon}_t$ are assumed to be nondegenerate and SCEP permits you to estimate them under such an assumption.

Indeed, equation (4) can depict all the restrictions under which equations (1)-(3) reduce to the fixed- β regression of \tilde{y}_t on \mathbf{x}_t . These and other restrictions are depicted in Appendix B to this manual.

Since equations (1)-(3) are more general than their fixed- β version, they can be justified on the grounds that general cases can be true even when the particular cases are false. In any case, it is a good practice to consider equations (1)-(3) and their special cases and compare the results.

Note that if the vector \mathbf{z}_t is null, then the coefficient vector of equation (1) is pure noise with mean zero. For this reason, you may not want to assume that \mathbf{z}_t is a null vector and p is equal to zero. In situations, where the relevant nonconstant z_{jt} are unknown, you can set $z_{1t} = 1$ and $z_{jt} = 0$ with $j \neq 1$ for all t , $p = 1$ and $L = I_K$, in which case equation (2) becomes $\tilde{\beta}_t = \pi_1 + \tilde{\epsilon}_t$, where π_1 is the first column of Π . **You can use this value for z_t by telling SCEP to use the default value for z_t .**

You can also set $\Phi = 0$ if you are estimating equation (1) from cross-section data.

However, you cannot set $\Phi = I_m$ because of the restriction on the eigenvalues of Φ .⁸

When equations (3) and (5) hold, SCEP uses the appropriate formulas to estimate the covariance matrices of the weighted and unweighted least squares estimators of Π . These formulas are presented in Appendix C to this manual.

A stochastic coefficient approach to forecasting the values of \tilde{y}_t is described in Appendix D to this manual. It accounts for more forecast error sources than its fixed- β counterpart.

PARAMETERS TO BE ESTIMATED

⁸The value I_m for Φ has serious econometric implications which are discussed in Case X of Appendix B to this manual.

$$\Pi, \Phi, \Delta_a, \sigma_a^2$$

SCEP uses an iterative method to estimate these parameters and, therefore, needs some initial values for Φ and Δ_a . If you do not know what values to use as the initial values for these parameters, then SCEP lets you use the default values $\Phi = 0$ and $\Delta_a = I_m$ as the initial values.

STABILIZATION CONDITIONS

The sequence of parameter estimates generated by SCEP may stabilize at a value because

- (1) Φ is estimated subject to the restriction that its eigenvalues are less than or equal to an upper bound **EVLTOL** (assigned by you, e.g., 0.995) in absolute value and
- (2) the matrix Δ_a is estimated subject to the restrictions that it is nonnegative definite and its Frobenius norm $\| \Delta_a \| = \sqrt{\sum_{i=1}^m \sum_{j=1}^m |\delta_{ij}|^2}$ is less than or equal to the Frobenius norm $\| \hat{\Delta}_a \|$, where $\hat{\Delta}_a$ is a nonnegative definite estimate of Δ_a obtained in iteration 1.

The upper bound on the eigenvalues of Φ in condition (1) above can be changed by assigning a positive real value that is less than 1.0 to the parameter EVLTOL in the data file PARAMTR. Sometimes changing EVLTOL from 0.995 to 0.99 can lead to a more stable overall approximation of the model. The diagonality restriction on Φ can sometimes provided more stability in the calculation, unless the elements of $\hat{\Delta}_a$ get very small, in which case certain parameters used in the estimation of Φ need to be changed to increase the likelihood of stabilization. If you need help in making these changes, you can contact

I-Lok Chang

Department of Mathematics and Statistics

Gray Hall, American University

4400 Massachusetts Avenue NW

Washington, DC 20016-8001, U.S.A.

Home Phone Number: 301-530-5865

Work Phone Number: 202-885-3132.

VALUES TO BE FORECASTED

$$\epsilon_t, \beta_t, \text{ and } y_t, \text{ for } t = T + 1, T + 2, \dots, T + F \quad (7)$$

MODEL FITTING AND VALIDATION

SCEP permits you to do the following:

- Select a subset of the observations on $\tilde{y}_t, \tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{z}}_t$ to be used in estimating model (5) and its special cases.
- Hold out the remaining subset of the observations on these variables for these models' validation.
- Forecast the latter values of \tilde{y}_t with the fitted models, and compare the actuals with the forecasts.

2.2 Multiple-Group Random Coefficient Model

If the entire set of observations is partitioned into G groups, the groups are identified with an integer index g , where $1 \leq g \leq G$. The data and the computed parameters for the different groups are now distinguished with the index g .

The global model (pooled model) is given by the following set of equations

$$\tilde{y}_{tg} = \sum_{j=1}^K x_{jtg} \tilde{\beta}_{jtg} = \underline{x}_{tg}' \tilde{\beta}_{\underline{t}g} \quad (g = 1, 2, \dots, G; t = 1, 2, \dots, T_g) \quad , \quad (8)$$

where $x_{1tg} = 1$. The coefficients $\tilde{\beta}_{jtg}$ is given the following decomposition

$$\tilde{\beta}_{jtg} = \bar{\beta}_j + \sum_{l=1}^{p-1} \pi_{jl} \tilde{z}_{ltg} + \tilde{\mu}_{jg} + \sum_{i=1}^M l_{ji} \hat{\epsilon}_{itg} \quad , \quad (9)$$

where $L = [l_{ji}]$ is a $K \times m$ matrix in (3), now serving as a global weighting matrix. The indices in (9) vary over the range $g = 1, 2, \dots, G$, $t = 1, 2, \dots, T_g$, $j = 1, 2, \dots, K$, $i = 1, 2, \dots, m$. In matrix forms, (9) can be expressed as

$$\begin{array}{ccccccc} \tilde{\underline{\beta}}_{tg} & = & \underline{\bar{\beta}} & + & \hat{\Pi} & \tilde{\underline{z}}_{tg} & + & \underline{\tilde{\mu}}_g & + & L & \tilde{\underline{\epsilon}}_{tg} , \\ K \times 1 & & K \times 1 & & K \times (p-1) & (p-1) \times 1 & & K \times 1 & & K \times m & m \times 1 \end{array}$$

$$\tilde{\underline{\beta}}_{tg} = \begin{bmatrix} \tilde{\beta}_{1tg} \\ \tilde{\beta}_{2tg} \\ \vdots \\ \tilde{\beta}_{Ktg} \end{bmatrix} = \begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \\ \vdots \\ \bar{\beta}_K \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{p-1} \pi_{1i} \tilde{z}_{itg} \\ \sum_{i=1}^{p-1} \pi_{2i} \tilde{z}_{itg} \\ \vdots \\ \sum_{i=1}^{p-1} \pi_{Ki} \tilde{z}_{itg} \end{bmatrix} + \begin{bmatrix} \tilde{\mu}_{1g} \\ \tilde{\mu}_{2g} \\ \vdots \\ \tilde{\mu}_{Kg} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^m l_{1i} \tilde{\epsilon}_{itg} \\ \sum_{i=1}^m l_{2i} \tilde{\epsilon}_{itg} \\ \vdots \\ \sum_{i=1}^m l_{Ki} \tilde{\epsilon}_{itg} \end{bmatrix} .$$

Stochastic assumptions on $\tilde{\underline{\epsilon}}_{tg}$ and $\underline{\tilde{\mu}}_g$ are given by the equations

$$\tilde{\underline{\epsilon}}_{tg} = \Phi_g \tilde{\underline{\epsilon}}_{(t-1)g} + \tilde{\underline{a}}_{tg} ,$$

where

$$E(\tilde{\underline{a}}_{tg}) = 0 , \quad E(\underline{\tilde{\mu}}_g) = 0 , \quad \text{and} \quad E(\underline{\tilde{\mu}}_g \tilde{\underline{\epsilon}}_{tg}) = 0$$

for all t and g, and

$$\begin{aligned} E(\tilde{\underline{a}}_{tg} \tilde{\underline{a}}'_{t'g'}) &= \begin{cases} \sigma_g^2 \Delta_{ag} & \text{if } g = g' \quad \text{and} \quad t = t' \\ 0 & \text{if } g \neq g' \quad \text{or} \quad t \neq t' \end{cases} , \\ E(\underline{\tilde{\mu}}_g \underline{\tilde{\mu}}'_{g'}) &= \begin{cases} \sigma_g^2 \Delta & \text{if } g = g' \\ 0 & \text{if } g \neq g' \end{cases} . \end{aligned}$$

The matrices Φ_g , Δ_{ag} , and Δ have the dimensions

$$\begin{array}{ccc} \Phi_g & \Delta_{ag} & \Delta \\ m \times m & m \times m & K \times K \end{array} .$$

The eigenvalues of each matrix Φ_g have moduli less than 1. The matrices Δ_{ag} and Δ are positive semi-definite matrices.

In the first pass of the calculation, linear constraints can be specified for each local model. The specification is global in the sense that the constraints in (5) are specified only once, to be applied identically to each group in the calculation of its local model. At this point, no corresponding linear constraints can be imposed on the global model (8)calculated in the second pass.

Forecasting within each group assumes the same format as (7). The upper bound T is now T_g , and the upperbound F is now F_g . For each group, forecasts are made relative to its local model, and then recomputed relative to the final global model.

3 Setting Up Data Files for SCEP

SCEP is a FORTRAN program that is currently running on Intel Pentium desktop computers under Windows 2000, or Windows XP. To run the program, place all the necessary data files and the executable module in the same directory on the hard disk. Then enter the name of the executable module to run the program. The output of the program is saved on two files SCEPOUT and SCEPLOG in the same directory.

INPUTS REQUIRED BY SCEP

Data required to construct a single-group model (with $G = 1$) or a multiple-group model (with $G > 1$):

$y_{tg}, \mathbf{x}_{tg}, \mathbf{z}_{tg}$, $g = 1, 2, \dots, G, t = 1, 2, \dots, \dots T_g, L$, initial values of Φ_g and Δ_{ag} , the values of T_g, K, p , and m . (Repetition: I_K is the default value of L , the default values of \mathbf{z}_{tg} are: for all t and g , $z_{jtg} = 1$ if $j = 1$ and $= 0$ if $j \neq 1$. The estimation method coded in SCEP is iterative and the values $\Phi_g = 0$ and $\Delta_{ag} = I_m$ are taken as the default initial values.)

Data required to compute the out-of-sample values of \tilde{y}_{tg} :

$\mathbf{x}_{tg}, \mathbf{z}_{tg}$, $g = 1, 2, \dots, G, t = T_g + 1, T_g + 2, \dots, T_g + F_g$, and the value of F_g , where F_g is the number of out-of-sample periods for which the forecasts of y_{tg} are needed. (This information is, of course, not needed if $F_g = 0$.)

Out-of-sample values of \tilde{y}_{tg} for comparison with their forecasts:

y_{tg} , $g = 1, 2, \dots, t = T_g + 1, T_g + 2, \dots, T_g + F_g$. (If these values are unknown to you at the time of forecasting them, then code the value -9999.99 for each of them.)

Data to impose the restrictions (4):

$$\begin{array}{lll} R_{\Pi}^e & \mathbf{r}_{\Pi}^e & \ell_{\Pi}^e \\ R_{\Pi}^i & \mathbf{r}_{\Pi}^i & \ell_{\Pi}^i \\ R_{\Phi} & \mathbf{r}_{\Phi} & \ell_{\Phi} \\ R_{\Delta_a} & \mathbf{r}_{\Delta_a} & \ell_{\Delta_a} \end{array}$$

LIMITATIONS OF SCEP

$$T_g \leq 150, \quad K \leq 15, \quad p \leq 15, \quad m \leq 15$$

$$\text{Number of groups} \leq 120$$

If you need to run a job that some of the upper limits above need to be changed, you may contact I-Lok Chang.

DATA FILES REQUIRED BY SCEP

The following list is a comprehensive list of data files that SCEP uses. Only three files must be present in any run of SCEP: PARAMTRS, YXDKEY, and YXDATA1. The other files are for data that will be needed for certain specified options in the calculation of the model. In each run of SCEP, all the required data files must reside in the same disk directory as the executable module of SCEP.

INPUT FILES FOR EXECUTING SCEP

PARAMTRS

YXDKEY

YXDATA1

YXDATA2

ZDATA1

ZDATA2

LDATA1

LDATA2

PDATA1

PDATA2

DDATA1

DDATA2

RCFPH

RVCPH

RCFDT

RVCDT

RCFPI

RVCPI

PSAVE.(GROUP ID)

DSAVE.(GROUP ID)

A RULE IN SETTING UP A DATA FILE

In all the data files, with the exception of PSAVE.(GROUP ID) and DSAVE.(GROUP ID), the final two lines of the file must be the following two lines:

END OF DATA

//

SETTING UP THE PARAMTRS FILE

In this file, you specify the dimensions of your data matrices and your computational and printing options. The maximum length of each line is 80 characters, starting from the first space of the line. A blank line or a line starting with an exclamation mark “!” is ignored by the program. All global options – options to be applied to every group – must be entered between the two lines “BEGIN GROUP(ALL)” and “END GROUP(ALL)”. An option that is to be applied only to Group K must be entered between the two lines “BEGIN GROUP(K)” and “END GROUP(K)”. Such a local option overrides any corresponding global option. The only local option that may be applied at this writing is the option of assigning the number of iterations for computing a local model. The following example is a PARAMTRS FILE for a problem analyzed at the U.S. Office of the Comptroller of the Currency. BEGIN GROUP(ALL)

```

T = 108
M = 2
K = 2
P = 3
AUTOMATIC CONSTANT 1 = NO
SAVE BETA = NO
ZERO PHI = NO
DIAGONAL PHI = YES
DIAGONAL DELTA = YES
INITIAL PHI MATRIX = DEFAULT
INITIAL DELTA MATRIX = DEFAULT
Z-VECTORS = USER'S
NUMBER OF ROWS IN LINEAR EQUALITY RESTRICTION ON PHI = 0
NUMBER OF ROWS IN LINEAR EQUALITY RESTRICTION ON PI = 6
NUMBER OF ROWS IN LINEAR INEQUALITY RESTRICTION ON PI = 1
NUMBER OF ROWS IN LINEAR EQUALITY RESTRICTION ON DELTA = 0
SAVE PHI MATRIX = NO
SAVE DELTA MATRIX = NO
NUMBER OF ITERATIONS = 30
PRINT LEVEL = 5
DATA COLUMN WIDTH = 10
OLS = NO
GLS = NO
PROGRAM TITLE = SWAMSLEY DEFICITS DATA
EVL TOL = 0.995
END GROUP(ALL)
!
BEGIN GROUP(2)
    NUMBER OF ITERATIONS = 5
END GROUP(2)
!
!BEGIN GROUP(1)
!    NUMBER OF ITERATIONS = 10
!END GROUP(1)
END OF DATA
//

```

Detailed instructions for creating the PARAMTRS FILE are given below:

$T =$ [Enter here the total number of observations on \tilde{y}_{tg} , \mathbf{x}'_{tg} , and \mathbf{z}'_{tg} . T is $(T_1 + F_1) + \dots + (T_g + F_g) + \dots + (T_G + F_G)$.]

$K =$	[Enter here the number of explanatory variables including the constant term in equation (1).]
$m =$	[Enter here the column dimension of L in equation (2). This dimension is K if L is an identity matrix.]
$p =$	[Enter here the number of explanatory variables including the constant term in equation (2). This number is equal to 1 when for all t and g , $z_{1tg} = 1$ and $z_{jtg} = 0$ with $j \neq 1$.]
AUTOMATIC CONSTANT 1 FOR X AND Z =	[Enter here YES if you want to set $x_{1tg} = z_{1tg} = 1$ for all t and g , but do not want to type a vector of 1's in the YXDATA1 and ZDATA1 files. Enter NO otherwise. You should also enter NO here if you do not want to set $x_{1tg} = 1$ for all t and g but want to type a vector of 1's in the ZDATA1 file.]
Z-VECTORS =	[Enter here DEFAULT if you want to set $z_{1tg} = 1$, $z_{jtg} = 0$ for $j \neq 1$ and all t and g . Enter USER'S otherwise.]
L-MATRIX =	[Enter here DEFAULT if you want to set $L = I_K$. Enter USER'S otherwise.]
SAVE BETA =	[Enter here YES if you want to save the β_{tg} vectors obtained in each iteration of in the calculation of each local model and all the β_t vectors of the final global model (for a MSRCM). Enter NO otherwise. To conserve disk space, the β -vectors are saved in binary format.]
ZERO PHI =	[Enter here YES if you want to restrict each Φ_g to be 0 and NO otherwise.]
DIAGONAL PHI =	[Enter here YES if you want to restrict each Φ_g to be diagonal and NO otherwise.]

DIAGONAL DELTA =

[Enter here YES if you want to restrict each Δ_{ag} to be diagonal and NO otherwise.]

INITIAL PHI MATRIX =

[Enter here DEFAULT, USER'S, or SWAMSLEY SAVE-FILE according as the initial value of Φ_g you want to use is equal to 0, your own value, or the estimate obtained in the last iteration of a previous run.]

INITIAL DELTA MATRIX =

[Enter here DEFAULT, USER'S, or SWAMSLEY SAVE-FILE according as the initial value of Δ_{ag} you want to use is equal to I_m , your own value, or the estimate obtained in the last iteration of a previous run.]

NUMBER OF ROWS IN LINEAR
EQUALITY RESTRICTION ON
PHI =

[Enter here the value of ℓ_Φ if you want to impose the restrictions described in (5) and 0 otherwise.]

NUMBER OF ROWS IN LINEAR
EQUALITY RESTRICTION ON PI
=

[Enter here the value of ℓ_Π^e if you want to impose the restrictions described in (5) and 0 otherwise.]

NUMBER OF ROWS IN LINEAR
INEQUALITY RESTRICTION ON
PI =

[Enter here the value of ℓ_Π^i if you want to impose the restrictions described in (5) and 0 otherwise.]

NUMBER OF ROWS IN LINEAR
EQUALITY RESTRICTION ON
DELTA =

[Enter here the value of $\ell_{\Delta_{ag}}$ if you want to impose the restrictions described in (5) and 0 otherwise.]

SAVE PHI MATRIX =

[Enter here YES if you want to save the estimate of Φ_g obtained in the last iteration in the calculation of a local model. Enter NO otherwise. You have to say, "YES," here if you want to say, "SWAMSLEY SAVE-FILE," on the right-hand side of the equation, INITIAL PHI MATRIX =, in the PARAMTRS file for a subsequent run.]

SAVE DELTA MATRIX =

[Enter here YES if you want to save the estimate of Δ_{ag} obtained in the last iteration in the calculation of a local model. Enter NO otherwise. You have to say, "YES," here if you want to say, "SWAMSLEY SAVE-FILE," on the right-hand side of the equation, INITIAL DELTA MATRIX =, in the PARAMTRS file for a subsequent run.]

NUMBER OF ITERATIONS =

[You may enter here a number less than or equal to 50. This number is the number of iterations executed in the calculation of a local model. In cases where Φ_g is diagonal, convergence of the iterative scheme coded in SCEP may not require more than 50 iterations. The first iteration is always labeled 0. This global upper limit may be overridden in specifying local options for individual groups.]

PRINT LEVEL =

[Enter here any one of the integers 1, 2, 3, 4, 5. Each of these integers denotes a particular level of SCEP's output in the file SCEPOUT. A detailed description of these levels is given in Section 5 below.]

STATISTICS LEVEL = 1

[The right entry here is always 1 because this option is not currently operative.]

DATA COLUMN WIDTH =

[Enter here the number of horizontal spaces (column width) in each column of data in a input file. Data in all the input files except the PARAMTRS FILE and the YXDKEY FILE must be entered in columns of this width.]

PROGRAM TITLE =

[Enter here any name which is up to 60 characters long.]

OLS = [Enter here YES if you are using SCEP to obtain the least squares estimates of the fixed- β version of equation (1) and NO otherwise.]

GLS = [Enter here YES if you are using SCEP to obtain the generalized least squares estimates of the fixed- β version of equation (1) with AR (1) errors and NO otherwise.]

EVLTOL = [Enter here a positive real number strictly less than 1.0. This number is the upper bound for the absolute values of the eigenvalues of the matrix Φ_g .]

SETTING UP THE YXDKEY FILE

The data files YXDATA1, YXDATA2, ZDATA1, and ZDATA2 contains observations in the form of one observation occupying one row in a file. There is no designation alongside each row to indicate the group where this row of data is to reside, nor any designation on whether the row of data is to be used for model calculation or for forecasting. The role of the file YXDKEY is to provide such association information for all the observations. The following example is a simple YXDKEY file.

```

1
1
1
1
2
2
2
1
2
2
2      YF
1      YF
1      YF
2      F
1      F
END OF DATA
//

```

Each row contains an positive integer followed possibly by a character string consisting of F or YF. At least one blank space separates the integer and any character string that follows. The entries in each row must appear within the first 80 spaces of the row. There is no further restriction on the position of each entry in a row.

The integer in a row is the group identification number of the corresponding row in a data file containing observations. The characters YF indicates that the row of observation is to be used in forecasting calculation and comparison y-value is provided in the observation. The characters F indicates the row of observation is to be used for forecasting but no comparison y-value is provided. If the character strings F and YF are absent, the row of observation is to be used for model calculation. The forecasting method implemented in SCEP require that within each group, all the forecasting observations must be a final segment of the data sequence. In the example above, GROUP 1 is assigned 4 observations for model calculation and 3 observations for forecast calculation. For GROUP 2, the numbers are 5 and 2 respectively.

SETTING UP SCEP DATA SETS

The information and options written out to the PARAMTRS file and the YXDKEY file do not provide all the numerical inputs needed by SCEP. You need to set up these files next. Data and prior information about the observables and unobservables, respectively, of equations (1) and (2) are brought into SCEP from the input (data) files other than the PARAMTRS file and the YXDKEY file. They are organized as follows:

- You have to write your data series for the \tilde{y}_t and x_{jt} to the YXDATA1 and YXDATA2 files.
- If you set $p =$ a value greater than 1 and say, “USER’S,” on the right-hand side of the equation, Z-VECTORS =, in the PARAMTRS file, then you have to write out your data series for the z_{jt} to the ZDATA1 and ZDATA2 files.
- If you say, “USER’S,” on the right-hand side of the equation, L-MATRIX =, in the PARAMTRS file, then you have to write out your value of L other than an identity matrix to the LDATA1 and LDATA2 files.
- If you say, “USER’S,” on the right-hand side of the equation, INITIAL PHI MATRIX =, in the PARAMTRS file, then you have to write your value of Φ to the files PDATA1 and PDATA2. The two files are to provide the same initial Φ_g for all the groups.
- If you say, “USER’S,” on the right-hand side of the equation, INITIAL DELTA MATRIX =, in the PARAMTRS file, then you have to write out your value of Δ_a to the DDATA1 and DDATA2 files. The two files are to provide the same initial Δ_{ag} for all the groups.

You write out data to all these files under the same format. Each row of data occupies at most the first 80 spaces of the row. If you want to split these spaces into several columns, then all the columns should have the same width and there should be no spaces left empty between any two consecutive columns. If the column width you chose is not equal to a fraction of 80 or if 80 divided by the column width you chose is not an integer, then you should leave all the spaces between the right edge of the last column and the right edge of the page empty. For example, if you decided to mark columns of width equal to 9 spaces each, then you should have 8 columns occupying the first 72 spaces and the last 8 spaces blank in each input file. The column width is called DATA COLUMN WIDTH (DCW). The product of DCW and the number of columns into which 80 spaces are split should be less than or equal to 80. Recall that the value of DCW is entered in the PARAMTRS file. A data value written to an input (data) file is either a numerical value or a character label and it must reside completely within a column. There is no restriction on the position of the value within a column. All the columns in a file must be filled before going to its continuation file. For example, the data which cannot fit in YXDATA1, ZDATA1, LDATA1, PDATA1, and DDATA1 files can be written to YXDATA2, ZDATA2, LDATA2, PDATA2, and DDATA2 files, respectively. The former files have to be filled out completely first before going to the latter. Examples of these files are given in the next few pages:

YXDATA1

80 spaces				
DCW ^a spaces	DCW ^b spaces	DCW spaces	...	DCW ^c spaces
<i>y</i> -label ^d	<i>x</i> -label 1 ^d	<i>x</i> -label 2 ^d	...	<i>x</i> -label <i>i</i> ^d
<i>y</i> ₁ ^e	1	<i>x</i> ₂₁	...	<i>x</i> _{<i>i</i>1}
<i>y</i> ₂	1	<i>x</i> ₂₂	...	<i>x</i> _{<i>i</i>2}
<i>y</i> ₃	1	<i>x</i> ₂₃	...	<i>x</i> _{<i>i</i>3}
⋮	⋮	⋮	...	⋮
<i>y</i> _{<i>T</i>}	1	<i>x</i> _{2<i>T</i>}	...	<i>x</i> _{<i>i</i><i>T</i>}

END OF DATA

//

^aThe first column should always contain observations on the dependent variable. If a row of observation is to be used for forecasting and the comparison *y*-value is not available, code the *y*-value as **−9999.99**. In these cases, SCEP computes the forecast of the actual value but cannot compute the root mean squared error for this forecast.

^bYou can eliminate this column by writing, “YES,” on the right-hand side of the equation, AUTOMATIC CONSTANT 1 FOR X AND Z =, in the PARAMTRS file.

^cContinued in YXDATA2 file.

^dIf you do not want to type in labels for the variables $\tilde{y}_t, x_{1t}, x_{2t}, \dots, x_{Kt}$, then leave this row blank.

^e SCEP matches the first row of numerical data in this file with the first row in the YXDKEY file. The matching process continues to the second row, the third row, etc..

YXDATA2

←────────────────── 80 spaces ─────────────────→		
← DCW spaces →	...	← DCW spaces →
<i>x</i> -label <i>i</i> + 1	...	<i>x</i> -label <i>K</i>
<i>x</i> _{<i>i</i>+1,1}	...	<i>x</i> _{<i>K</i>1}
<i>x</i> _{<i>i</i>+1,2}	...	<i>x</i> _{<i>K</i>2}
<i>x</i> _{<i>i</i>+1,3}	...	<i>x</i> _{<i>K</i>3}
⋮	...	⋮
<i>x</i> _{<i>i</i>+1,<i>T</i>}	...	<i>x</i> _{<i>K</i><i>T</i>}

END OF DATA
//

ZDATA1

80 spaces			
DCW ^a spaces	DCW spaces	...	DCW ^b spaces
z-label 1	z-label 2	...	z-label <i>i</i>
1	z_{21}	...	z_{i1}
1	z_{22}	...	z_{i2}
1	z_{23}	...	z_{i3}
\vdots	\vdots	...	\vdots
1	z_{2T}	...	z_{iT}

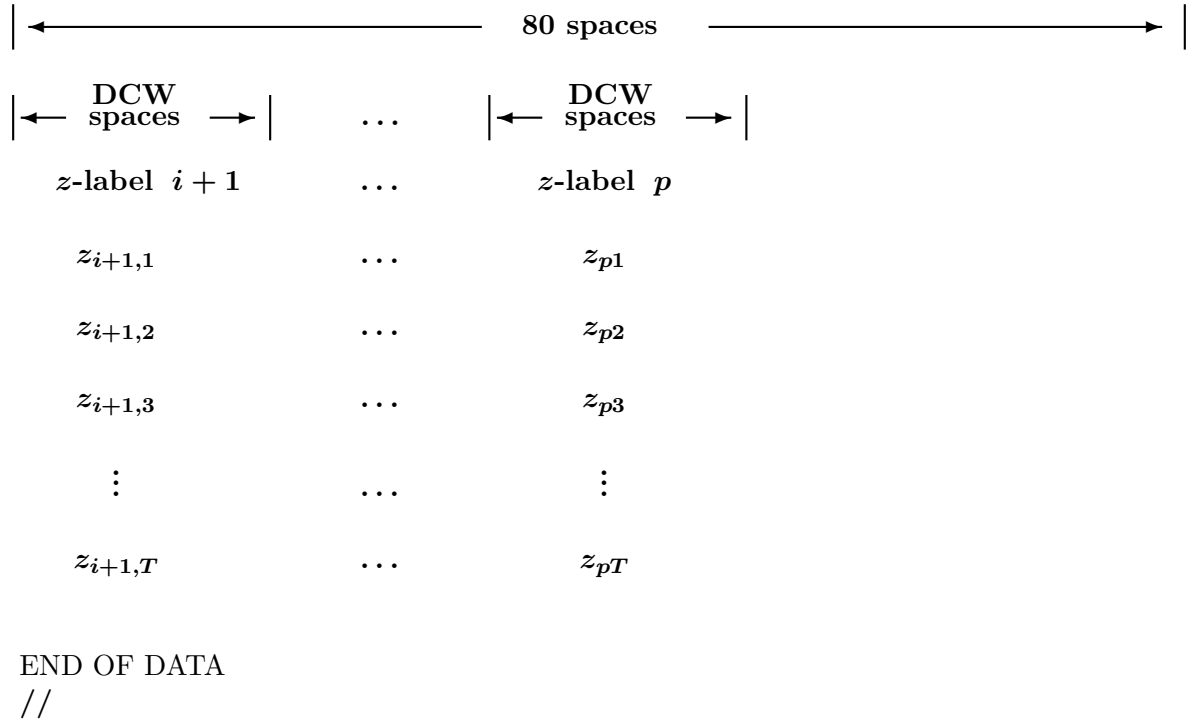
END OF DATA

//

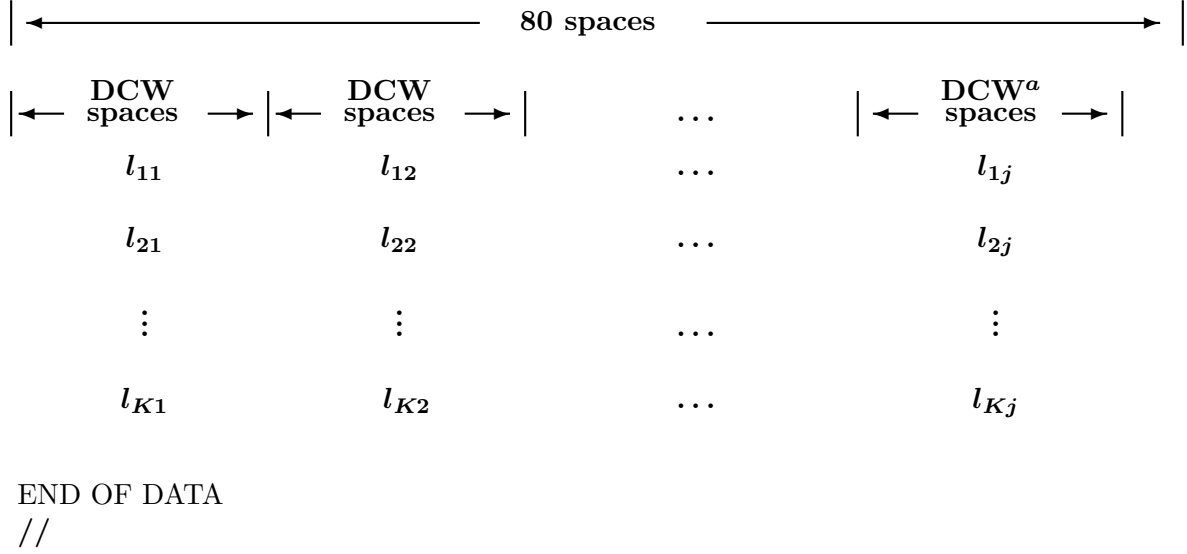
^aYou can eliminate this column by writing, "YES," on the right-hand side of the equation, AUTOMATIC CONSTANT 1 FOR X AND Z = , in the PARAMTRS file.

^bContinued in ZDATA2.

ZDATA2

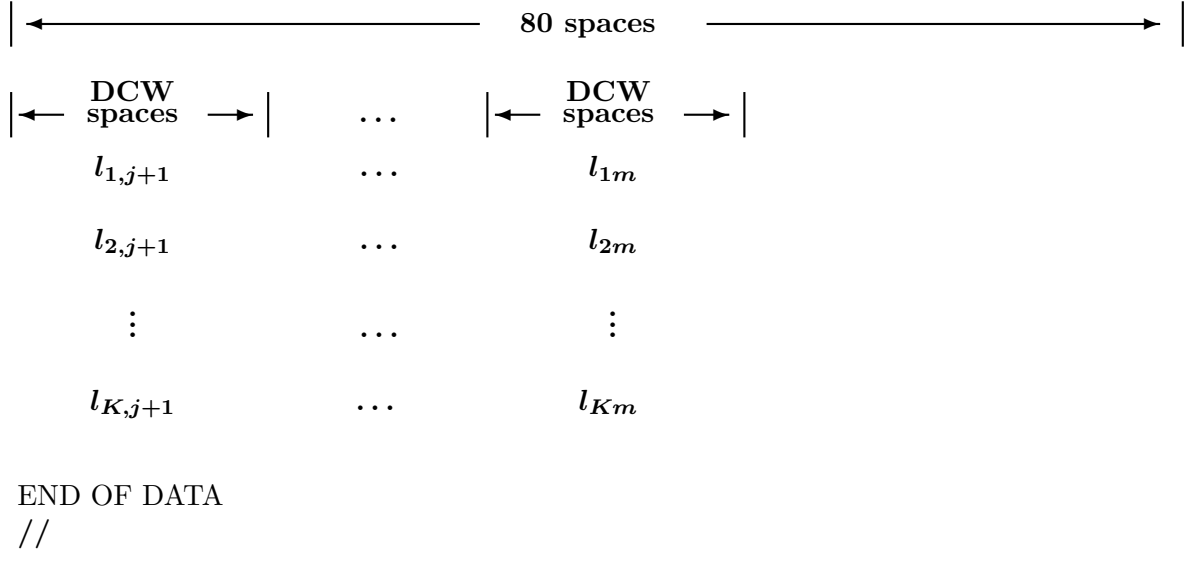


LDATA1

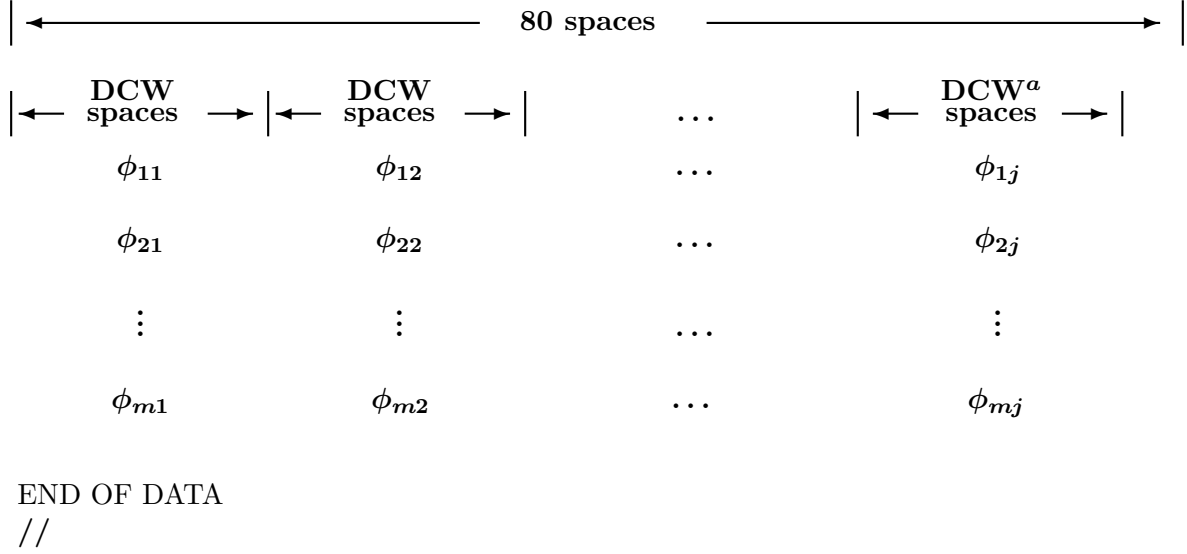


^aContinued in LDATA2.

LDATA2

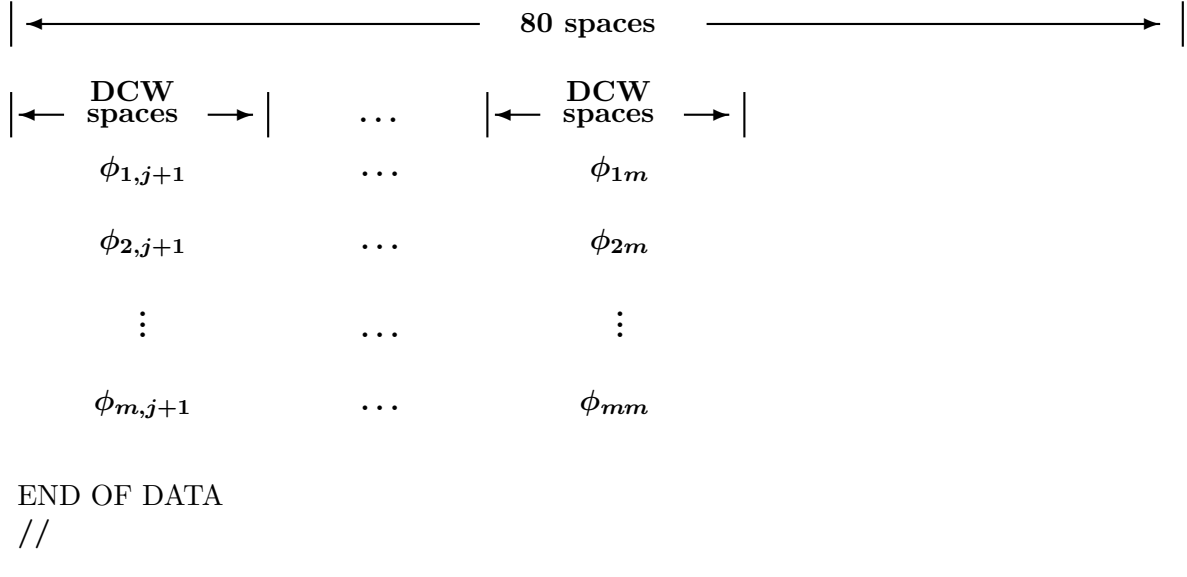


PDATA1



^aContinued in PDATA2.

PDATA2



DDATA1

<div style="display: flex; justify-content: space-between; align-items: center;"> ← 80 spaces → </div>			
<div style="display: flex; justify-content: center; align-items: center;"> ← DCW^a spaces → </div>	<div style="display: flex; justify-content: center; align-items: center;"> ← DCW spaces → </div>	...	<div style="display: flex; justify-content: center; align-items: center;"> ← DCW^b spaces → </div>
σ_a^2			
δ_{11}	δ_{12}	...	δ_{1j}
δ_{21}	δ_{22}	...	δ_{2j}
\vdots	\vdots	...	\vdots
δ_{m1}	δ_{m2}	...	δ_{mj}

END OF DATA
//

^aThe element at the intersection of the first row and the first column must be the value of σ_a^2 . You can use 1.0 as the value of σ_a^2 .

^bContinued in DDATA2.

DDATA2

<div style="display: flex; justify-content: space-between; align-items: center;"> ← 80 spaces → </div>		
<div style="display: flex; justify-content: center; align-items: center;"> ← DCW spaces → </div>	...	<div style="display: flex; justify-content: center; align-items: center;"> ← DCW spaces → </div>
$\delta_{1,j+1}$...	δ_{1m}
$\delta_{2,j+1}$...	δ_{2m}
\vdots	...	\vdots
$\delta_{m,j+1}$...	δ_{mm}

END OF DATA
//

WRITING RESTRICTION MATRICES TO FILES

The restrictions defined by equations in (5) can be imposed by setting up the following files:

- If you entered a positive integer on the right-hand side of the equation, NUMBER OF ROWS IN LINEAR EQUALITY RESTRICTIONS ON PHI =, in the PARAMTRS file, then you need to set up the RCFPH and RVCPI files.
- If you entered a positive integer on the right-hand side of the equation, NUMBER OF ROWS IN LINEAR EQUALITY RESTRICTIONS ON PI =, in the PARAMTRS file, then you need to set up the RCFPI and RVCPI files.
- If you entered a positive integer on the right-hand side of the equation, NUMBER OF ROWS IN LINEAR INEQUALITY RESTRICTIONS ON PI =, in the PARAMTRS file, then you need to set up the RCFPI and RVCPI files.
- **The file RCFPI contains data for the composite matrix**

$$\begin{bmatrix} R_{\Pi}^e \\ R_{\Pi}^i \end{bmatrix}$$

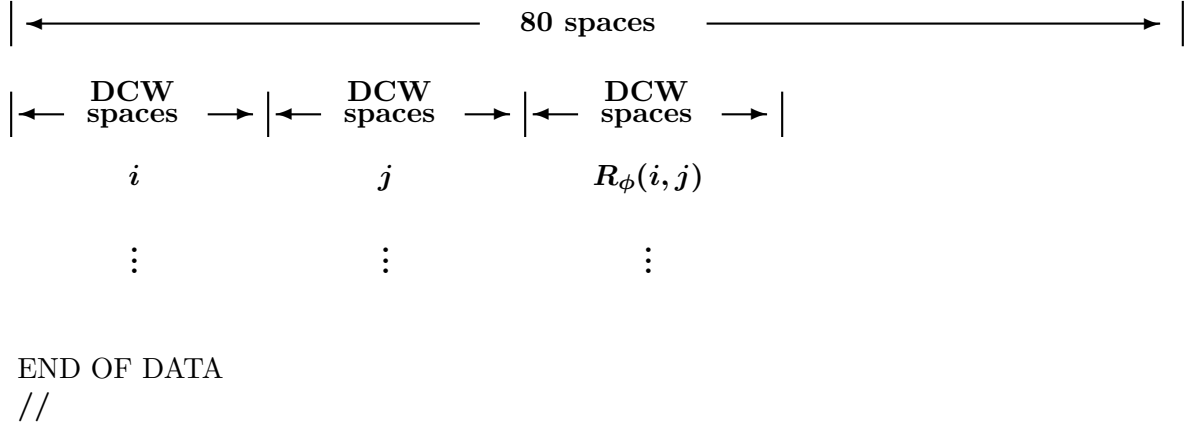
- **The file RVCPI contains data for the composite vector**

$$\begin{bmatrix} r_{\Pi}^e \\ r_{\Pi}^i \end{bmatrix}$$

- If you entered a positive integer on the right-hand side of the equation, NUMBER OF ROWS IN LINEAR EQUALITY RESTRICTIONS ON DELTA =, in the PARAMTRS file, then you need to set up the RCFDT and RVCDT files.

The RCFPH file specifies the number of each restriction imposed on the elements of \mathbf{R}_{Φ} (see (5)) and the numbers of the rows of \mathbf{R}_{Φ} in which these elements appear and the coefficients of these elements in the restriction. For example, if the i th restriction says that the element in the j th row of \mathbf{R}_{Φ} , denoted by $\mathbf{R}_{\Phi}(j)$, is zero, then enter the value of i in the first column, the value of j in the second column, and the value 1.0 in the third column of the RCFPH file. (Again, there is no restriction on the position of a value within a column.)

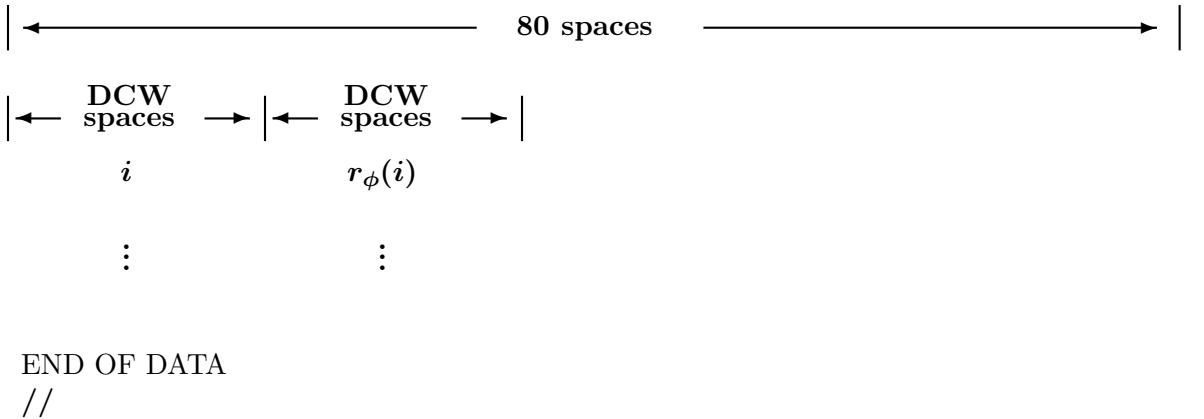
RCFPH



Here i and j should be entered as integers and $R_\Phi(j)$ as a floating-point number even when it is an integer. For example, the value 1 of the coefficient of $R_\Phi(j)$ should be entered as **1.0**. It should also be noted that corresponding to each restriction, there will be as many rows in the RCFPH file as there are elements of R_Φ in the restriction with nonzero coefficients. For example, if the r th restriction says that the i th and j th elements of R_Φ are equal, i.e., that $R_\Phi(i) - R_\Phi(j) = 0$, then there will be two rows in the RCFPH file corresponding to this restriction. The value of r appears in both the rows of the first column, the two rows of the second column contain the positions of $R_\Phi(i)$ and $R_\Phi(j)$ in R_Φ , respectively, and the two rows of the third column contain **1.0** and **-1.0**, respectively. Consequently, the number of rows of the RCFPH file may exceed that of R_Φ . The values of i should be entered in the RCFPH file in the same order in which the rows of R_Φ are written in (5).

The RVCPH file specifies the elements of \mathbf{r}_Φ defined in (?). If $\mathbf{r}_\Phi(i)$ denotes the i th element of \mathbf{r}_Φ , then the RVCPH file looks as follows:

RVCPH



Here i is an integer and $\mathbf{r}_{\Phi}(i)$ is a floating-point number. The number of rows of the RVCPH file will be equal to ℓ_{Φ} .

The procedure of setting up the RCFPI and RVCPI files is analogous to that of RCFPH and RVCPH, respectively. The procedure of setting up the RCFDT and RVCDT files is also analogous to that of RCFPH and RVCPH, respectively, but in setting up RCFDT and RVCDT, you should remember that \mathbf{R}_{Δ_a} and \mathbf{r}_{Δ_a} refer only to the diagonal and below diagonal elements of Δ_a .

SAVE FILES

You have the option of either saving or not saving the estimate of Φ_g computed in the last iteration of the calculation of a local model. If you want to save this estimate, then write YES on the right-hand side of the equation, SAVE PHI MATRIX =, in the PARAMTRS file. This estimate will be written to the PSAVE.(GROUP ID) file. In a subsequent run, you can use this estimate as the initial value of Φ_g by writing SWAMSLEY SAVE-FILE on the right-hand side of the equation, INITIAL PHI MATRIX =, in the PARAMTRS file.

Similarly, your entry, “YES,” on the right-hand side of the equation, SAVE DELTA MATRIX =, in the PARAMTRS file tells SCEP to write the estimate of Δ_{ag} computed in the last iteration of a run to the DSAVE.(GROUP ID) file. You can use this estimate as the initial value of Δ_{ag} in a subsequent run by writing SWAMSLEY SAVE-FILE on the right-hand side of the equation, INITIAL DELTA MATRIX =, in the PARAMTRS file.

You also have the option of saving or not saving all the computed β_{tg} of the local models and the global model. The forecasts of β_{tg} are not saved in this version of SCEP. If you write YES on the right-hand side of the equation, SAVE BETA =, in the PARAMTRS file, then SCEP writes all the β_{tg} (all groups, all iterations) of the local models to the BSAVE.(GROUP ID) files, in binary format. The β_{tg} of the global model are saved in the files BETAVECP.(GROUP ID), also in binary format.

4 How to Execute SCEP?

SCEP is a collection of several fortran routines. The steps required to execute SCEP on a Intel PC system under Windows 2000 or Windows XP are given below. The main program is called SCEPMAIN.FOR. A compiler is used to compile the routines to obtain an module with a name such as SCEPMAIN.OBJ. A linker is then invoked to produce an executable module with a name such as SCEPMAIN.EXE. This executable module is then run in a directory where all the necessary data files reside. The program creates two output files, SCEPOUT and SCEPLOG. The file SCEPOUT contains all the results on the computed model. The file SCEPLOG contains working messages from the program on the operational status of each of the file used by the program. Messages in SCEPLOG are mainly for checking the integrity of the run and for debugging problems related to input or output encountered in a run.

5 Five Levels of Output

The size of the output file SCEPOUT depends on the number you wrote on the right-hand side of the equation, PRINT LEVEL =, in the PARAMTRS file. You may write in here any one of the integers 1-5, for there are five levels of output. We describe each of these levels below. We use the notations adopted in the flow diagram given in the next section to indicate how the results are sequenced in the output. The printing of most of the results is performed by printing subroutines that have names of the form PR****.

PRINT LEVEL

THE RESULTS PRINTED

1, 2, 3, 4, 5

●the parameter values and options read in from the PARAMTRS file;



IF (SPIRWU .EQ. NEQPIU) THEN

The condition for this IF-block is that there is no linear inequality restriction imposed on the $\mathbf{\Pi}$ matrix for each group.

If linear equality restriction is to be imposed on the $\mathbf{\Pi}$ matrix of each group, the restriction is of the form $\mathbf{R}^* \text{vec}(\mathbf{\Pi}) = \mathbf{r}$. In the pooled calculation, a matrix \mathbf{C} with the property that $\mathbf{RC}' = \mathbf{0}$ and a vector $\mathbf{v} = \mathbf{R}'(\mathbf{RR}')^{-1}\mathbf{r}$ are computed. If no linear equality is to be imposed on $\mathbf{\Pi}$, the matrix \mathbf{C} is set to an identity matrix and the vector \mathbf{v} is set to the zero vector.

1, 2, 3, 4, 5

●the numerical rank of the restriction matrix \mathbf{R} ;
●the matrix \mathbf{C} and the vector \mathbf{v} ;

ENDIF



DO 1 IG=1,NGROUP

In each loop, a local stochastic coefficient model is computed for the IGth group. The end of this loop is marked by the statement **1 CONTINUE**.

1, 2, 3, 4, 5

- the data read in from the YXDATA1 and YXDATA2 files for the estimation period;
- the data read in from the YXDATA1 and YXDATA2 files for the forecast period (if supplied);
- the data read in from the ZDATA1 and ZDATA2 files for the estimation period (if supplied);
- the data read in from the ZDATA1 and ZDATA2 files for the forecast period (if supplied);
- the L -matrix used;
- the linear restriction matrices and vectors on Π , Φ , and Δ_a (if supplied);



1, 2, 3, 4, 5

- the initial value of Φ used;
- whether an adjustment of this value was necessary to make its eigenvalues less than one in absolute value;
- the lengths of the eigenvalues of the initial value of Φ before it was adjusted;
- the method of adjusting the initial value if adjustments were needed;
- the adjusted initial value if adjustments were made;
- the lengths of the eigenvalues of the adjusted initial value if adjustments were made;



1, 2, 3, 4, 5

- the initial value of Δ_a used;
- the value of the initial σ_a^2 used if it was different from the default value;
- whether the initial value of Δ_a was adjusted to make it non-negative definite;



3, 4, 5

- the eigenvalues of the unadjusted initial Δ_a if adjustments were to be made;
- the adjusted initial Δ_a if adjustments were made;
- the eigenvalues of the adjusted initial Δ_a if adjustments were made;



- 5 ●the observation vectors on the \mathbf{x}_{jt} premultiplied by \mathbf{L}' for the estimation period;
- the observation vectors on the \mathbf{x}_{jt} premultiplied by \mathbf{L}' for the forecast period;



- 5 ●the observation vectors on $(\mathbf{z}'_t \otimes \mathbf{x}'_t)$;



ITER = 0

1 IF (ITER .LE. MAXITR) THEN

The condition of this IF-block is that the iteration count ITER has not exceeded the upper limit MAXITR. This IF-block is a loop where a local stochastic coefficient model for the IGth group is iteratively refined. The end of this loop is marked by the three statements **ITER = ITER +1, GO TO 1, ENDIF**.



- 3, 4, 5 ●the method used to estimate Φ (if this iteration is not the initial iteration);
- an estimate of Φ (if this iteration is not the initial iteration);



- 3, 4, 5 ●the method used to estimate Δ_a (if this iteration is not the initial iteration);
- an estimate of Δ_a (if this iteration is not the initial iteration);



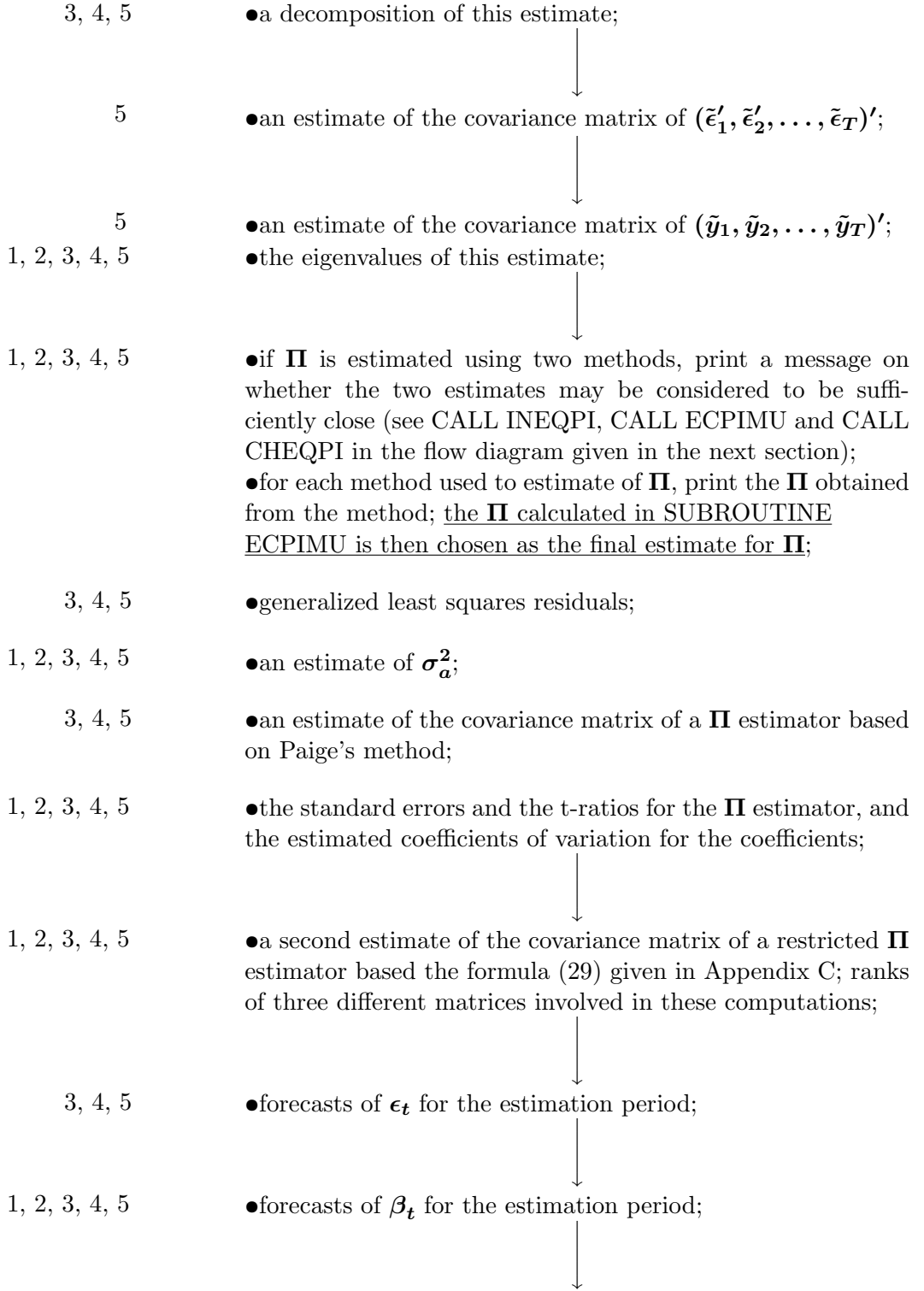
- 1, 2, 3, 4, 5 ●whether this estimate of Δ_a was adjusted to make it non-negative definite;

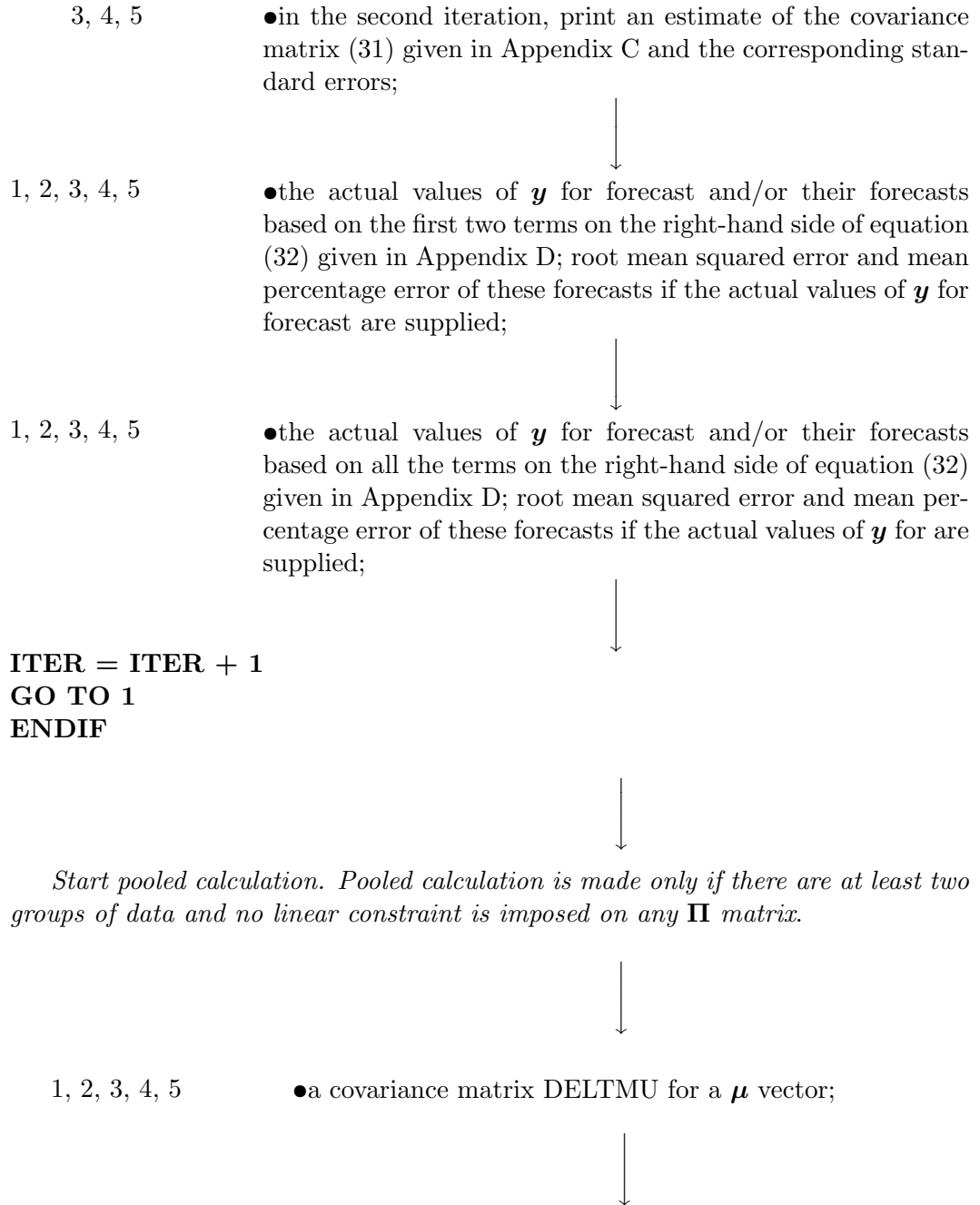


- 3, 4, 5 ●the eigenvalues of the unadjusted estimate of Δ_a if adjustment were made;
- an adjusted estimate of Δ_a matrix if adjustments were made;
- the eigenvalues of the adjusted estimate of Δ_a if adjustments were made;



- 4, 5 ●an estimate of the covariance matrix of $\tilde{\epsilon}_1$;





1, 2, 3, 4, 5

- a message stating that no adjustment was made on DELTMU if the matrix is already non-negative definite;
- the adjusted DELTMU if adjustments were made on DELTMU;
- the eigenvalues of DELTMU if adjustments were made on DELTMU;
- an upper Cholesky factor of DELTMU if adjustments were made on DELTMU;



1, 2, 3, 4, 5

- the pooled estimate for the Π matrix;
- an estimate of the covariance matrix for Π ;
- the standard errors and the t -ratios for the Π estimator, and the estimated coefficient of variation for the coefficients;



1, 2, 3, 4, 5

- for each group of data, print the components in the updated stochastic coefficient model for the group: the vector $\hat{\mu}_g$, the vectors $\bar{\epsilon}_{tg}$ and the vectors $\bar{\beta}_{tg}$. See formula *** in Section 2.2 of this manual.



1, 2, 3, 4, 5

- for each group of data, print pooled forecasts of ϵ_t , pooled forecasts of β_t ; print the updated forecasts for y together with any actual y -values supplied by the user; print updated error measurements for the forecasts.

6 Computational Steps Coded in SCEP

MAIN PROGRAM (SCEPMAIN.FOR)

SCEP follows the computational scheme described in this section. The code is written in Fortran 77. The file name for the main program is SCEPMAIN.FOR.

The first part of the main program contains declarations. Variables that end with the letter “u” are in general global variables which remain constant from one group of data to another group of data. Variables that are to be used for specific group of data are mostly variables in an arrays. The names of such an arrays usually contain the letter “G”. For instance, the array declaration of REAL*8 NTOTG(150) allocates one hundred and fifty variables for the purpose of saving the number (the count) of observations for the Ith group in the variable NTOTG(I).



Following the declarations, values are assigned to several variables for controlling the amount of memory to be used: MAXWR, MAXWI, MAXWC, MAXKEY, MAXNGR, MAXGID, MAXITU, HIPHRU, HIPIRU, HIDTRU. Further such initializations are made at several subsequent points of SCEPMAIN, assigning memory sizes to the variables SBATCH, LDRPHU, LDRDTU, LDRPIU, LDMRPI, LDMXDB, LDMZDB, LDMLU, LDORCU, LDMDLM, LDMSCV, LDSBTB, LDMPMP.



SCEP then commences to construct components for the model through sequences of subroutine calls described below. A call to a major subroutine is usually preceded by a call to a subprogram with a name of the form AL****, and followed by a call to a subprogram with name of the form PR****. An AL-subprogram allocates memory and initializes switching variables for the major subroutine that follows. A PR-subprogram prints the results computed by the preceding major subroutine(s). The first subroutine called is Subroutine SETFA.

CALL SETFA

Assigned file names and file unit numbers to disk files that are global. This set of files are shared by all groups.



Open the Files SCEPOUT and SCEPLOG. Rewind the files.



CALL GETPRM

Get the basic parameters and options from the file PARAMTR.

CALL PRTPRM**CALL GETKEY**

Read in data keys from the file YXDKEY. Sort the keys and generate counts to give information on the grouping structure of the data set.

**CALL SETFB**

Assign file names and file unit numbers to disk files that are used by individual groups.

**CALL IDXFC**

Create arrays of pointers and indices for the subsequent calculation and forecast within each group.

**CALL ALCSDT****CALL SETDAT**

Get the numerical data and separate the data into groups. A set of data files is created for each group.

**CALL OPGSET**

Process options specified for selected groups.



IF (SPIRWU .EQ. NEQPIU) THEN

The condition for executing this IF-block is that there is no linear inequality restriction imposed on the Π matrix within each group. In this case, the final pooled calculation is made. The following matrix ORCU and vector TRIVRU are needed in the pool calculation. If linear inequality constraints are imposed on the Π matrix in each group, no pooled calculation will be performed.



CALL ALPORC

CALL ORCRRP

Calculate the matrix ORCU and the vector TRIVRU that will be needed in final pooled calculation.

CALL PRPOCR



ENDIF



Initialize leading dimensions such as LDMDLM for use in SUBROUTINE DLTMDUP within SUBROUTINE CGROUP.



DO 1 IG=1,NGROUP

In each loop, a local stochastic coefficient model is computed for the IGth group.



CALL ALCCGR

Allocate memory and set up options for the computation of a stochastic coefficient model for the IGth group. Calculate upper bounds for several memory requirements. If any upper bound exceeds the allocation size specified in the main program SCEPMAIN, program execution is halted.

CALL CGROUP

SUBROUTINE CGROUP computes a stochastic coefficient model the IGth group. This subroutine is the main computation component of SCEP. The steps executed by CGROUP are described in a separate section that is to follow.



1 CONTINUE



IF ((NGROUP .GT. 1) .AND. (SPIRWU .EQ. NEQPIU)) THEN

The condition for the execution of this IF-block is that there are at least two groups and no linear inequality constraint is imposed on the Π matrix in each group.



CALL RWRITE

Print a covariance matrix DELTMU for a μ -vector. The covariance matrix is to be used in the pooled calculation. This covariance matrix is computed in SUBROUTINE DLTMUP (see the flow diagram for SUBROUTINE CGROUP).



CALL ALPADS

CALL ADPSYM

Adjust the covariance matrix DLTMUP for non-negative definiteness and determine an upper Cholesky factor for the matrix.

CALL PRPADS



DO 5 IG=1,NGROUP

For each group, calculate vectors and matrices which are basic components of the pooled estimates.



CALL ALPXZY

CALL CDXZYP

Calculate components for the $\hat{\Pi}$ matrix, the vectors $\bar{\epsilon}_{tg}$ and the vectors $\bar{\mu}_g$. Calculate the vector $\bar{\beta}$. See formula *** of Section 2.2 in this manual for the roles of such matrix and vectors.



5 CONTINUE



CALL ALPPIC

CALL CPICVP

Calculate the pooled estimate for $\hat{\Pi}$ matrix and an estimate for its covariance matrix.

CALL PRPPIC



DO 10 IG=1,NGROUP

Calculate and print pooled results for each group.



CALL ALPEMB

CALL EMBTAP

Calculate the vector $\hat{\mu}_g$, the vectors $\bar{\epsilon}_{tg}$ and the vectors $\bar{\beta}_{tg}$. See formula *** in Section 2.2 of this manual for the roles of the various vectors.

CALL PRPEMB



CALL ALCPFC

CALL PFCAST

Calculate the requested forecasts of y_t based on the pooled results for this group. Compute the root mean square error and the mean absolute error if the user has provided actual y_t values.

CALL PRTPFC



10 CONTINUE



CALL DELETF

Delete designated disk files such as temporary files that were used for intermediate calculations.



Close the files SCEPOUT and SCEPLOG.



ENDIF

This is the end of the IF-block where the pooled estimates are calculated.



STOP
END

COMPUTATIONAL STEPS IN SUBROUTINE CGROUP (FILE CGROUP.FOR)

Declaration of variables and assign leading dimensions to arrays.



Give comments for each significant variable.



Declare local variables and give comments on most of the local variables.



IF (SPIRWU .GT. 0) THEN

The condition for this IF-block is that linear restriction is to be imposed on the matrix Π .



CALL INIRPI

Reset a system of parameters that will be used if linear constraints are to be imposed on the Π matrix. This set of parameters may have been changed by the previous call to SUBROUTINE INEQPI.



ENDIF



CALL GETDTA

Get numerical data for the computation of the stochastic coefficient model for this group.

CALL PRTGDT



CALL ALCINP

CALL INTPHI

Initialize the Φ matrix either by reading values from a disk file or setting Φ to the zero matrix. Modify the matrix, if necessary, so that the eigenvalues of the matrix have moduli not exceeding a fixed value EVLTOL. EVLTOL is a positive value less than 1.0 provided by the user in the file PARAMTR. The results of this initialization are printed.

**CALL ALCIND****CALL INTDLT**

Initialize the Δ_a matrix by either reading values from a disk file or by setting Δ_a to the identity matrix. Modify the matrix, if necessary, so that the eigenvalues of the matrix have are non-negative. The results of this initialization are printed.

**CALL ALCTLX****CALL CTLXS**

Compute a matrix of the form $L'X$ for the computation described in equation **** in Section 2.1 of this manual. Each column of the matrix X is a vector x_t .

IF (NUMXF .GT. 0) THEN

The condition of this IF-block is that data have been provided for the forecasting of y_t .

CALL CTLXS

Compute a matrix of the form $L'X$ for calculating the forecasts of y_t .

ENDIF**CALL PRTTLX****CALL ALCXZ****CALL CXZ**

Compute the matrix XZ having $z'_t \otimes x'_t$ as its t th row.

CALL PRTXZ

↓

ITER = 0

↓

1 IF (ITER .LE. MAXITR) THEN

The condition of this IF-block is that the iteration count ITER has not exceeded the upper limit MAXITR. This IF-block is a loop where a local stochastic coefficient model for the IGth group is iteratively refined. The end of this loop is marked by the three statements **ITER = ITER +1, GO TO 1, ENDIF**.

↓

CALL PGRESS

Show the progress of the iteration on the monitor screen.

↓

IF ((SPIRWU .GT. 0) .AND. (SPIRWU .GT. NEQPIU)) THEN

The condition for this IF-block is that linear inequality constraints have been imposed on the **Π** matrix.

↓

CALL INIRPI

Reset a system of parameters to be used when applying linear constraints to the **Π** matrix. This set of parameters may have been changed by SUBROUTINE INEQPI in the previous loop.

↓

ENDIF

↓

IF (ITER .GT. 0) THEN

The condition for this IF-block is that the iteration counter is 1 or higher. This loop is not the initial loop where ITER is 0. Within this IF-block,

calculate the Φ -matrix using one of the available algorithms. Algorithms available are “LM” for a Levenberg–Maquart method, “QP” for a quadratic programming, and “RN” for a monte carlo method. The character string ALGCPH contains the designation “LM”, “QP” or “RN”. The matrix Δ_a is then computed in the same manner, using one of the three available methods.



CALL HOWCPH

Assign either “LM”, ”QP” or “RN” to the character string ALGCPH, according to a strategy implemented in this subroutine.



IF (INDEX(ALGCPH, 'LM') .GT. 0) THEN

The condition for this IF-block is that the method “LM” has been chosen.



CALL ALLMPH

CALL LMCPI

Calculate the matrix Φ using a Levenberg–Marquardt method.

CALL PRLMPH



ENDIF



IF (INDEX(ALGCPH, 'QP') .GT. 0) THEN

The condition for this IF-block is that the method “QP” has been chosen.

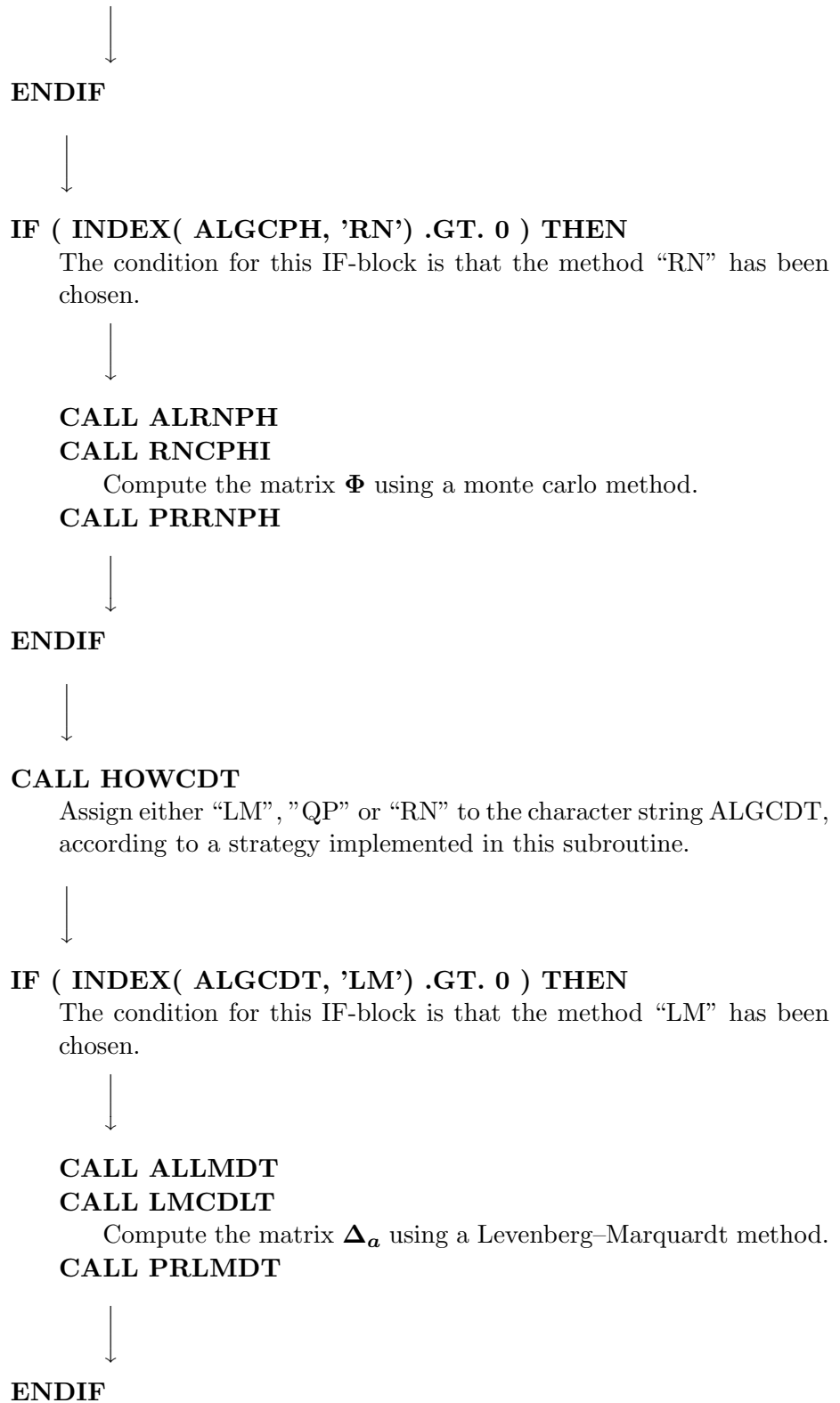


CALL ALQPPH

CALL QPCPI

Compute the matrix Φ using a method of quadratic programming.

CALL PRQPPH



↓

IF (INDEX(ALGCDT, 'QP') .GT. 0) THEN

The condition for this IF-block is that the method “QP” has been chosen.

↓

CALL ALQPDT

CALL QPCDLT

Compute the matrix Δ_a using a quadratic programming method.

CALL PRQPDT

↓

ENDIF

↓

IF (INDEX(ALGCDT, 'RN') .GT. 0) THEN

The condition for this IF-block is that the method “RN” has been chosen.

↓

CALL ALRNDT

CALL RNCDLT

Compute the matrix Δ_a using a monte carlo method.

CALL PRRNDT

↓

ENDIF

↓

IF ((.NOT. OLS) .AND. (.NOT. GLS)) THEN

The condition for this IF-block is that the stochastic coefficient model to be computed is neither an OLS model nor an AR(1) model.

↓

IF (.NOT. DDIAG) THEN

The condition for this IF-block is that Δ_a is not to be a diagonal matrix.

↓

CALL ALCADN

CALL ADJDND

Modify the computed Δ_a matrix, if necessary, to a non-negative definite matrix. Compute an upper Cholesky factor of Δ_a .

CALL PRTADN

↓

ELSE

Otherwise Δ_a is to be a diagonal matrix.

↓

CALL ALCADD

CALL ADJDD

Modify the diagonal matrix Δ_a , if necessary, to a non-negative definite matrix. Compute an upper Cholesky factor of Δ_a .

CALL PRTADD

↓

ENDIF

↓

ELSE

The ELSE-condition here is that the stochastic coefficient model is to be an OLS model or an AR(1) model.

↓

CALL MZERO

Set the upper Cholesky factor for the matrix Δ_a to the zero matrix.

↓

IF (DELTA(1,1) .LT. 0.0D0) THEN

If the element DELTA(1,1) of Δ_a is negative, set DELTA(1,1) to zero and print a message.

ENDIF



UDELTA(1,1) = SQRT(DELTA(1,1))

Assign the square root of DELTA(1,1) to the element UDELTA(1,1)
in the upper Cholesky factor for Δ_a .



ENDIF

This is the end of the IF-block of “IF ((.NOT. OLS) .AND. (.NOT.
GLS)) THEN”



ENDIF

This is the end of the IF-block of “IF (ITER .GT. 0) THEN”.



CALL ALCGMA

CALL CGAMMA

Compute an estimate for the matrix Γ_0 .

CALL PRTGMA



CALL ALCYE

CALL CSIGYE

Compute estimates for the matrices Σ_ϵ and Σ_y . Calculate an upper
Cholesky factor for the computed Σ_y .

CALL PRTSGE

CALL PR TYSG



IF (ITER .EQ. 0) THEN

The condition for this IF-block is that the present iteration is the first
iteration – ITER is 0.



CALL SAVYSG

Save the estimate of Σ_y in a temporary memory block YSGMA0. This estimate will be used later when evaluating a covariance matrix of the weighted or unweighted least squares estimate of the Π -matrix in the second iteration (ITER = 1).

↓

ENDIF

↓

IF (ITER .EQ. 1) THEN

↓

CALL SAVYSG

Save an upper Cholesky factor for Σ_y in a temporary memory block YSGMA1.

↓

ENDIF

↓

IF (SPIRWU .GT. NEQPIU) THEN

The condition for this IF-block is that linear inequality constraints have been imposed on the Π -matrix.

↓

CALL ALCEIP

↓

CALL INEQPI

Calculate a Π -matrix using a quadratic programming method. The linear constraints include possible equality constraints and at least one linear inequality constraint. At the end of the computation, the system of constraints may be modified to include only the linear constraints where equality has been achieved by the computed Π -matrix. This step prepares for the subsequent covariance calculation

for the Π -matrix.

↓

ENDIF

↓

CALL ALEPMU

CALL ECPIMU

Calculate a Π -matrix, perhaps for the second time in this loop. If linear restrictions are present, the constraints are only equality constraints. This system of equality constraints include those specified by and user and possibly several specified by SUBROUTINE INEQPI in the previous step of the computation of Π . The method used in this subroutine is Paige's method. Other estimates computed are σ_a^2 , a covariance matrix for Π , and the residual vectors $y_t - x_t' \Pi z_t$.

↓

IF (SPIRWU .GT. NEQPIU) THEN

The condition for this IF-block is that linear inequality constraints are present for the matrix Π .

↓

CALL CHEQPI

The Π -matrix computed in SUBROUTINE INEQPI and the Π -matrix computed in SUBROUTINE ECPIMU are expected to be equal to within the range of round-off errors. SUBROUTINE CHEQPI check for such equality. The results of the checks are printed.

↓

ENDIF

↓

CALL SQCVAR

Multiply the estimate of the covariance matrix of the generalized least squares estimate of Π by the estimate of σ_a^2 . The estimate of Π calculated in SUBROUTINE ECPIMU was formed without the factor σ_a^2 .

↓

CALL PRTPMU

↓

CALL ALCORC

CALL ORCVAR

Use a method of orthogonal decomposition to compute another estimate for the covariance matrix associated with $\mathbf{\Pi}$.

CALL PRTORC

↓

CALL ALCEPS

CALL CEPS

Compute the forecasts of ϵ_t . See equation **** of section 2.1 in this manual.

CALL PRTEPS

↓

IF ((NGROUP .GT. 1) .AND. (ITER .EQ. MAXITR) .AND.

(SPIRWU .EQ. NEQPIU)) THEN

The three conditions for this IF-block are: more than one group is involved in this model, this iteration is the last iteration, and no linear inequality constraint is imposed on $\mathbf{\Pi}$.

↓

CALL SVSG2G

Save the σ_a^2 computed in the last iteration for use in the subsequent pooled calculation.

↓

CALL ALPDMU

CALL DLTMUP

Accumulate matrices to form the matrix DELTMU, to be used in

the pooled calculation.



ENDIF



CALL ALCBTA

CALL CBETA

Compute forecasts of β_t . See equation **** in Section 2.1 of this manual.

CALL PRTBTA



IF (SAVBTA) THEN

The condition for this IF-block is that the vectors β_t computed in this iteration is to be saved on a disk file.



CALL PUTBTA

Write the vectors β_t computed in this iteration to a disk file.



ENDIF



IF (ITER .EQ. 1) THEN

The condition for this IF-block is that this loop is the second loop of the iteration ($ITER = 1$).



CALL ALCSCV

CALL SCOVAR

Compute a special covariance matrix associated with Π .

CALL PRTSCV



ENDIF



IF (NUMXF .GT. 0) THEN

The condition for this IF-block is that forecasts of y_t have been requested.



CALL ALCYFC

CALL YFCAST

Compute forecasts of y_t . If the user supplies actual values of y_t for the forecast period, compute also the root mean square error and the mean absolute error.

CALL PRTYFC



ENDIF



ITER = ITER + 1



GO TO 1



ENDIF



CALL PUTPHI

Write the last estimate of Φ to a disk file.



IF (SAVDLT) THEN

The condition for this IF-block is that the user wants to save the last estimate of Δ_a .



CALL PUTDLT

Save the last estimates of σ_a^2 and Δ_a to a disk file.



ENDIF



CALL FDELET

Delete temporary files used by SUBROUTINE CGROUP in this call to the subroutine.



RETURN

Return execution control to the main program SCEPMAIN.

END

7 An Example

In recent years there has been considerable interest in the issue of the effects of federal budget deficits on interest rates. To measure the effect of such deficits on a specific interest rate, the following model was applied to U.S. economic data:

$$\tilde{y}_t = \tilde{\beta}_{1t} + \tilde{\beta}_{2t}x_{2t} + \tilde{\beta}_{3t}z_{1t} + \tilde{\beta}_{4t}z_{2t}, \quad (10)$$

where \tilde{y}_t = the 3-month Treasury bill rate, $-x_{2t}$ = the ratio of net surplus of federal government (National Income and Product Accounts basis) to nominal gross national product, $z_{1t} = \sum_{j=0}^5 w_j PGNPDOT_{t-j}$, $PGNPDOT$ = the annual rate of change of the implicit deflator for GNP (1982 = 100), and z_{2t} = the annual rate of change of the narrow (M1) measure of money supply. This model is considered here because it was previously found to perform well in prediction over a broad range of conditions by Swamy, Kolluri and Singamsetti.⁹

Let $\tilde{\beta}_t = (\tilde{\beta}_{1t}, \tilde{\beta}_{2t}, \tilde{\beta}_{3t}, \tilde{\beta}_{4t})'$. Then for $t = 1, 2, \dots, T$ we assume that

$$\tilde{\beta}_t = \bar{\beta} + \tilde{\epsilon}_t, \quad (11)$$

where $\bar{\beta} = (\bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3, \bar{\beta}_4)'$, $\tilde{\epsilon}_t = (\tilde{\epsilon}_{1t}, \tilde{\epsilon}_{2t}, \tilde{\epsilon}_{3t}, \tilde{\epsilon}_{4t})'$ satisfies equation (3) with $m = 4$, and $\tilde{a}_t = (\tilde{a}_{1t}, \tilde{a}_{2t}, \tilde{a}_{3t}, \tilde{a}_{4t})'$. Equation (7) is the case $K = m = 4$, $L = I_4$, and all the elements of z_t other than the first element are equal to zero of equation (2).

Since we want to discern the effect on \tilde{y}_t of only \tilde{x}_{2t} , we follow Pratt and Schlaifer¹⁰ in assuming that \tilde{x}_{2t} is the only factor (simply put, a factor is a variable whose value can be manipulated) and \tilde{z}_{1t} and \tilde{z}_{2t} are concomitants (a concomitant is a variable, the manipulation of whose value does not make sense). In terms of these variables, equation (5) can be written as

$$\tilde{y}_t = (1, x_{2t}) \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \end{pmatrix} \begin{pmatrix} 1 \\ z_{1t} \\ z_{2t} \end{pmatrix} + (1, x_{2t}) \begin{pmatrix} \tilde{\epsilon}_{1t} \\ \tilde{\epsilon}_{2t} \end{pmatrix}, \quad (12)$$

where $x_t = (1, x_{2t})'$, $z_t = (1, z_{1t}, z_{2t})'$, $\tilde{y}_t, x_{2t}, z_{1t}$, and z_{2t} are as defined in equation (6), $K = 2$, $p = 3$, $m = 2$, $\Pi = (\pi_{ij})$ is a 2×3 matrix, and $L = I_2$.

Note that equation (6) is not an example of equation (1) but of an equation obtained by including concomitants in equation (1) rather than in equation (2). While equation (6) decomposes the coefficient on \tilde{x}_{2t} into $\bar{\beta}_2 + \tilde{\epsilon}_{2t}$, equation (8) decomposes the same coefficient into $\pi_{21} + \pi_{22}z_{1t} + \pi_{23}z_{2t} + \tilde{\epsilon}_{2t}$, as in equation (2), where π_{21} is the coefficient on x_{2t} and π_{22} and π_{23} are the coefficients on the interactions of x_{2t} with z_{1t} and z_{2t} , respectively. Since z_{1t} and z_{2t} are substantially influenced by the Federal Reserve's policies, equation (8) attempts to measure the

⁹Swamy, P.A.V.B., B.R. Kolluri and R.N. Singamsetti (1990), "What Do Regressions of Interest Rates on Deficits Imply?," *Southern Economic Journal*, 56, 1010-1028.

¹⁰Pratt, J.W. and R. Schlaifer (1984), "On the Nature and Discovery of Structure," *Journal of the American Statistical Association*, 79, 9-21, 29-33 and Pratt, J.W. and R. Schlaifer (1988), "On the Interpretation and Observation of Laws," *Journal of Econometrics, Annals*, 39, 23-52.

effect of budget deficits on a short-term interest rate, which is realizable in an economic environment influenced by the Federal Reserve.

Two viewpoints dominate most analyses of the effects of federal deficits on interest rates. The conventional view is that large federal deficits drive up interest rates. This view is supported by equation (6) if $\bar{\beta}_2$ is positive and does not contain any proxy effects and by equation (8) if π_{21} is positive and the sum $\pi_{22}z_{1t} + \pi_{23}z_{2t} + \bar{e}_{2t}$ accurately measures the proxy effects contained in $\bar{\beta}_{2t}$. The “Ricardian equivalence proposition” which challenges the conventional view, is supported by equation (6) if $\bar{\beta}_2$ measures the direct effect of \tilde{x}_{2t} on \tilde{y}_t and is equal to zero and by equation (8) if $\pi_{21} = 0$ and measures accurately the direct effect of \tilde{x}_{2t} on \tilde{y}_t .

To estimate and validate equations (6) and (8), we use Swamy et al.’s¹¹ quarterly data on the variables $\tilde{y}_t, \tilde{x}_{2t}, \tilde{z}_{1t}$, and \tilde{z}_{2t} for the U.S. for the period 1960:Q1-1986:Q4. These data are presented in Table 1. The data on \tilde{z}_{1t} were generated by using the estimates of a fixed- β model which performed better than several other fixed- β models in prediction. The total period, 1960:Q1-1986:Q4, was split into two subperiods, 1960:Q1-1983:Q4 and 1984:Q1-1986:Q4. The first subperiod was used as the fitting period and the second subperiod was used as the forecast period.

If assumption (3) is true, then there is the problem of precisely determining the true values of the nuisance parameters Φ and Δ_a . Obtaining a satisfactory solution to this problem is of first order importance because the nuisance parameters substantially affect the accuracy with which we can estimate (or predict) the parameters of interest which are the elements of $\bar{\beta}$ and Π (or the $\tilde{\beta}_t$). The absence of any strong beliefs about the true values of the nuisance parameters has led us to estimate them from data under a variety of judgments concerning simplification. These simplifications are described in Appendix B. We may be able to discover the usefulness of some of these simplifications after seeing how certain constrained and unconstrained estimates of $\bar{\beta}$, Π , Φ , and $\sigma_a^2 \Delta_a$ perform in terms of explanation and prediction in different periods. For each of several simplifications and for each of 12 quarters in the forecast period, the forecasts of y_t were obtained. To generate these forecasts, models (6) and (8) were estimated using SCEP and using all the 96 observations covering the fitting period. The results of these computations are presented in Tables 2-6.

The root mean square error (RMSE) $\left[\frac{1}{12} \sum_{s=1}^{12} (\text{a forecast of } y_{T+s} - y_{T+s})^2 \right]^{\frac{1}{2}}$ is used as a measure of the inaccuracy of the forecasts of y_{T+s} , $s = 1, 2, \dots, 12$, with T referring to the last quarter of the year 1983 and s referring to a quarter in the forecast period. The mean of the observations on \tilde{y}_t in the forecast period is 7.6590.

Different sequences of parameter estimates for models (6) and (8) under different restrictions on Φ and Δ_a were obtained by an *iterative* application of an operational version of the generalized least squares method, in which Swamy and Tinsley’s¹² estimates of Φ and Δ_a were used in place of their unknown population values. This iterative scheme converges if Φ and Δ_a are estimated subject to certain upper-bound constraints on the eigenvalues of Φ and the Frobenius norm of Δ_a . In practice we have taken its convergence to occur when no changes in the estimates of $\bar{\beta}$, t -ratios,

¹¹Swamy, Kolluri and Singamsetti, op. cit.

¹²Chang *et al.*, pp. 46-51.

σ_a^2 , and RMSE's printed upto 4 decimal places are observed. It is possible that the terminal estimates of $\bar{\beta}$, Φ , Δ_a , and σ_a^2 given by this iterative procedure are closest to the maximum likelihood estimates whenever the latter exist.

For estimation in models (6) and (8), we do not consider sequential methods because forecasting without sequential estimation is more appropriate for exposing and isolating potential weaknesses inherent in fixed- β modeling, since sequentially updated parameter estimates would cloud the distinction between *revisions* of estimates due to changes in sample values (like adding a recent data point with or without dropping an earlier one) involved in sequential estimation and the *behavior* of parameters driven by potentially time-dependent processes. Thus, we do not know whether we are estimating a fixed parameter or a variable parameter whenever we follow sequential methods to estimate it. Moreover, Swamy and Schinasi¹³ have shown that the mean square errors of predictors about actual realizations are not necessarily minimized when fixed- β models are sequentially re-estimated and used for prediction.

Tables 2-5 report estimates of the parameters of model (6) obtained in the following cases. (The symbol $\lambda_{\max}(\Phi)$ is used below to denote the maximum eigenvalue of Φ .)

Case 1. $\Phi = 0$, and for $i = 2, 3, 4$ and for all t , the random variable $\tilde{\epsilon}_{it}$ is degenerate at 0.

Case 2. $\Phi = 0$ and $\Delta_a = I$.

Case 3. $\Phi = 0$ and Δ_a is diagonal with nonnegative diagonal elements.

Case 4. $\Phi = 0$ and Δ_a is nondiagonal with an upper-bound constraint on its Frobenius norm.

Case 5. $(\tilde{\beta}_{1t} - \bar{\beta}_1) = \phi_{11}(\tilde{\beta}_{1t-1} - \bar{\beta}_1) + \tilde{a}_{1t}$, where $|\phi_{11}| < 1$, $E\tilde{a}_{1t} = 0$, $E\tilde{a}_{1t}\tilde{a}_{1s} = \sigma_a^2$ if $t = s$ and $= 0$ if $t \neq s$, and for $i = 2, 3, 4$ and all t , the random variable $\tilde{\epsilon}_{it}$ is degenerate at 0.

Case 6. Φ is diagonal with $|\lambda_{\max}(\Phi)| \leq 0.95, 0.99$, or 0.995 and Δ_a is diagonal with nonnegative diagonal elements.

Case 7. Φ is diagonal with $|\lambda_{\max}(\Phi)| \leq 0.99$ or 0.995 and Δ_a is nondiagonal with an upper-bound constraint on its Frobenius norm.

Case 8. Φ is nondiagonal with $|\lambda_{\max}(\Phi)| \leq 0.99$ or 0.995 and Δ_a is diagonal with nonnegative diagonal elements.

Case 9. Φ is nondiagonal with $|\lambda_{\max}(\Phi)| \leq 0.8, 0.95, 0.99$, or 0.995 and Δ_a is nondiagonal with an upper-bound constraint on its Frobenius norm.

The following remarks clarify our estimation procedure further.

1. Since the least-squares criterion is appropriate in Case 1, the first row of Table 2 gives the least squares estimates of the parameters of model (6). These estimates were used as starting values for generating a sequence of parameter estimates in Case 5. Note that in Case 5, model (6) reduces to a fixed- β model with first-order autoregressive (denoted by AR(1)) errors. Since SCEP is based on a derivative-free approach, the estimates given in Table 2 for Case

¹³Swamy, P.A.V.B. and G.J. Schinasi (1989), "Should Fixed Coefficients be Re-estimated Every Period for Extrapolation?" *Journal of Forecasting*, 8, 1-17.

- 5 are more accurate than those given by a nonlinear least squares program which uses analytical or numerical derivatives.
2. The estimates given in the second row of Table 2 were used as starting values for generating sequences of parameter estimates in Cases 3-9, excluding Case 5. Some of these sequences converged to either one point or several points in the parameter space. We indicate this by putting a superscript c on the numbers of the final iterations of these sequences. Initial iterations are labeled "0."
 3. In Case 6 with the restriction that $|\lambda_{\max}(\Phi)| \leq 0.99$, the sequence of parameter estimates generated by our iterative scheme converged to 5 different points. It can be seen from the entries in rows, labeled "6," of Table 2 that these points are not very far apart. These points did not change very much when the estimates obtained in Case 6 and iteration 3 were used as starting values. (Compare the estimates given in rows, labeled "6," of Tables 2 and 3.) When we imposed the restriction that $|\lambda_{\max}(\Phi)| \leq 0.995$, the sequence of parameter estimates generated by our iterative scheme converged to one point in the parameter space, as can be seen from Table 4.
 4. When a prespecified upper bound for the number of iterations was 30, convergence of our iterative scheme did not occur in Cases 8 and 9 (see Table 2). It occurred when $|\lambda_{\max}(\Phi)|$ is constrained not to lie very close to 1 (see Table 5).
 5. The estimates of σ_a^2 given in the last columns of Tables 2-5 do not measure the goodness of fit of equation (6) because the error covariance matrix of this equation is $\sigma_a^2 \Sigma_y$ which is shown in Appendix C. The smaller the estimates of the elements of Σ_y , the bigger the estimates of σ_a^2 and vice versa.

The entries in the RMSE columns of Tables 2-5 show that the points to which different estimate sequences converged do not provide the smallest RMSE's. A striking implication of our results (some of which are not presented in Tables 2-5) is that with a few exceptions, model (6), when fitted by an iterative application of Swamy and Tinsley's operational generalized least squares procedure subject to the restrictions that (i) Φ is diagonal with $|\lambda_{\max}(\Phi)| \leq 0.99$ and (ii) Δ_a is diagonal with nonnegative diagonal elements, has produced in every iteration starting with the second, a RMSE of forecasts which is smaller than that produced in any iteration in any other case we considered. The range of RMSE's obtained in iterations 2-49 in Case 6 with $|\lambda_{\max}(\Phi)| \leq 0.99$ is (0.3680-0.4843). We can conclude from Table 2 that the model given by equations (6) and (7) does not generally perform very well in prediction whenever the estimates of $\bar{\beta}_2$ are positive. Perhaps this result is a consequence of the fact that model (6) does not rule out negative values for the variable $\tilde{y}_t - \bar{\beta}_{3t} z_{1t}$ which can be interpreted as a real rate. Negative values for a real rate are meaningless. In any case, we cannot conclude from any of these results, however, that the conventional view concerning the effects of government deficits on

interest rates is incorrect because our assumption that $\bar{\beta}_2$ does not contain any terms contributing to proxy effects may be false.

Tables 6 and 7 report estimates of the parameters of model (8) for two different upper-bound constraints on $|\lambda_{\max}(\Phi)|$. It can be seen from the entries in these tables that several estimates of π_{21} including those which have yielded good forecasts are positive. Even though a sequence of estimates obtained in Case 6 converges to three points given in Table 7, it is inconclusive because it converges to both negative and positive estimates of π_{21} . The estimate sequence obtained in Case 7 converges to a point in the parameter space which performs well in prediction and the estimate of π_{21} implied by this point is positive. This result supports the conventional view if π_{21} measures accurately the direct effect of \tilde{x}_{2t} on \tilde{y}_t and if the t -ratio given in parenthesis below the estimate of π_{21} does not provide strong evidence against the hypothesis that $\pi_{21} \neq 0$.

Tables 2 and 6 might be compared to determine whether one of models (6) and (8) has a clear advantage over the other in terms of explanation and prediction. From the preceding discussion it follows that in making these comparisons, cases other than Case 6 for model (6) and Case 7 for model (8) can be ignored. Relative to model (8) in Case 7, the RMSE's for forecasts from model (6) in Case 6 are generally substantially lower. To the extent that $\bar{\beta}_2$ is not as good a measure of the direct effect of \tilde{x}_2 on \tilde{y} as π_{21} is, then this gain in predictive accuracy offered by model (6) in Case 6 occurs at the cost of a loss in explanatory power. The question of what form of equation (6) or (8) performs well in both explanation and prediction is still open. However, it appears that the inclusion in equation (2) of concomitants that absorb proxy effects can have a payoff in terms of explanatory power.

As a final remark, we point out that one out of 21 estimates of $\bar{\beta}_2$ given in Table 2 and one out of 17 estimates of π_{21} given in Table 6 are significantly different from zero on the convention that a t -ratio is greater than 2 in absolute value. This result does not imply that the variable \tilde{x}_{2t} must be excluded from equations (6) and (8). The concomitants that substantially reduce the omitted variables biases in the estimates of $\bar{\beta}_2$ and π_{21} are likely to be highly correlated with the \tilde{x}_{2t} and this high correlation results in low t -ratios. As Pratt and Schlaifer¹⁴ show, exclusion of a variable because its estimated coefficient is not significantly different from zero is wrong if the variable is a factor. It is better to include a concomitant that substantially reduces the bias in an estimate of a factor's effect than to exclude it even if its inclusion means reducing the t -ratio of the estimate.

¹⁴Pratt and Schlaifer 1984, *op. cit.*, p. 19.

Appendices

Appendix A: Estimation of Direct Effects—A Step-by-Step Argument

Question: *Why should any economist use SCEP?*

To answer this question, consider the usual case where an economist wants to estimate an equation relating \tilde{y}_t to $\tilde{x}_{1t}, \dots, \tilde{x}_{Kt}$. If such a relationship exists, then it can be represented by:

$$\tilde{y}_t = f_t(\tilde{x}_{1t}, \dots, \tilde{x}_{Kt}, \tilde{e}_{1t}, \tilde{e}_{2t}, \dots, \tilde{e}_{rt}), \quad (13)$$

where $\tilde{e}_{1t}, \tilde{e}_{2t}, \dots, \tilde{e}_{rt}$ represent unknown and/or unobservable variables that possibly affect \tilde{y}_t .

If, in empirical econometrics that relies on equation (9), no assumption other than equation (9) is needed, then econometrics is a user-friendly subject. Unfortunately, in this research, the following questions arise:

Set I: What are the conditions under which equation (9) exists? How can one verify these conditions?

Set II: What is the correct parametric form of f if equation (9) exists? If this form is unknown, as it usually is, can one specify it so that the specified function coincides with a stochastic law? What are the conditions under which such a coincidence occurs? How can one decide whether these conditions are or are not satisfied?

Set III: Are the \tilde{x}_{jt} independent of the \tilde{e}_{it} ? If they are not, then what is the right way to estimate and interpret the parameters of f ?

Set IV: Are the coefficients on $\tilde{x}_{1t}, \tilde{x}_{2t}, \dots, \tilde{x}_{Kt}, \tilde{e}_{1t}, \tilde{e}_{2t}, \dots, \tilde{e}_{rt}$ in the true parametric form of f constant?

Set V: If some random variables are used as proxies for the \tilde{e}_{it} , then what are their statistical consequences?

Set I of questions is important because any model which does not exist could not have generated our data and should not be used for their analysis.¹⁵ Specification problems exist in econometrics because the answers to all these sets of questions are unknown. However, as we argue below, they become less serious if equation (5) rather than its fixed- β version is adopted.

To answer Set II of questions, we need to consider Pratt and Schlaifer's¹⁶ conditions under which an observed distribution conforms to a stochastic law. These conditions which are stated at the end of this section are not automatically satisfied if we interpret equation (9) to refer to the sum of a parametric form of $f(\mathbf{x}_{1t}, \dots, \mathbf{x}_{Kt}; \alpha)$ and random disturbance. Yet it is a common practice in econometric model-building to derive a mathematical function of $\mathbf{x}_{1t}, \dots, \mathbf{x}_{Kt}$, and then at some final stage to simply tack on an error term without any discussion on the possible meaning and real-world sources of the probabilities implied by the assumed error distribution.

¹⁵For further discussion of this point, see Swamy, P.A.V.B. and P. von zur Muehlen (1988), "Further Thoughts on Testing for Causality with Econometric Models," *Journal of Econometrics, Annals*, 39, 105-147.

¹⁶Pratt and Schlaifer 1988, *op. cit.*, p. 37

In laying out a tentative model, let us deviate from this practice. Without imposing an arbitrary functional form, equation (9) can be written as

$$\tilde{y}_t = \sum_{j=1}^K x_{jt} \tilde{\alpha}_{jt} + \sum_{i=1}^r \tilde{e}_{it} \tilde{\delta}_{it}, \quad (14)$$

where $\mathbf{x}_{1t} = \mathbf{1}$ for all t . For certain variations in the $\tilde{\alpha}_{jt}$ and $\tilde{\delta}_{it}$, this equation coincides with equation (9) if the latter exists. More recently, Klein¹⁷ has argued that “Random parameters and systematic changes in parameters may be evidence of nonlinearities that have not been adequately captured in a model’s specification.” (p. 290). For example, if equation (9) is a constant elasticity of substitution production function, then the variations in the $\tilde{\alpha}_{jt}$ and $\tilde{\delta}_{it}$ for which equation (10) coincides with equation (9) are shown in Narasimham, Swamy and Reed.¹⁸ A matrix formulation of equation (10) is

$$\tilde{y}_t = \tilde{\alpha}_{1t} + \mathbf{x}'_{2t} \tilde{\alpha}_{2t} + \tilde{\mathbf{e}}'_t \tilde{\boldsymbol{\delta}}_t, \quad (15)$$

where the scalar $\tilde{\alpha}_{1t}$ represents the time-varying intercept, $\mathbf{x}_{2t} = (x_{2t}, \dots, x_{Kt})'$ is a $(K-1)$ -vector of observations on $(K-1)$ regressors, $\tilde{\alpha}_{2t} = (\tilde{\alpha}_{2t}, \dots, \tilde{\alpha}_{Kt})'$ is a $(K-1)$ -vector of coefficients, $\tilde{\mathbf{e}}_t = (\tilde{e}_{1t}, \dots, \tilde{e}_{rt})'$ is a r -vector of excluded variables, and $\tilde{\boldsymbol{\delta}}_t = (\tilde{\delta}_{1t}, \dots, \tilde{\delta}_{rt})'$ is a r -vector of coefficients. Assume that the data on \tilde{y}_t and $\tilde{\mathbf{x}}_{2t}$ for $t = 1, 2, \dots, T$ are available.

Equation (11) yields perfect predictions of as yet unobserved values of \tilde{y}_t if it is deterministic. The condition under which equation (11) is deterministic is of course the usual one: For all t , the sum of $\tilde{\alpha}_{1t}$, $\mathbf{x}'_{2t} \tilde{\alpha}_{2t}$, and $\tilde{\mathbf{e}}'_t \tilde{\boldsymbol{\delta}}_t$ predicts y_t with zero mean square error. According to Pratt and Schlaifer,¹⁹ if this condition is satisfied, then $\tilde{\mathbf{e}}_t$ is “a sufficient set.” The $\tilde{\alpha}_{1t}$, $\tilde{\alpha}_{2t}$, and $\tilde{\boldsymbol{\delta}}_t$ can be assumed to be constants (or degenerate) as in Pratt and Schlaifer²⁰ if the true relation of \tilde{y}_t to \mathbf{x}_{2t} and $\tilde{\mathbf{e}}_t$ is known to be linear.²¹ Otherwise, equation (11), quite possibly, provides a better approximation to the true relation than an exact parametric form of the nonlinear regression $\tilde{y}_t = f_t(\tilde{\mathbf{x}}_{2t}, \tilde{\mathbf{e}}_t; \boldsymbol{\alpha})$ because with the true parametric form being unknown, there is, after all, no guarantee that any of the “exact” forms will in fact be exact. Thus, by assigning time-varying coefficients to \mathbf{x}_{2t} and $\tilde{\mathbf{e}}_t$ we are not assuming any particular parametric form not implied by the theory.

However, the unobservables of equation (11) are not unique. To see this, consider the usual case, where r is greater than K , and let $\tilde{\boldsymbol{\delta}}_{1t}$ be a $(K-1)$ -vector consisting of any $(K-1)$ elements of $\tilde{\boldsymbol{\delta}}_t$. Then adding and subtracting the term $\mathbf{x}'_{2t} \tilde{\boldsymbol{\delta}}_{1t}$ on the right-hand side of equation (11) give $\tilde{y}_t = \tilde{\alpha}_{1t} + \mathbf{x}'_{2t} (\tilde{\alpha}_{2t} + \tilde{\boldsymbol{\delta}}_{1t}) + \tilde{\mathbf{e}}'^*_t \tilde{\boldsymbol{\delta}}_t$, where $\tilde{\mathbf{e}}'^*_t$ is obtained by subtracting \mathbf{x}_{2t} from the subvector of $\tilde{\mathbf{e}}_t$ whose coefficient vector is $\tilde{\boldsymbol{\delta}}_{1t}$.

¹⁷Klein, L.R. (1989), “Developments and prospects in macroeconomic modeling,” *Eastern Economic Journal*, XV, 287-304.

¹⁸Narasimham, G.V.L., P.A.V.B. Swamy and R.C. Reed (1988), “Productivity Analysis of U.S. Manufacturing Using a Stochastic-Coefficients Production Function,” *Journal of Business & Economic Statistics*, 6, 339-349.

¹⁹Pratt and Schlaifer 1984, *op. cit.*, p. 13.

²⁰Pratt and Schlaifer 1984, *op. cit.*, p. 13.

²¹A random variable is said to be degenerate if it takes a single value with probability 1.

This result can also be written as $\tilde{\mathbf{y}}_t = \tilde{\alpha}_{1t} + \mathbf{x}'_{2t}(\tilde{\alpha}_{2t} + \tilde{\delta}_{1t}) + (\tilde{\mathbf{e}}_t^* \mathbf{P})(\mathbf{P}^{-1}\tilde{\delta}_t)$, where \mathbf{P} is any nonsingular matrix. Since the $\tilde{\mathbf{e}}_t$ and the coefficients of equation (11) are unknown, equation (11) and the equations, $\tilde{\mathbf{y}}_t = \tilde{\alpha}_{1t} + \mathbf{x}'_{2t}(\tilde{\alpha}_{2t} + \tilde{\delta}_{1t}) + \tilde{\mathbf{e}}_t^* \tilde{\delta}_t$ and $\tilde{\mathbf{y}}_t = \tilde{\alpha}_{1t} + \mathbf{x}'_{2t}(\tilde{\alpha}_{2t} + \tilde{\delta}_{1t}) + (\tilde{\mathbf{e}}_t^* \mathbf{P})(\mathbf{P}^{-1}\tilde{\delta}_t)$, are (observationally) equivalent implying the same distribution for $\tilde{\mathbf{y}}_t$. From this result it follows that (i) equation (11) has an infinite number of equivalent representations, (ii) in the regressions of $\tilde{\mathbf{y}}_t$ on \mathbf{x}_{2t} , the vector $\tilde{\mathbf{e}}_t^*$ (or $\mathbf{P}'\tilde{\mathbf{e}}_t^*$) is just as much an excluded variable as $\tilde{\mathbf{e}}_t$, and (iii) the fixed-coefficients model $\tilde{\mathbf{y}}_t = \mathbf{f}(\mathbf{x}_{2t}; \alpha) + \tilde{\xi}_t$ cannot describe every equivalent representation of equation (11) unless it is both equivalent to equation (11) and not invariant under changes in the parametrization of equation (11). Unlike equation (11), real-world relations are unique, as Basmann²² has shown. That is, real-world relations do not change merely because of changes in the parametrization of equation (11). This proves that the correspondence between a real-world relation and equation (11) is one-to-many if one of the parametrization of the latter is true and does not exist otherwise. The fixed-coefficients model $\tilde{\mathbf{y}}_t = \mathbf{f}(\mathbf{x}_{2t}; \alpha) + \tilde{\xi}_t$ is true if it coincides with a parametrization of equation (11) which is true. These results have led Basmann²³ to say that econometricians should not restrict their analysis to one particular parametrization of equation (11).

Because of their nonuniqueness, assumptions about the unobservables of equation (11) may be logically inconsistent. Swamy and von zur Muehlen²⁴ show that specification errors arise if the assumptions underlying a representation of model (9) are not consistent. Zellner²⁵ agrees with this view, emphasizing that an analysis or theory that involves logical inconsistencies is not acceptable as a scientific explanation of observed phenomena. The following is an example of an inconsistent set of assumptions. Since at least one of $\tilde{\mathbf{e}}_t$ and $\tilde{\mathbf{e}}_t^*$ must be correlated with $\tilde{\mathbf{x}}_{2t}$, the assumption that the $\tilde{\mathbf{x}}_{2t}$ be uncorrelated with ‘the’ unknown excluded variables which could be any one of the vectors $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{e}}_t^*$, and $\mathbf{P}'\tilde{\mathbf{e}}_t^*$, is meaningless and the stronger assumption that the $\tilde{\mathbf{x}}_{2t}$ be uncorrelated with any of $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{e}}_t^*$, and $\mathbf{P}'\tilde{\mathbf{e}}_t^*$ is false, as Pratt and Schlaifer²⁶ show. According to them,²⁷ “A condition for consistent estimation stated in virtually every book on econometrics is meaningless in one common form, impossible to satisfy in another.” This proves that any regression of $\tilde{\mathbf{y}}_t$ on \mathbf{x}_{2t} does not satisfy the logical consistency condition if its error term is thought of as the joint effect of ‘the’ unknown excluded variables and if $\tilde{\mathbf{x}}_{2t}$ or any of its subvectors is assumed to be uncorrelated with those excluded variables.

To avoid the restriction that \mathbf{x}_{2t} be uncorrelated with any of $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{e}}_t^*$, and $\mathbf{P}'\tilde{\mathbf{e}}_t^*$, we combine equation (11) with the plausible equation

$$\tilde{\mathbf{e}}_t = \tilde{\Psi}_t \mathbf{x}_{2t} + \tilde{\mathbf{v}}_t, \quad (16)$$

²²See pp. 74-76 of Basmann, R.L. (1988), “Causality Tests and Observationally Equivalent Representations of Econometric Models,” *Journal of Econometrics, Annals*, 39, 69-104.

²³Basmann, *op. cit.*, pp. 75-76.

²⁴Swamy and von zur Muehlen, *op. cit.*, p. 119.

²⁵Zellner, A. (1988), “Causality and Causal Laws in Economics,” *Journal of Econometrics, Annals*, 39, 7-21.

²⁶Pratt and Schlaifer 1984, *op. cit.*, pp. 11-12.

²⁷Pratt and Schlaifer 1984, *op. cit.*, p. 9.

where the $r \times (K - 1)$ matrix $\tilde{\Psi}_t$ and the r -vector $\tilde{\mathbf{v}}_t$ are possibly random. For certain variations in $\tilde{\Psi}_t$ and $\tilde{\mathbf{v}}_t$, this equation coincides with the true relation of $\tilde{\mathbf{e}}_t$ to $\tilde{\mathbf{x}}_{2t}$ if the latter exists. Substituting (12) for $\tilde{\mathbf{e}}_t$ in (11) gives

$$\begin{aligned}\tilde{\mathbf{y}}_t &= (\tilde{\alpha}_{1t} + \tilde{\mathbf{v}}_t' \tilde{\delta}_t) + \mathbf{x}_{2t}' (\tilde{\alpha}_{2t} + \tilde{\Psi}_t' \tilde{\delta}_t) \\ &= \mathbf{x}_t' \tilde{\beta}_t \quad (t = 1, 2, \dots, T),\end{aligned}\tag{17}$$

where $\mathbf{x}_t = (\mathbf{1}, \mathbf{x}_{2t}')'$ and $\tilde{\beta}_t = [(\tilde{\alpha}_{1t} + \tilde{\mathbf{v}}_t' \tilde{\delta}_t), (\tilde{\alpha}_{2t} + \tilde{\Psi}_t' \tilde{\delta}_t)']'$. This equation is the same as equation (1). The terms $\tilde{\alpha}_{2t}$ and $\tilde{\Psi}_t' \tilde{\delta}_t$ of $\tilde{\beta}_t$ are needed to measure the direct and proxy effects of \mathbf{x}_{2t} on $\tilde{\mathbf{y}}_t$, respectively.

Alternatively, if the equation $\tilde{\mathbf{y}}_t = \tilde{\alpha}_{1t} + \mathbf{x}_{2t}' (\tilde{\alpha}_{2t} + \tilde{\delta}_{1t}) + \tilde{\mathbf{e}}_t^* \tilde{\delta}_t$ is considered, then it can be combined with the equation $\tilde{\mathbf{e}}_t^* = \tilde{\mathbf{e}}_t - \mathbf{c}_t = \tilde{\Psi}_t \mathbf{x}_{2t} + \tilde{\mathbf{v}}_t - \mathbf{c}_t = \tilde{\Psi}_t \mathbf{x}_{2t} + \tilde{\mathbf{v}}_t^*$, where the elements of \mathbf{c}_t are equal to the elements of \mathbf{x}_{2t} if the latter have $\tilde{\delta}_{1t}$ as their coefficient vector, and are equal to zero otherwise. This combination results in the equation $\tilde{\mathbf{y}}_t = (\tilde{\alpha}_{1t} + \tilde{\mathbf{v}}_t^* \tilde{\delta}_t) + \mathbf{x}_{2t}' (\tilde{\alpha}_{2t} + \tilde{\delta}_{1t} + \tilde{\Psi}_t' \tilde{\delta}_t)$ which is equivalent to equation (13), since $\mathbf{x}_{2t}' \tilde{\delta}_{1t} = \mathbf{c}_t' \tilde{\delta}_t$.

The vector $\tilde{\mathbf{v}}_t$ (or $\tilde{\mathbf{v}}_t^*$) is the *remainder* of $\tilde{\mathbf{e}}_t$ (or $\tilde{\mathbf{e}}_t^*$) after the effect $\tilde{\Psi}_t \mathbf{x}_{2t}$ of \mathbf{x}_{2t} on $\tilde{\mathbf{e}}_t$ (or $\tilde{\mathbf{e}}_t^*$) has been subtracted out. Although $\tilde{\mathbf{x}}_{2t}$ cannot be uncorrelated with every excluded variable that affects $\tilde{\mathbf{y}}_t$, it can be uncorrelated with the *remainder* of every such variable.²⁸ This shows that equation (13) containing $\tilde{\mathbf{v}}_t$ but not $\tilde{\mathbf{e}}_t$ has an advantage over equation (11). It can also be seen that an effect of changing the representation in (11) to $\tilde{\mathbf{y}}_t = \tilde{\alpha}_{1t} + \mathbf{x}_{2t}' (\tilde{\alpha}_{2t} + \tilde{\delta}_{1t}) + \tilde{\mathbf{e}}_t^* \tilde{\delta}_t$ is to change the decomposition $(\tilde{\alpha}_{1t}, \tilde{\alpha}_{2t})' + (\tilde{\mathbf{v}}_t' \tilde{\delta}_t, \tilde{\delta}_t' \tilde{\Psi}_t)'$ of $\tilde{\beta}_t$ to $(\tilde{\alpha}_{1t}, \tilde{\alpha}_{2t} + \tilde{\delta}_{1t})' + (\tilde{\mathbf{v}}_t^* \tilde{\delta}_t, \tilde{\delta}_t' \tilde{\Psi}_t)'$. Thus, the decomposition of $\tilde{\beta}_t$ into the terms that are needed to measure direct and proxy effects of \mathbf{x}_{2t} on $\tilde{\mathbf{y}}_t$ is not unique if it is not known whether $\tilde{\mathbf{e}}_t$ or $\tilde{\mathbf{e}}_t^*$ represents excluded variables. Note that equation (11) accounts for all the explanatory variables in equation (9) and hence does not suffer from a specification error due to omitted explanatory variables. It has another virtue. Unlike the parametric forms of linear and nonlinear regressions of $\tilde{\mathbf{y}}_t$ on \mathbf{x}_t , equation (11) may be appropriate even when the functional form of the true relationship between $\tilde{\mathbf{y}}_t$ and \mathbf{x}_t does not correspond to any familiar form.²⁹

A well known axiom in the literature is that the real effects of a deliberate change in the value of $\tilde{\mathbf{x}}_{2t}$ depend not only on the magnitude of the change but also on the timing of the change. Model (13) says exactly that. More specifically, it implies that the ‘total’ effect on $\tilde{\mathbf{y}}_t$ of \mathbf{x}_t relative to \mathbf{x}_{t-1} is $\mathbf{y}_t - \mathbf{y}_{t-1} = (\mathbf{x}_t - \mathbf{x}_{t-1})' \beta_t + \mathbf{x}_{t-1}' (\beta_t - \beta_{t-1})$ which depends on both the magnitude of the change $(\mathbf{x}_t - \mathbf{x}_{t-1})$ and the time index of β_t . Another important implication of model (13) is that only the part of this total effect contributed by α_{2t} represents a ‘direct’ effect and the other part which is due to $\Psi_t' \delta_t$ arises as a direct consequence of the fact that $\tilde{\mathbf{x}}_{2t}$ proxies for $\tilde{\mathbf{e}}_t$. Therefore, a ‘direct’ effect of \mathbf{x}_t on $\tilde{\mathbf{y}}_t$ is not equal to $(\mathbf{x}_t - \mathbf{x}_{t-1})' \beta_t + \mathbf{x}_{t-1}' (\beta_t - \beta_{t-1})$ unless equation (13) is applied to situations,

²⁸See Pratt and Schlaifer 1984, *op. cit.*, p. 14.

²⁹By contrast, a simultaneous-equation model for $\tilde{\mathbf{y}}_t$ does not account for all possible functional forms of equation (9), does not account for the nonuniqueness of the unobservables of equation (11), and assumes, perhaps inconsistently, that some of the \mathbf{x} ’s are uncorrelated with the \mathbf{e} ’s.

where $\tilde{\mathbf{x}}_t$ does not proxy for $\tilde{\mathbf{e}}_t$. Characterizations of such situations are given by Pratt and Schlaifer.³⁰

Zellner³¹ makes an additional point that model (13) must be shown to work well in explanation and prediction of $\tilde{\mathbf{y}}$ over a broad range of conditions and data before it is elevated to the status of a causal law. It is not difficult to see that equation (13) always performs well in explanation and prediction if $\tilde{\alpha}_{2t} \neq \mathbf{0}$ almost surely for all t and can be separated from $\tilde{\Psi}_t' \tilde{\delta}_t$ whenever the latter is different from a null vector with probability 1 for all t . Also, equation (13) with $\tilde{\alpha}_{2t} = \mathbf{0}$ with probability 1 for all t cannot provide any explanation of $\tilde{\mathbf{y}}$ but is of use in predicting $\tilde{\mathbf{y}}$, provided $\tilde{\Psi}_t \neq \mathbf{0}$ with probability 1 for all t . Thus, a variable that does not affect $\tilde{\mathbf{y}}$ may nevertheless be of use in predicting $\tilde{\mathbf{y}}$ because it proxies for an excluded variable that does affect $\tilde{\mathbf{y}}$.

To separate $\tilde{\Psi}_t' \tilde{\delta}_t$ from $\tilde{\alpha}_{2t}$, we assume that $\tilde{\beta}_t$ satisfies the equation

$$\tilde{\beta}_t = \Pi \mathbf{z}_t + L \tilde{\epsilon}_t, \quad (18)$$

which is the same as equation (2). This completes the derivation of equation (5) which reveals that the stochastic coefficients model given by equations (1) and (2) can be transformed into a fixed-coefficient model, the source whose disturbance term is the second term on the right-hand side of equation (2). Quite possibly, this second term is a vector and not a scalar because it is difficult to find \mathbf{z}_t such that all the elements of $\tilde{\beta}_t$ other than the first element are exactly equal to the corresponding elements of $\Pi \mathbf{z}_t$. Thus, the fixed-coefficient model (5) is different from the fixed- β version of equation (1). This version is usually obtained by adding (or multiplying) random disturbance to (or by) a mathematical formula. This random disturbance is enigmatic because its real-world sources and interpretations are not clear.

The usefulness of equation (14) becomes apparent when Π is partitioned as $\Pi = \begin{pmatrix} \pi_{11} & \pi_{12}' \\ \pi_{21} & \Pi_{22} \end{pmatrix}$, where π_{11} is a scalar, π_{21} and π_{12} are $(K-1)$ - and $(p-1)$ -vectors, respectively, and Π_{22} is $(K-1) \times (p-1)$. Partitioning \mathbf{z}_t in conformity with the partitioning of Π as $\mathbf{z}_t = (z_{1t}, \mathbf{z}_{2t}')'$ gives $\mathbf{x}_t' \tilde{\beta}_t = (\tilde{\alpha}_{1t} + \tilde{\mathbf{v}}_t' \tilde{\delta}_t) + \mathbf{x}_{2t}' (\tilde{\alpha}_{2t} + \tilde{\Psi}_t' \tilde{\delta}_t) = \mathbf{x}_t' \Pi \mathbf{z}_t + \mathbf{x}_t' L \tilde{\epsilon}_t = (\pi_{11} + \pi_{12}' \mathbf{z}_{2t}) + \mathbf{x}_{2t}' (\pi_{21} + \Pi_{22} \mathbf{z}_{2t}) + \mathbf{x}_t' L \tilde{\epsilon}_t$. This formulation shows that the vector π_{21} measures the mean of $\tilde{\alpha}_{2t}$ if the $\tilde{\mathbf{z}}_{2t}$ are highly correlated with the $\tilde{\Psi}_t' \tilde{\delta}_t$ but independent of the $\tilde{\alpha}_{2t}$. The \mathbf{z}_{2t} need not be included in equation (14) if the expected values of $\tilde{\mathbf{v}}_t' \tilde{\delta}_t$ and $\tilde{\Psi}_t' \tilde{\delta}_t$ are zero and the expected value of $\tilde{\alpha}_{2t}$ is nonzero. Note that Pratt and Schlaifer's³² approach differs from our approach in that it includes concomitants in equation (13) rather than in equation (14). Inclusion of concomitants in equation (13) was enough because they were dealing with linear models with constant coefficients.

The model given by equations (13) and (14) may be neither too broad nor too narrow if the error vector $\{\tilde{\epsilon}_t\}$ is assumed to satisfy the difference equation

$$\tilde{\epsilon}_t = \Phi \tilde{\epsilon}_{t-1} + \tilde{\mathbf{a}}_t, \quad (19)$$

³⁰Pratt and Schlaifer 1984, *op. cit.*, p. 15-16.

³¹Zellner, *op. cit.*, pp. 8-9.

³²Pratt and Schlaifer 1984, *op. cit.*, p. 17.

which is the same as equation (3).

Assumption 1 *The variables $\tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{z}}_t$ are statistically independent of $\tilde{\mathbf{e}}_t$.*

Pratt and Schlaifer's³³ definition of genetic distributions should suffice to suggest the various ways in which Assumption 1 can be satisfied in the linear case. For example, $\tilde{\mathbf{x}}_t$ will be uncorrelated with $\tilde{\mathbf{v}}_t$ if the conditional distribution of $\tilde{\mathbf{e}}_t$ given a value of $\tilde{\mathbf{x}}_t$ implied by equation (12) coincides with the distribution that would be realized if $\tilde{\mathbf{x}}_t$ were held fixed at that value. Since $\mathbf{L}\tilde{\mathbf{e}}_t$ is the remainder of $\tilde{\beta}_t$ after the effect $\mathbf{\Pi}\mathbf{z}_t$ of $\tilde{\mathbf{z}}_t$ on $\tilde{\beta}_t$ has been subtracted out, it can be independent of $\tilde{\mathbf{z}}_t$. If $\tilde{\mathbf{x}}_t$ is assumed to be correlated with $\tilde{\beta}_t$, then it can be included in equation (14) as a subvector of \mathbf{z}_t . This shows that unlike the absence-of-correlation assumption between $\tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{e}}_t$ (or $\tilde{\mathbf{e}}_t^*$), Assumption 1 can be true.

Equations (14) and (15) are the expressions of our beliefs about the coefficient vector of model (13) whose structure does not involve logical or mathematical contradictions. The parameter matrices of model (13) are $\mathbf{\Pi}$, $\mathbf{\Phi}$, and $\sigma_a^2\Delta_a$. Certainly, it would be a mistake to think of these parameters as a description of a reality when, in fact, they are Greek letters, divorced from the reality they are supposed to represent.³⁴ Such mistakes can be avoided if model (5) coincides with a stochastic law defined by Pratt and Schlaifer.³⁵

In preparation for stating their definition of a law, we need to replace the observed outcome notation \mathbf{y}_t by a potential outcome notation because it is the existence of potential values that distinguishes a law from a statistical association. Let $\tilde{\mathbf{x}}_t$ denote a vector of factors whose values can be deliberately manipulated. Let $\tilde{\mathbf{z}}_{xt}$ denote a vector of concomitants. Suppose that $\tilde{\mathbf{y}}_{xzt}$ denotes a set of random variables, one for each possible value of $\tilde{\mathbf{x}}_t$, each realized value of $\tilde{\mathbf{z}}_{xt}$ and each t and that there exists a vector of random variables, denoted by $\tilde{\mathbf{a}}_{xzt}$, whose distribution is conditional on the realized value of $\tilde{\mathbf{z}}_{xt}$ but is defined for each possible value of $\tilde{\mathbf{x}}_t$ whether or not this possible value is realized. Suppose also that the \mathbf{z} 's and \mathbf{a} 's are generated by a single random process following a joint distribution in which for each \mathbf{x} , the marginal distribution of the $\tilde{\mathbf{z}}$'s is arbitrary but given any one pair of values of $\tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{z}}_{xt}$, the $\tilde{\mathbf{a}}_{xzt}$ are identically and independently distributed (iid).

Then according to Pratt and Schlaifer,³⁶ $\tilde{\mathbf{y}}$ is related to $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{z}}$ by a law with concomitants if there exists a function f which on the t th observation associates with every possible value of $\tilde{\mathbf{x}}_t$ and realized value of $\tilde{\mathbf{z}}_{xt}$ a random variable, $\tilde{\mathbf{y}}_{xzt} = f(\mathbf{x}_t, \mathbf{z}_t, \tilde{\mathbf{a}}_{xzt})$, and a potential value, $\mathbf{y}_{xzt} = f(\mathbf{x}_t, \mathbf{z}_{xt}, \mathbf{a}_{xzt})$. This value is called 'potential' because it is not always realized. It will be realized on any one observation t if it corresponds to the realized value \mathbf{x} of $\tilde{\mathbf{x}}_t$.

Pratt and Schlaifer³⁷ prove that when the value and the variance of $\tilde{\mathbf{a}}_{xzt}$ do not depend on \mathbf{x}_t , the law, $\tilde{\mathbf{y}}_{xzt} = f(\mathbf{x}_t, \mathbf{z}_t, \tilde{\mathbf{a}}_{xzt})$, is observable if $\tilde{\mathbf{x}}_t$ and $\tilde{\mathbf{a}}_{xzt}$ are

³³Pratt and Schlaifer 1984, *op. cit.*, p. 15.

³⁴See p. 55 of Lindley, D.V. (1990), "The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics," *Statistical Science*, 5, 44-89.

³⁵Pratt and Schlaifer 1988, *op. cit.*, pp. 28 and 35.

³⁶Pratt and Schlaifer 1988, *op. cit.*, pp. 28 and 35.

³⁷Pratt and Schlaifer 1988, *op. cit.*, p. 37.

statistically independent given $\tilde{\mathbf{z}}_{\mathbf{x}t} = \mathbf{z}_{\mathbf{x}t}$ for all t . Since the value and the variance of the error in equation (5) depend on \mathbf{x}_t , the law is observable if Assumption 1 is true.³⁸ Therefore, under Assumption 1, model (5) coincides with the law, in which case Π and $\tilde{\epsilon}_t$ are unique because the law specifies a unique distribution. It should be noted that the derivation of the condition for the observability of a law entails two steps:

- Recognizing that causal inferences can be soundly drawn from data obtained by randomized experimentation.
- Showing that non-randomized circumstances correspond to randomized circumstances if they adequately conform to the observability conditions which are satisfied in randomized studies.³⁹

For example, from a strong sample multiple correlation between \mathbf{y} and \mathbf{x} we are permitted to conclude that the value of \mathbf{y} is largely determined by \mathbf{x} only when the sample is generated by a randomized experiment. In non-randomized circumstances, this permission is granted to us only when these circumstances adequately conform to critical assumptions which are satisfied in randomized circumstances.

The total number of parameters in model (5) is $\mathbf{n} = K\mathbf{p} + \mathbf{m}^2 + [\mathbf{m}(\mathbf{m} + 1)/2] + 1$. Nothing in the definition of the law requires that some of these parameters be equal to zero. This fact is usually obscured by the practice of using the a priori restrictions that will reduce model (5) to a fixed- β model. When $\mathbf{n} > \mathbf{T}$, such practices, though ad hoc, may be unavoidable. But whatever statistical estimates of the parameters of a special case of (5), their signs, their ‘large sample t -statistics’ seem to tell us about a reality, may be false, and is due partly to the imposition of a set of arbitrary restrictions on model (5). When $\mathbf{n} < \mathbf{T}$, there are no grave difficulties in implementing a general stochastic coefficients approach, as shown in Section 7.

³⁸Pratt and Schlaifer 1984, *op. cit.*, p. 33.

³⁹Pratt and Schlaifer 1988, *op. cit.*, pp. 23-24.

Appendix B: The Special Cases of a Stochastic Coefficients Model

The framework outlined in Section 2 is useful for estimating not only the very general fixed-coefficients model (5) but also a variety of models that have been considered in theoretical and applied econometrics.

Special Case I of Equation (5): Classical Linear Fixed- β Model

To get a classical linear fixed- β model from equation (5), impose the following restrictions:

$$\begin{aligned}
 (Ia) \quad \mathbf{z}_t &= (1, 0, \dots, 0)', \\
 (Ib) \quad p &= 1, \\
 (Ic) \quad m &= K, \\
 (Id) \quad \Phi &= 0, \\
 (Ie) \quad L &= I_K, \\
 (If) \quad \Delta_a &= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.
 \end{aligned} \tag{20}$$

Linear restrictions on the first column of $\mathbf{\Pi}$ will be present if some of the first K rows of $\mathbf{R}_{\mathbf{\Pi}}$ defined in equation (4) are not null.

SCEP imposes restrictions $(\mathbf{Ia}) - (\mathbf{If})$ on equation (5) if you fill out the PARAMTR, DDATA1, DDATA2, RCFDT, and RVCDT files as follows:

PARAMTRS FILE

M = K

(Note: Here \mathbf{K} stands for the number of independent variables in equation (1).)

P = 1

Z-VECTORS = DEFAULT

L-MATRIX = DEFAULT

ZERO PHI = YES

DIAGONAL DELTA = NO

INITIAL PHI MATRIX = DEFAULT

INITIAL DELTA MATRIX = USER'S

NUMBER OF ROWS IN LINEAR RESTRICTION ON PHI = 0

NUMBER OF ROWS IN LINEAR RESTRICTION ON DELTA = $K(K+1)/2$

OLS = YES

END OF DATA

//

(Note: Other fields in the PARAMTRS file should also be filled appropriately.)

FILE DDATA1

← 80 spaces →			
DCW spaces	DCW spaces	...	DCW spaces ^a
1.0			
1.0	0.0	...	0.0
0.0	0.0	...	0.0
⋮	⋮		⋮
0.0	0.0	...	0.0
END OF DATA			
//			

^aContinued in DDATA2.

FILE DDATA2

← 80 spaces →		
DCW spaces	...	DCW spaces
0.0	...	0.0
0.0	...	0.0
⋮		⋮
0.0	...	0.0
END OF DATA		
//		

FILE RCFDT

← 80 spaces →		
DCW spaces	DCW spaces	DCW spaces
1	1	1.0
2	2	1.0
3	3	1.0
⋮	⋮	⋮
K(K+1)/2	K(K+1)/2	1.0
END OF DATA		
//		

FILE RVCDD

← 80 spaces →		
DCW spaces	DCW spaces	
1	1.0	
2	0.0	
3	0.0	
⋮	⋮	
K(K+1)/2	0.0	
END OF DATA		
//		

You can use only the restrictions (**Ic**) – (**Ie**) without the restrictions (**Ia**) and (**Ib**) by writing **P** = the number of z's in equation (2), Z-VECTORS = USER'S to the PARAMTRS file, and typing the data on the z's in the ZDATA1 and ZDATA2 files.

Special Case II of Equation (5): A Linear Fixed- β Model with AR(1) Errors

A linear fixed- β model in which the disturbances follow an autoregressive process of order 1 (AR(1)) is given by equation (5) when this equation is subjected to the following restrictions:

$$\begin{aligned}
 (IIa) \quad \mathbf{z}_t &= (1, 0, \dots, 0)', \\
 (IIb) \quad p &= 1, \\
 (IIc) \quad m &= K, \\
 (IIId) \quad \Phi &= \begin{bmatrix} \phi_{11} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \\
 (IIe) \quad L &= I_K, \\
 (IIIf) \quad \Delta_a &= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.
 \end{aligned} \tag{21}$$

Linear restrictions on the first column of $\mathbf{\Pi}$ will be present if some of the first K rows of $\mathbf{R}_{\mathbf{\Pi}}$ defined in equation (4) are not null.

To impose the restrictions **(IIa)–(II f)** on equation (5), fill out the PARAMTR, DDATA1, DDATA2, RCFDT, RVCDT, RCFPH, and RVCPH files as follows:

PARAMTR SFILE

M = K

(Note: Here **K** stands for the number of independent variables in equation (1).)

P = 1

Z-VECTORS = DEFAULT

L-MATRIX = DEFAULT

ZERO PHI = NO

DIAGONAL PHI = NO

DIAGONAL DELTA = NO

INITIAL PHI MATRIX = DEFAULT

INITIAL DELTA MATRIX = USER'S

NUMBER OF ROWS IN LINEAR RESTRICTION ON PHI = **$K^2 - 1$**

NUMBER OF ROWS IN LINEAR RESTRICTION ON DELTA = $K(K+1)/2$

OLS = NO

GLS = YES

END OF DATA

//

(Note: Other fields in the PARAMTRS file should also be filled appropriately.)

The files DDATA1, DDATA2, RCFDT, and RVCDT are the same as those in Case I.

FILE RCFPH

	80 spaces	
←	→	
DCW spaces	DCW spaces	DCW spaces
1	2	1.0
2	3	1.0
3	4	1.0
⋮	⋮	⋮
$K^2 - 1$	K^2	1.0
END OF DATA		
//		

FILE RVCPH

	80 spaces	
←	→	
DCW spaces	DCW spaces	
1	0.0	
2	0.0	
3	0.0	
⋮	⋮	
$K^2 - 1$	0.0	
END OF DATA		
//		

You can impose the restrictions **(IIc)** – **(II f)** without the restrictions **(IIa)** and **(IIb)** by writing **P** = the number of z's in equation (2), Z-VECTORS = USER'S to the PARAMTRS file, and typing the data on the z's in the ZDATA1 and ZDATA2 files.

Special Case III of Equation (5): A Linear Fixed- β Model with Heteroscedastic Disturbances

If only heteroscedasticity is present in a linear fixed- β regression model, then the following restrictions may be imposed on equation (5):

$$\begin{aligned}
 (IIIa) \quad \mathbf{z}_t &= (1, 0, \dots, 0)', \\
 (IIIb) \quad p &= 1, \\
 (IIIc) \quad m &= K, \\
 (IIId) \quad \Phi &= \mathbf{0}, \\
 (IIIe) \quad L &= I_K, \\
 (III f) \quad \Delta_a &= (1/\sigma_a^2) \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{KK} \end{bmatrix}.
 \end{aligned} \tag{22}$$

Linear restrictions on the first column of $\mathbf{\Pi}$ will be present if some of the first K rows of $\mathbf{R}_{\mathbf{\Pi}}$ defined in equation (4) are not null.

To impose the restrictions **(IIIa)–(III f)** on equation (5), fill out the PARAMTRS file as follows:

PARAMTRS FILE

M = K

(Note: Here **K** stands for the number of independent variables in equation (1).)

P = 1

Z-VECTORS = DEFAULT

L-MATRIX = DEFAULT

ZERO PHI = YES

DIAGONAL PHI = NO

DIAGONAL DELTA = YES

INITIAL PHI MATRIX = DEFAULT

INITIAL DELTA MATRIX = DEFAULT

NUMBER OF ROWS IN LINEAR RESTRICTION ON PHI = 0

NUMBER OF ROWS IN LINEAR RESTRICTION ON DELTA = 0

OLS = NO

GLS = NO

END OF DATA

//

(Note: Other fields in the PARAMTRS file should also be filled appropriately.)

If you relax the restriction **(III f)** to make Δ_a non-diagonal, then you obtain a general type of heteroscedasticity.

Special Case IV of Equation (5): A Shifting Fixed- β Model

Equations (1)-(3) represent temporal shifts in the regression of $\tilde{\mathbf{y}}_t$ on \mathbf{x}_t if

$$(IVa) \quad \mathbf{z}_t = (d_{1t}, d_{2t}, \dots, d_{pt})', \quad (23)$$

where $d_{1t}, d_{2t}, \dots, d_{pt}$ are a set of p dummy variables. This definition of \mathbf{z}_t may be combined with any of the definitions of Φ , L , and Δ_a given in Cases I, II, and III.

In the conventional approach, dummy variables are sometimes used to represent a finite number of shifts in some or all the coefficients of a regression model without allowing any shifts in the variance of the error distribution. Not all of these shifts may be temporary but some of them may persist indefinitely. Typically, the timing of any of these shifts will be unknown to the modeler. Use of dummy variables will reliably detect a shift only if the preselected location of the break is close to the true break, if the model used to represent this break is true, and if there are sufficient observations before and after the true break. By contrast, the approach based on equation (1) permits model coefficients to vary every period or continuously over time. This variation introduces continuous temporal variation in both the mean and the variances and covariances of a dependent variable, as can be seen from equation (5). The variation in the mean is consonant with the variation in the variances and covariances. The stochastic coefficient analysis employs less-restrictive assumptions—notably, that the timing of a coefficient change is generally unknown and that coefficient changes may be intrinsically evolutionary, with some coefficients subject to changes every period and others to infrequent changes. This analysis assumes also that coefficient changes may be partly transient and may provide only partial predictions of future values of the model coefficients (see Fuhrer and Tinsley).⁴⁰

Furthermore, shift adjustments typically represent our ability to accommodate events after they happen. In any research program, the problem of finding evidence against a theory, or equivalently, the problem of finding outcomes that have a small probability of occurring whenever a theory is true can really not be solved by use of dummy variables in post hoc adjustments.

Special Case V of Equation (5): A Non-Linear Fixed-Coefficients Model

If the elements of \mathbf{z}_t are some functions of \mathbf{x}_t , then equation (5) represents a non-linear relationship. Flexible functional forms introduced in Diewert and Wales⁴¹ also fall out as special cases of equation (5). For example, equation (5) represents a flexible functional form if

⁴⁰Fuhrer, J.C. and P.A. Tinsley (1990), "Continuous versus One-time Changes in Policy Transmission Channels," *Federal Reserve Bulletin*, 76, 1004-8.

⁴¹Diewert, W.E. and T.J. Wales (1987), "Flexible Functional Forms and Global Curvature Conditions," *Econometrica*, 55, 43-68.

$$\begin{aligned}
(Va) \quad m &= K \\
(Vb) \quad p &= K \\
(Vc) \quad \mathbf{z}_t &= \mathbf{x}_t, \\
(Vd) \quad \Pi &= \Pi', \\
(Ve) \quad \Phi &= \begin{bmatrix} \phi_{11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \\
(Vf) \quad L &= I_K, \\
(Vg) \quad \Delta_a &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.
\end{aligned} \tag{24}$$

The model satisfying these restrictions may be compared with model (5). Imposing the restrictions (Vf) and (Vg) on equation (5) amounts to adding an error term to $\mathbf{x}_t' \Pi \mathbf{z}_t$. It is shown in Appendix A that model (5) can approximate the true function between $\tilde{\mathbf{y}}_t$ and \mathbf{x}_t better than the equation, $\tilde{\mathbf{y}}_t = \mathbf{x}_t' \Pi \mathbf{z}_t$ plus an error term, even when $\mathbf{x}_t' \Pi \mathbf{z}_t$ is a second order differential approximation to an arbitrary twice continuously differential function at any point in an admissible domain.

To impose the restrictions (Va)–(Vg) on equation (5), fill out the PARAMTR, DDATA1, DDATA2, RCFDT, RVC DT, RCFPH, RVC PH, RCFPI, and RVCPI files as follows:

PARAMTRS FILE

```

M = K
(Note: Here K stands for the number of independent variables in equation (1).)
P = K
Z-VECTORS = USER'S
L-MATRIX = DEFAULT
ZERO PHI = NO
DIAGONAL PHI = NO
DIAGONAL DELTA = NO
INITIAL PHI MATRIX = DEFAULT
INITIAL DELTA MATRIX = USER'S
NUMBER OF ROWS IN LINEAR RESTRICTION ON PHI =  $K^2 - 1$ 
NUMBER OF ROWS IN LINEAR RESTRICTION ON PI =  $K(K - 1)/2$ 
NUMBER OF ROWS IN LINEAR RESTRICTION ON DELTA =  $K(K+1)/2$ 
OLS = NO
GLS = YES

```


END OF DATA

//

(Note: Other fields in the PARAMTRS file should also be filled appropriately.)

The files DDATA1, DDATA2, RCFDT, RVCDT, RCFPH, and RVCPH are the same as those in Case II.

FILE RCFPI

80 spaces		
DCW spaces	DCW spaces	DCW spaces
1	2	1.0
1	$K + 1$	-1.0
2	3	1.0
2	$2K + 1$	-1.0
3	4	1.0
3	$3K + 1$	-1.0
\vdots	\vdots	\vdots
$K - 1$	K	1.0
$K - 1$	$K(K - 1) + 1$	-1.0
K	$K + 3$	1.0
K	$2(K + 1)$	-1.0
\vdots	\vdots	\vdots
$2K - 3$	$2K$	1.0
$2K - 3$	$K(K - 1) + 2$	-1.0
\vdots	\vdots	\vdots
$K(K - 1)/2$	$K(K - 1)$	1.0
$K(K - 1)/2$	$(K + 1)(K - 1)$	-1.0
END OF DATA		
//		

This file specifies the subscripts and values of only the nonzero elements of \mathbf{R}_{Π} defined in equation (4). If the ij th element of \mathbf{R}_{Π} , denoted by $\mathbf{R}_{\Pi}(i, j)$, is nonzero, then in the above file the value of i is entered in the first column, the value of j is entered in the second column, and the value of $\mathbf{R}_{\Pi}(i, j)$ is entered in the third column. Unfortunately, the pattern of values presented in the second column of the above file may not be as clear as it should be. The easiest way to figure out this pattern is to write down explicitly the column stack of Π on paper.

FILE RVCPI

80 spaces	
DCW spaces	DCW spaces
1	0.0
2	0.0
3	0.0
\vdots	\vdots
$K(K - 1)/2$	0.0

END OF DATA

//

If the restrictions $\mathbf{V}(b)$, $\mathbf{V}(c)$, $(\mathbf{V}d)$, $(\mathbf{V}e)$ and $(\mathbf{V}g)$ are dropped, then SCEP provides a general method of estimating flexible functional forms.

Special Case VI of Equation (5): Structural Restrictions on Equation (5)

Pratt and Schlaifer⁴² provide a very elegant proof of the proposition that the econometrician's definitions of exogenous and instrumental variables do not lead to consistent estimates of the direct effects of \mathbf{x}_t on $\tilde{\mathbf{y}}_t$. Therefore, it is correct to ignore such definitions. If these definitions are dropped, then each coefficient in equation (1) is a sum of direct and proxy effects. Direct effects can be separated from proxy effects by experimenting with different sets of concomitants, denoted by the \mathbf{z}_t .⁴³

Case VII (A Generalization of Equation (5)): A Fixed-Coefficients Model with ARMA Errors

$$\begin{aligned}
 (VIIa) \quad m &= (p+q)K, \\
 (VIIb) \quad \Phi &= \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p & \theta_1 & \theta_2 & \cdots & \theta_{q-1} & \theta_q \\ I_K & 0 & \cdots & 0 & 0 & & & & & \\ 0 & I_K & \cdots & 0 & 0 & & & 0 & & \\ \vdots & \vdots & & \vdots & \vdots & & & & & \\ 0 & 0 & \cdots & I_K & 0 & & & & & \\ & & & & & 0 & 0 & \cdots & 0 & 0 \\ & & & & & I_K & 0 & \cdots & 0 & 0 \\ & & & & & 0 & I_K & \cdots & 0 & 0 \\ & & & & & \vdots & \vdots & & \vdots & \vdots \\ & & & & & 0 & 0 & \cdots & I_K & 0 \end{bmatrix},
 \end{aligned}$$

where each of the

$$\Phi_i = \begin{bmatrix} \phi_{i11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad (i = 1, 2, \dots, p)$$

is a $K \times K$ matrix with all of its elements other than the leading diagonal element equal to zero, and each of the

$$\theta_j = \begin{bmatrix} \theta_{j11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad (j = 1, 2, \dots, q)$$

is a $K \times K$ matrix with all of its elements other than the leading diagonal element equal to zero,

$$(VIIc) \quad L = [I_K, 0, \dots, 0],$$

⁴²Pratt and Schlaifer 1988, *op. cit.*, p. 48.

⁴³Swamy, Mehta and Singamsetti, *op. cit.*

where the number of null matrices of order $K \times K$, denoted by 0, is $(p + q - 1)$,

$$(VIIId) \quad L\Delta_a L' = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad (25)$$

is a $K \times K$ matrix with all of its elements other than the leading diagonal element equal to zero.

The restrictions (VIIa) – (VIIId) are appropriate if all the coefficients of equation (1) except the first coefficient are constant and the deviation of the first coefficient from its fixed mean follows a univariate autoregressive and moving average (ARMA) process of orders p and q .⁴⁴ SCEP does not yet have the capability of estimating equation (5) subject to these restrictions.

Case VIII (ARCH Restrictions on Equation (5)): An ARCH Model

The autoregressive conditional heteroscedastic (ARCH) model developed by Engle⁴⁵ specifies a set of restrictions on equations (1)-(3). These restrictions are

$$\begin{aligned} (VIIIa) \quad & \text{replace } \tilde{y}_t \text{ by } \tilde{y}_t^* = (\tilde{y}_t - \mathbf{x}_t' \boldsymbol{\beta}), \text{ where } \boldsymbol{\beta} \text{ is fixed,} \\ (VIIIb) \quad & \text{replace } \tilde{\mathbf{x}}_t' \text{ by } (1, \tilde{y}_{t-1}^*, \dots, \tilde{y}_{t-K+1}^*), \\ (VIIIc) \quad & \boldsymbol{\Pi} = \mathbf{0}, \\ (VIId) \quad & m = K, \\ (VIIIe) \quad & \boldsymbol{\Phi} = \mathbf{0}, \\ (VIIf) \quad & L = I_K, \\ (VIIIg) \quad & \Delta_a = (1/\sigma_a^2) \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{KK} \end{bmatrix}, \\ (VIIIh) \quad & \tilde{\boldsymbol{\epsilon}}_t \text{ and } \tilde{\mathbf{x}}_t \text{ are independent.} \end{aligned} \quad (26)$$

These restrictions do not have any obvious connection with equation (9) and are not designed to satisfy the conditions for the observability of a law. Therefore, any parameter of ARCH models does not have obvious economic interpretations. For example, if the $\tilde{\mathbf{x}}_t$ are correlated with the $\tilde{\boldsymbol{\epsilon}}_t$'s of equation (9), as they usually are, then $\boldsymbol{\beta}$ does not measure the direct effect on \tilde{y}_t of \mathbf{x}_t . Real-world interpretations of $\tilde{\boldsymbol{\epsilon}}_t$ are also not obvious. Furthermore, the restriction (VIIIa) eliminates all the advantages equation (5) has over its fixed- $\boldsymbol{\beta}$ version because the assumption of the constancy of $\boldsymbol{\beta}$ is as severe as the assumption that the coefficients on the nonconstant regressors of equation (1) are constant.

⁴⁴Swamy and Tinsley, *op. cit.*, p. 106.

⁴⁵Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1008.

Case IX: Co-integrated Variables and Error-Correction Models

Let B denote the backward shift operator, i.e., for any variable x_t , $x_{t-i} = B^i x_t$. Granger⁴⁶ defines that a vector, $\tilde{\mathbf{y}}_t = (\tilde{y}_{1t}, \tilde{y}_{2t}, \dots, \tilde{y}_{Kt})'$, of K variables is *co-integrated* of orders d, b if:

1. each $\tilde{\mathbf{y}}_{jt}^{**} \equiv (\mathbf{1} - B)^d \tilde{\mathbf{y}}_{jt}$ has no deterministic component and is covariance stationary for $d > 0$;
2. there exists a set of constants $\alpha_1, \alpha_2, \dots, \alpha_K$ (not all zero) such that $(\mathbf{1} - B)^{d-b} (\sum_{j=1}^K \alpha_j \tilde{\mathbf{y}}_{jt})$ has no deterministic component and is covariance stationary for $b > 0$. (Caution: Do not confuse these α 's with the α 's of equation (11).)

The essential differences between the co-integrated variables and equation (13) are: (i) Whereas knowledge of the stationarity producing transformation is required to specify a vector of co-integrating variables, no similar requirement is imposed on equation (13). (ii) Whereas by assumption the coefficients (α 's) on the co-integrated variables do not contain any effects of omitted variables, the coefficients of equation (13) may contain both direct and proxy effects.

In Granger's definition, the *co-integrating* vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)'$, is the same as an eigenvector corresponding to a zero eigenvalue of the sum of unknown coefficient matrices in a moving average model of $\tilde{\mathbf{y}}_t^{**} = (\tilde{y}_{1t}^{**}, \tilde{y}_{2t}^{**}, \dots, \tilde{y}_{Kt}^{**})'$. Even when $\tilde{y}_{1t}, \tilde{y}_{2t}, \tilde{y}_{3t}, \dots, \tilde{y}_{Kt}$ are the same as $\tilde{y}_t, \tilde{x}_{2t}, \tilde{x}_{3t}, \dots, \tilde{x}_{Kt}$, respectively, there is no obvious connection between any of these co-integrating models and equation (13) or (5) which, under very general conditions, coincides with a stochastic law, as shown in Appendix A. Co-integrating models are not designed to coincide with stochastic laws and, hence, any of their parameters do not have economic interpretations, in general (see Swamy and Tavlas).⁴⁷

Special Case X of Equation (5): A Version of Kalman's Model

If, in equations (1) and (2),

$$\begin{aligned} (Xa) \quad \mathbf{z}_t &= (\mathbf{1}, 0, \dots, 0)', \\ (Xb) \quad p &= \mathbf{1}, \\ (Xc) \quad m &= K, \\ (Xd) \quad L &= I_K, \end{aligned}$$

⁴⁶Granger, C. W. J. (1983), "Co-integrating Variables and Error-Correcting Models," Unpublished discussion paper 83-13a, University of California, San Diego, CA.

⁴⁷Swamy, P.A.V.B. and G.S. Tavlas (1992), "Is it Possible to Find an Econometric Law that Works Well in Explanation and Prediction? The Case of Australian Money Demand," *Journal of Forecasting*, 11, 17-33.

$$\begin{aligned}
(Xe) \quad \Phi &= \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \phi_{22} & \cdots & \phi_{2K} \\ \vdots & \vdots & & \vdots \\ 0 & \phi_{K2} & \cdots & \phi_{KK} \end{bmatrix}, \\
(Xf) \quad \Delta_a &= (1/\sigma_a^2) \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & \sigma_{2K} \\ \vdots & \vdots & & \vdots \\ 0 & \sigma_{K2} & \cdots & \sigma_{KK} \end{bmatrix}, \tag{27}
\end{aligned}$$

then we obtain a version of Kalman's model. Assuming that the first column of Π is zero, Φ of the above form is known, $\tilde{\epsilon}_t$ is Gaussian, that Δ_a of the above form is known, and that at some initial period (say $t = 0$), β has a prior distribution in the form of a multivariate normal distribution with specified mean and covariance matrix, Harrison and Stevens proposed a Bayesian forecasting method which is based on the Kalman algorithm for predicting the y_t from equations (1) and (2).⁴⁸

No doubt, Harrison and Stevens' insistence on complete specification of model (5) (including the values of its parameters) prior to observation of the data would lessen the risk of selecting, while looking at the data, of models or priors that are overly special. (Berger⁴⁹ calls such a selection 'purposeful or unintended "cheating"'.) This security, however, comes with an impossibly heavy price, as Berger⁵⁰ points out. For example, the values $\Pi = 0$, $L = I$, and $\Phi = I$ are used in the literature; consider, in this connection, what price the user of these values pays. When equation (2) is supplemented by $\Pi = 0$ and $L = I$, it amounts only to stating that the mean of the coefficient vector of the relationship between \tilde{y}_t and \tilde{x}_t is zero, so on average, this relationship does not exist. In other words, when equation (2) is equated to equation (3), equation (1) is assumed to provide no signal. When further supplemented by the prior $\Phi = I$, it amounts only to announcing that equation (1) is equivalent to $\tilde{y}_t = \sum_{j=1}^K x_{jt} \tilde{\beta}_{jt}$, where $\tilde{\beta}_{jt} - \pi_{j1} = \tilde{\beta}_{j,t-1} - \pi_{j1} + \tilde{a}_{jt}$. Here the subtraction of π_{j1} from both sides of the equation $\tilde{\beta}_{jt} = \tilde{\beta}_{j,t-1} + \tilde{a}_{jt}$ does not change the equation unless the β_{j0} are known, in which case the π_{j1} can be set equal to zero. The requirement that the β_{j0} be known must in most cases be unreasonably demanding. All the time profiles of β_{jt} generated in the literature using the Kalman algorithm are probably wrong. Therefore, the β_{j0} employed in the literature are only loosely connected with a relation of \tilde{y}_t to \tilde{x}_t when such a relation exists and hence equation (1) with $\Phi = I$ is arbitrary. In this case, Δ_a is a Greek letter. As de Finetti once said to Lindley: "Stop thinking about Greek letters."⁵¹ The reason is that stochastic laws are not arbitrary, as shown by Pratt and Schlaifer⁵² and

⁴⁸See Priestley, M. B. (1981), *Spectral Analysis and Time Series*, New York: Academic Press, p. 814.

⁴⁹Berger, J.O. (1990), "Comment on 'The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics' by D.V. Lindley," *Statistical Science*, 5, 71-75.

⁵⁰Berger, *op. cit.*, p. 73.

⁵¹See p. 55 of Lindley, D.V. (1990): "The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics," *Statistical Science*, 5, 44-89.

⁵²Pratt and Schlaifer 1988, *op. cit.*, p. 35.

Basmann,⁵³ and arbitrary equations do not coincide with nonarbitrary laws except by accident. The above set of values of $\mathbf{\Pi}$, \mathbf{L} , and $\mathbf{\Phi}$ is unrealistic if the theories implying the relationship (1) are realistic. Thus, the values $\mathbf{\Pi} = \mathbf{0}$ and $\mathbf{\Phi} = \mathbf{I}$, like the fixed- β assumption, imply overly special prior distributions for the parameters of model (5).

Because of our inability to make arbitrarily fine distinctions in judgment (e.g., we cannot decide whether our subjective probability of rain tomorrow is .80 or .79), we cannot specify a priori highly accurate values of all relevant probabilities. The probabilities that will matter are determined by the eventual data and, therefore, the degree of accuracy necessary in the values of $\mathbf{\Pi}$, $\mathbf{\Phi}$, and $\sigma_a^2 \Delta_a$ to be used in the Bayesian forecasting method will often depend on the data that are impossible to assess a priori, as Berger⁵⁴ points out. This practical difficulty is eliminated by SCEP which does not require its users to supply the a priori values of $\mathbf{\Pi}$, $\mathbf{\Phi}$, and $\sigma_a^2 \Delta_a$. SCEP estimates all these parameters from the data. Using the estimates of these parameters given by several iterations of SCEP for several cases, you can consider a variety of $\mathbf{\Pi}$, $\mathbf{\Phi}$, and $\sigma_a^2 \Delta_a$ for your data, all of which (together with your own knowledge of the subject) convince you that as many reasonable possibilities have been covered as can be expected. Out of all these possibilities, a few yield excellent forecasts of the out-of-sample values of \mathbf{y} . We now ask you Berger's⁵⁵ question: Would you be happy, even if the development of the models and priors utilized your data in a certain way? Berger⁵⁶ said that he would be, as long as he felt that your models and priors covered the range of reasonable possibilities, and you did not purposefully fail to disclose models or priors that were reasonable and yet supported different conclusions.

The estimates of $\mathbf{\Pi}$, $\mathbf{\Phi}$, and $\sigma_a^2 \Delta_a$ given by several iterations of SCEP for several cases can show you how general the distributions of coefficients in equation (1) would have to be (within a certain class of distributions) in order to obtain good forecasts of \mathbf{y}_t in a certain period. If such distributions are not in the ballpark of your prior for the $\tilde{\beta}_t$, then the data on $\tilde{\mathbf{y}}_t$, \mathbf{x}_t , and \mathbf{z}_t in conjunction with your prior are incapable of giving a clear indication as to which model is preferable. Thus, you can use SCEP to calibrate your prior beliefs about the coefficients of equation (1).

Case XI: Cooley and Prescott's Stochastic Coefficients Model

Cooley and Prescott's⁵⁷ assumption about the β_{it} in equation (1) is

$$\begin{aligned} (XIa) \quad \mathbf{\Pi} &= \mathbf{0}, \\ (XIb) \quad \mathbf{m} &= \mathbf{K}, \\ (XIc) \quad \mathbf{L} &= \mathbf{I}_K, \end{aligned}$$

⁵³Basmann, *op. cit.*, p. 73.

⁵⁴Berger, *op. cit.*, p. 73.

⁵⁵Berger, *op. cit.*, p. 73.

⁵⁶Berger, *op. cit.*, p. 73.

⁵⁷Cooley, T. F. and E. C. Prescott (1976), "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica*, 44, 167-184.

$$\begin{aligned}
(XId) \quad \tilde{\beta}_t &= \tilde{\beta}_t^p + \tilde{\mathbf{u}}_t, \\
\tilde{\beta}_t^p &= \tilde{\beta}_{t-1}^p + \tilde{\mathbf{v}}_t,
\end{aligned} \tag{28}$$

where the $\tilde{\mathbf{u}}_t$ and $\tilde{\mathbf{v}}_t$ are identically and independently distributed normal variables with mean vectors 0 and covariance structures known up to scale factors.

Since the probability density function for $\tilde{\mathbf{y}}_t$, considered for fixed \mathbf{y}_t and $\mathbf{x}_t, t = 1, 2, \dots, T$, as a function of parameters, is called the likelihood function, the unconditional likelihood function for Cooley and Prescott's model is not defined unless the unconditional probability density function for $\tilde{\mathbf{y}}_t$ is defined. Under the restrictions $(XIa) - (XIc)$, this unconditional probability density function is not defined. However, the density conditional on a given initial value, β_0^p , is defined but the analysis based on this conditional density will heavily depend on the initial value. Cooley and Prescott's model has the same problems as the version of Kalman model considered in Case X.

Case XII: The Errors-In-Variables Nature of Equation (5)

Consider the case where the dependent and independent variables of an equation are measured with error. Let $\tilde{\mathbf{y}}_t = \mathbf{y}_t^* + \tilde{\mathbf{u}}_t$ be the dependent variable and $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it}^* + \tilde{\mathbf{v}}_{it}, i = 2, 3, \dots, K$, be the independent variables of this equation. Here we distinguish a true measurement from its observed value by an asterisk. For example, \mathbf{y}_t is the observed counterpart of the true value \mathbf{y}_t^* . The deviation of an observed value from the corresponding true measurement represents a measurement error. Suppose that the relationship between the true variables is given by

$$\mathbf{y}_t^* = \beta_1 + \beta_2 \mathbf{x}_{2t}^* + \dots + \beta_K \mathbf{x}_{Kt}^* + \tilde{\epsilon}_t. \tag{29}$$

Replacing each true variable in this equation by the corresponding observable variable minus its measurement error gives

$$\begin{aligned}
\tilde{\mathbf{y}}_t &= (\beta_1 + \tilde{\mathbf{u}}_t + \tilde{\epsilon}_t) + \beta_2 \left(1 - \frac{\tilde{\mathbf{v}}_{2t}}{\tilde{\mathbf{x}}_{2t}}\right) \tilde{\mathbf{x}}_{2t} + \dots + \beta_K \left(1 - \frac{\tilde{\mathbf{v}}_{Kt}}{\tilde{\mathbf{x}}_{Kt}}\right) \tilde{\mathbf{x}}_{Kt}, \\
&= \tilde{\beta}_{1t} + \tilde{\beta}_{2t} \tilde{\mathbf{x}}_{2t} + \dots + \tilde{\beta}_{Kt} \tilde{\mathbf{x}}_{Kt},
\end{aligned} \tag{30}$$

which is observationally equivalent to equation (1). This equivalence was previously noted by Swamy, et al.⁵⁸ SCEP can estimate equation (26) if the $\tilde{\mathbf{z}}_{it}$ in equation (2) are highly correlated with the $\left(1 - \frac{\tilde{\mathbf{v}}_{it}}{\tilde{\mathbf{x}}_{it}}\right)$. Thus, equation (5) is appropriate when the independent variables of equation (1) are measured with error, though these errors introduce dependence between the coefficients and the independent variables of equation (1).

⁵⁸Swamy, Kolluri and Singamsetti, *op. cit.*

Case XIII: An Equation with Smoothly Varying Coefficients

In some applications of the model given by equations (1)-(3), the estimated coefficients may change too abruptly, given prior knowledge. In this section, we attempt to state, in Bayesian terms, what this prior knowledge is, and then use this prior knowledge to derive posterior means.

Following Shiller,⁵⁹ we first define

$$\tilde{\epsilon}_{jt} = (1 - B)^{d+1} \tilde{\beta}_{jt} \quad (j = 1, 2, \dots, K; t = 1, 2, \dots, T) \quad (31)$$

where B is the backward shift operator operating only on t , i.e., $\tilde{\beta}_{j,t-i} = B^i \tilde{\beta}_{jt}$, d refers to the degree of smoothness priors, and the $\tilde{\epsilon}_{jt}$ with different t and the same j are independently and normally distributed with zero mean and constant variance, σ_{jj} . To incorporate the prior (27), we combine the procedures of Swamy and Tinsley⁶⁰, Thurman, Swamy, and Mehta⁶¹ and Kashyap, Swamy, Mehta, and Porter.⁶² SCEP does not yet have the capability of incorporating the prior (27).

⁵⁹Shiller, R. J. (1973), "A Distributed Lag Estimator Derived From Smoothness Priors," *Econometrica*, 41, 775-788.

⁶⁰Swamy and Tinsley, *op. cit.*

⁶¹Thurman, S. S., P. A. V. B. Swamy, and J. S. Mehta (1986), "An Examination of Distributed Lag Model Coefficients Estimated with Smoothness Priors," *Communications in Statistics: Theory and Methods*, 15, 1723-1750.

⁶²Kashyap, A. K., P. A. V. B. Swamy, J. S. Mehta, and R. D. Porter (1988), "Further Results on Estimating Linear Regression Models with Partial Prior Information," *Economic Modelling*, 5, 49-57.

Appendix C: A Numerically Stable Method of Computing Generalized Least Squares Estimators and their Standard Errors

For $t = 1, 2, \dots, T$, equation (5) can be compactly written as

$$\tilde{\mathbf{y}} = \mathbf{X}_z \text{vec}(\mathbf{\Pi}) + \tilde{\mathbf{u}}, \quad (32)$$

where $\tilde{\mathbf{y}}$ is a T -vector having $\tilde{\mathbf{y}}_t$ as its t th element, \mathbf{X}_z is a $T \times Kp$ matrix having $(\mathbf{z}'_t \otimes \mathbf{x}'_t)$ as its t th row, and $\tilde{\mathbf{u}}$ is a T -vector having $\mathbf{x}'_t \mathbf{L} \tilde{\boldsymbol{\epsilon}}_t$ as its t th element.

It follows from equation (3) that the covariance matrix of $\tilde{\boldsymbol{\epsilon}} = (\tilde{\boldsymbol{\epsilon}}'_1, \tilde{\boldsymbol{\epsilon}}'_2, \dots, \tilde{\boldsymbol{\epsilon}}'_T)'$ is

$$\mathbf{E} \tilde{\boldsymbol{\epsilon}} \tilde{\boldsymbol{\epsilon}}' = \sigma_a^2 \boldsymbol{\Sigma}_\epsilon = \sigma_a^2 \begin{bmatrix} \Gamma_0 & \Gamma_0 \Phi' & \Gamma_0 \Phi'^2 & \dots & \Gamma_0 \Phi'^{T-1} \\ \Phi \Gamma_0 & \Gamma_0 & \Gamma_0 \Phi' & \dots & \Gamma_0 \Phi'^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Phi^{T-1} \Gamma_0 & \Phi^{T-2} \Gamma_0 & \Phi^{T-3} \Gamma_0 & \dots & \Gamma_0 \end{bmatrix} \quad (33)$$

where $\sigma_a^2 \Gamma_0 = \sigma_a^2 \Phi \Gamma_0 \Phi' + \sigma_a^2 \Delta_a$, and the covariance matrix of $\tilde{\mathbf{y}}$ is

$$\sigma_a^2 \boldsymbol{\Sigma}_y = \mathbf{D}_x (\mathbf{I}_T \otimes \mathbf{L}) \sigma_a^2 \boldsymbol{\Sigma}_\epsilon (\mathbf{I}_T \otimes \mathbf{L}') \mathbf{D}_x', \quad (34)$$

where $\mathbf{D}_x = \text{diag}[\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_T]$ is a $T \times TK$ block diagonal matrix and \mathbf{I}_T is a $T \times T$ identity matrix.

When the restrictions on $\text{vec}(\mathbf{\Pi})$ given by equation (4) are true, the minimum variance linear unbiased estimator of $\text{vec}(\mathbf{\Pi})$ is

$$\begin{aligned} & \mathbf{R}'_\Pi (\mathbf{R}_\Pi \mathbf{R}'_\Pi)^{-1} \mathbf{r}_\Pi + \mathbf{C}'_\Pi (\mathbf{C}_\Pi \mathbf{X}'_z \boldsymbol{\Sigma}_y^{-1} \mathbf{X}_z \mathbf{C}'_\Pi)^{-1} \mathbf{C}_\Pi \mathbf{X}'_z \boldsymbol{\Sigma}_y^{-1} \\ & \times (\tilde{\mathbf{y}} - \mathbf{X}_z \mathbf{R}'_\Pi (\mathbf{R}_\Pi \mathbf{R}'_\Pi)^{-1} \mathbf{r}_\Pi), \end{aligned} \quad (35)$$

where \mathbf{C}_Π is such that $\mathbf{R}_\Pi \mathbf{C}'_\Pi = \mathbf{0}$.

The covariance matrix of the estimator (31) is

$$\sigma_a^2 \mathbf{C}'_\Pi (\mathbf{C}_\Pi \mathbf{X}'_z \boldsymbol{\Sigma}_y^{-1} \mathbf{X}_z \mathbf{C}'_\Pi)^{-1} \mathbf{C}_\Pi. \quad (36)$$

The formulas (31) and (32) corresponding to the values $\mathbf{R}_\Pi = \mathbf{0}$, $\mathbf{r}_\Pi = \mathbf{0}$, and $\mathbf{C}_\Pi = \mathbf{I}$ are the unrestricted estimator of $\text{vec}(\mathbf{\Pi})$ and its covariance matrix, respectively. SCEP evaluates both these formulas at the sample estimates of Φ , Δ_a and σ_a^2 for the general and special cases of equations (1)-(3). Clearly, the formulas (31) and (32) are not correct if $\boldsymbol{\Sigma}_y$ is singular and/or \mathbf{X}_z has less than full column rank. Even when they are correct, Kourouklis and Paige⁶³ have shown how some computational algorithms based on them can be numerically unreliable. These algorithms introduce large numerical errors when some of the nonzero eigenvalues of $\boldsymbol{\Sigma}_y$ are much smaller than others. A much more stable way of evaluating the formulas (31) and (32) using orthogonal transformations has been suggested by Paige.⁶⁴ It

⁶³Kourouklis, S. and C.C. Paige (1981), "A Constrained Least Squares Approach to the General Gauss-Markov Linear Model," *Journal of the American Statistical Association*, 76, 620-625.

⁶⁴Kourouklis and Paige, *op. cit.*, p. 622.

can be applied whether or not Σ_y is nonsingular and whether or not X_z has full column rank. The attractiveness of Paige's method is that all potential ill-conditioning is concentrated in certain triangular systems. Moreover, Paige⁶⁵ has shown that his procedure is numerically stable, something that is not true of any method that explicitly forms the formula (31). SCEP adopts Paige's method and compares the value of the matrix (32) with the value of the same matrix given by Paige's method. In principle, these two values should agree if Σ_y is nonsingular and X_z has full column rank.⁶⁶ The estimates of (30) given by SCEP are always nonnegative definite because SCEP computes only the lower triangular factor of $\sigma_a^2 \Sigma_y$ subject to the restriction that the diagonal elements of this factor are nonnegative. An algebraic form of this lower triangular factor is given in Chang et al.⁶⁷

In the initial iteration of SCEP, the estimator (31) reduces to

$$(X'_z[D_x(I_T \otimes LL')D'_x]^{-1}X_z)^{-1}X'_z[D_x(I_T \otimes LL')D'_x]^{-1}\tilde{y} \quad (37)$$

if $R_\Pi = 0$, $r_\Pi = 0$ and 0 and I are used as the initial values of Φ and Δ_a , respectively.

A general estimator of the covariance matrix of the estimator (33) is

$$(X'_z[D_x(I_T \otimes LL')D'_x]^{-1}X_z)^{-1}X'_z[D_x(I_T \otimes LL')D'_x]^{-1}D_x(I_T \otimes L)\hat{\sigma}_a^2\hat{\Sigma}_\epsilon \\ \times (I_T \otimes L')D'_x[D_x(I_T \otimes LL')D'_x]^{-1}X_z(X'_z[D_x(I_T \otimes LL')D'_x]^{-1}X_z)^{-1}X_z \quad (38)$$

where $\hat{\sigma}_a^2\hat{\Sigma}_\epsilon$ is evaluated by SCEP by replacing σ_a^2 , Φ , and Δ_a in (30) by their sample estimates. SCEP computes the square roots of the diagonal elements of the matrix (34). This method is much more accurate than White's⁶⁸ method. The asymptotic distribution of the estimator (33) is derived by Havenner and Swamy.⁶⁹

The formula (34) gives the standard errors of the unweighted least squares estimator of $vec(\Pi)$ if $D_x(I_T \otimes LL')D'_x$ is set equal to I_T . This can be done by using I_K in place of L and the values of Φ and Δ_a given in Case I as the initial values of Φ and Δ_a , respectively.

Thus, the standard errors of the estimates of Π depend on the formula used to compute them. Since it is not known which of these formulas give the standard errors that are close to the true values of the square roots of the diagonal elements of (32), it is difficult to interpret the results of tests based on these standard errors. So there are limits to the usefulness of hypothesis testing. Furthermore, rejection of a null hypothesis, say H_0 , by a test is not proof that H_0 is false or acceptance of H_0 is not proof that H_0 is true. In any test of H_0 against an alternative hypothesis, say H_1 , finding the value of the test statistic which lies in the critical region should not be taken to be strong evidence against H_0 and for H_1 if both H_0 and H_1 are false. This is because in this case, the probabilities of the critical region calculated

⁶⁵See Kourouklis and Paige, *op. cit.*, and the references cited therein.

⁶⁶For further discussion, see Swamy, Mehta and Singamsetti, *op. cit.*

⁶⁷See Chang et al., *op. cit.*, p. 45.

⁶⁸White, H. (1978), "A Heteroscedastic Consistent Covariance Matrix and a Direct Test for Heteroscedasticity," *Econometrica*, 46, 817-838.

⁶⁹Havenner, A. and P.A.V.B. Swamy (1981), "A Random Coefficient Approach to Seasonal Adjustment of Economic Time Series," *Journal of Econometrics*, 15, 177-210.

under H_0 and H_1 are *not* true. Alternatively stated, a test of any null hypothesis against an alternative provides fictitious evidence against the null hypothesis or the alternative if both the null and alternative hypotheses are false.

Appendix D: A Stochastic Coefficients Approach to Forecasting

Equations (1)-(3) provide a useful approach for the investigation of forecast error sources. For this purpose, partition several of the vectors and matrices of these equations as follows:

$$\mathbf{x}'_t = (1, \mathbf{z}'_{2t}), \quad \mathbf{z}'_t = (1, \mathbf{z}'_{2t}), \quad \mathbf{\Pi} = (\boldsymbol{\pi}_1, \mathbf{\Pi}_2), \quad \mathbf{L}' = (L_1, L_2),$$

where $\boldsymbol{\pi}_1$ is the first column of $\mathbf{\Pi}$ and L'_1 is the first row of \mathbf{L} . Equation (5) may then be expressed as

$$\begin{aligned} \tilde{\mathbf{y}}_t &= \mathbf{x}'_t(\boldsymbol{\pi}_1, \mathbf{\Pi}_2) \begin{pmatrix} 1 \\ \mathbf{z}_{2t} \end{pmatrix} + (1, \mathbf{z}'_{2t}) \begin{pmatrix} L'_1 \tilde{\epsilon}_t \\ L'_2 \tilde{\epsilon}_t \end{pmatrix} \\ &= \mathbf{x}'_t \boldsymbol{\pi}_1 + \mathbf{x}'_t \mathbf{\Pi}_2 \mathbf{z}_{2t} + L'_1 \tilde{\epsilon}_t + \mathbf{z}'_{2t} L'_2 \tilde{\epsilon}_t. \end{aligned} \quad (39)$$

Equation (1) with fixed coefficients implies a forecast of \mathbf{y} in some future period s after T , given by $\hat{\mathbf{y}}_{T+s}$,

$$\hat{\mathbf{y}}_{T+s} = \hat{\mathbf{x}}'_{T+s} \hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_{T+s}. \quad (40)$$

The forecast **error**—the difference between the future realization, \mathbf{y}_{T+s} , and the prediction, $\hat{\mathbf{y}}_{T+s}$ —that arises from using any one of the fixed-coefficients versions of equation (1) may be decomposed as

$$\begin{aligned} \hat{\mathbf{y}}_{T+s} - \mathbf{y}_{T+s} &= \hat{\mathbf{x}}'_{T+s}(\hat{\boldsymbol{\beta}} - \boldsymbol{\pi}_1) + (\hat{\mathbf{x}}'_{T+s} - \mathbf{x}'_{T+s})\boldsymbol{\pi}_1 + (\hat{\mathbf{u}}_{T+s} - L'_1 \epsilon_{T+s}) \\ &\quad - \mathbf{x}'_{T+s} \mathbf{\Pi}_2 \mathbf{z}_{2,T+s} - \mathbf{z}'_{2,T+s} L'_2 \epsilon_{T+s}, \end{aligned} \quad (41)$$

which, in order of appearance, is the sum of (i) a linear combination of the sampling errors and bias in the fixed-coefficient estimates, (ii) a linear combination of the errors in predicting future values of the independent variables, (iii) the errors in predicting stochastic shifts in the intercept, (iv) the failure to predict deterministic shifts in intercept and nonconstant regressors' coefficients, and (v) the failure to predict stochastic shifts in nonconstant regressors' coefficients. Except for (ii), all these forecast error sources are accounted for when equation (5) is used. Observe that an accounting of forecast error sources based on fixed-coefficient estimates is limited to (i) and (iii), alone. The remaining error sources may not be diagnosed using fixed-coefficients models. The error resulting from (ii) is, of course, beyond the reach of any of the above models, since they originate from errors in forecasting independent variables.⁷⁰

SCEP uses the estimates of $\boldsymbol{\pi}_1, \mathbf{\Pi}_2, E(\tilde{\epsilon}_{T+s} | \mathbf{y}_1, \dots, \mathbf{y}_T, \mathbf{x}_1, \dots, \mathbf{x}_{T+s}, \mathbf{z}_1, \dots, \mathbf{z}_{T+s})$ and the values of \mathbf{x}_{T+s} and \mathbf{z}_{T+s} in place of the true values used in equation (35) to forecast the value of $\tilde{\mathbf{y}}$ in $T + s$. It prints the forecasts of both $(\mathbf{x}'_{T+s} \boldsymbol{\pi}_1 + \mathbf{x}'_{T+s} \mathbf{\Pi}_2 \mathbf{z}_{2,T+s})$ and \mathbf{y}_{T+s} . It also computes the root mean squared error of these

⁷⁰For further discussion, see Swamy, P.A.V.B., R.K. Conway, and M.R. LeBlanc (1989), "The Stochastic Coefficients Approach to Econometric Modeling, Part III: Estimation, Stability, and Prediction," *The Journal of Agricultural Economics Research*, 41, 4-20.

forecasts. However, it does not yet have the capability of computing the terms on the right-hand side of equation (37).

At the end of Section 2, it is mentioned that SCEP permits you to validate model (5) and its special cases. Suppose that use of one of these models produced improvement in terms of root mean squared error of forecasts relative to all the other models you considered. Then you can argue that this model is useful. You cannot, however, *prove* that this model is a true model. To indicate why this is the case, make the proviso that you know that one of a finite number of models is true, but you do not know which one. Suppose also that specification errors lead to poor predictions. Then it is correct to argue that focusing on the comparisons of out-of-sample forecasts is preferable to statistical testing of hypotheses because specification errors lead to poor predictions which are more likely to be discerned with out-of-sample forecasting methods than with statistical tests. The practical fault in this reasoning is that the above proviso is often not present in any realistic fashion if the set of alternative models you consider is very small. It is also incorrect to suppose that specification errors lead to poor predictions because sometimes naive models predict quite well. The above proviso may be present in a realistic fashion if you consider a very general model of the form (5) which covers a variety of fixed-coefficient models as special cases. The reason is that a general model can be true even though its special cases are false.