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
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

Abstract 2

Acknowledgments 4

List of Tables 10

CHAPTER 1 - INTRODUCTION 14

CHAPTER 2 - STRUCTURAL TIME SERIES MODELS 20

1. Introduction

2. State Space Form Models

   SUTSE Models

   Prediction Error Decomposition

   Kalman Filter Equations

   The Link Between Cholesky Decomposition and State Space Techniques

   Conditional Likelihood Function

3. Estimation in the frequency domain

   The Spectral Likelihood Function

   Asymptotic Information Matrix

   Maximum Likelihood Estimation

4. Asymptotic Theory

Appendix 2.1 - Optimisation Procedures
CHAPTER 3 - TIME SERIES REGRESSION MODELS

1. Introduction
2. Regression Models with Stochastic Components
   Multivariate Time Series Regression Models
   Maximum Likelihood Estimation
3. Frequency Domain Estimation
   Asymptotic Information Matrix
   Spectral Maximum Likelihood Estimates
4. Asymptotic Theory of Time Series Regression Models
Appendix 3.1 - Proof of Asymptotic Results

CHAPTER 4 - INSTRUMENTAL VARIABLE ESTIMATION

1. Introduction
2. Instrumental Variable Principle
   Serially Uncorrelated Disturbances
   Serially Correlated Disturbances
   Asymptotic Properties of the GLS-IV Analog
3. Instrumental Variable Estimation for Models with Stochastic Trends
   Comparison of Estimators
   Initial Consistent Estimator
4. Instrumental Variable Estimation in the Frequency Domain
   Asymptotic Properties
   Estimator Suggested by Hannan and Terrell
   Relative Asymptotic Efficiency of IIV/ML Compared with 2SLS

CHAPTER 5 - SIMULTANEOUS EQUATION SYSTEMS WITH STOCHASTIC TREND COMPONENTS

1. Introduction
2. Model Formulation

An Alternative Formulation

3. Identifiability

The Concept of Identifiability.

The Identification Problem in Simultaneous Equation Systems

The Identification Problem in Simultaneous Equation Systems with Stochastic Trends

Local Linear Trends

4. The Likelihood Function

CHAPTER 6 - FREQUENCY DOMAIN APPROACH OF SIMULTANEOUS EQUATION SYSTEMS WITH STOCHASTIC TREND COMPONENTS

1. Introduction

2. The Spectral Likelihood Function

The First Order Conditions

The Hessian Matrix

3. A Computational Method for Finding the Spectral ML Estimator

4. Asymptotic Information and Covariance Matrices

Asymptotic Information Matrix

Asymptotic Covariance Matrix

5. Asymptotically Efficient 2-step Estimators

A $/T$-consistent Estimator for Coefficients

A $/T$-consistent Estimator for Hyperparameters

The 2-step Estimator for Coefficients

The 2-step Estimator for Hyperparameters

Appendix 6.1 - Proof of Expression (6.3.2)

Appendix 6.2 - Asymptotic Covariance Matrix

Appendix 6.3 - $/T$-consistency of Coefficient Estimators
CHAPTER 7 - THE LIMITED INFORMATION MAXIMUM LIKELIHOOD ESTIMATOR
OF A SINGLE EQUATION IN A SIMULTANEOUS EQUATION SYSTEM
WITH STOCHASTIC TRENDS

1. Introduction
2. The LIML Estimator
3. LIML Viewed as a Special Case of FIML
4. Computational Method
5. Asymptotic Theory of LIML
6. Efficiency Comparisons
   - Explanatory Variables Following a Multivariate Random Walk
7. Stochastic Trend in First Equation Only
Appendix 7.1 - Result on Matrix Inversion
Appendix 7.2 - Asymptotic Covariance Matrix When Only the First
   Equation has a Stochastic Trend

CHAPTER 8 - AN APPLICATION TO REAL ECONOMIC TIME SERIES

1. Introduction
2. Estimation of the Model
3. Conclusions
Appendix 8.1 - Tables

CHAPTER 9 - MONTE CARLO STUDY

1. Introduction
2. The Choice of the Parameters
3. Data Generation
4. Estimation Procedures
5. Asymptotic Standard Errors

6. Discussion of the Results

Appendix 9.1 - Tables

CHAPTER 10 - CONCLUSIONS 222

REFERENCES 225
LIST OF TABLES

Appendix 8.1

Table 1 - GLS estimates for equation with stochastic trend
Table 2 - OLS estimates for first differenced equation
Table 3 - IV estimates for first differenced equation
Table 4 - IV/QML estimates with transformed instruments, case A
Table 5 - IIV/ML estimates with transformed instruments, case A
Table 6 - IV/QML estimates with untransformed instruments, case A
Table 7 - IIV/ML estimates with untransformed instruments, case A
Table 8 - IV/QML estimates with transformed instruments, case B
Table 9 - IIV/ML estimates with transformed instruments, case B
Table 10 - IV/QML estimates with untransformed instruments, case B
Table 11 - IIV/ML estimates with untransformed instruments, case B
Table 12 - IV estimates for first differenced equation, case C'
Table 13 - IV estimates for first differenced equation, case C
Table 14 - IV estimates for first differenced equation, case D'
Table 15 - IV estimates for first differenced equation, case D

Appendix 9.1

Table 1 - Estimated biases, standard deviations and RMSE's for
   Experiment A, T = 50
Table 2 - Estimated biases, standard deviations and RMSE's for
   Experiment A, T = 200
Table 3 - Estimated biases, standard deviations and RMSE's for
   Experiment B, T = 50
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Estimated biases, standard deviations and RMSE's for Experiment B, T = 200</td>
</tr>
<tr>
<td>5</td>
<td>Estimated biases, standard deviations and RMSE's for Experiment C, T = 50</td>
</tr>
<tr>
<td>6</td>
<td>Estimated biases, standard deviations and RMSE's for Experiment C, T = 200</td>
</tr>
<tr>
<td>7</td>
<td>Estimated biases, standard deviations and RMSE's for Experiment D, T = 50</td>
</tr>
<tr>
<td>8</td>
<td>Estimated biases, standard deviations and RMSE's for Experiment A', T = 50</td>
</tr>
<tr>
<td>9</td>
<td>Estimated biases, standard deviations and RMSE's for Experiment A', T = 200</td>
</tr>
<tr>
<td>10</td>
<td>Estimated biases, standard deviations and RMSE's for Experiment C', T = 50</td>
</tr>
<tr>
<td>11</td>
<td>Estimated biases, standard deviations and RMSE's for Experiment C', T = 200</td>
</tr>
<tr>
<td>12</td>
<td>Percentiles and first ID for $\hat{\sigma}_\epsilon^2$ for experiment A, T=50</td>
</tr>
<tr>
<td>13</td>
<td>Percentiles and first ID for $\hat{\sigma}_\eta^2$ for experiment A, T=50</td>
</tr>
<tr>
<td>14</td>
<td>Percentiles and first ID for $\hat{\beta}_1$ for experiment A, T=50</td>
</tr>
<tr>
<td>15</td>
<td>Percentiles and first ID for $\hat{\gamma}_1$ for experiment A, T=50</td>
</tr>
<tr>
<td>16</td>
<td>Percentiles and first ID for $\hat{\sigma}_\epsilon^2$ for experiment A, T=200</td>
</tr>
<tr>
<td>17</td>
<td>Percentiles and first ID for $\hat{\sigma}_\eta^2$ for experiment A, T=200</td>
</tr>
<tr>
<td>18</td>
<td>Percentiles and first ID for $\hat{\beta}_1$ for experiment A, T=200</td>
</tr>
<tr>
<td>19</td>
<td>Percentiles and first ID for $\hat{\gamma}_1$ for experiment A, T=200</td>
</tr>
<tr>
<td>20</td>
<td>Percentiles and first ID for $\hat{\sigma}_\epsilon^2$ for experiment B, T=50</td>
</tr>
<tr>
<td>21</td>
<td>Percentiles and first ID for $\hat{\sigma}_\eta^2$ for experiment B, T=50</td>
</tr>
<tr>
<td>22</td>
<td>Percentiles and first ID for $\hat{\beta}_1$ for experiment B, T=50</td>
</tr>
<tr>
<td>23</td>
<td>Percentiles and first ID for $\hat{\gamma}_1$ for experiment B, T=50</td>
</tr>
</tbody>
</table>
Table 24 - Percentiles and first ID for $\sigma_\varepsilon^2$ for experiment B, T-200
Table 25 - Percentiles and first ID for $\sigma_\eta^2$ for experiment B, T-200
Table 26 - Percentiles and first ID for $\hat{\beta}_1$ for experiment B, T-200
Table 27 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment B, T-200
Table 28 - Percentiles and first ID for $\sigma_\varepsilon^2$ for experiment C, T-50
Table 29 - Percentiles and first ID for $\sigma_\eta^2$ for experiment C, T-50
Table 30 - Percentiles and first ID for $\hat{\beta}_1$ for experiment C, T-50
Table 31 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment C, T-50
Table 32 - Percentiles and first ID for $\sigma_\varepsilon^2$ for experiment C, T-200
Table 33 - Percentiles and first ID for $\sigma_\eta^2$ for experiment C, T-200
Table 34 - Percentiles and first ID for $\hat{\beta}_1$ for experiment C, T-200
Table 35 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment C, T-200
Table 36 - Percentiles for $\sigma_\varepsilon^2$ for experiment A', T-50
Table 37 - Percentiles for $\sigma_\eta^2$ for experiment A', T-50
Table 38 - Percentiles for $\hat{\beta}_1$ for experiment A', T-50
Table 39 - Percentiles for $\hat{\gamma}_1$ for experiment A', T-50
Table 40 - Percentiles for $\sigma_\varepsilon^2$ for experiment A', T-200
Table 41 - Percentiles for $\sigma_\eta^2$ for experiment A', T-200
Table 42 - Percentiles for $\hat{\beta}_1$ for experiment A', T-200
Table 43 - Percentiles for $\hat{\gamma}_1$ for experiment A', T-200
Table 44 - Percentiles for $\sigma_\varepsilon^2$ for experiment A', T-200
Table 45 - Percentiles for $\sigma_\eta^2$ for experiment A', T-200
Table 46 - Percentiles for $\hat{\beta}_1$ for experiment A', T-200
Table 47 - Percentiles for $\hat{\gamma}_1$ for experiment A', T-200
Table 48 - Percentiles for $\sigma_\varepsilon^2$ for experiment C', T-50
Table 49 - Percentiles for $\sigma_\eta^2$ for experiment C', T-50
Table 50 - Percentiles for $\hat{\beta}_1$ for experiment C', T-50
Table 51 - Percentiles for $\hat{\gamma}_1$ for experiment C', T-50
Table 52 - Percentiles for $\hat{\sigma}_n^2$ for experiment C', T=200
Table 53 - Percentiles for $\hat{\beta}_1$ for experiment C', T=200
Table 54 - Percentiles for $\hat{\delta}_1$ for experiment C', T=200
Table 55 - Percentiles for $\hat{\gamma}_1$ for experiment C', T=200
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

\[ y_t = Z \alpha_t + \epsilon_t \quad \text{(measurement equation)} \]  
\[ \alpha_t = T \alpha_{t-1} + \eta_t \quad \text{(transition equation)} \]

\( t=1, \ldots, T \), where \( y_t \) is a px1 vector of observable variables, \( \alpha_t \) is a
mxl vector of unobservable variables, known as the state vector, Z is a pxm matrix, T is a mxm matrix, \( \epsilon_t \) is a px1 vector of serially uncorrelated disturbances with mean zero and covariance matrix \( \Sigma_\epsilon \) and \( \eta_t \) is a mxl vector of serially uncorrelated disturbances with mean zero and covariance matrix \( \Sigma_\eta \). We also assume that \( \epsilon_t \) and \( \eta_t \) are normally distributed and uncorrelated with each other for all periods of time and with the initial state vector \( \alpha_0 \) which is assumed to have a normal distribution with mean \( a_0 \) and covariance matrix \( P_0 \).

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, \( y_{it} \), may be modelled as

a) a local linear trend model, that is,

\[
\begin{align*}
Y_{it} &= \mu_{it} + \epsilon_{it} , \\
\mu_{it} &= \mu_{i,t-1} + \beta_{i,t-1} + \eta_{it} , \\
\beta_{it} &= \beta_{i,t-1} + \zeta_{it} .
\end{align*}
\]

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

\[
\begin{align*}
Y_{it} &= \mu_{it} + \epsilon_{it} , \\
\mu_{it} &= \mu_{i,t-1} + \eta_{it} ;
\end{align*}
\]

c) or simply as a sequence of independent variables.

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

\[
\begin{align*}
Y_{it} &= z_i'\alpha_{it} + \epsilon_{it} , \\
\alpha_{it} &= T_i\alpha_{i,t-1} + \eta_{it} ,
\end{align*}
\]
where for the random walk plus noise, $\alpha_t = \mu_t$, $z_{t}' = 1$ and $T_t = 1$.

whereas for the local linear trend

$$\alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} ; \quad z_{t}' = [1, 0] ; \quad T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$ 

Thus for each series, $z_{t}'$ and $T_t$ 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

$$y_t = (z'@I_p) \alpha_t + \epsilon_t \quad \text{(4a)}$$

$$\alpha_t = (T@I_p) \alpha_{t-1} + \eta_t \quad \text{(4b)}$$

where $\alpha_{t}$ and $\eta_{t}$, are of dimension $m \times 1$. We remark that in (4b), $T$ is $m \times m$, while in (1b) $T$ is $\bar{m} \times \bar{m}$, 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 $p \times p$ covariance matrix $\Sigma_\epsilon$ and the $pm \times pm$ covariance matrix $\Sigma_\eta$. The distinct elements of these matrices are known as the hyperparameters and will be denoted by the vector $\psi$. 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
\[ y_t = \begin{bmatrix} \mu_{1t} \\ \vdots \\ \mu_{pt} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{pt} \end{bmatrix} \tag{5a} \]

\[ \mu_t = \begin{bmatrix} \mu_{1t-1} \\ \vdots \\ \mu_{pt-1} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \vdots \\ \eta_{pt} \end{bmatrix} \tag{5b} \]

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

\[ \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix} \psi = \text{vec}(\Sigma_\epsilon: \Sigma_\eta) \tag{6} \]

and

\[ \psi = \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \text{vec}(\Sigma_\epsilon: \Sigma_\eta) \tag{7} \]

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 \(\alpha_0\) and on the disturbances \(\epsilon_t\) and \(\eta_t\), \(y = \text{vec} [y_1, \ldots, y_T]\) will have a multivariate normal distribution with mean \(\mu\) (\(T \times 1\)) and covariance matrix \(\Omega\) (\(T \times T\)). Now the density of \(y\) can be written in terms of the conditional densities, that is,
where $Y_{t-1}=(y_{t-1},...,y_1)$. Therefore the density of $y$ becomes

$$f(y|\mu,\Omega) = \frac{1}{(2\pi)^{\frac{d}{2}}|F_t|^{-\frac{1}{2}}} e^{-\frac{1}{2}(y_t-Ey_t)F_t^{-1}(y_t-Ey_t)}$$

where $F_t$ is the conditional covariance matrix of $y_t$ given $Y_{t-1},Y_{t-2},...,Y_1$.

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 $\theta$ are elements of $a_0$, $P_0$ and $\psi$, where $\psi$ is the vector containing the distinct hyperparameters, rather than $(\mu,\Omega)$. Clearly there is a relation between $\theta$ and $(\mu,\Omega)$. 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 $\alpha_t$ given $Y_t$ and the information at time $t=0$ is given by

$$a_t = E[\alpha_t / Y_t]$$

(10a)

with associated MSE

$$P_t = E((\alpha_t - a_t)(\alpha_t - a_t)')$$

(10b)

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

$$a_{t/t-1} = E[\alpha_t / Y_{t-1}]$$

(11a)

with associated MSE

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

(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

$$a_{t/t-1} = T a_{t-1}$$

(12a)

$$P_{t/t-1} = T P_{t-1} T' + \Sigma_\eta$$

(12b)

and the updating equations

$$a_t = a_{t/t-1} + P_{t/t-1} Z' F_t^{-1} r_t$$

(13a)

$$P_t = P_{t/t-1} - P_{t/t-1} Z' F_t^{-1} Z P_{t/t-1}$$

(13b)
Note that from (1a) and (11a) we have

\[ y_t - E_t(Y_{t-1}) = y_t - Z_{t-1} = r_t \]

and

\[ E_t(Y_{t-1}) = E_t(Y_{t-1} - E_t(Y_{t-1}))' = E_t(Z_{t-1} + \epsilon_t)' - F_t \]

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

\[ \ell(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \log |F_t| - \frac{1}{2} \sum_{t=1}^{T} r_t'F_t^{-1}r_t \]

where \( \theta = (\alpha_{t-1}, \phi, \psi) \), and \( r_t = r_t(\theta) \) and \( F_t = F_t(\theta, \phi) \) are obtained from the Kalman filter equations with starting \( a_0 \) and \( P_0 \).

The Link between Cholesky Decomposition and State Space Techniques

If \( \Omega \) is positive definite it can be factorized (Cholesky decomposition) in such a way that \( \Omega^{-1} = L'F'IL \), 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 \( \text{diag}(I_p, \ldots, I_p) \)
on its main block diagonal and $F$ is a block diagonal matrix, i.e.,

$$F = \text{diag}(F_1 \ldots F_T).$$

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 $r_t$, $t=1, \ldots, T$ 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 $\mu_t$ 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 $\Omega$ and $P_o$ and the hyperparameters, $\psi = (\sigma^2, \sigma^2)$, is

$$\sigma_{ii} = P_o + i \sigma^2 + \sigma^2, \quad i=1, \ldots, T,$$

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

Substituting these values in $\Omega$, constructing the Cholesky decomposition of $\Omega$ and inverting the triangular matrix yields
where $P_{t-1}$ 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 $(\mu, \Omega)$ and $(a_o, P_o, \psi)$ can be easily established for the local linear trend, as well as for SUTSE models. We write

$$Y_t = (z'@I_p)(T@I_p)^t a_o + (z'@I_p)\sum_{j=1}^{T} (T@I_p)^{t-j} \eta_j + \epsilon_t,$$

hence

$$E_y = X_o a_o$$

where

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

and

\begin{align}
L &= \begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 0 & 0 \\\np_{t-1} & 1 & 0 & 0 & \ldots & 0 & 0 \\
p_{t-1} [1 - p_{t+1}] & p_{t+1} & 1 & 0 & \ldots & 0 & 0 \\
p_{t-1} [1 - p_{t+1}] [1 - p_{t+2}] & p_{t+1} [1 - p_{t+2}] & p_{t+2} & 1 & \ldots & 0 & 0 \\
p_{t-1} [1 - p_{t+1}] [1 - p_{t+2}] [1 - p_{t+3}] & p_{t+1} [1 - p_{t+2}] [1 - p_{t+3}] & p_{t+2} [1 - p_{t+3}] & p_{t+3} & 1 & \ldots & 0 \\
p_{t-1} [1 - p_{t+1}] [1 - p_{t+2}] [1 - p_{t+3}] [1 - p_{t+4}] & & & & & & 1 \\
\vdots & & & & & & \ddots & \\
\vdots & & & & & & \ddots & \\
\ddots & & & & & & \ddots & \\
1 & & & & & & \ddots & \\
\end{bmatrix}
\end{align}

(17a)
and

\[ X_0 = [(z'T^@Ip)', ..., (z'T^T@Ip)']'. \tag{20} \]

and analogously as before the lower triangular and the block diagonal matrices \( L \) and \( F \) can be constructed with elements being functions of \( Z, T, \psi \) and \( P_0 \). (Note that the superscript \( T \) in (2) is the sample size). Therefore (16) can be also written as

\[
\ell(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \log |F_t| - \frac{1}{2} L(y-X_0\alpha_0)'F^{-1}L(y-X_0\alpha_0) \tag{21}
\]

where \( \theta=(\alpha_0, P_0, \psi) \) and \( y=\text{vec}[y_1, y_2, ..., y_T] \).

**Conditional Likelihood Function**

As it stands the loglikelihood function given in (21) is a function of \( \theta \), the distinct parameters which enter into \( (\alpha_0, P_0, \psi) \). 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 \( Y_T=(y_1, ..., y_T) \) via the prediction error decomposition. For the models considered here however, no prior information is available. de Jong (1988) derived an expression for the likelihood function of \( Y_T \), not conditional on \( \alpha_0 \), where the place of both \( \alpha_0 \) and \( P_0 \) is made explicit in the likelihood function. He pointed out that it is not possible to find the ML estimates of both \( \alpha_0 \) and \( P_0 \). He also justifies the specification of the
unconditional distribution of \( \alpha_o \) 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 \( \alpha_t \) in the transition equation yields

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

The diffuse prior of \( \alpha_o \) is obtained as \( s \to -\infty \). 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 \( y_{T-1}, \ldots, y_{m+1} \), 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 \( (T-m) \times 1 \) mean vector being \( Z_{T-m}^T a_m \). The \( (T-m) \times (T-m) \) covariance matrix is \( (L'F^{-1}L)^{-1} \) where \( L \) 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,

\[
\begin{align*}
a_1 & \to y_1 & & (22a) \\
p_1 & \to \sigma^2 & & (22b)
\end{align*}
\]
and for the local trend model

\[ a_2 \rightarrow \begin{bmatrix} y_2 \\ y_2 - y_1 \end{bmatrix} \] \hspace{1cm} (23a)

\[ P_2 \rightarrow \begin{bmatrix} \sigma^2_{e} & \sigma_{e}^2 \\ \sigma_{e}^2 & 2\sigma_{e}^2 + \sigma_{\eta}^2 + \sigma_{\zeta}^2 \end{bmatrix} \] \hspace{1cm} (23b)

On the other hand, rather than assuming a diffuse prior for the initial state we could construct a proper prior for \( \alpha_m \) from the first observations. It turns out, however, that the resulting estimators of the mean and variance of \( \alpha_m \) 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 \( a_0 \) and \( P_0 = K I_m \) where \( K \) is a large finite number, \( I_m \) is the \( m \times m \) 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 \( pxl \) vector \( y_1 \) while \( P_1 \) the \( pxp \) matrix \( \Sigma_e \). From the above discussion the resulting conditional loglikelihood function
becomes

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

(24)

where $\bar{y} = \text{vec}(y_{m+1}, \ldots y_T)$. 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 $z'Tt^{-2}$ for the local trend model.

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

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

(25)

where $r_t$ and $F_t$, $t=m+1, \ldots T$, 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 $\psi$ 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 $P_m^* = P_m/\sigma_x^2$ and $\psi$ by $(1, \psi^*)$ where $\sigma_x^2$ is the hyperparameter whose place is made explicit and $\psi^*$ 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

$$\ell(\sigma^*, \psi*) = -\frac{1}{2} \sum_{t=m+1}^{T} \log f_t - \frac{T-m}{2} \log \sigma^2 - \frac{1}{2} \sigma^{-2} \sum_{t=m+1}^{T} v_t^2 / f_t$$

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

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 $u_t$, $t=0, \pm 1, \pm 2, \ldots$ be a $p$-variate stationary, zero mean, Gaussian
process. Let F(λ,ψ) be the spectral matrix of the process, where ψ belongs to the parameter set Θ and λ∈[-π,π]. F(λ,ψ) is defined by

$$F(λ,ψ) = (2π)^{-1} \sum_{τ=-∞}^{∞} Γ(τ,ψ) e^{-iλτ}$$  \hspace{1cm} (1)

where

$$Γ(τ,ψ) = E u_t u_{t-τ}'$$  \hspace{1cm} (2)

is the autocovariance matrix at lag τ, see Harvey (1989, page 428). The diagonal elements of F(λ), 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≠i. The spectral matrix, see Fuller (1976), is Hermitian, that is, F(λ) = F*(λ) and positive semidefinite, that is, w*F(λ)w > 0 for any complex vector such that w*w>0, where here and onwards [·]* denotes the complex conjugate transpose of a matrix or of a vector.

Let F(λj), j=0,...,T-1, denote the spectral matrices at frequency λj, where

$$λ_j = \frac{2πj}{T}, \hspace{1cm} j = 0,...,T-1.$$  

Let

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

and let I_{U'U}(λ_j) be the Hermitian matrix of periodograms and crossperiodograms of U', or as we shall say, the periodogram matrix, given by
We also introduce, \( I_T(\lambda) = I_U'U(\lambda) \), which we shall need later, the periodogram matrix defined for all \( \lambda \) in \([-\pi, \pi)\). Of course we cannot evaluate \( I_T(\lambda) \) numerically as a continuous function of \( \lambda \).

As is well known, see among others Robinson (1978), the frequency-domain or spectral likelihood function for \( \text{vec}U' \) is given by

\[
\ell(\psi) = -\frac{1}{2} \sum_{j=0}^{T-1} \log \left| F(\lambda_j) \right| - \frac{1}{2} \sum_{j=0}^{T-1} \text{tr} \left[ F^{-1}(\lambda_j) I_U'U(\lambda_j) \right].
\]  (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 \( \text{vec}U' \) 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 $\sigma_\xi^2 > 0$. 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 $\sigma_\xi^2 = 0$ 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 $\Sigma_\varepsilon$ and $\Sigma_\eta$. Therefore if $\Sigma_\eta$ is positive definite then differencing once yields a multivariate stationary and invertible process with spectral matrix

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

where

$$c(\lambda_j) = 2(1 - \cos \lambda_j), \quad (5b)$$

and $\psi$ is the $p(p+1)$ vector containing the distinct elements of $\Sigma_\varepsilon$ and $\Sigma_\eta$. We note that $F(\lambda_j)$, $j=0,...,T-1$ 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

$$I_{U'U}(\lambda_j) = R_{I_{U'U}(\lambda_j)} + i \text{Im}I_{U'U}(\lambda_j)$$

(6)

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

$$\text{tr}[F(\lambda_j)^{-1}\text{Im}I_{U'U}(\lambda_j)] = -\text{tr}[\text{Im}I_{U'U}(\lambda_j)F(\lambda_j)^{-1}]$$

$$= -\text{tr}[F(\lambda_j)^{-1} \text{Im}I_{U'U}(\lambda_j)] = 0.$$ 

Hence,

$$\text{tr} [(F(\lambda_j)^{-1}I_{U'U}(\lambda_j)) = \text{tr}[F(\lambda_j)^{-1} R_{I_{U'U}(\lambda_j)}],$$

(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

$$R_{I_{U'U}(\lambda_j)} = U'\Psi_j U,$$

(8)

where the $T \times T$ matrix $\Psi_j$, $j=1,...T-1$, is real, symmetric, with $(\ell,k)$ entry being

$$\Psi_j(\ell,k) = \cos(\lambda_j(\ell-k)), \ell,k=1,...T.$$

(9)

Using (9), the spectral likelihood takes the form
\[
\ell(\psi) = -\frac{1}{T} \sum_{j=0}^{T-1} \log F(\lambda_j) - \frac{1}{T} \sum_{j=0}^{T-1} \text{tr} \left[ F^{-1}(\lambda_j) U' \psi_j U \right].
\]  
(10)

Asymptotic Information Matrix

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

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

where \( \psi_o \) 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 \( \psi \) is given in (2.7). Now the first and second derivatives of (10) with respect to \( \psi \) are given in Harvey (1989) or in Fernandez (1986) and are

\[
\frac{\partial}{\partial \psi} \ell = - \frac{1}{2} \sum_{j=0}^{T-1} S_j \text{vec} \left[ F_{j-1}^{-1} - F_{j-1}^{-1} U' \psi_j U F_{j-1}^{-1} \right]
\]

(11)

and

\[
\frac{\partial^2 \ell}{\partial \psi \partial \psi'} = \sum_{j=0}^{T-1} S_j \left[ \frac{1}{2} (F_{j-1}^{-1} \otimes F_{j-1}^{-1}) - (F_{j-1}^{-1} \otimes F_{j-1}^{-1} U' \psi_j U F_{j-1}^{-1}) \right] S_j
\]

(12)

where

\[
S_j = \frac{\partial \text{vec} F_j}{\partial \psi},
\]

and

\[
F_j = F(\lambda_j, \psi) = (2\pi)^{-1} \left[ \Sigma_\xi : \Sigma_\eta \right]^{c_j 1_p},
\]

\[
c_j = c(\lambda_j) = 2(1 - \cos \lambda_j).
\]
Vectoring $F_j$ and using (2.6) we have

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

hence $S_j$ becomes

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

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

$$E[\ell(T(k))] = F(k, \psi_0)_{1k} + O(T^{-1}\log T), \quad 1, k = 1, \ldots, p \quad (14)$$

we have

$$IA(\psi_0) = \frac{1}{4\pi} \int_{-\pi}^{\pi} S(\lambda)' \left[ F^{-1}(\lambda) \otimes F^{-1}(\lambda) \right] S(\lambda) \, d\lambda \quad (15)$$

where $F(\lambda) = F(\lambda, \psi_0)$, $S(\lambda) = [c(\lambda)D : D]$, $c(\lambda) = 2(1-\cos \lambda)$ 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

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

where $g(\hat{\psi}_k)$ is the first derivative of $\ell$ with respect to $\psi$ 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 \( \tilde{H}(\hat{\psi}_k) \) yields the scoring algorithm, where for the multivariate random walk plus noise, \( T \) times \( \tilde{H}(\psi) \) is a finite approximation of the integral given in (15). Hence \( \tilde{H}(\hat{\psi}_k) \) is given by

\[
\tilde{H}(\hat{\psi}_k) = \frac{1}{T} \sum_{j=0}^{T-1} \left( \hat{F}_j^{-1} \otimes \hat{F}_j^{-1} \right) S_j,
\]

where \( \hat{F}_j = F_j(\hat{\psi}_k) \). As shown in Robinson (1978), under certain regularity conditions \(-T^{-1}\tilde{H}(\psi)\) 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
\[ c(\tau) = \frac{1}{T} \sum_{t=1}^{T} u_t u_{t+\tau}; \quad (1) \]
\[ \tilde{\psi} = G(c(0), c(1), \ldots, c(\tau)); \quad (2) \]
where \( G \) is continuously differentiable in a neighbourhood of \((\gamma(0), \gamma(1), \ldots, \gamma(\tau))\);
and \( \tilde{\psi} \) a single iteration of the scheme given in (3.16) with initial value \( \tilde{\psi} \).

Under suitable conditions, we have
\[ \lim_{T \to \infty} (c(\tau) - \gamma(\tau)) = 0 \text{ a.s. for all fixed } \tau; \quad (3) \]
\[ T \delta(c(0) - \gamma(0), c(1) - \gamma(1), \ldots, c(\tau) - \gamma(\tau)) \overset{d}{\to} N(0, \cdot), \text{ for all } \tau > 1; \quad (4) \]
\[ \lim_{T \to \infty} (\tilde{\psi} - \psi_0) = 0 \text{ a.s.}; \quad (5) \]
\[ T \delta(\tilde{\psi} - \psi_0) \overset{d}{\to} N(0, \cdot); \quad (6) \]
\[ \lim_{T \to \infty} (\tilde{\psi} - \psi_0) = 0 \text{ a.s.}; \quad (7) \]
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 straightforward 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 $\ell$, $\ell > \frac{1}{2}$, the Lipschitz class of degree $\ell$.

(A3) $\psi_0$ is an interior point of $\Theta$, which is the compact closure of an open submanifold in a twice-differentiable p-dimensional manifold.

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

(A5) $f(\lambda, \psi) \neq f(\lambda, \psi_0)$ for all $\psi \neq \psi_0$, $\psi \in \Theta$. 

\[ T_i(\hat{\psi} - \psi_0) \overset{d}{\rightarrow} N(0, IA^{-1}(\psi_0)). \quad (8) \]
(A6) Within a neighborhood of \( \psi_0 \), \( f(\lambda, \psi) \) has first and second derivatives with respect to \( \psi \), these being continuous in \( \lambda \) and \( \psi \).

(A7) \( f(\lambda, \psi_0) > 0; f'(\lambda, \psi_0) \in \text{Lip}, i>1 \).

(A8) \( IA(\psi_0) \) 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

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

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 \( \psi_0 \) is obtained by a single use of the iterative scheme provided that the current estimate is consistent, is particularly 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) \( f(\lambda, \psi_0) > 0 \) is replaced by \( F(\lambda, \psi_0) \) positive definite.
Appendix 2.1

Optimisation Procedures

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

\[ \frac{\partial}{\partial \theta} L(\theta | x) = 0 \]  

(1)

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

\[ \theta_{k+1} = \theta_k - G_k^{-1} g_k \]  

(2)

where \( \theta_1 \) is a vector of initial values, \( \theta_k, k=1, \ldots \) is the current estimate of \( \theta \), \( G_k \) 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
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 $\delta$ given $\psi$ 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 $\theta$. Such a procedure which induces separate optimisation of the parameters in $\theta$ 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 $\delta$ and $\psi$ respectively. Thus the iterative scheme is for $j=2,\ldots$

\[
\delta j(k+1) - \delta j k + H_{11}(\delta j k, \psi j) - 1 g_\delta(\delta j k, \psi j) \quad k=0,\ldots \quad (3a)
\]

\[
\psi j(k+1) - \psi j k + H_{22}(\delta j k, \psi j k) - 1 g_\psi(\delta j k, \psi j k) \quad k=0,\ldots \quad (3b)
\]

where $\delta_1$ and $\psi_1$ are initial values, $\delta j0 = \delta j$, $\psi j0 = \psi j$ are the final estimates computed at step $j-1$. $g_\delta$, the first derivatives with respect to $\delta$, and $H_{11}$ are evaluated at the current estimate of $\delta$ and at the final estimate of $\psi$ at step $j-1$; similarly $g_\psi$ and $H_{22}$. The procedure is bound to converge, see Oberhofer and Kmenta (1974). Clearly if the likelihood equations are linear in $\delta$ given $\psi$ then $\delta j$
can be obtained directly, and vice-versa.

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

\begin{align}
Y_t &= z_t' \delta + \omega_t \\
\omega_t &= \mu_t + \epsilon_t \\
\mu_t &= \mu_{t-1} + \eta_t
\end{align}

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

Now conditionally on \( \omega_1 \) the loglikelihood function for \( \omega_2, \ldots, \omega_T \) 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
\[ \ell(\theta) = -\frac{1}{2} \sum_{t=2}^{T} \log f_t - \frac{1}{2} \left[ \sum_{t=2}^{T} \log f_t - (T-1) \log \sigma^2 + \frac{1}{2} \sigma^{-2} \sum_{t=2}^{T} r_t^2 f_t^{-1} \right] \]

where \( y' = (y_2, \ldots, y_T), \ Z' = (z_2, \ldots, z_T) \) and \( I \) is a \( T \times 1 \) 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 \( P_1 = \sigma^2 \). However, since

\[ \tilde{L}[y_1 - Iy_1 - (Z - Iz_1) \delta] = \tilde{L}[y_1 - y_1] - \tilde{L}[(Z - Iz_1) \delta], \]

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 \( [z_{11}, \ldots, z_{1K}] \). Hence, if the hyperparameters are scaled as in (2.2.26) the resulting loglikelihood function becomes

\[ \ell(\delta, \sigma^2, \psi) = -\frac{1}{2} \sum_{t=2}^{T} \log f_t - \frac{1}{2} (T-1) \log \sigma^2 + \frac{1}{2} \sigma^{-2} \sum_{t=2}^{T} r_t^2 f_t^{-1} \]

where

\[ \sum_{t=2}^{T} r_t^2 f_t^{-1} = \sum_{t=2}^{T} (\tilde{y}_t - \tilde{z}_t' \delta)^2 / f_t = (\tilde{Y}' - \tilde{Z} \delta)' \tilde{P}^{-1} (\tilde{Y}' - \tilde{Z} \delta) \]

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, \( \lambda \) must be specified. However if \( z_t \) is weakly exogenous for \( \theta \) 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, $\lambda$ consists of the mean vector and the covariance matrix. Usually the elements of $\theta$, the parameters of interest, do not coincide with those in $\lambda$. Thus, let $\theta^t$ be a reparameterization of $\lambda$ such that $\theta^t = (\theta, \theta_z)$, where $\theta$ and $\theta_z$ are variation free, i.e. $\theta$ and $\theta_z$ are not subject to cross restrictions so that for any admissible value of $\theta_z$, $\theta$ can take any value in its parameter space and vice versa. Then $z_t$ is weakly exogenous for $\theta$ if the joint distribution of $[Y_t, Z_t]$ can be factorized as

$$
\begin{align*}
  f(Y,Z; \theta^t) &= \prod_{t=2}^{T} f(y_t | z_t, Y_{t-1}, Z_{t-1}; \theta) f(y_{11} | z_1; \theta) f(z_1 | \theta_z) \\
  &\text{Hence all sample information concerning $\theta$ can be obtained from the first term in the RHS of (4). For prediction purposes we have to assume that $z_t$ is strongly exogenous. We note that $z_t$ is strongly exogenous for $\theta$ if it is weakly exogenous and in addition past values of $y_t$ does not Granger-cause $z_t$, that is, conditionally on $z_{t-1}, z_{t-2}, \ldots, z_t$ is independent of past values of $y_t$.}
\end{align*}
$$

Multivariate Time Series Regression Models

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

$$
y_t = B'z_t + \omega_t
$$
where \( y_t \) is a px1 vector, \( B' \) is a pxK matrix, \( z_t \) is a Kx1 vector and \( \text{plim} T^{-1} Z'Z \) is positive definite, where \( Z'=(z_1,\ldots,z_T) \). \( \omega_t \) follows a multivariate random walk plus noise model. Let \( \theta=(\beta,\psi) \) the parameters of interest, where \( \beta=\text{vec}B \) and \( \psi \) is the \( p(p+1)\times 1 \) vector containing the unrestricted elements of \( \Sigma_{\epsilon} \) and \( \Sigma_{\eta} \). Using rules on Kronecker products, see Magnus and Neudecker (1988, page 47), (5) can be rewritten as

\[
y_t = (I_p \otimes z_t') \beta + \omega_t. \tag{6}
\]

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

\[
y_t = \begin{bmatrix} Z_1t' & 0 & 0 & 0 \\ 0 & Z_2t' & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & Z_{pt}' \end{bmatrix} \beta^* \tag{7}
\]

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

Now conditionally on the first observations the loglikelihood function for \( \psi=\text{vec}(\omega_2,\ldots,\omega_T) \) 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

\[
\ell(\theta) = -\frac{1}{2} \sum_{t=2}^{T} \log |F_t| - \frac{1}{2} \sum_{t=2}^{T} r_t' F_t^{-1} r_t, \tag{8a}
\]

where

\[
r_t = y_t - z_t' \beta, \tag{8b}
\]
and $\tilde{y}_t$ is obtained by applying the multivariate Kalman filter to $y_t$, $t=2,\ldots,T$, with starting value $y_1$ (p×1) and $\tilde{z}_t'$ is obtained by applying the multivariate Kalman filter separately to each column of $(I_p @ z_t')$, $t=2,\ldots,T$, with starting values being the respective column of $(I_p @ z_1')$.

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 $(\sigma_x^2, \psi_y)$ and one for $\delta$. The part for $(\sigma_x^2, \psi_y)$ consists of finding the point that maximises (3) with respect to $(\sigma_x^2, \psi_y)$ conditionally on a given $\delta$. We shall denote this point by $(\hat{\sigma}_x^2, \hat{\psi}_y)$, $\hat{\sigma}_x^2 = \hat{\sigma}_x^2(\delta)$, $\hat{\psi}_y = \hat{\psi}_y(\delta)$. Clearly, $(\hat{\sigma}_x^2, \hat{\psi}_y)$ is
obtained as it would be for a model without explanatory variables. The part for $\delta$ consists in finding the point that maximises (3b) with respect to $\delta$, conditionally on a given $\psi_*$. This is simply the GLS (Generalized Least Squares) estimator, that is,

$$\hat{\delta}(\psi_*) = \left[ \sum_{t=2}^{T} f_t^{-1} \tilde{z}_t \tilde{z}_t' \right]^{-1} \left[ \sum_{t=2}^{T} [f_t^{-1} \tilde{z}_t y_t] \right]. \quad (9)$$

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

On the other hand replacing (9) in (3b) yields $\ell_c(\sigma_*^2, \psi_*)$, the concentrated loglikelihood function with respect to $(\sigma_*^2, \psi_*)$. 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 $\psi$.

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

$$\hat{\beta}(\psi) = \left[ \sum_{t=2}^{T} \tilde{z}_t F_t^{-1} \tilde{z}_t' \right]^{-1} \sum_{t=2}^{T} \tilde{z}_t F_t^{-1} y_t, \quad (10)$$

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 6 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

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

where the vectors $y_t$ (px1) and $z_t$ (kxl) are the observable variables and the vector $\omega_t$ (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 $\omega_t$ follows a multivariate random walk plus noise differencing once $y_t$ and $z_t$ yields

$$y_t = B'z_t + u_t,$$  \hspace{1cm} (1)
where we deliberately omitted the differencing operator $\Delta$ 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 $\Delta y_t$ or $\Delta z_t$. 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,

$$\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} z_t z_t' = \Gamma_Z(\tau) \text{ a.s., } \tau = 0, \pm 1, \pm 2, \ldots$$  \hspace{1cm} (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 $\lim_{T \to \infty} \sum_{t=1}^{T} z_t$ exists.

In matrix notation (1) can be written as

$$Y' = B'Z' + U'$$  \hspace{1cm} (3)

where

$$Y' = [y_1 \ldots y_T] \text{ where } y_t \text{ is } p \times 1$$

$$Z' = [z_1 \ldots z_T] \text{ where } z_t \text{ is } K \times 1$$

$$U' = [u_1 \ldots u_T] \text{ where } u_t \text{ is } p \times 1$$

Now the spectral likelihood for vec$U'$ is given in (2.3.4) and since $Z$
and $U$ are totally independent the spectral likelihood function for $\text{vec}(Y')$ becomes

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

(4)

where $\theta=(\beta, \psi)$, $\beta = \text{vec}B$ and $\psi$ is as in (2.2.7), that is, the $p(p+1)\times 1$ vector obtained from $\text{vec}(\Sigma_\epsilon; \Sigma_\eta)$ by eliminating the supradiagonal elements of $\Sigma_\epsilon$ and $\Sigma_\eta$. $U' \Psi_j U = U_\lambda \lambda_j$ is the real part of the periodogram matrix of $U$ where $U'$ is expressed in terms of the observations through (3). Thus,

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

(5)

where

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

(6)

is the real part of the periodogram matrix of the augmented process $[y_t', z_t']'$, $\Psi_j$ 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 \( \beta = \text{vec}B \) rather than \( \beta = \text{vec}B' \) will become clear in the simultaneous equation system context.

Thus

\[
\frac{\partial}{\partial \beta} \ell = \text{vec}\left[ \sum_{j=0}^{T-1} Z_j' \Psi_j U_j F_j^{-1} \right] \tag{7a}
\]

\[
\frac{\partial}{\partial \psi} \ell = -\frac{1}{2} \sum_{j=0}^{T-1} S_j \text{vec}\left[ (F_j^{-1} - F_j^{-1} U_j' \Psi_j U_j F_j^{-1}) \right] \tag{7b}
\]

and

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

\[
\frac{\partial^2}{\partial \psi \partial \beta} \ell = -\sum_{j=0}^{T-1} S_j \left[ (F_j^{-1} \otimes \Psi_j Z \otimes F_j^{-1}) K_p \right] \tag{8b}
\]

\[
\frac{\partial^2}{\partial \psi \partial \psi'} \ell = \sum_{j=0}^{T-1} S_j \left[ (F_j^{-1} \otimes F_j^{-1}) - (F_j^{-1} \otimes F_j^{-1} U_j' \Psi_j U_j F_j^{-1}) \right] S_j \tag{8c}
\]

where \( K_{kp} \) is a \( K_p \times K_p \) commutation matrix and \( S_j \) is given in (2.3.13), that is,

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

Since \( u_\ell \) and \( z_\ell \) are totally independent and \( u_\ell \) has zero mean we have \( E Z_j' U(\lambda_j) = 0 \) and because of (2) and (2.3.14) we have the asymptotic information matrix,

\[
\text{IA}(\theta_0) = \lim_{T \to \infty} T^{-1} \left[ -E \frac{\partial^2 \ell}{\partial \theta \partial \theta'} \right] \theta_0 .
\]
\[ \Theta(\theta_0) = \begin{bmatrix} \frac{1}{2\pi} \int_{-\pi}^{\pi} (F(\lambda)^{-1} \Theta F(\lambda)) \, d\lambda & 0 \\ 0 & \frac{1}{4\pi} \int_{-\pi}^{\pi} S(\lambda)^{'} \left[ F(\lambda)^{-1} \Theta F(\lambda)^{-1} \right] S(\lambda) \, d\lambda \end{bmatrix} \]  

(9)

where \( \theta_0 \) is the true value, \( F(\lambda) = F_u(\lambda, \theta_0), S(\lambda) = (2\pi)^{-1} [c(\lambda)D: D], \)
\( c(\lambda) = 2(1 - \cos \lambda) \) and \( D \) is the \( p^{2} \times p(p+1) \) duplication matrix.

**Spectral Maximum Likelihood Estimates**

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

\[ \hat{\beta}(\psi) = \left[ \sum_{j=0}^{T_{2}} [F_{j}^{-1} \otimes I_{Z^{'}Y}(j)]^{-1} \left[ \sum_{j=0}^{T_{2}} [F_{j}^{-1} \otimes I_{K}] \text{vec}I_{Z^{'}Y}(j) \right] \right]^{-1} \left[ \sum_{j=0}^{T_{2}} [F_{j}^{-1} \otimes I_{K}] \text{vec}I_{Z^{'}Y}(j) \right] \]  

(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 \( (\beta, \psi) \) can be found as follows.

**Step 1** - An initial value for \( \beta \) is evaluated, say \( \hat{\beta} = \text{vec} \hat{B}_{LS} \), where \( \hat{B}_{LS} \) is the least squares estimator, that is,

\[ \hat{B}_{LS} = [Z^{'}Z]^{-1}Z^{'}Y ; \]  

(11)

**Step 2** - Conditionally on \( \hat{\beta} \), (4) is maximised with respect to \( \psi \);
Step 3: Making use of (10), \( \hat{\beta} = \hat{\beta}(\psi) \) 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 \( \beta \) no gain in efficiency will be achieved if the procedure is repeated. Also, in obtaining \( \hat{\psi} = \hat{\psi}(\hat{\beta}) \), we do not need to iterate until convergence, if we start with a consistent estimator of \( \psi \). 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 \( (\beta, \psi) \) could be obtained by concentrating \( \beta \) out of the spectral likelihood function. The concentrated likelihood function becomes

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

where \( \hat{U} = Y - Z\hat{B} \), and \( \hat{B} \) is such that \( \text{vec}\hat{B} = \hat{\beta} \), 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 straightforward 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 \( \theta, \theta=(\beta, \psi) \) 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 \( u_t = \eta_t + \Delta \varepsilon_t \) and \( f_u(\lambda) = \sigma \varepsilon^2 (1-\cos \lambda) + \sigma^2 \eta \). Under the condition on \( z_t \) given in (3.2), we have from Hannan (1971) that

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{j=0}^{J} \Phi_j I_z(j) = 0
\]

and

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{j=0}^{J} \Phi_j I_u(j) = 0
\]

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

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

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{j=0}^{J} [f_j^{-1} I_z(j)] = N \left[ 0, \frac{1}{2\pi} \int_{-\pi}^{\pi} f_u^{-1}(\lambda) F_z(\lambda) d\lambda \right]
\]
We note that, if $\tilde{\psi}$ is $\mathcal{T}$-consistent, that is, $T^1(\tilde{\psi} - \psi_o) = O_p(1)$ then, because $f_j(\lambda)$ satisfies the regularity conditions given in section 2.4, we can expand $f_j^{-1}$ in a Taylor's series for random functions, see Fuller(1976, page 191), and write

$$T^{-1} \left[ \sum_{j=0}^{T-1} f_j^{-1} I_z' u(j) - \sum_{j=0}^{T-1} f_j^{-1} I_z' u(j) \right]$$

$$- T^{-1} \sum_{j=0}^{T-1} \left[ \frac{\partial}{\partial \tilde{\psi}} f_j^{-1} \right] I_z' u(j) T^1(\tilde{\psi} - \psi_o) + O_p(T^{-1})$$

where the derivative is evaluated at the true parameter vector $\psi_o$, $f_j = f_j(\psi_o)$ and $\tilde{f}_j = \tilde{f}_j(\tilde{\psi})$. 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

$$T^1(\tilde{\psi} - \psi_o) \xrightarrow{d} N \left[ 0, IA^{-1}(\psi_o) \right]$$

(4)

and

$$T^1(\hat{\delta} - \delta_o) \xrightarrow{d} N \left[ 0, IA^{-1}(\delta_o) \right],$$

(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;

$$\hat{\psi} - \tilde{\psi} + \bar{H}^{-1}(\tilde{\psi}) g_{\psi}(\hat{\delta}, \tilde{\psi}),$$

(6)

$$\bar{H}(\tilde{\psi}) = \frac{1}{2} \sum_{j=0}^{T-2} S_j \tilde{f}_j^{-2} S_j,$$
where $\hat{\delta} = f_j(\tilde{\delta})$ and $g_j(\hat{\delta}, \tilde{\psi})$ is a specialization of (3.7b) for the univariate model evaluated at $(\hat{\delta}, \tilde{\psi})$:

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

(7)

$$T_{\delta}(\hat{\delta}, \delta_0) = O_p(1),$$

$$T_{\psi}(\tilde{\psi}, \psi_0) = O_p(1).$$

We note that because $\psi' = [-\gamma(1): \gamma(0) + 2\gamma(1)]'$, a $T$-consistent estimator of $\psi$, say $\tilde{\psi} = \tilde{\psi}(\hat{\delta})$, can be constructed from $\tilde{c}(\tau)$, where $\tilde{c}(\tau) = T^{-1} \sum \tilde{u}_t \tilde{u}_{t-\tau}^T$, $\tilde{u}_t = y_t - \tilde{\eta} \tilde{e}_t$ and $\tilde{\eta}$ is the OLS estimator of $\tilde{\eta}$. From the discussion in section 2.4, we have that $T_{\psi}(\tilde{c}(\tau), \gamma(\tau)) \rightarrow N(0, \cdot)$, where $c(\tau) = T^{-1} \sum u_t u_{t-\tau}$. However since $T_{\delta}(\hat{\delta}, \delta_0) = O_p(1)$, 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

$$\lim T^{-1} \sum_{j=0}^{T-1} \left[ F_j^{-1} @ I_{Z'}Z(j) \right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_{u^{-1}}(\lambda) @ F_{z}(d\lambda),$$

(8)

$$\lim T^{-1} \sum_{j=0}^{T-1} \left[ (F_j^{-1} @ I_{k}) \text{vec} I_{Z'}Y(j) \right] = 0,$$

(9)

$$T^{-1} \sum_{j=0}^{T-1} \left[ (F_j^{-1} @ I_{p}) \text{vec} I_{Z'}Y(j) \right] \rightarrow N \left[ 0, \frac{1}{2\pi} \int_{-\pi}^{\pi} F_{u^{-1}}(\lambda) @ F_{z}(d\lambda) \right],$$

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

\[
T\bar{\psi} \left[ \hat{\beta} \cdot \beta_o \right] \overset{d}{\rightarrow} N(0, \text{IA}^{-1}(\theta_o)),
\]

where \( \bar{\psi} \) 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 \( \psi \). \( \theta_o = (\beta_o', \psi_o')' \) and IA(\( \theta_o \)) is given in (3.9).

Finally 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,

\[
\hat{\psi} = \psi + \frac{1}{H(\psi)} g(\delta, \psi),
\]

where

\[
H(\psi) = \frac{1}{2} \sum_{j=0}^{T-1} S_j f_j^{-2} S_j,
\]

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

\[
g(\delta, \psi) = g(\delta, \psi_o) + H_{\psi}(\delta, \psi_o)(\psi - \psi_o) + O_p(T^{-1})
\]

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

\[
T^{\frac{1}{2}}(\hat{\psi} - \psi_o) = \left[ \frac{H(\psi)}{T} \right]^{-1} \left[ \frac{H(\hat{\psi})}{T} + \frac{H_{\psi}(\delta, \psi_o)}{T} \right] T^{\frac{1}{2}}(\psi - \psi_o)
\]

\[
+ \left[ \frac{H(\psi)}{T} \right]^{-1} T^{\frac{1}{2}} g(\delta, \psi_o) + \left[ \frac{H(\psi)}{T} \right]^{-1} T^{\frac{1}{2}} O_p(T^{-1}).
\]

Now

\[
\frac{1}{T} H_{\psi}(\delta, \psi_o) = \frac{1}{2T} \sum_{j=0}^{T-1} S_j f_j^{-2} S_j - \frac{1}{T} \sum_{j=0}^{T-1} S_j f_j^{-3} 1_{U_j}(\psi) S_j
\]
\[- \frac{1}{2T} \sum_{j=0}^{T-1} s_j f_j^{-1} s_j - \frac{1}{T} \sum_{j=0}^{T-1} s_j f_j^{-3} [I_{U^jU(j)} - I_{U^jU(j)}]s_j \]

\[- \frac{1}{T} \sum_{j=0}^{T-1} s_j f_j^{-3} I_{U^jU(j)} s_j \quad (4)\]

where

\[I_{U^jU(j)} - I_{U^jU(j)} = -(\delta - \delta') I_{ZU(j)} - I_{UZ(j)}(\delta - \delta') + (\delta - \delta') I_{ZZ(j)}(\delta - \delta') \quad (5)\]

Clearly, the first term in (4) converges to \(IA(\psi_0)\). 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 \(-2IA(\psi_0)\). Hence

\[\text{plim } T^{-1} H_\psi(\tilde{\delta}, \psi_0) = -IA(\psi_0).\]

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

\[\text{plim } T^{-1} H(\tilde{\psi}) = IA(\psi_0),\]

hence the first term in (3) converges in probability to zero. Writing \(T^{-\frac{1}{2}} g_\psi(\tilde{\delta}, \psi_0)\) as

\[T^{-\frac{1}{2}} g_\psi(\tilde{\delta}, \psi_0) = - \frac{1}{2} T^{-\frac{1}{2}} \sum_{j=0}^{T-1} s_j [f_j - f_j^{-2} I_{U^jU(j)}] \]

\[+ \frac{1}{2} T^{-\frac{1}{2}} \sum_{j=0}^{T-1} s_j f_j^{-2} [I_{U^jU(j)} - I_{U^jU(j)}], \quad (7)\]

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 $N(0, \text{IA}(\psi_0))$.

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

$$T^d(\hat{\psi} - \psi_0) \Rightarrow N(0, \text{IA}^{-1}(\psi_0)) \quad (8)$$

On the other hand subtracting $\delta_0$ and multiplying by $T^d$ both sides of (4.7) we have

$$T^d(\hat{\delta} - \delta_0) = \left[ T^{-1}^d \sum_{j=0}^{T} [\hat{f}_j^{-1}Z(j)] \right]^{-1} T^d \sum_{j=0}^{T} f_j^{-1}Z(j), \quad (9)$$

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

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

$$\text{plim} \ T^{-1} T^{-1}^d \sum_{j=0}^{T} [\hat{f}_j^{-1}Z(j)] = \text{IA}(\delta_0)$$

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

$$T^d \sum_{j=0}^{T} f_j^{-1}Z(j) \Rightarrow N \left[ 0, \text{IA}(\delta_0) \right].$$

Hence, using Slutsky's theorem we have

$$T^d(\hat{\delta} - \delta_0) \Rightarrow N \left[ 0, \text{IA}^{-1}(\delta_0) \right]. \quad (10)$$
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

\[ y_{lt} = y_{lt}' \beta_1 + Z_{lt}' \gamma_1 + w_{lt} - x_{lt}' \delta_1 + w_{lt} \]  
\[ y_{lt}' = Z_{lt}' \Pi_1 + Z_{2t}' \Pi_2 + v_{lt}' = z_{t}' \Pi + v_{lt}' \]  

where \( y_{lt}' \) is a \( l \times p_1 \) row vector of observations on the endogenous variables and \( Z_{lt}' \) is a \( l \times k_1 \) row vector of observations on the exogenous variables appearing in the first equation. \( z_{t}' = (Z_{lt}', Z_{2t}') \) is the \( l \times K \) row vector of observations on all exogenous variables appearing in the system. The \( l \times (p_1+1) \) row vector of disturbances, \([w_{lt} \, v_{lt}]\), is assumed to follow a multivariate random walk plus noise model with associated covariance matrices \( \Sigma_{\epsilon}^\dagger \) and \( \Sigma_{\eta}^\dagger \).

The estimator of the parameter vector \( \theta = (\beta_1, \gamma_1, \sigma^2_{\epsilon}, \sigma^2_{\eta}) \), where \( \sigma^2_{\epsilon} \) and \( \sigma^2_{\eta} \) are respectively the top left hand elements of \( \Sigma_{\epsilon}^\dagger \) 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 $\Sigma_\epsilon^\dagger$ 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 $\text{plim} \frac{1}{t} y_1' \Omega^{-1} w_1 \to 0$, where $\Omega$ 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
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,

\[ y = X \delta + w \]  \hspace{1cm} (1)

where \( X = (Y_1: Z_1) \) is a \( T \times (p_1+k_1) \) matrix of observations on the regressors and \( w \) is a \( T \times 1 \) vector of disturbances which we shall, initially, assume to be independent with zero mean and variance \( \sigma^2 \).

Let \( Z \) be a \( T \times k \) matrix containing the instruments. Pre-multiplying (1) by \( Z' \) yields

\[ Z' y = Z' X \delta + Z' w. \]  \hspace{1cm} (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

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

or

\[ \hat{\delta} = (M'X)^{-1} M'y \]  \hspace{1cm} (4)

where
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, $Ew'w = \sigma^2 \Omega$ where $\Omega$ 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 $\delta$ all based on the Instrumental Variable Principle. These are:

i) The OLS analog

The Ordinary Least Squares analog of $\delta$ 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

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

(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 \( F^{-1}L \) 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

\[
F^{-1}Ly = F^{-1}LX\hat{\delta} + F^{-1}Lw. \tag{7}
\]

Now pre-multiplying (7) by \( Z'L'F^{-1} \), regarding \( \sigma^2Z'\Omega^{-1}Z \) 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

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

where \( P_1 \) is the idempotent projection matrix given by

\[
P_1 = F^{-1} LZ (Z'\Omega^{-1}Z)^{-1} Z'L'F^{-1}. \tag{10}
\]

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
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 primarily 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

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

or

$$\hat{\delta}_2 = \left[X'LF^{-\frac{1}{2}}P_2F^{-\frac{1}{2}}LX\right]^{-1}X'LF^{-\frac{1}{2}}P_2F^{-\frac{1}{2}}Ly \quad (12)$$

where

$$P_2 = Z(Z'Z)^{-\frac{1}{2}}Z'. \quad (13)$$

The only difference between (12) and (9) is the replacement of $P_1$ by
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 $\delta$ 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 $\delta_0$, where $\delta_0$ is the true parameter vector, from (9) and multiplying by $T^\dagger$ yields

$$T^\dagger(\hat{\delta} - \delta_0) = \frac{X'\Omega^{-1}Z}{T} \left[ \frac{Z'\Omega^{-1}Z}{T} \right]^{-1} \frac{Z'\Omega^{-1}X}{T} \left[ \frac{Z'\Omega^{-1}Z}{T} \right]^{-1} T^\dagger Z'\Omega^{-1}w.$$

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

(i) $\text{plim} \ T^{-1}X'\Omega^{-1}Z$ exists and has full column rank;
(ii) $\text{plim} \ T^{-1}(Z'\Omega^{-1}Z)$ exists and is positive definite;
(iii) $\text{plim} \ T^{-1}(Z'\Omega^{-1}w) = 0.$
Furthermore if (i)-(iii) hold and in addition

(vi) $T^{-1} Z' \Omega^{-1} w$ has a limiting normal distribution with zero mean and covariance matrix $\text{plim} \sigma^2 T^{-1} Z' \Omega^{-1} Z$, then $T^{-1} (\hat{\delta} - \delta_0)$ has a limiting normal distribution with zero mean and covariance matrix $V$, where

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

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

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

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,

$$y_t = X_t' \delta + w_t. \quad (1)$$

where $w_t$ follows a random walk plus noise process with associated parameters $\sigma^2_\varepsilon$ and $\sigma^2_\eta$. Again we have omitted the subscript indicating that (1) is a single equation from a simultaneous equation.
Generalizations to the local linear trend model are immediate. As in chapters 2 and 3 the initial state $\mu_0$ 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 $\sigma^2=\sigma^2_\epsilon$ if $\psi$, the signal-noise ratio, is $\sigma^2_\eta/\sigma^2_\epsilon$ and $\sigma^2_\epsilon=\sigma^2_\eta/\sigma^2_\eta$.

When the hyperparameters $\sigma^2$ and $\psi$ 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 $\sigma^2$ and $\psi$ to be known, the three possible IV estimators that we shall consider onwards can be regarded as the ones obtained from the minimand.
\[ S_i(\delta, \psi_\star) = (\tilde{y} - \tilde{X}\delta)' P_i^\dagger P_i P_i^\dagger (\tilde{y} - \tilde{X}\delta), \quad i=1,2,3 \] (2)

where

\[ P_1 = F_i^\dagger Z(Z'Z)^{-1} Z' P_i \] (3)
\[ P_2 = Z(Z'Z)^{-1}Z' \] (4)
\[ P_3 = A[Z(AZ'AZ)^{-1}]AZ' \] (5)

We shall now construct estimators of the regression coefficients 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

\[
\ell(\delta, \sigma^2_\star, \psi_\star) = -\frac{1}{2} \sum_{t=2}^{T} \log f_t - \frac{1}{2} (T-1) \log \sigma^2_\star - \frac{1}{2} \sigma^{-2}_\star S(\delta, \psi_\star) \] (6a)

where

\[ S(\delta, \psi_\star) = (\tilde{y} - \tilde{X}\delta)' P_1 (\tilde{y} - \tilde{X}\delta), \] (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 \( \sigma^2_\star, \psi_\star \). 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 \( \delta \), say \( \tilde{\delta} \), is available. Later we shall discuss the construction of such an estimator. We replace \( \delta \) in (6b) by \( \tilde{\delta} \) and optimise (6a) with respect to \( \sigma^2_\star \) and \( \psi_\star \). We note that \( \sigma^2_\star \) can be
concentrated out of (6a), hence the maximisation is nonlinear only with respect to $\psi$. The resulting estimators of $\sigma^2_*$ and $\psi_*$ 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 $\delta$ and $\sigma^2_*$, $\psi_*$ 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 $\delta$ 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 $\psi_*$. The resulting concentrated criterion function becomes

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

(7)

where

$$S_c (\psi_*) = (\tilde{y} - \tilde{X} \hat{\delta})' \tilde{f}^1 (\tilde{y} - \tilde{X} \hat{\delta})$$

(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 $\psi_\ast$. In practice $\sigma_\ast^2$ can be concentrated out. Once we have found the optimal point $\hat{\psi}_\ast$, $\hat{\delta}$ 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 $\hat{\theta}=\hat{(\delta,\sigma_\ast^2,\psi_\ast)}$ that optimises (6) subject to the restriction

$$\hat{\delta} - [\tilde{X}' P^{-1} X]^{-1} \tilde{X}' P^{-1} y = 0,$$

where $P$ is any of the projection matrices given before. On the other hand the IIIV/ML estimates for the hyperparameters are those which optimise (6) conditional on a given $\delta$. Let $\hat{(\sigma_\ast^2,\psi_\ast)}$ be the point that optimises (6) conditional on $\hat{\delta}$, where $\hat{\delta}$ is the IV/QML estimate. Of course, $\hat{(\sigma_\ast^2,\psi_\ast)}$ will be different from $\hat{(\sigma_\ast^2,\psi_\ast)}$ 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 \( \delta \). 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

\[
\Delta y_t = \Delta X_t' \delta + u_t,
\]

(10)

where \( u_t = \eta_t + \Delta \epsilon_t \). The 2SLS of \( \delta \) is then given by

\[
\hat{\delta}_{2SLS} = \left[ \Delta Z' \Delta Z \right]^{-1} \Delta Z' \Delta y.
\]

(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.,

\[
T^\frac{1}{2} \left( \hat{\delta}_{2SLS} - \delta_0 \right) \xrightarrow{d} N \left( 0, V \right),
\]

where

\[
V = \text{plim} \left[ \begin{array}{c} \Delta Z' \Delta Z \\ R' T R \end{array} \right]^{-1} \left[ \begin{array}{c} \Delta Z' \Omega_u \Delta Z \\ R' T R \end{array} \right] \left[ \begin{array}{c} \Delta Z' \Delta Z \\ R' T R \end{array} \right]^{-1},
\]

(12)

where

\[
\Omega_u = \text{Eu}u' \quad \text{and} \quad R = \begin{bmatrix} \Pi_1 I & k \end{bmatrix}.
\]
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 $\Sigma_z$, then

$$\lim_{T \to \infty} \frac{\Delta Z' \Omega_u \Delta Z}{T} = (2\sigma^2_\varepsilon + \sigma^2_\eta) \Sigma_z,$$

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

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

There are 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 $\delta$ can be regarded as the
resulting estimator obtained from (3.2.9), by making use of asymptotically equivalent expressions similar to

\[ \bar{x}'F^{-1}\bar{x} \approx \sum_{j=0}^{T-2} I_{x'x}(j)/f_j, \]  

(1)

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

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

(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 \(\delta\).

Asymptotic Properties

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

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

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 $\delta_o$ and multiplying by $T^\dagger$ yields

$$ T^\dagger (\hat{\delta} - \delta_o) = \hat{H} T^\dagger \sum_{j=0}^{T-2} \frac{I_z'u(j)}{\hat{f}_j} $$

(3)

where

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

(4)

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

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

$$\Delta X = \Delta Z R + (\Delta Y_1: 0)$$  \hspace{1cm} (5)

where

$$R = \begin{bmatrix} \Pi_1 & I_k \\ \Pi_2 & 0 \end{bmatrix}. \hspace{1cm} (6)$$

It now follows from (3.4.2) and (3.4.1) that if $\Delta Z$ is totally independent from $\Delta Y_1$, $\hat{H}$ converges in probability to $H$, where $H$ is given by

$$H = \left[ \frac{\Pi_1}{\Pi_2} \right]^{-1} \int_{-\pi}^{\pi} f_u^{-1}(\lambda)F_z(\lambda) d\lambda \left[ \frac{\Pi_1}{\Pi_2} \right]. \hspace{1cm} (7)$$

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

$$T \delta \overset{d}{\rightarrow} N(0, V^{-1})$$

where

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

We note that if the exogenous variables, $z_t$, follow a multivariate random walk process with disturbance covariance matrix $\Sigma_z$ 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
\[
Avar \hat{\delta} = \frac{4\pi^2}{T} \left[ R'S_{\hat{z}} \int_{-\pi}^{\pi} \left( f_{u}(\lambda) \right)^{-1} d\lambda \right]^{-1},
\]

(9)

which can be written as

\[
Avar \hat{\delta} = T^{-1} \left[ \sigma_{\eta}^4 + 4\sigma_{\eta}^2\sigma_{\varepsilon}^2 \right]^\frac{1}{2} \left[ R'S_{\hat{z}} R \right]^{-1},
\]

(10)

since

\[
\int_{-\pi}^{\pi} \left( f_{u}(\lambda) \right)^{-1} d\lambda = 2\pi \int_{-\pi}^{\pi} \left[ 2\sigma_{\varepsilon}^2 + \sigma_{\eta}^2 - 2\sigma_{\varepsilon}^2 \cos \lambda \right]^{-\frac{1}{2}} d\lambda
\]

\[
= 4\pi^2 \left[ \sigma_{\eta}^4 + 4\sigma_{\eta}^2\sigma_{\varepsilon}^2 \right]^{-\frac{1}{2}}.
\]

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 \( \delta \) then becomes

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

(11)

Making use of (11) rather than (2) we can obtain \( \hat{\theta} = (\hat{\delta}, \hat{\psi}) \) in the same way as we have obtained the spectral IIIV/ML. Clearly the resulting estimator is as efficient as the spectral IIIV/ML. We note that a consistent estimator of \( \hat{R} \) can be obtained estimating \( \Pi \) 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 \( \Pi \) given in (3.3.11) is already consistent.

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

\[
\delta = [R'(\Delta Z'\Delta Z)R]^{-1} R'\Delta Z'\delta_y.
\]

Now if \( \hat{R} \) is estimated by means of LS then \( \hat{\delta} \) 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

\[
\hat{\delta}(\psi^*) = \left[ R' \sum_{t=2}^{T} f_t^{-1} z_t \hat{z}_t R \right]^{-1} R' \sum_{t=2}^{T} f_t^{-1} z_t y_t,
\]

and similarly 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
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, \( \sigma_{\varepsilon}^2 \) goes to zero, 2SLS tends to the same distribution as IIV/ML as the disturbances in the differenced observations are tending to white noise.
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

\[ x_t' A - y_t' \beta + z_t' \Gamma - \omega_t' , \quad t = 1, \ldots, T, \quad (1a) \]
where \( x_t' = (y_t', z_t') \) and \( A = (B', \Gamma')' \). \( y_t \) is a pxl vector of observed endogenous variables, \( z_t \) is a Kxl vector of observed exogenous variables. The non-singular pxp matrix \( B \) and the Kxp matrix \( \Gamma \) are unknown fixed parameters matrices of the endogenous and of the exogenous variables respectively. The pxl 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.

\[
\begin{align*}
\omega_t &= \mu_t + \epsilon_t, \\
\mu_t &= \mu_{t-1} + \eta_t,
\end{align*}
\]

with covariances matrices \( \Sigma_\epsilon \) and \( \Sigma_\eta \). The reduced form of (1) is

\[
y_t' = z_t' \Pi + \nu_t',
\]

where

\[
\begin{align*}
\Pi &= - \Gamma B^{-1}, \\
\nu_t' &= w_t' B^{-1} - \mu_t^* + \epsilon_t^*, \\
\mu_t^* &= \mu_{t-1}^* + \eta_t^*.
\end{align*}
\]

The covariance matrices of \( \epsilon_t^* \) and \( \eta_t^* \) are respectively

\[
\Sigma_\epsilon^* = B^{-1} \Sigma_\epsilon B^{-1} \quad \text{and} \quad \Sigma_\eta^* = B^{-1} \Sigma_\eta B^{-1}.
\]

Combining the observations we define \( X = [Y : Z] \).

\[
Y = \begin{bmatrix} y_{1T}' \\ \vdots \\ y_{T}' \end{bmatrix} = \begin{bmatrix} y_{11} \ldots y_{1p} \\ \vdots \\ y_{T1} \ldots y_{Tp} \end{bmatrix}, \quad Z = \begin{bmatrix} z_{1T}' \\ \vdots \\ z_{T}' \end{bmatrix} = \begin{bmatrix} z_{11} \ldots z_{1K} \\ \vdots \\ z_{T1} \ldots z_{KT} \end{bmatrix}
\]

and similarly
\[
\begin{align*}
W &= \begin{bmatrix} w_1' \\ \\ w_T' \end{bmatrix}, & \mu &= \begin{bmatrix} \mu_1' \\ \\ \mu_T' \end{bmatrix}, & \epsilon &= \begin{bmatrix} \epsilon_1' \\ \\ \epsilon_T' \end{bmatrix}, & \eta &= \begin{bmatrix} \eta_1' \\ \\ \eta_T' \end{bmatrix}.
\end{align*}
\]

We use the notation \( X_i \) for any submatrix of \( X \). Thus, \( X_i = [Y_i:Z_i] \), where \( Y_i \) is a \( T \times p_i \) submatrix of \( Y \) and \( Z_i \) is a \( T \times k_i \) 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

\[
XA = YB + Z \Gamma = W,
\]

where

\[
W = \mu + \epsilon, \quad (3a)
\]

\[
\mu = \mu(-1) + \eta, \quad (3b)
\]

and its reduced form as

\[
Y = Z \Pi + V, \quad (4a)
\]

where

\[
V = \mu^* + \epsilon^*, \quad (4b)
\]

\[
\mu^* = \mu^*(-1) + \eta^*. \quad (4c)
\]

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 θ. We shall only consider zero restrictions, that is, the corresponding variable is excluded from the equation in question, plus normalisation constraints, that is $β_i = 1$, for $i = 1, ..., p$, 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 $r \times 1$ vector $δ$,

$$δ' = [δ_1' \ldots δ_p'], \tag{5}$$

where

$$δ_1' = [β_1' : γ_1']. \tag{6}$$

and the $p_i \times 1$ 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 $θ$, so that $r = ∑ (p_i + k_i)$. We define the $r \times p(p+K)$ selection matrix $S_A'$ such that

$$δ = - S_A' vec A. \tag{7}$$

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

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

$$S_A δ = - vec(A^+), \tag{9}$$
vec(A) = - S_A \delta + s, \quad (10)

where \( A^+ = A - J \), \( J = [I_p : O_{p \times k}]' \) and \( s = \text{vec}J \). Also

\[ S_A = - \frac{\partial}{\partial \delta} \text{vec}(A), \quad (11) \]

and

\[ (I_p \otimes X)S_A = \begin{bmatrix} X_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots \\ 0 & 0 & 0 & X_p \end{bmatrix}, \quad (12) \]

where \( X_i = [Y_i : Z_i] \) is the \( T x(p_i + k_i) \) 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 explanatory 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

\[ y_{it} = X_{it}'\delta_i + w_{it}, \quad t=1,\ldots,T, \quad i=1,\ldots,p \quad (13) \]

where \( X_{it}' = [Y_{it}' : Z_{it}'] \) and \( \delta_i \) is given in \((6)\). The entire system can then be written as

\[
\begin{bmatrix}
X_1 & 0 & 0 & 0 \\
0 & X_2 & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots \\
0 & 0 & 0 & X_p
\end{bmatrix} \delta + \text{vec} W. \quad (14)
\]

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 identifiability 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 \( \theta \) is a p-dimensional parameter vector. Suppose we intend to estimate \( \theta \) by maximum likelihood. The identification assumption states there cannot exist \( \theta^1 \neq \theta \) such that \( \ell(\theta^1; x) = \ell(\theta; x) \) for all \( x \), where \( \ell(\theta; x) \) is the loglikelihood function. If two points, \( \theta^1 \) and \( \theta \), 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 \( \theta^+ \) can be produced by premultiplying the transpose of (2.1a), that is,

\[
B'y_t + \Gamma'z_t = w_t,
\]

by a nonsingular \( pxp \) matrix \( F \); see Hsaio(1983) for a full discussion.

We note that if no restrictions are placed on \( B \) and \( \Gamma \), 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 assumptions. Hence \( B \) and \( \Gamma \) cannot be determined from the reduced form. However if \textit{a priori} restrictions on \( B \) and \( \Gamma \) are placed then of course \( F \) must be such that \( FB' \) satisfies the same \textit{a priori} restrictions as \( B' \), and \( FF' \) the same as \( \Gamma' \). The model is identified if the only matrix \( F \) which yields a system satisfying the same \textit{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 \( \beta_i \) and \( \gamma_i \), the unknown elements of the \( i \)-th rows of \( B' \) and \( \Gamma' \), 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 $\Gamma'$ appear last, we have

\[
\begin{bmatrix}
1 & \beta_1' & 0' \\
- & - & - \\
B_{21}' & B_{22}'
\end{bmatrix}
\begin{bmatrix}
y_t \\
z_t
\end{bmatrix} + 
\begin{bmatrix}
\gamma_1' & 0' \\
- & - & - \\
\Gamma_{21}' & \Gamma_{22}'
\end{bmatrix}
\begin{bmatrix}
\gamma_t \\
y_{t-1}
\end{bmatrix} = \omega_t
\]  

(2)

where $\beta_1'$ is $1 \times p_1$, $B_{21}'$ is $(p-1) \times (p_1+1)$, $B_{22}'$ is $(p-1) \times (p-p_1-1)$, $\gamma_1'$ is $1 \times k_1$, $\Gamma_{21}'$ is $(p-1) \times k_1$ and $\Gamma_{22}'$ is $(p-1) \times (K-k_1)$. Thus the matrices $B_{22}'$ and $\Gamma_{22}'$ are submatrices of $B'$ and $\Gamma'$ 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

\[
y_{1t} = X_{1t}' \delta_1 + w_{1t}, \quad t=1, \ldots, T,
\]  

(3)

is identifiable, where $X_{1t}' = [Y_{1t}' : Z_{1t}']$ 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 $(p-1) \times 1$ vector $f$ is null. In the classical model, it is well known that a necessary condition for identifiability is the order condition, $K > p_1 + k_1$. A necessary and sufficient condition, the rank condition, requires that

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

(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 \( F' = 0 \), that is, no exogenous variable are included in the model. We also assume \( B' \) normalised. The model is then

\[
B'y_t = \mu_t + \epsilon_t \quad (5a)
\]

\[
\mu_t = \mu_{t-1} + \eta_t, \quad t = 1, \ldots, T. \quad (5b)
\]

with associated parameters contained in \( \theta \), \( \theta = (B', \Sigma_\epsilon, \Sigma_\eta) \).

Premultiplying (5a) by \( B'^{-1} \) we obtain the reduced form

\[
y_t = \mu^*_t + \epsilon^*_t \quad (6a)
\]

\[
\mu^*_t = \mu^*_{t-1} + \eta^*_t \quad (6b)
\]

with associated parameters being the covariances matrices \( \Sigma^*_\epsilon = B'^{-1}\Sigma_\epsilon B^{-1} \) 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 \( \theta^f = (FB', F\Sigma_\epsilon F', F\Sigma_\eta F') \). 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 \( \mu_{i0} \), \( i=1,2 \), is fixed,
we can express \( \mu_t \) in terms of a deterministic and a stochastic part.
The model is then

\[
\begin{bmatrix}
1 & \beta_1 \\
\beta_2 & 1 \\
\end{bmatrix}
\begin{bmatrix}
y_{1t} \\
y_{2t} \\
\end{bmatrix}
= \begin{bmatrix}
\mu_{10} \\
\mu_{20} \\
\end{bmatrix} + \begin{bmatrix}
\mu_{1t}^\dagger \\
\mu_{2t}^\dagger \\
\end{bmatrix} + \begin{bmatrix}
e_{1t} \\
e_{2t} \\
\end{bmatrix}, \quad t=1,\ldots,T
\]
\]

(7a)

where

\[
\mu_{it}^\dagger = \mu_{i,t-1}^\dagger + \eta_{it}, \quad t=1,\ldots,T ,
\]

(7b)

with \( \mu_{i0}^\dagger = 0 \) 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 reparameterised as

\[
\begin{bmatrix}
1 & \beta_1 \\
\beta_2 & 1 \\
\end{bmatrix}
\begin{bmatrix}
y_{1t} \\
y_{2t} \\
\end{bmatrix}
= \begin{bmatrix}
\mu_{10} \\
\mu_{20} \\
\end{bmatrix} + \begin{bmatrix}
0 \\
\bar{s} \mu_{2t}^\dagger \\
\end{bmatrix} + \begin{bmatrix}
e_{1t} \\
e_{2t} \\
\end{bmatrix}, \quad t=1,\ldots,T
\]

\]

(8a)

where

\[
\bar{\mu}_{2t}^\dagger = \bar{\mu}_{2,t-1}^\dagger + \bar{\eta}_{2t}, \quad t=1,\ldots,T
\]

(8b)

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

Now the reduced form of model (8) is

\[
y_{1t} = \pi_{11} + \pi_{12} \bar{\mu}_{2t}^\dagger + \nu_{1t}, \quad t=1,\ldots,T
\]

(9a)

\[
y_{2t} = \pi_{21} + \pi_{22} \bar{\mu}_{2t}^\dagger + \nu_{2t}, \quad t=1,\ldots,T
\]

(9b)
where
\[
\begin{align*}
\tau_{11} & = (\mu_{10} - \beta_1 \mu_{20})/(1 - \beta_1 \beta_2) \\
\tau_{21} & = (\mu_{20} - \beta_2 \mu_{10})/(1 - \beta_1 \beta_2) \\
\tau_{12} & = -s \beta_1/(1 - \beta_1 \beta_2) \\
\tau_{22} & = -s/(1 - \beta_1 \beta_2).
\end{align*}
\]

This reduced form contains a common stochastic trend component \( \mu_{2t} \).

Estimators of the parameters \( \tau_{11}, \tau_{21}, \tau_{12}, \tau_{22} \) can be computed by ML and unique estimators of the structural parameters \( \mu_{10} \) and \( \beta_1 \) obtained by noting that

\[
\beta_1 = -\tau_{12}/\tau_{22}
\]

and

\[
\mu_{10} = \tau_{11} - \tau_{12} \tau_{21}/\tau_{22}.
\]

The first equation is therefore exactly identified. If it were known that \( \mu_{10} \) were zero, it would be overidentified as \( \beta_1 \) could also be estimated from \( \tau_{11}/\tau_{12} \). 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 \( \mu_{1t} \). If (8a) is written as

\[
\begin{bmatrix}
[1 & 0]
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
\mu_{10} \\
\mu_{20}
\end{bmatrix}
\begin{bmatrix}
1 \\
s
\end{bmatrix}
+ \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}, \quad t=1,\ldots,T
\]

(10)
and $\mu_{2t}^\dagger$ is regarded as an explanatory variable the rank condition given in (4) is verified provided that $s$ 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 initial 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 $\mu_{20}$ and $\mu_{30}$. The model is then

$$B' \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_{2t} \\ \mu_{3t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix}, \quad t = 1, \ldots, T \quad (11)$$

where $S$ is any matrix such that $S' S = \Sigma_\eta$, and $(\mu_{2t}, \mu_{3t})'$ 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

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

and

$$\Sigma_\eta^* = B' \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma_\eta & 0 \\ 0 & 0 & \Sigma_\eta \end{bmatrix} B^{-1}.$$
Now conditionally on the first observations the covariance matrices $\Sigma^*_c$ and $\Sigma^*_\eta$ can be estimated by ML, and unique estimators of the structural form parameters appearing in the first equation can be obtained by noting that since

$$B'\Sigma^*_\eta = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma^*_\eta & 0 \end{bmatrix} B^{-1}$$  \hspace{1cm} (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 $\Sigma^*_\eta$, the covariance matrix associated with the stochastic trends appearing in all equations other than the first equation has full rank, or equivalently if $\Sigma^*_\eta$ has full rank, where $\Sigma^*_\eta = \Sigma$.

We now consider the identifiability of the first equation in a general model with exogenous variables. Assuming a diffuse prior for $\mu_0 = (\mu_{20}, \mu_{30}, \ldots, \mu_{p0})'$ and no stochastic trend component at all in the first equation, the model becomes
where $\mu_t$ is a $(p-1) \times 1$ multivariate random walk with associated covariance matrix being the identity matrix. $S$ is any $(p-1) \times (p-1)$ matrix such that $S'S = \Sigma$. A necessary and sufficient condition for the identifiability of the first equation is that

$$\text{Rank } [B_{22}' \Gamma_{22}' S'] = p-1.$$ (14)

On the other hand if the initial state $\mu_0$ is regarded as being fixed, the necessary and sufficient condition is that

$$\text{Rank } [B_{22}' \Gamma_{22}' \mu_0 S'] = p-1.$$ (15)

Local Linear Trends

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

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

it follows from (2.2.4) that $w_t$ is given by

$$w_t = \mu_t + \epsilon_t,$$ (16a)

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

$$\delta_t = \delta_{t-1} + \xi_t.$$ (16c)
where $\eta_t$ and $\xi_t$ are assumed to be independent of each other with associated covariance matrices $\Sigma_\eta$ and $\Sigma_\xi$ respectively. The initial state is assumed to be fixed. Let the $p \times p$ positive definite matrices $S_\eta$, $S_\xi$ and $S_*$ such that $S_\eta' S_\eta = \Sigma_\eta$, $S_\xi' S_\xi = \Sigma_\xi$ and $S_* S_\eta = S_\xi$. Clearly $S_\eta$, $S_\xi$ and $S_*$ are not unique. We reparametrise (16) as

\begin{align*}
\nu_t &= S_\eta' \mu_t + \epsilon_t, \quad (17a) \\
\bar{\mu}_t &= \bar{\mu}_{t-1} + S_*' \delta_{t-1} + \eta_t, \quad (17b) \\
\bar{\delta}_t &= \delta_{t-1} + \xi_t, \quad (17c)
\end{align*}

where $\text{Var} \eta_t = I_p$ and $\text{Var} \xi_t = I_p$.

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

\begin{align*}
\nu_t &= \mu_0 + \delta_0 t + S_\eta' \mu_t + S_\xi' \delta_t + \epsilon_t, \quad (18a) \\
\bar{\mu}_t &= \bar{\mu}_{t-1} + \eta_t, \quad \bar{\mu}_0 = 0, \quad t=1,\ldots,T \quad (18b) \\
\bar{\delta}_t &= 2\delta_{t-1} - \delta_{t-2} + \xi_t, \quad \delta_0 = \delta_1 = 0, \quad t=2,\ldots,T \quad (18c)
\end{align*}

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

\begin{align*}
\begin{bmatrix}
1 & \beta_1' & 0' \\
& \cdots & \cdots \\
& & \cdots & \cdots \\
B_{21}' & B_{22}'
\end{bmatrix}
\begin{bmatrix}
y_t \\
\vdots \\
\cdots \\
z_t
\end{bmatrix}
= \begin{bmatrix}
\gamma_1' & 0' \\
& \cdots & \cdots \\
& & \cdots & \cdots \\
\Gamma_{21}' & \Gamma_{22}'
\end{bmatrix}
\begin{bmatrix}
y_t \\
\vdots \\
\cdots \\
z_t
\end{bmatrix}
+ \begin{bmatrix}
0 \\
0 \\
0 \\
\delta_0
\end{bmatrix}
\begin{bmatrix}
1 \\
\mu_{t-1} \\
\end{bmatrix}
+ \begin{bmatrix}
0 \\
\mu_0 \\
\eta_t \\
\xi_t
\end{bmatrix}
+ \begin{bmatrix}
0' \\
S_{\eta}' \\
S_{\xi}'
\end{bmatrix}
\begin{bmatrix}
\mu_t \\
\delta_t
\end{bmatrix}
+ \epsilon_t, \quad (19)
\end{align*}
where \( \bar{\mu}_t \) and \( \bar{\delta}_t \) are \((p-1)\times 1\) vectors obeying the equations of the form (18b) and (18c). \( \bar{S}_\eta \) and \( \bar{S}_\tau \) are \((p-1)\times(p-1)\) matrices such that the covariance matrices of the \((p-1)\times 1\) disturbance vectors \( \bar{\eta}_t \) and \( \bar{\tau}_t \) are identity matrices. Hence the necessary and sufficient condition for the identifiability of the first equation is that

\[
\text{Rank } [B_{22}' \Gamma_{22}' \bar{\mu}_0 \bar{\delta}_0 \bar{S}_\eta \bar{S}_\tau] = p-1.
\]

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 \( \bar{S}_\tau \) 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 \( \bar{\mu}_t \) and \( \bar{\delta}_t \), 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 \( \eta_t \) and \( \tau_t \) 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, \( \theta^* = (\Pi, \Sigma_e^*, \Sigma_\eta^*) \), is given by

\[
\ell(\theta^*) = -\frac{1}{2} \sum_{t=2}^{T} \log \| F_t^* \| - \frac{1}{2} \sum_{t=2}^{T} V_t^* (F_t^*)^{-1} V_t^* ,
\]

(1)

where \( V_t^* \) 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 \( \theta = (B, \Gamma, \Sigma_e, \Sigma_\eta) \) we note that premultiplying (2.2.13a), specialized for the multivariate random walk plus noise model, by \( B' \) yields

\[
B' a_t^* = B' a_{t-1}^* + B' P_{t-1}^* B (B' F_t^* B)^{-1} B' V_t^* ,
\]

(2)

where the superscript \(*\) indicates that we are dealing with the reduced form, and \( v_t^* = V_t - a_{t-1}^* \). 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

\[ a_t = a_{t-1} + P_{t/t-1}F_{t-1}^{-1}r_t, \quad (3) \]

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

\[ \sum_{t=2}^{T} r_t'F_{t-1}^{-1}r_t = \sum_{t=2}^{T} r_t^* (F_t^*)^{-1} r_t^*, \]

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

\[ \ell(\theta) = (T-1)\log B + \sum_{t=2}^{T} \log I_{F_{t-1}} + \sum_{t=2}^{T} r_t'F_{t-1}^{-1}r_t. \quad (4) \]

As it stands, to obtain the prediction errors \( r_t \) 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 \( \Sigma_\varepsilon \) and \( \Sigma_\eta \). 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

\[ y_t'B + z_t'\Gamma = u_t' \]  \hspace{1cm} (1a) \\
\[ u_t = \eta_t + \Delta \varepsilon_t \]  \hspace{1cm} (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 \( \Delta \) in front of \( y_t' \) and \( z_t' \). In matrix notation (1) becomes

\[ XA = YB + Z\Gamma - U \]  \hspace{1cm} (2)

Let \( \delta \) denote the vector containing the unknown elements of the matrix A. From chapter 5 we have

\[ \delta = - S_A'\text{vec}(A), \]  \hspace{1cm} (3)
\[ S_{A\delta} = \text{vec}J' \cdot \text{vec}A \]  

where

\[ S_A = - \frac{\partial}{\partial \delta} \text{vec}(A). \]  

and

\[ J = [I_p:0(p \times k)]. \]

As regards the hyperparameters, we have from (2.2.7) that the distinct elements of \( \Sigma_\epsilon \) and \( \Sigma_\eta \) are contained in the vector

\[ \psi = [D^+ \ 0] \text{vec}[\Sigma_\epsilon : \Sigma_\eta] \]  

and from (2.3.13) that

\[ S_j = \frac{\partial \text{vec}F_j}{\partial \psi} = (2\pi)^{-1}[c_jD : D] \]

Now, as the reduced form of (2) is

\[ Y = Z\Pi + UB^{-1}, \]

the spectral likelihood function of \( \text{vec}Y' \), 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

\[ \ell(\theta) = T \log \|B\| - \frac{1}{2} \sum_{j=0}^{T-1} \log |F(\lambda_j)| - \frac{1}{2} \sum_{j=0}^{T-1} \text{tr}F^{-1}(\lambda_j)I_{U'U}(\lambda_j), \]  

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 $A'X\Psi_jXA$ where $\Psi_j$ is defined in (2.3.9) and denoting $F(\lambda_j)$ as $F_j$, (9) can be written as

$$\ell(\theta) = T\log|B| - \frac{1}{4}\sum_{j=0}^{T-1} \log|F_j| + \frac{1}{4}\sum_{j=0}^{T-1} \text{tr} F_j^{-1}(XA)'\Psi_j(XA).$$  \hspace{1cm} (10)

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

$$\frac{\partial}{\partial \theta^T} \ell(\theta) = 0. \hspace{1cm} (11)$$

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 $\psi$ containing the hyperparameters cannot be concentrated out of the spectral likelihood function, except in a rather special case, $\Sigma_\eta$ proportional to $\Sigma_\epsilon$ (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 too much difficulty.

Applying the following results

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

\[(ii) \quad dA^{-1} = -A^{-1} dAA'^{-1}, \quad (12b)\]

\[(iii) \quad \text{tr}(ABCD) = \text{vec}'(D')(C'A)\text{vec}B, \quad (12c)\]

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

\[dQ = T \text{vec}'B^{-1}\text{vec}dB + \frac{1}{2} \sum_{j=0}^{T-1} \text{vec}'[F_j^{-1}A'X'\Psi_jXAF_j^{-1}]\text{vec}dF_j\]

Now because \(\delta\) and \(\psi\) are functionally independent we have from (5) and (7) that

\[\frac{\partial \text{vec}A}{\partial \delta^i} = -(S_A'^0); \quad \frac{\partial \text{vec}B}{\partial \delta^i} = -(I_p @ J)(S_A'^0); \quad \frac{\partial \text{vec}F_j}{\partial \delta^i} = (0:S_j);\]
or

\[ \text{dvecA} = - (S^A_0)d\theta \quad \quad \text{dvecA}' = - K_{(p+k)}(S^A_0)d\theta; \]
\[ \text{dvecB} = - (I_p @ J)(S^A_0)d\theta \quad \quad \text{dvecB}' = - K_p(I_p @ J)(S^A_0)d\theta; \quad (13) \]
\[ \text{dvecF}_j = (0:S_j)d\theta \quad \quad \text{dvecF}_j' = (0:S_j)d\theta. \]

Hence, the first order conditions are

\[ \frac{\partial}{\partial \delta} \ell : - S^\prime \text{vec}\left[ T \prime B^{-1} - \sum_{j=0}^{T-1} X' \Psi_j(XA)F_j^{-1}\right] = 0 \quad (14a) \]
\[ \frac{\partial}{\partial \psi} \ell : - \frac{1}{2} \sum_{j=0}^{T-1} S_j \text{vec}\left[ (F_j^{-1} - F_j^{-1}(XA)' \Psi_j(XA)F_j^{-1})\right] = 0 \quad (14b) \]

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

\[ d^2\ell = - T \text{tr } (B^{-1}dB)^2 + T \text{tr } B^{-1}d^2B + \frac{1}{2} \sum \text{tr } (F_j^{-1}dF)^2 \]
\[ - \frac{1}{2} \sum \text{tr } F_j^{-1}d^2F_j - \sum \text{tr } F_j^{-1}(dA)'X'\Psi_j XdA \]
\[ + 2 \sum \text{tr } F_j^{-1}(dF_j)F_j^{-1}A'X'\Psi_j XdA - \sum \text{tr } F_j^{-1}A'X'\Psi_j X d^2A \]
\[ - \sum \text{tr } A'X'\Psi_j XA(F_j^{-1}dF_j)^2F_j^{-1} + \frac{1}{2} \sum \text{tr } F_j^{-1}A'X'\Psi_j XAF_j^{-1}d^2F_j. \]

Because \( d^2F_j = 0 \) and \( d^2A = 0 \), using (12c) and (13) the expression above can be rewritten as
\[ d^2 \ell = -T (d\theta)' [S_{A:0}'] (I_p \otimes B^{-1}) K_p (B^{-1} \otimes B^{-1}) (I_p \otimes J) [S_{A:0}] \, d\theta \\
+ \frac{1}{2} \sum (d\theta)' [0:S_j]' (F_j^{-1} \otimes F_j^{-1}) [0:S_j] \, d\theta \\
- \sum (d\theta)' [S_{A:0}]' K_{(p+k)p}' (X' \Psi_j X^{-1} A' X' \Psi_j X^{-1} J) K_{(p+k)p} [S_{A:0}] \, d\theta \\
- 2 \sum (d\theta)' [S_{A:0}]' K_{(p+k)p}' (X' \Psi_j X F_j^{-1} A' X' \Psi_j X F_j^{-1} J) [0:S_j] \, d\theta \\
- \sum (d\theta)' [0:S_j]' (I_p \otimes F_j^{-1}) (F_j^{-1} \otimes A' X' \Psi_j X F_j^{-1} J) [0:S_j] \, d\theta \]

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

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

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

\[ d^2 \ell = -(d\theta)' [S_{A:0}]' \left[ T (I_p \otimes B^{-1}) K_p (I_p \otimes B^{-1} J) + \sum (F_j^{-1} \otimes X' \Psi_j X) \right] [S_{A:0}] \, d\theta \\
+ (d\theta)' \sum [0:S_j]' \left[ \frac{1}{2} (F_j^{-1} \otimes F_j^{-1}) - (F_j^{-1} \otimes F_j^{-1} A' \Psi_j X F_j^{-1}) \right] [0:S_j] \, d\theta \\
- 2 (d\theta)' [S_{A:0}]' K_{(p+k)} \sum \left[ (X' \Psi_j X F_j^{-1} \otimes F_j^{-1}) [0:S_j] \right] \, d\theta. \]

Now using the second identification theorem in Magnus and Neudecker (1988, page 189) we have
3. A Computational Method for Finding the Spectral ML Estimator

In order to obtain the spectral FIML (Full Information Maximum Likelihood) for \( \delta \) 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 \( \hat{\psi}(\delta) \), we could construct the concentrated spectral likelihood with respect to \( \delta \) and use the Newton-Raphson method to find the spectral FIML estimator for \( \delta \) by solving

\[
\frac{\partial}{\partial \delta} \ell_c(\delta) = 0
\]  

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 \( \psi \) and so we cannot construct the concentrated likelihood function. Nevertheless, because we only need the first derivatives of \( \ell_c \) at a given \( \delta \), say \( \hat{\delta}_k \), we can still use the iterative procedure. We note that these derivatives, evaluated
at $\hat{d}_k$, are the derivatives of the unconcentrated likelihood with respect to $\delta$ at $(\hat{d}_k, \hat{\psi})$ where $\hat{\psi} = \hat{\psi}(\hat{d}_k)$ is the solution of (14b) with $\delta$ replaced by $\hat{d}_k$. Thus,

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

and whether $\hat{\psi}(\hat{d}_k)$ is obtained analytically or numerically makes no difference. The only issue is with respect to the Hessian matrix of $\ell_c$ 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

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

where $\hat{F}_j = F_j(\hat{\psi})$ and $\hat{\psi} = \hat{\psi}(\delta)$ solves (2.14b) for a given $\delta$. 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

\[ T J B'^{-1} = \sum_{j=0}^{T-1} J'B'^{-1} A' X' \Psi_j XA \hat{F}_j^{-1}, \]  

(3)

where \( J = [I_p : 0_{p \times k}] \). 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

\[ q(\delta) = \frac{\partial}{\partial \delta} \phi(\delta, \hat{\psi}) = - \text{vec} [ T \sum_{j=0}^{T-1} [J'B'^{-1} A' X' \Psi_j XA \hat{F}_j^{-1} - X' \Psi_j XA \hat{F}_j^{-1} ]] \]

\[ = - \text{vec} [ T \sum_{j=0}^{T-1} [Y' \Psi_j XA \hat{F}_j^{-1} + B'^{-1} \Gamma' Z' \Psi_j XA \hat{F}_j^{-1} ] - [Y' \Psi_j XA \hat{F}_j^{-1} ] ] \]

\[ = \text{vec} [ T \sum_{j=0}^{T-1} [B'^{-1} \Gamma' Z' ] \Psi_j XA \hat{F}_j^{-1} ] . \]  

(4)

Hence, the spectral FIML estimator of \( \delta \) emerges as a solution of

\[ q(\delta) = \text{vec} [ T \sum_{j=0}^{T-1} P' Z' \Psi_j XA \hat{F}_j^{-1} ] = 0 \]  

(5)

where

\[ P = [ \Pi : I_k ] \]  

(6)

and

\[ \Pi = - \Gamma B^{-1} . \]  

(7)

The solution of (5) can be obtained by the Newton-Raphson iterative procedure described in appendix 2.1. The first differential, \( dq(\delta) \)
\[ dq(\delta) = S_A' \text{vec} \sum_{j=0}^{T-1} \left[ dP'Z'\Psi_jX\hat{F}_j^{-1} + P'Z'\Psi_jX dA \hat{F}_j^{-1} + P'Z'\Psi_jX d\hat{P}_j^{-1} \right] \]

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

\[ \text{plim} \ T^{-1} \text{vec} \sum_{j=0}^{T-1} dP'Z'\Psi_jX\hat{F}_j^{-1} - (I_p @ dP') \text{plim} \ T^{-1} \sum_{j=0}^{T-1} (\hat{F}_j^{-1} @ I_k) \text{vec} IZ'U(j) = 0 \]

and

\[ \text{plim} \ T^{-1} \text{vec} \sum_{j=0}^{T-1} P'Z'\Psi_jX d\hat{F}_j^{-1} - (I_p @ P') \text{plim} \ T^{-1} \sum_{j=0}^{T-1} (d\hat{F}_j @ I_k) \text{vec} IZ'U(j) = 0. \]

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

\[ H = - S_A' \left[ \sum_{j=0}^{T-1} \hat{F}_j^{-1} @ P'Z'\Psi_jX \right] S_A, \quad (8) \]

in the sense that

\[ \text{plim} \ T^{-1} dq(\delta) = \text{plim} \ T^{-1} S_A' \text{vec} \left[ \sum_{j=0}^{T-1} P'Z'\Psi_jX(dA)\hat{F}_j^{-1} \right] S_A \]

\[ = \text{plim} \ T^{-1} Hd\delta, \]

where the last equality above follows from (2.13).

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

\[ \text{plim} \ T^{-1} dq(\delta) = - \text{plim} \ T^{-1} S_A' \left[ \sum_{j=0}^{T-1} (\hat{F}_j^{-1} @ P'Z'\Psi_jZP) \right] S_A d\delta \]

\[ - \text{plim} \ T^{-1} S_A' \left[ \sum_{j=0}^{T-1} (\hat{F}_j^{-1} @ P'Z'\Psi_j[V;0]) \right] S_A d\delta. \quad (9) \]
Now, since \( z \) and \( v \) 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 \( H \), where

\[
H = - S_A' \left[ \sum_{j=0}^{T-1} (\hat{F}_j^{-1} \hat{P}' Z' \hat{Z} P) \right] S_A .
\] (10)

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

\[
\hat{\delta}_{k+1} = \hat{\delta}_k + \left[ S_A' \sum_{j=0}^{T-1} (\hat{F}_j^{-1} \hat{P}' Z' \hat{Z} P) S_A \right]^{-1} \left[ S_A' \text{vec} \sum_{j=0}^{T-1} (\hat{P}' Z' \hat{Z} X \hat{A} \hat{F}_j^{-1}) \right]
\] (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 \( Z' \hat{\psi} Z = I_{Z'Z}(\lambda_j) \) and \( Z' \hat{\psi} X = I_{Z'X}(\lambda_j) \) 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

\[
\hat{\delta}_{k+1} = \hat{\delta}_k + \left[ S_A' \sum_{j=0}^{T-1} (\hat{F}_j^{-1} \hat{P}' I_{Z'Z}(\lambda_j) \hat{P}) S_A \right]^{-1} \left[ S_A' \text{vec} \sum_{j=0}^{T-1} (\hat{P}' I_{Z'X}(\lambda_j) \hat{X} \hat{A} \hat{F}_j^{-1}) \right]
\] (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 \( \hat{U} = \hat{X} \hat{A} \). Thus,

\[
\hat{F}_j(\lambda_j) = \frac{1}{m_j} \sum_{\omega \in A_j} I_{\hat{U}}(\omega) \quad j=1, \ldots, 2M
\]

where \( A_j \) consists of \( m_j \) of the \( \omega \) frequencies clustered around
\[\lambda_j = \omega/M, \ j = 1, \ldots, 2M.\] 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

\[
\hat{\delta}_{k+1} = \left[-H\right]^{-1} \left[ S_A' \left[ T_{j=0}^{t-1} \Psi_j X \right] S_A \hat{\delta}_k + S_A' \text{vec} \sum_{j=0}^{T} \Psi_j \Phi_j \Psi_j \right]. \tag{13}
\]

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

\[
- S_A' \left[ T_{j=0}^{t-1} \left( \Psi_j P' \right) \Psi_j \right] S_A \hat{\delta}_k + S_A' \text{vec} \sum_{j=0}^{T} \left( \hat{\psi}_j \right) \vec{Z} \Psi_j Y.
\]

Therefore (13) takes the form

\[
\hat{\delta}_{k+1} = \left[ S_A' \left[ T_{j=0}^{t-1} \left( \Psi_j P' \right) \right] S_A \right]^{-1} S_A' \left[ T_{j=0}^{t-1} \left( \hat{\psi}_j \right) \vec{Z} \right] Y, \tag{14}
\]

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

\[
\hat{\delta}_{k+1} = \left[ S_A' \left[ T_{j=0}^{t-1} \left( \hat{\psi}_j \right) \right] S_A \right]^{-1} S_A' \left[ T_{j=0}^{t-1} \left( \hat{\psi}_j \right) \vec{Z} \right] Y(\lambda_j). \tag{15}
\]

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 \( \psi \).
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), \( \hat{\psi}(\hat{\delta}) \) will be the spectral FIML estimate of \( \psi \).

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

4. Asymptotic Information and Covariance Matrices

Asymptotic Information Matrix

Let \( \theta_0 = (\delta_0, \psi_0) \) the true parameter vector. The asymptotic information matrix for \( \theta_0 \) is

\[
IA(\theta_0) = \lim_{T \to \infty} T^{-1} E \frac{\partial^2 Q}{\partial \theta \partial \theta'} \bigg|_{\theta_0} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix},
\]

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

\[
X = ZP_0 + UB_0^{-1}J,
\]

and

\[
X'\Psi_j X = P_0'Z'\Psi_j ZP_0 + P_0'Z'\Psi_j UB_0^{-1}J + J'B_0^{-1}U'\Psi_j ZP_0 + J'B_0^{-1}U'\Psi_j UB_0^{-1}J.
\]

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

\[
E X'\Psi_j X = E P_0'Z'\Psi_j ZP_0 + E J'B_0^{-1}U'\Psi_j UB_0^{-1}J.
\]

(1)
Also
\[ EA_0'X'\psi_j^1X = EU_0'\psi_j^1U\beta_0^{-1}J. \quad (2) \]

Now let
\[ \bar{S} = (I_p \otimes \beta_0^{-1}J)S_A. \quad (3) \]

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

\[ f_{11} = \bar{S}'K_p\bar{S} + S_A \frac{1}{2\pi} \int [F(\lambda)^{-1} \otimes P_0']F_\lambda(d\lambda)P_0]S_A + \bar{S}' \frac{1}{2\pi} \int [F(\lambda)^{-1} \otimes F(\lambda)]d\lambda \bar{S} \quad (4a) \]

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

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

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 \( \text{Avar}(\hat{\theta}) \) for the asymptotic covariance matrix of \( \hat{\theta} \). We remark that \( \text{Avar}(\hat{\theta}) \) relates to the distribution of \( \hat{\theta} \) and not \( T\hat{\theta} \). Thus,

\[ \text{Avar}(\hat{\theta}) = T^{-1}IA^{-1}(\theta_0). \]
In appendix 6.2 we show that the inverse of the asymptotic information matrix is given by

\[
IA^{-1}(\theta_0) = \begin{bmatrix}
    f_{11} & f_{12} \\
    f_{21} & f_{22}
\end{bmatrix},
\]

where

\[
f_{11} = \left[ \frac{1}{2\pi} S_A' \int (F(\lambda)^{-1} \otimes P' F_{2}(d\lambda) P) S_A \right]^{-1}
\]

(5a)

\[
f_{12} = -2 \left[ \frac{1}{2\pi} S_A' \int (F(\lambda)^{-1} \otimes P' F_{2}(d\lambda) P) S_A \right]^{-1} \left[ (I_p \otimes \Sigma'_\varepsilon) D^{+} : (I_p \otimes \Sigma'_\eta) D^{+} \right]
\]

(5b)

\[
f_{22} = 16\pi^3 \left[ \frac{D^+ \Delta^{-1}}{D^+ \Delta^{-1} c(\lambda)} \right] \left[ F(\lambda)^{-1} \otimes F(\lambda)^{-1} \right] \left[ \frac{D^+}{D^+ c(\lambda)} \right] d\lambda
\]

\[+ 4 \left[ \frac{D^+(I_p \otimes \Sigma'_\varepsilon)}{D^+(I_p \otimes \Sigma'_\eta)} \right] S f_{11} S' \left[ (I_p \otimes \Sigma'_\varepsilon) D^{+} : (I_p \otimes \Sigma'_\eta) D^{+} \right]
\]

(5c)

where

\[
\Delta = \int (F(\lambda)^{-1} F(\nu)^{-1} \otimes F(\lambda)^{-1} F(\nu)^{-1}) (c(\lambda)^2 - c(\lambda) c(\nu)) d\lambda d\nu
\]

(6)

and \( F(\lambda)=F(\lambda, \theta_0), \ \theta_0=(\delta_0, \psi_0), \ \Sigma_\varepsilon=\Sigma_\varepsilon(\psi_0), \ \Sigma_\eta=\Sigma_\eta(\psi_0), \ P=P(\delta_0) \).

5. Asymptotically Efficient 2-step Estimators

We shall now deal with the problem of constructing asymptotically efficient 2-step estimators for \( \delta \) and \( \psi \). We begin presenting \( \mathcal{F} \)-consistent estimators for \( \delta \) and \( \psi \), since it is well known that the first step in such estimators consists in finding \( \mathcal{F} \)-consistent
A $\gamma_T$-consistent Estimator for Coefficients

A $\gamma_T$-consistent estimator for $\delta$ can be constructed as follows. Let

$$Y = Z\Pi + V$$

be the reduced form of the model with associated covariances matrices $\Sigma_\varepsilon^* = B^{-1}\Sigma_\varepsilon B^{-1}$ and $\Sigma_\eta^* = B^{-1}\Sigma_\eta B^{-1}$. Let $\psi^*$ be the $p(p+1) \times 1$ 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 $\Pi$ and $\psi^*$. Let these estimators be $\hat{\Pi}$ and $\hat{\psi}$. 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

$$\tilde{\delta} = \left[ S_A \sum_{j=0}^{T-1} (F_v(j)^{-1} \tilde{P}'IZ(j)^{P}S_A)^{-1} S_A' \sum_{j=0}^{T-1} [F_v(j)^{-1} \tilde{P}'] \text{vec}IZ_Y(j) \right] (\text{1})$$

Proceeding in the same way as Hannan and Terrell (1973), we can show that $T^\dagger(\delta - \delta_0) = O_p(1)$. For details, see appendix 6.3.

A $\gamma_T$-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 $\delta$ given in
A natural estimator for \( \psi \), the structural form hyperparameters, is then given by

\[
\tilde{\psi} = \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \text{vec}\{\tilde{B}'\tilde{\Sigma}^x\tilde{B} : \tilde{B}'\tilde{\Sigma}^\eta \tilde{B}\},
\]

where \( \tilde{B} = \tilde{B}(\delta) \), \( \tilde{\Sigma}^x = \tilde{\Sigma}^x(\tilde{\psi}) \) and \( \tilde{\Sigma}^\eta = \tilde{\Sigma}^\eta(\tilde{\psi}) \). Later we shall discuss the asymptotic efficiency of the 2-step estimator for \( \psi \). Using the same approach it can be easily verified that \( \tilde{\psi} \) is \( \mathcal{A} \)-consistent.

The 2-step Estimator for Coefficients

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

\[
\hat{\delta} = \left[ S_A, \sum_{j=0}^{T-1} [F_j^{-1} \otimes P' \text{vec} Z, \chi(j)] S_A \right]^{-1},
\]

where \( \tilde{P} = P(\tilde{\delta}) \), \( \tilde{F}_j = F_j(\tilde{\psi}) \), \( \tilde{\delta} \) and \( \tilde{\psi} \) are \( \mathcal{A} \)-consistent estimators for \( \delta \) and \( \psi \) respectively.

We shall show that

\[
\begin{array}{c}
d \left( T^4(\delta, \delta_0) \right) \rightarrow N(0, f^{11}) \quad (4)
\end{array}
\]

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

Subtracting \( \delta_0 \) and multiplying by \( T^4 \) both sides of (3) yields

\[
T^4(\delta, \delta_0) = \{ -HT^{-1} \} T^4 \left[ S_A, (I, \otimes P') \sum_{j=0}^{T-1} [F_j^{-1} \otimes I_k] \text{vec} Z, \chi(j) + HT^{-1} \delta_0 \} \right]
\]

(5)
where
\[ \hat{H} = \left[ S_A \sum_{j=0}^{T-1} (F_j^{-1} \circ P' I_{Z'} X(j)) S_A \right]. \tag{6} \]

Now from (6) and (2.4) we have that
\[ \hat{H}^{-1}\delta_0 = S_A' \sum_{j=0}^{T-1} (F_j^{-1} \circ P' I_{Z'} X(j)) \text{vec} A_0 - S_A' \sum_{j=0}^{T-1} (F_j^{-1} \circ P') \text{vec} I_{Z'} Y(j), \tag{7} \]

hence (5) becomes
\[ T! (\hat{\delta} - \delta_0) = [-\hat{H}^{-1}]^{-1} S_A' (I_p \circ P') \left[ T^{-1} \sum_{j=0}^{T-1} (F_j^{-1} \circ I_k) \text{vec} I_{Z'} U(j) \right]. \tag{8} \]

Now, since \( \hat{\delta} \) and \( \hat{\psi} \) are \( \Delta \)-consistent, arguing as in (3.10) we have that
\[ \text{plim} \hat{H}^{-1} = - S_A' (I_p \circ P') \text{plim} \left[ T^{-1} \sum_{j=0}^{T-1} (F_j^{-1} \circ I_{Z'} X(j)) \right] (I_p \circ P) S_A \]
\[ = - S_A' (I_p \circ P') Q_u (I_p \circ P) S_A, \tag{9} \]

where the last equality follows from (3.4.8), \( Q_u \) being
\[ Q_u = \frac{1}{2\pi} \int F_u^{-1}(\lambda) \circ F_Z(\lambda) d\lambda. \tag{10} \]

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
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 \( \Pi \) which takes into account the restrictions on \( B \) and \( \Gamma \). 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

\[ T^t \text{vec}(\hat{\Pi} - \Pi_0) = O_p(1) \]

and therefore as discussed in section 3.4

\[ T^t(\hat{\psi}^* - \psi^*) \rightarrow N(0, IA^{-1}(\psi^*)), \]

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

\[ IA^{-1}(\psi^*) = 16\pi^3 \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \begin{bmatrix} c^2(F_v^{-1} @ F_v^{-1}) & c(F_v^{-1} @ F_v^{-1}) \\ c(F_v^{-1} @ F_v^{-1}) & (F_v^{-1} @ F_v^{-1}) \end{bmatrix}^{-1} \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \]

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

Now let

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

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

\[ \text{vec}[\hat{\Sigma}^*_\varepsilon : \hat{\Sigma}^*_\eta] = \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix} \hat{\psi}^* \].
We shall now show that
\[
\text{d} \quad T^4(\hat{\psi} - \psi_o) \to N(0, f^{22})
\]  
(15)

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

First we note that
\[
[(\hat{\Sigma}_{\hat{\epsilon}} : \hat{\Sigma}_\eta) - (\Sigma_{\epsilon}: \Sigma_\eta)_o] = [\hat{B}'(\hat{\Sigma}_{\hat{\epsilon}}^* - \Sigma_{\epsilon o}^*) \hat{B}] - [B_o' \Sigma_{\epsilon o} B_o : B_o' \Sigma_{\eta o} B_o] \\
- [\hat{B}'(\hat{\Sigma}_\eta^* - \Sigma_{\eta o}^*) \hat{B}] + [(\hat{B} - B_o)' \Sigma_{\epsilon o} B_o : (\hat{B} - B_o)' \Sigma_{\eta o} B_o] \\
+ [(\hat{B}' \Sigma_{\epsilon o}(\hat{B} - B_o) : \hat{B}' \Sigma_{\eta o}(\hat{B} - B_o)].
\]

Hence,
\[
\text{vec}[(\hat{\Sigma}_{\hat{\epsilon}} : \hat{\Sigma}_\eta) - (\Sigma_{\epsilon}: \Sigma_\eta)_o] = \\
- [\text{vec} B'(\hat{\Sigma}_{\hat{\epsilon}}^* - \Sigma_{\epsilon o}^*) B] + [\text{vec} B'(\Sigma_{\eta}^* - \Sigma_{\eta o}^*) B] + [\text{vec}(B - B_o)' \Sigma_{\epsilon o} B_o] \\
+ [\text{vec}(B - B_o)' \Sigma_{\eta o} B_o] + [(B - B_o)'(\hat{\Sigma}_{\eta o} B) : B'(\hat{\Sigma}_{\eta o} B)].
\]

or
\[
- [\hat{B}' \hat{\Sigma}_{\epsilon o} B_o] \quad \text{vec} [(\hat{\Sigma}_{\hat{\epsilon}}^* - \Sigma_{\epsilon o}^*) : (\hat{\Sigma}_\eta^* - \Sigma_{\eta o}^*)] \\
+ [\text{vec}(\hat{B} - B_o) + [B_o' \Sigma_{\epsilon o} \Theta_1 \text{vec}(B - B_o) + [B_o' \Sigma_{\eta o} \Theta_2 \text{vec}(B - B_o)]
\]

and therefore
\[
T^4(\hat{\psi} - \psi_o) = [D^+ \ 0] \quad [\hat{B}' \hat{\Sigma}_{\epsilon o} B_o \ 0 \ \hat{B}' \hat{\Sigma}_\eta B_o] \quad [D^+ \ 0] \quad T^4(\hat{\psi}^* - \psi_o^*) \\
+ [D^+ \ 0] \quad [\text{vec}(\hat{B} - B_o)] \quad T^4 \text{vec}(\hat{B} - B_o) \\
+ [D^+ \ 0] \quad [\text{vec}(\hat{B} - B_o)] \quad T^4 \text{vec}(\hat{B} - B_o).
\]  
(16)
Since \( \text{plim} \mathbb{B} = \mathbb{B}_0 \), 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 \( B_0 \) as \( B_0 = J A_0 \), \( J = [I_p: 0_{p \times k}] \), from (2.4) and (4.3) it follows that

\[
\text{vec}(\mathbb{B} - \mathbb{B}_0) = \text{vec}J(\hat{A} - A_0) \\
\quad = -(I_p \otimes J) S_{\hat{A}}(\hat{\delta} - \delta_0) \\
\quad = -(I_p \otimes \mathbb{B}_0) S(\hat{\delta} - \delta_0). \tag{17}
\]

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 \( \hat{\Pi} \) and \( \hat{\psi}^* \) are asymptotically independent. Now, because \( \hat{\Pi} = \hat{\Pi}(\hat{\delta}) \) takes into account the restrictions on \( B \) and \( \Gamma \)

\[
\text{Avar}(\hat{\Pi}, \hat{\psi}^*) - \text{Avar}(\hat{\Pi}, \hat{\psi}^*) > 0. \tag{18}
\]

However, because \( \hat{\psi}^* \) 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

\[ T^{-1} \sum_{j=0}^{T-1} I_j \hat{F}_j^{-1} = I_P, \]  

(1)

where \( I_j \) is the real part of the crossperiodogram matrix of \( U' \) \((p \times T)\) and \( \hat{F}_j \) is the estimated spectral matrix. Thus,

\[ \hat{F}_j = (2\pi)^{-1} \left( c_j \hat{\Sigma}_\epsilon + \hat{\Sigma}_\eta \right), \]  

(2)

\[ \hat{\Sigma}_\epsilon = \Sigma_\epsilon(\psi), \hat{\Sigma}_\eta = \Sigma_\eta(\psi), \]  

and \( \psi \) is the solution of

\[ (2\pi)^{-1} \sum_{j=0}^{T-1} S_j \text{vec}[(F_j^{-1} - F_j^{-1}I_jF_j^{-1})] = 0 \]  

(3)

where \( S_j \) is given in (2.7).

Replacing \( S_j \), (3) becomes

\[ (2\pi)^{-1} \sum_{j=0}^{T-1} \left[ D'c_j \right] \text{vec}[(F_j^{-1} - F_j^{-1}I_jF_j^{-1})] = 0. \]  

(4)

Now since \( D \) has full column rank, (4) implies that

\[ \sum_j \left[ c_jF_j^{-1} - c_jF_j^{-1}I_jF_j^{-1} \right] = 0 \]  

(5a)

and

\[ \sum_j \left[ F_j^{-1} - F_j^{-1}I_jF_j^{-1} \right] = 0 \]  

(5b)
Solving (5) for \( \psi \), it does not matter if it is numerically, we have

\[
\sum_j (c_j \hat{F}_j^{-1} - c_j \hat{F}_j^{-1} I_j \hat{F}_j^{-1}) = 0
\]

(6a)

and

\[
\sum_j (\hat{F}_j^{-1} - \hat{F}_j^{-1} I_j \hat{F}_j^{-1}) = 0
\]

(6b)

Now premultiplying (2) by \( \hat{\Sigma}_{\varepsilon}^{-1} \) yields

\[
\hat{\Sigma}_{\varepsilon}^{-1} \hat{F}_j = (2\pi)^{-1} (c_j I_p + Q)
\]

where

\[
\hat{Q} = \hat{\Sigma}_{\varepsilon}^{-1} \hat{\Sigma}_\eta.
\]

Thus,

\[
(2\pi) \hat{\Sigma}_{\varepsilon}^{-1} = (c_j I_p + \hat{Q}) \hat{F}_j^{-1}.
\]

(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 (6a) we have

\[
\sum_j (c_j I_p + Q) \hat{F}_j^{-1} - \hat{Q} \sum_j \hat{F}_j^{-1} = \sum_j (c_j I_p + Q) \hat{F}_j^{-1} I_j \hat{F}_j^{-1} - \hat{Q} \sum_j \hat{F}_j^{-1} I_j \hat{F}_j^{-1}.
\]

Using (7) we have

\[
2\pi \hat{\Sigma}_{\varepsilon}^{-1} - \hat{Q} \sum_j \hat{F}_j^{-1} = 2\pi \hat{\Sigma}_{\varepsilon}^{-1} \sum_j I_j \hat{F}_j^{-1} - \hat{Q} \sum_j \hat{F}_j^{-1} I_j \hat{F}_j^{-1}
\]

and from (6.b) we have

\[
2\pi \hat{\Sigma}_{\varepsilon}^{-1} - 2\pi \hat{\Sigma}_{\varepsilon}^{-1} \sum_j I_j \hat{F}_j^{-1} = 0,
\]

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

\[
f_{12} = \frac{1}{(2\pi)^{\frac{3}{2}}} \left[ \int c(F^{-1} @ I_p) : \int (F^{-1} @ I_p) \right] \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix}
\]

and as

\[
f_{22} = \frac{1}{16\pi^3} \begin{bmatrix} D' & 0 \\ 0 & D' \end{bmatrix} \left[ \int c^2(F^{-1} @ F^{-1}) & \int c(F^{-1} @ F^{-1}) \\ \int c(F^{-1} @ F^{-1}) & \int (F^{-1} @ F^{-1}) \right] \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix}
\]

where in (1) and (2) \(c=c(\lambda)\) and \(F=F(\lambda)\). We shall often omit the argument variable \(\lambda\) in situations where there will be no risk of confusion.

Now using the fact that the square matrices \(f c^2(F^{-1} @ F^{-1})\), \(f (F^{-1} @ F^{-1})\) and \(f(F^{-1} @ F^{-1})\) commute with each other we have

\[
f_{22}^{-1} = 16\pi^3 \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix} \left[ \int (F^{-1} @ F^{-1}) & -\int c(F^{-1} @ F^{-1}) \\ -\int c(F^{-1} @ F^{-1}) & \int c^2(F^{-1} @ F^{-1}) \right] \begin{bmatrix} D^+ & 0 \\ 0 & D^+ \end{bmatrix}
\]
where \( D^+ \) is the Moore-Penrose inverse of \( D \) and \( \Delta \) is given by

\[
\Delta = \int c(\lambda)^2 (F(\lambda)^{-1}@F(\lambda)^{-1}) d\lambda \int (F(\lambda)^{-1}@F(\lambda)^{-1}) d\lambda
- \int c(\lambda)(F(\lambda)^{-1}@F(\lambda)^{-1}) d\lambda \int c(\lambda)(F(\lambda)^{-1}@F(\lambda)^{-1}) d\lambda
- \int (F(\lambda)^{-1}F(\sigma)^{-1}@F(\lambda)^{-1}F(\sigma)^{-1}) (c(\lambda)^2-c(\lambda)c(\sigma)) d\lambda d\sigma . \tag{4}
\]

Let \( f_{12}' \) be partitioned as

\[
f_{12}' = [B_1:B_2] .
\]

From (1) and (3) and from the properties of \( D \) we obtain

\[
B_1 = 2\pi \bar{S}' \int c(F^{-1}@I_p)(I+K_p)\Delta^{-1} \int (F^{-1}@F^{-1})
- \int (F^{-1}@I_p)(I+K_p)\Delta^{-1} \int c(F^{-1}@F^{-1}) \]d\'+

which can be rewritten as

\[
- 4\pi \bar{S}' \Delta^{-1} \int c(F^{-1}@I_p) \int (F^{-1}@F^{-1}) - \int (F^{-1}@I_p) \int c(F^{-1}@F^{-1}) \]d\'+

\[
-4\pi \bar{S}' \Delta^{-1} \int \int (F(\lambda)^{-1}F(\sigma)^{-1}@F(\lambda)^{-1}F(\sigma)^{-1})(I_p@F(\lambda))(c(\lambda)-c(\sigma)) d\lambda d\sigma \]d\'+

by noting that \( \Delta^{-1} \) commutes with \( K_p \) and with \( f(F^{-1}@I_p) \) and \( K_pD' = D' \).

Now because

\[
(I_p@F(\lambda))(c(\lambda)-c(\sigma)) = (2\pi)^{-1} \int (I_p@c(\lambda)(\Sigma_\epsilon + \Sigma_\eta))(c(\lambda) - c(\sigma))
\]
\[ -(2\pi)^{-1}\left[ (I_p \otimes \epsilon) (c(\lambda)^2 - c(\lambda)c(\nu)) + (I_p \otimes \eta)(c(\lambda) - c(\nu)) \right] \]

we have
\[ B_1 = 2\bar{S}'(I_p \otimes \epsilon)D^+. \]

Similary we can show
\[ B_2 = 2\bar{S}'(I_p \otimes \eta)D^+. \]

and therefore
\[ f_{12f_{22}^{-1}} = 2\bar{S}' \left[ (I_p \otimes \epsilon) : (I_p \otimes \eta) \right] \left[ D^+ 0 \right] \left[ 0 D^+ \right] . \]

Now
\[ f_{12f_{22}^{-1}f_{21}} = \frac{2}{4\pi^2} \bar{S}'(I_p \otimes \epsilon : I_p \otimes \eta) \left[ D^+ 0 \right] \left[ 0 D^+ \right] \left[ \int c(F^{-1} \otimes I_p) \right] \bar{S} \]

\[ - \frac{1}{4\pi^2} \bar{S}'(I_p \otimes \epsilon : I_p \otimes \eta) \left[ I+K_p 0 \right] \left[ 0 I+K_p \right] \left[ \int (F^{-1} \otimes I_p) \right] \bar{S} \]

\[ - \frac{1}{4\pi^2} \bar{S}' \left[ (I_p \otimes \epsilon) \left\{ (I_p \otimes \epsilon) + (I_p \otimes \eta) \right\} \right] \bar{S} \]
\[ + \frac{1}{4\pi^2} \bar{S}' \left[ (I_p \otimes \epsilon)K_p \left\{ (I_p \otimes \epsilon) + (I_p \otimes \eta) \right\} K_p \right] \bar{S} \]

\[ - \frac{1}{4\pi^2} \bar{S}' \left[ \right] \bar{S} + \frac{1}{4\pi^2} \bar{S}' \left[ K_p \left\{ (c(\epsilon) + \epsilon(\eta) \right\} F^{-1} \otimes I_p \right] \bar{S} \]

\[ - \frac{1}{2\pi} \bar{S}' \left[ \right] d\lambda \bar{S} + \bar{S}'K_p \bar{S} . \]

Hence
\[ W = f_{11} - f_{12} f_{22}^{-1} f_{21} = \frac{1}{2\pi} S_A \int (F(\lambda)^{-1} \otimes F_2(d\lambda) P) S_A \]

and

\[ f_{11} = \left[ \frac{1}{2\pi} S_A \int (F(\lambda)^{-1} \otimes F_2(d\lambda) P) S_A \right]^{-1} \]

(8)

Now using (8) and (7) we have

\[ f_{12} = -2 \left[ \frac{1}{2\pi} S_A \int (F(\lambda)^{-1} \otimes F_2(d\lambda) P) S_A \right]^{-1} \frac{1}{S} \left[ (I_p \otimes \Sigma) D^+ : (I_p \otimes \Sigma) D^+ \right] \]

(9)

and using (9) and (7) and (3) we have

\[ f_{22} = 16\pi^3 \left[ \begin{array}{ccc} D^+ \Delta^{-1} & 0 & \int (F^{-1} \otimes F^{-1}) \ -c(F^{-1} \otimes F^{-1}) \ 
 \frac{c}{c} \int c(F^{-1} \otimes F^{-1}) \ 
 \frac{c}{c} \int c^2(F^{-1} \otimes F^{-1}) \end{array} \right] \left[ \begin{array}{c} D^+ \ 0 \\
 0 \\
 0 \\
 0 \end{array} \right] + 4 \left[ \begin{array}{c} D^+ \ 0 \\
 0 \\
 0 \\
 0 \end{array} \right] \left[ \begin{array}{c} I_p \otimes \Sigma \ \\
 I_p \otimes \Sigma \ \\
 W^{-1}S \cdot (I_p \otimes \Sigma) : (I_p \otimes \Sigma) D^+ : (I_p \otimes \Sigma) D^+ \ \\
 0 \end{array} \right] . \]

or, after some algebra,

\[ f_{22} = 16\pi^3 \left[ \begin{array}{c} D^+ \Delta^{-1} \\
 D^+ \Delta^{-1} c(\lambda) \\
 \end{array} \right] \left[ \begin{array}{c} F(\lambda)^{-1} \otimes F(\lambda)^{-1} \\
 D^+ c(\lambda) \end{array} \right] d\lambda \]

\[ + 4 \left[ \begin{array}{c} D^+ (I_p \otimes \Sigma) \ \\
 D^+ (I_p \otimes \Sigma) \ \\
 \end{array} \right] \left[ \int W^{-1}S \cdot (I_p \otimes \Sigma) D^+ : (I_p \otimes \Sigma) D^+ \right] . \]

(10)
Appendix 6.3

/J T/-consistency of Coefficient Estimators

We shall show that $\tilde{\delta}$ given in (5.1) is a $/J T/$-consistent estimator for $\delta$, that is

$$T \delta (\tilde{\delta} - \delta_0) = O_p(1)$$  \hspace{1cm} (1)

Subtracting $\delta_0$ from both sides of (5.1) we have

$$(\tilde{\delta} - \delta_0) = H_V^{-1} S_A' \left[ \sum_{j=0}^{T-1} [F_v^{-1}(j) \otimes \tilde{P}']\text{vec} I_Z'\tilde{U}(j) \right]$$  \hspace{1cm} (2)

where

$$H_V = S_A' \sum_{j=0}^{T-1} (F_v^{-1}(j) \otimes \tilde{P}') I_Z'Z(j)\tilde{P}S_A.$$  \hspace{1cm} (3)

and

$$\text{vec} I_Z'\tilde{U}(j) = \text{vec} I_Z'Y(j) - (I_p \otimes I_Z'Z(j)\tilde{P}) S_A \delta_0.$$  \hspace{1cm} (4)

Now from (2.4), after some algebra, (4) becomes

$$\text{vec} I_Z'\tilde{U}(j) = \text{vec} I_Z'Y(j) - [(I_p - B_0)' \otimes I_Z'Z(j)] \text{vec}(\tilde{\Pi} - \Pi_0).$$  \hspace{1cm} (5)

Now from (3.3.10) we have that

$$\text{vec}(\tilde{\Pi} - \Pi_0) = A_V^{-1} \sum_{j=0}^{T-1} [F_v^{-1}(j)' \otimes I_k] \text{vec} I_Z'Y(j),$$  \hspace{1cm} (6)

where

$$A_V = \sum_{j=0}^{T-1} [F_v^{-1}(j) \otimes I_Z'Z(j)].$$

Thus (5) becomes
\begin{align*}
\text{vec } I_Z \tilde{y}(j) &= \text{vec } I_Z y(j) - [I_p \otimes I_Z(y(j))] A_v^{-1} \sum_{j=0}^{T-1} [F_v^{-1}(j) @ I_k] \text{vec } I_Z y(j) \\
&+ [B_o \otimes I_Z(y(j))] A_v^{-1} \sum_{j=0}^{T-1} [F_v^{-1}(j) @ I_k] \text{vec } I_Z y(j). \quad (7)
\end{align*}

Hence, replacing (7) in (2) we have after some algebra

\begin{align*}
\tilde{\delta} - \delta_0 &= H_v^{-1} S_A'(I_p \otimes \hat{P}') A_v (B_o \otimes I_k) A_v^{-1} \sum_{j=0}^{T-1} [F_v^{-1}(j) @ I_k] \text{vec } I_Z y(j). \quad (7)
\end{align*}

Now from (3.4.8) and (3.4.9) we have

\[ \text{plim } T^{-1} A_v = Q_v \]

and

\[ T^{-1} \sum_{j=0}^{T-1} (F_v^{-1}(j) @ I_k) \text{vec } I_Z y(j) \overset{d}{\rightarrow} N(0, Q_v) \]

where

\[ Q_v = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_v^{-1}(\lambda) @ F_z(d\lambda). \]

Also because \( \text{plim } \hat{P} = P \),

\[ \text{plim } T^{-1} H_v = \text{plim } T^{-1} S_A'(I_p \otimes \hat{P}') A_v (I_p \otimes \hat{P}') S_A \]

\[ = S_A'(I_p \otimes \hat{P}') Q_v (I_p \otimes P) S_A. \]

Hence making use of all of this, (1) follows from Slutsky's theorem.
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,

\[ y_t'B + z_t'\Gamma = w_t', \]  

(1)

where \( B \) is \( pxp \), \( \Gamma \) is \( kxp \) and \( w_t \) follows a multivariate random walk plus noise model with associated covariance matrices \( \Sigma_\varepsilon \) and \( \Sigma_\eta \). 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 \( 1x(p-1-p_1) \), \( Z_{1t}' \) is \( 1xk_1 \), \( Z_{2t}' \) is \( 1xk_2 \) and

\[ B = \begin{bmatrix} 1 & B_{12} & B_{13} \\ \beta & B_{22} & B_{23} \\ 0 & B_{32} & B_{33} \end{bmatrix} \]
\[ \Gamma = \begin{bmatrix} \gamma & \Gamma_{12} & \Gamma_{13} \\ 0 & \Gamma_{22} & \Gamma_{23} \end{bmatrix} \]

Within this notation, the first equation, the equation of interest, may be then written as

\[ y_{1t} = y_{1t}'\beta - Z_{1t}'\gamma + \mu_{1t} + \epsilon_{1t}, \]  

(2a)
\[ \mu_{1t} = \mu_{1,t-1} + \eta_{1t}. \]  

(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

\[
[y_{1t} y_{1t}' y_{2t}'] = [Z_{1t}' Z_{2t}'] \Pi + [v_{1t} v_{1t}' v_{2t}']
\]

(3)

where

\[
\Pi = \Gamma B^{-1} = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} \\ \Pi_{21} & \Pi_{22} & \Pi_{23} \end{bmatrix}
\]

and

\[
[v_{1t} v_{1t}' v_{2t}'] = [w_{1t} w_{1t}' w_{2t}'] \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix}
\]

(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

\[
[y_{1t} y_{1t}'] = [Z_{1t}' Z_{2t}'] \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} + [v_{1t} v_{1t}']
\]

(5)

where

\[
[v_{1t} v_{1t}'] = [w_{1t} w_{1t}' w_{2t}'] \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{bmatrix}
\]

(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

\[
\Pi_{21} + \Pi_{22} \beta = 0
\]

(7a)
and

\[ \Pi_{11} + \Pi_{12} \beta = -\gamma \]  \hspace{1cm} (7b)

As pointed out in Hall and Pagan (1981), (7b) is not a restriction as such, but it will hold, and we can deduce \( \gamma \) 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 than the reduced form subject to restrictions.

The new system may be written as

\[
\begin{bmatrix}
y_{1t} & y_{1t}'
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
\beta I_{p1} & \end{bmatrix}
+ \begin{bmatrix}
z_{1t}'z_{2t}'
\end{bmatrix}
\begin{bmatrix}
\gamma & -\Pi_{12} \\
0 & -\Pi_{22}
\end{bmatrix}
- \begin{bmatrix}
w_{1t} & v_{1t}'
\end{bmatrix}
\]  \hspace{1cm} (1)

where

\[
\begin{bmatrix}
w_{1t} & v_{1t}'
\end{bmatrix}
- \begin{bmatrix}
w_{1t} & w_{1t}'w_{2t}'
\end{bmatrix}
\begin{bmatrix}
1 & B_{12} \\
0 & E_{22}
\end{bmatrix}
\begin{bmatrix}
0 & B_{32}
\end{bmatrix}
\]  \hspace{1cm} (2)
The reduced form of (1) is given by

\[ [y_{1t} \ y_{1t}] = [Z_{1t}Z_{2t}] \Pi^\dagger + [v_{1t}^* \ v_{1t}^*] \]  \quad (3)

where

\[ \Pi^\dagger = \begin{bmatrix} \gamma & -\Pi_{12} \\ 0 & -\Pi_{22} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \beta & I_{p1} \end{bmatrix}^{-1} \quad \begin{bmatrix} \Pi_{11}^\dagger & \Pi_{12} \\ \Pi_{21}^\dagger & \Pi_{22} \end{bmatrix} \]  \quad (4)

and

\[ [v_{1t}^* \ v_{1t}^*] = [v_{1t} \ v_{1t}] \begin{bmatrix} 1 & 0 \\ \beta & I_{p1} \end{bmatrix}^{-1}. \]  \quad (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 \( \Sigma_e \) and \( \Sigma_n \), 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.
\[ \Sigma_{\varepsilon}^\dagger = \begin{bmatrix} 1 & 0 & 0 \\ B^{12} & B^{22} & B^{32} \end{bmatrix}, \quad \Sigma_{\varepsilon} = \begin{bmatrix} 1 & B^{12} \\ 0 & B^{22} \\ 0 & B^{32} \end{bmatrix} \quad (6a) \]

\[ \Sigma_{\eta}^\dagger = \begin{bmatrix} 1 & 0 & 0 \\ B^{12} & B^{22} & B^{32} \end{bmatrix}, \quad \Sigma_{\eta} = \begin{bmatrix} 1 & B^{12} \\ 0 & B^{22} \\ 0 & B^{32} \end{bmatrix} \quad (6b) \]

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 \( \Theta \) 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 \( \epsilon_{1t} \) and \( \eta_{1t} \) 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

\[
\begin{bmatrix}
    y_{1t} \\
    y_{1t}'
\end{bmatrix}
B^\dagger +
\begin{bmatrix}
    z_{1t}' \\
    z_{2t}'
\end{bmatrix}
\Gamma^\dagger = w_t^\dagger ,
\]

(1)

where the \( \hat{p} \times \hat{p} \) matrix \( B^\dagger \) and the \( K \times \hat{p} \) matrix \( \Gamma^\dagger \), where \( \hat{p} = p_1 + 1, K = k_1 + k_2 \), are

\[
B^\dagger = \begin{bmatrix}
    1 \\
    \beta I_{p_1}
\end{bmatrix}
\quad \text{and} \quad
\Gamma^\dagger = \begin{bmatrix}
    \gamma \\
    0
\end{bmatrix}
\begin{bmatrix}
    -\Pi_{12} \\
    -\Pi_{22}
\end{bmatrix} .
\]

As we pointed out before the \( \hat{p} \times 1 \) vector \( w_t^\dagger = [w_{1t} \; y_{1t}']' \) follows a \( \hat{p} \)-variate random walk plus noise model with associated covariance matrices \( \Sigma_{\epsilon^\dagger} \) and \( \Sigma_{\eta^\dagger} \) given in (3.6a) and (3.6b).

Let \( \alpha \) denote the unrestricted elements of \( \text{-vec } [B^\dagger \; \Gamma^\dagger]' \). Thus
Conditional on the first observations the loglikelihood function of (1) is given in (5.4.4), that is,

\[
\ell(\alpha, \psi^\dagger) = (T-1) \log |B^\dagger| - \frac{1}{2} \sum_{t=2}^{T} \log |F_t| - \frac{1}{2} \sum_{t=2}^{T} r_t' F_t^{-1} r_t ,
\]

(2)

where \( \psi^\dagger \) is the \( p(p+1) \times 1 \) vector obtained from \( \text{vec}(\Sigma_e^\dagger : \Sigma_\eta^\dagger) \) by eliminating the supradiagonal elements of \( \Sigma_e^\dagger \) 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

\[
r_t = \tilde{y}_t - \tilde{X}_t' \alpha
\]

(3)

where \( \tilde{y}_t \) is obtained by applying the multivariate Kalman filter to the \( \tilde{p} \)-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 \( \tilde{X}_t' \), where

\[
\tilde{X}_t' = \begin{bmatrix} y_{1t}' & Z_{1t}' & 0 \\ 0 & (I_p \otimes z_t') \end{bmatrix}
\]

(4)

Replacing (3) in (2) yields

\[
\ell(\alpha, \psi^\dagger) = - \frac{1}{2} \sum_{t=2}^{T} \log |F_t| - \frac{1}{2} \sum_{t=2}^{T} (\tilde{y}_t - \tilde{X}_t' \alpha)' F_t^{-1} (\tilde{y}_t - \tilde{X}_t' \alpha)
\]

(5)
which has to be maximised with respect to \( \alpha \) and \( \psi^\dagger \).

As in the univariate case the maximisation of (5) can be carried out in two different ways. Firstly \( \alpha \) can be concentrated out of the loglikelihood function. Thus, solving the likelihood equations for \( \alpha \) we have

\[
\hat{\alpha} = \hat{\alpha}(\psi^\dagger) = \left[ \sum_{t=2}^{T} \tilde{X}_t F_{t-1} \tilde{X}_t' \right]^{-1} \sum_{t=2}^{T} \tilde{X}_t F_{t-1} \tilde{y}_t.
\] (6)

Replacing (6) in (5) yields the concentrated loglikelihood function

\[
\mathcal{L}_c(\psi^\dagger) = -\frac{1}{2} \sum_{t=2}^{T} \log |F_{t-1}| - \frac{1}{2} \sum_{t=2}^{T} (\tilde{y}_t - \tilde{X}_t' \hat{\alpha})' F_{t-1}^{-1} (\tilde{y}_t - \tilde{X}_t' \hat{\alpha})
\] (7)

and the ML estimate of \( \psi^\dagger \) is obtained by maximising (7) nonlinearly with respect to \( \psi^\dagger \). Once we have found \( \hat{\psi}^\dagger \), \( \hat{\alpha} = \hat{\alpha}(\hat{\psi}) \) is obtained from (6).

Alternatively we could maximise (5) in a stepwise fashion. A consistent estimator of \( \alpha \) 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, \( \hat{w}_t^\dagger = y_t - \tilde{X}_t' \hat{\alpha} \), are computed and the matrices \( \hat{\Sigma}_\epsilon^\dagger \) and \( \hat{\Sigma}_\eta^\dagger \) are estimated maximising the loglikelihood of \( \hat{w}_t^\dagger \). \( \alpha \) is reestimated using (6). As noted in Lahiri and Schmidt (1978) the two-step estimator of \( \alpha \) 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 asymptotic 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 $T^\dagger(\hat{\alpha} - \alpha_0)$ 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

$$A\text{var}(\hat{\alpha}) = \frac{1}{T} S_A^\prime (I_p \otimes P^\dagger) \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} [F(\lambda)^{-1} \otimes F_x(\lambda)] d\lambda (I_p \otimes P^\dagger) S_A \right]^{-1}$$

(1)

where

$$F(\lambda) = F(\lambda, \psi^\dagger) = (2\pi)^{-1}[2(1-\cos \lambda)\Sigma_\epsilon^\dagger + \Sigma_\eta^\dagger],$$

with $\Sigma_\epsilon^\dagger$ and $\Sigma_\eta^\dagger$ as given in (3.6), $\Sigma_\eta^\dagger$ positive definite; $F_x(\lambda)$ is the spectral matrix of the process generating the (differenced) exogenous variables;

$$P^\dagger = [P^\dagger I_k], \quad P^\dagger = \begin{bmatrix} \Pi_11^\dagger & \Pi_12^\dagger \\ \Pi_21^\dagger & \Pi_22^\dagger \end{bmatrix}$$

as given in (3.4).
where \( \Pi_{11}^{\dagger} \) is \( k_1 \times 1 \), \( \Pi_{12} \) is \( k_1 \times p_1 \), \( \Pi_{21}^{\dagger} \) is \( k_2 \times 1 \) and \( \Pi_{22} \) is \( k_2 \times p_1 \); \( S_A \) is the selection matrix as given in (5.2.7). For LIML \( S_A \) becomes

\[
S_A = \begin{bmatrix}
0 & 1 x(p_1+k_1) & \cdots \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0 \\
0 & k_2 x(p_1+k_1) & \cdots \\
0 & \cdots & 0
\end{bmatrix}
\]

and so

\[
(I_{p} \otimes \Pi^{\dagger}) S_A = \begin{bmatrix}
R & 0 \\
0 & I_{p_1} \otimes I_k
\end{bmatrix}
\]

where

\[
R = \begin{bmatrix}
\Pi_{12} \\
\Pi_{22}
\end{bmatrix}
\]

Therefore (1) becomes

\[
\text{Avar} \hat{\alpha} = \frac{2\pi}{T} \left[ \begin{array}{cccc}
R & 0 \\
0 & (I_{p_1} \otimes I_k)
\end{array} \right] \int_{-\pi}^{\pi} (F(\lambda)^{-1} \otimes F_{\pi}(\lambda)) \, d\lambda \left[ \begin{array}{cccc}
R & 0 \\
0 & (I_{p_1} \otimes I_k)
\end{array} \right]^{-1}
\]

Because the parameters in (5) are unknown, the asymptotic covariance matrix of \( \hat{\alpha} \) can be consistently estimated by

\[
\text{Avar} \hat{\alpha} = \left[ \begin{array}{cccc}
\hat{R} & 0 \\
0 & (I_{p_1} \otimes I_k)
\end{array} \right] \int_{\mathcal{J}=0}^{\mathcal{J}} (\hat{F}_j^{-1} \otimes I_2(\hat{\theta})) \left[ \begin{array}{cccc}
\hat{R} & 0 \\
0 & (I_{p_1} \otimes I_k)
\end{array} \right]^{-1}
\]

where \( \hat{R} \) is (4) with \( \Pi_{12} \) and \( \Pi_{22} \) replaced by their ML estimators; \( \hat{F}_j = F(\lambda_j, \hat{\theta}^{\dagger}) \), where \( \hat{\theta}^{\dagger} \) is the \( p(p+1) \times 1 \) vector containing the ML
estimators of the unrestricted elements in \( \text{vec}(\Sigma_\epsilon^\dagger \Sigma_\eta^\dagger) \); \( I_Z(j) \) is the periodogram matrix of the differenced exogenous variables.

Our main interest is centered on the asymptotic covariance matrix of \( \hat{\theta} \), where \( \hat{\theta}' = (\hat{\beta}', \hat{\gamma}') \) 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

\[
A_{\text{var}} \hat{\theta} = \frac{2\pi}{T} \left[ \begin{array}{c}
R' \int F(\lambda)^{11} @ F_Z(\lambda) d\lambda \quad R' \int F(\lambda)^{12} @ F_Z(\lambda) d\lambda \\
\int F(\lambda)^{21} @ F_Z(\lambda) d\lambda \quad \int F(\lambda)^{22} @ F_Z(\lambda) d\lambda
\end{array} \right]^{-1}
\]

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 \( \hat{\theta} \) is given by

\[
A_{\text{var}} \hat{\theta} = \frac{2\pi}{T} \left[ \begin{array}{c}
R' \left[ \int F(\lambda)^{11} @ F_Z(\lambda) d\lambda - \left( \int F(\lambda)^{12} @ F_Z(\lambda) d\lambda \right) \left( \int F(\lambda)^{22} @ F_Z(\lambda) d\lambda \right)^{-1} \left( \int F(\lambda)^{21} @ F_Z(\lambda) d\lambda \right) \right] R' \right]^{-1}
\]

which can be written as

\[
A_{\text{var}} \hat{\theta} = \frac{2\pi}{T} \left[ \left( \int F(\lambda)^{-1} @ F_Z(\lambda) d\lambda \right)^{11} \right]^{-1} R' \left( \int F(\lambda)^{-1} @ F_Z(\lambda) d\lambda \right)^{-1} \left( \int F(\lambda)^{11} @ F_Z(\lambda) d\lambda \right) R' \left( \int F(\lambda)^{-1} @ F_Z(\lambda) d\lambda \right)^{-1}
\]

where here \( [\cdot]^{11} \) denotes the kxk top left submatrix of the inverse of the \( p_k \times p_k \) matrix in squared brackets. As in (6), (9) can be consistently estimated as
Explanatory Variables Following a Multivariate Random Walk

If we assume that the explanatory variables follow a multivariate random walk with disturbance covariance matrix $\Sigma_z$, then the differenced variables will have constant spectrum, that is,

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

and expression (9) becomes

$$\begin{align*}
\operatorname{Avar} \hat{\delta} &= \frac{2\pi}{T} \left[ R \left[ \int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \otimes (2\pi)^{-1} \Sigma_z \right]_{\Sigma_z}^{-1} \right]^{-1}. \tag{12}
\end{align*}$$

Now

$$\left[ \int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \otimes (2\pi)^{-1} \Sigma_z \right]_{\Sigma_z}^{-1} = 2\pi \left[ \int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \right]_{\Sigma_z}^{-1}$$

hence

$$\left[ \int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \otimes (2\pi)^{-1} \Sigma_z \right]_{\Sigma_z}^{-1} = 2\pi \left[ \int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \right]_{\Sigma_z}^{11} \Sigma_z^{-1}. \tag{13}$$

Making use of (13), (12) becomes after some algebra

$$\begin{align*}
\operatorname{Avar} \hat{\delta} &= \frac{4\pi^2}{T} \left[ \int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \right]_{\Sigma_z}^{11} \left[ R \Sigma_z R \right]_{\Sigma_z}^{-1} \tag{14}
\end{align*}$$

where $F(\lambda) = F(\lambda, \psi^\top)$.

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

$$F(\lambda, \psi^+) = (2\pi)^{-\frac{1}{2}} \left[ c(\lambda) \Sigma_\epsilon^+ + \Sigma_\eta^+ \right] .$$  \hspace{1cm} (15a)

where

$$c(\lambda) = 2(1 - \cos \lambda).$$  \hspace{1cm} (15b)

After some algebra the inverse of (15a) becomes

$$[F(\lambda, \psi^+)]^{-1} = 2\pi \left[ I_p - B \cos \lambda \right]^{-1} \left[ 2\Sigma_\epsilon^+ + \Sigma_\eta^+ \right]^{-1}$$

where

$$B = \left[ 2\Sigma_\epsilon^+ + \Sigma_\eta^+ \right]^{-1} 2\Sigma_\epsilon^+. $$

We note that if \( \Sigma_\eta^+ \) 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

$$ (I_p - B \cos \lambda)^{-1} = \sum_{k=0}^\infty B^k \cos^k \lambda. $$

Therefore

$$\int_{-\pi}^\pi [F(\lambda, \psi^+)]^{-1} d\lambda = 2\pi \sum_{k=0}^\infty B^k \int_{-\pi}^\pi \cos^k \lambda d\lambda \left[ 2\Sigma_\epsilon^+ + \Sigma_\eta^+ \right]^{-1}. $$

Now

$$\int_{-\pi}^\pi \cos^k \lambda d\lambda = 0 \quad \text{for} \ k \ \text{odd,}$$

$$\int_{-\pi}^\pi \cos^2 \lambda d\lambda = \pi$$

and using the fact that for \( j > 2 \)
\[
\int_{-\pi}^{\pi} \cos^2 j \cdot 2 \lambda \sin^2 \lambda d\lambda = \frac{1}{2j} \int_{-\pi}^{\pi} \cos^2 \lambda d\lambda
\]

we have

\[
\int_{-\pi}^{\pi} \cos^2 \lambda d\lambda = \int_{-\pi}^{\pi} \cos^2 j \cdot 2 \lambda d\lambda - \int_{-\pi}^{\pi} \cos^2 j \cdot 2 \lambda \sin^2 \lambda d\lambda
\]

\[
= \frac{2j-1}{2j} \int_{-\pi}^{\pi} \cos^2 \lambda d\lambda.
\]

Thus,

\[
\int_{-\pi}^{\pi} F(\lambda, \eta^\dagger)^{-1} d\lambda = 4\pi^2 \sum_{j=0}^{\infty} c_j B^2 j [2\Sigma e^\dagger + \Sigma_\eta^\dagger]^{-1}
\]

where

\[
c_j = \frac{2j-1}{2j} c_{j-1} \quad j=1, \ldots
\]

\[
c_0 = 1,
\]

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 ratio of the determinants of the asymptotic covariance matrices.

The asymptotic covariance matrix of LIML is given in (5.9), namely
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

\[
A\text{var} \hat{\delta}_{IIV/ML} = \frac{2\pi}{T} \left\{ R' \left[ \int_{-\pi}^{\pi} \left( F(\lambda)^{-1} \otimes F_{Z}(\lambda) \right) d\lambda \right]_{11}^{1} \right\}^{-1} R^{-1}^{-1}
\]  

(1)

with \( f(\lambda) = [F(\lambda, \psi^{T})]_{11} \) - \( [F(\lambda, \psi)]_{11} \), where the last equality follows from (3.6), and \([\cdot]_{11}\) is the top left element of the matrix in square brackets.

Thus,

\[
\frac{1}{A\text{var} \hat{\delta}_{LIML}} = \frac{1}{A\text{var} \hat{\delta}_{IIV/ML}} = \frac{\left| \begin{array}{c} R' \left[ \int_{-\pi}^{\pi} f(\lambda)^{-1} F_{Z}(\lambda) \right]_{11}^{1} R \\ \left| R' \left[ \int_{-\pi}^{\pi} (F(\lambda)^{-1} \otimes F_{Z}(\lambda)) d\lambda \right]_{11}^{1} R \right| \end{array} \right|}{\lim_{T \to \infty} \left| \begin{array}{c} R' \sum_{j=0}^{T} f(\lambda_{j})^{-1} F_{Z}(\lambda_{j}) R \\ \left| R' \left[ \sum_{j=0}^{T} F(\lambda_{j})^{-1} \otimes F_{Z}(\lambda_{j}) \right]_{11}^{1} R \right| \end{array} \right|}.
\]  

(3)

(4)

Now let

\[
B = \left[ \sum_{j=0}^{T} F(\lambda_{j})^{-1} \otimes F_{Z}(\lambda_{j}) \right]_{11}^{1} \quad \text{and} \quad C = \left[ \sum_{j=0}^{T} f(\lambda_{j})^{-1} F_{Z}(\lambda_{j}) \right]^{-1}.
\]

Because \( B \) and \( C \) are positive definite and \( C - B \) is positive semidefinite, see appendix 7.1, we have that \( B^{-1} \cdot C^{-1} \) 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 $\delta$ is at least as efficient as the IIV/ML estimator of $\delta$.

**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, $\Sigma_\eta = q\Sigma_\epsilon$ and of course $\Sigma_\eta^\dagger = q\Sigma_\epsilon^\dagger$, we have

$$F(\lambda, \psi^\dagger) = (2\pi)^{-1}[c(\lambda)\Sigma_\epsilon^\dagger + \Sigma_\eta^\dagger] = (2\pi)^{-1}\Sigma_\epsilon^\dagger \overline{c}(\lambda),$$

where $\overline{c}(\lambda) = c(\lambda) + q$. Therefore

$$\left[\int_\pi^{\pi} F(\lambda, \psi^\dagger) \cdot 1 F_Z(\lambda) d\lambda\right]^{11} = \left[\int_\pi^{\pi} (\Sigma_\epsilon^\dagger)^{-1} \overline{c}(\lambda)^{-1} F_Z(\lambda) d\lambda\right]^{11}$$

$$= \left[2\pi (\Sigma_\epsilon^\dagger)^{-1} \overline{c}(\lambda)^{-1} F_Z(\lambda) d\lambda\right]^{11}$$

$$= (2\pi)^{-1} \left[\Sigma_\epsilon(1,1)^{-1} \left[\int_\pi^{\pi} \overline{c}(\lambda)^{-1} F_Z(\lambda) d\lambda\right]^{-1}\right]$$

$$= \left[\int_\pi^{\pi} f(\lambda)^{-1} F_Z(\lambda) d\lambda\right]^{-1},$$  

(5)
where the last equality follows because

$$2\pi [\Sigma_\varepsilon^{(1,1)} \varepsilon_\lambda]^{-1} - 2\pi [\Sigma_\varepsilon^{(1,1)} c(\lambda) + \Sigma_\eta^{(1,1)}]^{-1} f(\lambda)^{-1}. $$

Hence, making use of (5) it follows that (1), the asymptotic covariance matrix of \( \hat{\delta}_{LIML} \), coincides with the asymptotic covariance of \( \hat{\delta}_{IV/ML} \) 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

$$F(\lambda, \psi^\dagger) = (2\pi)^{-1} [\Sigma_\varepsilon^{\dagger} c(\lambda) + \Sigma_\eta^{\dagger}] = \bar{B}'F(\lambda, \psi)\bar{B},$$

where

$$\bar{B} = \begin{bmatrix} 1 & B_{12} \\ 0 & B_{22} \end{bmatrix}$$

is square and positive definite.

Therefore,

$$\left[ \int_{-\pi}^{\pi} F(\lambda, \psi^\dagger)^{-1} @ F_Z(\lambda) d\lambda \right]^{-1} = \left( \bar{B}' @ I_k \right) \left[ \int_{-\pi}^{\pi} F(\lambda, \psi)^{-1} @ F_Z(\lambda) d\lambda \right]^{-1} \left( \bar{B} @ I_k \right). \quad (6)$$

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, \( \epsilon_{1t} \) and \( \eta_{1t} \), 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

\[
\left[ \int_{-\infty}^{\infty} F(\lambda, \psi) \cdot 1_{\mathbb{R}^2}(\lambda) d\lambda \right]_{11} = \left[ \int_{-\infty}^{\infty} f(\lambda) \cdot 1_{\mathbb{R}^2}(\lambda) d\lambda \right]_{11}
\] (7)

and again (1) reduces to (2).

7. Stochastic Trend in First Equation Only

Up to this point we have been assuming that \( \Sigma \eta \) 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 assumptions mean that if stochastic trends were, in
fact, present in all the structural equations, the asymptotic covariance matrix of the LIML estimator of \( \hat{\delta} \) would be given by (5.14). Because of (6.6), \( F(\lambda, \psi) \) can be replaced by \( F(\lambda, \psi^t) \). Thus

\[
Avar\hat{\delta} = \frac{4 \pi^2}{T} \left[ \int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \right]^{11} \left[ R' \Sigma_z R \right]^{-1},
\]

where \( F(\lambda) = F(\lambda, \psi) \), and because \( F(\lambda) = F(-\lambda) \), (1) can be written as

\[
Avar\hat{\delta} = \frac{2 \pi}{T} \frac{1}{2} \left[ \int_{-\pi}^{\pi} F(\lambda)^{-1} d\lambda \right]^{11} \left[ R' \Sigma_z R \right]^{-1}.
\]

In the case we are interested in \( F(\lambda) \) is not positive definite at \( \lambda = 0 \). Based on heuristic arguments we take

\[
Avar\hat{\delta} = \lim_{\lambda \to 0^+} \frac{2 \pi}{T} \frac{1}{2} \left[ \int_{-\lambda}^{\lambda} F(\lambda)^{-1} d\lambda \right]^{11} \left[ R' \Sigma_z R \right]^{-1}.
\]

as an expression for the asymptotic covariance matrix of \( \hat{\delta}_{\text{LIML}} \). Evaluating the limit, see appendix 7.2, (3) becomes

\[
Avar\hat{\delta} = T^{-1} \left[ \sigma_\eta^4 + 4 \sigma_\eta^2 \sigma_\xi^2 (1 - \rho^2) \right] \left[ R' \Sigma_z R \right]^{-1},
\]

where \( \sigma_\xi^2 \) and \( \sigma_\eta^2 \) are the variances of the disturbances appearing in the first equation and \( \rho^2 \), as given in appendix 7.2, is the coefficient of determination between \( \epsilon_1t \) and the other \( p_1 \) elements in the vector \( \epsilon_t \).
In terms of the signal-noise ratio \( q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2} \), (4) becomes

\[
\text{Avar} \hat{\delta} = T^{-1} \sigma_\varepsilon^2 \left[ q^2 + 4q(1 - \rho^2) \right]^\frac{1}{2} \left[ R' \Sigma_{x} R \right]^{-1}.
\]  

(5)

Again demonstrating the asymptotic superiority of LIML compared with IIV/ML, we have from (4.4.10) that

\[
\text{Avar} \hat{\delta}_{\text{IIV/ML}} = T^{-1} \sigma_\varepsilon^2 \left[ q^2 + 4q \right]^\frac{1}{2} \left[ R' \Sigma_{x} R \right]^{-1},
\]

(6)

hence

\[
\frac{\text{Avar} \hat{\delta}_{\text{LIML}}}{\text{Avar} \hat{\delta}_{\text{IIV/ML}}} = \frac{\left[ q^2 + 4q(1 - \rho^2) \right]^\frac{1}{2}}{\left[ q^2 + 4q \right]^\frac{1}{2}} < 1.
\]

When \( \rho^2 \) 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 \( \rho^2 \) tends to unity. Hence the only thing affecting the asymptotic distribution of the LIML estimator of \( \delta \) is the correlation between \( \varepsilon_{1t} \) and the other elements in \( \varepsilon_{t} \). 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 \( m_1 \times m_1 \), \( A_{12} \) is \( m_1 \times m_2 \), \( A_{21} \) is \( m_2 \times m_1 \), \( A_{22} \) is \( m_2 \times m_2 \) and \( m_1 + m_2 = m \).

We want to show that

\[
\left[ \sum_{j=1}^{T} A_{11}(j)^{-1} \right]^{-1} - \left[ \sum_{j=1}^{T} A(j)^{-1} \right]^{11} > 0,
\]

(1)

where \([ \cdot ]^{11}\) is the \( m_1 \times m_1 \) 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

\[
\left[ \left( \sum_{j=1}^{k-1} A(j)^{-1} \right)^{11} \right]^{-1} + A_{11}(k)^{-1}
\]

to the LHS of (1) for \( T=k \). This yields
Using the fact that for $D>0$ and $F>0$, $D-F>0$ if and only if $F^{-1}D^{-1}>0$, we have from (1) that

\[
\begin{align*}
\left( \sum_{j=1}^{k} A_{11}(j)^{-1} \right)^{-1} - \left[ \sum_{j=1}^{k} A(j)^{-1} \right]_{11}^{-1} - (2)
\end{align*}
\]

or

\[
\begin{align*}
\left( \sum_{j=1}^{k} A(j)^{-1} \right)^{-1} + A_{11}(k)^{-1} - \left[ \sum_{j=1}^{k} A_{11}(j)^{-1} \right] > 0
\end{align*}
\]

or

\[
\begin{align*}
\left[ \sum_{j=1}^{k} A_{11}(j)^{-1} \right]^{-1} - \left( \left( \sum_{j=1}^{k} A_{11}(j)^{-1} \right)^{-1} + A_{11}(k)^{-1} \right]^{-1} > 0.
\end{align*}
\] (3)

On the other hand, noting that

\[
\begin{align*}
\left[ \sum_{j=1}^{k} A(j)^{-1} \right]_{11}^{-1} = \left[ \left( \sum_{j=1}^{k} A(j)^{-1} \right)_{11}^{-1} \right]
\end{align*}
\]

we can write

\[
\begin{align*}
\left( \left( \sum_{j=1}^{k} A_{11}(j)^{-1} \right)^{-1} + A_{11}(k)^{-1} \right]^{-1} - \left[ \sum_{j=1}^{k} A(j)^{-1} \right]_{11}^{-1} - (2)
\end{align*}
\]

or

\[
\begin{align*}
\left[ \sum_{j=1}^{k} A_{11}(j)^{-1} \right]^{-1} - \left( \left( \sum_{j=1}^{k} A_{11}(j)^{-1} \right)^{-1} + A_{11}(k)^{-1} \right]^{-1} > 0.
\end{align*}
\] (3)
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

\[
\int_\lambda^{\lambda} \left[ 2\pi F(\lambda) \right]^{-1} d\lambda = \begin{bmatrix} \int_\lambda^{\lambda} \frac{c}{\Delta_1-\Delta_2 \cos \lambda} d\lambda & -\int_\lambda^{\lambda} \frac{b}{\Delta_1-\Delta_2 \cos \lambda} d\lambda \\ -\int_\lambda^{\lambda} \frac{b}{\Delta_1-\Delta_2 \cos \lambda} d\lambda & \int_\lambda^{\lambda} \frac{1}{c(1-\cos \lambda)} d\lambda + \int_\lambda^{\lambda} \frac{b^2}{c(\Delta_1-\Delta_2 \cos \lambda)} d\lambda \end{bmatrix}
\]
Because
\[
\det \left[ \int_{x}^{x} \frac{d\lambda}{2\pi F(\lambda)} \right]^{-1} = \int_{x}^{x} \frac{d\lambda}{\Delta \lambda \Delta \cos \lambda} = \int_{x}^{x} \frac{d\lambda}{1 - \cos \lambda}
\]
we have
\[
\left[ \int_{x}^{x} \frac{d\lambda}{2\pi F(\lambda)} \right]^{11} = \frac{1}{c \int_{x}^{x} \frac{d\lambda}{\Delta \lambda \Delta \cos \lambda}} + \frac{b^2}{c \int_{x}^{x} \frac{d\lambda}{1 - \cos \lambda}}.
\]

Now
\[
\lim_{x \to 0^+} \int_{x}^{x} \frac{d\lambda}{1 - \cos \lambda} = \lim_{x \to 0^+} \cot \left[ \frac{x}{2} \right] = \infty
\]
and
\[
\lim_{x \to 0^+} \int_{x}^{x} \frac{d\lambda}{\Delta \lambda \Delta \cos \lambda} = \lim_{x \to 0^+} \frac{2}{(\Delta \lambda \Delta \cos \lambda)^{11}} \left[ \frac{\pi}{2} - \arctan \left( \frac{\Delta \lambda \Delta \cos \lambda}{2} \right) \tan \left( \frac{x}{2} \right) \right]
\]
\[
= \frac{2\pi}{2(\Delta \lambda \Delta \cos \lambda)^{11}},
\]
and so
\[
\lim_{x \to 0^+} \left[ \int_{x}^{x} \frac{d\lambda}{2\pi F(\lambda)} \right]^{11} = \frac{2(\Delta \lambda \Delta \cos \lambda)^{11}}{c \cdot 2\pi}.
\]

In terms of the covariance matrices $\Sigma_\epsilon$ and $\Sigma_\eta$, we have that
\[
\Delta \lambda \Delta \cos \lambda = \Delta \lambda \Delta \cos \lambda \left[ \sigma_\eta^2 \Omega_\epsilon + \sigma_\epsilon^2 \Omega_\epsilon - \omega^2 \right],
\]
and after some algebra
\[
\frac{(\Delta \lambda \Delta \cos \lambda)^{11}}{c} = \left[ \sigma_\eta^2 + 4\sigma_\eta^2 \left[ \sigma_\epsilon^2 - \omega^2/\Omega_\epsilon \right] \right]^{11}.\]

Since we are assuming $p=2$, $\Omega_\epsilon$ and $\omega$ are scalar. For $p>2$ and $p_1=p-1$,
\( \Omega_\varepsilon \) is a \( p_1 \times p_1 \) matrix and \( \omega \) is a \( p_1 \times 1 \) vector. It can be verified that (2) holds with \( \omega^2/\Omega_\varepsilon \) replaced by \( \omega' \Omega_\varepsilon^{-1} \omega \).

Finally, using the notation \( \rho^2 \) for the coefficient of determination between \( \varepsilon_{1t} \) and the other \( p_1 \) elements in the vector \( \varepsilon_t \), i.e.

\[
\rho^2 = \sigma_\varepsilon^{-2}_t \omega' \Omega_\varepsilon^{-1} \omega,
\]

(3)

(1) becomes

\[
\lim_{x \to 0^+} \left[ \int \left[ \frac{2 \pi F(\lambda)}{x} \right]^{-1} d\lambda \right]^{11} = \frac{2}{2\pi} \left[ \sigma_\xi^4 + 4\sigma_\xi^2 \sigma_\varepsilon^2 [1 - \rho^2] \right]^{\frac{1}{2}}.
\]

(4)
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

\[ n_t = \lambda n_{t-1} + \delta_1 q_t + \delta_2 q_{t-1} + \mu_t + \epsilon_t \]  
\[ \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \]  
\[ \beta_t = \beta_{t-1} + \xi_t \]

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 $\sigma_\eta^2$ is concentrated out, and the starting values for the hyperparameters were $\sigma_\epsilon^2=0.000001, \sigma_\eta^2=0.000005 \sigma_t^2=0.00000$. 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
169

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 IIIV/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 $\sigma_t^2$ is equal to zero, we estimate the model in differences by means of the TSP program for comparisons purposes. We note that when $\sigma_t^2 = 0$, differencing once model (1.1) yields

$$\Delta n_t = \beta + \lambda \Delta n_{t-1} + \delta_1 q_t + \delta_2 q_{t-1} + \omega_t \quad (1a)$$

where

$$\omega_t = \eta_t + \Delta \epsilon_t \quad (1b)$$

Because the disturbances, $\omega_t$, are serially correlated and $\Delta n_{t-1}$ is a lagged dependent variable, consistent estimation is achieved based on instrumental variable estimators with $\Delta n_{t-2}$ acting as an instrument for $\Delta n_{t-1}$. As before we shall consider two cases. These are:

Case C - where the instruments are

CONST AOUT-1 AEMP-2 AOUT-2 AEXP AWT T AQDK AEXP-1 AWT T-1 AQDK-1

and

Case D - where in addition to the instruments given in C, we also consider AEMP-3 AOUT-3 AEXP-2 AWT T-2 AQDK-2 as instruments.

However, since $\eta_t$ dominates $\epsilon_t$, using AEMP-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' - $\Delta EMP-1$ is included in the set of instruments given in C and

Case D' - $\Delta EMP-1$ 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 IIIV/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 $\Delta EMP\cdot 1$ 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

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.355372</td>
<td>-0.001603</td>
<td>0.111968</td>
<td>0.752064</td>
<td>0.059587</td>
</tr>
</tbody>
</table>

(0.290319) (0.000318) (0.013930) (0.037596) (0.015755)

\[ \sigma_e^2 = 0.00000013880 \quad \sigma_\eta^2 = 0.0000048945 \quad \sigma_T^2 = 0.0000000000 \]

Table 2 - OLS estimates for first differenced equation

<table>
<thead>
<tr>
<th>CONSTR.</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.001717</td>
<td>0.106017</td>
<td>0.728599</td>
<td>0.063339</td>
</tr>
</tbody>
</table>

(0.000396) (0.014522) (0.045486) (0.016119)

Table 3 - IV estimates for first differenced equation

INSTR.: CONSTR. AOUT AOUT-1 AEMP-2

<table>
<thead>
<tr>
<th>CONSTR.</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.001547</td>
<td>0.103886</td>
<td>0.761677</td>
<td>0.057802</td>
</tr>
</tbody>
</table>

(0.000411) (0.014632) (0.050002) (0.016533)
### Table 4 - IV/QML estimates with transformed instruments, case A

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.378326</td>
<td>-0.001627</td>
<td>0.114613</td>
<td>0.748057</td>
<td>0.059503</td>
</tr>
<tr>
<td>(0.297369)</td>
<td>(0.000335)</td>
<td>(0.026826)</td>
<td>(0.039626)</td>
<td>(0.016234)</td>
</tr>
</tbody>
</table>

\[ \sigma^2_{\varepsilon} = 0.0000012152 \quad \sigma^2_{\eta} = 0.0000051951 \quad \sigma^2_{\tau} = 0.0000000000 \]

### Table 5 - IIIV/ML estimates with transformed instruments, case A

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.331485</td>
<td>-0.001614</td>
<td>0.120290</td>
<td>0.751617</td>
<td>0.057341</td>
</tr>
<tr>
<td>(0.284708)</td>
<td>(0.000318)</td>
<td>(0.026123)</td>
<td>(0.037767)</td>
<td>(0.016626)</td>
</tr>
</tbody>
</table>

\[ \sigma^2_{\varepsilon} = 0.0000015553 \quad \sigma^2_{\eta} = 0.0000046499 \quad \sigma^2_{\tau} = 0.0000000000 \]

### Table 6 - IV/QML estimates with untransformed instruments, case A

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.384608</td>
<td>-0.001594</td>
<td>0.101688</td>
<td>0.751917</td>
<td>0.063755</td>
</tr>
<tr>
<td>(0.296205)</td>
<td>(0.000336)</td>
<td>(0.028813)</td>
<td>(0.039913)</td>
<td>(0.016473)</td>
</tr>
</tbody>
</table>

\[ \sigma^2_{\varepsilon} = 0.0000012986 \quad \sigma^2_{\eta} = 0.0000050924 \quad \sigma^2_{\tau} = 0.0000000000 \]

### Table 7 - IIIV/ML estimates with untransformed instruments, case A

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.398523</td>
<td>-0.001602</td>
<td>0.101428</td>
<td>0.750384</td>
<td>0.063871</td>
</tr>
<tr>
<td>(0.300016)</td>
<td>(0.000341)</td>
<td>(0.028836)</td>
<td>(0.040413)</td>
<td>(0.016347)</td>
</tr>
</tbody>
</table>

\[ \sigma^2_{\varepsilon} = 0.0000011987 \quad \sigma^2_{\eta} = 0.0000052636 \quad \sigma^2_{\tau} = 0.0000000000 \]
### Table 8 - IV/QML estimates with transformed instruments, case B

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.366334</td>
<td>-0.001623</td>
<td>0.115793</td>
<td>0.749059</td>
<td>0.059051</td>
</tr>
<tr>
<td>(0.294011)</td>
<td>(0.000327)</td>
<td>(0.021076)</td>
<td>(0.038571)</td>
<td>(0.016032)</td>
</tr>
</tbody>
</table>

\[ \sigma^2 = 0.0000013011 \quad \sigma^2 = 0.0000050496 \quad \sigma^2 = 0.0000000000 \]

### Table 9 - IIV/ML estimates with transformed instruments, case B

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.338629</td>
<td>-0.001611</td>
<td>0.117879</td>
<td>0.751696</td>
<td>0.058039</td>
</tr>
<tr>
<td>(0.286208)</td>
<td>(0.000317)</td>
<td>(0.020985)</td>
<td>(0.037552)</td>
<td>(0.016176)</td>
</tr>
</tbody>
</table>

\[ \sigma^2 = 0.0000015047 \quad \sigma^2 = 0.0000047175 \quad \sigma^2 = 0.0000000000 \]

### Table 10 - IV/QML estimates with untransformed instruments, case B

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.361570</td>
<td>-0.001603</td>
<td>0.107679</td>
<td>0.751692</td>
<td>0.063237</td>
</tr>
<tr>
<td>(0.287485)</td>
<td>(0.000320)</td>
<td>(0.024287)</td>
<td>(0.037928)</td>
<td>(0.016375)</td>
</tr>
</tbody>
</table>

\[ \sigma^2 = 0.0000014860 \quad \sigma^2 = 0.0000047482 \quad \sigma^2 = 0.0000000000 \]

### Table 11 - IIV/ML estimates with untransformed instruments, case B

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>TREND</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.391155</td>
<td>-0.001614</td>
<td>0.105319</td>
<td>0.749175</td>
<td>0.063884</td>
</tr>
<tr>
<td>(0.296412)</td>
<td>(0.000331)</td>
<td>(0.024349)</td>
<td>(0.039110)</td>
<td>(0.016174)</td>
</tr>
</tbody>
</table>

\[ \sigma^2 = 0.0000012613 \quad \sigma^2 = 0.0000051292 \quad \sigma^2 = 0.0000000000 \]
Table 12 - IV estimates for first differenced equation, case C

<table>
<thead>
<tr>
<th></th>
<th>CONST</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.001654</td>
<td>0.092805</td>
<td>0.736950</td>
<td>0.062869</td>
</tr>
<tr>
<td>stderr</td>
<td>0.000420</td>
<td>0.031269</td>
<td>0.048963</td>
<td>0.016238</td>
</tr>
</tbody>
</table>

Table 13 - IV estimates for first differenced equation, case C

<table>
<thead>
<tr>
<th></th>
<th>CONST</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.001591</td>
<td>0.115784</td>
<td>0.756682</td>
<td>0.057791</td>
</tr>
<tr>
<td>stderr</td>
<td>0.000422</td>
<td>0.033890</td>
<td>0.050279</td>
<td>0.016511</td>
</tr>
</tbody>
</table>

Table 14 - IV estimates for first differenced equation, case D'

<table>
<thead>
<tr>
<th></th>
<th>CONST</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.001708</td>
<td>0.104078</td>
<td>0.7298243</td>
<td>0.063270</td>
</tr>
<tr>
<td>stderr</td>
<td>0.000405</td>
<td>0.022519</td>
<td>0.0467747</td>
<td>0.016132</td>
</tr>
</tbody>
</table>

Table 15 - IV estimates for first differenced equation, case D

<table>
<thead>
<tr>
<th></th>
<th>CONST</th>
<th>OUT</th>
<th>EMP-1</th>
<th>OUT-1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.001600</td>
<td>0.117573</td>
<td>0.7555205</td>
<td>0.057860</td>
</tr>
<tr>
<td>stderr</td>
<td>0.000412</td>
<td>0.023756</td>
<td>0.0490003</td>
<td>0.016502</td>
</tr>
</tbody>
</table>
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

\[ Y_{1t} = \beta_1 Y_{2t} + \delta Y_{1t-1} + \gamma_1 Z_{it} + \mu_{1t} + \epsilon_{1t} \] (1a)

\[ Y_{2t} = \beta_1 Y_{1t} + \sum_{i=2}^{4} \gamma_i Z_{it} + \mu_{2t} + \epsilon_{2t} \] (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 \( \epsilon_t \) and \( \eta_t \) are respectively \( \Sigma_\epsilon \) and \( \Sigma_\eta \). The exogenous variables \( Z_{it}, i=1,4 \) are independent random walks generated by

\[ Z_{it} = Z_{it-1} + \nu_{it} \] (2)
where $v_{1t}$ is a Gaussian white noise process with variance $\sigma_{v_{1t}}^2$, $i=1,4$. Our interest centres on estimation of the first equation. This is overidentified.

2. The Choice of the Parameters

The values chosen for $\beta_1, \beta_2$ and $\gamma_i$, $i=1,4$ 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). $\sigma_{v_{1t}}^2$ is fixed at 2.0.

The values for $\Sigma_\epsilon$ and $\Sigma_\eta$ were fixed according to the following argument. Differencing (1.1) yields, in matrix notation,

$$\Delta y_t' B + \Delta z_t' \Gamma = u_t'$$

(1)

where $u_t=(\eta_{1t}+\Delta \epsilon_{1t}, \eta_{2t}+\Delta \epsilon_{2t})'$ is a vector MA(1) process with disturbances covariance matrix $\Omega$,

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} = 2\Sigma_\epsilon + \Sigma_\eta.$$

Models with stationary disturbances were studied in Campos (1986b). So we have chosen the values for the diagonal elements of $\Omega$, 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 $\Omega$ 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 $\Omega$ into $2\Sigma_\varepsilon$ and $\Sigma_\eta$ 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 $\varepsilon_t$ 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 \( i \), \( i=1, \ldots, p(p+1) \) 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 \( y_t', t=1, \ldots, T \) are obtained by solving the reduced form of (1.1), that is,
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 \(\nu\) 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 \(\nu\) 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 \(\eta\) is generated as a bivariate Gaussian white noise process with covariance matrix \(\Sigma_{\eta}\), in order to obtain the \((T+30)x2\) matrix \(\mu\) containing the stochastic trends. Again we have set \(\mu(0,1) = \mu(0,2) = 0\).

Finally the \((T+30)x2\) matrix \(\epsilon\) is formed with each row generated as a bivariate Gaussian white noise process with covariance \(\Sigma_{\epsilon}\).

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 \(\Gamma\), 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 \(y_{1t-1}\). 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 \( \sigma^2_\eta \). In a preliminary study we have concentrated out \( \sigma^2_\epsilon \). Although the results were not significantly different from the ones obtained when \( \sigma^2_\eta \) is concentrated out, it proved more appropriate to concentrate \( \sigma^2_\eta \) out for numerical reasons.

b) IV/QML - as described in section 4.3. The criterion function is given in (4.3.7). \( \sigma^2_\eta \) is concentrated out. The three estimators considered differ according to the transformation applied to the instruments. They are:

IV/QML\(^1\) - 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) IIIV/ML - as described in section 4.3. As for the IV/QML estimation procedures we have considered three estimators, these are:

IIIV/ML\(^1\) - Kalman filter is applied to the instruments. That is, at each step (4.3.6) is optimised with respect to \(\sigma^2_\eta\) and \(\psi_\ast=\sigma^2_\epsilon/\sigma^2_\eta\), with \(\delta\) replaced by \(\hat{\delta}(\hat{\psi}_\ast)\), where \(\hat{\delta}(\hat{\psi}_\ast)\) 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). \(\hat{\psi}_\ast\) was obtained in the previous step;

IIIV/ML\(^2\) - differencing the instruments once. As above but with projection matrix as given in (4.3.5);

IIIV/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 \(\sigma^2_\epsilon\) and \(\sigma^2_\eta\) 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 $\beta_1$ and $\gamma_1$ 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 and 2SLS estimators of $(\beta_1, \gamma_1)$ 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

$$\text{Avar}[\hat{\beta}_1]_{2SLS} = T^{-1}[2\sigma_\xi^2 + \sigma_\eta^2] [R'\Sigma R]^{-1}$$

(1)

$$\text{Avar}[\hat{\gamma}_1]_{IIV/ML} = T^{-1}[\sigma_\eta^4 + 4\sigma_\eta^2\sigma_\xi^2] [R'\Sigma R]^{-1}$$

(2)

The asymptotic covariance matrix for the LIML estimators of $\beta_1$ and $\gamma_1$ 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

\[
\text{Avar}[\hat{\gamma}_1^{\text{LIML}}] = T^{-1}[\sigma^2 + 4\sigma^2(1-\rho^2)]^2 [R'\Sigma R]^{-1},
\]  

(3)

where \( \rho^2 = \sigma^2/\Omega = 0.25 \), 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

\[
\rho^2 = \sigma^2/\Omega = 0.25,
\]  

(4)

and if no lagged dependent variable is included

\[
B = \begin{bmatrix} 1 & -0.4 \\ -0.2 & 1 \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 0.92 & 0.4/0.92 \\ 0.2/0.92 & 1/0.92 \end{bmatrix}, \quad \text{and} \quad \Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\]

Thus,

\[
\Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} = \begin{bmatrix} 1/0.92 & 0.4/0.92 \\ 0.2/0.92 & 1/0.92 \end{bmatrix}
\]

and from (7.5.4)

\[
R = \begin{bmatrix} \Pi_{12} & 1 \\ \Pi_{22} & 0 \end{bmatrix} = \begin{bmatrix} 0.4/0.92 & 1/0.92 \\ 1/0.92 & 0 \\ 1/0.92 & 0 \end{bmatrix},
\]

and so

\[
(R'\Sigma R)^{-1} = \begin{bmatrix} 0.14107 & -0.06134 \\ -0.06134 & 0.52665 \end{bmatrix}
\]  

(5)
Thus, for T=50, for experiment A, the homogeneous case, from (5), (2) and (1) we have

\[
\begin{align*}
\text{ASE}_{\text{IIV/ML}}(\hat{\beta}_1) &= \text{ASE}_{\text{LIML}}(\hat{\beta}_1) = 0.05616 \quad (6a) \\
\text{ASE}_{\text{IIV/ML}}(\hat{\gamma}_1) &= \text{ASE}_{\text{LIML}}(\hat{\gamma}_1) = 0.10852 \quad (6b) \\
\text{ASE}_{2\text{SLS}}(\hat{\beta}_1) &= 0.06505 \quad (7a) \\
\text{ASE}_{2\text{SLS}}(\hat{\gamma}_1) &= 0.12570 \quad (7b)
\end{align*}
\]

For experiment B, the ASEs of the LIML estimators of \( \beta_1 \) and \( \gamma_1 \) are

\[
\begin{align*}
\text{ASE}_{\text{LIML}}(\hat{\beta}_1) &= 0.05557 \quad (8a) \\
\text{ASE}_{\text{LIML}}(\hat{\gamma}_1) &= 0.10737 \quad (8b)
\end{align*}
\]

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 \( \beta_1 \) and \( \gamma_1 \) 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 \( \beta_1 \) and \( \gamma_1 \) are

\[
\begin{align*}
\text{ASE}_{\text{LIML}}(\hat{\beta}_1) &= 0.05312 \quad (9a) \\
\text{ASE}_{\text{LIML}}(\hat{\gamma}_1) &= 0.10263 \quad (9b)
\end{align*}
\]

Again the ASEs of the IIV/ML and 2SLS estimators of \( \beta_1 \) and \( \gamma_1 \) are as in (6) and (7).

Finally for experiment D we have
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 \( \sigma_e^2, \sigma_\eta^2, \beta_1, \delta, \gamma_1 \), 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.
### Table 1 - Estimated biases, standard deviations and RMSE's for Experiment A, T = 50

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_e^2 = 0.500$</th>
<th>$\sigma_{\pi}^2 = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GLS</strong></td>
<td>-0.13399</td>
<td>-0.03944</td>
<td>0.11384</td>
<td>-0.09763</td>
</tr>
<tr>
<td></td>
<td>0.171 0.217</td>
<td>0.247 0.251</td>
<td>0.035 0.120</td>
<td>0.107 0.145</td>
</tr>
<tr>
<td></td>
<td>0.00446</td>
<td>-0.00604</td>
<td>0.01585</td>
<td>-0.03153</td>
</tr>
<tr>
<td><strong>IV/QML(^1)</strong></td>
<td>0.272 0.272</td>
<td>0.371 0.371</td>
<td>0.050 0.053</td>
<td>0.123 0.127</td>
</tr>
<tr>
<td></td>
<td>0.00117</td>
<td>-0.00661</td>
<td>0.00985</td>
<td>-0.01323</td>
</tr>
<tr>
<td><strong>IV/QML(^2)</strong></td>
<td>0.244 0.244</td>
<td>0.321 0.322</td>
<td>0.052 0.053</td>
<td>0.121 0.122</td>
</tr>
<tr>
<td></td>
<td>0.09806</td>
<td>0.00151</td>
<td>0.03975</td>
<td>-0.00919</td>
</tr>
<tr>
<td><strong>IV/QML(^3)</strong></td>
<td>0.412 0.424</td>
<td>0.431 0.431</td>
<td>0.094 0.103</td>
<td>0.246 0.266</td>
</tr>
<tr>
<td><strong>2SLS(^1)</strong></td>
<td>-0.02806</td>
<td>0.03265</td>
<td>0.00349</td>
<td>-0.01113</td>
</tr>
<tr>
<td></td>
<td>0.279 0.280</td>
<td>0.370 0.371</td>
<td>0.053 0.053</td>
<td>0.136 0.136</td>
</tr>
<tr>
<td></td>
<td>0.02243</td>
<td>-0.03343</td>
<td>0.00200</td>
<td>-0.01491</td>
</tr>
<tr>
<td><strong>IIV/ML(^1)</strong></td>
<td>0.224 0.226</td>
<td>0.267 0.269</td>
<td>0.048 0.048</td>
<td>0.122 0.123</td>
</tr>
<tr>
<td></td>
<td>-0.00282</td>
<td>-0.00462</td>
<td>0.00412</td>
<td>-0.01101</td>
</tr>
<tr>
<td><strong>IIV/ML(^2)</strong></td>
<td>0.214 0.214</td>
<td>0.259 0.259</td>
<td>0.052 0.052</td>
<td>0.124 0.125</td>
</tr>
<tr>
<td></td>
<td>0.01731</td>
<td>0.18881</td>
<td>0.03344</td>
<td>-0.09106</td>
</tr>
<tr>
<td><strong>2SLS(^2)</strong></td>
<td>0.459 0.460</td>
<td>0.555 0.586</td>
<td>0.118 0.123</td>
<td>0.340 0.352</td>
</tr>
<tr>
<td></td>
<td>0.08108</td>
<td>0.04847</td>
<td>0.02993</td>
<td>-0.08752</td>
</tr>
<tr>
<td><strong>IIV/ML(^3)</strong></td>
<td>0.352 0.361</td>
<td>0.450 0.453</td>
<td>0.106 0.110</td>
<td>0.277 0.291</td>
</tr>
<tr>
<td></td>
<td>0.05248</td>
<td>-0.01783</td>
<td>-0.00847</td>
<td>-0.00499</td>
</tr>
<tr>
<td><strong>LIML</strong></td>
<td>0.276 0.281</td>
<td>0.308 0.309</td>
<td>0.057 0.057</td>
<td>0.133 0.133</td>
</tr>
<tr>
<td>Asymptotic 2SLS(^1)</td>
<td>0.0650</td>
<td>0.1257</td>
<td>0.1086</td>
<td>0.1086</td>
</tr>
<tr>
<td>standard IIV/ML(^1)</td>
<td>0.0562</td>
<td>0.1086</td>
<td>0.1086</td>
<td>0.1086</td>
</tr>
<tr>
<td>errors LIML</td>
<td>0.0562</td>
<td>0.1086</td>
<td>0.1086</td>
<td>0.1086</td>
</tr>
</tbody>
</table>
Table 2 - Estimated biases, standard deviations and RMSE's for Experiment A, T = 200

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_\epsilon^2 = 0.500$</th>
<th>$\sigma_\eta^2 = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLS</td>
<td>-0.15635</td>
<td>-0.00960</td>
<td>0.12106</td>
<td>-0.07152</td>
</tr>
<tr>
<td>IV/QML$^1$</td>
<td>0.080 0.176</td>
<td>0.125 0.126</td>
<td>0.015 0.123</td>
<td>0.044 0.084</td>
</tr>
<tr>
<td>IV/QML$^2$</td>
<td>-0.00782</td>
<td>-0.00996</td>
<td>0.00342</td>
<td>-0.01810</td>
</tr>
<tr>
<td>IV/QML$^3$</td>
<td>0.133 0.133</td>
<td>0.176 0.176</td>
<td>0.032 0.032</td>
<td>0.052 0.055</td>
</tr>
<tr>
<td>IV/QML$^4$</td>
<td>0.00595</td>
<td>-0.00525</td>
<td>0.00137</td>
<td>-0.01640</td>
</tr>
<tr>
<td>IV/QML$^5$</td>
<td>0.121 0.121</td>
<td>0.148 0.149</td>
<td>0.032 0.032</td>
<td>0.053 0.056</td>
</tr>
<tr>
<td>IV/QML$^6$</td>
<td>0.02374</td>
<td>0.16474</td>
<td>0.00589</td>
<td>-0.01427</td>
</tr>
<tr>
<td>IV/QML$^7$</td>
<td>0.244 0.245</td>
<td>0.409 0.441</td>
<td>0.071 0.071</td>
<td>0.257 0.258</td>
</tr>
<tr>
<td>IV/QML$^8$</td>
<td>-0.00068</td>
<td>0.00184</td>
<td>0.00055</td>
<td>-0.01657</td>
</tr>
<tr>
<td>2SLS$^1$</td>
<td>0.129 0.129</td>
<td>0.152 0.152</td>
<td>0.033 0.033</td>
<td>0.061 0.063</td>
</tr>
<tr>
<td>2SLS$^2$</td>
<td>0.01006</td>
<td>-0.01420</td>
<td>0.00078</td>
<td>-0.01604</td>
</tr>
<tr>
<td>IIV/ML$^1$</td>
<td>0.109 0.109</td>
<td>0.124 0.125</td>
<td>0.032 0.032</td>
<td>0.052 0.055</td>
</tr>
<tr>
<td>IIV/ML$^2$</td>
<td>0.00652</td>
<td>-0.00634</td>
<td>0.00012</td>
<td>-0.01584</td>
</tr>
<tr>
<td>IIV/ML$^3$</td>
<td>0.110 0.110</td>
<td>0.122 0.122</td>
<td>0.032 0.032</td>
<td>0.054 0.056</td>
</tr>
<tr>
<td>IIV/ML$^4$</td>
<td>-0.00877</td>
<td>0.22684</td>
<td>0.00067</td>
<td>-0.00934</td>
</tr>
<tr>
<td>IIV/ML$^5$</td>
<td>0.172 0.172</td>
<td>0.424 0.482</td>
<td>0.075 0.075</td>
<td>0.267 0.267</td>
</tr>
<tr>
<td>IIV/ML$^6$</td>
<td>0.00130</td>
<td>0.19885</td>
<td>0.00255</td>
<td>-0.01141</td>
</tr>
<tr>
<td>IIV/ML$^7$</td>
<td>0.158 0.158</td>
<td>0.422 0.467</td>
<td>0.074 0.074</td>
<td>0.264 0.264</td>
</tr>
<tr>
<td>LIML</td>
<td>0.01237</td>
<td>-0.01252</td>
<td>-0.00044</td>
<td>-0.01525</td>
</tr>
<tr>
<td>Asymptotic 2SLS$^1$</td>
<td>0.0325</td>
<td>0.0628</td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard errors</td>
<td>IIV/ML$^1$</td>
<td>0.0281</td>
<td>0.0543</td>
<td></td>
</tr>
<tr>
<td>errors</td>
<td>LIML</td>
<td>0.0281</td>
<td>0.0543</td>
<td></td>
</tr>
</tbody>
</table>
Table 3 - Estimated biases, standard deviations and RMSE's for Experiment B, T = 50

<table>
<thead>
<tr>
<th></th>
<th>$\sigma^2_{\varepsilon} = 0.500$</th>
<th>$\sigma^2_{\theta} = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLS</td>
<td>-0.10000</td>
<td>0.01679</td>
<td>0.08435</td>
<td>-0.07470</td>
</tr>
<tr>
<td></td>
<td>0.191</td>
<td>0.279</td>
<td>0.038</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>0.279</td>
<td>0.325</td>
<td>0.049</td>
<td>0.122</td>
</tr>
<tr>
<td>IV/QML1</td>
<td>0.01294</td>
<td>-0.01358</td>
<td>0.01025</td>
<td>-0.02275</td>
</tr>
<tr>
<td></td>
<td>0.249</td>
<td>0.324</td>
<td>0.049</td>
<td>0.122</td>
</tr>
<tr>
<td>IV/QML2</td>
<td>0.00295</td>
<td>-0.00226</td>
<td>0.00675</td>
<td>-0.01015</td>
</tr>
<tr>
<td></td>
<td>0.227</td>
<td>0.289</td>
<td>0.052</td>
<td>0.123</td>
</tr>
<tr>
<td>IV/QML3</td>
<td>0.18983</td>
<td>-0.00473</td>
<td>0.01729</td>
<td>-0.06999</td>
</tr>
<tr>
<td></td>
<td>0.470</td>
<td>0.394</td>
<td>0.098</td>
<td>0.259</td>
</tr>
<tr>
<td>2SLS1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.274</td>
<td>0.365</td>
<td>0.054</td>
<td>0.136</td>
</tr>
<tr>
<td>IIIV/ML1</td>
<td>0.02510</td>
<td>-0.03249</td>
<td>0.00059</td>
<td>-0.01223</td>
</tr>
<tr>
<td></td>
<td>0.225</td>
<td>0.256</td>
<td>0.049</td>
<td>0.122</td>
</tr>
<tr>
<td>IIIV/ML2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00149</td>
<td>0.00090</td>
<td>0.00259</td>
<td>-0.00863</td>
</tr>
<tr>
<td></td>
<td>0.212</td>
<td>0.249</td>
<td>0.052</td>
<td>0.125</td>
</tr>
<tr>
<td>2SLS2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.05378</td>
<td>0.23686</td>
<td>0.01378</td>
<td>-0.05233</td>
</tr>
<tr>
<td></td>
<td>0.439</td>
<td>0.560</td>
<td>0.119</td>
<td>0.345</td>
</tr>
<tr>
<td>IIIV/ML3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.13154</td>
<td>0.07854</td>
<td>0.01140</td>
<td>-0.05569</td>
</tr>
<tr>
<td></td>
<td>0.352</td>
<td>0.385</td>
<td>0.110</td>
<td>0.290</td>
</tr>
<tr>
<td>LIML</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.259</td>
<td>0.263</td>
<td>0.054</td>
<td>0.133</td>
</tr>
</tbody>
</table>

Asymptotic standard errors
2SLS1 0.0650 0.1257
IIIV/ML1 0.0562 0.1086
LIML 0.0556 0.1074
Table 4 - Estimated biases, standard deviations and RMSE's for Experiment B, T = 200

<table>
<thead>
<tr>
<th></th>
<th>$\sigma^2 = 0.500$</th>
<th>$\sigma_{\eta}^2 = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLS</td>
<td>-0.11778</td>
<td>0.04541</td>
<td>0.08760</td>
<td>-0.05683</td>
</tr>
<tr>
<td></td>
<td>0.088 0.147</td>
<td>0.130 0.138</td>
<td>0.018 0.090</td>
<td>0.049 0.075</td>
</tr>
<tr>
<td></td>
<td>0.00789</td>
<td>-0.01201</td>
<td>0.00263</td>
<td>-0.01737</td>
</tr>
<tr>
<td>IV/QML1</td>
<td>0.120 0.120</td>
<td>0.139 0.140</td>
<td>0.031 0.031</td>
<td>0.052 0.055</td>
</tr>
<tr>
<td></td>
<td>0.00604</td>
<td>-0.00716</td>
<td>0.00101</td>
<td>-0.01625</td>
</tr>
<tr>
<td>IV/QML2</td>
<td>0.114 0.114</td>
<td>0.123 0.123</td>
<td>0.031 0.031</td>
<td>0.054 0.056</td>
</tr>
<tr>
<td>IV/QML3</td>
<td>0.3240</td>
<td>0.15206</td>
<td>-0.00075</td>
<td>-0.00953</td>
</tr>
<tr>
<td></td>
<td>0.188 0.191</td>
<td>0.265 0.306</td>
<td>0.068 0.068</td>
<td>0.255 0.256</td>
</tr>
<tr>
<td>2SLS1</td>
<td>-0.00113</td>
<td>0.00211</td>
<td>0.00044</td>
<td>-0.01661</td>
</tr>
<tr>
<td></td>
<td>0.127 0.127</td>
<td>0.146 0.146</td>
<td>0.033 0.033</td>
<td>0.061 0.063</td>
</tr>
<tr>
<td></td>
<td>0.00974</td>
<td>-0.01504</td>
<td>0.00085</td>
<td>-0.01598</td>
</tr>
<tr>
<td>IIV/ML1</td>
<td>0.107 0.108</td>
<td>0.108 0.109</td>
<td>0.031 0.031</td>
<td>0.052 0.055</td>
</tr>
<tr>
<td></td>
<td>0.00600</td>
<td>-0.00727</td>
<td>0.00018</td>
<td>-0.01588</td>
</tr>
<tr>
<td>IIV/ML2</td>
<td>0.109 0.109</td>
<td>0.108 0.109</td>
<td>0.031 0.031</td>
<td>0.054 0.056</td>
</tr>
<tr>
<td></td>
<td>-0.00216</td>
<td>0.21816</td>
<td>-0.00451</td>
<td>-0.00756</td>
</tr>
<tr>
<td>2SLS2</td>
<td>0.161 0.161</td>
<td>0.315 0.384</td>
<td>0.072 0.072</td>
<td>0.265 0.265</td>
</tr>
<tr>
<td></td>
<td>0.00911</td>
<td>0.18861</td>
<td>-0.00266</td>
<td>-0.00811</td>
</tr>
<tr>
<td>IIV/ML3</td>
<td>0.150 0.151</td>
<td>0.305 0.359</td>
<td>0.070 0.070</td>
<td>0.262 0.262</td>
</tr>
<tr>
<td></td>
<td>0.01021</td>
<td>-0.01482</td>
<td>0.00043</td>
<td>-0.01549</td>
</tr>
<tr>
<td>LIML</td>
<td>0.107 0.108</td>
<td>0.108 0.109</td>
<td>0.031 0.031</td>
<td>0.053 0.055</td>
</tr>
</tbody>
</table>

Asymptotic standard errors

|                     | 2SLS1               | 0.0325 | 0.0628 |
|                     | IIV/ML1             | 0.0281 | 0.0543 |
|                     | LIML                | 0.0278 | 0.0537 |
Table 5 - Estimated biases, standard deviations and RMSE's for Experiment C, T = 50

<table>
<thead>
<tr>
<th></th>
<th>$\sigma^2 = 0.500$</th>
<th>$\sigma^2 = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLS</td>
<td>-0.12197</td>
<td>0.02450</td>
<td>0.09825</td>
<td>-0.08556</td>
</tr>
<tr>
<td>IV/QML1</td>
<td>0.186 0.223</td>
<td>0.282 0.283</td>
<td>0.041 0.107</td>
<td>0.114 0.142</td>
</tr>
<tr>
<td></td>
<td>0.01124</td>
<td>-0.02228</td>
<td>0.01108</td>
<td>-0.02487</td>
</tr>
<tr>
<td>IV/QML2</td>
<td>0.246 0.247</td>
<td>0.320 0.320</td>
<td>0.051 0.052</td>
<td>0.122 0.125</td>
</tr>
<tr>
<td></td>
<td>0.00040</td>
<td>-0.00955</td>
<td>0.00730</td>
<td>-0.01034</td>
</tr>
<tr>
<td>IV/QML3</td>
<td>0.223 0.223</td>
<td>0.283 0.283</td>
<td>0.054 0.054</td>
<td>0.124 0.124</td>
</tr>
<tr>
<td></td>
<td>0.17828</td>
<td>-0.02210</td>
<td>0.01852</td>
<td>-0.07559</td>
</tr>
<tr>
<td>2SLS1</td>
<td>-0.02701</td>
<td>0.03113</td>
<td>0.00204</td>
<td>0.01000</td>
</tr>
<tr>
<td></td>
<td>0.272 0.274</td>
<td>0.355 0.356</td>
<td>0.056 0.056</td>
<td>0.137 0.138</td>
</tr>
<tr>
<td></td>
<td>0.02026</td>
<td>-0.03813</td>
<td>0.00223</td>
<td>-0.01460</td>
</tr>
<tr>
<td>IIV/ML1</td>
<td>0.216 0.217</td>
<td>0.249 0.252</td>
<td>0.050 0.050</td>
<td>0.122 0.123</td>
</tr>
<tr>
<td></td>
<td>-0.00189</td>
<td>-0.00862</td>
<td>0.00331</td>
<td>-0.00899</td>
</tr>
<tr>
<td>IIV/ML2</td>
<td>0.211 0.211</td>
<td>0.246 0.246</td>
<td>0.054 0.054</td>
<td>0.126 0.126</td>
</tr>
<tr>
<td></td>
<td>0.04760</td>
<td>0.17870</td>
<td>0.01830</td>
<td>-0.07324</td>
</tr>
<tr>
<td>2SLS2</td>
<td>0.386 0.389</td>
<td>0.497 0.529</td>
<td>0.118 0.119</td>
<td>0.329 0.337</td>
</tr>
<tr>
<td></td>
<td>0.12473</td>
<td>0.05159</td>
<td>0.01379</td>
<td>-0.06883</td>
</tr>
<tr>
<td>IIV/ML3</td>
<td>0.351 0.372</td>
<td>0.348 0.352</td>
<td>0.114 0.114</td>
<td>0.293 0.301</td>
</tr>
<tr>
<td></td>
<td>0.01194</td>
<td>-0.03861</td>
<td>0.00164</td>
<td>-0.01166</td>
</tr>
<tr>
<td>LIML</td>
<td>0.221 0.222</td>
<td>0.237 0.241</td>
<td>0.047 0.047</td>
<td>0.114 0.115</td>
</tr>
<tr>
<td>Asymptotic 2SLS1</td>
<td>0.0650</td>
<td>0.1257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>standard  IIV/ML1</td>
<td>0.0562</td>
<td>0.1086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>errors    LIML</td>
<td>0.0531</td>
<td>0.1026</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6 - Estimated biases, standard deviations and RMSE's for Experiment C, T = 200

<table>
<thead>
<tr>
<th>Method</th>
<th>$\sigma^2 = 0.500$</th>
<th>$\sigma^2 = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLS</td>
<td>-0.13778</td>
<td>0.04631</td>
<td>0.10645</td>
<td>-0.06518</td>
</tr>
<tr>
<td>IV/QML1</td>
<td>0.085 0.162</td>
<td>0.132 0.140</td>
<td>0.021 0.109</td>
<td>0.048 0.081</td>
</tr>
<tr>
<td>IV/QML2</td>
<td>0.00798</td>
<td>-0.01519</td>
<td>0.00270</td>
<td>-0.01731</td>
</tr>
<tr>
<td>IV/QML3</td>
<td>0.119 0.119</td>
<td>0.141 0.142</td>
<td>0.032 0.032</td>
<td>0.052 0.055</td>
</tr>
<tr>
<td>2SLS1</td>
<td>0.00611</td>
<td>-0.01072</td>
<td>0.00113</td>
<td>-0.01612</td>
</tr>
<tr>
<td>2SLS2</td>
<td>0.175 0.178</td>
<td>0.274 0.312</td>
<td>0.068 0.068</td>
<td>0.257 0.257</td>
</tr>
<tr>
<td>IIV/ML1</td>
<td>-0.00142</td>
<td>-0.00067</td>
<td>0.00041</td>
<td>-0.00807</td>
</tr>
<tr>
<td>IIV/ML2</td>
<td>0.127 0.127</td>
<td>0.146 0.146</td>
<td>0.033 0.033</td>
<td>0.060 0.063</td>
</tr>
<tr>
<td>IIV/ML3</td>
<td>0.00932</td>
<td>-0.01760</td>
<td>0.00081</td>
<td>-0.01645</td>
</tr>
<tr>
<td>2SLS2</td>
<td>0.03121</td>
<td>0.14903</td>
<td>-0.00430</td>
<td>-0.00807</td>
</tr>
<tr>
<td>IIV/ML3</td>
<td>0.107 0.108</td>
<td>0.108 0.110</td>
<td>0.032 0.032</td>
<td>0.052 0.055</td>
</tr>
<tr>
<td>IIV/ML2</td>
<td>0.109 0.109</td>
<td>0.108 0.109</td>
<td>0.032 0.032</td>
<td>0.053 0.056</td>
</tr>
<tr>
<td>LIML</td>
<td>0.149 0.149</td>
<td>0.302 0.366</td>
<td>0.071 0.071</td>
<td>0.265 0.265</td>
</tr>
<tr>
<td>2SLS1</td>
<td>0.00559</td>
<td>-0.01015</td>
<td>0.00023</td>
<td>-0.01573</td>
</tr>
<tr>
<td>2SLS2</td>
<td>0.149 0.149</td>
<td>0.302 0.366</td>
<td>0.071 0.071</td>
<td>0.265 0.265</td>
</tr>
<tr>
<td>IIV/ML3</td>
<td>0.109 0.109</td>
<td>0.108 0.109</td>
<td>0.032 0.032</td>
<td>0.053 0.056</td>
</tr>
<tr>
<td>LIML</td>
<td>0.00559</td>
<td>-0.01015</td>
<td>0.00023</td>
<td>-0.01573</td>
</tr>
</tbody>
</table>

Asymptotic standard errors

<table>
<thead>
<tr>
<th>Method</th>
<th>2SLS1</th>
<th>IIV/ML1</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0325</td>
<td>0.0281</td>
<td>0.0266</td>
</tr>
<tr>
<td></td>
<td>0.0628</td>
<td>0.0543</td>
<td>0.0513</td>
</tr>
</tbody>
</table>
Table 7 - Estimated biases, standard deviations and RMSE's for Experiment D, T = 50

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_\varepsilon^2 = 0.600$</th>
<th>$\sigma_\eta^2 = 0.300$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLS</td>
<td>-0.10639</td>
<td>0.02544</td>
<td>0.07611</td>
<td>-0.08045</td>
</tr>
<tr>
<td>IV/QML¹</td>
<td>0.188 0.216</td>
<td>0.208 0.210</td>
<td>0.037 0.085</td>
<td>0.104 0.132</td>
</tr>
<tr>
<td></td>
<td>0.01039 -0.01393</td>
<td>0.01018</td>
<td></td>
<td>-0.02875</td>
</tr>
<tr>
<td>IV/QML²</td>
<td>0.222 0.222</td>
<td>0.227 0.228</td>
<td>0.046 0.047</td>
<td>0.110 0.113</td>
</tr>
<tr>
<td></td>
<td>-0.01110</td>
<td>0.01500</td>
<td></td>
<td>-0.01032</td>
</tr>
<tr>
<td>IV/QML³</td>
<td>0.214 0.214</td>
<td>0.209 0.209</td>
<td>0.050 0.050</td>
<td>0.115 0.115</td>
</tr>
<tr>
<td></td>
<td>0.12496 0.00032</td>
<td>0.01475</td>
<td></td>
<td>-0.05643</td>
</tr>
<tr>
<td>2SLS¹</td>
<td>0.379 0.399</td>
<td>0.315 0.315</td>
<td>0.078 0.079</td>
<td>0.196 0.205</td>
</tr>
<tr>
<td></td>
<td>-0.03551</td>
<td>0.05182</td>
<td>0.00095</td>
<td></td>
</tr>
<tr>
<td>IIV/ML¹</td>
<td>0.284 0.286</td>
<td>0.339 0.343</td>
<td>0.055 0.055</td>
<td>0.136 0.136</td>
</tr>
<tr>
<td></td>
<td>0.02619 -0.03127</td>
<td>0.00124</td>
<td></td>
<td>-0.01531</td>
</tr>
<tr>
<td>IIV/ML²</td>
<td>0.219 0.221</td>
<td>0.173 0.176</td>
<td>0.045 0.045</td>
<td>0.110 0.111</td>
</tr>
<tr>
<td></td>
<td>-0.00755 0.01140</td>
<td>0.00250</td>
<td></td>
<td>-0.00808</td>
</tr>
<tr>
<td>2SLS²</td>
<td>0.206 0.206</td>
<td>0.175 0.175</td>
<td>0.051 0.051</td>
<td>0.119 0.119</td>
</tr>
<tr>
<td></td>
<td>0.03002 0.17192</td>
<td>0.01248</td>
<td></td>
<td>-0.04395</td>
</tr>
<tr>
<td>IIV/ML³</td>
<td>0.421 0.422</td>
<td>0.447 0.479</td>
<td>0.101 0.102</td>
<td>0.282 0.286</td>
</tr>
<tr>
<td></td>
<td>0.11103 0.02054</td>
<td>0.00760</td>
<td></td>
<td>-0.03920</td>
</tr>
<tr>
<td>LIML</td>
<td>0.315 0.334</td>
<td>0.241 0.241</td>
<td>0.087 0.087</td>
<td>0.217 0.221</td>
</tr>
<tr>
<td></td>
<td>0.04547 -0.03024</td>
<td>-0.00403</td>
<td></td>
<td>-0.00819</td>
</tr>
<tr>
<td>Asymptotic 2SLS¹</td>
<td>0.225 0.230</td>
<td>0.177 0.179</td>
<td>0.051 0.051</td>
<td>0.124 0.125</td>
</tr>
<tr>
<td>standard errors</td>
<td>IIV/ML¹</td>
<td>0.0650 0.1257</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LIML</td>
<td>0.0499 0.0965</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8 - Estimated biases, standard deviations and RMSE's for Experiment A', T = 50

<table>
<thead>
<tr>
<th></th>
<th>$\sigma^2 = 0.500$</th>
<th>$\sigma^2 = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\delta = 0.500$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GLS</strong></td>
<td>-0.23436</td>
<td>0.04252</td>
<td>0.12214</td>
<td>-0.08136</td>
<td>-0.05555</td>
</tr>
<tr>
<td></td>
<td>0.152 0.280</td>
<td>0.237 0.241</td>
<td>0.036 0.128</td>
<td>0.055 0.099</td>
<td>0.121 0.133</td>
</tr>
<tr>
<td><strong>IV/QML</strong></td>
<td>-0.14920</td>
<td>0.12038</td>
<td>0.03292</td>
<td>-0.04209</td>
<td>-0.02682</td>
</tr>
<tr>
<td></td>
<td>0.323 0.356</td>
<td>0.410 0.427</td>
<td>0.055 0.064</td>
<td>0.078 0.089</td>
<td>0.141 0.144</td>
</tr>
<tr>
<td><strong>IV/QML</strong></td>
<td>-0.16948</td>
<td>0.14838</td>
<td>0.02856</td>
<td>-0.04554</td>
<td>-0.02242</td>
</tr>
<tr>
<td></td>
<td>0.281 0.329</td>
<td>0.398 0.425</td>
<td>0.057 0.064</td>
<td>0.077 0.090</td>
<td>0.135 0.137</td>
</tr>
<tr>
<td><strong>IV/QML</strong></td>
<td>0.14961</td>
<td>0.09049</td>
<td>0.02667</td>
<td>-0.00652</td>
<td>0.01013</td>
</tr>
<tr>
<td></td>
<td>0.550 0.570</td>
<td>0.648 0.654</td>
<td>0.106 0.109</td>
<td>0.088 0.088</td>
<td>0.323 0.323</td>
</tr>
<tr>
<td><strong>2SLS</strong></td>
<td>-0.09561</td>
<td>0.06463</td>
<td>0.02612</td>
<td>-0.04538</td>
<td>-0.01719</td>
</tr>
<tr>
<td></td>
<td>0.377 0.389</td>
<td>0.424 0.429</td>
<td>0.070 0.075</td>
<td>0.127 0.135</td>
<td>0.148 0.149</td>
</tr>
<tr>
<td><strong>2SLS</strong></td>
<td>-0.00617</td>
<td>-0.04788</td>
<td>0.01509</td>
<td>-0.01650</td>
<td>-0.01692</td>
</tr>
<tr>
<td></td>
<td>0.329 0.329</td>
<td>0.275 0.279</td>
<td>0.062 0.064</td>
<td>0.072 0.074</td>
<td>0.138 0.139</td>
</tr>
<tr>
<td><strong>IV/ML</strong></td>
<td>-0.05092</td>
<td>-0.01019</td>
<td>0.02057</td>
<td>-0.02286</td>
<td>-0.02739</td>
</tr>
<tr>
<td></td>
<td>0.284 0.288</td>
<td>0.275 0.276</td>
<td>0.061 0.065</td>
<td>0.074 0.077</td>
<td>0.139 0.142</td>
</tr>
<tr>
<td><strong>2SLS</strong></td>
<td>0.17037</td>
<td>0.11827</td>
<td>0.02034</td>
<td>0.02754</td>
<td>-0.07046</td>
</tr>
<tr>
<td></td>
<td>0.500 0.529</td>
<td>0.689 0.699</td>
<td>0.113 0.115</td>
<td>0.142 0.145</td>
<td>0.418 0.424</td>
</tr>
<tr>
<td><strong>IV/ML</strong></td>
<td>0.19785</td>
<td>0.11125</td>
<td>0.01252</td>
<td>-0.01062</td>
<td>0.04698</td>
</tr>
<tr>
<td></td>
<td>0.545 0.580</td>
<td>0.672 0.681</td>
<td>0.121 0.122</td>
<td>0.112 0.113</td>
<td>0.357 0.360</td>
</tr>
<tr>
<td><strong>LIML</strong></td>
<td>-0.00310</td>
<td>0.38073</td>
<td>0.00627</td>
<td>-0.01702</td>
<td>0.01477</td>
</tr>
<tr>
<td></td>
<td>0.515 0.515</td>
<td>0.990 1.061</td>
<td>0.060 0.061</td>
<td>0.132 0.133</td>
<td>0.142 0.143</td>
</tr>
</tbody>
</table>
Table 9 - Estimated biases, standard deviations and RMSE's for Experiment A', T = 200

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_\varepsilon^2 = 0.500$</th>
<th>$\sigma_\eta^2 = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\delta = 0.500$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLS</td>
<td>-0.19696</td>
<td>0.03769</td>
<td>0.12148</td>
<td>-0.05599</td>
<td>-0.03809</td>
</tr>
<tr>
<td></td>
<td>0.079</td>
<td>0.213</td>
<td>0.127</td>
<td>0.132</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.09623</td>
<td>0.10816</td>
<td>0.00612</td>
<td>-0.01672</td>
<td>0.00530</td>
</tr>
<tr>
<td></td>
<td>0.133</td>
<td>0.165</td>
<td>0.211</td>
<td>0.238</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.05910</td>
<td>0.05851</td>
<td>0.00472</td>
<td>-0.00934</td>
<td>-0.00472</td>
</tr>
<tr>
<td></td>
<td>0.128</td>
<td>0.141</td>
<td>0.178</td>
<td>0.187</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02670</td>
<td>0.28384</td>
<td>0.00731</td>
<td>0.00464</td>
<td>0.06085</td>
</tr>
<tr>
<td></td>
<td>0.267</td>
<td>0.268</td>
<td>0.579</td>
<td>0.645</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02627</td>
<td>0.01728</td>
<td>0.00542</td>
<td>-0.00932</td>
<td>-0.00476</td>
</tr>
<tr>
<td>2SLS¹</td>
<td>0.174</td>
<td>0.176</td>
<td>0.208</td>
<td>0.209</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.00255</td>
<td>-0.01651</td>
<td>0.00146</td>
<td>-0.00050</td>
<td>-0.00139</td>
</tr>
<tr>
<td>IIV/ML¹</td>
<td>0.125</td>
<td>0.125</td>
<td>0.137</td>
<td>0.138</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00020</td>
<td>-0.01476</td>
<td>0.00303</td>
<td>0.00108</td>
<td>-0.00758</td>
</tr>
<tr>
<td></td>
<td>0.132</td>
<td>0.132</td>
<td>0.141</td>
<td>0.142</td>
<td></td>
</tr>
<tr>
<td>IIV/ML²</td>
<td>0.26159</td>
<td>0.32176</td>
<td>-0.02355</td>
<td>0.10303</td>
<td>-0.11751</td>
</tr>
<tr>
<td></td>
<td>0.454</td>
<td>0.525</td>
<td>0.824</td>
<td>0.885</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02946</td>
<td>0.24897</td>
<td>0.00390</td>
<td>-0.00297</td>
<td>0.06171</td>
</tr>
<tr>
<td>IIV/ML³</td>
<td>0.324</td>
<td>0.326</td>
<td>0.572</td>
<td>0.624</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02429</td>
<td>0.47919</td>
<td>0.00279</td>
<td>-0.00112</td>
<td>0.00463</td>
</tr>
<tr>
<td>LIIML</td>
<td>0.172</td>
<td>0.173</td>
<td>0.758</td>
<td>0.898</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.032</td>
<td>0.032</td>
<td>0.044</td>
<td>0.044</td>
<td>0.102</td>
</tr>
</tbody>
</table>
Table 10 - Estimated biases, standard deviations and RMSE's for Experiment C', T = 50

<table>
<thead>
<tr>
<th>Method</th>
<th>$\sigma_2 - 0.500$</th>
<th>$\sigma^2 - 0.500$</th>
<th>$\beta_1 - 0.200$</th>
<th>$\delta - 0.500$</th>
<th>$\gamma_1 - 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLS</td>
<td>-0.22268</td>
<td>0.10812</td>
<td>0.10959</td>
<td>-0.08120</td>
<td>-0.04979</td>
</tr>
<tr>
<td>IV/QML</td>
<td>0.183 0.289</td>
<td>0.276 0.297</td>
<td>0.045 0.119</td>
<td>0.062 0.102</td>
<td>0.124 0.134</td>
</tr>
<tr>
<td></td>
<td>-0.14938</td>
<td>0.12936</td>
<td>0.03000</td>
<td>-0.04621</td>
<td>-0.01756</td>
</tr>
<tr>
<td>IV/QML</td>
<td>0.286 0.323</td>
<td>0.390 0.411</td>
<td>0.056 0.064</td>
<td>0.077 0.090</td>
<td>0.138 0.139</td>
</tr>
<tr>
<td></td>
<td>-0.14514</td>
<td>0.12406</td>
<td>0.02570</td>
<td>-0.04498</td>
<td>-0.02055</td>
</tr>
<tr>
<td>IV/QML</td>
<td>0.271 0.308</td>
<td>0.348 0.369</td>
<td>0.060 0.066</td>
<td>0.077 0.089</td>
<td>0.137 0.139</td>
</tr>
<tr>
<td></td>
<td>0.15922</td>
<td>0.06275</td>
<td>0.00967</td>
<td>-0.00643</td>
<td>0.04059</td>
</tr>
<tr>
<td>IV/QML</td>
<td>0.478 0.504</td>
<td>0.466 0.471</td>
<td>0.111 0.112</td>
<td>0.088 0.089</td>
<td>0.302 0.305</td>
</tr>
<tr>
<td></td>
<td>-0.09055</td>
<td>0.06402</td>
<td>0.02456</td>
<td>-0.04723</td>
<td>-0.01646</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.355 0.366</td>
<td>0.414 0.419</td>
<td>0.067 0.072</td>
<td>0.125 0.132</td>
<td>0.147 0.148</td>
</tr>
<tr>
<td></td>
<td>-0.01986</td>
<td>-0.02892</td>
<td>0.01538</td>
<td>-0.02318</td>
<td>-0.01162</td>
</tr>
<tr>
<td>IV/ML</td>
<td>0.316 0.317</td>
<td>0.278 0.280</td>
<td>0.062 0.064</td>
<td>0.073 0.076</td>
<td>0.138 0.138</td>
</tr>
<tr>
<td></td>
<td>-0.06182</td>
<td>0.00955</td>
<td>0.02046</td>
<td>-0.02934</td>
<td>-0.02358</td>
</tr>
<tr>
<td>IV/ML</td>
<td>0.271 0.278</td>
<td>0.278 0.278</td>
<td>0.061 0.065</td>
<td>0.073 0.079</td>
<td>0.139 0.141</td>
</tr>
<tr>
<td></td>
<td>0.18887</td>
<td>0.13490</td>
<td>-0.00038</td>
<td>0.03395</td>
<td>-0.04310</td>
</tr>
<tr>
<td>2SLS</td>
<td>0.528 0.561</td>
<td>0.631 0.645</td>
<td>0.119 0.119</td>
<td>0.141 0.145</td>
<td>0.407 0.410</td>
</tr>
<tr>
<td></td>
<td>0.19562</td>
<td>0.06507</td>
<td>-0.00094</td>
<td>-0.00729</td>
<td>0.06880</td>
</tr>
<tr>
<td>IV/ML</td>
<td>0.500 0.537</td>
<td>0.475 0.479</td>
<td>0.116 0.116</td>
<td>0.103 0.104</td>
<td>0.345 0.352</td>
</tr>
<tr>
<td></td>
<td>-0.08790</td>
<td>0.46775</td>
<td>0.01241</td>
<td>-0.03388</td>
<td>0.01447</td>
</tr>
<tr>
<td>LIML</td>
<td>0.462 0.470</td>
<td>1.067 1.166</td>
<td>0.060 0.061</td>
<td>0.150 0.154</td>
<td>0.131 0.132</td>
</tr>
</tbody>
</table>
### Table 11 - Estimated biases, standard deviations and RMSE's for Experiment C', T = 200

<table>
<thead>
<tr>
<th></th>
<th>$\sigma_{\varepsilon}^2 = 0.500$</th>
<th>$\sigma_{\eta}^2 = 0.500$</th>
<th>$\beta_1 = 0.200$</th>
<th>$\delta = 0.500$</th>
<th>$\gamma_1 = 1.000$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GLS</strong></td>
<td>-0.18731</td>
<td>0.10717</td>
<td>0.10728</td>
<td>-0.05535</td>
<td>-0.02976</td>
</tr>
<tr>
<td></td>
<td>0.090 0.209</td>
<td>0.145 0.181</td>
<td>0.018 0.109</td>
<td>0.029 0.063</td>
<td>0.056 0.064</td>
</tr>
<tr>
<td><strong>IV/QML</strong></td>
<td><strong>1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.07903</td>
<td>0.08485</td>
<td>0.00569</td>
<td>-0.01906</td>
<td>0.00891</td>
</tr>
<tr>
<td></td>
<td>0.127 0.150</td>
<td>0.179 0.198</td>
<td>0.025 0.026</td>
<td>0.034 0.039</td>
<td>0.062 0.063</td>
</tr>
<tr>
<td><strong>IV/QML</strong></td>
<td><strong>2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.05346</td>
<td>0.05220</td>
<td>0.00462</td>
<td>-0.01414</td>
<td>-0.00254</td>
</tr>
<tr>
<td></td>
<td>0.128 0.139</td>
<td>0.160 0.168</td>
<td>0.026 0.027</td>
<td>0.035 0.037</td>
<td>0.064 0.064</td>
</tr>
<tr>
<td><strong>IV/QML</strong></td>
<td><strong>3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.04888</td>
<td>0.13835</td>
<td>0.00433</td>
<td>-0.00426</td>
<td>0.07436</td>
</tr>
<tr>
<td></td>
<td>0.238 0.243</td>
<td>0.232 0.270</td>
<td>0.082 0.083</td>
<td>0.131 0.131</td>
<td>0.178 0.193</td>
</tr>
<tr>
<td><strong>2SLS</strong></td>
<td><strong>1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02357</td>
<td>0.01916</td>
<td>0.00503</td>
<td>-0.00807</td>
<td>-0.00482</td>
</tr>
<tr>
<td></td>
<td>0.174 0.176</td>
<td>0.205 0.206</td>
<td>0.028 0.029</td>
<td>0.056 0.056</td>
<td>0.070 0.070</td>
</tr>
<tr>
<td><strong>IV/ML</strong></td>
<td><strong>1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.01139</td>
<td>-0.00145</td>
<td>0.00250</td>
<td>-0.00796</td>
<td>0.00449</td>
</tr>
<tr>
<td></td>
<td>0.126 0.126</td>
<td>0.134 0.134</td>
<td>0.025 0.025</td>
<td>0.031 0.032</td>
<td>0.062 0.062</td>
</tr>
<tr>
<td><strong>IV/ML</strong></td>
<td><strong>2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.01514</td>
<td>0.00404</td>
<td>0.00347</td>
<td>-0.00736</td>
<td>-0.00442</td>
</tr>
<tr>
<td></td>
<td>0.130 0.131</td>
<td>0.140 0.140</td>
<td>0.027 0.027</td>
<td>0.034 0.035</td>
<td>0.064 0.064</td>
</tr>
<tr>
<td><strong>2SLS</strong></td>
<td><strong>2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.27976</td>
<td>0.42809</td>
<td>-0.03283</td>
<td>0.10215</td>
<td>-0.09191</td>
</tr>
<tr>
<td></td>
<td>0.453 0.534</td>
<td>1.418 1.482</td>
<td>0.084 0.090</td>
<td>0.233 0.255</td>
<td>0.437 0.447</td>
</tr>
<tr>
<td><strong>IV/ML</strong></td>
<td><strong>3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01514</td>
<td>0.23796</td>
<td>0.00509</td>
<td>-0.01138</td>
<td>0.09536</td>
</tr>
<tr>
<td></td>
<td>0.307 0.308</td>
<td>0.448 0.508</td>
<td>0.085 0.085</td>
<td>0.152 0.152</td>
<td>0.228 0.248</td>
</tr>
<tr>
<td><strong>LIML</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.02286</td>
<td>0.46751</td>
<td>0.00185</td>
<td>0.00040</td>
<td>0.00652</td>
</tr>
<tr>
<td></td>
<td>0.150 0.152</td>
<td>0.672 0.820</td>
<td>0.030 0.030</td>
<td>0.044 0.044</td>
<td>0.099 0.100</td>
</tr>
</tbody>
</table>
Table 12 - Percentiles and first ID for \( \hat{\sigma}_e^2 \) for experiment A, T-50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>GLS</th>
<th>IIV/QML(^1)</th>
<th>IIV/QML(^1)</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0675</td>
<td>0.0783</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0983</td>
<td>0.0527</td>
<td>0.2095</td>
<td>0.2078</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.1657</td>
<td>0.1651</td>
<td>0.2758</td>
<td>0.2806</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.2289</td>
<td>0.3335</td>
<td>0.3729</td>
<td>0.3819</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.3462</td>
<td>0.5098</td>
<td>0.4958</td>
<td>0.5208</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.4920</td>
<td>0.6679</td>
<td>0.6121</td>
<td>0.6449</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>0.5877</td>
<td>0.8017</td>
<td>0.8581</td>
<td>0.8742</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>0.6519</td>
<td>1.0677</td>
<td>0.9781</td>
<td>1.0374</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>0.8720</td>
<td>1.3239</td>
<td>1.1843</td>
<td>2.0586</td>
<td></td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>GLS</th>
<th>IIV/QML(^1)</th>
<th>IIV/QML(^1)</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>6.224E-5</td>
<td>6.276E-5</td>
<td>3.122E-5</td>
<td>1E-8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0644</td>
<td>6.493E-3</td>
<td>0.0856</td>
<td>0.0809</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.1420</td>
<td>0.0716</td>
<td>0.1451</td>
<td>0.1351</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.2767</td>
<td>0.2260</td>
<td>0.2821</td>
<td>0.2770</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.4343</td>
<td>0.4219</td>
<td>0.4396</td>
<td>0.4341</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.6185</td>
<td>0.7048</td>
<td>0.6066</td>
<td>0.6175</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>0.8137</td>
<td>1.1028</td>
<td>0.8461</td>
<td>0.8975</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>0.9186</td>
<td>1.2452</td>
<td>0.9869</td>
<td>1.0566</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>1.1048</td>
<td>1.4830</td>
<td>1.5056</td>
<td>1.7695</td>
<td></td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2063</td>
<td>0.0662</td>
<td>0.0659</td>
<td>-5.675E-3</td>
</tr>
<tr>
<td>5</td>
<td>0.2557</td>
<td>0.1338</td>
<td>0.1306</td>
<td>0.0972</td>
</tr>
<tr>
<td>10</td>
<td>0.2654</td>
<td>0.1536</td>
<td>0.1465</td>
<td>0.1273</td>
</tr>
<tr>
<td>25</td>
<td>0.2884</td>
<td>0.1794</td>
<td>0.1684</td>
<td>0.1532</td>
</tr>
<tr>
<td>50</td>
<td>0.3153</td>
<td>0.2154</td>
<td>0.2051</td>
<td>0.1943</td>
</tr>
<tr>
<td>75</td>
<td>0.3425</td>
<td>0.2529</td>
<td>0.2318</td>
<td>0.2292</td>
</tr>
<tr>
<td>90</td>
<td>0.3594</td>
<td>0.2842</td>
<td>0.2673</td>
<td>0.2640</td>
</tr>
<tr>
<td>95</td>
<td>0.3643</td>
<td>0.2990</td>
<td>0.2802</td>
<td>0.2783</td>
</tr>
<tr>
<td>maximum</td>
<td>0.3910</td>
<td>0.3067</td>
<td>0.3064</td>
<td>0.3033</td>
</tr>
</tbody>
</table>

| ID          | 0.1208 | 0.1231 | 0.1367 | 0.1461    |
| ID(theor.)  |       |        |        |           |

Table 15 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment A, T-50

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.6643</td>
<td>0.6561</td>
<td>0.7158</td>
<td>0.7146</td>
</tr>
<tr>
<td>5</td>
<td>0.7256</td>
<td>0.7874</td>
<td>0.7946</td>
<td>0.7951</td>
</tr>
<tr>
<td>10</td>
<td>0.7700</td>
<td>0.8176</td>
<td>0.8256</td>
<td>0.8236</td>
</tr>
<tr>
<td>25</td>
<td>0.8278</td>
<td>0.8708</td>
<td>0.8911</td>
<td>0.8967</td>
</tr>
<tr>
<td>50</td>
<td>0.8901</td>
<td>0.9602</td>
<td>0.9727</td>
<td>0.9713</td>
</tr>
<tr>
<td>75</td>
<td>0.9751</td>
<td>1.0554</td>
<td>1.0727</td>
<td>1.0900</td>
</tr>
<tr>
<td>90</td>
<td>1.0504</td>
<td>1.1424</td>
<td>1.1473</td>
<td>1.1633</td>
</tr>
<tr>
<td>95</td>
<td>1.0804</td>
<td>1.1731</td>
<td>1.1891</td>
<td>1.2411</td>
</tr>
<tr>
<td>maximum</td>
<td>1.1788</td>
<td>1.2725</td>
<td>1.3490</td>
<td>1.3454</td>
</tr>
</tbody>
</table>

| ID          | 0.3217 | 0.3397 |        |        |
| ID(theor.)  | 0.3128 | 0.3410 |        |        |
Table 16 - Percentiles and first ID for $\hat{\sigma}_e^2$ for experiment A, T-200

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>II/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.1256</td>
<td>0.1573</td>
<td>0.2734</td>
<td>0.2734</td>
</tr>
<tr>
<td>5</td>
<td>0.2129</td>
<td>0.2875</td>
<td>0.3373</td>
<td>0.3270</td>
</tr>
<tr>
<td>10</td>
<td>0.2591</td>
<td>0.3412</td>
<td>0.3800</td>
<td>0.3752</td>
</tr>
<tr>
<td>25</td>
<td>0.2954</td>
<td>0.4324</td>
<td>0.4390</td>
<td>0.4382</td>
</tr>
<tr>
<td>50</td>
<td>0.3380</td>
<td>0.5079</td>
<td>0.5018</td>
<td>0.5094</td>
</tr>
<tr>
<td>75</td>
<td>0.3951</td>
<td>0.5627</td>
<td>0.5683</td>
<td>0.5710</td>
</tr>
<tr>
<td>90</td>
<td>0.4449</td>
<td>0.6760</td>
<td>0.6547</td>
<td>0.6565</td>
</tr>
<tr>
<td>95</td>
<td>0.4807</td>
<td>0.7423</td>
<td>0.6866</td>
<td>0.7001</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.5904</td>
<td>0.9981</td>
<td>0.9060</td>
<td>0.9041</td>
</tr>
</tbody>
</table>

Table 17 - Percentiles and first ID for $\hat{\sigma}_u^2$ for experiment A, T-200

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>II/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.1976</td>
<td>0.1451</td>
<td>0.1864</td>
<td>0.1878</td>
</tr>
<tr>
<td>5</td>
<td>0.3133</td>
<td>0.2575</td>
<td>0.2932</td>
<td>0.2902</td>
</tr>
<tr>
<td>10</td>
<td>0.3506</td>
<td>0.3018</td>
<td>0.3298</td>
<td>0.3220</td>
</tr>
<tr>
<td>25</td>
<td>0.3939</td>
<td>0.3426</td>
<td>0.3809</td>
<td>0.3825</td>
</tr>
<tr>
<td>50</td>
<td>0.4823</td>
<td>0.4842</td>
<td>0.4928</td>
<td>0.4978</td>
</tr>
<tr>
<td>75</td>
<td>0.5778</td>
<td>0.6003</td>
<td>0.5808</td>
<td>0.5855</td>
</tr>
<tr>
<td>90</td>
<td>0.6604</td>
<td>0.7278</td>
<td>0.6313</td>
<td>0.6321</td>
</tr>
<tr>
<td>95</td>
<td>0.6924</td>
<td>0.7881</td>
<td>0.6898</td>
<td>0.6982</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.3812</td>
<td>1.0178</td>
<td>0.7607</td>
<td>0.7538</td>
</tr>
</tbody>
</table>

ID          | 0.3015 | 0.3101 |
ID (theor.) | 0.3179 | 0.3231 |
Table 18 - Percentiles and first ID for $\hat{\theta}_1$ for experiment A, T=200

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML$_{1}$</th>
<th>IIIV/ML$_{1}$</th>
<th>LIIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.2923</td>
<td>0.0989</td>
<td>0.0969</td>
<td>0.0907</td>
</tr>
<tr>
<td>5</td>
<td>0.3000</td>
<td>0.1439</td>
<td>0.1430</td>
<td>0.1377</td>
</tr>
<tr>
<td>10</td>
<td>0.3008</td>
<td>0.1623</td>
<td>0.1592</td>
<td>0.1596</td>
</tr>
<tr>
<td>25</td>
<td>0.3108</td>
<td>0.1826</td>
<td>0.1817</td>
<td>0.1792</td>
</tr>
<tr>
<td>50</td>
<td>0.3190</td>
<td>0.2083</td>
<td>0.2046</td>
<td>0.2035</td>
</tr>
<tr>
<td>75</td>
<td>0.3303</td>
<td>0.2278</td>
<td>0.2252</td>
<td>0.2245</td>
</tr>
<tr>
<td>90</td>
<td>0.3432</td>
<td>0.2415</td>
<td>0.2367</td>
<td>0.2352</td>
</tr>
<tr>
<td>95</td>
<td>0.3514</td>
<td>0.2457</td>
<td>0.2431</td>
<td>0.2416</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.3609</td>
<td>0.2561</td>
<td>0.2553</td>
<td>0.2539</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML$_{1}$</th>
<th>IIIV/ML$_{1}$</th>
<th>LIIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.8254</td>
<td>0.8593</td>
<td>0.8643</td>
<td>0.8633</td>
</tr>
<tr>
<td>5</td>
<td>0.8542</td>
<td>0.8917</td>
<td>0.8948</td>
<td>0.8932</td>
</tr>
<tr>
<td>10</td>
<td>0.8682</td>
<td>0.9082</td>
<td>0.9094</td>
<td>0.9127</td>
</tr>
<tr>
<td>25</td>
<td>0.8970</td>
<td>0.9434</td>
<td>0.9446</td>
<td>0.9466</td>
</tr>
<tr>
<td>50</td>
<td>0.9279</td>
<td>0.9900</td>
<td>0.9921</td>
<td>0.9936</td>
</tr>
<tr>
<td>75</td>
<td>0.9658</td>
<td>1.0192</td>
<td>1.0198</td>
<td>1.0225</td>
</tr>
<tr>
<td>90</td>
<td>0.9821</td>
<td>1.0439</td>
<td>1.0501</td>
<td>1.0504</td>
</tr>
<tr>
<td>95</td>
<td>0.9972</td>
<td>1.0633</td>
<td>1.0677</td>
<td>1.0681</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.0130</td>
<td>1.1051</td>
<td>1.1018</td>
<td>1.1047</td>
</tr>
</tbody>
</table>

| ID          | 0.1407 | 0.1377       |
| ID(theor.)  | 0.1333 | 0.1333       |
Table 20 - Percentiles and first ID for $\hat{\sigma}_e^2$ for experiment B, T=50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML(^1)</th>
<th>IIV/ML(^1)</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.000</td>
<td>0.0000</td>
<td>0.0935</td>
<td>0.0961</td>
</tr>
<tr>
<td>5</td>
<td>0.1027</td>
<td>0.1168</td>
<td>0.2022</td>
<td>0.2431</td>
</tr>
<tr>
<td>10</td>
<td>0.1722</td>
<td>0.2371</td>
<td>0.2695</td>
<td>0.2697</td>
</tr>
<tr>
<td>25</td>
<td>0.2576</td>
<td>0.3409</td>
<td>0.3681</td>
<td>0.3826</td>
</tr>
<tr>
<td>50</td>
<td>0.3929</td>
<td>0.5089</td>
<td>0.4974</td>
<td>0.5050</td>
</tr>
<tr>
<td>75</td>
<td>0.5534</td>
<td>0.6465</td>
<td>0.6188</td>
<td>0.6431</td>
</tr>
<tr>
<td>90</td>
<td>0.6358</td>
<td>0.8005</td>
<td>0.8577</td>
<td>0.8685</td>
</tr>
<tr>
<td>95</td>
<td>0.7210</td>
<td>1.0081</td>
<td>0.9603</td>
<td>1.0054</td>
</tr>
<tr>
<td>maximum</td>
<td>0.9901</td>
<td>1.2686</td>
<td>1.2375</td>
<td>1.7751</td>
</tr>
</tbody>
</table>

| ID (theor.) | 0.5882| 0.5988        |
| ID          | 0.5769| 0.6641        |

Table 21 - Percentiles and first ID for $\hat{\sigma}_\eta^2$ for experiment B, T=50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML(^1)</th>
<th>IIV/ML(^1)</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>6.252E-5</td>
<td>6.31E-5</td>
<td>3.106E-5</td>
<td>1E-8</td>
</tr>
<tr>
<td>5</td>
<td>0.0973</td>
<td>0.0446</td>
<td>0.0810</td>
<td>0.0787</td>
</tr>
<tr>
<td>10</td>
<td>0.1673</td>
<td>0.0896</td>
<td>0.1637</td>
<td>0.1379</td>
</tr>
<tr>
<td>25</td>
<td>0.3087</td>
<td>0.2543</td>
<td>0.2985</td>
<td>0.2798</td>
</tr>
<tr>
<td>50</td>
<td>0.5062</td>
<td>0.4285</td>
<td>0.4462</td>
<td>0.4553</td>
</tr>
<tr>
<td>75</td>
<td>0.6810</td>
<td>0.6379</td>
<td>0.6045</td>
<td>0.6572</td>
</tr>
<tr>
<td>90</td>
<td>0.8992</td>
<td>1.0181</td>
<td>0.8226</td>
<td>0.8289</td>
</tr>
<tr>
<td>95</td>
<td>1.0086</td>
<td>1.1405</td>
<td>0.9257</td>
<td>0.9118</td>
</tr>
<tr>
<td>maximum</td>
<td>1.4105</td>
<td>1.4027</td>
<td>1.2725</td>
<td>1.3966</td>
</tr>
</tbody>
</table>

| ID (theor.) | 0.6589| 0.6910        |
| ID          | 0.6564| 0.6820        |
Table 22 - Percentiles and first ID for $\hat{\beta}_1$ for experiment B, T-50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.1947</td>
<td>0.0718</td>
<td>0.0708</td>
<td>-0.0255</td>
</tr>
<tr>
<td>5</td>
<td>0.2209</td>
<td>0.1330</td>
<td>0.1313</td>
<td>0.1059</td>
</tr>
<tr>
<td>10</td>
<td>0.2314</td>
<td>0.1507</td>
<td>0.1435</td>
<td>0.1307</td>
</tr>
<tr>
<td>25</td>
<td>0.2558</td>
<td>0.1723</td>
<td>0.1680</td>
<td>0.1610</td>
</tr>
<tr>
<td>50</td>
<td>0.2870</td>
<td>0.2149</td>
<td>0.2013</td>
<td>0.1931</td>
</tr>
<tr>
<td>75</td>
<td>0.3141</td>
<td>0.2652</td>
<td>0.2328</td>
<td>0.2273</td>
</tr>
<tr>
<td>90</td>
<td>0.3327</td>
<td>0.2766</td>
<td>0.2657</td>
<td>0.2617</td>
</tr>
<tr>
<td>95</td>
<td>0.3430</td>
<td>0.2916</td>
<td>0.2837</td>
<td>0.2798</td>
</tr>
<tr>
<td>maximum</td>
<td>0.3575</td>
<td>0.3137</td>
<td>0.3134</td>
<td>0.3129</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.6669</td>
<td>0.6832</td>
<td>0.7123</td>
<td>0.6655</td>
</tr>
<tr>
<td>5</td>
<td>0.7356</td>
<td>0.7954</td>
<td>0.7994</td>
<td>0.8047</td>
</tr>
<tr>
<td>10</td>
<td>0.7803</td>
<td>0.8249</td>
<td>0.8308</td>
<td>0.8326</td>
</tr>
<tr>
<td>25</td>
<td>0.8410</td>
<td>0.8798</td>
<td>0.8969</td>
<td>0.8991</td>
</tr>
<tr>
<td>50</td>
<td>0.9180</td>
<td>0.9712</td>
<td>0.9745</td>
<td>0.9915</td>
</tr>
<tr>
<td>75</td>
<td>1.0012</td>
<td>1.0674</td>
<td>1.0717</td>
<td>1.0924</td>
</tr>
<tr>
<td>90</td>
<td>1.0917</td>
<td>1.1503</td>
<td>1.1509</td>
<td>1.1617</td>
</tr>
<tr>
<td>95</td>
<td>1.1131</td>
<td>1.1721</td>
<td>1.1968</td>
<td>1.2398</td>
</tr>
<tr>
<td>maximum</td>
<td>1.2148</td>
<td>1.2819</td>
<td>1.3419</td>
<td>1.3469</td>
</tr>
</tbody>
</table>

| ID          | 0.3201 | 0.3291 |
| ID(theor.)  | 0.3128 | 0.3410 |
Table 24 - Percentiles and first ID for $\hat{\sigma}_e^2$ for experiment B, T=200

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.1436</td>
<td>0.2057</td>
<td>0.2658</td>
<td>0.2628</td>
</tr>
<tr>
<td>5</td>
<td>0.2454</td>
<td>0.3250</td>
<td>0.3354</td>
<td>0.3205</td>
</tr>
<tr>
<td>10</td>
<td>0.2932</td>
<td>0.3616</td>
<td>0.3765</td>
<td>0.3732</td>
</tr>
<tr>
<td>25</td>
<td>0.3268</td>
<td>0.4332</td>
<td>0.4395</td>
<td>0.4435</td>
</tr>
<tr>
<td>50</td>
<td>0.3705</td>
<td>0.5024</td>
<td>0.5028</td>
<td>0.5115</td>
</tr>
<tr>
<td>75</td>
<td>0.4418</td>
<td>0.5678</td>
<td>0.5678</td>
<td>0.5695</td>
</tr>
<tr>
<td>90</td>
<td>0.4907</td>
<td>0.6683</td>
<td>0.6491</td>
<td>0.6436</td>
</tr>
<tr>
<td>95</td>
<td>0.5419</td>
<td>0.7234</td>
<td>0.6818</td>
<td>0.6753</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.6046</td>
<td>0.9577</td>
<td>0.8946</td>
<td>0.8985</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.2341</td>
<td>0.1966</td>
<td>0.2099</td>
<td>0.2168</td>
</tr>
<tr>
<td>5</td>
<td>0.3496</td>
<td>0.2866</td>
<td>0.3063</td>
<td>0.2986</td>
</tr>
<tr>
<td>10</td>
<td>0.3930</td>
<td>0.3344</td>
<td>0.3391</td>
<td>0.3372</td>
</tr>
<tr>
<td>25</td>
<td>0.4423</td>
<td>0.3832</td>
<td>0.3893</td>
<td>0.3960</td>
</tr>
<tr>
<td>50</td>
<td>0.5550</td>
<td>0.4823</td>
<td>0.4986</td>
<td>0.4929</td>
</tr>
<tr>
<td>75</td>
<td>0.6419</td>
<td>0.5821</td>
<td>0.5623</td>
<td>0.5670</td>
</tr>
<tr>
<td>90</td>
<td>0.7100</td>
<td>0.6627</td>
<td>0.5976</td>
<td>0.6026</td>
</tr>
<tr>
<td>95</td>
<td>0.7477</td>
<td>0.6951</td>
<td>0.6375</td>
<td>0.6543</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.9266</td>
<td>0.9664</td>
<td>0.7730</td>
<td>0.7870</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.2468</td>
<td>0.1012</td>
<td>0.1005</td>
<td>0.0999</td>
</tr>
<tr>
<td>5</td>
<td>0.2590</td>
<td>0.1447</td>
<td>0.1440</td>
<td>0.1416</td>
</tr>
<tr>
<td>10</td>
<td>0.2645</td>
<td>0.1626</td>
<td>0.1607</td>
<td>0.1612</td>
</tr>
<tr>
<td>25</td>
<td>0.2754</td>
<td>0.1811</td>
<td>0.1802</td>
<td>0.1792</td>
</tr>
<tr>
<td>50</td>
<td>0.2864</td>
<td>0.2043</td>
<td>0.2036</td>
<td>0.2016</td>
</tr>
<tr>
<td>75</td>
<td>0.2991</td>
<td>0.2268</td>
<td>0.2249</td>
<td>0.2272</td>
</tr>
<tr>
<td>90</td>
<td>0.3121</td>
<td>0.2396</td>
<td>0.2371</td>
<td>0.2352</td>
</tr>
<tr>
<td>95</td>
<td>0.3199</td>
<td>0.2446</td>
<td>0.2445</td>
<td>0.2447</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.3273</td>
<td>0.2581</td>
<td>0.2577</td>
<td>0.2562</td>
</tr>
<tr>
<td>ID</td>
<td></td>
<td></td>
<td></td>
<td>0.0764</td>
</tr>
<tr>
<td>ID( theor.)</td>
<td></td>
<td></td>
<td></td>
<td>0.0795</td>
</tr>
</tbody>
</table>

### Table 27 - Percentiles and first ID for $\hat{\gamma}_1$ for experiment B, T-200

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.8339</td>
<td>0.8590</td>
<td>0.8633</td>
<td>0.8611</td>
</tr>
<tr>
<td>5</td>
<td>0.8623</td>
<td>0.8924</td>
<td>0.8941</td>
<td>0.8951</td>
</tr>
<tr>
<td>10</td>
<td>0.8789</td>
<td>0.9092</td>
<td>0.9100</td>
<td>0.9140</td>
</tr>
<tr>
<td>25</td>
<td>0.9078</td>
<td>0.9435</td>
<td>0.9437</td>
<td>0.9436</td>
</tr>
<tr>
<td>50</td>
<td>0.9442</td>
<td>0.9924</td>
<td>0.9934</td>
<td>0.9915</td>
</tr>
<tr>
<td>75</td>
<td>0.9795</td>
<td>1.0203</td>
<td>1.0207</td>
<td>1.0232</td>
</tr>
<tr>
<td>90</td>
<td>1.0037</td>
<td>1.0464</td>
<td>1.0506</td>
<td>1.0487</td>
</tr>
<tr>
<td>95</td>
<td>1.0206</td>
<td>1.0652</td>
<td>1.0680</td>
<td>1.0708</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.0378</td>
<td>1.0975</td>
<td>1.0954</td>
<td>1.0894</td>
</tr>
<tr>
<td>ID</td>
<td></td>
<td></td>
<td></td>
<td>0.1406</td>
</tr>
<tr>
<td>ID( theor.)</td>
<td></td>
<td></td>
<td></td>
<td>0.1333</td>
</tr>
</tbody>
</table>
Table 28 - Percentiles and first ID for $\hat{\sigma}_\varepsilon^2$ for experiment C, T=50

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0947</td>
<td>0.1326</td>
</tr>
<tr>
<td>5</td>
<td>0.0748</td>
<td>0.1118</td>
<td>0.2083</td>
<td>0.2038</td>
</tr>
<tr>
<td>10</td>
<td>0.1393</td>
<td>0.2262</td>
<td>0.2803</td>
<td>0.2593</td>
</tr>
<tr>
<td>25</td>
<td>0.2561</td>
<td>0.3700</td>
<td>0.3710</td>
<td>0.3638</td>
</tr>
<tr>
<td>50</td>
<td>0.3648</td>
<td>0.5147</td>
<td>0.4922</td>
<td>0.5078</td>
</tr>
<tr>
<td>75</td>
<td>0.5234</td>
<td>0.6499</td>
<td>0.6270</td>
<td>0.6141</td>
</tr>
<tr>
<td>90</td>
<td>0.5933</td>
<td>0.7694</td>
<td>0.8634</td>
<td>0.7605</td>
</tr>
<tr>
<td>95</td>
<td>0.6114</td>
<td>1.0034</td>
<td>0.9608</td>
<td>0.9361</td>
</tr>
<tr>
<td>maximum</td>
<td>1.0877</td>
<td>1.3131</td>
<td>1.1473</td>
<td>1.3751</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>6.252E-5</td>
<td>6.279E-5</td>
<td>3.14E-5</td>
<td>0.0245</td>
</tr>
<tr>
<td>5</td>
<td>0.1180</td>
<td>0.0615</td>
<td>0.0896</td>
<td>0.1044</td>
</tr>
<tr>
<td>10</td>
<td>0.1824</td>
<td>0.1108</td>
<td>0.1681</td>
<td>0.1927</td>
</tr>
<tr>
<td>25</td>
<td>0.3033</td>
<td>0.2444</td>
<td>0.2847</td>
<td>0.2902</td>
</tr>
<tr>
<td>50</td>
<td>0.4821</td>
<td>0.4089</td>
<td>0.4184</td>
<td>0.4393</td>
</tr>
<tr>
<td>75</td>
<td>0.6966</td>
<td>0.6398</td>
<td>0.6037</td>
<td>0.6011</td>
</tr>
<tr>
<td>90</td>
<td>0.8921</td>
<td>0.9038</td>
<td>0.8235</td>
<td>0.7915</td>
</tr>
<tr>
<td>95</td>
<td>1.0414</td>
<td>1.1362</td>
<td>0.9004</td>
<td>0.9077</td>
</tr>
<tr>
<td>maximum</td>
<td>1.2836</td>
<td>1.5110</td>
<td>1.3033</td>
<td>1.2616</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>Model</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td></td>
<td>0.1979</td>
<td>0.0555</td>
<td>0.0548</td>
<td>0.0753</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.2274</td>
<td>0.1324</td>
<td>0.1318</td>
<td>0.1300</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.2482</td>
<td>0.1498</td>
<td>0.1454</td>
<td>0.1487</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>0.2716</td>
<td>0.1736</td>
<td>0.1660</td>
<td>0.1659</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0.2953</td>
<td>0.2149</td>
<td>0.2035</td>
<td>0.2048</td>
</tr>
<tr>
<td>75</td>
<td></td>
<td>0.3263</td>
<td>0.2498</td>
<td>0.2342</td>
<td>0.2301</td>
</tr>
<tr>
<td>90</td>
<td></td>
<td>0.3523</td>
<td>0.2838</td>
<td>0.2673</td>
<td>0.2651</td>
</tr>
<tr>
<td>95</td>
<td></td>
<td>0.3693</td>
<td>0.2925</td>
<td>0.2852</td>
<td>0.2870</td>
</tr>
<tr>
<td>Maximum</td>
<td></td>
<td>0.3829</td>
<td>0.3327</td>
<td>0.3310</td>
<td>0.3225</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>Model</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td></td>
<td>0.6598</td>
<td>0.6830</td>
<td>0.7162</td>
<td>0.7312</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.7235</td>
<td>0.7892</td>
<td>0.7995</td>
<td>0.8046</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.7756</td>
<td>0.8105</td>
<td>0.8180</td>
<td>0.8292</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>0.8390</td>
<td>0.8838</td>
<td>0.8938</td>
<td>0.9084</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0.9060</td>
<td>0.9658</td>
<td>0.9706</td>
<td>0.9855</td>
</tr>
<tr>
<td>75</td>
<td></td>
<td>0.9996</td>
<td>1.0598</td>
<td>1.0783</td>
<td>1.0724</td>
</tr>
<tr>
<td>90</td>
<td></td>
<td>1.0666</td>
<td>1.1376</td>
<td>1.1460</td>
<td>1.1212</td>
</tr>
<tr>
<td>95</td>
<td></td>
<td>1.1071</td>
<td>1.1605</td>
<td>1.1949</td>
<td>1.1874</td>
</tr>
<tr>
<td>Maximum</td>
<td></td>
<td>1.1866</td>
<td>1.2981</td>
<td>1.3421</td>
<td>1.2906</td>
</tr>
</tbody>
</table>

| ID          | 0.3280         | 0.2920 |
| ID(theor.)  | 0.3128         | 0.2923 |
Table 32 - Percentiles and first ID for $\hat{\sigma}_e^2$ for experiment C, T-200

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QMLl</th>
<th>IIV/MLl</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.1352</td>
<td>0.1995</td>
<td>0.2635</td>
<td>0.2607</td>
</tr>
<tr>
<td>5</td>
<td>0.2197</td>
<td>0.3190</td>
<td>0.3345</td>
<td>0.3379</td>
</tr>
<tr>
<td>10</td>
<td>0.2772</td>
<td>0.3630</td>
<td>0.3742</td>
<td>0.3853</td>
</tr>
<tr>
<td>25</td>
<td>0.3135</td>
<td>0.4343</td>
<td>0.4420</td>
<td>0.4436</td>
</tr>
<tr>
<td>50</td>
<td>0.3502</td>
<td>0.5070</td>
<td>0.4969</td>
<td>0.4924</td>
</tr>
<tr>
<td>75</td>
<td>0.4121</td>
<td>0.5678</td>
<td>0.5703</td>
<td>0.5587</td>
</tr>
<tr>
<td>90</td>
<td>0.4796</td>
<td>0.6712</td>
<td>0.6489</td>
<td>0.6171</td>
</tr>
<tr>
<td>95</td>
<td>0.5139</td>
<td>0.7200</td>
<td>0.6849</td>
<td>0.6857</td>
</tr>
<tr>
<td>maximum</td>
<td>0.5476</td>
<td>0.9198</td>
<td>0.8714</td>
<td>0.8673</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QMLl</th>
<th>IIV/MLl</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2350</td>
<td>0.1923</td>
<td>0.2016</td>
<td>0.2050</td>
</tr>
<tr>
<td>5</td>
<td>0.3396</td>
<td>0.2854</td>
<td>0.3083</td>
<td>0.3343</td>
</tr>
<tr>
<td>10</td>
<td>0.3961</td>
<td>0.3250</td>
<td>0.3358</td>
<td>0.3494</td>
</tr>
<tr>
<td>25</td>
<td>0.4286</td>
<td>0.3783</td>
<td>0.3910</td>
<td>0.4041</td>
</tr>
<tr>
<td>50</td>
<td>0.5474</td>
<td>0.4769</td>
<td>0.4906</td>
<td>0.4964</td>
</tr>
<tr>
<td>75</td>
<td>0.6413</td>
<td>0.5871</td>
<td>0.5656</td>
<td>0.5551</td>
</tr>
<tr>
<td>90</td>
<td>0.7100</td>
<td>0.6710</td>
<td>0.5972</td>
<td>0.6050</td>
</tr>
<tr>
<td>95</td>
<td>0.7418</td>
<td>0.7013</td>
<td>0.6349</td>
<td>0.6415</td>
</tr>
<tr>
<td>maximum</td>
<td>0.9257</td>
<td>0.9740</td>
<td>0.7726</td>
<td>0.7518</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2584</td>
<td>0.0985</td>
<td>0.0976</td>
<td>0.1096</td>
</tr>
<tr>
<td>5</td>
<td>0.2761</td>
<td>0.1456</td>
<td>0.1449</td>
<td>0.1536</td>
</tr>
<tr>
<td>10</td>
<td>0.2802</td>
<td>0.1626</td>
<td>0.1609</td>
<td>0.1619</td>
</tr>
<tr>
<td>25</td>
<td>0.2932</td>
<td>0.1810</td>
<td>0.1796</td>
<td>0.1816</td>
</tr>
<tr>
<td>50</td>
<td>0.3042</td>
<td>0.2074</td>
<td>0.2048</td>
<td>0.2038</td>
</tr>
<tr>
<td>75</td>
<td>0.3201</td>
<td>0.2275</td>
<td>0.2253</td>
<td>0.2254</td>
</tr>
<tr>
<td>90</td>
<td>0.3361</td>
<td>0.2399</td>
<td>0.2386</td>
<td>0.2377</td>
</tr>
<tr>
<td>95</td>
<td>0.3434</td>
<td>0.2458</td>
<td>0.2449</td>
<td>0.2411</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.3521</td>
<td>0.2570</td>
<td>0.2563</td>
<td>0.2587</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.8261</td>
<td>0.8584</td>
<td>0.8628</td>
<td>0.8616</td>
</tr>
<tr>
<td>5</td>
<td>0.8511</td>
<td>0.8899</td>
<td>0.8950</td>
<td>0.9006</td>
</tr>
<tr>
<td>10</td>
<td>0.8725</td>
<td>0.9088</td>
<td>0.9097</td>
<td>0.9139</td>
</tr>
<tr>
<td>25</td>
<td>0.9036</td>
<td>0.9458</td>
<td>0.9454</td>
<td>0.9498</td>
</tr>
<tr>
<td>50</td>
<td>0.9398</td>
<td>0.9929</td>
<td>0.9938</td>
<td>0.9942</td>
</tr>
<tr>
<td>75</td>
<td>0.9689</td>
<td>1.0192</td>
<td>1.0197</td>
<td>1.0179</td>
</tr>
<tr>
<td>90</td>
<td>0.9933</td>
<td>1.0461</td>
<td>1.0499</td>
<td>1.0480</td>
</tr>
<tr>
<td>95</td>
<td>1.0129</td>
<td>1.0663</td>
<td>1.0679</td>
<td>1.0710</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.0309</td>
<td>1.1028</td>
<td>1.1012</td>
<td>1.1037</td>
</tr>
</tbody>
</table>

ID | 0.1402 | 0.1341 |
ID(theor.) | 0.1333 | 0.1308 |
Table 36 - Percentiles for \( \hat{\sigma}_e^2 \) for experiment A', T=50

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.000E-8</td>
</tr>
<tr>
<td>5</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1268</td>
<td>1.000E-8</td>
</tr>
<tr>
<td>10</td>
<td>0.0818</td>
<td>0.0000</td>
<td>0.1747</td>
<td>1.762E-3</td>
</tr>
<tr>
<td>25</td>
<td>0.1632</td>
<td>0.11832</td>
<td>0.2537</td>
<td>0.1344</td>
</tr>
<tr>
<td>50</td>
<td>0.2545</td>
<td>0.2994</td>
<td>0.4107</td>
<td>0.3298</td>
</tr>
<tr>
<td>75</td>
<td>0.3597</td>
<td>0.5063</td>
<td>0.6054</td>
<td>0.6812</td>
</tr>
<tr>
<td>90</td>
<td>0.4739</td>
<td>0.6558</td>
<td>0.9467</td>
<td>1.2178</td>
</tr>
<tr>
<td>95</td>
<td>0.5059</td>
<td>1.0748</td>
<td>1.0931</td>
<td>1.6415</td>
</tr>
<tr>
<td>maximum</td>
<td>0.8242</td>
<td>1.3842</td>
<td>1.6559</td>
<td>2.3128</td>
</tr>
</tbody>
</table>

Table 37 - Percentiles for \( \hat{\sigma}_\eta^2 \) for experiment A', T=50

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>1.2486E-4</td>
<td>1.486E-4</td>
<td>1.0997E-4</td>
<td>8.51161E-3</td>
</tr>
<tr>
<td>5</td>
<td>0.1747</td>
<td>0.0616</td>
<td>0.1031</td>
<td>0.0846</td>
</tr>
<tr>
<td>10</td>
<td>0.2809</td>
<td>0.1491</td>
<td>0.1424</td>
<td>0.1861</td>
</tr>
<tr>
<td>25</td>
<td>0.3850</td>
<td>0.3268</td>
<td>0.2578</td>
<td>0.2565</td>
</tr>
<tr>
<td>50</td>
<td>0.5025</td>
<td>0.5426</td>
<td>0.3999</td>
<td>0.5464</td>
</tr>
<tr>
<td>75</td>
<td>0.7072</td>
<td>0.8809</td>
<td>0.6078</td>
<td>1.1624</td>
</tr>
<tr>
<td>90</td>
<td>0.8681</td>
<td>1.1761</td>
<td>0.8994</td>
<td>1.8702</td>
</tr>
<tr>
<td>95</td>
<td>1.0230</td>
<td>1.3914</td>
<td>0.9743</td>
<td>3.1720</td>
</tr>
<tr>
<td>maximum</td>
<td>1.1307</td>
<td>1.7970</td>
<td>1.2745</td>
<td>6.1585</td>
</tr>
</tbody>
</table>
Table 38 - Percentiles for $\hat{\beta_1}$ for experiment A', T=50

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2247</td>
<td>0.0158</td>
<td>6.4977E-3</td>
<td>0.0648</td>
</tr>
<tr>
<td>5</td>
<td>0.2703</td>
<td>0.1587</td>
<td>0.1114</td>
<td>0.0846</td>
</tr>
<tr>
<td>10</td>
<td>0.2821</td>
<td>0.1707</td>
<td>0.1406</td>
<td>0.1229</td>
</tr>
<tr>
<td>25</td>
<td>0.2963</td>
<td>0.2012</td>
<td>0.1758</td>
<td>0.1654</td>
</tr>
<tr>
<td>50</td>
<td>0.3190</td>
<td>0.2426</td>
<td>0.2300</td>
<td>0.2104</td>
</tr>
<tr>
<td>75</td>
<td>0.3460</td>
<td>0.2641</td>
<td>0.2544</td>
<td>0.2573</td>
</tr>
<tr>
<td>90</td>
<td>0.3714</td>
<td>0.2950</td>
<td>0.2832</td>
<td>0.2782</td>
</tr>
<tr>
<td>95</td>
<td>0.3829</td>
<td>0.3186</td>
<td>0.3002</td>
<td>0.2891</td>
</tr>
<tr>
<td>maximum</td>
<td>0.4056</td>
<td>0.3506</td>
<td>0.3438</td>
<td>0.3401</td>
</tr>
</tbody>
</table>

Table 39 - Percentiles for $\hat{\delta_1}$ for experiment A', T=50

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2795</td>
<td>0.2753</td>
<td>0.3048</td>
<td>0.1774</td>
</tr>
<tr>
<td>5</td>
<td>0.3123</td>
<td>0.3277</td>
<td>0.3724</td>
<td>0.2895</td>
</tr>
<tr>
<td>10</td>
<td>0.3368</td>
<td>0.3532</td>
<td>0.3874</td>
<td>0.3129</td>
</tr>
<tr>
<td>25</td>
<td>0.3881</td>
<td>0.4168</td>
<td>0.4438</td>
<td>0.3672</td>
</tr>
<tr>
<td>50</td>
<td>0.4251</td>
<td>0.4514</td>
<td>0.4735</td>
<td>0.4718</td>
</tr>
<tr>
<td>75</td>
<td>0.4537</td>
<td>0.5072</td>
<td>0.5374</td>
<td>0.6006</td>
</tr>
<tr>
<td>90</td>
<td>0.4889</td>
<td>0.5670</td>
<td>0.5741</td>
<td>0.6621</td>
</tr>
<tr>
<td>95</td>
<td>0.5041</td>
<td>0.5880</td>
<td>0.6038</td>
<td>0.6813</td>
</tr>
<tr>
<td>maximum</td>
<td>0.5412</td>
<td>0.6388</td>
<td>0.6743</td>
<td>0.7285</td>
</tr>
</tbody>
</table>

Table 40 - Percentiles for $\gamma_1$ for experiment A', T=50

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.6674</td>
<td>0.6730</td>
<td>0.6837</td>
<td>0.6610</td>
</tr>
<tr>
<td>5</td>
<td>0.7474</td>
<td>0.7097</td>
<td>0.7409</td>
<td>0.7417</td>
</tr>
<tr>
<td>10</td>
<td>0.7711</td>
<td>0.7742</td>
<td>0.7936</td>
<td>0.8360</td>
</tr>
<tr>
<td>25</td>
<td>0.8585</td>
<td>0.8942</td>
<td>0.8991</td>
<td>0.9251</td>
</tr>
<tr>
<td>50</td>
<td>0.9391</td>
<td>0.9804</td>
<td>0.9856</td>
<td>1.0153</td>
</tr>
<tr>
<td>75</td>
<td>1.0294</td>
<td>1.0511</td>
<td>1.0605</td>
<td>1.1319</td>
</tr>
<tr>
<td>90</td>
<td>1.0963</td>
<td>1.1351</td>
<td>1.1415</td>
<td>1.2044</td>
</tr>
<tr>
<td>95</td>
<td>1.1583</td>
<td>1.2082</td>
<td>1.2352</td>
<td>1.2381</td>
</tr>
<tr>
<td>maximum</td>
<td>1.2567</td>
<td>1.3691</td>
<td>1.4081</td>
<td>1.2930</td>
</tr>
</tbody>
</table>
Table 41 - Percentiles for $\hat{\sigma}_e^2$ for experiment A', T=200

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML$^1$</th>
<th>IIV/ML$^1$</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.1194</td>
<td>0.1089</td>
<td>0.1858</td>
<td>0.0230</td>
</tr>
<tr>
<td>5</td>
<td>0.1842</td>
<td>0.2086</td>
<td>0.3231</td>
<td>0.2165</td>
</tr>
<tr>
<td>10</td>
<td>0.2129</td>
<td>0.2319</td>
<td>0.3480</td>
<td>0.2421</td>
</tr>
<tr>
<td>25</td>
<td>0.2533</td>
<td>0.3181</td>
<td>0.4256</td>
<td>0.3613</td>
</tr>
<tr>
<td>50</td>
<td>0.2936</td>
<td>0.3885</td>
<td>0.4948</td>
<td>0.4710</td>
</tr>
<tr>
<td>75</td>
<td>0.3573</td>
<td>0.4889</td>
<td>0.5694</td>
<td>0.5870</td>
</tr>
<tr>
<td>90</td>
<td>0.4117</td>
<td>0.5903</td>
<td>0.6546</td>
<td>0.6975</td>
</tr>
<tr>
<td>95</td>
<td>0.4415</td>
<td>0.6261</td>
<td>0.7071</td>
<td>0.7589</td>
</tr>
<tr>
<td>maximum</td>
<td>0.5068</td>
<td>0.7995</td>
<td>1.0138</td>
<td>0.9232</td>
</tr>
</tbody>
</table>

Table 42 - Percentiles for $\hat{\sigma}_\eta^2$ for experiment A', T=200

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML$^1$</th>
<th>IIV/ML$^1$</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2328</td>
<td>0.2138</td>
<td>0.2184</td>
<td>0.2472</td>
</tr>
<tr>
<td>5</td>
<td>0.3516</td>
<td>0.3042</td>
<td>0.2612</td>
<td>0.3248</td>
</tr>
<tr>
<td>10</td>
<td>0.3859</td>
<td>0.3735</td>
<td>0.3261</td>
<td>0.3870</td>
</tr>
<tr>
<td>25</td>
<td>0.4604</td>
<td>0.4428</td>
<td>0.3894</td>
<td>0.5157</td>
</tr>
<tr>
<td>50</td>
<td>0.5259</td>
<td>0.5841</td>
<td>0.4753</td>
<td>0.7188</td>
</tr>
<tr>
<td>75</td>
<td>0.6097</td>
<td>0.7464</td>
<td>0.5498</td>
<td>1.2011</td>
</tr>
<tr>
<td>90</td>
<td>0.7104</td>
<td>0.8835</td>
<td>0.6675</td>
<td>1.7448</td>
</tr>
<tr>
<td>95</td>
<td>0.7724</td>
<td>0.9965</td>
<td>0.7460</td>
<td>2.8731</td>
</tr>
<tr>
<td>maximum</td>
<td>0.8516</td>
<td>1.3278</td>
<td>0.9200</td>
<td>4.4243</td>
</tr>
</tbody>
</table>
Table 43 - Percentiles for $\hat{\beta}_1$ for experiment A', T-200

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2767</td>
<td>0.1374</td>
<td>0.1373</td>
<td>0.1056</td>
</tr>
<tr>
<td>5</td>
<td>0.2963</td>
<td>0.1624</td>
<td>0.1545</td>
<td>0.1488</td>
</tr>
<tr>
<td>10</td>
<td>0.2999</td>
<td>0.1755</td>
<td>0.1708</td>
<td>0.1578</td>
</tr>
<tr>
<td>25</td>
<td>0.3109</td>
<td>0.1918</td>
<td>0.1885</td>
<td>0.1854</td>
</tr>
<tr>
<td>50</td>
<td>0.3222</td>
<td>0.2052</td>
<td>0.2027</td>
<td>0.2062</td>
</tr>
<tr>
<td>75</td>
<td>0.3328</td>
<td>0.2234</td>
<td>0.2180</td>
<td>0.2253</td>
</tr>
<tr>
<td>90</td>
<td>0.3408</td>
<td>0.2385</td>
<td>0.2319</td>
<td>0.2399</td>
</tr>
<tr>
<td>95</td>
<td>0.3444</td>
<td>0.2441</td>
<td>0.2402</td>
<td>0.2504</td>
</tr>
<tr>
<td>maximum</td>
<td>0.3527</td>
<td>0.2641</td>
<td>0.2559</td>
<td>0.2802</td>
</tr>
</tbody>
</table>

Table 44 - Percentiles for $\hat{\beta}_1$ for experiment A', T-200

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.3803</td>
<td>0.3939</td>
<td>0.4222</td>
<td>0.3796</td>
</tr>
<tr>
<td>5</td>
<td>0.3839</td>
<td>0.4212</td>
<td>0.4419</td>
<td>0.4104</td>
</tr>
<tr>
<td>10</td>
<td>0.4120</td>
<td>0.4334</td>
<td>0.4610</td>
<td>0.4311</td>
</tr>
<tr>
<td>25</td>
<td>0.4243</td>
<td>0.4605</td>
<td>0.4805</td>
<td>0.4732</td>
</tr>
<tr>
<td>50</td>
<td>0.4444</td>
<td>0.4873</td>
<td>0.5031</td>
<td>0.5090</td>
</tr>
<tr>
<td>75</td>
<td>0.4644</td>
<td>0.5075</td>
<td>0.5190</td>
<td>0.5322</td>
</tr>
<tr>
<td>90</td>
<td>0.4753</td>
<td>0.5256</td>
<td>0.5333</td>
<td>0.5462</td>
</tr>
<tr>
<td>95</td>
<td>0.4827</td>
<td>0.5401</td>
<td>0.5454</td>
<td>0.5507</td>
</tr>
<tr>
<td>maximum</td>
<td>0.5186</td>
<td>0.5800</td>
<td>0.5710</td>
<td>0.5820</td>
</tr>
</tbody>
</table>

Table 45 - Percentiles for $\hat{\gamma}_1$ for experiment A', T-200

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.8352</td>
<td>0.8491</td>
<td>0.8333</td>
<td>0.7699</td>
</tr>
<tr>
<td>5</td>
<td>0.8787</td>
<td>0.9086</td>
<td>0.8980</td>
<td>0.8071</td>
</tr>
<tr>
<td>10</td>
<td>0.8929</td>
<td>0.9213</td>
<td>0.9165</td>
<td>0.8468</td>
</tr>
<tr>
<td>25</td>
<td>0.9228</td>
<td>0.9646</td>
<td>0.9535</td>
<td>0.9431</td>
</tr>
<tr>
<td>50</td>
<td>0.9546</td>
<td>1.0039</td>
<td>0.9989</td>
<td>1.0179</td>
</tr>
<tr>
<td>75</td>
<td>0.9971</td>
<td>1.0370</td>
<td>1.0305</td>
<td>1.0766</td>
</tr>
<tr>
<td>90</td>
<td>1.0378</td>
<td>1.0918</td>
<td>1.0846</td>
<td>1.1270</td>
</tr>
<tr>
<td>95</td>
<td>1.0532</td>
<td>1.1118</td>
<td>1.1044</td>
<td>1.1578</td>
</tr>
<tr>
<td>maximum</td>
<td>1.1039</td>
<td>1.1926</td>
<td>1.1896</td>
<td>1.2211</td>
</tr>
</tbody>
</table>
Table 46 - Percentiles for $\hat{\sigma}_\epsilon^2$ for experiment C', T=50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.000E-8</td>
</tr>
<tr>
<td>5</td>
<td>0.0127</td>
<td>0.0000</td>
<td>0.1085</td>
<td>1.000E-8</td>
</tr>
<tr>
<td>10</td>
<td>0.0454</td>
<td>5.550E-3</td>
<td>0.1704</td>
<td>1.000E-8</td>
</tr>
<tr>
<td>25</td>
<td>0.1460</td>
<td>0.1284</td>
<td>0.2611</td>
<td>0.0563</td>
</tr>
<tr>
<td>50</td>
<td>0.2453</td>
<td>0.3016</td>
<td>0.3900</td>
<td>0.3074</td>
</tr>
<tr>
<td>75</td>
<td>0.3769</td>
<td>0.4757</td>
<td>0.6222</td>
<td>0.5216</td>
</tr>
<tr>
<td>90</td>
<td>0.5226</td>
<td>0.6644</td>
<td>0.9396</td>
<td>1.0251</td>
</tr>
<tr>
<td>95</td>
<td>0.5818</td>
<td>1.0192</td>
<td>1.0783</td>
<td>1.2576</td>
</tr>
<tr>
<td>maximum</td>
<td>0.9950</td>
<td>1.1403</td>
<td>1.7989</td>
<td>2.7206</td>
</tr>
</tbody>
</table>

Table 47 - Percentiles for $\hat{\sigma}_\eta^2$ for experiment C', T=50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>1.9132E-4</td>
<td>1.9704E-4</td>
<td>1.3983E-4</td>
<td>0.0101</td>
</tr>
<tr>
<td>5</td>
<td>0.2161</td>
<td>0.1065</td>
<td>0.0546</td>
<td>0.1570</td>
</tr>
<tr>
<td>10</td>
<td>0.2755</td>
<td>0.1921</td>
<td>0.1407</td>
<td>0.2128</td>
</tr>
<tr>
<td>25</td>
<td>0.3891</td>
<td>0.3239</td>
<td>0.2528</td>
<td>0.3106</td>
</tr>
<tr>
<td>50</td>
<td>0.5646</td>
<td>0.5792</td>
<td>0.4284</td>
<td>0.6094</td>
</tr>
<tr>
<td>75</td>
<td>0.7946</td>
<td>0.8456</td>
<td>0.6120</td>
<td>1.1861</td>
</tr>
<tr>
<td>90</td>
<td>1.0372</td>
<td>1.1554</td>
<td>0.8710</td>
<td>2.2638</td>
</tr>
<tr>
<td>95</td>
<td>1.1782</td>
<td>1.3426</td>
<td>1.0000</td>
<td>3.3483</td>
</tr>
<tr>
<td>maximum</td>
<td>1.2044</td>
<td>1.9310</td>
<td>1.2358</td>
<td>6.1755</td>
</tr>
</tbody>
</table>
### Table 48 - Percentiles for \( \hat{\beta}_1 \) for experiment C', T-50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.1636</td>
<td>0.0574</td>
<td>5.53021E-3</td>
<td>0.0739</td>
</tr>
<tr>
<td>5</td>
<td>0.2443</td>
<td>0.1408</td>
<td>0.1083</td>
<td>0.1301</td>
</tr>
<tr>
<td>10</td>
<td>0.2639</td>
<td>0.1608</td>
<td>0.1361</td>
<td>0.1420</td>
</tr>
<tr>
<td>25</td>
<td>0.2822</td>
<td>0.1945</td>
<td>0.1746</td>
<td>0.1691</td>
</tr>
<tr>
<td>50</td>
<td>0.3095</td>
<td>0.2406</td>
<td>0.2295</td>
<td>0.2058</td>
</tr>
<tr>
<td>75</td>
<td>0.3346</td>
<td>0.2612</td>
<td>0.2518</td>
<td>0.2587</td>
</tr>
<tr>
<td>90</td>
<td>0.3749</td>
<td>0.2996</td>
<td>0.2931</td>
<td>0.2858</td>
</tr>
<tr>
<td>95</td>
<td>0.3914</td>
<td>0.3167</td>
<td>0.3061</td>
<td>0.3152</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.4109</td>
<td>0.3529</td>
<td>0.3465</td>
<td>0.3512</td>
</tr>
</tbody>
</table>

### Table 49 - Percentiles for \( \hat{\alpha}_1 \) for experiment C', T-50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.2687</td>
<td>0.2618</td>
<td>0.3075</td>
<td>0.1693</td>
</tr>
<tr>
<td>5</td>
<td>0.3091</td>
<td>0.3203</td>
<td>0.3558</td>
<td>0.2391</td>
</tr>
<tr>
<td>10</td>
<td>0.3321</td>
<td>0.3533</td>
<td>0.3845</td>
<td>0.2562</td>
</tr>
<tr>
<td>25</td>
<td>0.3879</td>
<td>0.4117</td>
<td>0.4291</td>
<td>0.3364</td>
</tr>
<tr>
<td>50</td>
<td>0.4225</td>
<td>0.4456</td>
<td>0.4701</td>
<td>0.4700</td>
</tr>
<tr>
<td>75</td>
<td>0.4572</td>
<td>0.5015</td>
<td>0.5260</td>
<td>0.5992</td>
</tr>
<tr>
<td>90</td>
<td>0.4937</td>
<td>0.5690</td>
<td>0.5756</td>
<td>0.6629</td>
</tr>
<tr>
<td>95</td>
<td>0.5159</td>
<td>0.5820</td>
<td>0.5942</td>
<td>0.6873</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.5641</td>
<td>0.6402</td>
<td>0.6545</td>
<td>0.7388</td>
</tr>
</tbody>
</table>

### Table 50 - Percentiles for \( \gamma_1 \) for experiment C', T-50

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>0.6587</td>
<td>0.6721</td>
<td>0.6738</td>
<td>0.6970</td>
</tr>
<tr>
<td>5</td>
<td>0.7388</td>
<td>0.7431</td>
<td>0.7435</td>
<td>0.7793</td>
</tr>
<tr>
<td>10</td>
<td>0.7678</td>
<td>0.7925</td>
<td>0.7946</td>
<td>0.8394</td>
</tr>
<tr>
<td>25</td>
<td>0.8843</td>
<td>0.9047</td>
<td>0.9060</td>
<td>0.9362</td>
</tr>
<tr>
<td>50</td>
<td>0.9546</td>
<td>0.9922</td>
<td>0.9949</td>
<td>1.0026</td>
</tr>
<tr>
<td>75</td>
<td>1.0294</td>
<td>1.0646</td>
<td>1.0722</td>
<td>1.1090</td>
</tr>
<tr>
<td>90</td>
<td>1.0875</td>
<td>1.1458</td>
<td>1.1413</td>
<td>1.2000</td>
</tr>
<tr>
<td>95</td>
<td>1.1821</td>
<td>1.2146</td>
<td>1.2229</td>
<td>1.2289</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.2605</td>
<td>1.3940</td>
<td>1.4244</td>
<td>1.3277</td>
</tr>
</tbody>
</table>
Table 51 - Percentiles for $\hat{\sigma}_e^2$ for experiment C', T-200

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.1058</td>
<td>0.1210</td>
<td>0.1738</td>
<td>0.1256</td>
</tr>
<tr>
<td>5</td>
<td>0.1703</td>
<td>0.2241</td>
<td>0.3142</td>
<td>0.2292</td>
</tr>
<tr>
<td>10</td>
<td>0.2157</td>
<td>0.2556</td>
<td>0.3358</td>
<td>0.3069</td>
</tr>
<tr>
<td>25</td>
<td>0.2486</td>
<td>0.3270</td>
<td>0.4085</td>
<td>0.3820</td>
</tr>
<tr>
<td>50</td>
<td>0.2951</td>
<td>0.4107</td>
<td>0.4898</td>
<td>0.4695</td>
</tr>
<tr>
<td>75</td>
<td>0.3764</td>
<td>0.4981</td>
<td>0.5565</td>
<td>0.5634</td>
</tr>
<tr>
<td>90</td>
<td>0.4398</td>
<td>0.5880</td>
<td>0.6320</td>
<td>0.6589</td>
</tr>
<tr>
<td>95</td>
<td>0.4610</td>
<td>0.6193</td>
<td>0.6908</td>
<td>0.7493</td>
</tr>
<tr>
<td>maximum</td>
<td>0.5277</td>
<td>0.7974</td>
<td>0.9902</td>
<td>0.9207</td>
</tr>
</tbody>
</table>

Table 52 - Percentiles for $\hat{\sigma}_\eta^2$ for experiment C', T-200

<table>
<thead>
<tr>
<th>percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2378</td>
<td>0.2147</td>
<td>0.2299</td>
<td>0.2547</td>
</tr>
<tr>
<td>5</td>
<td>0.3599</td>
<td>0.2997</td>
<td>0.2772</td>
<td>0.3551</td>
</tr>
<tr>
<td>10</td>
<td>0.4414</td>
<td>0.3796</td>
<td>0.3366</td>
<td>0.4224</td>
</tr>
<tr>
<td>25</td>
<td>0.5088</td>
<td>0.4538</td>
<td>0.4134</td>
<td>0.4978</td>
</tr>
<tr>
<td>50</td>
<td>0.6118</td>
<td>0.5760</td>
<td>0.4840</td>
<td>0.7217</td>
</tr>
<tr>
<td>75</td>
<td>0.6795</td>
<td>0.6940</td>
<td>0.5663</td>
<td>1.2456</td>
</tr>
<tr>
<td>90</td>
<td>0.8051</td>
<td>0.8342</td>
<td>0.6827</td>
<td>1.7964</td>
</tr>
<tr>
<td>95</td>
<td>0.8724</td>
<td>0.9013</td>
<td>0.7769</td>
<td>2.2773</td>
</tr>
<tr>
<td>maximum</td>
<td>1.0019</td>
<td>1.1373</td>
<td>0.8458</td>
<td>3.7368</td>
</tr>
</tbody>
</table>
Table 53 - Percentiles for \( \hat{\beta}_1 \) for experiment C', T-200

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.2578</td>
<td>0.1393</td>
<td>0.1376</td>
<td>0.1267</td>
</tr>
<tr>
<td>5</td>
<td>0.2822</td>
<td>0.1589</td>
<td>0.1551</td>
<td>0.1537</td>
</tr>
<tr>
<td>10</td>
<td>0.2866</td>
<td>0.1736</td>
<td>0.1710</td>
<td>0.1619</td>
</tr>
<tr>
<td>25</td>
<td>0.2960</td>
<td>0.1909</td>
<td>0.1878</td>
<td>0.1832</td>
</tr>
<tr>
<td>50</td>
<td>0.3059</td>
<td>0.2038</td>
<td>0.2026</td>
<td>0.2024</td>
</tr>
<tr>
<td>75</td>
<td>0.3200</td>
<td>0.2220</td>
<td>0.2189</td>
<td>0.2220</td>
</tr>
<tr>
<td>90</td>
<td>0.3302</td>
<td>0.2372</td>
<td>0.2329</td>
<td>0.2478</td>
</tr>
<tr>
<td>95</td>
<td>0.3355</td>
<td>0.2453</td>
<td>0.2432</td>
<td>0.2428</td>
</tr>
<tr>
<td>maximum</td>
<td>0.3455</td>
<td>0.2663</td>
<td>0.2584</td>
<td>0.2768</td>
</tr>
</tbody>
</table>

Table 54 - Percentiles for \( \hat{\delta}_1 \) for experiment C', T-200

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.3810</td>
<td>0.3962</td>
<td>0.4150</td>
<td>0.3832</td>
</tr>
<tr>
<td>5</td>
<td>0.3992</td>
<td>0.4164</td>
<td>0.4382</td>
<td>0.4147</td>
</tr>
<tr>
<td>10</td>
<td>0.4053</td>
<td>0.4347</td>
<td>0.4483</td>
<td>0.4326</td>
</tr>
<tr>
<td>25</td>
<td>0.4260</td>
<td>0.4596</td>
<td>0.4737</td>
<td>0.4750</td>
</tr>
<tr>
<td>50</td>
<td>0.4434</td>
<td>0.4831</td>
<td>0.4935</td>
<td>0.5075</td>
</tr>
<tr>
<td>75</td>
<td>0.4655</td>
<td>0.5024</td>
<td>0.5125</td>
<td>0.5356</td>
</tr>
<tr>
<td>90</td>
<td>0.4778</td>
<td>0.5190</td>
<td>0.5320</td>
<td>0.5483</td>
</tr>
<tr>
<td>95</td>
<td>0.4938</td>
<td>0.5396</td>
<td>0.5442</td>
<td>0.5580</td>
</tr>
<tr>
<td>maximum</td>
<td>0.5224</td>
<td>0.5648</td>
<td>0.5669</td>
<td>0.5716</td>
</tr>
</tbody>
</table>

Table 55 - Percentiles for \( \hat{\gamma}_1 \) for experiment C', T-200

<table>
<thead>
<tr>
<th>Percentages</th>
<th>GLS</th>
<th>IV/QML</th>
<th>IIV/ML</th>
<th>LIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum</td>
<td>0.8468</td>
<td>0.8576</td>
<td>0.8489</td>
<td>0.7723</td>
</tr>
<tr>
<td>5</td>
<td>0.8770</td>
<td>0.9153</td>
<td>0.9069</td>
<td>0.8255</td>
</tr>
<tr>
<td>10</td>
<td>0.9035</td>
<td>0.9234</td>
<td>0.9230</td>
<td>0.8554</td>
</tr>
<tr>
<td>25</td>
<td>0.9307</td>
<td>0.9706</td>
<td>0.9634</td>
<td>0.9479</td>
</tr>
<tr>
<td>50</td>
<td>0.9633</td>
<td>1.0054</td>
<td>1.0032</td>
<td>1.0117</td>
</tr>
<tr>
<td>75</td>
<td>1.0013</td>
<td>1.0470</td>
<td>1.0384</td>
<td>1.0782</td>
</tr>
<tr>
<td>90</td>
<td>1.0520</td>
<td>1.0965</td>
<td>1.0905</td>
<td>1.1338</td>
</tr>
<tr>
<td>95</td>
<td>1.0660</td>
<td>1.1119</td>
<td>1.1057</td>
<td>1.1737</td>
</tr>
<tr>
<td>maximum</td>
<td>1.1375</td>
<td>1.1984</td>
<td>1.1952</td>
<td>1.1942</td>
</tr>
</tbody>
</table>
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


Durbin, J (1988), Maximum Likelihood Estimation of the Parameters of
a System of Simultaneous Regression Equations, *Econometric Theory*, 4, 159-170


Fuller, W A (1976), *Introduction to Statistical Time Series*, New York: John Wiley and Sons


Hannan, E J and R D Terrell (1973), Multiple Equation Systems with
Stationary Errors, *Econometrica*, 41, 299-320


Amsterdam: North-Holland


Measurement, 2, 399-428


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


Theil, H (1961), Economic Forecasts and Policy, 2nd ed. Amsterdam: North Holland