STOCHASTIC TRENDS IN SIMULTANEOUS EQUATION SYSTEMS

by

Mariane Streibel

1992

Thesis submitted for the Degree of Doctor of Philosophy

at the London School of Economics and Political Science,

University of London

ABSTRACT

The estimation of univariate and multiple regression models with stochastic trend components has been considered in the time domain and in the frequency domain. Such models assume as regressors weakly exogenous variables. However if the regression equations are part of a simultaneous equation system some of the regressors will no longer be weakly exogenous and estimators obtained by ignoring this fact will be inconsistent.

One way of proceeding in such situations is to estimate the whole system, that is, to construct full information maximum (FIML) estimators. Alternatively, single equation estimators such as limited information maximum likelihood (LIML) can be constructed, as well as estimators based on the instrumental variable (IV) principle which possess the merit of consistency.

As in the analogous situation in classical simultaneous equation systems, within this class of limited information estimators, LIML is asymptotically efficient. Hence it is appropriate to study the asymptotic properties of LIML and review the possibility of alternative consistent estimators, using LIML as a benchmark.

The purpose of the thesis is thus:

- to examine the issues of identifiability when stochastic trends

are present in simultaneous equation systems;

- to examine the computational issues associated with FIML, LIML and various IV estimators in simultaneous equation systems with stochastic trends and derive the asymptotic properties in the frequency domain of these estimators;
- to compare the performance of IV and LIML via Monte Carlo experiments;
- to apply the methods to real data.

ACKNOWLEDGMENTS

I am grateful for the constant support and encouragement I have received from many friends during this research. I would particularly like to thank my supervisor Professor A.C.Harvey for his invaluable help and advice.

Many of my colleagues at the Institute of Mathematics and Statistics of the University of São Paulo have given initial encouragement. I would like to thank them all, in particular Dr. P.A.Morettin who was my MSc supervisor. I also would like to thank Dr. H.G.Calhau who first introduced me to statistics and Dr. G.B.Silveira for his helpful comments on this thesis.

Finally, I gratefully acknowledge partial financial support from CNPQ and from FAPESP.

LIST OF CONTENTS

LIST OF TABLES

Appendix 8.1

Appendix 9.1

CHAPTER 1

INTRODUCTION

Stochastic trend components are introduced into econometric equations when the level of a nonstationary dependent variable cannot be completely explained by observable explanatory variables. The presence of a stochastic trend can often be rationalised by the fact that a variable has been excluded from the equation because it is difficult, or even impossible, to measure. Thus in Harvey *et al*(1986) and in Slade(1989) a stochastic trend is used as a proxy for technical progress, while in the demand equation for UK spirits estimated by Kohn and Ansley(1989) the stochastic trend can be thought of as picking up changes in tastes. Such rationalisation not only lends support to the specification of the model, but it also means that the estimated stochastic trend can be analysed and interpreted.

Economic theory often suggests the appearance of stochastic trend components in particular equations within a simultaneous equation system. Indeed many published econometric models contain a time trend. For example the wage equation in the textbook Klein model has a time trend which is included to account for union pressure. As in single equations, such effects are more appropriately modelled by stochastic trends. If they are not explicitly modelled, their effects will be picked up indirectly by time trends and lags on the variables. This can lead to a proliferation of lags which have no economic meaning, and which are subject to common factors and problems of inference associated with unit roots; see Harvey *et al*(1986). Thus if economic theory suggests the presence of stochastic trends there are likely to be considerable gains from estimating the implied structural relationships directly.

The focus of this thesis is on models where the behaviour of a dependent variable is explained by observable explanatory variables and unobservable components. The unobservable components are modelled using the ideas of structural time series. Thus the components have a direct interpretation, see Harvey(1989).

When the explanatory variables are weakly exogenous variables we shall refer to the model as a time series regression model. Examples include the seat belt study of Harvey and Durbin(1986) as well as the application by Harvey *et al* referred to earlier. Multivariate structural time series, in particular seemingly unrelated time series equations (SUTSE) models were studied in Fernandez(1986) and Fernandez and Harvey(1990), while the inclusion of explanatory variables in SUTSE models is examined in Marshall(1990) and Harvey and Marshall (1991).

Our interest centres on a single equation within a simultaneous equation system with stochastic trend components. The estimation of time series regression models is based on the maximum likelihood principle and the assumption that the regressors are weakly exogenous is crucial. However, if some of the regressors are not assumed to be weakly exogenous variables, the maximum likelihood criterion function will not be a valid basis for inference. In simultaneous equation systems some of the regressors are endogenous variables to the system and estimators obtained by ignoring this fact will be inconsistent.

In order to obtain consistent estimators we have to proceed as we would in the classical simultaneous equation systems, that is, without stochastic trends. So, if the complete system of equations can be specified, a full information maximum likelihood (FIML) procedure may be employed. If only a subsystem is specified, but all the predetermined variables are named, a limited information maximum likelihood (LIML) procedure is appropriate. When the rest of the system has not been specified at all, ML methods cannot be applied, but a valid instrumental variable (IV) estimator can be obtained.

As in the analogous situation in classical simultaneous equation systems, within this class of limited information estimators, LIML is asymptotically efficient. Hence it is appropriate to study the asymptotic properties of LIML and review the possibility of alternative consistent estimators, using LIML as a benchmark.

A well known result in classical simultaneous equation systems is that LIML can be obtained by applying FIML to a ''new'' system formed from the structural equation of interest and the reduced form corresponding to the endogenous variables included in the equation of interest. It turns out that this new system is a triangular one. This is also true for models with stochastic trends components. The estimation of triangular systems is somewhat easier since such systems can be formulated as a set of seemingly unrelated regression equations (SURE) with stochastic trend components and can be carried out in the time domain framework.

Unfortunately, the triangular property is not helpful in deriving the asymptotic properties of LIML. In order to obtain the asymptotic properties we have to study the properties of FIML. The frequency domain framework turns out to be most appropriate.

The purpose of the thesis is thus:

(a) to examine the computational issues associated with FIML and LIML in simultaneous equation systems with stochastic trends;

(b) to derive the asymptotic properties of FIML and LIML;

(c) to examine the computational issues arising with various IV estimators;

(d) to derive asymptotic properties of viable IV procedures;

(e) to compare IV and LIML on the basis of asymptotic theory and Monte Carlo experiments;

(f) to examine the issues of identifiability when stochastic trends are present;

(g) to apply the methods to real data.

The plan of the thesis is as follows.

In chapter 2 we review some standard results which are needed to handle multivariate structural time series models. We look at state space form models and discuss estimation in the time domain and in the frequency domain, as well as asymptotic properties of the estimators. Chapter 3 provides a basis for the estimation of a single equation within a simultaneous equation system, as well as of the whole system.

Chapter 4 contains material on alternative limited information estimators based on the instrumental variable principle. Several time-domain instrumental variable estimators for single equations with stochastic trend are presented. We also deal with frequency-domain instrumental variable estimators and their asymptotic properties.

In chapter 5 we introduce simultaneous equation systems with stochastic trend components and discuss the role played by stochastic trends in helping to identify a single equation in the system.

The purpose of chapter 6 is to derive the asymptotic properties of FIML. As mentioned earlier, LIML is a special case of FIML and so to obtain the asymptotic properties of LIML we have to consider those of FIML. Again these properties are derived in the frequency domain. We also present a computational method for FIML itself, based on the nonparametric approach of Hannan and Terrell(1973), and asymptotically efficient two-step full information estimators.

However we have not computed such estimators. The reason is because from our experience with LIML we thought that in order to able to make meaningful comparisons a complete study of FIML should be done.

In chapter 7, we extend the results given in Hall and Pagan (1981) in order to provide a computational method for LIML when the system contains stochastic trends. We also compare the asymptotic distribution of LIML with that of our preferred IV estimators. We determine the conditions under which the IV estimator has the same efficiency as LIML.

An application to the employment-output equation is presented in chapter 8. A series of Monte Carlo experiments are reported in chapter 9. Finally the conclusions are presented in chapter 10.

CHAPTER 2

STRUCTURAL TIME SERIES MODELS

1. Introduction

In this chapter we review some standard results which are needed to handle multivariate structural time series models. We look at state space form models and discuss the estimation in the time domain and in the frequency domain, as well as asymptotic properties of the estimators. We also present in appendix, a brief review of optimisation procedures.

2. State Space Form Models

The models that will be considered here have a time invariant state space form given by

$$y_t = Z\,\alpha_t + \varepsilon_t \qquad \text{(measurement equation)} \qquad (1a)$$

$$\alpha_t = T\,\alpha_{t-1} + \eta_t \qquad \text{(transition equation)} \qquad (1b)$$

$t=1,\ldots T$, where y_t is a px1 vector of observable variables, α_t is a

$\bar{m}$x1 vector of unobservable variables, known as the state vector, Z is a px$\bar{m}$ matrix , T is a $\bar{m}$x$\bar{m}$ matrix, ϵ_t is a px1 vector of serially uncorrelated disturbances with mean zero and covariance matrix Σ_ϵ and η_t is a $\bar{m}$x1 vector of serially uncorrelated disturbances with mean zero and covariance matrix Σ_η. We also assume that ϵ_t and η_t are normally distributed and uncorrelated with each other for all periods of time and with the initial state vector α_o which is assumed to have a normal distribution with mean a_o and covariance matrix P_o.

Although ARMA models can be cast in the space state form we shall only consider nonstationary structural time series models. Specifically, the i-th series, y_{it} i=1,...p, may be modelled as

a) *a local linear trend model*, that is,

$$y_{it} = \mu_{it} + \epsilon_{it} , \tag{2a}$$

$$\mu_{it} = \mu_{i,t-1} + \beta_{i,t-1} + \eta_{it} , \tag{2b}$$

$$\beta_{it} = \beta_{i,t-1} + \zeta_{it} ; \tag{2c}$$

b) *a random walk plus noise model*, that is,

$$y_{it} = \mu_{it} + \epsilon_{it} , \tag{3a}$$

$$\mu_{it} = \mu_{i,t-1} + \eta_{it} ; \tag{3b}$$

c) or simply as a sequence of independent variables.

We note that (2) and (3) may be formulated as

$$y_{it} = z_i'\alpha_{it} + \epsilon_{it} ,$$

$$\alpha_{it} = T_i\alpha_{i,t-1} + \eta_{it} ,$$

where for the random walk plus noise, $\alpha_{it}=\mu_{it}$, $z_i'=1$ and $T_i=1$, whereas for the local linear trend

$$\alpha_{it} = \begin{bmatrix} \mu_{it} \\ \beta_{it} \end{bmatrix} \quad ; \quad z_i' = [1 \ , \ 0] \quad ; \quad T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Thus for each series, z_i' and T_i are known and fixed.

SUTSE Models

If all series have the same state form, that is, $z_1'=z_2'=\ldots=z_p'=z'$ and $T_1=T_2=\ldots=T_p=T$, (1) becomes

$$y_t = (z' \otimes I_p)\ \alpha_t + \varepsilon_t \tag{4a}$$

$$\alpha_t = (T \otimes I_p)\alpha_{t-1} + \eta_t \tag{4b}$$

where α_t and η_t, are of dimension $\bar{m}x1$. We remark that in (4b), T is mxm, while in (1b) T is $\bar{m}x\bar{m}$, where $\bar{m}=pm$, and m=1 if each series follows a random walk plus noise and m=2 if each series follows a local linear trend model. The associated parameters are the pxp covariance matrix Σ_ε and the pmxpm covariance matrix Σ_η. The distinct elements of these matrices are known as the hyperparameters and will be denoted by the vector ψ. Such models are known as SUTSE (Seemingly Unrelated Time Series Equations), see Harvey (1989, page 432) for a comprehensive study. The simplest SUTSE model is the multivariate random walk plus noise, obtained when m=1. Thus

$$y_t = \begin{bmatrix} \mu_{1t} \\ \cdot \\ \mu_{pt} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \cdot \\ \varepsilon_{pt} \end{bmatrix} \tag{5a}$$

$$\mu_t = \begin{bmatrix} \mu_{1t-1} \\ \cdot \\ \mu_{pt-1} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \cdot \\ \eta_{pt} \end{bmatrix} \tag{5b}$$

and both associated covariance matrices are of order (pxp). Because the matrices Σ_ε and Σ_η are symmetric it will prove convenient to define ψ, as being the p(p+1)x1 vector obtained from $\mathrm{vec}(\Sigma_\varepsilon:\Sigma_\eta)$ by eliminating all supradiagonal elements of Σ_ε and Σ_η. Following Magnus and Neudecker (1988,page 49) we have

$$\begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix} \psi = \mathrm{vec}[\Sigma_\varepsilon:\Sigma_\eta] \tag{6}$$

and

$$\psi = \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \mathrm{vec}[\Sigma_\varepsilon:\Sigma_\eta] \tag{7}$$

where the $p^2 \times \frac{1}{2}p(p+1)$ matrix D is the duplication matrix and D^+ is the Moore-Penrose inverse of D, given by

$$D^+ = (D'D)^{-1}D'. \tag{8}$$

Prediction Error Decomposition

Given the normality assumption on the initial state α_0 and on the disturbances ε_t and η_t, $y = \mathrm{vec}\,[y_1, \ldots, y_T]$ will have a multivariate normal distribution with mean μ (Tpx1) and covariance matrix Ω (TpxTp). Now the density of y can be written in terms of the conditional densities, that is,

$$f(y|\mu,\Omega) = \prod_{t=2}^{T} \left[f(y_t|Y_{t-1},\mu,\Omega)\right] f(y_1|\mu,\Omega),$$

where $Y_{t-1}=(y_{t-1},\ldots,y_1)$. Therefore the density of y becomes

$$f(y|\mu,\Omega) = (2\pi)^{-\frac{1}{2}T} \prod_{t=1}^{T} |F_t|^{-\frac{1}{2}} \, e^{-\frac{1}{2}\sum_{t=1}^{T}(y_t-Ey_t|Y_{t-1})'F_t^{-1}(y_t-Ey_t|Y_{t-1})} \qquad (9)$$

where F_t is the conditional covariance matrix of y_t given $y_{t-1},y_{t-2},\ldots y_1$.

It can be shown, see Harvey (1981,page 13), that $Ey_t|Y_{t-1}$ and F_t are respectively the MMSE (Minimum Mean Square Estimator) of y_t given Y_{t-1} and its MSE (Mean Square Error) matrix.

Once the model is formulated in a state space form these prediction errors can be obtained from the Kalman filter equations, see Schweppe (1965) and among others Harvey (1981). The Kalman filter equations will be given in next sub section.

Kalman Filter Equations

When the model is cast in the state space form the parameters θ are elements of a_o, P_o and ψ, where ψ is the vector containing the distinct hyperparameters, rather than (μ,Ω). Clearly there is a relation between θ and (μ,Ω). The assumption of normality of the initial state and disturbances implies that the process $\{y_t,\alpha_t\}$

t=1,... is jointly Gaussian, and therefore the MMSE of α_t given Y_t and the information at time t=0 is given by

$$a_t = E\left[\alpha_{t/Y_t}\right] \tag{10a}$$

with associated MSE

$$P_t = E_t(\alpha_t - a_t)(\alpha_t - a_t)' \tag{10b}$$

and the MMSE of α_t given Y_{t-1} and the information at time t=0 is given by

$$a_{t/t-1} = E\left[\alpha_{t/Y_{t-1}}\right] \tag{11a}$$

with associated MSE

$$P_{t/t-1} = E_{t-1}(\alpha_{t/t-1} - a_{t/t-1})(\alpha_{t/t-1} - a_{t/t-1})' . \tag{11b}$$

The notation $E_t[\cdot]$ indicates the conditional expectation given Y_t and the information at time t=0. The necessary equations to compute these quantities are known as the Kalman filter equations and are:

the *prediction equations*

$$a_{t/t-1} = T\, a_{t-1} \tag{12a}$$

$$P_{t/t-1} = T\, P_{t-1}T' + \Sigma_\eta \qquad t=1,\dots T \tag{12b}$$

and *the updating equations*

$$a_t = a_{t/t-1} + P_{t/t-1}\, Z'F_t^{-1}\, \nu_t \tag{13a}$$

$$P_t = P_{t/t-1} - P_{t/t-1}\, Z'F_t^{-1}\, Z\, P_{t/t-1} \qquad t=1,\dots T \tag{13b}$$

where

$$F_t = Z\, P_{t/t-1}\, Z' + \Sigma_\varepsilon \qquad (14a)$$

$$\nu_t = y_t - Z\, a_{t/t-1}. \qquad t=1,\dots T \qquad (14b)$$

Note that from (1a) and (11a) we have

$$y_t - Ey_t|Y_{t-1} = y_t - Za_{t|t-1} = \nu_t \qquad (15a)$$

and

$$E_{t-1}(y_t - Ey_t|Y_{t-1})(y_t - Ey_t|Y_{t-1})'$$

$$= E_{t-1}[Z(\alpha_t - a_{t|t-1}) + \varepsilon_t][Z(\alpha_t - a_{t|t-1}) + \varepsilon_t]' = F_t \qquad (15b)$$

where the last equality in (15b) follows from (11b).

Thus, omitting additive constants that do not depend on the parameters, the loglikelihood function of y takes the form

$$\ell(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \log |F_t| - \frac{1}{2} \sum_{t=1}^{T} \nu_t' F_t^{-1} \nu_t \qquad (16)$$

where $\theta = (a_o, P_o, \psi)$, and $\nu_t = \nu_t(\theta)$ and $F_t = F_t(P_o, \psi)$ are obtained from the Kalman filter equations with starting a_o and P_o.

The Link between Cholesky Decomposition and State Space Techniques

If Ω is positive definite it can be factorized (Cholesky decomposition) in such a way that $\Omega^{-1} = L'F^{-1}L$, where L is a lower triangular matrix with *1*'s on the diagonal and F is a diagonal matrix. In multivariate models it turns out that L has $\mathrm{diag}(I_p,\dots,I_p)$

on its main block diagonal and F is a block diagonal matrix, i.e., $F = \mathrm{diag}\{F_1 \ldots F_T\}$.

We shall now derive the matrix L for the univariate random walk plus noise model in order to show the obvious result that the prediction errors ν_t, $t=1,\ldots T$ delivered by the Kalman filter can be written as $V=L(y-\mu)$.

The matrix L will be derived for notational rather than computational purposes. In fact, in practice we never perform the Cholesky decomposition, since the major advantage of the Kalman filter is exactly to avoid the storage of a high dimensional matrix such as L.

By repeated substitution of μ_t in the measurement equation we have

$$y_t = \mu_o + \sum_{j=1}^{t} \eta_j + \epsilon_t .$$

Hence, the mean of y_t is constant and equal to a_o and the relationship between Ω and P_o and the hyperparameters, $\psi = (\sigma_\epsilon^2, \sigma_\eta^2)$, is

$$\sigma_{ii} = P_o + i\,\sigma_\eta^2 + \sigma_\epsilon^2, \qquad i=1,\ldots T$$

$$\sigma_{ij} = P_o + k\,\sigma_\eta^2\ , \qquad k=\min(i,j),\ i,j=1,\ldots T.$$

Substituting these values in Ω, constructing the Cholesky decomposition of Ω and inverting the triangular matrix yields

$$L = \begin{bmatrix} 1 & 0 & 0 & 0\ldots0 & 0 \\ -\frac{P_{1/0}}{f_1} & 1 & 0 & 0\ldots0 & 0 \\ -\frac{P_{1/0}}{f_1}\left[1-\frac{P_{2/1}}{f_2}\right] & -\frac{P_{2/1}}{f_2} & 1 & 0\ldots0 & 0 \\ -\frac{P_{1/0}}{f_1}\left[1-\frac{P_{2/1}}{f_2}\right]\left[1-\frac{P_{3/2}}{f_3}\right] & -\frac{P_{2/1}}{f_2}\left[1-\frac{P_{3/2}}{f_3}\right] & -\frac{P_{3/2}}{f_3} & 1\ldots0 & 0 \\ \vdots & & & \vdots & \vdots \\ -\frac{P_{1/0}}{f_1}\left[1-\frac{P_{2/1}}{f_2}\right]\ldots\left[1-\frac{P_{T-1/T-2}}{f_{T-1}}\right] & \ldots & \ldots & -\frac{P_{T-1/T-2}}{f_{T-1}} & 1 \end{bmatrix} \tag{17a}$$

and

$$F = \mathrm{diag}(f_1,\ldots,f_T), \tag{17b}$$

where $P_{t/t-1}$ is as in (12b) and f_t is as in (14a). It is easy to verified that $V = L(Y - 1a_o)$ is the T-dimensional vector containing the prediction errors given in (14b). It is interesting to note that when the initial state is regarded as fixed these prediction errors are identical to the prediction errors obtained by means of the Rosenberg (1973) algorithm.

The relationship between (μ,Ω) and (a_o,P_o,ψ) can be easily established for the local linear trend, as well as for SUTSE models. We write

$$y_t = (z'\otimes I_p)(T\otimes I_p)^t\,\alpha_o + (z'\otimes I_p)\sum_{j=1}^{t}(T\otimes I_p)^{t-j}\,\eta_j + \varepsilon_t\ , \tag{18}$$

hence

$$E\ y = X_o a_o \tag{19}$$

where

$$y = \mathrm{vec}[y_1,\ldots,y_T]$$

and

$$X_o = [(z'T^1 \otimes I_p)', \ldots, (z'T^T \otimes I_p)']', \tag{20}$$

and analogously as before the lower triangular and the block diagonal matrices L and F can be constructed with elements being functions of Z, T, ψ and P_o. (Note that the superscript T in (2) is the sample size). Therefore (16) can be also written as

$$\ell(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2} L(y-X_o a_o)'F^{-1}L(y-X_o a_o) \tag{21}$$

where $\theta=(a_o, P_o, \psi)$ and $y = \text{vec}[y_1, y_2, \ldots, y_T]$.

Conditional Likelihood Function

As it stands the loglikelihood function given in (21) is a function of θ, the distinct parameters which enter into (a_o, P_o, ψ). Often prior information on the initial state is available and of course should be taken into account. Typically prior information arises when the process is stationary or when the initial state may be regarded as fixed. When this is the case the Kalman filter yields the exact likelihood function for $Y_T=(y_1, \ldots, y_T)$ via the prediction error decomposition. For the models considered here however, no prior information is avaiable. de Jong (1988) derived an expression for the likelihood function of Y_T, not conditional on α_o, where the place of both a_o and P_o is made explicit in the likelihood function. He pointed out that it is not possible to find the ML estimates of both a_o and P_o. He also justifies the specification of the unconditional distribution of α_o in terms of a diffuse or noninformative prior. This kind of specification can be interpreted as if the process has started in the remote past. In particular, for the random walk plus noise model, assuming that the process has started at time s, s<0, repeated substitution for α_t in the transition equation yields

$$\alpha_o = \sum_{j=s+1}^{0} \eta_j + \alpha_s .$$

The diffuse prior of α_o is obtained as $s \rightarrow -\infty$. Certain caution, however, should be taken when we say that the process has started in the remote past. First because there might be no physical interpretation, usually an economic time series has started in some finite time and second Y_T will have an improper distribution since all elements of Y_T will have infinite variance.

Nevertheless, although the unconditional distribution of Y_T is not defined, the conditional distribution of $y_T,\ldots y_{m+1}$, given $y_1,\ldots y_m$ is defined. We note that in univariate models, if P_o is bounded, then conditionally on $y_1,\ldots y_m$, $y_T,\ldots y_{m+1}$ is normally distributed, with t-th element of the (T-m)x1 mean vector being $ZT^{t-m}a_m$. The (T-m)x(T-m) covariance matrix is $(\bar{L}'\bar{F}^{-1}\bar{L})^{-1}$ where $\bar{L}$ and $\bar{F}$ are obtained by eliminating the first m rows and columns of L, and F, and $L'F^{-1}L$ is the covariance matrix of the unconditional distribution of Y_T. Now if P_o goes to infinity, it is easy to verify that for the univariate random walk plus noise model,

$$a_1 \rightarrow y_1 \tag{22a}$$

$$P_1 \rightarrow \sigma_\varepsilon^2 \tag{22b}$$

and for the local trend model

$$a_2 \rightarrow \begin{bmatrix} y_2 \\ y_2 - y_1 \end{bmatrix} \tag{23a}$$

$$P_2 \rightarrow \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 & 2\sigma_\varepsilon^2+\sigma_\eta^2+\sigma_\zeta^2 \end{bmatrix} . \tag{23b}$$

On the other hand, rather than assuming a diffuse prior for the initial state we could construct a proper prior for α_m from the first observations. It turns out, however, that the resulting estimators of the mean and variance of α_m are the same as the converging values given in (22) and (23) for the random walk and local trend models respectively. In other words, the use of a diffuse prior is equivalent to constructing a proper prior from the first m observations, in the sense that either would result in the same conditional likelihood function, see Harvey(1989, pages 120-128).

For complex multivariate models it is not always clear how to construct a proper prior from the first observations. Therefore the conditional likelihood function is obtained starting the filter with a_o and P_o=$KI_{\bar{m}}$ where K is a large finite number, $I_{\bar{m}}$ is the $\bar{m}$x$\bar{m}$ identity matrix, and $\bar{m}$ is the dimension of the state. Initial observations are discarded. Alternatively the conditional likelihood function can be computed by means of an algorithm devised by de Jong (1988,1991). Nevertheless, since the multivariate model that we are primarily interested in is the multivariate random walk plus noise given in section 2, (22) can be generalized straightforwardly. Thus a_1 becomes the px1 vector y_1 while P_1 the pxp matrix Σ_ε. From the above discussion the resulting conditional loglikelihood function becomes

$$\ell(\psi) = -\frac{1}{2}\sum_{t=m+1}^{T}\log |F_t| - \frac{1}{2}(\bar{L}(\bar{y} - X_o a_m)'\ \bar{F}^{-1}\ \bar{L}(\bar{y} - X_o a_m), \qquad (24)$$

where $\bar{y}=\text{vec}(\tilde{y}_{m+1},..\tilde{y}_T)$. For the models we shall consider onwards the $p(T-m)xpm$ matrix X_o, becomes

- a T-1 vector of ones for the univariate random walk;
- T-1 identity matrices of order p stacked together for the multivariate random walk ;
- a T-2 vector with t-th element equal to $z'T^{t-2}$ for the local trend model.

For a given ψ, $\ell(\psi)$ is evaluated applying the Kalman filter to y_t, $t=m,...T$, with starting values a_m and P_m as discussed above. Often we shall write (24) as

$$\ell(\psi) = -\frac{1}{2}\sum_{t=m+1}^{T}\log |F_t| - \frac{1}{2}\sum_{t=m+1}^{T}\nu_t' F_t^{-1}\nu_t \qquad (25)$$

where ν_t and F_t ,$t=m+1,...T$, are respectively the prediction errors and MSEs delivered by the Kalman filter.

In univariate models the place of one of the m+1 hyperparameters contained in ψ can be made explicit in the loglikelihood function. This can be done by scaling the hyperparameters. That is, in the Kalman filter equations, P_m is replaced by $P_m^*=P_m/\sigma_*^2$ and ψ by $(1,\psi_*)$ where σ_*^2 is the hyperparameter whose place is made explicit and ψ_* is the mx1 vector containing the remaining m scaled hyperparameters. It turns out that the prediction errors delivered by the Kalman filter with scaled hyperparameters will be unaffected whereas their MSEs, f_t^*, will also be scaled. The resulting loglikelihood function then becomes

$$\ell(\sigma_*^2, \psi_*) = -\frac{1}{2}\sum_{t=m+1}^{T}\log f_t - \frac{T-m}{2}\log\sigma_*^2 - \frac{1}{2}\sigma_*^{-2}\sum_{t=m+1}^{T}\nu_t^2/f_t \qquad (26)$$

where we have omitted the star on f_t which indicates that these MSEs are delivered by the Kalman filter with scaled hyperparameters.

For multivariate models unless the system is homogeneous, it is not possible to reparametrize in terms of an entire covariance matrix. However one element can always be made explicit in the loglikelihood function.

Clearly the MLE of (σ_*^2, ψ_*) is the point, $(\hat{\sigma}_*^2, \hat{\psi}_*)$, that maximises the loglikelihood function. Since σ_*^2 can be concentrated out the maximisation of (26) is nonlinear only with respect to ψ_*.

3. Estimation in the Frequency Domain

We shall now turn to the frequency-domain approach for estimating structural models. We introduce the spectral likelihood function and derive the asymptotic information matrix.

The Spectral Likelihood Function

Let u_t, $t=0,\pm1,\pm2,..$ be a p-variate stationary, zero mean, Gaussian process. Let $F(\lambda,\psi)$ be the spectral matrix of the process, where ψ belongs to the parameter set Θ and $\lambda\epsilon[-\pi,\pi)$. $F(\lambda,\psi)$ is defined by

$$F(\lambda,\psi) = (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \Gamma(\tau,\psi)\ e^{-i\lambda\tau} \tag{1}$$

where

$$\Gamma(\tau,\psi) = Eu_t u_{t-\tau}', \quad \tau=0,\pm1,.. \tag{2}$$

is the autocovariance matrix at lag τ, see Harvey (1989,page 428). The diagonal elements of $F(\lambda)$, where we have omitted the argument ψ and often shall do so, are the power spectra of the individual processes. The ij-th element is the cross-spectrum between the i-th and the j-th variable for $j\neq i$. The spectral matrix, see Fuller (1976), is Hermitian, that is, $F(\lambda) = F^*(\lambda)$ and positive semidefinite, that is, $w^*F(\lambda)w \geq 0$ for any complex vector such that $w^*w>0$, where here and onwards $[\cdot]^*$ denotes the complex conjugate transpose of a matrix or of a vector.

Let $F(\lambda_j)$, $j=0,...T-1$, denote the spectral matrices at frequency λ_j, where

$$\lambda_j = \frac{2\pi j}{T}, \quad j = 0,...T-1.$$

Let

$$U' = [u_1,u_2,...,u_T],$$

and let $I_{U'U}(\lambda_j)$ be the Hermitian matrix of periodograms and crossperiodograms of U', or as we shall say, the periodogram matrix, given by

$$I_{U'U}(\lambda_j) = \frac{1}{2\pi T} \sum_{t=1}^{T} u_t \, e^{i\lambda_j t} \sum_{t=1}^{T} u_t' \, e^{-i\lambda_j t} . \qquad (3)$$

We also introduce, $I_T(\lambda)=I_{U'U}(\lambda)$, which we shall need later, the periodogram matrix defined for all λ in $[-\pi,\pi)$. Of course we cannot evaluate $I_T(\lambda)$ numerically as a continuous function of λ.

As is well known, see among others Robinson (1978), the frequency-domain or spectral likelihood function for vecU' is given by

$$\ell(\psi) = -\tfrac{1}{2}\sum_{j=0}^{T-1}\log|F(\lambda_j)| - \tfrac{1}{2}\sum_{j=0}^{T-1}\mathrm{tr}\,[F^{-1}(\lambda_j)I_{U'U}(\lambda_j)]. \qquad (4)$$

If the process u_t is a non-zero mean one then $j=0$ has to be excluded from the sum to mean-correct the process. We remark that if the covariance matrix of vecU' has the form of a circulant matrix then (4) is the exact time domain loglikelihood function, otherwise (4) has to be regarded as an approximate version of the time domain loglikelihood function, see Harvey (1989,page 193). Because the periodogram matrix does not depend on the parameters, changes when a new estimate of these parameters is produced in an iterative optimisation scheme only affect the estimates of the spectral matrix. As we shall see below in structural time series models the spectrum can be easily evaluated. Hence the optimisation procedure can be carried out quite rapidly.

The structural processes introduced in the previous section are clearly nonstationary. Nevertheless for a univariate process stationarity can be achieved by differencing the process, once, if it is assumed that the process follows a random walk plus noise with $\sigma_\eta^2>0$ and twice if it is assumed to follow a local linear trend with $\sigma_\zeta^2>0$. It can be shown that the differenced processes are respectively restricted MA(1) and restricted MA(2). Moreover they are invertible and therefore the respective spectral densities are strictly positive over $[-\pi,\pi)$.

We note that in the random walk plus noise model if $\sigma_\eta^2 = 0$ the process is already stationary with mean different from zero and in the local linear trend model if $\sigma_\zeta^2=0$ we only need to difference once to obtain a stationary non zero mean process. Overdifferencing will yield a strictly non-invertible process with non-strictly positive spectrum over $[-\pi,\pi)$.

The multivariate processes that we shall consider are the SUTSE models introduced in section 2, in particular the multivariate random walk plus noise, given in (2.5) with associated covariance matrices Σ_ε and Σ_η. Therefore if Σ_η is positive definite then differencing once yields a multivariate stationary and invertible process with spectral matrix

$$F(\lambda_j) = F(\lambda_j,\psi) = (2\pi)^{-1}\,(c(\lambda_j)\Sigma_\varepsilon + \Sigma_\eta), \tag{5a}$$

where

$$c(\lambda_j) = 2(1 - \cos\lambda_j), \tag{5b}$$

and ψ is the p(p+1) vector containing the distinct elements of Σ_ε and Σ_η. We note that $F(\lambda_j)$, $j=0,\ldots T-1$ are real, positive definite, symmetric matrices, and therefore the determinant of $F(\lambda_j)$ is strictly positive.

It should also be noted that because $I_{U'U}(\lambda_j)$ is a Hermitian matrix it can be expressed as

$$I_{U'U}(\lambda_j) = \mathit{R}I_{U'U}(\lambda_j) + i\ \mathit{Im}I_{U'U}(\lambda_j) \tag{6}$$

where $\mathit{R}I_{U'U}(\lambda_j)$ is a real symmetric matrix and $\mathit{Im}I_{U'U}(\lambda_j)$ is a real skew symmetric matrix, that is, $(\mathit{Im}I_{U'U}(\lambda_j))' = -\ \mathit{Im}I_{U'U}(\lambda_j)$. Now since $F(\lambda_j)$ is symmetric we have that

$$\begin{aligned} \mathrm{tr}[F(\lambda_j)^{-1}\mathit{Im}I_{U'U}(\lambda_j)] &= -\ \mathrm{tr}[\mathit{Im}I_{U'U}(\lambda_j)F(\lambda_j)^{-1}] \\ &= -\ \mathrm{tr}[F(\lambda_j)^{-1}\ \mathit{Im}I_{U'U}(\lambda_j)] = 0. \end{aligned}$$

Hence,

$$\mathrm{tr}\ [(F(\lambda_j)^{-1}I_{U'U}(\lambda_j)] = \mathrm{tr}[F(\lambda_j)^{-1}\ \mathit{R}I_{U'U}(\lambda_j)], \tag{7}$$

and therefore the periodogram matrix in (4) is in fact only the real part of the periodogram matrix. We shall however keep the notation. It can easily verified that the real part of the periodogram matrix can be written as

$$\mathit{R}I_{U'U}(\lambda_j) = U'\Psi_j U\ , \tag{8}$$

where the TxT matrix Ψ_j, $j=1,\ldots T-1$, is real, symmetric, with (ℓ,k) entry being

$$\Psi_j(\ell,k) = \cos(\lambda_j(\ell-k)),\ \ell,k=1,\ldots T. \tag{9}$$

Using (9), the spectral likelihood takes the form

$$\ell(\psi) = -\tfrac{1}{2}\sum_{j=0}^{T-1}\log|F(\lambda_j)| - \tfrac{1}{2}\sum_{j=0}^{T-1}\mathrm{tr}\,[F^{-1}(\lambda_j)U'\Psi_j U] \,. \tag{10}$$

Asymptotic Information Matrix

It is well known that the asymptotic information matrix is determined by

$$IA(\psi_o) = -\lim_{T\to\infty} T^{-1}\, E\left[\frac{\partial^2\ell(\psi)}{\partial\psi\partial\psi'}\right]_{\psi_o} ,$$

where ψ_o is the true parameter vector and $\ell(\psi)$ is the loglikelihood function of the T observations. We shall only consider the case where the observations are generated by a multivariate random walk plus noise process. Hence ψ is given in (2.7). Now the first and second derivatives of (10) with respect to ψ are given in Harvey (1989) or in Fernandez(1986) and are

$$\frac{\partial}{\partial\psi}\,\ell = -\frac{1}{2}\sum_{j=0}^{T-1} S_j'\mathrm{vec}\left[\, F_j^{-1} - F_j^{-1}U'\Psi_j UF_j^{-1} \,\right] \tag{11}$$

and

$$\frac{\partial^2\ell}{\partial\psi\partial\psi'} = \sum_{j=0}^{T-1} S_j' \left[\tfrac{1}{2}(F_j^{-1}\otimes F_j^{-1}) - (F_j^{-1}\otimes F_j^{-1}U'\Psi_j UF_j^{-1})\right] S_j \tag{12}$$

where

$$S_j = \frac{\partial \mathrm{vec}F_j}{\partial\psi'} ,$$

and

$$F_j = F(\lambda_j,\psi) = (2\pi)^{-1}\,[\Sigma_\varepsilon:\Sigma_\eta]\begin{bmatrix} c_j I_p \\ I_p \end{bmatrix} ,$$

$$c_j = c(\lambda_j) = 2(1-\cos\lambda_j).$$

Vectoring F_j and using (2.6) we have

$$\mathrm{vec}F_j = (2\pi)^{-1}[(c_jI_p:I_p)\otimes I_p] \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix} \psi ,$$

hence S_j becomes

$$S_j = (2\pi)^{-1} [c_jD : D] . \tag{13}$$

Using the fact, see Priestley (1981,page 418), that

$$EI_T(\lambda)_{lk} = F(\lambda,\psi_o)_{lk} + O(T^{-1}\log T), \quad l,k=1,\ldots,p \tag{14}$$

we have

$$IA(\psi_o) = \frac{1}{4\pi} \int_{-\pi}^{\pi} S(\lambda)' \, [F^{-1}(\lambda)\otimes F^{-1}(\lambda)] \, S(\lambda)d\lambda \tag{15}$$

where $F(\lambda) = F(\lambda,\psi_o)$, $S(\lambda) = [c(\lambda)D : D]$, $c(\lambda) = 2(1-\cos\lambda)$ and D is the duplication matrix.

Maximum Likelihood Estimation

In order to find the point $\hat{\psi}$ that maximises $\ell(\psi)$ given in (4) we need a numerical method. Now from appendix 2.1, where we briefly discuss optimisation procedures, we have the iterative scheme defined by

$$\hat{\psi}_{k+1} = \hat{\psi}_k - H^{-1}(\hat{\psi}_k)\, g(\hat{\psi}_k), \tag{16}$$

where $g(\hat{\psi}_k)$ is the first derivative of ℓ with respect to ψ evaluated at $\hat{\psi}_k$ and $H(\hat{\psi}_k)$ is the Hessian matrix evaluated at $\hat{\psi}_k$.

Alternatively, replacing $H(\hat{\psi}_k)$ by minus $\overline{H}(\hat{\psi}_k)$ yields the scoring algorithm, where for the multivariate random walk plus noise, T times $\overline{H}(\psi_o)$ is a finite approximation of the integral given in (15). Hence $\overline{H}(\hat{\psi}_k)$ is given by

$$\overline{H}(\hat{\psi}_k) = \tfrac{1}{2}\sum_{j=0}^{T-1} S_j'(\hat{F}_j^{-1} \otimes \hat{F}_j^{-1})S_j \ , \qquad (17)$$

where $\hat{F}_j = F_j(\hat{\psi}_k)$. As shown in Robinson (1978), under certain regularity conditions $-T^{-1}\overline{H}(\psi)$ and $T^{-1}H(\psi)$ are asymptotically equivalent.

4. Asymptotic Theory

The spectral likelihood function given in (3.4) has to be regarded as an approximation to the time-domain likelihood function. However the spectral ML estimators have the same asymptotic properties of the time-domain ML estimators. Therefore we are only going to consider the asymptotic properties of spectral estimators. This issue was considered in Fernandez (1986). Since, after differencing, the structural models considered here are vector MA's, the underlying asymptotic theory in Fernandez (1986) is regarded as a specialization of the asymptotic theory for stationary vector time series given in Deistler *et al* (1978), Dunsmuir and Hannan(1976) and Dunsmuir(1979). However, since we are primarily interested in the asymptotic properties of estimators obtained by the iterative scheme given in (3.16), for example the two-step estimator obtained as $\hat{\psi}_2$ when $\hat{\psi}_1$ is consistent, the asymptotic theory given in Robinson (1978) seems to be more appealing.

In what follows we state without proof asymptotic results for estimators of the parameters involved in a scalar structural process. These results are a specialization of the ones given in Robinson (1978) for more general stationary processes.

Let

$$c(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} u_t u_{t+\tau} \quad ; \tag{1}$$

$$\tilde{\psi} = G(c(o), c(1), \ldots, c(\tau)), \tag{2}$$

where G is continuously differentiable in a neighbourhood of $(\gamma(o), \gamma(1), \ldots, \gamma(\tau))$;

and $\hat{\psi}$ a single iteration of the scheme given in (3.16) with initial value $\tilde{\psi}$.

Under suitable conditions, we have

$$\lim_{T\to\infty} (c(\tau) - \gamma(\tau)) = 0 \text{ a.s. for all fixed } \tau \ ; \tag{3}$$

$$T^{\frac{1}{2}}(c(0)-\gamma(0), c(1)-\gamma(1), \ldots, c(\tau)-\gamma(\tau)) \xrightarrow{d} N(0,\cdot), \text{ for all } \tau \geq 1; \tag{4}$$

$$\lim_{T\to\infty} (\tilde{\psi} - \psi_o) = 0 \quad \text{a.s.} \ ; \tag{5}$$

$$T^{\frac{1}{2}}(\tilde{\psi} - \psi_o) \xrightarrow{d} N(0, \cdot) \ ; \tag{6}$$

$$\lim_{T\to\infty} (\hat{\psi} - \psi_o) = 0 \quad \text{a.s.} \ ; \tag{7}$$

$$T^{\frac{1}{2}}(\hat{\psi} - \psi_o) \xrightarrow{d} N(0, IA^{-1}(\psi_o)). \tag{8}$$

The condition under which (3) holds is

(A1) u_t is zero mean, Gaussian, second order stationary and $\sum\gamma(\tau)^2 < \infty$.

We remark that Gaussianity can be replaced by conditions on the fourth cumulant. If the mean of the process is unknown then $c(\tau)$ in (1) must be mean-corrected.

Result (4) can be shown for Gaussian processes, see Priestley (1981,page 339). In Robinson (1978) result (4) is imposed plus additional conditions on u_t.

Results (5) and (6) follow from a straighforward application of the mean value theorem.

Results (7) and (8) hold under (A1) and the following additional conditions.

(A2) u_t has spectrum $f(\lambda,\psi)$ belonging to Lip ζ, $\zeta>\frac{1}{2}$, the Lipschitz class of degree ζ.

(A3) ψ_o is an interior point of Θ, which is the compact closure of an open submanifold in a twice-differentiable p-dimensional manifold.

(A4) $f(\lambda,\psi)$ is continuous in $\lambda\epsilon[-\pi,\pi]$, $\psi\epsilon\Theta$.

(A5) $f(\lambda,\psi) \neq f(\lambda,\psi_o)$ for all $\psi \neq \psi_o$, $\psi\epsilon\Theta$.

(A6) Within a neighborhood of ψ_o, $f(\lambda,\psi)$ has first and second derivatives with respect to ψ, these being continuous in λ and ψ.

(A7) $f(\lambda,\psi_o) > 0$; $f'(\lambda,\psi_o) \in \mathrm{Lip}\zeta$, $\zeta>\frac{1}{2}$.

(A8) $IA(\psi_o)$ is positive definite.

It can be easily verified that these conditions will be satisfied for the models considered here. For the random walk plus noise model, since

$$\psi'=[\sigma_\varepsilon^2 : \sigma_\eta^2] = [-\gamma(1) : \gamma(0)+2\gamma(1)],$$

the estimator suggested in (2) becomes

$$\tilde{\psi}' = [-c(1) : c(0)+2c(1)]. \tag{9}$$

We also note that result (8), i.e., an asymptotically efficient estimator for ψ_o is obtained by a single use of the iterative scheme provided that the current estimate is consistent, is particulary useful when we consider time series regression models.

Results (3-8) can be generalized to vector processes. The conditions are basically the same but are on the elements of the spectrum matrix $F(\lambda,\psi)$. Also in (A7) $f(\lambda,\psi_o)>0$ is replaced by $F(\lambda,\psi_o)$ positive definite.

Appendix 2.1

Optimisation Procedures

The maximum likelihood estimates emerge as a solution of the likelihood equations

$$\frac{\partial}{\partial \theta} L(\theta | x) = 0 \tag{1}$$

These equations are often nonlinear in θ, hence they must be solved numerically. The basic procedure to solve a nonlinear set of equations is the well known Newton-Raphson method

$$\theta_{k+1} = \theta_k - G_k^{-1} g_k \tag{2}$$

where θ_1 is a vector of initial values, θ_k, k=1,... is the current estimate of θ , G_k and g_k are respectively the Hessian matrix and the gradient vector evaluated at the current estimate.

Different subroutines have been written for the implementation of the Newton's method or variations of it. Hence from the computational point of view we do not need to worry about the actual calculation of the maximum likelihood estimates. We do not even need to derive the analytic form of the first and second derivatives since they can also be computed numerically. However for a large number of parameters, as is the case in simultaneous equation systems, the computational time burden is formidable and it might be worth to examining different iterative procedures by exploiting the structure of the model.

Variations of Newton-Raphson method arise by replacing G by a positive definite matrix H close to G. In the particular case where G is replaced by the information matrix the procedure is known as the scoring algorithm.

Sometimes the parameter set can be partitioned as $\theta = (\delta, \psi)$ such that the likelihood equations are linear in δ given ψ and vice versa. To exploit this property Sargan (1964) introduced the stepwise optimisation procedure. From the theoretical point of view, however, the procedure is valid for any partition of θ. Such a procedure which induces separate optimisation of the parameters in θ can be regarded as (2) with G^* replacing G where $G^* = \text{diag}\{ H_{11}, H_{22} \}$ and where H_{11} and H_{22} are the submatrices of the Hessian matrix corresponding to the second derivatives with respect to δ and ψ respectively. Thus the iterative scheme is for j=2,..

$$\delta_{j(k+1)} = \delta_{jk} + H_{11}(\delta_{jk}, \psi_j)^{-1} g_\delta(\delta_{jk}, \psi_j) \qquad k=0,\ldots \qquad (3a)$$

$$\psi_{j(k+1)} = \psi_{jk} + H_{22}(\delta_j, \psi_{jk})^{-1} g_\psi(\delta_j, \psi_{jk}) \qquad k=0,\ldots \qquad (3b)$$

where δ_1 and ψ_1 are initial values, $\delta_{j0} = \delta_j$, $\psi_{j0} = \psi_j$ are the final estimates computed at step j-1. g_δ, the first derivatives with respect to δ, and H_{11} are evaluated at the current estimate of δ and at the final estimate of ψ at step j-1; similary g_ψ and H_{22}. The procedure is bound to converge, see Oberhofer and Kmenta (1974). Clearly if the likelihood equations are linear in δ given ψ then δ_j

can be obtained directly, and vice-versa.

On the other hand linearity in a subset of parameters given the second set, say, in δ given ψ can be explored by concentrating δ out of the likelihood function. In other words, δ is replaced in the likelihood function by the solution of the equations for δ, say, $\hat{\delta}=\hat{\delta}(\psi)$ yielding the concentrated likelihood function which has to be maximised nonlinearly with respect to ψ. Once we have obtained $\hat{\psi}$ the maximum likelihood estimator of δ is given by $\hat{\delta}=\hat{\delta}(\hat{\psi})$.

CHAPTER 3

TIME SERIES REGRESSION MODELS

1. Introduction

This chapter is a continuation of chapter 2 in the sense that more known material is presented. It provides a basis for our discussion of a single equation from a simultaneous equation system, as well as for the whole system, to be given in the next chapters. Also its notational content will be relevant for our purposes.

Although this material could have been presented in the remaining chapters together with new material, we have chosen to bring it together in this chapter, so that it might be skipped by the well informed reader, and also to avoid burdening the reading of the related chapters.

Our main purpose is the discussion of time series regression models, more specifically, the inclusion of regressors in the models formerly handled. Since the generalization from scalar to vector processes is straighforward we shall consider in some detail only scalar models. Such models can be formulated in state space form. Two possibilities are open. Firstly, augmenting the state to include the regression coefficients, and secondly via the Kohn and Ansley (1985) approach. Our emphasis relies on the latter. In either case, the Kalman filter plays an important role.

2. Regression Models with Stochastic Components

We shall now consider the inclusion of explanatory variables in structural time series models. Since the generalization to more complex models is straightforward, for simplicity, only the local level model will be considered. Thus, let the model be

$$y_t = z_t'\delta + w_t \tag{1a}$$

$$w_t = \mu_t + \varepsilon_t \tag{1b}$$

$$\mu_t = \mu_{t-1} + \eta_t \tag{1c}$$

where z_t' is a 1xK row vector of explanatory variables which we initially assume to be non-stochastic. The assumptions on w_t are as in chapter 2, that is, ε_t and η_t independent and normally distributed. No information about the initial state is available. We also assume that rank(Z)=K, where $Z'=(z_1,\ldots,z_T)$. Let $\theta = (\delta',\psi')'$, $\psi = (\sigma_\varepsilon^2,\sigma_\eta^2)$ denote the parameters of interest.

Now conditionally on w_1 the loglikelihood function for $w_2,\ldots,w_t$ is given in (2.2.24). Hence, since z_t is non-stochastic, it follows immediately that the loglikelihood function for $y_2,\ldots,y_T$ conditionally on y_1 is given by

$$\ell(\theta) = -\tfrac{1}{2}\sum_{t=2}^{T}\log f_t - \tfrac{1}{2}[\bar{L}[y-\mathit{1}y_1-(Z-\mathit{1}z_1)\delta]]'F^{-1}\bar{L}[y-\mathit{1}y_1-(Z-\mathit{1}z_1)\delta] \qquad (2)$$

where $y'=(y_2,\ldots,y_T)$, $Z'=(z_2,\ldots,z_T)$ and $\mathit{1}$ is a T-1 vector of ones. As discussed above (2.2.25), $\ell(\theta)$ is evaluated applying the Kalman filter to $y_t-z_t'\delta$ with starting $y_1-x_1'\delta$ and $P_1=\sigma_\varepsilon^2$. However, since

$$\bar{L}[y-\mathit{1}y_1-(Z-\mathit{1}z_1)\delta] = \bar{L}[y-\mathit{1}y_1]-\bar{L}[(Z-\mathit{1}z_1)\delta],$$

applying the Kalman filter to $y_t-z_t\delta$, $t=2,\ldots,T$ is equivalent to applying the Kalman filter separately to y_t and to each column of z_t' with respective starting values y_1 and $[z_{11},\ldots,z_{K1}]$. Hence, if the hyperparameters are scaled as in (2.2.26) the resulting loglikelihood function becomes

$$\ell(\delta,\sigma_*^2,\psi_*) = -\tfrac{1}{2}\sum_{t=2}^{T}\log f_t - \tfrac{1}{2}(T-1)\log\sigma_*^2 - \tfrac{1}{2}\,\sigma_*^{-2}\sum_{t=2}^{T}\nu_t^2 f_t^{-1} \qquad (3a)$$

where

$$\sum_{t=2}^{T}\nu_t^2 f_t^{-1} = \sum_{t=2}^{T}(\tilde{y}_t-\tilde{z}_t\delta)^2/f_t = (\tilde{Y}-\tilde{Z}\delta)'F^{1}\ (\tilde{Y}-\tilde{Z}\delta) \qquad (3b)$$

and $\tilde{y}_t$, $\tilde{z}_t'$ are the 'innovations' delivered by the Kalman filter with scaled hyperparameters.

We now consider the case where the exogenous variable are stochastic. Clearly, if this is the case, in principle, the whole distribution of y and Z, which depends on the full set of parameters, say, λ must be specified. However if z_t is weakly exogenous for θ then (3) is a valid basis for inferences purposes since in this case z_t could be regarded as being fixed. We shall confine ourselves to the concept of weak exogeneity given in Engle, Hendry and Richard (1983). In the normal framework, λ consists of the mean vector and the covariance matrix. Usually the elements of θ, the parameters of interest, do not coincide with those in λ. Thus, let $\theta^{\dagger}$ be a reparameterization of λ such that $\theta^{\dagger}=(\theta,\theta_z)$, where θ and θ_z are variation free, i.e. θ and θ_z are not subject to cross restrictions so that for any admissible value of θ_z, θ can take any value in its parameter space and vice versa. Then z_t is weakly exogenous for θ if the joint distribution of $[Y_T,Z_T]$ can be factorized as

$$f(Y,Z;\theta^{\dagger})=\prod_{t=2}^{T} f(y_t|z_t,Y_{t-1},Z_{t-1};\theta)f(y_1|z_1;\theta)\prod_{t=2}^{T} f(z_t|Y_{t-1},Z_{t-1};\theta_z)f(z_1|\theta_z) \tag{4}$$

Hence all sample information concerning θ can be obtained from the first term in the RHS of (4). For prediction purposes we have to assume that z_t is strongly exogenous. We note that z_t is strongly exogenous for θ if it is weakly exogenous and in addition past values of y_t does not Granger-cause z_t, that is, conditionally on $z_{t-1},z_{t-2},\ldots$, z_t is independent of past values of y_t.

Multivariate Time Series Regression Models

Multivariate models can be handled in the same way. Thus, let the model be

$$y_t = B'z_t + w_t \tag{5}$$

where y_t is a pxl vector, B' is a pxK matrix, z_t is a Kxl vector and $\text{plim}T^{-1}Z'Z$ is positive definite, where $Z'=(z_1,\ldots,z_T)$. w_t follows a multivariate random walk plus noise model. Let $\theta=(\beta,\psi)$ the parameters of interest, where $\beta=\text{vec}B$ and ψ is the p(p+1)xl vector containing the unrestricted elements of Σ_ϵ and Σ_η. Using rules on Kronecker products, see Magnus and Neudecker (1988,page 47), (5) can be rewritten as

$$y_t = (I_p \otimes z_t')\beta + w_t. \tag{6}$$

We note that if some of the elements of β are constrained to be zero then (6) can be written as

$$y_t = \begin{bmatrix} Z_{1t}' & 0 & 0 & 0 \\ 0 & Z_{2t}' & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & Z_{pt}' \end{bmatrix} \beta^* \tag{7}$$

where Z_{it}' are the explanatory variables in the i-th equation and β^* contains the unrestricted elements of $\beta=\text{vec}B$.

Now conditionally on the first observations the loglikelihood function for $w=\text{vec}(w_2,\ldots,w_t)$ is given in (2.2.25). Hence arguing as in the univariate case it follows that the conditional loglikelihood function for $y=\text{vec}(y_2,\ldots,y_T)$ takes the form

$$\ell(\theta) = -\tfrac{1}{2}\sum_{t=2}^{T}\log|F_t| - \tfrac{1}{2}\sum_{t=2}^{T}\nu_t'F_t^{-1}\nu_t\ , \tag{8a}$$

where

$$\nu_t = \tilde{y}_t - \tilde{Z}_t'\beta\ , \tag{8b}$$

and $\tilde{y}_t$ is obtained by applying the multivariate Kalman filter to y_t, $t=2,\ldots T$, with starting value y_1 (px1) and $\tilde{Z}_t'$ is obtained by applying the multivariate Kalman filter separately to each column of $(I_p \otimes z_t')$, $t=2,\ldots T$, with starting values being the respective column of $(I_p \otimes z_1')$.

Maximum Likelihood Estimation

In the next section we are going to derive the asymptotic information matrix in the frequency domain. We are going to show that the asymptotic information matrix is block diagonal with respect to regression coefficients and hyperparameters. As pointed out in chapter 2 the asymptotic properties in the frequency domain are the same as in the time domain, hence the time domain asymptotic information matrix must be block diagonal. Therefore the stepwise algorithm seems to be a natural way to obtain the ML estimates. Alternatively the ML estimates can be obtained by concentrating the vector containing the regression coefficients out of the loglikelihood function.

From the discussion of the optimisation procedures given in appendix 2.1, the optimisation of (3) by means of the stepwise algorithm consist of two parts, one for (σ_*^2, ψ_*) and one for δ. The part for (σ_*^2, ψ_*) consists of finding the point that maximises (3) with respect to (σ_*^2, ψ_*) conditionally on a given δ. We shall denote this point by $(\hat{\sigma}_*^2, \hat{\psi}_*)$, $\hat{\sigma}_*^2 = \hat{\sigma}_*^2(\delta)$, $\hat{\psi}_* = \hat{\psi}_*(\delta)$. Clearly, $(\hat{\sigma}_*^2, \hat{\psi}_*)$ is obtained as it would be for a model without explanatory variables. The part for δ consists in finding the point that maximises (3b) with respect to δ, conditionally on a given ψ_*. This is simply the GLS (Generalized Least Squares) estimator, that is,

$$\hat{\delta}(\psi_*) = \Big[\sum_{t=2}^{T} f_t^{-1} \tilde{z}_t \tilde{z}_t' \Big]^{-1} \Big[\sum_{t=2}^{T} [f_t^{-1} \tilde{z}_t \tilde{y}_t] \Big] . \tag{9}$$

Hence, given an initial value, say $\tilde{\delta}$, $\hat{\sigma}_*^2 = \hat{\sigma}_*^2(\tilde{\delta})$, $\hat{\psi}_* = \hat{\psi}_*(\tilde{\delta})$ and $\hat{\delta} = \hat{\delta}(\hat{\psi}_*)$ are evaluated as described above. Then, making use of $\hat{\delta}$, $\hat{\sigma}_*^2$, $\hat{\psi}_*$ and $\hat{\delta}$ are updated. The procedure is repeated until convergence is attained. We mention that the OLS (Ordinary Least Squares) estimator of δ of the differenced model may be used as a starting value for δ.

On the other hand replacing (9) in (3b) yields $\ell_c(\sigma_*^2, \psi_*)$, the concentrated loglikelihood function with respect to (σ_*^2, ψ_*). Once the point that optimises $\ell_c(\sigma_*^2, \psi_*)$ is found, say $(\hat{\sigma}_*^2, \hat{\psi}_*)$, $\hat{\delta}$ is evaluated by means of (9) with $\hat{\psi}_*$ replacing ψ.

Finally we mention that for multivariate models, with loglikelihood function given in (8), $\hat{\beta}(\psi)$ becomes

$$\hat{\beta}(\psi) = \Big[\sum_{t=2}^{T} \tilde{Z}_t F_t^{-1} \tilde{Z}_t' \Big]^{-1} \sum_{t=2}^{T} \tilde{Z}_t F_t^{-1} \tilde{y}_t , \tag{10}$$

and the optimisation procedures described for the univariate case can be generalized straightforwardly.

From the computational point of view, for a large number of regressors concentrating δ out of the loglikelihood function might be more time consuming. This will be so, because the iterative scheme activated to optimise $\ell_c(\sigma_*^2, \psi_*)$ requires the computation of $\hat{\delta}$ at each iteration. For disturbances following a univariate random walk plus noise model, $\ell_c(\sigma_*^2, \psi_*)$ is optimised nonlinearly with respect to one parameter only. In this case the NAG subroutine E04JBF calls $\ell_c(\sigma_*^2, \psi_*)$ approximately 50 times. So 50 times a large matrix must be inverted. In the stepwise procedure $\hat{\delta}$ is computed only at each step. Our experience shows that only a few steps are required for convergence to be attained.

3. Frequency Domain Estimation

In the previous section we have considered the model

$$y_t = B'z_t + w_t \ ,$$

where the vectors y_t (px1) and z_t (kx1) are the observable variables and the vector w_t (px1) is the non observable process following a multivariate random walk plus noise. We shall now discuss the frequency-domain approach to handling such model. The first step is to transform to a regression model with stationary disturbances. Because w_t follows a multivariate random walk plus noise differencing once y_t and z_t yields

$$y_t = B'z_t + u_t \ , \tag{1}$$

where we deliberately omitted the differencing operator Δ in front of y_t and z_t and will do so in the rest of this chapter, to avoid overelaborate notation. Thus whenever we refer to y_t or z_t we are in fact referring to Δy_t or Δz_t. The spectrum matrix of u_t was given in (2.3.5). The exogenous variables z_t are assumed totally independent of the process u_t and it is assumed that the following limit exists,

$$\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} z_t \, z_{t+\tau}' = \Gamma_z(\tau) \quad \text{a.s.}, \quad \tau=0,\pm 1,\pm 2,\ldots \tag{2}$$

with $\Gamma_z(0)$ nonsingular. The existence of the limit implies that there exists a spectrum matrix $F_z(\lambda)$ such that

$$\Gamma_z(\tau) = \int_{-\pi}^{\pi} e^{i\tau\lambda} \, F_z(d\lambda) \ ,$$

see Hannan (1970, ch 2) for details concerning this assertion. We also assume that $\lim T^{-1}\sum z_t$ exists.

In matrix notation (1) can be written as

$$Y' = B'Z' + U' \tag{3}$$

where

$Y' = [y_1 \ldots y_T]$ where y_t is pxl

$Z' = [z_1 \ldots z_T]$ where z_t is Kxl

$U' = [u_1 \ldots u_T]$ where u_t is pxl

Now the spectral likelihood for vecU' is given in (2.3.4) and since Z and U are totally independent the spectral likelihood function for vec(Y') becomes

$$\ell(\theta) = -\frac{1}{2}\sum_{j=0}^{T-1}\log|F_j| - \frac{1}{2}\sum_{j=0}^{T-1}\mathrm{tr}[F_j^{-1}U'\Psi_j U] \;, \tag{4}$$

where $\theta=(\beta,\psi)$, β = vecB and ψ is as in (2.2.7), that is, the p(p+1)x1 vector obtained from $\mathrm{vec}(\Sigma_\varepsilon:\Sigma_\eta)$ by eliminating the supradiagonal elements of Σ_ε and Σ_η. $U'\Psi_j U = I_{U'U}(\lambda_j)$ is the real part of the periodogram matrix of U where U' is expressed in terms of the observations through (3). Thus,

$$I_{U'U}(\lambda_j) = I_{Y'Y}(\lambda_j) - I_{Y'Z}(\lambda_j)B - B'I_{Z'Y}(\lambda_j) + B'I_{Z'Z}(\lambda_j)B \tag{5}$$

where

$$\begin{bmatrix} I_{Y'Y}(\lambda_j) & I_{Z'Y}(\lambda_j) \\ I_{Y'Z}(\lambda_j) & I_{Z'Z}(\lambda_j) \end{bmatrix} = \begin{bmatrix} Y'\Psi_j Y & Z'\Psi_j Y \\ Y'\Psi_j Z & Z'\Psi_j Z \end{bmatrix} \tag{6}$$

is the real part of the periodogram matrix of the augmented process $[y_t'z_t']'$, Ψ_j as given in (2.3.9).

The spectral ML estimates of $\theta=(\beta',\psi')'$ are those which maximise (4). Before discussing the maximisation of (4) we shall derive the asymptotic information matrix.

Asymptotic Information Matrix

The first and second derivatives of (4) can be obtained as a specialization of the ones given in chapter 6 when we shall handle simultaneous equation systems. They are also given in Fernandez (1986) although with respect to vecB'. The reason why we prefer to define β=vecB rather than β=vecB' will become clear in the simultaneous equation system context.

Thus

$$\frac{\partial}{\partial\beta}\,\ell = \mathrm{vec}\Big[\sum_{j=0}^{T-1} Z'\Psi_j U\, F_j^{-1}\Big] \tag{7a}$$

$$\frac{\partial}{\partial\psi}\,\ell = -\frac{1}{2}\sum_{j=0}^{T-1} S_j'\,\mathrm{vec}\Big[(F_j^{-1} - F_j^{-1}\,U'\Psi_j U\,F_j^{-1})\Big] \tag{7b}$$

and

$$\frac{\partial^2}{\partial\beta\partial\beta'}\,\ell = -\sum_{j=0}^{T-1}\Big[(F_j^{-1}\otimes Z'\Psi_j Z)\Big] \tag{8a}$$

$$\frac{\partial^2}{\partial\psi\partial\beta'}\,\ell = -\sum_{j=0}^{T-1} S_j'\,\Big[(F_j^{-1}U'\Psi_j Z\otimes F_j^{-1})K_{kp}\Big] \tag{8b}$$

$$\frac{\partial^2}{\partial\psi\partial\psi'}\,\ell = \sum_{j=0}^{T-1} S_j'\,\Big[\tfrac{1}{2}(F_j^{-1}\otimes F_j^{-1}) - (F_j^{-1}\otimes F_j^{-1}U'\Psi_j U F_j^{-1})\Big]S_j \tag{8c}$$

where K_{Kp} is a KpxKp commutation matriz and S_j is given in (2.3.13), that is,

$$S_j = (2\pi)^{-1}\,[c_j D : D]\,.$$

Since u_t and z_t are totally independent and u_t has zero mean we have $EI_{Z'U}(\lambda_j)=0$ and because of (2) and (2.3.14) we have the asymptotic information matrix,

$$IA(\theta_o) = \lim_{T\to\infty} T^{-1}\Big[-E\,\frac{\partial^2\ell}{\partial\theta\partial\theta'}\Big]_{\theta_o}$$

$$IA(\theta_o) = \begin{bmatrix} \frac{1}{2\pi}\int_{-\pi}^{\pi}(F(\lambda)^{-1}\otimes F_Z(d\lambda)) & 0 \\ 0 & \frac{1}{4\pi}\int_{-\pi}^{\pi} S(\lambda)'\,[F(\lambda)^{-1}\otimes F(\lambda)^{-1})]\,S(\lambda)d\lambda \end{bmatrix} \quad (9)$$

where θ_o is the true value, $F(\lambda) = F_u(\lambda,\theta_o)$, $S(\lambda) = (2\pi)^{-1}[c(\lambda)D:D]$, $c(\lambda) = 2(1-\cos\lambda)$ and D is the $p^2 x \frac{1}{2}p(p+1)$ duplication matrix.

Spectral Maximum Likelihood Estimates

As in the time domain, the spectral ML estimates of (β,ψ) can be obtained by means of the stepwise algorithm. From (7a), we have that conditionally on ψ, the spectral ML estimator of β, is given by

$$\hat{\beta}(\psi) = \left[\sum_{j=0}^{T-1}[F_j^{-1}\otimes I_{Z'Z}(j)]\right]^{-1}\left[\sum_{j=0}^{T-1}[F_j^{-1}\otimes I_k]\ \mathrm{vec}I_{Z'Y}(j)\right] \quad (10)$$

where $I_{Z'Z}(j) = Z'\psi_j Z$ and $I_{Z'Y}(j) = Z'\psi_j Y$ are the real part of the respective periodogram matrices. Hence the point that optimises (4) with respect to (β,ψ) can be found as follows.

Step 1- An initial value for β is evaluated, say $\tilde{\beta}$=vec $\tilde{B}_{LS}$, where $\tilde{B}_{LS}$ is the least squares estimator, that is,

$$\tilde{B}_{LS} = [Z'Z]^{-1}Z'Y \; ; \quad (11)$$

Step 2- Conditionally on $\tilde{\beta}$, (4) is maximised with respect to ψ;

Step 3- Making use of (10), $\hat{\beta}=\hat{\beta}(\hat{\psi})$ is evaluated, where $\hat{\psi}$ was obtained in step 2.

Steps 2 and 3 are repeated until convergence is attained. However as we shall see in the next section, provided that we start with a consistent estimator of β no gain in efficiency will be achieved if the procedure is repeated. Also, in obtaining $\hat{\psi}=\hat{\psi}(\tilde{\beta})$, we do not need to iterate until convergence, if we start with a consistent estimator of ψ. This efficiency, however is asymptotic and in practice it will be best if we iterate until convergence is attained. Therefore any starting value might suffice.

Alternatively, the ML estimates of (β,ψ) could be obtained by concentrating β out of the spectral likelihood function. The concentrated likelihood function becomes

$$\ell_c(\psi) = -\frac{1}{2}\sum_{j=0}^{T-1}\log|F_j| - \frac{1}{2}\sum_{j=0}^{T-1}\mathrm{tr}[\ F_j^{-1}\hat{U}'\Psi_j\hat{U}\ ]\ .$$

where $\hat{U}=Y-Z\hat{B}$, and $\hat{B}$ is such that $\mathrm{vec}\hat{B}=\hat{\beta}$, and $\hat{\beta}$ is given in (10).

4. Asymptotic Theory of Time Series Regression Models

The asymptotic theory of regression time series models is discussed in Fernandez (1986). However as we find that some of the results may not be as straighforward as they appear and because we will need these results in the next chapter when we shall handle Instrumental Variable estimation a brief discussion is in order. We shall restrict ourselves to the asymptotic properties of the stepwise estimator of θ, $\theta=(\beta,\psi)$ outlined in the previous section. Details will be omitted since our results follow directly from the results given in Robinson (1978), Hannan (1973), Hannan (1971). Moreover as the vector case is essentially the same as the scalar case, for simplicity of presentation we shall handle only the scalar case. Thus, let the model be

$$y_t = z_t'\delta + u_t$$

where $u_t = \eta_t + \Delta\epsilon_t$ and $f_u(\lambda) = \sigma_\epsilon^2 2(1-\cos\lambda) + \sigma^2_\eta$. Under the condition on z_t given in (3.2), we have from Hannan (1971) that

$$\operatorname{plim} \ T^{-1}\sum_{j=0}^{T-1}[\Phi_j \ I_{Z'Z}(j)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(\lambda)F_Z(d\lambda) \tag{1}$$

and

$$\operatorname{plim} \ T^{-1}\sum_{j=0}^{T-1}[\Phi_j \ I_{Z'U}(j)] = 0 \tag{2}$$

where $\Phi(\lambda)$ is a continuous, even, function of λ, satisfying $\Phi(\lambda)\geq 0$, $\lambda\epsilon[0,\pi]$. Clearly these results hold if $\Phi_j=f_j$, however as in Hannan (1971) we need these results to remain valid if Φ_j is replaced by $\tilde{f}_j^{-1}$, $\tilde{f}_j=f_j(\tilde{\psi})$, where $\tilde{\psi}$ is a consistent estimator of ψ.

From Hannan (1973), for more general processes generating the stationary disturbances than the ones considered here, we have the following central limit theorem

$$T^{-\frac{1}{2}}\sum_{j=0}^{T-1}[f_j^{-1} \ I_{Z'U}(j)] \xrightarrow{d} N\left[0 \ , \ \frac{1}{2\pi}\int_{-\pi}^{\pi} f_u^{-1}(\lambda)F_Z(d\lambda)\ \right] \tag{3}$$

We note that, if $\tilde{\psi}$ is $\sqrt{T}$-consistent, that is, $T^{\frac{1}{2}}(\tilde{\psi}-\psi_o)=O_p(1)$ then, because $f_u(\lambda)$ satisfies the regularity conditions given in section 2.4, we can expand $\tilde{f}_j^{-1}$ in a Taylor's series for random functions, see Fuller(1976,page 191), and write

$$T^{-\frac{1}{2}}\Big[\sum_{j=0}^{T-1}\tilde{f}_j^{-1}I_{Z'U}(j) - \sum_{j=0}^{T-1}f_j^{-1}I_{Z'U}(j)\Big]$$

$$= T^{-1}\sum_{j=0}^{T-1}\Big[\frac{\partial}{\partial\psi'}f_j^{-1}\Big]I_{Z'U}(j)\ T^{\frac{1}{2}}(\tilde{\psi}-\psi_o)+O_p(T^{-1})$$

where the derivative is evaluated at the true parameter vector ψ_o, $f_j=f_j(\psi_o)$ and $\tilde{f}_j=f_j(\tilde{\psi})$. Now from (2) it follows that the RHS of the above expression converges in probability to zero. Thus (3) holds if f_j is replaced by $\tilde{f}_j$.

In appendix 3.1 we show that

$$T^{\frac{1}{2}}(\hat{\psi}-\psi_o) \xrightarrow{d} N\,[0,\ IA^{-1}(\psi_o)] \tag{4}$$

and

$$T^{\frac{1}{2}}(\hat{\delta}-\delta_o) \xrightarrow{d} N\,[0,\ IA^{-1}(\delta_o)]\ , \tag{5}$$

where $IA(\delta_o)$ and $IA(\psi_o)$ are respectively the top left block and the bottom right block of the asymptotic information matrix given in (3.9), specialized for the univariate model;

$$\hat{\psi} = \tilde{\psi} + \overline{H}^{-1}(\tilde{\psi})\ g_{\psi}(\tilde{\delta},\tilde{\psi}), \tag{6}$$

$$\overline{H}(\tilde{\psi}) = \frac{1}{2}\ \sum_{j=0}^{T} S_j'\tilde{f}_j^{-2}\ S_j\ ,$$

where $\tilde{f}_j = f_j(\tilde{\psi})$ and $g_\psi(\tilde{\delta},\tilde{\psi})$ is a specialization of (3.7b) for the univariate model evaluated at $(\tilde{\delta},\tilde{\psi})$;

$$\hat{\delta}(\hat{\psi}) = \left[\sum_{j=0}^{T-1}[\hat{f}_j^{-1} I_{Z'Z}(j)]\right]^{-1}\left[\sum_{j=0}^{T-1}[\hat{f}_j^{-1} I_{Z'Y}(j)\right]; \tag{7}$$

$$T^{\frac{1}{2}}(\tilde{\delta}-\delta_o) = O_p(1),$$
$$T^{\frac{1}{2}}(\tilde{\psi}-\psi_o) = O_p(1).$$

We note that because $\psi' = [-\gamma(1):\gamma(0)+2\gamma(1)]'$, a $\sqrt{T}$-consistent estimator of ψ, say $\tilde{\psi}=\psi(\tilde{\delta})$, can be constructed from $\tilde{c}(\tau)$, where $\tilde{c}(\tau) = T^{-1}\sum \tilde{u}_t\tilde{u}_{t-\tau}$, $\tilde{u}_t = y_t - z_t\tilde{\delta}$ and $\tilde{\delta}$ is the OLS estimator of δ. From the discussion in section 2.4, we have that $T^{\frac{1}{2}}(c(\tau)-\gamma(\tau)) \to N(0,\cdot)$, where $c(\tau) = T^{-1}\sum u_t u_{t-\tau}$. However since $T^{\frac{1}{2}}(\tilde{\delta}-\delta_o) = O_p(1)$, it can be verified that the central limit theorem above holds if $c(\tau)$ is replaced by $\tilde{c}(\tau)$.

For the vector process given in (3.1), the results (1-3) take the form

$$\text{plim}\ T^{-1}\sum_{j=0}^{T-1}[F_j^{-1}\otimes I_{Z'Z}(j)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_u^{-1}(\lambda)\otimes F_Z(d\lambda), \tag{8}$$

$$\text{plim}\ T^{-1}\sum_{j=0}^{T-1}[(F_j^{-1}\otimes I_k)\ \text{vec} I_{Z'U}(j)] = 0, \tag{9}$$

$$T^{-\frac{1}{2}}\sum_{j=0}^{T-1}[(F_j^{-1}\otimes I_p)\text{vec}I_{Z'U}(j)] \xrightarrow{d} N\left[0\ ,\ \frac{1}{2\pi}\int_{-\pi}^{\pi}F_u^{-1}(\lambda)\otimes F_Z(d\lambda)\ \right], \tag{10}$$

and as discussed below (2) and (3) results (8-10) hold if F_j is replaced by $\tilde{F}_j=F_j(\tilde{\psi})$, where $T^{\frac{1}{2}}(\tilde{\psi} - \psi_o) = O_p(1)$. Hence proceeding in an analogous way to the scalar case and bearing in mind that the information matrix is block diagonal, we have

$$T^{\frac{1}{2}}\begin{bmatrix} \hat{\beta}-\beta_o \\ \hat{\psi}-\psi_o \end{bmatrix} \xrightarrow{d} N(0, IA^{-1}(\theta_o)),$$

where $\hat{\psi}$ is a single iteration of the Newton-Raphson scheme with initial value constructed from $\tilde{u}_t = y_t-\tilde{B}'z_t$, and $\hat{\beta}$ is given in (3.10) with $\hat{\psi}$ replacing ψ. $\theta_o=(\beta_o',\psi_o')'$ and $IA(\theta_o)$ is given in (3.9).

Finnaly we mention that the asymptotic results can be extended to the case where the z_t's satisfy the Grenander conditions given in Hannan (1970,page 77).

Appendix 3.1

Proof of Asymptotic Results

We start discussing the asymptotic properties of $\hat{\psi}$ given in (4.4), that is,

$$\hat{\psi} = \tilde{\psi} + \overline{H}^{-1}(\tilde{\psi})\; g_{\psi}(\tilde{\delta},\tilde{\psi}), \tag{1}$$

where

$$\overline{H}(\tilde{\psi}) = \frac{1}{2} \sum_{j=0}^{T-1} S_j{}'\tilde{f}_j{}^{-2}\; S_j \;,$$

and $\tilde{\delta},\tilde{\psi}$ are $\sqrt{T}$-consistent estimators of δ and ψ respectively. $g_{\psi}(\tilde{\delta},\tilde{\psi})$ is (3.7b) evaluated at $(\tilde{\delta},\tilde{\psi})$. We note that the first order Taylor expansion of $g_{\psi}(\tilde{\delta},\tilde{\psi})$ around $(\tilde{\delta},\psi_o)$ is

$$g_{\psi}(\tilde{\delta},\tilde{\psi}) = g_{\psi}(\tilde{\delta},\psi_o) + H_{\psi\psi}(\tilde{\delta},\psi_o)(\tilde{\psi}-\psi_o)+0_p(T^1\;) \tag{2}$$

where $H_{\psi\psi}(\tilde{\delta},\psi_o)$ is (3.8c) evaluated at $(\tilde{\delta},\psi_o)$. Hence subtracting ψ_o and multiplying by $T^{\frac{1}{2}}$ both sides of (1), from (2) we have

$$\begin{aligned} T^{\frac{1}{2}}(\hat{\psi}-\psi_o) &= \left[\frac{\overline{H}(\tilde{\psi})}{T}\right]^{-1}\left[\;\frac{\overline{H}(\tilde{\psi})}{T} + \frac{H_{\psi\psi}(\tilde{\delta},\psi_o)}{T}\right] T^{\frac{1}{2}}(\tilde{\psi}-\psi_o) \\ &+ \left[\frac{\overline{H}(\tilde{\psi})}{T}\right]^{-1} T^{-\frac{1}{2}}\; g_{\psi}(\tilde{\delta},\psi_o) + \left[\;\frac{\overline{H}(\tilde{\psi})}{T}\right]^{-1} T^{-\frac{1}{2}} 0_p(T^{-1}). \end{aligned} \tag{3}$$

Now

$$\frac{1}{T}\, H_{\psi\psi}(\tilde{\delta},\psi_o) = \frac{1}{2T} \sum_{j=0}^{T-1} S_j{}'f_j{}^{-2}\; S_j \;-\; \frac{1}{T} \sum_{j=0}^{T-1} S_j{}'f_j{}^{-3}\; I_{\tilde{U}'\tilde{U}}(j)\; S_j$$

$$- \frac{1}{2T}\sum_{j=0}^{T-1} s_j' f_j^{-2} s_j - \frac{1}{T}\sum_{j=0}^{T-1} s_j' f_j^{-3}\,[I_{\tilde{U}'\tilde{U}}(j) - I_{U'U}(j)]s_j$$

$$- \frac{1}{T}\sum_{j=0}^{T-1} s_j' f_j^{-3}\, I_{U'U}(j)\, s_j \tag{4}$$

where

$$I_{\tilde{U}'\tilde{U}}(j) - I_{U'U}(j)] = -(\tilde{\delta}-\delta_o)' I_{ZU}(j) - I_{UZ}(j)(\tilde{\delta}-\delta_o) + (\tilde{\delta}-\delta_o)' I_{ZZ}(j)(\tilde{\delta}-\delta_o) \tag{5}$$

Clearly, the first term in (4) converges to $IA(\psi_o)$. Because $\tilde{\delta}$ is $\sqrt{T}$-consistent, using (4.1) and (4.2), the second term in (4) is $o_p(T^{-\frac{1}{2}})$, and from Robinson (1978,th 2) the third term in (4) converges in probability to $-2IA(\psi_o)$. Hence

$$\text{plim } T^{-1} H_{\psi\psi}(\tilde{\delta},\psi_o) = -IA(\psi_o).$$

Now, because $\tilde{\psi}$ is $\sqrt{T}$-consistent,

$$\text{plim } T^{-1}\bar{H}(\tilde{\psi}) = IA(\psi_o),$$

hence the first term in (3) converges in probability to zero. Writing $T^{-\frac{1}{2}} g_{\psi}(\tilde{\delta},\psi_o)$ as

$$T^{-\frac{1}{2}} g_{\psi}(\tilde{\delta},\psi_o) = -\tfrac{1}{2} T^{-\frac{1}{2}} \sum_{j=0}^{T-1} s_j'[f_j - f_j^{-2} I_{U'U}(j)]$$

$$+ \tfrac{1}{2} T^{-\frac{1}{2}} \sum_{j=0}^{T-1} s_j' f_j^{-2}[I_{\tilde{U}'\tilde{U}}(j) - I_{U'U}(j)], \tag{7}$$

and arguing as above, the second term in (7) converges in probability to zero. From Robinson (1978,th 4) the limit distribution of the first term is $N(0, IA(\psi_o))$.

Obviously the third term in (3) converges in probability to zero and therefore making use of Slutsky's theorem we have that

$$T^{\frac{1}{2}}(\hat{\psi} - \psi_o) \xrightarrow{d} N(0, IA^{-1}(\psi_o)) \ . \tag{8}$$

On the other hand subtracting δ_o and multiplying by $T^{\frac{1}{2}}$ both sides of (4.7) we have

$$T^{\frac{1}{2}}(\hat{\delta} - \delta_o) = \left[T^{-1}\sum_{j=0}^{T-1}[\hat{f}_j^{-1} I_{Z'Z}(j)]\right]^{-1} T^{-\frac{1}{2}}\sum_{j=0}^{T-1}\hat{f}_j^{-1} I_{Z'U}(j), \tag{9}$$

where $\hat{f}_j = f_j(\hat{\psi})$ and $\hat{\psi} = \hat{\psi}(\tilde{\delta})$.

Now from (4.1) and from the discussion below (4.2) we have

$$\text{plim} \ T^{-1}\sum_{j=0}^{T-1}[\hat{f}_j^{-1} I_{Z'Z}(j)] = IA(\delta_o)$$

and from (4.2) and the discussion below (4.2) we have

$$T^{-\frac{1}{2}}\sum_{j=0}^{T-1}\hat{f}_j^{-1} I_{Z'U}(j) \xrightarrow{d} N\,[0 \ , \ IA(\delta_o)].$$

Hence, using Slutsky's theorem we have

$$T^{\frac{1}{2}}(\hat{\delta} - \delta_o) \xrightarrow{d} N\,[0 \ , \ IA^{-1}(\delta_o)]. \tag{10}$$

CHAPTER 4

INSTRUMENTAL VARIABLE ESTIMATION

1. Introduction

In this chapter we shall deal with the problem of estimation and asymptotic properties of the estimators for the parameters involved in a regression equation with stochastic trend components. Differently from the previous chapter some of the regressors are not assumed to be weakly exogenous variables. Hence, as discussed previously the maximum likelihood criterion function will not be a valid basis for inference. A typical situation where some of the regressors are not weakly exogenous variables arises when the equation to be estimated is a single equation from a simultaneous equation system. In other words some of the regressors are endogenous variables to the system. One way of proceeding in such situations is by estimating the whole system, that is, to construct full information estimators. We defer a detailed discussion of simultaneous equation systems to chapters 5 and 6.

As an alternative, to full information estimators, limited estimators can be derived. Such estimators are highly attractive if the other equations in the system have not been specified. All we have specified is the equation of interest, say the first, and the reduced form of the right hand endogenous variables of this equation. That is, we have the following equations

$$y_{1t} = Y_{1t}'\beta_1 + Z_{1t}'\gamma_1 + w_{1t} = X_{1t}'\delta_1 + w_{1t} \tag{1a}$$

$$Y_{1t}' = Z_{1t}'\Pi_1 + Z_{2t}'\Pi_2 + V_{1t}' = z_t'\Pi + V_{1t}' \tag{1b}$$

where Y_{1t}' is a $1xp_1$ row vector of observations on the endogenous variables and Z_{1t}' is a $1xk_1$ row vector of observations on the exogenous variables appearing in the first equation. $z_t' = (Z_{1t}', Z_{2t}')$ is the 1xK row vector of observations on all exogenous variables appearing in the system. The $1x(p_1+1)$ row vector of disturbances, $[w_{1t}\ V_{1t}']$, is assumed to follow a multivariate random walk plus noise model with associated covariance matrices $\Sigma_\varepsilon^\dagger$ and $\Sigma_\eta^\dagger$.

The estimator of the parameter vector $\theta = (\beta_1, \gamma_1, \sigma^2_\varepsilon, \sigma^2_\eta)$, where σ^2_ε and σ^2_η are respectively the top left hand elements of $\Sigma_\varepsilon^\dagger$ and $\Sigma_\eta^\dagger$, is known as a limited information estimator, limited because we do not impose the complete specification of all equations. The most efficient estimator within this class is the LIML (Limited Information Maximum Likelihood). However because LIML can be viewed as a special case of FIML (Full Information Maximum Likelihood) we defer the discussion of LIML to chapter 7, after we have discussed FIML.

Alternative limited information estimators can be obtained based on the instrumental variable principle. The construction of these estimators is examined in Harvey (1989). We present alternative estimators and discard those which have unsatisfactory properties.

The chapter is divided as follows. Section 2 contains standard material on instrumental variable estimation for a single equation. It is mainly drawn from Bowden and Turkington (1984). The reader familiar with the instrumental variables technique may skip this section. In section 3 several time-domain instrumental variable estimators for single equations with stochastic trend are presented. Section 4 deals with frequency-domain instrumental variable estimators as well as asymptotic properties of the constructed estimators.

2. Instrumental Variable Principle

In general, because $\Sigma_\varepsilon^\dagger$ and $\Sigma_\eta^\dagger$ are not diagonal matrices, Y_{1t} is correlated with w_{1t}. Now, the ML estimator of the regression coefficients derived in previous chapter is the GLS (Generalized Least Squares). Hence, because Y_1 is correlated with w_1 we have that $\text{plim}\ T^{-1}X_1'\Omega^{-1}w_1 \neq 0$, where Ω is the covariance matrix of the disturbances. Therefore, the GLS estimator will be inconsistent.

On the other hand, if the equation of interest contains no endogenous variables but a lagged dependent variable and the disturbances are serially correlated we also have a situation where regressors and disturbances are correlated. However, because in this case plim $T^{-1}Y_{-1}\Omega^{-1}w = 0$, the ML procedure provides consistent estimators for the associated parameters. Although alternative estimators based on the instrumental variable principle may be constructed, since ML methods can be applied we shall consider instrumental variable procedures mainly inside the simultaneous equation context.

The instrumental variable principle exploits the fact that even when disturbances and regressors are correlated it is often possible to use economic theory to find other variables that are uncorrelated with the disturbances, in large samples. These variables are admissible instruments in allowing us to estimate the parameter of interest. To be useful, the instruments must also be closely enough related to the regressors. The choice of the instruments is in general suggested by the structure of the model. In the case of the simultaneous equation systems a useful choice consists of the excluded exogenous variables from the equation of interest. If there are no exogenous variables excluded from the equation of interest, we have no instruments, but this is a problem of identification and will be discussed in next chapter.

In the case of serial correlation in the presence of lagged dependent variables, a useful choice is the remaining exogenous variables in the equation lagged once or twice. With a very large sample we can add as many instruments as we please. In small samples, however, a large set of instruments is in itself undesirable.

In what follows we shall assume that such admissible instruments exist and present a review of the instrumental variable estimation procedure .

Serially Uncorrelated Disturbances

To avoid overelaborate notation we shall omit all the subscripts indicating that the equation of interest is the first equation from a simultaneous equation system. The equation is then, in matrix notation,

$$y = X\,\delta + w \tag{1}$$

where $X = (Y_1 \colon Z_1)$ is a $Tx(p_1+k_1)$ matrix of observations on the regressors and w is a Tx1 vector of disturbances which we shall, initially, assume to be independent with zero mean and variance σ^2. Let Z be a Txk matrix containing the instruments. Pre-multiplying (1) by Z' yields

$$Z'y = Z'X\,\delta + Z'w. \tag{2}$$

Now if $\sigma^2 Z'Z$ is the estimated covariance matrix of the new disturbances Z'w, then applying the standard GLS formula to (2) we obtain

$$\hat{\delta} = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}\,Z'y \tag{3}$$

or

$$\hat{\delta} = (M'X)^{-1}\,M'y \tag{4}$$

where

$$M = Z(Z'Z)^{-1}Z'X. \tag{5}$$

Formula (4) is the standard textbook IV estimator. In a simultaneous equation system if the matrix Z contains the full set of exogenous variables, (3) is the two stage least squares estimator (2SLS). We also note that while Z is the matrix containing the instruments, M is known as the instrumental variable.

Serially Correlated Disturbances

We shall now consider the case where the disturbances in (1) are serially correlated, that is, $Eww' = \sigma^2\Omega$ where Ω is positive definite which we shall initially assume to be known. In handling this kind of model Bowden and Turkington (1984, Ch.3) present different estimators for δ all based on the Instrumental Variable Principle. These are:

i) The OLS analog

The Ordinary Least Squares analog of δ is obtained by applying the GLS formula to (2) with $\sigma^2 Z'\Omega Z$ as an estimate for the covariance matrix of the new disturbances $Z'w$. Thus

$$\hat{\delta} = \left[X'Z(Z'\Omega Z)^{-1}Z'X\right]^{-1}X'Z(Z'\Omega Z)^{-1}\ Z'y \tag{6}$$

(ii) The GLS Analog

In order to obtain the so called Generalized Least Squares analog the first step is to transform the serially correlated disturbances into uncorrelated disturbances. This can be achieved by pre-multiplying (1) by $F^{-\frac{1}{2}}L$ where F is a diagonal matrix and L is a lower triangular matrix with ones on the main diagonal such that $\Omega^{-1} = L'F^{-1}L$. The resulting equation to be estimated becomes

$$F^{-\frac{1}{2}}Ly = F^{-\frac{1}{2}}LX\delta + F^{-\frac{1}{2}}Lw. \tag{7}$$

Now pre-multiplying (7) by $Z'L'F^{-\frac{1}{2}}$, regarding $\sigma^2 Z'\Omega^{-1}Z$ as an estimate of the new disturbance $Z'L'F^{-1}Lw$ and applying the GLS formula yields

$$\hat{\delta}_1 = \left[X'\Omega^{-1}Z\ (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X\right]^{-1}X'\Omega^{-1}Z(Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y. \tag{8}$$

Formula (8) can be rewritten as

$$\hat{\delta}_1 = \left[X'L'F^{-\frac{1}{2}}\ P_1\ F^{-\frac{1}{2}}\ LX\ \right]^{-1}X'L'F^{-\frac{1}{2}}\ P_1F^{-\frac{1}{2}}\ Ly, \tag{9}$$

where P_1 is the idempotent projection matrix given by

$$P_1 = F^{-\frac{1}{2}}\ LZ\ (Z'\Omega^{-1}Z)^{-1}\ Z'L'F^{-\frac{1}{2}}. \tag{10}$$

If the matrix Z contains the full set of exogenous variables in the system the estimator given in (8) is also known as G2SLS (Generalized 2 Stage Least Squares) and was first used by Theil (1961).

The nomenclature OLS and GLS analog arises from the fact that if the number of instruments is the same as the number of regressors then (6) and (9) reduce respectively to

$$\hat{\delta} = [Z'X]^{-1}Z'y \qquad \text{and} \qquad \hat{\delta}_1 = [Z'\Omega^{-1}X]^{-1}Z'\Omega^{-1}y.$$

It is now readily apparent that the former can be regarded as an IV-OLS analog while the latter is an IV-GLS analog.

Bowden and Turkington (1984) also explored the relative efficiency of the IV–OLS and IV–GLS analogs. No firm conclusion is available. Nevertheless it is the IV–GLS that provides the interpretation of LIML and FIML as iterated IV estimators. Moreover, since we are primarly interested in models with stochastic trend components the estimation of the hyperparameters is somewhat simpler by means of the GLS analog. We shall therefore not consider the OLS analog any further.

iii) Alternative IV estimator

An alternative IV estimator can be obtained by pre-multiplying equation (7) by the Z', that is, without transforming the matrix of instruments. The resulting estimator is then

$$\hat{\delta}_2 = \left[X'L'F^{-\frac{1}{2}}\ Z(Z'Z)^{-1}Z'F^{-\frac{1}{2}}\ LX\right]^{-1} X'L'F^{-\frac{1}{2}}\ Z(Z'Z)^{-1}Z'F^{-\frac{1}{2}}\ Ly \tag{11}$$

or

$$\hat{\delta}_2 = \left[X'L'F^{-\frac{1}{2}}\ P_2\ F^{-\frac{1}{2}}\ LX\ \right]^{-1}X'L'F^{-\frac{1}{2}}\ P_2\ F^{-\frac{1}{2}}\ Ly \tag{12}$$

where

$$P_2 = Z(Z'Z)^{-1}Z'. \tag{13}$$

The only difference between (12) and (9) is the replacement of P_1 by

P_2. Campos (1986a) uses untransformed instruments in the estimation of single equation with ARMA disturbances. However as it is pointed out in Bowden and Turkington (1984) the estimator given in (11) is usually dominated in efficiency by one or both of the IV–OLS or IV-GLS analogs.

Finally we note that if the process generating the disturbances is stationary, an estimator of δ can be obtained using the standard IV estimator given in (3), that is, the 2SLS. If the number of instruments is bigger than the number of regressors in general the OLS analog will be more efficient than the standard IV estimator, otherwise they are exactly the same.

Asymptotic Properties of the GLS-IV Analog

The asymptotic properties of the IV–GLS estimator can be obtained straightforwardly under certain regularity conditions. Subtracting δ_o, where δ_o is the true parameter vector, from (9) and multiplying by $T^{\frac{1}{2}}$ yields

$$T^{\frac{1}{2}}(\hat{\delta} - \delta_o) = \left[\frac{X'\Omega^{-1}Z}{T}\left[\frac{Z'\Omega^{-1}Z}{T}\right]^{-1}\frac{Z'\Omega^{-1}X}{T}\right]^{-1}\frac{X'\Omega^{-1}Z}{T}\left[\frac{Z'\Omega^{-1}Z}{T}\right]^{-1} T^{-\frac{1}{2}}Z'\Omega^{-1}w.$$

Hence $\hat{\delta}$ is consistent provided that

(i) plim $T^{-1}X'\Omega^{-1}Z$ exists and has full column rank;

(ii) plim $T^{-1}(Z'\Omega^{-1}Z)$ exists and is positive definite;

(iii) $\text{plim}T^{-1}(Z'\Omega^{-1}w) = 0$.

Furthermore if (i)-(iii) hold and in addition

(vi) $T^{-\frac{1}{2}} Z'\Omega^{-1}w$ has a limiting normal distribution with zero mean and covariance matrix plim $\sigma^2 T^{-1}Z'\Omega^{-1}Z$, then $T^{\frac{1}{2}} (\hat{\delta} - \delta_o)$ has a limiting normal distribution with zero mean and covariance matrix V, where

$$V = \sigma^2 \text{ plim} \left[\frac{X'\Omega^{-1}Z}{T} \left[\frac{Z'\Omega^{-1}Z}{T} \right]^{-1} \frac{Z'\Omega^{-1}X}{T} \right]^{-1}. \tag{14}$$

We observe that, for model (1.1), V is asymptotically equivalent to

$$V = \sigma^2 \text{ plim} \left[R'\frac{Z'\Omega^{-1}Z}{T} R\right]^{-1}, \tag{15}$$

where

$$R = \begin{bmatrix} \Pi_1 & I_{k1} \\ \Pi_2 & 0 \end{bmatrix},$$

see Wickens (1969).

3. Instrumental Variable Estimation for Models with Stochastic Trends

We shall now consider the equation given in (1.1a), that is,

$$y_t = X_t'\delta + w_t, \tag{1}$$

where w_t follows a random walk plus noise process with associated parameters σ_ε^2 and σ_η^2. Again we have omitted the subscript indicating that (1) is a single equation from a simultaneous equation system.Generalizations to the local linear trend model are immediate. As in chapters 2 and 3 the initial state μ_o will be modelled in terms of a diffuse prior, that means, all estimators considered are conditional on the first observation. The estimation problem is therefore concerned with respect to $\theta=(\delta,\sigma_*^2,\psi_*)$ where $\sigma_*^2=\sigma_\varepsilon^2$ if ψ_*, the signal-noise ratio, is $\sigma_\eta^2/\sigma_\varepsilon^2$ and $\sigma_*^2=\sigma_\eta^2$ if $\psi_*=\sigma_\varepsilon^2/\sigma_\eta^2$.

When the hyperparameters σ_*^2 and ψ_* are known, the IV-GLS analog given in (2.9) can be obtained by means of the Kalman filter. That is, the Kalman filter is applied separately to y and to each column of X and Z with respective starting values y_1 and the first row of the matrices X and Z.

Similarly $\hat{\delta}_2$ given in (2.11) can be constructed by applying the Kalman filter only to y and each column of X. As already noted, the fact that the nonstationary process, w_t, is converted into a stationary one suggests that $\hat{\delta}_2$ may not be very attractive. Even though the elements of X and Z may be highly correlated, applying the Kalman filter only to X could result in the correlation becoming much smaller. On the other hand, if the instruments are differenced by the degree of differencing needed to make the stochastic trend stationary, once for the random walk plus noise model, they are likely to be more effective instruments.

In summary, denoting by $\tilde{y}$, $\tilde{X}$ and $\tilde{Z}$ the 'innovations' delivered by the Kalman filter, and assuming σ_*^2 and ψ_* to be known, the three possible IV estimators that we shall consider onwards can be regarded as the ones obtained from the minimand

$$S_i(\delta,\psi_*)^{\dagger} = (\tilde{y} - \tilde{X}\delta)'F^{-\frac{1}{2}}P_iF^{-\frac{1}{2}}(\tilde{y}-\tilde{X}\delta), \qquad i=1,2,3 \tag{2}$$

where

$$P_1 = F^{-\frac{1}{2}}\tilde{Z}[\tilde{Z}'F^{-1}\tilde{Z}]^{-1}\tilde{Z}'F^{\frac{1}{2}} \tag{3}$$

$$P_2 = Z(Z'Z)^{-1}Z' \tag{4}$$

$$P_3 = \Delta Z[\Delta Z'\Delta Z]^{-1}\Delta Z' \tag{5}$$

We shall now construct estimators of the regression coeficients as well as of the hyperparameters by bringing together both procedures, the IV and the ML discussed in chapter 3. For this, we rewrite the ML criterion function given in (3.2.3) as

$$\ell(\delta,\sigma_*^2,\psi_*) = -\tfrac{1}{2}\sum_{t=2}^{T}\log f_t - \tfrac{1}{2}(T-1)\log\sigma_*^2 - \tfrac{1}{2}\sigma_*^{-2}S(\delta,\psi_*) \tag{6a}$$

where

$$S(\delta,\psi_*) = (\tilde{y} - \tilde{X}\delta)'F^{-1}(\tilde{y}-\tilde{X}\delta), \tag{6b}$$

We recall that the optimisation of (6) can be carried out by means of the stepwise algorithm and by means of the concentrated likelihood function, with respect to σ_*^2,ψ_*. Based on these two optimisation procedures two alternative estimators can be obtained. These are:

(i) The IIV/ML

The IIV/ML (Iterated Instrumental Variable / Maximum Likelihood) estimator is closely related to the ML estimator obtained by means of the stepwise algorithm. We assume that an initial consistent estimator of δ, say $\tilde{\delta}$, is available. Later we shall discuss the construction of such an estimator. We replace δ in (6b) by $\tilde{\delta}$ and optimise (6a) with respect to σ_*^2 and ψ_*. We note that σ_*^2 can be concentrated out of (6a), hence the maximisation is nonlinear only with respect to ψ_*. The resulting estimators of $\sigma^2{}_*$ and ψ_* can be used to construct each of the feasible IV estimators, $\hat{\delta}_i$, $i = 1,2,3$ minimising (2). The procedure is then iterated until convergence is attained. Although iterating will not change the asymptotic properties of the estimators of δ and $\sigma^2{}_*$, ψ_* when there are no lagged dependent variables it may yield estimators with better small sample properties.

(ii) The IV/QML

The IV/QML (Instrumental Variable / Quasi Maximum Likelihood) estimator, as suggested in Harvey (1989), is closely related to the ML estimator obtained by optimising the concentrated likelihood function with respect to $(\sigma_*{}^2,\psi_*)$. So δ in (6b) is replaced by $\hat{\delta}$, where $\hat{\delta}$ is one of the IV estimators obtained minimising (2). Thus here $\hat{\delta}$, $\tilde{y}$, $\tilde{X}$, and f_t are all functions of the same ψ_*. The resulting concentrated criterion function becomes

$$\ell_c\ (\sigma^2{}_*,\psi_*) = -\tfrac{1}{2}\sum_{t=2}^{T}\log f_t - \tfrac{1}{2}(T-1)\log\sigma^2{}_* - \tfrac{1}{2}\sigma^2{}_*\ S_c(\psi_*) \qquad (7)$$

where

$$S_c(\psi_*) = (\tilde{y} - \tilde{X}\hat{\delta})'F^{-1}(\tilde{y}-\tilde{X}\hat{\delta}). \qquad (8)$$

We note that (7) is not the concentrated likelihood function as it was in chapter 3. It is simply a criterion function that we have obtained when proceeding in a similar way as in the case where the matrix X contains only weakly exogenous variables.

The optimisation of (7) is carried out nonlinearly with respect to ψ_*. In practice σ_*^2 can be concentrated out. Once we have found the optimal point $\hat{\psi}_*$, $\hat{\delta}$ is obtained from the minimand given in (2).

Comparison of Estimators

At first sight the estimates obtained from these two procedures might be seen as being numerically equivalent. However this is not the case and can be justified as follows. First we note that the IV/QML can be regarded as being the point $\hat{\theta}=(\hat{\delta},\hat{\sigma}_*^2,\hat{\psi}_*)$ that optimises (6) subject to the restriction

$$\hat{\delta} - [\tilde{X}F^{-\frac{1}{2}}\ P\ F^{-\frac{1}{2}}\tilde{X}]^{1}\ \tilde{X}F^{\frac{1}{2}}\ P\ F^{\frac{1}{2}}\tilde{y} = 0, \tag{9}$$

where P is any of the projection matrices given before. On the other hand the IIV/ML estimates for the hyperparameters are those which optimise (6) conditional on a given δ. Let $(\tilde{\sigma}^2{}_*,\tilde{\psi}_*)$ be the point that optimises (6) conditional on $\hat{\delta}$, where $\hat{\delta}$ is the IV/QML estimate. Of course, $(\tilde{\sigma}^2{}_*,\tilde{\psi}_*)$ will be different from $(\hat{\sigma}^2{}_*,\hat{\psi}_*)$ since the optimisation is now unrestricted.

Initial Consistent Estimator

As we have already mentioned in section 2, before discussing the asymptotic properties of the GLS-IV estimator, provided that the process generating these disturbances is stationary we may use the

2SLS as an estimator for δ. Clearly the models considered here are nonstationary, however, the variables may be differenced so as to make the disturbances stationary. Thus, after differencing once, model (1) becomes

$$\Delta y_t = \Delta X_t'\delta + u_t, \tag{10}$$

where $u_t = \eta_t + \Delta\varepsilon_t$. The 2SLS of δ is then given by

$$\tilde{\delta}_{2SLS} = \left[\Delta X'\Delta Z\,[\Delta Z'\Delta Z]^{-1}\Delta Z'\Delta X\right]^{-1}\Delta X'\Delta Z\,[\Delta Z'\Delta Z]^{-1}\Delta Z'\Delta y. \tag{11}$$

This estimator will be consistent provided that X_t does not contain lagged values of the endogenous variables. If it does, the instruments should exclude lagged values of these (differenced) variables which are correlated with u_t. (For u_t as below (10) then only those at lag one are inadmissible instruments.)

It is well known, see Wickens (1969) that the 2SLS given in (11) has a limiting normal distribution, i.e,

$$T^{\frac{1}{2}}\,(\tilde{\delta}_{2SLS} - \delta_o) \xrightarrow{d} N\,(0\,,\,V),$$

where

$$V = \operatorname{plim}\left[R'\frac{\Delta Z'\Delta Z}{T}R\right]^{-1}\left[R'\frac{\Delta Z'\Omega_u\Delta Z}{T}R\right]\left[R'\frac{\Delta Z'\Delta Z}{T}R\right]^{-1}, \tag{12}$$

where

$$\Omega_u = Euu' \quad \text{and} \quad R = \begin{bmatrix}\Pi_1 & I_{k1}\\ \Pi_2 & 0\end{bmatrix}.$$

It can be easily verified that if z_t, the K-dimensional vector containing the exogenous variables, follows a multivariate random walk model with associated disturbance covariance matrix Σ_z, then

$$\operatorname{plim} \frac{\Delta Z'\Omega_u \Delta Z}{T} = (2\sigma_\varepsilon{}^2+\sigma_\eta{}^2)\Sigma_z, \tag{13}$$

and the asymptotic covariance matrix of $\tilde{\delta}_{2SLS}$ becomes

$$\text{Avar } \tilde{\delta}_{2SLS} = T^{-1} \ (2\sigma_\varepsilon{}^2+\sigma_\eta{}^2)[R'\Sigma_z R]^{-1} \ . \tag{14}$$

There a number of ways of estimating the hyperparameters. In the simple cases, closed form expressions based on the residual autocorrelations are available as discussed in chapter 2.

4. Instrumental Variable Estimation in the Frequency Domain

The frequency domain estimation procedure which we have discussed in section 3.3 can be conveniently adapted to handle the model given in (3.1). After differencing once the equation of interest is as in (3.10), namely

$$\Delta y_t = \Delta X_t{}'\delta + u_t.$$

Comparing (3.2.9) with (3.3.10) specialized for the univariate case, we observe that the spectral ML estimator of δ can be regarded as the resulting estimator obtained from (3.2.9), by making use of asymptotically equivalent expressions similar to

$$\tilde{X}'F^{-1}\tilde{X} \approx \sum_{j=0}^{T-2} I_{x'x}(j)/f_j \,, \tag{1}$$

where the (i,k) entry of the matrix $I_{x'x}(j)$ is the crossperiodogram between Δx_i and Δx_k. Hence, using similar expressions, it follows immediately that the spectral IV estimator of δ corresponding to the time-domain IV estimator obtained from the minimand given in (3.2) with weighting matrix P_1 becomes

$$\hat{\delta}(\psi) = \left[\sum_{j=0}^{T-2} \frac{I_{x'z}(j)}{f_j} \left[\sum_{j=0}^{T-2} \frac{I_{z'z}(j)}{f_j}\right]^{-1} \sum_{j=0}^{T-2} \frac{I_{z'x}(j)}{f_j}\right]^{-1}$$
$$x \quad \sum_{j=0}^{T-2} \frac{I_{x'z}(j)}{f_j} \left[\sum_{j=0}^{T-2} \frac{I_{z'z}(j)}{f_j}\right]^{-1} \sum_{j=0}^{T-2} \frac{I_{z'y}(j)}{f_j} \tag{2}$$

where $\psi=(\sigma_\varepsilon^2,\sigma_\eta^2)$. We mentioned that if we do not transform the instruments or if we use differenced instruments we cannot have an expression for $\hat{\delta}$ in terms of the periodogram.

We can now proceed as described in section 3.3, but with (2) replacing (3.3.10) to find the spectral IIV/ML. The 2SLS estimator given in (3.12) can be used as an initial consistent estimator for δ.

Asymptotic Properties

The asymptotic theory given in section 3.4 can be straightforwardly extended to handle the case we are interested in. Because $\tilde{\delta}_{2SLS}$ given in (3.12) is $\sqrt{T}$-consistent, arguing as below (3.4.7) a $\sqrt{T}$-consistent estimator of ψ, say $\tilde{\psi}=\tilde{\psi}(\tilde{\delta}_{2SLS})$, can be obtained from the autocorrelations of the residuals $y_t-z_t\tilde{\delta}$. Moreover, from appendix 3.1, it follows immediately that $T^{\frac{1}{2}}(\hat{\psi}-\psi_o)$, where $\hat{\psi}=\hat{\psi}(\tilde{\delta})$ is as in (3.4.6) but with $\tilde{\psi}$ and $\tilde{\delta}$ as above, has normal limiting distribution with zero mean and covariance matrix $IA(\psi_o)^{-1}$, where

$$IA(\psi_o) = \frac{1}{4\pi}\int_{-\pi}^{\pi}[c(\lambda),1]'[c(\lambda),1]f_u^{-2}(\lambda)d\lambda \quad ,$$

and $c(\lambda)=2(1-\cos\lambda)$.

We now turn to the limiting distribution of $\hat{\delta}$, where $\hat{\delta}$ is given in (2) but with $f_j=f_j(\psi)$ replaced by $\hat{f}_j=f_j(\hat{\psi})$. So $\hat{\delta}$ is a feasible IV estimator. Subtracting δ_o and multiplying by $T^{\frac{1}{2}}$ yields

$$T^{\frac{1}{2}}(\hat{\delta}-\delta_o) = \hat{H}\; T^{-\frac{1}{2}}\sum_{j=0}^{T-2}\frac{I_{z'u}(j)}{\hat{f}_j} \tag{3}$$

where

$$\hat{H}=\left[\sum_{j=0}^{T-2}\frac{I_{x'z}(j)}{T\hat{f}_j}\left[\sum_{j=0}^{T-2}\frac{I_{z'z}(j)}{T\hat{f}_j}\right]^{-1}\sum_{j=0}^{T-2}\frac{I_{z'x}(j)}{T\hat{f}_j}\right]^{-1}\sum_{j=0}^{T-2}\frac{I_{x'z}(j)}{T\hat{f}_j}\left[\sum_{j=0}^{T-2}\frac{I_{z'z}(j)}{T\hat{f}_j}\right]^{-1} \tag{4}$$

Now, since $\hat{\psi}$ is a consistent estimator of ψ, assuming ΔZ totally independent from u it follows from (3.4.3) that

$$T^{-\frac{1}{2}}\sum_{j=0}^{T-2}\frac{I_{z'u}(j)}{\hat{f}_j} \xrightarrow{d} N\left[0,\ \frac{1}{2\pi}\int_{-\pi}^{\pi}f_u^{-1}(\lambda)F_z(\lambda)d\lambda\right].$$

Moreover, differencing (1.1b), that is, the reduced form of the endogenous variables included in the equation of interest, $\Delta X = (\Delta Y_1 : \Delta Z_1)$ can be written as

$$\Delta X = \Delta Z\ R + (\Delta V_1 : 0) \tag{5}$$

where

$$R = \begin{bmatrix} \Pi_1 & I_{k1} \\ \Pi_2 & 0 \end{bmatrix}. \tag{6}$$

It now follows from (3.4.2) and (3.4.1) that if ΔZ is totally independent from ΔV_1, $\hat{H}$ converges in probability to H, where H is given by

$$H = \left[\frac{1}{2\pi} R' \int_{-\pi}^{\pi} f_u^{-1}(\lambda) F_z(\lambda) d\lambda\ R \right]^{-1} R' \quad . \tag{7}$$

Hence, making use of Slutsky's theorem we have that

$$T^{\frac{1}{2}} (\hat{\delta} - \delta_o) \xrightarrow{d} N (0 , V^{-1})$$

where

$$V = \frac{1}{2\pi} R' \int_{-\pi}^{\pi} f_u^{-1}(\lambda) F_z(\lambda) d\lambda\ R \ . \tag{8}$$

We note that if the exogenous variables, z_t, follow a multivariate random walk process with disturbance covariance matrix Σ_z then the differenced variables will have constant spectrum matrix, that is,

$$F_z(\lambda) = (2\pi)^{-1} \Sigma_z$$

and the asymptotic covariance of $\hat{\delta}$ becomes

$$\text{Avar}\hat{\delta} = \frac{4\pi^2}{T} \left[R'\Sigma_z \int_{-\pi}^{\pi} f_u(\lambda)^{-1} d\lambda \; R \right]^{-1}, \tag{9}$$

which can be written as

$$\text{Avar}\hat{\delta} = T^{-1}[\sigma_\eta^4 + 4\sigma_\eta^2\sigma_\varepsilon^2]^{\frac{1}{2}} \; [R'\Sigma_z R]^{-1}, \tag{10}$$

since

$$\int_{-\pi}^{\pi} f_u(\lambda)^{-1} \, d\lambda = 2\pi \int_{-\pi}^{\pi} [2\sigma_\varepsilon^2 + \sigma_\eta^2 - 2\sigma_\varepsilon^2 \cos\lambda]^{-\frac{1}{2}} \, d\lambda$$

$$= 4\pi^2 \; [\sigma_\eta^4 + 4\sigma_\eta^2\sigma_\varepsilon^2]^{-\frac{1}{2}} \, .$$

Estimator Suggested by Hannan and Terrell

Proceeding as in Hannan and Terrell (1973) we could replace, in (2), $I_{x'z}(j)$ by $\hat{R}'I_{z'z}(j)$, where $\hat{R}$ is a consistent estimator of R given in (6). The resulting estimator of δ then becomes

$$\hat{\delta}_{HT}(\psi) = \left[\hat{R}' \sum_{j=0}^{T-2} \frac{I_{z'z}(j)}{f_j} \hat{R} \right]^{-1} \hat{R}' \sum_{j=0}^{T-2} \frac{I_{z'y}(j)}{f_j} \quad . \tag{11}$$

Making use of (11) rather then (2) we can obtain $\hat{\theta}=(\hat{\delta},\hat{\psi})$ in the same way as we have obtained the spectral IIV/ML. Clearly the resulting estimator is as efficient as the spectral IIV/ML. We note that a consistent estimator of $\hat{R}$ can be obtained estimating Π by means of the stepwise optimisation procedure described in section 3.3, although it is not necessary to iterate since the Least Squares estimator of Π given in (3.3.11) is already consistent.

When σ_ε^2 is zero, so that the disturbances have constant spectrum, (11) becomes

$$\tilde{\delta} = [\hat{R}'(\Delta Z'\Delta Z)\hat{R}]^{-1}\hat{R}'\Delta Z'\Delta y.$$

Now if $\hat{R}$ is estimated by means of LS then $\tilde{\delta}$ collapses to 2SLS, as given in (2.3). In our case, since the spectrum is not constant the estimators given in (11) and (2) are not numerically equivalent.

We mention that making use of expressions similar to (1), the time domain expression for (11) becomes

$$\hat{\delta}(\psi_*) = \left[\hat{R}'\sum_{t=2}^{T} f_t^{-1}\tilde{z}_t\tilde{z}_t'\hat{R}\right]^{-1}\hat{R}'\sum_{t=2}^{T} f_t^{-1}\tilde{z}_t\tilde{y}_t\ , \tag{12}$$

and similary as we have constructed the IIV/ML and the IV/QML we can construct these modified estimators.

Relative Asymptotic Efficiency of IIV/ML Compared with 2SLS

In the special case when the exogenous variables follow a random walk we can see that each element of the asymptotic covariance matrix of 2SLS, as given in (3.14), is greater than the corresponding element of the asymptotic covariance matrix of IIV/ML, as given in (4.10), by a factor of

$$F = \frac{(2+q)}{(q^2+4q)^{\frac{1}{2}}} ,$$

where $q=\sigma_\eta^2/\sigma_\varepsilon^2$. The closer q is to zero, the more inefficient is 2SLS. For example if q=1, F=1.34, while if q=0.01, F=10.04. As q goes to infinity, that is, σ_ε^2 goes to zero, 2SLS tends to the same distribution as IIV/ML as the disturbances in the differenced observations are tending to white noise.

CHAPTER 5

SIMULTANEOUS EQUATION SYSTEMS WITH STOCHASTIC TREND COMPONENTS

1. Introduction

In this chapter we shall introduce simultaneous equation systems with stochastic trend components and discuss the role played by stochastic trends in helping to identify a single equation in the system.

We start by specifying the model. In section 3 we present a brief review of the issue of identifiability in simultaneous equation systems with no stochastic trends, and generalize the classical rank condition to simultaneous equation systems with stochastic trends. In section 4 we show how the multivariate Kalman filter can be used to deliver the likelihood function.

2. Model Formulation

We shall consider the complete simultaneous equation system given by

$$x_t'A = y_t'B + z_t'\Gamma = w_t', \qquad t = 1,\ldots,T, \qquad (1a)$$

where $x_t' = (y_t', z_t')$ and $A = (B', \Gamma')'$. y_t is a px1 vector of observed endogenous variables, z_t is a Kx1 vector of observed exogenous variables. The non-singular pxp matrix B and the Kxp matrix Γ are unknown fixed parameters matrices of the endogenous and of the exogenous variables respectively. The px1 vector w_t contains the unobserved stochastic components and is assumed to follow a multivariate random walk plus noise model as introduced in (2.2.5) i.e.

$$w_t = \mu_t + \epsilon_t, \tag{1b}$$

$$\mu_t = \mu_{t-1} + \eta_t, \tag{1c}$$

with covariances matrices Σ_ϵ and Σ_η. The reduced form of (1) is

$$y_t' = z_t'\Pi + v_t', \tag{2a}$$

where

$$\Pi = -\Gamma B^{-1}, \tag{2b}$$

$$v_t' = w_t' B^{-1} = \mu_t^{*\prime} + \epsilon_t^{*\prime}, \tag{2c}$$

$$\mu_t^* = \mu_{t-1}^* + \eta_t^*. \tag{2d}$$

The covariance matrices of ϵ_t^* and η_t^* are respectively $\Sigma_\epsilon^* = B^{-1\prime}\Sigma_\epsilon B^{-1}$ and $\Sigma_\eta^* = B^{-1\prime}\Sigma_\eta B^{-1}$.

Combining the observations we define $X = [Y : Z]$,

$$Y = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix} = \begin{bmatrix} y_{11} \cdots y_{p1} \\ \vdots \quad\quad \vdots \\ y_{1T} \cdots y_{pT} \end{bmatrix}, \qquad Z = \begin{bmatrix} z_1' \\ \vdots \\ z_T' \end{bmatrix} = \begin{bmatrix} z_{11} \cdots z_{K1} \\ \vdots \quad\quad \vdots \\ z_{1T} \cdots z_{KT} \end{bmatrix}$$

and similary

$$W = \begin{bmatrix} w_1' \\ \vdots \\ w_T' \end{bmatrix}, \qquad \mu = \begin{bmatrix} \mu_1' \\ \vdots \\ \mu_T' \end{bmatrix}, \qquad \varepsilon = \begin{bmatrix} \varepsilon_1' \\ \vdots \\ \varepsilon_T' \end{bmatrix}, \qquad \eta = \begin{bmatrix} \eta_1' \\ \vdots \\ \eta_T' \end{bmatrix}.$$

We use the notation X_i for any submatrix of X. Thus, $X_i = [Y_i : Z_i]$, where Y_i is a $T x p_i$ submatrix of Y and Z_i is a $T x k_i$ submatrix of Z. We use the notation Y_{it}' and Z_{it}' for the t-th row of Y_i and Z_i respectively. Thus, $X_{it}' = [Y_{it}' : Z_{it}']$. We can then write (1) as

$$XA = YB + Z\Gamma = W, \tag{3a}$$

where

$$W = \mu + \varepsilon, \tag{3b}$$

$$\mu = \mu(-1) + \eta, \tag{3c}$$

and its reduced form as

$$Y = Z\Pi + V, \tag{4a}$$

where

$$V = \mu^* + \varepsilon^*, \tag{4b}$$

$$\mu^* = \mu^*(-1) + \eta^* . \tag{4c}$$

The distribution of the endogenous variables is determined by the reduced form, and in order to be able to make any statistical inference about the structural parameters, these parameters must be identifiable. As we shall see in section 3, stochastic trends play an active role in helping to identify a single equation in a simultaneous equation system. Therefore, before discussing the identification issue concerning simultaneous equation systems with stochastic trends, we present an alternative formulation for system given in (1) which is more convenient for handling individual equations.

An Alternative Formulation

Let us assume that we have *a priori* restrictions on B and Γ. We shall only consider zero restrictions, that is, the corresponding variable is excluded from the equation in question, plus normalisation constraints, that is $\beta_{ii}=1$, for $i=1,\ldots p$, and of course the symmetry constraint on Σ_ε and Σ_η. Such constraints can best be handled if we introduce a selection matrix. We define the rx1 vector δ,

$$\delta' = [\delta_1' \ldots \delta_p'], \tag{5}$$

where

$$\delta_i' = - [\beta_i' : \gamma_i'], \tag{6}$$

and the p_ix1 vector β_i and the k_i vector γ_i consist of the unknown elements in the i-th columns of the matrices B and Γ, so that $r = \sum (p_i+k_i)$. We define the $rxp(p+K)$ selection matrix S_A' such that

$$\delta = - S_A' \mathrm{vec} A. \tag{7}$$

Thus S_A' may be interpreted as a selection matrix to choose only the elements corresponding to unknown elements of A. It is easy to verify that

$$S_A' S_A = I_r, \tag{8}$$

$$S_A \delta = - \mathrm{vec}(A^+), \tag{9}$$

$$\mathrm{vec}(A) = - S_A\delta + s, \tag{10}$$

where $A^+ = A - J$, $J = [I_p : 0_{pxK}]'$ and $s = \mathrm{vec}J$. Also

$$S_A = - \frac{\partial}{\partial\delta'}\mathrm{vec}(A), \tag{11}$$

and

$$(I_p \otimes X)S_A = \begin{bmatrix} X_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & X_p \end{bmatrix}, \tag{12}$$

where $X_i=[Y_i \vdots Z_i]$ is the $Tx(p_i+k_i)$ submatrix of X formed by the p_i included endogenous variables other than the dependent variable, and by the k_i included exogenous variables, considered to be the explanatories variables included in the i-th equation. We note that while S_A' chooses the unrestricted elements of A, S_A chooses the columns of $(I_p \otimes X)$ which correspond to the included variables in each equation, other than the dependent variable.

Bearing in mind that we only have zero constraints plus normalisation constraints the i-th equation in (1a) may be written as

$$y_{it} = X_{it}'\delta_i + w_{it}, \qquad t=1,\ldots,T, \qquad i=1,\ldots,p \tag{13}$$

where $X_{it}' = [Y_{it}' \vdots Z_{it}']$ and δ_i is given in (6). The entire system can then be writen as

$$\mathrm{vec}\, Y = \begin{bmatrix} X_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & X_p \end{bmatrix} \delta + \mathrm{vec}\, W. \tag{14}$$

We note that (14) could have been obtained directly by vectoring (3a) and making use of the properties of the selection matrix S_A given in (10) and (12).

3. Identifiability

In this section we seek to extend the identifiablity conditions for simultaneous equation models in the classical case in order to take account of the role played by stochastic trends. We start the discussion from the concept of identifiability.

The Concept of Identifiability.

Let $x = (x_1, x_2, \ldots x_T)$ be a vector of random variables with continuous density function $f(x,\theta)$ where θ is a p-dimensional parameter vector. Suppose we intend to estimate θ by maximum likelihood. The identification assumption states there cannot exist $\theta^\dagger \neq \theta$ such that $\ell(\theta^\dagger;x) = \ell(\theta,x)$ for all x, where $\ell(\theta;x)$ is the loglikelihood function. If two points, $\theta^\dagger$ and θ, lead to the same loglikelihood they are said to be observationally equivalent.

The Identification Problem in Simultaneous Equation Systems

In a classical model with no stochastic trends, under the assumption that the rank of Z is K, the reduced form parameters are always identified, see Magnus and Neudecker (1988, page 333). However in the majority of the situations one is interested in the structural form parameters, and these are identified if and only if their values can be deduced from the reduced form parameters.

The general approach to identifiability is then to determine whether any observationally equivalent parameter vector $\theta^{\dagger}$ can be produced by premultiplying the transpose of (2.1a), that is,

$$B'y_t + \Gamma'z_t = w_t, \tag{1}$$

by a nonsingular *pxp* matrix F; see Hsaio(1983) for a full discussion. We note that if no restrictions are placed on B and Γ, then for any matrix F, in the reduced form, the expectation of y_t and its covariance matrix are identical for any t, which implies identical distributions under normality asssumptions. Hence B and Γ cannot be determined from the reduced form. However if *a priori* restrictions on B and Γ are placed then of course F must be such that FB' satisfies the same *a priori* restrictions as B', and $F\Gamma'$ the same as Γ'. The model is identified if the only matrix F which yields a system satisfying the same *a priori* restrictions is the identity matrix. Thus identification is achieved by imposing restrictions on the structural parameters.

Often we may be interested only in a subset of parameters, say β_i and γ_i, the unknown elements of the i-th rows of B' and Γ', which correspond to the parameters of the i-th equation. Without lost of generality we suppose that is the first equation we wish to identify. If we order the variables so that the zero coefficients in the first row of B' and Γ' appear last, we have

$$\begin{bmatrix} 1 & \beta_1' & 0' \\ - & - & - \\ B_{21}' & & B_{22}' \end{bmatrix} y_t + \begin{bmatrix} \gamma_1' & 0' \\ - & - \\ \Gamma_{21}' & \Gamma_{22}' \end{bmatrix} z_t = w_t \tag{2}$$

where β_1' is $1xp_1$, B_{21}' is $(p\text{-}1)x(p_1+1)$, B_{22}' is $(p\text{-}1)x(p\text{-}p_1\text{-}1)$, γ_1' is $1xk_1$, Γ_{21}' is $(p\text{-}1)x\ k_1$ and Γ_{22}' is $(p\text{-}1)x(K\text{-}k_1)$. Thus the matrices B_{22}' and Γ_{22}' are submatrices of B' and Γ' corresponding to the coefficients of the variables in the equations other than the first which do not appear in the first equation. Using the notation given in the previous section for single equation this leads to the question whether

$$y_{1t} = X_{1t}'\delta_1 + w_{1t}, \qquad t=1,\dots,T, \tag{3}$$

is identifiable, where $X_{1t}' = [Y_{1t}' \vdots Z_{1t}']$ and $\delta_1' = (\beta_1', \gamma_1')$.

Now equation (3) is identifiable if (2) premultiplied by (1 f') only yields an equation satisfying the same *a priori* restrictions if the $(p\text{-}1)x1$ vector f is null. In the classical model, it is well known that a necessary condition for identifiability is the order condition, $K \geq p_1+k_1$. A necessary and sufficient condition, the rank condition, requires that

$$\text{Rank}\,[B_{22}' : \Gamma_{22}'] = p\text{-}1. \tag{4}$$

If all equations are identified the system is identified.

The Identification Problem in Simultaneous Equation Systems with Stochastic Trends

We shall now consider the identification problem in simultaneous equation systems with stochastic trends. The simplest kind of stochastic trend is a random walk, and a good deal of insight into the problem can be obtained by considering this case first. Initially we consider $\Gamma' = 0$, that is, no exogenous variable are included in the model. We also assume B' normalised. The model is then

$$B'y_t = \mu_t + \varepsilon_t \tag{5a}$$

$$\mu_t = \mu_{t-1} + \eta_t, \qquad t = 1,\ldots,T. \tag{5b}$$

with associated parameters contained in θ, $\theta = (B',\Sigma_\varepsilon,\Sigma_\eta)$. Premultiplying (5a) by B'^{-1} we obtain the reduced form

$$y_t = \mu_t^* + \varepsilon_t^* , \tag{6a}$$

$$\mu_t^* = \mu_{t-1}^* + \eta_t^*, \tag{6b}$$

with associated parameters being the covariances matrices $\Sigma_\varepsilon^* = B'^{-1}\Sigma_\varepsilon B^{-1}$ and $\Sigma_\eta^* = B'^{-1}\Sigma_\eta B^{-1}$. Clearly premultiplying (5a) by F, where F is any *pxp* positive definite matrix, would result in a model with the same reduced form parameters as model (5), but with structural form parameters $\theta^\dagger = (FB',F\Sigma_\varepsilon F',F\Sigma_\eta F')$. Therefore model (5) is not identifiable. However, if, say the first equation, does not contain a stochastic trend and all the other equations contain stochastic trends, then the first equation is identifiable, since any linear combination involving the other equations would yield a stochastic trend, hence violating the distributional assumptions of the first equation. In other words the first row of matrix F must be (1,0').

In what follows we shall give some details concerning the identifiability of a single equation. We initially consider a two-equation system. Under the assumption that μ_{i0}, i=1,2, is fixed, we can express μ_t in terms of a deterministic and a stochastic part. The model is then

$$\begin{bmatrix} 1 & \beta_1 \\ \beta_2 & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix} + \begin{bmatrix} \mu_{1t}^{\dagger} \\ \mu_{2t}^{\dagger} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \quad t=1,\dots,T \tag{7a}$$

where

$$\mu^{\dagger}_{it} = \mu^{\dagger}_{i,t-1} + \eta_{it}, \quad t=1,\dots,T , \tag{7b}$$

with $\mu^{\dagger}_{io}=0$ for i=1,2.

Now if we assume that $\Sigma_\eta(1,1) = \Sigma_\eta(1,2) = 0$, that is, $\mu_{1t}^{\dagger}$ is excluded from the model, (7) can be reparametrised as

$$\begin{bmatrix} 1 & \beta_1 \\ \beta_2 & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{s} \end{bmatrix} \bar{\mu}_{2t}^{\dagger} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \quad t=1,\dots,T \tag{8a}$$

where

$$\bar{\mu}^{\dagger}_{2t} = \bar{\mu}^{\dagger}_{2,t-1} + \bar{\eta}_{2t}, \quad t=1,\dots,T \tag{8b}$$

$Var(\bar{\eta}_{2t}) = 1$ and $\bar{s}^2 = \Sigma_\eta(2,2)$.

Now the reduced form of model (8) is

$$y_{1t} = \pi_{11} + \pi_{12}\bar{\mu}_{2t}^{\dagger} + v_{1t}, \tag{9a}$$

$$y_{2t} = \pi_{21} + \pi_{22}\bar{\mu}_{2t}^{\dagger} + v_{2t}, \quad t=1,\dots,T \tag{9b}$$

where

$$\pi_{11} = (\mu_{10} - \beta_1\mu_{20})/(1 - \beta_1\beta_2)$$

$$\pi_{21} = (\mu_{20} - \beta_2\mu_{10})/(1 - \beta_1\beta_2)$$

$$\pi_{12} = -\bar{s}\,\beta_1/(1 - \beta_1\beta_2)$$

$$\pi_{22} = \bar{s}\,/(1 - \beta_1\beta_2).$$

This reduced form contains a common stochastic trend component $\bar{\mu}_{2t}^{\dagger}$. Estimators of the parameters π_{11}, π_{21}, π_{12} and π_{22} can be computed by ML and unique estimators of the structural parameters μ_{10} and β_1 obtained by noting that

$$\beta_1 = -\pi_{12}/\pi_{22}$$

and

$$\mu_{10} = \pi_{11} - \pi_{12}\pi_{21}/\pi_{22}.$$

The first equation is therefore exactly identified. If it were known that μ_{10} were zero, it would be overidentified as β_1 could also be estimated from π_{11}/π_{12}. Thus both the deterministic and the stochastic part of a stochastic trend can help in identification, but as will be seen in the general case they do not count in quite the same way.

We have just seen that identification of the parameters in the first equation is achieved because of the exclusion of the stochastic component $\mu_{1t}^{\dagger}$. If (8a) is written as

$$\begin{bmatrix} 1 & \beta_1 \\ \beta_2 & 1 \end{bmatrix}\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{10} & 0 \\ \mu_{20} & s \end{bmatrix}\begin{bmatrix} 1 \\ \bar{\mu}_{2t}^{\dagger} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \quad t=1,\dots,T \tag{10}$$

and $\bar{\mu}_{2t}{}^{\dagger}$ is regarded as an explanatory variable the rank condition given in (4) is verified provided that $\bar{s}$ is strictly positive. Thus the exclusion of the stochastic component is similar in its effect to the exclusion of an explanatory variable.

We observe that the assumption that the intitial state is fixed is not necessary. In the context of (8) identifiability of the first equation is also achieved if the initial state is modelled in terms of a diffuse prior.

The generalization to $p>2$ is straightforward. Suppose $p=3$, and there is no stochastic component at all in the first equation. For simplicity we also assume a diffuse prior for μ_{20} and μ_{30}. The model is then

$$B'\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \multicolumn{2}{c}{\bar{S}'} \end{bmatrix} \begin{bmatrix} \bar{\mu}_{2t} \\ \bar{\mu}_{3t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}, \qquad t=1,\dots,T \tag{11}$$

where $\bar{S}$ is any matrix such that $\bar{S}'\bar{S} = \bar{\Sigma}_\eta$, and $(\bar{\mu}_{2t}, \bar{\mu}_{3t})'$ follows a multivariate random walk with associated covariance matrix being the identity matrix. The reduced form of (11) is as in (6) with associated covariance matrices

$$\Sigma_\varepsilon{}^* = B'^{-1}\Sigma_\varepsilon B^{-1}$$

and

$$\Sigma_\eta{}^* = B'^{-1}\begin{bmatrix} 0 & 0 & 0 \\ 0 & \multicolumn{2}{c}{\multirow{2}{*}{$\bar{\Sigma}_\eta$}} \\ 0 & & \end{bmatrix} B^{-1}.$$

Now conditionally on the first observations the covariance matrices $\Sigma_\varepsilon{}^*$ and $\Sigma_\eta{}^*$ can be estimated by ML, and unique estimators of the srtuctural form parameters appearing in the first equation can be obtained by noting that since

$$B'\Sigma_\eta{}^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & & \\ 0 & \multicolumn{2}{c}{\bar{\Sigma}_\eta} \end{bmatrix} B^{-1} \tag{12}$$

the first row of $B'\Sigma_\eta{}^*$ is zero, and so we can recover the first row of B' from $\Sigma_\eta{}^*$.

However if one of equations two and three in (11) does not contain a stochastic trend, it can be easily verified that $\Sigma_\eta{}^*$ has rank equal to one. This in turn implies that the three equations obtained equating the first row of the matrix in the LHS of (12) to zero are identical, and therefore the first equation is not identifiable. This is also the case when the trends are perfectly correlated, in other words if they are what Engle and Granger (1987) call co-integrated. To summarize, in the context of model (11), the first equation is identifiable if and only if $\bar{\Sigma}_\eta$, the covariance matrix associated with the stochastic trends appearing in all equations other than the first equation has full rank, or equivalently if $\bar{S}$ has full rank, where $\bar{S}'\bar{S}=\bar{\Sigma}_\eta$.

We now consider the identifiability of the first equation in a general model with exogenous variables. Assuming a diffuse prior for $\bar{\mu}_0=(\mu_{20},\mu_{30},\ldots\mu_{p0})'$ and no stochastic trend component at all in the first equation, the model becomes

$$\begin{bmatrix} 1 & \beta_1' & 0' \\ \hline B_{21}' & & B_{22}' \end{bmatrix} y_t + \begin{bmatrix} \gamma_1' & 0' \\ \hline \Gamma_{21}' & \Gamma_{22}' \end{bmatrix} z_t = \begin{bmatrix} 0 \\ \bar{S}' \end{bmatrix} \bar{\mu}_t + \epsilon_t, \tag{13}$$

where $\bar{\mu}_t$ is a $(p\text{-}1)$x1 multivariate random walk with associated covariance matrix being the identity matrix . $\bar{S}$ is any $(p\text{-}1)x(p\text{-}1)$ matrix such that $\bar{S}'\bar{S} = \bar{\Sigma}_\eta$. A necessary and sufficient condition for the identifiability of the first equation is that

$$\text{Rank } [B_{22}' \; \Gamma_{22}' \; \bar{S}'] = p\text{-}1. \tag{14}$$

On the other hand if the initial state $\bar{\mu}_0$ is regarded as being fixed, the necessary and sufficient condition is that

$$\text{Rank } [B_{22}' \; \Gamma_{22}' \; \bar{\mu}_0 \; \bar{S}'] = p\text{-}1. \tag{15}$$

Local Linear Trends

Consider the p-dimensional process w_t following a multivariate local linear trend model. Taking

$$z' = (1,0) \quad \text{and} \quad T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

it follows from (2.2.4) that w_t is given by

$$w_t = \mu_t + \epsilon_t, \tag{16a}$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + \eta_t, \tag{16b}$$

$$\delta_t = \delta_{t-1} + \zeta_t, \tag{16c}$$

where η_t and ζ_t are assumed to be independent of each other with associated covariance matrices Σ_η and Σ_ζ respectively. The initial state is assumed to be fixed. Let the *pxp* positive definite matrices S_η, S_ζ and S_* such that $S_\eta'S_\eta=\Sigma_\eta$, $S_\zeta'S_\zeta= \Sigma_\zeta$ and $S_*S_\eta=S_\zeta$. Clearly S_η, S_ζ and S_* are not unique. We reparametrise (16) as

$$w_t = S_\eta'\bar{\mu}_t + \epsilon_t, \tag{17a}$$

$$\bar{\mu}_t = \bar{\mu}_{t-1} + S_*'\bar{\delta}_{t-1} + \bar{\eta}_t, \tag{17b}$$

$$\bar{\delta}_t = \bar{\delta}_{t-1} + \bar{\zeta}_t, \tag{17c}$$

where Var $\bar{\eta}_t = I_p$ and Var $\bar{\zeta}_t = I_p$.

We split up the model (17) into a deterministic and a stochastic component by writing

$$w_t = \mu_0 + \delta_0 t + S_\eta'\bar{\mu}_t^\dagger + S_\zeta'\bar{\delta}_t^\dagger + \epsilon_t \tag{18a}$$

where

$$\bar{\mu}_t^\dagger = \bar{\mu}_{t-1}^\dagger + \bar{\eta}_t, \quad \bar{\mu}_0^\dagger = 0, \quad t=1,\dots,T \tag{18b}$$

$$\bar{\delta}_t^\dagger = 2\bar{\delta}_{t-1}^\dagger - \bar{\delta}_{t-2}^\dagger + \bar{\zeta}_t, \quad \bar{\delta}_0^\dagger = \bar{\delta}_1^\dagger = 0, \quad t=2,\dots,T \tag{18c}$$

We now consider a general simultaneous equation system with unobservable components modelled as (18). If the first equation does not contain a trend component, the system in question is given by

$$\left[\begin{array}{cc|c} 1 & \beta_1' & 0' \\ \hline \multicolumn{2}{c|}{B_{21}'} & B_{22}' \end{array}\right] y_t + \left[\begin{array}{c|c} \gamma_1' & 0' \\ \hline \Gamma_{21}' & \Gamma_{22}' \end{array}\right] z_t =$$

$$\begin{bmatrix} 0 & 0 \\ \bar{\mu}_0 & \bar{\delta}_0 \end{bmatrix} \begin{bmatrix} 1 \\ t \end{bmatrix} + \begin{bmatrix} 0' \\ \bar{S}_\eta' \end{bmatrix} \bar{\mu}_t^\dagger + \begin{bmatrix} 0' \\ \bar{S}_\zeta' \end{bmatrix} \bar{\delta}_t^\dagger + \epsilon_t, \tag{19}$$

where $\bar{\mu}_t^\dagger$ and $\bar{\delta}_t^\dagger$ are $(p\text{-}1)\text{x}1$ vectors obeying the equations of the form (18b) and (18c). $\bar{S}_\eta$ and $\bar{S}_\zeta$ are $(p\text{-}1)x(p\text{-}1)$ matrices such that the covariance matrices of the $(p\text{-}1)\text{x}1$ disturbance vectors $\bar{\eta}_t$ and $\bar{\zeta}_t$ are identity matrices. Hence the necessary and sufficient condition for the identifiability of the first equation is that

$$\text{Rank}\left[B_{22}'\ \Gamma_{22}'\ \bar{\mu}_0\ \bar{\delta}_0\ \bar{S}_\eta\ \bar{S}_\zeta\right] = p\text{-}1. \tag{20}$$

The appearance of some kind of trend component in the first equation leads to a modification of (20). For example, if the first equation contains a stochastic trend which is a random walk plus drift, then $\bar{\mu}_0$, $\bar{\delta}_0$ and $\bar{S}_\eta$ disappear from (20) and only $\bar{S}_\zeta$ can help identifiability.

Note that in (19), the deterministic components, one and t, are treated in exactly the same way as the exogenous variables in z_t. The stochastic components $\bar{\mu}_t^\dagger$ and $\bar{\delta}_t^\dagger$, both contribute to identifiability since the first is white noise in the first differences while the second is white noise in second differences and so they cannot be confused. Indeed if the model contains no lagged endogenous variables, it is possible to let η_t and ζ_t be stationary, invertible stochastic processes without affecting the identifiability conditions.

4. The Likelihood Function

As in the classical case where the disturbances are serially uncorrelated, the loglikelihood function of $y = \text{vec}Y'$ can be expressed in terms of the reduced form parameters or in terms of the structural form parameters. Since the transpose of the reduced form given in (2.2a) is as in (3.2.5), it follows from (3.2.8) that conditional on the first observations the loglikelihood function of y, expressed in terms of the reduced form parameters, $\theta^* = (\Pi, \Sigma_\varepsilon^*, \Sigma_\eta^*)$, is given by

$$\ell(\theta^*) = -\tfrac{1}{2}\sum_{t=2}^{T}\log|F_t^*| - \tfrac{1}{2}\sum_{t=2}^{T}\nu_t^{*\prime}(F_t^*)^{-1}\nu_t^*, \tag{1}$$

where ν_t^* and F_t^* are delivered by the Kalman filter applied to the

$$v_t = y_t - \Pi' z_t ,$$

with starting values $a_1^* = y_1 - \Pi' z_1$ and $P_1^* = \Sigma_\varepsilon^*$.

In order to obtain the loglikelihood function in terms of the structural parameters $\theta = (B, \Gamma, \Sigma_\varepsilon, \Sigma_\eta)$ we note that premultiplying (2.2.13a), specialized for the multivariate random walk plus noise model, by B' yields

$$B'a_t^* = B'a_{t-1}^* + B'P_{t/t-1}^* B(B'F_t^* B)^{-1} B'\nu_t^*, \tag{2}$$

where the superscript * indicates that we are dealing with the reduced form, and $\nu_t^* = v_t - a_{t-1}^*$. Now if the Kalman filter is applied to the structural form

$$w_t = B'y_t + \Gamma'z_t \ ,$$

with starting values $a_1 = B'y_1 + \Gamma'z_1$ and $P_1 = \Sigma_\varepsilon$, we have

$$a_t = a_{t-1} + P_{t/t-1}F_t^{-1}\nu_t, \tag{3}$$

where $\nu_t = w_t - a_{t-1}$. Comparing (3) with (2), we note that $a_t = B'a_t^*$, $P_{t/t-1} = B'P_{t/t-1}^*B$, $F_t = B'F_t^*B$ and $\nu_t = B'\nu_t^*$. Hence

$$\sum_{t=2}^{T}\nu_t'F_t^{-1}\nu_t = \sum_{t=2}^{T}\nu_t^{*\prime}(F_t^*)^{-1}\ \nu_t^*,$$

and the loglikelihood function in terms of the structural form parameters becomes

$$\ell(\theta) = (T-1)\log|B| - \sum_{t=2}^{T}\log|F_t| \ - \sum_{t=2}^{T}\nu_t'F_t^{-1}\nu_t. \tag{4}$$

As it stands, to obtain the prediction errors ν_t via the Kalman filter we have to construct first $w_t = A'x_t = B'y_t + \Gamma'z_t$ for each t, $t=1,\ldots,T$, and then apply the Kalman filter to $A'x_t$ with hyperparameters Σ_ε and Σ_η. Alternatively, using results on matrices given in Magnus and Neudecker (1988, page 47) we can write w_t as

$$w_t = A'x_t = (I_p \otimes x_t')\text{vec}A,$$

and using (2.10) we have

$$w_t = -\ (I_p \otimes x_t')S\delta + (I_p \otimes x_t')s \ = \ y_t - \bar{X}_t'\delta,$$

where

$$\bar{X}_t' = \begin{bmatrix} X_{1t}' & 0 & 0 & 0 \\ 0 & X_{2t}' & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & X_{pt}' \end{bmatrix},$$

and X_{it}' are the endogenous variables other than the dependent variable and the exogenous variables included in the i-th equation. Hence the prediction errors can be obtained by applying the multivariate Kalman filter separately to y_t and to each column of $\bar{X}_t'$. Such formulation will be more convenient for handling LIML, as we shall see in chapter 7.

CHAPTER 6

FREQUENCY DOMAIN APPROACH TO SIMULTANEOUS EQUATION SYSTEMS WITH STOCHASTIC TREND COMPONENTS

1. Introduction

In this chapter we present a computational method for finding the spectral FIML (Full Information Maximum Likelihood) estimators, and asymptotically efficient 2-step estimators of the parameters involved in a SES (Simultaneous Equation System) with disturbances following a multivariate random walk plus noise process.

The computational method that we present is an adaptation of the Hannan and Terrell (1973) approach for simultaneous equation systems with stationary disturbances. It may also be interpreted as a reflection of the Durbin (1988) iterative scheme for the classical case, that is, serially uncorrelated, normally distributed disturbances. We shall speak of this as the uniform error-spectrum case. (We mention that although recently published, copies of Durbin's paper have been circulated since 1963). In the uniform error-spectrum case the procedure is also known as iterated 3SLS, see Hendry (1976).

As pointed out in Hausman (1983) it is unlikely that these sort of iterative schemes will have good numerical properties. However they give us an insight into the construction of asymptotically equivalent estimators.

2. The Spectral Likelihood Function

As in chapter 2, the first step is to transform the system given in (5.2.1) into a system with stationary disturbances. Thus, let the model be

$$y_t'B + z_t'\Gamma = u_t' \qquad (1a)$$

$$u_t = \eta_t + \Delta\varepsilon_t \qquad (1b)$$

where the vectors y_t, px1, and z_t, kx1, contain the differenced observable variables and the px1 vector u_t contains the unobservable components in the model. Again, for presentational convenience we have omitted the differencing operator Δ in front of y_t' and z_t'. In matrix notation (1) becomes

$$XA = YB + Z\Gamma = U \ . \qquad (2)$$

Let δ denote the vector containing the unknown elements of the matrix A. From chapter 5 we have

$$\delta = - S_A'\mathrm{vec}(A), \qquad (3)$$

$$S_A \delta = \text{vec}J' - \text{vec}A \tag{4}$$

where

$$S_A = -\frac{\partial}{\partial \delta'}\,\text{vec}(A). \tag{5}$$

and

$$J = [I_p : 0_{(pxk)}].$$

As regards the hyperparameters, we have from (2.2.7) that the distinct elements of Σ_ε and Σ_η are contained in the vector

$$\psi = \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \text{vec}[\Sigma_\varepsilon : \Sigma_\eta] \tag{6}$$

and from (2.3.13) that

$$S_j = \frac{\partial \text{vec}F_j}{\partial \psi'} = (2\pi)^{-1}[c_j D : D] \tag{7}$$

Now, as the reduced form of (2) is

$$Y = Z\Pi + UB^{-1}, \tag{8}$$

the spectral likelihood function of vecY', in terms of the reduced form parameters, is as in (3.3.4). It can be easily verified that the spectral likelihood function in terms of the structural form parameters is given by

$$\ell(\theta) = T\log\|B\| - \tfrac{1}{2}\sum_{j=0}^{T-1}\log|F(\lambda_j)| - \tfrac{1}{2}\sum_{j=0}^{T-1}\text{tr}F^{-1}(\lambda_j)I_{U'U}(\lambda_j), \tag{9}$$

where $\theta' = (\delta', \psi')$ is the vector containing the unknown parameters and

U is expressed in terms of the observations through (2).

Alternatively, replacing $I_{U'U}(\lambda_j)$ by $A'X'\Psi_j XA$ where Ψ_j is defined in (2.3.9) and denoting $F(\lambda_j)$ as F_j, (9) can be written as

$$\ell(\theta) = T\log\|B\| - \tfrac{1}{2}\sum_{j=0}^{T-1}\log|F_j| - \tfrac{1}{2}\sum_{j=0}^{T-1}\mathrm{tr}\, F_j^{-1}(XA)'\Psi_j(XA). \qquad (10)$$

Clearly, the spectral maximum likelihood estimates emerge as a numerical solution of the likelihood equations

$$\frac{\partial}{\partial\theta'}\,\ell(\theta) = 0\ . \qquad (11)$$

Numerical methods to solve (11) are described in appendix 2.1. In the classical case, that is, when the disturbances are serially uncorrelated, the standard way to obtain FIML is by concentrating the covariance matrix of the disturbances out of the likelihood function, see Hendry (1976), Hausman (1983) or Rothenberg and Leenders (1964) among others. Reinsel (1979) in handling SES models with ARMA errors also concentrated out the covariance of the disturbances of the white noise process. For the models considered here the vector ψ containing the hyperparameters cannot be concentrated out of the spectral likelihood function, except in a rather special case, Σ_η proportional to Σ_ϵ (homogeneity). Nevertheless the iterative scheme for the uniform error-spectrum case can be used as a basis to solve (11).

In preparation for the numerical solution of (11) we derive the first order conditions and the Hessian matrix.

The First Order Conditions

Magnus and Neudecker (1988,theorem 16.5) derived the information matrix for SES models with uncorrelated disturbances. Their approach is very elegant and we can generalize it to the case in which we are interested without to much difficulty.

Applying the following results

$$\text{(i)}\quad d\log|A| = \mathrm{tr}(A^{-1})dA\ , \tag{12a}$$

$$\text{(ii)}\quad dA^{-1} = -\,A^{-1}dAA'^{-1}\ , \tag{12b}$$

$$\text{(iii)}\quad \mathrm{tr}(ABCD) = \mathrm{vec}'(D')(C'\otimes A)\mathrm{vec}B\ , \tag{12c}$$

the first differential of $\ell(\theta)$ given in (10) becomes

$$d\ell = T\mathrm{vec}'B^{-1}{}'\mathrm{vec}dB + \tfrac{1}{2}\sum_{j=0}^{T-1}\mathrm{vec}'[F_j{}^{-1}A'X'\Psi_jXAF_j{}^{-1}]\mathrm{vec}dF_j$$
$$-\sum_{j=0}^{T-1}\mathrm{vec}'[(X'\Psi_jXA)F_j{}^{-1}]\mathrm{vec}dA\ -\tfrac{1}{2}\sum_{j=0}^{T-1}\mathrm{vec}'F_j{}^{-1}\mathrm{vec}dF_j.$$

Now because δ and ψ are functionally independent we have from (5) and (7) that

$$\frac{\partial\mathrm{vec}A}{\partial\theta'} = -\,(S_A{:}0)\ ;\quad \frac{\partial\mathrm{vec}B}{\partial\theta'} = -\,(I_p\otimes J)(S_A{:}0)\ ;\ \frac{\partial\mathrm{vec}F_j}{\partial\theta'}j = (0{:}S_j);$$

or

$$\begin{aligned} \mathrm{dvec}A &= -(S_A:0)d\theta\,; & \mathrm{dvec}A' &= -K_{(p+k)p}(S_A:0)d\theta; \\ \mathrm{dvec}B &= -(I_p\otimes J)(S_A:0)d\theta\,; & \mathrm{dvec}B' &= -K_p(I_p\otimes J)(S_A:0)d\theta; \\ \mathrm{dvec}F_j &= (0:S_j)d\theta\,; & \mathrm{dvec}F_j' &= (0:S_j)d\theta. \end{aligned} \tag{13}$$

Hence, the first order conditions are

$$\frac{\partial}{\partial\delta}\,\ell\ :\ -S_A'\ \mathrm{vec}\Big[\ TJ'B'^{-1}-\sum_{j=0}^{T-1}X'\Psi_j(XA)F_j^{-1})\Big]=0 \tag{14a}$$

$$\frac{\partial}{\partial\psi}\,\ell\ :\ -\frac{1}{2}\sum_{j=0}^{T-1}S_j'\mathrm{vec}\Big[(F_j^{-1}-F_j^{-1}(XA)'\Psi_j(XA)F_j^{-1})\Big]=0 \tag{14b}$$

The Hessian Matrix

The second differential of the time domain loglikelihood function for SES with normal independent disturbances is derived in Magnus and Neudecker (1988, page 339). Since we can pass the differential operator under the summation operator we end up in our case, with

$$\begin{aligned} d^2\ell = &-T\,\mathrm{tr}\,(B^{-1}dB)^2 + T\,\mathrm{tr}\,B^{-1}d^2B + \tfrac{1}{2}\sum \mathrm{tr}\,(F_j^{-1}dF)^2 \\ &-\tfrac{1}{2}\sum \mathrm{tr}\,F_j^{-1}d^2F_j - \sum\mathrm{tr}\,F_j^{-1}(dA)'X'\Psi_jXdA \\ &+2\sum \mathrm{tr}\,F_j^{-1}(dF_j)F_j^{-1}A'X'\Psi_jXdA - \sum \mathrm{tr}\,F_j^{-1}A'X'\Psi_jX\,d^2A \\ &-\sum \mathrm{tr}\,A'X'\Psi_jXA(F_j^{-1}dF_j)^2F_j^{-1} + \tfrac{1}{2}\sum \mathrm{tr}\,F_j^{-1}A'X'\Psi_jXAF_j^{-1}d^2F_j\ . \end{aligned}$$

Because $d^2F_j=0$ and $d^2A=0$, using (12c) and (13) the expression above can be rewritten as

$$d^2\ell = - T\,(d\theta)'\,[S_A:0]'(I_p\otimes J')K_p(B^{-1\prime}\otimes B^{-1})(I_p\otimes J)[S_A:0]\,d\theta$$
$$+ \tfrac{1}{2}\sum (d\theta)'\,[0:S_j]'(F_j^{-1}\otimes F_j^{-1})[0:S_j]\,d\theta$$
$$- \sum (d\theta)'\,[S_A:0]'K_{(p+k)p}'(X'\Psi_j X\otimes F_j^{-1})K_{(p+k)p}[S_A:0]\,d\theta$$
$$- 2\sum (d\theta)'[S_A:0]'K_{(p+k)p}'\,(X'\Psi_j XAF_j^{-1}\otimes F_j^{-1})[0:S_j]\,d\theta$$
$$- \sum (d\theta)'\,[0:S_j]'(I_p\otimes F_j^{-1})(F_j^{-1}\otimes A'X'\Psi_j XAF_j^{-1}[0:S_j]\,d\theta$$

where the square matrices K_p and $K_{(p+k)p}$ are commutation matrices of order p^2 and $(p+k)p$ respectively.

Now using standard rules on Kronecker products given in Magnus and Neudecker (1988,page 47) the first and the third term in the above expression can be written more compactly as

$$- T\,(d\theta)'\,[S_A:0]'(I_p\otimes J'B^{-1\prime})K_p(I_p\otimes B^{-1}J)[S_A:0]\,d\theta$$

and as

$$-(d\theta)'[S_A:0]'\,\sum(F_j^{-1}\otimes X'\Psi_j X\,[S_A:0]d\theta.$$

Thus the second differential becomes

$$d^2\ell =$$

$$-(d\theta)'[S_A:0]'\Big[T(I_p\otimes J'B^{-1\prime})K_p(I_p\otimes B^{-1}J) + \sum(F_j^{-1}\otimes X'\Psi_j X)\Big][S_A:0]d\theta$$

$$+ (d\theta)'\sum[0:S_j]'\Big[\tfrac{1}{2}(F_j^{-1}\otimes F_j^{-1}) - (F_j^{-1}\otimes F_j^{-1}A'X'\Psi_j XAF_j^{-1}\Big][0:S_j]\,d\theta$$

$$- \quad 2\,(d\theta)'\,[S_A:0]'K_{p(p+k)}\sum\Big[(X'\Psi_j XAF_j^{-1}\otimes F_j^{-1})[0:S_j]\Big]\,d\theta\ .$$

Now using the second identification theorem in Magnus and Neudecker (1988,page 189) we have

$$\frac{\partial^2 \ell}{\partial\delta\partial\delta'} = -\sum S_A'\left[(I_p \otimes J'B^{-1'})K_p(I_p \otimes B^{-1}J) \;+\; (F_j^{-1} \otimes X'\Psi_j X)\right] S_A \tag{15a}$$

$$\frac{\partial^2 \ell}{\partial\psi\partial\delta'} = -\sum S_j'\left[(F_j^{-1}A'X'\Psi_j X \otimes F_j^{-1})K_{(p+k)p}\right] S_A \tag{15b}$$

$$\frac{\partial^2 \ell}{\partial\psi\partial\psi'} = \sum S_j'\left[\tfrac{1}{2}(F_j^{-1} \otimes F_j^{-1}) - (F_j^{-1} \otimes F_j^{-1}A'X'\Psi_j XAF_j^{-1})\right] S_j. \tag{15c}$$

3. A Computational Method for Finding the Spectral ML Estimator

In order to obtain the spectral FIML (Full Information Maximum Likelihood) for δ we are going to proceed in a similar way to Hendry (1976) and Hausman (1983) when they considered simultaneous equation systems with uniform error-spectrum.

First we note that if (14b) had an analytic solution say $\hat{\psi}=\hat{\psi}(\delta)$, we could construct the concentrated spectral likelihood with respect to δ and use the Newton-Raphson method to find the spectral FIML estimator for δ by solving

$$\frac{\partial}{\partial\delta}\,\ell_c(\delta) = 0 \tag{1}$$

where $\ell_c(\delta)$ is the concentrated likelihood function, i.e., $\ell_c(\delta)=\ell(\delta,\hat{\psi}(\delta))$. However this is not the case here. The likelihood equations are non-linear in ψ and so we cannot construct the concentrated likelihood function. Nevertheless, because we only need the first derivatives of ℓ_c at a given δ, say $\hat{\delta}_k$, we can still use the iterative procedure. We note that these derivatives, evaluated at $\hat{\delta}_k$, are the derivatives of the unconcentrated likelihood with respect to δ at $(\hat{\delta}_k, \hat{\psi})$ where $\hat{\psi}=\hat{\psi}(\hat{\delta}_k)$ is the solution of (14b) with δ replaced by $\hat{\delta}_k$. Thus,

$$q(\hat{\delta}_k) = \frac{\partial}{\partial\delta} \ell_c(\delta)\Big|_{\hat{\delta}_k} = \frac{\partial}{\partial\delta} \ell(\delta,\psi)\Big|_{\hat{\delta}_k,\hat{\psi}(\hat{\delta}_k)}$$

and whether $\hat{\psi}(\hat{\delta}_k)$ is obtained analytically or numerically makes no difference. The only issue is with respect to the Hessian matrix of ℓ_c since, as pointed out before, because $\hat{\psi}$ is obtained numerically we do not have a closed form for $q(\delta)$ and therefore we obviously cannot compute the second derivatives. However we can replace the Hessian by an asymptotically equivalent matrix. We return to this point later.

We shall now derive an expression for $q(\delta)$ close to the one given in Hendry (1976,page 53). We keep the notation $q(\delta)$, although $q(\delta)$ does not have a functional form. For this purpose we need the following identity

$$\frac{1}{T}\sum_{j=0}^{T-1} A'X'\Psi_j XA\hat{F}_j^{-1} = I_p \tag{2}$$

where $\hat{F}_j = F_j(\hat{\psi})$ and $\hat{\psi}=\hat{\psi}(\delta)$ solves (2.14b) for a given δ. Clearly, (2) is trivially satisfied in the uniform error-spectrum case since in this case $\hat{\Sigma} = U'U/T$ solves (2.14b). Also, in the nonparametric framework (2) is satisfied by construction, see Espasa (1977). However in our case is not as immediate as it might appear. We leave the proof to appendix 6.1 to avoid discontinuity. Now, making use of

(2) we have

$$TJ'B'^{-1} = \sum_{j=0}^{T-1} J'B'^{-1}A'X'\Psi_j XA\hat{F}_j^{-1} \ , \tag{3}$$

where $J=[I_p:0_{pxk})$. Therefore, if in equation (2.14a), we replace F_j by $\hat{F}_j$, and the first term within the brackets by (3) we have

$$q(\delta) = \frac{\partial}{\partial\delta}\ell\Big|_{(\delta,\hat{\psi})} = - S_A'\mathrm{vec}\left\{\sum_{j=0}^{T-1}[J'B'^{-1}A'X'\Psi_j XA\hat{F}_j^{-1} - X'\Psi_j XA\hat{F}_j^{-1}]\right\}$$

$$= - S_A'\mathrm{vec}\left[\sum_{j=0}^{T-1}[J'B'^{-1}(B':\Gamma')\begin{bmatrix}Y'\\Z'\end{bmatrix}\Psi_j XA\ \hat{F}_j^{-1} - \begin{bmatrix}Y'\\Z'\end{bmatrix}\Psi_j XA\ \hat{F}_j^{-1}]\right]$$

$$= - S_A'\mathrm{vec}\left[\sum_{j=0}^{T-1}\begin{bmatrix}Y'\Psi_j XA\hat{F}_j^{-1} + B'^{-1}\Gamma'Z'\Psi_j XA\ \hat{F}_j^{-1}\\ 0\end{bmatrix} - \begin{bmatrix}Y'\Psi_j XA\hat{F}_j^{-1}\\ Z'\Psi_j XA\hat{F}_j^{-1}\end{bmatrix}\right]$$

$$= S_A'\mathrm{vec}\left[\sum_{j=0}^{T-1}\begin{bmatrix}-B'^{-1}\Gamma'Z'\\ Z'\end{bmatrix}\Psi_j XA\hat{F}_j^{-1}\right]. \tag{4}$$

Hence, the spectral FIML estimator of δ emerges as a solution of

$$q(\delta) = S_A'\ \mathrm{vec}\left[\sum_{j=0}^{T-1} P'Z'\Psi_j\ XA\ \hat{F}_j^{-1}\right] = 0 \tag{5}$$

where

$$P = [\ \Pi : I_k\ ] \tag{6}$$

and

$$\Pi = - \Gamma B^{-1}. \tag{7}$$

The solution of (5) can be obtained by the Newton-Raphson iterative procedure described in appendix 2.1. The first differential, $dq(\delta)$

is

$$dq(\delta) = S_A' \operatorname{vec}\sum_{j=0}^{T-1}\left[dP'Z'\Psi_j XA\hat{F}_j^{-1} + P'Z'\Psi_j X\, dA\, \hat{F}_j^{-1} + P'Z'\Psi_j XAd\hat{F}_j^{-1}\right]$$

However, as we do not have a closed form for $\hat{F}_j(\delta)$ we cannot obtain its differential. Nevertheless because z_t and u_t are totally independent, from the results given in section 3.4, we have

$$\operatorname{plim} T^{-1}\operatorname{vec}\sum_{j=0}^{T-1} dP'Z'\Psi_j XA\hat{F}_j^{-1} = (I_p \otimes dP')\operatorname{plim} T^{-1}\sum_{j=0}^{T-1}(\hat{F}_j^{-1}\otimes I_k)\operatorname{vec} I_{Z'U}(j) = 0$$

and

$$\operatorname{plim} T^{-1}\operatorname{vec}\sum_{j=0}^{T-1} P'Z'\Psi_j Xd\hat{F}_j^{-1} = (I_p \otimes P')\operatorname{plim} T^{-1}\sum_{j=0}^{T-1}(d\hat{F}_j\otimes I_k)\operatorname{vec} I_{Z'U}(j) = 0.$$

Therefore $\partial q(\delta)/\partial\delta$ is asymptotically equivalent to H, where

$$H = -S_A'\left[\sum_{j=0}^{T-1}\hat{F}_j^{-1}\otimes P'Z'\Psi_j X\right]S_A, \tag{8}$$

in the sense that

$$\begin{aligned}\operatorname{plim} T^{-1} dq(\delta) &= \operatorname{plim} T^{-1} S_A'\operatorname{vec}\left[\sum_{j=0}^{T-1} P'Z'\Psi_j X(dA)\hat{F}_j^{-1}\right]S_A \\ &= \operatorname{plim} T^{-1} H d\delta,\end{aligned}$$

where the last equality above follows from (2.13).

Moreover, replacing X by ZP + [V:0] in (8) we have

$$\begin{aligned}\operatorname{plim} T^{-1} dq(\delta) = &- \operatorname{plim} T^{-1} S_A'\left[\sum_{j=0}^{T-1}(\hat{F}_j^{-1}\otimes P'Z'\Psi_j ZP)\right]S_A d\delta \\ &- \operatorname{plim} T^{-1} S_A'\left[\sum_{j=0}^{T-1}(\hat{F}_j^{-1}\otimes P'Z'\Psi_j [V:0])\right]S_A d\delta.\end{aligned} \tag{9}$$

Now, since z_t and v_t are totally independent, from the results given in section 3.4, it follows that the last term in (9) is equal to zero. Hence $\partial q(\delta)/\partial\delta$ is also asymptotically equivalent to $\bar{H}$, where

$$\bar{H} = - S_A' \Big[\sum_{j=0}^{T-1} (\hat{F}_j^{-1} \otimes P'Z'\Psi_j ZP) \Big] S_A . \tag{10}$$

Now if we make use of (10) the iterative scheme becomes

$$\hat{\delta}_{k+1} = \hat{\delta}_k + \Big[S_A' \sum_{j=0}^{T-1} [\hat{F}_j^{-1} \otimes \hat{P}'Z'\Psi_j \hat{Z}P]\ S_A \Big]^{-1} \Big[S_A' \mathrm{vec} \sum_{j=0}^{T-1} [\hat{P}'Z'\Psi_j \hat{X}\hat{A}F_j^{1}\ ] \Big] \tag{11}$$

where here $\hat{F}_j = F_j(\hat{\psi})$, $\hat{\psi} = \hat{\psi}(\hat{\delta}_k)$. $\hat{P}$ and $\hat{A}$ are also constructed from $\hat{\delta}_k$. Because $Z'\Psi_j Z = I_{Z'Z}(\lambda_j)$ and $Z'\Psi_j X = I_{Z'X}(\lambda_j)$ where $I_{Z'Z}(\lambda_j)$ and $I_{Z'X}(\lambda_j)$ are the real part of the respective periodogram matrices (11) could also be written as

$$\hat{\delta}_{k+1} = \hat{\delta}_k + \Big[S_A' \sum_{j=0}^{T-1} [\hat{F}_j^{-1} \otimes \hat{P}' I_{Z'Z}(\lambda_j) \hat{P}] S_A \Big]^{-1} \Big[S_A' \mathrm{vec} \sum_{j=0}^{T-1} [\hat{P}' I_{Z'X}(\lambda_j) \hat{A}F_j^{1}\ ] \Big] \tag{12}$$

For SES models with stationary disturbances Hannan and Terrell (1973) and Espasa (1977) considered the same iterative scheme as given in (12) but with $\hat{F}_j$ replaced at each iteration by a nonparametric estimate of the spectrum of the process U based on $\hat{U} = X\hat{A}_k$. Thus,

$$\hat{F}_k(\lambda_j) = \frac{1}{m_j} \sum_{w_t \in A_j} I_{\hat{u}}(w_t) \qquad\qquad j=1,\ldots 2M$$

where A_j consists of m_j of the w_t frequencies clustered around

$\lambda_j = -\pi + \pi j/M$, $j=1,\ldots 2M$. Our method thus differs from the Hannan and Terrell (1973) iterative scheme in the use of a parametric estimator of the spectrum. Assuming our model to be correct, we would expect our estimator to have better small sample properties.

On the other hand if we make use of (8) the iterative scheme becomes

$$\hat{\delta}_{k+1} = [-H]^{-1}\Big[S_A'[\sum_{j=0}^{T-1}\hat{F}_j^{-1} \otimes \hat{P}'Z'\Psi_j X]S_A\hat{\delta}_k + S_A'\mathrm{vec}\sum_{j=0}^{T-1}P'Z'\Psi_j\hat{X}\hat{A}F_j^{-1}\Big]. \quad (13)$$

Now vectoring and making use of (2.4), the second term in square brackets in (13) becomes

$$-\; S_A'\Big[\sum_{j=0}^{T-1}[\hat{F}_j^{-1}\otimes P'Z'\Psi_j X]\Big]\; S_A\hat{\delta}_k + S_A'\sum_{j=0}^{T-1}[\;\hat{F}_j^{-1}\otimes P']\mathrm{vec}Z'\Psi_j Y\;.$$

Therefore (13) takes the form

$$\hat{\delta}_{k+1} = \Big[S_A'\sum_{j=0}^{T-1}[\hat{F}_j^{-1}\otimes \hat{P}'Z'\Psi_j X]\;S_A\Big]^{-1} S_A'\sum_{j=0}^{T-1}[\hat{F}_j^{-1}\otimes\hat{P}']\mathrm{vec}Z'\Psi_j Y, \quad (14)$$

where $\hat{F}_j$ and $\hat{P}$ are formed from $\hat{\delta}_k$. In terms of the periodogram notation (14) can be written as

$$\hat{\delta}_{k+1} = \Big[S_A'\sum_{j=0}^{T-1}[\hat{F}_j^{-1}\otimes \hat{P}'I_{Z'X}(\lambda_j)]\;S_A\Big]^{-1} S_A'\sum_{j=0}^{T-1}[\hat{F}_j^{-1}\otimes\hat{P}']\mathrm{vec}I_{Z'Y}(\lambda_j). \quad (15)$$

Clearly both procedures (12) and (15) are numerically equivalent. In the uniform error-spectrum case, if $\hat{\delta}_1$ is the 3SLS (Three Stages Least Squares) estimator, the iterative scheme given in (15) is known as iterated 3SLS, see Hendry (1976).

It remains to say something about the spectral ML estimator for ψ. We have, for both iterative procedures, (12) and (15), that at each iteration $\hat{\psi}(\hat{\delta}_k)$ solves (2.14b). Thus, if $\hat{\delta}$ is the convergence point, i.e., $\hat{\delta}$ solves (5), $\hat{\psi}=\hat{\psi}(\hat{\delta})$ will be the spectral FIML estimate of ψ.

Before presenting asymptotically efficient 2-Step estimators for θ, we shall derive the asymptotic information and covariance matrices.

4. Asymptotic Information and Covariance Matrices

Asymptotic Information Matrix

Let $\theta_o=(\delta_o,\psi_o)$ the true parameter vector. The asymptotic information matrix for θ_o is

$$IA(\theta_o) = - \lim_{T\to\infty} T^{-1} E \left.\frac{\partial^2 \ell}{\partial\theta\partial\theta'}\right|_{\theta_o} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} ,$$

where the second derivatives are given in (2.15). We write X as

$$X = ZP_o + UB_o{}^{-1}J \ , \text{ where } \ J = [I_p:0_{pxk}],$$

and

$$X'\Psi_jX = P_o'Z'\Psi_jZP_o + P_o'Z'\Psi_jUB_o^{-1}J + J'B_o^{-1}{}'U'\Psi_jZP_o + J'B_o^{-1}{}'U'\Psi_jUB_o^{-1}J.$$

Since z_t and u_t are totally independent and u_t has zero mean we have

$$E\ X'\Psi_jX = E\ P_o'Z'\Psi_jZP_o + E\ J'B_o^{-1}{}'U'\Psi_jUB_o^{-1}J \ . \tag{1}$$

Also

$$EA_o'X'\Psi_jX = EU'\Psi_jUB_o^{-1}J \quad . \tag{2}$$

Now let

$$\bar{S} = (I_p \otimes B_o^{-1}J)S_A. \tag{3}$$

Making use of rules on Kronecker products, it follows immediately from (1), (2), (3), (2.3.13), (2.3.14) that

$$f_{11} = \bar{S}'K_p\bar{S} + S_A'\frac{1}{2\pi}\int[F(\lambda)^{-1}\otimes P_o'F_z(d\lambda)P_o]S_A + \bar{S}'\frac{1}{2\pi}\int[F(\lambda)^{-1}\otimes F(\lambda)]d\lambda\bar{S} \tag{4a}$$

$$f_{21} = f_{12}' = \frac{1}{2\pi}\int S(\lambda)'\ [F(\lambda)^{-1}\otimes I_p]\ d\lambda\ \ \bar{S} \tag{4b}$$

$$f_{22} = \frac{1}{4\pi}\int S(\lambda)'\ [F(\lambda)^{-1}\otimes F(\lambda)^{-1}]\ S(\lambda)\ d\lambda\ , \tag{4c}$$

where

$S(\lambda) = (2\pi)^{-1}[c(\lambda)D : D]$,

$c(\lambda) = 2(1-\cos\lambda)$,

and D is the duplication matrix.

Asymptotic Covariance Matrix

We shall use the notation $\mathrm{Avar}(\hat{\theta})$ for the asympotic covariance matrix of $\hat{\theta}$. We remark that $\mathrm{Avar}(\hat{\theta})$ relates to the distribution of $\hat{\theta}$ and not $T^{\frac{1}{2}}\hat{\theta}$. Thus,

$$\mathrm{Avar}(\hat{\theta}) = T^{-1}IA^{-1}(\theta_o)\ .$$

In appendix 6.2 we show that the inverse of the asymptotic information matrix is given by

$$IA^{-1}(\theta_o) = \begin{bmatrix} f^{11} & f^{12} \\ f^{21} & f^{22} \end{bmatrix},$$

where

$$f^{11} = \left[\frac{1}{2\pi} S_A'\int (F(\lambda)^{-1}\otimes P'F_z(d\lambda)P)\ S_A\ \right]^{-1} \tag{5a}$$

$$f^{12} = -2\left[\frac{1}{2\pi} S_A'\int (F(\lambda)^{-1}\otimes P'F_z(d\lambda)P)S_A\right]^{-1}\bar{S}'\left[(I_p\otimes\Sigma_\varepsilon)D^{+'}:(I_p\otimes\Sigma_\eta)D^{+'}\right] \tag{5b}$$

$$f^{22} = 16\pi^3\int \begin{bmatrix} D^+\Delta^{-1} \\ D^+\Delta^{-1}c(\lambda) \end{bmatrix} \left[\ F(\lambda)^{-1}\otimes F(\lambda)^{-1}\ \right] \begin{bmatrix} D^{+'} \\ D^{+'}c(\lambda) \end{bmatrix} d\lambda$$

$$+\quad 4\begin{bmatrix} D^+(I_p\otimes\Sigma_\varepsilon) \\ D^+(I_p\otimes\Sigma_\eta) \end{bmatrix}\bar{S}\ f^{11}\bar{S}'\ \left[(I_p\otimes\Sigma_\varepsilon)D^{+'}:(I_p\otimes\Sigma_\eta)D^{+'}\right] \tag{5c}$$

where

$$\Delta = \iint (F(\lambda)^{-1}F(\nu)^{-1} \otimes F(\lambda)^{-1}F(\nu)^{-1})\ (c(\lambda)^2-c(\lambda)c(\nu))d\lambda d\nu\ , \tag{6}$$

and $F(\lambda)=F(\lambda,\theta_o)$, $\theta_o=(\delta_o,\psi_o)$, $\Sigma_\varepsilon=\Sigma_\varepsilon(\psi_o)$, $\Sigma_\eta=\Sigma_\eta(\psi_o)$, $P=P(\delta_o)$.

5. Asymptotically Efficient 2-step Estimators

We shall now deal with the problem of constructing asymptotically efficient 2-step estimators for δ and ψ. We begin presenting $\sqrt{T}$-consistent estimators for δ and ψ, since it is well known that the first step in such estimators consists in finding $\sqrt{T}$-consistent estimators.

A $\sqrt{T}$-consistent Estimator for Coefficients

A $\sqrt{T}$-consistent estimator for δ can be constructed as follows. Let

$$Y = Z\Pi + V$$

be the reduced form of the model with associated covariances matrices $\Sigma_\varepsilon{}^* = B^{-1\prime}\Sigma_\varepsilon B^{-1}$ and $\Sigma_\eta{}^* = B^{-1\prime}\Sigma_\eta B^{-1}$. Let ψ^* be the $p(p+1)\text{x}1$ vector obtained from $\text{vec}(\Sigma_\varepsilon{}^*:\Sigma_\eta{}^*)$ by eliminating the supradiagonal elements of $\Sigma_\varepsilon{}^*$ and $\Sigma_\eta{}^*$. In section 3.3 we have seen how to construct an estimator of Π and ψ^*. Let these estimators be $\tilde{\Pi}$ and $\tilde{\psi}^*$. Let $\tilde{F}_V(j) = F_V(\lambda_j,\tilde{\psi}^*)$ be the estimated spectrum matrix of the reduced form disturbances. Let $\tilde{P} = [\tilde{\Pi}:I_k]$. The formula for $\tilde{\delta}$ suggested by Hannan and Terrell (1973) is

$$\tilde{\delta} = \left[S_A{}'\sum_{j=0}^{T-1}(\tilde{F}_V(j)^{-1}\otimes\tilde{P}'I_{Z'Z}(j)\tilde{P})S_A\right]^{-1}S_A{}'\sum_{j=0}^{T-1}[\tilde{F}_V(j)^{-1}\otimes\tilde{P}']\text{vec}I_{Z'Y}(j) \quad (1)$$

Proceeding in the same way as Hannan and Terrell (1973), we can show that $T^{\frac{1}{2}}(\tilde{\delta} - \delta_o) = O_p(1)$. For details, see appendix 6.3.

A $\sqrt{T}$-consistent Estimator for Hyperparameters

Let $\tilde{\psi}^*$ be the estimator of the reduced form hyperparameters as discussed in section 3.3, and let $\tilde{\delta}$ be the estimator of δ given in

(1). A natural estimator for ψ, the structural form hyperparameters, is then given by

$$\tilde{\psi} = \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \mathrm{vec}[\tilde{B}'\tilde{\Sigma}_\epsilon^*\tilde{B} : \tilde{B}'\tilde{\Sigma}_\eta^*\tilde{B}] \ , \tag{2}$$

where $\tilde{B} = \tilde{B}(\tilde{\delta})$, $\tilde{\Sigma}_\epsilon^* = \Sigma_\epsilon^*(\tilde{\psi}^*)$ and $\tilde{\Sigma}_\eta^* = \Sigma_\eta^*(\tilde{\psi}^*)$. Later we shall discuss the asymptotic efficiency of the 2-step estimator for ψ. Using the same approach it can be easily verified that $\tilde{\psi}$ is $\surd T$-consistent.

The 2-step Estimator for Coefficients

Let $\hat{\delta}$ as given in (3.15), that is,

$$\hat{\delta} = \Big[S_A'\sum_{j=0}^{T-1}[\tilde{F}_j^{-1} \otimes \tilde{P}'I_{Z'X}(j)]\ S_A\ \Big]^{-1} S_A'\sum_{j=0}^{T-1}[\tilde{F}_j^{-1} \otimes \tilde{P}']\mathrm{vec}I_{Z'Y}(j) \tag{3}$$

where $\tilde{P} = P(\tilde{\delta})$, $\tilde{F}_j = F_j(\tilde{\psi})$, $\tilde{\delta}$ and $\tilde{\psi}$ are $\surd T$-consistent estimators for δ and ψ respectively.

We shall show that

$$T^{\frac{1}{2}}(\hat{\delta}-\delta_o) \xrightarrow{d} N(0, f^{11}), \tag{4}$$

where f^{11} is given in (4.5a).

Subtracting δ_o and multiplying by $T^{\frac{1}{2}}$ both sides of (3) yields

$$T^{\frac{1}{2}}(\hat{\delta}-\delta_o) = [-\tilde{H}T^{-1}]^{-1} T^{-\frac{1}{2}} \Big\{S_A'(I_p \otimes \tilde{P}')\sum_{j=0}^{T-1}[\tilde{F}_j^{-1} \otimes I_k]\mathrm{vec}I_{Z'Y}(j) + \tilde{H}T^{-1}\ \delta_o\Big]\Big\} \tag{5}$$

where

$$\tilde{H} = -\left[S_A' \sum_{j=0}^{T-1} [\tilde{F}_j^{-1} \otimes \tilde{P}' I_{Z'X}(j)]\ S_A\right]. \tag{6}$$

Now from (6) and (2.4) we have that

$$\tilde{H}T^{-1}\delta_o = S_A' \sum_{j=0}^{T-1} [\tilde{F}_j^{-1} \otimes \tilde{P}' I_{Z'X}(j)] \mathrm{vec} A_o - S_A' \sum_{j=0}^{T-1} (\tilde{F}_j^{-1} \otimes \tilde{P}') \mathrm{vec}\ I_{Z'Y}(j), \tag{7}$$

hence (5) becomes

$$T^{\frac{1}{2}}(\hat{\delta} - \delta_o) = [-\tilde{H}T^{-1}]^{-1}\ S_A'(I_p \otimes \tilde{P}')\left[T^{\frac{1}{2}} \sum_{j=0}^{T-1} (\tilde{F}_j^{-1} \otimes I_k) \mathrm{vec} I_{Z'U}(j)\right]. \tag{8}$$

Now, since $\tilde{\delta}$ and $\tilde{\psi}$ are $\sqrt{T}$-consistent, arguing as in (3.10) we have that

$$\begin{aligned} \mathrm{plim}\ \tilde{H}T^{-1} &= -S_A'(I_p \otimes P')\ \mathrm{plim}\left[T^{-1} \sum_{j=0}^{T-1} [\tilde{F}_j^{-1} \otimes I_{Z'Z}(j)]\right] (I_p \otimes P) S_A \\ &= -S_A'(I_p \otimes P')\ Q_u\ (I_p \otimes P) S_A, \end{aligned} \tag{9}$$

where the last equality follows from (3.4.8), Q_u being

$$Q_u = \frac{1}{2\pi} \int F_u^{-1}(\lambda) \otimes F_z(\lambda) d\lambda. \tag{10}$$

From (3.4.10) we also have that the term in squared brackets in (8) converges in distribution to $N(0, Q_u)$. Hence (4) follows from Slutsky's theorem.

Finally we mention that the asymptotic covariance matrix of $\hat{\delta}$ can be consistently estimated by

$$\mathrm{avar}(\hat{\delta}) = T^{-1}\left[S_A'(I_p\otimes\tilde{P}')\sum_{j=0}^{T-1}\frac{[\tilde{F}_j^{-1}\otimes I_{Z'Z(j)}]}{T}(I_p\otimes\tilde{P})S_A\right]^{-1}. \tag{11}$$

The 2-step Estimator for Hyperparameters

Let $\hat{\delta}$ as in (3) and $\hat{\Pi} = \hat{\Pi}(\hat{\delta})$. Hence $\hat{\Pi}$ is an estimator of Π which takes into account the restrictions on B and Γ. Let $\hat{\psi}^* = \hat{\psi}^*(\hat{\Pi})$ be an estimator of the reduced form hyperparameters. Thus $\hat{\psi}^*$ is obtained by maximising the spectral likelihood function for vec Y', in terms of the reduced form parameters, conditional on $\hat{\Pi}$. Clearly $T^{\frac{1}{2}}\mathrm{vec}(\hat{\Pi}-\Pi_o) = O_p(1)$ and therefore as discussed in section 3.4

$$T^{\frac{1}{2}}(\hat{\psi}^* - \psi_o^*) \to N(0, IA^{-1}(\psi_o^*)), \tag{12}$$

where $IA(\psi_o^*)$ is the bottom right block of (3.3.9), that is,

$$IA^{-1}(\psi_o^*) = 16\pi^3\begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix}\begin{bmatrix} \int c^2(F_v^{-1}\otimes F_v^{-1}) & \int c(F_v^{-1}\otimes F_v^{-1}) \\ \int c\,(F_v^{-1}\otimes F_v^{-1}) & \int (F_v^{-1}\otimes F_v^{-1}) \end{bmatrix}^{-1}\begin{bmatrix} D^{+\prime} & 0 \\ 0 & D^{+\prime} \end{bmatrix} \tag{13}$$

where $F_v = F_v(\lambda)$ is the spectrum of the reduced form disturbances and $c = c(\lambda) = 2(1-\cos\lambda)$.

Now let

$$\hat{\psi} = \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix}\mathrm{vec}[\hat{B}'\hat{\Sigma}_\varepsilon^*\hat{B} : \hat{B}'\hat{\Sigma}_\eta^*\hat{B}] \tag{14}$$

where $\hat{B} = \hat{B}(\hat{\delta})$ and

$$\mathrm{vec}[\hat{\Sigma}_\varepsilon^* : \hat{\Sigma}_\eta^*] = \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix}\hat{\psi}^*.$$

We shall now show that

$$T^{\frac{1}{2}}(\hat{\psi}-\psi_o) \xrightarrow{d} N(0,f^{22}) \tag{15}$$

where f^{22} is given in (4.5c).

First we note that

$$[(\hat{\Sigma}_\epsilon:\hat{\Sigma}_\eta) - (\Sigma_\epsilon:\Sigma_\eta)_o] = [\hat{B}'\hat{\Sigma}_\epsilon^*\hat{B} : \hat{B}'\hat{\Sigma}_\eta^*\hat{B}] - [B_o'\Sigma_{\epsilon o}^* B_o : B_o'\Sigma_{\eta o}^* B_o]$$

$$\begin{aligned} &= [(\hat{B}'(\hat{\Sigma}_\epsilon^*-\Sigma_{\epsilon o}^*)\hat{B} \quad : \hat{B}'(\hat{\Sigma}_\eta^*-\Sigma_{\eta o}^*)\hat{B}] \\ &+ [(\hat{B}-B_o)'\Sigma_{\epsilon o}^*B_o \quad : (\hat{B}-B_o)'\Sigma_{\eta o}^*B_o] \\ &+ [(\hat{B}'\Sigma_{\epsilon o}^*(\hat{B}-B_o) \quad : \hat{B}'\Sigma_{\eta o}^*(\hat{B}-B_o)] \,. \end{aligned}$$

Hence,

$\mathrm{vec}[(\hat{\Sigma}_\epsilon:\hat{\Sigma}_\eta) - (\Sigma_\epsilon:\Sigma_\eta)_o] =$

$$= \begin{bmatrix} \mathrm{vec}\ \hat{B}'(\hat{\Sigma}_\epsilon^*-\Sigma_{\epsilon o}^*)\hat{B} \\ \mathrm{vec}\ \hat{B}'(\hat{\Sigma}_\eta^*-\Sigma_{\eta o}^*)\hat{B} \end{bmatrix} + \begin{bmatrix} \mathrm{vec}\ \hat{B}'\Sigma_{\epsilon o}^*(\hat{B}-B_o) \\ \mathrm{vec}\ \hat{B}'\Sigma_{\eta o}^*(\hat{B}-B_o) \end{bmatrix} + \begin{bmatrix} \mathrm{vec}(\hat{B}-B_o)'\Sigma_{\epsilon o}^* B_o \\ \mathrm{vec}(\hat{B}-B_o)'\Sigma_{\eta o}^* B_o \end{bmatrix}$$

or

$$= \begin{bmatrix} \hat{B}'\otimes\hat{B}' & 0 \\ 0 & \hat{B}'\otimes\hat{B}' \end{bmatrix} \mathrm{vec}\,[(\hat{\Sigma}_\epsilon^*-\Sigma_{\epsilon o}^*) : (\hat{\Sigma}_\eta^*-\Sigma_{\eta o}^*)]$$

$$+ \begin{bmatrix} I_p\otimes\hat{B}'\Sigma_{\epsilon o}^* \\ I_p\otimes\hat{B}'\Sigma_{\eta o}^* \end{bmatrix} \mathrm{vec}(\hat{B}-B_o) + \begin{bmatrix} B_o'\Sigma_{\epsilon o}^*\otimes I_p \\ B_o'\Sigma_{\eta o}^*\otimes I_p \end{bmatrix} K_p \mathrm{vec}(\hat{B}-B_o)$$

and therefore

$$\begin{aligned} T^{\frac{1}{2}}(\hat{\psi}-\psi_o) &= \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \begin{bmatrix} \hat{B}'\otimes\hat{B}' & 0 \\ 0 & \hat{B}'\otimes\hat{B}' \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix} T^{\frac{1}{2}}(\hat{\psi}^*-\psi_o^*) \\ &+ \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \begin{bmatrix} I_p\otimes\hat{B}'\Sigma_{\epsilon o}^* \\ I_p\otimes\hat{B}'\Sigma_{\eta o}^* \end{bmatrix} T^{\frac{1}{2}}\mathrm{vec}(\hat{B}-B_o) \\ &+ \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \begin{bmatrix} I_p\otimes B_o'\Sigma_{\epsilon o}^* \\ I_p\otimes B_o'\Sigma_{\eta o}^* \end{bmatrix} T^{\frac{1}{2}}\mathrm{vec}(\hat{B}-B_o) \,. \end{aligned} \tag{16}$$

Since $\text{plim}\hat{B}=B_o$, making use of Slutsky's theorem, we have that the first term in (16) converges in distribution to $N(0,f_{22}^{-1})$, where f_{22} is given in (4.4c). Now, rewriting B_o as $B_o=JA_o$, $J=[I_p:0_{pxk}]$, from (2.4) and (4.3) it follows that

$$
\begin{aligned}
\text{vec}(\hat{B}-B_o) &= \text{vec}J(\hat{A}-A_o) \\
&= -(I_p \otimes J)S_A(\hat{\delta}-\delta_o) \\
&= -(I_p \otimes B_o)\bar{S}\,(\hat{\delta}-\delta_o). \qquad (17)
\end{aligned}
$$

Hence, making use of Slutsky's theorem it follows that the sum of the last two terms in (16) has limiting normal distribution with zero mean and covariance matrix being the second term in (4.5c). Thus, provided that $\hat{\psi}^*$ and $\hat{\delta}$ are asymptotically independent (15) follows.

The asymptotic independence between $\hat{\psi}^*$ and $\hat{\delta}$ can be justified as follows. We have seen in section 3.4 that $\tilde{\Pi}$ and $\tilde{\psi}^*$ are asymptotically independent. Now, because $\hat{\Pi}=\hat{\Pi}(\hat{\delta})$ takes into account the restrictions on B and Γ

$$
\text{Avar}(\tilde{\Pi},\tilde{\psi}^*) - \text{Avar}(\hat{\Pi},\hat{\psi}^*) \geq 0. \qquad (18)
$$

However, because $\tilde{\psi}^*$ and $\hat{\psi}^*$ are asymptotically equivalent, (18) holds only if $\text{Avar}(\hat{\Pi},\hat{\psi}^*)$ is diagonal. Thus $\hat{\Pi}$ and $\hat{\psi}^*$ are asymptotically independent and, since $\hat{\delta}$ is uniquely determined from $\hat{\Pi}$, $\hat{\delta}$ and $\hat{\psi}^*$ are asymptotically independent.

Appendix 6.1

Proof of Expression (6.3.2)

We shall show that

$$T^{-1}\sum_{j=0}^{T-1} I_j . \hat{F}_j^{-1} = I_p, \tag{1}$$

where I_j is the real part of the crossperiodogram matrix of U' (pxT) and $\hat{F}_j$ is the estimated spectral matrix. Thus,

$$\hat{F}_j = (2\pi)^{-1} \; (c_j\hat{\Sigma}_\varepsilon + \hat{\Sigma}_\eta), \tag{2}$$

$\hat{\Sigma}_\varepsilon = \Sigma_\varepsilon(\hat{\psi})$, $\hat{\Sigma}_\eta = \Sigma_\eta(\hat{\psi})$, and $\hat{\psi}$ is the solution of

$$\sum_{j=0}^{T-1} S_j' \; \mathrm{vec}[(F_j^{-1} - F_j^{-1}I_jF_j^{-1})] = 0 \tag{3}$$

where S_j is given in (2.7).

Replacing S_j, (3) becomes

$$(2\pi)^{-1}\sum_{j=0}^{T-1} \begin{bmatrix} D'c_j \\ D' \end{bmatrix} \mathrm{vec}[(F_j^{-1} - F_j^{-1}I_jF_j^{-1})] = 0. \tag{4}$$

Now since D has full column rank, (4) implies that

$$\sum_j [c_jF_j^{-1} - c_jF_j^{-1}I_jF_j^{-1}] = 0 \tag{5a}$$

and

$$\sum_j [\; F_j^{-1} - \; F_j^{-1}I_jF_j^{-1} \;] = 0 \tag{5b}$$

Solving (5) for ψ, it does not matter if it is numerically, we have

$$\sum_j (c_j \hat{F}_j^{-1} - c_j \hat{F}_j^{-1} I_j \hat{F}_j^{-1}) = 0 \tag{6a}$$

and

$$\sum_j (\hat{F}_j^{-1} - \hat{F}_j^{-1} I_j \hat{F}_j^{-1}) = 0 \tag{6b}$$

Now premultiplying (2) by $\hat{\Sigma}_\epsilon^{-1}$ yields

$$\hat{\Sigma}_\epsilon^{-1}\hat{F}_j = (2\pi)^{-1} (c_j I_p + \hat{Q})$$

where

$$\hat{Q} = \hat{\Sigma}_\epsilon^{-1}\hat{\Sigma}_\eta .$$

Thus,

$$(2\pi)\hat{\Sigma}_\epsilon^{-1} = (c_j I_p + \hat{Q})\hat{F}_j^{-1} . \tag{7}$$

Adding and subtracting $\hat{Q}\hat{F}_j^{-1}$ to the LHS and $\hat{Q}\hat{F}_j^{-1} I_j \hat{F}_j^{-1}$ to the RHS of (6.a) we have

$$\sum_j (c_j I_p + \hat{Q})\hat{F}_j^{-1} - \hat{Q} \sum_j \hat{F}_j^{-1} = \sum_j (c_j I_p + \hat{Q})\hat{F}_j^{-1} I_j \hat{F}_j^{-1} - \hat{Q} \sum_j \hat{F}_j^{-1} I_j \hat{F}_j^{-1} .$$

Using (7) we have

$$2\pi T\hat{\Sigma}_\epsilon^{-1} - \hat{Q} \sum_j \hat{F}_j^{-1} = 2\pi\hat{\Sigma}_\epsilon^{-1} \sum_j I_j \hat{F}_j^{-1} - \hat{Q} \sum_j \hat{F}_j^{-1} I_j \hat{F}_j^{-1}$$

and from (6.b) we have

$$2\pi T\hat{\Sigma}_\epsilon^{-1} - 2\pi\hat{\Sigma}_\epsilon^{-1} \sum_j I_j \hat{F}_j^{-1} = 0,$$

hence (1) follows.

Appendix 6.2

Asymptotic Covariance Matrix

We have

$$F^{-1} = \begin{bmatrix} W^{-1} & -W^{-1}f_{12}f_{22}^{-1} \\ -f_{22}^{-1}f_{21}W^{-1} & f_{22}^{-1}+f_{22}^{-1}f_{21}W^{-1}f_{12}f_{22}^{-1} \end{bmatrix}$$

where

$$W = f_{11} - f_{12}f_{22}^{-1}f_{21}$$

and f_{ij}, i,j=1,2 are given in (4.4).

We rewrite f_{12} and f_{22} as

$$f_{12} = \frac{1}{(2\pi)^2}\bar{S}' \left[\int c(F^{-1}\otimes I_p) \ : \int (F^{-1}\otimes I_p)\right]\begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix} \tag{1}$$

and as

$$f_{22} = \frac{1}{16\pi^3}\begin{bmatrix} D' & 0 \\ 0 & D' \end{bmatrix} \begin{bmatrix} \int c^2(F^{-1}\otimes F^{-1}) & \int c(F^{-1}\otimes F^{-1}) \\ \int c\ (F^{-1}\otimes F^{-1}) & \int (F^{-1}\otimes F^{-1}) \end{bmatrix}\begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix}, \tag{2}$$

where in (1) and (2) $c=c(\lambda)$ and $F=F(\lambda)$. We shall often omit the argument variable λ in situations where there will be no risk of confusion.

Now using the fact that the square matrices $\int c^2(F^{-1}\otimes F^{-1})$, $\int c(F^{-1}\otimes F^{-1})$ and $\int(F^{-1}\otimes F^{-1})$ commute with each other we have

$$f_{22}^{-1} = 16\pi^3\begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix}\begin{bmatrix} \Delta^{-1} & 0 \\ 0 & \Delta^{-1} \end{bmatrix}\begin{bmatrix} \int (F^{-1}\otimes F^{-1}) & -\int c(F^{-1}\otimes F^{-1}) \\ -\int c(F^{-1}\otimes F^{-1}) & \int c^2(F^{-1}\otimes F^{-1}) \end{bmatrix}\begin{bmatrix} D^{+\prime} & 0 \\ 0 & D^{+\prime} \end{bmatrix} \tag{3}$$

where D^+ is the Moore-Penrose inverse of D and Δ is given by

$$\Delta = \int c(\lambda)^2(F(\lambda)^{-1}\otimes F(\lambda)^{-1})d\lambda \int(F(\lambda)^{-1}\otimes F(\lambda)^{-1})d\lambda$$

$$- \int c(\lambda)(F(\lambda)^{-1}\otimes F(\lambda)^{-1})d\lambda\int c(\lambda)(F(\lambda)^{-1}\otimes F(\lambda)^{-1})d\lambda$$

$$=\iint(F(\lambda)^{-1}F(\nu)^{-1} \otimes F(\lambda)^{-1}F(\nu)^{-1})\ (c(\lambda)^2-c(\lambda)c(\nu))d\lambda d\nu\ . \tag{4}$$

Let $f_{12}f_{22}^{-1}$ be partitioned as

$$f_{12}f_{22}^{-1} = [B_1:B_2]\ .$$

From (1) and (3) and from the properties of D we obtain

$$B_1 = 2\pi\ \bar{S}'\Big[\int c(F^{-1}\otimes I_p)(I+K_p)\Delta^{-1}\int(F^{-1}\otimes F^{-1})$$

$$- \int(F^{-1}\otimes I_p)(I+K_p)\Delta^{-1} \int c(F^{-1}\otimes F^{-1})\ \Big]D^{+'}$$

which can be rewritten as

$$= 4\pi\ \bar{S}'\Delta^{-1}\Big[\int c(F^{-1}\otimes I_p) \int(F^{-1}\otimes F^{-1}) - \int(F^{-1}\otimes I_p) \int c(F^{-1}\otimes F^{-1})\Big]D^{+'}$$

$$=4\pi\bar{S}'\Delta^{-1}\Big[\iint(F(\lambda)^{-1}F(\nu)^{-1}\otimes F(\nu)^{-1}F(\lambda)^{-1})(I_p\otimes F(\lambda))(c(\lambda)-c(\nu))d\lambda d\nu\Big]D^{+'} \tag{5}$$

by noting that Δ^{-1} commutes with K_p and with $\int(F^{-1}\otimes I_p)$ and $K_pD^{+'}=D^{+'}$.

Now because

$$(I_p\otimes F(\lambda))\ (c(\lambda)-c(\nu)) = (2\pi)^{-1}[(I_p\otimes(c(\lambda)\Sigma_\varepsilon + \Sigma_\eta)](c(\lambda) - c(\nu))$$

$$= (2\pi)^{-1}\Big[(I_p \otimes \Sigma_\varepsilon)(c(\lambda)^2 - c(\lambda)c(\nu)) + (I_p \otimes \Sigma_\eta)(c(\lambda) - c(\nu))\Big] \tag{6}$$

we have

$$B_1 = 2\bar{S}'(I_p \otimes \Sigma_\varepsilon)D^{+\prime}.$$

Similary we can show

$$B_2 = 2\bar{S}'(I_p \otimes \Sigma_\eta)D^{+\prime},$$

and therefore

$$f_{12}f_{22}^{-1} = 2\bar{S}'\Big[(I_p \otimes \Sigma_\varepsilon) : (I_p \otimes \Sigma_\eta)\Big]\begin{bmatrix} D^{+\prime} & 0 \\ 0 & D^{+\prime} \end{bmatrix} . \tag{7}$$

Now

$$f_{12}f_{22}^{-1}f_{21} = \frac{2}{4\pi^2}\,\bar{S}'(I_p \otimes \Sigma_\varepsilon : I_p \otimes \Sigma_\eta)\begin{bmatrix} D^{+\prime} & 0 \\ 0 & D^{+\prime} \end{bmatrix}\begin{bmatrix} D' & 0 \\ 0 & D' \end{bmatrix}\begin{bmatrix} \int c(F^{-1} \otimes I_p) \\ \int (F^{-1} \otimes I_p) \end{bmatrix}\bar{S}$$

$$= \frac{1}{4\pi^2}\,\bar{S}'(I_p \otimes \Sigma_\varepsilon : I_p \otimes \Sigma_\eta))\begin{bmatrix} I+K_p & 0 \\ 0 & I+K_p \end{bmatrix}\begin{bmatrix} \int c(F^{-1} \otimes I_p) \\ \int (F^{-1} \otimes I_p) \end{bmatrix}\bar{S}$$

$$= \frac{1}{4\pi^2}\,\bar{S}'\Big[(I_p \otimes \Sigma_\varepsilon)\int c(F^{-1} \otimes I_p) + (I_p \otimes \Sigma_\eta)\int (F^{-1} \otimes I_p)\Big]\bar{S}$$

$$+ \frac{1}{4\pi^2}\,\bar{S}'\Big[(I_p \otimes \Sigma_\varepsilon)K_p\int c(F^{-1} \otimes I_p) + (I_p \otimes \Sigma_\eta)K_p\int (F^{-1} \otimes I_p)\Big]\bar{S}$$

$$= \frac{1}{4\pi^2}\bar{S}'\Big[\int F^{-1} \otimes (c\Sigma_\varepsilon + \Sigma_\eta)\Big]\bar{S} + \frac{1}{4\pi^2}\bar{S}'\Big[K_p\int [(c\Sigma_\varepsilon + \Sigma_\eta)F^{-1} \otimes I_p]\Big]\bar{S}$$

$$= \frac{1}{2\pi}\,\bar{S}'\int (F(\lambda)^{-1} \otimes F(\lambda))d\lambda\;\bar{S} + \bar{S}'K_p\bar{S} .$$

Hence

$$W = f_{11} - f_{12}f_{22}^{-1}f_{21} = \frac{1}{2\pi} S_A'\int(F(\lambda)^{-1}\otimes P'F_z(d\lambda)P)S_A$$

and

$$f^{11} = \left[\frac{1}{2\pi} S_A'\int(F(\lambda)^{-1}\otimes P'F_z(d\lambda)P)\ S_A\right]^{-1} \tag{8}$$

Now using (8) and (7) we have

$$f^{12} = -2\left[\frac{1}{2\pi} S_A'\int(F(\lambda)^{-1}\otimes P'F_z(d\lambda)P)S_A\right]^{-1}\bar{S}'\left[(I_p\otimes\Sigma_\varepsilon)D^{+\prime}:(I_p\otimes\Sigma_\eta)D^{+\prime}\right] \tag{9}$$

and using (9) and (7) and (3) we have

$$f^{22} = 16\pi^3\begin{bmatrix} D^+\Delta^{-1} & 0 \\ 0 & D^+\Delta^{-1}\end{bmatrix}\begin{bmatrix}\int(F^{-1}\otimes F^{-1}) & -\int c(F^{-1}\otimes F^{-1}) \\ -\int c(F^{-1}\otimes F^{-1}) & \int c^2(F^{-1}\otimes F^{-1})\end{bmatrix}\begin{bmatrix} D^{+\prime} & 0 \\ 0 & D^{+\prime}\end{bmatrix}$$

$$+ \quad 4\begin{bmatrix} D^+ & 0 \\ 0 & D^+\end{bmatrix}\begin{bmatrix} I_p\otimes\Sigma_\varepsilon \\ I_p\otimes\Sigma_\eta\end{bmatrix}\bar{S}\ W^{-1}\bar{S}'(I_p\otimes\Sigma_\varepsilon):(I_p\otimes\Sigma_\eta)\begin{bmatrix} D^{+\prime} & 0 \\ 0 & D^{+\prime}\end{bmatrix},$$

or, after some algebra,

$$f^{22} = 16\pi^3\int\begin{bmatrix} D^+\Delta^{-1} \\ D^+\Delta^{-1}c(\lambda)\end{bmatrix}\left[F(\lambda)^{-1}\otimes F(\lambda)^{-1}\right]\begin{bmatrix} D^{+\prime} \\ D^{+\prime}c(\lambda)\end{bmatrix}d\lambda$$

$$+ \quad 4\begin{bmatrix} D^+(I_p\otimes\Sigma_\varepsilon) \\ D^+(I_p\otimes\Sigma_\eta)\end{bmatrix}\bar{S}\ W^{-1}\bar{S}'\left[(I_p\otimes\Sigma_\varepsilon)D^{+\prime}:(I_p\otimes\Sigma_\eta)D^{+\prime}\right]. \tag{10}$$

Appendix 6.3

$\sqrt{T}$-consistency of Coefficient Estimators

We shall show that $\tilde{\delta}$ given in (5.1) is a $\sqrt{T}$-consistent estimator for δ, that is

$$T^{\frac{1}{2}}(\tilde{\delta}-\delta_o)=O_p(1) \tag{1}$$

Subtracting δ_o from both sides of (5.1) we have

$$(\tilde{\delta} - \delta_o) = H_V^{-1}\ S_A'\ \Big[\sum_{j=0}^{T-1}[\tilde{F}_V^{-1}(j)\otimes \tilde{P}']\mathrm{vec}I_{Z'\tilde{U}}(j)\ \Big] \tag{2}$$

where

$$H_V = S_A'\sum_{j=o}^{T-1}(\tilde{F}_V^{-1}(j)\ \otimes \tilde{P}'I_{Z'Z}(j)\tilde{P})S_A, \tag{3}$$

and

$$\mathrm{vec}\ I_{Z'\tilde{U}}(j) = \mathrm{vec}\ I_{Z'Y}(j) - (I_p\otimes I_{Z'Z}(j)\tilde{P})S_A\delta_o. \tag{4}$$

Now from (2.4), after some algebra, (4) becomes

$$\mathrm{vec}\ I_{Z'\tilde{U}}(j) = \mathrm{vec}\ I_{Z'V}(j) - [(I_p-B_o)'\otimes I_{Z'Z}(j)]\ \mathrm{vec}(\tilde{\Pi}-\Pi_o). \tag{5}$$

Now from (3.3.10) we have that

$$\mathrm{vec}(\tilde{\Pi}-\Pi_o) = A_V^{-1}\sum_{j=0}^{T-1}[\tilde{F}_V(j)^{-1}\otimes I_k]\ \mathrm{vec}I_{Z'V}(j), \tag{6}$$

where

$$A_V = \sum_{j=0}^{T-1}[\tilde{F}_V^{-1}(j)\otimes I_{Z'Z}(j)]\ .$$

Thus (5) becomes

$$\text{vec } I_{Z'\tilde{U}}(j) = \text{vec } I_{Z'V}(j) - [I_p \otimes I_{Z'Z}(j)]\ A_V^{-1} \sum_{j=0}^{T-1} [\tilde{F}_V^{-1}(j) \otimes I_k] \text{vec} I_{Z'V}(j)$$

$$+ [B_o' \otimes I_{Z'Z}(j)]\ A_V^{-1} \sum_{j=0}^{T-1} [\tilde{F}_V^{-1}(j) \otimes I_k]\ \text{vec} I_{Z'V}(j)\ . \qquad (7)$$

Hence, replacing (7) in (2) we have after some algebra

$$\tilde{\delta} - \delta_o = H_V^{-1} S_A'(I_p \otimes \tilde{P}')\ A_V(B_o' \otimes I_k)\ A_V^{-1} \sum_{j=0}^{T-1} [\tilde{F}_V^{-1}(j) \otimes I_k] \text{vec} I_{Z'V}(j). \quad (7)$$

Now from (3.4.8) and (3.4.9) we have

$$\text{plim } T^{-1} A_V = Q_V$$

and

$$T^{-\frac{1}{2}} \sum_{j=0}^{T-1} (\tilde{F}_V^{-1}(j) \otimes I_k\ ) \text{vec} I_{Z'V}(j) \ \xrightarrow{d}\ N(0, Q_V)$$

where

$$Q_V = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_V^{-1}(\lambda) \otimes F_Z(d\lambda).$$

Also because plim $\tilde{P} = P$,

$$\text{plim } T^{-1} H_V = \text{plim } T^{-1} S_A'(I_p \otimes \tilde{P}')\ A_V(I_p \otimes \tilde{P}') S_A$$

$$= S_A'(I_p \otimes P')\ Q_V(I_p \otimes P) S_A.$$

Hence making use of all of this, (1) follows from Slutsky's theorem.

CHAPTER 7

THE LIMITED INFORMATION MAXIMUM LIKELIHOOD ESTIMATOR OF A SINGLE EQUATION IN A SIMULTANEOUS EQUATION SYSTEM WITH STOCHASTIC TRENDS

1. Introduction

In this chapter we are going to consider the LIML (Limited Information Maximum Likelihood) estimator of the parameters in a single equation of a simultaneous equation system with stochastic trend components.

The LIML estimator was developed by Anderson and Rubin (1949) in order to estimate a single overidentified equation from a system of equations with uncorrelated normally distributed disturbances. It is obtained by considering only the portion of the system that relates the endogenous variables in the equation of interest. Because it is hard to grasp the theory underlying LIML, many different procedures have been derived, which are numerically or asymptotically equivalent, e.g. least variance ratio, instrumental variables. It is sometimes referred as the least generalized residual variance estimator. For a comprehensive study see, among others, Hausman (1983), and Hendry (1976).

A relatively easy way to understand LIML is by considering it as FIML applied to the new system formed by the equation of interest in its structural form and the reduced equations corresponding to the endogenous included in the equation of interest. As pointed out by Hall and Pagan (1981) this result can be found in a number of places in the literature. It has the interesting property that the new system is a triangular system and therefore, based on Lahiri and Schmidt's theorem (1978) concerning FIML estimation of triangular systems, LIML can be interpreted as an iterated version of the SURE (Seemingly Unrelated Regression Equations) estimator. Following these lines Hall and Pagan (1981) investigated the situation where the disturbances follow a multivariate MA(1) process.

Our task is to study the situation where the disturbances in the complete system follow a multivariate random walk. Proceeding in an analogous manner to Hall and Pagan (1981) we show that, as in the classical case, the LIML estimator of the parameters in the equation of interest can be obtained by applying FIML to the new system consisting of the first (structural) equation and the reduced form for the endogenous appearing in this equation. We derive a computational method for LIML in the time domain via the multivariate Kalman filter and consider the asymptotic properties in the frequency domain as a specialization of the results given in chapter 6. We finish by examining the efficiency of the estimators studied in chapter 4 relative to LIML.

2. The LIML Estimator

We shall consider the complete system as in chapter 5 , that is,

$$y_t'B + z_t'\Gamma = w_t', \tag{1}$$

where B is pxp ,Γ is kxp and w_t follows a multivariate random walk plus noise model with associated covariance matrices Σ_ϵ and Σ_η. Let

$$y_t' = [y_{1t}:Y_{1t}':Y_{2t}']$$
$$z_t' = [Z_{1t}':Z_{2t}']$$
$$w_t' = [w_{1t}:W_{1t}':W_{2t}']$$

where y_{1t} and w_{1t} are scalars, Y_{1t}' is $(1xp_1)$, Y_{2t}' is $1x(p\text{-}1\text{-}p_1)$, Z_{1t}' is $1xk_1$, Z_{2t}' is $1xk_2$ and

$$B = \begin{bmatrix} 1 & B_{12} & B_{13} \\ \beta & B_{22} & B_{23} \\ 0 & B_{32} & B_{33} \end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} \gamma & \Gamma_{12} & \Gamma_{13} \\ & & \\ 0 & \Gamma_{22} & \Gamma_{23} \end{bmatrix}$$

Within this notation, the first equation, the equation of interest, may be then written as

$$y_{1t} = - Y_{1t}'\beta - Z_{1t}'\gamma + \mu_{1t} + \epsilon_{1t} , \tag{2a}$$
$$\mu_{1t} = \mu_{1,t-1} + \eta_{1t} . \tag{2b}$$

Thus, Y_{1t}' and Z_{1t}' contain, respectively, the endogenous and exogenous variables included in the first equation, while Y_{2t}' and Z_{2t}' contain, respectively, the endogenous and exogenous variables excluded from this equation.

The reduced form of (1) is given by

$$[y_{1t}\; Y_{1t}'Y_{2t}'] = [Z_{1t}'Z_{2t}']\,\Pi + [v_{1t}\; V_{1t}'V_{2t}'] \tag{3}$$

where

$$\Pi = -\Gamma\, B^{-1} = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} \end{bmatrix}$$

and

$$\left[\, v_{1t}\; V_{1t}'V_{2t}' \right] = \left[w_{1t}\; W_{1t}'W_{2t}' \right] \begin{bmatrix} B^{11} & B^{12} & B^{13} \\ B^{21} & B^{22} & B^{23} \\ B^{31} & B^{32} & B^{33} \end{bmatrix} \tag{4}$$

where B^{ij} $i,j=1,3$, are the ij-th submatrices of the inverse of B. Now the LIML estimator of the parameters in (2) is obtained by maximising the loglikelihood function of the system

$$\left[y_{1t}\; Y_{1t}' \right] = \left[Z_{1t}'Z_{2t}' \right] \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} + \left[\, v_{1t}\; V_{1t}' \right] \tag{5}$$

where

$$\left[v_{1t}\; V_{1t}' \right] = \left[\, w_{1t}\; W_{1t}'W_{2t}' \right] \begin{bmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \\ B^{31} & B^{32} \end{bmatrix} . \tag{6}$$

The loglikelihood function of (5) is as in (5.4.1). However the maximisation is subject to constraints guaranteeing that only the exogenous variables contained in Z_{1t} are included in the first equation. These constraints are

$$\Pi_{21} + \Pi_{22}\beta = 0 \tag{7a}$$

and

$$\Pi_{11} + \Pi_{12}\beta = - \gamma . \tag{7b}$$

As pointed out in Hall and Pagan (1981), (7b) is not a restriction as such, but it will hold, and we can deduce γ from the reduced form parameters.

3. LIML Viewed as a Special Case of FIML

We are now going to proceed as in Hall and Pagan (1981) in order to show that the ML estimator of (2.5) subject to the restrictions given in (2.7) is numerically equivalent to the FIML estimator of the new incomplete system consisting of the first (structural) equation, given in (2.2), and the reduced form for the endogenous variables appearing in this equation. We remark that recognizing LIML as a special case of FIML, will allow us to obtain the ML estimates from the structural form, which, from the computational point of view is easier to handle then the reduced form subject to restrictions.

The new system may be written as

$$\begin{bmatrix} y_{1t} & Y_{1t}' \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta & I_{p1} \end{bmatrix} + \begin{bmatrix} Z_{1t}' Z_{2t}' \end{bmatrix} \begin{bmatrix} \gamma & -\Pi_{12} \\ 0 & -\Pi_{22} \end{bmatrix} = \begin{bmatrix} w_{1t} & V_{1t}' \end{bmatrix} \tag{1}$$

where

$$\begin{bmatrix} w_{1t} & V_{1t}' \end{bmatrix} = \begin{bmatrix} w_{1t} & W_{1t}' W_{2t}' \end{bmatrix} \begin{bmatrix} 1 & B^{12} \\ 0 & B^{22} \\ 0 & B^{32} \end{bmatrix} \tag{2}$$

The reduced form of (1) is given by

$$[y_{1t}\ Y_{1t}'] = [Z_{1t}' Z_{2t}']\,\Pi^{\dagger} + [v_{1t}^{*}\ V_{1t}'^{*}] \tag{3}$$

where

$$\Pi^{\dagger} = -\begin{bmatrix} \gamma & -\Pi_{12} \\ 0 & -\Pi_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta & I_{p1} \end{bmatrix}^{-1} = \begin{bmatrix} \Pi_{11}{}^{\dagger} & \Pi_{12} \\ \Pi_{21}{}^{\dagger} & \Pi_{22} \end{bmatrix} \tag{4}$$

and

$$\left[v_{1t}^{*}\ V_{1t}^{*\prime}\right] = \left[w_{1t}\ V_{1t}'\right] \begin{bmatrix} 1 & 0 \\ \beta & I_{p1} \end{bmatrix}^{-1}. \tag{5}$$

Now the restrictions on $\Pi^{\dagger}$ are the same as the one given in (2.7). Moreover, noting that

$$\begin{bmatrix} 1 & B^{12} \\ 0 & B^{22} \\ 0 & B^{32} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta & I_{p1} \end{bmatrix}^{-1} = \begin{bmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \\ B^{31} & B^{32} \end{bmatrix},$$

we have that (5) is identical to (2.6): hence the equivalence of the two procedures.

Finally we note that if we assume that the disturbances of the complete system follow a multivariate random walk plus noise model, with associated covariance matrices Σ_{ε} and Σ_{η}, then the disturbances $(w_{1t}, V_{1t}')'$ of the new incomplete system given in (1) will also follow a multivariate random walk plus noise but with associated covariance matrices

$$\Sigma_\varepsilon{}^\dagger = \begin{bmatrix} 1 & 0 & 0 \\ B^{12\prime} & B^{22\prime} & B^{32\prime} \end{bmatrix} \Sigma_\varepsilon \begin{bmatrix} 1 & B^{12} \\ 0 & B^{22} \\ 0 & B^{32} \end{bmatrix} \tag{6a}$$

$$\Sigma_\eta{}^\dagger = \begin{bmatrix} 1 & 0 & 0 \\ B^{12\prime} & B^{22\prime} & B^{32\prime} \end{bmatrix} \Sigma_\eta \begin{bmatrix} 1 & B^{12} \\ 0 & B^{22} \\ 0 & B^{32} \end{bmatrix} \tag{6b}$$

This fact has the relevant property of allowing us to make use of the estimation techniques and asymptotic properties for FIML with random walk plus noise disturbances as discussed in the previous chapter.

For models with disturbances following an unrestricted multivariate MA(1) process, if we proceed as above and consider only the incomplete system with reduced form equations for the endogenous variables in the first equation, the MA structure of the disturbances will not be maintained. In other words if w_t' in (2.1) is given by

$$w_t' = \varepsilon_t' + \varepsilon_{t-1}' \Theta,$$

then the disturbances in (1) will take the form

$$\begin{bmatrix} w_{1t} & V_{1t}' \end{bmatrix} = \varepsilon_t' \begin{bmatrix} 1 & B^{21} \\ 0 & B^{22} \\ 0 & B^{32} \end{bmatrix} + \varepsilon_{t-1}' \Theta \begin{bmatrix} 1 & B^{21} \\ 0 & B^{22} \\ 0 & B^{32} \end{bmatrix},$$

and clearly will not follow a vector MA(1), since the matrix in square brackets and Θ do not commute. Hence in this case, in order to be able to use the literature, such as Reinsel (1979), concerning FIML estimation of simultaneous equation systems with vector ARMA disturbances, Hall and Pagan (1981) considered LIML as FIML applied to the system consisting of the (structural) equation of interest and the reduced form for *all* endogenous variables in the system apart from y_{1t}. Moreover constraints on the MA coefficient matrix must be imposed. These constraints in turn restrict us to models in which all disturbances follow a vector MA, apart from the one in the first equation, which follows a univariate MA process. Such models do not seem to be very natural. Structural models in which both ϵ_{1t} and η_{1t} may be correlated with the corresponding disturbances in the other structural equations are more appealing.

4. Computational Method

We are now going to derive the FIML estimates of the system given in (3.1), which may be written as

$$\left[y_{1t}\ Y_{1t}'\right] B^{\dagger} + \left[Z_{1t}'Z_{2t}'\right] \Gamma^{\dagger} = w_t^{\dagger\prime} , \qquad (1)$$

where the $\bar{p}x\bar{p}$ matrix $B^{\dagger}$ and the $Kx\bar{p}$ matrix $\Gamma^{\dagger}$, where $\bar{p}=p_1+1$, $K=k_1+k_2$, are

$$B^{\dagger} = \begin{bmatrix} 1 & 0 \\ \beta & I_{p1} \end{bmatrix} \quad \text{and} \quad \Gamma^{\dagger} = \begin{bmatrix} \gamma & -\Pi_{12} \\ 0 & -\Pi_{22} \end{bmatrix} .$$

As we pointed out before the $\bar{p}x1$ vector $w_t^{\dagger} = [w_{1t}\ V_{1t}']'$ follows a $\bar{p}$-variate random walk plus noise model with associated covariance matrices $\Sigma_\varepsilon^{\dagger}$ and $\Sigma_\eta^{\dagger}$ given in (3.6a) and (3.6b).

Let α denote the unrestricted elements of -vec $[B^{\dagger\prime}\Gamma^{\dagger\prime}]'$. Thus

$$\alpha = \begin{bmatrix} -\ \beta \\ -\ \gamma \\ \mathrm{vec}\begin{bmatrix}\Pi_{12} \\ \Pi_{22}\end{bmatrix} \end{bmatrix} .$$

Conditional on the first observations the loglikelihood function of (1) is given in (5.4.4), that is,

$$\ell(\alpha,\psi^{\dagger}) = (T-1)\log|B^{\dagger}| - \tfrac{1}{2}\sum_{t=2}^{T}\log|F_t| - \tfrac{1}{2}\sum_{t=2}^{T}\nu_t{}'F_t{}^{-1}\nu_t , \qquad (2)$$

where $\psi^{\dagger}$ is the $\bar{p}(\bar{p}+1)x1$ vector obtained from $\mathrm{vec}(\Sigma_{\epsilon}{}^{\dagger}:\Sigma_{\eta}{}^{\dagger})$ by eliminating the supradiagonal elements of $\Sigma_{\epsilon}{}^{\dagger}$ and $\Sigma_{\eta}{}^{\dagger}$. Now because $B^{\dagger}$ is a triangular matrix, $|B^{\dagger}|=1$, hence $\log|B^{\dagger}|$ is absent in the loglikelihood function. Moreover from the discussion given below (5.4.4) we have that

$$\nu_t = \tilde{y}_t - \tilde{X}_t{}'\alpha \qquad (3)$$

where $\tilde{y}_t$ is obtained by applying the multivariate Kalman filter to the $\bar{p}$-dimensional process, $y_t = [y_{1t}\ Y_{1t}{}']'$ and $\tilde{X}_t{}'$ is obtained by applying the multivariate Kalman filter separately to each column of $\bar{X}_t{}'$, where

$$\bar{X}_t{}' = \begin{bmatrix} Y_{1t}{}'\ Z_{1t}{}' & 0 \\ 0 & (I_{p1}\otimes z_t{}') \end{bmatrix}. \qquad (4)$$

Replacing (3) in (2) yields

$$\ell(\alpha,\psi^{\dagger}) = -\tfrac{1}{2}\sum_{t=2}^{T}\log|F_t| - \tfrac{1}{2}\sum_{t=2}^{T}(\tilde{y}_t - \tilde{X}_t{}'\alpha)'F_t{}^{-1}(\tilde{y}_t - \tilde{X}_t{}'\alpha) \qquad (5)$$

which has to be maximised with respect to α and $\psi^\dagger$.

As in the univariate case the maximisation of (5) can be carried out in two different ways. Firstly α can be concentrated out of the loglikelihood function. Thus, solving the likelihood equations for α we have

$$\hat{\alpha} = \hat{\alpha}\ (\psi^\dagger) = \left[\sum_{t=2}^{T} \tilde{X}_t F_t^{-1} \tilde{X}_t'\right]^{-1} \sum_{t=2}^{T} \tilde{X}_t F_t^{-1} \tilde{y}_t \quad . \tag{6}$$

Replacing (6) in (5) yields the concentrated loglikelihood function

$$\ell_c(\psi^\dagger) = -\tfrac{1}{2} \sum_{t=2}^{T} \log|F_t| - \tfrac{1}{2} \sum_{t=2}^{T} (\tilde{y}_t - \tilde{X}_t'\hat{\alpha})' F_t^{-1} (\tilde{y}_t - \tilde{X}_t'\hat{\alpha}) \tag{7}$$

and the ML estimate of $\psi^\dagger$ is obtained by maximising (7) nonlinearly with respect to $\psi^\dagger$. Once we have found $\hat{\psi}^\dagger$, $\hat{\alpha}=\hat{\alpha}(\hat{\psi}^\dagger)$ is obtained from (6).

Alternatively we could maximise (5) in a stepwise fashion. A consistent estimator of α is constructed by differencing all the variables, and then applying 2SLS to the first equation and least squares to the second set of equations given in (1). The residuals, $\hat{w}_t^\dagger = y_t - \bar{X}_t'\hat{\alpha}$, are computed and the matrices $\hat{\Sigma}_\epsilon^\dagger$ and $\hat{\Sigma}_\eta^\dagger$ are estimated maximising the loglikelihood of $\hat{w}_t^\dagger$. α is reestimated using (6). As noted in Lahiri and Schmidt (1978) the two-step estimator of α is asymptotically inefficient. However, if the procedure is repeated until convergence, apart from computational restrictions, the same estimator is obtained.

5. Asymptotic Theory of LIML

Although the fact that the determinant of $B^{\dagger}$ is equal to unity helps regarding the computation of LIML it does not help regarding the asymtotic properties. Thus, in order to obtain the asymptotic properties of LIML we have to proceed in the same way as we would in obtaining the asymptotic properties of FIML in the general case where $B^{\dagger}$ is not a triangular matrix. This has nothing to do with the fact that the disturbances are serially correlated, since even in the classical case we have to proceed in this way. Now the asymptotic properties of FIML were studied in the frequency domain in the previous chapter. From section (6.5) we have that $T^{\frac{1}{2}}(\hat{\alpha}-\alpha_o)$ has a normal limiting distribution with zero mean and covariance matrix given in (6.4.5a). Hence the asymptotic covariance matrix of $\hat{\alpha}$ is

$$\operatorname{Avar}(\hat{\alpha}) = \frac{1}{T}\left[S_A'(I_{\bar{p}}\otimes P^{\dagger\prime})\ \frac{1}{2\pi}\int_{-\pi}^{\pi}[F(\lambda)^{-1}\otimes F_z(\lambda)]d\lambda(I_{\bar{p}}\otimes P^{\dagger})S_A\right]^{-1} \qquad (1)$$

where

$$F(\lambda) = F(\lambda,\psi^{\dagger}) = (2\pi)^{-1}[2(1-\cos\lambda)\Sigma_{\varepsilon}{}^{\dagger} + \Sigma_{\eta}{}^{\dagger}],$$

with $\Sigma_{\varepsilon}{}^{\dagger}$ and $\Sigma_{\eta}{}^{\dagger}$ as given in (3.6), $\Sigma_{\eta}{}^{\dagger}$ positive definite; $F_z(\lambda)$ is the spectral matrix of the process generating the (differenced) exogenous variables;

$$P^{\dagger} = [\Pi^{\dagger}\ I_K], \qquad \Pi^{\dagger} = \begin{bmatrix}\Pi_{11}{}^{\dagger} & \Pi_{12}\\ \Pi_{21}{}^{\dagger} & \Pi_{22}\end{bmatrix} \qquad \text{as given in (3.4),}$$

where $\Pi_{11}{}^{\dagger}$ is $k_1 x1$, Π_{12} is $k_1 x p_1$, $\Pi_{21}{}^{\dagger}$ is $k_2 x1$ and Π_{22} is $k_2 x p_1$; S_A is the selection matrix as given in (5.2.7). For LIML S_A becomes

$$S_A = \left[\begin{array}{c:c} 0_{1x(p1+k1)} & \\ \hdashline I_{p1+k1} & 0 \\ \hdashline 0_{k2x(p1+k1)} & \\ \hdashline 0 & I_{p1} \otimes \left[\begin{array}{c} 0_{(p1+1)x(k1+k2)} \\ \hdashline I_{k1+k2} \end{array}\right] \end{array}\right] \tag{2}$$

and so

$$(I_{\bar{p}} \otimes P^{\dagger}) S_A = \left[\begin{array}{c:c} R & 0 \\ \hdashline 0 & I_{p1} \otimes I_k \end{array}\right] \tag{3}$$

where

$$R = \begin{bmatrix} \Pi_{12} & I_{k1} \\ \Pi_{22} & 0 \end{bmatrix} . \tag{4}$$

Therefore (1) becomes

$$\mathrm{Avar}\hat{\alpha} = \frac{2\pi}{T}\left[\left[\begin{array}{c:c} R' & 0 \\ \hdashline 0 & (I_{p1} \otimes I_k) \end{array}\right] \int_{-\pi}^{\pi} (F(\lambda)^{-1} \otimes F_z(\lambda)) d\lambda \left[\begin{array}{c:c} R & 0 \\ \hdashline 0 & (I_{p1} \otimes I_k) \end{array}\right] \right]^{-1} \tag{5}$$

Because the parameters in (5) are unknown, the asymptotic covariance matrix of $\hat{\alpha}$ can be consistently estimated by

$$\mathrm{avar}\hat{\alpha} = \left[\left[\begin{array}{c:c} \hat{R}' & 0 \\ \hdashline 0 & (I_{p1} \otimes I_k) \end{array}\right] \sum_{j=0}^{T} (\hat{F}_j^{-1} \otimes I_z(j)) \left[\begin{array}{c:c} \hat{R} & 0 \\ \hdashline 0 & (I_{p1} \otimes I_k) \end{array}\right] \right]^{-1} \tag{6}$$

where $\hat{R}$ is (4) with Π_{12} and Π_{22} replaced by their ML estimators; $\hat{F}_j = F(\lambda_j, \hat{\vartheta}^{\dagger})$, where $\hat{\vartheta}^{\dagger}$ is the $\bar{p}(\bar{p}+1)x1$ vector containing the ML estimators of the unrestricted elements in $\mathrm{vec}(\Sigma_\epsilon^\dagger\ \Sigma_\eta^\dagger)$; $I_z(j)$ is the periodogram matrix of the differenced exogenous variables.

Our main interest is centered on the asymptotic covariance matrix of $\hat{\delta}$, where $\hat{\delta}'=(\hat{\beta}',\hat{\gamma}')$ contains the estimators of the regression coefficients appearing in the first equation. The relevant expression is given by the left top $(p_1+k_1 \times p_1+k_1)$ submatrix of (5). Writing (5) as

$$\mathrm{Avar}\hat{\alpha} = \frac{2\pi}{T}\begin{bmatrix} R'\int_{-\pi}^{\pi} F(\lambda)^{11}\otimes F_z(\lambda)d\lambda\, R & R'\int_{-\pi}^{\pi} F(\lambda)^{12}\otimes F_z(\lambda)d\lambda \\ \int_{-\pi}^{\pi} F(\lambda)^{21}\otimes F_z(\lambda)\, d\lambda\, R & \int_{-\pi}^{\pi} F(\lambda)^{22}\otimes F_z(\lambda)d\lambda \end{bmatrix}^{-1} \tag{7}$$

where $F(\lambda)^{ij}$,i,j=1,2 are the (i,j)-th submatrices of the inverse of $F(\lambda)$, and using the partitioned form of an inverse, it turns out that the asymptotic covariance matrix of $\hat{\delta}$ is given by

$$\mathrm{Avar}\hat{\delta} =$$

$$\frac{2\pi}{T}\left\{R'\left[\int_{-\pi}^{\pi} F(\lambda)^{11}\otimes F_z(\lambda) - \int_{-\pi}^{\pi} F(\lambda)^{12}\otimes F_z(\lambda)\left[\int_{-\pi}^{\pi} F(\lambda)^{22}\otimes F_z(\lambda)\right]^{-1}\int_{-\pi}^{\pi} F(\lambda)^{21}\otimes F_z(\lambda)\right]R\right\}^{-1} \tag{8}$$

which can be written as

$$\mathrm{Avar}\hat{\delta} = \frac{2\pi}{T}\left\{R'\left[\ \left[\int_{-\pi}^{\pi} F(\lambda)^{-1}\otimes F_z(\lambda)d\lambda\right]^{11}\ \right]^{-1} R\right\}^{-1}, \tag{9}$$

where here $[\cdot]^{11}$ denotes the kxk top left submatrix of the inverse of the $\bar{p}k\mathrm{x}\bar{p}k$ matrix in squared brackets. As in (6), (9) can be consistently estimated as

$$\text{avar}\hat{\delta} = \left\{\hat{R}'\left[\sum_{j=0}^{T}\hat{F}_j^{-1}\otimes I_z(j)\right]^{11}\right]^{-1}\hat{R}\Big\}^{-1} . \tag{10}$$

Explanatory Variables Following a Multivariate Random Walk

If we assume that the explanatory variables follow a multivariate random walk with disturbance covariance matrix Σ_z, then the differenced variables will have constant spectrum, that is,

$$F_z(\lambda) = (2\pi)^{-1}\,\Sigma_z\,, \qquad -\pi \leq \lambda \leq \pi \tag{11}$$

and expression (9) becomes

$$\text{Avar}\hat{\delta} = \frac{2\pi}{T}\left\{R'\left[\left[\int_{-\pi}^{\pi}F(\lambda)^{-1}d\lambda \otimes (2\pi)^{-1}\Sigma_z\right]^{11}\right]^{-1} R\right\}^{-1}. \tag{12}$$

Now

$$\left[\int_{-\pi}^{\pi}F(\lambda)^{-1}d\lambda \otimes (2\pi)^{-1}\Sigma_z\right]^{-1} = 2\pi\left[\int_{-\pi}^{\pi}F(\lambda)^{-1}d\lambda\right]^{-1}\otimes \Sigma_z^{-1}\,,$$

hence

$$\left[\int_{-\pi}^{\pi}F(\lambda)^{-1}d\lambda \otimes (2\pi)^{-1}\Sigma_z\right]^{11} = 2\pi\left[\int_{-\pi}^{\pi}F(\lambda)^{-1}d\lambda\right]^{11} \Sigma_z^{-1}\,. \tag{13}$$

Making use of (13), (12) becomes after some algebra

$$\text{Avar}\hat{\delta} = \frac{4\pi^2}{T}\left[\int_{-\pi}^{\pi}F(\lambda)^{-1}d\lambda\right]^{11}\left[R'\Sigma_z R\right]^{-1}, \tag{14}$$

where $F(\lambda)=F(\lambda,\psi^{\dagger})$.

This being the case, the integral in (14) can be computed in the following way. For disturbances following a multivariate random walk plus noise process, we have

$$F(\lambda,\psi^{\dagger}) = (2\pi)^{-1}\,[c(\lambda)\Sigma_{\varepsilon}^{\dagger} + \Sigma_{\eta}^{\dagger}\ ], \tag{15a}$$

where

$$c(\lambda) = 2(1 - \cos\lambda). \tag{15b}$$

After some algebra the inverse of (15a) becomes

$$[F(\lambda,\psi^{\dagger})]^{-1} = 2\pi\,[I_{\bar{p}} - B\cos\lambda]^{-1}[2\Sigma_{\varepsilon}^{\dagger} + \Sigma_{\eta}^{\dagger}]^{-1}$$

where

$$B = [2\Sigma_{\varepsilon}^{\dagger} + \Sigma_{\eta}^{\dagger}]^{-1}\,2\Sigma_{\varepsilon}^{\dagger}.$$

We note that if $\Sigma_{\eta}^{\dagger}$ is positive definite then the eigenvalues of B are less than one, see Magnus and Neudecker (1988,page 25), and because the norm of a positive semidefinite matrix is its largest eigenvalue, it follows that $\|B\|<1$, and so

$$(I_{\bar{p}} - B\cos\lambda)^{-1} = \sum_{k=0}^{\infty} B^{k}\cos^{k}\lambda.$$

Therefore

$$\int_{-\pi}^{\pi} F(\lambda,\psi^{\dagger})^{-1}d\lambda = 2\pi\sum_{k=0}^{\infty} B^{k}\int_{-\pi}^{\pi}\cos^{k}\lambda d\lambda\ [2\Sigma_{\varepsilon}^{\dagger} + \Sigma_{\eta}^{\dagger}]^{-1}.$$

Now

$$\int_{-\pi}^{\pi}\cos^{k}\lambda d\lambda = 0 \qquad \text{for k odd,}$$

$$\int_{-\pi}^{\pi}\cos^{2}\lambda d\lambda = \pi$$

and using the fact that for $j>2$

$$\int_{-\pi}^{\pi}\cos^{2j-2}\lambda\sin^{2}\lambda d\lambda = \frac{1}{2j}\int_{-\pi}^{\pi}\cos^{2j-2}\lambda d\lambda$$

we have

$$\int_{-\pi}^{\pi}\cos^{2j}\lambda d\lambda = \int_{-\pi}^{\pi}\cos^{2j-2}\lambda d\lambda - \int_{-\pi}^{\pi}\cos^{2j-2}\lambda\sin^{2}\lambda d\lambda$$

$$= \frac{2j-1}{2j}\int_{-\pi}^{\pi}\cos^{2j-2}\lambda d\lambda.$$

Thus,

$$\int_{-\pi}^{\pi} F(\lambda,\psi^{\dagger})^{-1}d\lambda = 4\pi^{2}\sum_{j=0}^{\infty}c_{j}B^{2j}[2\Sigma_{\varepsilon}{}^{\dagger} + \Sigma_{\eta}{}^{\dagger}]^{-1} \tag{16}$$

where

$$c_j = \frac{2j-1}{2j}c_{j-1} \qquad j=1,..$$

$$c_0 = 1,$$

and the integral can be evaluated numerically with a desired precision.

6. Efficiency Comparisons

We shall now examine the efficiency of the estimators studied in chapter 4 relative to LIML. The criterion used here is the the ratio of the determinants of the asymptotic covariance matrices.

The asymptotic covariance matrix of LIML is given in (5.9), namely

$$\text{Avar}\,\hat{\delta}_{LIML} = \frac{2\pi}{T}\left\{R'\left[\left[\int_{-\pi}^{\pi} F(\lambda)^{-1}\otimes F_z(\lambda)d\lambda\right]^{11}\right]^{-1} R\right\}^{-1} \tag{1}$$

while the asymptotic covariance matrix of the $\hat{\delta}_{IIV/ML}$ estimator studied in chapter 4 is T^{-1} times the inverse of (4.4.8), namely

$$\text{Avar}\;\hat{\delta}_{IIV/ML} = \frac{2\pi}{T}\left[R'\int_{-\pi}^{\pi} f(\lambda)^{-1}F_z(\lambda)d\lambda\;\;R\right]^{-1}, \tag{2}$$

with $f(\lambda) = [F(\lambda,\psi^{\dagger})]_{11} = [F(\lambda,\psi)]_{11}$, where the last equality follows from (3.6), and $[\cdot]_{11}$ is the top left element of the matrix in square brackets.

Thus,

$$\frac{|\text{Avar}\;\hat{\delta}_{LIML}|}{|\text{Avar}\;\hat{\delta}_{IIV/ML}|} = \frac{\left| R'\int_{-\pi}^{\pi} f(\lambda)^{-1}F_z(\lambda)d\lambda\;\;R \right|}{\left| R'\left[\left[\int_{-\pi}^{\pi}(F(\lambda)^{-1}\otimes F_z(\lambda))d\lambda\right]^{11}\right]^{-1}R\right|} \tag{3}$$

$$= \lim_{T\to\infty} \frac{\left| R'\sum_{j=0}^{T} f(\lambda_j)^{-1}F_z(\lambda_j)\;R\right|}{\left| R'\left[\left[\sum_{j=0}^{T}F(\lambda_j)^{-1}\otimes F_z(\lambda_j)\right]^{11}\right]^{-1}R\right|}. \tag{4}$$

Now let

$$B = \left[\sum_{j=0}^{T}F(\lambda_j)^{-1}\otimes F_z(\lambda_j)\right]^{11} \quad \text{and} \quad C = \left[\sum_{j=0}^{T}f(\lambda_j)^{-1}F_z(\lambda_j)\right]^{-1}.$$

Because B and C are positive definite and C-B is positive semidefinite, see appendix 7.1, we have that B^{-1}-C^{-1} is positive semidefinite. Therefore $R'B^{-1}R - R'C^{-1}R$ is positive semidefinite and so

$$|R'B^{-1}R| \geq |R'C^{-1}R|.$$

Hence the LIML estimator of δ is at least as efficient as the IIV/ML estimator of δ.

Cases Where IIV/ML is Efficient

There are two cases where IIV/ML is efficient in the sense that it has the same asymptotic distribution as LIML.

a) Homogeneity

In the homogeneous case, that is, $\Sigma_\eta = q\Sigma_\varepsilon$ and of course $\Sigma_\eta{}^\dagger = q\Sigma_\varepsilon{}^\dagger$, we have

$$F(\lambda,\psi^\dagger) = (2\pi)^{-1}[\ c(\lambda)\Sigma_\varepsilon{}^\dagger + \Sigma_\eta{}^\dagger\ ] = (2\pi)^{-1}\Sigma_\varepsilon{}^\dagger\bar{c}(\lambda),$$

where $\bar{c}(\lambda) = c(\lambda)+q$. Therefore

$$\begin{aligned}
\left[\int_{-\pi}^{\pi} F(\lambda,\psi^\dagger)^{-1}\otimes F_z(\lambda)d\lambda\right]^{11} &= \left[\int_{-\pi}^{\pi} 2\pi(\Sigma_\varepsilon{}^\dagger)^{-1}\bar{c}(\lambda)^{-1}\otimes F_z(\lambda)d\lambda\right]^{11} \\
&= \left[2\pi(\Sigma_\varepsilon{}^\dagger)^{-1}\otimes\int_{-\pi}^{\pi}\bar{c}(\lambda)^{-1}F_z(\lambda)d\lambda\right]^{11} \\
&= (2\pi)^{-1}\left[\Sigma_\varepsilon(1,1)^{-1}\int_{-\pi}^{\pi}\bar{c}(\lambda)^{-1}F_z(\lambda)d\lambda\right]^{-1} \\
&= \left[\ \int_{-\pi}^{\pi} f(\lambda)^{-1}F_z(\lambda)d\lambda\right]^{-1}, \qquad (5)
\end{aligned}$$

where the last equality follows because

$$2\pi \left[\Sigma_\varepsilon(1,1)\ \bar{c}(\lambda)\right]^{-1} = 2\pi \left[\Sigma_\varepsilon(1,1)c(\lambda) + \Sigma_\eta(1,1)\right]^{-1} = f(\lambda)^{-1}.$$

Hence, making use of (5) it follows that (1), the asymptotic covariance matrix of $\hat{\delta}_{LIML}$, coincides with the asymptotic covariance of $\hat{\delta}_{IIV/ML}$ given in (2).

b) Disturbances in First Equation Uncorrelated with Those in Other Equations

Suppose $p_1 = p-1$, that is the first equation contains all the endogenous variables in the system. This being the case

$$F(\lambda,\psi^\dagger) = (2\pi)^{-1}\left[\Sigma_\varepsilon{}^\dagger c(\lambda) + \Sigma_\eta{}^\dagger\right] = \bar{B}'F(\lambda,\psi)\bar{B},$$

where

$$\bar{B} = \begin{bmatrix} 1 & B^{12} \\ 0 & B^{22} \end{bmatrix} \quad \text{is square and positive definite.}$$

Therefore,

$$\left[\int_{-\pi}^{\pi} F(\lambda,\psi^\dagger)^{-1} \otimes F_z(\lambda)d\lambda\right]^{-1} = (\bar{B}'\otimes I_k)\left[\int_{-\pi}^{\pi} F(\lambda,\psi)^{-1} \otimes F_z(\lambda)d\lambda\right]^{-1}(\bar{B}\otimes I_k). \quad (6)$$

The top left hand kxk submatrix of the matrix on the left hand side of (6) appears in (1), but on evaluating the right hand side we find that $F(\lambda,\psi^\dagger)$ can be replaced by $F(\lambda,\psi)$. Now if the disturbances in the first equation, ε_{1t} and η_{1t}, are uncorrelated with the disturbances in the other structural equations, then $F(\lambda,\psi)$ is block diagonal and so the top left submatrix of the matrix on the left hand side of (6) becomes

$$\Big[\int_{-\pi}^{\pi} F(\lambda,\psi^{\dagger})^{-1} \otimes F_z(\lambda)d\lambda\Big]^{11} = \Big[\int_{-\pi}^{\pi} f(\lambda)^{-1} F_z(\lambda)d\lambda\Big]^{-1}, \tag{7}$$

and again (1) reduces to (2).

7. Stochastic Trend in First Equation Only

Up to this point we have been assuming that $\Sigma_{\eta}{}^{\dagger}$ given in (3.6b) is positive definite. If $p_1=p-1$ this is the same as requiring that stochastic trends be present in all the structural equations in the system. There is no reason why this should be true in general. If the assumption is not true, LIML can still be calculated in the time domain since the exact likelihood is still produced by the Kalman filter. The question concerns its asymptotic distribution, since differencing the observations in the system (4.1) will yield a strictly noninvertible model and so the conditions for the asymptotic distribution theory used to obtain (5.9) will no longer hold. Note that the distribution of IIV/ML is unaffected and so (6.2) remains true irrespective of whether or not equations other than the first contain stochastic trends.

We shall investigate the properties of LIML when only the first structural equation contains a stochastic trend. It will be also assumed that (i) $p_1=p-1$ and (ii) the exogenous variables are random walks. These two assmptions mean that if stochastic trends were, in fact, present in all the structural equations, the asymptotic covariance matrix of the LIML estimator of δ would be given by (5.14). Because of (6.6), $F(\lambda,\psi^{\dagger})$ can be replaced by $F(\lambda,\psi)$. Thus

$$\text{Avar}\hat{\delta} = \frac{4\pi^2}{T} \left[\int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \right]^{11} \left[R'\Sigma_z R \right]^{-1}, \tag{1}$$

where $F(\lambda)=F(\lambda,\psi)$, and because $F(\lambda)=F(-\lambda)$, (1) can be written as

$$\text{Avar}\hat{\delta} = \frac{2\pi}{T} \frac{1}{2} \left[\int_{0}^{\pi} [2\pi F(\lambda)]^{-1} d\lambda \right]^{11} \left[R'\Sigma_z R \right]^{-1}. \tag{2}$$

In the case we are interested in $F(\lambda)$ is not positive definite at $\lambda=0$. Based on heuristic arguments we take

$$\text{Avar}\hat{\delta} = \lim_{x\to 0^+} \frac{2\pi}{T} \frac{1}{2} \left[\int_{x}^{\pi} [2\pi F(\lambda)]^{-1} d\lambda \right]^{11} \left[R'\Sigma_z R \right]^{-1}. \tag{3}$$

as an expression for the asymptotic covariance matrix of $\hat{\delta}_{LIML}$. Evaluating the limit, see appendix 7.2, (3) becomes

$$\text{Avar}\hat{\delta} = T^{-1}[\sigma_\eta^4 + 4\sigma_\eta^2\sigma_\varepsilon^2(1-\rho^2)]^{\frac{1}{2}} [R'\Sigma_z R]^{-1}, \tag{4}$$

where σ_ε^2 and σ_η^2 are the variances of the disturbances appearing in the first equation and ρ^2, as given in appendix 7.2, is the coefficient of determination between ε_{1t} and the other p_1 elements in the vector ε_t.

In terms of the signal-noise ratio $q = \sigma_\eta^2/\sigma_\varepsilon^2$, (4) becomes

$$\mathrm{Avar}\hat{\delta} = T^{-1}\sigma_\varepsilon^2[q^2 + 4q(1 - \rho^2)]^{\frac{1}{2}} \quad [R'\Sigma_z R]^{-1}. \tag{5}$$

Again demonstrating the asymptotic superiority of LIML compared with IIV/ML, we have from (4.4.10) that

$$\mathrm{Avar}\hat{\delta}_{IIV/ML} = T^{-1}\sigma_\varepsilon^2[q^2 + 4q]^{\frac{1}{2}} \quad [R'\Sigma_z R]^{-1}, \tag{6}$$

hence

$$\frac{\mathrm{Avar}\hat{\delta}_{LIML}}{\mathrm{Avar}\hat{\delta}_{IIV/ML}} = \frac{[q^2 + 4q(1 - \rho^2)]^{\frac{1}{2}}}{[q^2 + 4q]^{\frac{1}{2}}} \leq 1.$$

When ρ^2 is zero, we have a special case of the result given at the end of section 6 showing the LIML and IIV/ML have the same asymptotic distribution when the disturbances in the first equation are uncorrelated with those in the others. Conversely, the maximum gain from using LIML comes as ρ^2 tends to unity. Hence the only thing affecting the asymptotic distribution of the LIML estimator of δ is the correlation between ε_{1t} and the other elements in ε_t. When there is correlation present, there is a gain in efficiency over IIV/ML since the IIV/ML does not use this information.

Appendix 7.1

Result on Matrix Inversion

Let $A(j)$, $j=1,\ldots T$, be $m\times m$ matrices partitioned as

$$A(j) = \begin{bmatrix} A_{11}(j) & A_{12}(j) \\ A_{21}(j) & A_{22}(j) \end{bmatrix}$$

where A_{11} is $m_1 x m_1$, A_{12} is $m_1 x m_2$, A_{21} is $m_2 x m_1$, A_{22} is $m_2 x m_2$ and $m_1+m_2=m$.

We want to show that

$$\left[\sum_{j=1}^{T} A_{11}(j)^{-1}\right]^{-1} - \left[\sum_{j=1}^{T} A(j)^{-1}\right]^{11} \geq 0, \qquad (1)$$

where $[\cdot]^{11}$ is the $m_1 x m_1$ top left submatrix of the inverse of the matrix in squared brackets.

Proof. Let us proceed by induction. It can be shown that (1) is true for T=2; see Harvey,Neudecker and Streibel (1991). Assume (1) to be true for T=k-1. We then have to show that (1) is true for T=k. The argument is as follows. We subtract and add

$$\left[\left\{\left[\sum_{j=1}^{k-1} A(j)^{-1}\right]^{11}\right\}^{-1} + A_{11}(k)^{-1}\right]^{-1}$$

to the LHS of (1) for T=k. This yields

$$\Big[\sum_{j=1}^{k} A_{11}(j)^{-1}\Big]^{-1} - \Big[\sum_{j=1}^{k} A(j)^{-1}\Big]^{11} =$$

$$= \Big[\sum_{j=1}^{k-1} A_{11}(j)^{-1}\Big]^{-1} - \Big[\Big\{\Big[\sum_{j=1}^{k-1} A(j)^{-1}\Big]^{11}\Big\}^{-1} + A_{11}(k)^{-1}\Big]^{-1}$$

$$+ \Big[\Big\{\Big[\sum_{j=1}^{k-1} A(j)^{-1}\Big]^{11}\Big\}^{-1} + A_{11}(k)^{-1}\Big]^{-1} - \Big[\sum_{j=1}^{k} A(j)^{-1}\Big]^{11} \quad . \qquad (2)$$

Using the fact that for $D>0$ and $F>0$, $D-F\geq 0$ if and only if $F^{-1}-D^{-1}\geq 0$, we have from (1) that

$$\Big\{\Big[\sum_{j=1}^{k-1} A(j)^{-1}\Big]^{11}\Big\}^{-1} - \sum_{j=1}^{k-1} A_{11}(j)^{-1} \geq 0$$

or

$$\Big\{\Big[\sum_{j=1}^{k-1} A(j)^{-1}\Big]^{11}\Big\}^{-1} + A_{11}(k)^{-1} - \sum_{j=1}^{k} A_{11}(j)^{-1} \geq 0$$

or

$$\Big[\sum_{j=1}^{k} A_{11}(j)^{-1}\Big]^{-1} - \Big[\Big\{\Big[\sum_{j=1}^{k-1} A(j)^{-1}\Big]^{11}\Big\}^{-1} + A_{11}(k)^{-1}\Big]^{-1} \geq 0. \quad (3)$$

On the other hand, noting that

$$\Big[\sum_{j=1}^{k} A(j)^{-1}\Big]^{11} = \Big[\Big\{\sum_{j=1}^{k} A(j)^{-1}\Big\}^{-1}\Big]_{11}$$

we can write

$$\Big[\Big\{\Big[\sum_{j=1}^{k-1} A(j)^{-1}\Big]^{11}\Big\}^{-1} + A_{11}(k)^{-1}\Big]^{-1} - \Big[\sum_{j=1}^{k} A(j)^{-1}\Big]^{11}$$

$$= \Big[\Big[\Big\{\Big[\sum_{j=1}^{k-1} A(j)^{-1}\Big]^{-1}\Big\}_{11}\Big]^{-1} + A_{11}(k)^{-1}\Big]^{-1} - \Big[\sum_{j=1}^{k-1} A(j)^{-1} + A(k)^{-1}\Big]^{11}$$

and because (1) is true for T-2, we have that the RHS matrix of the above expression is positive semidefinite. Hence (2) is a sum of two positive semidefinite matrices and therefore (2) is also positive semidefinite.

Appendix 7.2

Asymptotic Covariance Matrix When Only the First Equation has a Stochastic Trend

For simplicity we consider a two-equation system. Thus p=2 and p_1=1.

Let

$$\Sigma_\varepsilon = \begin{bmatrix} \sigma_\varepsilon^2 & \omega \\ \omega & \Omega_\varepsilon \end{bmatrix} \quad \text{and} \quad \Sigma_\eta = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

For convenience we write

$$2\,\Sigma_\varepsilon = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad \text{and} \quad \Sigma_\eta = \begin{bmatrix} d-a & 0 \\ 0 & 0 \end{bmatrix},$$

where $d = \sigma_\eta^2 + 2\sigma_\varepsilon^2$. Hence

$$2\pi F(\lambda) = \begin{bmatrix} d & b \\ b & c \end{bmatrix} - \begin{bmatrix} a & b \\ b & c \end{bmatrix} \cos\lambda = \begin{bmatrix} d-a\cos\lambda & b(1-\cos\lambda) \\ b(1-\cos\lambda) & c(1-\cos\lambda) \end{bmatrix}.$$

Now as

$$\det[2\pi F(\lambda)] = (1-\cos\lambda)(\Delta_1-\Delta_2\cos\lambda),$$

where

$$\Delta_1 = dc - b^2 = 2(\sigma_\eta^2 + 2\sigma_\varepsilon^2)\Omega_\varepsilon - (2\omega)^2$$

and

$$\Delta_2 = ac - b^2 = 4\sigma_\varepsilon^2\Omega_\varepsilon - (2\omega)^2,$$

and as

$$\frac{c(d-a\cos\lambda)}{c\det[2\pi F(\lambda)]} = \frac{1}{c(1-\cos\lambda)} + \frac{b^2}{c(\Delta_1-\Delta_2\cos\lambda)},$$

we have

$$\int_x^\pi [2\pi F(\lambda)]^{-1} d\lambda = \begin{bmatrix} \int_x^\pi \frac{c}{\Delta_1-\Delta_2\cos\lambda} d\lambda & -\int_x^\pi \frac{b}{\Delta_1-\Delta_2\cos\lambda} d\lambda \\ -\int_x^\pi \frac{b}{\Delta_1-\Delta_2\cos\lambda} d\lambda & \int_x^\pi \frac{1}{c(1-\cos\lambda)} d\lambda + \int_x^\pi \frac{b^2}{c(\Delta_1-\Delta_2\cos\lambda)} d\lambda \end{bmatrix}$$

Because

$$\det\left[\int_x^{\pi}[2\pi F(\lambda)]^{-1}d\lambda\right] = \int_x^{\pi}\frac{d\lambda}{\Delta_1-\Delta_2\cos\lambda}\int_x^{\pi}\frac{d\lambda}{1-\cos\lambda}$$

we have

$$\left[\int_x^{\pi}[2\pi F(\lambda)]^{-1}d\lambda\right]^{11} = \frac{1}{c\int_x^{\pi}\frac{d\lambda}{\Delta_1-\Delta_2\cos\lambda}} + \frac{b^2}{c\int_x^{\pi}\frac{d\lambda}{1-\cos\lambda}} .$$

Now

$$\lim_{x\to o^+}\int_x^{\pi}\frac{d\lambda}{1-\cos\lambda} = \lim_{x\to o^+}\cot\left[\frac{x}{2}\right] = \infty$$

and

$$\lim_{x\to o^+}\int_x^{\pi}\frac{d\lambda}{\Delta_1-\Delta_2\cos\lambda} = \lim_{x\to o^+}\frac{2}{(\Delta_1^2-\Delta_2^2)^{\frac{1}{2}}}\left[\frac{\pi}{2} - \arctan\frac{\Delta_1-\Delta_2}{(\Delta_1^2-\Delta_2^2)^{\frac{1}{2}}}\tan\left[\frac{x}{2}\right]\right]$$

$$= \frac{2\pi}{2(\Delta_1^2-\Delta_2^2)^{\frac{1}{2}}},$$

and so

$$\lim_{x\to o^+}\left[\int_x^{\pi}[2\pi F(\lambda)]^{-1}d\lambda\right]^{11} = \frac{2(\Delta_1^2-\Delta_2^2)^{\frac{1}{2}}}{c\;2\pi} . \qquad (1)$$

In terms of the covariance matrices Σ_ε and Σ_η, we have that

$$\Delta_1^2-\Delta_2^2 = 4\,\sigma_\eta^2\Omega_\varepsilon\,[\sigma_\eta^2\Omega_\varepsilon + 4(\sigma_\varepsilon^2\Omega_\varepsilon - \omega^2)],$$

and after some algebra

$$\frac{(\Delta_1^2-\Delta_2^2)^{\frac{1}{2}}}{c} = \left[\sigma_\eta^4 + 4\sigma_\eta^2\,[\sigma_\varepsilon^2 - \omega^2/\Omega_\varepsilon]\,\right]^{\frac{1}{2}} . \qquad (2)$$

Since we are assuming $p=2$, Ω_ε and ω are scalar. For $p>2$ and $p_1=p-1$,

Ω_ε is a $p_1 x p_1$ matrix and ω is a $p_1 x1$ vector. It can be verified that (2) holds with $\omega^2/\Omega_\varepsilon$ replaced by $\omega'\Omega_\varepsilon^{-1}\omega$.

Finally, using the notation ρ^2 for the coefficient of determination between ε_{1t} and the other p_1 elements in the vector ε_t, i.e.

$$\rho^2 = \sigma_\varepsilon^{-2}\omega'\Omega_\varepsilon^{-1}\omega, \tag{3}$$

(1) becomes

$$\lim_{x\to o^+}\Big[\int_x^\pi [2\pi F(\lambda)]^{-1}d\lambda\Big]^{11} = \frac{2}{2\pi}\Big[\sigma_\eta^4 + 4\sigma_\eta^2\sigma_\varepsilon^2[1-\rho^2]\Big]^{\frac{1}{2}}. \tag{4}$$

CHAPTER 8

AN APPLICATION TO REAL ECONOMIC TIME SERIES

1. Introduction

As an illustration of the estimation techniques described in the previous chapters we now focus our attention on the employment-output relationship. Our model is in the same spirit as the one analysed in Harvey *et al*(1986), namely

$$n_t = \lambda n_{t-1} + \delta_1 q_t + \delta_2 q_{t-1} + \mu_t + \epsilon_t \quad (1a)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad (1b)$$

$$\beta_t = \beta_{t-1} + \zeta_t \quad (1c)$$

where n_t refers to employment and q_t refers to output. However while the approach adopted in Harvey *et al*(1986) treats output as a weakly exogenous variable here we treat it as an endogenous one. This seems more reasonable, since the assumption that output is fixed in advance and employment simply adapts to it is rather a strong one. It seems more plausible that output and employment are jointly determined. But since we do not wish to specify a model for the determination of output, single equation estimation of the employment equation is appropriate.

We estimated (1) using various IV procedures. The variables we have chosen to act as instruments are: world trade, UK investment and UK government expenditure.

Data Definitions

The five data series used run from 1963Q1 to 1983Q3 and are seasonally adjusted. They are:

a)EMP = logarithm of UK manufacturing employment in thousands;

b)OUT = logarithm of UK manufacturing output index, with 1980=100;

c)WTT = logarithm of total world trade index, with 1980=100;

d)QDK = logarithm of UK total investment, 1980 prices;

e)GEXP= logarithm of UK government expenditure, 1980 prices.

2. Estimation of the Model

Our estimation of (1.1) by means of instrumental variable estimators is conducted by considering two different sets of instruments. So we have

Case A - where the instruments used are the predetermined variables EMP-1 and OUT-1, acting as their own instruments, WTT, QDK and GEXP, and all these variables lagged once;

Case B - as above plus all variables lagged twice .

So two observations are lost in case A and three in case B. However, for comparison purposes we have discarded the first observation when estimating case A.

The organisation of the study is the following. We begin reestimating (1.1) where we have discarded the first 2 observations, without taking into account the endogeneity of output. Our estimated version of (1.1) is given in table 1. The estimates obtained are slightly different from those reported in Harvey *et al*(1986), (after correction the coefficient of lagged output in equation 17 of Harvey *et al* should read 0.058 and the constant term in equation 18 should read -0.00159). The differences are primarily because we have less observations.

Next we estimate the model by the instrumental variable procedures outlined in section 4.3. For each case, A and B, we consider four different procedures. These are:

1)IIV/ML with transformed instruments;

2)IIV/ML with untransformed instruments;

3)IV/QML with transformed instruments;

4)IV/QML with untransformed instruments.

In all procedures σ_η^2 is concentrated out, and the starting values for the hyperparameters were $\sigma_\varepsilon^2=0.000001$, $\sigma_\eta^2=0.000005$ $\sigma_\zeta^2=0.00000$. For IIV/ML, the starting values for the regression coefficients on (n_t, q_t, q_{t-1}) are (0.10 0.76 0.06). In preliminary estimation we have considered different starting values and we find out that the final results seem not to be affected by the choice of the starting values. The program was run on the VAX and the optimisation subroutine was E04JBF from the NAG library. The CPU time consumed by the IIV/ML procedure is considerably less then by the IV/QML one. The results are reported in tables 4 to 11.

Model in First Differences

Since, in all cases, the estimate of σ_ζ^2 is equal to zero, we estimate the model in differences by means of the TSP program for comparisons purposes. We note that when $\sigma_\zeta^2 = 0$, differencing once model (1.1) yields

$$\Delta n_t = \beta + \lambda \Delta n_{t-1} + \delta_1 \Delta q_t + \delta_2 \Delta q_{t-1} + w_t \tag{1a}$$

where

$$w_t = \eta_t + \Delta \epsilon_t \tag{1b}$$

Because the disturbances, w_t, are serially correlated and Δn_{t-1} is a lagged dependent variable, consistent estimation is achieved based on instrumental variable estimators with Δn_{t-2} acting as an instrument for Δn_{t-1}. As before we shall consider two cases. These are:

Case C - where the instruments are

CONST ΔOUT-1 ΔEMP-2 ΔOUT-2 ΔGEXP ΔWTT ΔQDK ΔGEXP-1 ΔWTT-1 ΔQDK-1

and

Case D - where in addition to the instruments given in C, we also consider ΔEMP-3 ΔOUT-3 ΔGEXP-2 ΔWTT-2 ΔQDK-2 as instruments.

However, since η_t dominates ϵ_t, using ΔEMP-1 as an instrument for itself should also give coefficients close to those obtained in cases

A and B. Therefore, two more cases are considered:

Case C'- ΔEMP-1 is included in the set of instruments given in C

and

Case D'- ΔEMP-1 is included in the set of instruments given in D.

The results are reported in tables 12 to 15. These results may be compared with the OLS and IV estimates of the first differenced equation reported in tables 2 and 3.

3. Conclusions

Although the results obtained do not allow us to draw dramatic conclusions about the endogeneity of output, it is worth paying attention to the following points.

i) Whether the equation is formulated in levels as in (1.1), and IV estimation is carried out with transformed instruments, or in first differences as in (2.1), there is, although small, a systematic increase in the coefficient for output which varies from 2.8% to 8% according to the estimation procedure and a minimum decrease in the lagged output coefficient. The changes in the lagged employment coefficient are negligible. Compare the results given in tables 4,5,8 and 9, and in tables 13 and 15 with those in table 1.

ii) Taking lagged once or lagged twice instruments seems not to affect the results systematically. Compare tables 4,5,6 and 7 with tables 8,9,10 and 11, respectively.

iii)Differences between the two estimation procedures, IV/QML and IIV/ML seem to be dominant. Different estimates were expected, since, as pointed out in earlier chapters, the two estimation procedures are not numerically equivalent.

iv)There is no longer an increase in the the coefficient of output when untransformed instruments are used, see tables 6,7,8 and 9.

v)When the model is formulated in first differences and ΔEMP-1 is taken as an instrument for itself, all coefficients are affected, compare table 12 with 13, and 14 with 15.

Appendix 8.1 - Tables

Table 1 - GLS estimates for equation with stochastic trend

LEVEL	TREND	OUT	EMP-1	OUT-1
1.355372	-0.001603	0.111968	0.752064	0.059587
(0.290319)	(0.000318)	(0.013930)	(0.037596)	(0.015755)

$\sigma_{\varepsilon}^2 = 0.0000013880$ $\sigma_{\eta}^2 = 0.0000048945$ $\sigma_{\zeta}^2 = 0.0000000000$

Table 2 - OLS estimates for first differenced equation

CONST	OUT	EMP-1	OUT-1
-0.001717	0.106017	0.728599	0.063339
(0.000396)	(0.014522)	(0.045486)	(0.016119)

Table 3 - IV estimates for first differenced equation
INSTR.: CONST ΔOUT ΔOUT-1 ΔEMP-2

CONST	OUT	EMP-1	OUT-1
-0.001547	0.103886	0.761677	0.057802
(0.000411)	(0.014632)	(0.050002)	(0.016533)

Table 4 - IV/QML estimates with transformed instruments, case A

LEVEL	TREND	OUT	EMP-1	OUT-1
1.378326	-0.001627	0.114613	0.748057	0.059503
(0.297369)	(0.000335)	(0.026826)	(0.039626)	(0.016234)

$\sigma_\varepsilon^2 = 0.0000012152$ $\sigma_\eta^2 = 0.0000051951$ $\sigma_\zeta^2 = 0.0000000000$

Table 5 - IIV/ML estimates with transformed instruments, case A

LEVEL	TREND	OUT	EMP-1	OUT-1
1.331485	-0.001614	0.120290	0.751617	0.057341
(0.284708)	(0.000318)	(0.026123)	(0.037767)	(0.016626)

$\sigma_\varepsilon^2 = 0.0000015553$ $\sigma_\eta^2 = 0.0000046499$ $\sigma_\zeta^2 = 0.0000000000$

Table 6 - IV/QML estimates with untransformed instruments, case A

LEVEL	TREND	OUT	EMP-1	OUT-1
1.384608	-0.001594	0.101688	0.751917	0.063755
(0.296205)	(0.000336)	(0.028813)	(0.039913)	(0.016473)

$\sigma_\varepsilon^2 = 0.0000012986$ $\sigma_\eta^2 = 0.0000050924$ $\sigma_\zeta^2 = 0.0000000000$

Table 7 - IIV/ML estimates with untransformed instruments, case A

LEVEL	TREND	OUT	EMP-1	OUT-1
1.398523	-0.001602	0.101428	0.750384	0.063871
(0.300016)	(0.000341)	(0.028836)	(0.040413)	(0.016347)

$\sigma_\varepsilon^2 = 0.0000011987$ $\sigma_\eta^2 = 0.0000052636$ $\sigma_\zeta^2 = 0.0000000000$

Table 8 - IV/QML estimates with transformed instruments, case B

LEVEL	TREND	OUT	EMP-1	OUT-1
1.366334	-0.001623	0.115793	0.749059	0.059051
(0.294011)	(0.000327)	(0.021076)	(0.038571)	(0.016032)

$\sigma_\epsilon^2 = 0.0000013011$ $\sigma_\eta^2 = 0.0000050496$ $\sigma_\zeta^2 = 0.0000000000$

Table 9 - IIV/ML estimates with transformed instruments, case B

LEVEL	TREND	OUT	EMP-1	OUT-1
1.338629	-0.001611	0.117879	0.751696	0.058039
(0.286208)	(0.000317)	(0.020985)	(0.037552)	(0.016176)

$\sigma_\epsilon^2 = 0.0000015047$ $\sigma_\eta^2 = 0.0000047175$ $\sigma_\zeta^2 = 0.0000000000$

Table 10 - IV/QML estimates with untransformed instruments, case B

LEVEL	TREND	OUT	EMP-1	OUT-1
1.361570	-0.001603	0.107679	0.751692	0.063237
(0.287485)	(0.000320)	(0.024287)	(0.037928)	(0.016375)

$\sigma_\epsilon^2 = 0.0000014860$ $\sigma_\eta^2 = 0.0000047482$ $\sigma_\zeta^2 = 0.0000000000$

Table 11 - IIV/ML estimates with untransformed instruments, case B

LEVEL	TREND	OUT	EMP-1	OUT-1
1.391155	-0.001614	0.105319	0.749175	0.063884
(0.296412)	(0.000331)	(0.024349)	(0.039110)	(0.016174)

$\sigma_\epsilon^2 = 0.0000012613$ $\sigma_\eta^2 = 0.0000051292$ $\sigma_\zeta^2 = 0.0000000000$

Table 12 - IV estimates for first differenced equation, case C'

CONST	OUT	EMP-1	OUT-1
-0.001654	0.092805	0.736950	0.062869
(0.000420)	(0.031269)	(0.048963)	(0.016238)

Table 13 - IV estimates for first differenced equation, case C

CONST	OUT	EMP-1	OUT-1
-0.001591	0.115784	0.756682	0.057791
(0.000422)	(0.033890)	(0.050279)	(0.016511)

Table 14 - IV estimates for first differenced equation, case D'

CONST	OUT	EMP-1	OUT-1
-0.001708	0.104078	0.7298243	0.063270
(0.000405)	(0.022519)	(0.0467747)	(0.016132)

Table 15 - IV estimates for first differenced equation, case D

CONST	OUT	EMP-1	OUT-1
-0.001600	0.117573	0.7555205	0.057860
(0.000412)	(0.023756)	(0.0490003)	(0.016502)

CHAPTER 9

MONTE CARLO STUDY

1. Introduction

We conducted a series of Monte Carlo experiments to examine the performance of the estimators and their small sample behaviour. The basic model is a two equation simultaneous system as in Hendry and Harrison (1974). However the unobserved part follows a multivariate random walk plus noise, and the exogenous variables are non stationary. Thus the form of the model is

$$y_{1t} = \beta_1 y_{2t} + \delta y_{1t-1} + \gamma_1 z_{1t} + \mu_{1t} + \epsilon_{1t} \tag{1a}$$

$$y_{2t} = \beta_1 y_{1t} + \sum_{i=2}^{4} \gamma_i z_{it} + \mu_{2t} + \epsilon_{2t} \tag{1b}$$

where $\mu_t = (\mu_{1t}, \mu_{2t})'$ is a Gaussian multivariate random walk and $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})'$ is a Gaussian multivariate white noise process. The covariance matrices of the disturbance vectors ϵ_t and η_t are respectively Σ_ϵ and Σ_η. The exogenous variables $z_{it}, i=1,4$ are independent random walks generated by

$$z_{it} = z_{it-1} + \nu_{it} , \tag{2}$$

where ν_{it} is a Gaussian white noise process with variance $\sigma_{\nu i}^2$, $i=1,4$. Our interest centres on estimation of the first equation. This is overidentified.

2. The Choice of the Parameters

The values chosen for β_1, β_2 and γ_i, $i=1,4$ are at $\beta_1=0.2$, $\beta_2=0.4$ and $\gamma_1=\gamma_2=\gamma_3=\gamma_4=1.0$. When a lagged dependent variable is included, $\delta=0.5$. These values are close to the ones in Hendry and Harrison (1974). $\sigma_{\nu i}^2$ is fixed at 2.0.

The values for Σ_ε and Σ_η were fixed according to the following argument. Differencing (1.1) yields, in matrix notation,

$$\Delta y_t'B + \Delta z_t'\Gamma = u_t' \qquad (1)$$

where $u_t=(\eta_{1t}+\Delta\varepsilon_{1t}\ ,\ \eta_{2t}+\Delta\varepsilon_{2t})'$ is a vector MA(1) process with disturbances covariance matrix Ω,

$$\Omega = \begin{bmatrix}\omega_{11} & \omega_{12}\\ \omega_{21} & \omega_{22}\end{bmatrix} = 2\Sigma_\varepsilon + \Sigma_\eta.$$

Models with stationary disturbances were studied in Campos (1986b). So we have chosen the values for the diagonal elements of Ω, proportional to the ones given in Campos (1986b). The values for the off diagonal elements are such that the correlation between u_{1t} and

u_{2t} is 0.5. This yields

$$\Omega = \begin{bmatrix} 1.5 & 1.5 \\ 1.5 & 6.0 \end{bmatrix} .$$

With Ω given above, the asymptotic reduced form multiple correlation coefficient of the differenced model when no lagged variable is included is equal to 0.7. Next we split Ω into $2\Sigma_\varepsilon$ and Σ_η in such a way that we have the homogeneous case with q=1. That is,

$$\Sigma_\varepsilon = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 2.0 \end{bmatrix} \quad \text{and} \quad \Sigma_\eta = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 2.0 \end{bmatrix} .$$

We call this experiment A.

We also consider experiment B where only the measurement disturbances ε_t are correlated, so

$$\Sigma_\varepsilon = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 2.0 \end{bmatrix} \quad \text{and} \quad \Sigma_\eta = \begin{bmatrix} 0.5 & 0 \\ 0 & 2.0 \end{bmatrix} ;$$

experiment C, where the stochastic trend is absent in the second equation, so

$$\Sigma_\varepsilon = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 2.0 \end{bmatrix} \quad \text{and} \quad \Sigma_\eta = \begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix} ;$$

and experiment D where the variability is more in the measurement disturbances than in the stochastic trend, so

$$\Sigma_\varepsilon = \begin{bmatrix} 0.6 & 0.55 \\ 0.55 & 2.0 \end{bmatrix} \quad \text{and} \quad \Sigma_\eta = \begin{bmatrix} 0.3 & 0 \\ 0 & 2.0 \end{bmatrix}.$$

If a lagged dependent variable is included in the first equation we consider experiment A', the homogeneous case, and experiment C', when there is no stochastic trend in the second equation.

We mention that we have tried several different sets of parameters not reported here. We found in some cases that when performing LIML around 10% of the replications gave anomalous results. In other estimation procedures this did not happen. We note that in order to obtain LIML we have to optimise the criterion function nonlinearly with respect to $\psi^\dagger$, where the p(p+1)x1 vector $\psi^\dagger$ contains the elements of the triangular matrices obtained when $\Sigma_\varepsilon{}^\dagger$ and $\Sigma_\eta{}^\dagger$ are factored by means of the Cholesky decomposition. The reason might be explained by the fact that if $|\psi_i{}^\dagger|>>1$ for some i, i=1,...,p(p+1) numerical problems might arise when optimising by means of the NAG subroutine. It is possible that this problem could be overcome by suitable re-scaling.

3. Data Generation

The simulations are carried out with sample sizes T=50 and T=200. Values for y_t', t=1,...,T are obtained by solving the reduced form of (1.1), that is,

$$y_t' = z_t'\Gamma B^{-1} + (\mu_t + \epsilon_t)'B^{-1} .$$

The (T+30)x4 matrix Z containing the exogenous variables is generated as follows. Using the NAG subroutine G05DDF, for t running from 1 to T+30, the (4x1) vector ν_t is generated as a Gaussian white noise process with covariance matrix $\Omega_\nu = \text{diag}\{2.0\ 2.0\ 2.0\ 2.0\}$ and the t-th row of the matrix Z is obtained by adding the transpose of ν_t to the previous row of Z. We have set Z(0,1)=Z(0,2)=Z(0,3)=Z(0,4)=0.

Similarly, for t running from 1 to T+30, the 2x1 vector η_t is generated as a bivariate Gaussian white noise process with covariance matrix Σ_η, in order to obtain the (T+30)x2 matrix μ containing the stochastic trends. Again we have set $\mu(0,1)=\mu(0,2)=0$.

Finally the (T+30)x2 matrix ϵ is formed with each row generated as a bivariate Gaussian white noise process with covariance Σ_ϵ.

We then form the Tx2 matrix Y, where the t-th row is obtained adding the t-th row of $\epsilon+\mu$ to the t-th row of Z post-multiplied by Γ, and postmultiplied the sum by B^{-1}. The first 30 observations are discarded.

If a lagged dependent variable is included in the first equation the matrix Z is augmented to include $y_1(-1)$. Thus each row of Z is given by

$$z_t' = [y_{1t-1}, z_{1t}, z_{2t}, z_{3t}, z_{4t}]$$

where y_{1t-1} is obtained from the previous row vector $[y_{1t-1}, y_{2t-2}]$. We set $y_{10} = 0$ and again discard the 30 first observations.

In subsequent replications we do not change the exogenous variables.

4. Estimation Procedures

The following estimation procedures were considered in order to obtain estimates of the parameters in the first equation:

a)GLS - is the ML estimate under the assumption that all the explanatory variables in the first equation are exogenous. This is described in section 3.2. The criterion function is given in (3.2.3) which is optimised by means of the stepwise algorithm. The hyperparameter concentrated out is σ_η^2. In a preliminary study we have concentrated out σ_ε^2. Although the results were not significantly different from the ones obtained when σ_η^2 is concentrated out, it proved more appropriate to concentrate σ_η^2 out for numerical reasons.

b)IV/QML - as described in section 4.3. The criterion function is given in (4.3.7). σ_η^2 is concentrated out. The three estimators considered differ according to the transformation applied to the instruments. They are:

IV/QML1 - Kalman filter is applied to the instruments. Thus, $\hat{\delta}$ in (4.3.8) is obtained from the minimand given in (4.3.2) with projection matrix P_1 as given in (4.3.3);

IV/QML^2 - differencing the instruments once. As above but with projection matrix as given in (4.3.5);

IV/QML^3 - untransformed instruments. As above but with projection matrix as given in (4.3.4).

c)IIV/ML - as described in section 4.3. As for the IV/QML estimation procedures we have considered three estimators, these are:

IIV/ML^1 - Kalman filter is applied to the instruments. That is, at each step (4.3.6) is optimised with respect to σ_η^2 and $\psi_*=\sigma_\varepsilon^2/\sigma_\eta^2$, with δ replaced by $\hat{\delta}(\hat{\psi}_*)$, where $\hat{\delta}(\hat{\psi}_*)$ is the feasible IV estimator obtained from the minimand given in (4.3.2) with projection matrix P_1 as given in (4.3.3). $\hat{\psi}_*$ was obtained in the previous step;

IIV/ML^2 - differencing the instruments once. As above but with projection matrix as given in (4.3.5);

IIV/ML^3 - untransformed instruments. As above but with projection matrix as given in (4.3.4).

d)2SLS^1 - as given in (4.3.11), that is, 2SLS is applied to the first equation after all the variables have been differenced. The estimates of σ_ε^2 and σ_η^2 were obtained from the variance and first-order autocovariance of the residuals. These estimators are inefficient, but they are useful for an iterative procedure.

2SLS^2 - as above, but without differencing the instruments.

e)LIML - as described in section 7.4, that is the criterion function given in (7.4.7) is optimised with respect to the vector $\psi^\dagger$

containing the 6 parameters. The LIML estimate of β_1 and γ_1 are the first elements of the vector $\hat{\alpha}$, which is given in (7.4.6).

The criterion functions were optimised by means of the NAG subroutine E04JBF. We have chosen as starting values the true hyperparameters to avoid extra computing time. However we had carried out a preliminary study to check whether the optimal point is affected by the choice of the initial values and we found that this seems not to be the case.

5. Asymptotic Standard Errors

We shall now report the asymptotic standard errors (ASE), given by the square roots of the diagonal elements of the Avar matrix, for the LIML, IIV/ML[1] and 2SLS[1] estimators of (β_1, γ_1) as outlined above. The relevant formulae for the 2SLS and IIV/ML estimators are given in (4.3.14) and (4.4.10), respectively, and are

$$\text{Avar}\begin{bmatrix}\hat{\beta}_1\\ \gamma_1\end{bmatrix}_{2SLS} = T^{-1}[2\sigma_\varepsilon^2 + \sigma_\eta^2]\,[R'\Sigma_z R]^{-1} \tag{1}$$

$$\text{Avar}\begin{bmatrix}\hat{\beta}_1\\ \gamma_1\end{bmatrix}_{IIV/ML} = T^{-1}[\sigma_\eta^4 + 4\sigma_\eta^2\sigma_\varepsilon^2]^{\frac{1}{2}}\,[R'\Sigma_z R]^{-1} \tag{2}$$

The asymptotic covariance matrix for the LIML estimators of β_1 and γ_1 varies according to the experiment. So in experiment A, the homogeneous case, IIV/ML is as efficient as LIML, see section 7.6.; in experiment C, where no stochastic trend appears in the second equation, we have from (7.7.4) that

$$\text{Avar}\begin{bmatrix}\hat{\beta}_1\\ \gamma_1\end{bmatrix}_{\text{LIML}} = T^{-1}[\sigma_\eta^4 + 4\sigma_\eta^2\sigma_\varepsilon^2(1-\rho^2)]^{\frac{1}{2}}\,[R'\Sigma_z R]^{-1}\ , \tag{3}$$

where $\rho^2 = \sigma_\varepsilon^{-2}\omega^2/\Omega_\varepsilon$; for experiments B and D, the asymptotic covariance is evaluated as outlined at the end of section 7.5.

Now, for the parameter values given in section 2, we have

$$\rho^2 = \sigma_\varepsilon^{-2}\omega^2/\Omega_\varepsilon = 0.25, \tag{4}$$

and if no lagged dependent variable is included

$$B = \begin{bmatrix} 1 & -0.4 \\ -0.2 & 1 \end{bmatrix} \quad B^{-1} = \begin{bmatrix} 1/0.92 & 0.4/0.92 \\ 0.2/0.92 & 1/0.92 \end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

Thus,

$$\Pi = \begin{bmatrix}\Pi_{11} & \Pi_{12}\\ \Pi_{21} & \Pi_{22}\end{bmatrix} = \begin{bmatrix} 1/0.92 & 0.4/0.92 \\ \hdashline 0.2/0.92 & 1/0.92 \\ 0.2/0.92 & 1/0.92 \\ 0.2/0.92 & 1/0.92 \end{bmatrix}$$

and from (7.5.4)

$$R = \begin{bmatrix}\Pi_{12} & 1\\ \Pi_{22} & 0\end{bmatrix} = \begin{bmatrix} 0.4/0.92 & 1 \\ 1/0.92 & 0 \\ 1/0.92 & 0 \\ 1/0.92 & 0 \end{bmatrix},$$

and so

$$(R'\Sigma_z R)^{-1} = \begin{bmatrix} 0.14107 & -0.06134 \\ -0.06134 & 0.52665 \end{bmatrix} \tag{5}$$

Thus, for T=50, for experiment A, the homogeneous case, from (5), (2) and (1) we have

$$ASE_{IIV/ML}(\hat{\beta}_1) = ASE_{LIML}(\hat{\beta}_1) = 0.05616 \quad (6a)$$

$$ASE_{IIV/ML}(\hat{\gamma}_1) = ASE_{LIML}(\hat{\gamma}_1) = 0.10852 \quad (6b)$$

$$ASE_{2SLS}(\hat{\beta}_1) = 0.06505 \quad (7a)$$

$$ASE_{2SLS}(\hat{\gamma}_1) = 0.12570 \quad (7b)$$

For experiment B, the ASEs of the LIML estimators of β_1 and γ_1 are

$$ASE_{LIML}(\hat{\beta}_1) = 0.05557 \quad (8a)$$

$$ASE_{LIML}(\hat{\gamma}_1) = 0.10737 \quad (8b)$$

Clearly, because the hyperparameters, associated with the disturbances in the first equation, are the same as the ones considered in experiment A, the ASEs of the IIV/ML and 2SLS estimators of β_1 and γ_1 are as given in (6) and (7) respectively.

For experiment C, from (3) and (5), it follows that the ASE's of the LIML estimators of β_1 and γ_1 are

$$ASE_{LIML}(\hat{\beta}_1) = 0.05312 \quad (9a)$$

$$ASE_{LIML}(\hat{\gamma}_1) = 0.10263 \quad (9b)$$

Again the ASEs of the IIV/ML and 2SLS estimators of β_1 and γ_1 are as in (6) and (7).

Finally for experiment D we have

$$ASE_{LIML}(\hat{\beta}_1) = 0.04995 \tag{10a}$$

$$ASE_{LIML}(\hat{\gamma}_1) = 0.09652 \tag{10b}$$

$$ASE_{IIV/ML}(\hat{\beta}_1) = 0.05039 \tag{11a}$$

$$ASE_{IIV/ML}(\hat{\gamma}_1) = 0.09737 \tag{11b}$$

The ASEs for the 2SLS estimators are as in (7).

6. Discussion of the Results

The results, which are based on 100 replications in each experiment are shown in appendix 9.1. In Tables 1 to 7 and 12 to 35 are reported the results from experiments A, B, C and D, that is, no lagged dependent variable in the first equation while tables 8 to 11 and 36 to 55 contain the results from experiments A' and C', that is, a lagged dependent variable is included in the first equation. Initially we analysed the results for experiments A, B, C and D.

The top entry in each box of tables 1 to 7 gives the estimate of the bias while the two figures below give the standard deviation and root mean square error (RMSE) respectively. The ASEs given in the previous section are reported at the bottom of these tables. The main findings may be summarised as follows.

(i) The GLS estimator is, as expected, biased, and this more than offsets the relatively small variance when the MSE is calculated.

(ii) The ASEs seem to give a reliable guide to the performance of the 2SLS^1, IIV/ML^1 and LIML for T=200. For experiment A, the homogeneous case, and T=200, the RMSEs of the two estimators LIML and IIV/ML^1 are, as expected, roughly the same. The RMSEs are slightly larger than the corresponding ASEs. This is not uncommon in econometrics, although here the discrepancy may be because of holding the explanatory variables constant throughout the simulations. The LIML and IIV/ML^1 procedures are also similar for experiment B, although from the theoretical point of view LIML should yield smaller RMSEs. However, the difference between the ASEs is so small that we cannot expect a significant difference between the estimators. In experiment C, the ASEs are slightly smaller for LIML and this also is shown in the simulations.

(iii) In experiments A and B, for T=50, IIV/ML^1 tends to be slightly better than LIML. The comparison between LIML and IV/QML^1 is not conclusive. While LIML has smaller RMSE for $\hat{\sigma}_\eta^2$, IV/QML^1 has smaller RMSE for $\hat{\beta}_1$ and $\hat{\gamma}_1$. In experiment C, LIML has smaller RMSEs than has IIV/ML^1 which in turn has smaller RMSEs than IV/QML^1 and 2SLS.

(iv) In experiment D, we find that the performance of IIV/ML^1 is relatively better when compared with 2SLS than it is in experiments A, B and C. This is also expected since the 2SLS is optimal under the assumption that $q=\infty$, hence for $q=0.5$ (experiment D) we expected a relative better performance of IIV/ML^1 than for $q=1$ (experiments A,B and C). In a general way, however, the overall performance of the 2SLS is quite good, although there are clearly gains to be had from using IIV/ML^1 and LIML. It certainly seems reasonable to recommend using 2SLS estimates of both the explanatory variable coefficients and the hyperparameters as starting values for an iterative procedure.

(v) The question of using transformed or untransformed instruments is well illustrated. All estimation procedures with untransformed instruments yield estimates with considerable larger RMSEs for all estimates of the parameters. This is not surprising since the untransformed instruments are integrated of order one (non-stationary) while the transformed variables are all stationary. Therefore the correlation between them and the transformed explanatory variables will tend to be smaller. Using first difference instruments rather than transformed via Kalman filter seems to be appropriate.

(vi) 2SLS and IIV/ML have finite moments up to the order of overidentification, which in our model is 3. However LIML does not have any moments, and so one must be careful in drawing conclusions based on RMSEs. However tables 12 to 35 indicate that we can be confident in using the estimated RMSEs as a basis for comparison. These tables contain the minimum and maximum values of the estimates, various percentiles, the first interdecile range and the theoretical first interdecile range under the assumption of normality with standard deviation being the standard deviation obtained from the simulations. We find that for T=50 the observed interdecile range is slightly smaller than we would expect if the distribution were normal. Thus there is only a small tendency towards heavy tails, and it seems that extreme observations are very unlikely to arise in practice. This is consistent with the fact that the random numbers are generated as truncated normals, and, as Sargan (1982) has argued, MSEs calculated from simulations can still provide a good guide in such cases. Overall the differences between these two interdecile ranges are small, indicating our Monte Carlo study gives a reasonably reliable indication of the variability of our estimators in small samples.

(vii) The apparent superiority of IIV/ML^1 over LIML indicated by comparison of RMSEs in the homogeneous case for T=50 is also confirmed in tables 14 and 15, by noting that the range of the first interdecile is slightly smaller for IIV/ML^1 then for LIML. For T=200 they are more or less the same.

First Equation Containing a Lagged Dependent Variable

We now turn our attention to the results obtained from experiments A' and C', that is, the case where a lagged dependent variable is included in the first equation. The estimated biases, standard deviations and root mean square errors are given in tables 8 and 9 for experiment A' and in tables 10 and 11 for experiment C'. The percentiles for experiments A' and C' are given in tables 36 to 45 and 46 to 55 respectively. No substantial differences between the two experiments were encountered, neither for T=200 nor for T=50. The important finding is the admirable performance of IIV/ML^1 in the presence of a lagged dependent variable in the first equation. Our experience with LIML was somewhat disappointing. Extreme LIML observations where encountered for $\hat{\sigma}_\eta^2$, see tables 36, 41, 46 and 51. Also for the LIML estimates of σ_ε^2, σ_η^2, β_1, δ, γ_1, in experiment A', the ratio of the RMSEs for T=50 to that for T=200 ranges from 1.19 to 3.02. For the IIV/ML^1 procedure the ratio ranges from 2.19 to 2.63. Thus, as expected, the RMSEs for IIV/ML^1 are halved. IIV/ML^1 behaves in a consistent fashion with respect to transformations. Its superiority over other IV estimators is still apparent despite the fact that there is no firm theoretical foundation for this when a lagged dependent variable is present. Since LIML is so erratic further investigation is necessary to check if the bad performance arises due to the presence of a lagged dependent variable or due to computational difficulties. From the results obtained its use seems risky.

Appendix 9.1 - Tables

Table 1 - Estimated biases, standard deviations and RMSE's for Experiment A, T = 50

	$\sigma_\epsilon^2 = 0.500$		$\sigma_\eta^2 = 0.500$		$\beta_1 = 0.200$		$\gamma_1 = 1.000$	
GLS	-0.13399		-0.03944		0.11384		-0.09763	
	0.171	0.217	0.247	0.251	0.035	0.120	0.107	0.145
IV/QML[1]	0.00446		-0.00604		0.01585		-0.03153	
	0.272	0.272	0.371	0.371	0.050	0.053	0.123	0.127
IV/QML[2]	0.00117		-0.00661		0.00985		-0.01323	
	0.244	0.244	0.321	0.322	0.052	0.053	0.121	0.122
IV/QML[3]	0.09806		0.00151		0.03975		-0.09919	
	0.412	0.424	0.431	0.431	0.094	0.103	0.246	0.266
2SLS[1]	-0.02806		0.03265		0.00349		-0.01113	
	0.279	0.280	0.370	0.371	0.053	0.053	0.136	0.136
IIV/ML[1]	0.02243		-0.03343		0.00200		-0.01491	
	0.224	0.226	0.267	0.269	0.048	0.048	0.122	0.123
IIV/ML[2]	-0.00282		-0.00462		0.00412		-0.01101	
	0.214	0.214	0.259	0.259	0.052	0.052	0.124	0.125
2SLS[2]	0.01731		0.18881		0.03344		-0.09106	
	0.459	0.460	0.555	0.586	0.118	0.123	0.340	0.352
IIV/ML[3]	0.08108		0.04847		0.02993		-0.08752	
	0.352	0.361	0.450	0.453	0.106	0.110	0.277	0.291
LIML	0.05248		-0.01783		-0.00847		-0.00499	
	0.276	0.281	0.308	0.309	0.057	0.057	0.133	0.133

Asymptotic standard errors	2SLS[1]	0.0650	0.1257
	IIV/ML[1]	0.0562	0.1086
	LIML	0.0562	0.1086

Table 2 - Estimated biases, standard deviations and RMSE's for Experiment A, T = 200

	$\sigma_\varepsilon^2 = 0.500$	$\sigma_\eta^2 = 0.500$	$\beta_1 = 0.200$	$\gamma_1 = 1.000$
GLS	-0.15635 0.080 0.176	-0.00960 0.125 0.126	0.12106 0.015 0.123	-0.07152 0.044 0.084
IV/QML[1]	0.00782 0.133 0.133	-0.00996 0.176 0.176	0.00342 0.032 0.032	-0.01810 0.052 0.055
IV/QML[2]	0.00595 0.121 0.121	-0.00525 0.148 0.149	0.00137 0.032 0.032	-0.01640 0.053 0.056
IV/QML[3]	0.02374 0.244 0.245	0.16474 0.409 0.441	0.00589 0.071 0.071	-0.01427 0.257 0.258
2SLS[1]	-0.00068 0.129 0.129	0.00184 0.152 0.152	0.00055 0.033 0.033	-0.01657 0.061 0.063
IIV/ML[1]	0.01006 0.109 0.109	-0.01420 0.124 0.125	0.00078 0.032 0.032	-0.01604 0.052 0.055
IIV/ML[2]	0.00652 0.110 0.110	-0.00634 0.122 0.122	0.00012 0.032 0.032	-0.01584 0.054 0.056
2SLS[2]	-0.00877 0.172 0.172	0.22684 0.424 0.482	0.00067 0.075 0.075	-0.00934 0.267 0.267
IIV/ML[3]	0.00130 0.158 0.158	0.19885 0.422 0.467	0.00255 0.074 0.074	-0.01141 0.264 0.264
LIML	0.01237 0.109 0.110	-0.01252 0.126 0.126	-0.00044 0.032 0.032	-0.01525 0.052 0.055
Asymptotic standard errors		2SLS[1] IIV/ML[1] LIML	0.0325 0.0281 0.0281	0.0628 0.0543 0.0543

Table 3 - Estimated biases, standard deviations and RMSE's for Experiment B, T = 50

	$\sigma_\varepsilon^2 = 0.500$		$\sigma_\eta^2 = 0.500$		$\beta_1 = 0.200$		$\gamma_1 = 1.000$	
GLS	-0.10000		0.01679		0.08435		-0.07470	
	0.191	0.216	0.279	0.279	0.038	0.093	0.116	0.138
IV/QML[1]	0.01294		-0.01358		0.01025		-0.02275	
	0.249	0.249	0.324	0.325	0.049	0.050	0.122	0.125
IV/QML[2]	0.00295		-0.00226		0.00675		-0.01015	
	0.227	0.227	0.289	0.289	0.052	0.052	0.123	0.123
IV/QML[3]	0.18983		-0.00473		0.01729		-0.06999	
	0.470	0.507	0.394	0.394	0.098	0.100	0.259	0.268
2SLS[1]	-0.02701		0.04075		0.00144		-0.00987	
	0.274	0.276	0.365	0.367	0.054	0.054	0.136	0.137
IIV/ML[1]	0.02510		-0.03249		0.00059		-0.01223	
	0.225	0.226	0.256	0.258	0.049	0.049	0.122	0.123
IIV/ML[2]	-0.00149		0.00090		0.00259		-0.00863	
	0.212	0.212	0.249	0.249	0.052	0.052	0.125	0.125
2SLS[2]	0.05378		0.23686		0.01378		-0.05233	
	0.439	0.442	0.560	0.609	0.119	0.120	0.345	0.349
IIV/ML[3]	0.13154		0.07854		0.01140		-0.05569	
	0.352	0.376	0.385	0.393	0.110	0.110	0.290	0.296
LIML	0.04466		-0.02715		-0.00594		-0.00446	
	0.259	0.263	0.266	0.267	0.054	0.054	0.133	0.133

Asymptotic standard errors	2SLS[1]	0.0650	0.1257
	IIV/ML[1]	0.0562	0.1086
	LIML	0.0556	0.1074

Table 4 - Estimated biases, standard deviations and RMSE's for Experiment B, T = 200

	$\sigma_\varepsilon^2 = 0.500$			$\sigma_\eta^2 = 0.500$			$\beta_1 = 0.200$			$\gamma_1 = 1.000$		
GLS	-0.11778	0.088	0.147	0.04541	0.130	0.138	0.08760	0.018	0.090	-0.05683	0.049	0.075
IV/QML[1]	0.00789	0.120	0.120	-0.01201	0.139	0.140	0.00263	0.031	0.031	-0.01737	0.052	0.055
IV/QML[2]	0.00604	0.114	0.114	-0.00716	0.123	0.123	0.00101	0.031	0.031	-0.01625	0.054	0.056
IV/QML[3]	0.03240	0.188	0.191	0.15206	0.265	0.306	-0.00075	0.068	0.068	-0.00953	0.255	0.256
2SLS[1]	-0.00113	0.127	0.127	0.00211	0.146	0.146	0.00044	0.033	0.033	-0.01661	0.061	0.063
IIV/ML[1]	0.00974	0.107	0.108	-0.01504	0.108	0.109	0.00085	0.031	0.031	-0.01598	0.052	0.055
IIV/ML[2]	0.00600	0.109	0.109	-0.00727	0.108	0.109	0.00018	0.031	0.031	-0.01588	0.054	0.056
2SLS[2]	-0.00216	0.161	0.161	0.21816	0.315	0.384	-0.00451	0.072	0.072	-0.00756	0.265	0.265
IIV/ML[3]	0.00911	0.150	0.151	0.18861	0.305	0.359	-0.00266	0.070	0.070	-0.00811	0.262	0.262
LIML	0.01021	0.107	0.108	-0.01482	0.108	0.109	0.00043	0.031	0.031	-0.01549	0.053	0.055

		β_1	γ_1
Asymptotic standard errors	2SLS[1]	0.0325	0.0628
	IIV/ML[1]	0.0281	0.0543
	LIML	0.0278	0.0537

Table 5 - Estimated biases, standard deviations and RMSE's for Experiment C, T = 50

	$\sigma_\varepsilon^2 = 0.500$	$\sigma_\eta^2 = 0.500$	$\beta_1 = 0.200$	$\gamma_1 = 1.000$
GLS	-0.12197	0.02450	0.09825	-0.08556
	0.186 0.223	0.282 0.283	0.041 0.107	0.114 0.142
IV/QML[1]	0.01124	-0.02228	0.01108	-0.02487
	0.246 0.247	0.320 0.320	0.051 0.052	0.122 0.125
IV/QML[2]	0.00040	-0.00955	0.00730	-0.01034
	0.223 0.223	0.283 0.283	0.054 0.054	0.124 0.124
IV/QML[3]	0.17828	-0.02210	0.01852	-0.07559
	0.405 0.443	0.361 0.361	0.107 0.108	0.266 0.277
2SLS[1]	-0.02701	0.03113	0.00204	-0.01000
	0.272 0.274	0.355 0.356	0.056 0.056	0.137 0.138
IIV/ML[1]	0.02026	-0.03813	0.00223	-0.01460
	0.216 0.217	0.249 0.252	0.050 0.050	0.122 0.123
IIV/ML[2]	-0.00189	-0.00862	0.00331	-0.00899
	0.211 0.211	0.246 0.246	0.054 0.054	0.126 0.126
2SLS[2]	0.04760	0.17870	0.01830	-0.07324
	0.386 0.389	0.497 0.529	0.118 0.119	0.329 0.337
IIV/ML[3]	0.12473	0.05159	0.01379	-0.06883
	0.351 0.372	0.348 0.352	0.114 0.114	0.293 0.301
LIML	0.01194	-0.03861	0.00164	-0.01166
	0.221 0.222	0.237 0.241	0.047 0.047	0.114 0.115

Asymptotic standard errors	2SLS[1]	0.0650	0.1257
	IIV/ML[1]	0.0562	0.1086
	LIML	0.0531	0.1026

Table 6 - Estimated biases, standard deviations and RMSE's for Experiment C, T = 200

	$\sigma_\epsilon^2 = 0.500$		$\sigma_\eta^2 = 0.500$		$\beta_1 = 0.200$		$\gamma_1 = 1.000$	
GLS	-0.13778		0.04631		0.10645		-0.06518	
	0.085	0.162	0.132	0.140	0.021	0.109	0.048	0.081
IV/QML[1]	0.00798		-0.01519		0.00270		-0.01731	
	0.119	0.119	0.141	0.142	0.032	0.032	0.052	0.055
IV/QML[2]	0.00611		-0.01072		0.00113		-0.01612	
	0.113	0.113	0.123	0.124	0.032	0.032	0.053	0.056
IV/QML[3]	0.03121		0.14903		-0.00430		-0.00807	
	0.175	0.178	0.274	0.312	0.068	0.068	0.257	0.257
2SLS[1]	-0.00142		-0.00067		0.00041		-0.01645	
	0.127	0.127	0.146	0.146	0.033	0.033	0.060	0.063
IIV/ML[1]	0.00932		-0.01760		0.00081		-0.01589	
	0.107	0.108	0.108	0.110	0.032	0.032	0.052	0.055
IIV/ML[2]	0.00559		-0.01015		0.00023		-0.01573	
	0.109	0.109	0.108	0.109	0.032	0.032	0.053	0.056
2SLS[2]	0.00043		0.20463		-0.00820		-0.00680	
	0.149	0.149	0.302	0.366	0.071	0.071	0.265	0.265
IIV/ML[3]	0.01083		0.18072		-0.00660		-0.00591	
	0.141	0.142	0.301	0.351	0.069	0.070	0.264	0.264
LIML	0.00255		-0.01510		0.00169		-0.01392	
	0.101	0.101	0.100	0.101	0.030	0.030	0.051	0.053

		β_1	γ_1
Asymptotic standard errors	2SLS[1]	0.0325	0.0628
	IIV/ML[1]	0.0281	0.0543
	LIML	0.0266	0.0513

Table 7 - Estimated biases, standard deviations and RMSE's for Experiment D, T = 50

	$\sigma_\varepsilon^2 = 0.600$	$\sigma_\eta^2 = 0.300$	$\beta_1 = 0.200$	$\gamma_1 = 1.000$
GLS	-0.10639 0.188 0.216	0.02544 0.208 0.210	0.07611 0.037 0.085	-0.08045 0.104 0.132
IV/QML[1]	0.01039 0.222 0.222	-0.01393 0.227 0.228	0.01018 0.046 0.047	-0.02875 0.110 0.113
IV/QML[2]	-0.01110 0.214 0.214	0.01500 0.209 0.209	0.00658 0.050 0.050	-0.01032 0.115 0.115
IV/QML[3]	0.12496 0.379 0.399	0.00032 0.315 0.315	0.01475 0.078 0.079	-0.05643 0.196 0.205
2SLS[1]	-0.03551 0.284 0.286	0.05182 0.339 0.343	0.00095 0.055 0.055	-0.00918 0.136 0.136
IIV/ML[1]	0.02619 0.219 0.221	-0.03127 0.173 0.176	0.00124 0.045 0.045	-0.01531 0.110 0.111
IIV/ML[2]	-0.00755 0.206 0.206	0.01140 0.175 0.175	0.00250 0.051 0.051	-0.00808 0.119 0.119
2SLS[2]	0.03002 0.421 0.422	0.17192 0.447 0.479	0.01248 0.101 0.102	-0.04395 0.282 0.286
IIV/ML[3]	0.11103 0.315 0.334	0.02054 0.241 0.241	0.00760 0.087 0.087	-0.03920 0.217 0.221
LIML	0.04547 0.225 0.230	-0.03024 0.177 0.179	-0.00403 0.051 0.051	-0.00819 0.124 0.125
Asymptotic standard errors		2SLS[1] IIV/ML[1] LIML	0.0650 0.0504 0.0499	0.1257 0.0974 0.0965

Table 8 - Estimated biases, standard deviations and RMSE's for Experiment A', T = 50

	$\sigma_\varepsilon^2 = 0.500$		$\sigma_\eta^2 = 0.500$		$\beta_1 = 0.200$		$\delta = 0.500$		$\gamma_1 = 1.000$	
	-0.23436		0.04252		0.12214		-0.08136		-0.05555	
GLS										
	0.152	0.280	0.237	0.241	0.036	0.128	0.055	0.099	0.121	0.133
	-0.14920		0.12038		0.03292		-0.04209		-0.02682	
IV/QML[1]										
	0.323	0.356	0.410	0.427	0.055	0.064	0.078	0.089	0.141	0.144
	-0.16948		0.14838		0.02856		-0.04554		-0.02242	
IV/QML[2]										
	0.281	0.329	0.398	0.425	0.057	0.064	0.077	0.090	0.135	0.137
	0.14961		0.09049		0.02667		-0.00652		0.01013	
IV/QML[3]										
	0.550	0.570	0.648	0.654	0.106	0.109	0.088	0.088	0.323	0.323
	-0.09561		0.06463		0.02612		-0.04538		-0.01719	
2SLS[1]										
	0.377	0.389	0.424	0.429	0.070	0.075	0.127	0.135	0.148	0.149
	-0.00617		-0.04788		0.01509		-0.01650		-0.01692	
IIV/ML[1]										
	0.329	0.329	0.275	0.279	0.062	0.064	0.072	0.074	0.138	0.139
	-0.05092		-0.01019		0.02057		-0.02286		-0.02739	
IIV/ML[2]										
	0.284	0.288	0.275	0.276	0.061	0.065	0.074	0.077	0.139	0.142
	0.17037		0.11827		0.02034		0.02754		-0.07046	
2SLS[2]										
	0.500	0.529	0.689	0.699	0.113	0.115	0.142	0.145	0.418	0.424
	0.19785		0.11125		0.01252		-0.01062		0.04698	
IIV/ML[3]										
	0.545	0.580	0.672	0.681	0.121	0.122	0.112	0.113	0.357	0.360
	-0.00310		0.38073		0.00627		-0.01702		0.01477	
LIML										
	0.515	0.515	0.990	1.061	0.060	0.061	0.132	0.133	0.142	0.143

Table 9 - Estimated biases, standard deviations and RMSE's for Experiment A', T = 200

	$\sigma_\epsilon^2 = 0.500$	$\sigma_\eta^2 = 0.500$	$\beta_1 = 0.200$	$\delta = 0.500$	$\gamma_1 = 1.000$
GLS	-0.19696	0.03769	0.12148	-0.05599	-0.03809
	0.079 0.213	0.127 0.132	0.015 0.123	0.028 0.063	0.053 0.066
IV/QML[1]	-0.09623	0.10816	0.00612	-0.01672	0.00530
	0.133 0.165	0.211 0.238	0.024 0.025	0.035 0.039	0.063 0.063
IV/QML[2]	-0.05910	0.05851	0.00472	-0.00934	-0.00472
	0.128 0.141	0.178 0.187	0.026 0.026	0.034 0.036	0.064 0.064
IV/QML[3]	-0.02670	0.28384	0.00731	-0.00464	0.06085
	0.267 0.268	0.579 0.645	0.080 0.081	0.129 0.129	0.222 0.230
2SLS[1]	-0.02627	0.01728	0.00542	-0.00932	-0.00476
	0.174 0.176	0.208 0.209	0.028 0.029	0.055 0.055	0.070 0.070
IIV/ML[1]	0.00255	-0.01651	0.00146	-0.00050	-0.00139
	0.125 0.125	0.137 0.138	0.024 0.025	0.030 0.030	0.063 0.063
IIV/ML[2]	-0.00020	-0.01476	0.00303	0.00108	-0.00758
	0.132 0.132	0.141 0.142	0.026 0.026	0.034 0.034	0.064 0.065
2SLS[2]	0.26159	0.32176	-0.02355	0.10303	-0.11751
	0.454 0.525	0.824 0.885	0.079 0.082	0.196 0.222	0.365 0.384
IIV/ML[3]	0.02946	0.24897	0.00390	-0.00297	0.06171
	0.324 0.326	0.572 0.624	0.083 0.084	0.143 0.143	0.255 0.262
LIML	-0.02429	0.47919	0.00279	-0.00112	0.00463
	0.172 0.173	0.758 0.898	0.032 0.032	0.044 0.044	0.102 0.102

Table 10 - Estimated biases, standard deviations and RMSE's for Experiment C', T = 50

	$\sigma_\varepsilon^2 = 0.500$		$\sigma_\eta^2 = 0.500$		$\beta_1 = 0.200$		$\delta = 0.500$		$\gamma_1 = 1.000$	
	-0.22268		0.10812		0.10959		-0.08120		-0.04979	
GLS	0.183	0.289	0.276	0.297	0.045	0.119	0.062	0.102	0.124	0.134
	-0.14938		0.12936		0.03000		-0.04621		-0.01756	
IV/QML[1]	0.286	0.323	0.390	0.411	0.056	0.064	0.077	0.090	0.138	0.139
	-0.14514		0.12406		0.02570		-0.04498		-0.02055	
IV/QML[2]	0.271	0.308	0.348	0.369	0.060	0.066	0.077	0.089	0.137	0.139
	0.15922		0.06275		0.00967		-0.00643		0.04059	
IV/QML[3]	0.478	0.504	0.466	0.471	0.111	0.112	0.088	0.089	0.302	0.305
	-0.09055		0.06402		0.02456		-0.04423		-0.01646	
2SLS[1]	0.355	0.366	0.414	0.419	0.067	0.072	0.125	0.132	0.147	0.148
	-0.01986		-0.02892		0.01538		-0.02318		-0.01162	
IIV/ML[1]	0.316	0.317	0.278	0.280	0.062	0.064	0.073	0.076	0.138	0.138
	-0.06182		0.00955		0.02046		-0.02934		-0.02358	
IIV/ML[2]	0.271	0.278	0.278	0.278	0.061	0.065	0.073	0.079	0.139	0.141
	0.18887		0.13490		-0.00038		0.03395		-0.04310	
2SLS[2]	0.528	0.561	0.631	0.645	0.119	0.119	0.141	0.145	0.407	0.410
	0.19562		0.06507		-0.00094		-0.00729		0.06880	
IIV/ML[3]	0.500	0.537	0.475	0.479	0.116	0.116	0.103	0.104	0.345	0.352
	-0.08790		0.46775		0.01241		-0.03388		0.01447	
LIML	0.462	0.470	1.067	1.166	0.060	0.061	0.150	0.154	0.131	0.132

Table 11 - Estimated biases, standard deviations and RMSE's for Experiment C', T = 200

	$\sigma_\varepsilon^2 = 0.500$		$\sigma_\eta^2 = 0.500$		$\beta_1 = 0.200$		$\delta = 0.500$		$\gamma_1 = 1.000$	
	-0.18731		0.10717		0.10728		-0.05535		-0.02976	
GLS										
	0.090	0.209	0.145	0.181	0.018	0.109	0.029	0.063	0.056	0.064
	-0.07903		0.08485		0.00569		-0.01906		0.00891	
IV/QML[1]										
	0.127	0.150	0.179	0.198	0.025	0.026	0.034	0.039	0.062	0.063
	-0.05346		0.05220		0.00462		-0.01414		-0.00254	
IV/QML[2]										
	0.128	0.139	0.160	0.168	0.026	0.027	0.035	0.037	0.064	0.064
	0.04888		0.13835		0.00433		-0.00426		0.07436	
IV/QML[3]										
	0.238	0.243	0.232	0.270	0.082	0.083	0.131	0.131	0.178	0.193
	-0.02357		0.01916		0.00503		-0.00807		-0.00482	
2SLS[1]										
	0.174	0.176	0.205	0.206	0.028	0.029	0.056	0.056	0.070	0.070
	-0.01139		-0.00145		0.00250		-0.00796		0.00449	
IIV/ML[1]										
	0.126	0.126	0.134	0.134	0.025	0.025	0.031	0.032	0.062	0.062
	-0.01514		0.00404		0.00347		-0.00736		-0.00442	
IIV/ML[2]										
	0.130	0.131	0.140	0.140	0.027	0.027	0.034	0.035	0.064	0.064
	0.27976		0.42809		-0.03283		0.10215		-0.09191	
2SLS[2]										
	0.453	0.534	1.418	1.482	0.084	0.090	0.233	0.255	0.437	0.447
	0.01514		0.23796		0.00509		-0.01138		0.09536	
IIV/ML[3]										
	0.307	0.308	0.448	0.508	0.085	0.085	0.152	0.152	0.228	0.248
	-0.02286		0.46751		0.00185		0.00040		0.00652	
LIML										
	0.150	0.152	0.672	0.820	0.030	0.030	0.044	0.044	0.099	0.100

Table 12 - Percentiles and first ID for $\hat{\sigma}_\varepsilon{}^2$ for experiment A, T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.0000	0.0000	0.0675	0.0783
5	0.0983	0.0527	0.2095	0.2078
10	0.1657	0.1651	0.2758	0.2806
25	0.2289	0.3335	0.3729	0.3819
50	0.3462	0.5098	0.4958	0.5208
75	0.4920	0.6679	0.6121	0.6449
90	0.5877	0.8017	0.8581	0.8742
95	0.6519	1.0677	0.9781	1.0374
maximum	0.8720	1.3239	1.1843	2.0586
ID			0.5823	0.5936
ID(theor.)			0.5743	0.7077

Table 13 - Percentiles and first ID for $\hat{\sigma}_\eta{}^2$ for experiment A, T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	6.224E-5	6.276E-5	3.122E-5	1E-8
5	0.0644	6.493E-3	0.0856	0.0809
10	0.1420	0.0716	0.1451	0.1351
25	0.2767	0.2260	0.2821	0.2770
50	0.4343	0.4219	0.4396	0.4341
75	0.6185	0.7048	0.6066	0.6175
90	0.8137	1.1028	0.8461	0.8975
95	0.9186	1.2452	0.9869	1.0566
maximum	1.1048	1.4830	1.5056	1.7695
ID			0.7010	0.7624
ID(theor.)			0.6846	0.7897

Table 14 - Percentiles and first ID for $\hat{\beta}_1$ for experiment A, T=50

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
			MODEL	
minimum	0.2063	0.0662	0.0659	-5.675E-3
5	0.2557	0.1338	0.1306	0.0972
10	0.2654	0.1536	0.1465	0.1273
25	0.2884	0.1794	0.1684	0.1532
50	0.3153	0.2154	0.2051	0.1943
75	0.3425	0.2529	0.2318	0.2292
90	0.3594	0.2842	0.2673	0.2640
95	0.3643	0.2990	0.2802	0.2783
maximum	0.3910	0.3067	0.3064	0.3033
ID			0.1208	0.1367
ID(theor.)			0.1231	0.1461

Table 15 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment A, T=50

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
			MODEL	
minimum	0.6643	0.6561	0.7158	0.7146
5	0.7256	0.7874	0.7946	0.7951
10	0.7700	0.8176	0.8256	0.8236
25	0.8278	0.8708	0.8911	0.8967
50	0.8901	0.9602	0.9727	0.9713
75	0.9751	1.0554	1.0727	1.0900
90	1.0504	1.1424	1.1473	1.1633
95	1.0804	1.1731	1.1891	1.2411
maximum	1.1788	1.2725	1.3490	1.3454
ID			0.3217	0.3397
ID(theor.)			0.3128	0.3410

Table 16 - Percentiles and first ID for $\hat{\sigma}_{\varepsilon}^2$ for experiment A, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1256	0.1573	0.2734	0.2734
5	0.2129	0.2875	0.3373	0.3270
10	0.2591	0.3412	0.3800	0.3752
25	0.2954	0.4324	0.4390	0.4382
50	0.3380	0.5079	0.5018	0.5094
75	0.3951	0.5627	0.5683	0.5710
90	0.4449	0.6760	0.6547	0.6565
95	0.4807	0.7423	0.6866	0.7001
maximum	0.5904	0.9981	0.9060	0.9041
ID			0.2747	0.2813
ID(theor.)			0.2795	0.2795

Table 17 - Percentiles and first ID for $\hat{\sigma}_{\eta}^2$ for experiment A, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1976	0.1451	0.1864	0.1878
5	0.3133	0.2575	0.2932	0.2902
10	0.3506	0.3018	0.3298	0.3220
25	0.3939	0.3426	0.3809	0.3825
50	0.4823	0.4842	0.4928	0.4978
75	0.5778	0.6003	0.5808	0.5855
90	0.6604	0.7278	0.6313	0.6321
95	0.6924	0.7881	0.6898	0.6982
maximum	0.3812	1.0178	0.7607	0.7538
ID			0.3015	0.3101
ID(theor.)			0.3179	0.3231

Table 18 - Percentiles and first ID for $\hat{\beta}_1$ for experiment A, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2923	0.0989	0.0969	0.0907
5	0.3000	0.1439	0.1430	0.1377
10	0.3008	0.1623	0.1592	0.1596
25	0.3108	0.1826	0.1817	0.1792
50	0.3190	0.2083	0.2046	0.2035
75	0.3303	0.2278	0.2252	0.2245
90	0.3432	0.2415	0.2367	0.2352
95	0.3514	0.2457	0.2431	0.2416
maximum	0.3609	0.2561	0.2553	0.2539
ID			0.0775	0.0756
ID(theor.)			0.0820	0.0820

Table 19 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment A, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.8254	0.8593	0.8643	0.8633
5	0.8542	0.8917	0.8948	0.8932
10	0.8682	0.9082	0.9094	0.9127
25	0.8970	0.9434	0.9446	0.9466
50	0.9279	0.9900	0.9921	0.9936
75	0.9658	1.0192	1.0198	1.0225
90	0.9821	1.0439	1.0501	1.0504
95	0.9972	1.0633	1.0677	1.0681
maximum	1.0130	1.1051	1.1018	1.1047
ID			0.1407	0.1377
ID(theor.)			0.1333	0.1333

Table 20 - Percentiles and first ID for $\hat{\sigma}_\varepsilon^2$ for experiment B, T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.0000	0.0000	0.0935	0.0961
5	0.1027	0.1168	0.2022	0.2431
10	0.1722	0.2371	0.2695	0.2697
25	0.2576	0.3409	0.3681	0.3826
50	0.3929	0.5089	0.4974	0.5050
75	0.5534	0.6465	0.6188	0.6431
90	0.6358	0.8005	0.8577	0.8685
95	0.7210	1.0081	0.9603	1.0054
maximum	0.9901	1.2686	1.2375	1.7751
ID			0.5882	0.5988
ID(theor.)			0.5769	0.6641

Table 21 - Percentiles and first ID for $\hat{\sigma}_\eta^2$ for experiment B, T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	6.252E-5	6.31E-5	3.106E-5	1E-8
5	0.0973	0.0446	0.0810	0.0787
10	0.1673	0.0896	0.1637	0.1379
25	0.3087	0.2543	0.2985	0.2798
50	0.5062	0.4285	0.4462	0.4553
75	0.6810	0.6379	0.6045	0.6572
90	0.8992	1.0181	0.8226	0.8289
95	1.0086	1.1405	0.9257	0.9118
maximum	1.4105	1.4027	1.2725	1.3966
ID			0.6589	0.6910
ID(theor.)			0.6564	0.6820

Table 22 - Percentiles and first ID for $\hat{\beta}_1$ for experiment B, T=50

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1947	0.0718	0.0708	-0.0255
5	0.2209	0.1330	0.1313	0.1059
10	0.2314	0.1507	0.1435	0.1307
25	0.2558	0.1723	0.1680	0.1610
50	0.2870	0.2149	0.2013	0.1931
75	0.3141	0.2452	0.2328	0.2273
90	0.3327	0.2766	0.2657	0.2617
95	0.3430	0.2916	0.2837	0.2798
maximum	0.3575	0.3137	0.3134	0.3129
ID			0.1222	0.1310
ID(theor.)			0.1256	0.1385

Table 23 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment B, T=50

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.6669	0.6832	0.7123	0.6655
5	0.7356	0.7954	0.7994	0.8047
10	0.7803	0.8249	0.8308	0.8326
25	0.8410	0.8798	0.8969	0.8991
50	0.9180	0.9712	0.9745	0.9915
75	1.0012	1.0674	1.0717	1.0924
90	1.0917	1.1503	1.1509	1.1617
95	1.1131	1.1721	1.1968	1.2398
maximum	1.2148	1.2819	1.3419	1.3469
ID			0.3201	0.3291
ID(theor.)			0.3128	0.3410

Table 24 - Percentiles and first ID for $\hat{\sigma}_\epsilon^2$ for experiment B, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1436	0.2057	0.2658	0.2628
5	0.2454	0.3250	0.3354	0.3205
10	0.2932	0.3616	0.3765	0.3732
25	0.3268	0.4332	0.4395	0.4435
50	0.3705	0.5024	0.5028	0.5115
75	0.4418	0.5678	0.5678	0.5695
90	0.4907	0.6683	0.6491	0.6436
95	0.5419	0.7234	0.6818	0.6753
maximum	0.6046	0.9577	0.8946	0.8985
ID			0.2726	0.2704
ID(theor.)			0.2743	0.2743

Table 25 - Percentiles and first ID for $\hat{\sigma}_\eta^2$ for experiment B, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2341	0.1966	0.2099	0.2168
5	0.3496	0.2866	0.3063	0.2986
10	0.3930	0.3344	0.3391	0.3372
25	0.4423	0.3832	0.3893	0.3960
50	0.5550	0.4823	0.4986	0.4929
75	0.6419	0.5821	0.5623	0.5670
90	0.7100	0.6627	0.5976	0.6026
95	0.7477	0.6951	0.6375	0.6543
maximum	0.9266	0.9664	0.7730	0.7870
ID			0.2585	0.2654
ID(theor.)			0.2769	0.2769

Table 26 - Percentiles and first ID for $\hat{\beta}_1$ for experiment B, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2468	0.1012	0.1005	0.0999
5	0.2590	0.1447	0.1440	0.1416
10	0.2645	0.1626	0.1607	0.1612
25	0.2754	0.1811	0.1802	0.1792
50	0.2864	0.2043	0.2036	0.2016
75	0.2991	0.2268	0.2249	0.2272
90	0.3121	0.2396	0.2371	0.2352
95	0.3199	0.2446	0.2445	0.2417
maximum	0.3273	0.2581	0.2577	0.2562
ID			0.0764	0.0740
ID(theor.)			0.0795	0.0795

Table 27 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment B, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.8339	0.8590	0.8633	0.8611
5	0.8623	0.8924	0.8941	0.8951
10	0.8789	0.9092	0.9100	0.9140
25	0.9078	0.9435	0.9437	0.9436
50	0.9442	0.9924	0.9934	0.9915
75	0.9795	1.0203	1.0207	1.0232
90	1.0037	1.0464	1.0506	1.0487
95	1.0206	1.0652	1.0680	1.0708
maximum	1.0378	1.0975	1.0954	1.0894
ID			0.1406	0.1347
ID(theor.)			0.1333	0.1359

Table 28 - Percentiles and first ID for $\hat{\sigma}_{\varepsilon}^2$ for experiment C, T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.0000	0.0000	0.0947	0.1326
5	0.0748	0.1118	0.2083	0.2038
10	0.1393	0.2262	0.2803	0.2593
25	0.2561	0.3700	0.3710	0.3638
50	0.3648	0.5147	0.4922	0.5078
75	0.5234	0.6499	0.6270	0.6141
90	0.5933	0.7694	0.8634	0.7605
95	0.6114	1.0034	0.9608	0.9361
maximum	1.0677	1.3131	1.1473	1.3751
ID			0.5831	0.5012
ID(theor.)			0.5538	0.5666

Table 29 - Percentiles and first ID for $\hat{\sigma}_{\eta}^2$ for experiment C, T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	6.252E-5	6.279E-5	3.14E-5	0.0245
5	0.1180	0.0615	0.0896	0.1044
10	0.1824	0.1108	0.1681	0.1927
25	0.3033	0.2444	0.2847	0.2902
50	0.4821	0.4089	0.4184	0.4393
75	0.6966	0.6398	0.6037	0.6011
90	0.8921	0.9038	0.8235	0.7915
95	1.0414	1.1362	0.9004	0.9077
maximum	1.2836	1.5110	1.3033	1.2616
ID			0.6554	0.5988
ID(theor.)			0.6384	0.6077

Table 30 - Percentiles and first ID for $\hat{\beta}_1$ for experiment C, T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1979	0.0555	0.0548	0.0753
5	0.2274	0.1324	0.1318	0.1300
10	0.2482	0.1498	0.1454	0.1487
25	0.2716	0.1736	0.1660	0.1659
50	0.2953	0.2149	0.2035	0.2048
75	0.3263	0.2498	0.2342	0.2301
90	0.3523	0.2838	0.2673	0.2651
95	0.3693	0.2925	0.2852	0.2870
maximum	0.3829	0.3327	0.3310	0.3225
ID			0.1219	0.1164
ID(theor.)			0.1282	0.1205

Table 31 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment C, T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.6598	0.6830	0.7162	0.7312
5	0.7235	0.7892	0.7995	0.8046
10	0.7756	0.8105	0.8180	0.8292
25	0.8390	0.8838	0.8938	0.9084
50	0.9060	0.9658	0.9706	0.9855
75	0.9996	1.0598	1.0783	1.0724
90	1.0666	1.1376	1.1460	1.1212
95	1.1071	1.1605	1.1949	1.1874
maximum	1.1866	1.2981	1.3421	1.2906
ID			0.3280	0.2920
ID(theor.)			0.3128	0.2923

Table 32 - Percentiles and first ID for $\hat{\sigma}_{\epsilon}^2$ for experiment C, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1352	0.1995	0.2635	0.2607
5	0.2197	0.3190	0.3345	0.3379
10	0.2772	0.3630	0.3742	0.3853
25	0.3135	0.4343	0.4420	0.4436
50	0.3502	0.5070	0.4969	0.4924
75	0.4121	0.5678	0.5703	0.5587
90	0.4796	0.6712	0.6489	0.6171
95	0.5139	0.7200	0.6849	0.6857
maximum	0.5476	0.9198	0.8714	0.8673
ID			0.2747	0.2318
ID(theor.)			0.2743	0.2590

Table 33 - Percentiles and first ID for $\hat{\sigma}_{\eta}^2$ for experiment C, T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2350	0.1923	0.2016	0.2050
5	0.3396	0.2854	0.3083	0.3343
10	0.3961	0.3250	0.3358	0.3494
25	0.4286	0.3783	0.3910	0.4041
50	0.5474	0.4769	0.4906	0.4964
75	0.6413	0.5871	0.5656	0.5551
90	0.7100	0.6710	0.5972	0.6050
95	0.7418	0.7013	0.6349	0.6415
maximum	0.9257	0.9740	0.7726	0.7518
ID			0.2614	0.2556
ID(theor.)			0.2769	0.2564

Table 34 - Percentiles and first ID for $\hat{\beta}_1$ for experiment C, T=200

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
	MODEL			
minimum	0.2584	0.0985	0.0976	0.1096
5	0.2761	0.1456	0.1449	0.1536
10	0.2802	0.1626	0.1609	0.1619
25	0.2932	0.1810	0.1796	0.1816
50	0.3042	0.2074	0.2048	0.2038
75	0.3201	0.2275	0.2253	0.2254
90	0.3361	0.2399	0.2386	0.2377
95	0.3434	0.2458	0.2449	0.2411
maximum	0.3521	0.2570	0.2563	0.2587
ID			0.0777	0.0758
ID(theor.)			0.0820	0.0769

Table 35 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment C, T=200

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
	MODEL			
minimum	0.8261	0.8584	0.8628	0.8616
5	0.8511	0.8899	0.8950	0.9006
10	0.8725	0.9088	0.9097	0.9139
25	0.9036	0.9458	0.9454	0.9498
50	0.9398	0.9929	0.9938	0.9942
75	0.9689	1.0192	1.0197	1.0179
90	0.9933	1.0461	1.0499	1.0480
95	1.0129	1.0663	1.0679	1.0710
maximum	1.0309	1.1028	1.1012	1.1037
ID			0.1402	0.1341
ID(theor.)			0.1333	0.1308

Table 36 - Percentiles for $\hat{\sigma}_\epsilon^2$ for experiment A', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.0000	0.0000	0.0000	1.000E-8
5	0.0000	0.0000	0.1268	1.000E-8
10	0.0818	0.0000	0.1747	1.762E-3
25	0.1632	0.11832	0.2537	0.1344
50	0.2545	0.2994	0.4107	0.3298
75	0.3597	0.5063	0.6054	0.6812
90	0.4739	0.6558	0.9467	1.2178
95	0.5059	1.1748	1.0931	1.6415
maximum	0.8242	1.3842	1.6559	2.3128

Table 37 - Percentiles for $\hat{\sigma}_\eta^2$ for experiment A', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	1.2486E-4	1.486E-4	1.0997E-4	8.51161E-3
5	0.1747	0.0616	0.1031	0.0846
10	0.2809	0.1491	0.1424	0.1861
25	0.3850	0.3268	0.2578	0.2565
50	0.5025	0.5426	0.3999	0.5464
75	0.7072	0.8809	0.6078	1.1624
90	0.8681	1.1761	0.8994	1.8702
95	1.0230	1.3914	0.9743	3.1720
maximum	1.1307	1.7970	1.2745	6.1585

Table 38 - Percentiles for $\hat{\beta}_1$ for experiment A', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2247	0.0158	6.49778E-3	0.0648
5	0.2703	0.1587	0.1114	0.0846
10	0.2821	0.1707	0.1406	0.1229
25	0.2963	0.2012	0.1758	0.1654
50	0.3190	0.2426	0.2300	0.2104
75	0.3460	0.2641	0.2544	0.2573
90	0.3714	0.2950	0.2832	0.2782
95	0.3829	0.3186	0.3002	0.2891
maximum	0.4056	0.3506	0.3438	0.3401

Table 39 - Percentiles for $\hat{\delta}_1$ for experiment A', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2795	0.2753	0.3048	0.1774
5	0.3123	0.3277	0.3724	0.2895
10	0.3368	0.3532	0.3874	0.3129
25	0.3881	0.4168	0.4438	0.3672
50	0.4251	0.4514	0.4735	0.4718
75	0.4537	0.5072	0.5374	0.6006
90	0.4889	0.5670	0.5741	0.6621
95	0.5041	0.5880	0.6038	0.6813
maximum	0.5412	0.6388	0.6743	0.7285

Table 40 - Percentiles for $\hat{\gamma}_1$ for experiment A', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.6674	0.6730	0.6837	0.6610
5	0.7474	0.7097	0.7409	0.7417
10	0.7711	0.7742	0.7936	0.8360
25	0.8585	0.8942	0.8991	0.9251
50	0.9391	0.9804	0.9856	1.0153
75	1.0294	1.0511	1.0605	1.1319
90	1.0963	1.1351	1.1415	1.2044
95	1.1583	1.2082	1.2352	1.2381
maximum	1.2567	1.3691	1.4081	1.2930

Table 41 - Percentiles for $\hat{\sigma}_{\varepsilon}^2$ for experiment A', T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1194	0.1089	0.1858	0.0230
5	0.1842	0.2086	0.3231	0.2165
10	0.2129	0.2319	0.3480	0.2421
25	0.2533	0.3181	0.4256	0.3613
50	0.2936	0.3885	0.4948	0.4710
75	0.3573	0.4889	0.5694	0.5870
90	0.4117	0.5903	0.6546	0.6975
95	0.4415	0.6261	0.7071	0.7589
maximum	0.5068	0.7995	1.0138	0.9232

Table 42 - Percentiles for $\hat{\sigma}_{\eta}^2$ for experiment A', T=200

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2328	0.2138	0.2184	0.2472
5	0.3516	0.3042	0.2612	0.3248
10	0.3859	0.3735	0.3261	0.3870
25	0.4604	0.4428	0.3894	0.5157
50	0.5259	0.5841	0.4753	0.7188
75	0.6097	0.7464	0.5498	1.2011
90	0.7104	0.8835	0.6675	1.7448
95	0.7724	0.9965	0.7460	2.8731
maximum	0.8516	1.3278	0.9200	4.4243

Table 43 - Percentiles for $\hat{\beta}_1$ for experiment A', T=200

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
		MODEL		
minimum	0.2767	0.1374	0.1373	0.1056
5	0.2963	0.1624	0.1545	0.1488
10	0.2999	0.1755	0.1708	0.1578
25	0.3109	0.1918	0.1885	0.1854
50	0.3222	0.2052	0.2027	0.2062
75	0.3328	0.2234	0.2180	0.2253
90	0.3408	0.2385	0.2319	0.2399
95	0.3444	0.2441	0.2402	0.2504
maximum	0.3527	0.2641	0.2559	0.2802

Table 44 - Percentiles for $\hat{\delta}_1$ for experiment A', T=200

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
		MODEL		
minimum	0.3803	0.3939	0.4222	0.3796
5	0.3839	0.4212	0.4419	0.4104
10	0.4120	0.4334	0.4610	0.4311
25	0.4243	0.4605	0.4805	0.4732
50	0.4444	0.4873	0.5031	0.5090
75	0.4644	0.5075	0.5190	0.5322
90	0.4753	0.5256	0.5333	0.5462
95	0.4827	0.5401	0.5454	0.5507
maximum	0.5186	0.5800	0.5710	0.5820

Table 45 - Percentiles for $\hat{\gamma}_1$ for experiment A', T=200

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
		MODEL		
minimum	0.8352	0.8491	0.8333	0.7699
5	0.8787	0.9086	0.8980	0.8071
10	0.8929	0.9213	0.9165	0.8468
25	0.9228	0.9646	0.9535	0.9431
50	0.9546	1.0039	0.9989	1.0179
75	0.9971	1.0370	1.0305	1.0766
90	1.0378	1.0918	1.0846	1.1270
95	1.0532	1.1118	1.1044	1.1578
maximum	1.1039	1.1926	1.1896	1.2211

Table 46 - Percentiles for $\hat{\sigma}_{\varepsilon}^2$ for experiment C', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.0000	0.0000	0.0000	1.000E-8
5	0.0127	0.0000	0.1085	1.000E-8
10	0.0454	5.550E-3	0.1704	1.000E-8
25	0.1460	0.1284	0.2611	0.0563
50	0.2453	0.3016	0.3900	0.3074
75	0.3769	0.4757	0.6222	0.5216
90	0.5226	0.6644	0.9396	1.0251
95	0.5818	1.0192	1.0783	1.2576
maximum	0.9950	1.1403	1.7989	2.7206

Table 47 - Percentiles for $\hat{\sigma}_{\eta}^2$ for experiment C', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	1.9132E-4	1.9704E-4	1.3983E-4	0.0101
5	0.2161	0.1065	0.0546	0.1570
10	0.2755	0.1921	0.1407	0.2128
25	0.3891	0.3239	0.2528	0.3106
50	0.5646	0.5792	0.4284	0.6094
75	0.7946	0.8456	0.6120	1.1861
90	1.0372	1.1554	0.8710	2.2638
95	1.1782	1.3426	1.0000	3.3483
maximum	1.2044	1.9310	1.2358	6.1755

Table 48 - Percentiles for $\hat{\beta}_1$ for experiment C', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1636	0.0574	5.53021E-3	0.0739
5	0.2443	0.1408	0.1083	0.1301
10	0.2639	0.1608	0.1361	0.1420
25	0.2822	0.1945	0.1746	0.1691
50	0.3095	0.2406	0.2295	0.2058
75	0.3346	0.2612	0.2518	0.2587
90	0.3749	0.2996	0.2931	0.2858
95	0.3914	0.3167	0.3061	0.3152
maximum	0.4109	0.3529	0.3465	0.3512

Table 49 - Percentiles for $\hat{\delta}_1$ for experiment C', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2687	0.2618	0.3075	0.1693
5	0.3091	0.3203	0.3558	0.2391
10	0.3321	0.3533	0.3845	0.2562
25	0.3879	0.4117	0.4291	0.3364
50	0.4225	0.4456	0.4701	0.4700
75	0.4572	0.5015	0.5260	0.5992
90	0.4937	0.5690	0.5756	0.6629
95	0.5159	0.5820	0.5942	0.6873
maximum	0.5641	0.6402	0.6545	0.7388

Table 50 - Percentiles for $\hat{\gamma}_1$ for experiment C', T=50

percentages	MODEL GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.6587	0.6721	0.6738	0.6970
5	0.7388	0.7431	0.7435	0.7793
10	0.7678	0.7925	0.7946	0.8394
25	0.8843	0.9047	0.9060	0.9362
50	0.9546	0.9922	0.9949	1.0026
75	1.0294	1.0646	1.0722	1.1090
90	1.0875	1.1458	1.1413	1.2000
95	1.1821	1.2146	1.2229	1.2289
maximum	1.2605	1.3940	1.4244	1.3277

Table 51 - Percentiles for $\hat{\sigma}_{\epsilon}^2$ for experiment C', T=200

percentages	MODEL			
	GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.1058	0.1210	0.1738	0.1256
5	0.1703	0.2241	0.3142	0.2292
10	0.2157	0.2556	0.3358	0.3069
25	0.2486	0.3270	0.4085	0.3820
50	0.2951	0.4107	0.4898	0.4695
75	0.3764	0.4981	0.5565	0.5634
90	0.4398	0.5880	0.6320	0.6589
95	0.4610	0.6193	0.6908	0.7493
maximum	0.5277	0.7974	0.9902	0.9207

Table 52 - Percentiles for $\hat{\sigma}_{\eta}^2$ for experiment C', T=200

percentages	MODEL			
	GLS	IV/QML[1]	IIV/ML[1]	LIML
minimum	0.2378	0.2147	0.2299	0.2547
5	0.3599	0.2997	0.2772	0.3551
10	0.4414	0.3796	0.3366	0.4224
25	0.5088	0.4538	0.4134	0.4978
50	0.6118	0.5760	0.4840	0.7217
75	0.6795	0.6940	0.5663	1.2456
90	0.8051	0.8342	0.6827	1.7964
95	0.8724	0.9013	0.7769	2.2773
maximum	1.0019	1.1373	0.8458	3.7368

Table 53 - Percentiles for $\hat{\beta}_1$ for experiment C', T=200

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
	MODEL			
minimum	0.2578	0.1393	0.1376	0.1267
5	0.2822	0.1589	0.1551	0.1537
10	0.2866	0.1736	0.1710	0.1619
25	0.2960	0.1909	0.1878	0.1832
50	0.3059	0.2038	0.2026	0.2024
75	0.3200	0.2220	0.2189	0.2220
90	0.3302	0.2372	0.2329	0.2478
95	0.3355	0.2453	0.2432	0.2428
maximum	0.3455	0.2663	0.2584	0.2768

Table 54 - Percentiles for $\acute{\delta}_1$ for experiment C', T=200

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
	MODEL			
minimum	0.3810	0.3962	0.4150	0.3832
5	0.3992	0.4164	0.4382	0.4147
10	0.4053	0.4347	0.4483	0.4326
25	0.4260	0.4596	0.4737	0.4750
50	0.4434	0.4831	0.4935	0.5075
75	0.4655	0.5024	0.5125	0.5356
90	0.4778	0.5190	0.5320	0.5483
95	0.4938	0.5396	0.5442	0.5580
maximum	0.5224	0.5648	0.5669	0.5716

Table 55 - Percentiles for $\hat{\gamma}_1$ for experiment C', T=200

percentages	GLS	IV/QML[1]	IIV/ML[1]	LIML
	MODEL			
minimum	0.8468	0.8576	0.8489	0.7723
5	0.8770	0.9153	0.9069	0.8255
10	0.9035	0.9234	0.9230	0.8554
25	0.9307	0.9706	0.9634	0.9479
50	0.9633	1.0054	1.0032	1.0117
75	1.0013	1.0470	1.0384	1.0782
90	1.0520	1.0965	1.0905	1.1338
95	1.0660	1.1119	1.1057	1.1737
maximum	1.1375	1.1984	1.1952	1.1942

CHAPTER 10

CONCLUSIONS

1. Introduction

The aim of the thesis was to examine the estimation of a single equation which contains a stochastic trend and is part of a system.

We looked at identification issues and extended the classical rank condition to show the role played by stochastic trends. Basically a deterministic trend in another equation in the system contributes to identification in the same way as any exogenous variables. Identification can also be achieved if the trend in another equation is stochastic, while in the equation of interest it is deterministic.

As regards estimation, a wide range of IV estimators were studied. The basis of these estimators were suggested in Harvey(1989). We found that some of these estimators have unsatisfactory properties and so cannot be recommended. The best approach is based on feasible G2SLS estimators. Such estimators can be obtained by concentrating the vector containing the regression coefficients out of the criterion function, as suggested in Harvey(1989). This estimator we called IV/QML. We suggested adopting a stepwise approach as an alternative. This estimator we called IIV/ML. The Monte Carlo experiments showed that IIV/ML performs better than IV/QML, as regard small sample properties. Moreover, IIV/ML is less time consuming to compute than IV/QML.

For a single equation efficiency can be defined in terms of the LIML estimator. We were able to work out expressions for the asymptotic variances of IIV/ML and LIML. This all was done in the frequency domain. We were also able to show that when the stochastic part of the model is homogeneous, IIV/ML is asymptotically efficient, i.e., it has the same distribution as LIML. Other cases when IIV/ML is asymptotically equivalent to LIML were noted.

Particularly interesting is the case where no stochastic trend is present in any equation other than first. Asymptotic theory for LIML is not straightforward to work out, but by using limiting arguments we succeeded in obtaining an expression for the asymptotic covariance matrix of LIML. We were then able to compare this with the asymptotic variance of IIV/ML.

The Monte Carlo experiments were not intended to be a comprehensive guide to small sample distribution. However we learned a good deal. In the homogeneous case, for small samples, we found that IIV/ML does better, relative to LIML, than asymptotic theory would suggest. In cases far from homogeneity, as is the case when no stochastic trend is present in the second equation, the superiority of LIML is apparent.

As regards computation, IIV/ML is more reliable. LIML can occasionally give non-convergence problems and implausible results, whereas IIV/ML almost always gives sensible results.

Our practical recommendation is to adopt IIV/ML with 2SLS providing initial values. There may be gains to be had from LIML in certain circumstances. However it should only be adopted if it converges to what seems to be a reasonable answer.

Although we have worked with random walk stochastic trends, extensions to local linear trends seem to be easy to handle.

We have also conducted Monte Carlo experiments for models with lagged dependent variables. The performance of LIML was somewhat disappointing. More work is needed to obtain computationally reliable procedures. However IIV/ML worked well and we are quite content to recommend it.

Other areas of future research would include constructing a computationally reliable and efficient method for FIML, the implementation of system IV estimators such as feasible G3SLS, and a thorough Monte Carlo study to analyse the performance of the estimators.

REFERENCES

Anderson, T W and H Rubin (1949), Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations, Annals of Mathematical Statistics, 20, 46-63

Bowden, R J and D A Turkington (1984), Instrumental Variables, Cambridge University Press, Cambridge

Campos, J (1986a), Instrumental Variables Estimation of Dynamic Simultaneous Systems with ARMA Errors, Review of Economics Studies, 53, 125-138

Campos, J (1986b), Finite-sample Properties of the Instrumental-Variables Estimator for Dynamic Simultaneous Equations Subsystems with ARMA Disturbances Journal of Econometrics, 32, 297-366

Deistler, M, W Dunsmuir and E J Hannan (1978), Vector Linear Time Series Models: Corrections and Extensions, Advances Applied Probability, 10, 360-372

de Jong, P (1988), The Likelihood for a State Space Model, Biometrika, 75, 165-169

de Jong P (1991), The Diffuse Kalman Filter, Annals of Statistics, 19, 1073-83

Dunsmuir, W (1979), A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed with Noise, The Annals of Statistics, 7, 490-506

Dunsmuir, W and Hannan, E J (1976), Vector Linear Time Series Models, Advances in Applied Probability, 8, 339-364

Durbin, J (1988), Maximum Likelihood Estimation of the Parameters of a System of Simultaneous Regression Equations, Econometric Theory, 4, 159-170

Engle, R F and C W J Granger (1987), Co-integration and Error Correction: Representation, Estimation, and Testing, Econometrica, 55, 251-276

Engle, R F, Hendry D F and J-F Richard (1983), Exogeneity, Econometrica, 51, 277-304

Espasa, A (1977), The Spectral Maximum Likelihood Estimation of Econometric Models with Stationary Errors, Vandenhoeck & Ruprecht:Gottingen

Fernandez, F J (1986), Estimation and Testing of Multivariate Structural Time Series Models, Unpublished Ph.D Thesis, University of London

Fernandez, F J and A C Harvey (1990), Seemingly Unrelated Time Series Equations and Test for Homogeneity, Journal of Business and Economic Statistics, 8, 71-82

Fuller, W A (1976), Introduction to Statistical Time Series, New York: John Wiley and Sons

Hall, A D and A R Pagan (1981), The LIML and Related Estimators of an Equation with Moving Average Disturbances, International Economic Review, 22, 719-730

Hannan, E J (1970), Multiple Time Series, New York: John Wiley

Hannan, E J (1971), Non-linear Time Series Regression, Journal of Applied Probability, 8, 767-780

Hannan, E J (1973), The Asymptotic Theory of Linear Time-series Models, Journal of Applied Probability, 10, 130-145

Hannan, E J and R D Terrell (1973), Multiple Equation Systems with

Stationary Errors, Econometrica, 41, 299-320

Harvey, A C (1981), Time Series Models, Deddington, Oxford: Philip Allan and Atlantic Highlands, N.J.: Humanities Press

Harvey, A C (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge: Cambridge University Press

Harvey, A C, Henry, B, Peters S and S Wren-Lewis (1986), Stochastic Trends in Dynamic Regression Models: An Application to the Output-Employment Equation,Economic Journal, 96, 975 -985

Harvey, A C and J Durbin (1986), The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling, Journal of the Royal Statistical Society, Series A, 149, 187-227

Harvey, A C and P Marshall (1991), Inter-Fuel Substitution, Technical Change and the Demand for Energy in the UK Economy, Applied Economics, 23, 1077-86

Harvey, A C, Neudecker, H and M Streibel (1991), Problem in Matrix Inversion, Econometric Theory, 7

Hausman, J A (1983), Specification and Estimation of Simultaneous Equations Models, Ch 7 in Z Griliches and M D Intriligator, (eds), Handbook of Econometrics, Volume I, Amsterdam: North Holland

Hendry, D F (1976), The Structure of Simultaneous Equations Estimators, Journal of Econometrics, 4, 51-88

Hendry, D F and R Harrison (1974), Monte Carlo Methodology and the Finite Sample Properties of Ordinary and Two-Stage Least Squares Estimators, Journal of Econometrics, 2, 151-74

Hsaio, C (1983), Identification, pp. 223-283 in Z Griliches and M D Intrilligator (eds), Handbook of Econometrics, Volume I, Amsterdam: North-Holland

Kohn, R and C F Ansley (1985), Efficient Estimation and Prediction in Time Series Regression Models, Biometrika, 72, 694-697

Kohn, R and C F Ansley (1989), A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State Space Models, Biometrika, 76, 65-79

Lahiri, K and P Schmidt (1978), On Estimation of Triangular Structural Systems, Econometrica, 46, 1217-1222

Magnus,J R and H Neudecker (1988), Matrix Differential Calculus with Applications in Statistics and Econometrics, John Wiley:Essex

Marshall P R (1990), Analysis of Cross–Section of Time Series using Structural Time Series Models, Unpublished Ph.D Thesis, University of London

Oberhofer, W and J Kmenta (1974), A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models, Econometrica, 42,579-590

Priestley, M B (1981), Spectral Analysis and Time Series, London: Academic Press

Reinsel, G (1979), FIML Estimation of the Dynamic Simultaneous Equations Model with ARMA Disturbances, Journal of Econometrics, 9, 263-281

Robinson, P M (1978), Alternative Models for Stationary Stochastic Processes, Stochastic Processes and their Application 8, 141-145

Rosenberg, B (1973), Random Coefficient Models: the Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression, Annals of Economic and Social

Measurement, 2, 399-428

Rothenberg, T J and C T Leenders (1964), Efficient Estimation of Simultaneous Equation Systems, Econometrica, 32, 57-76

Sargan, J D (1964), Wages and Prices in the United Kingdom: A Study in Econometric Methodology, in P E Hart, G Mills and J K Whitaker (eds), Econometric Analysis for NationalEconomic Planning, Butterworth, pp 25-54

Sargan, J D (1982), On Monte Carlo Estimates of Moments that are Infinite, in Advances in Econometrics, Vol 1, R L Basmann and G F Rhodes, Jr (eds), 267-99

Schweppe, F (1965), Evaluation of Likelihood Functions for Gaussian Signals, IEEE Transactions on Information Theory, 11, 61-70

Slade, M E (1989), Modelling Stochastic and Cyclical Components of Structural Change: An Application of the Kalman Filter, Journal of Econometrics, 41, 363-83

Theil, H (1961), Economic Forecasts and Policy, 2nd ed. Amsterdam: North Holland

Wickens, M R (1969), The Consistency and Efficiency of Generalized Least Squares in Simultaneous Equation Systems with Autocorrelated Errors, Econometrica, 37, 651-659