ANALYSIS OF A CROSS-SECTION OF TIME SERIES USING STRUCTURAL TIME SERIES MODELS

by

Pablo Marshall Rivera

London School of Economics and Political Science

1990

Submitted to the University of London for the degree of Doctor of Philosophy.

ABSTRACT

This study deals with multivariate structural time series models, and in particular, with the analysis and modelling of cross-sections of time series. In this context, no cause and effect relationships are assumed between the time series, although they are subject to the same overall environment.

The main motivations in the analysis of cross-sections of time series are (i) the gains in efficiency in the estimation of the irregular, trend and seasonal components; and (ii) the analysis of models with common effects.

The study contains essentially two parts. The first one considers models with a general specification for the correlation of the irregular, trend and seasonal components across the time series. Four structural time series models are presented, and the estimation of the components of the time series, as well as the estimation of the parameters which define this components, is discussed.

The second part of the study deals with dynamic error components models where the irregular, trend and seasonal components are generated by common, as well as individual, effects. The extension to models for multivariate observations of cross-sections is also considered.

Several applications of the methods studied are presented. Particularly relevant is an econometric study of the demand for energy in the U. K.

CONTENTS

Introduction	10
1. Multivariate Structural Time Series Models	16
1.1 Introduction	16
1.2 Definitions and Basic Statistical Properties of the Models	19
1.3 The Kalman Filter	28
1.4 The Diffuse Kalman Filter	38
Appendix 1.1 Matrix Algebra Results for the Kalman Filter	42
2. The Likelihood Function	44
2.1 Introduction and General Results	44
2.2 Diagonalisation of a Symmetric r-Toeplitz Matrix	49
2.3 A Simple Expression for the Likelihood Function	54
2.4 Comparison with the Frequency Domain Approach	64
3. Maximum Likelihood Estimation of the Parameters and Testing	67
3.1 Introduction	67
3.2 Estimation of the Coefficients of the Exogenous Variables	70
3.3 Estimation of the Variance Covariance Matrices	78
3.4 Estimation Strategy	82
3.5 Tests of Hypotheses	85
4. The Demand for Energy in the U. K.: An Application	87
4.1 Introduction	87
4.2 The Econometric Model	90
4.3 Estimation of the Model and Results	97
4.4 Forecasting the Demand for Energy	112

4.5 Conclusions 119

Appendix 4.1 Definition of the Data 121

Appendix 4.2 Constrained Dependent Variables 122

5. Dynamic Error Components Models 124

5.1 Introduction and Formulation 124

5.2 Basic Properties of the Models 135

5.3 Extension to Multivariate Observations 142

5.4 Extension to Factor Analysis 145

6. Estimation of Dynamic Error Components Models Type I 149

6.1 Introduction 149

6.2 Estimation of the Unobserved Components 151

6.3 Maximum Likelihood Estimation of the Parameters 167

6.4 Analysis of Labour Cost Time Series: An Application 180

7. Estimation of Dynamic Error Components Models Type II 185

7.1 Introduction 185

7.2 Estimation of the Unobserved Components 188

7.3 Maximum Likelihood Estimation of the Parameters 207

7.4 Analysis of Labour Cost Time Series: An Application 213

8. Multivariate Dynamic Error Components Models 219

8.1 Introduction 219

8.2 Estimation of the Unobserved Components 224

8.3 Maximum Likelihood Estimation of the Parameters 227

8.4 The Demand for Energy in the U. K.: An Application 234

References 241

LIST OF TABLES

2.3.1: Coefficients of the Eigenvalues of the Powers of F to Form $\lambda_{c,t-p}$ 59

4.3.1: Estimates of Price Effects 101

4.3.2: Slopes of the Biases in Technical Progress 102

4.3.3: Error Structure of the Models 103

4.3.4: Test of Hypotheses 104

4.3.5: Demand Elasticities 107

4.4.1: Estimates in Total Quantity of Energy Models 116

4.4.2: Demand Forecasts for 1987 117

4.4.3: Absolute Value of Prediction Errors for 1987 118

5.4.1: Eigenvalues and Eigenvectors of Σ_ϵ and Σ_η 148

6.2.1: Parameter Values in the Evaluation of the Efficiency 160

6.2.2 to 6.2.11: Relative Efficiency of Estimators Using a Single Time Series Compared to Estimators Using n Times Series 162-166

6.4.1: Estimation Results for Univariate Models 183

6.4.2: Estimation Results for the Multivariate Model 184

7.2.1: Parameter Values in the Evaluation of the Efficiency 200

7.2.2 to 7.2.11: Relative Efficiency of Estimators Using a Single Time Series Compared to Estimators Using n Time Series 202-206

7.4.1: Ordinary Least Square Regressions 214

7.4.2: Estimation Results for Univariate Models 216

7.4.3: Estimation Results for the Multivariate Model 218

8.3.1: Number of Parameters in Variance Covariance Matrices in Model Type I 233

8.3.2: Number of Parameters in Variance Covariance Matrices in Model Type II 233

8.4.1: Estimates of Price Effects 237

8.4.2: Slopes of the Biases in Technical Progress 238

8.4.3: Error Structure of the Model 239

8.4.4: Demand Elasticities 240

LIST OF FIGURES

4.3.1: Biases in Technical Progress: Other Industry Sector 108

4.3.2: Biases in Technical Progress: Domestic Sector 109

4.3.3: Biases in Technical Progress: Other Final Users Sector 110

4.3.4: Biases in Technical Progress: Transport Sector 111

4.4.1: Price Index of Energy (1971 Q1 = 100) 115

6.4.1: Labour Cost Time Series (in logarithms) 182

7.4.1: Labour Cost Time Series (in logarithms) 215

To the memory of my father

ACKNOWLEDGMENTS

First of all I wish to espress my gratitude to my supervisor, Professor Andrew C. Harvey, for his constant support and guidance at all stages of this work.

I am grateful to the participants at the Econometrics Workshop and the members of the Statistics Department at the London School of Economics for their comments and suggestions.

I am also grateful to Guido del Pino, Osvaldo Ferreiro, German Rodriguez and Alvaro Vial who encouraged me to undertake postgraduate studies.

Financial contributions by the British Council, the Pontificia Universidad Catolica de Chile and the Instituto Nacional de Estadisticas de Chile is acknowledged. A teaching assistantship at the London School of Economics is also acknowledged.

Finally, I would like to thank my wife Carmen Gloria for a good deal of encouragement and endless patience during these years.

INTRODUCTION

It has long been recognised that many economic time series can be decomposed as a sum of trend, seasonal and irregular components,

$$(1) \qquad y_t = \mu_t + \gamma_t + \epsilon_t, \qquad t = 1,\ldots,T,$$

where y_t denotes, perhaps after some transformation, the t-th observation and μ_t, γ_t and ϵ_t are the trend, seasonal and irregular components.

Structural time series modelling deals with the specification and estimation of these components, and of the parameters which define them. The historical development of structural time series models and a review of the specifications usually considered in the literature can be found in Harvey and Durbin (1986) and Harvey (1990). In the basic structural model, the trend and seasonal components are defined by

$$(2) \qquad \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t,$$

$$(3) \qquad \beta_t = \beta_{t-1} + \delta_t,$$

$$(4) \qquad \gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \cdots - \gamma_{t-s+1} + \omega_t,$$

where μ_t is the level of the trend at time t, β_t is its slope, γ_t is the seasonal component at time t, s is the seasonal period, and η_t, δ_t, ω_t, and the irregular component ϵ_t in (1), are random shocks assumed to be mutually and serially uncorrelated, with expected values equal to zero and variances σ_η^2, σ_δ^2, σ_ω^2 and σ_ϵ^2 respectively. Thus, the trend and seasonal effects in the time series are assumed random variables changing over time. If the variances of the random shocks η_t, δ_t and ω_t are equal to zero, model (1) collapses to a deterministic linear trend with fixed dummy seasonal variables. Allowing those variances to be

greater than zero gives flexibility to the trend and seasonal components to evolve over time. Model (1)-(4) can be written in state space form and as a result the Kalman filter algorithm can be used to estimate the unobserved components. The variances of the random shocks, usually called hyperparameters, can be estimated by maximum likelihood.

Although the above specification may be extended to include cyclical components as in Harvey (1985a), or general ARMA structures as defined by Box and Jenkins (1976), there is substantial empirical evidence supporting specifications like the one presented above, or particular cases of it, for many economic time series; see Harvey and Todd (1983), Kitagawa and Gersh (1984), Harvey and Durbin (1986), Fernandez-Macho (1986), and Fernandez-Macho and Harvey (1989).

This study is concerned with the construction of structural time series models for cross-sections of time series. In this context, an n-dimensional multivariate version of model (1)-(4) can be naturally constructed. It is assumed that there is no cause and effect relationships between the n time series. However, as these time series are subject to the same overall environment, the multivariate white noise random shocks which drive the irregular, trend, and seasonal components: ϵ_t, η_t, δ_t and ω_t, have variance covariance matrices: Σ_ϵ, Σ_η, Σ_δ and Σ_ω respectively.

The two main motivations in the study of multivariate models for cross-sections of time series are (i) the gains in efficiency in the estimation of the trend and seasonal components, and (ii) the analysis of models with common trend and seasonal effects.

Although, as mentioned earlier, the especification (1)-(4) can be extended straightforwardly to a multivariate model, the estimation and the analysis of a model with more than four or five time series becomes extremely complicated. In that sense, an important objective in the

formulation and estimation of multivariate models is the specification and testing of hypotheses which reduce the dimensionality of the problem and simplify the analysis and the interpretation. Related to these ideas is the development of simple estimation procedures. That is also an important objective in the study of multivariate models.

Previous studies on multivariate structural time series models include Jones (1966), Enns et all (1982), Harvey (1985b), Fernandez-Macho (1986), Fernandez-Macho, Harvey and Stock (1987), Fernandez-Macho and Harvey (1989), Fernandez-Macho (1989), and Harvey (1990). Harvey (1990) and Fernandez-Macho (1986) are excellent reviews.

Jones (1966) considered a multivariate smoothing model which corresponds to the multivariate version of (1)-(4) but with no slope or seasonal terms. Enns et all (1982) studied the same model as Jones and assumed that the two variance covariance matrices in the model, Σ_ϵ and Σ_η, were proportional; what is called the homogeneity hypothesis. They proposed a method to estimate the parameters of the model by maximum likelihood and showed how estimates of the trends could be computed by the Kalman filter. Harvey (1985b) extended the multivariate exponential smoothing model to include slope and seasonal components as in (1)-(4), studied the estimation of the unobserved components by the Kalman filter, and formed the likelihood function by using the Kalman filter and the prediction error decomposition. Fernandez-Macho (1986) and Fernandez-Macho (1989) developed the frequency domain maximum likelihood estimation of the parameters in the multivariate version of (1)-(4), and in particular cases of it. Fernandez-Macho and Harvey (1989) developed a test for the homogeneity hypothesis used by Enns et all (1982). Finally, Fernandez-Macho, Harvey and Stock (1987) and also Fernandez-Macho (1986) studied dynamic factor analysis models.

The present study contains, basically, two parts. The first one,

which includes the first four chapters, is concerned with the multivariate structural time series models defined above. The main contributions here are (i) a more detailed study of the model with exogenous variables, as it is usually the case in analysing economic time series; (ii) a new expression for the likelihood function; and (iii) an econometric application where the unobserved trend components have an economic interpretation. The second part of the study includes chapters 5 to 8 and is concerned with the formulation and the estimation of dynamic error components models. Although these kind of models have received significant attention in the econometric literature, the approach developed here is new.

The following lines present a detailed account of the content of this study, with special reference to the new material.

Chapter 1 defines a multivariate regression model where the residuals follow a structural time series models. Four time series models, and basic statistical properties of them, are presented in Section 1.2. Section 1.3 presents the standard Kalman filter and Section 1.4 the diffuse Kalman filter. These two filters are alternative algorithms to obtain estimates of the unobserved components in the model. They differ in the way they deal with the initial estimates used to start the recursions of the filter. The main contributions of this chapter are the development of formulas to start the standard Kalman filter in Section 1.3, and the simple and direct derivation of the diffuse Kalman filter in Section 1.4.

In Chapter 2, Section 2.1 presents three asymptotically equivalent expressions for the likelihood function: the one obtained using the standard Kalman filter and the prediction error decomposition, the frequency domain likelihood, and the diffuse likelihood which is based on the diffuse Kalman filter. Sections 2.2 and 2.3 develop a new

alternative expression for the likelihood function which is simple and has some advantages over the more standard approaches. Section 2.4 compares the new expression with the frequency domain likelihood.

Chapter 3 deals with the maximum likelihood estimation of the parameters in the regression model defined in Chapter 1. Section 3.2 considers the estimation of the vector of exogenous variables, and Section 3.3 the estimation of the parameters in the variance covariance matrices of the random shocks. Section 3.4 analyses the estimation strategy and Section 3.5 presents results for the asymptotic behaviour of the estimators. Most of the results in this chapter have already been developed in the literature, although the results in Section 3.2 are considerably more general than the ones found in previous studies.

Chapter 4 presents an econometric study for the demand for energy in the U. K. for the period 1971-1986. The study illustrates the techniques and procedures of the first three chapters. Using a translog cost equation, Section 4.2 presents an econometric model where the technical progress takes the factor augmenting form. Assuming that these factors follow stochastic trends, the reduced form of the model has the form defined in Chapter 1. Section 4.3 estimates the model separately for each of four economic sectors: other industry, domestic, other final users, and transport; which use mainly four fuels: gas, electricity, oil and coal. Section 4.4 obtains forecasts of the individual demands.

Chapter 5 formulates dynamic error components models. In Section 5.1 the standard specification of static error components models considered in the literature is extended to dynamic models. Two kind of models are defined depending on whether the time series share the same trend and seasonal components or not. In the terminology of Engle and Granger (1987), the distinction between the two models is based on the

concept of cointegration. Basic statistical properties of the models are presented in Section 5.2, while sections 5.3 and 5.4 generalise the specifications to multivariate observations and factor analysis models respectively. The contents of this chapter, as well as of the ones which follows constitute a completely new approach to dynamic error components models.

Chapter 6 studies the estimation of error components models type I where the time series are not assumed to share the same trend and seasonal components. Chapter 7 deals with the estimation of error components models type II, where the time series have the same trend and seasonal components. Finally, Chapter 8 presents results for the estimation of multivariate error components models.

Some final comments with respect to the presentation of the material follows. Equations, tables and figures are numbered according to the section. The chapter number is omitted except when referring to an equation in another chapter. Finally, the presentation uses extensively definitions and results in matrix algebra. A good reference is Magnus and Neudecker (1988, chs. 1, 2, and 3).

CHAPTER 1 : MULTIVARIATE STRUCTURAL TIME SERIES MODELS

1.1 Introduction

This chapter defines a multivariate regression model with stochastic trend and seasonal components, reviews basic statistical properties of multivariate structural time series models, and presents the Kalman filter algorithm for the estimation of the unobserved components. For most of the definitions and results stated, Fernandez-Macho (1986) and Harvey (1990) are good references.

The multivariate regression model with stochastic trend and seasonal components is defined as

$$(1.1a) \qquad y_t = B\, z_t + \alpha_t, \qquad t = 1,\ldots,T,$$

where y_t is an (n x 1) vector of observations, z_t is an (r x 1) vector of exogenous variables, B is an (n x r) matrix of fixed parameters which satisfy the restrictions

$$(1.1b) \qquad vec(B) = S\, \beta,$$

and α_t is a residual component which follows a structural time series model. In (1.1b), S is a known (nr x k) selection matrix and β is a (k x 1) vector representing the functionally independent parameters in B. Alternatively, the model can be written as

$$(1.2) \qquad y_t = X_t\, \beta + \alpha_t, \qquad t = 1,\ldots,T,$$

with X_t the (n x k) matrix $(z_t' \otimes I_n)S$, where $\otimes$ represents the Kronecker product and I_n is the identity matrix of order n. For the residuals α_t, the following four alternative specifications are considered : local level, local linear trend, seasonal local level, and basic structural

model. Although detailed definitions are presented in Section 1.2, all the mentioned models for the vector α_t can be written in the state space form

$$\alpha_t = (Z \otimes I_n)\ \theta_t + \epsilon_t, \qquad t = 1,\ldots,T, \tag{1.3a}$$

$$\theta_t = (T \otimes I_n)\ \theta_{t-1} + (R \otimes I_n)\ \kappa_t, \tag{1.3b}$$

where Z, T and R are time invariant and known matrices of dimension (1 x p), (p x p) and (p x u) respectively, θ_t is the state vector which contains the trend and seasonal components, and ϵ_t and κ_t are (n x 1) and (nu x 1) dimensional random shocks assumed to be serially and mutually uncorrelated, normally distributed, with expected values equal to zero and variance covariance matrices Σ_ϵ and Σ_κ respectively. The random shock κ_t has u subcomponents of dimension (n x 1) each, which correspond to the random shocks of the trend and seasonal effects. These u subcomponents are also assumed to be mutually uncorrelated, and then, the matrix Σ_κ is block diagonal. In the state space representation (1.3), (1.3a) is called the measurement equation and (1.3b) the transition equation.

Model (1.3) is said to be homogeneous if

$$\Sigma_\kappa = D_\kappa \otimes \Sigma_\epsilon, \tag{1.4}$$

where D_κ is a diagonal matrix of dimension u. Thus, the model is homogeneous if the variance covariance matrices of the irregular random shock ϵ_t, and the variance covariance matrices of the u subcomponents in the vector κ_t are proportional.

Apart from this introduction, the chapter contains three more sections and an appendix. Section 1.2 defines the four time series models over α_t and presents some basic statistical properties of the models. The standard Kalman filter formulas, which are needed to handle

the state space model (1.3), with some additional results concerning: (i) the initialisation of the algorithm, (ii) the Kalman filter for an homogeneous model, and (iii) the treatment of the vector of coefficients β, are presented in Section 1.3. Section 1.4 deals with the diffuse Kalman filter which extends the standard filter to handle state space models with very general diffuse or semi diffuse initial state vectors. Finally, Appendix 1.1 presents some basic matrix algebra results required in Section 1.3.

1.2 Definitions and Basic Statistical Properties of the Models

This section defines four structural time series models for the (n x 1) vector of residuals α_t defined in Section 1.1. The models are: local level, local linear trend, seasonal local level, and basic structural model. Basic statistical properties are also presented. Specific analyses for each model are considered in what follows.

Local Level Model

The local level model is defined as

$$\alpha_t = \mu_t + \epsilon_t, \qquad t = 1,\ldots,T, \tag{2.1a}$$

$$\mu_t = \mu_{t-1} + \eta_t, \tag{2.1b}$$

where μ_t is an (n x 1) vector representing the level or trend of α_t, and ϵ_t and η_t are (n x 1) random shocks assumed to be normally distributed, serially and mutually uncorrelated, with expected values equal to zero and variance covariance matrices Σ_ϵ and Σ_η respectively.

The model is in the state space form (1.3) with p = u = Z = T = R = 1, and $\kappa_t = \eta_t$. Model (2.1) is not stationary given the presence of a random walk trend; however, the stationary form of the model is obtained by taking first differences in (2.1a). That gives

$$\nu_t \equiv (1\text{-}L)\ \alpha_t = \eta_t + (1\text{-}L)\ \epsilon_t, \qquad t = 2,\ldots,T, \tag{2.2}$$

where L is the lag operator. From this expression, the autocovariance function of ν_t is given by

$$\Gamma(0) = \Sigma_\eta + 2\ \Sigma_\epsilon, \tag{2.3a}$$

(2.3b) $\Gamma(\pm 1) = - \Sigma_\epsilon,$

and

(2.3c) $\Gamma(\pm k) = 0,$ if $k \succ 1.$

It then follows that the differences of α_t can be written as a restricted MA(1) model, which is called the reduced form of the local level model. The autocovariance generating function evaluated at $e^{-i\lambda}$, which is equal to 2π times the spectral density evaluated at the frequency λ, is given by

(2.4) $G(e^{-i\lambda}) = \Sigma_\eta + (2 - 2\cos(\lambda))\,\Sigma_\epsilon, \qquad -\pi \leq \lambda \leq \pi.$

Thus, the spectral density is real and the model is strictly invertible if Σ_η is positive definite.

The local level model can be extended to include a fixed slope in the transition equation (2.1b). In that situation,

(2.5) $\mu_t = \mu_{t-1} + \beta + \eta_t,$

and, by repeated substitutions, the trend μ_t can be expressed as

(2.6) $\mu_t = \mu_t^* + t\,\beta,$

with μ_t^* satisfying (2.1b). Replacing (2.6) in (2.1a) shows that the model with a fixed slope β has the standard form (2.1) with t as an exogenous variable and β as its coefficient. Fixed seasonal components can also be included in the model as exogenous variables in the measurement equation (2.1a).

Local Linear Trend Model

The local linear trend model is defined as

$$\alpha_t = \mu_t + \epsilon_t, \qquad t = 1,\ldots,T, \tag{2.7a}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \tag{2.7b}$$

$$\beta_t = \beta_{t-1} + \delta_t, \tag{2.7c}$$

where the (n x 1) vector μ_t is the trend of α_t and the (n x 1) vector β_t is its slope. The (n x 1) dimensional random shocks ϵ_t, η_t and δ_t are assumed to be normally distributed, serially and mutually uncorrelated, with expected values equal to zero and variance covariance matrices Σ_ϵ, Σ_η and Σ_δ respectively.

The model is in the state space form (1.3), with p = u = 2, and with

$$Z = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{2.8a}$$

$$T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \tag{2.8b}$$

$$R = I_2, \tag{2.8c}$$

$$\kappa_t' = (\eta_t' , \delta_t'), \tag{2.8d}$$

and

$$\theta_t' = (\mu_t' , \beta_t'). \tag{2.8e}$$

That is, Z is a (1 x 2) matrix, T and R are (2 x 2) matrices, and κ_t and θ_t are (2n x 1) vectors. Notice that in the state vector θ_t, the first n elements correpond to the levels or trends μ_t while the last n elements correspond to the slopes β_t. That is the way the components are ordered in the state vector and the same principle will be used in the following models. The stationary form of the model is obtained by

taking second differences in (2.7a). That is,

$$(2.9) \qquad \nu_t \equiv (1-L)^2 \alpha_t = \delta_t + (1-L)\,\eta_t + (1-L)^2\,\epsilon_t, \qquad t = 3,\ldots,T.$$

From where, the autocovariance function of ν_t is given by

$$(2.10a) \qquad \Gamma(0) = \Sigma_\delta + 2\,\Sigma_\eta + 6\,\Sigma_\epsilon,$$

$$(2.10b) \qquad \Gamma(\pm 1) = -\,\Sigma_\eta - 4\,\Sigma_\epsilon,$$

$$(2.10c) \qquad \Gamma(\pm 2) = \Sigma_\epsilon,$$

and

$$(2.10d) \qquad \Gamma(\pm k) = 0, \qquad \text{if } k \succ 2.$$

It follows that the reduced form of the local linear trend model is a restricted MA(2). The autocovariance generating function evaluated at $e^{-i\lambda}$ is given by

$$(2.11) \qquad G(e^{-i\lambda}) = \Sigma_\delta + (2 - 2\cos(\lambda))\,\Sigma_\eta + (2 - 2\cos(\lambda))^2\,\Sigma_\epsilon,$$

where $-\pi \leq \lambda \leq \pi$. The local linear trend model is strictly invertible if the matrix Σ_δ is a positive definite matrix.

Notice that when Σ_δ is equal to zero and the component β_t at time zero is defined as a fixed parameter, the local linear trend model reduces to the local level model with a fixed slope in the trend μ_t; see equation (2.5).

Seasonal Local Level Model

The seasonal local level model, with seasonal period s, is defined as

(2.12a) $\alpha_t = \mu_t + \gamma_t + \epsilon_t, \qquad t = 1,\ldots,T,$

(2.12b) $\mu_t = \mu_{t-1} + \eta_t,$

(2.12c) $S(L)\ \gamma_t = \omega_t,$

where the (n x 1) vector μ_t is the level or trend of α_t, and the (n x 1) vector γ_t the seasonal component. The polynomial in the lag operator L, S(L), is defined as $S(L) = (1 + L +\ldots+ L^{s-1})$; and, the (n x 1) random shocks ϵ_t, η_t and ω_t are assumed to be normally distributed, serially and mutually uncorrelated, with expected values equal to zero and variance covariance matrices Σ_ϵ, Σ_η and Σ_ω respectively.

The model is transformed into the state space form (1.3) by defining auxiliary components $\gamma_{1t},\ldots,\gamma_{s-2,t}$, and by setting p = s, u = 2, and,

(2.13a) $Z = [\ 1,\ 1,\ 0,\ \ldots,\ 0\ ],$

(2.13b) $$T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & -1 & \cdots & -1 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

(2.13c) $$R = \begin{bmatrix} I_2 \\ 0 \end{bmatrix},$$

(2.13d) $\kappa_t' = (\ \eta_t'\ ,\ \omega_t'\ ),$

and

(2.13e) $\theta_t' = (\ \mu_t'\ ,\ \gamma_t'\ ,\ \gamma_{1t}'\ ,\ \cdots\ ,\ \gamma_{s-2t}'\ ).$

That is, Z is a (1 x s) matrix, T is an (s x s) matrix, R is an (s x 2) matrix, κ_t is a (2n x 1) vector, and the state vector θ_t has dimension (sn x 1). The auxiliary components $\gamma_{1t},\ldots,\gamma_{s-2,t}$ correspond to the seasonal effects at times (t-1),...,(t-s+2) respectively. The stationary form of the model is obtained by taking seasonal differences in (2.12a) and using the fact that (1-L) S(L) = $(1-L^s)$. That is,

$$(2.14) \qquad \nu_t \equiv (1-L^s)\,\alpha_t = S(L)\,\eta_t + (1-L)\,\omega_t + (1-L^s)\,\epsilon_t,$$

for t = s+1,...,T. The autocovariance function of ν_t is then given by

$$(2.15a) \qquad \Gamma(0) = s\,\Sigma_\eta + 2\,\Sigma_\omega + 2\,\Sigma_\epsilon,$$

$$(2.15b) \qquad \Gamma(\pm 1) = (s-1)\,\Sigma_\eta - \Sigma_\omega,$$

$$(2.15c) \qquad \Gamma(\pm 2) = (s-2)\,\Sigma_\eta,$$

$$(2.15d) \qquad \Gamma(\pm(s-1)) = \Sigma_\eta,$$

$$(2.15e) \qquad \Gamma(\pm s) = -\,\Sigma_\epsilon,$$

and

$$(2.15f) \qquad \Gamma(\pm k) = 0, \qquad \text{if } k \succ s.$$

It follows that the reduced form of the seasonal local level is a restricted MA(s) model. The autocovariance generating function evaluated at $e^{-i\lambda}$ is

$$(2.16) \qquad G(e^{-i\lambda}) = \begin{cases} s^2\,\Sigma_\eta, & \text{if } \lambda = 0, \\ (c_s/c_1)\,\Sigma_\eta + c_1\,\Sigma_\omega + c_s\,\Sigma_\epsilon, & \text{if } \lambda \neq 0, \end{cases}$$

where $c_r = (2 - 2\cos(\lambda r))$ and $-\pi \leq \lambda \leq \pi$. The model is strictly invertible if both Σ_η and Σ_ω are positive definite matrices.

As in the local level model, a fixed slope β can be added to the trend component μ_t in (2.12b), and that is equivalent to include the time as an exogenous variable with β as its coefficient. On the other hand, if Σ_ω is equal to zero and the seasonal components at time zero are defined as fixed parameters, the model reduces to the local level model with fixed seasonal dummies as exogenous variables.

Basic Structural Model

The basic structural model, with seasonal period s, is defined as

$$\alpha_t = \mu_t + \gamma_t + \epsilon_t, \qquad t = 1,\ldots,T, \tag{2.17a}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \tag{2.17b}$$

$$\beta_t = \beta_{t-1} + \delta_t, \tag{2.17c}$$

$$S(L)\ \gamma_t = \omega_t, \tag{2.17d}$$

where the (n x 1) vector μ_t is the trend of α_t, the (n x 1) vector β_t is its slope and the (n x 1) vector γ_t is the seasonal component. $S(L) = (1 + L + L^2 + \ldots + L^{s-1})$, with L the lag operator; and the n dimensional random shocks ϵ_t, η_t, δ_t and ω_t are assumed to be normally distributed, serially and mutually uncorrelated, with expected values equal to zero and variance covariance matrices Σ_ϵ, Σ_η, Σ_δ and Σ_ω respectively.

The model is written in the state space form by defining auxiliary components $\gamma_{1t},\ldots,\gamma_{s-2t}$ as in the seasonal local level model. In terms of the general form (1.3), p = (s+1), u = 3, and

$$Z = [1\ 0\ 1\ 0\ \ldots\ 0], \tag{2.18a}$$

$$(2.18b)\qquad T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & . & . & 0 \\ 0 & 1 & 0 & 0 & 0 & . & . & 0 \\ 0 & 0 & -1 & -1 & -1 & . & . & -1 \\ 0 & 0 & 1 & 0 & 0 & . & . & 0 \\ 0 & 0 & 0 & 1 & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & . & . & 1 & 0 \end{bmatrix},$$

$$(2.18c)\qquad R = \begin{bmatrix} I_3 \\ \\ 0 \end{bmatrix},$$

$$(2.18d)\qquad \kappa_t' = (\eta_t' , \delta_t' , \omega_t'),$$

and

$$(2.18e)\qquad \theta_t' = (\mu_t' , \beta_t' , \gamma_t' , \gamma_{1t}' , \cdots \gamma_{s-2t}').$$

That is, Z is a (1 x s+1) matrix, T is an (s+1 x s+1) matrix, R is an (s+1 x 3) matrix, κ_t is a (3n x 1) vector and the state vector θ_t has dimension (n(s+1) x 1). The stationary form of the basic structural model is obtained by taking seasonal and regular differences in (2.17a), and using also the relation (1-L) S(L) = (1-L^s). That is,

$$(2.19)\qquad \nu_t \equiv (1-L)(1-L^s)\,\alpha_t = S(L)\,\delta_t + (1-L^s)\,\eta_t + (1-L)^2\,\omega_t$$

$$+ (1-L)(1-L^s)\,\epsilon_t, \qquad t = s+2,\ldots,T.$$

From where the autocovariance function of ν_t is given by

$$(2.20a)\qquad \Gamma(0) = s\,\Sigma_\delta + 2\,\Sigma_\eta + 6\,\Sigma_\omega + 4\,\Sigma_\epsilon,$$

$$(2.20b)\qquad \Gamma(\pm1) = (s-1)\,\Sigma_\delta - 4\,\Sigma_\omega - 2\,\Sigma_\epsilon,$$

$$(2.20c)\qquad \Gamma(\pm2) = (s-2)\,\Sigma_\delta + \Sigma_\omega,$$

$$(2.20d)\qquad \Gamma(\pm3) = (s-3)\,\Sigma_\delta,$$

(2.20e) $\Gamma(\pm(s-1)) = \Sigma_\delta + \Sigma_\epsilon,$

(2.20f) $\Gamma(\pm s) = - \Sigma_\eta - 2\ \Sigma_\epsilon,$

(2.20g) $\Gamma(\pm(s+1)) = \Sigma_\epsilon,$

and

(2.20h) $\Gamma(\pm k) = 0,$ if $k \succ s+1.$

Hence, the differences of α_t follow a restricted MA(s+1) model. The autocovariance generating function evaluated at $e^{-i\lambda}$ becomes

$$(2.21)\quad G(e^{-i\lambda}) = \begin{cases} s^2\ \Sigma_\delta, & \text{if } \lambda = 0, \\ (c_s/c_1)\ \Sigma_\delta + c_s\ \Sigma_\eta + c_1^2\ \Sigma_\omega + c_1 c_s\ \Sigma_\epsilon, & \text{if } \lambda \neq 0, \end{cases}$$

where $c_r = (2 - 2\ \cos(\lambda r))$ and $-\pi \leq \lambda \leq \pi$. The model is strictly invertible if both Σ_δ and Σ_ω are positive definite matrices.

When Σ_δ is equal to zero, and the component β_t at time zero is defined as a fixed parameter, the basic structural model reduces to the seasonal local level with the time as an exogenous variable. If also Σ_ω is equal to zero and the seasonal components at time zero are defined as fixed parameters, the model reduces to the local level model with the time and seasonal dummies as exogenous variables.

1.3 The Kalman Filter

This section presents the standard Kalman filter formulas for the state space model (1.3). The filter is used to obtain estimates of the unobserved components θ_t and to form the likelihood of the model. The general formulas are presented assuming that the vectors $\alpha_1,\ldots,\alpha_T$ are observed and that the variance covariance matrices of the random shocks are known. This assumptions are of course unrealistic but the rationality is the following. To obtain estimates of the unknown parameters in the model, the likelihood function is maximised using a nonlinear optimisation procedure which requires the evaluation of this function. This evaluation is provided by the Kalman filter and, obviously, initial values for the parameters are needed. On the other hand, once the parameters of the model have been estimated, the Kalman filter presented in this section yields the estimates of the unobserved trend and seasonal components in the state vector θ_t, as well as their mean square errors.

The section presents first general formulas for the Kalman filter, and studies the construction of the initial quantities required in the filter recursions. Particular cases of interest are considered later: the Kalman filter for an homogeneous model and the treatment of the vector of exogenous variables β when the vectors α_t, $t = 1,\ldots,T$, are not observed.

To present the general Kalman filter recursions, the following definitions are required:

(3.1a) $$\Psi_t = \{\alpha_1, \ldots, \alpha_t\},$$

(3.1b) $$m_t = E(\theta_t / \Psi_t),$$

(3.1c) $\bar{P}_t = V(\theta_t / \Psi_{t-1})$,

(3.1d) $P_t = V(\theta_t / \Psi_t)$,

(3.1e) $v_t = \alpha_t - E(\alpha_t / \Psi_{t-1})$,

(3.1f) $F_t = V(v_t / \Psi_{t-1})$,

where E(x) and V(x) represent the expected value and variance covariance matrix of the random variable x. That is, Ψ_t is the information set up to time t, m_t is the estimator of the state vector at time t with information up to time t while v_t is the one step ahead prediction error of α_t. P_t, $\bar{P}_t$ and F_t are condicional variances. With these definitions, the Kalman filter formulas for the general model (1.3), are

(3.2a) $\bar{P}_t = (T \otimes I_n) P_{t-1} (T' \otimes I_n) + (R \otimes I_n) \Sigma_\kappa (R' \otimes I_n)$,

(3.2b) $F_t = (Z \otimes I_n) \bar{P}_t (Z' \otimes I_n) + \Sigma_\epsilon$,

(3.2c) $P_t = \bar{P}_t - \bar{P}_t (Z' \otimes I_n) F_t^{-1} (Z \otimes I_n) \bar{P}_t$,

(3.2d) $m_t = (T \otimes I_n) m_{t-1} + \bar{P}_t (Z' \otimes I_n) F_t^{-1} v_t$,

and

(3.2e) $v_t = \alpha_t - (Z\ T \otimes I_n) m_{t-1}$;

see for example Anderson and Moore (1979, sec. 3.1) or Harvey (1990, sec. 3.3). Under normality, the Kalman filter yields the minimum mean square error estimator of θ_t and α_t conditional on an information set. It also yields the mean square errors of these estimators. Without the normality assumption, the estimates obtained from the Kalman filter

equations minimise the mean square errors within the class of linear estimators. Formulas for smoothing and forecasting are found in Anderson and Moore (1979, ch. 7) and Harvey (1990, sec. 3.3).

Two important questions, with respect to the recursions (3.2) arise at this point. These are, the existence of a steady state Kalman filter and the specification of the initial values m_0 and P_0 needed to start the filter. Harvey (1990, sec. 3.3) showed that the models presented here always satisfy necessary conditions for a steady state Kalman filter. That is, if the above formulas are started with a positive semi definite matrix P_0, then

$$(3.3) \qquad \lim_{t\to\infty} \; P_t = P,$$

where P is unique, and independent of P_0. If also, the stationary form of the model is strictly invertible, the convergence is exponentially fast.

With respect to the initial values m_0 and P_0, as no prior information for these quantities is, in general, available, a diffuse prior for the state vector at time zero is defined. That is, the state vector at time zero is assumed to have a normal distribution with expected value zero and variance covariance matrix $(\tau\ I_{np})$, where τ is a large number. On the other hand, it may be the case that only some linear combinations of the components of the state vector are defined as diffuse, while for other linear combinations a proper prior at time zero can be defined. In that situation the state vector is said to have a semi diffuse prior.

The general solution to deal with the Kalman filter under a diffuse or semi diffuse prior specification is presented in the next section, and that involves some modifications to the standard Kalman filter equations presented above. An alternative solution, although not very

general, is presented in Harvey (1990, sec. 3.3) and that may be simpler in many practical situations. Harvey showed that, for the models considered here, a proper prior can be formed using the first p vector of observations, where p is the number of components in the state vector for each time series. The procedure is the following. From (1.3a), the first p vectors $\alpha_1,\ldots,\alpha_p$ can be written in terms of θ_p, by repeated substitution for θ_t from (1.3b). This yields

$$(3.4a)\quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_{p-1} \\ \alpha_p \end{bmatrix} = \left[\begin{bmatrix} Z\ T^{1-p} \\ Z\ T^{2-p} \\ \cdot \\ \cdot \\ \cdot \\ Z\ T^{-1} \\ Z \end{bmatrix} \otimes I_n\right] \theta_p - \left[\begin{bmatrix} Z\ T^{1-p}\ R \\ Z\ T^{2-p}\ R \\ \cdot \\ \cdot \\ \cdot \\ Z\ T^{-1}\ R \\ 0 \end{bmatrix} \otimes I_n\right] \kappa_p$$

$$- \left[\begin{bmatrix} Z\ T^{2-p}\ R \\ Z\ T^{3-p}\ R \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \otimes I_n\right] \kappa_{p-1} - \ldots\ldots - \left[\begin{bmatrix} Z\ T^{-1}\ R \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \otimes I_n\right] \kappa_2 + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_p \end{bmatrix},$$

which, with obvious notation, can be written as

$$(3.4b)\quad \alpha = (H \otimes I_n)\ \theta_p + (H_p \otimes I_n)\ \kappa_p + \ldots + (H_2 \otimes I_n)\ \kappa_2 + \epsilon$$

$$= (H \otimes I_n)\ \theta_p + e,$$

where α, θ_p and ϵ are vectors of dimension np, H is a (p x p) matrix, $H_p,\ldots,H_2$ are (p x u) matrices and e is a vector of dimension np and expected value zero. The generalised least squares estimator of θ_p is then given by

$$(3.5)\quad m_p = (H^{-1} \otimes I_n)\ \alpha,$$

and the variance covariance matrix of the estimation error by

$$(3.6)\quad P_p = (H^{-1} \otimes I_n)\ V(e)\ (H'^{-1} \otimes I_n),$$

where V(e) is the variance covariance matrix of the vector e; see also Duncan and Horn (1972). These quantities are then used to start the recursive formulas (3.2) from t = (p+1). The problem, however, is to find analytic expressions for H and V(e) in order to evaluate m_p and P_p. Harvey presented the solution for the local level model. In the following lines the solutions for the local linear trend, seasonal local level and basic structural models are also presented. The proofs, which require tedious matrix algebra, use equations (3.4) and the matrix algebra results presented in Appendix 1.1

In the local level model the value of p is the unity,

$$(3.7a) \qquad m_1 = \alpha_1,$$

and

$$(3.7b) \qquad P_1 = \Sigma_\epsilon.$$

In the local linear trend model the value of p is two,

$$(3.8a) \qquad m_2 = \begin{bmatrix} \alpha_2 \\ \alpha_2 - \alpha_1 \end{bmatrix},$$

and

$$(3.8b) \qquad P_2 = \begin{bmatrix} \Sigma_\epsilon & \Sigma_\epsilon \\ \Sigma_\epsilon & (\Sigma_\delta + \Sigma_\eta + 2\,\Sigma_\epsilon) \end{bmatrix}.$$

For the seasonal local level model the value of p is s, the seasonal period, and

$$(3.9a) \qquad m_s = \begin{bmatrix} \bar{\alpha}_s \\ \alpha_s - \bar{\alpha}_s \\ \cdot \\ \cdot \\ \alpha_2 - \bar{\alpha}_s \end{bmatrix},$$

where $\bar{\alpha}_s = (\alpha_1 + \ldots + \alpha_s)/s$. The matrix P_s is formed with (3.6) and

$$(3.9b) \qquad V(e) = (I_s \otimes \Sigma_\epsilon) + (A_1 \otimes \Sigma_\eta) + (A_2 \otimes \Sigma_\delta),$$

where the (s x s) matrices H^{-1} and A_1 are given by

$$(3.9c) \qquad H^{-1} = (1/s) \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 & 1 \\ -1 & -1 & -1 & \ldots & -1 & (s-1) \\ -1 & -1 & -1 & \ldots & (s-1) & 1 \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ -1 & (s-1) & -1 & \ldots & \cdot & \cdot \end{bmatrix},$$

and

$$(3.9d) \qquad A_1 = \begin{bmatrix} (s-1) & (s-2) & (s-3) & \ldots & 1 & 0 \\ (s-2) & (s-2) & (s-3) & \ldots & 1 & 0 \\ (s-3) & (s-3) & (s-3) & \ldots & 1 & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ 1 & 1 & 1 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{bmatrix},$$

while the (s x s) matrix A_2 has all its elements equal to zero with the exception of the element in the first column and first row which is the unity.

Finally, for the basic structural model with seasonal period s, the value of p is (s+1), and

$$(3.10a)\qquad m_{s+1} = \begin{bmatrix} \bar{\alpha}_s \\ \beta \\ \alpha_{s+1} - \bar{\alpha}_s \\ \alpha_s + \beta - \bar{\alpha}_s \\ \vdots \\ \alpha_3 + (s-2)\beta - \bar{\alpha}_s \end{bmatrix},$$

where,

$$(3.10b)\qquad \beta = (\alpha_{s+1} - \alpha_1) / s,$$

and

$$(3.10b)\qquad \bar{\alpha}_s = (1/s)\left[\alpha_{s+1} + (\alpha_s+\beta) + (\alpha_{s-1}+2\beta) + \ldots + (\alpha_2+(s-1)\beta)\right].$$

The matrix P_{s+1} is formed using (3.6) and

$$(3.10c)\qquad V(e) = (I_{s+1} \otimes \Sigma_\epsilon) + (A_1 \otimes \Sigma_\eta) + (A_2^2 \otimes \Sigma_\delta) + (A_3 \otimes \Sigma_\omega),$$

where the (s+1 x s+1) matrices H^{-1}, A_1, A_2, and A_3 are given by

$$(3.10d)\qquad H^{-1} = (1/2s)\begin{bmatrix} -(s-1) & 2 & 2 & \ldots & 2 & (s+1) \\ -2 & 0 & 0 & \ldots & 0 & 2 \\ (s-1) & -2 & -2 & \ldots & -2 & (s-1) \\ (s-3) & -2 & -2 & \ldots & 2(s-1) & -(s-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -(s-3) & -2 & 2(s-1) & \ldots & -2 & (s-5) \end{bmatrix},$$

$$(3.10e)\qquad A_1 = \begin{bmatrix} s & s-1 & s-2 & \ldots & 1 & 0 \\ s-1 & s-1 & s-2 & \ldots & 1 & 0 \\ s-2 & s-2 & s-2 & \ldots & 1 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{bmatrix},$$

$$(3.10f) \qquad A_2 = \begin{bmatrix} s & s-1 & s-2 & . & . & 2 & 1 & 0 \\ s-1 & s-2 & s-3 & . & . & 1 & 0 & 0 \\ s-2 & s-3 & s-4 & . & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & 0 & 0 & . & . & 0 & 0 & 0 \\ 0 & 0 & 0 & . & . & 0 & 0 & 0 \end{bmatrix},$$

and

$$(3.10g) \qquad A_3 = \begin{bmatrix} 2 & -1 & 0 & . & . & . & . & . & . & 0 \\ -1 & 1 & 0 & . & . & . & . & . & . & 0 \\ 0 & 0 & 0 & . & . & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & . & . & . & 0 \end{bmatrix}.$$

Homogeneous Models

It can be shown using an induction principle that under the homogeneity restriction (1.4),

$$(3.11a) \qquad \bar{P}_t = \bar{Q}_t \otimes \Sigma_\epsilon, \qquad t = p+1,\ldots,T,$$

$$(3.11b) \qquad P_t = Q_t \otimes \Sigma_\epsilon, \qquad t = p,\ldots,T,$$

and

$$(3.11c) \qquad F_t = f_t \Sigma_\epsilon, \qquad t = p+1,\ldots,T,$$

where $\bar{Q}_t$, Q_t and f_t have dimensions (p x p), (p x p), and (1 x 1) respectively; and they are evaluated, from t = (p+1), according with the recursions

$$(3.12a) \qquad \bar{Q}_t = T\, Q_{t-1}\, T' + R\, D_K\, R',$$

$$(3.12c) \qquad Q_t = \bar{Q}_t - \bar{Q}_t\, Z'\, f_t^{-1}\, Z\, \bar{Q}_t,$$

and

$$(3.12c) \qquad f_t = Z\,\bar{Q}_t\,Z' + 1,$$

which are exactly equivalent to the Kalman filter recursions of a univariate model with the variance of the irregular random shock ϵ_t equal to the unity and the variance covariance matrix of κ_t equal to D_κ. The estimates of the state vector θ_t, $t = p+1,\ldots,T$, are then obtained from

$$(3.13) \qquad m_t = (T \otimes I_n)\, m_{t-1} + (\bar{Q}_t\, Z'\, \bar{f}_t^{-1} \otimes I_n)\, v_t.$$

The vector m_t in (3.13) can also be computed running the Kalman filter equations for each time series in turn, with the variance of ϵ_t equal to the unity and the variance covariance matrix of κ_t equal to D_κ. Thus, the Kalman filter recursions for an homogeneous model can be computed separately for each time series as if the model were univariate. The full matrices $\bar{P}_t$, P_t and F_t are then formed using (3.11).

Exogenous Variables

To run the Kalman filter using equations (3.2), it was assumed that the vector of exogenous variable coefficients β was known. An important result proved by Kohn and Ansley (1985) for the univariate case, n = 1, is that the filter can be run conditional on β. That is, if the vector of coefficients β is unknown, the prediction errors and the estimators of the state vectors, can be expressed as explicit functions of β. The idea can be extended to the multivariate case, $n \geq 1$, in a straightforward form. Define the (np x 1) vector m_{yt} and the (np x k) matrix M_{xt} by

(3.14a) $$m_{yt} = (T \otimes I_n)\, m_{y,t-1} + K_t\,[y_t - (Z\ T \otimes I_n)\, m_{y,t-1}],$$

and

(3.14b) $$M_{xt} = (T \otimes I_n)\, M_{x,t-1} + K_t\,[X_t - (Z\ T \otimes I_n)\, M_{x,t-1}],$$

for $t = p+1,\ldots,T$; where K_t is the gain matrix defined as $K_t = [\bar{P}_t\ (Z' \otimes I_n)\ F_t^{-1}]$ Then, from (3.2d) and (3.2e), the estimator of the state vector at time t, m_t, can be written as

(3.15) $$m_t = m_{yt} - M_{xt}\,\beta, \qquad t = p+1,\ldots,T,$$

provide (3.15) holds for $t = p$; and that is immediat from (3.5). Using (3.2e), the prediction error v_t can be written as

(3.16) $$v_t = v_{yt} - V_{xt}\,\beta, \qquad t = p+1,\ldots,T,$$

where v_{yt} is an (n x 1) vector of pseudo innovations after running the Kalman filter over the time series y_t, and V_{xt} is an (n x k) matrix of pseudo innovations after running the filter over each column of X_t. Notice that the recursions for the estimators of the state vectors and the prediction errors are the only quantities of interest in this situation because the variance covariance matrix of the state vector and the variance covariance matrix of the prediction error do not depend on the observations and hence on β.

1.4 The Diffuse Kalman Filter

This section presents alternative and more general Kalman filter formulas for the situation in which the state vector in a state space model is defined as diffuse or semi diffuse. The results were developed by De Jong (1988, 1989) although the presentation here is somehow different and more direct. The main advantage of the diffuse Kalman filter, over the standard formulas presented in the previous section, is that the problem of the initial conditions needed to start the recursions is solved by considering an extended filter. The disadvantage is that the recursions are slightly more complicated.

Consider the state space model (1.3) with initial state vector given by

$$(4.1) \qquad \theta_0 = a + A\,\xi,$$

where a is an (np x 1) normal random variable with expected value zero and variance covariance matrix Σ_a, A is an (np x r) matrix of known fixed values, and ξ is an (r x 1) normal random variable with expected value zero and variance covariance matrix Σ_ξ. Even more, assume that a and ξ are uncorrelated between them and with ϵ_t and κ_t defined in (1.3). The vector ξ is said to be diffuse if $\Sigma\xi^1$ converges to zero in the Euclidean norm. In general, (4.1) specifies a semi diffuse state vector at time zero; but if a = 0, and A = I_{np}, θ_0 is said to be diffuse and the results presented here coincide with the ones in the previous section.

Consider the idea of running the Kalman filter conditional on ξ. Define m_t^ξ, $\bar{P}_t^\xi$, P_t^ξ, v_t^ξ and F_t^ξ exactly as in (3.1) but conditional on ξ. Then, the matrices $\bar{P}_t^\xi$, P_t^ξ and F_t^ξ are obtained, for t = 1,...,T, using the recursions (3.2a), (3.2b) and (3.2c) with starting value

(4.2) $\quad P_0^c = \Sigma_a.$

Using an algebraic manipulation similar to the one applied to the model with exogenous variables, it can be shown that m_t^ξ and v_t^ξ can be expressed as

(4.3a) $\quad m_t^\xi = m_{1t} - M_{2t}\,\xi, \qquad t = 1,\ldots,T,$

and

(4.3b) $\quad v_t^\xi = v_{1t} - V_{2t}\,\xi, \qquad t = 1,\ldots,T,$

where,

(4.4a) $\quad m_{1t} = (T \otimes I_n)\, m_{1,t-1} + K_t^\xi\, v_{1t},$

(4.4b) $\quad M_{2t} = (T \otimes I_n)\, M_{2,t-1} + K_t^\xi\, V_{2t},$

(4.4c) $\quad v_{1t} = \alpha_t - (Z\,T \otimes I_n)\, m_{1,t-1},$

(4.4d) $\quad V_{2t} = (Z\,T \otimes I_n)\, M_{2,t-1},$

$m_{10} = 0$, $M_{20} = -A$, and K_t^ξ is the gain matrix defined as $K_t^\xi = [\bar{P}_t^\xi\,(Z' \otimes I_n)\,(F_t^\xi)^{-1}]$; see also Rosenberg (1973). Thus, m_{1t} and v_{1t} are obtained after running the standard Kalman filter over the time series α_t with $\xi = 0$; while M_{2t} and V_{2t} are obtained after running the same filter over each column of an (n x r) matrix of zeroes with the initial estimate for the state vector given by the columns of -A.

At this stage, two general results concerning random variables are required. If ξ and z are random variables, then

(4.5a) $\quad E(z) = \underset{\xi}{E}\ E(z \,/\, \xi),$

and

(4.5b) $$V(z) = \underset{\xi}{E}\ V(z \mid \xi) + \underset{\xi}{V}\ E(z \mid \xi).$$

From (4.5) and (4.3) follows that the estimators of the state vector unconditional on ξ, the prediction errors unconditional on ξ, and their conditional variance covariance matrices are given by

(4.6a) $$E(\theta_t \mid \psi_t) = m_{1t} - M_{2t}\, E(\xi \mid \psi_t),$$

(4.6b) $$\alpha_t - E(\alpha_t \mid \psi_{t-1}) = v_{1t} - V_{2t}\, E(\xi \mid \psi_{t-1}),$$

(4.6c) $$V(\theta_t \mid \psi_t) = P_t^c + M_{2t}\, V(\xi \mid \psi_t)\, M_{2t}',$$

and

(4.6d) $$V(\alpha_t \mid \psi_{t-1}) = F_t^c + V_{2t}\, V(\xi \mid \psi_{t-1})\, V_{2t}',$$

for $t = 1,\ldots,T$. Thus, these quantities can be obtained from the formulas given by the conditional Kalman filter and with expressions for $E(\xi \mid \psi_t)$ and $V(\xi \mid \psi_t)$. To obtain the last two quantities notice first that V_{2t}, $t = 1,\ldots,T$, are non random and that all the sample information up to time t is contained in $v_{11},\ldots,v_{1t}$. From the Kalman filter and conditional on ξ,

(4.7) $$(v_{1s} + V_{2s}\,\xi) \sim N(0, F_s^c), \qquad s = 1,\ldots,t,$$

and then,

(4.8) $$v_{1s} \sim N(V_{2s}\,\xi, F_s^c), \qquad s = 1,\ldots,t.$$

Using (4.5), the joint distribution of $v = (v_{11}',\ldots,v_{1t}')'$ and ξ is

(4.9) $$\begin{bmatrix} v \\ \xi \end{bmatrix} \sim N\left[0, \begin{bmatrix} F + V\,\Sigma_\xi\,V' & V\,\Sigma_\xi \\ \Sigma_\xi\,V' & \Sigma_\xi \end{bmatrix}\right],$$

where F and V are (nt x nt) and (nt x r) matrices given by

$$F = \text{diag}[\, F_1^c, \ldots, F_t^c \,], \tag{4.10a}$$

and

$$V = [\, V_{21}', \ldots, V_{2t}' \,]'. \tag{4.10b}$$

Finally, from (4.9) and using standard matrix algebra results,

$$E(\xi \;/\; \psi_t) = [\, \Sigma_\xi^{-1} + V' \; F^{-1} \; V \,]^{-1} \; [\, V' \; F^{-1} \; v \,] \tag{4.11a}$$

$$= [\, \Sigma_\xi^{-1} + \sum_s V_{2s}' \; (F_s^c)^{-1} \; V_{2s} \,]^{-1} \; [\, \sum_s V_{2s}' \; (F_s^c)^{-1} \; v_{1s} \,],$$

and

$$V(\xi - E(\xi \;/\; \psi_t)) = [\, \Sigma_\xi^{-1} + V' \; F^{-1} \; V \,]^{-1} \tag{4.11b}$$

$$= [\, \Sigma_\xi^{-1} + \sum_s V_{2s}' \; (F_s^c)^{-1} \; V_{2s} \,]^{-1}.$$

The same result is obtained by minimising the mean square error; see Theil (1970, sec. 7.8). If ξ is diffuse, Σ_ξ^{-1} converges to zero in (4.11). Replacing (4.11) into (4.6) gives all the quantities of interest in the Kalman filter. The existence of the estimator presented in (4.11) depends on whether $(V' \; F^{-1} \; V)$ is non singular. Finally, notice that when the value of t in (4.11) is such that ξ has a proper distribution, further calculations of the state estimates can be made by the standard Kalman filter.

Appendix 1.1 : Matrix Algebra Results for the Kalman Filter

To obtain the initial quantities for the Kalman filter recursions using equations (3.4), the inverse and the powers of the matrix T are required. Consider the basic structural model which is the most general and from where can be obtained, with obvious substitutions, the results for the other three models.

Let, for convenience of the presentation, (k+2) the dimension of the square matrix T defined in (2.18b). That is,

$$(A1.1) \qquad T = \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix},$$

where T_1 has dimension (2 x 2) and T_2 has dimension (k x k). The matrix T_1 is given by

$$(A1.2) \qquad T_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

and it is not difficult to verify that the powers of T_1 are given by,

$$(A1.3) \qquad T_1^m = \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix}, \qquad \text{for all m.}$$

On the other hand, if e_i is the i-th row of I_k, the identity of order k, the matrix T_2 can be written as

$$(A1.4) \qquad T_2 = \begin{bmatrix} -(e_1 + e_2 + \ldots + e_k) \\ e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_{k-1} \end{bmatrix}.$$

The inverse of T_2 can be obtained by applying elementary operations over the rows of T_2 and I_k. That yields,

$$(A1.5) \qquad T_2^{-1} = \begin{bmatrix} & e_2 \\ & e_3 \\ & . \\ & . \\ & . \\ & e_k \\ -(e_1 + e_2 + \ldots + e_k) & \end{bmatrix},$$

while the powers of this matrix, which are easily obtained by direct multiplication, are given by

$$(A1.6) \qquad T_2^{-i} = \begin{bmatrix} & e_{i+1} \\ & . \\ & . \\ & e_k \\ -(e_1 + e_2 + \ldots + e_k) & \\ & e_1 \\ & . \\ & . \\ & e_{i-1} \end{bmatrix}, \qquad \text{for } 1 < i < k,$$

and $T_2^{-k} = T_2$, $T_2^{-k-1} = I_k$. Hence, the matrix H defined in (3.4) has the form,

$$(A1.7) \qquad H = \begin{bmatrix} 1 & -s & 1 & 0 & . & . & . & . & 0 \\ 1 & 1-s & -1 & -1 & . & . & . & . & -1 \\ . & . & & & & & & & 1 \\ . & . & & & & & . & & \\ . & . & & & . & & & & \\ 1 & -1 & & . & & & & & \\ 1 & 0 & 1 & & & & & & \end{bmatrix},$$

and the inverse of H is computed using the formula for a partioned matrix; see for example Magnus and Neudecker (1988, sec. 1.11). To obtain the variance of the residual e in (3.4), V(e), notice that the postmultiplication of the matrices in (3.4a), by the matrix R defined in (2.18c), is equivalent to the selection of the first three columns of matrices which are similar to H but with the final rows equal to zero.

CHAPTER 2 : THE LIKELIHOOD FUNCTION

2.1 Introduction and General Results

This chapter presents various expressions for the likelihood or, what is in practice equivalent, the log-likelihood function of the regression model (1.1.1) or (1.1.2), where the residuals α_t are assumed to satisfy one of the structural time series models presented in Section 1.2. The general state space representation for these models was given in (1.1.3). This introduction presents three expressions for the likelihood function which have been considered in the literature. Sections 2.2 to 2.4 develop a fourth expression which has some advantages and compare the results with the more standard approaches.

Consider first the situation where the state vector at time zero is defined as diffuse. That is, the (np x 1) vector θ_0 in (1.1.3) has a normal distribution with expected value zero and variance covariance matrix (τ I_{np}), where τ is a large number. Under this assumption the likelihood for the whole set of observations $y_1,\ldots,y_T$ is not defined. The likelihood is only defined for $y_{p+1},\ldots,y_T$ conditional on $y_1,\ldots,y_p$ as well as on X_t, $t = 1,\ldots,T$. The first p vectors of observations are used to form an initial estimate of the state vector; see Section 1.3. Two main approaches have been considered to forming the likelihood function. One of these obtains the likelihood by means of the prediction error decomposition and the standard Kalman filter presented in Section 1.3; see Harvey (1990, sec. 3.4). Alternatively, the likelihood function can be formed from the stationary form of the model presented in Section 1.2. The two approaches are necessarily equivalent since the differences needed to obtain the stationary form of the model can be seen as a transformation of the observations $y_{p+1},\ldots,y_T$, with

the Jacobian of the transformation equal to the unity.

The log-likelihood function of the model defined by (1.1.1) or (1.1.2) constructed by means of the Kalman filter is, apart from a constant, given by

$$(1.1) \qquad \ell = -\tfrac{1}{2} \sum_{t=p+1}^{T} \left[\log|F_t| + (v_{yt} - V_{xt}\,\beta)'\, F_t^{-1}\, (v_{yt} - V_{xt}\,\beta) \right],$$

where v_{yt} and V_{xt} are the (n x 1) vector and the (n x k) matrix defined in Section 1.3. That is, they are pseudo innovations after running the Kalman filter over the vector y_t and each column of X_t. The matrix F_t corresponds to the (n x n) prediction error variance covariance matrix defined in (1.3.1f). Expression (1.1) is, in general, a complicated function of the parameters in the variance covariance matrices of the random shocks in the model, and for the multivariate case, n ↘ 1, the non linear maximisation procedure needed to obtain the estimates becomes very time consuming. In fact, each evaluation of the log-likelihood (1.1) requires a run of the Kalman filter over y_t and each column of X_t.

Fernandez-Macho (1986, ch. 3) developed the frequency domain likelihood for multivariate structural time series models. This likelihood is obtained from the stationary form of the model and has a form which is easy to evaluate but, as it is based on asymptotic results, it is only an approximation for finite samples. The frequency domain log-likelihood is, apart from a constant, given by

$$(1.2) \qquad \ell = -\tfrac{1}{2} \sum_{t=p+1}^{T} \left[\log|G_t| + \operatorname{trace}(G_t^{-1}\, P_t) \right],$$

where G_t and P_t are (n x n) matrices representing the autocovariance generating function and 2π times the periodogram of the differences of

the residuals α_t, defined as ν_t in Section 1.2, at the frequency $2\pi(t-p-1)/(T-p)$. A general expression for the matrix G_t is given by

$$(1.3) \qquad G_t = g_{\epsilon t}\,\Sigma_\epsilon + g_{\eta t}\,\Sigma_\eta + g_{\delta t}\,\Sigma_\delta + g_{\omega t}\,\Sigma_\omega, \qquad t = p+1,\ldots,T,$$

where $g_{\epsilon t}$, $g_{\eta t}$, $g_{\delta t}$ and $g_{\omega t}$ are known scalars as defined in Section 1.2. From (1.1.2), the matrix P_t can be written as

$$(1.4) \qquad P_t = (w_{yt} - W_{xt}\,\beta)\,(w_{yt} - W_{xt}\,\beta)^*, \qquad t = 1,\ldots,T,$$

where w_{yt} and W_{xt} are $(2\pi)^{\frac{1}{2}}$ times the Fourier transform of the differences of y_t and X_t respectively, and the sign * represents the conjugate transpose. Alternatively, the matrix P_t can be obtained from (1.1.1). That gives

$$(1.5) \qquad P_t = P_{yyt} + B\,P_{zzt}\,B' - B\,P_{zyt} - P_{yzt}\,B',$$

where P_{yyt}, P_{zzt}, P_{zyt} and P_{yzt} are (n x n), (r x r), (r x n) and (n x r) matrices respectively which correspond to 2π times the own and cross periodograms of the differences of the vectors y_t and z_t. It can be shown that only the real part of P_t is needed to evaluate the log-likelihood (1.2), and that the sum in (1.2) can be redefined to run from t = (p+1) to [(T-p)/2] only; see Fernandez-Macho (1986, ch. 3) for details.

What makes the frequency domain approach attractive is the fact that G_t in (1.3) is an explicit function of the parameters in the variance covariance matrices of the random shocks, and that the Fourier transformation of the observations does not depend on the parameters of the model so it can be computed before the non linear optimisation procedure is carried out.

Consider now a more general situation where the state vector is defined as semi diffuse; see Section 1.4. Then, in general, an initial

estimate of the state vector θ_t can be formed with less than the full first p vectors of observations. Although (1.1) and (1.2) can still be evaluated, (1.1) is no longer the exact log-likelihood; now, both are approximations. The exact likelihood of the model in this situation was developed by De Jong (1988, 1989). Using the notation in Section 1.4, the log-likelihood of $y_1,\ldots,y_T$ can be obtained from the identity

$$(1.6) \qquad \ell(y_1,\ldots,y_T) = \ell(\xi) + \ell(y_1,\ldots,y_T \,/\, \xi) - \ell(\xi \,/\, y_1,\ldots,y_T),$$

which holds for all possible values of ξ; in particular it does for $\xi = 0$ which is the case of interest here. The first term on the right hand side of (1.6) is evaluated directly from the definition of ξ, the second term is given by the conditional Kalman filter defined in Section 1.4, and the third term is evaluated from (1.4.11). If ξ is diffuse, (1.6) is not defined but (1.6) plus $(\frac{1}{2} \log|\Sigma_\xi|)$ is, and apart from a constant, that expression takes the form

$$(1.7) \qquad \ell + \tfrac{1}{2} \log|\Sigma_\xi| = -\tfrac{1}{2} \log\Big|\sum_{t=1}^{T} v_{2t}' (F_t^\xi)^{-1} v_{2t}\Big| - \tfrac{1}{2} \sum_{t=1}^{T} \log|F_t^\xi|$$

$$+ \tfrac{1}{2} \Big[\sum_{t=1}^{T} v_{1t}' (F_t^\xi)^{-1} v_{2t}\Big] \Big[\sum_{t=1}^{T} v_{2t}' (F_t^\xi)^{-1} v_{2t}\Big]^{-1} \Big[\sum_{t=1}^{T} v_{2t}' (F_t^\xi)^{-1} v_{1t}\Big]$$

$$- \tfrac{1}{2} \Big[\sum_{t=1}^{T} v_{1t}' (F_t^\xi)^{-1} v_{1t}\Big],$$

which is called the diffuse log-likelihood; see De Jong (1988, 1989).

In the following sections of this chapter, an alternative expression for the likelihood of a structural time series model is presented. The proposed form is based on the stationary form of the model in which the state vector at time zero is diffuse. This new expression for the likelihood function is, in general, an asymptotic

approximation and it has some advantages over (1.1) and (1.2). It has a form which is easy to evaluate, with conditions which are apparently less restrictive than the ones in the frequency domain approach, and it reduces to the exact likelihood for the local level time series model. The remaining sections are organised in the following form. Section 2.2 considers the asymptotic and exact diagonalisation of a symmetric r-Toeplitz matrix. Section 2.3 deals with the application of this result to the formation of the likelihood for the four structural time series models considered in section 1.2. Finally, Section 2.4 compares the expression obtained with the frequency domain likelihood.

2.2 Diagonalisation of a Symmetric r-Toeplitz Matrix

This section defines a symmetric r-Toeplitz matrix. The results for the exact and asymptotic diagonalisation of it are presented in the following three lemmas.

Definition 2.2.1 : A symmetric matrix $A = [a_{ij}]$ such that

(i) $a_{ij} = a_{i+k,j+k}$, for all i, j, k,

and

(ii) $a_{i1} = 0$, for all $i > r$,

is called a symmetric r-Toeplitz matrix.

Lemma 2.2.1 : Apart from corner elements, any symmetric r-Toeplitz matrix A of dimension (T x T) can be expressed as a unique linear combination of $F^0, F^1, \ldots, F^{r-1}$; where $F = [f_{ij}]$, $f_{ij} = 1$ if $|i-j| = 1$ and $f_{ij} = 0$ otherwise.

Proof: Let $\gamma(A)$ be the (r x 1) vector formed with the first r rows of the first column of A. Clearly, $\gamma(A)$ has all the distinct elements of A and then, it defines the matrix A.

It is not difficult to see that without considering the elements in or on the triangles formed with the components (1,1), (1,k-1), (k-1,1) and (T,T), (T,T-k), (T-k,T), the matrix F^k is a symmetric (k+1)-Toeplitz matrix for all the values of $k > 1$. The result is clearly true for k = 2, and suppose it is also true for $k = m > 2$, then if $F^m = [f^m_{ij}]$,

$$(2.1)\qquad f_{ij}^{m+1} = \begin{cases} f_{i+1,j}^{m} + f_{i-1,j}^{m}, & 1 < i < T, \\ f_{i+1,j}^{m}, & i = 1, \\ f_{i-1,j}^{m}, & i = T. \end{cases}$$

Thus, apart from corner elements, F^{m+1} is a symmetric (m+2)-Toeplitz matrix. In the top left corner of F^{m+1}, the rows i = 2,...,m are obtained as the sum of the (i-1)-th and (i+1)-th rows of F^{m}; while the first row of F^{m+1} is equal to the second row of F^{m}. This implies that the triangle formed with the elements (1,1), (1,m-1) and (m-1,1) in F^{m}, which does not satisfy the definition of a Toeplitz matrix, is augmented to (1,1), (1,m) and (m,1) in the matrix F^{m+1}. The same analysis aplies to the bottom right corner.

Using (2.1) it is easy to obtain the matrices F^{m} for m = 1, 2,... For example, apart from corner elements and in terms of the vector γ defined above,

$$\gamma'(F^{0}) = 1,$$

$$\gamma'(F^{1}) = (0\ 1),$$

$$\gamma'(F^{2}) = (2\ 0\ 1),$$

and in general for k > 1 the vector $\gamma(F^{k})$ of dimension (k+1) can be formed recursively as follows. If $\gamma_i(F^{k})$ is the i-th element of the vector $\gamma(F^{k})$, then

$$(2.2a)\qquad \gamma_1(F^{k}) = 2\ \gamma_2(F^{k-1}),$$

$$(2.2b)\qquad \gamma_i(F^{k}) = \gamma_{i-1}(F^{k-1}) + \gamma_{i+1}(F^{k-1}), \qquad \text{for } i = 2,\ldots,(k-1),$$

$$(2.2c)\qquad \gamma_k(F^{k}) = 0,$$

and

(2.2d) $\gamma_{k+1}(F^k) = 1.$

Let the (r x r) matrix Γ be formed with the column vectors $\gamma(F^0),\ldots,\gamma(F^{r-1})$, and zeros to complete the dimension. Clearly Γ is an upper triangular matrix with all the elements in the diagonal equal to the unity. Given an arbitrary vector $\gamma(A)$ of dimension r, there is a unique (r x 1) vector x such that

(2.3) $\Gamma\, x = \gamma(A),$

which implies that $\gamma(A)$ can be formed as a linear combination of $\gamma(F^0),\ldots,\gamma(F^{r-1})$. The particular form of these matrices ensures that, apart from the mentioned triangles in the top left and bottom right corners, the matrix A can be expressed as the same linear combination of $F^0,\ldots,F^{r-1}$. The solution of the system of equations (2.3) provides the coefficients of this linear combination.

Lemma 2.2.1 implies that apart from top left and botton right elements, any symmetric r-Toeplitz matrix can be diagonalised by the eigenvectors of *F*. The corner elements, which does not satisfy the definition of a Toeplitz matrix, appear only because the matrix A has a finite dimension, and then, it is a natural idea to express an approximation of A as a linear combination of the powers of the matrix *F*. Using the same argument, the importance of these corner elements should dissipate as the dimension of the matrix A increases. Before presenting a formal result with respect to this question, the following lemma provides the eigenvalues and eigenvectors of the matrix *F*.

<u>Lemma 2.2.2 :</u> The eigenvalues λ_{Tt}, and eigenvectors h_{Tt}, t = 1,...,T, of the (T x T) matrix *F* defined in Lemma 2.2.1 are .

(i) $\lambda_{Tt} = 2\cos(\pi t/T+1),$

(ii) $h'_{Tt} = (2/T+1)^{\frac{1}{2}} [\sin(\pi t/T+1), \sin(2\pi t/T+1), \ldots, \sin(T\pi t/T+1)]$.

Proof : See Theorem 6.5.5 in Anderson (1971).

The only problem that remains with the diagonalisation of the symmetric r-Toeplitz matrix A is the presence of the corner elements. If A is a 2-Toeplitz matrix, the diagonalisation is exact because neither F^0 nor F have these distinct elements in the corners. However, when r is greater than 2, an exact diagonalisation of A using the eigenvectors defined in Lemma 2.2.2 is not possible; but, as the following lemma proves, when the dimension of A goes to infinity the diagonalisation result holds for any fixed value of r.

Lemma 2.2.3 : Let A be a (T x T) symmetric r-Toeplitz matrix, A^* the approximation of A obtained using Lemma 2.2.1, and $H_T = [h_{T1}, \ldots, h_{TT}]$, with $h_{T1}, \ldots, h_{TT}$ defined as in Lemma 2.2.2; then, if A and A^* are positive definite matrices,

(i) $\lim_{T\to\infty} (1/T) \text{ trace}(A\, A^{*-1}) = 1$,

(ii) $\lim_{T\to\infty} |A\, A^{*-1}|^{(1/T)} = 1$,

(iii) If α is a (T x 1) vector with bounded elements,

$$\lim_{T\to\infty} (1/T)\, \alpha' (A^{-1} - A^{*-1})\, \alpha = 0.$$

Proof : Notice first that $(A - A^*)$ has only $2(r-2)$ columns different from zero. Then, $\text{rank}(A - A^*) \leq 2(r-2)$. Now,

$$(1/T) \text{ trace}(A\, A^{*-1}) = (1/T) \text{ trace}(A\, A^{*-1} - I_T) + 1$$

$$= (1/T) \text{ trace}(A^{*-\frac{1}{2}} (A - A^*)\, A^{*-\frac{1}{2}}) + 1,$$

from where (i) follows because

$$\text{rank}(A^{*-\frac{1}{2}} (A - A^*) A^{*-\frac{1}{2}}) = \text{rank}(A - A^*).$$

Similarly,

$$\lim_{T\to\infty} (1/T) \text{ trace}(A^{-1} A^*) = 1,$$

and (ii) follows from the inequality

$$T \text{ trace}^{-1}(A^{-1} A^*) \leq |A A^{*-1}|^{(1/T)} \leq (1/T) \text{ trace}(A A^{*-1}).$$

Finally, (iii) follows because

$$\text{rank}(A^{-1} - A^{*-1}) = \text{rank}(A - A^*),$$

and then α' $(A^{-1} - A^{*-1})$ α is bounded.

Lemma 2.2.3 says that as the dimension of A and A^* goes to infinity, the harmonic, geometric and arithmetic mean of the eigenvalues of $(A\ A^{*-1})$ tends to the unity. In that sense, the matrix H_T diagonalises A asymptotically.

2.3 A Simple Expression for the Likelihood Function

This section considers the application of the results of Section 2.2 to the formation of the likelihood for the structural time series models defined in Section 1.2.

Define the (T-p x T-p) symmetric r-Toeplitz matrices A, B, C_s, D_s and E_s with r equal to 2, 3, s, (s+1) and (s+2) respectively, with the value p as defined for each time series model in Section 1.2, and with the γ vectors, formed with the elements in the first r rows of the first column, equal to

$$\gamma'(A) = (2 \quad -1), \tag{3.1a}$$

$$\gamma'(B) = (6 \quad -4 \quad 1), \tag{3.1b}$$

$$\gamma'(C_s) = (s \quad s-1 \quad s-2 \quad . \; . \; . \quad 1), \tag{3.1c}$$

$$\gamma'(D_s) = (2 \quad 0 \quad 0 \; . \; . \; . \; 0 \quad -1), \tag{3.1d}$$

and

$$\gamma'(E_s) = (4 \quad -2 \quad 0 \; . \; . \; . \; 0 \quad 1 \quad -2 \quad 1). \tag{3.1e}$$

Using the results of the previous section, A is exactly diagonalised by H_{T-p} defined in Lemma 2.2.3. The other four matrices are asymptotically diagonalised by H_{T-p} as T goes to infinity.

To form the likelihood of the model (1.1.1) or (1.1.2), it is assumed here a diffuse prior for the state vector θ_t in the state space representation (1.1.3). Under that situation, the likelihood can be obtained from the stationary form of the model which considers the differences of the original observations; see Section 2.1.

Let ν_{yt} denote the (n x 1) vector of differences of the

observations y_t, N_{xt} the (n x k) matrix of differences of X_t, and $\nu_t = (\nu_{yt} - N_{xt}\ \beta)$, for t = p+1,...,T. The log-likelihood of $\nu' = (\nu_{p+1},\ldots,\nu_T)$, apart from a constant, is equal to

$$(3.2) \qquad \ell = -\tfrac{1}{2}\log|\Omega| - \tfrac{1}{2}\,\nu'\Omega^{-1}\nu,$$

where Ω is the (n(T-p) x n(T-p)) variance covariance matrix of ν. Using the autocovariance structure of the vector ν presented in Section 1.2, the following lines show that for each of the four structural time series models considered in this chapter, the matrix Ω has, asymptotically, a simple form. This asymptotic expression for the matrix Ω is used later to simplify the general log-likelihood (3.2).

Local Level Model

For the local level model, the value of p is the unity and the variance covariance matrix of the vector of differences ν is given by

$$(3.3) \qquad \Omega = (I_{T-1} \otimes \Sigma_\eta) + (A \otimes \Sigma_\epsilon),$$

where A is the (T-1 x T-1) matrix defined in (3.1) and the (n x n) matrices Σ_η and Σ_ϵ are the variance covariance matrices of the level and irregular random shocks respectively. As A and I_{T-1} are diagonalised by H_{T-1} defined in Lemma 2.2.3, the matrix Ω can be written as

$$(3.4) \qquad \Omega = \sum_{t=2}^{T} (h_{T-1,t-1}\ h'_{T-1,t-1}) \otimes (\Sigma_\eta + \lambda_{a,t-1}\ \Sigma_\epsilon),$$

with $\lambda_{a,t-1}$ the (t-1)-th eigenvalue of A. Clearly,

$$(3.5) \qquad \lambda_{a,t-1} = 2 - \lambda_{T-1,t-1}, \qquad t = 2,\ldots,T,$$

where $\lambda_{T-1,t-1}$ was defined in Lemma 2.2.2.

Local Linear Trend Model

For the local linear trend model, the value of p is 2 and the variance covariance matrix of the vector of differences ν is given by

$$(3.6) \qquad \Omega = (I_{T-2} \otimes \Sigma_\delta) + (A \otimes \Sigma_\eta) + (B \otimes \Sigma_\epsilon),$$

where A and B are the (T-2 x T-2) matrices defined in (3.1); and Σ_δ, Σ_η and Σ_ϵ are (n x n) matrices which represent the variance covariance matrices of the slope, level and irregular random shocks respectively. The matrix A is diagonalised by H_{T-2} defined in Lemma 2.2.3; while for large T, B is also diagonalised by H_{T-2}, and hence, the matrix Ω is asymptotically equivalent to the matrix $\bar{\Omega}$ defined as

$$(3.7) \qquad \bar{\Omega} = \sum_{t=3}^{T} (h_{T-2,t-2}\, h_{T-2,t-2}') \otimes (\Sigma_\delta + \lambda_{a,t-2}\, \Sigma_\eta + \lambda_{b,t-2}\, \Sigma_\epsilon),$$

where $\lambda_{a,t-2}$ is the (t-2)-th eigenvalue of A and $\lambda_{b,t-2}$ is the (t-2)-th eigenvalue of a matrix which apart from the elements (1,1) and (T-2,T-2) is equal to B; and it is formed as a linear combination of F^0, F^1, and F^2 in accordance with the results in Lemma 2.2.1. By noticing that this matrix is equal to A^2, and using (3.5), it can be seen that

$$(3.8a) \qquad \lambda_{a,t-2} = 2 - \lambda_{T-2,t-2}, \qquad t = 3,\ldots,T,$$

and

$$(3.8b) \qquad \lambda_{b,t-2} = \lambda_{a,t-2}^2 = [2 - \lambda_{T-2,t-2}]^2, \qquad t = 3,\ldots,T,$$

where $\lambda_{T-2,t-2}$ was defined in Lemma 2.2.2.

Seasonal Local Level Model

For the seasonal local level model with seasonal period s, the value of p is equal to s and the variance covariance matrix of the vector of differences ν is given by

$$\Omega = (A \otimes \Sigma_\omega) + (C_s \otimes \Sigma_\eta) + (D_s \otimes \Sigma_\epsilon), \tag{3.9}$$

where A, C_s and D_s are the (T-s x T-s) matrices defined in (3.1) and the (n x n) matrices Σ_ω, Σ_η and Σ_ϵ are the variance covariance matrices of the seasonal, level and irregular random shocks respectively. A is exactly diagonalised by H_{T-s} defined in Lemma 2.2.3, while the matrices C_s and D_s are asymptotically diagonalised by H_{T-s}. Then, the matrix Ω is asymptotically equivalent to $\bar{\Omega}$ defined as

$$\bar{\Omega} = \sum_{t=s+1}^{T} [(h_{T-s,t-s}\ h_{T-s,t-s}^{\dagger}) \otimes (\lambda_{a,t-s}\ \Sigma_\omega + \lambda_{c,t-s}\ \Sigma_\eta + \lambda_{d,t-s}\ \Sigma_\epsilon)], \tag{3.10}$$

where $\lambda_{a,t-s}$ is defined, with obvious substitutions, as in (3.5) and (3.8a). The value $\lambda_{c,t-s}$ corresponds to the (t-s)-th eigenvalue of a matrix which, apart from corner elements, is equal to C_s, and it is formed as a linear combination of $F^0,\ldots,F^{s-1}$ in accordance with Lemma 2.2.1. The value $\lambda_{d,t-s}$ corresponds to the (t-s)-th eigenvalue of a matrix which, apart from corner elements, is equal to D_s, and it is formed as a linear combination of $F^0,\ldots,F^s$. It can be verified that for any value of the seasonal period s,

$$\lambda_{d,t-s} = \lambda_{a,t-s}\ \lambda_{c,t-s}, \qquad t = s+1,\ldots,T, \tag{3.11}$$

and although we do not have an analytic expression for $\lambda_{c,t-s}$ for the different values of s, these values can be computed for each value of s

in the following way. As $\lambda_{c,t-s}$ is the (t-s)-th eigenvalue of a matrix which is a linear combination of $F^0,\ldots,F^{s-1}$, it must be equal to the same linear combination of the (t-s)-th eigenvalues of the matrices $F^0,\ldots,F^{s-1}$. The coefficients of this linear combination are obtained by solving the system of equations (2.2). Table 2.3.1 gives the coefficients of the linear combinations of the eigenvalues of $F^0,\ldots,F^{s-1}$ to form $\lambda_{c,t-s}$ for values of s between 4 and 12. Thus, if the seasonal period is 4,

$$\lambda_{c,t-s} = 2\,(\lambda_{T-s,t-s})^2 + (\lambda_{T-s,t-s})^3, \qquad t = s+1,\ldots,T. \tag{3.12}$$

Table 2.3.1 : Coefficients of the Eigenvalues of the Power of F to form $\lambda_{c,t-p}$

	Seasonal Period								
Power of F	4	5	6	7	8	9	10	11	12
0		1	2	1		1	2	1	
1		-2	1	4		-4	1	6	
2	2	-1	-4	2	8	-2	-12	3	18
3	1	2	-2	-6	4	14	-6	-26	9
4		1	2	-3	-8	7	22	-13	-48
5			1	2	-4	-10	11	32	-24
6			1	2	-5	-12	16	44	
7				1	2	-6	-14	22	
8					1	2	-7	-16	
9						1	2	-8	
10							1	2	
11								1	

Basic Structural Model

For the basic structural model with seasonal period s, the value of p is equal to (s+1) and the variance covariance matrix of the vector of differences ν is given by

$$(3.13) \qquad \Omega = (B \otimes \Sigma_\omega) + (C_s \otimes \Sigma_\delta) + (D_s \otimes \Sigma_\eta) + (E_s \otimes \Sigma_\epsilon),$$

where B, C_s, D_s and E_s are the (T-s-1 x T-s-1) matrices defined in (3.1), and the (n x n) matrices Σ_ω, Σ_δ, Σ_η and Σ_ϵ are the variance covariance matrices of the seasonal, slope, level and irregular random shocks respectively. The matrices B, C_s, D_s, and E_s are asymptotically diagonalised by H_{T-s-1} defined in Lemma 2.2.3, and then, the matrix Ω is asymptotically equivalent to $\bar{\Omega}$ defined as

$$(3.14) \qquad \bar{\Omega} = \sum_{t=s+2}^{T} [(h_{T-s-1,t-s-1}\ h'_{T-s-1,t-s-1}) \otimes (\lambda_{b,t-s-1}\ \Sigma_\omega + \lambda_{c,t-s-1}\ \Sigma_\delta + \lambda_{d,t-s-1}\ \Sigma_\eta + \lambda_{e,t-s-1}\ \Sigma_\epsilon)],$$

where $\lambda_{b,t-s-1}$, $\lambda_{c,t-s-1}$ and $\lambda_{d,t-s-1}$ are defined, with obvious substitutions, as for the local linear trend and seasonal local level models. The value $\lambda_{e,t-s-1}$ corresponds to the (t-s-1)-th eigenvalue of a matrix which, apart from corner elements, is equal to E_s, and it is formed as a linear combination of $F^0,\ldots,F^{s+1}$ in accordance with Lemma 2.2.1. It can be shown that for any seasonal period s,

$$(3.15) \qquad \lambda_{e,t-s-1} = \lambda_{b,t-s-1}\ \lambda_{c,t-s-1}, \qquad t = s+2,\ldots,T.$$

A Simple Form for the Likelihood Function

From the analysis of the four models presented above, a general expression for $\bar{\Omega}$, which is taken to be equal to Ω for the local level

model, is given by

$$\bar{\Omega} = \sum_{t=p+1}^{T} (h_t h_t' \otimes R_t), \tag{3.16a}$$

with R_t the (n x n) matrix given by

$$R_t = r_{\epsilon t}\,\Sigma_\epsilon + r_{\eta t}\,\Sigma_\eta + r_{\delta t}\,\Sigma_\delta + r_{\omega t}\,\Sigma_\omega, \tag{3.16b}$$

where the known scalars r_{it}, $i = \epsilon, \eta, \delta$ and ω are chosen in accordance with the model and equations (3.5), (3.8), (3.11), (3.15) and Table 2.3.1. Replacing the value of Ω by $\bar{\Omega}$ in the log-likelihood (3.2), and using Lemma 2.2.3, yields the asymptotically equivalent expression

$$\ell = -\tfrac{1}{2} \log\left|\sum_t h_t h_t' \otimes R_t\right| - \tfrac{1}{2}\,\mathrm{trace}\left[\left(\sum_t h_t h_t' \otimes R_t\right)^{-1} \nu\nu'\right], \tag{3.17}$$

and using Lemma 2.1 in Magnus (1982), and the fact that the h_t are orthogonal vectors, this expression can be written as

$$\ell = -\tfrac{1}{2} \sum_t \left[\log|R_t| + \mathrm{trace}(R_t^{-1}\, Q_t)\right], \tag{3.18}$$

with R_t given by (3.16b) and the (n x n) matrix Q_t by

$$Q_t = (\nu_{p+1}, \ldots, \nu_T)\; h_t h_t'\; (\nu_{p+1}, \ldots, \nu_T)', \qquad t = p+1, \ldots, T. \tag{3.19}$$

Alternatively,

$$\ell = -\tfrac{1}{2} \sum_t \left[\log|R_t| + (\bar{\nu}_{yt} - \bar{N}_{xt}\,\beta)'\, R_t^{-1}\, (\bar{\nu}_{yt} - \bar{N}_{xt}\,\beta)\right], \tag{3.20}$$

where the (n x 1) vector $\bar{\nu}_{yt}$ is defined by

$$\bar{\nu}_{yt} = (\nu_{y,p+1}, \ldots, \nu_{yT})\; h_t, \qquad t = p+1, \ldots, T, \tag{3.21}$$

and each column of the (n x k) matrix $\bar{N}_{xt}$ is obtained by applying the transformation (3.21) to the corresponding column of N_{xt}. See below

equation (3.1e) for the definitions of v_t, v_{yt} and N_{xt}.

A number of observations, with respect to this expression for the log-likelihood, are in order. First, (3.18) is of essentially the same form as the frequency domain log-likelihood, with Q_t playing the role of 2π times the periodogram and R_t the role of the autocovariance generating function. However, unlike the frequency domain approach, the eigenvalues λ_{Tt} defined in Lemma 2.2.2 do not occur in pairs and so the sum in (3.18) cannot be reduced to run from zero to [(T-p)/2] as in the frequency domain approach.

Second, for the local level model, equation (3.18) corresponds to the exact log-likelihood under the assumption that the state vector at time zero, in the state space representation (1.1.3), is diffuse. The value of $r_{\eta t}$ is the unity and the value of $r_{\epsilon t}$, $(2 - \lambda_{T-1,t-1})$, is greater than zero for all t. Thus, the log-likelihood (3.18) is defined even if Σ_η has rank less than n provided Σ_ϵ has full rank. Essentially the same result is valid in the local linear trend model, where $r_{\delta t}$ is the unity, $r_{\eta t}$ is $(2 - \lambda_{T-2,t-2})$ and $r_{\epsilon t}$ is $r^2_{\eta t}$. For this model it is possible to have both Σ_η and Σ_δ less than full rank, and provided Σ_ϵ is full rank, the approximate log-likelihood (3.18) is perfectly well defined. These results are important for the estimation of common trends models; see Section 5.4.

Third, for the seasonal local level with seasonal period equal to 4, $r_{\omega t}$ is equal to $(2 - \lambda_{T-4,t-4})$, $r_{\eta t}$ is given by (3.12) and $r_{\epsilon t}$ is given by the product of $r_{\eta t}$ and $r_{\omega t}$. Clearly, $r_{\omega t}$ is never zero, and $r_{\eta t}$ and $r_{\epsilon t}$ are zero only when (T-4) is odd because the roots of $r_{\eta t} = 0$ are 0 and -1. Thus, if (T-4) is odd, Σ_ω should be positive definite; but if (T-4) is even, $r_{\eta t}$ and then $r_{\epsilon t}$ are never zero; and it is possible to have both Σ_ω and Σ_η less than full rank provided Σ_ϵ is strictly positive definite. Similar results apply to the basic

structural model.

Fourth, the matrix F defined in Lemma 2.2.1 has already been used in previous studies to form the likelihood of an MA(1) model; see Cooley and Prescott (1973, 1976), Balestra (1980), Pesaran and Slater (1980) and Enns et all (1982). The results of this and previous sections allow, with some obvious modifications, the formation of an asymptotically equivalent likelihood for a general MA(q) model.

Fifth, analytic first and second derivatives of the log-likelihood (3.18) can be obtained in exactly the same way as in the frequency domain approach; see Section 3.3.

2.4 Comparison with the Frequency Domain Approach

This section compares the restrictions imposed on the model to construct the likelihood (3.18) with the circularity restrictions imposed by the frequency domain approach. The two approaches add or substract some elements to the corners of the matrices A, B, C_s, D_s and E_s defined in (3.1). As these matrices are symmetric, only the top left and top right corners need to be analised.

For the local level model, (3.18) is the exact log-likelihood, while in the frequency domain approach the block

$$\begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}, \tag{4.1}$$

is added to the top right corner of the matrix A defined in (3.1). This restriction can be interpreted as

$$\epsilon_0 = \epsilon_T, \tag{4.2}$$

where ϵ_t is the irregular random shock.

For the local linear trend model, the log-likelihood (3.18) places no restrictions over A; while the block

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.3}$$

is substracted from the top left corner of the matrix B defined in (3.1). This restriction can be interpreted as

$$\epsilon_{-1} = \epsilon_T = 0, \tag{4.4}$$

with ϵ_t the irregular random shock. In the frequency domain approach, (4.1) is added to the top left corner of A, and the block

(4.5) $$\begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix},$$

to the same corner of B. The frequency domain restrictions can be interpreted as

(4.6a) $$\epsilon_{-1} = \epsilon_{T-1},$$

(4.6b) $$\epsilon_0 = \epsilon_T,$$

and

(4.6c) $$\eta_0 = \eta_T,$$

where ϵ_t and η_t are the irregular and level random shocks. Comparing the restrictions (4.4) with (4.6), it can be seen that if Σ_ϵ is zero, the log-likelihood (3.18) corresponds to the exact one, while the frequency domain log-likelihood still requires the condition (4.6c).

In the seasonal local level model with seasonal period 4, the log-likelihood (3.18) substract the blocks

(4.7) $$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix},$$

from the top left corner of C_4 and D_4 defined in (3.1). The frequency domain approach adds (4.1) and the blocks

(4.8) $$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

to the top right corner of A, C_4 and D_4 respectively.

Finally, for the basic structural model with seasonal period 4, the log-likelihood (3.18) substracts (4.3), (4.7) and

$$(4.9) \qquad \begin{bmatrix} 0 & 1 & -2 & 1 \\ 1 & -2 & 1 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

from the top left corner of B, C_4, D_4 and E_4 respectively. The frequency domain approach adds (4.1), (4.8) and

$$(4.10) \qquad \begin{bmatrix} 1 & -2 & 1 & 0 & -2 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

to the top right corner of A, C_4, D_4 and E_4 respectively.

Apart from the local level model, where the log-likelihood proposed in the previous section is exact, it can be seen from the analysis above that in general, the restrictions imposed to form the log-likelihood (3.18) seems to be less restrictive than the ones imposed by the frequency domain approach, although no conclusive evidence has been presented in order to establish that the log-likelihood (3.18) is a better approximation.

CHAPTER 3 : MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS AND TESTING

3.1 Introduction

This chapter considers the maximum likelihood estimation, and associated asymptotic tests of hypotheses, of the parameters in the regression model presented in Section 1.1. The parameters of the model are the vector of coefficients of the exogenous variables, β, and the variance covariance matrices of the random shocks in the structural time series model for the residuals α_t. These are Σ_ϵ, Σ_η, Σ_δ and Σ_ω.

It is assumed in this chapter, unless otherwise stated, that the appropriate model for the residuals is the basic structural model. The results for the other three time series models presented in Section 1.2 are analogous and can be obtained with obvious substitutions. It is also assumed here that the log-likelihood of the model is defined for the observations $y_{p+1},\ldots,y_T$, conditional on $y_1,\ldots,y_p$, as well as on X_t, $t = 1,\ldots,T$; where p is the number of components of the state vector θ_t in (1.1.3) for each time series. That is equivalent to assuming a diffuse prior for the initial state vector. If the state vector is defined as semi diffuse, the exact likelihood is based on the diffuse Kalman filter defined in Section 1.4. That leads to the expression (2.1.7) which is substantially more complicated to handle. The situation of a semi diffuse prior will not be considered although, as stated in Section 2.1, the likelihood conditional on $y_1,\ldots,y_p$ is asymptotically equivalent to the exact likelihood function. Finally, it is assumed that the stationary form of the model is strictly invertible; see Section 2.1. For the basic structural model this implies that both Σ_δ and Σ_ω are strictly positive definite matrices.

Three asymptotically equivalent expressions for the log-likelihood

function of the model (1.1.1) or (1.1.2) were defined in Chapter 2. A general form for this function, which includes the one obtained by means of the Kalman filter as well as the ones formed using the frequency domain approach or the alternative transformation developed in Section 2.3, is given by

$$(1.1) \qquad \ell = -\tfrac{1}{2} \sum_{t=p+1}^{T} [\log|G_t| + (w_{yt} - W_{xt}\,\beta)'\, G_t^{-1}\, (w_{yt} - W_{xt}\,\beta)],$$

or alternatively, by

$$(1.2) \qquad \ell = -\tfrac{1}{2} \sum_{t=p+1}^{T} [\log|G_t| + \operatorname{trace}(G_t^{-1}\, P_t)],$$

where $P_t = (w_{yt} - W_{xt}\,\beta)\,(w_{yt} - W_{xt}\,\beta)'$ and G_t are (n x n) matrices, and w_{yt} and W_{xt} have dimensions (n x 1) and (n x k) respectively. Although (1.1) and (1.2) use the notation given in Section 2.1 to the frequency domain log-likelihood, they represent the log-likelihood formed by means of the Kalman filter if G_t is the prediction error variance defined in (1.3.1f) and w_{yt} and W_{xt} are pseudo innovations after running the Kalman filter over y_t and each column of X_t respectively; see Section 1.3. Expressions (1.1) and (1.2) also represent the alternative log-likelihood defined in Section 2.3 if G_t corresponds to the matrix defined in (2.3.16b), w_{yt} corresponds to the vector defined in (2.3.21) and W_{xt} to a matrix obtained applying the transformation (2.3.21) to each column of the differences of X_t.

If the log-likelihood is formed by means of the Kalman filter, G_t is a function of the parameters in the variance covariance matrices of the random shocks and it does not depend on the vector of coefficients of the exogenous variables β; while w_{yt} and W_{xt} depend on the observations as well as on the matrices Σ_ϵ, Σ_η, Σ_δ and Σ_ω. Of course,

w_{yt} and W_{xt} are functionally independent of β.

On the other hand, if the log-likelihood (1.1) or (1.2) corresponds to either the frequency domain log-likelihood or the alternative expression developed in Section 2.3, the matrix G_t is an explicit function of the parameters in the variance covariance matrices of the random shocks, and it can be written as

$$(1.3) \qquad G_t = g_{\epsilon t}\,\Sigma_\epsilon + g_{\eta t}\,\Sigma_\eta + g_{\delta t}\,\Sigma_\delta + g_{\omega t}\,\Sigma_\omega,$$

where $g_{\epsilon t}$, $g_{\eta t}$, $g_{\delta t}$ and $g_{\omega t}$ are known scalars. Also, under the same two expressions for the log-likelihood, the (n x n) matriz P_t can alternatively be defined as

$$(1.4) \qquad P_t = P_{yyt} + B\,P_{zzt}\,B' - B\,P_{zyt} - P_{yzt}\,B',$$

where P_{yyt}, P_{zzt}, P_{yzt} and P_{zyt} are (n x n), (r x r), (n x r) and (r x n) matrices respectively which depend only on the observations. Equations (1.3) and (1.4) do not hold if the log-likelihood is formed by means of the Kalman filter. Finally, notice that for the frequency domain approach, the sign ' in (1.1) represents the conjugate transpose.

The remainder of this chapter is organised as follows. Section 3.2 presents the estimation of the vector of exogenous variables β, and Section 3.3 the estimation of the parameters in the variance covariance matrices Σ_ϵ, Σ_η, Σ_δ and Σ_ω. Section 3.4 discusses the estimation strategy. Finally, Section 3.5 presents asymptotic properties of the estimators and discusses the formulation of test of hypothesis.

3.2 Estimation of the Coefficients of the Exogenous Variables

This section presents formulas for the maximum likelihood estimator of the vector of coefficients of the exogenous variables, β, in the regression model (1.1.1) or (1.1.2). General expressions for the estimator of β are presented first. Then, particular cases of interest are analysed.

From (1.1), the maximum likelihood estimator of β minimises the objective function

$$(2.1) \qquad Q = \tfrac{1}{2} \sum_t (w_{yt} - W_{xt}\beta)' G_t^{-1} (w_{yt} - W_{xt}\beta),$$

and the solution for β is given by

$$(2.2) \qquad \tilde{\beta} = \Big[\sum_t W_{xt}' G_t^{-1} W_{xt}\Big]^{-1} \Big[\sum_t W_{xt}' G_t^{-1} w_{yt}\Big],$$

while the information matrix with respect to β, $I(\beta)$, has the form

$$(2.3) \qquad I(\beta) = \Big[\sum_t W_{xt}' G_t^{-1} W_{xt}\Big].$$

As Magnus (1978) showed that the information matrix for all the parameters in the model is block diagonal with respect to β and the parameters in the variance covariance matrices of the random shocks, the inverse of $I(\beta)$ can be associated with the variance covariance matrix of the estimators under certain regularity conditions. The expression for $\tilde{\beta}$ in (2.2) can be replaced in the log-likelihood (1.1) to obtain a concentrated log-likelihood. This takes the form

$$(2.4a) \qquad \ell_c = -\tfrac{1}{2} \sum_t \big[\log|G_t| + w_{yt}' G_t^{-1} w_{yt}\big] + \tfrac{1}{2}\lambda,$$

where

$$(2.4b) \qquad \lambda = \Big[\sum_t W'_{xt}\, G_t^{-1}\, w_{yt}\Big]' \Big[\sum_t W'_{xt}\, G_t^{-1}\, W_{xt}\Big]^{-1} \Big[\sum_t W'_{xt}\, G_t^{-1}\, w_{yt}\Big].$$

This expression has then to be maximised, using a nonlinear optimisation procedure, with respect to the parameters in the variance covariance matrices of the random shocks in the model; see Section 3.3. In the basic structural model, the number of parameters to be estimated using the non linear procedure is 2n(n+1).

Important simplifications in the estimation procedure and also in the number of parameters in the model can be achieved under the homogeneity restriction (1.1.4). Using the results in Section 1.3 for the Kalman filter, and from (1.3), the homogeneity restriction implies that, whichever the approach to form the log-likelihood,

$$(2.5) \qquad G_t = g_t\, \Sigma_\epsilon,$$

where Σ_ϵ is the variance covariance matrix of the irregular random shock and g_t is an scalar which, in the basic structural model, is a function of three unknown parameters: q_η, q_δ and q_ω. These scalar parameters represent the proportional factors needed to obtain the variance covariance matrices of the level, slope and seasonal random shock respectively, from the variance covariance matrix of the irregular random shock Σ_ϵ. Replacing (2.5) in (2.2), (2.3) and (2.4), yields the estimator of β, the information matrix with respect to β, and the concentrated log-likelihood. Under the homogeneity restriction (1.1.4) the number of parameters to be estimated using a non linear procedure are reduced to (n(n+1)/2 + 3) for the basic structural model.

In practice, if the variance covariance matrix of the irregular random shock is too small, the non linear optimisation procedure might exhibit numerical instability. In this situation, it may be convenient

to express G_t in (2.5) as a function of some of the other variance covariance matrices in the model. Particularly convenient for this purpose are the ones which, under the invertibility assumption, are restricted to be positive definite. In the basic structural model these are Σ_δ and Σ_ω.

An Alternative Expression for β

The formulas presented above are obtained from the log-likelihood (1.1) which was constructed from the model (1.1.2). That expression for the model considers implicitly the restrictions (1.1.1b) on B and it is very general in the sense that it can be used for the three procedures to obtain the log-likelihood. When the log-likelihood is formed using the frequency domain approach, or the alternative approach presented in Section 2.3, equivalent, although sometimes advantagous, formulas can be obtained by using the expression (1.4) for P_t. The idea is to treat the restrictions (1.1.1b) explicitly in the optimisation procedure. From (1.2), the objective function to be minimised in the estimation of β is

$$(2.6a) \qquad Q = \tfrac{1}{2} \sum_t \operatorname{trace}\left[G_t^{-1} \left(P_{yyt} + B\, Pzzt\, B' - 2\, B\, P_{zyt}\right)\right],$$

subject to the restrictions

$$(2.6b) \qquad \operatorname{vec}(B) = S\, \beta.$$

Using the procedure proposed by Magnus and Neudecker (1988, chs. 9 and 10) to obtain the derivatives of a function of matrices, the first two differentials of (2.6a), when G_t is known for all t, are

$$(2.7a) \qquad dQ = \sum_t \operatorname{trace}(G_t^{-1}\, B\, P_{zzt}\, dB') - \operatorname{trace}(G_t^{-1}\, P_{yzt}\, dB')$$

$$= \sum_t \left[\text{vec}'(G_t^{-1} B P_{zzt}) - \text{vec}'(G_t^{-1} P_{zyt}) \right] \text{vec}(dB),$$

and

$$(2.7b) \qquad d^2Q = \sum_t \text{trace}(G_t^{-1} dB P_{zzt} dB')$$

$$= \sum_t \text{vec}'(dB) (P_{zzt} \otimes G_t^{-1}) \text{vec}(dB),$$

respectively. Also, from (2.6b)

$$(2.8) \qquad \text{vec}(dB) = S \, d\beta,$$

and then, replacing (2.8) into (2.7) yields the desired derivatives. From these, the estimator of β and the information matrix with respect to β can be written as

$$(2.9) \qquad \tilde{\beta} = \left[S' \sum_t (P_{zzt} \otimes G_t^{-1}) S \right]^{-1} \left[S' \sum_t \text{vec}(G_t^{-1} P_{yzt}) \right],$$

and

$$(2.10) \qquad I(\beta) = \left[S' \sum_t (P_{zzt} \otimes G_t^{-1}) S \right].$$

The advantage of using these formulas instead of (2.2) and (2.3), is that the Fourier transformation used in the frequency domain approach, as well as the one defined in (2.3.19), are applied only over the differences of the (r x 1) vector of exogenous variables z_t and not over the differences of the whole (n x k) matrix X_t.

As before, if the variance covariance matrices of the random shocks are unknown, a concentrated log-likelihood can be obtained by replacing (2.9) into the original log-likelihood (1.2). That yields,

$$(2.11a) \qquad \ell_c = -\tfrac{1}{2} \sum_t \left[\log|G_t| + \text{trace}(G_t^{-1} P_{yyt}) \right] + \tfrac{1}{2} \lambda,$$

where

$$\lambda = \Big[S' \sum_t \mathrm{vec}(G_t^{-1} P_{yzt})\Big]' \Big[S' \sum_t (P_{zzt} \otimes G_t^{-1}) S\Big]^{-1} \Big[S' \sum_t \mathrm{vec}(G_t^{-1} P_{yzt})\Big]. \tag{2.11b}$$

This function has then to be maximised numerically with respect to the parameters in the variance covariance matrices of the random shocks. All the comments below equation (2.4) concerning the estimation of a homogeneous model apply here.

Same Regressors in Each Equation

Consider the case where the same regressors are present in all the equations of the model. That is, there is no restrictions on the parameters of the matrix B. In terms of the notation in Section 1.1, $S = I_{nr}$ and $\mathrm{vec}(B) = \beta$. If the frequency domain or the alternative approach presented in Section 2.3 are used to obtain the log-likelihood,

$$W_{xt} = (w_{zt}' \otimes I_n), \qquad t = p+1,\ldots,T. \tag{2.12}$$

However, (2.12) does not, in general, hold if the log-likelihood is formed by means of the Kalman filter. Simpler formulas for the estimator of β, its information matrix, and the concentrated log-likelihood are obtained if (2.12) is replaced in previous formulas of this section.

An interesting situation arises when the same regressors are present in all the equations and the model is homogeneous. Under this situation, (2.12) is also true if the log-likelihood is formed by means of the Kalman filter. The reason is that as W_{xt} are pseudo innovations

after running the Kalman filter over each column of X_t, the results in Section 1.3 concerning the formulas for the Kalman filter under homogeneity imply that this operation is equivalent to run the filter over each element of the matrix X_t. Then, (2.12) follows from the fact that $X_t = (z_t' \otimes I_n)$. Using (2.12) and (2.5), the estimator of $\beta = vec(B)$ in a homogeneous model is given by

$$(2.13) \qquad \tilde{\beta} = \Big[\sum_t (w_{zt}\, g_t^{-1}\, w_{zt}') \otimes \Sigma_\epsilon^{-1}\Big]^{-1} \Big[\sum_t (w_{zt}\, g_t^{-1} \otimes \Sigma_\epsilon^{-1})\, w_{yt}\Big]$$

$$= vec\Big\{ \Big[\sum_t w_{yt}\, g_t^{-1}\, w_{zt}'\Big] \Big[\sum_t w_{zt}\, g_t^{-1}\, w_{zt}'\Big]^{-1} \Big\}.$$

From which, the maximum likelihood estimator of the matrix B is

$$(2.14) \qquad \tilde{B} = \Big[\sum_t w_{yt}\, g_t^{-1}\, w_{zt}'\Big] \Big[\sum_t w_{zt}\, g_t^{-1}\, w_{zt}'\Big]^{-1}.$$

Thus, the estimator of B depends on the parameters q_η, q_δ and q_ω but not on the variance covariance matrix Σ_ϵ. From (2.3), the information matrix with respect to β is

$$(2.15) \qquad I(\beta) = \Big[\sum_t (w_{zt}\, g_t^{-1}\, w_{zt}')\Big] \otimes \Sigma_\epsilon^{-1},$$

while the concentrated log-likelihood takes the form

$$(2.16a) \qquad \ell_c = -\tfrac{1}{2}\Big[\sum_t \log|\Sigma_\epsilon\, g_t|\Big] - \tfrac{1}{2}\,\mathrm{trace}\big[\Sigma_\epsilon^{-1}\,\Lambda\big],$$

where the (n x n) matrix Λ is equal to

$$(2.16b) \qquad \Lambda = \Big[\sum_t w_{yt}\, g_t^{-1}\, w_{yt}'\Big] - \Big[\sum_t w_{yt}\, g_t^{-1}\, w_{zt}'\Big]$$

$$\Big[\sum_t w_{zt}\, g_t^{-1}\, w_{zt}'\Big]^{-1} \Big[\sum_t w_{yt}\, g_t^{-1}\, w_{zt}'\Big]'.$$

Analogous results are obtained from the expression (1.2) for the log-likelihood. That is, replacing $S = I_{nr}$ into (2.9), (2.10) and (2.11). The concentrated log-likelihood (2.16) has to be maximised with respect to the parameters q_η, q_δ, q_ω, and the ones in Σ_ϵ. However, it is clear from (2.16a) that Σ_ϵ can also be concentrated out of the log-likelihood; see Section 3.3. The conclusion is that when the same regressors are present in all the equations of a homogeneous model, it is possible to concentrate out of the log-likelihood both the matrix of coefficients B and the variance covariance matrix of the irregular random shock Σ_ϵ. The non linear optimisation procedure is carried out only over the parameters q_η, q_δ and q_ω.

Fixed Slopes

In many practical situations the slope component β_t, in the local linear trend or basic structural models, is time invariant and can be treated as a fixed vector of parameters; see Section 1.2. Under this situation, the local linear trend and basic structural models can be written as a local level and a seasonal local level model respectively, with the time as an exogenous variable and the slope β as its coefficient. Thus, if there is no other exogenous variables, the estimation of a fixed slope can be obtained from the formulas in this section for the local level and seasonal local level models; and with $z_t = t$ ot $X_t = t\ I_n$.

Fernandez-Macho (1986, ch. 4) obtained simple expressions for the frequency domain estimators of the slope parameter β and the information matrix with respect to β. For the local level model, the value of p is the unity and the estimator of β is

$$(2.17) \qquad \tilde{\beta} = (1/(T-1)) \sum_t (y_t - y_{t-1}),$$

with the information matrix given by

$$(2.18) \qquad I(\beta) = (T-1)\, \Sigma_\eta^{-1},$$

where Σ_η is the variance covariance matrix of the level random shock. For the seasonal local level model, with seasonal period s, the value of p is s and the estimator of β is

$$(2.19) \qquad \tilde{\beta} = (1/s)\, (1/(T-s)) \sum_t (y_t - y_{t-s}),$$

while the information matrix is given by

$$(2.20) \qquad I(\beta) = ((T-s)/s^2)\, \Sigma_\eta^{-1}.$$

Formulas (2.17) and (2.19) are attractive because they are simple functions of the observations. Furthermore, it can be shown that the concentrated log-likelihood has exactly the original form (1.2) but with $P_{p+1} = 0$, and $P_t = P_{yyt}$ for $t \succ (p+1)$.

No simple formulas seems to be possible if the log-likelihood is formed using the Kalman filter or the alternative expression developed in Section 2.3.

3.3 Estimation of the Variance Covariance Matrices

This section considers the maximum likelihood estimation of the parameters in the variance covariance matrices of the random shocks in the model: Σ_ϵ, Σ_η, Σ_δ and Σ_ω.

It will be assumed, unless otherwise stated, that the vector of coefficients of the exogenous variables, β, is known; and, that the log-likelihood is formed using either the frequency domain approach or the alternative procedure developed in Section 2.3. That implies (1.3) holds. If the log-likelihood is formed using the Kalman filter, (1.3) does not hold and first and second derivatives of the log-likelihood with respect to the parameters become much more difficult to obtain analytically.

To obtain the first two derivatives of the log-likelihood, define

$$\theta_i = v(\Sigma_i), \qquad i = \epsilon,\ \eta,\ \delta \text{ and } \omega, \tag{3.1}$$

$$\theta' = (\theta'_\epsilon,\ \theta'_\eta,\ \theta'_\delta,\ \theta'_\omega), \tag{3.2}$$

$$X'_t = D\left[(g_{\epsilon t},\ g_{\eta t},\ g_{\delta t},\ g_{\omega t}) \otimes I_{n(n+1)/2}\right], \tag{3.3}$$

where $v(\Sigma_i)$ is a $(\frac{1}{2}n(n+1) \times 1)$ vector obtained from $vec(\Sigma_i)$ by eliminating all supradiagonal elements of Σ_i, and D is the $(n^2 \times n(n+1)/2)$ duplication matrix defined in Magnus (1988, ch. 4) such that for any symmetric matrix A, $vec(A) = D\ v(A)$. The $(2n(n+1) \times 1)$ vector θ contains the functionally independent parameters. Using the results developed by Fernandez-Macho (1986, ch. 5), the first two derivatives of the log-likelihood (1.2) with respect to θ are

$$\frac{\partial \ell}{\partial \theta} = \tfrac{1}{2} \sum_t X_t \left[vec(G_t^{-1\prime}\ P_t\ G_t^{-1}) - vec(G_t^{-1\prime})\right], \tag{3.4}$$

and

$$(3.5) \qquad \frac{\partial^2 \ell}{\partial \theta \, \partial \theta'} = - \tfrac{1}{2} \sum_t X_t \left[(G_t^{-1} \otimes 2 \, G_t^{-1} \, P_t \, G_t^{-1}) - (G_t^{-1} \otimes G_t^{-1}) \right] X_t'.$$

The matrix

$$(3.6) \qquad I(\theta) = \tfrac{1}{2} \sum_t X_t \, (G_t^{-1} \otimes G_t^{-1}) \, X_t',$$

is asymptotically equivalent to the information matrix; and, under some regularity conditions, its inverse can be associated with the variance covariance matrix of the maximum likelihood estimator of θ; see Section 3.5.

In general, none of the variance covariance matrices of the random shocks can be concentrated out of the log-likelihood, and for the basic structural model, the number of parameters to be estimated are 2n(n+1). Also, as the parameters in the model form variance covariance matrices which must be positive (semi) definite, some kind of restrictions should be imposed on the non linear optimisation procedure. The solution proposed by Magnus (1982), which seems to work quite well in practice, is to write each variance covariance matrix as $\Sigma_i = (L_i \, L_i')$, $i = \epsilon, \eta, \delta$ and ω; where L_i represents a lower triangular matrix whose elements, apart from the sign of the main diagonal, are unrestricted. The non zero elements in the matrices L_i define the new set of parameters to be estimated. Under this new set of parameters, the formulas (3.4) to (3.6) can still be applied with the following changes in the definitions of θ and X_t:

$$(3.7) \qquad \theta_i = v(L_i), \qquad i = \epsilon, \eta, \delta \text{ and } \omega,$$

and

$$(3.8) \qquad X_t' = (g_{\epsilon t} M_\epsilon,\ g_{\eta t} M_\eta,\ g_{\delta t} M_\delta,\ g_{\omega t} M_\omega),$$

where,

$$(3.9) \qquad M_i = [(L_i \otimes I_n) + (I_n \otimes L_i) K] D,$$

and K is the $(n^2 \times n^2)$ commutation matrix defined in Magnus (1988, ch. 3) such that for any matrix A, K vec(A) = vec(A').

If the model is homogeneous, it is immediate from (1.2) and (2.5) that Σ_ϵ can be concentrated out of the log-likelihood irrespective of the procedure used to form the log-likelihood. The maximum likelihood estimator of Σ_ϵ is

$$(3.10) \qquad \tilde{\Sigma}_\epsilon = (1/(T-p)) \sum_t P_t \tilde{g_t^{-1}},$$

and the concentrated log-likelihood takes the form

$$(3.11) \qquad \ell_c = -\tfrac{1}{2} n \left[\sum_t \log(g_t)\right] - \tfrac{1}{2} (T-p) \log|\tilde{\Sigma}_\epsilon| - \tfrac{1}{2} n (T-p).$$

This function has then to be maximised with respect to the parameters q_η, q_δ and q_ω using a non linear optimisation procedure. If the vector of coefficients of the exogenous variables is unknown, $\tilde{\Sigma}_\epsilon$ in (3.11) is a function of β, and the concentrated log-likelihood has to be maximised over β as well as over q_η, q_δ and q_ω. However, when the same regressors are present in all the equations, the maximum likelihood estimator of Σ_ϵ, from (2.16), is $[(1/(T-p))\ \Lambda]$ which does not depend on β. In that situation, both β and Σ_ϵ are concentrated out of the log-likelihood.

For the homogeneous model, it is also possible to obtain analytic first and second derivatives of the log-likelihood with respect to the parameters $v(\Sigma_\epsilon)$ and $q' = (q_\eta, q_\delta, q_\omega)$. These are,

$$(3.12)\qquad \frac{\partial \ell}{\partial v(\Sigma_\epsilon)} = -\tfrac{1}{2}(T-p)\,D'\mathrm{vec}(\Sigma_\epsilon^{-1}) + \tfrac{1}{2}\sum_t D'\mathrm{vec}(\Sigma_\epsilon^{-1} P_t g_t^{-1} \Sigma_\epsilon^{-1}),$$

$$(3.13)\qquad \frac{\partial \ell}{\partial q} = \tfrac{1}{2}\sum_t x_t g_t^{-1}\left[\mathrm{trace}(\Sigma_\epsilon^{-1} g_t^{-1} P_t) - n\right],$$

$$(3.14)\qquad \frac{\partial^2 \ell}{\partial v(\Sigma_\epsilon)\,\partial v'(\Sigma_\epsilon)} = \tfrac{1}{2}(T-p)\,D'(\Sigma_\epsilon^{-1}\otimes\Sigma_\epsilon^{-1})D$$

$$-\sum_t D'(\Sigma_\epsilon^{-1}\otimes\Sigma_\epsilon^{-1} P_t g_t^{-1}\Sigma_\epsilon^{-1})D$$

$$(3.15)\qquad \frac{\partial^2 \ell}{\partial q\,\partial q'} = \sum_t x_t g_t^{-2}\left[\tfrac{1}{2}n - \mathrm{trace}(\Sigma_\epsilon^{-1} P_t g_t^{-1})\right] x_t'$$

$$(3.16)\qquad \frac{\partial^2 \ell}{\partial v(\Sigma_\epsilon)\,\partial q'} = -\sum_t D'\mathrm{vec}(\Sigma_\epsilon^{-1} P_t g_t^{-1}\Sigma_\epsilon^{-1})\, g_t^{-1} x_t'$$

where $x_t' = (g_{\eta t}, g_{\delta t}, g_{\omega t})$. The matrix

$$(3.17)\quad I(v(\Sigma_\epsilon),q) = \begin{bmatrix} \tfrac{1}{2}(T-p)\,D'(\Sigma_\epsilon^{-1}\otimes\Sigma_\epsilon^{-1})D & \sum_t D'\mathrm{vec}(\Sigma_\epsilon^{-1})\, g_t^{-1} x_t' \\ \sum_t x_t g_t^{-1}\,\mathrm{vec}'(\Sigma_\epsilon^{-1})D & \tfrac{1}{2}n\sum_t x_t g_t^{-2} x_t' \end{bmatrix}$$

is asymptotically equivalent to the information matrix; hence the inverse of it can be associated with the variance covariance matrix of the maximum likelihood estimators; see Section 3.5.

3.4 Estimation Strategy

The previous two sections considered separately the estimation of the two sets of parameters in the model: the coefficients of the exogenous variables β, and the functionally independent elements in the variance covariance matrices of the random shocks Σ_ϵ, Σ_η, Σ_δ and Σ_ω. This section analyses the alternative options to obtain the maximum likelihood estimators.

If the model is homogeneous and the same regressors are presented in all the equations, both β and Σ_ϵ can be concentrated out of the log-likelihood irrespective of the procedure used to form this log-likelihood; see (2.13) and (3.10). The concentrated log-likelihood (3.11) is then maximised with respect to the parameters q_η, q_δ and q_ω. This has to be done using a non linear optimisation procedure. Analytic derivatives are difficult to obtain from (3.11).

If the model is homogeneous but the regressors are not the same in all the equations, it is possible to concentrate out of the log-likelihood either β or Σ_ϵ but not both. If β is concentrated out, the estimator of β is given by (2.2) and the concentrated log-likelihood by (2.4). In both equations G_t is given by (2.5). On the other hand, if Σ_ϵ is concentrated out, the maximum likelihood estimator of Σ_ϵ was presented in (3.10), with the concentrated log-likelihood in (3.11). Under this two alternative estimation procedures, the parameters which are not concentrated out of the log-likelihood have to be estimated using a non linear optimisation procedure, with no analytic derivatives. A third option, which seems to be quite appropiated in these circumstances is a stepwise procedure; see Sargan (1964) and Oberhofer and Kmenta (1974). Given consistent initial estimates of β, Σ_ϵ, q_η, q_δ and q_ω, which can be obtained as indicated

below, a new estimate of β is obtained from (2.2) or (2.9). This value of $\tilde{\beta}$ is then used to evaluate and maximise the log-likelihood (1.2), where Σ_ϵ can be concentrated out as in (3.10). The procedure is repeated until convergence of the log-likelihood and the parameters. All these results are independent of the log-likelihood used.

If the model is not homogenous, the only set of parameters which can be concentrated out of the log-likelihood is β; while the parameters in Σ_ϵ, Σ_η, Σ_δ and Σ_ω have to be estimated using a non linear optimisation procedure. This is true even if the regressors are the same in all the equations. The problem in concentrating β out is that it is difficult to obtain first and second derivatives of the concentrated log-likelihood (2.4) with respect to the parameters in the variance covariance matrices of the random shocks. If none of the parameters is concentrated out, analytic first and second derivatives with respect to all the parameters in the model are relatively easy to evaluate for the frequency domain log-likelihood and for the alternative log-likelihood developed in Section 2.3; see Sections 3.2 and 3.3.

Although this chapter is basically concerned with the maximum likelihood estimation procedure, preliminary and consistent estimators of all the parameters can be obtained from the stationary form of the model defined in Section 1.2. As this stationary form can be seen as a regression model with autocorrelated residuals, the ordinary least squares procedure gives consistent estimates of the vector β, while the autocovariance matrices of the residuals give consistent estimates of the variance covariance matrices of the random shocks; see Hannan (1970, secs. 4.3 and 7.5). Of course, that suggests the possibility of a two step or scoring algorithm to obtain asymptotically efficient estimators. Fernandez-Macho (1986, sec. 3.3) developed the scoring

algorithm for the models presented here. As Magnus (1978) showed that the information matrix is block diagonal with respect to the two subsets of parameters: β and θ, the scoring algorithm is run in parallel.

3.5 Tests of Hypotheses

Hannan (1970), Dunsmuir and Hannan (1976), and Dunsmuir (1979) studied the asymptotic properties of the frequency domain maximum likelihood estimators of vector ARMA models. The reduced form of the models considered in this chapter correspond to restricted ARMA models and then the results of the mentioned studies can be applied to establish the asymptotic properties of the estimators of the models considered here.

The study by Dunsmuir and Hannan (1976) presents conditions for the strong consistency of the frequency domain maximum likelihood estimators. Kohn (1979) presents conditions for the strong consistency of the time domain maximum likelihood estimators.

The asymptotic normality of the frequency domain estimator of β can be obtained from the results in Hannan (1970, sec. 7.4). Besides regularity conditions, Hannan assumes the Grenander's conditions over the vector of exogenous variables z_t to obtain that $T^{\frac{1}{2}} (\tilde{\beta} - \beta)$ has, asymptotically, a normal distribution with expected value zero and a variance covariance matrix which is consistently estimated by $(T I^{-1}(\beta))$, where $I(\beta)$ defined in (3.2.4) is evaluated at the maximum likelihood estimator.

A central limit theorem for the frequency domain maximum likelihood estimators of θ is provided by Dunsmuir (1979). Under normality of the random shocks and certain regularity conditions, $T^{\frac{1}{2}} (\tilde{\theta} - \theta)$ has, asymptotically, a normal distribution with expected value zero and a variance covariance matrix which is consistently estimated by $(T I^{-1}(\theta))$, where $I(\theta)$ defined in (3.3.6) or (3.3.17) is evaluated at the maximum likelihood estimator.

Dunsmuir's central limit theorem is not the only one available for

vector ARMA models. The difference with other central limit theorems is that the parameters in the variance covariance matrix of the innovations and all the other parameters in the model are not partitioned into two subsets; see Dunsmuir and Hannan (1976) and Kohn (1979). That is exactly the situation of the structural time series models considered here where such partition is not, in general, possible.

Finally, for the univariate case, n = 1, Pagan (1980) presents conditions for the consistency and asymptotic normality of the time domain maximum likelihood estimator of (β , θ).

The above results enable the formulation of asymptotic tests of hypotheses. Particularly important is a test for the homogeneity hypothesis (1.1.4). As the model under homogeneity is substantially simpler to estimate, a test for homogeneity is better based on the LM principle. Fernandez-Macho (1986, sec. 3.5) formulated an LM test which has the form

$$(5.1) \qquad LM = \tfrac{1}{2} \Big[\sum_t \text{vec}'(q_t^{-2}\, \Sigma_\epsilon^{-1}\, P_t\, \Sigma_\epsilon^{-1} - q_t^{-1}\, \Sigma_\epsilon^{-1})\, X_t' \Big]$$

$$\Big[\sum_t X_t\, q_t^{-2}\, (\Sigma_\epsilon \otimes \Sigma_\epsilon)^{-1}\, X_t' \Big]^{-1}$$

$$\Big[\sum_t X_t\, \text{vec}(q_t^{-2}\, \Sigma_\epsilon^{-1}\, P_t\, \Sigma_\epsilon^{-1} - q_t^{-1}\, \Sigma_\epsilon^{-1}) \Big].$$

The degrees of freedom associated to this statistic are ($\frac{1}{2}$n(n+1)3-3). To test hypotheses concerning the parameters in the vector β, a Wald or a Likelihood Ratio test can be used.

CHAPTER 4 : THE DEMAND FOR ENERGY IN THE U. K. : AN APPLICATION

4.1 Introduction

This chapter presents an econometric study of the demand for energy in the U.K. economy. The study, which represents an application of the techniques presented in previous chapters, considers quarterly time series data for the period 1971-1986, four economic sectors: other industry, domestic, other final users and transport; and the four most important fuels : gas, electricity, oil and coal. The objectives of the study are (i) to construct an econometric model to explain the substitution possibilities between the four fuels in each economic sector, and (ii) to obtain forecasts of the individual demands.

A translog cost function is used to explain the production possibilities in each sector. Under the assumption of separability in energy inputs this leads to a share equation system in which the share of each energy input in the total cost of energy depends on the prices of all the energy inputs. Estimation of such a system enables estimates of substitution and demand elasticities to be made. There is a considerable literature on the use of translog production and cost functions; see the recent survey by Jorgenson (1986).

Technical progress enters into the model in two ways. Neutral technical progress affects the overall production, and hence the cost, irrespective of the mix of inputs employed and so is not associated with any particular input. On the other hand, factor augmenting technical progress affects output and cost via particular inputs. Such technical progress is clearly relevant in the case of energy where the changes in technology are often specific to particular inputs. The present study makes a methodological contribution to the way in which

factor augmenting technical progress is modelled by using stochastic, rather than deterministic, time trends. The use of stochastic trends to pick up the effects of technical progress has already proved to be quite effective in other contexts; see the studies by Harvey et all (1986) and Slade (1989). Similar improvements in parameter estimates and forecasts can likewise be expected here. From the statistical point of view, we draw on the notion of statistical homogeneity defined in Section 1.1.

Using the estimated coefficients from the share equations, a price index for energy may be constructed for each sector. Since the total cost of energy is decomposed into a price index and a quantity index, a quantity index may also be constructed. Following the assumed specification of the cost function, this quantity index is modelled in terms of output, temperature, and a stochastic trend, part of which can be interpreted as neutral technical progress. Forecasts of future values of the quantity index are made from this model and combining these with a price index based on hypothesised future prices leads to predictions for total costs. Predictions for individual energy demand are then made using the share equations.

Earlier work on U.K. energy demand by Pepper (1985) used a multivariate ARIMA modelling approach, and did not take account of the economic structure on production and cost functions. The attraction of the structural time series modelling approach adopted here is that it ties in much more naturally with the models suggested by economic theory.

The chapter is organized as follows. Section 4.2 presents the econometric model. Section 4.3 considers the estimation of the share equation system in each economic sector. Section 4.4 contains the estimation of the total cost equations and the forecasts. The

conclusions are presented in Section 4.5. Finally, in Appendix 4.1 is presented the definition of the data and Appendix 4.2 contains a result concerning the estimation of a multivariate regression model with stochastic trend and seasonal components when the vector of dependent variables is constrained.

4.2 The Econometric Model

This section considers the general specification of an econometric model for energy demand and discusses how forecasts can be made.

The Economic Model

Assume that a firm produces output Q using energy inputs X_1, $X_2,\ldots,$ X_n; non-energy inputs $\bar{X}_1$, $\bar{X}_2,\ldots,$ $\bar{X}_m$ and a level of technology T. Then,

$$(2.1) \qquad Q = F(X_1,\ldots,X_n;\ \bar{X}_1,\ldots,\bar{X}_m;\ T).$$

Assume that the level of technology can be represented by an index of technical progress A which is neutral, and indexes $A_1,\ldots,A_n$, $\bar{A}_1,\ldots,\bar{A}_m$ of relative technical progress which take the factor augmented form.

If the input prices $P_1,\ldots,P_n$, $\bar{P}_1,\ldots,\bar{P}_m$ are exogenous to the firm, and under cost minimising behaviour, the characteristics of production can be represented by a cost function C of the form

$$(2.2) \qquad C = C\left(\frac{P_1}{A_1},\ldots,\frac{P_n}{A_n};\ \frac{\bar{P}_1}{\bar{A}_1},\ldots,\frac{\bar{P}_m}{\bar{A}_m};\ A,\ Q\right).$$

The factor augmented form hypothesis for the technical progress has been used in empirical work; see Binswanger (1974) and Wills (1979). As shown in Solow (1967) it imposes a constraint on the form in which the technology affects the production possibilities. The idea behind the augmentation factors is that they transform the inputs from "physical" to "efficient" units. From this point of view, the function (2.2) can be seen as a standard cost function in prices of "efficient" units.

If the production function is homothetic weak separable (HWS) in

the energy inputs, then as shown by Shephard (1953), it is possible to write the total cost function (2.2) as

$$(2.3) \qquad C = C\left(\frac{P_e}{A_e} ; \frac{\bar{P}_1}{\bar{A}_1}, \ldots, \frac{\bar{P}_m}{\bar{A}_m} ; A, Q\right),$$

where

$$(2.4) \qquad \frac{P_e}{A_e} = C_e\left(\frac{P_1}{A_1}, \ldots, \frac{P_n}{A_n}\right).$$

Here, P_e is an aggregate price index of energy and P_e/A_e the price in augmented form.

The HWS assumption is a standard one in the econometric literature when the interest of the study lies in analysing the substitution possibilities of a subset of inputs; see Magnus and Woodland (1987), Fuss (1977) and Pindyck (1979). Shephard (1953) showed that HWS is a necessary and sufficient condition for a two-stage allocation. At the first stage, the optimal mix of energy inputs is chosen by minimising the cost per unit of energy given by P_e/A_e in equation (2.4); while at the second stage the mix of energy, as an aggregate, and non-energy inputs is chosen from the minimisation of the total cost in (2.3).

This study concentrates on analysing the substitution possibilities of the energy inputs. The translog second order approximation of (2.4) is

$$(2.5) \qquad \ell n \frac{P_e}{A_e} = \alpha_0 + \sum_i \alpha_i \ell n \frac{P_i}{A_i} + \tfrac{1}{2} \sum_i \sum_j \alpha_{ij} \ell n \frac{P_i}{A_i} \ell n \frac{P_j}{A_j},$$

where α_0, $\{\alpha_i,\ i = 1,\ldots,n\}$, $\{\alpha_{ij},\ i,j = 1,\ldots,n\}$ are fixed parameters. Differentiating (2.5) with respect to the logarithm of augmented prices, $\ell n\ P_i/A_i$, $i = 1,\ldots,n$, yields the share equation system

$$(2.6) \qquad S_i = \alpha_i + \sum_j \alpha_{ij} \ln \frac{P_j}{A_j}, \qquad i = 1,\ldots,n,$$

where, S_i is the share of the energy input i in the total energy cost.

If the production function associated with the cost function (2.4) is a well behaved production function, the following restrictions should be imposed on the parameters of equation (2.6):

(2.7a) Cost Exhaustion: (i) $\sum_i \alpha_i = 1,$

(ii) $\sum_i \alpha_{ij} = 0, \quad j = 1,\ldots,n.$

(2.7b) Homogeneity: $\sum_j \alpha_{ij} = 0, \quad i = 1,\ldots,n.$

(2.7c) Symmetry: $\alpha_{ij} = \alpha_{ji}, \quad i,j = 1,\ldots,n.$

(2.7d) Concavity: The price function (2.4) is concave in the input prices.

Two commonly used measures of price responsiveness are the Allen-Uzawa partial elasticity of substitution, s_{ij}, and the price elasticity of demand, e_{ij}. Berndt and Wood (1975) showed that for the translog cost function these measures are

$$(2.8a) \qquad s_{ij} = (\alpha_{ij} + S_iS_j)/S_iS_j, \qquad \text{if } i \neq j,$$

$$(2.8b) \qquad s_{ii} = (\alpha_{ii} + S_i^2 - S_i)/S_i^2,$$

$$(2.8c) \qquad e_{ij} = s_{ij} S_j, \qquad \text{if } i \neq j,$$

$$(2.8d) \qquad e_{ii} = s_{ii} S_i.$$

If $s_{ij} \prec 0$, the factors i and j are said to be complements. If $s_{ij} \succ 0$ the two factors are substitutes; and if $s_{ij} = 0$, they are independent.

To measure the biases of the technical progress, which represent the change in the shares at constant prices, we use the definition of Binswanger (1974). That is,

$$(2.9) \qquad B_i = \frac{d\bar{S}_i}{S_i}, \qquad i = 1,\ldots,n,$$

where $d\bar{S}_i$ is the change in the share of fuel i, with constant prices.

The Stochastic Specification

Assume that the shares of each fuel for times t = 1,...,T are observed. Adding a random disturbance term to system (2.6) gives

$$(2.10) \qquad S_{it} = \alpha_i + \sum_j \alpha_{ij} \ell n\,(P_{jt}/A_{jt}) + \epsilon_{it}, \qquad i = 1,\ldots,n,$$

where the vector $\epsilon_t' = (\epsilon_{1t} \ldots \epsilon_{nt})$ is assumed serially uncorrelated, with expected value zero and variance covariance matrix Σ_ϵ.

At this point, a specific form for the augmentation factors A_{it}, i = 1,...,n is needed. The logarithms of these factors are assumed to follow a random walk with drift. That is

$$(2.11) \qquad \ell n\, A_{it} = \ell n\, A_{it-1} + \bar{\beta}_i + \bar{\eta}_{it}, \qquad i = 1,\ldots,n,$$

where the slope parameters $\bar{\beta}_i$ are fixed, and the disturbance vector $\bar{\eta}_t'$ = $(\bar{\eta}_{1t},\ldots,\bar{\eta}_{nt})$ is serially uncorrelated, independent of ϵ_t in equation (2.10), with expected value zero and variance covariance matrix $\bar{\Sigma}_\eta$. If the disturbance term is removed from (2.11), $\ell n\, A_{it}$ reduces to a deterministic trend.

The system of share equations can now be written as

$$(2.12a)\qquad S_{it} = \mu_{it} + \sum_j \alpha_{ij} \ell n\, P_{jt} + \epsilon_{it}, \qquad \begin{matrix} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{matrix}$$

$$(2.12b)\qquad \mu_{it} = \mu_{i,t-1} + \beta_i + \eta_{it},$$

where, in terms of the system (2.10), the following relations are obtained

$$(2.13a)\qquad \mu_{it} = \alpha_i - \sum_j \alpha_{ij} \ell n\, A_{jt}, \qquad i = 1,\ldots,n,$$

$$(2.13b)\qquad \beta_i = - \sum_j \alpha_{ij} \bar{\beta}_j, \qquad i = 1,\ldots,n,$$

$$(2.13c)\qquad \eta_{it} = - \sum_j \alpha_{ij} \bar{\eta}_{jt}, \qquad i = 1,\ldots,n.$$

The system of equations (2.12) has the form of the multivariate regression model with stochastic trend components defined in Section 1.1. The slopes β_i in (2.12b) can be written in (2.12a) as the coefficients of the time variable; see Section 1.2. That is, in terms of the structural time series models defined in Chapter 1, the residuals of the regression model (2.12) follow a local level model. The maximum likelihood estimation of the parameters and the estimation of the unobserved components in this model were considered in previous chapters. Three comments are in order:

a) The sum of the dependent variables is unity for all t. It is shown in Appendix 4.2 that only (n-1) equations need to be estimated.

b) For the identification of the augmentation factors A_{it} from (2.13a), the following extra conditions are required

$$(2.14a)\qquad \ell n\, A_{i1} = 0, \qquad i = 1,\ldots,n$$

(2.14b) $\sum_i \ell n\ A_{it} = 0, \qquad t = 1,\ldots,T.$

The first of these restrictions is needed given the presence of the parameter α_i in (2.13a); while the second restriction is required from the fact that we can estimate independently only (n-1) of the components μ_{it}, i = 1,...,n.

c) A stochastic seasonal component can be added to the model if quarterly or monthly data is used at the estimation stage.

Forecasting

Suppose that forecasts of the individual energy demands $X_1,\ldots,X_n$ are required. The system of equations (2.12) gives forecast of the shares given input prices. To obtain forecasts of the individual demands, a forecast of the total cost of energy, TC_{et}, is needed. This total cost of energy decomposes as

(2.15) $\ell n\ TC_{et} = \ell n\ P_{et}/A_{et} + \ell n\ Q_{et}\ A_{et}, \qquad t = 1,\ldots,T,$

where P_{et}/A_{et} and $Q_{et}\ A_{et}$ are the price and quantity of energy in augmented form. In principle, we might forecast $\ell n\ TC_{et}$ directly. However, we prefer to form a forecast of a price index of energy; and then to forecast the quantity of energy which is in real terms and can be associated with other real variables.

The usual practice in the econometric literature is to use a Divisia index to form an aggregate of prices; see Fuss (1977) and Diewert (1976). For the model presented in this study, the Divisia index in differential form is

$$(2.16) \qquad d\ \ell n \frac{P_{et}}{A_{et}} = \sum_{i=1}^{n} S_{it}\ d\ \ell n \frac{P_{it}}{A_{it}}, \qquad t = 1,\ldots,T,$$

and a natural discrete approximation for this expression is

$$(2.17) \quad \ell n \frac{P_{et}}{A_{et}} - \ell n \frac{P_{et-1}}{A_{et-1}} = \sum_{i} \tfrac{1}{2} (S_{it} + S_{it-1}) \left(\ell n \frac{P_{it}}{A_{it}} - \ell n \frac{P_{it-1}}{A_{it-1}} \right).$$

Using this expression and (2.15) gives a quantity index of energy. This index may be predicted using a structural time series model with output and temperature as exogenous variables, and a stochastic trend component which can be partially associated with the neutral technical progress in energy, A_{et}.

The forecasting procedure as a whole consists then of the following steps.

a) Given prices of the energy inputs, forecast the shares using the system (2.12). Use the same system to forecast the components μ_{it}, i = 1,...,n; and, with the restrictions (2.14), the augmentation factors A_{it}, i = 1,...,n.

b) Forecast $\ell n\ (P_{et}/A_{et})$ using (2.17) and $\ell n\ (Q_{et}\ A_{et})$ using a univariate model. Form a forecast of $\ell n\ TC_{et}$ using (2.15).

c) Multiply the total cost TC_{et} by the shares to obtain individual costs. These values divided by the prices give the individual demand forecasts.

4.3 Estimation of the Model and Results

This section considers the estimation, for each sector, of the share equation system (2.12). The fuels examined are gas, electricity, oil and coal for the other industry, domestic and other final users sectors; and electricity and oil for the transport sector. The sample period goes from 1971 Q1 to 1986 Q4.

Maximum Likelihood Estimation

Adding a seasonal component to (2.12), the model can be written in matrix form as

$$S_t = \mu_t + \gamma_t + A\, p_t + \epsilon_t, \qquad t = 1,\ldots,T, \tag{3.1a}$$

$$\mu_t = \mu_{t-1} + \beta + \eta_t, \tag{3.1b}$$

$$(1 + L + L^2 + L^3)\, \gamma_t = \omega_t, \tag{3.1c}$$

where L is the lag operator. In accordance with the definitions in Section 1.2, (3.1) corresponds to a seasonal local level model. S_t is the vector of shares, μ_t is a vector of trends with i-th component μ_{it} as defined in (2.12b), β is a vector of fixed parameters representing the slopes of the trends, and γ_t is the vector of seasonal components. The logarithms of the input prices contained in the vector p_t are assumed to be exogenous and the matrix A has as its (i,j)-th element the fixed parameter α_{ij} defined in (2.5). Finally, the random shocks ϵ_t, η_t and ω_t are assumed to be serially and mutually uncorrelated, with expected values zero and variance covariance matrices Σ_ϵ, Σ_η and Σ_ω respectively.

As noted earlier, one of the equations in (3.1) is redundant

because the sum of the dependent variables is unity for all times. Thus, for estimation purposes, the system has dimension three in the other industry, domestic, and other final users sectors, while it reduces to a single equation in the transport sector.

Assuming the disturbances in (3.1) to be normally distributed allows estimation to be carried out by maximum likelihood. Although the normality assumption cannot be strictly valid for share equations, it is not unreasonable provided none of the shares is very small. The estimation can be simplified considerably by imposing the statistical homogeneity restriction (1.1.4). That is,

$$\Sigma_\epsilon = q_\epsilon \Sigma_\omega, \tag{3.2a}$$

$$\Sigma_\eta = q_\eta \Sigma_\omega, \tag{3.2b}$$

where q_ϵ and q_η are scalar parameters. The restriction (3.2) is an important one in the estimation of structural time series models. It not only reduces the number of parameters in the model; but also, it allows the matrix Σ_ω to be concentrated out of the likelihood, reducing significantly the number of parameters that have to be estimated using a nonlinear estimation procedure.

If the matrix of coefficients of the exogenous variables were unrestricted, it could also be concentrated out of the likelihood, leaving the nonlinear optimisation procedure to be carried out over the parameters q_ϵ and q_η only. However, in (3.1), the matrix A is subject to the economic restrictions (2.7a) to (2.7c). That is, A is a symmetric matrix and its rows sum to zero. The restriction (2.7d) was not imposed on the model; it can be checked afterwards using the matrix of substitution elasticities. The estimation of the model, for each sector, was done using the frequency domain approach; see chapter 3. Of course, for the transport sector where the system has only one

equation, the homogeneity restriction and the discussion which follows equations (3.2) is irrelevant.

Results

The results of the estimation procedure are presented in tables 4.3.1, 4.3.2 and 4.3.3. Most of the price coefficients, in Table 4.3.1, are significant at the 5% significance level. The estimates of the slopes of the trends, which are associated to the slopes of the biases in technical progress, are shown in Table 4.3.2. In the trends of the gas equations the slopes are greater than zero and clearly significant in all the economic sectors. For the electricity fuel, the slopes are greater than zero and significant in the other industry and other final users sectors; while in the domestic and transport sectors the estimates of the slopes are less than zero and not very significant. For the oil fuel, the estimates of the slopes are negative in all but the transport sector. However, its significance is important only in the other industry and other final users sectors. The estimates of the slopes in the coal equation are always less than zero and especially significant in the other industry and domestic sectors.

Likelihood Ratio and Lagrange Multiplier statistics were used to test the economic restrictions (2.7a) to (2.7c) and the statistical homogeneity (3.2). The results are shown in Table 4.3.4. At the 5% significance level, the economic restrictions are accepted in the other industry, domestic and other final users sectors. In the transport sector, the restrictions are accepted at the 3% level. Statistical homogeneity is accepted in the other industry sector at any reasonable level; while in the domestic and other final users sectors, the statistics are very close to the critical value at the 5% significance

level. In fact, the hypothesis of homogeneity is accepted at the 4.6% and 3.6% level respectively.

Tests for normality and serial correlation were applied to the residuals in all the four models. No evidence of misspecification was found.

Table 4.3.1: Estimates of Price Effects(1)

Parameter	Other Industry	Domestic	Other Final Users	Transport
α_{gg}	.033	.155	.060	-
	(.020)	(.039)	(.016)	
α_{ge}	-.014	-.117	-.047	-
	(.016)	(.021)	(.017)	
α_{go}	-.032	-.018	-.027	-
	(.010)	(.014)	(.008)	
α_{gc}	.013	-.020	.014	-
	(.017)	(.038)	(.008)	
α_{ee}	.220	.239	.253	.005
	(.025)	(.023)	(.026)	(.001)
α_{eo}	-.133	-.024	-.160	-.005
	(.010)	(.010)	(.013)	(.001)
α_{ec}	-.073	-.098	-.046	-
	(.022)	(.028)	(.012)	
α_{oo}	.196	.057	.191	.005
	(.013)	(.009)	(.012)	(.001)
α_{oc}	-.031	-.015	-.004	-
	(.010)	(.016)	(.004)	
α_{cc}	.091	.133	.036	-
	(.029)	(.054)	(.013)	

(1) Standard errors in parenthesis. Gas: g, Electricity: e, Oil: o, Coal: c.

Table 4.3.2: Slopes of the Biases in Technical Progress(1)

Parameter	Other Industry	Domestic	Other Final Users	Transport
$\beta_g(x10^{-2})$	.291	.354	.203	-
	(.078)	(.080)	(.062)	
$\beta_e(x10^{-2})$	.161	-.068	.196	-.003
	(.076)	(.049)	(.099)	(.002)
$\beta_o(x10^{-2})$	-.346	-.042	-.298	.003
	(.094)	(.036)	(.086)	(.002)
$\beta_c(x10^{-2})$	-.106	-.244	-.101	-
	(.069)	(.075)	(.085)	

(1) Standard errors in parenthesis. Gas: g, Electricity: e, Oil: o, Coal: c.

Table 4.3.3 : Error Structure of the Models(1)

Sector	Σ_ω (x10^{-5})	q_ϵ	q_η
Other Industry	$\begin{bmatrix} .767 & & & \\ -.090 & .692 & & \\ -.423 & -.493 & 1.147 & \\ -.254 & -.109 & -.231 & .594 \end{bmatrix}$	2.513 (1.805)	4.547 (1.727)
Domestic	$\begin{bmatrix} 3.294 & & & \\ -.943 & 1.265 & & \\ -.457 & .197 & .724 & \\ -1.894 & -.519 & -.464 & 2.877 \end{bmatrix}$	1.719 (.835)	1.017 (.380)
Other Final Users	$\begin{bmatrix} 1.587 & & & \\ -1.188 & 4.150 & & \\ -.363 & -2.762 & 3.225 & \\ -.036 & -.200 & -.100 & .336 \end{bmatrix}$	.210 (.815)	1.333 (.800)
Transport	$\begin{bmatrix} .00095 & \\ -.00095 & .00095 \end{bmatrix}$	2.947 (5.804)	3.648 (5.365)

(1) Standard errors in parenthesis.

Table 4.3.4: Test of Hypothesis

Hypothesis	d.f.	Chi-square Statistics
Economic Restrictions(1)		
Other Industry	6	12.160
Domestic	6	10.980
Other Final Users	6	4.060
Transport	1	4.700
Statistical Homogeneity(2)		
Other Industry	10	8.466
Domestic	10	18.329
Other Final Users	10	19.522

(1) Equations (2.7a) to (2.7c)

(2) Equations (3.2)

Given the estimates of the parameters in the model, and for a given vector of shares, we can obtain substitution and demand elasticities as defined in (2.8). As the model does not consider the long term effects of prices, these elasticities should be interpreted as short term elasticities. Table 4.3.5 presents the demand elasticities for the average values of the shares in the sample period. In the other industry sector, the results are, in general terms, consistent with the findings by Magnus and Woodland (1987) for the Dutch manufacturing industry, and by Fuss (1977) for Canada; that is, gas, electricity and oil are substitutes. The own demand elasticity for gas is -.62 while for electricity and oil these values are -.06 and -.07. In the domestic sector, it was also found that gas, electricity and oil are substitutes. The own demand elasticities for gas and oil are -.21 and -.28; while for the electricity and coal fuels, the own demand elasticities are very close to zero. The estimated elasticities for the other final users sector are, in general, similar to the ones in the other industry sector. The own demand elasticity for gas is -.41 and for electricity and oil these values are very small. The estimated own demand elasticity for coal in this sector was .23. The transport sector model considers only the electricity and oil fuel. They are substitutes and the own demand elasticities are -.49 for electricity and zero for oil.

Using estimates of the unobserved components μ_t, in model (3.1), it is possible to compute the biases in technical progress using a discrete approximation to (2.9). This is defined as,

$$(3.3) \qquad B_{it} = 100 \left[\tilde{\mu}_{it/T} - \tilde{\mu}_{i,t-1/T} \right] / S_{it}, \qquad \begin{matrix} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{matrix}$$

where B_{it} is the bias of input i at time t, S_{it} is the share of input i at time t, and $\tilde{\mu}_{it/T}$ is the smoothed estimate of μ_{it}; that is, the

estimate based on all the sample information.

Figures 4.3.1, 4.3.2, 4.3.3 and 4.3.4 show the annual averages of the biases in technical progress for the four economic sectors. In the other industry sector, Figure 4.3.1, the technology biases the use of fuels towards gas and electricity. The biases for gas are high but decreasing during the sample period. The large value of the 1971 bias for gas can be explained by the introduction of the North Sea gas. Figure 4.3.2 presents the biases for the domestic sector. The values for gas are greater than zero although smaller than in the other industry sector. The biases for electricity, oil and coal are, in general, less than zero. In the other final users sector, Figure 4.3.3, the biases are positive for gas and electricity, and negative for oil and coal. Finally, the biases for the two fuels in the transport sector are presented in Figure 4.3.4. The figure shows positive biases for oil and negative biases for electricity, although in both cases the values are very small.

The system of share equations were also estimated using deterministic trends and dummy variables to capture the seasonal effects. The results of this exercise showed substantial changes in the price coefficients, and very significant serial correlation in the residuals of the models for all the sectors. The ratio of the Akaike information criteria for the models with fixed trends and seasonals, to the Akaike information criteria for the systems with stochastic components, were 1.29 for the other industry sector, 1.57 for the domestic sector, 3.49 for the other final users sector, and 16.89 for the transport sector. These results show significant evidence in favour of the models with stochastic trends and seasonals, as presented in this study.

Table 4.3.5 : Demand Elasticities(1)

Elasticity	Other Industry	Domestic	Other Final Users	Transport
e_{gg}	-.62	-.21	-.41	-
e_{ge}	.36	.06	.21	-
e_{go}	.09	.04	.06	-
e_{gc}	.17	.11	.14	-
e_{eg}	.11	.06	.05	-
e_{ee}	-.06	.01	.01	-.49
e_{eo}	.03	.03	-.01	.49
e_{ec}	-.08	-.10	-.05	-
e_{og}	.04	.17	.03	-
e_{oe}	.05	.11	-.02	.00
e_{oo}	-.07	-.28	-.02	.00
e_{oc}	-.02	.01	.02	-
e_{cg}	.30	.24	.60	-
e_{ce}	-.45	-.23	-.96	-
e_{co}	-.07	.00	.14	-
e_{cc}	.22	-.01	.23	-

(1) e_{ij} represent the change in the demand input i given a unit change in the price of input j. Gas: g, Electricity: e, Oil: o, Coal: c.

Figure 4.3.1 : Biases in Technical Progress - Other Industry Sector

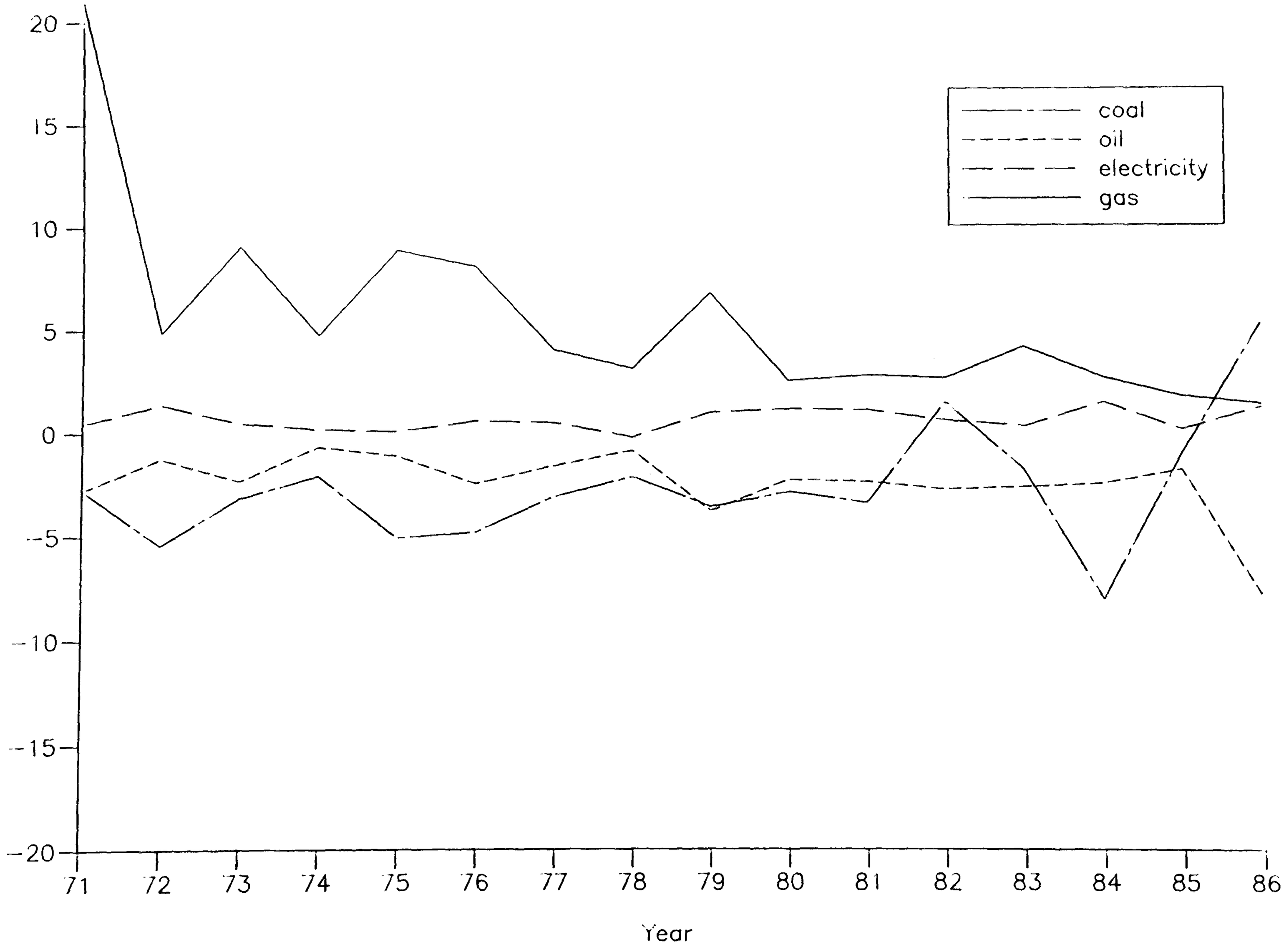


Figure 4.3.2 : Biases in Technical Progress : Domestic Sector

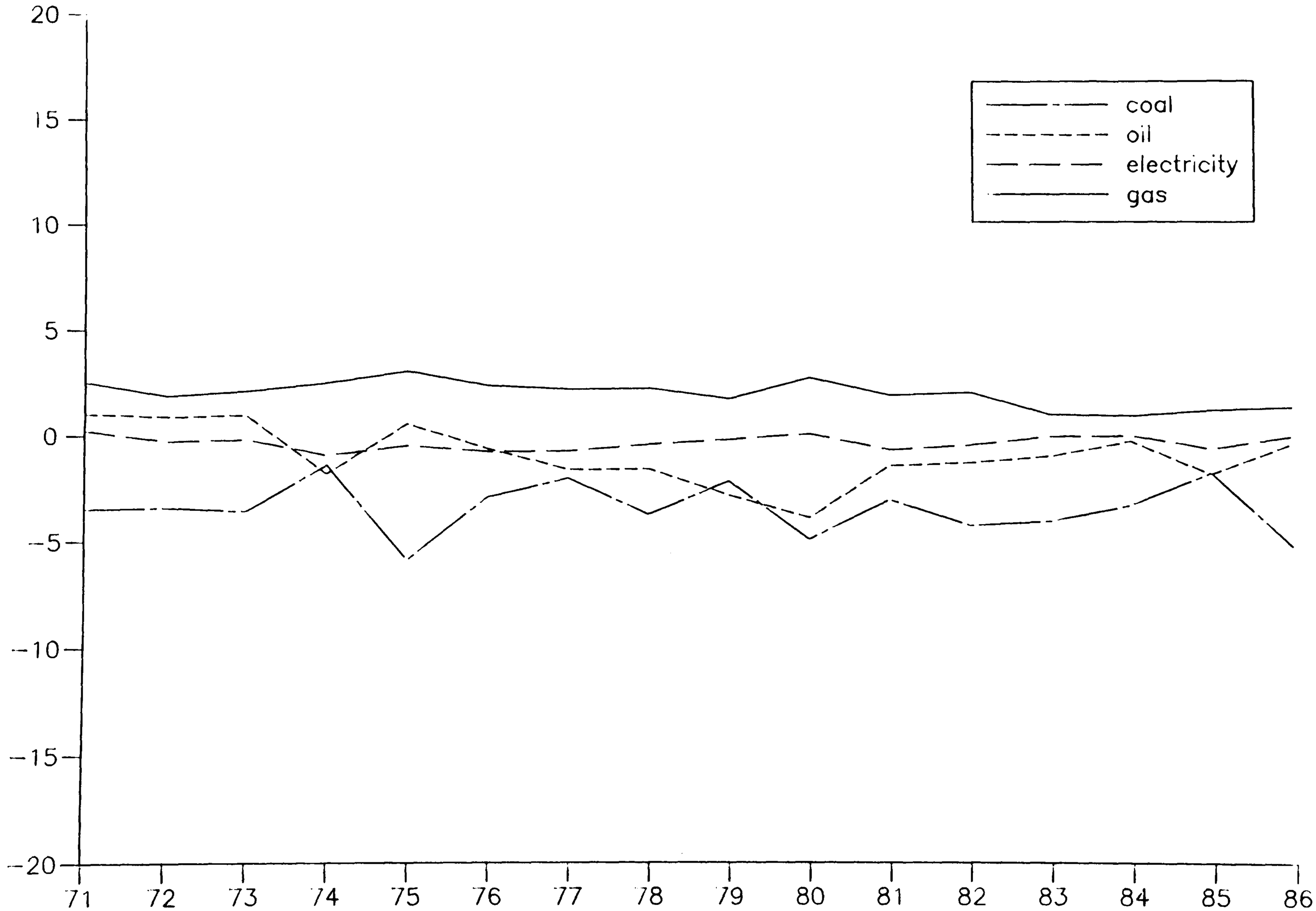


Figure 4.3.3 : Biases in Technical Progress : Other Final Users Sector

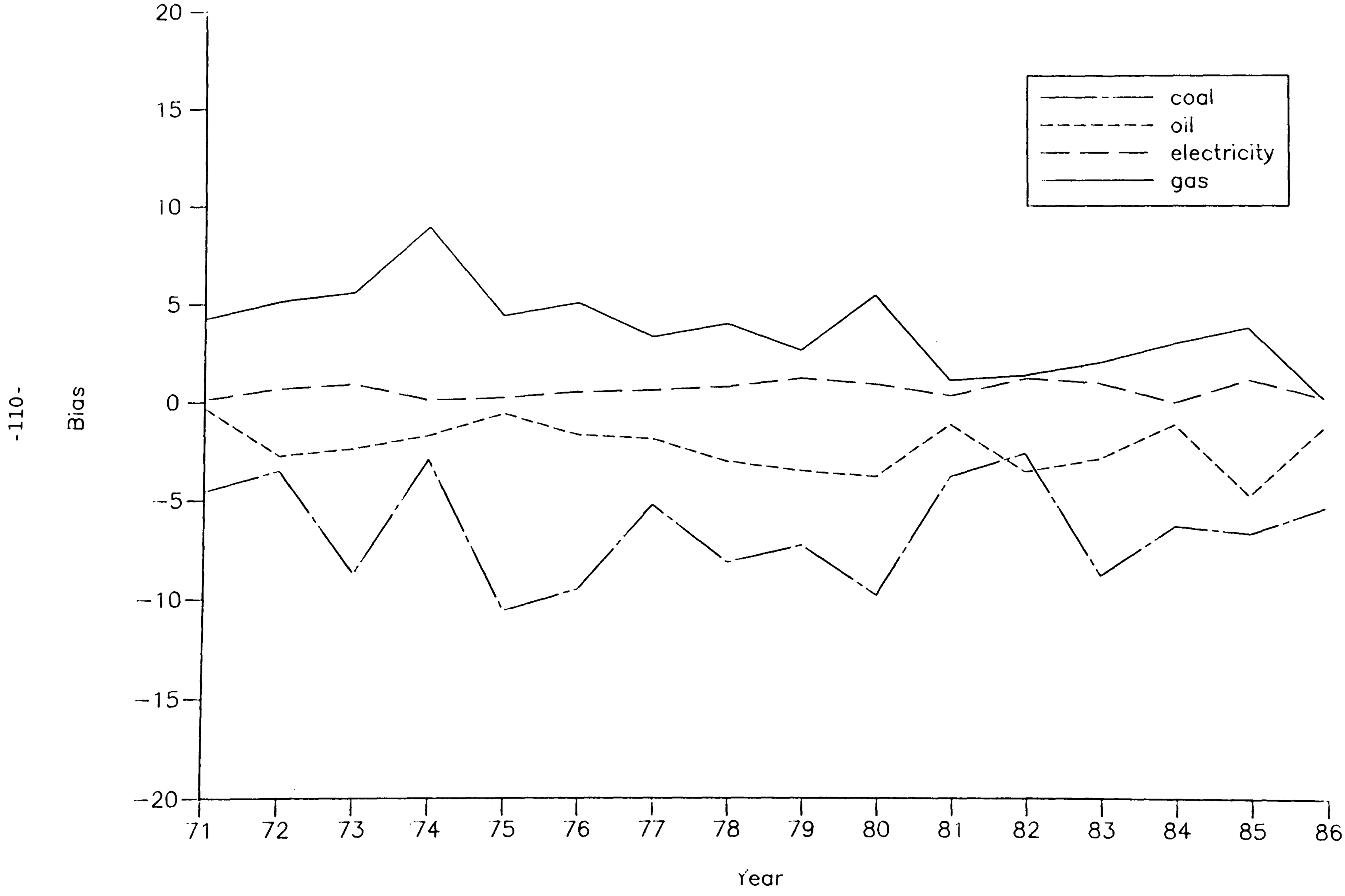
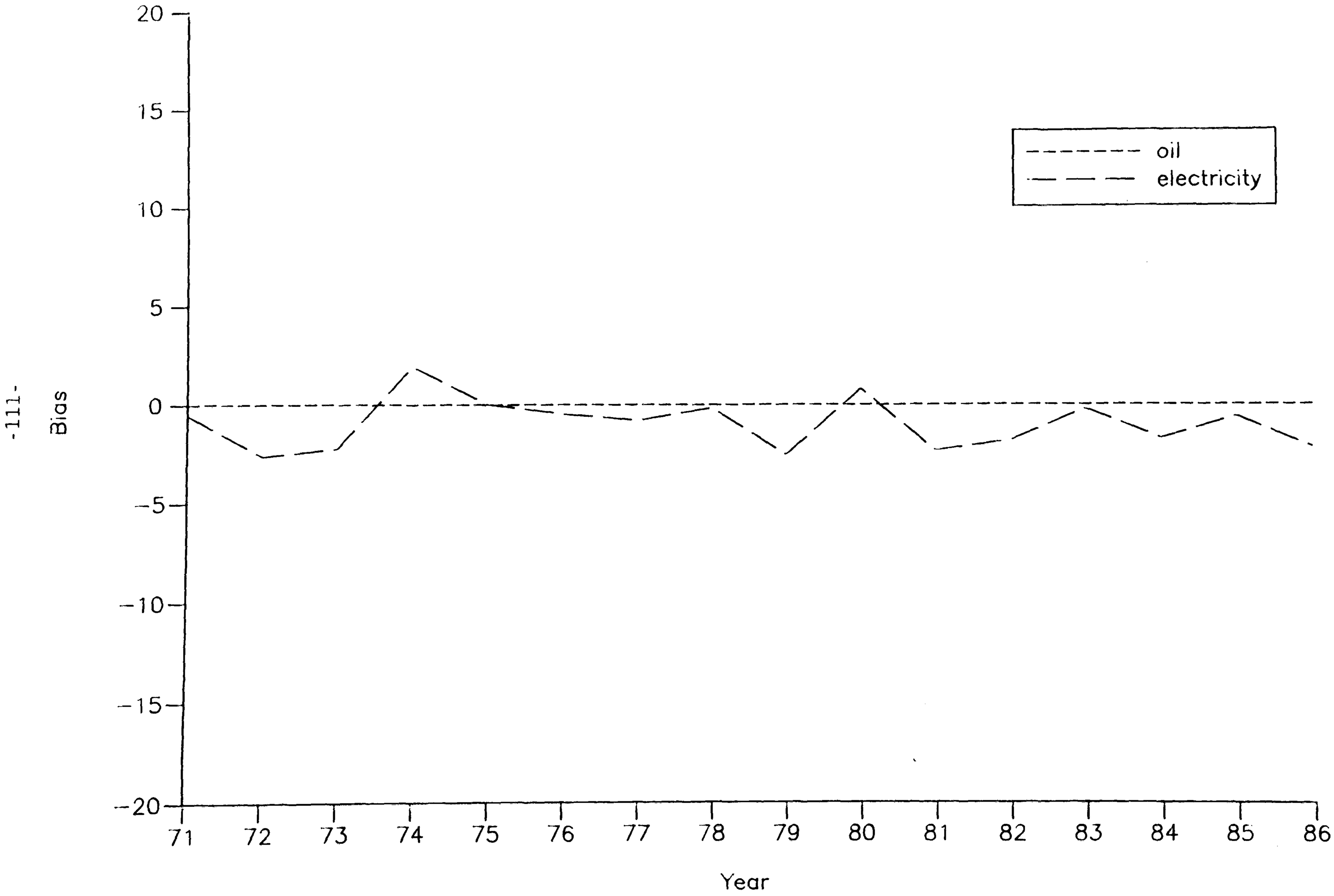


Figure 4.3.4 : Biases in Technical Progress : Transport Sector



4.4 Forecasting the Demand for Energy

This section presents the estimation results for the total cost of energy equations, and the forecasts of individual demands for the four economic sectors.

In each sector, the total cost of energy was decomposed into a price index and a quantity index, both in augmented form. The price index of energy was obtained from equation (2.17). For that, estimates of the augmentation factors A_{it} are needed, and these were computed using the smoothed estimates of the trends μ_t defined in Section 4.3, and solving, at each time t, the system of equations (2.13a) for A_{it}, i = 1,...,n. To solve this system, the extra conditions (2.14) are used.

Figure 4.4.1 presents the annual value of the estimated price indexes (1971 Q1 = 100), in augmented form, for the four economic sectors.

Using the price index of energy for each economic sector, a quantity index in augmented form was computed from (2.15). This quantity index, which is in real terms and can be associated with other real variables, was modelled using a univariate structural time series model with output and temperature as exogenous variables, that is,

$$(4.1a) \qquad \ell n\, Q_{et}\, A_{et} = \theta_1\, x_{1t} + \theta_2\, x_{2t} + \mu_t + \gamma_t + \epsilon_t, \qquad t = 1,\ldots,T,$$

$$(4.1b) \qquad \mu_t = \mu_{t-1} + \beta + \eta_t,$$

$$(4.1c) \qquad (1 + L + L^2 + L^3)\, \gamma_t = \omega_t,$$

where $(\ell n\, Q_{et}\, A_{et})$ is the quantity index in augmented form, x_{1t} and x_{2t} are the output and temperature variables, μ_t and γ_t are the trend and seasonal component, L is the lag operator, and θ_1, θ_2 and β are fixed parameters. The random shocks ϵ_t, η_t and ω_t are assumed to be normally

distributed and mutually and serially independent, with expected value zero and variances σ_ϵ^2, σ_η^2 and σ_ω^2. The trend component μ_t in (4.1) can be partially associated with the neutral technical progress in energy A_{et}.

The model was estimated by maximum likelihood using the frequency domain approach, with data for the sample period 1971 Q1 to 1986 Q4. The results are reported in Table 4.4.1. The estimate of the output elasticity was .43 in the other industry sector, .63 in the economic sector, .29 in the other final users sector, and .35 in the transport sector. In all the cases, the estimates are significant at the 5% level. Temperature has a negative, and significant, effect on the total quantity of energy in all the sectors except in transport where the estimate of the coefficient is greater than zero and not clearly significant. The estimates of the slopes of the trends are greater than zero in the other industry and transport sectors, and less than zero in the domestic and other final users sectors. The estimates, however, are not very significant in any economic sector. We applied tests for normality, serial correlation and heteroscedasticity to the residuals of the four models. At the 5% significance level, all the tests supported the model.

Forecasts of the quantity index for 1987 were made from these models and combining these with a price index based on 1987 prices led to predictions for total costs. Predictions for individual demands were then obtained using also the forecasts of the shares, from the share equation system estimated in section 3. Table 4.4.2 presents these predictions for the four fuels in each economic sector. The errors, defined as the per cent deviations of the forecasts from the observed demands, fluctuate between -10.5% , for the gas fuel in the other industry sector, and 14.3% for the coal fuel in the other final users

sector. The average of the absolute value of the errors in all the sector, and for all the fuels, is 5.1%.

For comparative purposes, forecasts of the individual demands were also made from univariate structural time series models, which used no information on prices, output or temperature. Table 4.4.3 compares the average of the absolute value of the errors for the two sets of predictions. One and five year ahead predictions were made. The one year ahead predictions , which use information on demands up to 1986, show no significant differences between the forecasts obtained from the econometric model and the ones from univariate models. The five year ahead predictions, are based on the same parameter estimates, but are constructed using information on demands up to 1982 only. They reveal that in the medium term the predictions using the econometric model are more accurate. Although in the other industry sector the differences between the errors are small, for the other three economic sectors the errors from univariate models are more than twice as high as the ones obtained from the econometric model.

Figure 4.4.1 : Price Index of Energy (1971 Q1 = 100)

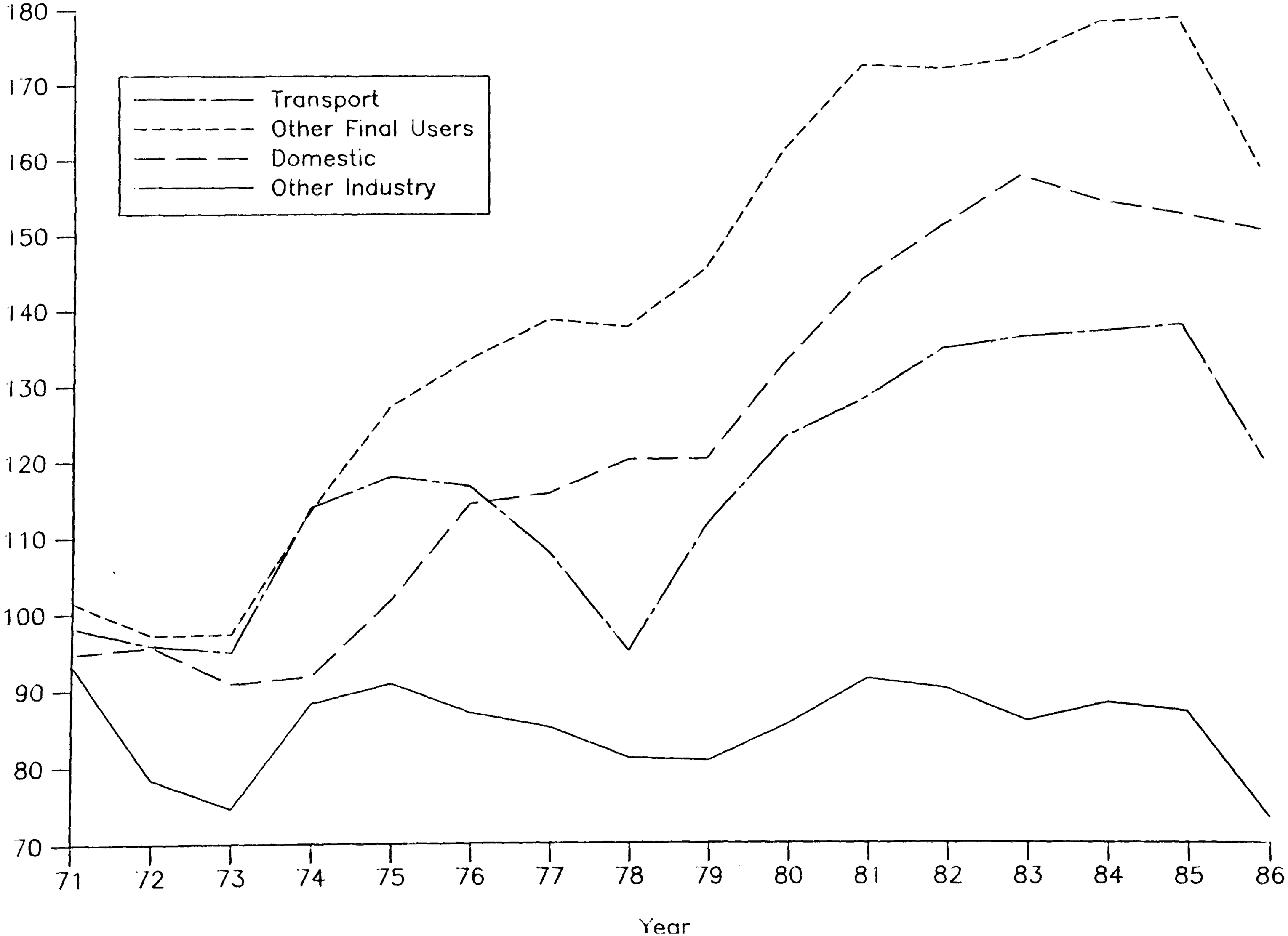


Table 4.4.1: Estimates in Total Quantity of Energy Models(1)

Parameter	Other Industry	Domestic	Other Final Users	Transport
θ_1(output)	.427	.632	.294	.353
	(.194)	(.254)	(.107)	(.151)
θ_2(temp.)	-.013	-.047	-.039	.003
	(.005)	(.005)	(.006)	(.002)
β(x10^{-2})	.410	-.544	-.158	.112
	(.366)	(.269)	(.090)	(.216)
σ^2_ϵ(x10^{-3})	.000	.245	.870	.014
	(-)	(.170)	(.212)	(.038)
σ^2_η(x10^{-3})	.835	.316	.044	.242
	(.218)	(.136)	(.035)	(.081)
σ^2_ω(x10^{-3})	.114	.107	.044	.003
	(.050)	(.055)	(.031)	(.002)

(1) Standard errors in parenthesis.

Table 4.4.2: Demand Forecasts for 1987(1)

Other Industry Sector	Gas	Elec.	Oil	Coal
Observed	4266	2602	3398	2017
Forecast	4714	2710	3294	2202
Error	-10.5	-4.2	3.1	-9.2
Domestic Sector	**Gas**	**Elec.**	**Oil**	**Coal**
Observed	10502	3183	980	2061
Forecast	10865	3515	982	1970
Error	-3.5	-10.4	-0.2	4.4
Other Final Users Sector	**Gas**	**Elec.**	**Oil**	**Coal**
Observed	2990	2370	1979	370
Forecast	3296	2375	1990	317
Error	-10.2	-0.2	-0.6	14.3
Transport sector	**Gas**	**Elec.**	**Oil**	**Coal**
Observed	-	105	16833	-
Forecast	-	105	16680	-
Error	-	0.0	0.9	-

(1) Error = 100(1 - Forecast/Observed).

Table 4.4.3 : Absolute Value of Prediction Errors for 1987(1)

Prediction	Other Industry	Domestic	Other Final Users	Transport
1 Year Ahead				
- Econometric	6.8	4.6	6.3	.5
- Univariate	5.3	5.3	5.4	.7
5 Year Ahead				
- Econometric	22.5	6.8	11.8	5.6
- Univariate	24.1	15.6	23.8	15.3

(1) See note (1) in Table 4.4.2

4.5 Conclusions

Section 4.2 of the chapter presented an econometric model to study the inter-fuel substitution possibilities between energy inputs. Using a translog cost function, and factor augmenting technology, the study made a contribution to the way in which the factor augmenting technical progress is modelled by using stochastic, rather than deterministic, time trends.

The econometric model led to a system of share equations. This system was estimated in Section 4.3 for four economic sectors: other industry, domestic, other final users and transport; for four fuels: gas, electricity, oil and coal; and using the sample period 1971 Q1 to 1986 Q4. The findings show significant price effects in the system of shares, although the resulting short term demand elasticities are, in general, quite small. Gas, electricity and oil are substitutes in all the sectors except in the other final users sector where electricity and oil were found to be complements. A study of the biases in technical progress for the four sectors, shows that at constant prices, the technology biases the use of fuels towards gas and against coal in all the sectors. The biases for electricity are positive in the other industry and other final users sectors and negative in the other two. For the oil fuel, the biases are positive in the transport sector only.

Two comments are necessary in connection with the econometric model. The first is that the use of only current values of prices means that the price effects captured by the model are short term. Over the longer term changes in prices will lead to changes in demand as consumers switch to using equipment which is appropiate for the fuels which have become relatively cheaper. Such changes could take several years to come into effect and building appropiate lag structures into

the model is extremely difficult given the limited amounts of data available. The omission of lagged prices from the model means that some of the long term price effects will be absorbed by the stochastic trends, and this should be borne in mind in interpreting the results. The second comment concerns the economic and statistical restrictions imposed on the model. In some cases the tests for the validity of these restrictions indicated a rejection of the null hypothesis, albeit very marginally, at the conventional 5% level of significance. The decision to impose the restrictions could therefore lead to some distortion on the estimates, but we believe that the attendent parsimony more than justifies this decision.

The estimation of a model for the total quantity of energy, using output and temperature as exogenous variables, and the computation of a price index of energy enabled predictions of individual demands to be made for 1987. These predictions were compared with the ones obtained from univariate structural time series models. The magnitudes of the errors were found to be very similar for the two sets of one year ahead predictions. However, for the five year ahead predictions, the econometric model was clearly superior.

Appendix 4.1 : Definition of the Data

The demands are in millions of therms, and adjusted to be on the SICC 1980 definition.

The prices are in units of pence per therm in 1975 money, deflacted by the Retails Price Index in the domestic and transport sectors and by the Producers Price Output Index in the other industry and other final users sectors. The prices of fuels in the other final users are assumed equal to the prices in the other industry sector. The price of oil in the other industry and other final users sectors is taken as the heavy fuel oil while the price of oil in the transport sector is based on a weighted average of four star petrol (weight .75) and DERV (weight .25). The price of all the other fuels are taken from published data in Energy Trends and in Digest of UK Energy Statistics.

The variable output is taken as the consumer expenditure in billion of pounds in 1980 money for the domestic and transport sectors. In the other industry and other final users sectors, the variable output is defined as the Manufacturing Output Index (1980 = 100).

The temperature variable is defined in terms of deviations from the seasonal mean and is in degrees Celsius.

Some observations on demands and prices were identified as outliers and removed prior to the analysis.

Appendix 4.2 : Constrained Dependent Variables

Consider the following multivariate structural time series model

$$(A2.1a) \qquad y_t = B\, x_t + \mu_t + \gamma_t + \epsilon_t, \qquad t = 1,\ldots,T,$$

$$(A2.1b) \qquad \mu_t = \mu_{t-1} + \beta + \eta_t,$$

$$(A2.1c) \qquad S(L)\, \gamma_t = \omega_t,$$

where y_t is an (n x 1) vector of dependent variables, B is an (n x k) matrix of fixed coefficients and x_t is a (k x 1) vector of exogenous variables. The (n x 1) vectors μ_t, β and γ_t represent the level, slope and seasonal components, with the seasonal period equal to s. The polynomial S(L) in the lag operator L is defined as in Section 1.2. Finally, the random shocks ϵ_t, η_t and ω_t are assumed serially and mutually uncorrelated, normal, with expected values equal to zero and variance covariance matrices Σ_ϵ, Σ_η and Σ_ω respectively.

Consider the following restriction on the dependent variable vector y_t:

$$(A2.2) \qquad \lambda' y_t = a, \qquad t = 1,\ldots,T,$$

where a is a fixed constant. Using (A2.1a),

$$(A2.3) \qquad \lambda' B\, x_t + \lambda' \mu_t + \lambda' \gamma_t + \lambda' \epsilon_t = a, \qquad t = 1,\ldots,T.$$

As x_t, μ_t, γ_t and ϵ_t are mutually uncorrelated, necessarily

$$(A2.4a) \qquad \lambda' B\, x_t = a_1, \qquad t = 1,\ldots,T,$$

$$(A2.4b) \qquad \lambda' \mu_t = a_2, \qquad t = 1,\ldots,T,$$

$$(A2.4b) \qquad \lambda' \gamma_t = a_3, \qquad t = 1,\ldots,T,$$

and

(A2.4c) $\lambda' \epsilon_t = a_4, \qquad t = 1,\ldots,T,$

with $a = a_1 + a_2 + a_3 + a_4$. Clearly (A2.4a) cannot hold if x_t changes with t unless $a_1 = 0$ and $\lambda'B = 0$. Similarly, $a_3 = a_4 = 0$ because the expected value of ω_t and ϵ_t are zero. That gives $a_2 = a$. Premultiplying (A2.1b) by λ', and with a similar argument, we obtain $\lambda'\beta = 0$ and $\lambda'\eta_t = 0$.

We have then that the variance covariance matrices of all the random shocks, and the prediction error variance are of rank (n-1). Following the argument in Cramer (1986, ch.7), the likelihood of the complete system is equivalent to the likelihood of any (n-1) components of the vector y_t.

After the estimation of the parameters for the (n-1) dimensional model, the original variance covariance matrices of the random shocks and the full matrix B can be recovered using the relations

(A2.5a) $\lambda' B = 0,$

(A2.5b) $\lambda' \beta = 0,$

and

(A2.5c) $\lambda' \Sigma_i = 0, \qquad i = \epsilon, \eta \text{ and } \omega.$

CHAPTER 5 : DYNAMIC ERROR COMPONENTS MODELS

5.1 Introduction and Formulation

This chapter considers the formulation and the basic properties of dynamic error components models based on the ideas of structural time series models. This first section presents some basic ideas and formulates the models. Section 5.2 studies basic statistical properties of the models; and sections 5.3 and 5.4 extend the specifications to multivariate observations and factor analysis respectively.

Error components models have been used to analyse data collected by observing a number of individuals or units over time, usually called panel data, since the early work by Balestra and Nerlove (1966). Suppose for start that the random variable α_{it} is observed over each unit $i = 1,\ldots,n$ and time $t = 1,\ldots,T$. The basic idea behind the error components models is that the random effects acting over α_{it} can be separated into three independent components: a unit specific effect which is the same for all times, a time specific effect which is the same for all units, and a time unit specific effect. The standard static model considered in the literature, without exogenous variables, is

$$\alpha_{it} = \lambda_i^{(1)} + \lambda_t^{(2)} + \lambda_{it}^{(3)}, \qquad \begin{matrix} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{matrix} \tag{1.1}$$

where the three stochastic components $\lambda_i^{(1)}$, $\lambda_t^{(2)}$ and $\lambda_{it}^{(3)}$ are assumed to be normally distributed, mutually and serially uncorrelated, with expected values equal to zero and variances σ_1^2, σ_2^2 and σ_3^2 respectively. Several generalisations or modifications can be introduced to (1.1). In a two error components model, either the effect $\lambda_i^{(1)}$ or $\lambda_t^{(2)}$ is not present; while in a fixed effect components model, either $\lambda_i^{(1)}$ or $\lambda_t^{(2)}$

is fixed rather than stochastic. Finally, the components $\lambda_{it}^{(3)}$ might have unit specific variances. Hsiao (1986) presents a review of the basic methods that have been used in the literature.

The dynamic version of (1.1) includes, in general, lagged dependent variables and autoregressive or more general ARMA structures for $\lambda_t^{(2)}$ and $\lambda_{it}^{(3)}$. Maddala (1971), Nerlove (1971), Trognon (1978) and Anderson and Hsiao (1981, 1982) studied different aspects of a model like (1.1) with lagged dependent variables. Lillard and Willis (1978) used (1.1) with $\lambda_t^{(2)} = 0$ and $\lambda_{it}^{(3)}$ following an AR(1) process, while Revankar (1979) studied the case where $\lambda_t^{(2)}$ follows an AR(1) process. Similar specifications were considered by Lillard and Weiss (1979), Hause (1980) and Anderson and Hsiao (1982). MaCurdy (1982) considered a more general time series process for the component $\lambda_{it}^{(3)}$.

Here, the approach used to transform (1.1) into a dynamic model is based on the ideas of structural time series models. To illustrate the concepts and facilitate the exposition in this introduction, the local level time series model defined in Section 1.2 is used. Several generalisations are considered later.

Allowing a specification like (1.1) for both the irregular and level random shocks in the local level model (1.2.1), yields

$$\alpha_{it} = \mu_{it} + \epsilon_t + \epsilon_{it}^*, \qquad i = 1,\ldots,n, \quad t = 1,\ldots,T, \tag{1.2a}$$

$$\mu_{it} = \mu_{i,t-1} + \eta_t + \eta_{it}^*, \tag{1.2b}$$

where the components ϵ_t, ϵ_{it}^*, η_t and η_{it}^* are assumed to be normally distributed, mutually and serially uncorrelated, with expected values equal to zero and variances σ_ϵ^2, σ_ϵ^{2*}, σ_η^2 and σ_η^{2*} respectively. The random shocks ϵ_t and η_t are common to all units, while ϵ_{it}^* and η_{it}^* are unit specific random shocks. The component $\lambda_i^{(1)}$, defined in (1.1), does

not appear in (1.2) because it has already been included in the stochastic unit specific trend μ_{it}.

Model (1.2) can be seen as a restricted n dimensional local level model, with the variance covariance matrices of the irregular and level random shocks equal to $(\sigma_\epsilon^2 \ \iota\iota' + \sigma_\epsilon^{2*} \ I_n)$ and $(\sigma_\eta^2 \ \iota\iota' + \sigma_\eta^{2*} \ I_n)$ respectively; with ι an (n x 1) vector of ones and I_n the identity matrix of dimension n. Thus, the error components model (1.2) imposes strong restrictions on the variance covariance matrices of the irregular and level random shocks. In the following subsections, where more general specifications are considered, some of these restrictions are relaxed and the unit specific random shocks ϵ_{it}^* and η_{it}^* are allowed to have unit specific variances.

Some comments with respect to the specification (1.2) are in order. When $\sigma_\eta^2 = \sigma_\eta^{2*} = 0$, (1.2) reduces to the static model (1.1); and when $\sigma_\eta^2 \searrow 0$ but $\sigma_\eta^{2*} = 0$, the n time series are cointegrated in the sense of Engle and Granger (1987). The distinction between the cases $\sigma_\eta^{2*} \searrow 0$ and $\sigma_\eta^{2*} = 0$ is important not only in terms of the behavior of the trends μ_{it} in the long run; but also because introduces significant modifications in the way the model should be handle and estimated. That is the reason why two kinds of error components models are defined. In the model type I, the trends and seasonal components are generated by both a random shock common to all units and a unit specific random shock whose variance is greater than zero; that is, in (1.2), $\sigma_\eta^{2*} \searrow 0$. On the other hand, in the model type II, the trend and seasonal components are, apart from a time invariant effect, the same for all units; that is, in (1.2), $\sigma_\eta^{2*} = 0$.

Two kinds of panel data sets have been considered in the literature. In one case the number of units n is large and the number of time periods T is typically small; the inference in this case is

usually based on n going to infinity with T fixed or going to infinity as well. In the second kind of data sets, which is the one in which we are most interested here, n is fixed although, in general, large and the inference is based on T going to infinity.

The remainder of this section defines more formally the dynamic error components models type I and II. Those definitions consider some generalisations: stochastic slopes and seasonal effects are incorporated into (1.2), and the unit specific random shocks are allowed to have unit specific variances. Finally, the components α_{it} are defined as residuals of a regression model rather than observed values.

Model Type I

This subsection defines a dynamic error components model type I where (1.2) is extended to include stochastic slopes and seasonal components and the variances of the unit specific random shocks are allowed to be unit specific. The general error components model type I is defined as

$$\alpha_{it} = Z\,\theta_{it} + \epsilon_t + \epsilon^*_{it}, \qquad i = 1,\ldots,n, \quad t = 1,\ldots,T, \tag{1.3a}$$

$$\theta_{it} = T\,\theta_{i,t-1} + R\,(\kappa_t + \kappa^*_{it}), \tag{1.3b}$$

where α_{it} is the observation for the unit i at time t; and Z, T and R are (1 x p), (p x p) and (p x u) matrices as defined, jointly with the values p and u, in Section 1.2 according with the time series model. The (p x 1) vector θ_{it} corresponds to the state vector which contains the trend and seasonal components; and the random shocks ϵ_t, ϵ^*_{it}, κ_t and κ^*_{it} have dimensions 1, 1, u, and u respectively; and they are

assumed to be normally distributed, serially and mutually uncorrelated, with expected values equal to zero and variances σ_ϵ^2, $\sigma_{\epsilon i}^{2*}$, D_κ and $D_{\kappa i}^*$; the last two being (u x u) diagonal matrices. The stationary form of the model, which is obtained by taking differences, is assumed to be strictly invertible; see Section 1.2. That restricts some elements of $D_{\kappa i}^*$ to be strictly positive. In the local level model, p = u = 1, $D_\kappa = \sigma_\eta^2$ and $D_{\kappa i}^* = \sigma_{\eta i}^{2*} \succ 0$. In the local linear trend model, p = u = 2, $D_\kappa = \text{diag}(\sigma_\eta^2 , \sigma_\delta^2)$ and $D_{\kappa i}^* = \text{diag}(\sigma_{\eta i}^{2*} , \sigma_{\delta i}^{2*})$ with $\sigma_\delta^{2*} \succ 0$. In the seasonal local level model, u = 2, $D_\kappa = \text{diag}(\sigma_\eta^2 , \sigma_\omega^2)$ and $D_{\kappa i}^* = \text{diag}(\sigma_{\eta i}^{2*} , \sigma_{\omega i}^{2*}) \succ 0$. Finally, in the basic structural model, u = 3, $D_\kappa = \text{diag}(\sigma_\eta^2, \sigma_\delta^2, \sigma_\omega^2)$ and $D_{\kappa i}^* = \text{diag}(\sigma_{\eta i}^{2*}, \sigma_{\delta i}^{2*}, \sigma_{\omega i}^{2*})$, with $\sigma_{\delta i}^{2*} \succ 0$ and $\sigma_{\omega i}^{2*} \succ 0$.

Model (1.3) can be written as a multivariate structural time series model of the form (1.1.3). That is,

$$(1.4a) \qquad \alpha_t = (Z \otimes I_n)\ \theta_t + \iota\ \epsilon_t + \epsilon_t^*, \qquad t = 1,\ldots,T,$$

$$(1.4b) \qquad \theta_t = (T \otimes I_n)\ \theta_{t-1} + (R \otimes I_n)\ ((I_u \otimes \iota)\ \kappa_t + \kappa_t^*),$$

where α_t is an (n x 1) vector with i-th component α_{it}, θ_t is the (np x 1) state vector, and ι is an (n x 1) vector of ones. The random shocks ϵ_t and κ_t are defined as in (1.3); while ϵ_t^* and κ_t^* are (n x 1) and (nu x 1) vectors with i-th components ϵ_{it}^* and κ_{it}^* respectively. The variance covariance matrices of ϵ_t^* and κ_t^* are defined as the (n x n) and (nu x nu) diagonal matrices D_ϵ^* and D_κ^* respectively. With these definitions, the variance covariance matrices of the random shocks in the measurement equation (1.4a) and in the transition equation (1.4b) can be written as

$$(1.5a) \qquad V(\iota\ \epsilon_t + \epsilon_t^*) = \iota\iota'\ \sigma_\epsilon^2 + D_\epsilon^*, \qquad t = 1,\ldots,T,$$

and

(1.5b) $$V((\kappa_t \otimes \iota) + \kappa_t^*) = (D_\kappa \otimes \iota\iota') + D_\kappa^*, \qquad t = 1,\ldots,T.$$

Model (1.4), with the variance covariances matrices of the random shocks as in (1.5), represents a very general specification where no restrictions are placed on the parameters of the models for each unit, but only on the parameters that capture the covariance structure across units.

The error components model type I is said to be restricted if it satisfies

(1.6a) $$D_\epsilon^* = D\ \sigma_\epsilon^{2*},$$

(1.6b) $$D_\kappa^* = D_\kappa^{**} \otimes D,$$

where D and D_κ^{**} are (n x n) and (u x u) diagonal matrices, and one restriction should be imposed for identifiability. Restriction (1.6) says that the unit specific variance covariance matrices of the random shocks of the level, trend and seasonal components, in κ_t^*, are proportional to the variance covariance matrix of the irregular random shock ϵ_t^*. This is exactly the idea used in the homogeneous model defined in Section 1.1; however, model (1.4) is not homogeneous under (1.6) because no restrictions have been introduced to the common random shocks ϵ_t and κ_t. A more detailed discussion of the relationship between the restriction (1.6) and the homogeneity restriction is presented in Section 5.3. As a special case of (1.6), the matrix D may be known or it may be set equal to the identity. Of course, (1.4) under (1.6) and $D = I_n$ reduces to (1.2) for the local level model.

Section 5.2 considers some basic statistical properties of the models considered here. Chapter 6 studies the minimum mean square error estimation of the unobserved components in the state vector θ_t in (1.4), and the maximum likelihood estimation of the parameters.

Model Type II

This subsection defines the error component model type II, which is a special case of the model type I defined above in that the unit specific random shocks which generate the trend and seasonals components are zero for all t = 1,...,T. Thus, in terms of the definitions in the previous subsection, $D_{\kappa}^{*} = 0$ and the trend and seasonal components are generated by the common to all units random shock κ_t. The general error components model type II is defined as

$$\alpha_{it} = Z\ \theta_{it} + \epsilon_t + \epsilon_{it}^{*}, \qquad i = 1,\ldots,n, \quad t = 1,\ldots,T, \tag{1.7a}$$

$$\theta_{it} = T\ \theta_{it-1} + R\ \kappa_t, \tag{1.7b}$$

where all the components in (1.7) are defined as in (1.3). In model (1.7), $\sigma_{\epsilon i}^{2*}$ is assumed to be greater than zero for all i and according to the time series model, some of the variances in D_{κ} are also assumed to be strictly positive. In the local level model $\sigma_{\eta}^{2} > 0$, in the local linear trend $\sigma_{\delta}^{2} > 0$, in the seasonal local level model $\sigma_{\eta}^{2} > 0$ and $\sigma_{\omega}^{2} > 0$, and in the basic structural model $\sigma_{\delta}^{2} > 0$ and $\sigma_{\omega}^{2} > 0$.

Model (1.7) can be written as a multivariate structural time series model of the form (1.1.3). That is,

$$\alpha_t = (Z \otimes I_n)\ \theta_t + \iota\ \epsilon_t + \epsilon_t^{*}, \qquad t = 1,\ldots,T, \tag{1.8a}$$

$$\theta_t = (T \otimes I_n)\ \theta_{t-1} + (R \otimes I_n)\ (I_u \otimes \iota)\ \kappa_t, \tag{1.8b}$$

where all the components in (1.8) are defined as in (1.4). The variance covariance matrices of the random shocks in the measurement equation (1.8a) and in the transition equation (1.8b) are given by

$$V(\iota\ \epsilon_t + \epsilon_t^{*}) = \iota\iota'\ \sigma_{\epsilon}^{2} + D_{\epsilon}^{*}, \qquad t = 1,\ldots,T, \tag{1.9a}$$

and

$$V(\kappa_t \otimes \iota) = D_\kappa \otimes \iota\iota', \qquad t = 1,\ldots,T. \tag{1.9b}$$

Although the trend and seasonal components contained in the state vector θ_{it} in (1.7) are generated by the common to all units random shocks κ_t, this does not mean that the trend and seasonal components are the same for all units because there may be initial differences between these trend and seasonal effects. Two alternative specifications for these initial differences are considered. First, the initial differences between the trend and seasonal components may be defined as fixed parameters; or second, the initial differences may be defined as random effects with a given prior distribution. Notice that these alternative specifications refer only to the differences between the components across units. Notice also that if $\kappa_t = 0$ in the local level model, (1.7) reduces to the static model (1.1) and the above two alternative specifications coincide with the two standard specifications in (1.1): the unit specific effect may be defined as fixed or random. The approach here is a straightforward generalisation.

Whichever the specification for the initial differences between the trend and seasonal components across units, the state vector for the unit i at time 0 can be written as

$$\theta_{i0} = \lambda_i + \theta_0^*, \qquad i = 1,\ldots,n, \tag{1.10}$$

where λ_i is a (p x 1) vector representing the unit specific effect, and θ_0^* is a (p x 1) vector which contains the common to all units components. Hence, at time t

$$\theta_{it} = T^t \lambda_i + \theta_t^*, \tag{1.11a}$$

where

(1.11b) $$\theta_t^* = T\ \theta_{t-1}^* + R\ \kappa_t,$$

and then equation (1.8a) can be written as

(1.12) $$\alpha_t = (Z\ T^t \otimes I_n)\ \lambda + \iota\ Z\ \theta_t^* + \iota\ \epsilon_t + \epsilon_t^*$$

$$= U_t\ \lambda + \iota\ Z\ \theta_t^* + \iota\ \epsilon_t + \epsilon_t^*,$$

where λ is an (np x 1) vector with p blocks of n elements each. These blocks of n elements correspond to the unit specific effects for the trend and seasonal components. It follows that the error components model type II can alternatively be written as in (1.12) and (1.11b). Notice that the state vector θ_t^* in (1.12) is only identifiable up to a time invariant effect because of the presence of λ. In the fixed effects model, and without loss in generality, the sum of the components λ_i across units may be set equal to zero. In that case the state vector θ_t^* is equal to the average of the state vectors θ_{it} in (1.7b).

The error components model type II is said to be restricted if it satisfies

(1.13) $$D_\epsilon^* = I_n\ \sigma_\epsilon^{2*},$$

with $\sigma_\epsilon^{2*} \searrow 0$. Thus, all the units have the same variance for the irregular unit specific random shock.

Section 5.2 considers some basic statistical properties of the model presented here; while Chapter 7 deals with the estimation of the unobserved components and the maximum likelihood estimation of the parameters.

Exogenous Variables

Until now, it has been assumed that the components α_{it} are observations. Consider the more general situation where the α_{it} are the residuals of a regression model of the form

$$y_{it} = z_{it}' \beta_i + \alpha_{it}, \qquad \begin{matrix} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{matrix} \tag{1.14}$$

where y_{it} is the dependent variable, z_{it} is an (r x 1) vector of exogenous variables, and β_i is an (r x 1) vector of fixed coefficients. Three special cases of (1.14), which are typically considered in the analysis of error components models, are the following

$$y_{it} = z_t' \beta_i + \alpha_{it}, \tag{1.15a}$$

$$y_{it} = z_{it}' \beta + \alpha_{it}, \tag{1.15b}$$

and

$$y_{it} = z_t' \beta + \alpha_{it}. \tag{1.15c}$$

In (1.15a), the exogenous variables are the same for all units, although the coefficients of these exogenous variables are unit specific. In (1.15b), the exogenous variables are time unit specific but the vector of coefficients is the same for all units. Finally, in (1.15c), both the exogenous variables and the vector of coefficients are the same for all units. In multivariate form,

$$y_t = X_t \beta + \alpha_t \tag{1.16}$$

where y_t is the (n x 1) vector of observations, X_t is an (n x k) matrix of exogenous variables, β is a (k x 1) vector of coefficients, and α_t is the (n x 1) vector of residuals. Under (1.15a), k = n r, $X_t = (I_n \otimes$

z_t') and the i-th component of β is β_i. Under (1.15b), k = r and the i-th row of X_t is z_{it}'. Finally, under (1.15c), k = r and $X_t = (\iota \otimes z_t')$, with ι a vector of ones.

5.2 Basic Properties of the Models

This section considers some basic properties of the two error components models defined in the previous section. Although the results presented in Section 1.2 for the unrestricted multivariate structural time series model can be applied here with obvious substitutions, this section presents results which are specific to the class of models studied in this chapter. Basic statistical properties of the error components model type I, under (1.6) and assuming that the diagonal matrix D defined there is equal to the identity matrix, are considered first. Results for the model type II under (1.13) follow.

Model Type I

The autocovariance function of the first differences in the local level error components model as defined in (1.4), and under (1.6) and D $= I_n$, is given by

$$\Gamma(0) = (2\,\sigma_\epsilon^2 + \sigma_\eta^2)\;\iota\iota' + (2\,\sigma_\epsilon^{2*} + \sigma_\eta^{2*})\;I_n, \tag{2.1a}$$

$$\Gamma(\pm 1) = -\,\sigma_\epsilon^2\;\iota\iota' - \sigma_\epsilon^{2*}\;I_n, \tag{2.1b}$$

and

$$\Gamma(\pm k) = 0, \qquad \text{if } k \geq 1, \tag{2.1c}$$

where ι is an (n x 1) vector of ones and I_n is the identity matrix of order n. From (2.1), the autocorrelation matrices at lags zero and one are given by

$$R(0) = \frac{1}{\varphi}\;[(2\,\sigma_\epsilon^2 + \sigma_\eta^2)\;\iota\iota' + (2\,\sigma_\epsilon^{2*} + \sigma_\eta^{2*})\;I_n], \tag{2.2a}$$

and

$$(2.2b) \qquad R(1) = \frac{1}{\varphi} \left[- \sigma_\epsilon^2 \, \iota\iota' - \sigma_\epsilon^{2*} \, I_n \right],$$

respectively; where $\varphi = (2 \, \sigma_\epsilon^2 + 2 \, \sigma_\epsilon^{2*} + \sigma_\eta^2 + \sigma_\eta^{2*})$. That is, the matrices R(0) and R(1) have the same form of the variance covariance matrices. The correlation coefficients at lag one are the same for any pair of units. Also, as σ_ϵ^2 tends to zero, R(1) approaches a matrix proportional to the identity, and as σ_ϵ^{2*} tends to zero, R(1) approaches a matrix proportional to $\iota\iota'$. It can be shown, using the results in Section 1.2, that the same kind of structure is found in the autocorrelation matrices of the more general models with stochastic slopes and seasonal components. For example, in the local linear trend model, the autocorrelation matrices at lags zero, one and two are given by

$$(2.3a) \qquad R(0) = \frac{1}{\varphi} \left[(6 \, \sigma_\epsilon^2 + 2 \, \sigma_\eta^2 + \sigma_\delta^2) \, \iota\iota' \right.$$

$$\left. + (6 \, \sigma_\epsilon^{2*} + 2 \, \sigma_\eta^{2*} + \sigma_\delta^{2*}) \, I_n \right],$$

$$(2.3b) \qquad R(1) = \frac{1}{\varphi} \left[- (4 \, \sigma_\epsilon^2 + \sigma_\eta^2) \, \iota\iota' - (4 \, \sigma_\epsilon^{2*} + \sigma_\eta^{2*}) \, I_n \right],$$

and

$$(2.3c) \qquad R(2) = \frac{1}{\varphi} \left[\sigma_\epsilon^2 \, \iota\iota' + \sigma_\epsilon^{2*} \, I_n \right],$$

respectively; where $\varphi = (6 \, \sigma_\epsilon^2 + 6 \, \sigma_\epsilon^{2*} + 2 \, \sigma_\eta^2 + 2 \, \sigma_\eta^{2*} + \sigma_\delta^2 + \sigma_\delta^{2*})$. The autocovariance generating function and the spectral density of these models can be obtained using the formulas above and the general results

in Section 1.2. Under the assumptions in Section 5.1, the autocovariance generating function is always positive definite and then the models considered here are strictly invertible.

The general model type I was defined in (1.4). The following lines consider an alternative expression for that model which presents some advantages in the estimation of the components and parameters. It is still assumed here that (1.6) holds with $D = I_n$. Consider the (n x n) matrix H defined as

$$(2.4) \qquad H = (1/n) \begin{bmatrix} 1 & . & . & . & . & . & . & 1 \\ n-1 & -1 & . & . & . & . & . & -1 \\ -1 & n-1 & -1 & . & . & . & . & -1 \\ . & & & . & & & & . \\ -1 & & . & & . & -1 & n-1 & -1 \end{bmatrix},$$

whose determinant is (1/n), and its inverse is given by

$$(2.5) \qquad H^{-1} = \begin{bmatrix} 1 & & & & & \\ 1 & & & I_{n-1} & & \\ . & & & & & \\ 1 & -1 & . & . & . & -1 \end{bmatrix}.$$

The premultiplication of an (n x 1) vector by this matrix H produces a transformed vector whose first component is the average of the elements in the original vector and the remaining (n-1) components correspond to deviations of the first (n-1) components of the original vector with respect to this average. Premultiplying (1.4a) by H yields

$$(2.6a) \qquad H\,\alpha_t = H\,(Z \otimes I_n)\,\theta_t + H\,(\iota\ \epsilon_t + \epsilon_t^*)$$

$$= (Z \otimes I_n)\,(I_p \otimes H)\,\theta_t + H\,(\iota\ \epsilon_t + \epsilon_t^*),$$

where, using (1.4b),

$$(2.6b) \qquad (I_p \otimes H)\,\theta_t = (T \otimes I_n)\,(I_p \otimes H)\,\theta_{t-1} +$$

$$(R \otimes I_n)\,(I_p \otimes H)\,((I_u \otimes \iota)\,\kappa_t + \kappa_t^*).$$

Thus, the transformation produces a structural time series model as defined in Section 1.1, where the new observations are (H α_t) and the state vector is given by ($I_p \otimes H$) θ_t. The advantage of this transformation is that the variance covariance matrices of the transformed random shocks H ($\iota\ \epsilon_t + \epsilon_t^*$) and ($I_p \otimes H$) (($I_u \otimes \iota$) $\kappa_t + \kappa_t^*$) have simple expressions. In fact,

$$(2.7a) \qquad V\left[H\ (\iota\ \epsilon_t + \epsilon_t^*)\right] = B_\epsilon,$$

and

$$(2.7b) \qquad V\left[(I_p \otimes H)\ ((I_u \otimes \iota)\ \kappa_t + \kappa_t^*)\right] = \mathrm{diag}[B_1, \ldots, B_p],$$

where B_k has the form

$$(2.8) \qquad B_k = \begin{bmatrix} \sigma_k^2 + \sigma_k^{2*}/n & 0 \\ 0 & (I_{n-1} - \iota\iota'/n)\ \sigma_k^{2*} \end{bmatrix},$$

and k takes some or all of the values ϵ, η, δ, ω. It can be seen from (2.7) and (2.8) that the first transformed time series, which corresponds to the average across units, is uncorrelated with the remaining (n-1); thus, the estimation of the state vector in that time series can be treated separately as a univariate model. The remaining (n-1) time series, which correspond to deviations from the average, form a homogeneous structural time series model in which each disturbance vector has a variance covariance matrix proportional to the known matrix (I_{n-1} - $\iota\iota'/n$); see Section 1.1 for the definition of a homogeneous model. As the transformation (2.6) is non singular, it can be exploited to estimate the unobserved components in the state vector and to form the likelihood of the model.

The univariate model for $\bar{\alpha}_t$, which corresponds to the first element

of the transformed observations, is given by

$$(2.9a) \qquad \bar{\alpha}_t = Z\,\bar{\theta}_t + \epsilon_t + \bar{\epsilon}^*_t, \qquad t = 1,\ldots,T,$$

$$(2.9b) \qquad \bar{\theta}_t = T\,\bar{\theta}_{t-1} + R\,(\kappa_t + \bar{\kappa}^*_t),$$

where, for any component x_{it}, $\bar{x}_t$ represents the average across units. The variances of the random shocks in (2.9a) and (2.9b) are given by $(\sigma^2_\epsilon + \sigma^{2*}_\epsilon/n)$ and $(D_\kappa + D^*_\kappa/n)$ respectively. If n is large, these variances approach σ^2_ϵ and D_κ respectively, and the state vector $\bar{\theta}_t$ in (2.9) becomes a vector with trend and seasonal components generated by the common to all units random shocks.

On the other hand, the model for the i-th deviation from the average is given by

$$(2.10a) \qquad (\alpha_{it} - \bar{\alpha}_t) = Z\,(\theta_{it} - \bar{\theta}_t) + (\epsilon^*_{it} - \bar{\epsilon}^*_t), \qquad t = 1,\ldots,T,$$

$$(2.10b) \qquad (\theta_{it} - \bar{\theta}_t) = T\,(\theta_{i,t-1} - \bar{\theta}_{t-1}) + R\,(\kappa^*_{it} - \bar{\kappa}^*_t).$$

The variances of the random shocks in (2.10a) and (2.10b) are given by $\sigma^{2*}_\epsilon(n-1)/n$ and $D^*_\kappa(n-1)/n$ respectively. The trend and seasonal components in $(\theta_{it} - \bar{\theta}_t)$ are independent of the common to all units random shocks, and if n is large, $(n-1)/n$ approaches unity, and $(\theta_{it} - \bar{\theta}_t)$ is generated by the unit specific random shocks.

The transformation (2.6) is attractive because it does not depend on the parameters of the model. However, it assumes (1.6) with $D = I_n$. If (1.6) holds but D is not equal to the identity matrix, and if without loss in generality the restriction $\iota' D^{-1} \iota = n$ is impossed, a transformation like (2.6) can still be applied with H replaced by the matrix

$$(2.11)\qquad H^{*} = (1/n)\begin{bmatrix} (1/d_1) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & (1/d_n) \\ (n-(1/d_1)) & -(1/d_2) & \cdot & \cdot & \cdot & \cdot & \cdot & & & -(1/d_n) \\ & \cdot & & & & & & & & \\ & & \cdot & & & & & & & \\ & & & \cdot & & & & & & \\ -(1/d_1) & \cdot & \cdot & \cdot & \cdot & \cdot & (n-(1/d_{n-1})) & & & -(1/d_n) \end{bmatrix},$$

where d_i is the i-th element in the diagonal of D. The determinant of H^{*} is $(n\ d_n)^{-1}$ and its inverse is given by

$$(2.12)\qquad H^{*-1} = \begin{bmatrix} 1 & & & & & \\ \cdot & & & I_{n-1} & & \\ \cdot & & & & & \\ 1 & -d_n/d_1 & \cdot & \cdot & \cdot & -d_n/d_{n-1} \end{bmatrix}.$$

The transformation (2.6) with H^{*} given by (2.11) gives a model where the variances of the random shocks are as in (2.7) but with matrices B_k given by

$$(2.13)\qquad B_k = \begin{bmatrix} \sigma_k^2 + \sigma_k^{2*}/n & 0 \\ 0 & (D^{*} - \iota\iota'/n)\ \sigma_k^{2*} \end{bmatrix},$$

where D^{*} is an (n-1 x n-1) diagonal matrix with i-th diagonal element equal to d_i.

Model Type II

The autocovariance and autocorrelation functions of the stationary form of the error components model type II can be obtained from those for the model type I with the unit specific variances of the trend and seasonal random shocks set equal to zero.

On the other hand, the transformation of the model by using the matrix H in (2.4) is particularly attractive. Indeed, premultiplying (1.12) by H yields

$$(2.14) \qquad H\,\alpha_t = H\,U_t\,\lambda + H\,\iota\,Z\,\theta_t^* + H\,(\iota\,\epsilon_t + \epsilon_t^*), \quad t = 1,\ldots,T,$$

where the first equation of this system is given by

$$(2.15a) \qquad \bar{\alpha}_t = Z\,T^t\,\bar{\lambda} + Z\,\theta_t^* + \epsilon_t + \bar{\epsilon}_t^*, \qquad t = 1,\ldots,T,$$

and the last (n-1) equations have the form

$$(2.15b) \qquad (\alpha_{it} - \bar{\alpha}_t) = Z\,T^t\,(\lambda_i - \bar{\lambda}) + (\epsilon_{it}^* - \bar{\epsilon}_t^*), \quad \begin{matrix} i = 1,\ldots,n-1, \\ t = 1,\ldots,T. \end{matrix}$$

The component $\bar{\lambda}$ in (2.15a) can be incorporated into the state vector θ_t^* to produce the univariate model

$$(2.16a) \qquad \bar{\alpha}_t = Z\,(\theta_t^* + \bar{\lambda}) + \epsilon_t + \bar{\epsilon}_t^*, \qquad t = 1,\ldots,T,$$

$$(2.16b) \qquad (\theta_t^* + \bar{\lambda}) = T\,(\theta_{t-1}^* + \bar{\lambda}) + R\,\kappa_t,$$

which has the standard form of a univariate structural time series model with the variance of the irregular random shock equal to $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n)$. Notice that equations (2.16) reveal the fact that θ_t^* is not in fact identifiable. The transformation (2.14) leads to a model where the trend and seasonal components are estimated from the univariate model (2.16) while (2.15b) is a multivariate stationary model. Although the above results were obtained assuming, if (1.13) does not hold, the main ideas here can still be applied with the transformation based on H^* instead of H; see equation (2.11).

5.3 Extension to Multivariate Observations

This section considers the extension of the models already presented in this chapter to the case where α_{it} is a (q x 1) vector rather than a scalar. A multivariate error components model was first estimated by Chamberlain and Griliches (1975). Magnus (1982) presents a full treatment of the static model, while a dynamic model with first order serial correlation over both the time specific and time unit specific effects was considered by Magnus and Woodland (1987). The dynamic error components model type I, defined in (1.3), is naturally extended to multivariate observations. The vector random shocks ϵ_t, ϵ^*_{it}, κ_t and κ^*_{it} defined there have now dimensions q, nq, qp and nqp respectively, and the variance covariance matrices are given by Σ_ϵ, $\Sigma^*_{\epsilon i}$, Σ_κ and $\Sigma^*_{\kappa i}$; where the last two matrices are block diagonal with generic (q x q) blocks Σ_k and Σ^*_{ki}, $k = \eta, \delta, \omega$. The model can also be written as an nq-dimensional structural time series model with the variance covariance matrices of the random shocks in the measurement and transition equation given by A_ϵ and $\text{diag}[A_1,\ldots,A_p]$ respectively; where

$$(3.1) \qquad A_k = (\iota\iota' \otimes \Sigma_k) + \text{diag}[\Sigma^*_{k1},\ldots,\Sigma^*_{kn}], \qquad k = \epsilon, \eta, \delta, \omega.$$

From (1.6), the error components model type I for multivariate observations is restricted if

$$(3.2) \qquad A_k = (\iota\iota' \otimes \Sigma_k) + (D \otimes \Sigma^{**}_k), \qquad k = \epsilon, \eta, \delta, \omega;$$

where ι is an (n x 1) vector of ones, D is an (n x n) diagonal matrix and Σ^{**}_k is a (q x q) symmetric matrix. A restriction like this, although in a different context, was used by Magnus and Woodland

(1987). As a special case, D may be known or equal to the identity matrix I_n.

Given the presence of a multivariate model for each unit, it is relevant to consider here the homogeneity restriction presented in Section 1.1. The following definition extends the idea of homogeneity to error components models. Two important results, whose proofs are trivial, are stated in the following lemmas.

Definition 5.3.1 : An error components model type I is partially homogeneous if

(i) $\Sigma_k = q_k \Sigma_\epsilon$, $k = \eta, \delta, \omega$,

(ii) $\Sigma^*_{ki} = q^*_{ki} \Sigma^*_{\epsilon i}$, $k = \eta, \delta, \omega$; $i = 1,\ldots,n$.

Lemma 5.3.1 : If an error components model is partially homogeneous and if $q^*_{ki} = q_k$ for all i and k, then

(i) The multivariate form of the error components model is homogeneous.

(ii) The model for each unit is homogeneous.

Lemma 5.3.2 : The error components model for univariate observations (1.3) or (1.4) is partially homogeneous.

It can be observed from these results that if the model for each unit is homogeneous, the whole multivariate model is not necessarily partially homogeneous even if the proportional factors of the variance covariance matrices are the same for all units. On the other hand, it should be noticed that for the multivariate error components models considered in this section, the kind of restrictions introduced by the idea of homogeneity are independent of the restriction (3.2); while the

homogeneity restriction defines variance covariance matrices proportional across components, (3.2) defines the proportionality across units.

As in the case of univariate observations, the error components model type I for multivariate observations, under (3.2) and $D = I_n$, can be transformed to a model which is simple to estimate. The transformation matrix is now given by $(H \otimes I_q)$ and the variance covariance matrices of the random shocks in the measurement and transition equations of the transformed model are given by B_ϵ and $\text{diag}[B_1, \ldots, B_p]$; where,

$$(3.3) \qquad B_k = \begin{bmatrix} \Sigma_k + (1/n)\,\Sigma_k^{**} & 0 \\ 0 & (I_{n-1} - \iota\iota'/n) \otimes \Sigma_k^{**} \end{bmatrix},$$

and k takes some or all of the values ϵ, η, δ, ω. The first block defines a q dimensional model for the average of the observations; while the second block defines an (n-1)q-dimensional model independent of the first one. If the model is partially homogeneous, the second block is homogeneous but not necessarily the first, and if also $q_k^* = q_k$ for all k, then the two models are homogeneous and with the same factors. The same kind of extension presented above can be used, jointly with the results in previous sections, to formulate a dynamic error components model type II.

Chapter 8 studies the estimation of the unobserved components and the maximum likelihood estimation of the parameters in the error components model for multivariate observations.

5.4 Extension to Factor Analysis

The use of factor analysis, and other related techniques as principal components and canonical analysis, has long been an important topic in the analysis of multivariate time series. Quenouille (1957), Anderson (1963), Box and Tiao(1977), Engle and Watson (1981), Peña and Box (1987) and Tiao and Tsay (1989) have all considered factor analysis methods for time series. The common idea behind all these studies is the use of ARIMA time series models. An alternative approach was developed by Brillinger (1981), who extended the standard principal components techniques to the frequency domain, while Geweke (1977) and Geweke and Singleton (1981) discussed a frequency domain version of factor analysis. Within the framework of structural time series models, factor analysis models have been studied by Fernandez-Macho (1986, ch. 7) and Fernandez-Macho, Harvey and Stock (1987).

Extending the ideas presented in Section 5.1, and for the local level time series model, a very general factor analysis model can be defined as

$$\alpha_t = \mu_t + \Gamma_\epsilon \epsilon_t + \epsilon_t^*, \qquad t = 1,\ldots,T, \tag{4.1a}$$

$$\mu_t = \mu_{t-1} + \Gamma_\eta \eta_t + \eta_t^*, \tag{4.1b}$$

where α_t is, as before, an (n x 1) vector of observations, while the component μ_t of dimension (n x 1) represents the stochastic trends, and the vector random shocks ϵ_t, ϵ_t^*, η_t and η_t^* have dimensions r_ϵ, n, r_η and n respectively. These random shocks are assumed to be normally distributed, serially and mutually uncorrelated, with expected values equal to zero, and variance covariance matrices $I_{r\epsilon}$, D_ϵ^*, $I_{r\eta}$ and D_η^* respectively; where D_ϵ^* and D_η^* are diagonal matrices. Finally, the matrices Γ_ϵ and Γ_η of dimension (n x r_ϵ) and (n x r_η) contain fixed

parameters, and to solve the usual identification problem in factor analysis, their elements $\gamma_{\epsilon ij}$ and $\gamma_{\eta ij}$ are restricted to be zero for i $\prec$ j; see for example Anderson (1984, sec. 14.2).

The random shocks ϵ_t and η_t represent the irregular and trend common factors, while ϵ_t^* and η_t^* are the specific factors. If $r_\epsilon = r_\eta = 1$, and $\Gamma_\epsilon = \Gamma_\eta = \iota$, with ι an (n x 1) vector of ones, (4.1) reduces to the error components model type I. If also $\eta_t^* = 0$ for all t, (4.1) becomes the error components model type II. On the other hand, if $r_\epsilon = r_\eta = n$ and $\epsilon_t^* = \eta_t^* = 0$ for all t, model (4.1) reduces to the unrestricted local level model defined in (1.2.1). Thus, the factor analysis model (4.1) can be seen as an intermediate class of models between the unrestricted structural time series models considered in the first chapter and the error components models of Section 5.1. Notice that if $\eta_t^* = 0$ for all t, and $r_\eta \prec n$, the n time series are cointegrated in the sense of Engle and Granger (1987). That is, the n trends μ_t are generated by only r_η random shocks.

The model studied by Fernandez-Macho (1986, ch. 7) assumes $\epsilon_t^* = \eta_t^* = 0$ for all t and $r_\epsilon = n$. Hence, no restrictions are placed on the variance covariance matrix of the irregular random shock, and the variance covariance matrix of the trend random shock has rank $r_\eta \preceq n$. Fernandez-Macho, Harvey and Stock (1987) considered the same model but allowed the irregular random shock to follow an autoregressive formulation.

Definition 5.3.1 can be extended to the factor analysis model. If $r_\epsilon = r_\eta$, (4.1) is said to be partially homogenous if Γ_ϵ is proportional to Γ_η.

With respect to the estimation of the unobserved components and the maximum likelihood estimation of the parameters, the results that will be presented in Chapters 6 and 7 are straightforwardly generalised to

(4.1). The Kalman filter equations are substantially simplified under partial homogeneity, while the maximum likelihood estimation of the parameters is obtained by following very closely the results in Section 6.3 when $D_\eta^* \succ 0$ and in Section 7.3 when $D_\eta^* = 0$. The generalisation of (4.1) to deal with stochastic slope and seasonal components is also straightforward.

An obvious practical problem in the specification of a model like (4.1) is the selection of the values r_ϵ and r_η. One possibility to solve this is to use the test of hypothesis developed by Stock and Watson (1988). A more descriptive approach is to apply principal components analysis. Starting from an unrestricted local level model, it is possible to obtain estimates of the variance covariance matrices of the irregular and trend random shocks, say Σ_ϵ and Σ_η respectively. This can be done using the maximum likelihood principle although consistent and preliminary estimates based on the autocorrelation function are also possible; see Section 3.4 and the references therein. Then, standard principal components analysis can be applied to each of the variance covariance matrices of the random shocks in order to choose the specific form of (4.1).

An illustration of this technique using the Flour Price data used by Tiao and Tsay (1989) is presented in what follows. An unretricted local level model with a fixed slope was fitted. The frequency domain maximum likelihood estimates of Σ_ϵ and Σ_η were

$$\Sigma_\epsilon = 10^{-4}\begin{bmatrix} .00 & .00 & .00 \\ .00 & .68 & .88 \\ .00 & .88 & 1.67 \end{bmatrix}, \quad \text{and} \quad \Sigma_\eta = 10^{-2}\begin{bmatrix} .24 & .25 & .24 \\ .25 & .27 & .26 \\ .24 & .26 & .28 \end{bmatrix}$$

Table 5.4.1 presents the eigenvalues and eigenvectors of these two matrices. The largest eigenvalues capture 93% of the variation in the irregular random shock, and 97% of the variation in the trend or level

random shock. Their associated eigenvectors are, approximately, the average of the second and third time series for the irregular random shock, and the average of the three time series for the level random shock. A model with one factor in both the irregular and trend components may fit the data well and provide a simple interpretation of the mechanism that generates the observations. Finally, notice that the variance of the irregular random shock is close to zero, and in that case, the average of the three time series follows a random walk which represents a common trend that captures most of the variability.

Table 5.4.1 : Eigenvalues and Eigenvectors of Σ_ϵ and Σ_η (1)

	Eigenvalues (x 10^3)	Eigenvectors		
	.218	.000	.504	.864
Σ_ϵ	.016	.000	-.864	.504
	.000	1.000	.000	.000
	7.638	-.551	-.590	-.590
Σ_η	.260	.355	.474	-.806
	.013	.755	-.654	-.052

(1) Eigenvectors in rows.

CHAPTER 6 : ESTIMATION OF DYNAMIC ERROR COMPONENTS MODELS TYPE I

6.1 Introduction

This chapter considers the estimation of the error components model type I defined in Section 5.1. Assume that the observation y_{it} for the unit i at time t is generated by

$$(1.1) \qquad y_{it} = z_{it}' \beta_i + \alpha_{it}, \qquad \begin{matrix} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{matrix}$$

where z_{it} is an (r x 1) vector of exogenous variables, β_i is an (r x 1) vector of fixed parameters and α_{it} is a residual defined below. Special cases of (1.1) include the situations where the vector of exogenous variables is the same for all units, $z_{it} = z_t$ for all i, and the situations where the vector of coefficients is the same for all units, $\beta_i = \beta$ for all i. In multivariate form,

$$(1.2) \qquad y_t = X_t \beta + \alpha_t, \qquad t = 1,\ldots,T,$$

where y_t and α_t are (n x 1) vectors with i-th component y_{it} and α_{it} respectively, X_t is an (n x k) matrix of exogenous variables and β is a (k x 1) vector of fixed parameters. The relationship between (X_t, β, k) and (z_{it}, β_i, r) was presented in Section 5.1. The vector of residuals α_t in (1.2) satisfies (5.1.4). That is,

$$(1.3a) \qquad \alpha_t = (Z \otimes I_n)\, \theta_t + \iota\, \epsilon_t + \epsilon_t^*, \qquad t = 1,\ldots,T,$$

$$(1.3b) \qquad \theta_t = (T \otimes I_n)\, \theta_{t-1} + (R \otimes I_n)\, ((I_u \otimes \iota)\, \kappa_t + \kappa_t^*),$$

where the (np x 1) vector θ_t, which corresponds to the state vector, contains the unobserved trend and seasonal components and the random shocks ϵ_t, ϵ_t^*, κ_t and κ_t^* have dimensions 1, n, u and nu respectively,

and they are assumed to be normally distributed, serially and mutually uncorrelated, with expected values equal to zero and variance covariance matrices σ_ϵ^2, D_ϵ^*, D_K and D_K^* respectively; the last three being diagonal matrices. According to the particular time series model, some elements of D_K^* are assumed to be strictly positive; see Section 5.1 for details.

Model (1.3) will also be studied under the restrictions on the variances of the random shocks given by (5.1.6). A restricted error components model type I satisfies

$$D_\epsilon^* = D\ \sigma_\epsilon^{2*}, \tag{1.4a}$$

$$D_K^* = D_K^{**} \otimes D, \tag{1.4b}$$

where D_K^{**} and D are (u x u) and (n x n) diagonal matrices and, of course, a restriction should be imposed on (1.4) for identifiability. As a special case of (1.4), D may be known or equal to the identity matrix.

The chapter, apart from this introduction, is organised as follows. Section 6.2 presents the estimation of the unobserved components in the state vector θ_t defined in (1.3). That section also analyses the efficiency of the estimates for the local level time series model in terms of both, the number of units and the number of time observations. Section 6.3 deals with the maximum likelihood estimation of the parameters in (1.1) and (1.3), and with the formulation of asymptotic tests of hypotheses. Finally, in Section 6.4, an application using time series of unit labour costs in Austria, Belgium and Luxemburg, and Netherlands is presented.

6.2 Estimation of the Unobserved Components

This section considers the estimation of the state vector θ_t defined in (1.3), assuming α_t is observed and all the variances of the random shocks defined in the model are known. The formulas for the estimators of the state vector θ_t, $t = 1,\ldots,T$, and their mean square errors, are given by the Kalman filter equations presented in Section 1.3. The results in that section concerning the steady state Kalman filter, and the formation of initial estimates, also apply to the models defined here. That follows from the fact that the only difference between the models here and the ones defined in Chapter 1 are the restrictions over the variance covariance matrices of the random shocks, and these restrictions do not compromise the mentioned results.

Equations (1.3.5) and (1.3.6) give the estimator of the state vector at time $t = p$, m_p, and its mean square error, P_p. Then, the Kalman filter recursions (1.3.2) are run from $t = (p+1)$ to obtain estimates of the state vectors, and their mean square errors, for all $t \geq p$. The expression for P_p in (1.3.6) and the Kalman filter equations (1.3.2) involve the variance covariance matrices of the random shocks which were defined in the introduction of this chapter.

Unfortunately, it seems that the Kalman filter formulas cannot be simplified for the general error components model (1.3). Although the variance covariance matrices of the random shocks have a simple form, the matrices P_t and F_t, which represent the mean square errors of the estimators of the state vectors and the prediction error variances, do not have simple expressions for all t.

The remainder of this section focuses on the estimation of the state vector under the restriction (1.4). It will be shown that the

Kalman filter equations can be simplified in that situation. Finally, The efficiency of the estimators, in terms of both the number of units and the number of time observations is studied.

The Restricted Model

This subsection presents simplified expressions for the Kalman filter formulas when applied over the model (1.3) under (1.4). The identifiability problem which arises when working with (1.4) can be solved by setting σ_ϵ^{2*} or $(\iota'\ D^{-1}\ \iota)$ equal to unity. The advantage of setting $(\iota'\ D^{-1}\ \iota)$ equal to unity will become clear later. Definitions (1.3.1) are assumed in what follows.

Using an induction principle, it can be shown that if the mean square error of the estimator of the state vector at time (t-1), P_{t-1}, has a determined simple expression, P_t also has the same expression and that the dimension of the recursions needed to obtain P_t do not depend on n. Suppose that at time (t-1)

$$(2.1) \qquad P_{t-1} = P_{1t-1} \otimes \iota\iota' + P_{2t-1} \otimes D,$$

where P_{1t-1} and P_{2t-1} are (p x p) matrices. Using (1.3.2a) and (1.3.2b), it follows that

$$(2.2) \qquad \bar{P}_t = (T\ P_{1t-1}\ T' + R\ D_K\ R') \otimes \iota\iota' + (T\ P_{2t-1}\ T' + R\ D_K^{**}\ R') \otimes D$$

$$= P_{1t} \otimes \iota\iota' + P_{2t} \otimes D,$$

and

$$(2.3) \qquad F_t = (Z\ \bar{P}_{1t}\ Z' + \sigma_\epsilon^2) \otimes \iota\iota' + (Z\ \bar{P}_{2t}\ Z' + \sigma_\epsilon^{2*}) \otimes D$$

$$= f_{1t} \otimes \iota\iota' + f_{2t} \otimes D,$$

where $\bar{P}_{1t}$ and $\bar{P}_{2t}$ are (p x p) matrices and f_{1t} and f_{2t} are scalars. To obtain P_t from (1.3.2c) the inverse of F_t is needed. Using a well known formula for the inverse of a matrix of the form (2.3),

$$F_t^{-1} = h_{1t} D^{-1} - h_{2t} D^{-1} \iota\iota' D^{-1}, \tag{2.4}$$

where,

$$h_{1t} = f_{2t}^{-1}, \tag{2.5a}$$

and

$$h_{2t} = f_{2t}^{-2} (f_{1t}^{-1} + \iota' D^{-1} \iota f_{2t}^{-1})^{-1}. \tag{2.5b}$$

Then, using (1.3.2c),

$$P_t = P_{1t} \otimes \iota\iota' + P_{2t} \otimes D, \tag{2.6}$$

where P_{1t} and P_{2t} are (p x p) matrices given by

$$\begin{aligned} P_{1t} = \bar{P}_{1t} &+ \bar{P}_{1t} Z' Z \bar{P}_{1t} (h_{2t} \theta^2 - h_{1t} \theta) \\ &+ \bar{P}_{1t} Z' Z \bar{P}_{2t} (h_{2t} \theta - h_{1t}) \\ &+ \bar{P}_{2t} Z' Z \bar{P}_{1t} (h_{2t} \theta - h_{1t}) \\ &+ \bar{P}_{2t} Z' Z \bar{P}_{2t} h_{2t}, \end{aligned} \tag{2.7a}$$

and

$$P_{2t} = \bar{P}_{2t} - \bar{P}_{2t} Z' Z \bar{P}_{2t} h_{1t}, \tag{2.7b}$$

where θ is defined as $(\iota' D^{-1} \iota)$. Thus, (2.1) is true for all t provide it is for the initial values, and from (1.3.6), it is not difficult to see that this is the case. It follows that given an estimate of the state vector at time t = p, and its mean square error, the estimates

from t = (p+1) and their mean square errors are obtained by running the recursions (2.7). In other words, the Kalman filter formulas can be run with formulas which are essentially univariate, and their dimensions do not depend on the number of units, n. Furthermore, there is no need to compute the inverse of the prediction error variance F_t numerically and so the estimates of the state vector and their mean square errors can be obtained without difficulty even when the model contains a large number of units.

The above results can, of course, be applied when $D = I_n$ in (1.4); but, as it was shown in Section 5.2, in that case the model can be estimated by running a univariate model for the average $\bar{\alpha}_t$ and a homogeneous model for a vector with (n-1) of the deviations $(\alpha_{it} - \bar{\alpha}_t)$. The advantage of running a univariate and a homogeneous model instead of the recursions (2.7) is that in the former case the recursions are exactly equivalent, and not similar as in (2.7), to univariate Kalman filter recursions. Although this equivalence was proved in Section 5.2 by using the non singular transformation (5.2.6), the following lines present an alternative proof in terms of the Kalman filter equations. From (2.1), (2.2) and (2.3)

$$(2.8a) \quad P_{t-1} = [P_{2,t-1} \otimes (I_n - \iota\iota'/n)] + [(P_{1,t-1} + P_{2,t-1}/n) \otimes \iota\iota'],$$

$$(2.8b) \quad \bar{P}_t = [\bar{P}_{2t} \otimes (I_n - \iota\iota'/n)] + [(\bar{P}_{1t} + \bar{P}_{2t}/n) \otimes \iota\iota'],$$

$$(2.8c) \quad F_t = [f_{2t}\,(I_n - \iota\iota'/n)] + [(f_{1t} + f_{2t}/n)\,\iota\iota'].$$

Also from (2.4) and (2.5)

$$(2.9) \quad F_t^{-1} = \frac{1}{f_{2t}}\,(I_n - \iota\iota'/n) + \frac{1}{(f_{1t} + f_{2t}/n)}\,(\iota\iota'/n^2),$$

and then, using (2.7) and

$$(2.10) \qquad h_{1t} - n\, h_{2t} = \frac{1/n}{(f_{1t} + f_{2t}/n)},$$

yields

$$(2.11a) \qquad P_{2t} = \bar{P}_{2t} - \frac{\bar{P}_{2t}\, Z'\, Z\, \bar{P}_{2t}}{f_{2t}},$$

$$(2.11b) \qquad (P_{1t} + P_{2t}/n) = (\bar{P}_{1t} + \bar{P}_{2t}/n) - \frac{(\bar{P}_{1t} + \bar{P}_{2t}/n)\, Z'\, Z\, (\bar{P}_{1t} + \bar{P}_{2t}/n)}{(f_{1t} + f_{2t}/n)}.$$

Thus, P_{2t} and $(P_{1t} + P_{2t}/n)$ are formed exactly as univariate Kalman filter recursions. The first corresponds to the recursions of a homogeneous model with variance covariance matrix $(I_n - \iota\iota'/n)$; while the second corresponds to the recursions for the average across units. The full matrices P_t, $\bar{P}_t$ and F_t are then formed using (2.8).

The estimates of the state vector θ_t are obtained from (1.3.2d); and by noticing that $(I_n - \iota\iota'/n)\, \iota\iota' = 0$, this formula reduces to

$$(2.12a) \qquad m_t = (T \otimes I_n)\, m_{t-1} + \left[\bar{P}_{2t}\, Z'\, f_{2t}^{-1} \otimes (I_n - \iota\iota'/n)\right] v_t + \left[(\bar{P}_{1t} + \bar{P}_{2t}/n)\, Z'\, (f_{1t} + f_{2t}/n)^{-1} \otimes \iota\iota'/n\right] v_t,$$

which can also be written as

$$(2.12b) \qquad m_t = m_{1t} \otimes \iota + m_{2t},$$

where m_{1t} and m_{2t} are (p x 1) and (np x 1) vectors defined by

$$(2.13a) \qquad m_{1t} = T\, m_{1,t-1} + \frac{(\bar{P}_{1t} + \bar{P}_{2t}/n)\, Z'}{(f_{1t} + f_{2t}\,/\,n)} \left[\bar{\alpha}_t - Z\, T\, m_{1,t-1}\right],$$

$$(2.13b)\qquad m_{2t} = (T \otimes I_n)\, m_{2,t-1} + \left(\frac{\bar{P}_{2t}\, Z'}{f_{2t}} \otimes I_n\right)$$

$$\left[(\alpha_t - \iota\, \bar{\alpha}_t) - (Z\, T \otimes I_n)\, m_{2,t-1}\right].$$

Formula (2.13a) coincides with the estimates of the state vector in the univariate model for the average, while (2.13b) corresponds to the estimates of the state vector in the homogenous model of the deviations $(\alpha_{it} - \bar{\alpha}_t)$. From Section 1.3, (2.13b) can also be written as n univariate recursions for each of the deviations $(\alpha_{it} - \bar{\alpha}_t)$. The proof is completed by showing that initial values m_p and P_p also satisfy the above results; and that is almost immediate from (1.3.5) and (1.3.6). It follows that a univariate Kalman filter run for the time series $\bar{\alpha}_t$ produces m_{1t} and $(P_{1t} + P_{2t}/n)$ for all $t \geq p$, and a univariate Kalman filter run for each of the deviations $(\alpha_{it} - \bar{\alpha}_t)$ produces the components m_{2t} and P_{2t} for all $t \geq p$. The full vector m_t is formed with (2.12) and the full matrix P_t with (2.8a).

When D is not equal to the identity matrix but (1.4) still holds, the transformation (5.2.6) with H replaced by H^* in (5.2.11) also leads to a decomposition like the above, although the fact that H^* is a function of the parameters in the model makes this approach less attractive.

Efficiency Analysis

The following lines analyse the efficiency of the estimates of the state vectors in model (1.3) under (1.4), $D = I_n$, and assuming the local level time series model. That is,

$$(2.14a)\qquad \alpha_{it} = \mu_{it} + \epsilon_t + \epsilon^*_{it}, \qquad \begin{array}{l} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{array}$$

(2.14b) $$\mu_{it} = \mu_{i,t-1} + \eta_t + \eta^*_{it},$$

where μ_{it} is the level or trend of α_{it} and the random shocks ϵ_t, ϵ^*_{it}, η_t and η^*_{it} are assumed to be serially and mutually uncorrelated, normally distributed, with espected values equal to zero and variances σ^2_ϵ, σ^{2*}_ϵ, σ^2_η and σ^{2*}_η respectively.

Previous results in this section presented formulas for the estimates of μ_{it} for i = 1,...,n and t = 1,...,T; and their mean square errors for given values of n and T. This subsection considers the effects of changes in the values of n and T on the mean square errors. The motivation is the following. Suppose we are interested in the estimation of μ_{iT} for specific values of i and T. The problem can be stated as how many units and how many previous time periods should be considered in the estimation procedure in order to obtain a desired mean square error for the estimator of μ_{iT}. As the number of previous time observations or the number of units increases, the mean square error of the estimator of μ_{iT} decreases and there is a trade-off between the number of units and the number of time observations. From an experimental design point of view, we may be interested in the pair (n,T) such that the desired mean square error is obtained at minimum cost.

From equations (2.6) and (2.7), the mean square error of the estimator of μ_{it} can be written as

(2.15) $$P_{it} = P_{1t}(n,\ t,\ \sigma^2_\epsilon,\ \sigma^{2*}_\epsilon,\ \sigma^2_\eta,\ \sigma^{2*}_\eta) + P_{2t}(t,\ \sigma^{2*}_\epsilon,\ \sigma^{2*}_\eta),$$

where P_{1t} and P_{2t} satisfies the recursions (2.7) which for the model (2.14) reduce to

(2.16a) $$P_{1t} = (P_{1,t-1} + \sigma^2_\eta)$$

$$- \frac{n\,(P_{1,t-1} + \sigma_\eta^2)^2 + 2\,(P_{2,t-1} + \sigma_\eta^{2*})\,(P_{1,t-1} + \sigma_\eta^2)}{n\,(P_{1,t-1} + \sigma_\eta^2 + \sigma_\epsilon^2) + (P_{2,t-1} + \sigma_\eta^{2*} + \sigma_\epsilon^{2*})}$$

$$+ \frac{(P_{2,t-1} + \sigma_\eta^{2*})^2\,(P_{1,t-1} + \sigma_\eta^2 + \sigma_\epsilon^2)}{n\,(P_{1,t-1} + \sigma_\eta^2 + \sigma_\epsilon^2)\,(P_{2,t-1} + \sigma_\eta^{2*} + \sigma_\epsilon^{2*}) + (P_{2,t-1} + \sigma_\eta^{2*} + \sigma_\epsilon^{2*})^2},$$

and

$$(2.16b) \qquad P_{2t} = (P_{2,t-1} + \sigma_\eta^{2*}) - \frac{(P_{2,t-1} + \sigma_\eta^{2*})^2}{(P_{2,t-1} + \sigma_\eta^{2*} + \sigma_\epsilon^{2*})},$$

for $t = 2,\ldots,T$; and with $P_{11} = \sigma_\epsilon^2$, $P_{21} = \sigma_\epsilon^{2*}$. In the above formulas, P_{1t} represents basically the variance of the components which are common to all units and hence, P_{1t} can be reduced by increasing the number of units; while P_{2t} represents the variance of the unit specific components, which are independent, and therefore P_{2t} cannot be reduced by increasing the number of units n.

It can be observed that P_{2t} in (2.16b) has exactly the form of the variance of a standard univariate local level model, with irregular and level random shock variances equal to σ_ϵ^{2*} and σ_η^{2*} respectively. Also, from (2.16a),

$$(2.17a) \qquad \lim_{n\to\infty} \quad P_{1t} = P_{1t}^*,$$

where

$$(2.17b) \qquad P_{1t}^* = (P_{1,t-1}^* + \sigma_\eta^2) - \frac{(P_{1,t-1}^* + \sigma_\eta^2)^2}{(P_{1,t-1}^* + \sigma_\eta^2 + \sigma_\epsilon^2)},$$

and then P_{1t}^* also has the form of the variance of a standard univariate local level model with random shock variances equal to σ_ϵ^2 and σ_η^2. It follows that as n goes to infinity, P_{it} is the sum of two standard univariate variances in a local level model. On the other hand, if only

the time series for unit i is considered in the estimation of μ_{iT}; that is, if n is the unity, P_t satisfies the formulas of a standard local level model with irregular and level random shock variances equal to $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*})$ and $(\sigma_\eta^2 + \sigma_\eta^{2*})$ respectively.

In what follows, the relative efficiency of the estimator of μ_{iT} with n = 1 and T = 1 compared to the estimator based on n time series with T observations each is computed. Some special cases are analysed first.

a) If $\sigma_\eta^2 = \sigma_\epsilon^2 = 0$, the n time series are uncorrelated, $P_{1t} = 0$ for all n and t, and no gains in efficiency can be obtained by increasing the number of units.

b) If the model is homogeneous,

$$(2.18) \qquad \rho_\epsilon \equiv \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_\epsilon^{2*}} = \rho_\eta \equiv \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\eta^{2*}},$$

and the relative efficiency is independent of the number of units n and decreases only with T. This is true because in a homogeneous model the estimates obtained using a single time series and the estimates obtained using the whole multivariate system coincide; see Section 1.3.

c) If T is the unity, $P_T = (\sigma_\epsilon^2 + \sigma_\epsilon^{2*})$ for all n, and the relative efficiency is the unity for all values of n.

Apart from these special cases, it seems that it is not possible to obtain analytic expressions for the relative efficiency given the complicated recursive formulas for P_{1t} and P_{2t} in (2.16). However, given values of n and T and given the variances of the random shocks,

values of P_{1t} and P_{2t} can be computed numerically. The results of this exercise for some specific values of the variances are presented below.

Table 6.2.1 : Parameter Values in the Evaluation of the Relative Efficiency(1)

Table	$\sigma_\eta^2 + \sigma_\eta^{2*}$	ρ_ϵ	ρ_η	σ_ϵ^2	σ_ϵ^{2*}	σ_η^2	σ_η^{2*}
6.2.2	.20	.20	.80	.20	.80	.16	.04
6.2.3	.50	.20	.80	.20	.80	.40	.10
6.2.4	1.00	.20	.80	.20	.80	.80	.20
6.2.5	2.00	.20	.80	.20	.80	1.60	.40
6.2.6	5.00	.20	.80	.20	.80	4.00	1.00
6.2.7	.20	.80	.20	.80	.20	.04	.16
6.2.8	.50	.80	.20	.80	.20	.10	.40
6.2.9	1.00	.80	.20	.80	.20	.20	.80
6.2.10	2.00	.80	.20	.80	.20	.40	1.60
6.2.11	5.00	.80	.20	.80	.20	1.00	4.00

(1) $\rho_i = \sigma_i^2 / (\sigma_i^2 + \sigma_i^{2*})$; $i = \epsilon, \eta$.

Table 6.2.1 shows the values of the parameters considered in the numerical evaluation of the relative efficiency. Without loss in generality, the value of $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*})$ was fixed as equal to unity. Five values of $(\sigma_\eta^2 + \sigma_\eta^{2*})$ were considered : .2, .5, 1, 2 and 5. These values represent the relative value of the variance of the level random shock compared to the variance of the irregular random shock. Finally, two values were considered for the correlation of the random shock across units : .2 and .8. The correlation was never set equal in both the irregular and level random shocks because in that case the model is homogeneous, and as stated before, for a given value of T the relative

efficiency is the same for all values of n.

The results of the numerical evaluation of the relative efficiency for the ten cases presented in Table 6.2.1 are shown in tables 6.2.2 to 6.2.11. In all the situations studied, the relative efficiency of the estimators based on more than one time observation are less than unity and decrease as n or T increases. From (2.17) and the results in Section 1.3 concerning the steady state Kalman filter, the relative efficiency approaches a limit as n or T increases, and this limit and the way in which it is reached depend heavily on the relative values of the variances of the random shocks. In tables 6.2.2 to 6.2.11, the relative efficiency fluctuates between .28 and .72 when n and T are equal to one hundred. The smaller the variance of the level or trend random shock, the smaller the relative efficiency as T increases. With respect to the effect of increasing the number of units, the results show that the gains in efficiency are more independent of the relative values of the variances. For example, when T is equal to one hundred, the reduction in the relative efficiency by increasing the number of units from one to one hundred fluctuates between 16% and 23%.

Table 6.2.2 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = .16 \quad \sigma_\eta^{2*} = .04$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.55	.39	.36	.36	.36	.36	.36	.36
2	1.00	.54	.38	.35	.34	.33	.33	.33	.33
4	1.00	.54	.37	.33	.32	.31	.31	.31	.31
6	1.00	.54	.36	.32	.31	.30	.30	.30	.30
8	1.00	.54	.36	.32	.30	.30	.29	.29	.29
10	1.00	.54	.36	.32	.30	.30	.29	.29	.29
20	1.00	.54	.35	.31	.29	.29	.28	.28	.28
50	1.00	.54	.35	.31	.29	.28	.28	.28	.28
100	1.00	.54	.35	.30	.29	.28	.28	.28	.28

Table 6.2.3 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = .40 \quad \sigma_\eta^{2*} = .10$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.60	.51	.50	.50	.50	.50	.50	.50
2	1.00	.59	.48	.46	.46	.46	.46	.46	.46
4	1.00	.59	.46	.44	.43	.43	.43	.43	.43
6	1.00	.58	.45	.42	.42	.42	.42	.42	.42
8	1.00	.58	.44	.42	.41	.41	.41	.41	.41
10	1.00	.58	.44	.41	.41	.41	.41	.41	.41
20	1.00	.58	.43	.40	.40	.40	.40	.40	.40
50	1.00	.58	.42	.40	.39	.39	.39	.39	.39
100	1.00	.57	.42	.39	.39	.39	.39	.39	.39

Table 6.2.4 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = .80 \quad \sigma_\eta^{2*} = .20$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.67	.62	.62	.62	.62	.62	.62	.62
2	1.00	.65	.58	.57	.57	.57	.57	.57	.57
4	1.00	.64	.54	.53	.53	.53	.53	.53	.53
6	1.00	.63	.53	.52	.52	.52	.52	.52	.52
8	1.00	.63	.52	.51	.51	.51	.51	.51	.51
10	1.00	.62	.52	.50	.50	.50	.50	.50	.50
20	1.00	.62	.51	.49	.49	.49	.49	.49	.49
50	1.00	.61	.50	.49	.48	.48	.48	.48	.48
100	1.00	.61	.50	.48	.48	.48	.48	.48	.48

Table 6.2.5 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = 1.60 \quad \sigma_\eta^{2*} = .40$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.75	.73	.73	.73	.73	.73	.73	.73
2	1.00	.72	.68	.67	.67	.67	.67	.67	.67
4	1.00	.70	.64	.63	.63	.63	.63	.63	.63
6	1.00	.69	.62	.62	.62	.62	.62	.62	.62
8	1.00	.68	.61	.61	.61	.61	.61	.61	.61
10	1.00	.68	.61	.60	.60	.60	.60	.60	.60
20	1.00	.67	.60	.59	.59	.59	.59	.59	.59
50	1.00	.66	.59	.59	.58	.58	.58	.58	.58
100	1.00	.66	.59	.58	.58	.58	.58	.58	.58

Table 6.2.6 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = 4.00 \quad \sigma_\eta^{2*} = 1.00$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.86	.85	.85	.85	.85	.85	.85	.85
2	1.00	.81	.80	.80	.80	.80	.80	.80	.80
4	1.00	.78	.76	.76	.76	.76	.76	.76	.76
6	1.00	.77	.75	.75	.75	.75	.75	.75	.75
8	1.00	.77	.74	.74	.74	.74	.74	.74	.74
10	1.00	.76	.74	.74	.74	.74	.74	.74	.74
20	1.00	.75	.73	.73	.73	.73	.73	.73	.73
50	1.00	.75	.72	.72	.72	.72	.72	.72	.72
100	1.00	.75	.72	.72	.72	.72	.72	.72	.72

Table 6.2.7 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = .04 \quad \sigma_\eta^{2*} = .16$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.55	.39	.36	.36	.36	.36	.36	.36
2	1.00	.54	.37	.34	.33	.33	.33	.33	.33
4	1.00	.54	.36	.32	.31	.31	.31	.31	.31
6	1.00	.54	.36	.32	.31	.30	.30	.30	.30
8	1.00	.54	.36	.31	.30	.30	.30	.30	.30
10	1.00	.54	.35	.31	.30	.29	.29	.29	.29
20	1.00	.54	.35	.31	.29	.29	.28	.28	.28
50	1.00	.54	.35	.30	.29	.28	.28	.28	.28
100	1.00	.54	.35	.30	.29	.28	.28	.28	.28

Table 6.2.8 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = .10 \quad \sigma_\eta^{2*} = .40$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.60	.51	.50	.50	.50	.50	.50	.50
2	1.00	.59	.48	.47	.46	.46	.46	.46	.46
4	1.00	.58	.45	.44	.43	.43	.43	.43	.43
6	1.00	.58	.44	.42	.42	.42	.42	.42	.42
8	1.00	.58	.44	.42	.41	.41	.41	.41	.41
10	1.00	.58	.43	.41	.41	.41	.41	.41	.41
20	1.00	.58	.43	.40	.40	.40	.40	.40	.40
50	1.00	.57	.42	.40	.39	.39	.39	.39	.39
100	1.00	.57	.42	.39	.39	.39	.39	.39	.39

Table 6.2.9 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = .20 \quad \sigma_\eta^{2*} = .80$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.67	.62	.62	.62	.62	.62	.62	.62
2	1.00	.65	.58	.58	.58	.58	.58	.58	.58
4	1.00	.63	.55	.54	.54	.54	.54	.54	.54
6	1.00	.63	.53	.52	.52	.52	.52	.52	.52
8	1.00	.62	.52	.52	.51	.51	.51	.51	.51
10	1.00	.62	.52	.51	.51	.51	.51	.51	.51
20	1.00	.62	.51	.50	.49	.49	.49	.49	.49
50	1.00	.61	.50	.49	.49	.48	.48	.48	.48
100	1.00	.61	.50	.48	.48	.48	.48	.48	.48

Table 6.2.10 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = .40 \quad \sigma_\eta^{2*} = 1.60$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.75	.73	.73	.73	.73	.73	.73	.73
2	1.00	.72	.69	.69	.69	.69	.69	.69	.69
4	1.00	.70	.65	.65	.65	.65	.65	.65	.65
6	1.00	.69	.63	.63	.63	.63	.63	.63	.63
8	1.00	.68	.62	.62	.62	.62	.62	.62	.62
10	1.00	.68	.62	.62	.62	.62	.62	.62	.62
20	1.00	.67	.60	.60	.60	.60	.60	.60	.60
50	1.00	.66	.59	.59	.59	.59	.59	.59	.59
100	1.00	.66	.59	.58	.58	.58	.58	.58	.58

Table 6.2.11 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = 1.00 \quad \sigma_\eta^{2*} = 4.00$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.86	.85	.85	.85	.85	.85	.85	.85
2	1.00	.83	.82	.82	.82	.82	.82	.82	.82
4	1.00	.80	.79	.79	.79	.79	.79	.79	.79
6	1.00	.78	.77	.77	.77	.77	.77	.77	.77
8	1.00	.78	.76	.76	.76	.76	.76	.76	.76
10	1.00	.77	.75	.75	.75	.75	.75	.75	.75
20	1.00	.76	.74	.74	.74	.74	.74	.74	.74
50	1.00	.75	.72	.72	.72	.72	.72	.72	.72
100	1.00	.75	.72	.72	.72	.72	.72	.72	.72

6.3 Maximum Likelihood Estimation of the Parameters

This section considers the maximum likelihood estimation and the formulation of tests of hypotheses for the vector of coefficients of the exogenous variables β and the variance covariance matrices of the random shocks in the model defined by (1.2) and (1.3). Following the results in Chapter 2, this introduction presents a general form for the log-likelihood function of the model. The estimation of the vector of coefficients β, the results for the estimation of the parameters in the variance covariance matrices of the random shocks, and the formulation of asymptotic tests of hypotheses are considered later.

Assuming that the (np x 1) state vector θ_t defined in (1.3) is initialised with a diffuse prior, a general form for the log-likelihood of the model (1.2)-(1.3), which includes the log-likelihood formed using the Kalman filter, the frequency domain log-likelihood, and the alternative expression developed in Section 2.3 is, apart from a constant, given by

$$(3.1a) \qquad \ell = -\tfrac{1}{2} \sum_{t=p+1}^{T} \left[\log|G_t| + (w_{yt} - W_{xt}\,\beta)'\, G_t^{-1}\, (w_{yt} - W_{xt}\,\beta) \right],$$

or, alternatively, by

$$(3.1b) \qquad \ell = -\tfrac{1}{2} \sum_{t=p+1}^{T} \left[\log|G_t| + \mathrm{trace}(G_t^{-1}\, P_t) \right],$$

where G_t is an (n x n) matrix which depends on the parameters in the variance covariance matrices of the random shocks, and w_{yt} and W_{xt} have dimensions (n x 1) and (n x k) respectively and they are functions of the observations. If the log-likelihood is formed using the Kalman filter, w_{yt} and W_{xt} depend also on the variance covariance matrices of

the random shocks. The (n x n) matrix P_t is defined as

$$(3.2) \qquad P_t = (w_{yt} - W_{xt}\,\beta)\,(w_{yt} - W_{xt}\,\beta)', \qquad t = p+1,\ldots,T.$$

If the log-likelihood is formed using the frequency domain approach, or the alternative procedure presented in Section 2.3, the matrix G_t has the form

$$(3.3) \qquad G_t = g_{\epsilon t}\,(\sigma_\epsilon^2\,\iota\iota' + D_\epsilon^*) + g_{\eta t}\,(\sigma_\eta^2\,\iota\iota' + D_\eta^*)$$

$$+ g_{\delta t}\,(\sigma_\delta^2\,\iota\iota' + D_\delta^*) + g_{\omega t}\,(\sigma_\omega^2\,\iota\iota' + D_\omega^*)$$

$$= g_t\,\iota\iota' + D_t^*, \qquad t = p+1,\ldots,T,$$

where $g_{\epsilon t}$, $g_{\eta t}$, $g_{\delta t}$ and $g_{\omega t}$ are known scalars. For details in the form of the log-likelihood and in the definitions above see Chapter 2.

Estimation of β

The maximum likelihood estimator of β, $\tilde{\beta}$, minimises the quadratic form

$$(3.4) \qquad Q = \tfrac{1}{2} \sum_t (w_{yt} - W_{xt}\,\beta)'\, G_t^{-1}\, (w_{yt} - W_{xt}\,\beta),$$

and the solution for β is given by

$$(3.5a) \qquad \tilde{\beta} = \Big[\sum_t W_{xt}'\, G_t^{-1}\, W_{xt}\Big]^{-1} \Big[\sum_t W_{xt}'\, G_t^{-1}\, w_{yt}\Big],$$

with the information matrix, $I(\beta)$, equal to

$$(3.5b) \qquad I(\beta) = \sum_t W_{xt}'\, G_t^{-1}\, W_{xt}.$$

In (3.5a), the maximum likelihood estimator of β depends on the

variance covariance matrices of the random shocks. As these variances are usually unknown, some kind of joint estimation procedure must be considered. A concentrated log-likelihood can be formed by replacing (3.5a) into (3.1). This concentrated log-likelihood has then to be maximised numerically with respect to the parameters in the variance covariance matrices.

If (3.3) holds, analytic expressions for the determinant and the inverse of G_t can be obtained using well known formulas. That reduces the computer time required in the evaluation of the log-likelihood (3.1). These expressions are,

$$|G_t| = |D_t^*| \, (1 + g_t \, (\iota' \, D_t^{*-1} \, \iota)), \tag{3.6a}$$

and

$$G_t^{-1} = D_t^{*-1} - \frac{D_t^{*-1} \, \iota\iota' \, D_t^{*-1}}{(g_t^{-1} + \iota' \, D_t^{*-1} \, \iota)}. \tag{3.6b}$$

Consider the estimation of β when the restriction (1.4) with $D = I_n$ holds, and when either the vector of exogenous variables or the vector of coefficients of the exogenous variables are the same for all units. Suppose first that the vector of exogenous variables is the same for all units. Premultiplying (1.2) by the (n x n) matrix H defined in (5.2.4), produces

$$\bar{y}_t = z_t' \, \bar{\beta} + \bar{\alpha}_t, \qquad t = 1,\ldots,T, \tag{3.7a}$$

$$(y_{it} - \bar{y}_t) = z_t' \, (\beta_i - \bar{\beta}) + (\alpha_{it} - \bar{\alpha}_t), \qquad \begin{matrix} i = 1,\ldots,n-1, \\ t = 1,\ldots,T, \end{matrix} \tag{3.7b}$$

where $\bar{x}_t$ represent the average of the component x_{it} across units. The joint likelihood of $\{\bar{y}_t, \; t = 1,\ldots,T\}$ and $\{(y_{it} - \bar{y}_t), \; i = 1,\ldots,n-1; \; t = 1,\ldots,T\}$ is, apart from a constant, equal to the likelihood of the

original observations because the determinant of H is equal to (1/n). Even more, as shown in Section 5.2, the random shocks in (3.7a) are uncorrelated with the random shocks in (3.7b), and the variance covariance matrices of the random shocks in (3.7b) are proportional to the matrix $(I_{n-1} - \iota\iota'/n)$; hence, (3.7b) represents a homogeneous model. It follows that the log-likelihood can be written as the sum of two log-likelihoods. One for the observations in (3.7a) and the other for the observations in (3.7b). The maximum likelihood estimator of $\bar{\beta}$ is obtained from the univariate model (3.7a), while the estimators of $(\beta_i - \bar{\beta})$, $i = 1,\ldots(n-1)$, are obtained from the homogeneous model (3.7b). In fact, to be able to estimate the two models in turn, a reparametarisation of the variances of the random shocks is also required. It will be shown later that this is possible. If the objective of the analysis is to test the hypothesis that the vector of coefficients is the same for all units, only (3.7b) needs to be estimated. If the hypothesis of equal coefficients is accepted, the common to all units vector of coefficients is estimated from (3.7a).

When the exogenous variables are unit specific but the vector of coefficients is the same for all units, premultiplying (1.2) by H yields

$$(3.8a) \qquad \bar{y}_t = \bar{z}_t' \beta + \bar{\alpha}_t, \qquad t = 1,\ldots,T,$$

$$(3.8b) \qquad (y_{it} - \bar{y}_t) = (z_{it} - \bar{z}_t)' \beta + (\alpha_{it} - \bar{\alpha}_t), \qquad i = 1,\ldots,n-1,\ t = 1,\ldots,T.$$

In this situation, the results presented below equations (3.7) concerning the log-likelihood still hold but now the vector of coefficients β appears in both models. The estimator of β obtained from the model (3.8a) is known in the literature as the "between groups" estimator, while the one which considers only (3.8b) is known as the

"within groups" estimator; see for example Hsiao (1986). Of course, none of these is the maximum likelihood estimator which minimises the sum of two quadratic forms; say

$$(3.9) \qquad Q = Q_1 + Q_2,$$

where Q_1 and Q_2 are quadratic forms of the form (3.4) but defined for the models (3.8a) and (3.8b) respectively. The "between" and "within" estimators have the form (3.5); hence, they can be written as

$$(3.10a) \qquad \tilde{\beta}_b = H_1^{-1} h_1,$$

and

$$(3.10b) \qquad \tilde{\beta}_w = H_2^{-1} h_2,$$

with obvious notation for H_1, H_2, h_1 and h_2. It can be shown without difficulty that the maximum likelihood estimator of β is formed as an average of the "between" and the "within" estimators. That is,

$$(3.11) \qquad \tilde{\beta} = (H_1 + H_2)^{-1} H_1 \tilde{\beta}_b + (H_1 + H_2)^{-1} H_2 \tilde{\beta}_w.$$

This result was shown by Maddala (1971) for the static model (5.1.1). The information matrix for β is given by

$$(3.12) \qquad I(\beta) = H_1 + H_2,$$

and then, if H_1 and H_2 are positive definite matrices and if the inverse of $I(\beta)$ represents an approximation of the variance covariance matrices of the estimators, the maximum likelihood estimator is more efficient than both the "between" and the "within" estimators.

Finally, consider the case where both the vector of exogenous variables and the vector of coefficients is the same for all units. Premultiplying (1.2) by H produces

(3.13a) $\bar{y}_t = z_t' \beta + \bar{\alpha}_t, \qquad t = 1,\ldots,T,$

(3.13b) $(y_{it} - \bar{y}_t) = (\alpha_{it} - \bar{\alpha}_t), \qquad i = 1,\ldots,n-1, \quad t = 1,\ldots,T.$

In this case, the maximum likelihood estimator of β is obtained simply from the univariate model (3.13a).

In sum, the maximum likelihood estimator of the vector of coefficients of the exogenous variables can be obtained as in Chapter 3 for the general structural time series model. If the number of units is large, several simplifications presented in this subsection avoid the inversion of large matrices. First, if the log-likelihood is formed by means of the frequency domain approach or the alternative transformation in Section 2.3, formulas (3.6) give analytic expressions for the determinant and the inverse of the (n x n) matrix G_t. Second, when the restriction (1.4) holds and either the vector of exogenous variables or the vector of coefficents of the exogenous variables is the same for all units, the model can be decomposed into two parts and the estimation of the coeficients of the exogenous variables becomes much simpler.

Estimation of the Variances

This subsection considers the maximum likelihood estimation of the variances of the random shocks. Unless otherwise stated, it is assumed that the log-likelihood is formed using the frequency domain approach or the alternative transformation developed in Section 2.3. Then (3.3) holds, and it is not difficult to obtain first and second derivatives of the log-likelihood (3.1b) with respect to the variances even for the unrestricted model. To simplify the exposition, the basic structural time series model is assumed. The necessary modifications for the other

three time series models are obvious.

From (3.1b), and the results in Magnus and Neudecker (1988), the first two differentials of the log-likelihood, when the only parameters in the model are contained in the matrices G_t, are

$$(3.14a) \qquad d\ell_t = vec'[\bar{G}_t^{-1} - \bar{G}_t^{-1} P_t \bar{G}_t^{-1}] vec[dG_t],$$

and

$$(3.14b) \qquad d^2\ell_t = vec'[dG_t][\bar{G}_t^{-1} \otimes (2(\bar{G}_t^{-1} P_t \bar{G}_t^{-1}) - \bar{G}_t^{-1})] vec[dG_t].$$

On the other hand, (3.3) implies that

$$(3.15) \qquad vec[G_t] = (\iota \; x_t') \theta_1 + S (x_t' \otimes I_n) \theta_2$$

$$= X_{1t}' \theta_1 + X_{2t}' \theta_2$$

$$= X_t' \theta,$$

where ι is an (n^2 x 1) vector with all the elements equal to the unity; x_t, θ_1 and θ_2 are (4 x 1), (4 x 1) and (4n x 1) vectors given by

$$(3.16a) \qquad x_t' = (g_{\epsilon t}, g_{\eta t}, g_{\delta t}, g_{\omega t}),$$

$$(3.16b) \qquad \theta_1' = (\sigma_\epsilon^2, \sigma_\eta^2, \sigma_\delta^2, \sigma_\omega^2),$$

and

$$(3.16c) \qquad \theta_2' = (\sigma_{\epsilon 1}^{2*}, \ldots, \sigma_{\epsilon n}^{2*}, \ldots, \sigma_{\omega 1}^{2*}, \ldots, \sigma_{\omega n}^{2*}),$$

and S is an (n^2 x n) selection matrix which transforms a vector of dimension n into the vector of an (n x n) diagonal matrix with the original vector in the diagonal. See Magnus (1988) for the form of this matrix. The vector $\theta' = (\theta_1' \; , \; \theta_2')$ contains the functionally independent parameters in the model. The matrices X_{1t}' and X_{2t}' have dimension (n^2 x

4) and (n^2 x 4n) respectively, and then $X_t^{\prime}$ has dimension (n^2 x 4n+4). Clearly,

$$\text{(3.17)} \qquad \text{vec}[\,dG_t\,] = X_t^{\prime}\, d\theta,$$

and then, the maximum likelihood estimator of θ satisfies the non linear system of equations

$$\text{(3.18)} \qquad \sum_t X_t \,\text{vec}[\,G_t^{-1} - G_t^{-1}\, P_t\, G_t^{-1}\,] = 0.$$

An asymptotically equivalent expression for the information matrix of $\theta' = (\theta_1^{\prime}\ ,\ \theta_2^{\prime})$ is given by,

$$\text{(3.19)} \qquad I(\theta) = \tfrac{1}{2} \sum_t X_t\,[\,G_t^{-1} \otimes G_t^{-1}\,]\, X_t^{\prime}.$$

If the restriction (1.4) holds, and D is known, all the above formulas can be applied with $X_{2t}^{\prime}$ and θ_2 replaced by $\bar{X}_{2t}^{\prime}$ and $\bar{\theta}_2$ of dimension (n^2 x 4) and (4 x 1) respectively, and given by

$$\text{(3.20a)} \qquad \bar{X}_{2t}^{\prime} = \text{vec(D)}\ x_t^{\prime},$$

and

$$\text{(3.20b)} \qquad \bar{\theta}_2^{\prime} = (\sigma_\epsilon^{2*},\ \sigma_\eta^{2*},\ \sigma_\delta^{2*},\ \sigma_\omega^{2*}).$$

Finally, if (1.4) holds but D is unknown, the above results can be formed following the same idea but from

$$\text{(3.21)} \qquad \text{vec}[\,G_t\,] = X_{1t}^{\prime}\,\theta_1 + \bar{X}_{2t}^{\prime}\,\bar{\theta}_2 + X_{3t}^{\prime}\,\theta_3, \qquad t = p+1,\ldots,T,$$

where $X_{3t}^{\prime} = (\bar{S}\ x_t^{\prime}\ \bar{\theta}_2)$, $\bar{S}$ is an (n^2 x n-1) selection matrix which transform θ_3 into vec(D), and θ_3 is the (n-1 x 1) vector of functionally independet parameters in D.

In the evaluation of (3.18) and (3.19), the analytic expression for

G_t^{-1} given by (3.6b) simplifies enormously the calculations.

In practice, β is unknown and has to be estimated jointly with the parameters θ in the variance covariance matrices. If β is concentrated out of the log-likelihood, by replacing (3.5a) into (3.1), it seems difficult to obtain first and second derivatives for the parameters θ. On the other hand, given the vector of first derivatives and the matrix of second derivatives with respect to β and θ, the results in this section suggest an iterative or stepwise procedure in the lines suggested in Section 3.4. As in the general multivariate structural time series model, the matrix of second derivatives in the models studied here is block diagonal with respect to the subsets of parameters β and θ; see Magnus (1978).

If (1.4) holds and $D = I_n$, the estimation procedure for the variances can be simplified. It was shown in Section 5.2, and also below equations (3.7), that in this case the log-likelihood can be decomposed as the sum of two log-likelihoods; one for the average of the observations $\bar{y}_t$ and the other for (n-1) deviations of the form $(y_{it} - \bar{y}_t)$. Using the results in Section 5.2,

$$(3.22) \qquad \ell = c + \ell_1 + \ell_2,$$

where c is a constant and

$$(3.23a) \qquad \ell_1 = -\tfrac{1}{2} \sum_t \left[\log(g_{1t}) + P_{1t}/g_{1t}\right],$$

$$(3.23b) \qquad \ell_2 = -\tfrac{1}{2} \sum_t \left[\log|(I_{n-1} - \iota\iota'/n)\, g_{2t}| + \operatorname{trace}((I_{n-1} - \iota\iota'/n)^{-1}\, g_{2t}^{-1}\, P_{2t})\right].$$

In (3.23), the scalar P_{1t} and the (n-1 x n-1) matrix P_{2t} are transformations of the residuals $\bar{\alpha}_t$ and $(\alpha_{it} - \bar{\alpha}_t)$, i = 1,...,n-1

respectively, as in (3.2); while the scalars g_{1t} and g_{2t} are defined by

$$(3.24a) \qquad g_{1t} = g_{\epsilon t}\,(\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n) + g_{\eta t}\,(\sigma_\eta^2 + \sigma_\eta^{2*}/n)$$

$$+ g_{\delta t}\,(\sigma_\delta^2 + \sigma_\delta^{2*}/n) + g_{\omega t}\,(\sigma_\omega^2 + \sigma_\omega^{2*}/n),$$

and

$$(3.24b) \qquad g_{2t} = g_{\epsilon t}\,(\sigma_\epsilon^{2*}) + g_{\eta t}\,(\sigma_\eta^{2*}) + g_{\delta t}\,(\sigma_\delta^{2*}) + g_{\omega t}\,(\sigma_\omega^{2*}).$$

Notice that by redefining the variances of the random shocks as $(\sigma_k^2 + \sigma_k^{2*}/n)$ and σ_k^{2*}, for $k = \epsilon,\ \eta,\ \delta$ and ω, the two log-likelihoods in (3.23) can be maximised separately. Expression (3.23) is the standard log-likelihood of a univariate model and then all the results in Chapter 3 can be applied. Basically, one of the redefined variances can be concentrated out of the log-likelihood and if (3.7) or (3.13) holds, the vector of exogenous variables can also be concentrated out.

Expression (3.23b) is the standard log-likelihood of a homogenous model but with a known common variance covariance matrix equal to $(I_{n-1} - \iota\iota'/n)$. This log-likelihood can alternatively be written as

$$(3.25) \qquad \ell_2 = c - \tfrac{1}{2} \sum_t \left[(n-1)\,\log(g_{2t}) + (1/g_{2t})\,\mathrm{trace}(P_{2t}^*) \right],$$

where c is a constant and P_{2t}^* is an (n x n) matrix which is obtained by applying the transformation (3.2) over the (n x 1) vector of differences of the form $(\alpha_{it} - \bar{\alpha}_t)$. Notice that the only difference between P_{2t} and P_{2t}^* is that the former applies to (n-1) of the deviations $(\alpha_{it} - \bar{\alpha}_t)$ while the later applies to the n deviations. Expression (3.25) is obtained from (3.23b) by using the fact that the inverse of $(I_{n-1} - \iota\iota'/n)$ is equal to $(I_{n-1} + \iota\iota')$; and then,

$$(3.26) \qquad \mathrm{trace}((I_{n-1} + \iota\iota')\,P_{2t}) = \mathrm{trace}(P_{2t}) + \iota'\,P_{2t}\,\iota$$

$$= \text{trace}(P^{*}_{2t}),$$

because the sum of the n deviations $(\alpha_{it} - \bar{\alpha}_t)$ is zero. It follows that the log-likelihood (3.25) has essentially the form of the log-likelihood of a univariate model. The results reported for (3.23a) also applies to (3.25).

One obvious problem with the above decomposition is that, although the transformed variances are always non negative, the original ones may well be negative. The above decomposition may still be useful to form the log-likelihood which is then maximised jointly with respect to all the variances. On the other hand, if both the exogenous variables and the vector of coefficients of the exogenous variables are unit specific, the decomposition does not lead to the independent estimation of the models. However, if a stepwise procedure is carried out, the joint maximisation of (3.23a) and (3.25) with respect to θ is always possible, and two of the variances can always be concentrated out of the likelihood.

Asymptotic Tests of Hypotheses

Under regularity conditions,

$$(3.27) \qquad T^{\frac{1}{2}} \begin{bmatrix} \tilde{\beta} - \beta \\ \\ \tilde{\theta} - \theta \end{bmatrix} \underset{\sim}{d} N \left[0 , \begin{bmatrix} T\, I^{*-1}(\beta) & 0 \\ \\ 0 & T\, I^{*-1}(\theta) \end{bmatrix} \right],$$

where the symbol d represent asymptotic distribution, $\tilde{\beta}$ and $\tilde{\theta}$ are the values that maximise the frequency domain log-likelihood, and $I^{*}(\beta)$ and $I^{*}(\theta)$ are consistently estimated by (3.5b) and (3.19) evaluated at $(\tilde{\beta}, \tilde{\theta})$ respectively.

For hypotheses concerning the vector of exogenous variable

coefficients β, a standard Wald statistic can be formed from (3.27). As a special case, when (1.4) holds with $D = I_n$ and the vector of exogenous variables is the same for all units, the hypothesis

$$(3.28) \qquad H_0 : \beta_1 = \beta_2 = \ldots = \beta_n,$$

is tested using only the (n-1)-dimensional homogenous model (3.7b). The hypothesis (3.28) being equivalent to the hypothesis that in the mentioned system, all the coefficients are zero. On the other hand, if both the vector of exogenous variables and the vector of coefficients are the same for all units, the Wald statistic to test a hypothesis concerning the common vector of coefficients β is obtained by estimating only the univariate model (3.13a).

With respect to the parameters in the variance covariance matrices of the random shocks in the model, there are two test of hypothesis of obvious interest. The first test compares the general error components model (1.3) with a completely unrestricted multivariate structural time series model. The second test takes the restriction (1.4) as the null hypothesis and the general model (1.3) as the alternative. In both cases it is convenient to formulate an LM test because it is always easy to estimate the restricted model. Using results in Section 3.5 and in Fernandez-Macho (1986, sec. 3.5), the LM statistic to test (1.3) against an unrestricted multivariate structural time series model is

$$(3.29a) \qquad LM = \tfrac{1}{2} \left[\sum_t h_t\right]' \left[\sum_t H_t\right]^{-1} \left[\sum_t h_t\right],$$

where,

$$(3.29b) \qquad h_t = X_t \, \mathrm{vec}\left[G_t^{-1} - G_t^{-1} P_t G_t^{-1}\right],$$

$$(3.29c) \qquad H_t = X_t \left[G_t^{-1} \otimes G_t^{-1}\right] X_t',$$

(3.29d) $$X_t^{\prime} = \bar{D}\ (x_t^{\prime} \otimes I_{n(n+1)/2})$$

$\bar{D}$ is the (n^2 x n(n+1)/2) duplication matrix defined in Magnus(1988, ch. 4), and G_t^{-1} and x_t were defined in (3.6) and (3.16a) respectively. In the basic structural model, (1.3) contains 4 variance covariance matrices and the degrees of freedom for the LM statistic (3.29) are 4(n+1)(n-2).

To test the restriction (1.4), the LM statistic has the same form (3.29) but with $X_t^{\prime}$ defined as in (3.15) and G_t^{-1} given by (3.6) with D_t^{*} = g_t^{*} D. If D in (1.4) is known, and for the basic structural model, the degrees of freedom for this LM statistic are 4(n-1).

6.4. Analysis of Labour Cost Time Series : An Application

This section illustrates the techniques presented in previous sections, using quarterly time series for the logarithms of the unit labour costs in Austria, Belgium and Luxemburg, and Netherlands for the period 1970 Q1 to 1987 Q4. The original data is in the index form with base 1985 = 100, and published in Department of Trade and Industry (1988). Figure 6.4.1 presents a graph of the three time series in logarithms.

Univariate structural time series models were first fitted to each series. The results are reported in Table 6.4.1. A local level model with a fixed slope was appropriate in all the cases, and the variances of the irregular and level random shocks were similar for the three time series. Table 6.4.1 also presents some diagnostics for the residuals. The Normality statistic defined by Bowman and Shenton (1975) has a chi-square distribution with 2 degrees of freedom. The standard serial correlation statistic of Ljung and Box (1978) is also reported and none of the diagnostics present evidence of misspecification. This preliminary analysis suggests that an error components local level model, with equal variances for the specific random shocks might be appropriate. The first order serial correlation of the deviations of the time series from the average were .63, .75 and .84. From the results in Section 5.2, these high correlations reveal that the error components model type I may be more appropriate than the error components model type II. In terms of the notation in previous sections, (1.4) with D equal to the identity matrix was assumed. The fixed slope in the model takes the form of the coefficients of an exogenous variable, time, which is the same for all units. The vector of coefficients was assumed unit specific.

The estimation results of the multivariate model, using the frequency domain approach, and some diagnostics for the residuals are reported in Table 6.4.2. The Normality statistics, and also the autocorrelation of the residuals, which are not reported in the table, show no significance evidence of misspecification in the model.

Turning now to the estimation of the parameters, the three slopes are significant at the 5% level and they are very similar to each other. With respect to the variances, the estimate of σ_ϵ^2 is zero, suggesting that the irregular components are not correlated. The other three estimated variances are significant, and the correlation of the level random shocks is estimated at .61.

Three test of hypothesis were considered. In first place, the estimated model was compared with a totally unrestricted multivariate local level model using the LM statistic (3.29), which in this case has 8 degrees of freedom. The value of the LM statistic was 11.66; and then, at the 5% significance level, the restricted model is accepted. The estimated model was also tested against the alternative hypothesis that D in (1.4) is not the identity matrix using a likelihood ratio test. The statistic was .28 and again the restricted model is accepted. Finally, the hypothesis that the three slopes are the same was tested. This test requires only the estimation of a bivariate homogeneous model of two deviation from the average of the observations across units. The Wald statistic for this hypothesis was .13 and then the hypothesis of equal slopes is accepted. From Section 6.3, the estimation of the common slope is obtained only from the univariate model for the average. The estimate of this common slope is 9.60×10^{-3}.

Estimates of the trends are easily obtained using the formulas in Section 6.2.

Figure 6.4.1 : Labour Cost Time Series (in logarithms)

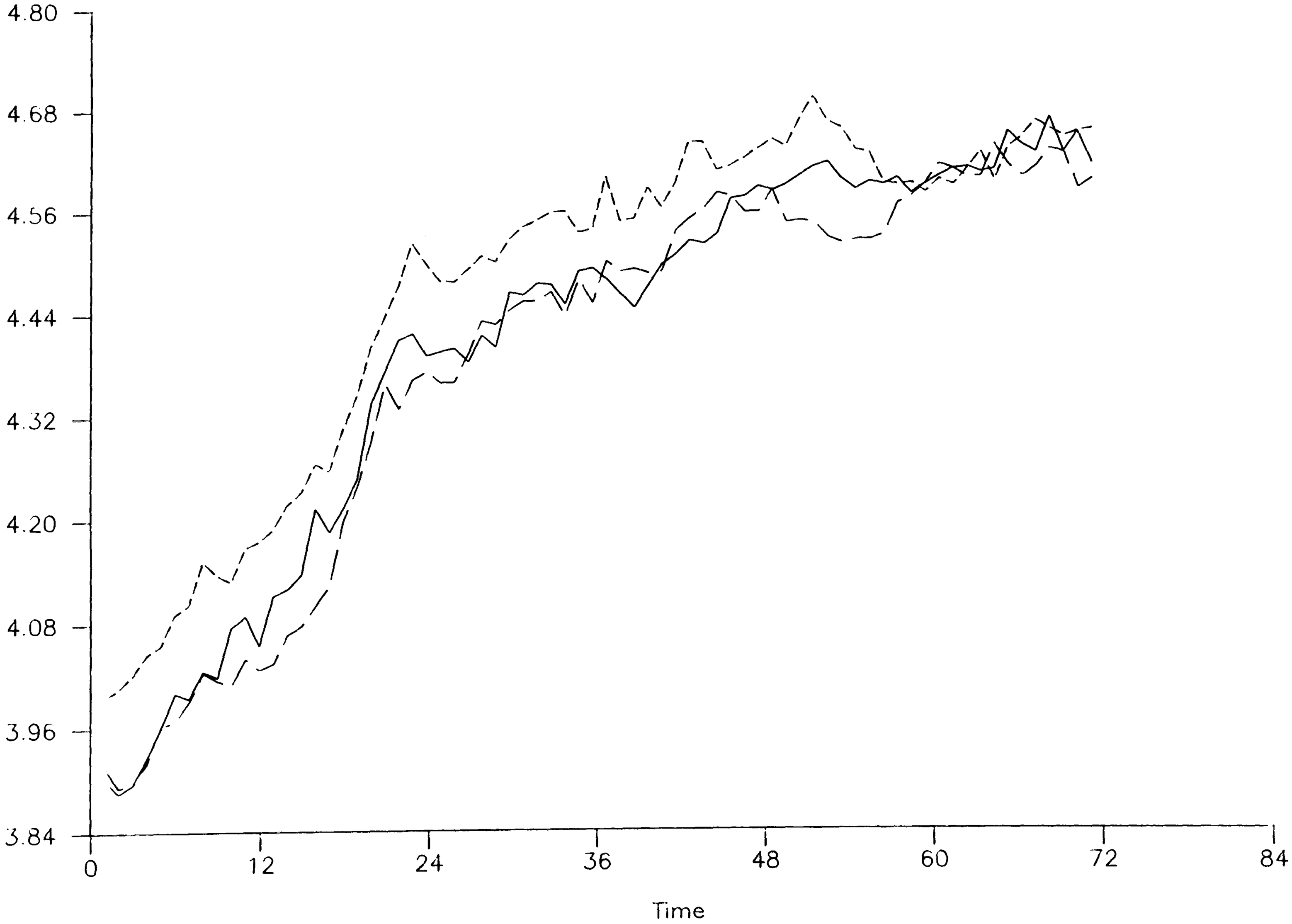


Table 6.4.1 : Estimation Results for Univariate Models(1)

Parameter	Austria	Belgium and Luxemburg	Netherlands
σ_ϵ^2 (x10^{-3})	.048	.001	.000
	(.079)	(.068)	(---)
σ_η^2 (x10^{-3})	.575	.575	.583
	(.176)	(.167)	(.098)
β (x10^{-2})	1.002	.956	.926
	(.284)	(.284)	(.286)
Normality	5.07	1.15	.38
Serial Correlation(12)	18.92	15.01	13.43

(1) Standard errors in parenthesis. Normality is chi-square(2) and Serial Correlation is chi-square(11) if β is known.

Table 6.4.2 : Estimation Results for the Multivariate Model(1)

Parameter	Estimate	S. Error
β_a (x10^{-2})	1.002	.234
β_b (x10^{-2})	.956	.234
β_n (x10^{-2})	.926	.234
σ_ϵ^2 (x10^{-3})	.000	--
σ_ϵ^{2*}(x10^{-3})	.115	.040
σ_η^2 (x10^{-3})	.249	.067
σ_η^{2*}(x10^{-3})	.159	.073
Normality (a)	3.57	
(b)	.83	
(n)	.72	

(1) Austria : a; Belgium and Luxemburg : b; Netherlands : n. Normality is chi-square (2).

CHAPTER 7 : ESTIMATION OF DYNAMIC ERROR COMPONENTS MODELS TYPE II

7.1 Introduction

This chapter considers the estimation of the dynamic error components model type II presented in Section 5.1. Assume that the observation y_{it}, for the unit i and time t, is generated by

$$(1.1) \qquad y_{it} = z_{it}^{1} \beta_i + \alpha_{it}, \qquad \begin{matrix} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{matrix}$$

where z_{it} is an (r x 1) vector of exogenous variables, β_i is an (r x 1) vector of fixed parameters and α_{it} is a residual defined below. Special cases of (1.1) include the situations where the vector of exogenous variables is the same for all units, $z_{it} = z_t$ for all i, and the situations where the vector of coefficients is the same for all units, $\beta_i = \beta$ for all i. Model (1.1) can be written in the multivariate form

$$(1.2) \qquad y_t = X_t \beta + \alpha_t, \qquad t = 1,\ldots,T,$$

where y_t and α_t are (n x 1) vectors with i-th component y_{it} and α_{it} respectively, X_t is an (n x k) matrix which contains the exogenous variables z_{it}, and β is a (k x 1) vector which contains the parameters β_i. The relationship between (X_t, β, k) and (z_{it}, β_i, r), for the general specification (1.1) as well as for the special cases mentioned below (1.1), was presented in Section 5.1.

The vector of residuals α_t in (1.2) is assumed to be generated by a structural time series model of the form (5.1.7). That is,

$$(1.3a) \qquad \alpha_t = (Z \otimes I_n)\, \theta_t + \iota\, \epsilon_t + \epsilon_t^{*}, \qquad t = 1,\ldots,T,$$

$$(1.3b) \qquad \theta_t = (T \otimes I_n)\, \theta_{t-1} + (R \otimes I_n)\, (I_u \otimes \iota)\, \kappa_t,$$

where the (np x 1) state vector θ_t contains the unobserved trend and seasonal components, ι is an (n x 1) vector of ones, I_n is the identity of orden n, and the vector random shocks ϵ_t, ϵ_t^*, and κ_t have dimensions 1, n and u; and they are assumed to be normally distributed, with expected values equal to zero and variances σ_ϵ^2, D_ϵ^* and D_κ respectively. The last two matrices being diagonal. The diagonal matrix D_ϵ^* is assumed to be strictly positive definite and, according to the time series model, some elements in the diagonal matrix D_κ are also assumed to be strictly positive; see Section 5.1 for details.

In (1.3), the trend and seasonal components for all units are generated by the common random shock κ_t. However, this does not mean that the seasonal and trend components are the same for all units because there may be initial differences between these trend and seasonal components. Two alternative specifications for the initial differences are studied: (i) the initial differences between the trend and seasonal components can be defined as fixed parameters, or (ii) the initial differences can be defined as random coefficients with a given distribution. Whichever the initial specification for the state vector θ_t, model (1.3) can be written as

$$\alpha_t = U_t \lambda + \iota Z \theta_t^* + \iota \epsilon_t + \epsilon_t^*, \qquad t = 1,\ldots,T, \tag{1.4a}$$

$$\theta_t^* = T \theta_{t-1}^* + R \kappa_t, \tag{1.4b}$$

where the state vector θ_t^* has now dimension (p x 1) and represents the common to all units trend and seasonal components. The (n x np) matrix U_t is known for all t and it was defined in (5.1.12), while the (np x 1) vector λ represents the initial specifications for the trend and seasonal components. Thus, according with this initial specification, λ may be a random component or a fixed vector of parameters. In the fixed effects case, model (1.4) requires a restriction for the

identifiability of the state vector θ_t^*. If the sum of the individual effects λ_i across units is chosen to be equal to zero, the state vector θ_t^* in (1.4b) is equal to the average of the state vectors θ_{it} defined in (1.3b); see (5.2.16).

Models (1.3) and (1.4) are said to be restricted if they satisfy the condition

$$(1.5) \qquad D_\epsilon^* = \sigma_\epsilon^{2*} I_n,$$

where σ_ϵ^{2*} is assumed to be greater than zero.

The chapter is organised as follows. Section 7.2 presents the estimation of the unobserved trend and seasonal components under the two initial specifications. That section also compares the results obtained under these two initial definitions and analyses the efficiency of the estimators with respect to both the number of units and the number of time observations. Section 7.3 considers the maximum likelihood estimation of the vector of coefficients β and of the variances of the random shocks. The formulation of asymptotic tests of hypotheses is also considered in that section. Finally, Section 7.4 illustrates the techniques presented with an empirical application.

7.2 Estimation of the Unobserved Components

This section considers the estimation of the unobserved trend and seasonal effects in the error components model type II defined by (1.3) or (1.4), assuming α_t is observed and all the variances of the random shocks defined in the model are known. The estimation of the unobserved components, by the Kalman filter, under the two initial conditions defined in the introduction of this chapter are considered first. Then, the results obtained under the two initial specifications are compared. Finally, the efficiency of the estimators as a function of both the number of units and the number of time observations is analysed.

Fixed Initial Differences

The initial differences between the trend and seasonal components across units may be defined as fixed constants to be estimated, jointly with the vector of coefficients β and the variances of the random shocks, using the maximum likelihood principle; see Section 7.3. In that case, these constants are assumed to be known at this stage and the objective in this subsection is to present estimates of the state vectors θ_t^*, t = 1,...,T, in (1.4).

Assume

$$(2.1) \qquad \theta_0^* = \xi,$$

where ξ is a (p x 1) random vector defined as diffuse. There are two ways in which estimates of θ_t^* can be obtained. One possibility is to apply the diffuse Kalman filter defined in Section 1.4. That procedure involves two steps. First, the standard Kalman filter for the state space model (1.4) is run conditional on $\xi = 0$. The second step computes

the estimates of θ_t^* unconditional on ξ, and their mean square errors. The standard Kalman filter can also be applied over (1.4) and that may be simpler. As (1.4) does not have the form of the models studied in Chapter 1, the results developed in Section 1.3 to form initial conditions, and the Kalman filter equations (1.3.2), cannot be applied here. The following lines show how initial values for the state vector are obtained, and present the Kalman filter recursions for (1.4).

For the model (1.4), equation (1.3.4) can be written as

$$(2.2a) \qquad \alpha = (H \otimes \iota)\ \theta_p + [(H_p\ \kappa_p + \ldots + H_2\ \kappa_2) \otimes \iota]$$

$$+ [\begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \epsilon_p \end{bmatrix} \otimes \iota] + \begin{bmatrix} \epsilon_1^* \\ \cdot \\ \cdot \\ \epsilon_p^* \end{bmatrix},$$

where α is an (np x 1) vector with k-th component α_k; and $H, H_p, \ldots, H_2$ are (p x p) matrices as defined in (1.3.4). In a more compact form,

$$(2.2b) \qquad \alpha = (H \otimes \iota)\ \theta_p + e,$$

where, $E(e) = 0$ and

$$(2.3) \qquad V(e) \equiv V = (P \otimes \iota\iota') + (I_p \otimes D_\epsilon^*).$$

From (2.2) and (2.3) follows that the minimum mean square error estimator of θ_p^*, and its mean square error, are given by

$$(2.4a) \qquad m_p^* = [(H' \otimes \iota')\ V^{-1}\ (H \otimes \iota)]^{-1}\ [(H' \otimes \iota')\ V^{-1}\ \alpha],$$

and

$$(2.4b) \qquad P_p^* = [(H' \otimes \iota')\ V^{-1}\ (H \otimes \iota)]^{-1}.$$

Although P in (2.3) is not strictly positive definite, it is semi positive definite, and so the inverse of V can be obtained using Lemma

2.2 in Magnus (1982). That is,

$$(2.5)\quad V^{-1} = (I_p \otimes D_\epsilon^{*-1}) - (1/\theta)\left[I_p - (Ip + \theta\ P)^{-1}\right] \otimes (D_\epsilon^{*-1}\ \iota\iota'\ D_\epsilon^{*-1}),$$

with $\theta = (\iota'\ D_\epsilon^{*-1}\ \iota)$. Thus, the estimates of the state vectors, and their mean square error are obtained applying the standard Kalman filter equations from t = (p+1) and with the starting values given by (2.4). From Anderson and Moore (1979, sec. 3.1) or Harvey (1989, sec. 3.2), and using the definitions (1.3.1) with a * for the state vector θ_t^* , the Kalman filter recursions for the model (1.4) are

$$(2.6a)\qquad \bar{P}_t^* = T\ P_{t-1}^*\ T' + R\ D_k\ R',$$

$$(2.6b)\qquad F_t^* = (Z\ \bar{P}_t^*\ Z' + \sigma_\epsilon^2)\ \iota\iota' + D_\epsilon^*,$$

$$(2.6c)\qquad P_t^* = \bar{P}_t^* - \bar{P}_t^*\ Z'\ \iota'\ F_t^{*-1}\ \iota\ Z\ \bar{P}_t^*,$$

$$(2.6d)\qquad m_t^* = T\ m_{t-1}^* + \bar{P}_t^*\ Z'\ \iota'\ F_t^{*-1}\ v_t^*,$$

$$(2.6e)\qquad v_t^* = \alpha_t - \iota\ Z\ T\ m_{t-1}^*,$$

where ι is a vector of ones, $\bar{P}_t^*$ and P_t^* are (p x p) matrices, F_t^* is an (n x n) matrix, m_t^* is a (p x 1) vector, and v_t^* is an (n x 1) vector. From the definitions (1.3.1), v_t^* is the one step ahead prediction error at time t and F_t^* its variance; while m_t^* is the estimate of the state vector at time t conditional on information up to time t and P_t^* its mean square error. Notice that the inverse of the (n x n) matrix F_t^* in (2.6b), which is needed to compute P_t^* in (2.6c), and m_t^* in (2.6d), can be obtained analytically; see equations (2.9) below.

If (1.5) holds, the estimators of the state vectors θ_t^* and their mean square errors can also be computed by running a univariate Kalman filter for the average across units $\bar{\alpha}_t$. This result can be shown by

using the transformation (5.2.17), or directly by using the Kalman filter equations above. With the second alternative, (2.5) under (1.5) is equal to

$$(2.7) \qquad V^{-1} = (I_p \otimes \frac{1}{\sigma_\epsilon^{2*}} I_n) - \frac{1}{n\,\sigma_\epsilon^{2*}} [I_p - (I_p + \frac{n}{\sigma_\epsilon^{2*}} P)^{-1}] \otimes \iota\iota'.$$

Then, replacing this matrix in (2.11), the initial conditions become

$$(2.8a) \qquad m_p^* = [H' (I_p \otimes \iota') V^{-1} (I_p \otimes \iota) H][H' (I_p \otimes \iota') V^{-1} \alpha]$$

$$= [H' (\frac{\sigma_\epsilon^{2*}}{n} I_p + P)^{-1} H]^{-1} [H' (\frac{\sigma_\epsilon^{2*}}{n} I_p + P)^{-1} \otimes \frac{1}{n} \iota] \alpha$$

$$= (H^{-1} \otimes (1/n)\ \iota)\ \alpha,$$

and

$$(2.8b) \qquad P_p^* = H^{-1} (\frac{\sigma_\epsilon^{2*}}{n} I_p + P)\ H'^{-1}.$$

The vector m_p^* and the matrix P_p^*, in (2.8), correspond exactly to the initial estimators in a univariate model for the average of the observations, with the variance of the irregular random shock equal to $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n)$. Hence, under (1.5), the initial conditions are formed using only the average of the observations from t = 1 to t = p. To show that this is also true for t = p+1,...,T, notice that from (2.6)

$$(2.9a) \qquad F_t^* = (Z\ \bar{P}_t^*\ Z' + \sigma_\epsilon^2)\ \iota\iota' + \sigma_\epsilon^{2*}\ I_n$$

$$= f_t\ \iota\iota' + \sigma_\epsilon^{2*}\ I_n,$$

and then,

$$(2.9b) \qquad F_t^{*-1} = \frac{1}{\sigma_\epsilon^{2*}} I_n - \frac{1}{(\sigma_\epsilon^{2*})^2 \, (f_t^{-1} + n \, (\sigma_\epsilon^{2*})^{-1})} \iota\iota',$$

from where it follows that

$$(2.10a) \qquad \iota' \, F_t^{*-1} \, \iota = \frac{1}{f_t + \sigma_\epsilon^{2*}/n},$$

and

$$(2.10b) \qquad \iota' \, F_t^{*-1} = \frac{1}{n \, (f_t + \sigma_\epsilon^{2*}/n)}.$$

Replacing (2.10a) and (2.10b) into (2.6c) and (2.6d) respectively, is easy to see that the recursions (2.6) are equal to the recursions of a univariate model for $\bar{\alpha}_t$. In fact, a similar result is obtained if (1.5) does not hold. In that case, the variance covariance matrix of the irregular random shock can be written as $(D \, \sigma_\epsilon^{2*})$, with $(\iota' \, D^{-1} \, \iota) = n$; and it can be shown that the estimates of the state vectors, and their mean square errors, are obtained by running a univariate Kalman filter over the time series $(\iota' D^{-1}/n) \, \alpha_t$.

Finally, an alternative approach to the estimation of the state vectors is obtained by transforming the observations α_t by the matrix H in (5.2.4) if (1.5) holds; or the matrix H^* in (5.2.11) if (1.5) does not hold. Under (1.5), and assuming that the sum of the unit effects λ_i, $i = 1,\ldots,n$, is equal to zero, that gives

$$(2.11a) \qquad \bar{\alpha}_t = Z \, \theta_t^* + \epsilon_t + \epsilon_t^*, \qquad t = 1,\ldots,T,$$

$$(2.11b) \qquad \theta_t^* = T \, \theta_{t-1}^* + R \, \kappa_t,$$

and

$$(2.11c) \qquad (\alpha_{it} - \bar{\alpha}_t) = Z\, T^t\, \lambda_i + (\epsilon^*_{it} - \bar{\epsilon}^*_t), \qquad \begin{matrix} i = 1,\ldots,n-1, \\ t = 1,\ldots,T; \end{matrix}$$

see equations (5.1.11). Alternatively, (2.11c) can be written in the state space form

$$(2.12a) \qquad (\alpha_{it} - \bar{\alpha}_t) = Z\, \lambda_{it} + (\epsilon^*_{it} - \bar{\epsilon}^*_t), \qquad \begin{matrix} i = 1,\ldots,n-1, \\ t = 1,\ldots,T, \end{matrix}$$

$$(2.12b) \qquad \lambda_{it} = T\, \lambda_{i,t-1}.$$

These expressions make clear that the estimates of the state vector θ^*_t require only the average of the observations; while the fixed effects λ_i are estimated from the stationary model (2.11c), or from the state space model (2.12). As noted in the introduction, if the sum of the effects λ_i across units is zero, θ^*_t is equal to the average of the state vectors θ_{it} defined in (1.3b).

Random Initial Differences

If the initial differences between the trend and seasonal components in model (1.3) are defined as random with a proper distribution, the estimates of the state vector, and their mean square errors, can be obtained by means of the diffuse Kalman filter although there are situations where the standard Kalman filter can also be applied. Assume that the state vector at time zero is defined as

$$(2.13) \qquad \theta_0 = \lambda + (I_p \otimes \iota)\, \xi,$$

where λ is an (np x 1) normal random vector with expected value zero and variance covariance matrix Σ_λ, I_p is the identity of order p, ι is an (n x 1) vector of ones, and ξ is a (p x 1) random vector defined as diffuse. The vector λ in (2.13) correponds exactly to the one in

(1.4a), and typically, its variance covariance matrix Σ_λ will be defined as block diagonal with blocks of the form $(\sigma_k^2 I_n)$, $k = 1,\ldots,p$. The vector ξ in (2.13) represents the diffuseness of the common trend and seasonal components, while the matrix that premultiply ξ considers the fact that these trend and seasonal components are common to all units. For example, in the local level model, $\theta_0 = \lambda + \iota\,\xi$, with ξ an scalar; and if $\xi = 0$ and $D_K = 0$, model (1.3) reduces to the static three error components model (5.1.1).

Estimates of θ_t in the general model defined by (1.3) and (2.13) are obtained by applying the filter procedure presented in Section 1.4. If Σ_λ is defined as

$$\Sigma_\lambda = D_\lambda \otimes I_n, \tag{2.14}$$

with D_λ a (p x p) diagonal matrix, the first step in the diffuse Kalman filter procedure is obtained by running the standard Kalman filter defined in equations (1.3.2) from t = 1 and with

$$m_0 = 0, \tag{2.15a}$$

$$P_0 = \Sigma_\lambda = D_\lambda \otimes I_n, \tag{2.15b}$$

$$\Sigma_\epsilon = (\sigma_\epsilon^2 \otimes \iota\iota') + D_\epsilon^*, \tag{2.15c}$$

and

$$\Sigma_K = D_K \otimes \iota\iota'. \tag{2.15d}$$

The estimators of the state vectors, unconditional on ξ, and their mean square errors, are then obtained with a straightforward application of the second step defined in Section 1.4. As in the fixed initial differences specification, the unconditional estimates of the state vector and their mean square errors exist only from t = p.

If (1.5) and (2.14) hold, the estimates of the state vectors, and their mean square errors, can also be obtained by means of the standard Kalman filter and that leads to simpler formulas. To show this, apply the transformation (5.2.17) to obtain

$$(2.17a) \qquad \bar{\alpha}_t = Z\,\bar{\theta}_t + \epsilon_t + \bar{\epsilon}^*_t, \qquad t = 1,\ldots,T,$$

$$(2.17b) \qquad \bar{\theta}_t = T\,\bar{\theta}_{t-1} + R\,\kappa_t,$$

$$(2.17c) \qquad (\alpha_{it} - \bar{\alpha}_t) = Z\,(\theta_{it} - \bar{\theta}_t) + (\epsilon^*_{it} - \bar{\epsilon}^*_t), \qquad \begin{matrix} i = 1,\ldots,n-1, \\ t = 1,\ldots,T, \end{matrix}$$

$$(2.17d) \qquad (\theta_{it} - \bar{\theta}_t) = T\,(\theta_{i,t-1} - \bar{\theta}_{t-1}).$$

As the stochastic components in (2.17a) and (2.17b) are uncorrelated with the stochastic components in (2.17c) and (2.17d), $\bar{\theta}_t$, $t = p,\ldots,T$ can be estimated using the standard Kalman filter for the univariate model (2.17a)-(2.17b); while $(\theta_{it} - \bar{\theta}_t)$, $t = 1,\ldots,T$ can also be estimated by means of the standard Kalman filter for the model (2.17c) and (2.17d) in which the state vector has initial expected value zero and mean square error equal to

$$(2.18) \qquad P_0 = D_\lambda \otimes (I_n - (1/n)\,\iota\iota').$$

Notice that, although $(\theta_{it} - \bar{\theta}_t)$ can be estimated for all t, θ_{it} is only estimable from t = p. In more general situations, where either (1.5) or (2.14) does not hold, the orthogonality in (2.17) is lost.

Comparison of the Initial Conditions

This subsection compares the results obtained in the estimation of the state vector θ_t in (1.3) under the two alternative initial conditions studied in previous subsections. To simplify the analysis

assume that (1.5) and (2.14) hold. From (1.3), take the average across units and (n-1) deviations to obtain

$$(2.19a) \qquad \bar{\alpha}_t = Z\,\bar{\theta}_t + \epsilon_t + \bar{\epsilon}^*_t, \qquad t = 1,\ldots,T,$$

$$(2.19b) \qquad \bar{\theta}_t = T\,\bar{\theta}_{t-1} + R\,\eta_t,$$

and

$$(2.20a) \qquad (\alpha_{it} - \bar{\alpha}_t) = Z\,(\theta_{it} - \bar{\theta}_t) + (\epsilon^*_{it} - \bar{\epsilon}^*_t), \qquad \begin{matrix} i = 1,\ldots,n-1, \\ t = 1,\ldots,T, \end{matrix}$$

$$(2.20b) \qquad (\theta_{it} - \bar{\theta}_t) = T\,(\theta_{it-1} - \bar{\theta}_{t-1}).$$

Previous results in this section showed that the estimates of the components θ_{it} in (1.3) can be obtained by estimating first $\bar{\theta}_t$ from (2.19) and then $(\theta_{it} - \bar{\theta}_t)$ from (2.20). From (2.19) follows that the estimates of $\bar{\theta}_t$ are the same under the two alternative initial specifications. The differences between the estimators under the two approaches are generated only by (2.20).

Model (2.20) can be written in the multivariate form

$$(2.21a) \qquad \alpha^{**}_t = (Z \otimes I_{n-1})\,\theta^{**}_t + \epsilon^{**}_t, \qquad t = 1,\ldots,T,$$

$$(2.21b) \qquad \theta^{**}_t = (T \otimes I_{n-1})\,\theta^{**}_{t-1},$$

where α^{**}_t and ϵ^{**}_t are ((n-1) x 1) vectors with i-th component equal to $(\alpha_{it} - \bar{\alpha}_t)$ and $(\epsilon^*_{it} - \bar{\epsilon}^*_t)$ respectively, and θ^{**}_t is the ((n-1)p x 1) state vector. Alternatively,

$$(2.22) \qquad \alpha^{**}_t = (Z\,T^t \otimes I_{n-1})\,\theta^{**}_0 + \epsilon^{**}_t$$

$$= X^{**}_t\,\theta^{**}_0 + \epsilon^{**}_t, \qquad t = 1,\ldots,T,$$

with X^{**}_t an ((n-1) x (n-1)p) matrix. If θ^{**}_0 has an initial distribution

with expected value zero and variance covariance matrix equal to Σ_θ, the minimum mean square estimator of θ_0^{**} conditional on information up to time $t \geq p$, m_t^{**}, and its mean square error, P_t^{**}, are given by

$$(2.23a) \qquad m_t^{**} = \left[\Sigma_\theta^{-1} + (\sigma_\epsilon^{2*})^{-1} \sum_{s=1}^{t} X_s^{*\prime} (I_{n-1} - \iota\iota'/n)^{-1} X_s^*\right]^{-1}$$

$$\left[(\sigma_\epsilon^{2*})^{-1} \sum_{s=1}^{t} X_s^{*\prime} (I_{n-1} - \iota\iota'/n)^{-1} \alpha_s^*\right],$$

and

$$(2.23b) \qquad P_t^{**} = \left[\Sigma_\theta^{-1} + (\sigma_\epsilon^{2*})^{-1} \sum_{s=1}^{t} X_s^{*\prime} (I_{n-1} - \iota\iota'/n) X_s^*\right]^{-1},$$

see for example Theil (1971, sec. 7.8). If θ_0^{**} is defined as fixed, $\Sigma_\theta^{-1} = 0$; while the estimator of θ_0^{**} under the random initial specification, and its mean square error are obtained in (2.26) if $\Sigma_\theta = (D_\lambda \otimes (I_{n-1} - \iota\iota'/n))$.

In the simple local level time series model, $p = 1$, $Z = T = 1$, $X_t^{**} = 1$, and formulas (2.23) for the fixed initial specification reduce to

$$(2.24a) \qquad m_t^{**} = (1/t) \sum_{s=1}^{t} \alpha_s^{**},$$

and

$$(2.24b) \qquad P_t^{**} = (\sigma_\epsilon^{2*}/t) \left[I_{n-1} - (1/n)\, \iota\iota'\right],$$

while, for the random initial specification $D_\lambda = \sigma_\lambda^2$ and m_t^{**} and P_t^{**} can be computed from

$$(2.25a) \qquad m_t^{**} = m_{t-1}^{**} + \frac{1}{t + (\sigma_\epsilon^{2*} / \sigma_\lambda^2)} (\alpha_t^{**} - m_{t-1}^{**}),$$

and

$$(2.25b) \qquad P_t^{**} = \frac{\sigma_\epsilon^{2*}}{t + (\sigma_\epsilon^{2*} / \sigma_\lambda^2)} \left[I_{n-1} - (1/n)\, \iota\iota' \right],$$

and with $m_0^{**} = 0$. As t goes to infinity, the mean square error of the estimator of θ_0^{**} converges to zero irrespective of the initial conditions. For finite values of t, P_t^{**} is always smaller when the initial state vector is defined as random; compare (2.25b) with (2.24b). Notice that as σ_λ^2 increases, the model under random initial conditions approaches the model under fixed initial conditions. On the other hand, as σ_λ^2 tends to zero, the model under random initial conditions becomes a model with fixed initial conditions. However, as these fixed initial conditions are equal to zero, P_t^{**} and m_t^{**} tend to zero for all t; see (2.25). Finally, notice that (2.24) are exactly equal to the formulas under a diffuse initial distribution.

Efficiency Analysis

The following lines analyse the efficiency of the estimates of the state vector in model (1.3) assuming the local level time series model under (1.5), and assuming also that the initial differences between the level components are fixed and equal to zero. That is, the model can be written as

$$(2.29a) \qquad \alpha_{it} = \mu_t + \epsilon_t + \epsilon_{it}^*, \qquad \begin{matrix} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{matrix}$$

$$(2.29b) \qquad \mu_t = \mu_{t-1} + \eta_t,$$

where μ_t is an scalar which represents the common to all units level, and ϵ_t, ϵ_{it}^* and η_t are scalar random shocks, assumed to be normally

distributed, serially and mutually uncorrelated, with expected values equal to zero, and variances σ_ϵ^2, σ_ϵ^{2*} and σ_η^2 respectively.

Earlier in this section were presented the minimum mean square estimator of μ_t for given values of n and T. The analysis here considers the fact that in practice n and T have to be chosen, and so, a relevant question is what are the values of n and T needed to obtain a desired prefixed mean square error for the estimator of μ_t. It will be assumed that the interes is place in the component μ_t at time t = T, that is μ_T.

In principle, an estimate of μ_T can be obtained using a single time series from t = 1, a cross-section of n units at time t = T, or a cross-section of n time series from t = 1. Other possibilities will not be considered here. Using a cross-section of n time series from t = 1, the mean square error of the estimator of μ_t in model (2.29) can be written as

$$(2.30) \qquad P_t = (P_{t-1} + \sigma_\eta^2) - \frac{(P_{t-1} + \sigma_\eta^2)^2}{(P_{t-1} + \sigma_\eta^2 + \sigma_\epsilon^2 + (\sigma_\epsilon^{2*}/n))},$$

for t = 2,...,T; and with $P_1 = (\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n)$. From (2.30) it can also be obtained the expression for the mean square error when μ_T is estimated using a single time series, n = 1, or when μ_T is estimated using a cross-section of units at time T, T = 1. It should be notice that P_t does not converge to zero as n or T go to infinity provide σ_ϵ^2 is greater than zero. If σ_ϵ^2 is equal to zero, the mean square error (2.30) converges to zero as n goes to infinity, but not necessarily if T goes to infinity with n fixed.

In what follows, the relative efficiency of the estimate of μ_T computed with n = 1 and T = 1 with respect to the estimate of μ_T computed with n cross-sections of T observations each is evaluated

numerically for different values of n and T and for the values of the parameters shown in Table 7.2.1.

Table 7.2.1 : Parameter Values in the Evaluation of the Relative Efficiency(1)

Table	ρ_ϵ	σ_ϵ^2	σ_ϵ^{2*}	σ_η^2
7.2.2	.20	.20	.80	.20
7.2.3	.20	.20	.80	.50
7.2.4	.20	.20	.80	1.00
7.2.5	.20	.20	.80	2.00
7.2.6	.20	.20	.80	5.00
7.2.7	.80	.80	.20	.20
7.2.8	.80	.80	.20	.50
7.2.9	.80	.80	.20	1.00
7.2.10	.80	.80	.20	2.00
7.2.11	.80	.80	.20	5.00

(1) $\rho_\epsilon = \sigma_\epsilon^2 \; / \; (\sigma_\epsilon^2 + \sigma_\epsilon^{2*})$

The value of $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*})$, which corresponds to the mean square error when n = T = 1, was set equal to the unity in all the cases. Two values for the correlation coefficient across units in the irregular random shock, defined as ρ_ϵ, were considered : .20 and .80. Finally, five values for the variance of the level random shock, σ_η^2, are studied : .2, .5, 1, 2 and 5.

The results of the numerical evaluation of the relative efficiency for the ten cases presented in Table 7.2.1 are shown in tables 7.2.2 to 7.2.11. In all the cases the relative efficiency decreases as n and T increase. When n and T are equal to one hundred, the values of the relative efficiency fluctuate between .13 and .61. The smaller $(\sigma_\epsilon^2 \; / \; \sigma_\epsilon^{2*})$, the smaller the relative efficiency as n increases; and for a

given values of n, the bigger $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n)/\sigma_\eta^2$, the smaller the relative efficiency as T increases. The last statement implies that the marginal contribution to the relative efficiency of new time observations decreases as n increases. For example, in Table 7.2.2, the relative efficiency decreases from 1.00 to .36 for n equal to unity, but from .21 to .13 only when n equals one hundred.

Comparing the case of a single time series, n = 1, with the case of a cross-section of units at time T, the increase of the number of time observations in a single time series may leads to more or less efficient estimators than the increase of the number of units in a cross-section at time T. In tables 7.2.2 to 7.2.6, where σ_ϵ^2 is small (.2) and σ_ϵ^{2*} is large (.8), the relative efficiency of the estimator obtained from a cross-section at time T reaches the value .21 as the number of units equals one hundred; while the relative efficiency of the estimator from a single time series is always greater than .36 for T equal to one hundred. In other words, one hundred observations are more efficient as a cross-section than as a single time series. However, if σ_ϵ^2 is large and σ_ϵ^{2*} is small, as in tables 7.2.7 to 7.2.11, the opposite happens and one hundred observations are more efficient as a single time series.

In all the cases studied and for a fixed value of n, a relative small number of time observations, which fluctuates between 4 and 6, is needed to reach, up to two decimal figures, the limiting relative efficiency. However, for a fixed value of T, the limiting relative efficiency up to two decimal figures is not reached, in most of the cases, even for values of n equal to one hundred.

Table 7.2.2 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = .20$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.55	.39	.36	.36	.36	.36	.36	.36
2	.60	.34	.27	.26	.26	.26	.26	.26	.26
4	.40	.24	.20	.20	.20	.20	.20	.20	.20
6	.33	.21	.18	.18	.18	.18	.18	.18	.18
8	.30	.19	.17	.16	.16	.16	.16	.16	.16
10	.28	.18	.16	.16	.16	.16	.16	.16	.16
20	.24	.16	.14	.14	.14	.14	.14	.14	.14
50	.22	.14	.13	.13	.13	.13	.13	.13	.13
100	.21	.14	.13	.13	.13	.13	.13	.13	.13

Table 7.2.3 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = .50$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.60	.51	.50	.50	.50	.50	.50	.50
2	.60	.39	.35	.35	.35	.35	.35	.35	.35
4	.40	.28	.26	.26	.26	.26	.26	.26	.26
6	.33	.24	.23	.23	.23	.23	.23	.23	.23
8	.30	.22	.21	.21	.21	.21	.21	.21	.21
10	.28	.21	.20	.20	.20	.20	.20	.20	.20
20	.24	.18	.18	.18	.18	.18	.18	.18	.18
50	.22	.17	.16	.16	.16	.16	.16	.16	.16
100	.21	.16	.16	.16	.16	.16	.16	.16	.16

Table 7.2.4 ; Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = 1.00$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.67	.62	.62	.62	.62	.62	.62	.62
2	.60	.44	.42	.42	.42	.42	.42	.42	.42
4	.40	.31	.31	.31	.31	.31	.31	.31	.31
6	.33	.27	.26	.26	.26	.26	.26	.26	.26
8	.30	.24	.24	.24	.24	.24	.24	.24	.24
10	.28	.23	.23	.23	.23	.23	.23	.23	.23
20	.24	.20	.20	.20	.20	.20	.20	.20	.20
50	.22	.18	.18	.18	.18	.18	.18	.18	.18
100	.21	.18	.18	.18	.18	.18	.18	.18	.18

Table 7.2.5 ; Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = 2.00$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.75	.73	.73	.73	.73	.73	.73	.73
2	.60	.49	.48	.48	.48	.48	.48	.48	.48
4	.40	.34	.34	.34	.34	.34	.34	.34	.34
6	.33	.29	.29	.29	.29	.29	.29	.29	.29
8	.30	.27	.26	.26	.26	.26	.26	.26	.26
10	.28	.25	.25	.25	.25	.25	.25	.25	.25
20	.24	.22	.22	.22	.22	.22	.22	.22	.22
50	.22	.20	.20	.20	.20	.20	.20	.20	.20
100	.21	.19	.19	.19	.19	.19	.19	.19	.19

Table 7.2.6 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .20 \quad \sigma_\epsilon^{2*} = .80 \quad \sigma_\eta^2 = 5.00$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.86	.85	.85	.85	.85	.85	.85	.85
2	.60	.54	.54	.54	.54	.54	.54	.54	.54
4	.40	.37	.37	.37	.37	.37	.37	.37	.37
6	.33	.31	.31	.31	.31	.31	.31	.31	.31
8	.30	.28	.28	.28	.28	.28	.28	.28	.28
10	.28	.27	.27	.27	.27	.27	.27	.27	.27
20	.24	.23	.23	.23	.23	.23	.23	.23	.23
50	.22	.21	.21	.21	.21	.21	.21	.21	.21
100	.21	.20	.20	.20	.20	.20	.20	.20	.20

Table 7.2.7 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = .20$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.55	.39	.36	.36	.36	.36	.36	.36
2	.90	.49	.36	.34	.34	.34	.34	.34	.34
4	.85	.47	.34	.33	.32	.32	.32	.32	.32
6	.83	.46	.34	.33	.32	.32	.32	.32	.32
8	.82	.46	.34	.32	.32	.32	.32	.32	.32
10	.82	.45	.33	.32	.32	.32	.32	.32	.32
20	.81	.45	.33	.32	.32	.31	.31	.31	.31
50	.80	.45	.33	.32	.31	.31	.31	.31	.31
100	.80	.45	.33	.31	.31	.31	.31	.31	.31

Table 7.2.8 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = .50$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.60	.51	.50	.50	.50	.50	.50	.50
2	.90	.55	.47	.47	.47	.47	.47	.47	.47
4	.85	.52	.45	.45	.45	.45	.45	.45	.45
6	.83	.51	.45	.44	.44	.44	.44	.44	.44
8	.82	.51	.44	.44	.44	.44	.44	.44	.44
10	.82	.51	.44	.44	.44	.44	.44	.44	.44
20	.81	.50	.44	.43	.43	.43	.43	.43	.43
50	.80	.50	.43	.43	.43	.43	.43	.43	.43
100	.80	.50	.43	.43	.43	.43	.43	.43	.43

Table 7.2.9 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = 1.00$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.67	.62	.62	.62	.62	.62	.62	.62
2	.90	.61	.57	.57	.57	.57	.57	.57	.57
4	.85	.58	.55	.55	.55	.55	.55	.55	.55
6	.83	.57	.54	.54	.54	.54	.54	.54	.54
8	.82	.57	.54	.54	.54	.54	.54	.54	.54
10	.82	.57	.53	.53	.53	.53	.53	.53	.53
20	.81	.56	.53	.53	.53	.53	.53	.53	.53
50	.80	.56	.53	.53	.53	.53	.53	.53	.53
100	.80	.55	.53	.53	.53	.53	.53	.53	.53

Table 7.2.10 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = 2.00$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.75	.73	.73	.73	.73	.73	.73	.73
2	.90	.69	.67	.67	.67	.67	.67	.67	.67
4	.85	.65	.64	.64	.64	.64	.64	.64	.64
6	.83	.64	.63	.63	.63	.63	.63	.63	.63
8	.82	.64	.63	.63	.63	.63	.63	.63	.63
10	.82	.64	.62	.62	.62	.62	.62	.62	.62
20	.81	.63	.62	.62	.62	.62	.62	.62	.62
50	.80	.62	.61	.61	.61	.61	.61	.61	.61
100	.80	.62	.61	.61	.61	.61	.61	.61	.61

Table 7.2.11 : Relative Efficiency of Estimates Using a Single Time Series Compared to Estimates Using n Time Series

($\sigma_\epsilon^2 = .80 \quad \sigma_\epsilon^{2*} = .20 \quad \sigma_\eta^2 = 5.00$)

	T								
n	1	2	4	6	8	10	20	50	100
1	1.00	.86	.85	.85	.85	.85	.85	.85	.85
2	.90	.78	.78	.78	.78	.78	.78	.78	.78
4	.85	.74	.74	.74	.74	.74	.74	.74	.74
6	.83	.73	.73	.73	.73	.73	.73	.73	.73
8	.82	.72	.72	.72	.72	.72	.72	.72	.72
10	.82	.72	.72	.72	.72	.72	.72	.72	.72
20	.81	.71	.71	.71	.71	.71	.71	.71	.71
50	.80	.71	.70	.70	.70	.70	.70	.70	.70
100	.80	.70	.70	.70	.70	.70	.70	.70	.70

7.3 Maximum Likelihood Estimation of the Parameters

This section considers the maximum likelihood estimation and the formulation of tests of hypotheses for the vector of coefficients β and the variances of the random shocks in the model defined by (1.2) and (1.3), or (1.2) and (1.4). Some general considerations are presented first. The analysis of some simple models, and their generalisation, follows.

The exact log-likelihood of an error components model type II is formed by means of the diffuse Kalman filter. That leads to an expression of the form (2.1.7). From there follows that it is always possible to concentrate out of the log-likelihood the vector of coefficients β. The solution for β can then be replaced into the original log-likelihood to form a concentrated expression which has to be maximised numerically with respect to the parameters in the variance covariance matrices of the random shocks. Analytic first and second derivatives of the log-likelihood with respect to these variances are, in general, difficult to obtain.

An alternative procedure to form the log-likelihood is to differentiate the n time series to obtain the stationary form of the model; see Section 1.2. From the stationary form of the model are obtained the frequency domain log-likelihood, as well as the one defined in Section 2.3. In these cases, the log-likelihood has the form (3.1.1) and results closely related to the ones in Section 6.3 are obtained. Two problems arise with a log-likelihood formed from the stationary form of the model. First, as the models considered here are not strictly invertible, the matrices G_t defined in (3.1.3) are not always positive definite and the log-likelihood become undefined. Fernandez-Macho (1986, ch. 7) proposed a solution to this problem for

the frequency domain approach; and exactly the same idea can be used in the log-likelihood formed by means of the transformation defined in Section 2.3; although the likelihood formed by means of this transformation is in many cases defined for strictly non invertible models. The second problem with the log-likelihood formed from the stationary form of the model is that it is based on only (T-p)n observations while the exact log-likelihood based on the diffuse Kalman filter uses Tn-p observations. If n is large, the efficiency of the estimators obtained from the mentioned approximations may be low in practical situations where T is not very large. Furthermore, from the stationary form of the model the initial differences λ, or its variance covariance matrix Σ_λ under the random effects specification, cannot be estimated because these differences cancel out.

The remainder of this section concentrates on the situation where (1.5) holds, and so, the variance of the irregular random shock is the same for all units. In the random effects model, (2.14) is also assumed.

Consider first the model under the fixed initial effects. Applying the transformation (5.2.6) with the matrix H defined in (5.2.4), the log-likelihood can be decomposed as

$$(3.1) \qquad \ell = c + \ell_1 + \ell_2,$$

where c is a constant, ℓ_1 is the log-likelihood of the average of the residuals $\bar{\alpha}_t$ and ℓ_2 is the log-likelihood of (n-1) of the deviations $(\alpha_{it} - \bar{\alpha}_t)$. The log-likelihood ℓ_1 has the form of a standard univariate structural time series model with the variance of the irregular random shock equal to $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n)$. This variance can be defined as a new parameter in the maximisation procedure, and so, all the results in Chapter 3 for an unrestricted model can be applied here. On the other

hand, ℓ_2 corresponds to the log-likelihood of (2.22) which is a stationary model. If the fixed initial differences are included in the vector of coefficients β, the exact form of ℓ_2 is, apart from a constant, given by

$$(3.2) \qquad \ell_2 = -\tfrac{1}{2} T \log|(I_{n-1} - \iota\iota'/n)\,\sigma_\epsilon^{2*}| - \tfrac{1}{2}(1/\sigma_\epsilon^{2*}) \sum_{t=1}^{T} \alpha_t^{**\prime} [I_{n-1} - \iota\iota'/n]^{-1} \alpha_t^{**},$$

where

$$(3.3) \qquad \alpha_t^{**\prime} = [(\alpha_{1t} - \bar{\alpha}_t), \ldots, (\alpha_{n-1,t} - \bar{\alpha}_t)].$$

From (3.2), and using the fact that the inverse of $(I_{n-1} - \iota\iota'/n)$ is $(I_{n-1} + \iota\iota')$, follows that the maximum likelihood estimator of σ_ϵ^{2*} is given by

$$(3.4) \qquad \tilde{\sigma}_\epsilon^{2*} = \frac{1}{T(n-1)} \sum_{t=1}^{T} \sum_{i=1}^{n} (\alpha_{it} - \bar{\alpha}_t)^2.$$

The maximum likelihood estimator of σ_ϵ^2 is obtained using (3.4) and the maximum likelihood estimator of $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n)$ from ℓ_1. A problem with this procedure to obtain the estimator of σ_ϵ^2 is that there is no guarantee that it will be greater or equal to zero. In any case, the decomposition (3.1) may be used to evaluate the whole log-likelihood which is then maximised with respect to all the variances. With respect to the vector of coefficients of the exogenous variables, essentially the same results obtained in Section 6.3 apply here. If either the exogenous variables or the vector of coefficients is the same for all units, the estimators can be obtained by maximising ℓ_1 and ℓ_2 in turn. In more general situations the decomposition (3.1) can still be applied if a stepwise optimisation procedure is implemented. Assuming the

log-likelihood ℓ_1 is formed by means of the frequency domain approach or the alternative procedure presented in Section 2.3 and a basic structural time series model, an espression which is asymptotically equivalent to the information matrix with respect to the parameters $\theta' = (\sigma_\epsilon^{2*}, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_\delta^2, \sigma_\omega^2)$ can be obtained from the results in chapter 3 and from (3.2). This matrix has the form

$$(3.5)\quad I(\theta) = \tfrac{1}{2}\begin{bmatrix} \frac{T(n-1)}{\sigma_\epsilon^{2*}} & 0 \\ & \\ 0 & 0 \end{bmatrix} + \tfrac{1}{2}\sum_t \begin{bmatrix} \frac{g_{\epsilon t}}{n} \\ \\ x_t \end{bmatrix} \frac{1}{g_t^2} \begin{bmatrix} \frac{g_{\epsilon t}}{n} \\ \\ x_t \end{bmatrix}',$$

where

$$(3.6a)\qquad g_t = g_{\epsilon t}\,(\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n) + g_{\eta t}\,\sigma_\eta^2 + g_{\delta t}\,\sigma_\delta^2 + g_{\omega t}\,\sigma_\omega^2,$$

$$(3.6b)\qquad x_t' = (g_{\epsilon t},\ g_{\eta t},\ g_{\delta t},\ g_{\omega t}),$$

and $g_{\epsilon t}$, $g_{\eta t}$, $g_{\delta t}$ and $g_{\omega t}$ are known constants; see Section 1.2 for the value of these constants in the frequency domain approach, and Section 2.3 for the values in the alternative approach. Analogous results are obtained for the other time series models. The matrix $I^{-1}(\theta)$ can be used as an approximation of the variances of the maximum likelihood estimators of θ.

Consider now the model under random initial differences. The decomposition (3.1) still holds with ℓ_1 exactly as before and with ℓ_2 representing the log-likelihood of the model (2.22). The difference with the previous situation is that now θ_0^{**} is random. The log-likelihood of that model is not difficult to form; for example, if only the levels of the trends have a unit specific effect, $X_t^{**} = 1$ for all t and the log-likelihood, apart from a constant, is given by

$$(3.7)\qquad \ell_2 = -\tfrac{1}{2}\log\left|(\iota\iota'\,\sigma_\lambda^2 + I_T\,\sigma_\epsilon^{2*}) \otimes (I_{n-1} - \iota\iota'/n)\right|$$

$$-\tfrac{1}{2}\,\alpha'\left[(\iota\iota'\,\sigma_\lambda^2 + I_T\,\sigma_\epsilon^{2*}) \otimes (I_{n-1} - \iota\iota'/n)\right]^{-1}\alpha,$$

where

$$(3.8)\qquad \alpha' = (\alpha_1^{**\prime}, \ldots, \alpha_T^{**\prime}).$$

Using known formulas for the determinant and the inverse of the matrices in (3.7), ℓ_2 can be written as

$$(3.9)\qquad \ell_2 = c - \tfrac{1}{2}(n-1)(T-1)\log(\sigma_\epsilon^{2*}) - \tfrac{1}{2}(n-1)\log\left[(\sigma_\epsilon^{2*}/T) + \sigma_\lambda^2\right]$$

$$-\tfrac{1}{2}(1/\sigma_\epsilon^{2*})\;\alpha'\left[(I_T - \iota\iota'/T) \otimes (I_n + \iota\iota')\right]\alpha$$

$$-\tfrac{1}{2}\left(1 \,/\, (\sigma_\epsilon^{2*}/T + \sigma_\lambda^2)\right)\;\alpha'\left[(\iota\iota'/T^2) \otimes (I_{n-1} + \iota\iota')\right]\alpha,$$

from where, the maximum likelihood estimators of σ_ϵ^{2*} and $(\sigma_\epsilon^{2*}/T + \sigma_\lambda^2)$ are given by

$$(3.10a)\qquad \tilde{\sigma}_\epsilon^{2*} = \frac{1}{(n-1)(T-1)}\sum_{t=1}^{T}\sum_{i=1}^{n}(\alpha_{it} - \bar{\alpha}_i - \bar{\alpha}_t + \bar{\alpha})^2,$$

and

$$(3.10b)\qquad (\tilde{\sigma}_\epsilon^{2*}/T + \tilde{\sigma}_\lambda^2) = \frac{1}{(n-1)}\sum_{i=1}^{n}(\bar{\alpha}_i - \bar{\alpha})^2,$$

where

$$(3.11)\qquad (\bar{\alpha}_i - \bar{\alpha}) = (1/T)\sum_{t=1}^{T}(\alpha_{it} - \bar{\alpha}_t), \qquad i = 1,\ldots,n.$$

In more general situations where also the slopes of the trends and the seasonal components have unit specific effects, the likelihood function can be formed using the same ideas although more complicated

expressions are obtained. The maximum likelihood estimators of the parameters σ_λ^2 and σ_ϵ^2 are then obtained by simple substitution. The former using (3.10), the later the estimator of $(\sigma_\epsilon^2 + \sigma_\epsilon^{2*}/n)$ from the maximisation of ℓ_1.

Finally, in the basic structural model, the information matrix with respect to the parameter vector $\theta' = (\sigma_\lambda^2,\ \sigma_\epsilon^{2*},\ \sigma_\epsilon^2,\ \sigma_\eta^2,\ \sigma_\delta^2,\ \sigma_\omega^2)$ is asymptotically equivalent to

$$(3.12)\qquad I(\theta) = \tfrac{1}{2}\begin{bmatrix} \frac{T}{\sigma_\lambda^{2*}} & 0 & 0 \\ 0 & \frac{T(n-1)}{\sigma_\epsilon^{2*}} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$+ \tfrac{1}{2}\sum_t \begin{bmatrix} 0 \\ \frac{g_{\epsilon t}}{n} \\ x_t \end{bmatrix} \frac{1}{g_t^2} \begin{bmatrix} 0 \\ \frac{g_{\epsilon t}}{n} \\ x_t \end{bmatrix}',$$

where g_t and x_t were defined in (3.6). The inverse of this matrix can be seen as an approximation of the variance covariance matrix of the maximum likelihood estimators as T goes to infinity.

Finally, if (1.5) does not hold, the transformation leading to (3.1) uses the matrix H^* in (5.2.11) instead of the matrix H. In this case, the likelihood function has essentially the forms above but an extra term, representing the Jacobian of the transformation should be added to (3.1); see Section 5.1.

7.4 Analysis of Labour Cost Time Series : An Application

The techniques presented in the previous sections are illustrated here using three time series of input labour costs in The United States industry. The data, obtained from Jorgenson, Gollop and Fraumeni (1987), is annual, for the period 1948 - 1971, and for the sectors

s1 : Nonmetallic mining and quarrying.

s2 : Food and kindred products.

s3 : Lumber and wood products, except furniture.

Figure 7.4.1 presents a graph of the three time series in logarithms. All the analysis below uses this logarithmic transformation. From Figure 7.4.1, it seems obvious that the three time series present, at least, similar trends. The first exercise in the analysis of this data consisted in some ordinary least square regressions of the input costs on the variable time as well as on one of the other costs. The results are presented in Table 7.4.1. In equations 1, 2 and 3 the independent variables are a constant and the time. The Durbin-Watson statistics show significant serial correlation in the residuals. That suggests the presence of stochastic trends. In equations 4, 5 and 6 the dependent variables s1, s2 and s3 were regressed against a constant and s2, s2 and s3 respectively. Stock (1987) showed that if the time series are co-integrated, the coefficients of these regressions are T-consistent; and so they represent, asymptotically, the relationship between the trends.

Three important facts are observed from these regressions. First,

the Durbin-Watson statistics are much higher and the hypothesis of no serial correlation in equations 4 and 5 is accepted at the 5% significance level. In equation 6, the hypothesis of no serial correlation is rejected at the same significance level, while at the 1% level the test is inconclusive. The second important observation from equations 4, 5 and 6 is that the constants are, at least in equations 4 and 5, clearly different from zero. Finally, it can be observed that the coefficients of the independent variables are close to unity. The above analysis, although preliminary, suggests that the three time series share, apart from a constant, the same stochastic trend.

Table 7.4.1 : Ordinary Least Square Regressions(1)

Equation	Dep. Variable	Constant	Time	s2	s3	D-W
1	s1	-12.205	.045			.442
		(.160)	(.001)			
2	s2	-11.490	.044			.368
		(.128)	(.001)			
3	s3	-12.080	.046			.697
		(.160)	(.001)			
4	s1	-.456		1.023		1.564
		(.067)		(.010)		
5	s1	-.371		.978		1.768
		(.118)		(.017)		
6	s2	.079		.956		1.131
		(.107)		(.015)		

(1) Standard errors in parenthesis.

Figure 7.4.1 : Labour Cost Time Series (in logarithms)

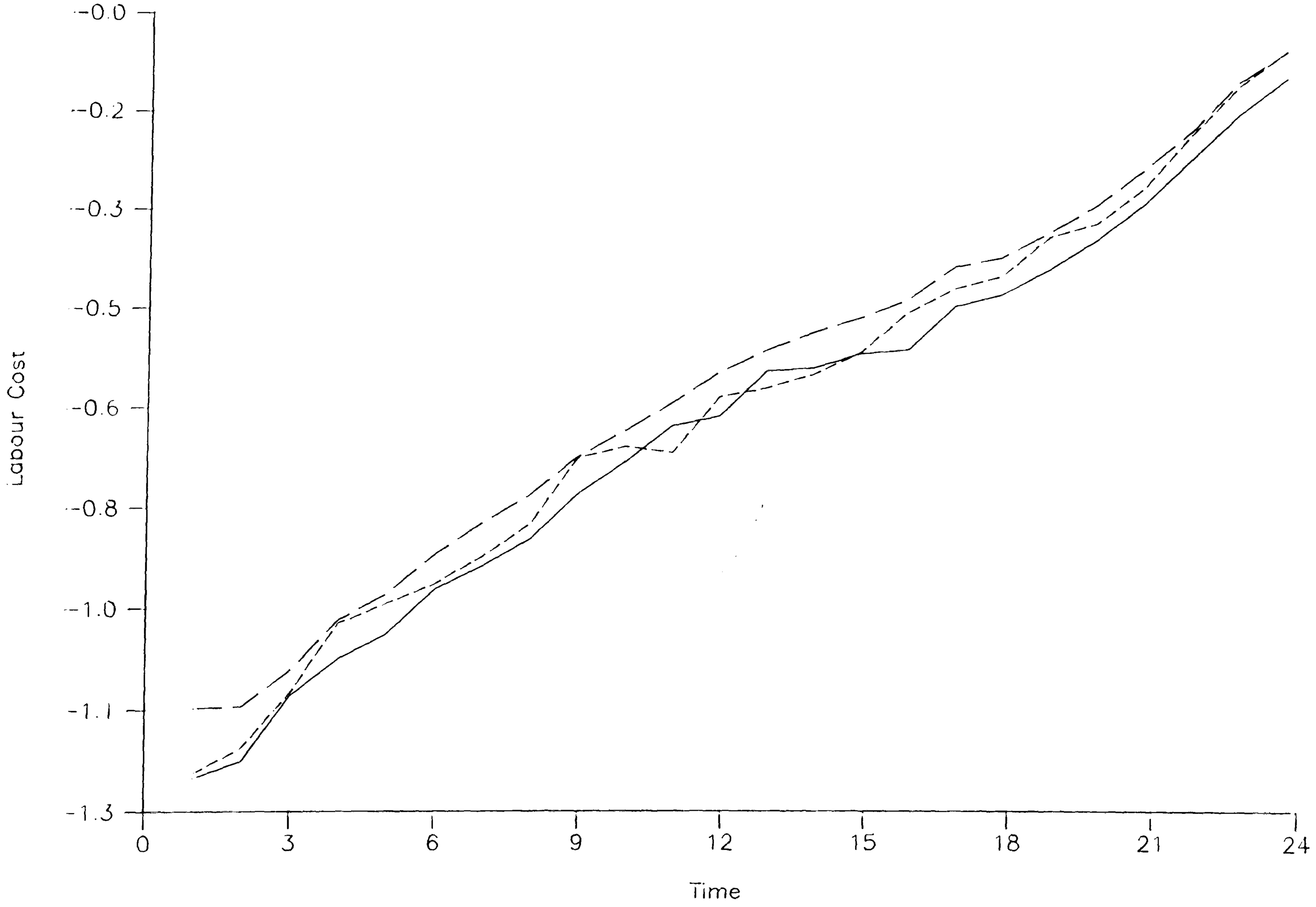


Table 7.4.2 presents the results of the estimation of univariate time series model for the three time series. The model fitted is the local level model plus a fixed slope. In all the cases the model seems to represent the data quite well. The Normality statistic of Bowman and Shenton (1975) has a chi-square distribution under the null hypothesis of normality in the residuals, and the Serial Correlation statistic is the standard Ljung and Box (1978) test for serial correlation. The fixed slopes are very similar between sectors and the variances of the irregular and level random shocks are also similar.

Table 7.4.2 : Univariate Time Series Models(1)

Parameter/Statistic	s1	s2	s3
σ_ϵ^2 (x10^{-3})	.003	.006	.001
	(.129)	(.071)	(.206)
σ_η^2 (x10^{-3})	.560	.328	.960
	(.296)	(.170)	(.495)
β	.049	.046	.050
	(.005)	(.004)	(.006)
Normality	.255	.264	.717
Serial Corr.(8)	4.461	6.883	3.538

(1) Standard errors in parenthesis. Normality is chi-square(2) and Serial Correlation(8) is chi-square(7) if β is known.

The above analysis suggests a multivariate model of the form

$$(4.1a) \qquad y_{it} = \lambda_i + \mu_t + \epsilon_t + \epsilon_{it}, \qquad \begin{matrix} i = 1, 2, 3, \\ t = 1,\ldots,24, \end{matrix}$$

$$(4.1b) \qquad \mu_t = \mu_{t-1} + \beta + \eta_t,$$

where y_{it} is the logarithm of the input labour cots in sector i at time t, μ_t is the common trend with fixed slope β, λ_i is the sector i fixed effect, and ϵ_t, ϵ_{it}, η_t are serially and mutually uncorrelated, normally distributed, with expected values equal to zero and variances σ^2_ϵ, $\sigma^{2*}_{\epsilon i}$, i = 1, 2, 3; and σ^2_η respectively. That is, in principle, the variances of the unit specific irregular random shocks were assumed unit specific.

The maximum likelihood estimates of the parameters in model (4.1) are presented in Table 7.4.3. To solve the identifiability problem with the constants λ_i and the common trend μ_t, λ_1 was set equal to zero. The diagnostics of the residuals of the models present no sign of misspecification. The Normality statistics were .570, 2.392 and .318 for the three sectors, and none of the auto or cross correlations was greater than or smaller than the usual critical values $\pm 2T^{-\frac{1}{2}}$. Two asymptotic test of hypothesis based on the Likelihood Ratio principle were performed. The hypothesis that the slopes were different in the three sectors gave a statistic equal to 4.82. This statistics has a chi-square distribution with 2 degrees of freedom and the common slope hypothesis is accepted. The hypothesis that the irregular unit specific random shocks have the same variance gave a statistic equal to 19.52. Again this has a chi-square distribution with 2 degrees of freedom and so the hypothesis of equal variances is clearly rejected.

Table 7.4.3 : Multivariate Time Series Model

Parameter	Estimate	Standard Error
σ_{ϵ}^2 (x10^{-3})	.000	--
$\sigma_{\epsilon_1}^{2*}$ (x10^{-3})	.157	.113
$\sigma_{\epsilon_2}^{2*}$ (x10^{-3})	.006	.101
$\sigma_{\epsilon_3}^{2*}$ (x10^{-3})	.523	.158
σ_{η}^2 (x10^{-3})	.311	.161
λ_2	.573	.028
λ_3	.240	.058
β	.046	.001

CHAPTER 8 : MULTIVARIATE DYNAMIC ERROR COMPONENTS MODELS

8.1 Introduction

This chapter extends the ideas of error components models to the case of multivariate observations. Although some basic concepts and definitions were presented in Section 5.3, this chapter defines the models more rigorously and studies the estimation procedure. Consider the regression model

$$(1.1) \qquad y_{it} = Z_{it}\,\beta_i + \alpha_{it}, \qquad \begin{matrix} i = 1,\ldots,n, \\ t = 1,\ldots,T, \end{matrix}$$

where y_{it} is a (q x 1) vector of observations for the unit i at time t, Z_{it} is a (q x r_i) matrix of exogenous variables, β_i is an (r_i x 1) vector of fixed coefficients, and α_{it} is a (q x 1) vector which satisfies a structural time series model. In multivariate form, model (1.1) can be written as

$$(1.2) \qquad y_t = X_t\,\beta + \alpha_t, \qquad t = 1,\ldots,T,$$

where y_t and α_t are (nq x 1) vectors with n components of dimension q each, X_t is an (nq x k) matrix which contains the exogenous variable matrices Z_{it}, i = 1,...,n, and β is the (k x 1) vector of coefficients formed with β_i, i = 1,...,n. In general, X_t is a block diagonal matrix with i-th block equal to Z_{it}, and β has as its i-th component the vector β_i. Special cases of (1.2) also include the situations where the matrices of exogenous variables Z_{it} are the same for all units, and the situations where the vector of coefficients are the same for all units. If $Z_{it} = Z_t$ for all i, then $r_i = r$, $X_t = (I_n \otimes Z_t)$, and k = q r. If $\beta_i = \beta$ for all i, then $r_i = r$, $X_t' = [Z_{1t}', \ldots, Z_{nt}']$, and k = r; and if also $Z_{it} = Z_t$ for all i, $X_t = (\iota \otimes Z_t)$, where ι is an (n x 1) vector of

ones.

The structural time series model for the vector of residuals α_t can be written in the state space form

$$(1.3a) \qquad \alpha_t = (Z \otimes I_{nq})\ \theta_t + (\iota \otimes I_q)\ \epsilon_t + \epsilon_t^*, \qquad t = 1,\ldots,T,$$

$$(1.3b) \qquad \theta_t = (T \otimes I_{nq})\ \theta_{t-1} + (R \otimes I_{nq})\ ((I_u \otimes \iota \otimes I_q)\ \kappa_t + \kappa_t^*),$$

where, for any m, I_m is the identity matrix of order m, ι is an (n x 1) vector of ones, and the matrices Z, T and R have dimensions (1 x p), (p x p) and (p x u) respectively, and they are defined, jointly with the values of p and u, as in Section 1.2. The (npq x 1) state vector θ_t contains the trend and seasonal components, while the normal random shocks ϵ_t, ϵ_t^*, κ_t, κ_t^* have dimensions (q x 1), (nq x 1), (uq x 1) and (nuq x 1), expected values equal to zero and variance covariance matrices Σ_ϵ, Σ_ϵ^*, Σ_κ and Σ_κ^* respectively; where Σ_ϵ^* is a block diagonal matrix with i-th block $\Sigma_{\epsilon i}^*$, Σ_κ is a block diagonal matrix as defined in Section 1.1, and Σ_κ^* is also a block diagonal matrix with nu blocks of dimension (q x q) each. That is,

$$(1.4a) \qquad \Sigma_\epsilon^* = \mathrm{diag}(\Sigma_{\epsilon 1}^*,\ldots,\Sigma_{\epsilon n}^*),$$

and for the basic structural model,

$$(1.4b) \qquad \Sigma_\kappa = \mathrm{diag}(\Sigma_\eta,\ \Sigma_\delta,\ \Sigma_\omega),$$

and

$$(1.4c) \qquad \Sigma_\kappa^* = \mathrm{diag}(\Sigma_{\eta 1}^*,\ldots,\Sigma_{\eta n}^*\ ;\ \Sigma_{\delta 1}^*,\ldots,\Sigma_{\delta n}^*\ ;\ \Sigma_{\omega 1}^*,\ldots,\Sigma_{\omega n}^*).$$

The random shock ϵ_t represents the common to all units irregular effects, while κ_t contains the common to all units trend and seasonal random shocks. The elements of the vectors ϵ_t^* and κ_t^* represent the unit specific random shocks.

The variance covariance matrix of the random shocks in the measurement equation (1.3a) is equal to

$$(1.5a) \qquad V[(\iota \otimes I_q)\,\epsilon_t + \epsilon_t^*] = A_\epsilon,$$

while the variance covariance matrices of the random shocks of the trend and seasonal components, in the transition equation (1.3b), have the form

$$(1.5b) \qquad V[(I_u \otimes \iota \otimes I_q)\kappa_t + \kappa_t^*] = \mathrm{diag}(A_1,\ldots,A_u),$$

where the (nq x nq) matrices A_k in (1.5a) and (1.5b) are defined as

$$(1.5c) \qquad A_k = (\iota\iota' \otimes \Sigma_k) + \mathrm{diag}(\Sigma_{k1}^*,\ldots,\Sigma_{kn}^*), \qquad k = \epsilon,\ \eta,\ \delta,\ \omega.$$

Model (1.3) is said to be a multivariate error components model type I if in the local level model $\Sigma_{\eta i}^*$ is positive definite for all i, in the local linear trend model $\Sigma_{\delta i}^*$ is positive definite for all i, in the seasonal local level $\Sigma_{\eta i}^*$ and $\Sigma_{\omega i}^*$ are positive definite for all i, and in the basic structural model $\Sigma_{\delta i}^*$ and $\Sigma_{\omega i}^*$ are positive definite for all i. That is, the stationary form of the error components model type I is strictly invertible; see Section 1.2.

Extending in a natural way the restrictions used in the error components model type I for univariate observations presented in Section 5.1, the error components model type I for multivariate observations is said to be restricted if

$$(1.6) \qquad A_k = (\iota\iota' \otimes \Sigma_k) + (D \otimes \Sigma_k^{**}), \qquad k = \epsilon,\ \eta,\ \delta,\ \omega,$$

where D is an (n x n) diagonal matrix and Σ_k^{**} is a (q x q) unrestricted variance covariance matrix. Thus, (1.6) implies that the variance covariance matrices of the unit specific random shocks differ acroos units only by an scalar. As a particular case, D may be equal to the

identity matrix of order n.

Using Definition 5.3.1, the error components model type I is said to be partially homogeneous if

$$(1.7) \qquad A_k = (\iota\iota' \otimes q_k \Sigma_\epsilon) + \text{diag}(q_{k1}^* \Sigma_{\epsilon 1}^*, \ldots, q_{kn}^* \Sigma_{\epsilon n}^*), \quad k = \eta, \delta, \omega,$$

and if the model is restricted and partially homogeneous,

$$(1.8) \qquad A_k = (\iota\iota' \otimes q_k \Sigma_\epsilon) + (D \otimes q_k^* \Sigma_\epsilon^{**}), \qquad k = \epsilon, \eta, \delta, \omega,$$

with $q_\epsilon = q_\epsilon^* = 1$. If also $q_k = q_k^*$ for all k, model (1.3) is homogeneous. Notice that the restrictions introduced by the idea of homogeneity are, in some way, independent of the restrictions (1.6). While the homogeneity restriction (1.7) defines variance covariance matrices proportional across random shocks, (1.6) defines matrices proportional across units.

Model (1.3) is said to be a multivariate error components model type II if $\kappa_t^* \equiv 0$ for all t. In that case, the variance covariance matrix of the random shocks in the measurement equation (1.3a) is always given by (1.5), while the variance covariance matrix of the random shocks in the transition equation (1.3b) is now given by

$$(1.9a) \qquad V[(I_u \otimes \iota \otimes I_q)\, \kappa_t] = \text{diag}(B_1, \ldots, B_u),$$

where

$$(1.9b) \qquad B_k = (\iota\iota' \otimes \Sigma_k), \qquad k = \eta, \delta, \omega.$$

The multivariate error components model type II assumes that the variance covariance matrix $\Sigma_{\epsilon i}^*$ is positive definite for all i, and also that Σ_η is positive definite in the local level model, Σ_δ is positive definite in the local linear trend, Σ_η and Σ_ω are positive definite in the seasonal local level, and Σ_δ and Σ_ω are positive definite in the

basic structural model. Finally, extending the restrictions used in Section 5.1 in the model for univariate observations, the multivariate error components model type II is said to be restricted if

$$(1.10) \qquad V\left[(\iota \otimes I_q)\ \epsilon_t + \epsilon_t^*\right] = (\iota\iota' \otimes \Sigma_\epsilon) + (D \otimes \Sigma_\epsilon^{**}),$$

where D is an (n x n) diagonal matrix, and Σ_ϵ^{**} is a (q x q) positive definite matrix. As a special case, D may be equal to the identity matrix of order n. Notice that, although the model type II is never homogeneous, it can be partially homogeneous.

The chapter, apart from this introduction, is organised as follows. Section 8.2 considers the estimation of the unobserved trend and seasonal components, while Section 8.3 studies the maximum likelihood estimation of the parameters in the model. Finally, Section 8.4 presents an illustration of the techniques presented, using the data and the econometric model in Chapter 4.

8.2 Estimation of the Unobserved Components

This section considers the estimation of the unobserved trend and seasonal components in the state space model (1.3) assuming α_t is observed for all t and the variance covariance matrices of the random shocks are known. As in all the models studied in previous chapters, the estimation of these unobserved components is carried out by means of the Kalman filter defined in sections 1.3 and 1.4.

Consider first the error components model type I. In the unrestricted model, the estimates of the state vectors θ_t, and their mean square errors are obtained by applying the recursions (1.3.2) from t = p+1, with the variance covariance matrices for the random shocks as defined in (1.5), and with initial estimate of the state vector, m_p, and mean square error, P_p, obtained from (1.3.5) and (1.3.6) respectively. As in the unrestricted model for univariate observations, it seems that the recursions needed to obtain the mean square error of the estimator of θ_t with information up to time t, P_t, and the prediction error variances, F_t, cannot be simplified. Although the variance covariance matrices of the random shocks in the model have simple forms, F_t and P_t do not.

The same seems to be true when the model is partially homogeneous. No simplifications in the Kalman filter recursions are, apparently, possible.

Consider now the restricted model where (1.6) holds with $D = I_n$. Premultiplying (1.3) by the matrix $(H \otimes I_q)$, where H is the (n x n) matrix defined in (5.2.4), yields a new state space model with observations $[(H \otimes I_q)\ \alpha_t]$, and state vector $[(I_p \otimes H \otimes I_q)\ \theta_t]$. It is not difficult to verify that the variance covariance matrix of the random shocks in the measurement equation is now equal to $\bar{A}_\epsilon$, and the

variance covariance matrix in the transition equation is $\mathrm{diag}(\bar{A}_1, \ldots, \bar{A}_u)$, where

$$(2.1) \qquad \bar{A}_k = \begin{bmatrix} \Sigma_k + (1/n)\,\Sigma_k^{**} & 0 \\ 0 & (I_{n-1} - \iota\iota'/n) \otimes \Sigma_k^{**} \end{bmatrix},$$

$k = \epsilon, \eta, \delta, \omega$. It follows that the variance covariance matrices of the random shocks in the transformed model are block diagonal matrices. The first blocks have dimensions (q x q) and correspond to a model for the average of the observations. The second blocks have dimensions ((n-1)q x (n-1)q) and correspond to a model for the first (n-1) deviations of the original observations with respect to the average. In other words, the unobserved trend and seasonal components in these two sub-models can be estimated separately. The first sub-model has only dimension q and the Kalman filter equations present no problems unless q is very large. The second sub-model has dimension (n-1)q and, although it is not in general homogeneous, it can be shown that the Kalman filter recursions have only dimension q. The idea is the following. Suppose the observations in the second sub-model are multiplied by the matrix $[(I_{n-1} - \iota\iota'/n)^{-\frac{1}{2}} \otimes I_q]$; hence the transformed observations follow a model where the variance covariance matrices of the random shocks have the form $(I_{n-1} \otimes \Sigma_k^{**})$, $k = \epsilon, \eta, \delta, \omega$. From there follows that the model for the transformed observations can be run for each unit in turn; that is, with a q-dimensional Kalman filter. The estimates of the original state vectors, and their mean square errors are then recovered using the transformation matrix above. If apart from the assumptions above the model is partially homogeneous, then the second sub-model becomes a homogeneous model.

If (1.6) holds but D is not equal to the identity matrix, exactly

the same idea can be used if the matrix H is replaced by H^* defined in (5.2.11). In that case, $\bar{A}_k$, k = ϵ, η, δ, ω, have the form

$$(2.2) \qquad \bar{A}_k = \begin{bmatrix} \Sigma_k + (1/n)\ \Sigma_k^{**} & 0 \\ 0 & (D^* - \iota\iota'/n) \otimes \Sigma_k^{**} \end{bmatrix},$$

where D^* is an (n-1 x n-1) diagonal matrix with the first (n-1) elements of D. As before, under partial homogeneity, the second sub-model is homogeneous.

Consider now the error components model type II under (1.7) and D = I_n. Premultiplying (1.3) by $(H \otimes I_q)$ yields,

$$(2.3a) \qquad \bar{\alpha}_t = (Z \otimes I_q)\ \bar{\theta}_t + (\epsilon_t + \epsilon_t^*/n), \qquad t = 1,\ldots,T,$$

$$(2.3b) \qquad \bar{\theta}_t = (T \otimes I_q)\ \bar{\theta}_{t-1} + (R \otimes I_q)\ \kappa_t,$$

and

$$(2.4a) \qquad (\alpha_{it} - \bar{\alpha}_t) = (Z \otimes I_{q(n-1)})\ (\theta_{it} - \bar{\theta}_t) + (\epsilon_{it}^* - \bar{\epsilon}_t^*),$$

$$(2.4b) \qquad (\theta_{it} - \bar{\theta}_t) = (T \otimes I_{q(n-1)})\ (\theta_{i,t-1} - \bar{\theta}_{t-1}),$$

for i = 1,...,(n-1) and t = 1,...,T. In (2.3) and (2.4), θ_{it} is the state vector which contains the trend and seasonal components for the unit i, and $\bar{\alpha}_t$ and $\bar{\theta}_t$ are the average of α_{it} and θ_{it} across units. In the fixed effects model, $(\theta_{i0} - \bar{\theta}_0)$ is a fixed vector of parameters which can be included in the vector of coefficients β. This vector of coeffcients is assumed known at this stage. Hence, $\bar{\theta}_t$ represents the common to all units trend and seasonal components at time t, and as the random shocks in (2.3 and (2.4) are uncorrelated, the estimation of $\bar{\theta}_t$ require only the q-dimensional model (2.3).

8.3 Maximum Likelihood Estimation of the Parameters

This section considers the maximum likelihood estimation of the vector of coefficients β in (1.2) and of the parameters in the variance covariance matrices of the random shocks. As the number of parameters in the model is large, even under the restrictions considered in the introduction, the log-likelihood functions obtained from the stationary form of the model are much more attractive than the one formed by means of the Kalman filter. Hence, this section assumes that the likelihood function is formed using the frequency domain approach or the alternative transformation developed in Section 2.3.

Consider first the error components model type I, and assume that a diffuse prior is defined over the state vector θ_t. From Section 3.1, the log-likelihood of the basic structural model has, apart from a constant, the form

$$(3.1) \qquad \ell = -\tfrac{1}{2} \sum_{t=p+1}^{T} [\log|G_t| + \text{trace}(G_t^{-1} P_t)],$$

where P_t is an (nq x nq) matrix given by

$$(3.2) \qquad P_t = (w_{yt} - W_{xt}\beta)(w_{yt} - W_{xt}\beta)', \qquad t = p+1,\ldots,T,$$

and G_t is an (nq x nq) matrix of the form

$$(3.3) \qquad G_t = g_{et} A_\epsilon + g_{\eta t} A_\eta + g_{\delta t} A_\delta + g_{\omega t} A_\omega, \qquad t = p+1,\ldots,T,$$

where A_k, $k = \epsilon, \eta, \delta$ and ω were defined in Section 8.1, and the (nq x 1) vector w_{yt}, the (nq x k) matrix W_{xt}, and the scalars $g_{\epsilon t}$, $g_{\eta t}$, $g_{\delta t}$ and $g_{\omega t}$, are defined as in Section 3.1. The matrix G_t in (3.3) can also be written as

$$(3.4) \qquad G_t = (\iota\iota' \otimes G_{1t}) + G_{2t}$$

$$= (\iota \otimes L_t)\ (\iota' \otimes L_t') + G_{2t}, \qquad t = p+1,\ldots,T,$$

where ι is an (n x 1) vector of ones, G_{1t} is a (q x q) matrix equal to $(L_t\ L_t')$, with L_t a lower triangular matrix; and G_{2t} is an (nq x nq) block diagonal matrix with n blocks of dimension (q x q) each.

The maximum likelihood estimator of β is obtained as in (3.2.2), with the information matrix of the form (3.2.3). A concentrated log-likelihood can be formed by replacing the estimator of β into (3.1). That concentrated function has to be maximised numerically with respect to the parameters in the variance covariance matrices of the random shocks. Alternatively, a stepwise procedure can be implemented as mentioned in Section 3.4, with the advantage that first and second derivatives of the log-likelihood with respect to all the parameters in the model can be computed analytically. The evaluation of the maximum likelihood estimator of β requires the inverse of the (nq x nq) matrix G_t, and the evaluation of (3.1) requires also the determinant of G_t. The following analytic expressions, which uses the inverse and the determinant of matrices of order (q x q) only, give those functions of the matrix G_t.

(3.5a) $$|G_t| = |G_{2t}|\ |I_q + (\iota' \otimes L_t')\ G_{2t}^{-1}\ (\iota \otimes L_t)|, \quad t = p+1,\ldots,T,$$

and

(3.5b) $$G_t^{-1} = G_{2t}^{-1} - G_{2t}^{-1}\ (\iota \otimes L_t)\ [I_q + (\iota' \otimes L_t')\ G_{2t}^{-1}\ (\iota \otimes L_t)]^{-1}$$

$$(\iota' \otimes L_t')\ G_{2t}^{-1}, \qquad t = p+1,\ldots,T.$$

Although formulas (3.5) imply a significant reduction in the calculations, the number of parameters in the model may still be too large to make the procedure feasible. In that sense, the partial homogeneity hypothesis is very attractive because although it does not

simplifies substantially the above formulas, the reduction in the number of parameters is very important; see Table 8.3.1. For example, if q = 3 and n = 3 as in the example in Section 8.4, the number of parameters in the variance covariance matrices of the random shocks in the unrestricted model are 48 in the local level model and 96 in the basic structural model. Under partial homogeneity, the number of parameters reduces to 28 for the local level model and to 36 for the basic structural model.

Consider now the model type I under the restriction (1.6) and with $D = I_n$. Premultiplying the model by $(H \otimes I_q)$, with H as in (5.2.4), the log-likelihood can be written as

$$\ell = c + \ell_1 + \ell_2, \tag{3.6}$$

where c is a constant, ℓ_1 is the log-likelihood of a q-dimensional structural time series model for the average of the observations with the variance covariance matrices of the random shocks of the form $(\Sigma_k + (1/n)\ \Sigma_k^{**})$, $k = \epsilon,\ \eta,\ \delta$ and ω; which can be defined as new variance covariance matrices. Thus ℓ_1 has the form (3.1) and the results in Chapter 3 can be applied here.

The expression ℓ_2 represents the log-likelihood of the first (n-1) deviations of the original observations with respect to the average. That is, a (n-1)q-dimensional structural time series model. The variance covariance matrices of the random shocks have the form $[(I_{n-1} - \iota\iota'/n) \otimes \Sigma_k^*]$, $k = \epsilon,\ \eta,\ \delta$ and ω; and then, the matrix G_t in (3.1) has the form

$$G_t = (I_{n-1} - \iota\iota'/n) \otimes G_t^*, \qquad t = p+1,\ldots,T, \tag{3.7}$$

where G_t^* is a (q x q) matrix. Using (3.1) and (3.7), ℓ_2 can be written as

$$(3.8) \qquad \ell_2 = -\tfrac{1}{2} \sum_{t=p+1}^{T} \left[(n-1) \log|G_t^*| + \operatorname{trace}(G_t^{*-1} P_t^*) \right],$$

where P_t^* is a (q x q) matrix defined by

$$(3.9) \qquad P_t^* = \sum_{i=1}^{n} P_{iit}, \qquad t = p+1, \ldots, T,$$

where P_{iit} is the i-th (q x q) diagonal block in P_t. That is, in the evaluation of ℓ_2, the determinant and the inverse of only (q x q) matrices is required. If the model is partially homogeneous, Σ_ϵ^{**} can be concentrated out of the log-likelihood ℓ_2.

With respect to the coefficients of the exogenous variables, essentially the same results developed in Section 6.3 apply here : if the exogenous variables are unit specific but the vector of coefficients is the same for all units, this vector of coefficients can be expressed as an average of the estimator obtained from the maximisation of ℓ_1 and the one which maximises ℓ_2. On the other hand, if the coefficients are unit specific and the exogenous variables are common for all units, the estimator of the average of the coefficients is given by ℓ_1 while ℓ_2 gives the estimators of the deviations of the coefficients with respect to the average. Finally, if both the vector of coefficients and the exogenous variables are the same for all units, this vector of coefficients is obtained only from the log-likelihood ℓ_1.

In the above cases, the variance covariance matrices in the model can be redefined for the estimation procedure and that implies that the two log-likelihoods in (3.6) can be maximised in turn. This procedure has the problem that the original variance covariance matrices are not necessarily positive (semi) definite; however, expression ℓ in (3.6) still has a simple form which can be maximised with respect to all the

parameters.

The number of parameters in the variance covariance matrices of the random shocks in the multivariate error components model type I under (1.6) and $D = I_n$ is presented in Table 8.3.1. The partial homogeneity is more attractive in models where q is large and n is small; while if the number of units is very large, restriction (1.6), with $D = I_n$, becomes very important because under this restriction the number of parameters does not depend on n.

For the multivariate error components model type II, under (1.10) and $D = I_n$, (3.6) also applies. The log-likelihood ℓ_1 is formed with the average of the observations acroos units, and the variance covariance matrices of the random shocks in the basic structural model are equal to $[\Sigma_\epsilon + (1/n)\ \Sigma_\epsilon^*]$, Σ_η, Σ_δ, and Σ_ω. Thus, the results in Chapter 3 can be applied. The log-likelihood ℓ_2 corresponds to the log-likelihood of a stationary model of dimension (n-1)q. Extending the results in Section 7.3, and for the fixed effects model, the maximum likelihood estimator of Σ_ϵ^* is

$$(3.10) \qquad \tilde{\Sigma}_\epsilon^* = 1/T(n-1) \sum_{i=1}^{n} \sum_{t=1}^{T} (\alpha_{it} - \bar{\alpha}_t)\ (\alpha_{it} - \bar{\alpha}_t)',$$

where $\bar{\alpha}_t$ is the average of α_{it} across units. All the comments below (3.9) with respect to the maximisation of $(\ell_1 + \ell_2)$, and the maximum likelihood estimators of the vector of coefficients β also apply to the multivariate error components model type II.

Table 8.3.2 presents the number of parameters in the variance covariance matrices of the random shocks for the model type II under the two restrictions considered: partial homogeneity and the one in (1.7). Restriction (1.7) is attractive when the number of units, n, is large because the number of parameters is independent of n. On the

other hand, when q is large but n is small, the homogeneity restriction may produce a more important reduction in the number of parameters.

Table 8.3.1 : Number of Parameters in Variance Covariance Matrices in Model Type I(1)

Model	Unrestricted	P.H.	R.
Local Level	$2q^*(n+1)$	$q^*(n+1)+(n+1)$	$4q^*$
Local Linear Trend	$3q^*(n+1)$	$q^*(n+1)+2(n+1)$	$6q^*$
Seasonal Local Level	$3q^*(n+1)$	$q^*(n+1)+2(n+1)$	$6q^*$
Basic Structural	$4q^*(n+1)$	$q^*(n+1)+3(n+1)$	$8q^*$

(1) $q^* = \frac{1}{2}q(q+1)$. P.H. is partial homegeneity and R. is restriction (1.6) with $D = I_n$.

Table 8.3.2 : Number of Parameters in Variance Covariance Matrices in Model Type II(1)

Model	Unrestricted	P.H.	R.
Local Level	$q^*(n+2)$	$q^*(n+1)+1$	$3q^*$
Local Linear Trend	$q^*(n+3)$	$q^*(n+1)+2$	$4q^*$
Seasonal Local Level	$q^*(n+3)$	$q^*(n+1)+2$	$4q^*$
Basic Structural	$q^*(n+4)$	$q^*(n+1)+3$	$5q^*$

(1) $q^* = \frac{1}{2}q(q+1)$. P.H. is partial homegeneity and R. is restriction (1.10) with $D = I_n$.

8.4 The Demand for Energy in the U. K.: An Application

This section illustrates the techniques presented in previous sections of this chapter using the data and the econometric model for the demand for energy developed in Chapter 4.

The econometric model presented in (4.3.1) was estimated in Chapter 4 separately for each economic sector: other industry, domestic, other final users and transport; for the four fuels: gas, electricity, oil and coal; and for the sample period 1971 Q1 to 1986 Q4. This section attempts the joint estimation of the model for the three most important economic sectors: other industry, domestic and other final users. The transport sector is not only smaller but uses basically two fuels while the other sectors use four, and that makes the joint formulation and estimation of the four sectors more complicated.

Model (4.3.1) for the economic sector i can be written as

$$(4.1a) \qquad S_{it} = \mu_{it} + \gamma_{it} + A_i\, p_{it} + \epsilon_{it}, \qquad \begin{matrix} i = 1, 2, 3, \\ t = 1, \ldots, T. \end{matrix}$$

$$(4.1b) \qquad \mu_{it} = \mu_{i,t-1} + \beta_i + \eta_{it},$$

$$(4.1c) \qquad (1 + L + L^2 + L^3)\, \gamma_{it} = \omega_{it},$$

where S_{it} is a (3 x 1) vector of shares, μ_{it} and γ_{it} are (3 x 1) vectors representing the trend and seasonal components, A_i is a (3 x 3) matrix of parameters satisfying the restrictions (4.2.7), and p_{it} is a (3 x 1) vector of exogenous prices. Finally, for each i, ϵ_{it}, η_{it} and ω_{it} are assumed to be serially and mutually uncorrelated random shocks, with expected values equal to zero.

Model (4.1) can be written as a 9-dimensional structural time series model of the form (1.1.3). That model contains 135 parameters in

the variance covariance matrices of these random shocks, and its estimation is extremely difficult. The idea of error components models presented in previous sections of this chapter provide a feasible alternative. Assume that the (3 x 1) random shocks ϵ_{it}, η_{it} and ω_{it} can be written as

$$\epsilon_{it} = \epsilon_t + \epsilon^*_{it}, \tag{4.2a}$$

$$\eta_{it} = \eta_t + \eta^*_{it}, \tag{4.2b}$$

and

$$\omega_{it} = \omega_t + \omega^*_{it}, \tag{4.2c}$$

for i = 1, 2, 3; and t = 1,...,T; where the random shocks ϵ_t, ϵ^*_{it}, η_t, η^*_{it}, ω_t, and ω^*_{it} have dimensions (3 x 1) and they are assumed to be serially and mutually uncorrelated, with expected values equal to zero and variances Σ_ϵ, $\Sigma^*_{\epsilon i}$, Σ_η, $\Sigma^*_{\eta i}$, Σ_ω and $\Sigma^*_{\omega i}$ respectively. Thus, ϵ_t, η_t and ω_t are common to all sectors random shocks, while ϵ^*_{it}, η^*_{it} and ω^*_{it}, i = 1, 2, 3, are uncorrelated sector specific random shocks. In terms of the augmentation factors defined in Section 4.2, (4.2b) says that there are random shocks which affect the augmentation factors in all the economic sectors, while there are also sector specific random shocks affecting these augmentation factors.

Model (4.1) under (4.2) has the form (1.2)-(1.3). Although the number of parameters in the variance covariance matrices has been reduced to 72 with the specification (4.2), it is still difficult to estimate. From Table 4.3.3, the partial homogeneity restriction (1.7) seems to be appropiate in this case because the estimation of the model for each sector, in Section 4.3, accepted the hypothesis of homogeneity. Although a partially homogeneous model for n units does not imply the model for each unit is homogeneous, it can be seen as an

approximation. Under (1.7) the number of parameters in the variance covariance matrices is reduced to 32, and from Section 8.3, the evaluation of the log-likelihood is more or less equivalent to the evaluation of a 3-dimensional structural time series model.

Assuming, as an approximation, normality for the random shocks in (4.2), the model was estimated using the frequency domain likelihood. The results are presented in tables 8.4.1, 8.4.2 and 8.4.3. Table 8.4.1 presents the coefficients of the price variables. Most of the estimates are significant at the 5% significance level. The estimates of the slopes of the trends, which represent the slopes in the biases of technical progress, are shown in Table 8.4.2. The slopes of the gas trends are greater than zero and clearly significant in all three economic sectors. The slopes of electricity are positive in the other industry and other final users sectors, and negative in the domestic sector. The slopes of the oil trends are negative in all sectors although in the domestic sector the coefficient is not very significant. Finally, the slopes of the coal trends are negative in all sectors. Table 8.4.3 presents the parameters in the variance covariance matrices of the random shocks.

Using (4.2.8), substitution and demand elasticities can be computed from estimates of the price coefficients and given values of shares. Table 8.4.4 presents the demand elasticities for the average value of the shares in the sample period. The results show that gas, electricity and oil are substitutes in all the economic sectors.

The estimated model may not represent complety the covariance structure of the random shocks across sectors; although the specification used might capture most of this covariance structure in a way that the estimation of the model is feasible and the interpretation easy.

Table 8.4.1: Estimates of Price Effects(1)

Parameter	Other Industry	Domestic	Other Final Users
α_{gg}	.038	.178	.061
	(.017)	(.033)	(.014)
α_{ge}	-.015	-.110	-.041
	(.014)	(.018)	(.016)
α_{go}	-.037	-.027	-.030
	(.009)	(.012)	(.007)
α_{gc}	.014	-.041	.010
	(.016)	(.036)	(.007)
α_{ee}	.225	.249	.244
	(.023)	(.020)	(.025)
α_{eo}	-.133	-.030	-.162
	(.010)	(.009)	(.013)
α_{ec}	-.077	-.109	-.041
	(.021)	(.026)	(.012)
α_{oo}	.205	.062	.196
	(.012)	(.008)	(.011)
α_{oc}	-.035	-.005	-.004
	(.010)	(.015)	(.004)
α_{cc}	.098	.155	.035
	(.028)	(.052)	(.012)

(1) g : gas, e : electricity, o : oil, c : coal. Standard errors in parenthesis.

Table 8.4.2 : Slopes of the Biases in Technical Progress(1)

Parameter	Other Industry	Domestic	Other Final Users
β_g (x10^{-2})	.289	.373	.207
	(.070)	(.068)	(.061)
β_e (x10^{-2})	.165	-.061	.187
	(.078)	(.049)	(.104)
β_o (x10^{-2})	-.346	-.049	-.299
	(.093)	(.036)	(.086)
β_c (x10^{-2})	-.108	-.263	-.095
	(.067)	(.074)	(.085)

(1) g : gas, e : electricity, o : oil, c : coal. Standard errors in parenthesis.

Table 8.4.3 : Error Structure of the Model(1)

	Variance Seasonal Random Shock ($x10^{-5}$)				q_ϵ	q_η
Other Industry	.119				13.786	16.953
	-.094	-.115	.264		(.875)	(.667)
	-.039	-.042	-.055	.136		
Domestic	1.967				3.576	.846
	-.447	.503			(.565)	(1.063)
	-.281	.315	.197			
	-1.239	-.371	-.231	1.841		
Other Final Users	.597				.066	2.143
	-.687	2.423			(.352)	(.386)
	-.035	-1.666	1.787			
	.125	-.070	-.086	.031		
Common to all Sectors	.657				.661	1.199
	-.250	.765			(1.370)	(1.516)
	-.230	-.338	.486			
	-.177	-.177	.082	.272		

(1) Standard errors in parenthesis. The order of the fuels is gas, electricity, oil and coal.

Table 8.4.4 : Demand Elasticities(1)

Elasticity	Other Industry	Domestic	Other Final Users
e_{gg}	-.59	-.15	-.40
e_{ge}	.36	.08	.25
e_{go}	.06	.02	.04
e_{gc}	.18	.05	.11
e_{eg}	.11	.08	.06
e_{ee}	-.05	.03	-.00
e_{eo}	.03	.01	-.01
e_{ec}	-.09	-.13	-.04
e_{og}	.02	.07	.02
e_{oe}	.04	.05	-.03
e_{oo}	-.04	-.22	-.00
e_{oc}	-.03	.10	.02
e_{cg}	.32	.11	.46
e_{ce}	-.50	-.30	-.80
e_{co}	-.12	.06	.14
e_{cc}	.31	.13	.20

(1) g : gas, e : electricity, o : oil, c : coal.

REFERENCES

Anderson, B.D.O. and J.B. Moore (1979), *Optimal Filtering*, New York: Prentice Hall.

Anderson, T.W. (1963), The use of factor analysis in the statistical analysis of multiple time series, *Psychometrika*, 28, 1-25.

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, New York : John Wiley.

Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis*, second edition, New York : John Wiley.

Anderson, T.W. and C. Hsiao (1981), Estimation of dynamic models with error components, *Journal of the American Statistical Association*, 76, 598-606.

Anderson, T.W. and C. Hsiao (1982), Formulation and estimation of dynamic models using panel data, *Journal of Econometrics*, 18, 47-82.

Balestra, P. (1980), A note on the exact transformation associated with the first order moving average process, *Journal of Econometrics*, 14, 381-394.

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

Berndt, E.R. and D.O. Wood (1975), Technology, prices and the derived demand for energy, *Review of Economics and Statistics*, 57, 259-268.

Binswanger, H. (1974), The measurement of technical change biases with many factors of production, *American Economic Review*, 64, 964-976.

Bowman, K.O. and B.R. Shenton (1975), Omnibus test contours for departures from normality based on b_1 and b_2, *Biometrika*, 62, 243-250.

Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, revised edition, San Francisco: Holden Day.

Box, G.E.P. and G.C. Tiao (1977), A canonical analysis of multiple time series, *Biometrika*, 64, 355-365.

Brillinger, D.R. (1981), *Time Series, Data Analysis and Theory*, San Francisco : Holden Day.

Chamberlain, G. and Z. Griliches (1975), Unobservables with variance-components structure: ability, schooling and the economic success of brothers, *International Economic Review*, 16, 422-450.

Cooley, T.F. and E.C. Prescott (1973), An adaptive regression model, *International Economic Review*, 14, 364-371.

Cooley, T.F. and E.C. Prescott (1976), Estimation in the presence of stochastic parameter variation, *Econometrica*, 44, 167-184.

Cramer, J.S. (1986), *Econometric Applications of Maximum Likelihood Methods*, Cambridge University Press.

De Jong, P. (1988), The likelihood for a state space model, *Biometrika*, 75, 165-169.

De Jong, P. (1989), *The Diffuse Kalman Filter*, Mimeo.

Department of Trade and Industry (1988), *Monthly Review of External Trade Statistics, Annual Supplement No. 9*, London: Department of Trade and Industry.

Diewert, W.E. (1976), Exact and superlative index numbers, *Journal of Econometrics*, 4, 115-145.

Duncan, D.B. and S.D. Horn (1972), Linear dynamic regression from the viewpoint of regression analysis, *Journal of the American Statistical Association*, 67, 815-821.

Dunsmuir, W. (1979), A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise, *The Annals of Statistics*, 7, 490-506.

Dunsmuir, W. and E.J. Hannan (1976), Vector linear time series models, *Adv. Appl. Probability*, 8, 339-364.

Engle, R.F. and M.W. Watson (1981), A one-factor multivariate time series model of metropolitan wage rates, *Journal of the American Statistical Association*, 76, 774-781.

Engle, R.F. and C.W.J. Granger (1987), Co-integration and error correction: representation, estimation and testing, *Econometrica*, 55, 251-276.

Enns, P.G., J.A. Machak, W.A. Spivey and W.J. Wrobleski (1982), Forecasting applications of an adaptive multiple exponential smoothing, *Management Science*, 28, 1035-1044.

Fernandez-Macho, F.J. (1986), *Estimation and Testing of Multivariate Structural Time Series Models*, Thesis, London School of Economics.

Fernandez-Macho, F.J. (1989), Estimation and testing of multivariate exponential smoothing models in the frequency domain. (forthcoming in *Journal of Time Series Analysis*).

Fernandez-Macho, F.J. and A.C. Harvey (1989), Seemingly unrelated time series equations and a test for homogeneity. (submitted for publication).

Fernandez-Macho, F.J., A.C. Harvey, and J. Stock (1987), Forecasting and interpolation using vector autoregressions with common trends, *Annales D'Economie et de Statistique*, N^0 6/7, 279-287.

Fuss, M.A. (1977), The demand for energy in canadian manufacturing, *Journal of Econometrics*, 5, 89-116.

Geweke, J. (1977), The dynamic factor analysis of economic time series models, in *Latent Variables in Socio-Economic Models*, eds. D.J. Aigner and A.S. Goldberger, Amsterdam: North Holland, 365-383.

Geweke, J. and K. Singleton (1981), Maximum likelihood confirmatory factor analysis of economic time series, *International Economic Review*, 22, 37-54.

Hannan, E.J. (1970), *Multiple Time Series*, New York: John Wiley.

Harvey, A.C. (1985a), Trends and cycles in macroeconomic time series, *Journal of Business, Economics and Statistics*, 3, 216-227.

Harvey, A.C. (1985b), Analysis and generalisation of a multivariate exponential smoothing model, *Management Sciences*, 32, 374-380.

Harvey, A.C. (1990), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, A.C. and J. Durbin (1986), The effects of seat belt legislation on british road casualties: a case study in structural time series modelling, *Journal of the Royal Statistical Society*, B-149, 187-227.

Harvey, A.C., S.G.B. Henry, S. Peters and S. Wren-Lewis (1986), Stochastic trends in dynamic regression models: an application to the employment-output equation, *The Economic Journal*, 96, 975-985.

Harvey, A.C. and P. Todd (1983), Forecasting economic time series with structural and Box-Jenkins models: a case study (with discussion), *Journal of Business, Economics and Statistics*, 1, 299-315.

Hause, J.C. (1980), The finite structure of earnings and the on-the-job training hypothesis, *Econometrica*, 48, 1013-1029.

Hsiao, C. (1986), *Analysis of Panel Data*, Econometric Society Monographs 11, Cambridge University Press.

Jones, R.H. (1966), Exponential smoothing for multivariate time series, *Journal of the Royal Statistical Society*, B-28, 241-251.

Jorgenson, D.W. (1986), Econometric methods for modelling producer behavior, in *Handbook of Econometrics III*, Edited by Z. Griliches and M.D. Intriligator. Amsterdam: North Holland, 1841-1915.

Jorgenson, D.W., F. Gollop and B. Fraumeni (1987), *Productivity and U. S. Economic Growth*, Cambridge: Harvard University Press.

Kitagawa, G. and W. Gersh (1984), A smoothness priors-state space modeling of time series with trend and seasonality, *Journal of the American Statistical Society*, 79, 378-389.

Kohn, R. (1979), Asymptotic estimation and hypothesis testing results for vector linear time series models, *Econometrica*, 47, 1005-1030.

Kohn, R. and C. Ansley (1985), Efficient estimation and prediction in time series regression models, *Biometrika*, 72, 694-697.

Lillard, L.A. and Y. Weiss (1979), Components of variation in panel data earnings: american scientists 1960-70, *Econometrica*, 47, 437-454.

Lillard, L.A. and R. Willis (1978), Dynamic aspects of earnings mobility, *Econometrica*, 46, 985-1012.

Ljung, G.M. and G.E.P. Box (1978), On a measure of lack of fit in time series models, *Biometrika*, 66, 297-303.

Ma Curdy, T.E. (1982), The use of time series processes to model the error structure of earnings in a longitudinal data analysis, *Journal of*

Econometrics, 18, 83-114.

Mc Cullagh, P. (1987), *Tensor Methods in Statistics*, Monographs on Statistics and Applied Probability ed. by D.R. Cox, D.V. Hinkley, D. Rubin and B.W. Silverman, Chapman and Hall.

Maddala, G.S. (1971), The use of variance components models in pooling cross-section and time series data, *Econometrica*, 39, 341-358.

Magnus, J.R. (1978), Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix, *Journal of Econometrics*, 7, 281-312.

Magnus, J.R. (1982), Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood, *Journal of Econometrics*, 19, 239-285.

Magnus, J.R. (1988), *Linear Structures*, London: Charles Griffin.

Magnus, J.R. and H. Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: John Wiley.

Magnus, J.R. and A.D. Woodland (1987), Inter-fuel substitution in dutch manufacturing, *Applied Economics*, 19, 1639-1664.

Nerlove, M. (1971), A note on error components models, *Econometrica*, 39, 383-396.

Oberhofer, W. and J. Kmenta (1974), A general procedure for obtaining maximum likelihood estimates in generalised regression models, *Econometrica*, 42, 579-590.

Peña, D. and G.E.P. Box (1987), Identifying a simplifying structure in time series, *Journal of the American Statistical Association*, 82, 836-843.

Pepper, M.P.G. (1985), Multivariate Box-Jenkins analysis: a case study in U.K. energy demand forecasting, *Energy Economics*, July, 168-178.

Pagan, A. (1980), Some identification and estimation results for regression models with stochastically varying coefficients, *Jounal of Econometrics*, 13, 341-363.

Pesaran, N.H. and L.J. Slater (1980), *Dynamic Regression : Theory and Algorithms*, Ellis Horwood.

Pindyck, R.S. (1979), Interfuel substitution and the industrial demand for energy: an International comparison, *Review of Economics and Statistics*, 61, 169-179.

Quenouille, M.H. (1957), *The Analysis of Multiple Time Series*, London: Charles W. Griffin.

Revankar, N.S. (1979), Error components models with serially correlated time effects, *Journal of the Indian Statistical Association*, 17, 137-160.

Rosenberg, B. (1973), The analysis of a cross-section of time series by stochastically convergent parameter regression, *Annals of Economic and Social Measurement*, 2/4, 399-428.

Sargan, J.D. (1964), Wages and prices in the United Kingdom: a study in econometric methodology, in *Econometric Analysis for National Economic Planning*, P.E. Hart, G. Mills and J. K. Whitaker eds., 25-54.

Shephard, R.W. (1953), *Cost and Production Functions*, Princeton University Press.

Slade, M.E. (1989), Modelling stochastic and cyclical components of technical change: an application of the Kalman filter, *Journal of Econometrics*, 41, 363-383.

Solow, R.M. (1967), Some recent developments in the theory of production, In M. Brown, ed. *The Theory and Empirical Analysis of Production*, Mat. Bur. Econ. Res. Stud. Income and Wealth, 31, 25-53.

Stock, J.H. (1987), Asymptotic properties of least squares estimators of cointegrating vectors, *Econometrica*, 55, 1035-1056.

Stock, J.H. and M.W. Watson (1988), Testing for common trends, *Journal of the American Statistical Association*, 83, 1097-1107.

Theil, H. (1971), *Principles of Econometrics*, New York: John Wiley.

Tiao, G.C. and R.S. Tsay (1989), Model specification in multivariate time series (with discussion), *Journal of the Royal Statistical Society*, B-51, 157-213.

Trognon, A. (1978) Miscellaneous asymptotic properties of ordinary least square and maximum likelihood estimators in dynamic error components models, *Annales de L'INSEE*, 30-1, 631-657.

Wills, J. (1979), Technical change in the U.S. primary metals industry, *Journal of Econometrics*, 10, 85-98.