Efficiency gains for seasonal adjustment by joint modelling of disaggregated series

Carole Birrell
University of Wollongong

Efficiency Gains for Seasonal Adjustment by Joint Modelling of Disaggregated Series

A thesis submitted in fulfillment of the
requirements for the award of the degree

Doctor of Philosophy

from

University of Wollongong

by

Carole Birrell BSc (UNSW), DipEd, MStat (Wollongong)

School of Mathematics and Applied Statistics

2008

CERTIFICATION

I, Carole Louise Birrell, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Mathematics and Applied Statistics, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Carole Louise Birrell

3 August, 2008

Abstract

Governments and businesses use data collected over time as indicators of the social, economic and business conditions of the country. These may then be used for policy and planning decisions, calculation of national accounts and monitoring of economic activity. The production and publication of seasonally adjusted series, in addition to unadjusted figures, is standard practice for government statistical agencies. In general, there are two main approaches to seasonal adjustment, namely a filter-based approach and a model-based approach. Filter-based methods estimate time series components, such as the trend and seasonal factors, by application of a set of filters to the original series. Model-based methods of seasonal adjustment are more specific to each series, and are thereby more flexible.

Time series resulting from aggregation of several sub-series can be seasonally adjusted directly or indirectly. With model-based seasonal adjustment, the sub-series may also be considered as a multivariate system of series and hence the analysis may be done jointly. This approach has considerable advantage over the indirect method, as it utilises the covariance structure between the sub-series.

The focus of this thesis is on examining how the accuracy of seasonally adjusted series can be improved by using the sub-series. A model-based approach to seasonally adjusting an aggregated series is carried out with two different methods. The first method utilises an univariate basic structural model (BSM) for the aggregated series. The second method utilises a multivariate basic structural model for the sub-series. In basic structural models, the series components are modelled individually, and then put into state space form. The Kalman filter is applied to obtain estimates of the aggregate series components and the prediction mean squared errors.

The variance of the seasonally adjusted series given by the two methods is studied through their relative efficiency. A particular emphasis of the thesis is on how the similarity of and differences between disaggregated series affect the efficiency of the two approaches to seasonal adjustment.

Results indicate that gains are attainable under specified conditions which rely on the values of the parameters of not only the seasonal component, but also the non-seasonal components. These results demonstrate the impact on relative efficiency of relationships among sub-series parameters, both between series (i.e. within components) and within series (i.e. between components).

The impact of the length of the time series on the accuracy of seasonally adjusted series is of particular interest. A simulation study investigates the parameter estimates obtained given varying series lengths and the subsequent effects on the accuracy of the time series components given by the Kalman filter. These effects are measured by the naïve bias in the prediction mean squared error and by the revision error. A bootstrap correction is applied to the estimated prediction mean squared error for both the univariate and multivariate approaches.

A single indicator measure is developed for predicting whether the properties of the disaggregated series (or sub-series) will lead to gains in the accuracy of the seasonally adjusted aggregated series. The quasi-likelihood method is applied to obtain the indicator measure of relative efficiency. It is shown to be directly related to the relative efficiency measure obtained with the Kalman filter.

Another application of the quasi-likelihood indicator is in identifying an appropriate grouping of the K sub-series into $r < K$ series. The grouping can considerably reduce the number of estimated parameters, while the accuracy of the seasonally adjusted series is maintained.

The integrated approach of this thesis to the seasonal adjustment of aggregated series thus provides a pathway to improved efficiency and an understanding of the conditions under which improvements may be achieved.

Acknowledgements

This thesis would not have been possible without the help and support of many people. I would firstly like to thank my supervisors, Professor David G. Steel and Associate Professor Yan-Xia Lin. Thank you David for your wisdom, knowledge, guidance and great patience. I appreciate that your comments have always been positive and for believing that I would finish this thesis one day, even when I had trouble believing it would happen. Thank you also for the opportunities to travel to Berlin, Southampton and Luxembourg. To Yan-Xia, thank you for the time you have spent working through most of the detail of this thesis. Your help with the theory has been invaluable and I have appreciated all the discussions we have had to get to this point. You have taught me so much.

This project has been funded by the Australian Research Council and the Australian Bureau of Statistics. I would especially like to thank Dr Mark Zhang and Dr Craig McLaren of the Time Series Analysis division for their advice early on in this project and for the time they made available to meet with me on my visits to Canberra.

On my visit to Southampton, I had the opportunity to visit academics including Professor Danny Pfeffermann. Thank you for your suggestions and help with the direction of this thesis.

The journey of a PhD is more enjoyable if it can be shared with others in the same situation. Thanks to all my fellow PhD students, particularly Sandy, Elahe, Maureen and Raed, for sharing the experiences. Thanks especially to Sandy for your friendship and all the coffees we had together when we needed an excuse to have a break. Your advice and help have been much appreciated.

I would also like to thank the staff in the Department of Mathematics and Applied Statistics for their support and encouraging words over many years. In particular, a special thank you to Professor David Griffiths who encouraged me to come back to study after a long break. Your friendship, support and encouragement is greatly valued. Your willingness and time taken to proof-read the thesis is very much appreciated.

To my friends at Figtree Anglican Church and Bible study, thanks for your fellowship and prayers, and for understanding when things were so busy. Your loyalty and support is what makes you such great friends. Special thanks to Alison and the Kerr family for the times we have shared and for the many retail therapy sessions! Thank you to Kristina, for although our lives have taken different paths over these last few years, our friendship has only grown.

To my parents and family, thanks for your understanding, for always believing in me and spurring me on to get to the finish line. I know I wouldn't have achieved this without your support.

Most importantly, to my husband Michael, for the unfailing love, patience, support and encouragement you have always given me. Thank you for your graciousness in helping with the cooking, cleaning, the occasional programming, and for setting up the computers at home. I greatly appreciate all the little things you have done, such as making me coffee and bringing me chocolate to keep me at my desk! Thanks for the talks, for the walks on the beach and for just being there at the end of every day.

I thank my Lord Jesus Christ for sustaining me throughout this time. In the words of Ecclesiastes 3:1, "There is a time for everything and a season for every activity under heaven" (Holy Bible, New International Version). I am looking forward to the next season in my life with Him at my side. I know that whatever it may be,"I can do everything through Him who gives me strength" (Holy Bible, New International Version, Philippians 4:13).

Contents

1 Introduction **1**

2 Issues in Seasonal Adjustment **7**

2.1 Current Methods 8

2.1.1 Filter-based Seasonal Adjustment: X-11 8

2.1.2 Model-based Seasonal Adjustment 10

2.1.3 Parameter Estimation and Bias 13

2.2 Comparison of X-11 with Model-Based Estimates 15

2.3 Measuring Revision Error 17

2.4 Seasonal Adjustment of an Aggregate Series 18

2.5 Applications of the Multivariate Basic Structural Model 23

2.6 Short Time Series 24

2.6.1 Short Time Series and Longitudinal Analysis 26

2.7 Modelling Survey Errors within a State Space Model Context 27

2.8 Intervention Analysis 28

2.9 Conclusion 30

3 Modelling Disaggregated Series **32**

3.1 Introduction 32

3.2 Univariate BSM 33

3.2.1 The Trend Component 33

3.2.2 The Seasonal Component 34

3.2.3 The Disturbance Terms 35

3.3 Multivariate BSM . . . 35
3.3.1 Placing Restrictions on the Variance Structure . . . 37
3.4 State Space Form . . . 42
3.4.1 Univariate BSM in State Space Form . . . 42
3.4.2 Multivariate BSM in State Space Form . . . 44
3.4.3 Applying a Transformation . . . 46
3.5 Application of the Kalman Filter . . . 49
3.5.1 Kalman Filter for the Univariate Model . . . 49
3.5.2 Kalman Filter for the Transformed Multivariate Model . . . 51
3.5.3 Kalman Smoother . . . 52
3.6 Comparison of Univariate and Multivariate Methods . . . 54
3.6.1 Variance of the Seasonally Adjusted Series . . . 54

4 Empirical Study with Exact Parameters 57
4.1 Introduction . . . 57
4.2 Model Structure . . . 58
4.2.1 State Space Form . . . 60
4.3 Design of the Study . . . 62
4.3.1 Setting the Parameters . . . 63
4.3.2 Application of Constraints . . . 66
4.4 Estimation of Components via the Kalman Filter . . . 67
4.5 Results . . . 68
4.5.1 The Effect of the C-ratios . . . 68
4.5.2 The Effect of Correlation Settings . . . 73
4.6 Other Aggregated Series . . . 81
4.6.1 Results for other Aggregated Series . . . 83
4.7 Conclusion . . . 88

5 Quasi-Likelihood Indicator of Relative Efficiency 92
5.1 Introduction . . . 92
5.2 Introduction to Quasi-Likelihood Method . . . 94

5.3 Quasi-Likelihood Approach to the Univariate BSM 95
5.4 Quasi-Likelihood Approach to the Multivariate BSM 100
5.4.1 Transforming the Multivariate System 102
5.5 The Quasi-Likelihood Indicator 109
5.6 Comparison with Kalman Filter Results 110
5.6.1 Homogeneous Model 111
5.6.2 Model 1: Compound Symmetry 114
5.6.3 Model 2: Unit-specific Variances 116
5.7 Grouping Application 121

6 Parameter Estimation: Varying Series Length and Accuracy of Estimated Components 127

6.1 Introduction 127
6.2 Estimation of Parameters in the State Space Model 128
6.2.1 Univariate Model 130
6.2.2 Multivariate Model 133
6.3 Simulation Experiment 139
6.3.1 Generation of Data Series 139
6.3.2 Estimation of Parameters 142
6.4 Results of Parameter Estimates for Varying Series Length 144
6.4.1 Design *A1a14* 145
6.4.2 Design *A1b33* 151
6.5 Effect of Parameter Estimation on MSE of Model Components 157
6.5.1 Naïve Bias 159
6.5.2 Bias Correction 164
6.6 Revision Error 170
6.6.1 Revision Error Results 172
6.7 Summary 175

7 Handling Several Sub-series: An Example 177

7.1 Introduction 177

7.2 Data Generation of the Eight Sub-series . . . 178
7.3 Grouping the Sub-series . . . 184
7.3.1 Calculation of Q for each Combination . . . 187
7.4 Relative Efficiency Results . . . 189
7.5 Summary . . . 191

8 Conclusions **193**
8.1 Summary . . . 193
8.2 Future Directions . . . 196

A The Kronecker Product **199**

B Trigonometric Seasonality **201**
B.1 Introduction . . . 201
B.2 Univariate BSM . . . 201
B.2.1 Univariate State Space Form . . . 203
B.3 Multivariate BSM . . . 204
B.3.1 Multivariate State Space Form . . . 205
B.4 Calculation of Relative Efficiency . . . 206
B.5 Relative Efficiency Results . . . 207

C Derivations for Chapter 5 **208**
C.1 Quasi-Likelihood Result for the Error Variance Matrix of the Seasonal Components . . . 208
C.2 Denominator of the Quasi-Likelihood Indicator of Relative Efficiency . . . 212
C.3 Stirling Numbers of the Second Kind . . . 215

D Derivations for Chapter 6 **217**
D.1 Diffuse Log-likelihood . . . 217
D.1.1 Exact Initial Kalman Filter . . . 217
D.2 Derivations for Section 6.2.1 . . . 219

D.2.1 Calculation of w_t . . . 219
D.2.2 Concentrated Diffuse Log-likelihood . . . 220
D.3 Derivations for Section 6.2.2 . . . 220
D.3.1 Calculation of w_t . . . 220
D.3.2 Concentrated Diffuse Log-likelihood . . . 222

E Program Files **223**
E.1 Chapter 4 Programs . . . 224
E.2 Chapter 5 Programs . . . 230
E.3 Chapter 6 Programs . . . 231
E.3.1 Programs for Section 6.4.1: Design *A1a14* . . . 231
E.3.2 Programs for Section 6.4.2: Design *A1b33* . . . 233
E.3.3 Programs for Section 6.5.1: Naïve Bias . . . 235
E.3.4 Programs for Section 6.5.2: Bootstrap Correction . . . 236
E.3.5 Programs for Section 6.6: Revision Error . . . 237
E.4 Chapter 7 Programs . . . 238
E.5 Appendix B Programs . . . 242
E.6 Appendix D Programs . . . 242

Bibliography **243**

List of Figures

4.1 $RE_t(M)$ for sub-series design *A1a*. 70

4.2 $RE_t(M)$ for sub-series design *A1b*. 71

4.3 $RE_t(M)$ for design *a12* with *A1 - A5*. 73

4.4 $RE_t(M)$ for design *a13* with *A1 - A5*. 74

4.5 $RE_t(M)$ for design *a14* with *A1 - A5*. 74

4.6 Seasonal correlation versus $RE_{40}(M)$ for designs *a11*, *a12*, *a13*, *a14* with *A1-A5*. 75

4.7 $RE_t(M)$ for design *a21* with *A1 - E1*. 76

4.8 $RE_t(M)$ for design *a31* with *A1 - E1*. 77

4.9 $RE_t(M)$ for design *a41* with *A1 - E1*. 77

4.10 Non-seasonal correlation versus $RE_{40}(M)$ for designs *a11*, *a21*, *a31*, *a41* with *A1 - E1*. 78

4.11 $RE_t(M)$ for designs *A1a11* to *A1a44* given by *Series 1, 2* and *3*. . . 86

4.12 $RE_t(M)$ for designs *A1b22* to *A1b44* given by *Series 1, 2* and *3*. . . . 87

5.1 Quasi-likelihood indicator vs KF relative efficiency 120

6.1 Box plot of the level parameter estimate for design *A1a14*. 146

6.2 Box plot of the seasonal parameter estimate for design *A1a14*. 147

6.3 Box plot of the measurement error parameter estimate for design *A1a14*.147

6.4 Box plot of the level parameter estimate for design *A1b33*. 151

6.5 Box plot of the seasonal parameter estimate for design *A1b33*. 152

6.6 Box plot of the measurement error parameter estimate for design *A1b33*.152

6.7 Box plot for relative efficiency of revision error variance 174

6.8 Box plot for relative efficiency of revision error variance with limited scale 174

7.1 Logarithms of quarterly wine sales by type 180

7.2 Simulated series of the logarithms of quarterly wine sales by type . . 183

List of Tables

4.1 Labels for sub-series design '*a*'. 64
4.2 Labels for sub-series design '*b*'. 65
4.3 Labels for correlation combinations. 65
4.4 Results of $RE_{40}(M)$ for sub-series design *A1a*. 70
4.5 Results of $RE_{40}(M)$ for sub-series design *A1b*. 71
4.6 Results of $RE_{40}(M)$ for design '*a*' correlation combinations. 79
4.7 Results of $RE_{40}(M)$ for design '*b*' correlation combinations. 81
4.8 *Series 1, 2*, and *3*: results of $RE_{40}(M)$ for design *A1a*. 85
4.9 *Series 1, 2*, and *3*: results of $RE_{40}(M)$ for design *A1b*. 85

5.1 Results of $RE_{40}(M)$ and Q for sub-series design '*a*': part 1. 117
5.2 Results of $RE_{40}(M)$ and Q for sub-series design '*a*': part 2. 119
5.3 Results of $RE_{40}(M)$ and Q for sub-series design '*b*'. 120

6.1 Parameters for sub-series design *A1a14* and *A1b33*. 140
6.2 Median parameter estimates for *Series 2* for design *A1a14*. 148
6.3 Means and standard errors of parameter estimates for design *A1a14*. . 149
6.4 Relative efficiency of the variances of the parameter estimates for design *A1a14* 150
6.5 Median parameter estimates for design *A1b33*. 154
6.6 Mean and standard errors of parameter estimates for design *A1b33*. . 155
6.7 Relative efficiency of the parameter estimates for design *A1b33*. . . . 156
6.8 Mean and RMSE of relative naïve bias for design *A1a14*. 163
6.9 Mean and RMSE of relative naïve bias for design *A1b33*. 163

6.10 Naïve bias results for design *A1b33* with $T = 28$. 169
6.11 Bootstrap correction for design *A1b33* with $T = 28$ 169
6.12 Relative efficiency of the revision error variance for design *A1a14* and *A1b33*. 173

7.1 Parameters used for generation of the eight sub-series 181
7.2 Initial values used for generation of the eight sub-series 182
7.3 C-ratios for pairs of series 186
7.4 Different combinations of the eight sub-series 187
7.5 Values of Q and relative efficiency for nine different combinations . . 189

B.1 Results for $RE_{40}(M)$ for trigonometric seasonal component. 207

C.1 Stirling numbers of the second kind 216

List of Abbreviations

ABS	Australian Bureau of Statistics
AMB	ARIMA-model-based
AR	autoregressive
ARIMA	autoregressive integrated moving average
ARMA	autoregressive moving average
BLS	Bureau of Labor Statistics
BSM	basic structural model
CPS	Current Population Survey
IGLS	iterative generalised least squares
KF	Kalman filter
KS	Kalman smoother
LLS	local level seasonal
MSE	mean squared error
QL	quasi-likelihood
PMSE	prediction mean squared error
RMSE	root mean squared error
SEA	survey error autocorrelations
SEATS	Signal Extraction in ARIMA Time Series
SSF	state space form
STAMP	Structural Time series Analyser, Modeller and Predictor
SUTSE	seemingly unrelated time series equations
TRAMO	time series regression with ARIMA noise, missing values, and outliers
UC	unobserved components

Chapter 1

Introduction

Governments and businesses use data collected over time as indicators of the social, economic and business conditions of the country, which may then be used for policy and planning decisions, calculation of national accounts and monitoring of economic activity. These time series are collected world-wide by government statistical agencies, including the Australian Bureau of Statistics (ABS). Some of the key time series include those of retail expenditure, employment and earnings, business turnover, and unemployment. Since analyses of these series are fundamental to the expenditure of billions of dollars that ultimately shape the direction of the national economy, it is vital to accurately estimate the underlying movement of the time series.

The production and publication of seasonally adjusted series, in addition to unadjusted figures, is standard practice for government statistical agencies. Three main objectives of seasonal adjustment are suggested by Bell and Hillmer (1984):

- to aid in short-term forecasting;
- to aid in relating time series to other series or extreme events;
- to allow comparability in the series from month to month.

Seasonally adjusted series are often quoted in the media, and are used by analysts in

business and governments in the development and monitoring of plans and policies. The first step in obtaining the seasonally adjusted series is to estimate the seasonal effects, which are the effects that regularly occur in the same month or quarter of each year. Secondly, the seasonal effects are removed from the original series so that the effects of other influences are more visible. A seasonally adjusted series contains a trend and an irregular component. For example, retail trade figures are higher in December than in other months of the year, largely due to the purchases associated with Christmas. This implies that there is a seasonal effect for December. Estimation and then removal of the seasonal factors for each month of the year result in the seasonally adjusted series.

It is important that seasonal factors be estimated accurately. Suppose, for example, that there is no underlying growth in the retail trade series to be analysed and that the true seasonal factor for in December is actually 30% higher than that for November. However, the estimated value is only 29%. If, for a particular December, the increase in the original series is 31%, then the estimated increase in the seasonally adjusted series is 2%, whereas the actual increase is only 1%. The increase is therefore estimated to be twice what it should be. This could be interpreted by analysts as showing stronger growth in the economy than expected, and could lead to incorrect decisions regarding economic direction and policy, such as the setting of interest rates.

The precision of seasonally adjusted series is also important. Recognition of the importance of the standard errors of seasonally adjusted series dates back to the early 1960's. Hausman and Watson (1985) report that in 1962, the (U.S.A) President's Committee to Appraise Employment and Unemployment Statistics (the Gordon Commission) recommended

> "*that estimates of the standard error of seasonally adjusted data be prepared and published as soon as the technical problems have been solved".*

The comment referred to the seasonally adjusted data which had been calculated by the filter-based X-11 procedure but is equally applicable to other methods of seasonal adjustment.

In general, there are two main approaches to seasonal adjustment, namely a filter-based approach and a model-based approach. Filter-based methods estimate time series components, such as the trend and seasonal factors, by application of a set of filters to the original series. This iterative technique is applied by the widely used X-11 package (Shiskin *et al.*, 1967). With filter-based methods, basically the same overall procedure is applied to all time series, although the specific filters used may vary according to the properties of the series. In this sense, it can be regarded as inflexible (Harvey, 1989), but convenient to apply to many series.

Model-based methods fit a particular statistical model to the series. The models employed could be ARIMA (Autoregressive Integrated Moving Average) models or structural time series models. The ARIMA-model-based (AMB) approach to seasonal adjustment (Hillmer and Tiao, 1982) is also widely used in statistical offices and institutions throughout the world. It is implemented with the TRAMO-SEATS (Time series Regression with ARIMA noise, Missing values, and Outliers - Signal Extraction in ARIMA Time Series) software (Gomez and Maravall, 1996). Structural time series models are models where each component of the series are modelled separately. They are appropriate for modelling time series from many disciplines such as economics, sociology, engineering and geography (Harvey, 1989, p xi). The Kalman filter (Kalman, 1960), first developed in the field of engineering, is the algorithm used to estimate the structural components of the model. Programs such as STAMP (Structural Time series Analyser, Modeller and Predictor) and others such as the set of functions collectively called the *SsfPack* (Koopman *et al.*, 1999) in the module `S+FinMetrics` may be used for the analysis. The AMB approach is generally not suitable for multivariate analysis, whereas in the structural time series model known as the basic structural model (BSM) (Harvey, 1989), multivariate analysis is a natural extension to the univariate model.

Model-based methods of seasonal adjustment are more specific to each series, and are thereby more flexible.

> "*The theoretical argument in favour of model-based seasonal adjustment is very strong.*" (Harvey, 1989, p306)

The basic structural model may be used for seasonal adjustment. As the structural components such as the level, slope and seasonal factors are modelled individually, the specification of the model may be different for different series.

> "*The use of modelling in connection with seasonal adjustment raises the basic question of whether seasonal adjustment should be done at all.*" (Bell and Hillmer, 1984, p100)

Harvey (1989) echoes this comment with reference to extrapolation as a reason for carrying out seasonal adjustment:

> "*However, if a series has been seasonally adjusted by fitting a model to it, this becomes pointless since optimal predictions can be made directly from the model.*" (Harvey, 1989, p308)

Harvey goes on to suggest that in some circumstances it is interesting to look at a series to understand its history and hence that seasonal adjustment is helpful. He also suggests that working with the seasonally adjusted series may assist in the detection of breaks and structural changes and this may be beneficial at the model formulation stage.

Seasonal adjustment of an aggregate series is usually of particular interest to statistical agencies. It may be carried out directly, with the analysis performed on the aggregate series itself. Alternatively, indirect seasonal adjustment is performed on each of the cross-sectional series (or sub-series) that make up the aggregate series. The seasonally adjusted sub-series are then summed to obtain the seasonally adjusted aggregate series. With filter-based methods, the results of these two approaches rarely agree. With model-based seasonal adjustment, the sub-series may be considered as a multivariate system of series and hence the analysis may be done jointly. This approach has considerable advantage over the indirect method as it utilises the covariance structure between the series. When the sub-series are assumed to be independent, the multivariate approach is equivalent to the indirect approach. Thus, the indirect approach can be regarded as a special case of a multi-

variate model-based approach. The aggregate series may also be analysed with an univariate model which corresponds to the direct method of seasonal adjustment.

Given the importance to statistical agencies and users of the data, the focus of this thesis is on examining the accuracy of seasonally adjusted series. The emphasis is on a model-based approach to seasonally adjusting an aggregate series via two different methods. Firstly, an univariate basic structural model (BSM) is applied to the aggregate series to obtain estimates of the series components. Secondly, the subseries are modelled jointly with a multivariate basic structural model to obtain the aggregate series components that utilises the information drawn from these series. The research will seek to elucidate conditions under which application of a multivariate model will improve the quality of estimated seasonal factors, and hence the seasonally adjusted series. A particular emphasis is on how the similarity of and the differences between disaggregated series affect the efficiency of seasonal adjustment approaches. The impact of the length of the time series on the accuracy of seasonally adjusted series will also be investigated for both univariate and multivariate approaches.

A review of the literature in Chapter 2 reveals many applications of model-based seasonal adjustment and its comparison to the more widely used filter-based method X-11 and associated variants. A bivariate basic structural model has been applied to improve estimates of the components of a target series by jointly modelling a related series. These and other relevant topics are reviewed and discussed in Chapter 2. This discussion leads to the multivariate model which is proposed in Chapter 3. By writing both the univariate and multivariate models in state space form, the Kalman filter (Anderson and Moore, 1979) can be applied, yielding estimates of the series components. The procedure is detailed in Chapter 3.

The estimation of a seasonally adjusted series is evaluated by several well known criteria. The accuracy of the seasonally adjusted series is measured by determining its variance. For a basic structural model, this is a by-product of the application of the Kalman filter. The variance of the seasonally adjusted series given by the univariate and multivariate models will be compared using their relative efficiency.

In Chapter 4, an empirical study will thoroughly investigate the conditions which affect relative efficiency. This will be carried out by fixing the known parameters of an aggregated series, but varying the parameters of the sub-series.

In Chapter 5, a single indicator measure is developed for predicting whether the properties of the sub-series will lead to gains in the accuracy of the seasonally adjusted aggregate series. A quasi-likelihood method is applied to the sub-series to obtain the indicator of relative efficiency, given the series parameters.

The length of a time series is crucial to the accuracy of model parameter estimates. For the univariate and multivariate approaches, a simulation study investigates the parameter estimates obtained given varying series length. Estimation of model parameters, and the subsequent effects on the estimates of the components and their mean squared errors, is the focus of Chapter 6.

Revision error is an important measure which attracts a lot of attention in government statistical agencies. It is the error associated with the degree of revision an adjusted series undergoes when new observations become available. The revision error for the two methods will also be compared in Chapter 6.

Lastly, if there are several sub-series, the question arises as to whether grouping them can maintain or improve the accuracy of the seasonally adjusted series. By grouping sub-series to form new series, the number of parameters to be estimated is reduced. The method of grouping the series is demonstrated by an example in Chapter 7.

If the joint modelling of the disaggregated series increases the reliability of the seasonally adjusted aggregate series, and the degree of revision is also reduced, then users will have more confidence in the published seasonally adjusted series. Analysts will be able to detect more accurately important changes in a series and hence to make more informed decisions regarding the direction of the economy. In particular, this research has application in the seasonal adjustment of short to moderate length time series where the component estimates are often considered experimental (ABS, 2007a). On a wider scale, it has implications for the area of linking two time series to improve the estimation of one of them by using a bivariate modelling approach.

Chapter 2

Issues in Seasonal Adjustment

The questions considered by this thesis relate various topics discussed in the statistical literature. Seasonal adjustment may be performed by several different methods, such as filter-based or model-based methods, using a single aggregate series, or by considering disaggregated series. Filter-based approaches such as X-11 have been compared to model-based approaches in terms of their capacity to estimate the series components such as the trend component, comprised of the level and slope of the series, and also the seasonal component. Component estimation in a multivariate model-based context will be reviewed with reference to not only the use of sub-series but also the use of a related series for the improvement of the component estimates of a target series. Another important factor in estimating series components is the length of the time series. A review of the literature will include the topic of estimating components of series with different lengths.

2.1 Current Methods

2.1.1 Filter-based Seasonal Adjustment: X-11

Many government statistical agencies use a filter-based or empirical approach to decompose an observed series into several unobserved components. The most widely used of these empirical methods is the X-11 method developed in 1965 by Shiskin (Shiskin *et al.*, 1967). X-11 evolved to include the use of ARIMA models to forecast the series in order to improve the estimation of components at the end of the series, thereby reducing revisions. Developments include X-11-ARIMA and X-11-ARIMA-88 developed by Dagum (Dagum, 1988) and X-12-ARIMA (U.S. Bureau of Census). For an historical review of seasonal adjustment see Bell and Hillmer (1984) or Ladiray and Quenneville (2001).

The X-11 method treats the observed value at a particular time, Y_t, as a function of three components; the trend (TC_t) which incorporates the overall cycle, the seasonal factor (S_t), and lastly, the irregular component (I_t). Assuming the effects are additive, the observed value is written as:

$$Y_t = TC_t + S_t + I_t \qquad \text{for} \qquad t = 1, \ldots, T \tag{2.1}$$

Smoothing techniques such as moving averages are applied iteratively to obtain estimates of the trend and seasonal components. The moving average of coefficients is defined in Ladiray and Quenneville (2001, Section 3.1.1) as

$$M(X_t) = \sum_{k=-p}^{+f} \theta_k X_{t+k}. \tag{2.2}$$

The quantity $p + f + 1$ is called the moving average order and $\{\theta_k\}$ is the set of coefficients.

A $P \times Q$ moving average is obtained by applying in succession a simple moving average of order P (with coefficients all equal to $1/P$) and a simple moving average of order Q (with coefficients all equal to $1/Q$). Thus, a 2×12 moving average for a monthly series has coefficients $\{1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1\}/24$ and therefore

requires 13 data points. A 3×3 moving average for a monthly series has coefficients $\{1, 2, 3, 2, 1\}/9$ and therefore requires 5 data points (Ladiray and Quenneville, 2001, Section 3.2).

The eight basic iterative steps in X-11 for monthly series are detailed in Ladiray and Quenneville (2001, Section 2.4) and summarised here as follows:

1. Estimate the trend using a 2×12 moving average (or 13 point) on the original series, which yields $TC_t^{(1)}$.

2. Remove the estimated trend to leave a series comprising a seasonal component and an irregular component, $(S_t + I_t)^{(1)}$.

3. Estimate the seasonal component using a 3×3 moving average (5-point) applied to $(S_t + I_t)^{(1)}$ and normalise the coefficients so that their sum is zero.

4. Remove the normalised seasonal component from the original series to obtain the estimated seasonally adjusted series, $Ysa_t^{(1)}$.

5. Re-estimate the trend using a 13-term Henderson moving average on $Ysa_t^{(1)}$ to give $TC_t^{(2)}$. (For information on Henderson moving averages used in X-11, refer to Ladiray and Quenneville, 2001, Section 3.2.2).

6. Remove the newly estimated trend from the original series to again leave seasonal and irregular components, $(S_t + I_t)^{(2)}$.

7. Estimate the seasonal component using a 3×5 moving average (7-point) applied to $(S_t + I_t)^{(2)}$ to give $S_t^{(2)}$. Normalise the coefficients to give $Snorm_t^{(2)}$.

8. Remove seasonal component $(Snorm_t^{(2)})$ from the original series to obtain the estimated seasonally adjusted series, $Ysa_t^{(2)}$.

Details of the X-11 method with a complete example can be found in Ladiray and Quenneville (2001). Other adjustments may also be included for outlier detection and removal, moving holidays (such as Easter) and trading day effects.

For many economic time series, a multiplicative model say, $Y_t = TC_t \times S_t \times I_t$ for $t = 1, \ldots, T$, is more appropriate than an additive model as given by (2.1), but the

same iterative approach may be applied. This thesis will concentrate on additive models.

The methods described above rely on the strength of the historical data. Data for at least five years are usually required to estimate seasonal factors. For example, if the series consisted of monthly data, then to calculate a 3×5 (or 7-point) moving average for December, seven data points for December (i.e. 7 years) are required. If instead a 3×3 (or 5-point) moving average was applied, only five data points (i.e. 5 years) would be required. Although data for at least five years are being utilised, the seasonal effects for a particular month are estimated by only five data points. Seasonally adjusted and trend series are likely to be considered experimental if fewer than five years of data are available for estimation of seasonal factors (ABS, 2007a).

2.1.2 Model-based Seasonal Adjustment

Model-based seasonal adjustment requires modelling the observed time series and the unobserved components such as the trend, seasonal and irregular components. There are two main approaches to modelling time series: the ARIMA model-based (AMB) approach (see Burman, 1980; Hillmer and Tiao, 1982) and the structural time series (STS) approach (see Engle, 1978; Harvey and Todd, 1983; and Harvey, 1989). The ARIMA model-based approach to seasonal adjustment involves fitting a seasonal ARIMA model to the overall series and then decomposing it into appropriate models for each of the components (Maravall, 1995). This approach is often called 'signal extraction' (Whittle, 1963 and Burman, 1980). The structural time series approach directly specifies models for each of the unobserved components which have a direct interpretation.

In the AMB approach, the estimation process requires two steps: estimating the ARIMA model and decomposing it into the additive unobserved components (signal extraction). In the development of ARIMA models, there was difficulty in dealing with series which had non-stationary trends and changing seasonality. Box and Jenkins (1970) use differencing to eliminate the trend and seasonal effects instead

of modelling the components separately. Differencing a series involves the lag or backward shift operator, B, which is defined as $B^k y_t = y_{t-k}$. The first-difference operator, Δ, is defined as $\Delta = 1-B$, and hence a first differenced series is determined by $\Delta y_t = (1 - B)y_t = y_t - y_{t-1}$. The seasonal difference operator, Δ_s, is defined as $\Delta_s = (1 - B^s)$, where s is the number of seasons in the data (Harvey, 1989, Section 2.1.2). Applying the seasonal difference operator yields $\Delta_s y_t = (1-B^s)y_t = y_t - y_{t-s}$. A series may need more than one application of differencing to attain stationarity. For a general review see Pena *et al.* (2001, Chapter 8).

Estimation of the unobserved components may be carried out with the Wiener-Kolmogorov filter (Burman, 1980 and for an illustration with economic time series refer to Maravall, 1995). Let $y_t = m_t + n_t$ where y_t is an observable time series and m_t and n_t are the unobservable signal and noise time series respectively. When m_t and n_t are independent and stationary, signal extraction may be carried out by applying the Wiener-Kolmogorov filter. Burman (1980) uses a partial fraction decomposition of the filter to obtain the estimate of the seasonal component. When either m_t or n_t or both are non-stationary, signal extraction may be performed by the method discussed in Bell (1984) which makes assumptions about the generation of the series m_t, n_t, y_t, and their starting values. Gomez (1999) also discusses the estimation of unobserved components in non-stationary time series. He shows that the same result is obtained with three different methods of filtering non-stationary time series under certain assumptions. The three methods applied are Kalman filtering plus smoothing, Wiener-Kolmogorov filtering plus Tunnicliffe Wilson algorithm (as given in Burman, 1980), and penalised least squares smoothing. He suggests that if the standard errors of the estimator are required, then the Kalman filter approach is the only approach which supplies this information under very general conditions (Gomez, 1999, p109).

In practice, estimation of the ARIMA model may be carried out using the TRAMO (Time series Regression with ARIMA noise, Missing values, and Outliers) program and the decomposition is carried out with the SEATS (Signal Extraction in ARIMA Time Series) program. The programs are often referred to jointly as

TRAMO-SEATS (developed by Gomez and Maravall, 1996), and is widely used in many statistical agencies and banks, particulary in Europe. The SEATS program implements the method proposed by Burman (1980). A helpful summary can be found in Pollock (2002). With TRAMO-SEATS, one has the advantage of an automatic model selection procedure (see Maravall, 2006 for a summary and example) which does not require prior knowledge about the structure of the series.

In general, STS models use an additive structural decomposition of the series: $y_t = TC_t + S_t + I_t$. In the case of multiplicative series ($y_t = TC_t \times S_t \times I_t$), this is handled by taking the logarithm of the series. The unobserved components are the linear trend cycle component TC_t, the seasonal component S_t, and an irregular or noise component I_t. These components may be modelled in various ways. In the Basic Structural Model (BSM), the unobserved components may be modelled by Markov processes (Harvey and Todd, 1983). The BSM is described in detail in Harvey (1989). The main feature of the BSM is that the trend is modelled as a local approximation to a linear trend defined by a level and slope component. Both the level and slope may change over time according to a random walk process. Two of the more commonly used models for the seasonal component are the dummy variable seasonal model and the trigonometric seasonal model. Both models allow for changing seasonality over time while ensuring that the sum of the seasonal components over any s time periods has an expected value of zero. The univariate BSM with a dummy seasonal component is described in Chapter 3, Section 3.2 and the trigonometric seasonal component is described in Appendix B. The BSM is written in state space form and application of the Kalman filter (Section 3.5.1) yields the optimal estimates of the individual components. Thus, by subtracting the estimated seasonal component, the seasonally adjusted series is readily obtained.

Riani (1998) compares the weights applied to observations in the two model-based approaches: AMB and STS. Theoretically, he shows that unless the variability in the seasonal component is very large with respect to the non-seasonal components, the two approaches produce very similar seasonally adjusted series. The main differences were found to refer to the trend.

There are many advantages in applying a structural model over an ARIMA model. The BSM and its state space form (Durbin and Koopman, 2001) are based upon the structural analysis of the series. Each component such as the trend and the seasonal factor is modelled explicitly, before being included in the state space model. This gives the state space approach great flexibility and transparency. Consequently, known changes in the structure over time can be allowed for with relative ease. Even though the estimates of the trend and seasonal components may be recovered from the differenced series in the AMB approach, it is more convenient to model them directly, as can be achieved using the structural model (Durbin and Koopman, 2001, Section 3.5). Seasonal adjustment in the STS approach is more straightforward.

The main advantage of applying a BSM, for the purposes of this thesis, is that multivariate series may be modelled as an extension to the univariate model. Other advantages are that the structural components are a direct result of the estimation via the Kalman filter and the mean squared errors of the estimators are also readily available.

2.1.3 Parameter Estimation and Bias

If the level, slope and seasonal components in a BSM are allowed to change over time, then the variances of their disturbance terms as well as the variance of the measurement error or irregular term, comprise the set of parameters for the model (see Section 3.2 for a description of the univariate BSM). In the literature, these parameters are sometimes referred to as the hyper-parameters of the model. When these parameters are unknown, maximum likelihood estimates can be obtained using the Kalman filter. These parameter estimates are then substituted into the state space model and application of the Kalman filter yields estimates of the series components and their associated mean squared errors (MSE). However, the resulting MSE is an underestimate of the true MSE (Ansley and Kohn, 1986). This has been termed the 'naïve' bias or the bias associated with naïve approximation. Naïve approximation allows for filter uncertainty but not for parameter uncertainty. The

underestimation is a result of not accounting for the variability resulting from the estimation of the model parameters (Ansley and Kohn, 1986; Durbin and Koopman, 2001; Quenneville and Singh, 2005; Pfeffermann and Tiller, 2005). This is discussed in more detail in Chapter 6.

Methods of calculating the naïve bias and procedures which reduce the bias are given in Quenneville and Singh (2005) and Pfeffermann and Tiller (2005). Both papers include a summary and a comparison of bias corrections proposed by other authors, including Hamilton (1986), and Ansley and Kohn (1986). An alternative bias correction technique is also proposed in each paper.

Quenneville and Singh (2005) performed a Monte Carlo study with series of lengths $T = 40$ and $T = 100$ for a model which comprised of a random walk plus noise. The average relative bias was calculated for six bias correction procedures and compared with the naïve bias. One of the main conclusions is related to the series length. The naïve bias was greater in absolute terms for the moderate length ($T = 40$), than for the longer series ($T = 100$). For the moderate length, there was a serious underestimation, with a relative bias of -21.2%, whereas for the series with $T = 100$, the relative bias was -9.0%. Using the corrected Ansley and Kohn approximation, the bias was reduced to -13.0% for $T = 40$, and -4.1% for $T = 100$. Another bias correction reduced these to -6.7% and -2.5% for moderate and long length respectively. Quenneville and Singh (2005) concluded that the corrected Ansley and Kohn approximation was the best compromise between bias, precision, theoretical exactness and computational requirements.

Pfeffermann and Tiller (2005) proposed a parametric and a non-parametric bootstrap method as bias corrections to the naïve bias. The bootstrap methods are compared to other bias correction methods discussed in Quenneville and Singh (2005). The robustness of the bootstrap method is examined with respect to non-normality of the model error terms. The distribution of the maximum likelihood parameter estimates can be skewed, especially for short series or when the parameters are close to the boundary value of zero. The results of an extensive simulation study show that the bootstrap methods are 'much superior to' the other methods in terms of bias,

and this is especially so for shorter series. In a study similar to that of Quenneville and Singh (2005), the naïve bias of -18.5% for a series with $T = 40$ was reduced to 0.6% with the parametric bootstrap method. Similarly, the naïve bias of -7.6% for a series with $T = 100$, was reduced to 1.6% (Pfeffermann and Tiller, 2005, p903).

These results have definite implications to the method proposed in this thesis. As shorter series are considered, and if parameters are estimated, then the bias in the mean squared error of the seasonal factor will need to be measured for both the univariate and multivariate approaches. This is discussed in Chapter 6.

2.2 Comparison of X-11 with Model-Based Estimates

The BSM can be fitted to series for which X-11 has been found to perform well (Maravall, 1985). This is further discussed in Harvey (1989, Section 6.2.3), where the autocorrelation function of a particular series obtained with X-11 is shown to be almost identical to that obtained by applying the BSM. However, for most series, X-11 tends to yield a smoother trend component than the BSM.

To obtain estimates of structural components analogous to those produced with the X-11 procedure, the smoothed estimates obtained from the Kalman filter are appropriate. Dagum and Quenneville (1993) as well as Pfeffermann *et al.* (1998) obtained model-based estimates using the Kalman smoother and compared them with the corresponding X-11 estimates for univariate time series.

Dagum and Quenneville (1993) compared a seasonally adjusted series estimated by a model-based procedure with that obtained by an X-11-ARIMA method. The two procedures were applied to the Total Sales for Department Stores in Canada from January 1975 to December 1985. The unobserved components model included second differences for the trend, a stochastic seasonal component comprised of dummy variables, and a deterministic or fixed trading day component, as well as an irregular component. It did not model survey errors. Initial estimates for the Kalman filter were supplied by estimates from the X-11-ARIMA decomposition. The results obtained with the unobserved components model were found to be very

close to those obtained with the standard options of X-11-ARIMA.

Pfeffermann *et al.* (1998) compared trend estimates from a state space model with corresponding estimates produced by X-11-ARIMA. They found that the model-based trends were much smoother than the X-11 trends. The series being investigated was the result of a survey and therefore included survey errors. By modelling the survey errors and hence subtracting them from the original data values, however, the trends produced by X-11 were very similar to the trends produced by the model.

> "*Thus, the use of X-11 fails to separate the genuine trend of the population values from the spurious trends induced by the correlation pattern of the survey errors.*" (Pfeffermann *et al.*, 1998, p346)

A similar result was obtained for the comparison of the seasonal effects. The seasonal effects estimated with X-11 for the survey error adjusted series were almost identical to the seasonal effects obtained under the state space model. (Pfeffermann *et al.*, 1998, p347).

Moosa and Lenten (2000) compared seasonal factors obtained by X-11 with those obtained by a basic structural model (BSM) on several Australian time series. The BSM parameter estimates for each series were obtained by maximum likelihood using the Kalman filter. The results varied with each series, where for some series, seasonal factors differed significantly between methods. The paper by Moosa and Lenten does not specify if any of the series results from a survey and does not specify if a survey error was built into the model as applied by Pfeffermann *et al.* (1998). If any of the series resulted from a survey, the differences found may well be due to the model misspecification given the survey error built into the data.

The model that has been shown to be appropriate for modelling economic time series and for which estimates of components such as seasonal factors are comparable to those using X-11, is the basic structural model (BSM) sometimes simply referred to as a structural time series model. By putting this model into state space form and using the Kalman filter, the time series components can be estimated. Discussion in Section 6.2.3 of Harvey (1989), mentions that the BSM is applicable to series for which X-11 is appropriate. From the research conducted by Pfeffermann *et al.*

(1998), to achieve comparability of estimates of components, it will be necessary to include the survey error in the model for those series which originate from a sample survey.

The differences between seasonally adjusting series using the X-12-ARIMA (Findley *et al.*, 1998) and TRAMO-SEATS (Gomez and Maravall, 1996) programs are discussed in Planas and Depoutot (2002). They compare the moving average filters of X-11 to the Wiener-Kolmogorov filters used in the AMB approach and notice that the X-11 filters have finite length whereas the Wiener-Kolmogorov filters are infinite. With reference to the airline model (Box and Jenkins, 1970), they show that the two programs can yield very similar adjustments.

2.3 Measuring Revision Error

As new observations become available in a series, the seasonally adjusted values at the end of the series need to be revised. This is most obvious for seasonal adjustment procedures which employ moving averages, as data points previous to, and after, the observation are required. These changes are called revisions, and are a major cause of concern, especially when the decomposition of the series is to be published. The revision error is commonly defined as the difference between the concurrent or real-time seasonally adjusted series (given by $Y^a_{t|t} = Y_t - S_{t|t}$) and the seasonally adjusted series calculated h periods ahead (given by $Y^a_{t|t+h} = Y_t - S_{t|t+h}$), where the 'a' superscript denotes 'adjusted'. Pierce (1980) developed a characterisation of seasonal revisions in terms of stationary and non-stationary linear time series models. He showed that a preliminary seasonal estimate and its successive revisions are mutually independent. This is primarily due to the mutual independence of the innovations (or disturbance terms) of the observable series (Pierce, 1980). He also examined revisions calculated with the X-11 procedure and compared them with those obtained with X-11-ARIMA.

Revisions are not usually applied to seasonally adjusted data which are more than three years old. When the variance of the final adjusted figure of an observa-

tion at time t, using h subsequent observations ($Y^a_{t|t+h}$), is reduced to almost half the variance of the current adjusted figure ($Y^a_{t|t}$), the period is referred to as the half-length. For X-11, the filters have a half-length of approximately seven years. However, due to the latter weights being very small, three years is usually adequate for revision purposes (Burridge and Wallis, 1984), after which the reductions are negligible.

A standard error may be applied to preliminary adjusted values as an indication of their reliability. For a basic structural model, the variance of the seasonally adjusted series is available and is the focus of the paper by Burridge and Wallis (1984). This is discussed in more detail in Section 3.6.1 of this thesis. Revision error, for both the univariate approach and multivariate approach, is detailed in Chapter 6.

Planas and Rossi (2004) investigate the reliability of real-time output gap measures and whether using inflation data improves the revision errors. They note that the standard error of the revisions rely on the model parameters, which in turn depend upon the length of the series. An empirical approach to the calculation of the revision error variance is therefore proposed. Their method is based on recursively estimating the model parameters as new observations become available. They conclude that for three out of the four cases tried, in comparison to univariate modelling, the bivariate approach of Phillips curve modelling substantially improved the reliability of the estimate of real-time output gap measures (Planas and Rossi, 2004, p128).

2.4 Seasonal Adjustment of an Aggregate Series

If an aggregate series is broken down into a several sub-series, then seasonal adjustment may be performed on the aggregate series alone, or using each of the sub-series.

The definitions of direct and indirect seasonal adjustment are reviewed:

1. In direct seasonal adjustment, the seasonal component is estimated from a single (aggregated) series and then removed.

2. In indirect seasonal adjustment, each of the sub-series is seasonally adjusted separately and then they are summed to obtain the seasonally adjusted aggregate series.

There is extensive debate on whether to use the indirect or the direct approach to seasonal adjustment (see Ghysels, 1997; Hood and Findley, 2003; Ladiray and Mazzi, 2003; and Otranto and Triacca, 2002). Most of the discussion focuses on filter-based methods such as X-11, and its subsequent variants, due to the fact that the seasonally adjusted series resulting from the two methods can, and usually do, differ (Hood and Findley, 2003). Questions arise as to which method produces the more accurate estimates, how to compare the methods, and, if one method is not always better than the other, under what conditions each method should be employed.

Both direct and indirect seasonal adjustment employ univariate analyses. Although the indirect method utilises all the sub-series, it does not do so jointly. By using a model-based approach, seasonal adjustment may be performed on the aggregate series using all the information in the sub-series by borrowing strength from the connections between the sub-series and their components. The indirect approach is not considered in this thesis, apart from taking into account the results from previous studies as a guide to designing the experimental study discussed in later chapters. However, the discussion of direct versus indirect seasonal adjustment provides insight into the possible situations which may benefit from a multivariate approach.

A thorough discussion of the direct versus indirect methods is given in Hood and Findley (2003) with reference to the X12-ARIMA and the SEATS (Signal Extraction in ARIMA Time Series) programs. In general, they comment that when the sub-series

> "*have quite distinct seasonal patterns and have adjustments of good quality, indirect seasonal adjustment is usually of better quality than direct adjustment. On the other hand, when the component series have similar seasonal patterns, then summing the series may result in noise*

> *cancelation, and the direct seasonal adjustment is usually of better quality than the indirect adjustment*" (Hood and Findley, 2003, p10).

Diagnostics of the two methods which measure the adequacy and quality of the adjustments can be calculated. The diagnostics include the spectral diagnostic for residual seasonality, and stability diagnostics such as sliding spans and revision history measures which test the stability of the estimates as new data are introduced. These diagnostics are produced by the X12-ARIMA program and it is suggested that they should be determined for the aggregate series as well as for the sub-series.

It is not an easy task to compare direct and indirect adjustment, due to the fact that the methods can produce different estimates. When looking for smoothness, and if the series are adjusted additively, Hood and Findley (2003) suggest looking at the differences of the seasonally adjusted aggregate series obtained with the direct and indirect methods. Alternatively, if the series are multiplicative, the ratio (rather than the difference) is appropriate.

Ladiray and Mazzi (2003, p40) state that indirect adjustment should be preferred to direct adjustment if the

> "*sub-components do not have similar characteristics or if the relative importance of the sub-series (in terms of weight) is changing very fast.*"

They clarify the idea of similarity of sub-series by saying that the direct approach could be more suited to 'horizontal' or geographical aggregation (e.g. by country) and the indirect approach to 'vertical' or sectorial aggregation, such as by sector, branch or product. In addition to the diagnostic measures described by Hood and Findley (2003), Ladiray and Mazzi (2003, p40) computed two measures called the mean and maximum of the Absolute Percentage Deviation; these both measure the relative difference between the direct and indirect adjusted series. They also considered the degree of consistency in growth rate as the growth rates should have the same sign and thus basically convey the same message.

Several different criteria are used in assessing the quality and adequacy of direct and indirect seasonal adjustment when using a filter-based method such as X12-

ARIMA (see Hood and Findley, 2003; Ladiray and Mazzi, 2003; and Otranto and Triacca, 2002). In essence, the indirect adjustment is favoured when the sub-series have different characteristics and direct adjustment is favoured when the sub-series are similar.

A comparison of three seasonal adjustment methods with respect to both temporal and sectoral aggregation is discussed in Geweke (1978). Using spectral densities, Geweke calculated the mean squared error (MSE) using the multivariate, indirect and direct methods of seasonal adjustment for stationary series. He then compared the MSEs of the different methods for several time series. His example featuring data on housing starts is particularly relevant. In this example, Geweke studied housing starts for four regions of the United States. He noted that the variances were largely different for the four series. He also noted that the correlation among the seasonal factors was higher than the correlation among the non-seasonal factors. The results of the comparison showed that the MSE achieved with the multivariate method was approximately half that of the other two approaches.

For model-based seasonal adjustment, Geweke (1978) concluded that the covariance structure between the series is crucial. He found that the joint ARIMA model was advantageous, as summarised by Taylor, when the sub-series are

> "*very heterogeneous, or where the stochastic structure of the non-seasonal and seasonal components are dissimilar*". (Taylor, 1978, p432)

The 'optimal procedure' referred to here is the joint or multivariate method. On the other hand, if the individual series are homogeneous, the efficiency gains are relatively small. The homogeneous model is is referred to in Section 3.3.1 and discussed in more detail in Section 5.6.1.

Planas and Campolongo (2001) used ARIMA models to confirm and extend the results in Geweke (1978). They studied the seasonal adjustment of contemporaneously aggregated series and compared the relative accuracy of the direct method with the indirect and multivariate methods. They confirmed Geweke's result that when the stochastic properties of the two series are even slightly dissimilar, the indirect adjustment is more precise than the direct adjustment. To compare the accuracy of

the seasonally adjusted series, Planas and Campolongo (2001) calculated the final estimation error of the non-seasonal component using the Wiener-Kolmogorov filter. The variance of the final estimation error was calculated following the procedure by Ghysels (1997) for the indirect case and by Geweke (1978) for the multivariate case. The multivariate adjustment was found to be the most accurate estimation in terms of the final estimation error.

Also using ARIMA model-based methodology, Maravall (2006), discusses the direct and indirect seasonal adjustment of an aggregate series with reference to an example: the exports, imports and balance of trade Japanese series. Using the TRAMO-SEATS program, he obtains the seasonally adjusted series of the balance of trade series with direct and indirect adjustment. No residual seasonality is found in either of the seasonally adjusted series produced. In the case of the computed revisions in the two seasonally adjusted series, the means are close to zero, but the RMSE for the indirect adjustment is found to be higher than for the direct adjustment. Maravall (2006) also shows that a smoother seasonally adjusted series and trend-cycle series is obtained with the direct adjustment. With these results, the conclusion points towards preference for direct seasonal adjustment at any level of aggregation:

> "*because aggregation modifies the dynamic structure of the series, and because seasonal adjustment is a non-linear transformation of the original series, aggregation constraints between the series cannot be expected to be preserved*". (Maravall, 2006, p2189)

Revision error is the error associated with the degree of revision a component estimate undergoes when new observations become available. Total revision errors of the three methods which had not been considered by Ghysels (1997) or Geweke (1978) are reviewed by Planas and Campolongo (2001). The indirect and direct methods performed better in many of the test cases than the multivariate method, thus giving no optimality if revision errors were considered (Planas and Campolongo, 2001). The estimation with the multivariate ARIMA model was noted as difficult to implement due to its complexity.

Studies described above by both Geweke (1978) and Planas and Campolongo (2001) cover the direct, indirect and multivariate seasonally adjusted series using ARIMA models. The work presented in this thesis focuses on the univariate and multivariate basic structural models and their comparison in relation to the variance of the seasonally adjusted series.

2.5 Applications of the Multivariate Basic Structural Model

Due to the flexibility of the basic structural model and its state space form, multiple time series can be modelled jointly with little difficulty. Extending this idea, a target series can be modelled jointly with one or more related series in order to obtain better estimates of the time series components of the target series.

Harvey and Chung (2000) calculated the filtered estimates in a bivariate BSM model and discussed the improvement in the root mean squared error (RMSE) of the slope component over that obtained from just using the univariate model. He found that the gains achieved in the estimation of the slope component using the bivariate model came primarily from the high correlation between the slopes of the two series.

The U.S. Bureau of Labor Statistics (BLS) applied state space models in estimating monthly employment and unemployment estimates for each of the 50 states and the District of Columbia. The models were fitted to the direct sample estimates obtained from the Current Population Survey (CPS) (Pfeffermann and Tiller, 2003). A filtering algorithm is developed for state space models with correlated measurement errors. The sampling errors are included as part of the observation (measurement) equation instead of the current practice of including them in the state vector. The states are grouped into a number of homogeneous groups (those States with similar labour force behaviour) and then the model is fitted jointly with added constraints. The empirical illustration fits the BLS model to the direct CPS unemployment estimators in 9 census divisions of the USA with the constraint that the Census divisions unemployment are benchmarked to the total national unemployment. Thus, use of

a multivariate state space model with added constraints leads to estimates which borrow strength from both past and disaggregated data (Pfeffermann and Tiller, 2003, p3).

Similarly, Marshall (1992) applies the multivariate BSM to a cross-section of time series. He investigates the relative efficiency of the filtered estimates as a function of both time and the number of units (or series). He concentrates on estimation of the time-dependent means by using the Kalman filter applied to the multivariate local level model, and reports that there are gains in using all the series to do so.

A specific focus of this thesis is in establishing whether there are gains in the estimation of the seasonal component of the total series by jointly modelling the sub-series. This topic extends the work carried out by Marshall through focussing on the seasonal component and simultaneously using the 'related series' idea of Harvey. In this case, the sub-series are treated as related series and the aggregate series is considered the target series.

The component estimates included in the state vector which apply to the sub-series are not of direct interest here but are calculated in order to obtain a better estimate of the structural components of the total series. It will be shown that under certain conditions, this procedure will enable the seasonal component estimates to be calculated with greater accuracy than if the total series was modelled by an univariate approach.

2.6 Short Time Series

A time series is referred to here as short if there are 5 or fewer years of data points available. An example is a time series with five years of monthly data, that is, 60 data points. Under a filter based system, seasonal adjustment requires at least three, but preferably five, years of data, due to constraints implied by the method of calculation of the moving averages, as described in Section 2.1.1.

Mir and Rondonotti (2003) studied the performance of the X-12-ARIMA procedure in the seasonal adjustment of short time series. They applied Monte Carlo

simulation techniques to compare the adjustment of a five year series (a short series) to one which has a fifteen year history (a long series). The first two years of the adjusted short time series were found to be 'seriously distorted'. This occurred especially when the original series had a highly variable seasonal pattern. However, when they compared the last two years of the adjusted short series to the adjusted long series, the former was 'only slightly inferior' to the latter.

When a survey undergoes a major innovation, the new and old surveys are usually both administered for an overlapping period of typically one to two years. Information regarding the effects of the change are required, however the new survey does not have the required number of time points to apply typical seasonal adjustment procedures. Gatto (2006) proposes a method of using the estimates of the components of the old survey to obtain estimates of the new series components. He allows each component in the new series to be a function of the estimated component of the old series. For the old survey, which has 12 years of quarterly data, he utilises the TRAMO-SEATS program, which is based upon a model-based canonical decomposition of the series to estimate the components. By minimising the sum of squared estimates for the irregular component, he obtains estimates of the coefficients linking the old and new survey components. By projecting the coefficients estimated in the overlapping period onto the components of the old survey over the historical data, a 'new' series is produced for the past data. This projected series is consistent with the new survey results and is long enough for seasonal adjustment procedures to be applied.

Mazzi and Savio (2003) compared the quality of the performance of two widely used programs for seasonal adjustment: TRAMO-SEATS (model-based) and X-12-ARIMA (filter-based) for different length time series. They apply the methods to short series (5 years of data), medium series (10 years of data) and long length series (20 years of data). The main finding is that the quality of seasonal adjustment reduces as the length of the series shortens. A variety of measures are calculated and compared given the output from the two programs. The measures of quality assessment are adapted from Ladiray and Mazzi (2003). More specifically, 'the

deterioration is found to be proportionally greater passing from long to medium, than from medium to short time series'. The results imply that greater instabilities exist in the seasonal adjustment process at the start of the series, than at the end of the series (Mazzi and Savio, 2003, p9). Further, such instabilities seem to be greater for the model-based approach (TRAMO-SEATS).

2.6.1 Short Time Series and Longitudinal Analysis

Another area of statistics which sometimes includes short series is the analysis of longitudinal data. In a study conducted by Feder *et al.* (2000), the data consist of longitudinal observations which are repeated measurements on the same units over a number of occasions with fixed or varying time periods between the occasions. The observations have a hierarchical structure in which individuals (first level) are within households (second level). Thus, each vector observation may be treated as a time series, usually of short length.

In order to model both the hierarchical and longitudinal nature of the data they propose 'a time series multi-level model which combines separate cross-sectional two-level models by modelling the evolution of the first and second level random effects over time' (Feder *et al.*, 2000, p57). The multi-level model may be written in state space form and the time series relationships can be proposed for the vector of coefficients and the random effects. An AR(1) relationship is assumed for both the first and second level random effects. Given the large number of parameters to be estimated, they propose a two-stage estimation process:

1. multi-level modelling estimation for the cross-sectional parameters; and

2. state space model estimation for the time series parameters.

The first stage involves fitting a multi-level model to each series separately to obtain iterative generalised least squares (IGLS) estimates of the time dependent fixed effects and the variances of the random effects. The second stage involves holding fixed the parameters estimated in first stage, and the remaining parameters of the combined model (three AR coefficients and corresponding residual variances)

are estimated by maximum likelihood estimation. The application of the Kalman filter is carried out after setting the required initial values.

As noted by Feder *et al.* (2000, p60), 'The Kalman filter is initialized by the unconditional means and variances of the random effects under the model, but at time $t = 1$ the moments holding for units in the sample can be different because of the sampling effects. For long enough series and under some regularity conditions, the estimates derived from the maximization of the likelihood are not sensitive to the initialization of the procedure. However, with short series, improper initialization under informative sampling could distort the estimation process'.

In a simulation study, Feder *et al.* (2000, p61) demonstrated that 'under non-informative sampling it is possible to successfully fit simple but non-trivial time series models to very short longitudinal series, provided that the number of observed series is sufficiently large'.

Thus, previous research shows that in appropriate circumstances, the Kalman filter may be applied to short series. In Chapter 6 of this thesis, the accuracy of the estimated time series components will be examined for different series lengths.

2.7 Modelling Survey Errors within a State Space Model Context

When analysing time series that have been constructed from panel estimates, it is important to account for the autocorrelations of the survey errors. Pfeffermann *et al.* (1998) developed a method of estimating the survey error autocorrelations (SEA) based on the raw panel estimates. These SEA estimates were then used 'to construct simple state space models that allow the separation of the trend of the population values from the spurious trend induced by the movement of the survey errors' (Pfeffermann *et al.*, 1998, p340). That is, by estimating the SEAs, researchers were able to identify and estimate time series models for the survey errors. Illustrations include both Australian Labour Force Survey and U.S. Labor Force Series (Pfeffermann *et al.*, 1998). An AR(2) model provided a good approximation for the survey errors in the Australian data. Fitting of the model to the observed series

of the direct survey estimates was carried out in two stages. This was necessary to deal with the identifiability problem caused by the stochastic structures within the population trend and the long-term behaviour of the survey errors. The two stages are:

1. Estimate the survey error model parameters by:

   - using the variance and autocorrelations computed from the distinct panel estimates;
   - solving the Yule Walker equations to obtain the estimators of the AR(2) coefficients and the residual variance.

2. Estimate the population model parameters by maximum likelihood with survey error model parameters held fixed at their estimated values. The Kalman filter was used to calculate the innovations and their variances. The OPTMUM routine in the GAUSS software was used.

Feder (2001) gives an example of using a multivariate BSM to estimate the number of households in Canada. The survey errors were estimated by an AR(3) process and the parameters were included in the state vector. The BSM includes a bias to allow for benchmarking of the filtered estimates on the Census figures. Using maximum-likelihood estimation, the autocorrelations of the survey errors were estimated by including them as hyper-parameters.

2.8 Intervention Analysis

Intervention analysis models the effect of a dynamic change on a time series at a known point in time (Pena *et al.*, 2001). The cause of the change may be a particular event or a result of a policy change on a time series (Harvey, 1989). It can be modelled by a dummy explanatory variable.

When there is a break in the series, or a change in the methodology used to produce a series, it may be viable to treat the series for the old and new surveys as

one continuous series. If τ is the time point at which the new survey begins, then intervention analysis could be used to determine the effect, if any, on the trend and seasonal factors after time τ. This analysis could be done within a basic structural model (BSM) framework.

Harvey and Durbin (1986) modelled the effect of seat belt legislation introduced in Britain on January 31, 1983 for front seat occupants of cars and light goods vehicles. They used intervention analysis based on a structural time series model to estimate the changes in casualty rates for different categories of road users following the introduction of the legislation. Within the BSM they used the trigonometric approach (described in Appendix B) to modelling seasonality and incorporated an intervention variable to model the change in the level of the series after the introduction of the law. They comment on the flexibility of the structural model, which allows a considerable amount of complexity to be accommodated within the model, such as explanatory and intervention variables. The model is written in state space form and the Kalman filter is applied for estimation of the time series components, the explanatory and intervention variables.

Durbin and Koopman (2001) also investigate the use of the structural model with intervention variables. They extend the study to a bivariate model for the front seat passengers and rear seat passengers killed and seriously injured. The rear seat series is used as the related series but in these circumstances it is also used as a control group. The bivariate model results in a more precise measure of the effect of the seat belt law on front seat passengers. In fact, the results show that the root mean squared error is almost halved for the estimated intervention coefficient for the level component of the front seat series.

A multivariate structural time series model with intervention analysis is also considered, among more traditional methods, by Sridharan *et al.* (2003). They examine the impact of new legislation regarding sentencing of felony offenders on reported crime rates committed on or after January 1, 1995. The multivariate BSM allows simultaneous consideration of a set of time series where some series are considered as treatment series and others as control series. When correlations between the

components are high, the components of the target series will be estimated with the combined use of the other control series rather than just using the univariate target series individually. Sridharan *et al.* (2003) found that this procedure leads to a more effective intervention analysis than other more traditional methods such as ARIMA models and regression models with correlated errors. It is also noteworthy that the data only included four years of post-intervention observations and that the intervention variable was modelled as a step variable to account for a level shift.

The work described above discusses intervention analysis which allows for level shifts in one or more time series. When there is a change in the survey methodology, (such as a change in the sample design or a change in the questionnaire design) or a break in a series, it is not only the trend of the series that may be affected but also the seasonal pattern. Penzer (2006) proposes a method of detecting seasonal shifts within the structural time series approach. He formulates diagnostic statistics which can be generated from the output of the smoothing algorithm associated with the Kalman filter. The flexibility and transparency of the basic structural model allows intervention variables to be added into the model. Within the ARIMA model-based approach, Kaiser and Maravall (2003) discuss different specifications for a seasonal outlier particulary with respect to the automatic procedures of outlier detection (such as in the TRAMO program). They propose an extended procedure which adds the seasonal level shift outlier to the standard method. The seasonal level shift, which can affect both the trend and the seasonal component, is discussed in detail through simulation and real data examples.

2.9 Conclusion

Seasonal adjustment may be performed by filter-based methods or model-based methods as described in this chapter. For an aggregate series, the adjustment can be made on the aggregate series directly, or it can be be done indirectly using the sub-series. Although each of the sub-series is utilised in the indirect approach, each series is seasonally adjusted individually rather than jointly. A multivariate

approach to seasonal adjustment is an alternative to both the direct and indirect methods. Limited research has been done in this area, and mostly with ARIMA models.

Previous studies have shown that the multivariate basic structural model has been effectively used to achieve gains in the estimation of series components of a target series. There is scope for additional research in the area of joint modelling of time series for the purpose of seasonal adjustment of an aggregated series. The next chapter describes the univariate and multivariate basic structural models and proposes a methodology for estimating time series components of an aggregated series.

Chapter 3

Modelling Disaggregated Series

3.1 Introduction

A structural time series model allows time series characteristics such as trend, seasonal and error components to be modelled specifically. The basic structural model will therefore be the model of choice for this thesis.

The series of observations of the aggregated series, $Y_1, \ldots, Y_T$, will be modelled by an univariate additive basic structural model (BSM). If the aggregated series is a sum of K sub-series, for $t = 1 \ldots T$, then,

$$Y_t = \sum_{k=1}^{K} Y_{kt} \tag{3.1}$$

and $Y_{1t}, \ldots, Y_{Kt}$, may be modelled jointly with a multivariate BSM.

This chapter details the models to be applied to the univariate aggregated (or total) series (Section 3.2) and the system of sub-series underlying the aggregated series (Section 3.3). The univariate and multivariate models are then converted into state space form (Section 3.4). Applying a transformation to the multivariate model of the sub-series, the series components for the aggregate series may be determined directly from the multivariate model (Section 3.4.3). The Kalman filter, which yields the estimates of the series components and their mean squared errors, is then detailed for each approach (Section 3.5). A measure of comparison of the mean squared errors of the seasonal component given by the two approaches is then proposed (Section

3.6).

3.2 Univariate BSM

For a single additive time series, the observations at time t, denoted by Y_t, may be written as the sum of a local linear trend, L_t, a seasonal component, S_t, and an irregular or disturbance term, $\varepsilon_{U,\,t}$. This basic structural model may be written in the notation adopted by Feder (2001), for $t = 1, \ldots, T$ as

$$Y_t = L_t + S_t + \varepsilon_{U,\,t}, \qquad \varepsilon_{U,\,t} \sim N(0, \sigma^2_{U,\,\varepsilon}), \tag{3.2}$$

where the U subscript in the serially independent $\varepsilon_{U,\,t}$ denotes the univariate model.

3.2.1 The Trend Component

The trend may be assumed to evolve stochastically over time, and may or may not include a slope term. If a slope term is included then the local linear trend may be written as

$$L_{t+1} = L_t + R_t + \eta_{U,\,t}, \qquad \eta_{U,\,t} \sim N(0, \sigma^2_{U,\,\eta}), \tag{3.3}$$

where L_t is the level of the series at time t and R_t is the local rate of change or slope of the series. The slope may evolve stochastically over time as a random walk:

$$R_{t+1} = R_t + \zeta_{U,\,t}, \qquad \zeta_{U,\,t} \sim N(0, \sigma^2_{U,\,\zeta}). \tag{3.4}$$

If a slope term is not included then the local level trend may be written as

$$L_{t+1} = L_t + \eta_{U,\,t}, \qquad \eta_{U,\,t} \sim N(0, \sigma^2_{U,\,\eta}). \tag{3.5}$$

The random errors in the univariate trend component, $\eta_{U,\,t}$ and $\zeta_{U,\,t}$, are mutually and serially independent.

For a comprehensive review of stochastic linear trends, which compares some of the trend models most often used in practice such as in the ARIMA model-based approach, the STS approach and the X11 procedure, see Maravall (1993).

3.2.2 The Seasonal Component

There are several ways of modelling the seasonal component, S_t, in (3.2). Let the number of seasons in the year be denoted by s. Thus, for quarterly data, $s = 4$, and for monthly data, $s = 12$. If the seasonal pattern is constant over time, the seasonal effects can be modelled by the constants $S_1^*, \ldots, S_s^*$ such that $\sum_{j=1}^{s} S_j^* = 0$. So the seasonal component for the jth season in year i is $S_t = S_j^*$ where $t = s(i-1) + j$ for $i = 1, 2, \ldots$ and $j = 1, \ldots, s$. The constraint to add to zero over s seasons can be written as $\sum_{j=1}^{s} S_j = 0$, or over any s time periods, $\sum_{j=0}^{s-1} S_{t+1-j} = 0$ with $t = s-1, s, \ldots$ (Durbin and Koopman, 2001, Section 3.2).

If the seasonal effects are allowed to change stochastically over time then a disturbance may be introduced such that

$$\sum_{j=0}^{s-1} S_{t+1-j} = \omega_{U,\,t}, \qquad \text{or} \qquad S_{t+1} = -\sum_{j=1}^{s-1} S_{t+1-j} + \omega_{U,\,t}, \tag{3.6}$$

where $\omega_{U,\,t}$ are independent over time and have distribution $\omega_{U,\,t} \sim N(0, \sigma^2_{U,\,\omega})$. Since the disturbance term has an expectation of zero, (3.6) still allows the expected value of the sum of the seasonal effects to be zero over any s time periods. The model for the seasonal component described in (3.6), is called the dummy seasonal model.

Alternatively, the seasonal component may be modelled by a set of trigonometric terms at the seasonal frequencies. This is described in detail in Appendix B. Other ways of expressing the seasonal component have been suggested. One method extends the dummy seasonal model (3.6) by letting each season evolve as a random walk, see Harrison and Stevens (1976).

Proietti (2000) looks at four different specifications of seasonal models, including the dummy seasonal and trigonometric seasonal models. He concludes that their performance depends particularly on the smoothness of the seasonal pattern. When the time series has seasonality which changes slowly, the dummy seasonal model is not preferred. The trigonometric specification allows for smoother changes in the seasonal components (Harvey and Scott, 1994). However, in one example in Proietti (2000), the dummy seasonal specification is found to be slightly superior in terms of forecasting performance. None of the four models studied could be viewed as the

optimal seasonal model. For a comprehensive review of modelling seasonal effects, see Proietti (2000).

In this thesis, the dummy seasonal model will be the specification used in the BSM due to its analytical simplicity. Part of the study in Chapter 4 is repeated by replacing the dummy seasonal model with trigonometric seasonal model. The results for the relative efficiency are almost indistinguishable (see Appendix B).

3.2.3 The Disturbance Terms

The disturbance terms $\eta_{U,\,t}, \zeta_{U,\,t}, \omega_{U,\,t}$ and $\varepsilon_{U,\,t}$, are assumed to be serially and mutually independent. Their respective variances, $\{\sigma^2_{U,\,\eta}, \sigma^2_{U,\,\zeta}, \sigma^2_{U,\,\omega}, \sigma^2_{U,\,\varepsilon}\}$, are the parameters of the univariate model, sometimes referred to in the literature as the hyper-parameters.

A univariate BSM may be defined by combining a model for the trend and a model for the seasonal component. A BSM with a level, slope, and dummy seasonal model is defined by (3.2) - (3.4) and (3.6). It will be used to model the aggregated series, Y_t, given by (3.1).

3.3 Multivariate BSM

If an univariate time series is disaggregated such that the sum of the K sub-series is the aggregated (or total) series (3.1), then a multivariate BSM can be applied to the sub-series, $Y_{1t}, \ldots, Y_{Kt}$.

The observation for series k at time t, denoted by Y_{kt}, is modelled by the level component, L_{kt}, the slope component, R_{kt}, and the seasonal component, S_{kt}, as described by

$$Y_{kt} = L_{kt} + S_{kt} + \varepsilon_{kt} \tag{3.7}$$

where

$$\begin{aligned} L_{k,\,t+1} &= L_{kt} + R_{kt} + \eta_{kt} \\ R_{k,\,t+1} &= R_{kt} + \zeta_{kt} \\ S_{k,\,t+1} &= -\sum_{j=1}^{s-1} S_{k,\,t+1-j} + \omega_{kt} \\ k &= 1, 2, \cdots, K \quad \text{and} \quad t = 1, 2, \cdots, T. \end{aligned}$$

The disturbance terms ε_{kt}, η_{kt}, ζ_{kt} and ω_{kt} are assumed to be serially and mutually independent. The disturbance terms for the same component, but for different series, may be correlated; for example, it may be that $\mathrm{Cov}(\eta_{1t}, \eta_{2t})$ may not be zero. The distributions of the disturbance terms are given by

$$\begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Kt} \end{pmatrix} \sim N(0, \mathbf{\Sigma}_\varepsilon), \quad \begin{pmatrix} \eta_{1t} \\ \vdots \\ \eta_{Kt} \end{pmatrix} \sim N(0, \mathbf{\Sigma}_\eta), \tag{3.8}$$

$$\begin{pmatrix} \zeta_{1t} \\ \vdots \\ \zeta_{Kt} \end{pmatrix} \sim N(0, \mathbf{\Sigma}_\zeta), \qquad \text{and} \qquad \begin{pmatrix} \omega_{1t} \\ \vdots \\ \omega_{Kt} \end{pmatrix} \sim N(0, \mathbf{\Sigma}_\omega). \tag{3.9}$$

Each covariance matrix will have dimension $K \times K$. If no restrictions are placed on the structure of each covariance matrix, apart from being positive definite, then for each matrix there are $K(K+1)/2$ parameters corresponding to K parameters for each of K sub-series and $K(K-1)/2$ covariance terms. Hence, if there are four components in the model (level, slope, seasonal, and error), then there will be, in total, $2K(K+1)$ parameters which may need to be estimated. Obviously, as K becomes large, the estimation process could become difficult due to the number of parameters to be estimated. It is possible to reduce the number of parameters by placing some restrictions on the structure of each covariance matrix.

Model (3.7) is referred to by Harvey (1989) as a system of 'seemingly unrelated time series equations' or a SUTSE model. They consist of sub-series which may be exposed to the same overall economic environment but not subject to any

cause-and-effect relationships. These sub-series may be linked via correlations of the disturbances driving the components of the series (level, slope, seasonal and error). This is referred to as 'contemporaneous correlation' (Harvey, 1989, p429).

3.3.1 Placing Restrictions on the Variance Structure

A multivariate BSM may be simplified by reducing the number of parameters to be estimated. This can be achieved by restricting the structure of some or all of the covariance matrices. Specific restrictions and their properties have been discussed in the literature. One restriction leads to the homogeneous SUTSE model, in which the covariance matrices are proportional to each other (see Harvey, 1989; Fernandez and Harvey, 1990). However, for a homogeneous model, there is no gain in using the sub-series to improve estimates of an aggregated series. In this case, 'the sum of the individually adjusted series is equal to the adjusted aggregate' (Harvey, 1989, Section p437). The properties of a homogeneous model are therefore not beneficial to the aim of this project. Other ways of restricting the structure of the covariance matrices will be examined.

Many areas of statistics employ the concept of random error terms which have been broken into more specific random components such as common and unit-specific components. These include longitudinal data analysis, multi-level data analysis, mixed effects models, and variance components models. More specific to the problem at hand, however, is the work carried out by Marshall (1992), who looks at cross-sections of time series. Within the framework of a BSM, Marshall (1992) decomposes the disturbance terms and relates these to the random error terms in a dynamic error components model (see Marshall, 1992; Harvey and Shephard, 1993). Marshall uses the example of a multivariate local level model:

$$\begin{aligned} Y_{kt} &= L_{kt} + \varepsilon_{kt} \\ L_{k,\,t+1} &= L_{kt} + \eta_{kt}. \end{aligned} \tag{3.10}$$

In this model, there are $K(K+1)/2$ parameters in the covariance matrix for $(\varepsilon_{1t}, \ldots, \varepsilon_{Kt})'$ and the covariance matrix for $(\eta_{1t}, \ldots, \eta_{Kt})'$. This gives a total of

$K(K+1)$ parameters.

Marshall (1992) then decomposes the disturbance terms into common effects (ε_t and η_t), which are time specific, and time-unit specific effects (ε^*_{kt} and η^*_{kt}). Model (3.10) is then rewritten as:

$$\begin{aligned} Y_{kt} &= L_{kt} + \varepsilon_t + \varepsilon^*_{kt} \\ L_{k,\,t+1} &= L_{kt} + \eta_t + \eta^*_{kt} \end{aligned} \tag{3.11}$$

where ε_t, ε^*_{kt}, η_t, and η^*_{kt} are assumed to be independent Normal random variables with zero mean and variances σ^2_{ε}, $\sigma^2_{\varepsilon^*}$, σ^2_{η}, $\sigma^2_{\eta^*}$ respectively. Therefore, the resulting two covariance matrices have a compound symmetry structure given by:

$$\boldsymbol{\Sigma}_{\varepsilon} = \sigma^2_{\varepsilon}\mathbf{J}_K + \sigma^2_{\varepsilon^*}\mathbf{I}_K \qquad \text{and} \qquad \boldsymbol{\Sigma}_{\eta} = \sigma^2_{\eta}\mathbf{J}_K + \sigma^2_{\eta^*}\mathbf{I}_K \tag{3.12}$$

where $\mathbf{I}_K$ is the $K \times K$ identity matrix and $\mathbf{J}_K$ is the $K \times K$ matrix of all ones. The number of parameters has therefore reduced from $K(K+1)$ as given in (3.10) to just four parameters, namely σ^2_{ε}, $\sigma^2_{\varepsilon^*}$, σ^2_{η}, $\sigma^2_{\eta^*}$.

Alternatively, by allowing the time unit-specific effects, ε^*_{kt} and η^*_{kt} to have unit-specific variances, $\sigma^2_{k\varepsilon^*}$ and $\sigma^2_{k\eta^*}$ respectively, the term including the identity matrix is replaced by a diagonal matrix $\mathbf{D}$:

$$\boldsymbol{\Sigma}_{\varepsilon} = \sigma^2_{\varepsilon}\mathbf{J}_K + \mathbf{D}_{\varepsilon^*} \qquad \text{and} \qquad \boldsymbol{\Sigma}_{\eta} = \sigma^2_{\eta}\mathbf{J}_K + \mathbf{D}_{\eta^*}, \tag{3.13}$$

where

$$\mathbf{D}_{\varepsilon^*} = \operatorname{diag}\left[\sigma^2_{1\varepsilon^*}, \ldots, \sigma^2_{K\varepsilon^*}\right] \qquad \text{and} \qquad \mathbf{D}_{\eta^*} = \operatorname{diag}\left[\sigma^2_{1\eta^*}, \ldots, \sigma^2_{K\eta^*}\right]. \tag{3.14}$$

Thus, there are $(K+1)$ parameters in each covariance matrix, giving $2(K+1)$ parameters to be estimated in a local level model with this structure.

The above two alternative covariance structures for the multivariate local level model may be easily extended to the multivariate BSM. The model for the observation, Y_{kt}, at time t and for series k, is given below with $k = 1, 2, \ldots, K$ representing

the K sub-series with dummy seasonal components:

$$\begin{aligned} Y_{kt} &= L_{kt} + S_{kt} + \varepsilon_t + \varepsilon^*_{kt}, \\ L_{k,\,t+1} &= L_{kt} + R_{kt} + \eta_t + \eta^*_{kt}, \\ R_{k,\,t+1} &= R_{kt} + \zeta_t + \zeta^*_{kt}, \\ S_{k,\,t+1} &= -\sum_{j=1}^{s-1} S_{k,\,t+1-j} + \omega_t + \omega^*_{kt}. \end{aligned} \tag{3.15}$$

The disturbance terms, $\varepsilon_t,\ \varepsilon^*_{kt},\ \eta_t,\ \eta^*_{kt},\ \zeta_t,\ \zeta^*_{kt},\ \omega_t,\ \omega^*_{kt}$ are assumed to be independent Normal random variables. The resulting four covariance matrices may have the following structure (Marshall, 1990):

$$\mathrm{Var}(x_t\mathbf{1}_K + \mathbf{x}^*_t) = \mathbf{\Sigma}_x = \sigma^2_x\mathbf{J}_K + \mathbf{D}_{x^*}, \tag{3.16}$$

where

$$\begin{aligned} x &\text{ stands for } \ \eta,\ \zeta,\ \omega, \text{ or } \varepsilon, \\ \mathbf{x}^*_t &\text{ stands for } (\eta^*_{1t},\ldots,\eta^*_{Kt})',\ (\zeta^*_{1t},\ldots,\zeta^*_{Kt})',\ (\omega^*_{1t},\ldots,\omega^*_{Kt})', \\ &\qquad \text{or } \ (\varepsilon^*_{1t},\ldots,\varepsilon^*_{Kt})' \end{aligned}$$

$$\begin{aligned} \mathbf{D}_{x^*} &= \sigma^2_{x^*}\mathbf{I}_K \quad \text{or} \quad \mathbf{D}_{x^*} = \mathrm{diag}\big[\sigma^2_{1x^*},\ldots,\sigma^2_{Kx^*}\big], \\ \mathbf{1}_K &\text{ is a } K \text{ dimensional vector of one's,} \\ \mathbf{I}_K &\text{ is a } K\times K \text{ identity matrix,} \\ \mathbf{J}_K &= \mathbf{1}_K\mathbf{1}_K^{'},\ (\text{a } K\times K \text{ matrix of all ones}). \end{aligned} \tag{3.17}$$

The first structure for the covariance matrix $\mathbf{D}_{x^*}$ proposed above in (3.17) will be referred to as *Model 1* and the second will be referred to as *Model 2*.

Model 1: Compound Symmetry

The first covariance structure, $\mathbf{D}_{x^*} = \sigma^2_{x^*}\mathbf{I}_K$, described above for the unit-specific disturbances, is just a multiple of the identity matrix, and has only one unknown parameter. Therefore, together with the common variance, σ^2_x, there are only two

parameters for each component (level, slope, seasonal, irregular) included in the model. The full BSM model would therefore have eight unknown parameters.

For example, if $K = 2$, the covariance matrix for the level component would be

$$\mathbf{\Sigma}_\eta = \begin{pmatrix} \sigma_\eta^2 + \sigma_{\eta^*}^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{\eta^*}^2 \end{pmatrix}. \tag{3.18}$$

If the correlation between the individual series for the level component is denoted by ρ_η, then, for the compound symmetry model, it is the same for each pair of series:

$$\rho_\eta = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_{\eta^*}^2}. \tag{3.19}$$

Similarly for ρ_ζ , ρ_ω and ρ_ε.

Since the aggregated series is the sum of the K sub-series (3.1), the following constraint can be applied to the variance of the level component for the aggregated series

$$\begin{aligned} \sigma_{tot,\eta}^2 &= \mathbf{1}_K^{'} \mathbf{\Sigma}_\eta \mathbf{1}_K \\ &= K^2 \sigma_\eta^2 + K \sigma_{\eta^*}^2. \end{aligned} \tag{3.20}$$

Application of this constraint will be similar for the other components (slope, seasonal and irregular) in Model 1, giving:

$$\begin{aligned} \sigma_{tot,\zeta}^2 &= K^2 \sigma_\zeta^2 + K \sigma_{\zeta^*}^2, \\ \sigma_{tot,\omega}^2 &= K^2 \sigma_\omega^2 + K \sigma_{\omega^*}^2, \\ \sigma_{tot,\varepsilon}^2 &= K^2 \sigma_\varepsilon^2 + K \sigma_{\varepsilon^*}^2. \end{aligned} \tag{3.21}$$

These constraints will be used when setting parameter values for the purpose of simulating the total series and sub-series for Model 1 in Chapter 4.

Model 2: Series-specific Variances

For Model 2, the K unit-specific variances are the variances which are specific to the K sub-series and will therefore be termed 'series-specific'. These are allowed

to differ as given by the definition for Model 2. The alternative structure for $\mathbf{D}_x^*$, has K different values on the diagonal, and hence would have $(K+1)$ unknown parameters for each of the 4 component covariance matrices. Thus, in total there would be $4(K+1)$ unknown parameters for the model.

For example, if $K = 2$, then the covariance matrix for the level component would be

$$\mathbf{\Sigma}_\eta = \begin{pmatrix} \sigma_\eta^2 + \sigma_{1\eta^*}^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{2\eta^*}^2 \end{pmatrix}. \tag{3.22}$$

If the correlation between series $k = 1$ and series $k = 2$ for the level component is denoted by $\rho_{1,\,2,\,\eta}$, then

$$\rho_{1,\,2,\,\eta} = \frac{\sigma_\eta^2}{\sqrt{(\sigma_\eta^2 + \sigma_{1,\,\eta^*}^2)(\sigma_\eta^2 + \sigma_{2,\,\eta^*}^2)}}. \tag{3.23}$$

Similarly for ρ_ζ , ρ_ω and ρ_ε.

The constraint for the level component for Model 2 would take the form

$$\sigma_{tot,\,\eta}^2 = K^2\sigma_\eta^2 + \sum_{k=1}^{K} \sigma_{k\eta^*}^2, \tag{3.24}$$

and similarly for $\sigma_{tot,\,\zeta}^2$, $\sigma_{tot,\,\omega}^2$, and $\sigma_{tot,\,\varepsilon}^2$. These constraints will be used when setting parameter values for the purpose of simulating the total series and sub-series for Model 2 in Chapter 4. When using the exact parameter values, the values of the parameters obtained from the multivariate model for the total series, namely $\sigma_{tot,\,\eta}^2$, $\sigma_{tot,\,\zeta}^2$, $\sigma_{tot,\,\omega}^2$ and $\sigma_{tot,\,\varepsilon}^2$ will be equal to the values of the parameters for the univariate model $\sigma_{U,\,\eta}^2$, $\sigma_{U,\,\zeta}^2$, $\sigma_{U,\,\omega}^2$ and $\sigma_{U,\,\varepsilon}^2$ respectively. When the parameters are estimated, this property will not necessarily hold.

Other Models

Other options for the variance structure include time-unit specific variances for one or more components with the remaining components keeping a compound symmetry structure. In another alternative structure, $\mathbf{D}_{x^*}$ is a $K \times K$ diagonal matrix but it

consists of fewer than K different variance components. This would correspond to having various groupings of sub-series where the variances for the sub-series within each group are equal. The number of unknown parameters would therefore be reduced, but would depend upon the number of groups. Examples of these alternative structures are discussed in Chapter 7.

3.4 State Space Form

Any basic structural model, whether univariate or multivariate, may be written more concisely in state space form (SSF). The Kalman filter and the Kalman smoother (see Harvey, 1989; Durbin and Koopman, 2001) may then be applied to the model to obtain estimates of the unobserved components, which may include the level, the slope and the seasonal components. The state space form for each approach is detailed in the following sections.

3.4.1 Univariate BSM in State Space Form

State space representation for the model described by (3.2) - (3.4) and (3.6), consists of a measurement (or observation) equation (3.25), and a transition (or state) equation (3.26). The measurement equation describes the observation at time t as a linear combination of the unobserved components included in the state vector, α_t. The transition equation describes the development of the state vector from one time point to the next. The state space model may be written (Durbin and Koopman, 2001, Section 3.1) as:

$$Y_t = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t}, \tag{3.25}$$

$$\alpha_{t+1} = \mathbf{T}\alpha_t + \mathbf{G}\gamma_t, \tag{3.26}$$

where, for quarterly data (s=4), and a dummy seasonal component,

$$\begin{aligned}
\alpha_t &= [L_t,\ R_t,\ S_t,\ S_{t-1},\ S_{t-2}]', & \alpha_1 &\sim \mathbf{N}(\mathbf{a}_1, \mathbf{P}_1),\\
\gamma_t &= [\eta_{U,\,t},\ \zeta_{U,\,t},\ \omega_{U,\,t}]', & \gamma_t &\sim \mathbf{N}(0, \mathbf{Q}),\\
& & \varepsilon_{U,\,t} &\sim \mathbf{N}(0, H),\\
\mathbf{Z} &= \begin{pmatrix} 1\ 0\ 1\ 0\ 0 \end{pmatrix}, & &
\end{aligned}$$

$$\mathbf{T} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{G} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathrm{Var}\,(\mathbf{G}\gamma_t) = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{U,\,\zeta} & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{U,\,\omega} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad H = \sigma^2_{U,\,\varepsilon}. \tag{3.27}$$

It is assumed that the initial state vector, α_1, has a mean and variance given by $\mathrm{E}(\alpha_1) = \mathbf{a}_1$, and $\mathrm{Var}(\alpha_1) = \mathbf{P}_1$.

In general, α_t is a $p \times 1$ vector, where p is the number of unobserved components to be estimated. Thus, a BSM with level, slope and a dummy seasonal component for quarterly data ($s = 4$), will have $p = 2+s-1 = 5$, as given above. With monthly data ($s = 12$), the same model will have $p = 2 + s - 1 = 13$. The vector, γ_t, has dimension $u \times 1$ and contains the disturbance terms which apply to the state vector. $\mathbf{Z}$ is a $1 \times p$ matrix, $\mathbf{T}$ is a $p \times p$ matrix and $\mathbf{G}$ is a $p \times u$ matrix. In the univariate model, H is a scalar.

3.4.2 Multivariate BSM in State Space Form

The multivariate BSM or SUTSE model (3.7) would usually be written in state space form in a similar way to the univariate state space form (refer to (3.25) and (3.26)), with the measurement errors separated from the state vector. This conventional format requires uncorrelated measurement errors. That is, the covariance matrix, $\mathbf{\Sigma}_\varepsilon$, is assumed to be diagonal. However, due to the common disturbance term, ε_t, the multivariate BSM contains correlated measurement errors, which cannot be handled by the standard Kalman filter or by standard software packages. To overcome this problem, Durbin and Koopman (2001, Section 6.4) suggest including the measurement errors in the state vector.

The state space form is amended to allow for these different dimensions for the multivariate system of sub-series $Y_{1t}, \ldots, Y_{Kt}$, as given below. The amended state space form has the (m) subscript to denote the multivariate model. The state vector is therefore denoted by $\alpha_{(m),\,t}$.

$$\mathbf{Y}_{(m),\,t} = (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),\,t}, \tag{3.28}$$

$$\alpha_{(m),\,t+1} = (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),\,t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),\,t}, \tag{3.29}$$

where $\otimes$ is the Kronecker product (for more detail refer to Appendix A). For quarterly data (s=4), and a dummy seasonal component for Model 2,

$$\begin{aligned}
\mathbf{Y}_{(m),\,t} &= [Y_{1t}, \ldots, Y_{Kt}]^{'}, \\
\alpha_{(m),\,t} &= [L_{1t}, \ldots, L_{Kt},\ R_{1t}, \ldots, R_{Kt},\ S_{1t}, \ldots, S_{Kt}, \\
&\qquad S_{1,\,t-1}, \ldots, S_{K,\,t-1},\ S_{1,\,t-2}, \ldots, S_{K,\,t-2}, \\
&\qquad (\varepsilon_t + \varepsilon^*_{1t}), \ldots, (\varepsilon_t + \varepsilon^*_{Kt})\ ]^{'}, \\
\gamma_{(m),\,t} &= [\ (\eta_t + \eta^*_{1t}), \ldots, (\eta_t + \eta^*_{Kt}),\ (\zeta_t + \zeta^*_{1t}), \ldots, (\zeta_t + \zeta^*_{Kt}), \\
&\qquad (\omega_t + \omega^*_{1t}), \ldots, (\omega_t + \omega^*_{Kt}),\ (\varepsilon_{t+1} + \varepsilon^*_{1,\,t+1}), \ldots, (\varepsilon_{t+1} + \varepsilon^*_{K,\,t+1})\ ]^{'}.
\end{aligned} \tag{3.30}$$

The system matrices are given by

$$\mathbf{Z}_{(m)} = \begin{pmatrix} 1\ 0\ 1\ 0\ 0\ 1 \end{pmatrix},$$

$$\mathbf{G}_{(m)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{T}_{(m)} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{3.31}$$

and hence the covariance matrix of the multivariate system is

$$\mathrm{Var}\left((\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),\,t}\right) = \begin{pmatrix} \boldsymbol{\Sigma}_\eta & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \boldsymbol{\Sigma}_\zeta & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \boldsymbol{\Sigma}_\omega & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \boldsymbol{\Sigma}_\varepsilon \end{pmatrix}, \tag{3.32}$$

where $\mathbf{0}_K$ represents a $K \times K$ matrix of zeroes.

Similarly to the state space form of the univariate model (3.27), the seasonal components contained in $\alpha_{(m),\,t}$ which have a subscript of $t-1$ or $t-2$ do not have disturbance terms associated with them, given the definition of the quarterly dummy seasonal variable. This means that, similarly to (3.27), the matrix (3.32) contains $2K$ rows and $2K$ columns which are zeroes. If $K = 2$, the covariance matrix for the level component, $\boldsymbol{\Sigma}_\eta$, is given by (3.22) for Model 2, and similarly for $\boldsymbol{\Sigma}_\omega$ and $\boldsymbol{\Sigma}_\varepsilon$.

By firstly writing the model in state space form, and letting x stand for $\eta, \zeta, \omega, \varepsilon$, the variance parameters σ_x^2, $\sigma_{1x^*}^2$, $\ldots, \sigma_{Kx^*}^2$, may be estimated by maximum likelihood estimation (Durbin and Koopman, 2001, Ch.7). These estimates would then be substituted into the Kalman filter equations to provide estimates of the unobserved components of the sub-series contained in the state vector $\alpha_{(m),\,t}$. However, the focus of this thesis is on the aggregated series and so it is desired to estimate

the model components for the aggregated series using the sub-series. This can be done more directly by applying a transformation to the above state space model.

3.4.3 Applying a Transformation

The aggregate (or total) series is the series of interest here rather than the individual sub-series. Therefore, the output from the analysis of the total series using both the univariate model and the multivariate model is required for comparison purposes. In standard software packages (such as in the `S-PLUS` module `S+FinMetrics`), only the diagonal elements of the mean squared error covariance matrix of the estimated components are available. Thus, if the individual sub-series data are used in the multivariate model, the required output for the total series using the multivariate model would not be available. The covariances would be required to manipulate the output into the desired form. A simple way of solving this problem, while continuing to use standard software, is to make a transformation on the multivariate state space model. This enables a direct estimation of the components of the total series, and their mean squared errors, within the multivariate framework.

The proposed transformation allows the total series to be included as one of the multivariate series. This means that the estimates of the components of the aggregated series and their mean squared errors are directly available from the output of the Kalman filter (see Section 3.5) with standard software. Thus, transforming the data before estimation, saves on manipulation of the results (all of which may or may not be provided in standard software) after estimation. As will be mentioned in Section 6.2.2, this transformation is not necessary for the estimation of the parameters.

Let $\mathbf{A}$ be a $K \times K$ transformation matrix:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} = \left(\begin{array}{ccc|c} 1 & 1 & & \dots & 1 \\ \hline & & & & 0 \\ & & \mathbf{I}_{(K-1)} & & \vdots \\ & & & & 0 \end{array}\right). \tag{3.33}$$

Applying $\mathbf{A}$ to obtain the transformed data, the total series becomes augmented to the data set comprising of series 1 to series $(K-1)$. The data for series K is no longer included in the data set. This data set will be referred to as the 'transformed' data,

$$\mathbf{A}\left(Y_{1t}, Y_{2t}, \ldots, Y_{Kt}\right)' = \left(Y_{tot,\,t}, Y_{1t}, \ldots, Y_{K-1,\,t}\right)'. \tag{3.34}$$

The transformation is applied to the state space model in (3.28) and (3.29) to give the following result. (Refer to Appendix A for properties of the Kronecker product).

$$\begin{aligned}
\mathbf{Y}_{(M),\,t} &= \mathbf{A}\mathbf{Y}_{(m),\,t} \\
&= \mathbf{A}(\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(m),\,t} \\
&= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),\,t} \\
&= (\mathbf{Z}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),\,t},
\end{aligned} \tag{3.35}$$

$$\begin{aligned}
\alpha_{(M),\,t+1} &= (\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),\,t+1} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_p \otimes \mathbf{A})\alpha_{(m),\,t} + (\mathbf{I}_p \otimes \mathbf{A})(\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(m),\,t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),\,t} + (\mathbf{I}_p\mathbf{G}_{(m)} \otimes \mathbf{A}\mathbf{I}_K)\gamma_{(m),\,t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),\,t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)(\mathbf{I}_u \otimes \mathbf{A})\gamma_{(m),\,t} \\
&= (\mathbf{T}_{(m)} \otimes \mathbf{I}_K)\alpha_{(M),\,t} + (\mathbf{G}_{(m)} \otimes \mathbf{I}_K)\gamma_{(M),\,t}.
\end{aligned} \tag{3.36}$$

The matrices $\mathbf{Z}_{(m)}$, $\mathbf{T}_{(m)}$ and $\mathbf{G}_{(m)}$ from (3.31) remain unchanged. However $\alpha_{(m),\,t}$, and $\gamma_{(m),\,t}$ are renamed with the (M) subscript to allow for the transformed elements. See (3.40) - (3.41) below.

The transformed model has the state space form:

$$\mathbf{Y}_{(M),\,t} = \mathbf{Z}_{(M)}\alpha_{(M),\,t}, \tag{3.37}$$

$$\alpha_{(M),\,t+1} = \mathbf{T}_{(M)}\alpha_{(M),\,t} + \mathbf{G}_{(M)}\gamma_{(M),\,t}, \tag{3.38}$$

where

$$\mathbf{Z}_{(M)} = \mathbf{Z}_{(m)} \otimes \mathbf{I}_K, \quad \mathbf{T}_{(M)} = \mathbf{T}_{(m)} \otimes \mathbf{I}_K, \quad \mathbf{G}_{(M)} = \mathbf{G}_{(m)} \otimes \mathbf{I}_K, \tag{3.39}$$

$$\begin{aligned}\alpha_{(M),\,t} = [&L_{tot,\,t}, L_{1t}, \ldots, L_{K-1,\,t},\ R_{tot,\,t}, R_{1t}, \ldots, R_{K-1,\,t},\\ &S_{tot,\,t}, S_{1t}, \ldots, S_{K-1,\,t},\ S_{tot,\,t-1}, S_{1,\,t-1}, \ldots, S_{K-1,\,t-1},\\ &S_{tot,\,t-2}, S_{1,\,t-2}, \ldots, S_{K-1,\,t-2},\\ &\varepsilon_{tot,\,t}, (\varepsilon_t + \varepsilon^*_{1t}), \ldots, (\varepsilon_t + \varepsilon^*_{K-1,\,t})\ \big]', \end{aligned} \tag{3.40}$$

$$\begin{aligned}\gamma_{(M),\,t} = \big[\ &\eta_{tot,\,t}, (\eta_t + \eta^*_{1t}), \ldots, (\eta_t + \eta^*_{K-1,\,t}),\ \zeta_{tot,\,t}, (\zeta_t + \zeta^*_{1t}), \ldots,\\ &(\zeta_t + \zeta^*_{K-1,\,t}),\ \omega_{tot,\,t}, (\omega_t + \omega^*_{1t}), \ldots, (\omega_t + \omega^*_{K-1,\,t}),\\ &\varepsilon_{tot,\,t+1}, (\varepsilon_{t+1} + \varepsilon^*_{1,\,t+1}), \ldots, (\varepsilon_{t+1} + \varepsilon^*_{K-1,\,t+1})\ \big]'. \end{aligned} \tag{3.41}$$

The distribution of $\gamma_{(M),\,t}$ is given by $\gamma_{(M),\,t} \sim N(0, \mathbf{Q}_{(M)})$, and this implies

$$\begin{aligned}\mathrm{Var}\left(\mathbf{G}_{(M)}\gamma_{(M),\,t}\right) &= \mathbf{G}_{(M)}\mathbf{Q}_{(M)}\mathbf{G}'_{(M)}\\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{(M),\,\eta} & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K\\ \mathbf{0}_K & \boldsymbol{\Sigma}_{(M),\,\zeta} & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K\\ \mathbf{0}_K & \mathbf{0}_K & \boldsymbol{\Sigma}_{(M),\,\omega} & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K\\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K\\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K\\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \boldsymbol{\Sigma}_{(M),\,\varepsilon} \end{pmatrix}. \end{aligned} \tag{3.42}$$

The disturbance terms of the total series are calculated by the following equations relating them to the disturbance terms of the sub-series:

$$\begin{aligned} \eta_{tot,\,t} &= K\eta_t + \sum_{k=1}^{K}\eta^*_{kt}, \qquad \zeta_{tot,\,t} = K\zeta_t + \sum_{k=1}^{K}\zeta^*_{kt},\\ \omega_{tot,\,t} &= K\omega_t + \sum_{k=1}^{K}\omega^*_{kt}, \qquad \varepsilon_{tot,\,t} = K\varepsilon_t + \sum_{k=1}^{K}\varepsilon^*_{kt}. \end{aligned} \tag{3.43}$$

If $K = 2$, the covariance matrix for the level component is given by

$$\boldsymbol{\Sigma}_{(M),\,\eta} = \mathbf{A}\boldsymbol{\Sigma}_\eta\mathbf{A}' = \begin{pmatrix} \sigma^2_{tot,\eta} & 2\sigma^2_\eta + \sigma^2_{1\eta^*}\\ 2\sigma^2_\eta + \sigma^2_{1\eta^*} & \sigma^2_\eta + \sigma^2_{1\eta^*} \end{pmatrix}, \tag{3.44}$$

with $\sigma^2_{tot,\,\eta} = 4\sigma^2_\eta + \sigma^2_{1\eta^*} + \sigma^2_{2\eta^*}$ from (3.24). Similarly for $\boldsymbol{\Sigma}_{(M),\,\zeta}$, $\boldsymbol{\Sigma}_{(M),\,\omega}$ and $\boldsymbol{\Sigma}_{(M),\,\varepsilon}$.

3.5 Application of the Kalman Filter

A linear Gaussian state space model may be analysed by applying the Kalman filter and Kalman smoother to the observations. The Kalman filter provides the optimal estimator of the state vector, α_{t+1}, taking into account observations up to time t, via a forward recursion. Denote the information provided by $Y_1, Y_2, \ldots, Y_t$, as $\mathbb{Y}_t$ when $t < T$. The Kalman smoother further improves the component estimates and provides the optimal estimator of the state vector, α_{t+1} at time $t < T$, taking into account all the observations in the sample, $Y_1, Y_2, \ldots, Y_T$, denoted by $\mathbb{Y}_T$.

Let the vector, $\mathbf{a}_{t+1|t}$, denote the conditional mean of the state vector, α_{t+1}, based on information available up to time t. Also, let the matrix, $\mathbf{P}_{t+1|t}$, denote the conditional variance for the estimation error of α_{t+1}, based on information available up to time t. $\mathbf{P}_{t+1|t}$ can also be referred to as the mean squared error (MSE) of the estimator $\mathbf{a}_{t+1|t}$ (Harvey, 1989). Therefore, the notation is given as (Durbin and Koopman, 2001):

$$\begin{aligned} \mathbf{a}_{t+1|t} &= \mathrm{E}(\alpha_{t+1}|\mathbb{Y}_t), \\ \mathbf{P}_{t+1|t} &= \mathrm{Var}(\alpha_{t+1}|\mathbb{Y}_t) \\ &= \mathrm{E}\left[(\alpha_{t+1} - \mathbf{a}_{t+1|t})(\alpha_{t+1} - \mathbf{a}_{t+1|t})'|\mathbb{Y}_t\right]. \end{aligned} \tag{3.45}$$

Note that when the assumption of normality of the disturbances and initial state vector (3.27) is dropped, the value of $\mathbf{a}_{t+1|t}$ provided by the Kalman filter minimises the mean square error within the class of all linear estimators (see Harvey, 1989, Section 3.2 and Durbin and Koopman, 2001, Section 4.2.1).

3.5.1 Kalman Filter for the Univariate Model

The standard set of filtering equations may be found in Chapter 4 of Durbin and Koopman (2001). For the univariate BSM in state space form, as described in (3.25) and (3.26), with corresponding system matrices (3.27), these are given by

$$\begin{aligned} \mathbf{a}_{t+1|t} &= \mathbf{T}\mathbf{a}_{t|t-1} + \mathbf{K}_t\nu_t, \\ \mathbf{P}_{t+1|t} &= \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{L}_t' + \mathbf{G}\mathbf{Q}\mathbf{G}', \end{aligned} \tag{3.46}$$

where

$$\begin{aligned}
\nu_t &= Y_t - \mathbf{Z}\mathbf{a}_{t|t-1} = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t} - \mathbf{Z}\mathbf{a}_{t|t-1}, \\
F_t &= \mathrm{Var}(\nu_t) = \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}^{'} + \mathrm{H}, \qquad\qquad \mathrm{H} = \mathrm{Var}(\varepsilon_{U,\,t}), \\
\mathbf{K}_t &= \mathbf{T}\mathbf{P}_{t|t-1}\mathbf{Z}^{'}F_t^{-1}, \\
\mathbf{L}_t &= \mathbf{T} - \mathbf{K}_t\mathbf{Z}.
\end{aligned} \tag{3.47}$$

Note that in the univariate model, H, F and ν are scalars. For example, if the state vector has dimension 5×1, as given by α_t in (3.27), then $\mathbf{Z}$ is a 1×5 matrix. $\mathbf{K}_t$ is a 5×1 vector and $\mathbf{L}_t$ is a 5×5 matrix.

The state vector estimator, $\mathbf{a}_{t|t}$, and its corresponding error variance matrix, $\mathbf{P}_{t|t}$, are defined by:

$$\begin{aligned}
\mathbf{a}_{t|t} &= \mathrm{E}(\alpha_t|\mathbb{Y}_t) = \mathbf{a}_{t|t-1} + \mathbf{M}_t F_t^{-1}\nu_t, \\
\mathbf{P}_{t|t} &= \mathrm{Var}(\alpha_t|\mathbb{Y}_t) = \mathbf{P}_{t|t-1} - \mathbf{M}_t F_t^{-1}\mathbf{M}_t^{'},
\end{aligned} \tag{3.48}$$

where $\mathbf{M}_t = \mathbf{P}_{t|t-1}\mathbf{Z}^{'}$ and has dimension 5×1. The updating equations in (3.46) may also be written in terms of the contemporaneous equations in (3.48) and are given by:

$$\begin{aligned}
\mathbf{a}_{t+1|t} &= \mathbf{T}\mathbf{a}_{t|t} \\
\mathbf{P}_{t+1|t} &= \mathbf{T}\mathbf{P}_{t|t}\mathbf{T}^{'} + \mathbf{G}\mathbf{Q}\mathbf{G}^{'}.
\end{aligned} \tag{3.49}$$

The variance matrix, $\mathbf{P}_1$, of the initial state vector α_1, is assumed to have the form:

$$\mathbf{P}_1 = \kappa\mathbf{P}_{\infty,1} + \mathbf{P}_{*,1}, \tag{3.50}$$

where κ is a large scalar value, $\mathbf{P}_{*,1}$ is the covariance matrix of the stationary components in α_1 and $\mathbf{P}_{\infty,1}$ is the covariance matrix of the non-stationary components in α_1, (Zivot and Wang, 2006).

Non-stationary components in the state vector require diffuse initialisation. A state is called diffuse if its covariance matrix is arbitrarily large. The problem of dealing with diffuse initial states in the Kalman filter was first solved by de Jong

(1991). An alternative approach of dealing with diffuse initial conditions is to apply the exact initial Kalman filter as described in detail in Koopman and Durbin (2000). The filter equations are derived as $\kappa \to \infty$, and the result is based on the expansion for F_t^{-1} as a power series in κ^{-1}. It is distinguished from the approximate filter obtained by choosing an arbitrary large value for κ and applying the standard Kalman filter (3.47). The exact approach is computationally more efficient in the process of parameter estimation when compared to other initialisation strategies such as that of de Jong (1991) (Koopman and Durbin, 2000, p293).

In particular, for the univariate BSM with a dummy seasonal component model, $\mathbf{a}_1 = E(\alpha_1)$ is a 5×1 zero vector, $\mathbf{P}_{\infty,1}$ is a 5×5 identity matrix and $\mathbf{P}_{*,1}$ is a 5×5 zero matrix. The term 'zero vector' is used in this thesis to describe a vector in which each element is zero. Similarly, the term 'zero matrix' is used to describe a matrix in which each element is zero. The exact initial Kalman filter can be applied using the `S+FinMetrics` software, in particular the set of functions collectively called the *SsfPack*, (Koopman *et al.*, 1999).

3.5.2 Kalman Filter for the Transformed Multivariate Model

The Kalman filter equations (3.46) to (3.49) need to be amended for the state space form given by (3.37) and (3.38), where the measurement error has been placed within the state vector. The corresponding system matrices $\mathbf{Z}_{(M)}$, $\mathbf{T}_{(M)}$, $\mathbf{G}_{(M)}$, and $\mathbf{Q}_{(M)}$ are included in (3.40) - (3.42). Only one equation listed in (3.47) requires modification, (apart from subscripts); F_t is now a matrix, which will be denoted by $\mathbf{F}_{(M),\,t}$, and its defining equation becomes

$$\mathbf{F}_{(M),\,t} = \mathbf{Z}_{(M)}\mathbf{P}_{(M),\,t|t-1}\mathbf{Z}_{(M)}^{'} \tag{3.51}$$

with H now a zero matrix, denoted by $\mathbf{H}$. The $\mathbf{Q}_{(M)} = \mathrm{Var}(\gamma_{(M),\,t})$ matrix now includes the variance matrix for the measurement error terms, $\mathbf{\Sigma}_{(M),\,\varepsilon}$, as shown in (3.42).

To compensate for this restructuring of the state vector, the set up of the exact initial conditions matrices described in Durbin and Koopman (2001, Section 5.2) is

amended. The $\mathbf{P}_{*,1}$ matrix, which holds the variance of the stationary part of α_1, is no longer a zero matrix, but now includes the $\mathbf{\Sigma}_{(M),\,\varepsilon}$ covariance matrix in the lower right $K \times K$ block diagonal. All other elements of the $(6K \times 6K)$ matrix are zero. The $\mathbf{P}_{\infty,1}$ matrix (also of dimension $6K \times 6K$) is no longer an identity matrix. The lower right $K \times K$ block diagonal is replaced by a $K \times K$ zero matrix. All other elements remain the same. For further details of the exact initialisation of the filter see Koopman and Durbin (2000).

A simplification of the exact initial Kalman filtering process (Koopman, 1997) for the multivariate model is described in detail in Koopman and Durbin (2000), where the elements of the observational vectors are brought into the analysis individually. This method basically converts the multivariate series into an univariate series and allows computational savings and simplifies the diffuse initialisation process. This method is applied in the *SsfPack* of functions in the `S-PLUS` module `S+FinMetrics`.

3.5.3 Kalman Smoother

The Kalman smoother takes into account all of the available observations in the series, not just the observations up to time t. The smoothed estimate of the state vector, α_t, is conditional on all observations, $(\mathbb{Y}_T)$, and is denoted by $\mathbf{a}_{t|T}$. It is the mean of the distribution of α_t given $\mathbb{Y}_T$ and is defined as (Harvey, 1989, Section 3.6):

$$\mathbf{a}_{t|T} = \mathrm{E}\left(\alpha_t|\mathbb{Y}_T\right). \tag{3.52}$$

The estimator, $\mathbf{a}_{t|T}$, is referred to as a smoother and is given by:

$$\begin{aligned} \mathbf{a}_{t|T} &= \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{r}_{t-1}, \quad t = T, \ldots, 1, \\ \mathbf{r}_{t-1} &= \mathbf{Z}^{'} F_t^{-1}\nu_t + \mathbf{L}_t^{'}\mathbf{r}_t, \qquad \text{with } \mathbf{r}_T = 0. \end{aligned} \tag{3.53}$$

The covariance matrix of α_t, conditional on all observations $(\mathbb{Y}_T)$, is defined as

$$\begin{aligned} \mathbf{V}_{t|T} &= \mathrm{Var}\left(\alpha_t|\mathbb{Y}_T\right) \\ &= E_T\left[\left(\alpha_t - \mathbf{a}_{t|T}\right)\left(\alpha_t - \mathbf{a}_{t|T}\right)^{'}\right] \end{aligned} \tag{3.54}$$

The matrix $\mathbf{V}_{t|T}$ is the MSE matrix of the estimator, $\mathbf{a}_{t|T}$, and is calculated by the recursion:

$$\begin{aligned}\mathbf{V}_{t|T} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{N}_{t-1}\mathbf{P}_{t|t-1} \\ \mathbf{N}_{t-1} &= \mathbf{Z}^{'} F_t^{-1}\mathbf{Z} + \mathbf{L}_t^{'}\mathbf{N}_t\mathbf{L}_t. \end{aligned} \tag{3.55}$$

Jointly, these smoothing equations (3.53) - (3.54) are referred to as the 'fixed interval smoother', (Durbin and Koopman, 2001, Section 4.3), and can be applied to the univariate state space model.

For the multivariate model, each vector and matrix is amended to have the (M) subscript. Let $\mathbb{Y}_{(M),\,T}$ denote the information provided by all observations in the transformed set of sub-series, that is, $\{Y_{tot,1}, Y_{tot,2}, \ldots, Y_{tot,T},\ Y_{11}, Y_{12}, \ldots, Y_{1T}, \ldots, Y_{K-1,1}, Y_{K-1,2}, \ldots, Y_{K-1,T}\}$. Then, the filter equations may be written as

$$\begin{aligned}\mathbf{a}_{(M),\,t|T} &= \mathrm{E}\left(\alpha_{(M),\,t}|\mathbb{Y}_{(M),\,T}\right) \\ &= \mathbf{a}_{(M),\,t|t-1} + \mathbf{P}_{(M),\,t|t-1}\ \mathbf{r}_{(M),\,t-1}, \qquad t = T, \ldots, 1. \\ \mathbf{r}_{(M),\,t-1} &= \mathbf{Z}_{(M)}^{'}\ \mathbf{F}_{(M),\,t}^{-1}\ \nu_{(M),\,t} + \mathbf{L}_{(M),\,t}^{'}\ \mathbf{r}_{(M),\,t}, \text{ with } \mathbf{r}_{(M),\,T} = 0. \end{aligned} \tag{3.56}$$

The covariance matrix of the state vector $\alpha_{(M),\,t}$ conditional on all T observations is given by:

$$\begin{aligned}\mathrm{V}_{(M),\,t|T} &= \mathrm{Var}\left(\alpha_{(M),\,t}|\mathbb{Y}_{(M),\,T}\right) \\ &= E_T\left[\left(\alpha_{(M),\,t} - \mathbf{a}_{(M),\,t|T}\right)\left(\alpha_{(M),\,t} - \mathbf{a}_{(M),\,t|T}\right)'\right], \end{aligned} \tag{3.57}$$

and is calculated by the recursion:

$$\begin{aligned}\mathrm{V}_{(M),\,t|T} &= \mathbf{P}_{(M),\,t|t-1} - \mathbf{P}_{(M),\,t|t-1}\ \mathbf{N}_{(M),\,t-1}\mathbf{P}_{(M),\,t|t-1}, \\ \mathbf{N}_{(M),\,t-1} &= \mathbf{Z}_{(M)}^{'}\ \mathbf{F}_{(M),\,t}^{-1}\ \mathbf{Z}_{(M)} + \mathbf{L}_{(M),\,t}^{'}\ \mathbf{N}_{(M),\,t}\ \mathbf{L}_{(M),\,t}. \end{aligned}$$

The smoother is based on at least as much information as the filtered estimator, and so the MSE of the smoother will be smaller than the MSE of the filtered estimator (Harvey, 1989, Section 3.6). The estimates of the model components given by the Kalman filter may be viewed as preliminary estimates, and the smoothed estimates may be viewed as the final estimates. Dagum and Quenneville (1993)

used the Kalman smoother estimates obtained from a BSM applied to the series of Total Department Stores in Canada and found that they were very close to estimates obtained with X-11-ARIMA.

3.6 Comparison of Univariate and Multivariate Methods

The main focus of this thesis is to determine whether the use of the underlying sub-series improves the estimates of the unobserved components of the aggregated series and hence the seasonally adjusted aggregated series. This chapter has considered two approaches, an univariate approach applied to the aggregated series, and a transformed multivariate approach which directly enables the components of the aggregated series to be estimated, taking into account all the sub-series. This section focuses on an appropriate measure that can be used to compare these two approaches. It involves the variance of the seasonally adjusted series. Also of interest to any seasonal adjustment procedure is the revision error. Revision error will be considered in Section 6.6.

3.6.1 Variance of the Seasonally Adjusted Series

The question arises as to how to calculate the accuracy of the seasonally adjusted series when a state space model is applied. Burridge and Wallis (1985) answer this question in more detail for the Kalman filter formulation of signal extraction methods. They note that the Kalman filter formulation is applicable to non-stationary time series and that, for stationary series it is equivalent to the classical Wiener-Kolmogorov theory as applied in Planas and Campolongo (2001). They state that the appropriate measure of the accuracy of the adjusted data is the error variance of the seasonal component estimate, conditional on the data. Further, Jain (2001) declares that an advantage of the structural model-based approach to seasonal adjustment is that it estimates the variance of the seasonally adjusted series as a by-product of estimating the seasonally adjusted series.

Harvey (1989) explains that when the optimal estimator of the seasonal component is obtained by the smoothing algorithm and subtracted from the original series to give the seasonally adjusted series,

$$Y^a_{t|T} = Y_t - \hat{S}_{t|T}, \qquad t = 1, \ldots, T, \tag{3.58}$$

then, the 'root mean squared error (RMSE) of $\hat{S}_{t|T}$, and hence $Y^a_{t|T}$, is also given by the smoother' (Harvey, 1989, p303). Hence, the RMSE of $\hat{S}_{t|t}$ will be given by the Kalman filter. The conditional mean, $\mathbf{a}_{t|t}$, (3.48) is unbiased and is a minimum mean square estimator of $\alpha_{t|t}$. The covariance matrix for the estimation error can be referred to as the mean squared error (MSE) matrix of the estimator (Harvey, 1989, Section 3.2.3). It is also referred to as the prediction mean squared error (PMSE) as in Pfeffermann and Tiller (2005). Thus, for the current-adjusted series, $Y^a_{t|t}$, the error variance of the seasonal component estimate, conditional on the data, is given by $\text{MSE}(\hat{S}_{t|t})$. This is the error variance of the state vector given by the Kalman filter in the matrix $\mathbf{P}_{t|t}$ (3.48), for the element pertaining to the seasonal component. The current-adjusted series, $Y^a_{t|t}$, can be viewed as the preliminary seasonally adjusted series as it is conditional on observations up to time t.

In this thesis, when using the multivariate transformed model, the error variance of the seasonal component estimate, conditional on the data for the aggregate series, $\text{MSE}(\hat{S}_{tot,t|t})$ will be denoted by $\text{MSE}(\hat{S}^M_{t|t})$. For the univariate method, $\text{MSE}(\hat{S}_{tot,t|t})$ will be denoted by $\text{MSE}(\hat{S}^U_{t|t})$. To compare the two values, the relative efficiency of the MSE obtained by the univariate method to that of the multivariate method will be calculated. This ratio is defined by:

$$RE_t(M) = \frac{\text{MSE}(\hat{S}^U_{t|t})}{\text{MSE}(\hat{S}^M_{t|t})}, \qquad t = 1, \ldots, T, \tag{3.59}$$

and can be considered as a preliminary estimate of $RE^{ks}_t(M)$, which uses the MSE of the smoothed seasonal component,

$$RE^{ks}_t(M) = \frac{\text{MSE}(\hat{S}^U_{t|T})}{\text{MSE}(\hat{S}^M_{t|T})}, \qquad t = 1, \ldots, T. \tag{3.60}$$

Note that at $t = T$, $RE_t(M) = RE^{ks}_t(M)$. The quantity of interest in the comparison of the multivariate method with the univariate method is $RE_t(M)$.

This is studied in more detail in Chapter 4. When the parameters of the model are unknown and have to be estimated, a downward bias in the MSE occurs. This will be discussed further in Chapter 6.

Chapter 4

Empirical Study with Exact Parameters

4.1 Introduction

To examine the question of whether, and under what conditions, the multivariate method is beneficial compared with the univariate method in determining a seasonally adjusted aggregated series, an experimental study is employed. A particular aggregated (or total) series with known (or exact) parameters is taken and then disaggregated into two sub-series. The two sub-series are determined by setting the parameters for each component and for each series. These parameters are used in determining the elements of the covariance matrices $\mathbf{\Sigma}_\eta$, $\mathbf{\Sigma}_\omega$ and $\mathbf{\Sigma}_\varepsilon$. The specification of their structure is referred to as the 'design' of the sub-series. By fixing the parameters for the total series, the parameters controlling the sub-series may be varied. This 'top-down' approach allows an experimental way of measuring the effect of the parameters of the sub-series on the accuracy of the seasonally adjusted total series.

In the previous chapter, it was noted that the error variance of the seasonal component estimate conditional on the data, $\text{MSE}(\hat{S}_{t|t})$, is an appropriate measure of the accuracy of the adjusted data. Thus, if the error variance of the seasonal component of the aggregated series is lower when the sub-series are modelled jointly, then the multivariate method is beneficial, and gains are achieved. The relative efficiency measure, $RE_t(M)$, defined in (3.59), is calculated for each different design

of the sub-series underlying the same total series.

By considering the ratio of the variances of the sub-series, and the correlation between the series for each of the level, seasonal and error components, the conditions under which gains are achievable with two sub-series are investigated in this chapter. An ideal situation is created whereby the exact or true parameters are used in an experiment so that the relative variances can be controlled. That is, using exact parameters maintains the relative values of the chosen seasonal and non-seasonal parameters. The results will serve as a guideline for identifying the conditions for which gains in efficiency occur, without the complication of estimating the parameters.

In a practical situation, the model parameters are unknown and need to be estimated. The estimation process could further influence the accuracy of the seasonally adjusted data. The estimation procedure and its implications are examined in Chapter 6.

The following section outlines the model structure and its state space form. Section 4.3 describes the experimental design and Section 4.3.2 explains the method of setting the parameters for the different sub-series designs. The application of the parameters into the Kalman filter to obtain the relative efficiency is detailed in Section 4.4, with the results in Section 4.5. In Section 4.6, the effect of altering the parameters of the total series is investigated with two examples.

4.2 Model Structure

For the experimental study of this chapter, the local level seasonal (LLS) model is chosen rather than the full BSM model. The local level seasonal model is a subset of the full BSM model with the restriction that the trend does not include a slope component. This means that the length of the state vector is reduced by K elements, and the number of parameters is reduced by $K + 1$. This reduces the complexity of the model and therefore simplifies the number of combinations of parameters to be set in the experimental study. A full BSM model is considered as another example

in Section 4.6

The multivariate model with unit-specific variances (Model 2) of a seasonal local level model with a dummy seasonal component is modified from (3.15) by removing the slope component (R_t) giving:

$$\begin{aligned} Y_{kt} &= L_{kt} + S_{kt} + \varepsilon_t + \varepsilon^*_{kt}, \\ L_{k,\,t+1} &= L_{kt} + \eta_t + \eta^*_{kt}, \\ S_{k,\,t+1} &= -\sum_{j=1}^{s-1} S_{k,\,t+1-j} + \omega_t + \omega^*_{kt}, \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} \varepsilon_t &\sim N(0, \sigma^2_\varepsilon), \qquad & \varepsilon^*_{kt} &\sim N(0, \sigma^2_{k\varepsilon^*}), \\ \eta_t &\sim N(0, \sigma^2_\eta), \qquad & \eta^*_{kt} &\sim N(0, \sigma^2_{k\eta^*}), \\ \omega_t &\sim N(0, \sigma^2_\omega), \qquad & \omega^*_{kt} &\sim N(0, \sigma^2_{k\omega^*}), \end{aligned} \tag{4.2}$$

and the aggregated series is given by:

$$\begin{aligned} Y_{tot,\,t} &= \sum_{k=1}^{K} Y_{kt} \\ &= L_{tot,\,t} + S_{tot,\,t} + \varepsilon_{tot,\,t}, \\ L_{tot,\,t+1} &= L_{tot,\,t} + \eta_{tot,\,t}, \\ S_{tot,\,t+1} &= -\sum_{j=1}^{s-1} S_{tot,\,t+1-j} + \omega_{tot,\,t}, \end{aligned} \tag{4.3}$$

where

$$\varepsilon_{tot,t} \sim N(0, \sigma^2_{tot,\varepsilon}), \quad \eta_{tot,t} \sim N(0, \sigma^2_{tot,\eta}), \quad \omega_{tot,t} \sim N(0, \sigma^2_{tot,\omega}). \tag{4.4}$$

There are three parameters to be set for the three components of the total series, namely $\sigma^2_{tot,\eta}$, $\sigma^2_{tot,\omega}$ and $\sigma^2_{tot,\varepsilon}$. The sample size, T, is set to be 40, allowing 10 years of quarterly data to be analysed. To simplify the number of parameters to be set, only two sub-series is disaggregated from the quarterly total series, giving $K = 2$ and $s = 4$.

Once the parameters for the total series have been set, the nine parameters for the two sub-series can be determined. Given $K = 2$, the constraints governing the

sub-series parameters, as described in (3.43), may be written as

$$\begin{aligned}\sigma^2_{tot,\varepsilon} &= 4\sigma^2_{\varepsilon} + \sigma^2_{1\varepsilon^*} + \sigma^2_{2\varepsilon^*},\\ \sigma^2_{tot,\eta} &= 4\sigma^2_{\eta} + \sigma^2_{1\eta^*} + \sigma^2_{2\eta^*},\\ \sigma^2_{tot,\omega} &= 4\sigma^2_{\omega} + \sigma^2_{1\omega^*} + \sigma^2_{2\omega^*}.\end{aligned} \tag{4.5}$$

4.2.1 State Space Form

Univariate: The state space form of the univariate LLS model is determined from the standard univariate form for the BSM (as described in Section 3.4.1), by excluding the slope term:

$$Y_t = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t}, \tag{4.6}$$

$$\alpha_{t+1} = \mathbf{T}\alpha_t + \mathbf{G}\gamma_t, \tag{4.7}$$

where, for quarterly data (s=4), and a dummy seasonal component,

$$\begin{aligned}\alpha_t &= \left[L_t,\ S_t,\ S_{t-1},\ S_{t-2}\right]', & \alpha_1 &\sim N(\mathbf{a}_1, \mathbf{P}_1),\\ \gamma_t &= \left[\eta_{U,\,t},\ \omega_{U,\,t}\right]', & \gamma_t &\sim N(0, \mathbf{Q}),\end{aligned}$$

$$\begin{aligned}\mathbf{Z} &= \begin{pmatrix}1 & 1 & 0 & 0\end{pmatrix}, & \varepsilon_{U,\,t} &\sim N(0, \mathrm{H}),\\ \mathbf{T} &= \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & -1 & -1 & -1\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{pmatrix}, & \mathbf{G} &= \begin{pmatrix}1 & 0\\ 0 & 1\\ 0 & 0\\ 0 & 0\end{pmatrix},\end{aligned}$$

$$\mathrm{Var}\left(\mathbf{G}\gamma_t\right) = \mathbf{G}\mathbf{Q}\mathbf{G}' = \begin{pmatrix}\sigma^2_{U,\,\eta} & 0 & 0 & 0\\ 0 & \sigma^2_{U,\,\omega} & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \quad \mathrm{H} = \sigma^2_{U,\,\varepsilon}. \tag{4.8}$$

The state space form of the model is required for the application of the Kalman filter. The standard Kalman filter is then applied to the univariate model with the

aggregated series data. This yields the estimates of the components for each time point and their MSEs.

Multivariate: The multivariate model given above (4.1) is transformed as described earlier in Section 3.4.3. As a result of the transformation, the two series in the system are the total series, $Y_{tot,\,t}$, and series 1, denoted by Y_{1t}. The transformation permits estimation of the seasonal component of the total series, $S_{tot,\,t}$, within a multivariate framework.

The state space model for the transformed system with $K = 2$ may be specified as follows, given that the measurement errors are placed within the state vector:

$$\begin{aligned} \mathbf{Y}_t &= (\mathbf{Z} \otimes \mathbf{I}_2)\alpha_{(M),\,t}, \\ \alpha_{(M),\,t+1} &= (\mathbf{T} \otimes \mathbf{I}_2)\alpha_{(M),\,t} + (\mathbf{G} \otimes \mathbf{I}_2)\gamma_{(M),\,t}, \end{aligned} \tag{4.9}$$

where $\mathbf{I}_2$ is a 2×2 identity matrix and

$$\begin{aligned} \alpha_{(M),\,t} = [&L_{tot,\,t}, L_{1t}, S_{tot,\,t}, S_{1t}, S_{tot,\,t-1}, S_{1,\,t-1}, S_{tot,\,t-2}, S_{1,\,t-2}, \\ &\varepsilon_{tot,\,t}, (\varepsilon_t + \varepsilon^*_{1t})]', \\ \gamma_{(M),\,t} = [&\eta_{tot,\,t}, (\eta_t + \eta^*_{1t}),\ \omega_{tot,\,t}, (\omega_t + \omega^*_{1t}), \varepsilon_{tot,\,t+1},\ (\varepsilon_{t+1} + \varepsilon^*_{1,\,t+1})]'. \end{aligned} \tag{4.10}$$

The system matrices are given by:

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{G} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\begin{aligned}\text{Var}\left((\mathbf{G}\otimes\mathbf{I}_2)\gamma_{(M),\,t}\right) &= \mathbf{G}_{(M)}\mathbf{Q}_{(M)}\mathbf{G}'_{(M)} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{(M),\,\eta} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} \\ \mathbf{0}_{(2\times2)} & \boldsymbol{\Sigma}_{(M),\,\omega} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} \\ \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} \\ \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} \\ \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \boldsymbol{\Sigma}_{(M),\,\varepsilon} \end{pmatrix}, \end{aligned} \tag{4.11}$$

where

$$\boldsymbol{\Sigma}_{(M),\,\eta} = \begin{pmatrix} \sigma^2_{tot,\eta} & 2\sigma^2_{\eta} + \sigma^2_{1\eta^*} \\ 2\sigma^2_{\eta} + \sigma^2_{1\eta^*} & \sigma^2_{\eta} + \sigma^2_{1\eta^*} \end{pmatrix}, \tag{4.12}$$

and similarly for $\boldsymbol{\Sigma}_{(M),\,\omega}$ and $\boldsymbol{\Sigma}_{(M),\,\varepsilon}$.

For the multivariate model, an adjustment to the standard Kalman filter equations is required due to the measurement errors being in the state vector. As well as the adjustment to the equation for F_t, an adjustment to the initial condition matrices is required to allow for the stationary nature of the measurement errors. This procedure is described in Section 3.5.2. The adjusted Kalman filter is applied to the state space model described above in (4.9) to (4.11), using the transformed data.

4.3 Design of the Study

As highlighted in the discussion on direct versus indirect adjustment in Section 2.4, when the series have similar seasonal patterns, direct adjustment is favoured. Alternatively, when the series have dissimilar patterns, indirect adjustment is favoured. Although neither the direct nor indirect seasonal adjustment involves a multivariate method, and the methods utilised are mainly filter-based X-12 ARIMA, or model-based ARIMA, as given in Planas and Campolongo (2001), the conclusions can still guide the setting of the sub-series parameters here. Planas and Campolongo (2001) found that when the stochastic properties of the two series were either slightly dissimilar or strongly dissimilar, the indirect adjustment became more precise than the direct adjustment. Further, it is the covariance structure between the series which

is crucial to the results found by Geweke (1978) who used a joint model. Taylor (1978, p432) summarises the result by stating that 'in cases where the series that are aggregated are very heterogeneous, or where the stochastic structure (sic) of the nonseasonal and seasonal components are dissimilar, the relative efficiency of the optimal procedure is quite high'. In application to this case study, the parameters for the first sub-series need to vary from those for the second. That is, the within components need to vary from being the same (as given by Model 1 in Section 3.3) to being very different (as given by Model 2 in Section 3.3). Also, the structure of the covariance matrices for the level component and the seasonal component need to be considered in relation to one another.

4.3.1 Setting the Parameters

A measure of between-series similarity, c, of the stochastic properties of the series is defined here to help quantify the comments above. Let c_η, c_ω and c_ε be the ratios of the variances of sub-series 1 and 2 within the level, seasonal and error components respectively. Using the level component as an example, the variance of the level component for sub-series k is given by:

$$\begin{aligned}\mathrm{Var}(L_{k,\,t+1} - L_{kt}) &= \mathrm{Var}(\eta_t + \eta^*_{kt}) \\ &= \sigma^2_\eta + \sigma^2_{k\eta^*}.\end{aligned}$$

Then the c-ratio for the level component is defined to be:

$$c_\eta = \frac{\mathrm{Var}(L_{1,\,t+1} - L_{1t})}{\mathrm{Var}(L_{2,\,t+1} - L_{2t})} = \frac{\sigma^2_\eta + \sigma^2_{1\eta^*}}{\sigma^2_\eta + \sigma^2_{2\eta^*}}. \tag{4.13}$$

Hence, analogously for the other components, the c-ratios for the seasonal and error components are given respectively by:

$$c_\omega = \frac{\sigma^2_\omega + \sigma^2_{1\omega^*}}{\sigma^2_\omega + \sigma^2_{2\omega^*}}, \quad c_\varepsilon = \frac{\sigma^2_\varepsilon + \sigma^2_{1\varepsilon^*}}{\sigma^2_\varepsilon + \sigma^2_{2\varepsilon^*}}. \tag{4.14}$$

If $c_\eta = c_\omega = c_\varepsilon = 1$ then Model 2 reverts to Model 1, the compound symmetry case, in which all diagonal elements have the same value and the same properties between series apply for each component. For the seasonal component, it does not

mean that the set of seasonal factors is the same for sub-series 1 and sub-series 2, but the degree of stability of the seasonal component is the same.

Planas and Campolongo (2001) examined three cases of stochastic behaviour in ARIMA models which they called 'similar', 'slightly dissimilar' and 'strongly dissimilar'. The model parameters for each component were varied such that the stochastic behaviour ranged from very stable (almost deterministic) to very unstable. In this study, the values of the c-ratios will be varied. For example, to obtain a strongly dissimilar case for the seasonal component, it would mean setting the value of c_ω to be much greater than one or much less than one.

In this study, the c-ratios vary in the set {1, 5, 10, 20} and their reciprocals {1, 0.2, 0.1, 0.05}. Furthermore, to set a design where the stochastic structures of the non-seasonal and seasonal components are different, the c-ratios need to differ between components, and so for one component, it could be greater than one, and for another component it could be less than one.

With this in mind, combinations of the c-ratios for the components are formulated and are labelled in the following tables. Table 4.1 shows design 'a' where all c-ratios are greater than or equal to one. Note that c_η and c_ε have been set to the same value in each design thereby reducing the number of combinations considered and setting the focus on the seasonal c-ratio, c_ω. Design 'b' given in Table 4.2, has $c_\omega > 1$ but has the reciprocal of these values for c_η and c_ε.

Table 4.1: Labels for sub-series design 'a': $c_\omega \geq 1$, and $c_\eta, c_\varepsilon \geq 1$.

Design 'a'		c_η and c_ε			
		1	5	10	20
c_ω	1	*a11*	*a12*	*a13*	*a14*
	5	*a21*	*a22*	*a23*	*a24*
	10	*a31*	*a32*	*a33*	*a34*
	20	*a41*	*a42*	*a43*	*a44*

In addition to setting c-ratios, the correlation between the series need to be considered for each component. For this study, the correlation values are necessarily

Table 4.2: Labels for sub-series design 'b': $c_\omega > 1$, and $c_\eta, c_\varepsilon < 1$.

Design 'b'		c_η and c_ε		
		0.2	0.1	0.05
c_ω	5	*b22*	*b23*	*b24*
	10	*b32*	*b33*	*b34*
	20	*b42*	*b43*	*b44*

positive, because of the structure of the covariance matrices, as given in (3.23). For the seasonal component, ρ_ω is set to one of the following values {0.1, 0.3, 0.5, 0.7, 0.9}. For the level and error components, the correlations (ρ_η and ρ_ε respectively), are set to be equal and take the values {0.2, 0.4, 0.6, 0.8}. The choice of these values means that the homogeneous case, in which the covariance matrices are proportional to one another, is avoided for certain combinations of the c-ratios described above (Harvey, 1989, Section 8.3). The design table, labelling the correlation combinations is given in Table 4.3.

Table 4.3: Labels for correlation design combinations of ρ_ω, ρ_η, and ρ_ε.

Correlations		ρ_η and ρ_ε				
		0.2	0.4	0.6	0.8	1.0
ρ_ω	0.1	*A1*	*B1*	*C1*	*D1*	*E1*
	0.3	*A2*	*B2*	*C2*	*D2*	*E2*
	0.5	*A3*	*B3*	*C3*	*D3*	*E3*
	0.7	*A4*	*B4*	*C4*	*D4*	*E4*
	0.9	*A5*	*B5*	*C5*	*D5*	*E5*

For example, as reference to Tables 4.1 and 4.3 makes clear, the design *A1a23* refers to the case when $c_\omega = 5$, $c_\eta, c_\varepsilon = 10$, $\rho_\omega = 0.1$, and $\rho_\eta, \rho_\varepsilon = 0.2$. Not all of the correlation designs are possible for each of the c-ratio design combinations due to the restrictions and constraints on the multivariate variance parameters, as explained in the next section.

The nine parameters in the multivariate model may be expressed in terms of the design of the model defined by $\{c_\eta,\ c_\omega,\ c_\varepsilon,\ \rho_\eta,\ \rho_\omega,\ \rho_\varepsilon,\ \sigma^2_{tot,\eta},\ \sigma^2_{tot,\omega},\ \sigma^2_{tot,\varepsilon}\}$.

4.3.2 Application of Constraints

In this study, the total series remains fixed but the properties of the underlying sub-series vary. The variance parameters for the total series are set with

$$\sigma^2_{tot,\eta} = 0.01, \quad \sigma^2_{tot,\omega} = 1, \quad \sigma^2_{tot,\varepsilon} = 1, \tag{4.15}$$

and will therefore be applied to the univariate method.

With these given univariate parameters, as well as the constraints given in (4.5), the c-ratios and the correlation for the required design, the multivariate parameters for each component are determined by solving a set of simultaneous equations. For example, the seasonal component equations are:

$$\begin{aligned} \sigma^2_{tot,\omega} &= 4\sigma^2_{\omega} + \sigma^2_{1\omega^*} + \sigma^2_{2\omega^*}, \qquad c_\omega = \frac{\sigma^2_{\omega} + \sigma^2_{1\omega^*}}{\sigma^2_{\omega} + \sigma^2_{2\omega^*}}, \\ \rho_\omega &= \frac{\sigma^2_{\omega}}{\sqrt{(\sigma^2_{\omega} + \sigma^2_{1\omega^*})(\sigma^2_{\omega} + \sigma^2_{2\omega^*})}}. \end{aligned} \tag{4.16}$$

Solving in terms of $\sigma^2_{tot,\omega}$, c_ω and ρ_ω, the seasonal parameters are

$$\begin{aligned} \sigma^2_{\omega} &= \frac{\rho_\omega \sqrt{c_\omega}\sigma^2_{tot,\omega}}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}, \\ \sigma^2_{1\omega^*} &= \frac{\sigma^2_{tot,\omega}(c_\omega - \rho_\omega\sqrt{c_\omega})}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}, \qquad \sigma^2_{2\omega^*} = \frac{\sigma^2_{tot,\omega}(1 - \rho_\omega\sqrt{c_\omega})}{1 + c_\omega + 2\rho_\omega\sqrt{c_\omega}}. \end{aligned} \tag{4.17}$$

Since $\sigma^2_{1\omega^*} \geq 0$, $\sigma^2_{2\omega^*} \geq 0$ and $\sigma^2_{\omega} \geq 0$, the restrictions on the correlations are such that,

$$\begin{aligned} \text{if} \quad c_\omega \geq 1, \quad \text{then} \quad 0 \leq \rho_\omega \leq \frac{1}{\sqrt{c_\omega}}, \\ \text{and if} \quad c_\omega < 1, \quad \text{then} \quad 0 \leq \rho_\omega \leq \sqrt{c_\omega}. \end{aligned} \tag{4.18}$$

Similar constraints apply to the level and error components.

4.4 Estimation of Components via the Kalman Filter

Analysis is performed with the `S-PLUS` software and, in particular, the set of functions collectively called the *SsfPack* (Koopman *et al.*, 1999) in the module `S+FinMetrics`. For this case study, the exact variance parameters are being used and therefore do not need to be estimated. The components of the state vector, $\alpha_{(M),t}$ (see first equation in (4.10)) are estimated using the Kalman filter recursions described in Section 3.5.1, with adjustments as given in Section 3.5.2. This gives $E(\alpha_{(M),t}|Y_t) = \mathbf{a}_{t|t}^M$, and the associated error variance matrix is given by $\mathbf{P}_{t|t}^M$ for $t = 1, \ldots, 40$. The error variance of the current-adjusted seasonal component, $\text{MSE}(\hat{S}_{tot,t|t})$, given by the Kalman filter, can be considered to be a preliminary estimate of the variance of the seasonally adjusted aggregate series, (Section 3.6.1). Note that the value of $\text{MSE}(\hat{S}_{tot,t|t})$ at $t = T = 40$ is equal to the MSE of the smoothed estimate $\text{MSE}(\hat{S}_{tot,t|T})$, since $T = 40$ for this example.

For the transformed multivariate model, the value of $\text{MSE}(\hat{S}_{tot,t|t})$, given the multivariate parameters, is denoted by $\text{MSE}(\hat{S}_{t|t}^M)$. For the univariate method, given the parameters for the total series (4.15), $\text{MSE}(\hat{S}_{tot,t|t})$ is denoted by $\text{MSE}(\hat{S}_{t|t}^U)$. Therefore, the relative efficiency of the filtered estimates for the univariate and multivariate models is given by

$$RE_t(M) = \frac{\text{MSE}(\hat{S}_{t|t}^U)}{\text{MSE}(\hat{S}_{t|t}^M)}, \qquad t = 1, \ldots, T. \tag{4.19}$$

This can be considered as a preliminary estimate of the equivalent measure which uses the MSE of the smoothed seasonal component. Here, $RE_t(M)$ is the quantity of interest in the comparison of the multivariate method with the univariate method.

Given the nine exact multivariate parameters, the data for Y_{1t} and Y_{2t} are generated from the multivariate model equations for $t = 40+T$ as described in (4.1), with starting values $L_1 = 5, S_1 = -1.5, S_0 = -1, S_{-1} = 0.5$ for both series. The first 40 data points of each series are discarded, leaving the $t = 1 \ldots T$ simulated quarterly observations required. For this study, T is set to 40, giving 10 years of quarterly data. The length of the series is therefore adequate to examine the behaviour of the

relative efficiency over time. Exact parameters are applied here so that the effect of the design on the relative efficiency is not obscured by the estimated parameter values. The effect of estimation is examined in Chapter 6.

The series are summed contemporaneously, $Y_{tot,t} = Y_{1t} + Y_{2t}$, to obtain the simulated aggregated series. Since exact parameters are applied to the model, no data are actually required to obtain the MSE of the seasonal component. For ease of computation however, one realisation of the data has been used with the exact parameters in the software `S+FinMetrics`. It is also possible to calculate the value of $RE_t(M)$ by substituting the parameter settings directly into its algebraic expression for a given value of t. This is because the theoretical expressions for the MSE values, which differ for each value of t, are a function of the model parameters only. In Chapter 5, a single measure has been derived for this model as an indicator of the value of $RE_t(M)$ using a quasi-likelihood approach.

4.5 Results

Relative efficiencies for simulations based on various parameter combinations are presented here.

4.5.1 The Effect of the C-ratios

The relative efficiency, $RE_t(M)$, is determined for each c-ratio combination specified in Tables 4.1 and 4.2 using the exact parameters. To obtain an overview of these results, the same correlation combination is chosen for each design, with $\rho_\omega = 0.1$ and $\rho_\eta = \rho_\varepsilon = 0.2$. This combination is labelled *A1* in Table 4.3.

Figure 4.1 shows the results over $t = 1, \ldots, 40$ for the 16 different *'a'* designs. The first noticeable feature is that for $t = 1, \ldots, 4$, the relative efficiency is exactly one. This result is due to the exact initial Kalman filter, as described in detail in Koopman (1997) and given in Appendix D, and to the fact that the exact parameter values have been used with constraints (Section 4.3.2). Using the exact initial Kalman filter for the local level seasonal BSM, the dependence of the covariance matrix $\mathbf{P}_{t|t}$ on κ

disappears after $t = 4$ (see Sections 3.5.1 and 3.5.2 with reference to Koopman and Durbin (2000, p287) and also Appendix D). However, from $t = 5$, gains using the multivariate method are achievable for some, but not all, of the *'a'* designs. These gains vary in magnitude and over time. For those designs which achieve gains, the gains climb in the next few time points to reach a steady value. The time until approximate convergence depends on the design. For example, design *a41* has the largest relative efficiency, $RE_{40}(M) = 1.29$, which translates to a reduction in MSE through using multivariate versus univariate of approximately 22.7%, but has the slowest rate of convergence.

The next few highest gains are for designs *a31*, *a14* and *a13* respectively. Note that *a41* has $c_\omega = 20$ and $c_\eta = c_\varepsilon = 1$, and *a31* has $c_\omega = 10$ and $c_\eta = c_\varepsilon = 1$, both with a high c-ratio for the seasonal component. For the design *a14*, $c_\omega = 1, c_\eta = c_\varepsilon = 20$ and for design *a13*, $c_\omega = 1, c_\eta = c_\varepsilon = 10$. Thus, the four *'a'* designs which give the highest $RE_t(M)$ result, have either a high between-series c-ratio for the seasonal component or a high between-series c-ratio for the non-seasonal components, but not both. The result is higher if the two c-ratios defining the design are at opposite ends of the scale. So, even when the variances for the two series are the same for the seasonal component ($c_\omega = 1$), if the variances of the non-seasonal components are very different ($c_\eta \geq 10, c_\varepsilon \geq 10$), a gain is still achievable (although not as large as when the variances differ) for the total seasonal component.

To explore the differences among the 'a' designs in more detail, the numerical results for $T = 40$ for each design are extracted. These results, which are equivalent to $RE_{40}(M)$, are found in Table 4.4. The lowest results for the relative efficiency belong to the designs which have $c_\omega = c_\eta = c_\varepsilon$, namely, *a11*, *a22*, *a33*, and *a44*. Note that for *a11*, where $c_\omega = c_\eta = c_\varepsilon = 1$, represents the compound symmetry case, (Model 1). As there is no gain in using the multivariate approach for Model 1, it is recommended to use the univariate approach. Even when all the c-ratios are high, as in *a33* and *a44*, where the series are largely dissimilar for all components, the fact that they are equal, overrides the between-series effect. Thus, when the c-ratios are equal, the structure of the covariance matrices become closer to a homogeneous state,

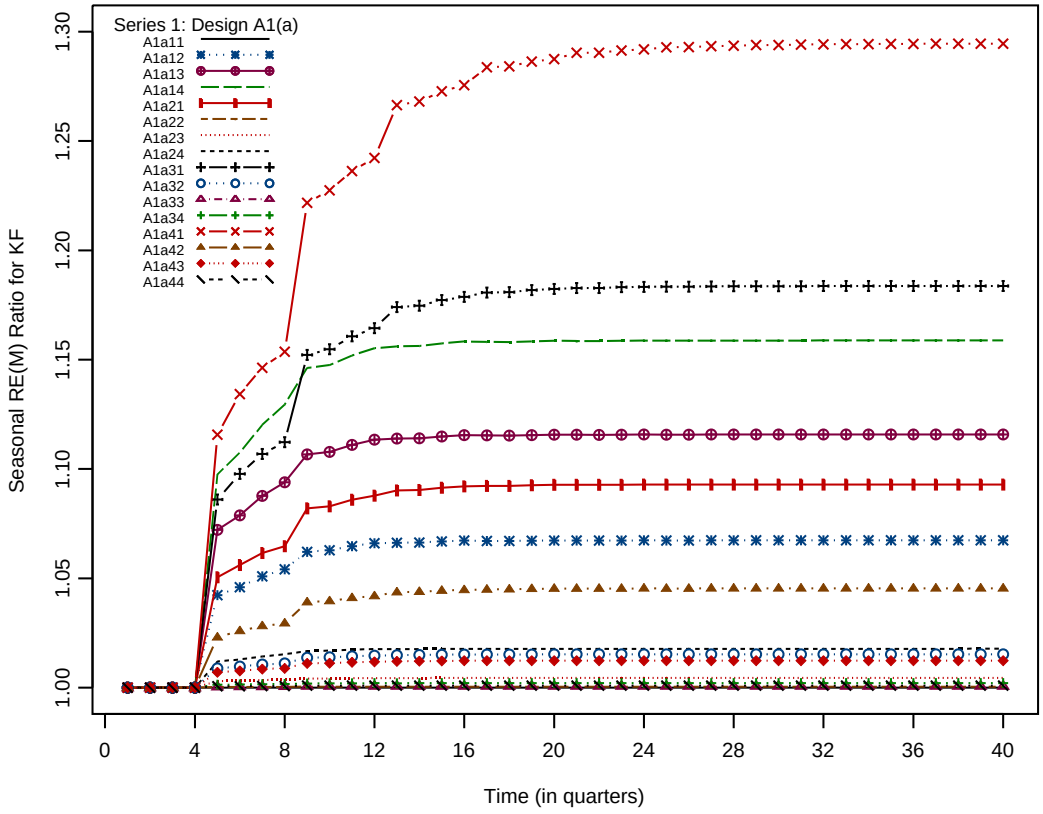


Figure 4.1: Results of $RE_t(M)$ for sub-series design 'a' with *A1* ($\rho_\omega = 0.1$, and $\rho_\eta = \rho_\varepsilon = 0.2$).

Table 4.4: Results of $RE_{40}(M)$ for sub-series design 'a' with *A1* ($\rho_\omega = 0.1$, and $\rho_\eta = \rho_\varepsilon = 0.2$).

Design 'a'		c_η and c_ε			
		1	5	10	20
c_ω	1	*a11* 1.0000	*a12* 1.0674	*a13* 1.1158	*a14* 1.1588
	5	*a21* 1.0929	*a22* 1.0005	*a23* 1.0045	*a24* 1.0178
	10	*a31* 1.1837	*a32* 1.0152	*a33* 1.0008	*a34* 1.0021
	20	*a41* 1.2945	*a42* 1.0454	*a43* 1.0124	*a44* 1.0010

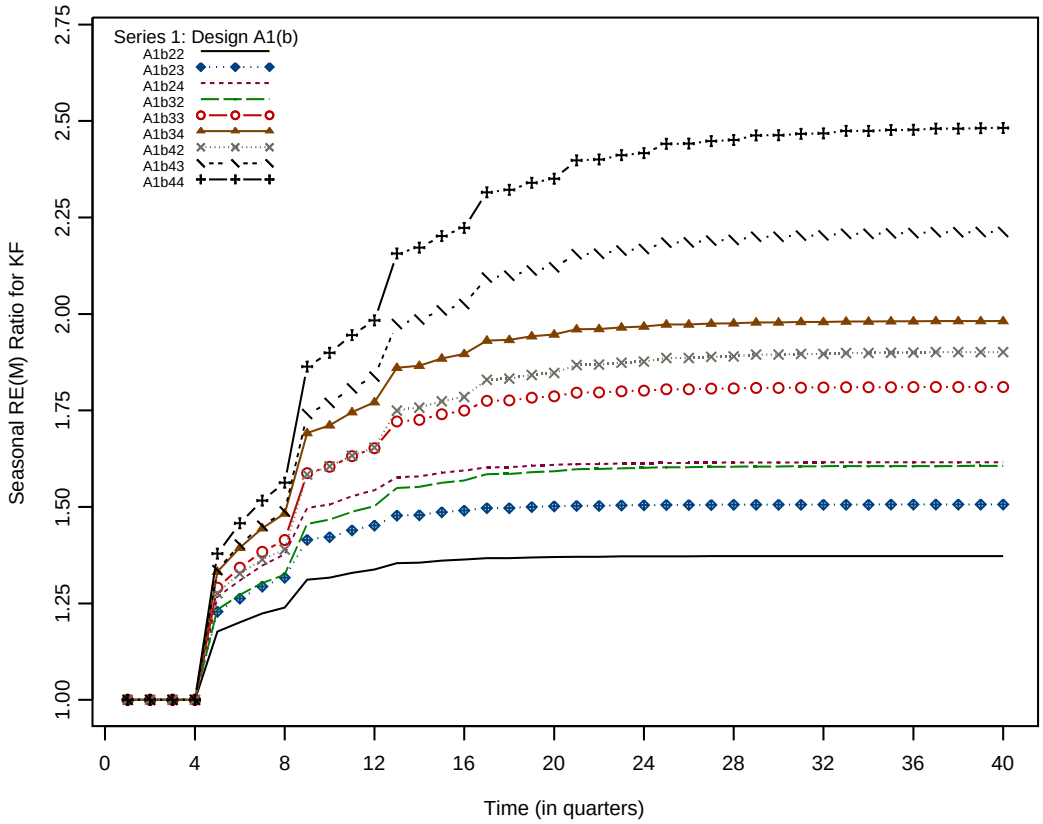


Figure 4.2: Results of $RE_t(M)$ for sub-series design 'b' with *A1* ($\rho_\omega = 0.1$, and $\rho_\eta = \rho_\varepsilon = 0.2$).

Table 4.5: Results of $RE_{40}(M)$ for sub-series design 'b' with *A1* ($\rho_\omega = 0.1$, and $\rho_\eta = \rho_\varepsilon = 0.2$).

Design 'b'		c_η and c_ε		
		0.2	0.1	0.05
c_ω	5	*b22* 1.3728	*b23* 1.5063	*b24* 1.6158
	10	*b32* 1.6060	*b33* 1.8108	*b34* 1.9818
	20	*b42* 1.9014	*b43* 2.2125	*b44* 2.4820

in which the covariance matrices are proportional to each other (see Harvey, 1989). Thus, for designs which have $c_\omega = c_\eta = c_\varepsilon$ where almost no gains are achieved, the univariate approach is preferable (due to simplicity) to the multivariate approach.

The 'b' designs use the reciprocal of the values of c_η, c_ε given in the 'a' designs and the results over time are shown in Figure (4.2). The results show a similar pattern for $RE_t(M)$. However, the magnitude is much greater than for the 'a' designs, with nine designs giving an $RE_{40}(M)$ of over 1.25 (or above 20% gain). The largest gain is achieved by design *b44* ($c_\omega = 20,\ c_\eta, c_\varepsilon = 0.05$), with $RE_{40}(M) = 2.48$. This translates to a gain of almost 60% for the multivariate method. Again, it can be seen that the designs where c_ω is very different to c_η and c_ε, for example *b44, b43, b34*, give the highest gains. The numerical results for $T = 40$ for each 'b' design are given in Table 4.5.

The results of the relative efficiency reported in Section 4.5.1 have been calculated for the local level seasonal model (LLS) with a quarterly dummy seasonal factor as described in (4.1) to (4.4). If the dummy seasonal factor is replaced by a trigonometric seasonal factor, the results of the relative efficiency calculations are very similar. For a description of the equivalent model with a trigonometric seasonal component and the results of the relative efficiency for three designs, refer to Appendix B.

4.5.2 The Effect of Correlation Settings

When the correlation settings are varied, as given in Table 4.3, the $RE_t(M)$ is also affected. The correlation combination in all 'a' and 'b' designs discussed so far, is identical, with $\rho_\omega = 0.1$ and $\rho_\eta = \rho_\varepsilon = 0.2$ (labelled *A1*). The results show that, even when the correlations between sub-series are small, large gains are attainable, with the size of the gain depending on the design structure.

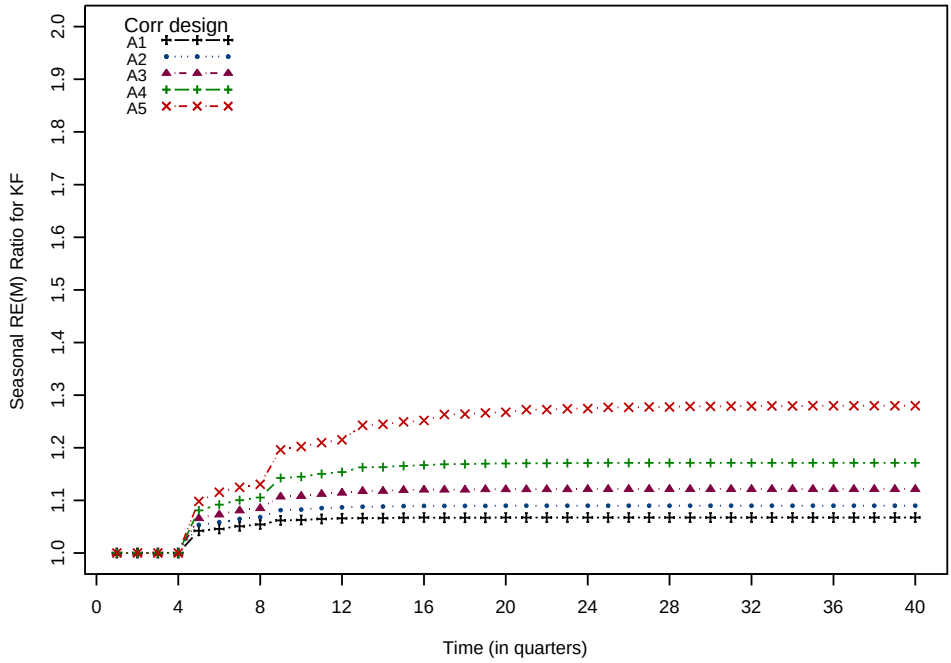


Figure 4.3: $RE_t(M)$ for sub-series design *a12* with *A1* - *A5*.

Three designs have been chosen to determine the effect of increasing the seasonal correlation for the 'a' design . Firstly, designs *a12*, *a13* and *a14* have been analysed with correlation combinations *A1* to *A5*, which keep the non-seasonal correlation coefficient low at 0.2, while allowing the seasonal correlation to be one of {0.1, 0.3, 0.5, 0.7, 0.9} as defined in Table 4.3. Figures 4.3 to 4.5 show these results for $t = 1, \ldots, 40$ and with the same vertical scale. The seasonal correlation affects $RE_t(M)$ for the seasonal component, as would be expected. The plots show that, as the seasonal correlation increases, the relative efficiency also increases, but the increase is dependent upon the design structure. As the non-seasonal c-ratio increases through

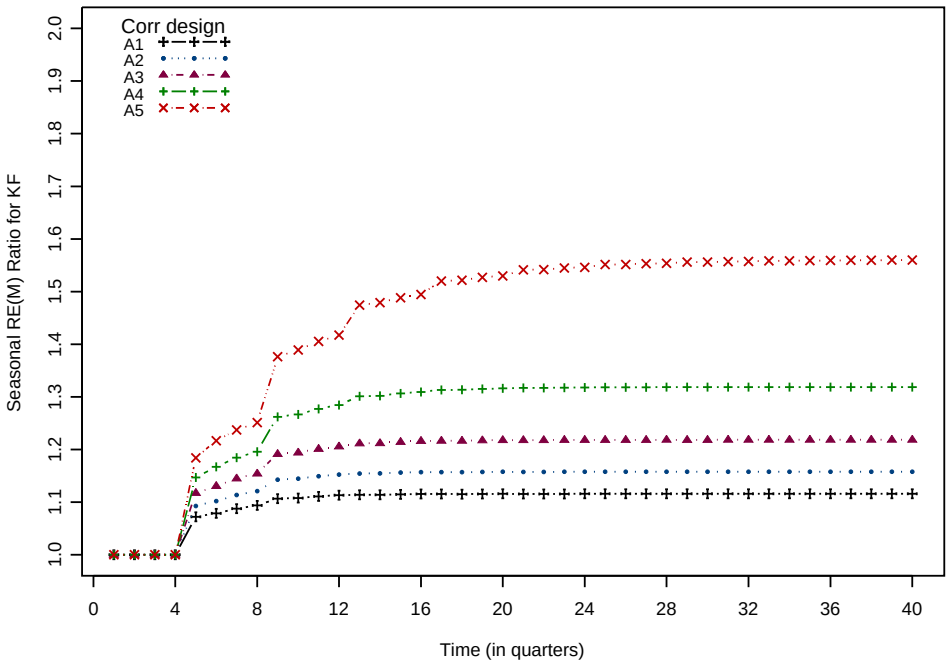


Figure 4.4: $RE_t(M)$ for sub-series design *a13* with *A1* - *A5*.

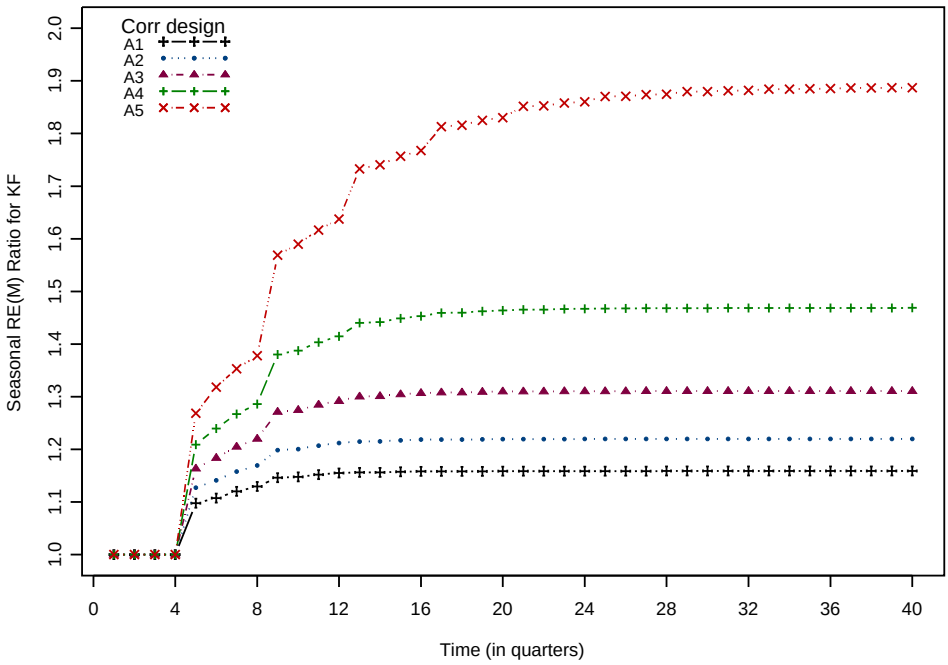


Figure 4.5: $RE_t(M)$ for sub-series design *a14* with *A1* - *A5*.

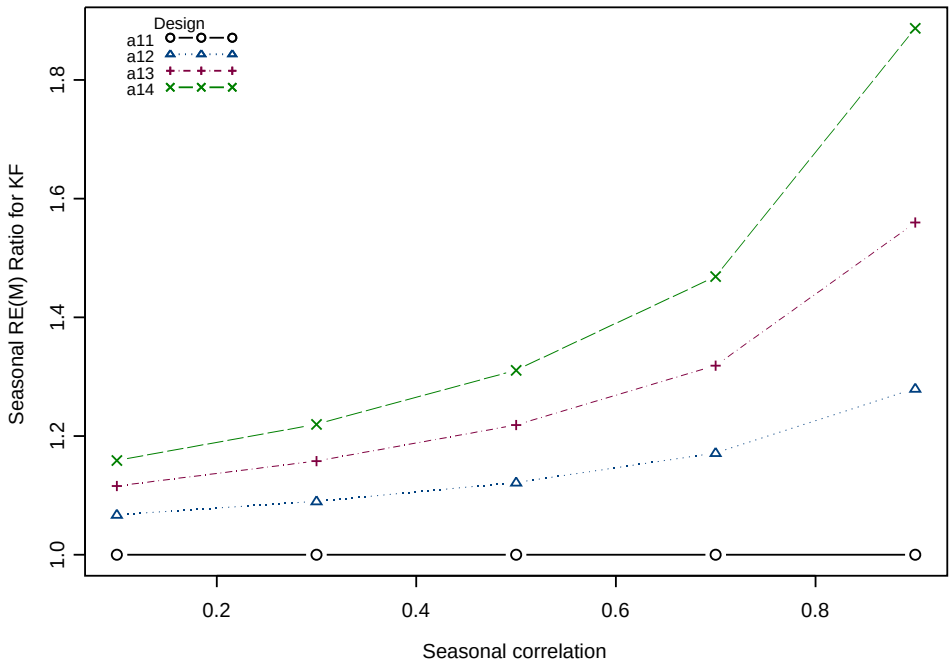


Figure 4.6: Seasonal correlation versus $RE_{40}(M)$ for designs *a11*, *a12*, *a13*, *a14* with *A1-A5*.

designs, from 5 (*a12*) to 10 (*a13*) and then to 20 (*a14*), the effect of the seasonal correlation coefficient intensifies, as shown in the following three plots (Figures 4.3 to 4.5).

By taking the results for the last time point (i.e. $T = 40$) from each time series within each of Figures 4.3, 4.4 and 4.5, the positive relationship between the correlation of the seasonal effects and the $RE_t(M)$ value is shown more clearly. Figure 4.6 shows these results, as well as those for design *a11*. Firstly, for the compound symmetry design (*a11*), the result for the relative efficiency remains constant at one as the seasonal correlation increases. The gradient of the curve increases from design *a12* to design *a13* to the steepest curve for design *a14*, thus as the non-seasonal c-ratios c_η, c_ε increase from 5 to 10 to 20.

The effect of increasing the non-seasonal correlation whilst keeping the seasonal correlation constant is now investigated. The designs *a21, a31, a41* are analysed for correlation combinations *A1* - *E1*. This means that the seasonal correlation is kept at $\rho_\omega = 0.1$, and the non-seasonal correlations ρ_η, ρ_ε are one of {0.2, 0.4, 0.6, 0.8, 1.0}. Figures 4.7 to 4.9 show the time series plots of $RE_t(M)$ for the three designs *a21, a31, a41* respectively with the same scale on the vertical axis. Note that here it is the non-seasonal correlation which is increasing, and the relative efficiency of the seasonal component is still affected. For design *a21*, for which $c_\omega = 5$ and $c_\eta = c_\varepsilon = 1$, the effect of increasing the non-seasonal correlation is quite small, as shown in Figure 4.7. For design *a31*, for which $c_\omega = 10$ and $c_\eta = c_\varepsilon = 1$, it can be seen from Figure 4.8 that the improvement is greater as the non-seasonal correlation increases. This becomes even more evident for design *a41*, for which $c_\omega = 20$ as shown in Figure 4.9.

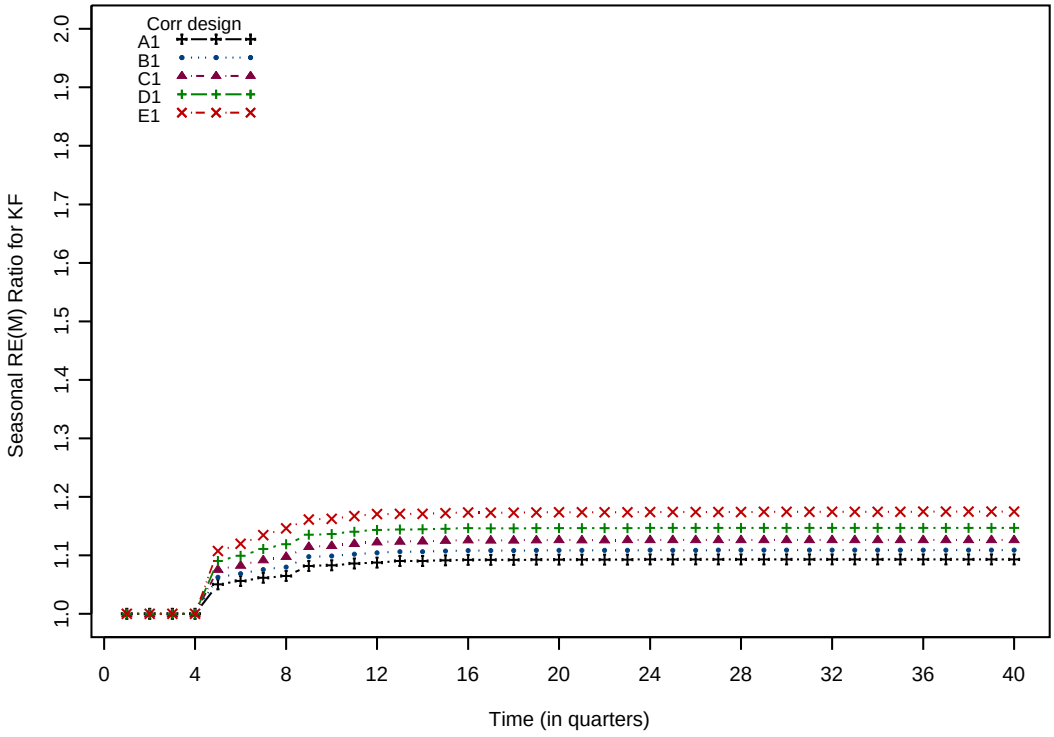


Figure 4.7: $RE_t(M)$ for sub-series design *a21* with *A1* - *E1*.

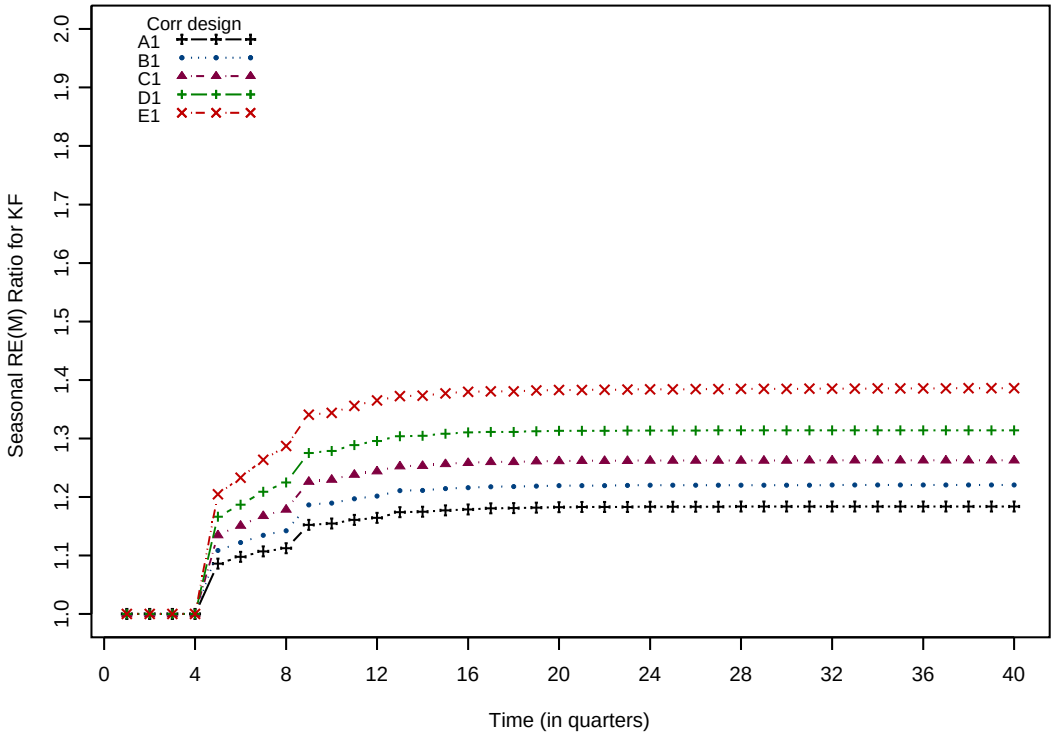


Figure 4.8: $RE_t(M)$ for sub-series design *a31* with *A1* - *E1*.

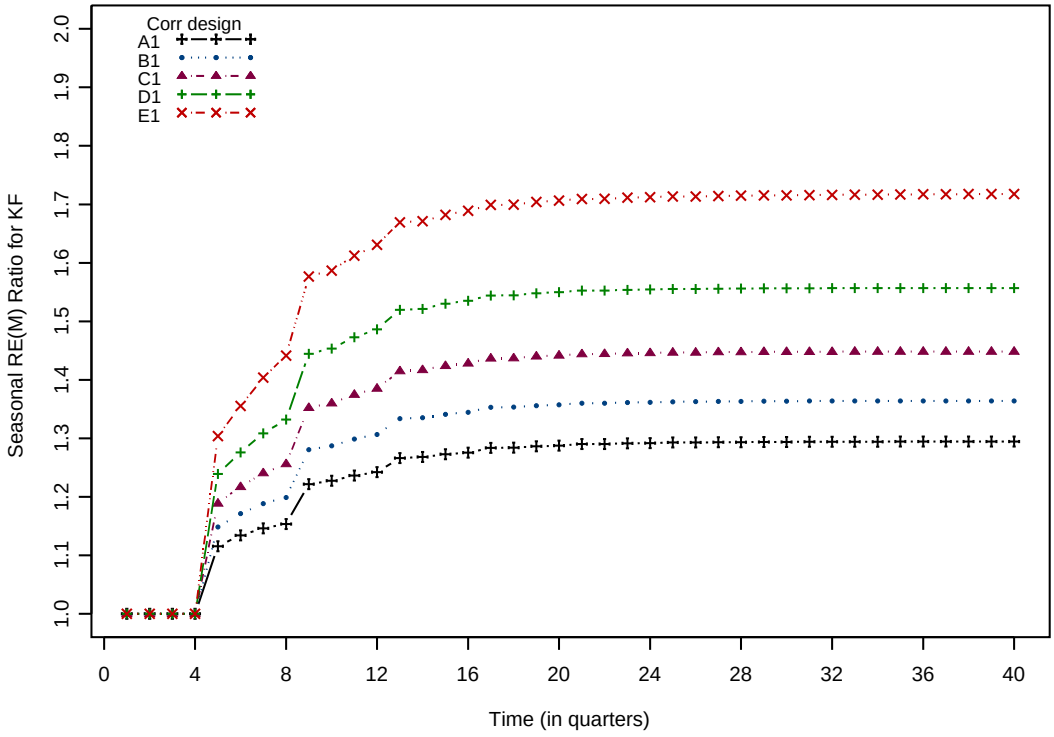


Figure 4.9: $RE_t(M)$ for sub-series design *a41* with *A1* - *E1*.

The results for $T = 40$ from each of the series shown in Figures 4.7 to 4.9, have been plotted against the non-seasonal correlation and given in Figure 4.10. The plot also shows the results for design *a11*. Overall, the results are similar to those shown in Figure 4.6. Again, it can be seen that the impact of the increasing non-seasonal correlation is dependent upon the parameter settings. There seems to be an interaction between the magnitude of c_ω and the non-seasonal correlation, since the gradient of the curve increases as both c_ω and $\rho_\eta, \rho_\varepsilon$ increase.

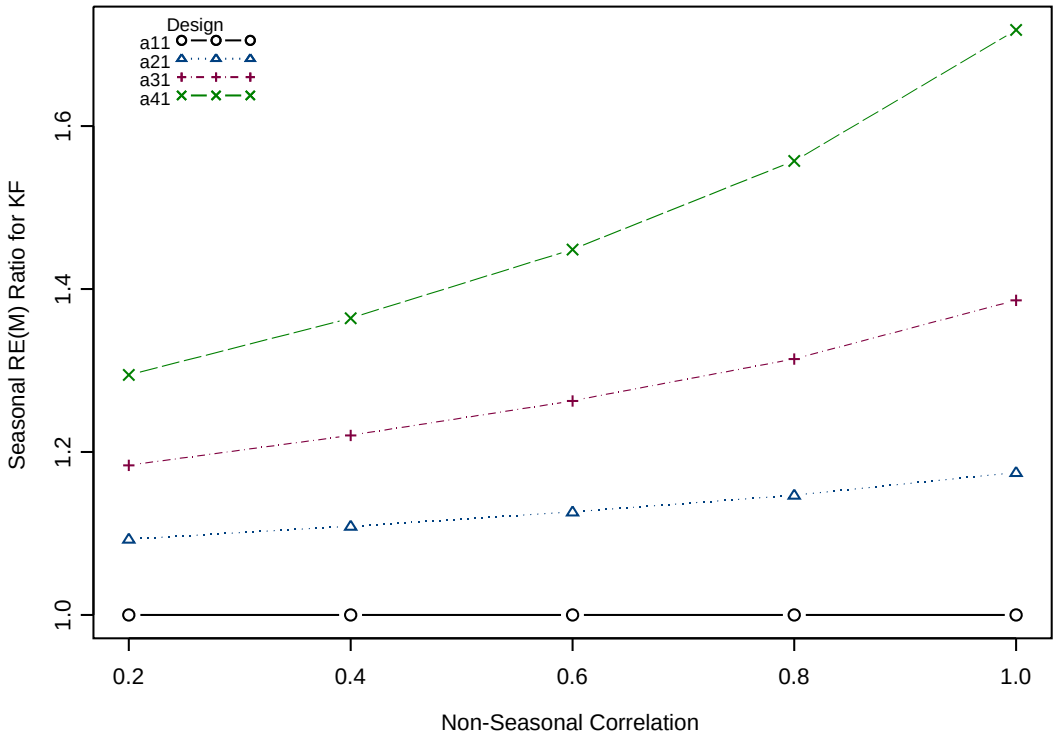


Figure 4.10: Non-seasonal correlation versus $RE_{40}(M)$ for designs *a11*, *a21*, *a31*, *a41* with *A1* - *E1*.

The correlation combinations given in Table 4.3 are not all available for each c-ratio combination. The restrictions on the correlation values are given in (4.18). The results of the relative efficiency at $T = 40$ for each possible combination for the '*a*' designs (except for *a11*) are presented in Table 4.6. For the '*b*' designs, analogous results are given in Table 4.7.

For the compound symmetry case (*a11*), the values of $RE_t(M)$ are all equal to

one for each correlation combination and are therefore not tabulated. Changing the correlation for this design has no effect on the relative efficiency. Thus, when all the c-ratios are unity, there is no gain using the multivariate approach in comparison to the univariate approach.

Table 4.6: Results of $RE_{40}(M)$ for design 'a' correlation combinations.

Design *a21*	$c_\eta = 1$ and $c_\varepsilon = 1$				
$c_\omega = 5$	*A1* 1.0929	*B1* 1.1088	*C1* 1.1263	*D1* 1.1469	*E1* 1.1748
	A2 1.0872	*B2* 1.1047	*C2* 1.1250	*D2* 1.1504	*E2* 1.1883
Design *a31*	$c_\eta = 1$ and $c_\varepsilon = 1$				
$c_\omega = 10$	*A1* 1.1837	*B1* 1.2204	*C1* 1.2626	*D1* 1.3141	*E1* 1.3860
	A2 1.1732	*B2* 1.2124	*C2* 1.2596	*D2* 1.3205	*E2* 1.4131
Design *a41*	$c_\eta = 1$ and $c_\varepsilon = 1$				
$c_\omega = 20$	*A1* 1.2945	*B1* 1.3641	*C1* 1.4484	*D1* 1.5570	*E1* 1.7178

	a12: $c_\eta, c_\varepsilon = 5$		*a13*: $c_\eta, c_\varepsilon = 10$	*a14*: $c_\eta, c_\varepsilon = 20$
$c_\omega = 1$	*A1* 1.0674	*B1* 1.0576	*A1* 1.1158	*A1* 1.1588
	A2 1.0899	*B2* 1.0787	*A2* 1.1578	*A2* 1.2198
	A3 1.1216	*B3* 1.1094	*A3* 1.2185	*A3* 1.3109
	A4 1.1714	*B4* 1.1598	*A4* 1.3187	*A4* 1.4687
	A5 1.2800	*B5* 1.2791	*A5* 1.5598	*A5* 1.8870
	a22: $c_\eta, c_\varepsilon = 5$		*a23*: $c_\eta, c_\varepsilon = 10$	*a24*: $c_\eta, c_\varepsilon = 20$
$c_\omega = 5$	*A1* 1.0005	*B1* 1.0039	*A1* 1.0045	*A1* 1.0178
	A2 1.0005	*B2* 1.0004	*A2* 1.0157	*A2* 1.0421
	a32: $c_\eta, c_\varepsilon = 5$		*a33*: $c_\eta, c_\varepsilon = 10$	*a34*: $c_\eta, c_\varepsilon = 20$
$c_\omega = 10$	*A1* 1.0152	*B1* 1.1088	*A1* 1.0008	*A1* 1.0021
	A2 1.0067	*B2* 1.1047	*A2* 1.0008	*A2* 1.0130
	a42: $c_\eta, c_\varepsilon = 5$		*a43*: $c_\eta, c_\varepsilon = 10$	*a44*: $c_\eta, c_\varepsilon = 20$
$c_\omega = 20$	*A1* 1.0454	*B1* 1.0813	*A1* 1.0124	*A1* 1.0010

For the remaining designs shown in Table 4.6, the correlation combination has an effect on the relative efficiency, as discussed previously. For design *a21* and *a31*, additional correlation combinations are *A2* to *E2*, for which $\rho_\omega = 0.3$ and $\rho_\eta, \rho_\varepsilon$ vary from 0.2 up to 1.0. As the non-seasonal correlation increases, the relative efficiency also increases, similarly to those already shown for *A1* to *E1*. Comparing the results for *A1* and *A2* and then *B1* and *B2*, it is seen that the relative efficiencies decrease slightly. Then for *C1* and *C2*, they are very similar and then for *D1* and *D2* and then *E1* and *E2*, they increase slightly.

Looking at results for design *a12* where $c_\omega = 1$ and $c_\eta = c_\varepsilon = 5$, when the seasonal correlation increases from 0.1 to 0.9 (*B1* to *B5*), the relative efficiencies also increase. This is similar to previously discussed results for *A1* to *A5*.

From Table 4.6, no simple patterns are evident. Nevertheless, it can be seen that when $c_\omega > c_\eta, c_\varepsilon$ (as in *a21*, *a31* and *a32*), for each t fixed, $RE_t(M)$ decreases from *A1* to *A2*, and from *B1* to *B2*. However, $RE_t(M)$ increases from *D1* to *D2*, and from *E1* to *E2*. On the other hand, when $c_\omega < c_\eta, c_\varepsilon$ (as in *a12*, *a13*, *a14*, *a23*, *a24*, and *a34*), $RE_t(M)$ increases from *A1* to *A2*, and from *B1* to *B2*. When $c_\omega = c_\eta = c_\varepsilon$ (as in *a11*, *a22*, and *a33*), $RE_t(M)$ remains stable for *A1* to *A2*, but decreases from *B1* to *B2*.

Table 4.7 shows the results of $RE_{40}(M)$ for design '*b*' correlation combinations. Design '*b*' has $c_\omega > 1$ and $c_\eta = c_\varepsilon < 1$. It is evident from Table 4.7 that $RE_{40}(M)$ increases from *A1* to *A2* and also from *B1* to *B2*, that is, as the seasonal correlation increases from 0.1 to 0.3.

Overall, apart from the compound symmetry model, the relative efficiency is affected by the correlation coefficients of both the seasonal and non-seasonal components. At first thought, it might be expected that as either of these correlations increase, the relative efficiency would also increase. It has been shown here, that while that is true for some cases, it is definitely not true for all cases. The two highest values of $RE_{40}(M)$ in Table 4.6, are 1.7178 for design *E1a41* and 1.8870 for design *A5a14*. Interestingly, these occur when the seasonal and non-seasonal correlations and c-ratios are at opposite ends of their given scales.

Table 4.7: Results of $RE_{40}(M)$ for design 'b' correlation combinations.

	b22: $c_\eta, c_\varepsilon = 0.2$		*b23*: $c_\eta, c_\varepsilon = 0.1$	*b24*: $c_\eta, c_\varepsilon = 0.05$
$c_\omega = 5$	*A1* 1.3728 *A2* 1.4076	*B1* 1.3689 *B2* 1.4071	*A1* 1.5063 *A2* 1.5711	*A1* 1.6158 *A2* 1.7099
	b32: $c_\eta, c_\varepsilon = 0.2$		*b33*: $c_\eta, c_\varepsilon = 0.1$	*b34*: $c_\eta, c_\varepsilon = 0.05$
$c_\omega = 10$	*A1* 1.6060 *A2* 1.6399	*B1* 1.6171 *B2* 1.6577	*A1* 1.8108 *A2* 1.8798	*A1* 1.9818 *A2* 2.0866
	b42: $c_\eta, c_\varepsilon = 0.2$		*b43*: $c_\eta, c_\varepsilon = 0.1$	*b44*: $c_\eta, c_\varepsilon = 0.05$
$c_\omega = 20$	*A1* 1.9014	*B1* 1.9431	*A1* 2.2125	*A1* 2.4820

4.6 Other Aggregated Series

The empirical results presented so far in this chapter are for a particular aggregated series with known parameters (4.15). Results for differing sets of parameters of the two sub-series or 'designs' have been determined using characteristics such as the c-ratio and the correlation coefficient for each component. For each different design, the sub-series are summed contemporaneously to obtain the same aggregated series. Thus, the aggregated (or total) series parameters have remained constant, whilst the sub-series parameters were varied.

There is an infinite set of total series for which this study could be repeated. To obtain an idea of what might occur if the total series had different parameters to those already chosen, some results are produced for two other series. For ease of reference, the total series previously considered is now referred to as *Series 1*, and consequently, the others are referred to as *Series 2*, and *Series 3*. The model and associated parameters for each will be described in the following subsections. A selection of results corresponding to those previously presented for *Series 1* is given in Section 4.6.1.

Series 1

As a reminder, *Series 1* is modelled by a seasonal local level BSM (given by (4.3) and (4.4)), which has a level component without a slope, and a dummy seasonal component for quarterly data. The parameters are the variances of the disturbance terms, specified as:

$$\textit{Series 1}: \qquad \sigma^2_{tot,\eta} = 0.01, \quad \sigma^2_{tot,\omega} = 1, \quad \sigma^2_{tot,\varepsilon} = 1. \tag{4.20}$$

The relationship between these univariate parameters may be described by the seasonal to non-seasonal ratio for the series:

$$\frac{\sigma^2_{tot,\omega}}{\sigma^2_{tot,\eta} + \sigma^2_{tot,\varepsilon}} = \frac{1.0}{0.01 + 1.0} = 0.99. \tag{4.21}$$

Series 2

For *Series 1*, the level component parameter ($\sigma^2_{tot,\eta} = 0.01$) is small in comparison with the other two parameters, giving a small signal to noise ratio of 0.01. For *Series 2*, the seasonal local level BSM model is retained but the level parameter is increased to 0.5. The resulting univariate parameters are:

$$\textit{Series 2}: \qquad \sigma^2_{tot,\eta} = 0.5, \quad \sigma^2_{tot,\omega} = 1, \quad \sigma^2_{tot,\varepsilon} = 1. \tag{4.22}$$

The seasonal to non-seasonal ratio for *Series 2* is given by:

$$\frac{\sigma^2_{tot,\omega}}{\sigma^2_{tot,\eta} + \sigma^2_{tot,\varepsilon}} = \frac{1.0}{0.5 + 1.0} = 0.67. \tag{4.23}$$

Series 3

The BSM model described in Section 3.2 was fitted to the well known data set, 'airline passengers' described by Harvey (1989, page 93). The monthly airline passengers data was aggregated to produce a quarterly time series and logarithms were taken. Harvey (1989, page 94) specifies the BSM parameters to be

$$\begin{aligned} \textit{Series 3}: \qquad & \sigma^2_{tot,\eta} = 5.32 \times 10^{-4}, \quad \sigma^2_{tot,\omega} = 1.32 \times 10^{-4}, \\ \text{and} \qquad & \sigma^2_{tot,\zeta} = 1.08 \times 10^{-6}, \quad \sigma^2_{tot,\varepsilon} = 0. \end{aligned} \tag{4.24}$$

Since the measurement error variance is zero, $\varepsilon_{U,\,t}$ is excluded from the model equations. In the model for *Series 1* and *Series 2*, the correlated measurement

error terms for the sub-series are put into the state vector due to the restriction in the software package. This is not necessary here. Thus, rather than the local level seasonal model, *Series 3* is generated from a BSM with level, slope and dummy seasonal components, and has no measurement error term. The parameters in (4.24), as well as some initial values, are used to generate data for *Series 3* from the model equations.

The seasonal to non-seasonal ratio for *Series 3* is given by:

$$\frac{\sigma^2_{tot,\omega}}{\sigma^2_{tot,\eta} + \sigma^2_{tot,\zeta}} = \frac{1.32 \times 10^{-4}}{(5.32 \times 10^{-4}) + (1.08 \times 10^{-6})} = 0.25. \tag{4.25}$$

Series 1, 2 and *3* have different parameters and as a result have different seasonal to non-seasonal ratios. *Series 2* and *Series 3* are both disaggregated into two sub-series, as carried out for *Series 1*, described in Section 4.3.

4.6.1 Results for other Aggregated Series

Results for *Series 2* and *Series 3* are presented here, and correspond to those for *Series 1* given in Section 4.5.1. The relative efficiency, $RE_t(M)$, has been calculated for each combination of the c-ratios given in Table 4.1 for design 'a' and Table 4.2 for design 'b' for *Series 2*. Only one correlation design has been considered here, namely *A1*, for which $\rho_\omega = 0.1$ and $\rho_\eta, \rho_\varepsilon = 0.2$. For *Series 3*, the c-ratio for the measurement error, c_ε, does not exist. It is replaced by c_ζ which is the analogous ratio of variances for the slope component between the series, thus

$$c_\zeta = \frac{\mathrm{Var}(R_{1t})}{\mathrm{Var}(R_{2t})} = \frac{\sigma^2_\zeta + \sigma^2_{1\zeta^*}}{\sigma^2_\zeta + \sigma^2_{2\zeta^*}}. \tag{4.26}$$

The results for $RE_{40}(M)$ for design 'a' are found in Table 4.8. For comparison purposes, the results for *Series 2* and *Series 3* are shown together with the previous results for *Series 1*. Looking at the results for the three series, it is clear that the parameters of the total series do not greatly affect the relative efficiency. There are some small differences, notably for designs *a13, a14*, and *a41*. For designs *a13* and *a14*, the relative efficiency increases from *Series 1* to *Series 3*. However, for some

designs which have a high c_ω and low non-seasonal c-ratios, such as design *a41*, the relative efficiency decreases; in this case it decreases from 1.29 to 1.26. Note that design *a41* has the highest relative efficiency for each of the total series.

Table 4.9 shows the results of $RE_{40}(M)$ for design '*b*'. The results of relative efficiency for *Series 2* and *3* show similar patterns as those for *Series 1*. Design *b44* has the highest relative efficiency for each series, with the largest ($RE_{40}(M) = 2.6$) being for *Series 3*, the airline series. For design *b24*, the relative efficiency increases from 1.62 to 1.77 from *Series 1* to *Series 3*.

The relative efficiency has also been plotted over time for *Series 2* and *Series 3*, as done previously for *Series 1*. Figure 4.11 shows the results of $RE_t(M)$ for design '*a*' for each total series. Each plot has the same vertical scale and the same legend for comparison. Comparing the three plots in Figure 4.11, it is noted that the rankings of the designs are almost identical, with the main differences being seen for designs *a14* and *a31*. For *Series 1*, the time plots intersect at $t = 8$ with *a31* eventually having the greater gain. For *Series 2*, the gain for design *a14* climbs more quickly but then eventually becomes almost the same as for *a31*. For *Series 3*, the gain for design *a14* is always larger than for *a31*.

For some designs, convergence rates also differ across the plots. A marked difference between *Series 1* and *Series 2* is that the level parameter increases from 0.01 to 0.5, thereby decreasing the seasonal to non-seasonal ratio from 0.99 to 0.67. *Series 2* seems to have a slower convergence rate than that for *Series 1*. For *Series 3*, which has a seasonal to non-seasonal ratio of only 0.25, the plot shows slower convergence than for the other two series.

Figure 4.12 shows the results for $RE_t(M)$ for design '*b*' for the three total series. Again, the three plots have the same vertical scale and legend. The rankings of the nine designs are again similar for each of the three series. The designs which differ here the most are *b43* and *b34*. Convergence rates are again slower for *Series 3* than for the other two series.

Table 4.8: *Series 1, 2*, and *3*: results of $RE_{40}(M)$ for design *A1a*.

Design '*a*'			c_η and c_ε (or c_ζ)							
			1		5		10		20	
c_ω	1	Series 1	*a11*	1.0000	*a12*	1.0674	*a13*	1.1158	*a14*	1.1588
		Series 2		1.0000		1.0732		1.1299		1.1840
		Series 3		1.0000		1.0769		1.1416		1.2092
	5	Series 1	*a21*	1.0929	*a22*	1.0005	*a23*	1.0045	*a24*	1.0178
		Series 2		1.0949		1.0005		1.0048		1.0199
		Series 3		1.0923		1.0005		1.0050		1.0216
	10	Series 1	*a31*	1.1837	*a32*	1.0152	*a33*	1.0008	*a34*	1.0021
		Series 2		1.1827		1.0156		1.0008		1.0022
		Series 3		1.1719		1.0151		1.0008		1.0023
	20	Series 1	*a41*	1.2945	*a42*	1.0454	*a43*	1.0124	*a44*	1.0010
		Series 2		1.2849		1.0450		1.0125		1.0010
		Series 3		1.2581		1.0422		1.0121		1.0010

Table 4.9: *Series 1, 2*, and *3*: results of $RE_{40}(M)$ for design *A1b*.

Design '*b*'			c_η and c_ε (or c_ζ)					
			0.2		0.1		0.05	
c_ω	5	Series 1	*b22*	1.3728	*b23*	1.5063	*b24*	1.6158
		Series 2		1.3946		1.5532		1.6963
		Series 3		1.4000		1.5839		1.7705
	10	Series 1	*b32*	1.6060	*b33*	1.8108	*b34*	1.9818
		Series 2		1.6246		1.8636		2.0849
		Series 3		1.6092		1.8769		2.1576
	20	Series 1	*b42*	1.9014	*b43*	2.2125	*b44*	2.4820
		Series 2		1.9004		2.2529		2.5939
		Series 3		1.8340		2.2030		2.6072

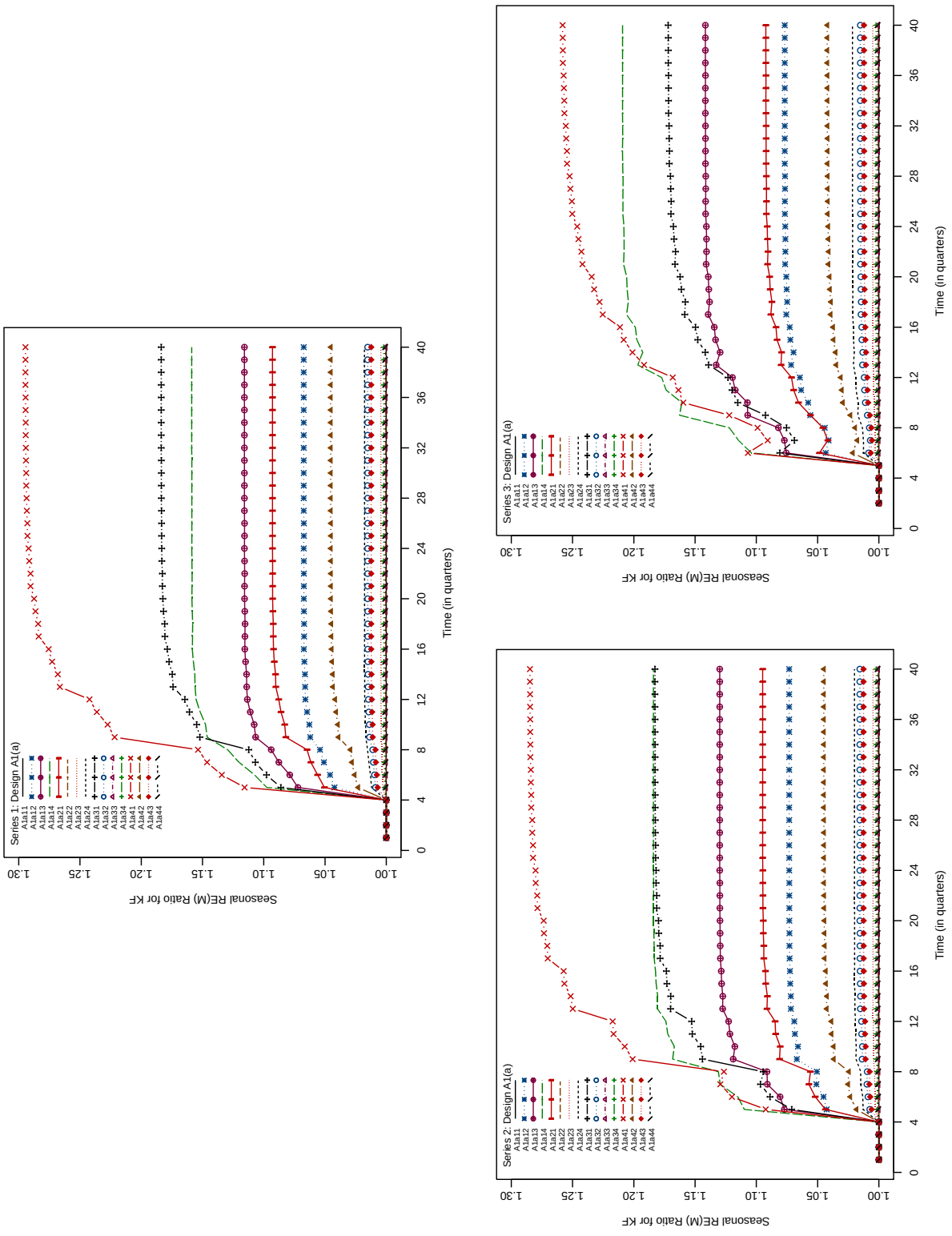


Figure 4.11: $RE_t(M)$ for designs *A1a11* to *A1a44* given by *Series 1 (top)*, *Series 2 (left)* and *Series 3 (right)*.

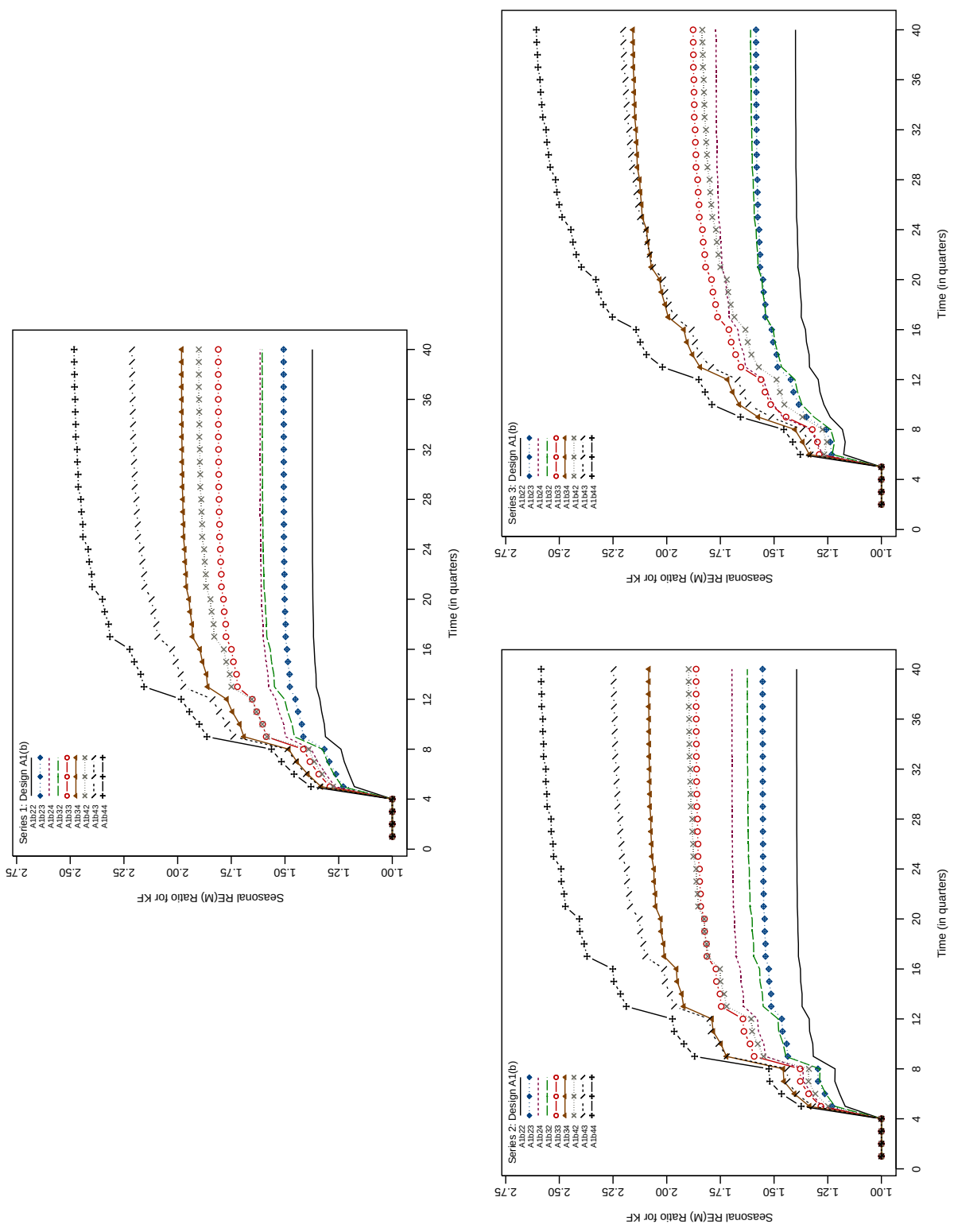


Figure 4.12: $RE_t(M)$ for designs $A1b22$ to $A1b44$ given by *Series 1 (top)*, *Series 2 (left)* and *Series 3 (right)*.

4.7 Conclusion

In this empirical study, the relative efficiency of the seasonally adjusted aggregated series has been investigated by using a multivariate structural time series model applied to the non-stationary sub-series. It is therefore an extension to the work of Geweke (1978), who used spectral densities in studying the accuracy of the seasonally adjusted series in reference to stationary time series. More recently, Planas and Campolongo (2001) applied some of the results for the multivariate case in Geweke (1978) but also used ARIMA models to describe the sub-series.

Although non-stationary time series have been considered here, and a different model structure has been used, the results from this study reflect quite similar conclusions to those made by Geweke (1978) and Planas and Campolongo (2001). It has been shown that gains in the accuracy of the seasonally adjusted series are possible by joint modelling of the sub-series.

This study focuses on one particular local level seasonal aggregate series and utilises a selection of designs for two sub-series. Keeping constraints for the aggregate parameters, the exact multivariate parameters are determined with reference to the ratios of the variances of the sub-series, and also the correlations for each of the seasonal and non-seasonal components. Gains are attainable under conditions which rely on the values of the parameters of the seasonal component and the non-seasonal components. The between-series (i.e. within components) and the within-series (i.e. between components) relationships for the two series have been studied and both affect the relative efficiency. The results are best summarised under five main points.

Firstly, when the two sub-series have the same variance parameters for both the seasonal and non-seasonal components (c-ratios are all equal to one), then there is no difference between the multivariate and the univariate methods. In addition, there is very little difference in the methods when the c-ratios are high, meaning that the series have very different variance parameters within components only if the c-ratios are equally high for both the seasonal and non-seasonal components. This is due to the design being close to the homogeneous system. This first point

confirms the 'similar' patterns case studied by Planas and Campolongo (2001, p21) who found that "the direct and multivariate adjustments tend to coincide and yield nearly equal estimation errors" when using ARIMA-based models.

Secondly, the relative efficiency is higher when the c-ratio for the seasonal component is very different to the c-ratio for the non-seasonal component, even if all c-ratios are greater than one, as in design '*a*'. The magnitude of the relative efficiency becomes much greater if the c-ratio is greater than one for one component (e.g. seasonal) but is less than one for other (i.e. non-seasonal) components, as in design '*b*'. This confirms the point made by Taylor in his comments on Geweke's paper (Geweke, 1978, p432): "where the stochastic structure of the non-seasonal and seasonal components are dissimilar, the relative efficiency of the optimal procedure is quite high". This study shows that even when the correlations between the series are low, this statement holds true.

Thirdly, if the c-ratios are held constant with non-seasonal correlation kept constant and low, when the seasonal correlation is increased incrementally, the relative efficiency improves, but the extent of the increase depends on the design structure. If the sub-series are described by Model 1, where c-ratios are all equal to one, then increasing the correlation has no effect on the relative efficiency. If the sub-series are quite similar, increasing seasonal correlation increases the relative efficiency, and this is magnified if the series have dissimilar c-ratios for the seasonal and non-seasonal components. A similar result holds if the seasonal correlation is kept constant and low, the c-ratios are held constant and the non-seasonal correlation is increased. Thus, the relative efficiency increases if the non-seasonal correlation increases away from the value of the seasonal correlation.

The last two points extend the work of Geweke (1978) and look more closely at the effect of different correlation combinations in addition to the effect of seasonal and non-seasonal stochastic structures. Fourthly, the results plotted over time show that the impact of increasing either the seasonal or non-seasonal correlation is greatest for the designs with very different c-ratios.

Lastly, this study also examines the evolution of relative efficiency over time, an

aspect not discussed by either of the previously mentioned authors. For the first 4 time points, the multivariate method and univariate method yield exactly the same MSEs for the filtered estimates. This is due to the application of the exact initial Kalman filter. For exact parameters, the theoretical expressions for the MSE of the seasonal component of the total series, for the univariate and multivariate methods, are equal for $t = 1, \ldots, 4$. As time progresses, the relative efficiency increases above one for each simulation carried out in this study. There are different rates of convergence but, on the whole, each plot shows a time series which reaches a steady state. Those with higher c-ratios for the seasonal component tend to be slowest to converge. Increasing the seasonal correlation also has an impact on the rate of convergence. As the correlation increases, convergence becomes slower. However, when the non-seasonal correlation is increased, the rate of convergence for relative efficiency seems to remain fairly constant.

If the aggregate series parameters are altered as in Section 4.6, the relative efficiency remains quite similar for the three series examined. The parameters of the aggregate series are shown to have an effect on the convergence of the relative efficiency over time. For the three series studied, *Series 3* has the smallest seasonal to non-seasonal ratio and also the slowest rate of convergence.

This empirical study highlights that efficiency gains are determined by different combinations of the parameters of the sub-series. As outlined in the five points above, the relative efficiency of the variance of the seasonally adjusted series, determined with the multivariate and univariate models, is highly dependent on the relative parameters within and between the sub-series, for both the seasonal and non-seasonal components. As there are many facets to the nature of this dependence, rather than a single general result, there are many conditions which may lead to gains in the efficiency.

The theoretical result given by the Kalman filter, which relies on the model parameters only, differs for each time point. Thus, given the parameters, it is difficult to make a decision about whether to use the multivariate or univariate model without applying the Kalman filter. However, by looking at the transition

from time $t-1$ to time t, and applying a quasi-likelihood approach to the BSM, a single measure may be developed for use as an indicator of the relative efficiency. This is derived in Chapter 5.

In this chapter, the focus has been on the relative efficiency when exact parameter values are used, and has not attempted to do any parameter estimation. A simulation study in Chapter 6 examines this extension for the two sub-series case. It also considers the effect of estimating the model parameters as the length of the series decreases. In Chapter 7, an example is given for the model which has more than two sub-series.

Chapter 5

Quasi-Likelihood Indicator of Relative Efficiency

5.1 Introduction

Due to the complexity of the iterative filtering process, the application of the Kalman filter can be difficult. The iterative nature of the process means that the theoretical expression for the estimates of the components contained in the state vector changes for each time point. This is also true of the elements in the error variance matrix of the state vector estimates. In the previous chapter, it was shown that the relative efficiency depends on the parameter values of the seasonal and non-seasonal components, and not on the observations. In particular, it was demonstrated that the relative efficiency is affected by the relationships among sub-series parameters, both between series (i.e. within components) and within series (i.e. between components). A simple expression for the relative efficiency ratio is not available, as the expression changes with each value of t. This makes it difficult to assess when gains from the multivariate approach will occur. It would therefore be helpful to have a single indicator which could be used to predict the relative efficiency for the Kalman filter given the variance parameters of both univariate and multivariate models. It would then also serve as a general indicator of when the multivariate approach is worthwhile.

In this chapter, a one-step-ahead indicator is derived using the quasi-likelihood method. Lin (2007) develops a quasi-likelihood approach to derive the gain matrix

for the state space model and shows that it is the same gain matrix defined in the Kalman filter process. Following Lin's procedure, a quasi-likelihood indicator may be determined for the relative efficiency; this compares the univariate and multivariate MSEs of the Kalman filter for the aggregate series. This measure is shown to have a direct relationship with the relative efficiency determined from the Kalman filter. The value of this single indicator can then be used to predict if a gain is achievable using the multivariate state space model rather than the univariate model. The indicator would therefore be useful in practice.

When there is a large number of sub-series, the multivariate state space form can be cumbersome to apply. The system of K sub-series may be reduced by aggregating the K original sub-series into $r < K$ new sub-series. This reduces the complexity of the model and also reduces the number of parameters.

There can be many ways of grouping the K original sub-series into the r new sub-series. The 18th century Scottish mathematician, James Stirling (1692 - 1770), determined the number of ways of partitioning a set of n elements into m non-empty sets. These are called 'Stirling numbers of the second kind'. Refer to Appendix C.3 for formulae and a table of values.

Take a simple example with $K = 4$ sub-series labeled A, B, C, D. There are seven ways to combine them into two groups. Three of the seven have two series in each group: {A+B, C+D} ; {A+C, B+D}; and {A+D, B+C}. The remaining four combinations have a single series in one group and three series in the other group: {A, B+C+D}; {B, A+C+D}; {C, A+B+D}; and {D, A+B+C}. The sub-series in each group are then aggregated contemporaneously to form the new r sub-series. The quasi-likelihood indicator for relative efficiency can then be calculated for each combination to identify which one will yield the greatest gain from the multivariate model.

Following a brief introduction to the quasi-likelihood method (Section 5.2), the method is applied to the state space form of both the univariate model (Section 5.3) and the multivariate model (Section 5.4). In the subsequent section, the quasi-likelihood indicator is derived for the case $K = 2$. In Section 5.6, the relationship

between the quasi-likelihood indicator and the Kalman filter relative efficiency is determined for the homogeneous model, the compound symmetry model (Model 1), and also for Model 2. Finally, in Section 5.7, the indicator is derived for the case where $K > 2$ original sub-series are grouped into $r < K$ new sub-series.

5.2 Introduction to Quasi-Likelihood Method

Unlike maximum likelihood, the quasi-likelihood is a method that only involves assumptions regarding moments up to second order. The principle of the quasi-likelihood (QL) method involves estimation of unknown parameters via a quasi-score estimating function. Estimating functions are functions of both data and the parameters (Heyde, 1997). The quasi-likelihood estimator is derived from setting the quasi-score estimating function to zero and solving. This is a method similar to that used to derive the maximum likelihood estimator from the likelihood score.

The primary rule is given by Godambe and Heyde (1987):

Definition: Let $\mathcal{G}_T$ be an estimating function space in R^n associated with unknown parameter $\theta \in \Theta$. Each element $\mathbf{G}_T(\theta)$ of $\mathcal{G}_T$ satisfies the conditions that $E[\mathbf{G}_T(\theta)] = 0$, $E[\mathbf{G}_T(\theta)\mathbf{G}_T'(\theta)]$, and $E[\dot{\mathbf{G}}_T(\theta)]$ are non-singular, for all $\theta \in \Theta$. Note that $\mathbf{G}_T'(\theta)$ denotes the transpose of $\mathbf{G}_T(\theta)$ and $\dot{\mathbf{G}}_T(\theta)$ denotes the derivative of $\mathbf{G}_T(\theta)$ with respect to the unknown parameter θ.

If $\mathbf{G}_T^*(\theta) \in \mathcal{G}_T$ such that

$$\left[E(\dot{\mathbf{G}}_T)\right]^{-1} E(\mathbf{G}_T\mathbf{G}_T') \left[E(\dot{\mathbf{G}}_T')\right]^{-1} - \left[E(\dot{\mathbf{G}}_T^*)\right]^{-1} E(\mathbf{G}_T^*\mathbf{G}_T^{*\prime}) \left[E(\dot{\mathbf{G}}_T^{*\prime})\right]^{-1} \quad (5.1)$$

is non-negative for all $\mathbf{G}_T(\theta) \in \mathcal{G}_T$ and $\theta \in \Theta$, then $\mathbf{G}_T^*(\theta)$ is called the quasi-score estimating function in $\mathcal{G}_T$. The following equation is called a quasi-score normal equation:

$$\mathbf{G}_T^*(\theta) = 0. \quad (5.2)$$

The solution to (5.2) is called a quasi-likelihood estimate of θ.

Lin (2007) provides an alternative derivation for the Kalman filter gain matrix for a state space model, using the quasi-likelihood approach. The following sections

apply Lin's technique to determine an alternative expression for the MSE of the seasonal component for both the univariate and multivariate methods. The unknown parameter θ, is a subset of the state vector α_t given in the univariate case by (3.27).

5.3 Quasi-Likelihood Approach to the Univariate BSM

For the univariate case, the error variance associated with the quasi-likelihood estimate of the seasonal component is derived here. The univariate BSM with level, slope and quarterly dummy seasonal components described in Section 3.2 may be rewritten with the one-step transition going from $t-1$ to t instead of t to $t+1$. This results in the following system of equations. Note that this does not change the model in any way, but is a more straight forward way of writing the model for the application of the quasi-likelihood method.

$$\begin{aligned}
Y_t &= L_t + S_t + \varepsilon_{U,\,t}, && \mathrm{E}(\varepsilon_{U,\,t}) = 0, \quad \mathrm{Var}(\varepsilon_{U,\,t}) = \sigma^2_{U,\,\varepsilon}, && (5.3)\\
L_t &= L_{t-1} + R_{t-1} + \eta_{U,\,t-1}, && \mathrm{E}(\eta_{U,\,t}) = 0, \quad \mathrm{Var}(\eta_{U,\,t}) = \sigma^2_{U,\,\eta},\\
R_t &= R_{t-1} + \zeta_{U,\,t-1}, && \mathrm{E}(\zeta_{U,\,t}) = 0, \quad \mathrm{Var}(\zeta_{U,\,t}) = \sigma^2_{U,\,\zeta},\\
S_t &= -\sum_{j=1}^{s-1} S_{t-j} + \omega_{U,\,t-1}, && \mathrm{E}(\omega_{U,\,t}) = 0, \quad \mathrm{Var}(\omega_{U,\,t}) = \sigma^2_{U,\,\omega}, && (5.4)
\end{aligned}$$

for $t = 1, \ldots, T$. In state space form, this is written as

$$Y_t = \mathbf{Z}\alpha_t + \varepsilon_{U,\,t}, \tag{5.5}$$

$$\alpha_t = \mathbf{T}\alpha_{t-1} + \mathbf{G}\gamma_{t-1}. \tag{5.6}$$

The system matrices, $\mathbf{Z}, \mathbf{G}$ and $\mathbf{T}$, remain unchanged and are given in equation (3.27). The state vector obtained for the Kalman filter approach, α_t, includes the level, slope and quarterly dummy seasonal components, such that

$$\alpha_t = \left[L_t,\ R_t,\ S_t,\ S_{t-1},\ S_{t-2}\right]'. \tag{5.7}$$

For the dummy seasonal model with $s = 4$, the state vector, α_t, defined in (3.27), has three seasonal elements. For a dummy seasonal component, only the

first element S_t has a disturbance term in the equation (5.4), which defines it. The last two elements, S_{t-1} and S_{t-2}, do not have disturbance terms connected to them. This is evident from the $\mathbf{G}$ matrix in the state space form which has zeros in its last two rows (3.27). When applying the quasi-likelihood method, S_{t-1} and S_{t-2} are therefore not required in the state vector. The state vector from the Kalman filter approach, α_t, is amended here by multiplying it by a selection matrix which is denoted by $\mathbf{U}$:

$$\mathbf{U} = \begin{pmatrix} 1\ 0\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0\ 0 \\ 0\ 0\ 1\ 0\ 0 \end{pmatrix}. \tag{5.8}$$

Hence, the state vector for the quasi-likelihood method, which will be denoted by $\alpha_t^{(q)}$, is given by:

$$\begin{aligned} \alpha_t^{(q)} &= \mathbf{U}\alpha_t \\ &= [L_t,\ R_t,\ S_t]^{'}. \end{aligned} \tag{5.9}$$

Before continuing with any derivations, some notation needs to be defined. In this chapter, $\mathcal{F}_{t-1}$ is defined by the following information:

$$\mathcal{F}_{t-1} = \{Y_1,\ Y_2, \ldots, Y_{t-1};\ L_1, \ldots, L_{t-1};\ R_1, \ldots, R_{t-1};\ S_1, \ldots, S_{t-1}\}. \tag{5.10}$$

Also, let $\mathcal{F}_{t-1}^*$ (5.11) denote the information provided by $\mathcal{F}_{t-1}$ in addition to the observation Y_t:

$$\mathcal{F}_{t-1}^* = \mathcal{F}_{t-1} \cup \{Y_t\}. \tag{5.11}$$

The quasi-likelihood (QL) approach assumes known values at the previous step, which includes not only the observations but also the estimated components of the state vector (Lin, 2007, p1630 and p1633). Thus, the aim is to find a one-step optimal estimator of $\alpha_t^{(q)}$ given all information provided by $\mathcal{F}_{t-1}^*$. Let $\mathbf{a}_t^{(q)}$ denote the optimal estimator of $\alpha_t^{(q)}$ based on the information $\mathcal{F}_{t-1}^*$. Also, let $\mathbf{a}_{t|t-1}^{(q)}$ denote

the optimal estimator of $\alpha_t^{(q)}$ based on $\mathcal{F}_{t-1}$, such that

$$\begin{aligned}
\mathbf{a}_{t|t-1}^{(q)} &= \mathrm{E}\left(\alpha_t^{(q)}|\mathcal{F}_{t-1}\right) \\
&= \mathrm{E}_{t-1}\left(\alpha_t^{(q)}\right) \\
&= \mathrm{E}_{t-1}\left(\mathbf{U}\alpha_t\right) \\
&= \mathbf{U}\,\mathrm{E}_{t-1}\left(\mathbf{T}\,\alpha_{t-1} + \mathbf{G}\gamma_{t-1}\right) \\
&= \mathbf{U}\,\mathbf{T}\,\alpha_{t-1}, \qquad (5.12)
\end{aligned}$$

where E_{t-1} denotes the expectation conditional on the information provided in $\mathcal{F}_{t-1}$.

The quasi-likelihood method can now be applied to the univariate BSM. The first step in this process is to determine a martingale difference (Lin, 2007, p1630) from the system of equations (5.3) to (5.4). All terms which are not disturbance terms are moved to the left hand side of the equation; the result is a vector:

$$\delta_t = \begin{pmatrix} Y_t - L_t - S_t \\ L_t - L_{t-1} - R_{t-1} \\ R_t - R_{t-1} \\ S_t + S_{t-1} + S_{t-2} + S_{t-3} \end{pmatrix} = \begin{pmatrix} \varepsilon_{U,\,t} \\ \eta_{U,\,t-1} \\ \zeta_{U,\,t-1} \\ \omega_{U,\,t-1} \end{pmatrix}. \qquad (5.13)$$

Thus, the martingale difference, for $t = 1, 2, \ldots, T$, may be written as

$$\delta_t = \begin{pmatrix} Y_t - \mathbf{Z}\mathbf{U}'\alpha_t^{(q)} \\ \alpha_t^{(q)} - \mathbf{U}\mathbf{T}\alpha_{t-1} \end{pmatrix} = \begin{pmatrix} \varepsilon_{U,\,t} \\ \alpha_t^{(q)} - \mathbf{a}_{t|t-1}^{(q)} \end{pmatrix}. \qquad (5.14)$$

The quasi-score estimating function $\mathbf{G}_t^*\left(\alpha_t^{(q)}\right)$ is given by (Lin, 2007, p1630):

$$\mathbf{G}_t^*\left(\alpha_t^{(q)}\right) = E_{t-1}\left(\frac{\partial\delta_t}{\partial\alpha_t^{(q)}}\right)\left[\mathrm{Var}_{t-1}(\delta_t)\right]^{-1}\delta_t. \qquad (5.15)$$

Let $\mathbf{B} = \frac{\partial\delta_t}{\partial\alpha_t^{(q)}}$, then

$$\mathbf{B} = \frac{\partial\delta_t}{\partial\alpha_t^{(q)}} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} = \left(-\mathbf{U}\mathbf{Z}',\ \ \mathbf{I}_3\right), \qquad (5.16)$$

where $\mathbf{I}_3$ is the 3×3 identity matrix. Also, let $\mathbf{V} = \mathrm{Var}_{t-1}(\delta_t)$ so that

$$\mathbf{V} = \mathrm{Var}_{t-1}(\delta_t) = \begin{pmatrix} \sigma^2_{U,\,\varepsilon} & 0 & 0 & 0 \\ 0 & \sigma^2_{U,\,\eta} & 0 & 0 \\ 0 & 0 & \sigma^2_{U,\,\zeta} & 0 \\ 0 & 0 & 0 & \sigma^2_{U,\,\omega} \end{pmatrix}. \tag{5.17}$$

The vector, δ_t, may then be rewritten to separate the components of $\alpha_t^{(q)}$,

$$\begin{aligned} \delta_t &= \begin{pmatrix} Y_t \\ -L_{t-1} - R_{t-1} \\ -R_{t-1} \\ S_{t-1} + S_{t-2} + S_{t-3} \end{pmatrix} + \begin{pmatrix} -1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} L_t \\ R_t \\ S_t \end{pmatrix} \\ &= \begin{pmatrix} Y_t \\ -\mathbf{a}^{(q)}_{t|t-1} \end{pmatrix} + \begin{pmatrix} -\mathbf{Z}\mathbf{U}' \\ \mathbf{I}_3 \end{pmatrix} \alpha_t^{(q)}. \end{aligned} \tag{5.18}$$

By letting $\mathbf{X}_1 = \begin{pmatrix} Y_t \\ -\mathbf{a}^{(q)}_{t|t-1} \end{pmatrix}$ in (5.18), the quasi-score estimating function, $\mathbf{G}_t^* \left(\alpha_t^{(q)}\right)$, given by (5.15) may be written as:

$$\begin{aligned} \mathbf{G}_t^* \left(\alpha_t^{(q)}\right) &= \mathbf{B}\mathbf{V}^{-1}\delta_t \\ &= \mathbf{B}\mathbf{V}^{-1}\left(\mathbf{X}_1 + \mathbf{B}'\alpha_t^{(q)}\right). \end{aligned} \tag{5.19}$$

The quasi-likelihood estimate, denoted by $\mathbf{a}_t^{(q)}$, of $\alpha_t^{(q)}$ is calculated by solving the equation $\mathbf{G}_t^* \left(\alpha_t^{(q)}\right) = 0$. The solution is

$$\mathbf{a}_t^{(q)} = -(\mathbf{B}\mathbf{V}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{V}^{-1}\mathbf{X}_1. \tag{5.20}$$

This estimate of the quasi-likelihood state vector is a function of the parameters of the model contained in $\mathbf{V}$ and the information contained in $\mathcal{F}^*_{t-1}$.

The error variance matrix of the state vector estimates can now be derived. If the conditional variance of $(\alpha_t^{(q)} - \mathbf{a}_t^{(q)})$ on the information set $\mathcal{F}_{t-1}$ is denoted by

$\mathbf{P}_t^{(q)}$ then,

$$\begin{aligned}\mathbf{P}_t^{(q)} &= E_{t-1}\left[\left(\alpha_t^{(q)} - \mathbf{a}_t^{(q)}\right)\left(\alpha_t^{(q)} - \mathbf{a}_t^{(q)}\right)'\right] \\ &= \mathrm{Var}_{t-1}\left[\alpha_t^{(q)} - \mathbf{a}_t^{(q)}\right] \\ &= \mathrm{Var}_{t-1}\left[\alpha_t^{(q)} - \mathcal{A}\mathbf{X}_1\right], \end{aligned} \tag{5.21}$$

where $\mathcal{A} = -(\mathbf{B}\mathbf{V}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{V}^{-1}$ from (5.20).

Now the vector $\mathbf{X}_1$ may be rewritten by substituting for Y_t, and then splitting it into parts:

$$\begin{aligned}\mathbf{X}_1 &= \begin{pmatrix} L_t + S_t + \varepsilon_{U,\,t} \\ -L_{t-1} - R_{t-1} \\ -R_{t-1} \\ S_{t-1} + S_{t-2} + S_{t-3} \end{pmatrix} \\ &= \begin{pmatrix} 1\ 0\ 1 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \end{pmatrix}\begin{pmatrix} L_t \\ R_t \\ S_t \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\varepsilon_{U,\,t} + \begin{pmatrix} 0 & 0\ 0\ 0\ 0 \\ -1 & -1\ 0\ 0\ 0 \\ 0 & -1\ 0\ 0\ 0 \\ 0 & 0\ 1\ 1\ 1 \end{pmatrix}\begin{pmatrix} L_{t-1} \\ R_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Z}\mathbf{U}' \\ \mathbf{0}_{(3\times 3)} \end{pmatrix}\alpha_t^{(q)} + \begin{pmatrix} 1 \\ \mathbf{0}_{(3\times 1)} \end{pmatrix}\varepsilon_{U,\,t} + \begin{pmatrix} \mathbf{0}_{(1\times 5)} \\ -\mathbf{U}\mathbf{T} \end{pmatrix}\alpha_{t-1} \\ &= \mathbf{X}_a\alpha_t^{(q)} + \mathbf{X}_b\varepsilon_{U,\,t} + \mathbf{X}_c\alpha_{t-1}, \end{aligned} \tag{5.22}$$

where $\mathbf{X}_a$ denotes the coefficient matrix of $\alpha_t^{(q)}$, $\mathbf{X}_b$ denotes the coefficient vector of $\varepsilon_{U,\,t}$, and $\mathbf{X}_c$ denotes the coefficient matrix of α_{t-1}. Thus, the error variance matrix associated with the quasi-likelihood state vector estimator may be determined as follows:

$$\begin{aligned}\mathrm{Var}_{t-1}\left[\alpha_t^{(q)} - \mathcal{A}\mathbf{X}_1\right] &= \mathrm{Var}_{t-1}\left[\mathbf{I}_3\alpha_t^{(q)} - \mathcal{A}\mathbf{X}_a\alpha_t^{(q)} - \mathcal{A}\mathbf{X}_b\varepsilon_{U,\,t} - \mathcal{A}\mathbf{X}_c\alpha_{t-1}\right] \\ &= \mathrm{Var}_{t-1}\left[(\mathbf{I}_3 - \mathcal{A}\mathbf{X}_a)\alpha_t^{(q)} - \mathcal{A}\mathbf{X}_b\varepsilon_{U,\,t} - \mathcal{A}\mathbf{X}_c\alpha_{t-1}\right] \\ &= (\mathbf{I}_3 - \mathcal{A}\mathbf{X}_a)\mathrm{Var}_{t-1}\left(\alpha_t^{(q)}\right)(\mathbf{I}_3 - \mathcal{A}\mathbf{X}_a)' \\ &\quad + \mathcal{A}\mathbf{X}_b\mathrm{Var}_{t-1}(\varepsilon_{U,\,t})(\mathcal{A}\mathbf{X}_b)'. \end{aligned} \tag{5.23}$$

From Section 3.4.1, the variance of the measurement error can be written as $\mathrm{Var}_{t-1}(\varepsilon_{U,\,t}) = \sigma^2_{U,\,\varepsilon}$ and also

$$\mathrm{Var}_{t-1}\left(\alpha_t^{(q)}\right) = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 & 0 \\ 0 & \sigma^2_{U,\,\zeta} & 0 \\ 0 & 0 & \sigma^2_{U,\,\omega} \end{pmatrix}. \tag{5.24}$$

Therefore, the error variance matrix for the univariate quasi-likelihood method is given by:

$$\mathbf{P}_t^{(q)} = \begin{pmatrix} \frac{\sigma^2_{U,\,\eta}(\sigma^2_{U,\,\varepsilon}+\sigma^2_{U,\,\omega})}{\sigma^2_{U,\,\eta}+\sigma^2_{U,\,\varepsilon}+\sigma^2_{U,\,\omega}} & 0 & \frac{-\sigma^2_{U,\,\eta}\sigma^2_{U,\,\omega}}{\sigma^2_{U,\,\eta}+\sigma^2_{U,\,\varepsilon}+\sigma^2_{U,\,\omega}} \\ 0 & \sigma^2_{U,\,\zeta} & 0 \\ \frac{-\sigma^2_{U,\,\eta}\sigma^2_{U,\,\omega}}{\sigma^2_{U,\,\eta}+\sigma^2_{U,\,\varepsilon}+\sigma^2_{U,\,\omega}} & 0 & \frac{\sigma^2_{U,\,\omega}(\sigma^2_{U,\,\eta}+\sigma^2_{U,\,\varepsilon})}{\sigma^2_{U,\,\eta}+\sigma^2_{U,\,\varepsilon}+\sigma^2_{U,\,\omega}} \end{pmatrix}. \tag{5.25}$$

The third diagonal element of the above matrix is the error variance associated with the quasi-likelihood seasonal component estimate for the aggregate series, given the information $\mathcal{F}_{t-1}$:

$$\mathrm{Var}_{t-1}\left(S_t^{(q)} - \hat{S}_t^{(q)}\right) = \frac{\sigma^2_{U,\,\omega}(\sigma^2_{U,\,\eta} + \sigma^2_{U,\,\varepsilon})}{\sigma^2_{U,\,\eta} + \sigma^2_{U,\,\varepsilon} + \sigma^2_{U,\,\omega}}. \tag{5.26}$$

This expression becomes the numerator of the indicator measure for the relative efficiency of the univariate and multivariate methods. As all the terms on the right-hand side of (5.26) are free from t, $\mathrm{Var}_{t-1}\left(S_t^{(q)} - \hat{S}_t^{(q)}\right)$ does not depend on t.

5.4 Quasi-Likelihood Approach to the Multivariate BSM

In this section, the quasi-likelihood method is used to derive the state vector for the multivariate system with two sub-series, along with its error variance matrix. The element associated with the seasonal component for the aggregate series will become the denominator of the indicator measure for the relative efficiency of the univariate and multivariate methods.

The multivariate BSM with level, slope and quarterly dummy seasonal components described in Section 3.3, may be rewritten with the one-step transition going from $t-1$ to t. For $K = 2$, the model becomes

$$Y_{1t} = L_{1t} + S_{1t} + \varepsilon_t + \varepsilon^*_{1t}, \qquad \mathrm{E}(\varepsilon^*_{1t}) = 0, \quad \mathrm{Var}(\varepsilon^*_{1t}) = \sigma^2_{1\varepsilon^*}, \tag{5.27}$$

$$Y_{2t} = L_{2t} + S_{2t} + \varepsilon_t + \varepsilon^*_{2t}, \qquad \mathrm{E}(\varepsilon^*_{2t}) = 0, \quad \mathrm{Var}(\varepsilon^*_{2t}) = \sigma^2_{2\varepsilon^*}, \tag{5.28}$$

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{Var}(\varepsilon_t) = \sigma^2_{\varepsilon},$$

$$\begin{aligned}
L_{1t} &= L_{1,\,t-1} + R_{1,\,t-1} + \eta_{t-1} + \eta^*_{1,\,t-1}, && \mathrm{E}(\eta^*_{1,\,t}) = 0, \quad \mathrm{Var}(\eta^*_{1,\,t}) = \sigma^2_{1\eta^*},\\
L_{2t} &= L_{2,\,t-1} + R_{2,\,t-1} + \eta_{t-1} + \eta^*_{2,\,t-1}, && \mathrm{E}(\eta^*_{2,\,t}) = 0, \quad \mathrm{Var}(\eta^*_{2,\,t}) = \sigma^2_{2\eta^*},\\
& && \mathrm{E}(\eta_t) = 0, \quad \mathrm{Var}(\eta_t) = \sigma^2_{\eta},\\
R_{1t} &= R_{1,\,t-1} + \zeta_{t-1} + \zeta^*_{1,\,t-1}, && \mathrm{E}(\zeta^*_{1,\,t}) = 0, \quad \mathrm{Var}(\zeta^*_{1,\,t}) = \sigma^2_{1\zeta^*},\\
R_{2t} &= R_{2,\,t-1} + \zeta_{t-1} + \zeta^*_{2,\,t-1}, && \mathrm{E}(\zeta^*_{2,\,t}) = 0, \quad \mathrm{Var}(\zeta^*_{2,\,t}) = \sigma^2_{2\zeta^*},\\
& && \mathrm{E}(\zeta_t) = 0, \quad \mathrm{Var}(\zeta_t) = \sigma^2_{\zeta},\\
S_{1t} &= -\sum_{j=1}^{s-1} S_{1,\,t-j} + \omega_{t-1} + \omega^*_{1,\,t-1}, && \mathrm{E}(\omega^*_{1,\,t}) = 0, \quad \mathrm{Var}(\omega^*_{1,\,t}) = \sigma^2_{1\omega^*},\\
S_{2t} &= -\sum_{j=1}^{s-1} S_{2,\,t-j} + \omega_{t-1} + \omega^*_{2,\,t-1}, && \mathrm{E}(\omega^*_{2,\,t}) = 0, \quad \mathrm{Var}(\omega^*_{2,\,t}) = \sigma^2_{2\omega^*},\\
& && \mathrm{E}(\omega_t) = 0, \quad \mathrm{Var}(\omega_t) = \sigma^2_{\omega},
\end{aligned}$$

for $t = 1, \ldots, T$.

In state space form, the state vector contains the components for each series as well as the measurement errors because standard software does not allow for correlated measurement errors (see Section 3.4.2). For the quasi-likelihood method, this adjustment is not necessary. So the state space form for $K = 2$ may be written with $(m*)$ subscripts as

$$\mathbf{Y}_{(m*),\,t} = (\mathbf{Z} \otimes \mathbf{I}_2)\alpha_{(m*),\,t} + \mathcal{E}_{(m*),\,t}, \tag{5.29}$$

$$\alpha_{(m*),\,t} = (\mathbf{T} \otimes \mathbf{I}_2)\alpha_{(m*),\,t-1} + (\mathbf{G} \otimes \mathbf{I}_2)\gamma_{(m*),\,t-1}, \tag{5.30}$$

where the system matrices $\mathbf{Z}$, $\mathbf{T}$ and $\mathbf{G}$ are the same as those for the univariate state space form (3.27). For this model, $\mathbf{Z}$, $\mathbf{T}$ and $\mathbf{G}$ have respective dimensions (1×5), (5×5) and (5×3). Separating the measurement errors from the state vector

requires adjustment of the system vectors:

$$\begin{aligned}
\mathbf{Y}_{(m*),\,t} &= [Y_{1t}, Y_{2t}]^{'}, \\
\alpha_{(m*),\,t} &= [L_{1t}, L_{2t},\ R_{1t}, R_{2t},\ S_{1t}, S_{2t},\ S_{1,\,t-1}, S_{2,\,t-1},\ S_{1,\,t-2}, S_{2,\,t-2}]^{'}, \\
\gamma_{(m*),\,t} &= [\ (\eta_t + \eta^*_{1t}), (\eta_t + \eta^*_{2t}),\ (\zeta_t + \zeta^*_{1t}), (\zeta_t + \zeta^*_{2t}),\ (\omega_t + \omega^*_{1t}), (\omega_t + \omega^*_{2t})\ ]^{'}, \\
\mathcal{E}_{(m*),\,t} &= \left[(\varepsilon_t + \varepsilon^*_{1,\,t}), (\varepsilon_t + \varepsilon^*_{2,\,t})\ \right]^{'}. \qquad (5.31)
\end{aligned}$$

5.4.1 Transforming the Multivariate System

Following the methodology described in Section 3.4.3, the transformation is applied to the multivariate system above with $K = 2$. The transformation matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}. \qquad (5.32)$$

The measurement equation (5.29) for the transformed state space system is now written:

$$\begin{aligned}
\mathbf{A}\mathbf{Y}_{(m*),\,t} &= \mathbf{A}(\mathbf{Z} \otimes \mathbf{I}_2)\alpha_{(m*),\,t} + \mathbf{A}\mathcal{E}_{(m*),\,t}, \\
&= (\mathbf{Z} \otimes \mathbf{I}_2)(\mathbf{I}_5 \otimes \mathbf{A})\alpha_{(m*),\,t} + \mathbf{A}\mathcal{E}_{(m*),\,t}. \qquad (5.33)
\end{aligned}$$

Letting $\mathbf{Y}_{(M),\,t} = \mathbf{A}\mathbf{Y}_{(m*),\,t}$ and simplifying remaining terms with the (M) subscript, the measurement equation (5.33) becomes:

$$\mathbf{Y}_{(M),\,t} = (\mathbf{Z} \otimes \mathbf{I}_2)\alpha_{(M),\,t} + \mathcal{E}_{(M),\,t}. \qquad (5.34)$$

The transition equation for the transformed state space system becomes

$$(\mathbf{I}_5 \otimes \mathbf{A})\alpha_{(m*),\,t} = (\mathbf{T} \otimes \mathbf{I}_2)(\mathbf{I}_5 \otimes \mathbf{A})\alpha_{(m*),\,t-1} + (\mathbf{I}_5 \otimes \mathbf{A})(\mathbf{G} \otimes \mathbf{I}_2)\gamma_{(m*),\,t-1},$$

and denoting $(\mathbf{I}_5 \otimes \mathbf{A})\alpha_{(m*),\,t}$ by $\alpha_{(M),\,t}$, this becomes

$$\alpha_{(M),\,t} = (\mathbf{T} \otimes \mathbf{I}_2)\alpha_{(M),\,t-1} + (\mathbf{I}_5 \otimes \mathbf{A})(\mathbf{G} \otimes \mathbf{I}_2)\gamma_{(m*),\,t-1}. \qquad (5.35)$$

The state vector for the multivariate system, $\alpha_{(M),\,t}$, requires modification similar to that for the univariate quasi-likelihood method. In the definition of the dummy

seasonal variable, only one season is defined to be stochastic. This means that the last four components in $\alpha_{(M),\,t}$, namely $S_{tot,\,t-1}$, $S_{1,\,t-1}$, $S_{tot,\,t-2}$ and $S_{1,\,t-2}$, do not have disturbance terms associated with them. The state vector is amended by multiplying it by the selection matrix appropriate for the $K = 2$ case, which is denoted by $\mathbf{U}_{(M)}$,

$$\mathbf{U}_{(M)} = \mathbf{U} \otimes \mathbf{I}_2, \tag{5.36}$$

where $\mathbf{U}$ is the selection matrix for the univariate case given in (5.8). Therefore, the multivariate state vector modified for the quasi-likelihood method is given by:

$$\begin{aligned} \alpha^{(q)}_{(M),\,t} &= \mathbf{U}_{(M)}\alpha_{(M),\,t} \\ &= \left[L_{tot,\,t}, L_{1t},\ R_{tot,\,t}, R_{1t},\ S_{tot,\,t}, S_{1t}\right]'. \end{aligned} \tag{5.37}$$

The notation for the multivariate quasi-likelihood approach is a simple extension of the notation for the univariate approach, as developed in the previous section. In the quasi-likelihood (QL) approach, the aim is to find a one-step optimal estimator of $\alpha^{(q)}_{(M),\,t}$ provided by the following set of information:

$$\begin{aligned} \mathcal{F}^{*}_{(M),\,t-1} = \ \{&Y_{11},\ Y_{12}, \ldots, Y_{1,\,t-1},\ Y_{1t}, \ldots, Y_{K1},\ Y_{K2}, \ldots, Y_{K,\,t-1},\ Y_{Kt}; \\ &L_{11},\ L_{12}, \ldots, L_{1,\,t-1}, \ldots,\ L_{K1}, \ldots, L_{K,\,t-1}; \\ &R_{11},\ R_{12}, \ldots, R_{1,\,t-1}, \ldots,\ R_{K1}, \ldots, R_{K,\,t-1}; \\ &S_{11},\ S_{12}, \ldots, S_{1,\,t-1}, \ldots,\ S_{K1}, \ldots, S_{K,\,t-1}\}. \end{aligned} \tag{5.38}$$

If the observations $Y_{1t}, \ldots, Y_{Kt}$ are excluded from (5.38), the result is the information denoted by $\mathcal{F}_{(M),\,t-1}$. Let $\mathbf{a}^{(q)}_{(M),\,t}$ denote the optimal estimator of $\alpha^{(q)}_{(M),\,t}$ based on the information $\mathcal{F}^{*}_{(M),\,t-1}$. Also, let $\mathbf{a}^{(q)}_{(M),\,t|t-1}$ denote the optimal estimator of $\alpha^{(q)}_{(M),\,t}$ based on $\mathcal{F}_{(M),\,t-1}$, which defines the conditional expectation of $\alpha^{(q)}_{(M),\,t}$:

$$\begin{aligned} \mathbf{a}^{(q)}_{(M),\,t|t-1} &= \mathrm{E}_{t-1}\left(\alpha^{(q)}_{(M),\,t}\right) \\ &= \mathrm{E}_{t-1}\left(\mathbf{U}_{(M)}\alpha_{(M),\,t}\right) \\ &= \mathbf{U}_{(M)}\left(\mathbf{T} \otimes \mathbf{I}_2\right)\alpha_{(M),\,t-1}. \end{aligned} \tag{5.39}$$

The martingale difference for the transformed multivariate system is determined by separating the disturbance terms as follows:

$$\delta_{(M),\,t} = \begin{pmatrix} Y_{tot,\,t} - L_{tot,\,t} - S_{tot,\,t} \\ Y_{1t} - L_{1t} - S_{1t} \\ L_{tot,\,t} - L_{tot,\,t-1} - R_{tot,\,t-1} \\ L_{1t} - L_{1,\,t-1} - R_{1,\,t-1} \\ R_{tot,\,t} - R_{tot,\,t-1} \\ R_{1t} - R_{1,\,t-1} \\ S_{tot,\,t} + \sum_{j=1}^{s-1} S_{tot,\,t-j} \\ S_{1t} + \sum_{j=1}^{s-1} S_{1,\,t-j} \end{pmatrix} = \begin{pmatrix} 2\varepsilon_t + \varepsilon^*_{1t} + \varepsilon^*_{2t} \\ \varepsilon_t + \varepsilon^*_{1t} \\ 2\eta_{t-1} + \eta^*_{1,\,t-1} + \eta^*_{2,\,t-1} \\ \eta_{t-1} + \eta^*_{1,\,t-1} \\ 2\zeta_{t-1} + \zeta^*_{1,\,t-1} + \zeta^*_{2,\,t-1} \\ \zeta_{t-1} + \zeta^*_{1,\,t-1} \\ 2\omega_{t-1} + \omega^*_{1,\,t-1} + \omega^*_{2,\,t-1} \\ \omega_{t-1} + \omega^*_{1,\,t-1} \end{pmatrix}. \tag{5.40}$$

This can be written in matrix notation as

$$\begin{aligned} \delta_{(M),\,t} &= \begin{pmatrix} \mathbf{A}\mathcal{E}_{(m*),\,t} \\ \mathbf{U}_{(M)}\alpha_{(M),\,t} - \mathbf{U}_{(M)}\mathbf{a}_{(M),\,t|t-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}\mathbf{Y}_{(m*),\,t} - (\mathbf{Z}\otimes\mathbf{I}_2)(\mathbf{I}_5\otimes\mathbf{A})\alpha_{(m*),\,t} \\ \mathbf{U}_{(M)}(\mathbf{I}_5\otimes\mathbf{A})\alpha_{(m*),\,t} - \mathbf{U}_{(M)}(\mathbf{T}\otimes\mathbf{I}_2)(\mathbf{I}_5\otimes\mathbf{A})\alpha_{(m*),\,t-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}\mathbf{Y}_{(m*),\,t} - (\mathbf{Z}\otimes\mathbf{I}_2)\mathbf{U}'_{(M)}\alpha^{(q)}_{(M),\,t} \\ \alpha^{(q)}_{(M),\,t} - \mathbf{U}_{(M)}(\mathbf{T}\otimes\mathbf{I}_2)(\mathbf{I}_5\otimes\mathbf{A})\alpha_{(m*),\,t-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}\mathbf{Y}_{(m*),\,t} \\ -\mathbf{U}_{(M)}(\mathbf{T}\otimes\mathbf{I}_2)(\mathbf{I}_5\otimes\mathbf{A})\alpha_{(m*),\,t-1} \end{pmatrix} + \begin{pmatrix} -(\mathbf{Z}\otimes\mathbf{I}_2)\mathbf{U}'_{(M)} \\ \mathbf{I}_6 \end{pmatrix} \alpha^{(q)}_{(M),\,t}. \end{aligned} \tag{5.41}$$

The quasi-score estimating function for the multivariate method is given by (Lin, 2007, p1630):

$$\mathbf{G}^*_t\left(\alpha^{(q)}_{(M),\,t}\right) = E_{t-1}\left(\frac{\partial\delta_{(M),\,t}}{\partial\alpha^{(q)}_{(M),\,t}}\right)[\mathrm{Var}_{t-1}(\delta_{(M),\,t})]^{-1}\delta_{(M),\,t}. \tag{5.42}$$

The derivative of $\delta_{(M),\,t}$ with respect to the quasi-likelihood state vector is denoted

by $\mathbf{B}_{(M)}$:

$$\begin{aligned}\mathbf{B}_{(M)} &= \frac{\partial \delta_{(M),\,t}}{\partial \alpha^{(q)}_{(M),\,t}} \\ &= \left(-\mathbf{U}_{(M)}(\mathbf{Z}' \otimes \mathbf{I}_2),\ \mathbf{I}_6\right) \\ &= (-\mathbf{U}\mathbf{Z}',\ \mathbf{I}_3) \otimes \mathbf{I}_2 \\ &= \mathbf{B} \otimes \mathbf{I}_2. \end{aligned} \tag{5.43}$$

Referring to (5.40) and to Section 3.4.3, in this case the variance term is:

$$\mathbf{V}_{(M)} = \mathrm{Var}_{t-1}(\delta_{(M),\,t}) = \begin{pmatrix} \boldsymbol{\Sigma}_{(M),\,\varepsilon} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \boldsymbol{\Sigma}_{(M),\,\eta} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \boldsymbol{\Sigma}_{(M),\,\zeta} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \boldsymbol{\Sigma}_{(M),\,\omega} \end{pmatrix}. \tag{5.44}$$

Since $K = 2$, the covariance matrix for the level component is given by

$$\boldsymbol{\Sigma}_{(M),\,\eta} = \begin{pmatrix} \sigma^2_{tot,\eta} & 2\sigma^2_{\eta} + \sigma^2_{1\eta^*} \\ 2\sigma^2_{\eta} + \sigma^2_{1\eta^*} & \sigma^2_{\eta} + \sigma^2_{1\eta^*} \end{pmatrix}, \tag{5.45}$$

with $\sigma^2_{tot,\,\eta} = 4\sigma^2_{\eta} + \sigma^2_{1\eta^*} + \sigma^2_{2\eta^*}$. Similarly for $\boldsymbol{\Sigma}_{(M),\,\zeta}$, $\boldsymbol{\Sigma}_{(M),\,\omega}$ and $\boldsymbol{\Sigma}_{(M),\,\varepsilon}$.

Let the first term on the right hand side of (5.41) be $\mathbf{X}_{(M),\,1}$. Then $\delta_{(M),\,t}$ may be written as

$$\delta_{(M),\,t} = \mathbf{X}_{(M),\,1} + \mathbf{B}'_{(M)}\alpha^{(q)}_{(M),\,t}. \tag{5.46}$$

The quasi-score estimating function for the multivariate case is therefore given by:

$$\begin{aligned}\mathbf{G}^*_t\left(\alpha^{(q)}_{(M),\,t}\right) &= \mathbf{B}_{(M)}\mathbf{V}^{-1}_{(M)}\delta_{(M),\,t} \\ &= \mathbf{B}_{(M)}\mathbf{V}^{-1}_{(M)}\left(\mathbf{X}_{(M),\,1} + \mathbf{B}'_{(M)}\alpha^{(q)}_{(M),\,t}\right). \end{aligned} \tag{5.47}$$

The quasi-likelihood estimate of the multivariate state vector is the solution of the equation $\mathbf{G}_t^* \left(\alpha_{(M),\,t}^{(q)} \right) = 0$:

$$\begin{aligned} \mathbf{a}_{(M),\,t}^{(q)} &= -(\mathbf{B}_{(M)} \mathbf{V}_{(M)}^{-1} \mathbf{B}_{(M)}^{'})^{-1} \mathbf{B}_{(M)} \mathbf{V}_{(M)}^{-1} \mathbf{X}_{(M),\,1} \\ &= \mathcal{A}_{(M)} \mathbf{X}_{(M),\,1}, \end{aligned} \tag{5.48}$$

where $\mathcal{A}_{(M)} = - \left(\mathbf{B}_{(M)} \mathbf{V}_{(M)}^{-1} \mathbf{B}_{(M)}^{'} \right)^{-1} \mathbf{B}_{(M)} \mathbf{V}_{(M)}^{-1}$. This produces an estimator of the quasi-likelihood state vector which is a function of the parameters of the multivariate model which are contained in $\mathbf{V}_{(M)}$, and of the information $\mathcal{F}_{(M),\,t-1}^*$.

To determine the denominator of the quasi-likelihood indicator for the relative efficiency of the univariate and multivariate methods, the error variance matrix of the multivariate quasi-likelihood estimate of the components is required. The conditional variance of $\left(\alpha_{(M),\,t}^{(q)} - \mathbf{a}_{(M),\,t}^{(q)} \right)$ is written as:

$$\begin{aligned} \mathbf{P}_{(M),\,t}^{(q)} &= \mathrm{Var}_{t-1} \left(\alpha_{(M),\,t}^{(q)} - \mathbf{a}_{(M),\,t}^{(q)} \right) \\ &= \mathrm{Var}_{t-1} \left(\alpha_{(M),\,t}^{(q)} - \mathcal{A}_{(M)} \mathbf{X}_{(M),\,1} \right). \end{aligned} \tag{5.49}$$

The vector $\mathbf{X}_{(M),\,1}$ may be rewritten by substituting for $Y_{tot,\,t}$ and Y_{1t},

$$\mathbf{X}_{(M),\,1} = \begin{pmatrix} L_{tot,\,t} + S_{tot,\,t} + 2\varepsilon_t + \varepsilon_{1t}^* + \varepsilon_{2t}^* \\ L_{1t} + S_{1t} + \varepsilon_t + \varepsilon_{1t}^* \\ -L_{tot,\,t-1} - R_{tot,\,t-1} \\ -L_{1,\,t-1} - R_{1,\,t-1} \\ -R_{tot,\,t-1} \\ -R_{1,\,t-1} \\ S_{tot,\,t-1} + S_{tot,\,t-2} + S_{tot,\,t-3} \\ S_{1,\,t-1} + S_{1,\,t-2} + S_{1,\,t-3} \end{pmatrix}. \tag{5.50}$$

Splitting this into three parts,

$$\mathbf{X}_{(M),\,1} = \begin{pmatrix} 1&0&0&0&1&0 \\ 0&1&0&0&0&1 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \end{pmatrix} \begin{pmatrix} L_{tot,\,t} \\ L_{1t} \\ R_{tot,\,t} \\ R_{1t} \\ S_{tot,\,t} \\ S_{1t} \end{pmatrix} + \begin{pmatrix} 1&0 \\ 0&1 \\ 0&0 \\ 0&0 \\ 0&0 \\ 0&0 \\ 0&0 \\ 0&0 \end{pmatrix} \begin{pmatrix} 2\varepsilon_t + \varepsilon^*_{1t} + \varepsilon^*_{2t} \\ \varepsilon_t + \varepsilon^*_{1t} \end{pmatrix}$$

$$+ \begin{pmatrix} 0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0 \\ -1&0&-1&0&0&0&0&0&0 \\ 0&-1&0&-1&0&0&0&0&0 \\ 0&0&-1&0&0&0&0&0&0 \\ 0&0&0&-1&0&0&0&0&0 \\ 0&0&0&0&1&0&1&0&1 \\ 0&0&0&0&0&1&0&1&0 \end{pmatrix} \begin{pmatrix} L_{tot,\,t-1} \\ L_{1,\,t-1} \\ R_{tot,\,t-1} \\ R_{1,\,t-1} \\ S_{tot,\,t-1} \\ S_{1,\,t-1} \\ S_{tot,\,t-2} \\ S_{1,\,t-2} \\ S_{tot,\,t-3} \\ S_{1,\,t-3} \end{pmatrix}. \tag{5.51}$$

In matrix notation, and using the same vectors obtained in the univariate solution for $\mathbf{X}_1$ in (5.22), this may be written as

$$\begin{aligned} \mathbf{X}_{(M),\,1} &= \left[\begin{pmatrix} \mathbf{Z}\mathbf{U}' \\ \mathbf{0}_{(3\times 3)} \end{pmatrix} \otimes \mathbf{I}_2 \right] \alpha^{(q)}_{(M),\,t} + \left[\begin{pmatrix} 1 \\ \mathbf{0}_{(3\times 1)} \end{pmatrix} \otimes \mathbf{I}_2 \right] \mathcal{E}_{(M),\,t} \\ &\quad + \left[\begin{pmatrix} \mathbf{0}_{(1\times 5)} \\ -\mathbf{U}\mathbf{T} \end{pmatrix} \otimes \mathbf{I}_2 \right] \alpha_{(M),\,t-1} \\ &= (\mathbf{X}_a \otimes \mathbf{I}_2)\, \alpha^{(q)}_{(M),\,t} + (\mathbf{X}_b \otimes \mathbf{I}_2)\, \mathcal{E}_{(M),\,t} + (\mathbf{X}_c \otimes \mathbf{I}_2)\, \alpha_{(M),\,t-1} \\ &= \mathbf{X}_{(M),\,a}\, \alpha^{(q)}_{(M),\,t} + \mathbf{X}_{(M),\,b}\, \mathcal{E}_{(M),\,t} + \mathbf{X}_{(M),\,c}\, \alpha_{(M),\,t-1}. \end{aligned} \tag{5.52}$$

where $\mathbf{X}_{(M),\,a} = \mathbf{X}_a \otimes \mathbf{I}_2$, and similarly for $\mathbf{X}_{(M),\,b}$ and $\mathbf{X}_{(M),\,c}$.

The error variance matrix associated with the quasi-likelihood multivariate state vector estimator may be determined as follows:

$$\begin{aligned}\mathbf{P}^{(q)}_{(M),\,t} &= \mathrm{Var}_{t-1}\left(\alpha^{(q)}_{(M),\,t} - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,1}\right) \\ &= \left[\mathbf{I}_6 - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,a}\right]\ \mathrm{Var}_{t-1}\left(\alpha^{(q)}_{(M),\,t}\right)\ \left[\mathbf{I}_6 - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,a}\right]' \\ &\quad + \mathcal{A}_{(M)}\mathbf{X}_{(M),\,b}\ \mathrm{Var}_{t-1}\left(\mathcal{E}_{(M),\,t}\right)\ \left(\mathcal{A}_{(M)}\mathbf{X}_{(M),\,b}\right)'. \end{aligned} \tag{5.53}$$

Therefore, the element corresponding to the variance of the seasonal component estimate of the aggregate series, $\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}\right)$, is the fifth diagonal element (since $K = 2$) of the matrix $\mathbf{P}^{(q)}_{(M),\,t}$.

An alternative to calculating the entire error variance matrix is to partition the $\mathcal{A}_{(M)}$ matrix and just do the calculations for the block of seasonal elements. For the $K = 2$ case, this block has dimensions 2×2. The 6×8 matrix $\mathcal{A}_{(M)}$ in (5.48) is given by:

$$\mathcal{A}_{(M)} = -\left(\mathbf{B}_{(M)}\mathbf{V}^{-1}_{(M)}\mathbf{B}'_{(M)}\right)^{-1}\ \mathbf{B}_{(M)}\mathbf{V}^{-1}_{(M)}, \tag{5.54}$$

and can be partitioned into 2×2 matrices such that

$$\mathcal{A}_{(M)} = \begin{pmatrix} \mathcal{A}_{11}\ \mathcal{A}_{12}\ \mathcal{A}_{13}\ \mathcal{A}_{14} \\ \mathcal{A}_{21}\ \mathcal{A}_{22}\ \mathcal{A}_{23}\ \mathcal{A}_{24} \\ \mathcal{A}_{31}\ \mathcal{A}_{32}\ \mathcal{A}_{33}\ \mathcal{A}_{34} \end{pmatrix}. \tag{5.55}$$

More specifically, this partitioned matrix can be used to obtain the results of (5.53) which correspond to the seasonal components only. The result is a 2×2 matrix:

$$\begin{aligned} &\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}, S^{(q)}_{1t} - \hat{S}^{(q)}_{1t}\right) \\ &= \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\eta}\mathcal{A}'_{31} + (\mathbf{I}_2 - \mathcal{A}_{31})\ \mathbf{\Sigma}_{(M),\,\omega}\ (\mathbf{I}_2 - \mathcal{A}_{31})' + \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}'_{31} \\ &= \mathcal{A}_{31}\left[\mathbf{\Sigma}_{(M),\,\eta} + \mathbf{\Sigma}_{(M),\,\varepsilon}\right]\mathcal{A}'_{31} + (\mathbf{I}_2 - \mathcal{A}_{31})\ \mathbf{\Sigma}_{(M),\,\omega}\ (\mathbf{I}_2 - \mathcal{A}_{31})'. \end{aligned} \tag{5.56}$$

The first diagonal element of the above matrix will be the variance of the seasonal component for the aggregated series given the two sub-series. For details of the derivation, see Appendix C.1.

To abbreviate the expression for the solution of the first element of (5.56), let

$$\begin{aligned} \sigma_\eta^2 + \sigma_{1\eta^*}^2 &= \mathrm{n}_1, & \sigma_\omega^2 + \sigma_{1\omega^*}^2 &= \mathrm{w}_1, & \sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 &= \mathrm{e}_1, \\ \sigma_\eta^2 + \sigma_{2\eta^*}^2 &= \mathrm{n}_2, & \sigma_\omega^2 + \sigma_{2\omega^*}^2 &= \mathrm{w}_2, & \sigma_\varepsilon^2 + \sigma_{2\varepsilon^*}^2 &= \mathrm{e}_2, \\ \sigma_{tot,\,\eta}^2 &= \mathrm{N}, & \sigma_{tot,\,\omega}^2 &= \mathrm{W}, & \sigma_{tot,\,\varepsilon}^2 &= \mathrm{E}. \end{aligned} \tag{5.57}$$

Then, the error variance of the estimate of the seasonal component of the aggregated series, using the quasi-likelihood method, is given by equation (5.58) below. See Appendix C.2 for details of the calculation.

$$\begin{aligned} &\mathrm{Var}_{t-1}\left(S_{tot,\,t}^{(q)} - \hat{S}_{tot,\,t}^{(q)}\right) \\ &\qquad = \frac{\mathrm{W}\left(\left[\sigma_\eta^2 + \sigma_\varepsilon^2\right]^2 - \left[\mathrm{n}_1 + \mathrm{e}_1\right]\left[\mathrm{n}_2 + \mathrm{e}_2\right]\right) + (\mathrm{N} + \mathrm{E})\left(\sigma_\omega^4 - \mathrm{w}_1\mathrm{w}_2\right)}{\left(\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\omega^2\right)^2 - \left(\mathrm{n}_1 + \mathrm{e}_1 + \mathrm{w}_1\right)\left(\mathrm{n}_2 + \mathrm{e}_2 + \mathrm{w}_2\right)}. \end{aligned} \tag{5.58}$$

This expression forms the denominator of the quasi-likelihood indicator measure.

5.5 The Quasi-Likelihood Indicator

The focus of this chapter up to this point has been on deriving the one-step ahead result for the conditional error variance of the estimator of the seasonal component of the aggregated series using the quasi-likelihood approach. This has been achieved for both the univariate model and the multivariate model with two time series. The purpose of the derivations is to determine an indicator measure (denoted by Q) which does not rely on an iterative process or on t, and can be used to predict the relative efficiency determined by the Kalman filter (3.59). Thus, Q will provide an indication of when the multivariate model, using the disaggregated series is more efficient than the univariate model.

The quasi-likelihood indicator (Q) is defined as the relative efficiency of the conditional error variance of the estimator of the seasonal component at time t, given by the univariate and multivariate QL methods. It is given by the ratio of the

expressions in (5.26) and (5.58):

$$Q = \frac{\mathrm{Var}_{t-1}\left(S_t^{(q)} - \hat{S}_t^{(q)}\right)}{\mathrm{Var}_{t-1}\left(S_{tot,\,t}^{(q)} - \hat{S}_{tot,\,t}^{(q)}\right)}. \tag{5.59}$$

With reference to the abbreviations in (5.57), and for $K = 2$, Q is given by:

$$Q = \frac{\sigma^2_{U,\,\omega}\left(\sigma^2_{U,\,\eta} + \sigma^2_{U,\,\varepsilon}\right)\left[\left(\sigma^2_\eta + \sigma^2_\varepsilon + \sigma^2_\omega\right)^2 - (\mathrm{n}_1 + \mathrm{e}_1 + \mathrm{w}_1)(\mathrm{n}_2 + \mathrm{e}_2 + \mathrm{w}_2)\right]}{\left[\sigma^2_{U,\,\eta} + \sigma^2_{U,\,\varepsilon} + \sigma^2_{U,\,\omega}\right]\left[\mathrm{W}\left(\left[\sigma^2_\eta + \sigma^2_\varepsilon\right]^2 - [\mathrm{n}_1 + \mathrm{e}_1][\mathrm{n}_2 + \mathrm{e}_2]\right) + (\mathrm{N} + \mathrm{E})\left(\sigma^4_\omega - \mathrm{w}_1\mathrm{w}_2\right)\right]}. \tag{5.60}$$

Note that the expression for Q involves all the parameters except those for the slope component, even though a slope component was included in the BSM for both the univariate and multivariate models. The solution for Q is the same whether or not the model contains a slope component. The solution for Q therefore applies to the local level seasonal model as described in Section 4.2.

It is also important to note that Q does not depend on t. Unlike the relative efficiency measure derived from the Kalman filter, the QL indicator does not change for different values of t. As it produces a single quantitative value, Q can be used as an indicator for the KF relative frequency. It is not intended here to replace the Kalman filter estimates of the state vector with the quasi-likelihood estimates but to use the calculated value of Q as a quick predictor tool only.

5.6 Comparison with Kalman Filter Results

In this section, the value of the quasi-likelihood indicator, Q, is compared to the relative efficiency determined by the Kalman filter, $RE_t(M)$ (refer to (3.59)). It is shown that these two measures are directly related for a homogeneous model. In addition, with examples from Model 1 and Model 2 from the experimental study in Chapter 4, it is also shown that Q has a positive linear relationship with $RE_t(M)$.

Application of the quasi-likelihood method results in an estimate of the quasi-likelihood state vector, which is a function of the parameters of the model, given

the information $\mathcal{F}_{t-1}^{*}$. Knowledge of all components up to and including those with subscripts $t-1$ is assumed. Hence the error variance matrix of the estimated state vector is also conditional on this information, as can be seen in (5.59).

On the other hand, the Kalman filter relative efficiency, as reported in (3.59), uses the appropriate element of $\mathbf{P}_{t|t} = \mathrm{E}(\alpha_t - \mathbf{a}_{t|t})(\alpha_t - \mathbf{a}_{t|t})'$, (Harvey, 1989, Section 3.2.3). This is the mean squared error (MSE) matrix of the estimator, $\mathbf{a}_{t|t}$. Also, since this matrix is independent of the observations, it can be described as the 'unconditional error covariance matrix associated with the conditional mean estimator' (Harvey, 1989, page 111).

It is expected that the value of Q will be directly related to the value of $RE_t(M)$, but may not be exactly equal, due to the differing conditional information.

5.6.1 Homogeneous Model

A multivariate BSM (or SUTSE model) is defined as homogeneous if the covariance matrix of each component is proportional to the covariance matrix of the measurement error. This special case is referred to in Section 3.3.1. More detail can be found in Harvey (1989, Section 8.3). Thus, in this context,

$$\boldsymbol{\Sigma}_{(M),\,\eta} = h_\eta \boldsymbol{\Sigma}_\varepsilon, \qquad \boldsymbol{\Sigma}_{(M),\,\zeta} = h_\zeta \boldsymbol{\Sigma}_\varepsilon, \qquad \boldsymbol{\Sigma}_{(M),\,\omega} = h_\omega \boldsymbol{\Sigma}_\varepsilon, \tag{5.61}$$

where h_η, h_ζ and h_ω are non-negative scalars.

Under these strict conditions, it is known that the Kalman filter becomes 'decoupled'. This means that the multivariate and univariate approach to the aggregated series are the same. In terms of the relative efficiency measure, $RE_t(M) = 1$. In this known case, since the model is simplified substantially, the quasi-likelihood indicator is easy to derive.

Rewriting the covariance matrices with simplified notation as given in (5.57), the covariance for the level component for $K = 2$ becomes

$$\boldsymbol{\Sigma}_{(m),\,\eta} = \begin{pmatrix} \mathrm{n}_1 & \sigma_\eta^2 \\ \sigma_\eta^2 & \mathrm{n}_2 \end{pmatrix}. \tag{5.62}$$

Hence, after the transformation, (5.45) is simplified to the following and the condition for the homogeneous case is applied:

$$\boldsymbol{\Sigma}_{(M),\,\eta} = \begin{pmatrix} \mathrm{N} & \sigma_\eta^2 + \mathrm{n}_1 \\ \sigma_\eta^2 + \mathrm{n}_1 & \mathrm{n}_1 \end{pmatrix} = h_\eta \begin{pmatrix} \mathrm{E} & \sigma_\varepsilon^2 + \mathrm{e}_1 \\ \sigma_\varepsilon^2 + \mathrm{e}_1 & \mathrm{e}_1 \end{pmatrix}. \tag{5.63}$$

Similarly, applying the condition for the seasonal component,

$$\boldsymbol{\Sigma}_{(M),\,\omega} = \begin{pmatrix} \mathrm{W} & \sigma_\omega^2 + \mathrm{w}_1 \\ \sigma_\omega^2 + \mathrm{w}_1 & \mathrm{w}_1 \end{pmatrix} = h_\omega \begin{pmatrix} \mathrm{E} & \sigma_\varepsilon^2 + \mathrm{e}_1 \\ \sigma_\varepsilon^2 + \mathrm{e}_1 & \mathrm{e}_1 \end{pmatrix}. \tag{5.64}$$

The condition $\boldsymbol{\Sigma}_{(M),\,\zeta} = h_\zeta \boldsymbol{\Sigma}_{(M),\,\varepsilon}$ also applies to complete the homogeneous model. Note, that in the equation for Q (5.60), the slope component parameters are not required and so the condition for the slope component is not written here.

By equating the corresponding terms, a set of equations is produced, determining each level, slope and seasonal parameter in terms of the measurement error terms. From (5.58), the numerator of $\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}\right)$ can be simplified by making the relevant substitutions:

$$\begin{aligned}
&\mathrm{W}\left(\left[\sigma_\eta^2 + \sigma_\varepsilon^2\right]^2 - \left[\mathrm{n}_1 + \mathrm{e}_1\right]\left[\mathrm{n}_2 + \mathrm{e}_2\right]\right) + (\mathrm{N} + \mathrm{E})\left(\sigma_\omega^4 - \mathrm{w}_1\mathrm{w}_2\right) \\
&= h_\omega \mathrm{E}\left(\left[h_\eta\sigma_\varepsilon^2 + \sigma_\varepsilon^2\right]^2 - \left[h_\eta\mathrm{e}_1 + \mathrm{e}_1\right]\left[h_\eta\mathrm{e}_2 + \mathrm{e}_2\right]\right) + (h_\eta\mathrm{E} + \mathrm{E})\left(h_\omega^2\sigma_\varepsilon^4 - h_\omega^2\mathrm{e}_1\mathrm{e}_2\right) \\
&= h_\omega \mathrm{E}\left(\sigma_\varepsilon^4\left[h_\eta + 1\right]^2 - \mathrm{e}_1\mathrm{e}_2\left[h_\eta + 1\right]^2\right) + \mathrm{E}\left[h_\eta + 1\right]h_\omega^2\left(\sigma_\varepsilon^4 - \mathrm{e}_1\mathrm{e}_2\right) \\
&= h_\omega \mathrm{E}\left[h_\eta + 1\right]^2\left(\sigma_\varepsilon^4 - \mathrm{e}_1\mathrm{e}_2\right) + h_\omega^2\mathrm{E}\left[h_\eta + 1\right]\left(\sigma_\varepsilon^4 - \mathrm{e}_1\mathrm{e}_2\right) \\
&= h_\omega \mathrm{E}\left[h_\eta + 1\right]\left(\sigma_\varepsilon^4 - \mathrm{e}_1\mathrm{e}_2\right)\left(h_\eta + 1 + h_\omega\right).
\end{aligned} \tag{5.65}$$

From (5.58), the denominator of $\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}\right)$ can also be simplified by making the relevant substitutions:

$$\begin{aligned}
&\left(\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\omega^2\right)^2 - \left(\mathrm{n}_1 + \mathrm{e}_1 + \mathrm{w}_1\right)\left(\mathrm{n}_2 + \mathrm{e}_2 + \mathrm{w}_2\right) \\
&= \left(h_\eta\sigma_\varepsilon^2 + \sigma_\varepsilon^2 + h_\omega\sigma_\varepsilon^2\right)^2 - \left(h_\eta\mathrm{e}_1 + \mathrm{e}_1 + h_\omega\mathrm{e}_1\right)\left(h_\eta\mathrm{e}_2 + \mathrm{e}_2 + h_\omega\mathrm{e}_2\right) \\
&= \sigma_\varepsilon^4\left(h_\eta + 1 + h_\omega\right)^2 - \mathrm{e}_1\mathrm{e}_2\left(h_\eta + 1 + h_\omega\right)^2 \\
&= \left(h_\eta + 1 + h_\omega\right)^2\left(\sigma_\varepsilon^4 - \mathrm{e}_1\mathrm{e}_2\right).
\end{aligned} \tag{5.66}$$

Therefore, the conditional error variance of the seasonal estimator for the homogeneous model is given as

$$\begin{aligned}\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t}-\hat{S}^{(q)}_{tot,\,t}\right) &= \frac{h_\omega\sigma^2_{tot,\,\varepsilon}\left[h_\eta+1\right]\left(\sigma^4_\varepsilon-\mathrm{e}_1\mathrm{e}_2\right)\left(h_\eta+1+h_\omega\right)}{\left(h_\eta+1+h_\omega\right)^2\left(\sigma^4_\varepsilon-\mathrm{e}_1\mathrm{e}_2\right)}\\ &= \frac{h_\omega\sigma^2_{tot,\,\varepsilon}\left[h_\eta+1\right]}{\left(h_\eta+1+h_\omega\right)}.\end{aligned} \tag{5.67}$$

To determine the numerator of Q, the conditional error variance of the seasonal estimate for the univariate model is required. From (5.26), this is given by

$$\mathrm{Var}_{t-1}\left(S^{(q)}_t-\hat{S}^{(q)}_t\right)=\frac{\sigma^2_{U,\,\omega}(\sigma^2_{U,\,\eta}+\sigma^2_{U,\,\varepsilon})}{\sigma^2_{U,\,\eta}+\sigma^2_{U,\,\varepsilon}+\sigma^2_{U,\,\omega}}. \tag{5.68}$$

Furthermore, if the exact parameters are being used in Q, then

$$\begin{aligned}\sigma^2_{U,\,\eta} &= \sigma^2_{tot,\,\eta}=h_\eta\sigma^2_{tot,\,\varepsilon},\\ \sigma^2_{U,\,\omega} &= \sigma^2_{tot,\,\omega}=h_\omega\sigma^2_{tot,\,\varepsilon},\\ \sigma^2_{U,\,\varepsilon} &= \sigma^2_{tot,\,\varepsilon}.\end{aligned} \tag{5.69}$$

By substituting for these in (5.68), the numerator of Q is simplified to

$$\begin{aligned}\mathrm{Var}_{t-1}\left(S^{(q)}_t-\hat{S}^{(q)}_t\right) &= \frac{h_\omega\sigma^2_{tot,\,\varepsilon}\left(h_\eta\sigma^2_{tot,\,\varepsilon}+\sigma^2_{tot,\,\varepsilon}\right)}{h_\eta\sigma^2_{tot,\,\varepsilon}+\sigma^2_{tot,\,\varepsilon}+h_\omega\sigma^2_{tot,\,\varepsilon}}\\ &= \frac{h_\omega\sigma^4_{tot,\,\varepsilon}\left(h_\eta+1\right)}{\sigma^2_{tot,\,\varepsilon}\left(h_\eta+1+h_\omega\right)}\\ &= \frac{h_\omega\sigma^2_{tot,\,\varepsilon}\left(h_\eta+1\right)}{\left(h_\eta+1+h_\omega\right)}.\end{aligned} \tag{5.70}$$

For the $K=2$ homogeneous case, the quasi-likelihood indicator, denoted by $Q^{(h)}$, is therefore given by the ratio of (5.70) and (5.67). Since these expressions are equal, it simplifies to 1.

$$Q^{(h)}=\frac{\mathrm{Var}_{t-1}\left(S^{(q)}_t-\hat{S}^{(q)}_t\right)}{\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t}-\hat{S}^{(q)}_{tot,\,t}\right)}=1. \tag{5.71}$$

Of course, if estimated parameters are used in $Q^{(h)}$, the estimated parameters given the univariate model are unlikely to be exactly equal (as shown in (5.69)) to the estimated total parameters given by the multivariate model. In the exact parameter

case, these are constrained to be equal, as described in Section 3.3. Therefore, $Q^{(h)}$ may not be exactly equal to one for estimated parameters, it is expected to be close to one.

In conclusion, for the homogeneous model with $K = 2$, the quasi-likelihood indicator is the same as that given for the Kalman filter relative efficiency ratio.

5.6.2 Model 1: Compound Symmetry

In Section 3.3, Model 1 is described as a BSM in which the covariance matrices have a compound symmetry structure. That is, each sub-series has the same series-specific variance. For example, when $K = 2$, the covariance matrix for the level component is

$$\mathbf{\Sigma}_{(m),\,\eta} = \begin{pmatrix} \sigma_\eta^2 + \sigma_{\eta^*}^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{\eta^*}^2 \end{pmatrix}, \tag{5.72}$$

and similarly for the slope, seasonal and error covariance matrices. Note that although this model is not strictly homogeneous, it would become so if the condition that $\rho_\eta = \rho_\zeta = \rho_\omega = \rho_\varepsilon$ is imposed. For this model, the c-ratios given in ((4.13) - (4.14)) are all equal to unity.

The abbreviations given in (5.57) are amended for the compound symmetry model:

$$\begin{aligned} \mathrm{n}_1 &= \sigma_\eta^2 + \sigma_{\eta^*}^2, & \mathrm{e}_1 &= \sigma_\varepsilon^2 + \sigma_{\varepsilon^*}^2, & \mathrm{w}_1 &= \sigma_\omega^2 + \sigma_{\omega^*}^2, \\ \mathrm{N} &= 2\sigma_\eta^2 + 2\mathrm{n}_1, & \mathrm{E} &= 2\sigma_\varepsilon^2 + 2\mathrm{e}_1, & \mathrm{W} &= 2\sigma_\omega^2 + 2\mathrm{w}_1, \end{aligned} \tag{5.73}$$

such that (5.72) becomes

$$\mathbf{\Sigma}_{(m),\,\eta} = \begin{pmatrix} \mathrm{n}_1 & \sigma_\eta^2 \\ \sigma_\eta^2 & \mathrm{n}_1 \end{pmatrix}. \tag{5.74}$$

Recalling the transformed matrix given in (5.45), the transformed version of (5.74) is given by:

$$\mathbf{\Sigma}_{(M),\,\eta} = \begin{pmatrix} \mathrm{N} & \sigma_\eta^2 + \mathrm{n}_1 \\ \sigma_\eta^2 + \mathrm{n}_1 & \mathrm{n}_1 \end{pmatrix}. \tag{5.75}$$

Similarly, the seasonal and error component covariance matrices become

$$\boldsymbol{\Sigma}_{(M),\,\omega} = \begin{pmatrix} \mathrm{W} & \sigma_\omega^2 + \mathrm{w}_1 \\ \sigma_\omega^2 + \mathrm{w}_1 & \mathrm{w}_1 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{(M),\,\varepsilon} = \begin{pmatrix} \mathrm{E} & \sigma_\varepsilon^2 + \mathrm{e}_1 \\ \sigma_\varepsilon^2 + \mathrm{e}_1 & \mathrm{e}_1 \end{pmatrix}. \quad (5.76)$$

From (5.58), the numerator of $\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}\right)$ (5.58) for Model 1 can be simplified by making the relevant substitutions and then factorising:

$$\begin{aligned} \mathrm{W}\left(\left[\sigma_\eta^2 + \sigma_\varepsilon^2\right]^2 - \left[\mathrm{n}_1 + \mathrm{e}_1\right]^2\right) + (\mathrm{N}+\mathrm{E})\left(\sigma_\omega^4 - \mathrm{w}_1^2\right) \\ = \mathrm{W}\left(\left[\sigma_\eta^2 + \sigma_\varepsilon^2 - \mathrm{n}_1 - \mathrm{e}_1\right]\left[\sigma_\eta^2 + \sigma_\varepsilon^2 + \mathrm{n}_1 + \mathrm{e}_1\right]\right) \\ + (\mathrm{N}+\mathrm{E})\left[\sigma_\omega^2 - \mathrm{w}_1\right]\left[\sigma_\omega^2 + \mathrm{w}_1\right]. \end{aligned} \quad (5.77)$$

Rearranging the terms in the expressions given in the second row of (5.73), it can be seen that

$$\begin{aligned} \sigma_\eta^2 + \mathrm{n}_1 &= \tfrac{1}{2}\mathrm{N}, \quad \text{then} \quad \sigma_\eta^2 - \mathrm{n}_1 = \tfrac{1}{2}(\mathrm{N} - 4\mathrm{n}_1); \\ \sigma_\varepsilon^2 + \mathrm{e}_1 &= \tfrac{1}{2}\mathrm{E}, \quad \text{then} \quad \sigma_\varepsilon^2 - \mathrm{e}_1 = \tfrac{1}{2}(\mathrm{E} - 4\mathrm{e}_1); \\ \sigma_\omega^2 + \mathrm{w}_1 &= \tfrac{1}{2}\mathrm{W}, \quad \text{then} \quad \sigma_\omega^2 - \mathrm{w}_1 = \tfrac{1}{2}(\mathrm{W} - 4\mathrm{w}_1). \end{aligned}$$

Substituting these expressions into (5.77), the numerator of $\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}\right)$ (5.58) for Model 1 becomes:

$$\begin{aligned} & \mathrm{W}\left[\frac{1}{2}(\mathrm{N} - 4\mathrm{n}_1) + \frac{1}{2}(\mathrm{E} - 4\mathrm{e}_1)\right]\left[\frac{\mathrm{N}}{2} + \frac{\mathrm{E}}{2}\right] + (\mathrm{N}+\mathrm{E})\left[\frac{1}{2}(\mathrm{W} - 4\mathrm{w}_1)\frac{\mathrm{W}}{2}\right] \\ &= \frac{\mathrm{W}}{4}\left[\mathrm{N} - 4\mathrm{n}_1 + \mathrm{E} - 4\mathrm{e}_1\right](\mathrm{N}+\mathrm{E}) + \frac{\mathrm{W}}{4}(\mathrm{W} - 4\mathrm{w}_1)(\mathrm{N}+\mathrm{E}) \\ &= \frac{\mathrm{W}}{4}(\mathrm{N}+\mathrm{E})\left[\mathrm{N} - 4\mathrm{n}_1 + \mathrm{E} - 4\mathrm{e}_1 + \mathrm{W} - 4\mathrm{w}_1\right]. \end{aligned} \quad (5.78)$$

Similarly, the denominator of $\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}\right)$ (5.58) can be simplified:

$$\begin{aligned} \left(\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\omega^2\right)^2 - \left(\mathrm{n}_1 + \mathrm{e}_1 + \mathrm{w}_1\right)^2 \\ = \left(\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\omega^2 - \mathrm{n}_1 - \mathrm{e}_1 - \mathrm{w}_1\right)\left(\sigma_\eta^2 + \sigma_\varepsilon^2 + \sigma_\omega^2 + \mathrm{n}_1 + \mathrm{e}_1 + \mathrm{w}_1\right) \\ = \frac{1}{4}(\mathrm{N} - 4\mathrm{n}_1 + \mathrm{E} - 4\mathrm{e}_1 + \mathrm{W} - 4\mathrm{w}_1)(\mathrm{N}+\mathrm{E}+\mathrm{W}). \end{aligned} \quad (5.79)$$

Therefore, the conditional error variance of the seasonal estimate under the multivariate approach for the compound symmetry model (Model 1), is given by the ratio of (5.78) and (5.79):

$$\begin{aligned}\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t}-\hat{S}^{(q)}_{tot,\,t}\right) &= \frac{\frac{\mathrm{W}}{4}\left(\mathrm{N}+\mathrm{E}\right)\left[\mathrm{N}-4\mathrm{n}_1+\mathrm{E}-4\mathrm{e}_1+\mathrm{W}-4\mathrm{w}_1\right]}{\frac{1}{4}\left(\mathrm{N}-4\mathrm{n}_1+\mathrm{E}-4\mathrm{e}_1+\mathrm{W}-4\mathrm{w}_1\right)\left(\mathrm{N}+\mathrm{E}+\mathrm{W}\right)} \\ &= \frac{\mathrm{W}\left(\mathrm{N}+\mathrm{E}\right)}{\left(\mathrm{N}+\mathrm{E}+\mathrm{W}\right)}.\end{aligned} \tag{5.80}$$

This is identical to the expression for the conditional error variance for the seasonal estimator under the univariate approach given by (5.26), if the exact parameters are used. Therefore, if $Q^{(M_1)}$ denotes the quasi-likelihood indicator for Model 1, then as in the homogeneous case,

$$Q^{(M_1)} = 1. \tag{5.81}$$

Referring back to the designs discussed in the previous chapter, the design which describes Model 1 with all c-ratios equal to one and with not all correlations equal, is labelled *a11*. The result of $RE_t(M)$ (the Kalman filter relative efficiency) for *a11* is reported to be 1.0 in Table (4.4). This result is consistent with the quasi-likelihood indicator measure as derived in this section.

5.6.3 Model 2: Unit-specific Variances

In this section, the quasi-likelihood indicator is calculated for the local level seasonal model described by (4.1). This model is a basic structural model with a local level trend (with no slope term) and a dummy quarterly seasonal component. It was analysed extensively in the last chapter. The analysis continues here with the calculation of the quasi-likelihood indicator measure (Q) for each different design identified in Chapter 4. It is then possible to determine a relationship between the relative efficiency from the Kalman filter method, $RE_t(M)$, and the associated value of Q.

The values for Q may be determined by the formula for the $K = 2$ case for Model 2, as given by (5.60). Although this formula has been derived from a BSM with a slope component, note that the formula remains the same for the local level seasonal model. Assuming exact parameters, once the parameter values are calculated for each design, they may be substituted into the formula for Q to determine its value.

Table 5.1: Results of $RE_{40}(M)$ for sub-series designs *a21, a31*, and *a41* with corresponding quasi-indicator, Q (in bold).

a21	$c_\eta = 1$ and $c_\varepsilon = 1$									
$c_\omega = 5$	*A1*	1.0929	*B1*	1.1088	*C1*	1.1263	*D1*	1.1469	*E1*	1.1748
		1.1682		**1.2127**		**1.2654**		**1.3286**		**1.4061**
	A2	1.0872	*B2*	1.1047	*C2*	1.1250	*D2*	1.1504	*E2*	1.1883
		1.1476		**1.1936**		**1.2528**		**1.3315**		**1.4417**
a31	$c_\eta = 1$ and $c_\varepsilon = 1$									
$c_\omega = 10$	*A1*	1.1837	*B1*	1.2204	*C1*	1.2626	*D1*	1.3141	*E1*	1.3860
		1.3114		**1.4146**		**1.5518**		**1.7429**		**2.0278**
	A2	1.1733	*B2*	1.2124	*C2*	1.25963	*D2*	1.3205	*E2*	1.4131
		1.2743		**1.3755**		**1.5192**		**1.7391**		**2.1180**
a41	$c_\eta = 1$ and $c_\varepsilon = 1$									
$c_\omega = 20$	*A1*	1.2945	*B1*	1.3641	*C1*	1.4484	*D1*	1.5570	*E1*	1.7178
		1.4512		**1.6327**		**1.9060**		**2.3639**		**3.2903**

For the specified aggregate series in Chapter 4, there are 16 different 'a' designs which have different combinations of c-ratios, as given in Table 4.1. With all possible correlation combinations (Table 4.3) there are altogether 90 different 'a' designs. For each combination, Q has been calculated and recorded in the Tables 5.1 and 5.2. Table 5.1 shows the results of Q in bold, and $RE_{40}(M)$ for the possible correlation combinations of designs *a21, a31*, and *a41*.

Note that design *a11* has $RE_t(M) = 1$ and $Q = 1$ for each correlation combination, as it represents a compound symmetry design as described in the previous section. It is therefore unnecessary to report the results here. For the 3 designs in Table 5.1, as $RE_{40}(M)$ increases across the rows, so does Q. Within each design, the trend in Q is similar to that of $RE_{40}(M)$. The only exception is for design '*a31*' with '*D1*' and '*D2*' where $RE_{40}(M)$ increases slightly but Q decreases slightly.

Table 5.2 shows the results of $RE_{40}(M)$ and Q for the remaining possible combinations of the 'a' designs. Again, it can be seen that within each design , as $RE_{40}(M)$ increases or decreases, Q does also. There are two exceptions here, both

within the '*a12*' design. Reading across from *A4* to *B4*, $RE_{40}(M)$ decreases slightly but Q increases slightly. Similarly for *A5* to *B5*. Again, in these cases and especially for *A5* to *B5*, the values of $RE_{40}(M)$ are very similar.

However, Q does not always show a consistent relationship with $RE_{40}(M)$. This is most likely due to the difference of the definitions. The values shown for $RE_{40}(M)$ are calculated with the Kalman filter specifically for the time point at $t = 40$ when $T = 40$, hence, for the last time point in the series. The quasi-likelihood value, Q, is not specific to a particular time point, but is a general measure which is derived using the one-step process from time $t - 1$ to t. For example, looking at *D1* for design '*a41*' in Table 5.1, $RE_{40}(M) = 1.557$ and $Q = 2.364$. Compare this with *A5* for design '*a13*' in Table 5.2 which has $RE_{40}(M) = 1.560$ and $Q = 1.872$. The $RE_{40}(M)$ are very similar in magnitude but the Q values are quite different. The majority however, do seem to be fairly consistent.

For the nine different 'b' designs (Table 4.2) which have different combinations of c-ratios, there are altogether 20 different possible correlation combinations. Table 5.3 shows the results of $RE_{40}(M)$ and Q for all the 'b' designs. The results are consistent with results for the 'a' designs, showing a direct relationship between the value of $RE_{40}(M)$ and the value of Q. The only exceptions are when movement is very slight between designs such as from *A1* to *B1* and also *A2* to *B2* in *b22*. Similarly, from *A1* to *A2* in *b32*.

This direct relationship can be seen more clearly when the value of $RE_{40}(M)$ is plotted against its corresponding value of Q for all 110 designs, as shown in Figure 5.1. The overall correlation coefficient is 0.98. For the 'a' designs only, the correlation is 0.96 and for the 'b' designs only, it is 0.97. The fitted regression line has an intercept of 0.66 and a slope of 0.37 and goes through (1, 1.03). For this example, Q is an overestimate for $RE_{40}(M)$ greater than one. The results indicate a very strong linear relationship between the two measures and shows that in general, Q preserves the relativity of $RE_{40}(M)$. The results presented here are for a particular aggregate series with 110 designs of two sub-series.

Table 5.2: Results of $RE_{40}(M)$ for remaining sub-series designs 'a' and corresponding quasi-indicator, Q (in bold).

	a12: $c_\eta, c_\varepsilon = 5$				**a13:** $c_\eta, c_\varepsilon = 10$		**a14:** $c_\eta, c_\varepsilon = 20$	
$c_\omega = 1$	*A1*	1.0674	*B1*	1.0576	*A1*	1.1158	*A1*	1.1588
		1.1379		**1.1207**		**1.2514**		**1.3609**
	A2	1.0899	*B2*	1.0788	*A2*	1.1578	*A2*	1.2198
		1.1787		**1.1619**		**1.3396**		**1.5076**
	A3	1.1216	*B3*	1.1094	*A3*	1.2185	*A3*	1.3109
		1.2282		**1.2162**		**1.4578**		**1.7238**
	A4	1.1714	*B4*	1.1598	*A4*	1.3187	*A4*	1.4687
		1.2892		**1.2903**		**1.6225**		**2.0709**
	A5	1.2800	*B5*	1.2791	*A5*	1.5598	*A5*	1.8870
		1.3671		**1.3991**		**1.8722**		**2.7312**
	a22: $c_\eta, c_\varepsilon = 5$				**a23:** $c_\eta, c_\varepsilon = 10$		**a24:** $c_\eta, c_\varepsilon = 20$	
$c_\omega = 5$	*A1*	1.0005	*B1*	1.0039	*A1*	1.0045	*A1*	1.0178
		1.0009		**1.0076**		**1.0088**		**1.0366**
	A2	1.0005	*B2*	1.0004	*A2*	1.0157	*A2*	1.0421
		1.0008		**1.0008**		**1.0301**		**1.0860**
	a32: $c_\eta, c_\varepsilon = 5$				**a33:** $c_\eta, c_\varepsilon = 10$		**a34:** $c_\eta, c_\varepsilon = 20$	
$c_\omega = 10$	*A1*	1.0152	*B1*	1.1088	*A1*	1.0008	*A1*	1.0021
		1.0268		**1.2127**		**1.0015**		**1.0041**
	A2	1.0067	*B2*	1.1047	*A2*	1.0008	*A2*	1.0130
		1.0111		**1.1936**		**1.0014**		**1.0252**
	a42: $c_\eta, c_\varepsilon = 5$				**a43:** $c_\eta, c_\varepsilon = 10$		**a44:** $c_\eta, c_\varepsilon = 20$	
$c_\omega = 20$	*A1*	1.0454	*B1*	1.08132	*A1*	1.0124	*A1*	1.0010
		1.0725		**1.14175**		**1.0213**		**1.0019**

Table 5.3: Results of $RE_{40}(M)$ for sub-series design 'b' with corresponding quasi-indicator, Q (in bold).

	b22: $c_\eta, c_\varepsilon = 0.2$				**b23:** $c_\eta, c_\varepsilon = 0.1$		**b24:** $c_\eta, c_\varepsilon = 0.05$	
$c_\omega = 5$	*A1*	1.3728	*B1*	1.3689	*A1*	1.5063	*A1*	1.6158
		1.8191		**1.8547**		**2.2554**		**2.6986**
	A2	1.4076	*B2*	1.4071	*A2*	1.5711	*A2*	1.7099
		1.8529		**1.9116**		**2.3794**		**2.9603**
	b32: $c_\eta, c_\varepsilon = 0.2$				**b33:** $c_\eta, c_\varepsilon = 0.1$		**b34:** $c_\eta, c_\varepsilon = 0.05$	
$c_\omega = 10$	*A1*	1.6060	*B1*	1.6171	*A1*	1.8108	*A1*	1.9818
		2.3146		**2.4548**		**3.0745**		**3.9374**
	A2	1.6399	*B2*	1.6577	*A2*	1.8798	*A2*	2.0866
		2.3094		**2.4791**		**3.1605**		**4.2008**
	b42: $c_\eta, c_\varepsilon = 0.2$				**b43:** $c_\eta, c_\varepsilon = 0.1$		**b44:** $c_\eta, c_\varepsilon = 0.05$	
$c_\omega = 20$	*A1*	1.9014	*B1*	1.9431	*A1*	2.2125	*A1*	2.4820
		2.8381		**3.1551**		**4.0469**		**5.6246**

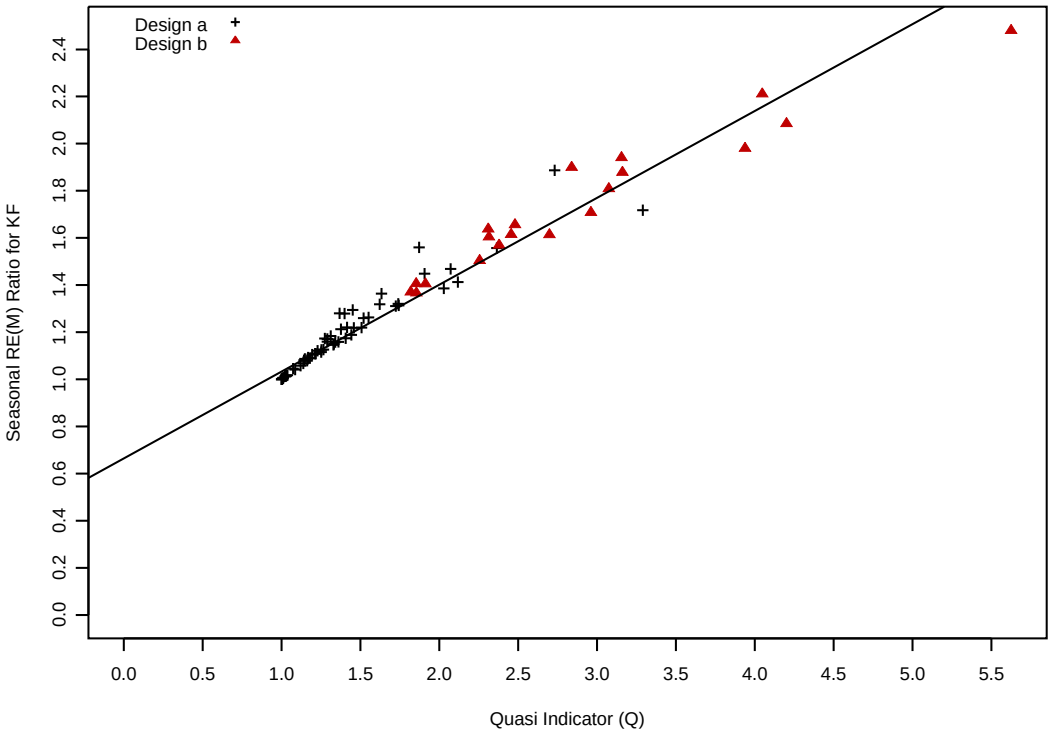


Figure 5.1: Quasi-likelihood indicator, Q, vs relative efficiency, $RE_{40}(M)$, for 110 possible sub-series designs with $K = 2$.

5.7 Grouping Application

The quasi-likelihood indicator, Q, may be useful in practice in identifying when gains are achievable using the multivariate model. In practice, if the main focus is the efficient seasonal adjustment of an aggregated series without particular interest in the seasonal adjustment of the sub-series, then grouping the sub-series may be beneficial to the estimation process. The quasi-likelihood indicator can provide useful information when there are different ways of combining sub-series to obtain the same aggregated series.

If there are several sub-series ($K > 2$), gains for the variance of the seasonally adjusted aggregated series may still be achievable using the multivariate model. However, as K becomes large, the number of parameters in the multivariate model increases and there may be instability in the model. By grouping the sub-series into $r < K$ groups, the number of parameters in the model is reduced. Within each group, the original sub-series are aggregated to form the new r sub-series. Depending on the value of K, there will be many different ways of grouping the sub-series. Application of the quasi-likelihood indicator measure may be helpful in identifying which grouping will achieve the better gains.

Until now, the aggregated series has been combined with the $K-1$ sub-series for use in the multivariate model using the transformation matrix given in (3.33). If the original K sub-series are grouped to have $r < K$ sub-series, a different transformation matrix applies. A simple example can be shown with $K = 4$ sub-series, $(Y_{1t},\ Y_{2t},\ Y_{3t},\ Y_{4t})$. The transformation matrix, A (3.33), for a system of four sub-series, has dimension 4×4, as follows:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{5.82}$$

The result is a system of four series, $(Y_{1t} + Y_{2t} + Y_{3t} + Y_{4t},\ Y_{1t},\ Y_{2t},\ Y_{3t})$, which includes the aggregated series and the first three series of the original four.

However, a different configuration may provide a similar gain with a more stable system for estimation. For example, $(Y_{1t} + Y_{2t},\, Y_{3t} + Y_{4t})$ results in $r = 2$ sub-series while $(Y_{1t} + Y_{2t},\, Y_{3t},\, Y_{4t})$ results in $r = 3$ sub-series. For the given examples, the corresponding transformation matrices are respectively:

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{\Lambda} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{5.83}$$

Thus, the transformation matrix, $\mathbf{\Lambda}$, for the first example in (5.83) has dimensions 2×4 and for the second example, the dimensions are 3×4. These are only two of 14 (refer to Appendix C.3) different possible ways in which four series can be grouped. As Q preserves the relativities of the relative efficiency, the configuration with the highest Q can identify the configuration which achieves the highest gain in the multivariate BSM. Note that calculation of Q depends upon the parameters for the multivariate model with K sub-series and the parameters for the univariate model, but not on t.

In this section, a general result for Q is derived for the case of $K > 2$ grouped into $r < K$ sub-series. The transformation matrix, $\mathbf{\Lambda}$, applied to the multivariate system of K series has K columns. The dimensions of $\mathbf{\Lambda}$ will be $r \times K$. If the BSM is further generalised to have univariate state vector with dimensions $1 \times p$, then the derivation in Sub-section 5.4.1 can be adapted for this general case. Note that in the previous sections, $p = 5$, since the state vector consisted of a level and slope component in addition to three seasonal elements.

The transformed state space system is now written as

$$\begin{aligned} \mathbf{\Lambda Y}_{(m*),\,t} &= \mathbf{\Lambda}(\mathbf{Z} \otimes \mathbf{I}_K)\alpha_{(m*),\,t} + \mathbf{\Lambda}\mathcal{E}_{(m*),\,t} \\ &= (\mathbf{Z} \otimes \mathbf{I}_r)(\mathbf{I}_p \otimes \mathbf{\Lambda})\alpha_{(m*),\,t} + \mathbf{\Lambda}\mathcal{E}_{(m*),\,t} \qquad (5.84) \\ (\mathbf{I}_p \otimes \mathbf{\Lambda})\alpha_{(m*),\,t} &= (\mathbf{I}_p \otimes \mathbf{\Lambda})(\mathbf{T} \otimes \mathbf{I}_K)\alpha_{(m*),\,t-1} + (\mathbf{I}_p \otimes \mathbf{\Lambda})(\mathbf{G} \otimes \mathbf{I}_K)\gamma_{(m*),\,t-1} \\ &= (\mathbf{T} \otimes \mathbf{I}_r)(\mathbf{I}_p \otimes \mathbf{\Lambda})\alpha_{(m*),\,t-1} + (\mathbf{I}_p \otimes \mathbf{\Lambda})(\mathbf{G} \otimes \mathbf{I}_K)\gamma_{(m*),\,t-1}. \end{aligned}$$

(5.85)

The multivariate state vector modified for the quasi-likelihood method is given by:

$$\begin{aligned}\alpha^{(q)}_{(M),\,t} &= \mathbf{U}_{(M)}\alpha_{(M),\,t} \\ &= \mathbf{U}_{(M)}(\mathbf{I}_p \otimes \mathbf{\Lambda})\alpha_{(m*),\,t}. \end{aligned} \tag{5.86}$$

If the univariate model remains the same as in Section 5.3, then $\mathbf{U}_{(M)} = \mathbf{U} \otimes \mathbf{I}_r$, in which $\mathbf{U}$ has dimensions $3 \times p$. Since a different model may require a different selection ($\mathbf{U}$) matrix, in general let $\mathbf{U}$ have dimensions $u \times p$.

The general expression for the martingale difference, $\delta_{(M),\,t}$, is adapted from (5.41) and becomes

$$\begin{aligned}\delta_{(M),\,t} &= \begin{pmatrix} \mathbf{\Lambda}\mathbf{Y}_{(m*),\,t} \\ -\mathbf{U}_{(M)}(\mathbf{T} \otimes \mathbf{I}_r)(\mathbf{I}_p \otimes \mathbf{\Lambda})\alpha_{(m*),\,t-1} \end{pmatrix} + \begin{pmatrix} -(\mathbf{Z} \otimes \mathbf{I}_r)\mathbf{U}^{'}_{(M)} \\ \mathbf{I}_{ur} \end{pmatrix} \alpha^{(q)}_{(M),\,t} \\ &= \mathbf{X}_{(M),\,1} + \mathbf{B}^{'}_{(M)}\alpha^{(q)}_{(M),\,t}. \end{aligned} \tag{5.87}$$

In this general case, $\mathbf{X}_{(M),\,1}$ has dimensions $(r + ur) \times 1$, and $\mathbf{B}_{(M)}$ has dimensions $ur \times (r + ur)$. The derivative of $\delta_{(M),\,t}$ with respect to the quasi-likelihood state vector is given by:

$$\begin{aligned}\mathbf{B}_{(M)} &= \frac{\partial \delta_{(M),\,t}}{\partial \alpha^{(q)}_{(M),\,t}} \\ &= \left(-\mathbf{U}_{(M)}(\mathbf{Z}' \otimes \mathbf{I}_r),\ \ \mathbf{I}_{ur}\right) \\ &= (-\mathbf{U}\mathbf{Z}',\ \ \mathbf{I}_u) \otimes \mathbf{I}_r \\ &= \mathbf{B} \otimes \mathbf{I}_r. \end{aligned} \tag{5.88}$$

The conditional variance matrix is denoted by $\mathbf{V}_{(M)}$,

$$\mathbf{V}_{(M)} = \mathrm{Var}_{t-1}(\delta_{(M),\,t}) = \begin{pmatrix} \mathbf{\Lambda}\mathbf{\Sigma}_{(m),\,\varepsilon}\mathbf{\Lambda}^{'} & \mathbf{0}_{(r\times r)} & \mathbf{0}_{(r\times r)} & \mathbf{0}_{(r\times r)} \\ \mathbf{0}_{(r\times r)} & \mathbf{\Lambda}\mathbf{\Sigma}_{(m),\,\eta}\mathbf{\Lambda}^{'} & \mathbf{0}_{(r\times r)} & \mathbf{0}_{(r\times r)} \\ \mathbf{0}_{(r\times r)} & \mathbf{0}_{(r\times r)} & \mathbf{\Lambda}\mathbf{\Sigma}_{(m),\,\zeta}\mathbf{\Lambda}^{'} & \mathbf{0}_{(r\times r)} \\ \mathbf{0}_{(r\times r)} & \mathbf{0}_{(r\times r)} & \mathbf{0}_{(r\times r)} & \mathbf{\Lambda}\mathbf{\Sigma}_{(m),\,\omega}\mathbf{\Lambda}^{'} \end{pmatrix}, \tag{5.89}$$

where, for example, $\mathbf{\Sigma}_{(m),\,\eta}$ is the level component covariance matrix associated with the original set of sub-series. If $K = 4$, it is given by

$$\mathbf{\Sigma}_{(m),\,\eta} = \begin{pmatrix} \sigma_\eta^2 + \sigma_{1\eta^*}^2 & \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{2\eta^*}^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{3\eta^*}^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 + \sigma_{4\eta^*}^2 \end{pmatrix}. \tag{5.90}$$

So, the transformed covariance matrix analogous to (5.45), but with the transformation matrix $\mathbf{\Lambda}$, is given by $\mathbf{\Sigma}_{(M),\,\eta} = \mathbf{\Lambda}\mathbf{\Sigma}_{(m),\,\eta}\mathbf{\Lambda}'$ with dimensions $r \times r$. Similar notation applies for the covariance matrices of the slope, seasonal and error components. Thus, the dimensions of $\mathbf{V}_{(M)}$ are $(1+u)r \times (1+u)r$ (or just $4r \times 4r$ since $u = 3$ here).

From (5.87), $\mathbf{X}_{(M),\,1}$ may be written as

$$\begin{aligned} \mathbf{X}_{(M),\,1} &= \begin{pmatrix} \mathbf{\Lambda}\mathbf{Y}_{(m*),\,t} \\ -\mathbf{U}_{(M)}(\mathbf{T}\otimes\mathbf{I}_r)(\mathbf{I}_p\otimes\mathbf{\Lambda})\alpha_{(m*),\,t-1} \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{Z}\otimes\mathbf{I}_r)\,\alpha_{(M),\,t} + \mathcal{E}_{(M),\,t} \\ -\mathbf{U}_{(M)}\,(\mathbf{T}\otimes\mathbf{I}_r)\,\alpha_{(M),\,t-1} \end{pmatrix} \\ &= \begin{bmatrix} (\mathbf{Z}\otimes\mathbf{I}_r)\,\mathbf{U}'_{(M)} \\ \mathbf{0}_{(ur\times ur)} \end{bmatrix} \alpha^{(q)}_{(M),\,t} + \begin{bmatrix} \mathbf{I}_r \\ \mathbf{0}_{(ur\times r)} \end{bmatrix} \mathcal{E}_{(M),\,t} + \begin{bmatrix} \mathbf{0}_{(r\times pr)} \\ -\mathbf{U}_{(M)}\,(\mathbf{T}\otimes\mathbf{I}_r) \end{bmatrix} \alpha_{(M),\,t-1}. \end{aligned} \tag{5.91}$$

By substituting $\mathbf{U}_{(M)} = \mathbf{U}\otimes\mathbf{I}_r$, the expression for $\mathbf{X}_{(M),\,1}$ may be written in terms of the univariate system matrices as in (5.52):

$$\begin{aligned} \mathbf{X}_{(M),\,1} &= \left[\begin{pmatrix} \mathbf{Z}\mathbf{U}' \\ \mathbf{0}_{(u\times u)} \end{pmatrix} \otimes \mathbf{I}_r\right] \alpha^{(q)}_{(M),\,t} + \left[\begin{pmatrix} 1 \\ \mathbf{0}_{(u\times 1)} \end{pmatrix} \otimes \mathbf{I}_r\right] \mathcal{E}_{(M),\,t} \\ &\quad + \left[\begin{pmatrix} \mathbf{0}_{(1\times p)} \\ -\mathbf{U}\mathbf{T} \end{pmatrix} \otimes \mathbf{I}_r\right] \alpha_{(M),\,t-1} \\ &= \mathbf{X}_{(M),\,a}\;\alpha^{(q)}_{(M),\,t} + \mathbf{X}_{(M),\,b}\;\mathcal{E}_{(M),\,t} + \mathbf{X}_{(M),\,c}\;\alpha_{(M),\,t-1}, \end{aligned} \tag{5.92}$$

where $\mathbf{X}_{(M),\,a}$ has dimensions $(r+ur)\times ur$, $\mathbf{X}_{(M),\,b}$ has dimensions $(r+ur)\times r$ and $\mathbf{X}_{(M),\,c}$ has dimensions $(r+ur)\times pr$.

The $ur \times (1+u)r$ matrix $\mathcal{A}_{(M)}$ is calculated using $\mathbf{B}_{(M)}$ from (5.88) and $\mathbf{V}_{(M)}$ from (5.89), and is given by:

$$\mathcal{A}_{(M)} = -\left(\mathbf{B}_{(M)}\mathbf{V}_{(M)}^{-1}\mathbf{B}_{(M)}'\right)^{-1}\ \mathbf{B}_{(M)}\mathbf{V}_{(M)}^{-1}. \tag{5.93}$$

The error variance matrix associated with the quasi-likelihood multivariate state vector estimator may be determined as follows:

$$\begin{aligned}\mathbf{P}_{(M),\,t}^{(q)} &= \mathrm{Var}_{t-1}\left(\alpha_{(M),\,t}^{(q)} - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,1}\right)\\ &= \left[\mathbf{I}_{ur} - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,a}\right]\ \mathrm{Var}_{t-1}\left(\alpha_{(M),\,t}^{(q)}\right)\ \left[\mathbf{I}_{ur} - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,a}\right]'\\ &\quad + \mathcal{A}_{(M)}\mathbf{X}_{(M),\,b}\,\mathrm{Var}_{t-1}\left(\mathcal{E}_{(M),\,t}\right)\ \left(\mathcal{A}_{(M)}\mathbf{X}_{(M),\,b}\right)'.\end{aligned} \tag{5.94}$$

Therefore, the element corresponding to the variance of the seasonal component estimate of the aggregated series, $\mathrm{Var}_{t-1}\left(S_{tot,\,t}^{(q)} - \hat{S}_{tot,\,t}^{(q)}\right)$, is the $(2r+1)$th diagonal element of the matrix $\mathbf{P}_{(M),\,t}^{(q)}$. This assumes the seasonal component is the 3rd component in the univariate state vector of dimension $p \times 1$. The element for the variance of the seasonal component estimate of the aggregated series, as calculated in (5.94), becomes the denominator in the formula for the quasi-likelihood indicator given in (5.59). The numerator remains the same, as the univariate model has not changed.

Using the procedure derived in this section, the value of Q can be obtained for the case where K sub-series are grouped into r new sub-series. Given the estimated parameters for the multivariate model with K sub-series, there are many different ways of grouping the series (as discussed in Appendix C.3) to obtain a more stable model with fewer parameters. To decide which grouping will yield the highest relative efficiency, the value of Q for each combination can be calculated and compared. The results of this procedure will assist in determining which grouping yields the highest relative efficiency if the Kalman filter method is applied. Further work on this topic of grouping sub-series is presented in Chapter 7. An example with $K = 8$ sub-series shows the application of Q in choosing the appropriate grouping into $r < K$ sub-series.

In the following chapter, the focus is on estimating the parameters of the univariate and multivariate models described in Chapter 4, and the implications of the

estimation on the MSE of the estimated components.

Chapter 6

Parameter Estimation: Varying Series Length and Accuracy of Estimated Components

6.1 Introduction

In Chapter 4, the variance of the seasonally adjusted series was calculated for the multivariate model and compared with that of the univariate model using the relative efficiency measure, $RE_t(M)$, defined in (3.59). The relative efficiency was calculated for different designs of the sub-series when the true parameters of the model were known. In this chapter, the model parameters are considered unknown and are estimated. Since the length of the series (T) is a determinant of the accuracy of the estimation process, the multivariate and univariate models are considered for different series lengths.

By introducing the estimation of the unknown model parameters, the process becomes two-staged. Firstly, the estimation of the parameters is carried out via maximum likelihood estimation. In this stage, the accuracy of the estimated parameters for both the multivariate and univariate models is investigated and linked to the length of the series. Secondly, the estimated parameters are substituted into the state space model. Subsequently, the Kalman filter is applied to the observations to yield estimates of the state vector components, such as the level and seasonal factors, and their mean squared errors. In this second stage, the question considered is

how the estimated parameters affect the accuracy of the components of the model, these being the values of interest.

When estimated parameters are utilised, apart from the relative efficiency, other measures such as the revision error are important to consider. The revision error is the change associated with the estimated components, such as the level and seasonal factors, when a new observation becomes available. This is an important measure to consider, since it relates to the data published by statistical agencies. Furthermore, large revisions can be unsettling for general users of the data.

The mean squared errors of the estimated state vector components are obtained with the application of the Kalman filter and are sometimes referred to as the prediction error variance or the prediction mean squared errors (PMSE). When the estimated parameters are substituted into the expressions for the Kalman filter and smoother, there is a so-called 'naïve bias' in the prediction error variance (Quenneville and Singh, 2005; Pfeffermann and Tiller, 2005). This bias will be described in more detail in Section 6.5.

Extending work described in Chapter 4, this chapter includes the estimation of parameters for different values of T. The number of sub-series is kept at two ($K = 2$), and the model is the local level seasonal model with quarterly dummy seasonal factors. Data for a selection of the sub-series designs described in Chapter 4 is simulated and studied for $T = 20,\ 21, \ldots,\ 27,\ 28$ and also for $T = 40,\ 80,\ 120,\ 240$. These values have been chosen for the purposes of studying the parameter estimates of short to medium length series and comparing these with parameter estimates of long series, and also for calculation of the revision error.

6.2 Estimation of Parameters in the State Space Model

Estimates of the unknown parameters of the state space model may be computed by maximising the log-likelihood function, ln L, which uses the output of the Kalman filter (Zivot and Wang, 2006, p552). When the initial state vector α_1 contains q elements which are diffuse, the log-likelihood is adjusted for the $t = 1 \ldots d$ terms for

which $P_{\infty,t} \neq \mathbf{0}$. The result is called the diffuse log-likelihood (Koopman, 1997) and is denoted by $\ln \mathrm{L}_d(Y;\psi)$:

$$\ln \mathrm{L}_d(Y;\psi) = -\frac{TK}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{d} w_t - \frac{1}{2}\sum_{t=d+1}^{T}\left(\ln|\mathbf{F}_t| + \nu_t'\mathbf{F}_t^{-1}\nu_t\right). \quad (6.1)$$

where w_t are constants, for which details can be found in Durbin and Koopman (2001, Section 7.2.2), or see Appendix D, and where ψ is the vector of unknown parameters of the state space model. For the models discussed in this chapter, the w_t terms are determined and given in the following sections. $\mathbf{F}_t$ and ν_t are defined by the Kalman filter in (3.51) and (3.47). Equation (6.1) is referred to as 'prediction error decomposition' of the log-likelihood function. For a complete derivation, see Harvey (1989, Section 3.4). The description of the estimation process will proceed in terms of the local level seasonal model.

For the univariate local level seasonal model given in state space form in (4.6) - (4.8), three parameters require estimation. The multivariate seasonal local level model described in (4.9) to (4.12), has two existing sub-series ($K = 2$), and hence nine parameters in the model.

It is possible to reduce the dimension of the parameter vector by concentrating out one parameter (σ_c^2) from the log-likelihood function. This has the advantage of reducing the dimensionality of the numerical search and it can improve the numerical stability of the optimisation (Zivot and Wang, 2006, p558). Harvey (1989, p183) states that it is "not only computationally efficient but is also likely to give more reliable results". For example, if the univariate model is considered, it is common practice to take the variance of the measurement error, $\sigma_{U,\,\varepsilon}^2$, to be the concentrated parameter, resulting in the signal to noise ratios as the parameters to be estimated. Alternatively, one of the variance terms from the diagonal of the covariance matrix $\mathbf{Q}$ (see (4.8)) could be used. Harvey (1989, p183) also suggests that the parameter chosen for (σ_c^2) should not be too close to zero as numerical problems may occur due to the relative variances becoming large.

The concentrated diffuse log-likelihood, which will be denoted by $\ln \mathrm{L}_{dc}$, is maximised with respect to the elements of ψ_c such that $\psi = \left(\psi_c',\ \sigma_c^2\right)'$. The elements

chosen for optimisation will depend on the model and choice of the parameter to be concentrated. The following subsections will describe ln L_{dc} in detail for both the univariate and multivariate models.

6.2.1 Univariate Model

For the univariate model, the log-likelihood function (6.1) may be simplified as $\mathbf{F}_t$ is scalar (denoted by F_t) and $K = 1$. For the local level seasonal model, $d = 4$, and the second term of equation (6.1) is calculated to be $-2\ln 2$, using the exact initial Kalman filter given in (Durbin and Koopman, 2001, Section 5.2.1). Equation (6.1) may therefore be rewritten as

$$\ln \mathrm{L}_d = -\frac{T}{2}\ln(2\pi) - 2\ln 2 - \frac{1}{2}\sum_{t=5}^{T}\ln F_t - \frac{1}{2}\sum_{t=5}^{T}\frac{\nu_t^2}{F_t} \tag{6.2}$$

This result is derived in Appendix D.2.

The model parameters for the univariate model are $\left(\sigma^2_{U,\,\eta},\ \sigma^2_{U,\,\omega},\ \sigma^2_{U,\,\varepsilon}\right)'$. If the measurement error variance parameter, $\sigma^2_{U,\,\varepsilon}$ is chosen as the concentrated parameter, then the covariance matrix $\mathbf{\Omega}^{(U)}$ for the local level seasonal model is defined to be (refer to equation (4.8))

$$\mathbf{\Omega}^{(U)} = \begin{pmatrix} \mathrm{Var}\,(\mathbf{G}\gamma_t) & \mathbf{0} \\ \mathbf{0} & \mathrm{H} \end{pmatrix} = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 & 0\,0 & 0 \\ 0 & \sigma^2_{U,\,\omega} & 0\,0 & 0 \\ 0 & 0 & 0\,0 & 0 \\ 0 & 0 & 0\,0 & 0 \\ 0 & 0 & 0\,0 & \sigma^2_{U,\,\varepsilon} \end{pmatrix} = \sigma^2_{U,\,\varepsilon} \begin{pmatrix} q_{U,\,\eta} & 0 & 0\,0 & 0 \\ 0 & q_{U,\,\omega} & 0\,0 & 0 \\ 0 & 0 & 0\,0 & 0 \\ 0 & 0 & 0\,0 & 0 \\ 0 & 0 & 0\,0 & 1 \end{pmatrix} \tag{6.3}$$

where $q_{U,\,\eta} = \sigma^2_{U,\,\eta}/\sigma^2_{U,\,\varepsilon}$ and $q_{U,\,\omega} = \sigma^2_{U,\,\omega}/\sigma^2_{U,\,\varepsilon}$. Thus, $q_{U,\,\eta}$ and $q_{U,\,\omega}$ are the signal-to-noise ratios for the level and seasonal components respectively. Recall that the definition of a dummy seasonal component (3.6), has one disturbance term $(\omega_{U,\,t})$ which is connected to the first seasonal factor in the state vector. Subsequently, for a quarterly series, there are two diagonal elements of $\mathbf{\Omega}^{(U)}$ which are zero as shown in (6.3). For an application of the concentrated log-likelihood function using the signal-to-noise ratios, see Dagum and Quenneville (1993).

In this case, it is also necessary to ensure that the model parameters to be estimated are positive. Rather than constraining the parameters, $q_{U,\,\eta}$ and $q_{U,\,\omega}$ in the estimation process, a square root transformation is applied, so that the maximisation is performed with respect to $\sqrt{q_{U,\,\eta}}$, and $\sqrt{q_{U,\,\omega}}$, where $-\infty < \sqrt{q_{U,\,\eta}}, \sqrt{q_{U,\,\omega}} < \infty$. An alternative transformation which is often used in practice is to take the logarithm of the parameters. This will yield estimates which are unconstrained. (See Durbin and Koopman, 2001, Section 7.3). In terms of the resulting values of the variance estimates, there seems to be little difference associated with using the alternate transformation. The square root transformation will be utilised here, to be consistent with the transformation required for the multivariate model parameters described in the following sub-section.

With the concentrated parameter $\sigma^2_{U,c} = \sigma^2_{U,\,\varepsilon}$, the parameters to be estimated are $\psi_c^{(U)} = \left(\sqrt{q_{U,\,\eta}},\ \sqrt{q_{U,\,\omega}}\right)'$. Hence, from (6.3) and referring to (4.8), it can be seen that

$$\text{Var}(\gamma_t) = \mathbf{Q} = \sigma^2_{U,\,\varepsilon}\mathbf{Q}^c \qquad \text{and} \qquad \text{Var}(\varepsilon_{U,\,t}) = H = \sigma^2_{U,\,\varepsilon} = \sigma^2_{U,\,\varepsilon}H^c. \tag{6.4}$$

where $H^c = 1$ and $\mathbf{Q}^c = \mathbf{Q}/\sigma^2_{U,\,\varepsilon}$.

The Kalman filter equations in (3.47) are amended such that

$$\begin{aligned} F_t &= \sigma^2_{U,\,\varepsilon}F_t^c \\ F_t^c &= \mathbf{Z}\mathbf{P}^c_{t|t-1}\mathbf{Z}^{'} + 1 \\ \mathbf{P}_{t|t-1} &= \sigma^2_{U,\,\varepsilon}\mathbf{P}^c_{t|t-1} \end{aligned} \tag{6.5}$$

Substituting F_t, as given by (6.5), into the log-likelihood function (6.2), the concentrated diffuse log-likelihood function for the univariate model is given as

$$\ln \text{L}_{dc} = -\frac{T}{2}\ln(2\pi) - 2\ln 2 - \frac{(T-4)}{2}\ln(\sigma^2_{U,\,\varepsilon}) - \frac{1}{2}\sum_{t=5}^{T}\ln F_t^c - \frac{1}{2\sigma^2_{U,\,\varepsilon}}\sum_{t=5}^{T}\frac{\nu_t^2}{F_t^c} \tag{6.6}$$

where ν_t and F_t^c do not depend on $\sigma^2_{U,\,\varepsilon}$. More detailed steps are given in Appendix D.2.

To determine the value of $\sigma^2_{U,\,\varepsilon}$ conditional on a given value of $\psi_c^{(U)}$, the concentrated diffuse log-likelihood ($\ln \mathrm{L}_{dc}$) is differentiated with respect to $\sigma^2_{U,\,\varepsilon}$ and solved to give

$$\hat{\sigma}^2_{U,\,\varepsilon}(\psi_c^{(U)}) = \frac{1}{(T-4)} \sum_{t=5}^{T} \frac{\nu_t^2}{F_t^c}. \tag{6.7}$$

By substituting this expression back into (6.6), the concentrated log-likelihood can be rewritten as

$$\ln \mathrm{L}_{dc}(\psi_c^{(U)}) = -\frac{T}{2}\ln(2\pi) - 2\ln 2 - \frac{(T-4)}{2}\left(1 + \ln[\hat{\sigma}^2_{U,\,\varepsilon}(\psi_c^{(U)})]\right) - \frac{1}{2}\sum_{t=5}^{T} \ln F_t^c \tag{6.8}$$

This function is maximised with respect to the elements of $\psi_c^{(U)}$ (Harvey, 1989, p127) to obtain the maximum likelihood estimates of $\sqrt{q_{U,\,\eta}}$ and $\sqrt{q_{U,\,\omega}}$. For more detail on taking the derivatives of the likelihood function with respect to $\psi_c^{(U)}$, refer to Harvey (1989, Section 3.4.6).

Finally, to retrieve the estimates of the original univariate parameters, the following calculations are required. Let $\hat{u}_1$ and $\hat{u}_2$ represent the maximum likelihood estimates of $\sqrt{q_{U,\,\eta}}$ and $\sqrt{q_{U,\,\omega}}$ respectively. Then the estimated variances for the disturbance terms for each component in the univariate model are given by

$$\begin{aligned} \text{Level:} \quad & \hat{\sigma}^2_{U,\,\eta} = \hat{\sigma}^2_{U,c}\hat{u}_1^2, \\ \text{Seasonal:} \quad & \hat{\sigma}^2_{U,\,\omega} = \hat{\sigma}^2_{U,c}\hat{u}_2^2, \\ \text{Error:} \quad & \hat{\sigma}^2_{U,\,\varepsilon} = \hat{\sigma}^2_{U,c} \end{aligned} \tag{6.9}$$

Let the vector of the original univariate parameters be denoted by $\psi^{(U)}$ so that the estimates may be written as

$$\hat{\psi}^{(U)} = \left(\hat{\sigma}^2_{U,\,\eta},\ \hat{\sigma}^2_{U,\,\omega},\ \hat{\sigma}^2_{U,\,\varepsilon}\right)' \tag{6.10}$$

These estimates may be substituted into the original state space model and then application of the Kalman filter provides estimates of the state vector.

6.2.2 Multivariate Model

For the multivariate model, the choice for the concentrated parameter is not as straightforward as for the univariate model. It is not possible to concentrate any of the covariance matrices out of the likelihood function as a block. However, it is possible to concentrate out one of the diagonal elements from one of the component covariance matrices (Harvey, 1989, Section 8.2.2). The state space form for either the untransformed model or the transformed model may be used for the estimation process. So the concentrated parameter can be one of the diagonal elements of one of the corresponding component covariance matrices. Refer to (4.9) - (4.12) for the details of the state space form of the transformed model.

In choosing an appropriate diagonal element to be the concentrated parameter, the parametrization of the resulting covariance matrices needs to be considered. If the untransformed model is used, each diagonal element is a sum of two of the original parameters. For two sub-series ($K = 2$), the Model 2 structure of the covariance matrix for the level component was defined in (3.22). The equivalent matrix for the measurement error is

$$\boldsymbol{\Sigma}_{(m),\,\varepsilon} = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 & \sigma_\varepsilon^2 \\ \sigma_\varepsilon^2 & \sigma_\varepsilon^2 + \sigma_{2\varepsilon^*}^2 \end{pmatrix}. \tag{6.11}$$

For example, the first diagonal element in (6.11) is $\sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2$ and could be assigned as the concentrated parameter.

Looking at the transformed model, each diagonal element is a linear combination of two or three of the original parameters. (Refer to (3.44) for the transformed covariance matrix of the level component). The transformed version of the measurement error covariance matrix (6.11) is given by

$$\boldsymbol{\Sigma}_{(M),\,\varepsilon} = \begin{pmatrix} 4\sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 + \sigma_{2\varepsilon^*}^2 & 2\sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 \\ 2\sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 & \sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2 \end{pmatrix} \tag{6.12}$$

Thus, if the first or second diagonal element is chosen as the concentrated parame-

ter, the resulting ratios within the covariance matrices would become unnecessarily complex compared to the result if the untransformed model was used. Therefore, the untransformed model is used for maximum likelihood estimation and the concentrated parameter is set as the first diagonal element of $\mathbf{\Sigma}_{(m),\,\varepsilon}$, from (6.11). Thus, if σ_c^2 denotes the concentrated parameter for the untransformed multivariate model, then

$$\sigma_c^2 = \sigma_\varepsilon^2 + \sigma_{1\varepsilon^*}^2. \tag{6.13}$$

The matrix $\mathbf{F}_{(M),t}$ given in (3.51) is for the transformed model. The corresponding matrix for the untransformed multivariate model is denoted by $\mathbf{F}_{(m),t}$. Taking the concentrated parameter into account, it is rewritten as

$$\begin{aligned} \mathbf{F}_{(m),t} &= \sigma_c^2 \mathbf{F}^c_{(m),t} \\ \mathbf{F}^c_{(m),t} &= \mathbf{Z}_{(m)} \mathbf{P}^c_{(m),t|t-1} \mathbf{Z}'_{(m)} \\ \mathbf{P}_{(m),t|t-1} &= \sigma_c^2 \mathbf{P}^c_{(m),t|t-1}. \end{aligned} \tag{6.14}$$

The untransformed multivariate model with level and quarterly dummy seasonal components is defined by equations (4.1) and (4.2). Using the exact initial Kalman filter given in (Durbin and Koopman, 2001, Section 7.2.2) with $K = 2$, $P_{\infty,t} \neq 0$ for the first four time points, giving $d = 4$. Hence, the second term of equation (6.1) may be determined and is calculated to be $-4\ln 2$. The concentrated diffuse log-likelihood function for the multivariate model is determined by substituting for (6.14) into (6.1) and is given by the following:

$$\begin{aligned} \ln \mathrm{L}_{(m),dc} = &-T\ln(2\pi) - 4\ln 2 - (T-4)\ln(\sigma_c^2) - \frac{1}{2}\sum_{t=5}^{T} \ln|\mathbf{F}^c_{(m),t}| \\ &-\frac{1}{2\sigma_c^2}\sum_{t=5}^{T} \nu_t' \left(\mathbf{F}^c_{(m),t}\right)^{-1} \nu_t. \end{aligned} \tag{6.15}$$

The result in (6.15) is derived in Appendix D.3. Note that for the multivariate model, there are Kq elements in the state vector which are non-stationary. For the

univariate model, $q = 4$, hence if $K = 2$, there are 8 non-stationary elements in the state vector $\alpha_{(m),t}$.

When ln $\mathrm{L}_{(m),c}$ (6.15) is differentiated with respect to σ_c^2 and set to zero, the result is

$$\hat{\sigma}_c^2 = \frac{1}{2(T-4)} \sum_{t=5}^{T} \nu_t' \left(\mathbf{F}_{(m),t}^c\right)^{-1} \nu_t. \tag{6.16}$$

For the derivation of the general case see Koopman *et al.* (1999, p140).

The covariance matrix of the state space form of the model which includes a slope component was given in (3.32). This is modified to remove the slope component parameters. Since the measurement error is moved into the state vector, the matrix $\mathbf{H}$ becomes a matrix of zeros. If the covariance matrix for the untransformed model is denoted by $\mathbf{\Omega}^{(m)}$ then:

$$\mathbf{\Omega}^{(m)} = \begin{pmatrix} \mathbf{\Sigma}_{(m),\,\eta} & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{\Sigma}_{(m),\,\omega} & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{\Sigma}_{(m),\,\varepsilon} & \mathbf{0}_K \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K \end{pmatrix}. \tag{6.17}$$

Let the original nine parameters in this multivariate model be denoted

$$\psi^{(m)} = \left(\sigma_\eta^2,\ \sigma_{1\eta^*}^2,\ \sigma_{2\eta^*}^2,\ \sigma_\omega^2,\ \sigma_{1\omega^*}^2,\ \sigma_{2\omega^*}^2,\ \sigma_\varepsilon^2,\ \sigma_{1\varepsilon^*}^2,\ \sigma_{2\varepsilon^*}^2\right)'. \tag{6.18}$$

For estimation, the parametrization of the component covariance matrices ($\mathbf{\Sigma}_{(m),\,\eta}$, $\mathbf{\Sigma}_{(m),\,\omega}$ and $\mathbf{\Sigma}_{(m),\,\varepsilon}$) requires two important characteristics. Firstly, each of the estimated component covariance matrices needs to be positive semi-definite, and secondly, the relationships between the elements within each matrix needs to be retained. It is also noted that the off-diagonal elements here are required to be positive.

To address the first characteristic, consider the resulting covariance matrix for

the level component, $\mathbf{\Sigma}_{(cm),\,\eta}$ such that,

$$\begin{aligned}\mathbf{\Sigma}_{(m),\,\eta} &= \sigma_c^2 \mathbf{\Sigma}_{(cm),\,\eta} \\ &= \sigma_c^2 \begin{pmatrix} \frac{\sigma_\eta^2}{\sigma_c^2} + \frac{\sigma_{1\eta^*}^2}{\sigma_c^2} & \frac{\sigma_\eta^2}{\sigma_c^2} \\ \frac{\sigma_\eta^2}{\sigma_c^2} & \frac{\sigma_\eta^2}{\sigma_c^2} + \frac{\sigma_{2\eta^*}^2}{\sigma_c^2} \end{pmatrix}\end{aligned} \tag{6.19}$$

A necessary and sufficient condition for a symmetric matrix $\mathbf{A}$ to be positive definite is (Harville, 1997, Theorem 14.9.5).

> 'The $n \times n$ matrix $\mathbf{A}$ is positive definite if and only if the determinants of all n of the leading principal sub-matrices $\mathbf{A}_1,\ \mathbf{A}_2, \ldots, \mathbf{A}_n$ are positive'.

If $\mathbf{A}$ is a 2×2 symmetric matrix:

$$\mathbf{A} = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \tag{6.20}$$

this definition simplifies to showing that $\mathbf{A}$ is positive definite if and only if a, b, and $ad - b^2$ are all positive (Harville, 1997, p280, (38b)). Looking at (6.19),

$$\begin{aligned} a &= \frac{\sigma_\eta^2}{\sigma_c^2} + \frac{\sigma_{1\eta^*}^2}{\sigma_c^2}, \qquad b = \frac{\sigma_\eta^2}{\sigma_c^2} \\ \text{and} \qquad ad - b^2 &= \left(\frac{\sigma_\eta^2}{\sigma_c^2} + \frac{\sigma_{1\eta^*}^2}{\sigma_c^2}\right) \times \left(\frac{\sigma_\eta^2}{\sigma_c^2} + \frac{\sigma_{2\eta^*}^2}{\sigma_c^2}\right) - \left(\frac{\sigma_\eta^2}{\sigma_c^2}\right)^2 \\ &= \frac{1}{\sigma_c^4}\left(\sigma_\eta^2\sigma_{1\eta^*}^2 + \sigma_\eta^2\sigma_{2\eta^*}^2 + \sigma_{1\eta^*}^2\sigma_{2\eta^*}^2\right). \end{aligned} \tag{6.21}$$

If the estimates of each term within each element is positive when recovered, then $a > 0$, $b > 0$ and $ad - b^2 > 0$. Therefore, the covariance matrix (6.19) will be positive definite. Thus, the parameters for maximum likelihood estimation need to be chosen such that each term is positive. For the level parameters, if

$$\frac{\sigma_\eta^2}{\sigma_c^2} > 0, \qquad \frac{\sigma_{1\eta^*}^2}{\sigma_c^2} > 0, \qquad \frac{\sigma_{2\eta^*}^2}{\sigma_c^2} > 0, \tag{6.22}$$

then the requirement will be met. This is done in a similar way for the seasonal parameters.

To ensure that the estimated terms are non-negative, the square root transformation is used in the the maximum likelihood estimation process. Let the elements

of the square root terms be labelled x_1 to x_8. Therefore, the concentrated log-likelihood is maximised with respect to the elements of $\psi_c^{(m)} = (x_1 \dots x_8)'$ which are:

$$\begin{aligned}
\text{Level:} \quad & x_1 = \sqrt{\frac{\sigma_\eta^2}{\sigma_c^2}}, \quad x_2 = \sqrt{\frac{\sigma_{1\eta^*}^2}{\sigma_c^2}}, \quad x_3 = \sqrt{\frac{\sigma_{2\eta^*}^2}{\sigma_c^2}} \\
\text{Seasonal:} \quad & x_4 = \sqrt{\frac{\sigma_\omega^2}{\sigma_c^2}}, \quad x_5 = \sqrt{\frac{\sigma_{1\omega^*}^2}{\sigma_c^2}}, \quad x_6 = \sqrt{\frac{\sigma_{2\omega^*}^2}{\sigma_c^2}} \\
\text{Error:} \quad & x_7 = \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_c^2}}, \qquad\qquad x_8 = \sqrt{\frac{\sigma_{2\varepsilon^*}^2}{\sigma_c^2}}.
\end{aligned} \tag{6.23}$$

To retrieve the estimates of the original multivariate parameters, the following calculations are required.

$$\begin{aligned}
\text{Level:} \quad & \hat{\sigma}_\eta^2 = \hat{\sigma}_c^2 \hat{x}_1^2, \quad \hat{\sigma}_{1\eta^*}^2 = \hat{\sigma}_c^2 \hat{x}_2^2, \qquad \hat{\sigma}_{2\eta^*}^2 = \hat{\sigma}_c^2 \hat{x}_3^2 \\
\text{Seasonal:} \quad & \hat{\sigma}_\omega^2 = \hat{\sigma}_c^2 \hat{x}_4^2, \quad \hat{\sigma}_{1\omega^*}^2 = \hat{\sigma}_c^2 \hat{x}_5^2, \qquad \hat{\sigma}_{2\omega^*}^2 = \hat{\sigma}_c^2 \hat{x}_6^2 \\
\text{Error:} \quad & \hat{\sigma}_\varepsilon^2 = \hat{\sigma}_c^2 \hat{x}_7^2, \quad \hat{\sigma}_{1\varepsilon^*}^2 = \hat{\sigma}_c^2 (1 - \hat{x}_7^2), \quad \hat{\sigma}_{2\varepsilon^*}^2 = \hat{\sigma}_c^2 \hat{x}_8^2
\end{aligned} \tag{6.24}$$

From (6.18), the estimate of $\psi^{(m)}$ is

$$\hat{\psi}^{(m)} = \left(\hat{\sigma}_\eta^2,\ \hat{\sigma}_{1\eta^*}^2,\ \hat{\sigma}_{2\eta^*}^2,\ \hat{\sigma}_\omega^2,\ \hat{\sigma}_{1\omega^*}^2,\ \hat{\sigma}_{2\omega^*}^2,\ \hat{\sigma}_\varepsilon^2,\ \hat{\sigma}_{1\varepsilon^*}^2,\ \hat{\sigma}_{2\varepsilon^*}^2\right)'. \tag{6.25}$$

For all except one of the estimates of the original multivariate parameters, this parametrization ensures that they are non-negative. To ensure that $\hat{\sigma}_{1\varepsilon^*}^2$ is non-negative, the maximum likelihood estimate for x_7 will need to be constrained so that $0 \leq x_7^2 \leq 1$. This can be achieved in the `S-PLUS` software by setting a lower bound of -1 and an upper bound of 1 on the estimate for x_7. With this constraint in place, the component covariance matrices will be positive semi-definite.

To obtain the values corresponding to the estimated univariate parameters contained in $\hat{\psi}^{(U)}$ (6.10), the expressions for the total parameter values are calculated from the estimated multivariate parameters:

$$\begin{aligned}
\text{Level:} \quad & \hat{\sigma}_{tot,\eta}^2 = 4\hat{\sigma}_\eta^2 + \hat{\sigma}_{1\eta^*}^2 + \hat{\sigma}_{2\eta^*}^2 \\
\text{Seasonal:} \quad & \hat{\sigma}_{tot,\omega}^2 = 4\hat{\sigma}_\omega^2 + \hat{\sigma}_{1\omega^*}^2 + \hat{\sigma}_{2\omega^*}^2 \\
\text{Error:} \quad & \hat{\sigma}_{tot,\varepsilon}^2 = 4\hat{\sigma}_\varepsilon^2 + \hat{\sigma}_{1\varepsilon^*}^2 + \hat{\sigma}_{2\varepsilon^*}^2
\end{aligned} \tag{6.26}$$

The vector containing the aggregate parameters in (6.26), together with the remaining multivariate parameters will be denoted by $\hat{\psi}^{(M)}$, and is given by

$$\hat{\psi}^{(M)} = \left(\hat{\sigma}^2_{tot,\eta},\ \hat{\sigma}^2_{\eta},\ \hat{\sigma}^2_{1\eta^*},\ \hat{\sigma}^2_{tot,\omega},\ \hat{\sigma}^2_{\omega},\ \hat{\sigma}^2_{1\omega^*},\ \hat{\sigma}^2_{tot,\varepsilon},\ \hat{\sigma}^2_{\varepsilon},\ \hat{\sigma}^2_{1\varepsilon^*}\right)'. \tag{6.27}$$

The corresponding vector containing the known quantities of the values in (6.27) will be denoted by $\psi^{(M)}$.

Alternative Concentrated Parameter

It may be necessary to change the element assigned to be the concentrated parameter, σ^2_c. This would be advisable when the value of $\sigma^2_{1\varepsilon^*} < \sigma^2_{2\varepsilon^*}$ which would occur when the c-ratio for the measurement error is below unity. In this case, to avoid having the concentrated parameter too close to zero, σ^2_c from (6.13) would become the larger of the two diagonal elements of $\mathbf{\Sigma}_{(m),\,\varepsilon}$ given in (6.11), thus

$$\sigma^2_c = \sigma^2_{\varepsilon} + \sigma^2_{2\varepsilon^*}. \tag{6.28}$$

The calculations applied in the previous section are performed similarly here, taking (6.28) into account. The parameters for the level and seasonal component remain the same as those in (6.23). However, the parameters associated with the error component become

$$\text{Error:} \qquad x_7 = \sqrt{\frac{\sigma^2_{\varepsilon}}{\sigma^2_c}}, \qquad x_8 = \sqrt{\frac{\sigma^2_{1\varepsilon^*}}{\sigma^2_c}}. \tag{6.29}$$

To retrieve the original error parameters, the following equations replace the last line in (6.24):

$$\text{Error:} \qquad \hat{\sigma}^2_{\varepsilon} = \hat{\sigma}^2_c \hat{x}^2_7, \qquad \hat{\sigma}^2_{1\varepsilon^*} = \hat{\sigma}^2_c \hat{x}^2_8, \qquad \hat{\sigma}^2_{2\varepsilon^*} = \hat{\sigma}^2_c (1 - \hat{x}^2_7). \tag{6.30}$$

6.3 Simulation Experiment

To investigate the behaviour of the univariate and multivariate models for varying series lengths, a simulation experiment will be carried out. Data will be simulated for two sub-series and aggregated to obtain the total series for 13 different series lengths, $T = 20, 21, \ldots, 27, 28, 40, 80, 120, 240$. For quarterly data, this translates to series of length 5 to 7 years, 10, 20, 30 and 60 years. Time series of length 5 to 7 years are often considered as short series, 10 to 20 years is considered moderate series length and 30 and 60 years is chosen for the long length. Maximum likelihood estimation of the parameters will be carried out as described in Section 6.2 for each series length. The results are summarised in Section 6.4.

6.3.1 Generation of Data Series

The parameters for the aggregate series are chosen as those defined for *Series 2* in (4.22) which are

$$\hat{\sigma}^2_{U,\eta} = 0.5, \quad \hat{\sigma}^2_{U,\omega} = 1.0, \quad \hat{\sigma}^2_{U,\varepsilon} = 1.0 \tag{6.31}$$

This choice avoids having the level parameter ($\hat{\sigma}^2_{U,\eta}$) too close to the boundary value of zero, which can cause complications for the maximum likelihood procedure.

To determine the parameters for the sub-series, firstly the relative values of the parameters given by the design need to be chosen from those defined in Section 4.3.1. Two sets of sub-series will be investigated. These are design *A1a14* and design *A1b33* which have relative efficiencies (with exact parameters) of 1.18 and 1.86 respectively (see Table 4.8 and Table 4.9 respectively). Design *A1a14* is chosen from the 'a' designs as it has a relative efficiency approximately midway between the lowest (1.0) and the highest (1.28) for the 'a' designs. Similarly for the 'b' designs, the lowest relative efficiency is 1.39 and the highest is 2.59 and so design *A1b33* is chosen. From Tables 4.1, 4.2 and 4.3, these two designs have different c-ratios but the same correlations:

$$\begin{aligned} &\textit{A1a14}: \ c_\omega = 1, \quad c_\eta = 20, \quad c_\varepsilon = 20, \quad \rho_\omega = 0.1, \quad \rho_\eta = 0.2, \quad \rho_\varepsilon = 0.2; \\ &\textit{A1b33}: \ c_\omega = 10, \quad c_\eta = 0.1, \quad c_\varepsilon = 0.1, \quad \rho_\omega = 0.1, \quad \rho_\eta = 0.2, \quad \rho_\varepsilon = 0.2. \end{aligned} \tag{6.32}$$

Given the c-ratios, the correlation coefficient for each component and the series parameters for the aggregate series, the multivariate seasonal parameters may be calculated from the equations given in (4.17), and similarly for the level and error parameters. Thus, the exact values of the nine multivariate parameters for each design are:

Table 6.1: Parameters for sub-series design *A1a14* and *A1b33*.

Parameter	*A1a14*	*A1b33*
σ^2_{η}	0.01962	0.02578
$\sigma^2_{1\eta^*}$	0.41919	0.01498
$\sigma^2_{2\eta^*}$	0.00232	0.38188
σ^2_{ω}	0.04545	0.02718
$\sigma^2_{1\omega^*}$	0.40909	0.83248
$\sigma^2_{2\omega^*}$	0.40909	0.05878
σ^2_{ε}	0.03925	0.05157
$\sigma^2_{1\varepsilon^*}$	0.83837	0.02997
$\sigma^2_{2\varepsilon^*}$	0.00463	0.76377

The data for each sub-series are generated from the model equations given by (4.1) for $T = 160$, to provide data for values of $T = 20$, 21,..., 27, 28, 40, 80, 120. For series length $T = 240$, data for $T = 280$ was generated. The following steps were used.

Level Component

Step 1: An $N \times 1$ vector of independent values is generated for the common disturbance term, η_t, from the $N(0, \sigma^2_{\eta})$ distribution. This vector is copied and appended so that it becomes a matrix with T rows and $K = 2$ columns where both columns are identical.

Step 2: The $T \times K$ values for η^*_{kt}, $(k = 1, 2)$ are generated from the multivariate normal distribution with zero mean and covariance matrix with $\sigma^2_{1\eta^*}$ and $\sigma^2_{2\eta^*}$ on the main diagonals with zero off-diagonal elements. This yields an $T \times K$ matrix. Note

that this is equivalent to generating a $T \times 1$ vector from the $N(0, \sigma^2_{1\eta^*})$ distribution and a $N \times 1$ vector from the $N(0, \sigma^2_{2\eta^*})$ distribution and then forming a $T \times 2$ matrix from the two vectors.

Step 3: The two matrices from Step 1 and Step 2 are then added together to form the set of random errors for the level component. The values in column 1 are the T random errors for sub-series 1 and the values in column 2 are the T random errors for sub-series 2.

Step 4: The initial values for each sub-series, ($L_{1,1}$ and $L_{2,1}$) are both set to a value of 5 and for convenience put into a 1×2 vector.

Step 5: The level component values for $t = 2, \ldots, T$ can now be calculated using $L_{k,t+1} = L_{k,t} + (\eta_t + \eta^*_{kt})$ where $(\eta_t + \eta^*_{kt})$ is a single value taken from row t and column k of the resulting matrix from Step 3. This generates the required T values for the level component for each sub-series.

Seasonal Component

Step 1: Step 1 above is repeated but with the common seasonal error term $\omega_t \sim N(0, \sigma^2_\omega)$ replacing $\eta_t \sim N(0, \sigma^2_\eta)$.

Step 2: Step 2 above is repeated with ω^*_{kt} replacing η^*_{kt}. Also $\sigma^2_{1\omega^*}$ and $\sigma^2_{2\omega^*}$ replace $\sigma^2_{1\eta^*}$ and $\sigma^2_{2\eta^*}$ respectively.

Step 3: Add the two matrices from Step 1 and Step 2 to form the set of random errors for the seasonal component.

Step 4: The initial values are set for the quarterly dummy seasonal component, S_{kt}. For $t = 1$, let $S_{1,1} = -1.5$ and $S_{2,1} = -1.5$. Let other initial values represent $t = 0$ and $t = -1$ then write $S_{1,0} = -1$ and $S_{2,0} = -1$ and $S_{1,-1} = 0.5$ and $S_{2,-1} = 0.5$.

Step 5: The seasonal component can be calculated for $t = 2, \ldots, T$ using $S_{k,t+1} = -\sum_{j=1}^{s-1} S_{k,t+1-j} + (\omega_t + \omega^*_{kt})$, where $(\omega_t + \omega^*_{kt})$ is a single value taken from row t and column k of the resulting matrix from Step 3. This generates the required T values for the seasonal component for each sub-series.

Data

To build the sub-series data, the equation for Y_{kt} is utilised:

$$Y_{kt} = L_{kt} + S_{kt} + (\varepsilon_t + \varepsilon^*_{kt}). \tag{6.33}$$

To obtain the measurement error values $(\varepsilon_t + \varepsilon^*_{kt})$, again, Steps 1 to 3 in Section 6.3.1 are repeated, but with ε replacing η. The values of Y_{1t} and Y_{2t} can now be generated by adding the resulting $T \times 2$ matrices of level, seasonal and error components. This generates two series of length $T = 160$. To obtain the series length of $T = 120$, the first 40 values are cut from the two generated series to remove any reliance on initial values. To obtain the required series with $T = 20$, 21,..., 27, 28, 40, 80, the values for the $t = 1, \ldots, T$ time points are retained from the cut series and the remaining $120 - T$ are discarded. This allows each series to be generated only once, and cut to required length. It also ensures that the observations remain the same for the corresponding time points.

To check asymptotic values, series of length 280 are generated and then the first 40 values are cut from the two generated series, to obtain the $T = 240$ as required.

The total series is then obtained by contemporaneously aggregating the two sub-series. This procedure is repeated to produce the required number of realisations of each sub-series and their aggregate.

6.3.2 Estimation of Parameters

Maximum likelihood estimation for both the univariate and multivariate models is performed using the `S+FinMetrics` function called *SsfFit*. Starting values are required to begin the iteration process for both models. Koopman *et al.* (1999) advise using the diagnostic of the estimated variance of the standardised prediction errors $(\hat{\sigma}^2)$ or scale factor in an unconcentrated model for a given value of ψ. This is produced as part of the output of the *SsfFit* function:

$$\hat{\sigma}^2(\psi) = \frac{1}{K(T-d)} \sum_{t=1}^{T} \nu_t' \left(\mathbf{F}_t\right)^{-1} \nu_t, \tag{6.34}$$

where Kd is the number of non-stationary elements and fixed regression effects in the state vector. Koopman *et al.* (1999, page 140) suggest choosing starting values such that $\hat{\sigma}^2(\psi_{start}) \approx 1$. Then, for models which are well specified, the value of $\hat{\sigma}^2(\hat{\psi}_{mle})$ should be very close to one (Zivot and Wang, 2006, page 553).

Zivot *et al.* (2004) discuss starting values for different examples. They state that the starting values may be chosen arbitrarily such as using a zero vector. Alternatively, the starting values are obtained by using estimates from the analysis of some other close model or are chosen to be "close to the true values" (Zivot *et al.*, 2004, p318, 322). For the purpose of this simulation exercise, starting values will be taken to be in the neighbourhood of the true values.

For the univariate model, starting values for the parameters to be estimated are required for $\psi_c^{(U)} = \left(\sqrt{q_{U,\,\eta}},\ \sqrt{q_{U,\,\omega}}\right)'$. The exact values for these expressions are calculated from the set of parameter values for the simulation (6.31) and rounded to 2 decimal places to be used as the starting values.

Similarly for the multivariate model, the concentrated log-likelihood is maximised with respect to the elements of $\psi_c^{(m)} = (x_1 \ldots x_8)'$ found in (6.23). The starting values are calculated using the exact parameter values given in Table 6.1 and rounded to 2 decimal places. For design *A1a14*, the concentrated parameter is assigned as in (6.13), and for design *A1b33* it is assigned as (6.28).

For each realisation, the maximum likelihood estimation process is performed for the univariate and multivariate models for each T, starting with the shortest series ($T = 20$). If relative function convergence is achieved in the maximum likelihood estimation process for this realisation, the process is repeated for the same realisation but with the next value of T in the set $T = 20,\ 21, \ldots, 27,\ 28,\ 40,\ 80,\ 120,\ 240$. If a particular optimisation process for a given realisation and T yields singular convergence, then that realisation for all T is discarded and the next realisation is started with $T = 20$. A valid run is obtained when one realisation for each length $T = 20,\ 21,\ \ldots,\ 27,\ 28,\ 40,\ 80,\ 120,\ 240$, produces relative function convergence. The process is repeated until 1000 valid runs are obtained and hence maximum likelihood estimates for both the univariate and the multivariate parameters for

each T are collected.

In this experiment, the methodology is tested with reference to the length of the series. Singular convergence can occur when there are too many degrees of freedom for the length of the series. The results are conditional on convergence for a realisation with values of T starting at $T = 20$. With nine parameters in the multivariate model and three in the univariate model, it is expected that with the shorter series, the proportion of realisations resulting in singular convergence will be higher than that with the longer series. Of course, if the number of parameters in the model is reduced, the estimation process may improve. However, the model will revert back to the compound symmetry model, for which it was found in Chapter 4 that there was no gain in using the multivariate model.

6.4 Results of Parameter Estimates for Varying Series Length

The results of the parameter estimates for 1000 realisations for each length of T will be given for the univariate and multivariate methods for the two chosen designs. The aim is to compare the results of the original univariate parameter estimates ($\hat{\sigma}^2_{U,\eta}$, $\hat{\sigma}^2_{U,\omega}$, $\hat{\sigma}^2_{U,\varepsilon}$) calculated by (6.9) with the corresponding multivariate parameter estimates for the aggregate series ($\hat{\sigma}^2_{tot,\eta}$, $\hat{\sigma}^2_{tot,\omega}$, $\hat{\sigma}^2_{tot,\varepsilon}$) given by (6.26). The results will be shown for each component of the two sub-series designs (*A1a14* and *A1b33*) for *Series 2*. The exact parameters for the aggregated series are equal for both designs. However, it is the relative values of the sub-series parameters that differ, as given by (6.32). The resulting exact parameter values are given in Table 6.1. It is of interest to determine whether the use of the sub-series yields more accurate parameter estimates of the aggregated series, for different values of T.

6.4.1 Design *A1a14*

The following box plots show the distribution of the estimated parameters of the level, seasonal and error components for the series lengths $T = 240$, 120, 80, 40, 28, 24, and 20. For the level component, the distribution of the parameter estimates is shown in Figure 6.1. The exact value of the parameter for the level component of the aggregate series is 0.5 (6.31), as shown by the horizontal dotted line.

It can be seen that the median estimate is an underestimate of the exact value for each value of T, for both the univariate and multivariate methods. This is a known bias which occurs with the estimation of parameters with the Kalman filter. The distributions are positively skewed for all $T < 120$. The range and interquartile range are slightly smaller for the parameters calculated using the multivariate model than with the univariate model. Therefore, for the level parameter, the multivariate model reduces the variability of the estimate, albeit only slightly. As expected, as T decreases, the distributions become more positively skewed with many outliers present. For $T \geq 120$, the distributions are almost symmetrical, with the medians approaching the true value of 0.5.

For the seasonal parameter, the results are shown in Figure 6.2. The exact parameter for the seasonal component is 1.0, as shown by the horizontal dotted line. In this plot, the difference between the multivariate and univariate models is more obvious. Again, it can be seen that the median estimate is an underestimate of the exact value for each value of T for both the univariate and multivariate methods. However, the median for the multivariate method is greater, and hence closer to the exact value for every value of T shown. The distributions are close to symmetrical for $T \geq 120$, and show that asymptotically the estimate approaches the true value. Again, as T decreases, the distributions become more positively skewed with outliers sometimes more extreme for the multivariate method. The range and interquartile range are smaller for most T for the multivariate method. In particular, the parameter distribution for the multivariate method for $T = 20$ is similar to that for the univariate method for $T = 40$. The medians are approximately equal, and the interquartile ranges are also similar, although there are more outliers for the

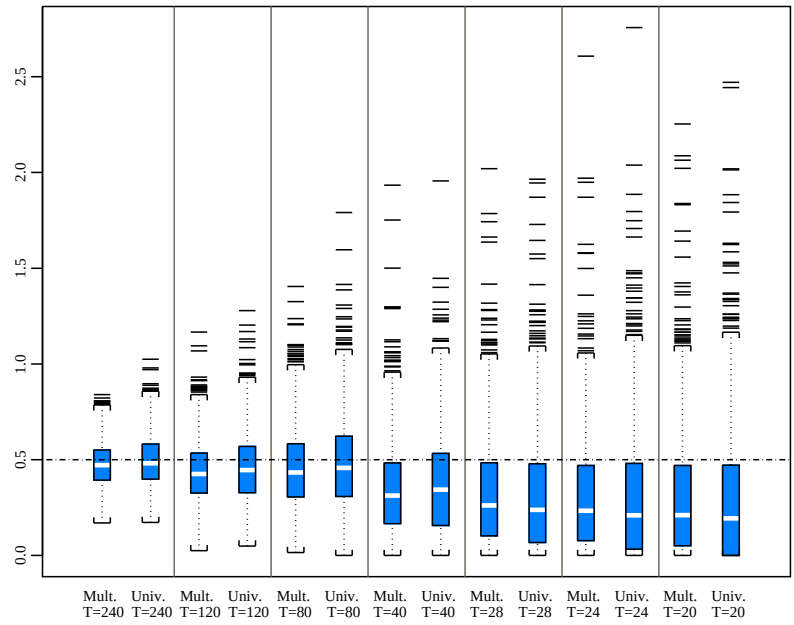


Figure 6.1: Distribution of the level parameter estimate for *Series 2* via univariate and multivariate methods for 1000 realisations of design *A1a14* with $T = 240$, 120, 80, 40, 28, 24, 20.

multivariate method for $T = 20$.

For the measurement error, the results are shown in Figure 6.3. The difference between this and the other two plots is that the results for the error parameter show an overestimate for each value of T. Similar to the results for the seasonal parameter, the medians for the multivariate method are closer to the exact value of 1.0 than those determined from the univariate method. The distributions also show less variability given the multivariate model as the interquartile range and range for each pair are smaller. Comparing the box plot for the multivariate model for $T = 20$ with the box plot for the univariate model for $T = 40$, it can be seen that the medians are similar and range and interquartile range are smaller for the multivariate model for $T = 20$ than for the univariate model for $T = 40$.

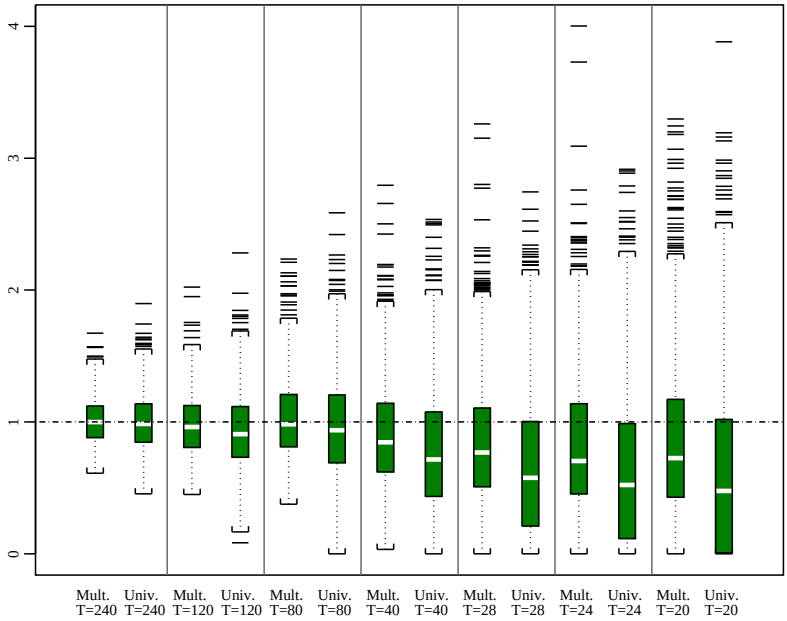


Figure 6.2: Distribution of the seasonal parameter estimate for *Series 2* via univariate and multivariate methods for 1000 realisations of design *A1a14* with $T = 240$, 120, 80, 40, 28, 24, 20.

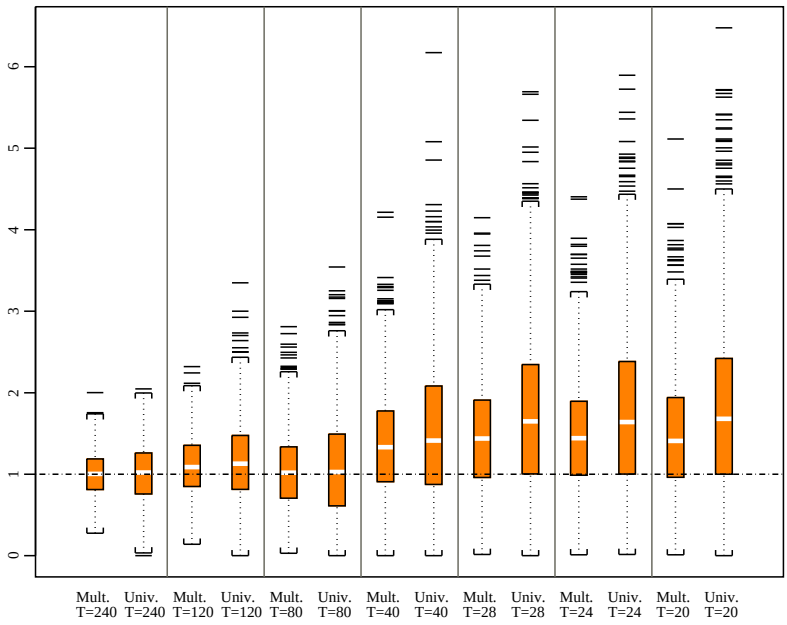


Figure 6.3: Distribution of the measurement error parameter estimate for *Series 2* via univariate and multivariate methods for 1000 realisations of design *A1a14* with $T = 240$, 120, 80, 40, 28, 24, 20.

Table 6.2: Medians of parameter estimates for *Series 2* with design *A1a14* for 1000 realisations with $T = 20, 21, \ldots, 27, 28, 40, 80, 120, 240$.

Series 2 Design *A1a14*	Level True=0.5		Seasonal True=1.0		Error True=1.0	
Median Estimate	Mult.	Univ.	Mult.	Univ.	Mult.	Univ.
T 20	0.210	0.194	0.726	0.477	1.410	1.680
21	0.221	0.185	0.694	0.476	1.422	1.656
22	0.228	0.192	0.699	0.494	1.435	1.620
23	0.233	0.214	0.703	0.515	1.441	1.624
24	0.234	0.210	0.704	0.523	1.442	1.641
25	0.246	0.215	0.734	0.531	1.451	1.663
26	0.244	0.220	0.746	0.572	1.441	1.656
27	0.245	0.226	0.769	0.566	1.433	1.654
28	0.261	0.238	0.768	0.576	1.438	1.649
40	0.313	0.344	0.846	0.716	1.332	1.413
80	0.434	0.457	0.983	0.937	1.019	1.026
120	0.426	0.446	0.962	0.908	1.088	1.130
240	0.472	0.481	0.999	0.984	1.003	1.020

The results for all values of T studied are given in Table 6.2. This table shows the median values for the estimated parameters of each component over the 1000 series generated. Not surprisingly, the estimates closest to the true parameter values are for $T = 240$, the series with the greatest length. However, it is seen here that the parameter estimates given by the multivariate method are closer to the exact value for every pair (univariate vs multivariate) for each value of T except for the level component for $T \geq 40$. That is, there is a marked improvement in the median estimate when the sub-series are used in the estimation process.

Table 6.3: Means and standard errors (in italics) of parameter estimates for design *A1a14* for 1000 realisations with $T = 20, 21, \ldots, 27, 28, 40, 80, 120, 240$.

Series 2 Design *A1a14*		Level True=0.5		Seasonal True=1.0		Error True=1.0	
Mean Estimate		Mult.	Univ.	Mult.	Univ.	Mult.	Univ.
T	20	0.314	0.308	0.862	0.639	1.499	1.804
		0.011	*0.012*	*0.019*	*0.021*	*0.024*	*0.035*
	21	0.312	0.308	0.848	0.635	1.520	1.813
		0.010	*0.012*	*0.018*	*0.021*	*0.024*	*0.033*
	22	0.314	0.309	0.840	0.643	1.522	1.807
		0.010	*0.011*	*0.017*	*0.020*	*0.023*	*0.033*
	23	0.316	0.315	0.836	0.647	1.515	1.787
		0.010	*0.011*	*0.016*	*0.020*	*0.023*	*0.033*
	24	0.313	0.314	0.844	0.650	1.499	1.777
		0.010	*0.011*	*0.017*	*0.020*	*0.023*	*0.032*
	25	0.312	0.310	0.849	0.654	1.498	1.788
		0.009	*0.011*	*0.016*	*0.019*	*0.022*	*0.032*
	26	0.315	0.314	0.845	0.660	1.501	1.772
		0.009	*0.011*	*0.015*	*0.019*	*0.023*	*0.032*
	27	0.319	0.318	0.849	0.670	1.492	1.748
		0.009	*0.010*	*0.015*	*0.018*	*0.022*	*0.031*
	28	0.324	0.322	0.843	0.673	1.494	1.747
		0.009	*0.010*	*0.015*	*0.018*	*0.022*	*0.031*
	40	0.355	0.372	0.906	0.784	1.379	1.516
		0.008	*0.009*	*0.013*	*0.016*	*0.021*	*0.027*
	80	0.457	0.481	1.017	0.962	1.041	1.093
		0.007	*0.007*	*0.009*	*0.012*	*0.014*	*0.020*
	120	0.439	0.460	0.981	0.936	1.102	1.148
		0.005	*0.006*	*0.008*	*0.010*	*0.011*	*0.016*
	240	0.476	0.494	1.011	0.999	1.004	1.013
		0.004	*0.004*	*0.005*	*0.007*	*0.008*	*0.012*

The means and standard errors of the parameter estimates are given in Table 6.3. The means are larger than the medians due to the positively skewed distributions. The means fluctuate slightly as T increases, probably due to the presence of outliers but overall become closer to their true parameter value as T reaches 240. As expected, the standard errors generally decrease as T increases. They are smaller for the multivariate method than for the univariate method. This is consistent with the variability shown in the box plots.

Table 6.4: Relative efficiency of the variances of the parameter estimates for design *A1a14* for 1000 realisations with $T = 20, 21, \ldots, 27, 28, 40, 80, 120, 240$.

A1a14		Level	Seasonal	Error
T	20	1.20	1.30	2.14
	21	1.27	1.31	2.00
	22	1.17	1.38	2.08
	23	1.26	1.42	2.00
	24	1.29	1.35	2.05
	25	1.34	1.37	2.04
	26	1.31	1.45	1.99
	27	1.21	1.39	1.89
	28	1.25	1.43	1.98
	40	1.23	1.49	1.78
	80	1.33	1.67	1.95
	120	1.35	1.63	1.99
	240	1.38	1.57	1.89

To compare the variability of the estimates for the two methods, the relative efficiencies of the parameter estimates have been calculated using the variances obtained from the results shown in Table 6.3. The results for each T are given in Table 6.4. This table highlights the gains which are achieved with the multivariate model. The relative efficiency varies between 1.17 and 1.38 for the level parameter. For the seasonal parameter, the relative efficiency shows greater gains than for the level parameter, with values ranging from 1.30 for $T = 20$ to 1.67 for $T = 80$. The gains are higher again for the measurement error parameter. These range from 1.78 for

$T = 40$ to 2.14 for $T = 20$. Thus the higher gains in the variance of the estimates are for the shorter series.

6.4.2 Design *A1b33*

The results for the parameter estimates for *Series 2*, design *A1b33*, are given in this section for the 1000 realisations for each length of T. Figure 6.4 shows the box plot of the distributions of the level parameter estimate for $T = 240$, 120, 80, 40, 28, 24 and 20, resulting from the univariate and multivariate models. The dotted horizontal line represents the exact parameter value.

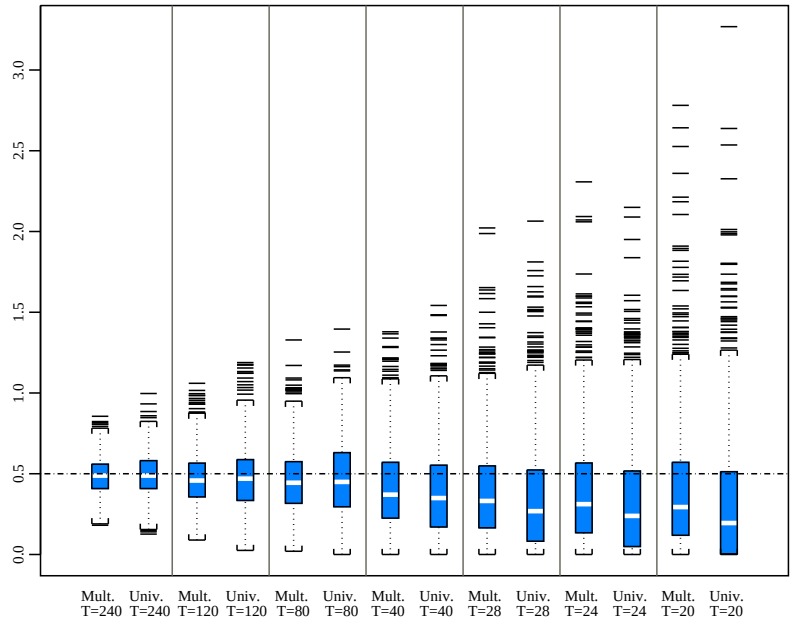


Figure 6.4: Distribution of the level parameter estimate for *Series 2* via univariate and multivariate methods for 1000 realisations of design *A1b33* with $T = 240$, 120, 80, 40, 28, 24, 20.

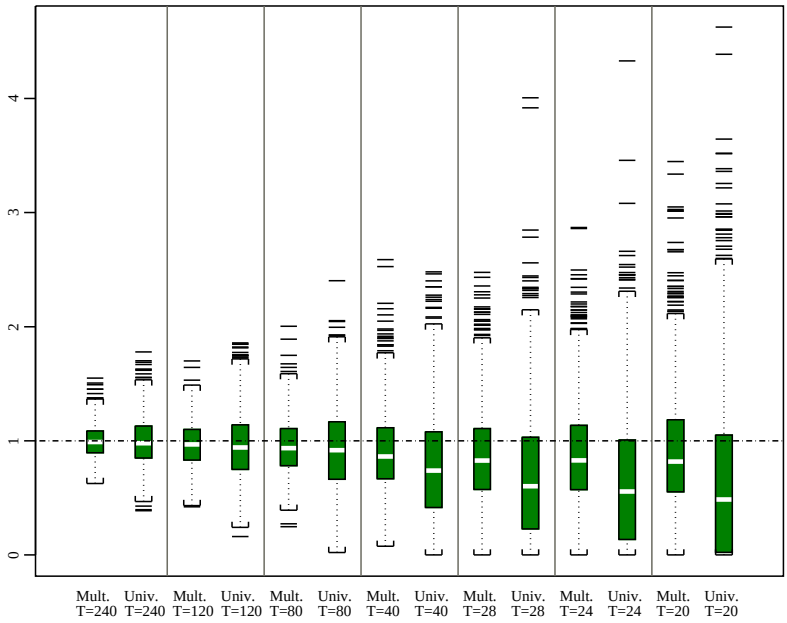


Figure 6.5: Distribution of the seasonal parameter estimate for *Series 2* via univariate and multivariate methods for 1000 realisations of design *A1b33* with $T = 240$, 120, 80, 40, 28, 24, 20.

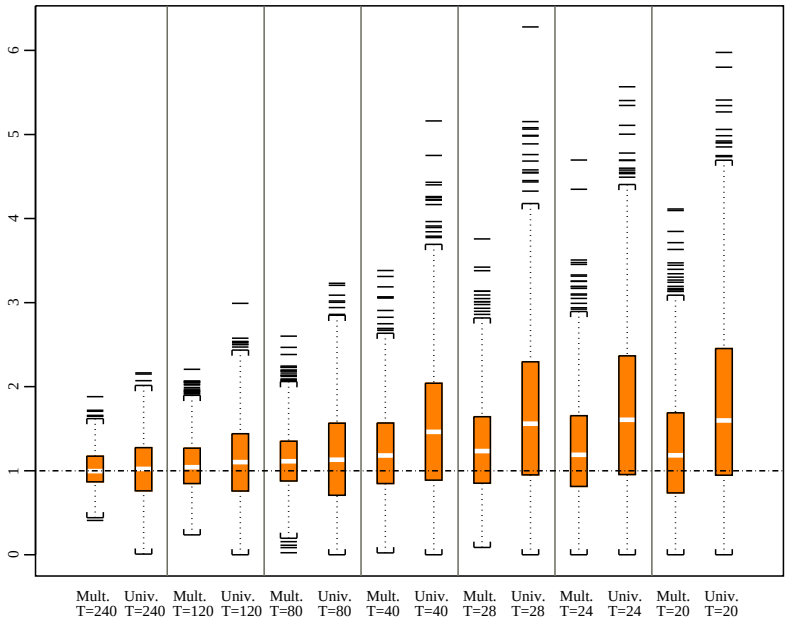


Figure 6.6: Distribution of the measurement error parameter estimate for *Series 2* via univariate and multivariate methods for 1000 realisations of design *A1b33* with $T = 240$, 120, 80, 40, 28, 24, 20.

As in the previous plots for design *A1a14*, Figure 6.4 shows that the median estimate is an underestimate of the true value and that the magnitude of the bias becomes larger as T decreases. The median for each T shown, except for $T \geq 120$, is closer to the true value if the multivariate model is used. Again, except for $T \geq 120$, the distributions are positively skewed, and become more so as the length of the series decreases. The interquartile range for the multivariate method is slightly smaller than for the univariate method. For this design, the dotted line for the true value crosses the section between the upper quartile and the median for each value of T displayed.

Figure 6.5 shows the distributions of the 1000 parameter estimates obtained for the seasonal component for different length series. There is a definite improvement in the precision of the estimates resulting from the multivariate model as shown by the medians. For all values of T displayed, the medians are much closer to the true value of 1.0. Notably, when $T = 28$, 24, and 20, the lower quartile of the multivariate method is approximately equal to the median for the corresponding T for the univariate method. The range and interquartile range are smaller for the multivariate method in each pair. Apart from $T = 40$, outliers are also less extreme for the multivariate method.

A similar result is shown for the measurement error component in Figure 6.6, except that the median of the estimates for each T is greater than the true value. Again, the medians given by the multivariate method are closer to the true value for each T displayed and the range and interquartile range are also obviously narrower.

More detail is given in Tables 6.5 and 6.6. Table 6.5 shows the median of the parameter estimates for each of the 13 values of T. For the level parameter, the median of the estimates at $T = 20$ for the multivariate model (0.29) is closer to the true value than the univariate estimate at $T = 28$ (0.27). For the seasonal parameter, the median of the estimates at $T = 20$ for the multivariate model (0.82) is closer to the true value than the univariate estimate (0.74) at $T = 40$.

Table 6.6 shows the means and standard errors for the parameter estimates for each component and for each series length. Due to the skewed distributions and

Table 6.5: Median of parameter estimates for design *A1b33* for 1000 realisations with $T = 20, 21, \ldots, 27, 28, 40, 80, 120, 240$.

Series 2 Design *A1b33*	Level True=0.5		Seasonal True=1.0		Error True=1.0	
Median Estimate	Mult.	Univ.	Mult.	Univ.	Mult.	Univ.
T 20	0.294	0.196	0.818	0.486	1.183	1.598
21	0.288	0.202	0.819	0.493	1.159	1.603
22	0.289	0.219	0.814	0.534	1.198	1.599
23	0.300	0.227	0.814	0.546	1.185	1.615
24	0.312	0.239	0.829	0.556	1.189	1.607
25	0.317	0.234	0.825	0.579	1.165	1.633
26	0.325	0.243	0.822	0.575	1.183	1.580
27	0.325	0.270	0.827	0.589	1.219	1.615
28	0.332	0.269	0.827	0.603	1.234	1.560
40	0.370	0.349	0.863	0.740	1.181	1.462
80	0.445	0.449	0.936	0.918	1.112	1.130
120	0.458	0.470	0.969	0.942	1.043	1.103
240	0.489	0.489	0.990	0.980	0.996	1.022

the outliers, the means vary slightly for values of T between 20 and 28. For the estimates of the level and seasonal parameters, the means increase as T increases from $T = 28$ to $T = 240$. The bias is negative for the level and seasonal parameter estimates but is positive for the measurement error parameter estimates for each value of T. This bias is reduced by using the multivariate model for every value of T except for the level parameter estimates for $T \geq 80$. These exceptions also occur for the medians in Table 6.5. Looking back to the results for design *A1a14*, Tables 6.3 and 6.2, show a similar result for $T \geq 40$. That is, the magnitude of the biases shown for the mean and median estimates for the level parameter is greater for the multivariate method. This is an interesting result which suggests there may be a series length, $T = T_L$, at which this 'cross-over' occurs. Thus, for $T > T_L$, the estimator of the level parameter has greater precision using the univariate model.

Table 6.6: Means and standard errors (in italics) of parameter estimates for design *A1b33* for 1000 realisations with $T = 20, 21, \ldots, 27, 28, 40, 80, 120, 240$.

Design *A1b33*		Level True=0.5		Seasonal True=1.0		Error True=1.0	
Mean Estimate		Mult.	Univ.	Mult.	Univ.	Mult.	Univ.
T	20	0.400	0.336	0.911	0.679	1.271	1.766
		0.012	*0.013*	*0.017*	*0.023*	*0.022*	*0.035*
	21	0.395	0.331	0.904	0.658	1.272	1.791
		0.012	*0.013*	*0.016*	*0.022*	*0.022*	*0.035*
	22	0.384	0.335	0.892	0.655	1.292	1.767
		0.011	*0.012*	*0.016*	*0.021*	*0.022*	*0.034*
	23	0.393	0.332	0.884	0.655	1.290	1.777
		0.011	*0.011*	*0.015*	*0.020*	*0.021*	*0.033*
	24	0.396	0.339	0.886	0.665	1.273	1.754
		0.011	*0.011*	*0.015*	*0.019*	*0.021*	*0.032*
	25	0.393	0.339	0.883	0.670	1.270	1.741
		0.010	*0.011*	*0.014*	*0.019*	*0.020*	*0.032*
	26	0.396	0.340	0.884	0.680	1.264	1.728
		0.010	*0.011*	*0.014*	*0.019*	*0.020*	*0.032*
	27	0.396	0.349	0.883	0.687	1.273	1.708
		0.010	*0.011*	*0.014*	*0.019*	*0.019*	*0.031*
	28	0.398	0.354	0.875	0.688	1.281	1.694
		0.010	*0.011*	*0.013*	*0.018*	*0.019*	*0.031*
	40	0.413	0.390	0.913	0.785	1.226	1.523
		0.008	*0.009*	*0.011*	*0.016*	*0.017*	*0.028*
	80	0.456	0.474	0.952	0.934	1.132	1.150
		0.006	*0.007*	*0.008*	*0.012*	*0.012*	*0.020*
	120	0.469	0.477	0.971	0.957	1.073	1.114
		0.005	*0.006*	*0.006*	*0.009*	*0.010*	*0.016*
	240	0.486	0.496	0.995	0.992	1.022	1.022
		0.004	*0.004*	*0.004*	*0.007*	*0.07*	*0.012*

The standard errors given in Table 6.6 decrease as T increases, and hence the estimates become more stable. To look at the comparison between the two methods more closely, the relative efficiencies of the variances have been determined and are given in Table 6.7. Since all relative efficiencies are greater than one, the multivariate method reduces the variability of the parameter estimates for each component. For this example, the relative efficiencies of the estimates of the level parameter range from 1.04 for $T = 23$ to 1.47 for $T = 120$. For the seasonal parameter, the gains are much greater, with values ranging from 1.70 for $T = 22$ to 2.28 for $T = 240$. For these two components, the relative efficiency increases for $T = 23$ up to $T = 120$. The lowest value of the relative efficiency for the measurement error parameter is 2.41 which is given for both $T = 22$ and $T = 120$. The highest value is 2.63 which is achieved for $T = 40$. For the measurement error, there is not a distinctive pattern across T.

Table 6.7: Relative efficiency of the parameter estimates for design *A1b33* for 1000 realisations with $T = 20, 21, \ldots, 27, 28, 40, 80, 120, 240$.

A1b33		Level	Seasonal	Error
T	20	1.13	1.79	2.50
	21	1.17	1.76	2.45
	22	1.18	1.70	2.41
	23	1.04	1.71	2.42
	24	1.06	1.73	2.48
	25	1.14	1.78	2.60
	26	1.15	1.83	2.60
	27	1.15	1.85	2.58
	28	1.19	1.89	2.61
	40	1.30	2.00	2.63
	80	1.44	2.27	2.70
	120	1.47	2.17	2.41
	240	1.37	2.28	2.68

The results shown in Table 6.7 for design *A1b33* are similar to those shown in Table 6.4 for design *A1a14*. For the estimate of the seasonal parameter, the relative efficiency of the variances are higher for *A1b33* than for *A1a14*, although all values are above unity. These results show that for the two examples studied, the use of the multivariate approach in the estimation of parameters improves the accuracy of the parameter estimates of the aggregated series. Similar results are expected for the parameter estimates of other designs, as these two examples (*A1a14* and *A1b33*) were chosen to be indicative of the '*a*' and '*b*' designs respectively (refer to Section 6.3.1).

6.5 Effect of Parameter Estimation on MSE of Model Components

In the previous section the model parameters were estimated by maximum likelihood using the prediction error decomposition. It has been shown that these estimates may be biased, especially for short to medium length time series. In this section, the effect of the two-stage process of the estimation of the state vector components and their corresponding mean squared errors are considered. The parameter estimates from the first stage are substituted for the true parameters in the second stage. When considering the variance of the seasonally adjusted series as discussed in Section 3.6.1, the prediction mean squared error (PMSE) of the seasonal component is required. It is therefore important to determine the effect of the parameter estimation on the PMSEs of the unobserved components.

The substitution of the estimated parameters in the theoretical expressions for the PMSE of the state vector components is known to produce an underestimate of the true PMSE in univariate models. This bias is often referred to as the 'naïve bias' and results from not accounting for the variability resulting from the estimation of the model parameters (Durbin and Koopman, 2001; Quenneville and Singh, 2005; Pfeffermann and Tiller, 2005). In this section, the naïve bias will be examined for different series lengths for the univariate model and for the multivariate model. It is not known how the naïve bias behaves in multivariate models, as previous literature

has focussed on univariate models such as the local level model as discussed in Quenneville and Singh (2005) and Pfeffermann and Tiller (2005). A trigonometric seasonal model is also considered in Pfeffermann and Tiller (2005).

Prior to defining the naïve bias, some notation is reviewed from Chapter 3, in light of the definitions given for $\psi^{(U)}$ (refer to (6.10)) and $\psi^{(m)}$ (6.18). In Section 3.5.3, the smoothed state vector estimator for the univariate model, $\mathbf{a}_{t|T}$, and its corresponding error variance matrix, $\mathbf{V}_{t|T}$, are defined in (3.52) and (3.54) respectively. These are calculated assuming the known parameter values, $\psi^{(U)}$. To be more specific as to whether known or estimated parameter values are used in the calculation, the definition will now include either $\psi^{(U)}$ (for known parameters) or $\hat{\psi}^{(U)}$ (for estimated parameters). That is, for known parameter values,

$$\begin{aligned} \mathbf{a}_{t|T} &= \mathrm{E}\left(\alpha_t|\, \mathbb{Y}_T,\ \psi^{(U)}\right) \\ \mathbf{V}_{t|T} &= \mathrm{Var}\left(\alpha_t|\, \mathbb{Y}_T,\ \psi^{(U)}\right) \end{aligned} \tag{6.35}$$

or, if estimated parameters are used in the application,

$$\begin{aligned} \hat{\mathbf{a}}_{t|T} &= \mathrm{E}\left(\alpha_t|\, \mathbb{Y}_T,\ \hat{\psi}^{(U)}\right) \\ \hat{\mathbf{V}}_{t|T} &= \mathrm{Var}\left(\alpha_t|\, \mathbb{Y}_T,\ \hat{\psi}^{(U)}\right). \end{aligned} \tag{6.36}$$

For the multivariate model, the smoothed state vector and its MSE matrix has an (M) subscript as defined in (3.56) and (3.57) respectively. The notation for the application of the smoother with known parameters is

$$\begin{aligned} \mathbf{a}_{(M),\,t|T} &= \mathrm{E}\left(\alpha_{(M),\,t}|\, \mathbb{Y}_{(M),\,T},\ \psi^{(M)}\right) \\ \mathbf{V}_{(M),\,t|T} &= \mathrm{Var}\left(\alpha_{(M),\,t}|\, \mathbb{Y}_{(M),\,T},\ \psi^{(M)}\right) \end{aligned} \tag{6.37}$$

and for estimated parameters, it is

$$\begin{aligned} \hat{\mathbf{a}}_{(M),\,t|T} &= \mathrm{E}\left(\alpha_{(M),\,t}|\, \mathbb{Y}_{(M),\,T},\ \hat{\psi}^{(M)}\right) \\ \hat{\mathbf{V}}_{(M),\,t|T} &= \mathrm{Var}\left(\alpha_{(M),\,t}|\, \mathbb{Y}_{(M),\,T},\ \hat{\psi}^{(M)}\right) \end{aligned} \tag{6.38}$$

The main focus of this thesis is on the variance of the seasonally adjusted aggregated series as discussed in Section 3.6.1. Recall that for the one-staged estimation

process, where the exact parameter values are used, the estimated seasonal component is denoted by $\hat{S}^U_{t|T}$, which is an element of $\mathbf{a}_{t|T}$ (6.35). For the two-staged estimation process, the univariate estimate of the seasonal component is denoted by $\hat{\mathcal{S}}^U_{t|T}$, which is an element of $\hat{\mathbf{a}}_{t|T}$. Its mean squared error, $MSE\left(\hat{\mathcal{S}}^U_{t|T}\right)$ is an element of $\hat{\mathbf{V}}_{t|T}$.

Similarly for the multivariate method, the estimate of the seasonal component from the two-staged process is denoted by $\hat{\mathcal{S}}^M_{t|T}$, and is an element of $\hat{\mathbf{a}}_{(M),\,t|T}$. Its mean squared error is given by by MSE $\left(\hat{\mathcal{S}}^M_{t|T}\right)$.

6.5.1 Naïve Bias

The naïve bias, which is the bias in the prediction mean squared error, will now be considered for the estimated seasonal component resulting from both the univariate and multivariate approaches. To calculate the naïve bias, the 'true' MSE is approximated using the simulated series component and the estimated series component as given in Pfeffermann and Tiller (2005, p903). The calculations here are based on 1000 simulated series for each series length $T = 20,\ 21,\ \ldots,27,\ 28,\ 40,\ 120$. For the seasonal component, it is

$$MSE^U_t = \sum_{i=1}^{1000} \frac{\left(\hat{\mathcal{S}}^U_{t|T,\,i} - S_{t,\,i}\right)^2}{1000}, \qquad i = 1, \ldots, 1000 \tag{6.39}$$

where $S_{t,i}$ is the simulated value of the seasonal component at time t for the ith generated data series and $\hat{\mathcal{S}}^U_{t|T,\,i}$ is the estimated seasonal component at time t determined by the Kalman smoother, given the vector of estimated parameters $\hat{\psi}^{(U)}$ for the ith generated data series of length T.

The multivariate form of (6.39) is given by

$$MSE^M_t = \sum_{i=1}^{1000} \frac{\left(\hat{\mathcal{S}}^M_{t|T,\,i} - S_{t,\,i}\right)^2}{1000}. \tag{6.40}$$

For the naïve bias, let $\widehat{MSE}^U_{i,t}$ denote the state variance calculated by the Kalman smoother with the estimated parameters ($\hat{\psi}^{(U)}$) given by the univariate model for the

ith simulated series. Similarly, $\widehat{MSE}^{M}_{i,t}$ is the notation adopted for the multivariate model. The biases of the estimators are therefore determined by

$$\text{Univariate:} \qquad d^U_{i,t} = \widehat{MSE}^{U}_{i,t} - MSE^U_t \tag{6.41}$$

$$\text{Multivariate:} \qquad d^M_{i,t} = \widehat{MSE}^{M}_{i,t} - MSE^M_t. \tag{6.42}$$

The mean bias over the 1000 realisations and its associated MSE are determined by

$$\text{Univariate:} \qquad \overline{d}^U_t = \sum_{i=1}^{1000} \frac{d^U_{i,t}}{1000}, \qquad \overline{d}^{U(2)}_t = \sum_{i=1}^{1000} \frac{(d^U_{i,t})^2}{1000} \tag{6.43}$$

$$\text{Multivariate:} \qquad \overline{d}^M_t = \sum_{i=1}^{1000} \frac{d^M_{i,t}}{1000}, \qquad \overline{d}^{M(2)}_t = \sum_{i=1}^{1000} \frac{(d^M_{i,t})^2}{1000}. \tag{6.44}$$

It is more informative to analyse a bias by considering its value relative to the true value. The relative values may be calculated for each $t = 1, \ldots, T$. If desired, the mean over T can be evaluated and expressed as a percentage. This is termed the 'mean percent relative bias' (Pfeffermann and Tiller, 2005). This value and the 'mean percent relative root mean squared error' are defined in Pfeffermann and Tiller (2005, p904) and are given here for both approaches as

Univariate:

$$\text{Rel-Bias} = \frac{100}{T} \sum_{t=1}^{T} \left(\frac{\overline{d}^U_t}{MSE^U_t} \right), \ \text{Rel-RMSE} = \frac{100}{T} \sum_{t=1}^{T} \left(\frac{\sqrt{\overline{d}^{U(2)}_t}}{MSE^U_t} \right); \tag{6.45}$$

Multivariate:

$$\text{Rel-Bias} = \frac{100}{T} \sum_{t=1}^{T} \left(\frac{\overline{d}^M_t}{MSE^M_t} \right), \ \text{Rel-RMSE} = \frac{100}{T} \sum_{t=1}^{T} \left(\frac{\sqrt{\overline{d}^{M(2)}_t}}{MSE^M_t} \right). \tag{6.46}$$

The naïve bias is calculated by Quenneville and Singh (2005) for a local level model for series of lengths $T = 40$ and $T = 100$. They found that the naïve bias is greater in absolute terms for the moderate length ($T = 40$) than for the longer series ($T = 100$). For the moderate length, the mean percent relative bias for 1000

realisations was reported to be -21.2%, whereas for the series with $T = 100$, the relative bias was -9.0%.

Pfeffermann and Tiller (2005) reproduce the experiment carried out by Quenneville and Singh (2005) with series of lengths $T = 40$ and $T = 100$ but extend the number of realisations to 5000. Their results for the naïve bias are slightly smaller, with the mean percent relative naïve bias for $T = 40$ reported as -18.5% and for $T = 100$, it is -7.6%.

Although these results are for a simpler model than is studied here, they show the effect of using the estimated parameters in the calculation of the predicted mean squared error of the state vector. It is expected that for this study, the naïve bias will be large and negative and that it will decrease (in absolute terms) as the length of the series is increased. The univariate model being investigated here is the local level seasonal model which differs to a local level model in that it also has a dummy seasonal component. The multivariate model is also considered here.

Results for Naïve Bias

The naïve bias for the seasonal component has been calculated for two different subseries designs for both the univariate and multivariate models for 1000 realisations. The results for design labelled *A1a14* for each value of $T = 20, 21, \ldots, 27, 28, 40, 120$ are given in Table 6.8. Similarly, the results for design labelled *A1b33* are given in Table 6.9. The Rel-Bias and Rel-RMSE for the univariate model are calculated with the equations in (6.45). The Rel-Bias and Rel-RMSE for the multivariate model are calculated with the equations in (6.46).

For design *A1a14*, Table 6.8 shows that the relative bias calculated for the univariate model ranges from -27.97 for $T = 24$, down to -6.71 for $T = 120$. The result for the relative bias varies slightly for T between 20 and 28, then decreases to -19.38 for $T = 40$, and then decreases markedly to -6.71 for $T = 120$. Note that although this model is different to the local level model studied by Quenneville and Singh (2005) and Pfeffermann and Tiller (2005), the results for $T = 40$ are of a similar magnitude.

The relative bias calculated for the multivariate model of design *A1a14* ranges from -28.16 for $T = 21$, down to -7.19 for $T = 120$. From $T = 20$ to $T = 28$ the results for the relative bias are very similar for the two approaches. It is not until $T = 40$ and $T = 120$ that a difference becomes more noticeable. That is, there is a slight increase in bias (in absolute terms) for the multivariate model compared to the univariate model. However, for this design, the Rel-RMSE is smaller for each value of T for the multivariate model. Thus, for values of T less than 40, the multivariate model performs slightly better than the univariate model.

For design *A1b33*, the results in Table 6.9 show a relative bias which ranges from -27.16 for $T = 23$ down (in absolute terms) to -5.57 for $T = 120$. The univariate values differ slightly from those in Table 6.8 due to the difference in the 1000 samples drawn. This is of no consequence. For this design, there is a large improvement in the Rel-Bias for the multivariate model especially for $T < 40$. For example, for $T = 20$ the Rel-Bias for the multivariate (-14.34) is just over half (53%) of the bias calculated with the univariate method (-27.04). Also, for $T < 40$, the Rel-RMSE is smaller for each value of T for the multivariate model. For $T = 40$, and 120, the Rel-RMSE is larger for the multivariate model, although the Rel-Bias is slightly smaller. For design *A1b33*, the naïve bias is greatly improved by applying the multivariate model to series with short to moderate length.

Table 6.8: Percent mean relative naïve bias and relative root mean squared error of smoothed seasonal component MSE for design *A1a14* for 1000 realisations with $T = 20, 21, \ldots, 27, 28, 40, 120$.

Design *A1a14*		Multivariate		Univariate	
Naïve Bias of Seasonal component		Rel-Bias	Rel-RMSE	Rel-Bias	Rel-RMSE
T	20	-26.97	41.14	-27.69	44.65
	21	-28.16	41.06	-27.31	43.69
	22	-26.74	40.17	-26.26	43.09
	23	-27.15	40.57	-26.87	43.22
	24	-27.57	40.09	-27.97	42.82
	25	-26.53	39.53	-27.77	42.56
	26	-26.81	39.39	-26.58	41.99
	27	-26.33	38.86	-25.03	41.00
	28	-25.78	38.58	-25.53	40.74
	40	-21.84	34.15	-19.38	35.20
	120	-7.19	18.32	-6.71	21.10

Table 6.9: Percent mean relative naïve bias and relative root mean squared error of smoothed seasonal component MSE for design *A1b33* for 1000 realisations with $T = 20, 21, \ldots, 27, 28, 40, 120$.

Design *A1b33*		Multivariate		Univariate	
Naïve Bias of Seasonal component		Rel-Bias	Rel-RMSE	Rel-Bias	Rel-RMSE
T	20	-14.34	42.09	-27.04	43.79
	21	-16.24	41.73	-28.41	44.16
	22	-15.41	41.22	-27.85	43.51
	23	-15.22	40.52	-27.16	42.76
	24	-15.92	40.02	-26.88	42.55
	25	-16.01	40.58	-26.94	42.08
	26	-15.01	40.16	-26.13	41.88
	27	-15.30	39.87	-26.07	41.81
	28	-14.92	39.80	-26.30	41.53
	40	-13.02	37.88	-18.60	35.15
	120	- 4.53	25.75	- 5.57	21.25

6.5.2 Bias Correction

As well as investigating the naïve bias, both Quenneville and Singh (2005) and Pfeffermann and Tiller (2005) examine various bias correction procedures. The study by Quenneville and Singh (2005) compares several different procedures including the corrected Ansley and Kohn approximation and a procedure which uses an approximation of the posterior distribution with initial priors ('Pc'). They show that the 'Pc' bias correction reduces the naïve bias from -21.2% to -6.7% for $T = 40$ and -9.0% to -2.5% for $T = 100$. However, this procedure 'can be more computationally demanding for more general structural models' (Quenneville and Singh, 2005, p229). They conclude that the corrected Ansley and Kohn approximation which reduces the relative naïve bias by 55% (down to -13% for $T = 40$ and -4% for $T = 100$) is the preferred method when other considerations such as theoretical exactness, bias, precision and computational requirements are taken into account.

Pfeffermann and Tiller (2005) carry out a similar comparative study for different bias correction procedures. They propose a new bootstrapping procedure which addresses both the filter uncertainty and the parameter uncertainty which constitute the predicted mean squared error. Their proposed method reduces the bias to be of order $(1/T^2)$ and does not require the assumption that the parameter estimators have, approximately, a normal distribution.

For the local level model example given in Pfeffermann and Tiller (2005), the bootstrap bias correction reduces the magnitude of the naïve bias from -18.5% to 0.6% for $T = 40$ and from -7.6% to 1.6% for $T = 100$. The bootstrap procedure is also applied to results from a simulation study which has a seasonal model for a monthly series with length $T = 84$. The extensive study generates 500 bootstrap series for each of 500 series selected from 10,000 primary series. The study then focuses on the last time point, $T = 84$. The naïve bias of -99 (-6%) at $T = 84$ is reduced to a bias of 12 (0.8%) (Pfeffermann and Tiller, 2005, Table IV, p910).

For the purpose of this study, the bootstrapping procedure detailed by Pfeffermann and Tiller (2005) will be carried out here as it has been shown to be effective, especially for shorter series. Previous estimators such as the estimator given by

Ansley and Kohn (1986), have a bias to the order of $(1/T)$, and rely on the asymptotic normality property of the maximum likelihood estimates of the parameters. In Section 6.4, the parameter estimates were shown to have quite skewed distributions, particularly for $T < 40$. For this reason, and since the proposed method by Pfeffermann and Tiller (2005) reduces the bias to be of order $(1/T^2)$, thereby substantially improving the bias, the bootstrap procedure will be implemented. A small simulation study will apply the bootstrapping procedure to the univariate and multivariate local level seasonal model for quarterly data with $T = 28$ for design *A1b33*.

The main equations and procedures from Pfeffermann and Tiller (2005) will be given in this section with notation adapted to comply with notation already defined in this thesis. For theoretical proofs and other details please refer to Pfeffermann and Tiller (2005). To save defining unnecessary additional notation, the procedure will be described in detail with reference to the univariate model only.

For the univariate model, the prediction error of the state vector when the parameter estimates are substituted for the known parameters may be written as

$$[\hat{\mathbf{a}}_{t|T} - \alpha_t] = [\hat{\mathbf{a}}_{t|T} - \mathbf{a}_{t|T}] + [\mathbf{a}_{t|T} - \alpha_t] \tag{6.47}$$

and the prediction mean squared error can therefore be written as the sum of two expectations, where the expectation is with respect to the joint distribution of α_t and $\mathbb{Y}_T$ (Pfeffermann and Tiller, 2005, p896):

$$\begin{aligned} MSE_t &= E[\hat{\mathbf{a}}_{t|T} - \alpha_t]^2 \\ &= E[\mathbf{a}_{t|T} - \alpha_t]^2 + E[\hat{\mathbf{a}}_{t|T} - \mathbf{a}_{t|T}]^2 \\ &= \text{filter uncertainty} + \text{parameter uncertainty}. \end{aligned} \tag{6.48}$$

Pfeffermann and Tiller (2005, p896) further explain that the covariance term is zero,

$$\begin{aligned} E\left([\hat{\mathbf{a}}_{t|T} - \mathbf{a}_{t|T}][\mathbf{a}_{t|T} - \alpha_t]\right) &= E\left(E\left([\hat{\mathbf{a}}_{t|T} - \mathbf{a}_{t|T}][\mathbf{a}_{t|T} - \alpha_t]|\mathbb{Y}_T\right)\right) \\ &= 0 \end{aligned} \tag{6.49}$$

since $[\hat{\mathbf{a}}_{t|T} - \mathbf{a}_{t|T}]$ is fixed when conditioning on $\mathbb{Y}_T$.

The first expression in (6.48) is the contribution to the prediction mean squared errors (PMSE) resulting from 'filter uncertainty'. The second expression is the contribution resulting from 'parameter uncertainty'. The naïve PMSE estimator accounts for 'filter uncertainty' but ignores 'parameter uncertainty' while the bootstrap procedure accounts for both.

Since the variance of the seasonally adjusted series is given by the error variance of the seasonal component (refer to Section 3.6.1), the formulae given in Pfeffermann and Tiller (2005, p897) will be adapted specifically to obtain results for the seasonal component.

The procedure consists of three steps (Pfeffermann and Tiller, 2005, p897):

1. Simulate a large number (B) of series of length $T = 28$ using the univariate model equations in (4.3) but with the estimated parameters, $\hat{\psi}^{(U)}$, instead of the true parameters, $\psi^{(U)}$. The simulated state vector series is denoted by α_t^b with the simulated seasonal component at time t denoted by S_t^b for $b = 1, \ldots, B$.

2. Re-estimate the model parameters for each of the B generated series using the same method as used for obtaining $\hat{\psi}^{(U)}$ (as described in 6.2). This process yields the bootstrap estimates $\hat{\psi}^{Ub}$ for $b = 1, \ldots, B$.

3. Estimate $MSE_t = E[\hat{\mathbf{a}}_{t|T} - \alpha_t]^2$ which is the error variance of the state vector given the estimated parameters. In particular, the interest is in $MSE_t^S = E[\hat{\mathcal{S}}_{t|T}^U - S_t]^2$. This is estimated with the following formulae:

$$\widehat{\mathrm{MSE}}_t^S = \mathrm{MSE}_t^{Sb} + 2\hat{\mathbf{V}}_{t|T}^S - \overline{V}_t^{Sb} \tag{6.50}$$

where

$$\mathrm{MSE}_t^{Sb} = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\mathcal{S}}_{t|T}^{bb} - \hat{\mathcal{S}}_{t|T}^{Ub}\right]^2; \qquad \overline{V}_t^{Sb} = \frac{1}{B}\sum_{b=1}^{B}\hat{\mathbf{V}}_{t|T}^{Sb}. \tag{6.51}$$

In equation (6.50),

- $\hat{\mathbf{V}}^{S}_{t|T}$ denotes the element of $\hat{\mathbf{V}}_{t|T}$ which corresponds to the seasonal component when estimated parameters $\hat{\psi}^{U}$ are substituted for exact parameters.

Thus, it is the naïve estimator of the mean squared error of the seasonal component given by the Kalman smoother, previously written as $MSE\left(\hat{\mathcal{S}}^{U}_{t|T}\right)$.
In equation (6.51),

- $\hat{\mathbf{V}}^{Sb}_{t|T}$ denotes the naïve estimator of the mean squared error of the seasonal component given by the Kalman smoother when estimated bootstrap parameters, $\hat{\psi}^{Ub}$, are used;

- $\hat{\mathcal{S}}^{bb}_{t|T}$ is the state moment estimator when the bootstrap series b, and the parameter estimates $\hat{\psi}^{Ub}$, are applied to the Kalman smoother;

- $\hat{\mathcal{S}}^{Ub}_{t|T}$ is the state moment estimator when the bootstrap series b, and the parameter estimates $\hat{\psi}^{U}$, are applied to the Kalman smoother.

The bootstrap procedure for the multivariate model is applied similarly to the above procedure, by substituting the corresponding multivariate parameter estimates $\hat{\psi}^{M}$ for $\hat{\psi}^{U}$. The data are generated as described in Section 6.3.1 with $\hat{\psi}^{M}$, and by applying the multivariate state space model as described in Section 4.2.1. The difference in this procedure of simulating the data and that for the previous simulation study is that the univariate data are simulated independently from the multivariate data. In the previous simulation study, the multivariate series were generated first and them summed to obtain the univariate series. However, because the estimated parameters are being used to generate the bootstrap series, the constraint that previously applied to the true parameters (4.5) does not apply here.

For the bootstrap simulation study, 100 series are randomly selected from the 1000 primary series generated to obtain the naïve bias results in Section 6.5.1. For each of the 100 selected series, 500 bootstrap series are generated, thus $B = 500$ and 100×500 series are produced. It is difficult to ascertain how many bootstrap

series are required. For the study involving the univariate local level model, Pfeffermann and Tiller (2005) generated 2000 bootstrap series for each of 1000 primary series, thus producing 1000×2000 series. For the univariate seasonal model, they produced 500×500 series. As the bootstrap method proved to be computationally intensive, this study will give results for 100×500 series for each of the univariate and multivariate models for sub-series design *A1b33*.

Results of Bootstrap Correction

The results of the bootstrapping study will be presented here focussing on the last time point of the series, $T = 28$. The results for the naïve and bootstrap estimators of the PMSE of the smoothed seasonal component are compared to the approximated true value.

In Table 6.9, the naïve bias for $T = 28$ showed a Rel-Bias of -14.92% for the multivariate model and -26.30% for the univariate model. These were calculated using (6.46) and (6.45) respectively. They are the result of the mean over $t = 1, \ldots, 28$. In order to compare the estimators, the value at $t = 28$ is recovered from the calculations using (6.45) and (6.46) to obtain

$$\text{Univariate:} \quad \text{Rel-Bias}_{28} = \frac{\overline{d}_{28}^{U}}{MSE_{28}^{U}}, \quad \text{Rel-RMSE}_{28} = \frac{\sqrt{\overline{d}_{28}^{U(2)}}}{MSE_{28}^{U}} \tag{6.52}$$

$$\text{Multivariate:} \quad \text{Rel-Bias}_{28} = \frac{\overline{d}_{28}^{M}}{MSE_{28}^{M}}, \quad \text{Rel-RMSE}_{28} = \frac{\sqrt{\overline{d}_{28}^{M(2)}}}{MSE_{28}^{M}}. \tag{6.53}$$

The expressions in (6.52) and (6.53) are calculated and the results are presented in Table 6.10. The Rel-Bias_{28} for the multivariate model (-25.87%) is smaller in magnitude than that for the univariate model (-33.11%). The value of the 'True' MSE for the univariate model has been approximated to be 1.19 by using MSE_{28}^{U} from (6.39). Similarly, the multivariate approximation of the true MSE at $t = 28$ is 0.65, determined by MSE_{28}^{M} (6.40). The value of the naïve estimator is given by $\hat{\mathbf{V}}_{28|28}^{S}$ (refer to Section 6.5.2) and hence the mean naïve bias for $t = 28$ is calculated

using $\overline{d}_t^U$ (6.43), for the univariate, and $\overline{d}_t^M$ (6.44), for the multivariate. The RMSE for the multivariate (42.59) is slightly smaller than that for the univariate (44.92).

The results of the 100×500 bootstrap simulation study are given in Table 6.11. For this study, the bootstrap correction method has markedly reduced the large negative bias associated with the naïve estimator. The Rel-Bias for the multivariate has reduced from -25.87% for the naïve estimator to 4.46% for the bootstrap estimator. For the univariate model, the bootstrap correction performs very well. The naïve Rel-Bias is reduced from -33.11% to just -1.62%. Thus, the univariate model seems to respond more quickly to the bootstrap method than for the multivariate model. This may be due to the larger number of parameters to be estimated in the multivariate model. It is expected that by increasing the number of primary series used for the multivariate bootstrap method, the bias may further reduce.

The RMSE for both the multivariate and univariate shown in Table 6.11 have increased slightly, and now show a similar value. This is most likely due to the addition of bias correction terms in the calculation of the bootstrap estimator (Pfeffermann and Tiller, 2005).

Table 6.10: Naïve bias for MSE of smoothed seasonal component with design *A1b33* for $t = T = 28$ calculated with 1000 realisations.

Design *A1b33*	'True' MSE	Naïve Est.	Naïve bias	Rel-Bias	Rel-RMSE
Multivariate	0.6487	0.4809	-0.1678	-25.87	42.59
Univariate	1.1919	0.7973	-0.3946	-33.11	44.92

Table 6.11: Bootstrap correction for MSE of smoothed seasonal component for design *A1b33* with $t = T = 28$ calculated with 100×500 series.

Design *A1b33*	'True' MSE	Boot Est.	Bootstrap bias	Rel-Bias	Rel-RMSE
Multivariate	0.6487	0.6776	0.0289	4.46	45.38
Univariate	1.1919	1.1726	-0.0193	-1.62	45.86

6.6 Revision Error

When government statistical agencies publish economic time series, preliminary series such as the seasonally adjusted series may be released. Then, as subsequent data become available, these series are revised. In practice, seasonally adjusted data are not usually revised if more than three years old (Burridge and Wallis, 1985). From a user's perspective, these revisions are cause for concern, as they tend to create a lack of trust in the data. The sizes of revisions are of particular concern when forecasts are being made from the data (Harvey, 1989, Section 6.4.4). It is preferable therefore to reduce the number and size of revisions.

The revision error may be defined as the difference between the concurrent (or real-time) estimate and the estimate calculated j time points later (Planas and Rossi, 2004). The current or preliminary seasonally adjusted value, $Y^a_{t|t}$, is given by

$$Y^a_{t|t} = Y_t - \hat{\mathcal{S}}_{t|t}. \tag{6.54}$$

An adjustment which is calculated with information available at time $t+j$ is denoted by $Y^a_{t|t+j}$ and can be written as,

$$Y^a_{t|t+j} = Y_t - \hat{\mathcal{S}}_{t|t+j}. \tag{6.55}$$

The revision error is therefore,

$$\begin{aligned} r_t^{t,t+j} &= Y^a_{t|t} - Y^a_{t|t+j} \\ &= Y_t - \hat{\mathcal{S}}_{t|t} - (Y_t - \hat{\mathcal{S}}_{t|t+j}) \\ &= \hat{\mathcal{S}}_{t|t+j} - \hat{\mathcal{S}}_{t|t}, \qquad j = 0, 1, ... \end{aligned} \tag{6.56}$$

The variance of the revision error is an important measure to consider, as it is used to calculate the width of the confidence interval around the concurrent estimates. Thus, the smaller the revision error variance, the narrower the confidence interval, which means the concurrent estimates are more reliable. Burridge and Wallis (1985) state that the variance of the revision error is a difference of variances, such that:

$$\text{Var}\left(r_t^{t,t+j}\right) = \text{Var}\left(\hat{\mathcal{S}}_{t|t}\right) - \text{Var}\left(\hat{\mathcal{S}}_{t|t+j}\right) \tag{6.57}$$

In the paper by Planas and Rossi (2004, p124), further explanation is given for this result. Firstly, the expression for the variance of (6.56) may be written as

$$\begin{aligned}
&\mathrm{E}\left([\hat{\mathcal{S}}_{t|t+j} - \hat{\mathcal{S}}_{t|t}][\hat{\mathcal{S}}_{t|t+j} - \hat{\mathcal{S}}_{t|t}]'\right) \\
&\qquad = \mathrm{E}\left([S_t - \hat{\mathcal{S}}_{t|t} - S_t + \hat{\mathcal{S}}_{t|t+j}][S_t - \hat{\mathcal{S}}_{t|t} - S_t + \hat{\mathcal{S}}_{t|t+j}]'\right) \\
&\qquad = \mathrm{Var}\left(\hat{\mathcal{S}}_{t|t}\right) + \mathrm{Var}\left(\hat{\mathcal{S}}_{t|t+j}\right) - 2\mathrm{E}\left([S_t - \hat{\mathcal{S}}_{t|t}][S_t - \hat{\mathcal{S}}_{t|t+j}]'\right) \qquad (6.58)
\end{aligned}$$

The covariance term in (6.58) may be expressed as

$$\begin{aligned}
&\mathrm{E}\left([S_t - \hat{\mathcal{S}}_{t|t}][S_t - \hat{\mathcal{S}}_{t|t+j}]'\right) \\
&\qquad = \mathrm{E}\left([S_t - \hat{\mathcal{S}}_{t|t+j} + \hat{\mathcal{S}}_{t|t+j} - \hat{\mathcal{S}}_{t|t}][S_t - \hat{\mathcal{S}}_{t|t+j}]'\right) \\
&\qquad = \mathrm{E}\left([S_t - \hat{\mathcal{S}}_{t|t+j}][S_t - \hat{\mathcal{S}}_{t|t+j}]'\right) + \mathrm{E}\left([\hat{\mathcal{S}}_{t|t+j} - \hat{\mathcal{S}}_{t|t}][S_t - \hat{\mathcal{S}}_{t|t+j}]'\right) \\
&\qquad = \mathrm{Var}\left(\hat{\mathcal{S}}_{t|t+j}\right) + 0 \qquad (6.59)
\end{aligned}$$

The second term is zero since the revision is independent of the error in the final estimate (see Pierce, 1980; Burridge and Wallis, 1985). Hence, by substituting the result for the covariance term from (6.59) into (6.58), the result for $\mathrm{Var}\left(r_t^{t,t+j}\right)$ given in (6.57) is obtained.

In Section 6.4, it was seen that the estimates of the model parameters depended upon the length of the series. As the error variance matrix of the state vector depends only on the values of the parameters and not on the observations, the terms on the right-hand side of (6.57) will depend on the parameters which are estimated with series length t or $t + j$. Therefore, as Planas and Rossi (2004) point out, (6.57) assumes that the parameters are held constant from time t to time $t + j$. As model parameters are re-estimated each time a new observation becomes available, (6.57) can be regarded as only an approximation to the revision error variance. Also another issue to consider with (6.57) is that the terms on the right hand side require the use of the mean squared error matrices due to estimated parameters. As shown by the naïve bias (Section 6.5.1), a serious underestimation of the elements of the mean squared error matrices exists when estimated parameters are applied.

To avoid these issues, Planas and Rossi (2004) propose an empirical method of calculating the variance of the revision error. Their method uses the revisions which

are actually observed. If $r_t^{t,t+J}$ denotes the total revisions after J additional periods, then the empirical estimate of variance of the revision, $\widehat{\text{Var}}\left(r_t^{t,t+J}\right)$, can be found by taking the mean over time t, of the sum of squared revisions calculated between time t and $t+J$. This is given by Planas and Rossi (2004, p125):

$$\widehat{\text{Var}}\left(r_t^{t,t+J}\right) = \frac{1}{T-J-\tau+1}\sum_{t=\tau}^{T-J}\sum_{j=1}^{J}\left(\hat{\mathcal{S}}_{t|t+j} - \hat{\mathcal{S}}_{t|t+j-1}\right)^2 \tag{6.60}$$

where τ is the time point at which the analysis of the concurrent estimate begins.

6.6.1 Revision Error Results

The revision error is calculated using (6.60) for both the univariate and multivariate models for the aggregate series with the two sub-series designs previously studied, namely *A1a14* and *A1b33*. For this study, the required values in (6.60) are set as $\tau = 20$, and $J = 4$. The calculation therefore involves the estimated seasonal components for the time period of $t = 20$ up to and including $t = 24$ with length of series varying between zero and 4 periods ahead of t. Thus, the focus is on the revision errors observed between 5 and 6 years of quarterly data, taking into account new observations one year ahead.

Estimates for the smoothed seasonal components using the estimated parameters are calculated for 1000 realisations of the univariate and multivariate models for $T = 20, 21, \ldots, 27, 28$. With $\tau = 20$, and $J = 4$, equation (6.60) simplifies to

$$\widehat{\text{Var}}\left(r_t^{t,t+4}\right) = \frac{1}{5}\sum_{t=20}^{24}\sum_{j=1}^{4}\left(\hat{\mathcal{S}}_{t|t+j} - \hat{\mathcal{S}}_{t|t+j-1}\right)^2, \tag{6.61}$$

and is calculated for each realisation and for each of the designs *A1a14* and *A1b33*. For the univariate model, this yields $\widehat{\text{Var}}_i^{U}\left(r_t^{t,t+4}\right)$ and for the multivariate model it yields $\widehat{\text{Var}}_i^{M}\left(r_t^{t,t+4}\right)$ for $i = 1, \ldots, 1000$.

To compare the revision error variance for the univariate model to that for the multivariate model, the relative efficiency will be calculated for each realisation, $i = 1, \ldots, 1000$ and for each sub-series design.

$$\text{RE}_i(r) = \frac{\widehat{\text{Var}}_i^{U}\left(r_t^{t,t+4}\right)}{\widehat{\text{Var}}_i^{M}\left(r_t^{t,t+4}\right)} \tag{6.62}$$

The distribution of $\text{RE}_i(r)$ for each design is summarised in Figure 6.7. Both box plots show a positively skewed distribution with many high outliers. To show more detail, the upper limit of the vertical scale of the box plots has been reduced to $\text{RE}_i(r) = 10$. This is shown in Figure 6.8.

Table 6.12 shows the 5-number summary and the mean for each design. The median for the relative efficiency of the revision error variance for design *A1b33* (1.546) is greater than that for design *A1a14* (1.101). However, for both designs, the median relative efficiency is above 1.0. This means that the multivariate design yields a smaller revision error variance than the univariate design. Thus, by applying the multivariate model, the confidence interval width for the revision error will be narrower than for the univariate model and the revisions will be also be smaller. For these two examples, there are good gains in using the multivariate model to improve the revision error.

Table 6.12: Five-number summary and mean of relative efficiency of the revision error variance for design *A1a14* and *A1b33* for 1000 realisations.

$\text{RE}_i(r)$	Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
Design *A1a14*	0.025	0.681	1.101	1.688	1.843	37.557
A1b33	0.018	0.820	1.546	2.470	2.922	26.747

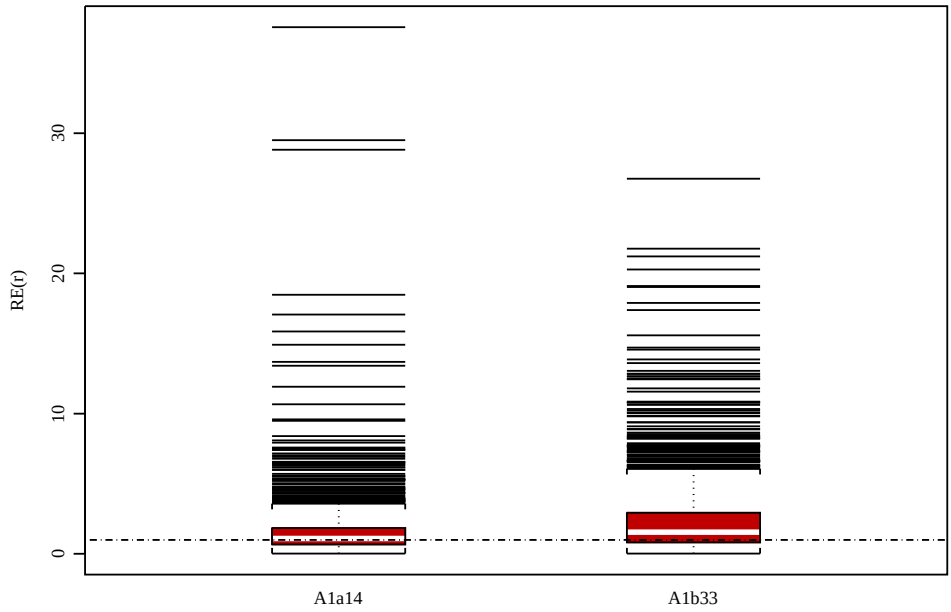


Figure 6.7: Box plot for relative efficiency of revision error variance for 1000 realisations of design *A1a14* and *A1b33*.

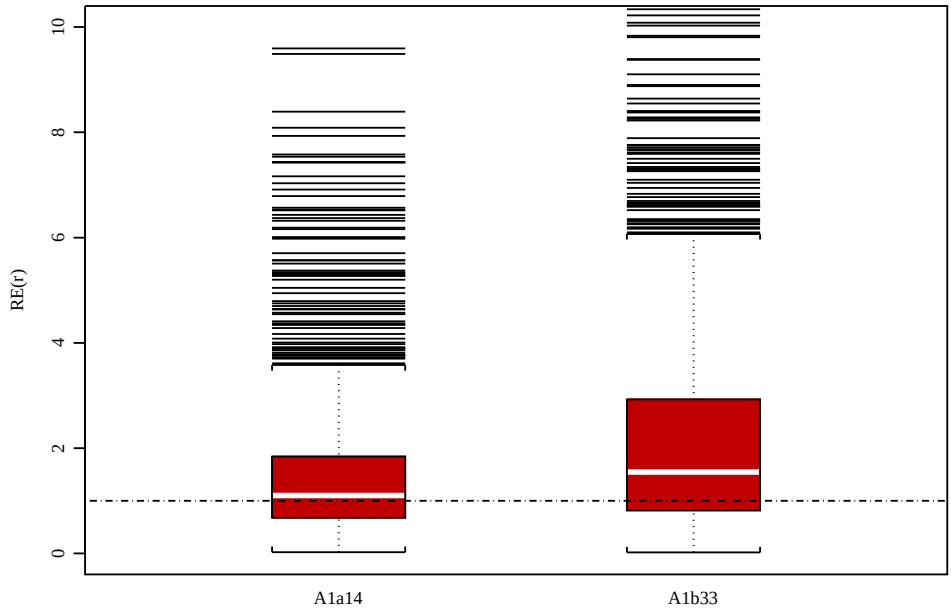


Figure 6.8: Box plot for relative efficiency of revision error variance for 1000 realisations of design *A1a14* and *A1b33* with limited scale.

6.7 Summary

In practice, a two-staged process is required in estimating state space model components. The first stage involved the estimation of the parameters using prediction error decomposition of the likelihood function. The second stage comprised the substituting of the estimated model parameters into the Kalman filter or smoother to obtain estimates of the components in the state vector and their mean squared errors. The effect of this substitution was studied in terms of the naïve bias for different series lengths. One method of bias correction, the bootstrap method, was investigated for both the univariate and multivariate models. Revision error was also considered using an empirical measure proposed by Planas and Rossi (2004).

In Section 6.4, the results of the estimation of the aggregate series parameters were shown for two different sub-series designs for series lengths $T = 20, 21, \ldots, 27, 28, 40, 120$. Applying the multivariate model decreased the bias associated with the estimates of the parameters of the aggregate series in both cases, especially for the parameters of the seasonal component (refer to Tables 6.3 and 6.6). The standard errors of the parameter estimates also reduced with the multivariate model. The results showed a greater improvement for design *A1b33*, than for design *A1a14*. In Section 4.6.1, the relative efficiency with known parameters for *Series 2* and *A1b33* was calculated to be 1.86, which is larger than that for *A1a14* (1.18). The results for the estimation of the parameters therefore suggest that the improvement in the estimated values of the aggregate series will be greater using the multivariate model under certain conditions which rely on the relative values of the sub-series model parameters.

In the second stage, the relative naïve bias and the relative root mean squared error were calculated for the two designs for both the univariate and multivariate models. The results showed that the naïve bias of the variance of the seasonally adjusted series could be decreased substantially with the use of the multivariate model. This was evident from the results for design *A1b33*, where the relative naïve bias was almost half of that obtained with the univariate model for short to medium

length time series. There was also a slight decrease in the relative RMSE.

One method of correcting the naïve bias, is the bootstrap correction proposed by Pfeffermann and Tiller (2005). This method was applied to design *A1b33* for both the univariate and multivariate models for the series with length $T = 28$. Application involved generating 500 bootstrap series for each of 100 primary series, re-estimating the model parameters and calculating the mean squared error of the seasonal component taking into account the filter uncertainty and the parameter uncertainty. The method reduced the relative naïve bias for the univariate model and for the multivariate model but was more effective for the univariate model.

In practice, the size of the revisions made to published series, as more observations are made available, is always a concern to government statistical agencies. The reliability of the concurrent estimates depends on the variance of the revision error. In Section 6.6, the variance of revision error was calculated empirically using the method proposed by Planas and Rossi (2004). The results for the univariate model were compared with the results for the multivariate model. It was shown that in both cases studied, a reduction in the variance of the revision error was evident when the multivariate model was used. The reduction was greater for design *A1b33* than for *A1a14*, again pointing to the importance of the relative values of the sub-series model parameters.

The focus of this chapter has been on the estimation of the model parameters and the effects of applying these in the calculation of the variance of the seasonally adjusted series and the revision error. In this chapter, as in previous chapters, the univariate model has been compared to the multivariate model when two sub-series are aggregated to obtain the total series. In the next chapter, an extension to this is given by way of an example in which the total series is an aggregate of eight sub-series.

Chapter 7

Handling Several Sub-series: An Example

7.1 Introduction

So far in this thesis, the case with two sub-series ($K = 2$) has been investigated in detail. In practice, the number of sub-series which are aggregated to obtain the total series is often more than two. In this chapter, the focus of attention is an example extending the work of Chapter 4 from two sub-series to eight sub-series and tieing this together with the quasi-likelihood theory developed in Chapter 5. A natural extension is to use the eight sub-series in the multivariate model and the aggregated series in the univariate model, thereby applying the same methodology as for $K = 2$ sub-series. However, as K increases, the multivariate model becomes more complex and the number of parameters grows quickly. For example, for the local level seasonal model with $K = 2$, there are $3(K + 1) = 9$ parameters in the model, and if $K = 4$, there are 15 parameters. When $K = 8$, the number of parameters increases to $3(K + 1) = 27$. If the series has a short to moderate length, the model could become unstable with so many parameters to fit. The main aim in this chapter is to determine a grouping strategy whereby similar gains can be achieved by aggregating the K original sub-series into $r < K$ new sub-series. The issue addressed is how to group the sub-series into a new set of sub-series, thereby reducing the complexity of the multivariate model and reducing the number of parameters to be estimated. To discuss this issue, the results from previous chapters are applied to an example.

In Chapter 4, the results for two sub-series showed that when the sub-series parameters are considerably different between series (i.e. within components) and within series (i.e. between components), gains in the relative efficiency were achieved. By using the multivariate model, the variance of the seasonally adjusted series could be improved. It would be too time intensive to consider every possible grouping combination if K is large. For example, if $K = 4$, there are 14 ways of grouping the series. If $K = 5$, the number of possible groupings increases to 51. (Refer to Table C.1 in Appendix C.3 for the calculation of these numbers using Stirling numbers of the second kind.) Therefore, using the results obtained in Chapter 4, several different combinations could be identified for investigation. This could be done empirically by calculating the c-ratios (4.14) for each pair of sub-series, given the multivariate parameters for the K series. The c-ratios for the seasonal and non-seasonal series components could then be compared in order to identify groupings with dissimilar c-ratios. In Section 5.7, the quasi-likelihood indicator for relative efficiency was developed for the situation where r new sub-series are formed from the K original sub-series. Given a number of different groupings, the value of Q (5.59) may be used to predict which grouping has the highest relative efficiency. Relative efficiencies can be determined via the Kalman filter for the chosen combinations. For the purposes of this study, both the Q indicator and the relative efficiency determined with the Kalman filter are calculated for nine different combinations of the sub-series.

7.2 Data Generation of the Eight Sub-series

In Chapter 4, data were simulated for two sub-series which summed to the same total series. Different combinations of parameters were determined for the sub-series in order to study the behaviour of the relative efficiency under different conditions. The parameters used were assumed to be exact parameters, thus requiring only one simulation of data for each case studied. In fact, the observations are not required at all, as the theoretical expressions for the MSE of the state vector component only rely on the parameters and not on the observations. The purpose of simulating data

values was to enable the use of `S-PLUS` software to more readily calculate the MSEs. Otherwise, the known parameters could be substituted into the derived theoretical expression for each value of t given by the Kalman filter.

In this chapter, a similar process is employed as in Chapter 4, but with $K = 8$ instead of $K = 2$. The parameters for the eight sub-series are assumed to be the exact parameters. In this example, the parameters for the sub-series remain fixed, and the interest lies in the grouping of the sub-series. Thus, parameters for the eight sub-series are required. Instead of simulating the data with artificially constructed sub-series (as carried out in Chapter 4), the data for the sub-series are simulated based on parameters determined by a real data series. In choosing an appropriate data set, two criteria had to be met. Firstly, it was required that the data have $K > 2$ sub-series, and enough to allow different groupings for comparison purposes. Secondly, to simplify the model, it was required that the data not be a result of a sample survey.

With this in mind, data from the Domestic Sales of Australian Wine and Brandy by Winemakers (Category number 8504.0, Table 2A) were obtained from the Australian Bureau of Statistics (ABS, 2007b). The series consists of 254 monthly sales of Australian wine in each of eight wine types from March 1985 to April 2006. The eight sub-series consist of monthly sales (in thousands of litres) of the following wine types: white table wine, red table wine, fortified wine, sparkling bottled wine, sparkling bulk wine, carbonated wine, flavoured wine and vermouth. Details of category definitions may be found in the associated technical report (ABS, 2007b).

The original data have been amended for the purposes of this example as follows:

- The first data point (March 1985) and the last data point (April 2006) were deleted, leaving a total of 252 data points, that is, 21 complete years.

- The monthly data were summed over each consecutive three month period to obtain 84 quarterly data points for each of the 8 sub-series.

- Logarithms of each series were then obtained in order to apply an additive basic structural model (since each series is considered multiplicative).

A plot of the amended data is given in Figure 7.1. It can be seen from the plot that some series have a fairly constant level with zero or very little slope and that all series have a seasonal component. The levels of some series vary over the 84 quarters. There are also breaks with strings of missing data for some series.

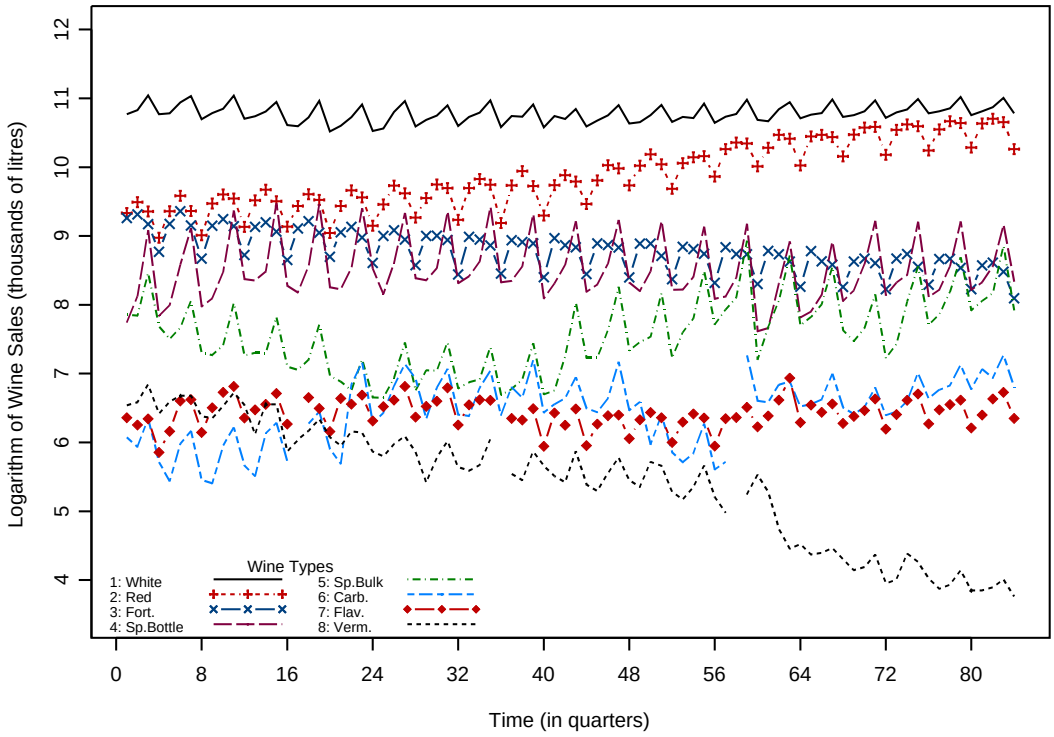


Figure 7.1: Logarithms of quarterly wine sales by type

The next step in constructing the simulated sub-series data was to obtain univariate parameter estimates for each sub-series individually. The univariate local level seasonal model was deemed to be appropriate for each sub-series for the purposes of this example. It is described in state space form in (4.6) to (4.8). The univariate model was applied to obtain maximum likelihood estimates of the parameters for each series. This produced three parameters for each series, giving a total of 24 estimated parameters. However, in order to simulate the sub-series, the multivariate parameters which include the common disturbance terms are also required. The univariate parameters can be used to obtain starting values for the

Table 7.1: Parameters used for generation of the eight sub-series

Parameter	Level ($x = \eta$)	Seasonal ($x = \omega$)	Error ($x = \varepsilon$)
σ_x^2	1.6e-013	0.00004	0.00043
$\sigma_{1x^*}^2$	0.00027	0.00001	0.00042
$\sigma_{2x^*}^2$	0.00178	0.00006	0.00016
$\sigma_{3x^*}^2$	0.00047	0.00007	0.00044
$\sigma_{4x^*}^2$	0.00397	0.00014	0.00602
$\sigma_{5x^*}^2$	0.01611	0.00096	0.00169
$\sigma_{6x^*}^2$	0.02291	0.00054	0.01172
$\sigma_{7x^*}^2$	0.00345	2.0e-017	0.00425
$\sigma_{8x^*}^2$	0.01409	6.6e-015	0.01226

estimation of the multivariate parameters.

The multivariate local level seasonal model with a quarterly dummy seasonal component for $K = 8$ has state space form similar to (4.9) but with $\mathbf{I}_2$ replaced by $\mathbf{I}_8$. Estimation of the parameters is performed by prediction error decomposition of the log-likelihood function, as described in Section 6.2.2. The 27 resulting multivariate parameter estimates are given in Table 7.1. These parameters are deemed to be the exact parameters for the purpose of this experiment. This construction allows a more realistic set of parameters to be used than if all sub-series were artificially constructed.

Given the 27 multivariate parameters shown in Table 7.1, the data for $Y_{1t}, \ldots, Y_{8t}$ are generated from the multivariate model equations according to the local level seasonal model for $K = 8$. The model is

$$\begin{aligned} Y_{kt} &= L_{kt} + S_{kt} + \varepsilon_t + \varepsilon_{kt}^*, \\ L_{k,\,t+1} &= L_{k,\,t} + \eta_t + \eta_{kt}^*, \\ S_{k,\,t+1} &= -\sum_{j=1}^{s-1} S_{k,\,t+1-j} + \omega_t + \omega_{kt}^*, \end{aligned} \tag{7.1}$$

where $k = 1, \ldots, 8$, and

$$\begin{aligned} \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), & \varepsilon_{kt}^* &\sim N(0, \sigma_{k\varepsilon^*}^2), \\ \eta_t &\sim N(0, \sigma_\eta^2), & \eta_{kt}^* &\sim N(0, \sigma_{k\eta^*}^2), \\ \omega_t &\sim N(0, \sigma_\omega^2), & \omega_{kt}^* &\sim N(0, \sigma_{k\omega^*}^2). \end{aligned} \tag{7.2}$$

Data is generated for 80 time points using starting values obtained from the logarithm of the original quarterly data for each series as given in Table 7.2. The first 40 data points of each series are then discarded, leaving $t = 1, \ldots, 40$ simulated quarterly observations for analysis.

Table 7.2: Initial values used for generation of the eight sub-series

Parameter	Level	Seasonal		
k	$L_{k,1}$	$S_{k,1}$	$S_{k,0}$	$S_{k,-1}$
1	11	0.06	0.21	-0.27
2	9	0.16	-0.13	-0.38
3	9	0.05	-0.14	-0.40
4	8	0.38	0.96	-1.25
5	8	-0.01	0.62	-0.78
6	6	-0.14	0.41	-0.63
7	6	-0.11	0.09	-0.49
8	7	0.05	0.26	-0.42

Data simulation and aggregation result in the set of eight sub-series (Y_{1t},. . . ,Y_{8t}), and their sum $Y_{tot,\,t}$, for $t = 1, \ldots, 40$. In practice, the aggregated series analysed would typically be the logarithm of the sum of the eight original (unlogged) series. As multiplicative series have not been considered in this thesis, the aggregated series has been constructed 'artificially' from the eight (logged) sub-series. The eight simulated sub-series are plotted in Figure 7.2.

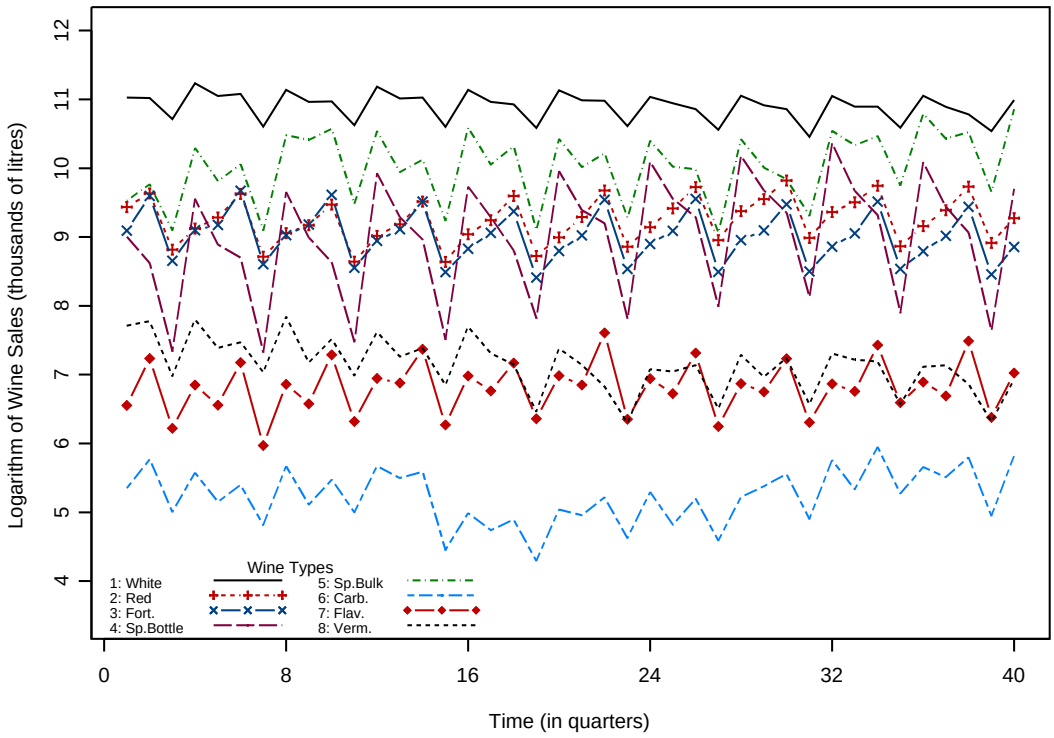


Figure 7.2: Simulated series of the logarithms of quarterly wine sales by type

The aggregated series is modelled by

$$\begin{aligned} Y_{tot,\,t} &= \sum_{k=1}^{K} Y_{kt} \\ &= L_{tot,\,t} + S_{tot,\,t} + \varepsilon_{tot,\,t}, \\ L_{tot,\,t+1} &= L_{tot,\,t-1} + \eta_{tot,\,t}, \\ S_{tot,\,t+1} &= -\sum_{j=1}^{s-1}, S_{tot,\,t+1-j} + \omega_{tot,\,t}, \end{aligned} \tag{7.3}$$

where

$$\varepsilon_{tot,t} \sim N(0, \sigma^2_{tot,\varepsilon}), \quad \eta_{tot,t} \sim N(0, \sigma^2_{tot,\eta}), \text{ and } \omega_{tot,t} \sim N(0, \sigma^2_{tot,\omega}). \tag{7.4}$$

Subsequently, the series parameters for the aggregated series are given by the fol-

lowing constraint formulae, evaluated with $K = 8$:

$$\sigma^2_{tot,\,\eta} = K^2\sigma^2_\eta + \sum_{k=1}^{K} \sigma^2_{k\eta^*} = 0.0630476,$$
$$\sigma^2_{tot,\,\omega} = K^2\sigma^2_\omega + \sum_{k=1}^{K} \sigma^2_{k\omega^*} = 0.0045557,$$
$$\sigma^2_{tot,\,\varepsilon} = K^2\sigma^2_\varepsilon + \sum_{k=1}^{K} \sigma^2_{k\varepsilon^*} = 0.0647954. \tag{7.5}$$

These parameters are substituted into the univariate model for the aggregated series.

7.3 Grouping the Sub-series

In Chapter 4, for the case with two sub-series, the relative efficiency of the multivariate model was calculated for different sub-series which summed to the same aggregate series. The between-series (i.e. within components) and the within-series (i.e. between components) relationships for the two series were studied and both were found to impact the relative efficiency.

In this example, the sub-series and the aggregate series are kept fixed, but the number of groups and size of the groups vary. To obtain some idea of how many combinations are possible, consider the case where only two groups ($r = 2$) are formed from the eight ($K = 8$) series. If the two groups consist of four series each, then there are 35 ($= {}^8C_4/2$) possible combinations. Two examples are $(Y_{1t} + Y_{2t} + Y_{3t}+Y_{4t},\ Y_{5t}+Y_{6t}+Y_{7t}+Y_{8t})$ and $(Y_{1t}+Y_{2t}+Y_{3t}+Y_{5t},\ Y_{4t}+Y_{6t}+Y_{7t}+Y_{8t})$. Alternatively, if one group consists of five series, then the other group will have 3 series. For this situation there are a 56 ($= {}^8C_5$) combinations. Another alternative is one group of 6 series and one group of 2 series. This results in another 28 ($= {}^8C_6$) combinations. Finally, there could be a group of 7 series and one single series, yielding another 8 ($= {}^8C_7$) combinations. Therefore, just for the case with $r = 2$, there are 127 possible combinations if $K = 8$. Then there are the cases where $r = 3$, 4, 5, 6, 7, and 8. Referring to Table C.1 in Appendix C.3, the number of possible combinations for these cases are 966, 1701, 1050, 266, 28, and 1 respectively. In total, there are 4139 possible groupings if $K = 8$.

To reduce the number of combinations to investigate, the c-ratios for each pair of series are calculated. The c-ratio definition is extended from the $K = 2$ case in Chapter 4 (see (4.13) -(4.14)) to the case with $K > 2$ groups. The c-ratio between the ith and jth sub-series is defined for each component. For the level, seasonal and error components, these are respectively,

$$c_{\eta\,(ij)} = \frac{\sigma_\eta^2 + \sigma_{i\eta^*}^2}{\sigma_\eta^2 + \sigma_{j\eta^*}^2}, \quad c_{\omega\,(ij)} = \frac{\sigma_\omega^2 + \sigma_{i\omega^*}^2}{\sigma_\omega^2 + \sigma_{j\omega^*}^2}, \quad c_{\varepsilon\,(ij)} = \frac{\sigma_\varepsilon^2 + \sigma_{i\varepsilon^*}^2}{\sigma_\varepsilon^2 + \sigma_{j\varepsilon^*}^2}. \tag{7.6}$$

For the example with eight sub-series, the c-ratios are shown in Table 7.3. By applying the conclusions determined in Chapter 4, decisions can be made as to which grouping combinations to test. For example, the 8 rows in the top left of Table 7.3 compare the parameters for Series 1 with those for Series 2 to 8 for each component. This shows that Series 1 is similar to Series 3, as the c-ratios are close to 1. However, the opposite is the case when comparing Series 1 parameters with Series 4 to 8. This suggests that Series 1 and 3 could be grouped together. Similarly, other observations can be made, such as possibly grouping Series 2 with Series 4, Series 5 with Series 6 and also grouping Series 7 with Series 8. This combination is just one of many possible combinations.

Guided by the results shown in Table 7.3, and for the purpose of demonstration, nine different combinations of the sub-series are investigated. Two different combinations will be considered for the cases where $r = 2$, 4, and 5. Three combinations will be considered for $r = 3$. Table 7.4 shows the grouping of the sub-series for each combination. A group may consist of the aggregate of two or more sub-series or it could simply be a single sub-series. For the first combination, there are 5 groups ($r = 5$). The first group contains the aggregate of Series 1 and 3 ($Y_{1t} + Y_{3t}$), the second group is the aggregate of Series 5 and 6 ($Y_{5t} + Y_{6t}$), the third group is the aggregate of Series 7 and 8 ($Y_{7t} + Y_{8t}$), and the remaining two groups are the individual series, Series 2 (Y_{2t}) and Series 4 (Y_{4t}) respectively.

Table 7.3: C-ratios for pairs of series

	$(\sigma_x^2+\sigma_{kx^*}^2)/(\sigma_x^2+\sigma_{1x^*}^2)$			$(\sigma_x^2+\sigma_{kx^*}^2)/(\sigma_x^2+\sigma_{2x^*}^2)$			$(\sigma_x^2+\sigma_{kx^*}^2)/(\sigma_x^2+\sigma_{3x^*}^2)$		
	Level	Seas.	Error	Level	Seas.	Error	Level	Seas.	Error
k	$x=\eta$	$x=\omega$	$x=\varepsilon$	$x=\eta$	$x=\omega$	$x=\varepsilon$	$x=\eta$	$x=\omega$	$x=\varepsilon$
1	1.0	1.0	1.0	0.2	0.5	1.4	0.6	0.5	1.0
2	6.6	1.9	0.7	1.0	1.0	1.0	3.8	0.9	0.7
3	1.7	2.2	1.0	0.3	1.2	1.5	1.0	1.0	1.0
4	14.7	3.4	7.5	2.2	1.8	10.8	8.5	1.5	7.4
5	59.6	18.8	2.5	9.1	9.9	3.6	34.4	8.5	2.4
6	84.8	11.0	14.2	12.9	5.8	20.4	48.9	5.0	13.9
7	12.8	0.8	5.5	1.9	0.4	7.9	7.4	0.4	5.4
8	52.1	0.8	14.8	7.9	0.4	21.3	30.1	0.4	14.5
	$(\sigma_x^2+\sigma_{kx^*}^2)/(\sigma_x^2+\sigma_{4x^*}^2)$			$(\sigma_x^2+\sigma_{kx^*}^2)/(\sigma_x^2+\sigma_{5x^*}^2)$			$(\sigma_x^2+\sigma_{kx^*}^2)/(\sigma_x^2+\sigma_{6x^*}^2)$		
	Level	Seas.	Error	Level	Seas.	Error	Level	Seas.	Error
k	$x=\eta$	$x=\omega$	$x=\varepsilon$	$x=\eta$	$x=\omega$	$x=\varepsilon$	$x=\eta$	$x=\omega$	$x=\varepsilon$
1	0.1	0.3	0.1	0.0	0.1	0.4	0.0	0.1	0.1
2	0.4	0.6	0.1	0.1	0.1	0.3	0.1	0.2	0.0
3	0.1	0.7	0.1	0.0	0.1	0.4	0.0	0.2	0.1
4	1.0	1.0	1.0	0.2	0.2	3.0	0.2	0.3	0.5
5	4.1	5.6	0.3	1.0	1.0	1.0	0.7	1.7	0.2
6	5.8	3.3	1.9	1.4	0.6	5.7	1.0	1.0	1.0
7	0.9	0.2	0.7	0.2	0.0	2.2	0.2	0.1	0.4
8	3.5	0.2	2.0	0.9	0.0	6.0	0.6	0.1	1.0
	$(\sigma_x^2+\sigma_{kx^*}^2)/(\sigma_x^2+\sigma_{7x^*}^2)$			$(\sigma_x^2+\sigma_{kx^*}^2)/(\sigma_x^2+\sigma_{8x^*}^2)$					
	Level	Seas.	Error	Level	Seas.	Error			
k	$x=\eta$	$x=\omega$	$x=\varepsilon$	$x=\eta$	$x=\omega$	$x=\varepsilon$			
1	0.1	1.2	0.2	0.0	1.2	0.1			
2	0.5	2.4	0.1	0.1	2.4	0.0			
3	0.1	2.7	0.2	0.0	2.7	0.1			
4	1.2	4.2	1.4	0.3	4.2	0.5			
5	4.7	23.2	0.5	1.1	23.2	0.2			
6	6.6	13.6	2.6	1.6	13.6	1.0			
7	1.0	1.0	1.0	0.2	1.0	0.4			
8	4.1	1.0	2.7	1.0	1.0	1.0			

Table 7.4: Different combinations of the eight sub-series

		Groups				
	r	1	2	3	4	5
1)	5	$Y_{1t} + Y_{3t}$	$Y_{5t} + Y_{6t}$	$Y_{7t} + Y_{8t}$	Y_{2t}	Y_{4t}
2)	5	$Y_{1t} + Y_{2t} + Y_{3t}$	Y_{4t}	$Y_{5t} + Y_{6t}$	Y_{7t}	Y_{8t}
3)	4	Y_{1t}	$Y_{6t} + Y_{7t} + Y_{8t}$	$Y_{2t} + Y_{3t}$	$Y_{4t} + Y_{5t}$	
4)	4	$Y_{1t} + Y_{3t}$	$Y_{2t} + Y_{4t}$	$Y_{5t} + Y_{6t}$	$Y_{7t} + Y_{8t}$	
5)	3	$Y_{1t} + Y_{2t} + Y_{3t}$	$Y_{4t} + Y_{7t} + Y_{8t}$	$Y_{5t} + Y_{6t}$		
6)	3	$Y_{1t} + Y_{3t}$	$Y_{2t} + Y_{4t} + Y_{7t} + Y_{8t}$	$Y_{5t} + Y_{6t}$		
7)	3	$Y_{1t} + Y_{3t}$	$Y_{5t} + Y_{6t} + Y_{7t} + Y_{8t}$	$Y_{2t} + Y_{4t}$		
8)	2	$Y_{1t} + Y_{6t} + Y_{7t} + Y_{8t}$	$Y_{2t} + Y_{3t} + Y_{4t} + Y_{5t}$			
9)	2	$Y_{1t} + Y_{2t} + Y_{3t} + Y_{4t}$	$Y_{5t} + Y_{6t} + Y_{7t} + Y_{8t}$			

7.3.1 Calculation of Q for each Combination

In order to compare the different combinations of sub-series, the relative efficiencies, $RE_t(M)$ (3.59), may be calculated. If a particular combination has a high relative efficiency, this means that the multivariate model is favoured over the univariate model for that combination. Hence, by comparing the relative efficiency, $RE_t(M)$, at some value of t for each combination, a decision can be made regarding which groupings to consider for trial in the estimation of series components. However, as shown in Chapter 4, the relative efficiency changes for each t and the Kalman filter needs to be applied to each combination.

In Section 5.7, the quasi-likelihood indicator for relative efficiency was developed for the case where r groups of sub-series are formed from the K original sub-series. The theory can be applied to this example with $K = 8$ to obtain the value of Q (5.59) for each combination. This will assist in determining which combinations have high relative efficiencies, $RE_t(M)$.

The value of Q has been evaluated for each combination by applying the theory in Section 5.7 using `Maple` software. The relevant transformation matrix ($\mathbf{\Lambda}$) is

determined by multiplying the $r \times r$ transformation matrix, $\mathbf{A}$ (3.33), by the selection matrix for that combination. For example, the transformation matrix for the first combination listed in Table 7.4 will be a $r \times K = 5 \times 8$ matrix, given by

$$\mathbf{\Lambda} = \begin{pmatrix} 1\,1\,1\,1\,1 \\ 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \end{pmatrix} \cdot \begin{pmatrix} 1\,0\,1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,1 \\ 0\,1\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0\,0\,0\,0 \end{pmatrix} = \begin{pmatrix} 1\,1\,1\,1\,1\,1\,1\,1 \\ 1\,0\,1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,1 \\ 0\,1\,0\,0\,0\,0\,0\,0 \end{pmatrix}. \tag{7.7}$$

The value of r, the number of groups formed for each combination, is taken from Table 7.4. Following the theory given in Section 5.7, the value of Q is obtained for each combination by substituting in the required values. For the local level seasonal model, these are $p = 4$, $u = 2$, and the parameter values given in Table 7.1. The results of Q are given in Table 7.5 alongside the calculated value of $RE_{40}(M)$. (For details of the matrix calculations performed in `Maple` refer to the CD enclosed with this thesis. Appendix E has a list of the filenames and descriptions of each program.)

The value of Q for the original 8 sub-series is calculated to be 1.026. As the number of groups (r) decreases, in general, the value of Q also decreases as shown in Table 7.5. The one exception is for the third combination, where $r = 4$. The value of Q is 1.022 which is higher than that for the first and second combinations where $r = 5$. These results give an indication of what to expect for the relative efficiencies produced by the Kalman filter. As the number of newly formed sub-series decreases (r is small), the relative efficiency is also expected to decrease. When sub-series are aggregated together to form a new sub-series, the within group differences between the series parameters no longer apply. This is because the parameters for the group aggregate now apply and not the parameters of the original sub-series. Some of the values of Q are very similar, indicating that the corresponding combinations will yield a similar relative efficiency.

7.4 Relative Efficiency Results

The variance of the seasonally adjusted series is determined by the Kalman filter, as described in Section 3.6.1 for both the univariate and the multivariate models. The relative efficiency is subsequently calculated at the last time point ($T = 40$), for each combination listed in Table 7.4. The values of $RE_{40}(M)$ are given in Table 7.5, together with those of Q.

Table 7.5: Values of Q and relative efficiency for nine different combinations

Combination	r	Q	$RE_{40}(M)$
Original	8	1.026	1.115
1	5	1.021	1.091
2	5	1.020	1.088
3	4	1.022	1.084
4	4	1.020	1.083
5	3	1.019	1.079
6	3	1.018	1.067
7	3	1.017	1.043
8	2	1.010	1.047
9	2	1.007	1.020

Using the original eight sub-series, the relative efficiency is 1.115. Thus, in this example with $K = 8$, there is a gain in using the multivariate model over the univariate model. As indicated by the value of Q, the value of $RE_{40}(M)$ is higher for the multivariate model with $K = 8$ than for any of the nine combinations tested. The first combination with $r = 5$ has $RE_{40}(M) = 1.09$, which is slightly lower than for the case with $K = 8$. The number of parameters in the model has reduced from 27 with $K = 8$ to 18 with $r = 5$. A model which achieves a similar relative efficiency with a smaller number of parameters may be a better choice in terms of stability of the model.

Generally, the value of $RE_{40}(M)$ decreases as r decreases. One exception to

this is in the comparison of combination 7 and 8. For combination 7, $r = 3$ and $RE_{40}(M) = 1.043$. However, for combination 8, where $r = 2$, the $RE_{40}(M)$ is higher at 1.047. This result suggests that combination 8 with only two groups is a better combination than that given by combination 7 with three groups. The increase was not indicated by Q. The value of Q preserves the relativities of $RE_{40}(M)$ within each value of r but not when comparing results across r. This requires further investigation but is not within the scope of this thesis.

The results for combination 7 and 8 prompt further investigation. It is possible to determine the c-ratios for the newly formed groups for each combination using the formula given in (7.6). Here, the i and j represent group i and group j. For combination 7, there are three c-ratios for each component, since $r = 3$. These are

$$\begin{aligned} c_{\eta\,(12)} &= 0.128, \quad c_{\eta\,(13)} = 0.013, \quad c_{\eta\,(23)} = 0.102, \\ c_{\omega\,(12)} &= 0.701, \quad c_{\omega\,(13)} = 0.117, \quad c_{\omega\,(23)} = 0.168, \\ c_{\varepsilon\,(12)} &= 0.328, \quad c_{\varepsilon\,(13)} = 0.071, \quad c_{\varepsilon\,(23)} = 0.215. \end{aligned} \tag{7.8}$$

For combination 8, there is only one c-ratio for each component, as $r = 2$:

$$\begin{aligned} c_{\eta\,(12)} &= 1.823, \\ c_{\omega\,(12)} &= 0.649, \\ c_{\varepsilon\,(12)} &= 2.333. \end{aligned} \tag{7.9}$$

The c-ratios in (7.8) for combination 7 show that there are differences between components, but that all c-ratios are below 1.0. For combination 8, however, the seasonal c-ratio, $c_{\omega\,(12)}$, is below 1.0, but the non-seasonal c-ratios are above 1.0 (7.9). This property is similar to the property defining the 'b' designs which yielded high relative efficiencies, studied in Chapter 4. Therefore, it is possible for a combination with, say r_1 groups to yield a higher relative efficiency than another combination with more groups, say r_2 ($r_2 > r_1$).

A further example demonstrates the use of c-ratios. A tenth combination was formed with the purpose of producing an example with a low relative efficiency. The objective this time is to combine together series with c-ratios that are not close to

1.0. Using the results in Table 7.3, Series 2 is combined with Series 6, as the c-ratios comparing these two series are high. Similarly, Series 3 and 4 are also combined with Series 6, thus forming one grouping with $Y_{2t}+Y_{3t}+Y_{4t}+Y_{6t}$. This leaves Series 1, 5, 7, and 8 to form another group, $Y_{1t}+Y_{5t}+Y_{7t}+Y_{8t}$.

For this tenth combination, $Q = 1$, thus indicating that the multivariate model will not achieve a gain over the univariate model, with respect to the relative efficiency. In fact, the relative efficiency is calculated to be very close to 1 ($RE_{40}(M) = 1.00003$). Thus, using the c-ratios to find a 'bad' example supports the notion that they may be used in an effective strategy of grouping the sub-series.

7.5 Summary

To reduce the complexity of the multivariate model, the K sub-series may be combined into several groups. These groups may be an aggregate of two or more series or may be a single series from the original K sub-series. The aim of this chapter was to determine a grouping strategy which would aid in the choice of grouping the sub-series into $r < K$ groups. The choice of grouping becomes a balance between reducing the number of parameters in the model and achieving a worthwhile gain in the relative efficiency. An example for which the aggregated series is the sum of eight sub-series, has been investigated.

Using the c-ratios as a guide to choosing the groupings, nine combinations of the eight sub-series were considered. For $r = 2$, 4 and 5, two combinations were considered. For $r = 3$, three combinations were considered. The quasi-likelihood indicator was calculated for each combination. The results showed that in general, the relative efficiency decreased as r decreased. There was one exception in the nine combinations studied. A combination of two groups ($r = 2$), had a higher relative efficiency for one of the combinations with three groups ($r = 3$). It was found that the seasonal c-ratio for the case with two groups had a value below 1.0 but the non-seasonal c-ratios were both above 1.0.

There may need to be a trade-off between the number of groups (r) chosen and

the desired relative efficiency. For example, it may be that the relative efficiency is higher with say five groups than it is for four groups, but that the difference is almost negligible. The model with the smaller number of groups (and hence parameters) may be more suitable in that it allows a more stable model. Further investigation is required into how the grouping may benefit the estimation process.

Chapter 8

Conclusions

8.1 Summary

The focus of this thesis has been on the reliability of a seasonally adjusted aggregated series. In particular, the research has examined whether gains are achievable in the seasonal adjustment of an aggregated series through joint modelling of the sub-series. If so, the aim was to identify the conditions under which these gains occurred. In addition, as estimates of the structural components are affected by the length of the series, the research also addressed examination of the impact of series length on the accuracy of seasonally adjusted aggregated series.

The aims were achieved by applying a multivariate basic structural model to a transformed set of sub-series, as discussed in Chapter 3. By including the aggregated series as one of the multivariate series in the model, estimates of the components such as the level and seasonal factors were obtained within the multivariate framework. For comparison, the variance of the seasonally adjusted aggregated series could then be calculated for both univariate and multivariate models.

Once the framework was established, an empirical study with two sub-series and known parameters was carried out in Chapter 4. From this study, it can be concluded that very good gains are possible and that the conditions rely on the relative parameters of the seasonal and non-seasonal components of the sub-series. To summarise results:

1. When the two sub-series have the same parameters for each component then

there is no gain using the multivariate model. Also, there is very little difference in the methods if the ratios of the series parameters within components are similar.

2. Gains are achievable when the ratio of the series parameters for the seasonal component is very different to that of the non-seasonal component. The gains are especially good when one of the ratios is greater than one and the other is less than one. These gains can be achieved even with low correlations between the series for each component.

3. Increasing the correlation between the sub-series for the seasonal component improves the gain but the magnitude is dependent upon the ratios of the series parameters within components. The more dissimilar the stochastic structure is between the components, the higher the magnitude of the gain from the seasonal correlation.

4. Increasing the correlation between the sub-series for the non-seasonal components improves the gain but not to the same magnitude as for the seasonal component.

5. Rates of convergence to a steady state of the relative efficiency varied with the magnitude of the ratios. Series with a high seasonal component ratio had a slower convergence rate than series with a low ratio.

Three alternative aggregated series were studied in Chapter 4. Changing the aggregated series parameters did not greatly affect the results of the relative efficiency. There was a difference in the rates of convergence of the relative efficiency. The aggregated series with the smallest seasonal to non-seasonal ratio of parameters had the slowest rate of convergence of the relative efficiency.

Because of the iterative nature of the Kalman filter, the theoretical expression for the relative efficiency changes for each time point. The expressions depend on the parameter values and not on the observations. An alternative to the relative efficiency calculated with the Kalman filter was derived using a quasi-likelihood

method. A single number indicator was found to be directly related to the relative efficiency given by the Kalman filter. Although it overestimated the value for the example studied, the quasi-likelihood indicator was shown to preserve the relativities. This alternative measure could be helpful in practice, especially when comparing different combinations of series. An example was given in Chapter 7.

In Chapter 6, a simulation study showed the effects of estimating the aggregate series parameters with the univariate and multivariate methods for different series lengths. For the two examples studied, it was shown that the bias of the estimated parameters was much less for the multivariate model than for the univariate model. This was especially the case for short to medium length time series. The relative efficiencies of the seasonally adjusted aggregated series also showed good gains for the multivariate model.

In practice, a two-staged process is required in estimating state space model components. Firstly, the parameters are estimated. Secondly, they are substituted into the model to obtain the estimates of the series components and their mean squared errors. The naïve bias in the mean squared error is a result of this substitution. For one of the examples, there was a substantial decrease in the naïve bias with the use of the multivariate model. The bootstrap method was applied to correct the naïve bias for both univariate and multivariate methods. The correction seemed to be more effective with the univariate model. Since the multivariate model has more parameters, it may be that to obtain comparable results, a larger number of bootstrap series is required for each primary series chosen.

The reliability of concurrent estimates was investigated in Chapter 6 by calculating the revision errors and the revision error variance. Large revision errors can be major contributors in a poor interpretation or forecast in the state of the economy. In both cases studied, there was a reduction in the revision error variance when the multivariate model was used. Thus, the confidence interval width for the revision error is narrower than for the univariate model and the revisions are also smaller. Overall, this means the concurrent estimates are more reliable.

In Chapter 7, the experience gained from Chapters 4, 5, and 6 was implemented

in a case study. When there are several sub-series, it may be required to decrease the number of parameters of the multivariate model by grouping the sub-series into $r \leq K$ new sub-series. The c-ratios can be used to identify possible groupings of similar series. Using these results, combinations of the sub-series may be identified for testing. The Q indicator can then be determined for each combination to predict that combination with the highest relative efficiency. The 10 combinations tested showed that in general, as r decreases, the relative efficiency decreases. The differences between the series parameters are lessened with the aggregation of the series within each grouping. However, it was demonstrated that it is possible for a combination with lower r to achieve a higher relative efficiency. There needs to be a balance between the required relative efficiency and the reduction of parameters which may help to stabilise the model in the estimation process.

In conclusion, applying the multivariate model proposed in this thesis to the system of sub-series of an aggregated series, allows gains to be achieved in the variance of the seasonally adjusted aggregate series, and the revision error variance.

8.2 Future Directions

During the course of this research, other issues were raised. While not in the scope of this thesis, they are a natural extension to it.

In this thesis, concentration has been on additive series. As many official time series are of a multiplicative nature, the methods proposed here could be extended to allow for multiplicative series. It is suggested that the logarithm of the aggregated series could be used in the multivariate system with the logarithm of each of the sub-series. In this way, there is a non-linear relationship between the K logged sub-series and the logged aggregate series. Further investigation is required in this area.

For those series which originate from a sample survey, it is necessary to include the survey error in the model. This can be incorporated into the basic structural model, as discussed in Pfeffermann *et al.* (1998).

Other options for the structure of the covariance matrix of each component could be investigated. Instead of having a common disturbance term in the definition of the model for each component, which forces the off-diagonal elements to be equal, this restriction could be relaxed. Alternative ways to reduce the number of parameters could be investigated in light of the results in Chapter 7.

There are many challenges in estimating the parameters of a time series when the length of the series is short. This thesis has examined the effects on parameter estimation while shortening the series. Although the multivariate method proposed here has shown to improve the accuracy of the estimates, there is great scope for further work in this area.

When the parameters of the model are estimated, the sampling errors due to the parameter uncertainty are not allowed for in the mean squared errors of the state vector estimates. (See comment by A. Harvey for Durbin and Koopman, 2000.) As previously mentioned, the bootstrap correction method, which allows for the parameter uncertainty, could be further investigated, especially for the multivariate model. Although 500 bootstrap series were generated for each of 100 primary series in this thesis, extending the study by increasing both the number of bootstrap series and the number of primary series would add further value in generalising the findings of this thesis.

In Chapter 7, the idea of grouping several sub-series to decrease the number of parameters was investigated with reference to the effects on the relative efficiency. The parameters in this study were assumed to be known. This example could be extended to examine the stability of the model for each chosen grouping, by carrying out the parameter estimation in each case. This would demonstrate whether the grouping benefits the estimation.

This research has implications in the area of using a related series to improve the estimates of a particular target series. This is an emerging area of research which is becoming popular in official statistical agencies worldwide. Although this thesis has concentrated on the accuracy of the aggregated series, the method could be similarly applied to focus on one of the individual series. If using a related series, the other

individual series in the multivariate system may not be cross-sectional series, and hence a transformation of the series would not be necessary.

Multivariate seasonal adjustment is becoming a topic of applied interest. The example referred to in Maravall (2006) is a good example for which the multivariate model could be applied. The data consists of the Japanese foreign trade series. The balance of trade series is the difference between the exports and the imports, thereby giving two series and the difference of these series. Maravall (2006) discusses the indirect versus the direct seasonal adjustment of the balance of trade series. It is planned to use this data in a forthcoming paper. Also, various research contract projects in the Centre for Statistical and Survey Methodology (CSSM) involve data which would benefit from this multivariate time series approach. If permission can be obtained from the client, this data could also be used for the purpose of future research.

Notwithstanding these issues that could be further pursued, this thesis identifies the conditions under which the multivariate model is beneficial to the seasonal adjustment of an aggregated series. The gains reported here are appreciable. Hence, when seasonal adjustment of an aggregated series of short to medium length is required, it is recommended that a multivariate framework be seriously considered.

Appendix A

The Kronecker Product

In Section 3.4.2 and Section 3.4.3, the Kronecker product is used in the calculations for the multivariate model. In this Appendix, the Kronecker product and some of the associated properties are defined. These properties and further information can be found in Harville (1997, Chapter 16).

Definition: Let $\mathbf{A}$ be an $m \times n$ matrix such that $\mathbf{A} = \{a_{ij}\}$ and let $\mathbf{B}$ be an $p \times q$ matrix such that $\mathbf{B} = \{b_{ij}\}$. The Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$ and will have dimensions $mp \times nq$:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{pmatrix}.$$

Property 1: The transpose of the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is given by

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}',$$

and it follows that if $\mathbf{A}$ and $\mathbf{B}$ are both symmetric,

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A} \otimes \mathbf{B}.$$

Property 2: For any $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, and any $p \times q$ matrix $\mathbf{C}$,

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C})$$
$$\mathbf{C} \otimes (\mathbf{A} + \mathbf{B}) = (\mathbf{C} \otimes \mathbf{A}) + (\mathbf{C} \otimes \mathbf{B}).$$

Property 3: For any $m \times n$ matrix $\mathbf{A} = \{a_{ij}\}$, $p \times q$ matrix $\mathbf{B} = \{b_{ij}\}$, $n \times u$ matrix $\mathbf{C} = \{c_{ij}\}$ and $q \times v$ matrix $\mathbf{D} = \{d_{ij}\}$,

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}).$$

Property 4: The Kronecker product, $\mathbf{A} \otimes \mathbf{B}$, of any $m \times m$ non-singular matrix $\mathbf{A}$, and any $p \times p$ non-singular matrix $\mathbf{B}$, is invertible, such that

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}.$$

Property 5: The Kronecker product, $\mathbf{A} \otimes \mathbf{B}$, of any $m \times n$ matrix $\mathbf{A}$, and $p \times q$ matrix $\mathbf{B}$, may be decomposed in two ways

$$\begin{aligned}\mathbf{A} \otimes \mathbf{B} &= (\mathbf{A} \otimes \mathbf{I}_p)(\mathbf{I}_n \otimes \mathbf{B}) \\ &= (\mathbf{I}_m \otimes \mathbf{B})(\mathbf{A} \otimes \mathbf{I}_q).\end{aligned}$$

Appendix B

Trigonometric Seasonality

B.1 Introduction

There are many ways of modelling the seasonal component in the basic structural model. The seasonal component, S_t, described in Section 3.2.2 is referred to as a dummy seasonal component. A common alternative is to express the seasonal component in trigonometric form, where it is represented as a sum of $s/2$ trigonometric terms (see Harvey, 1989, Section 2.3.4, or Durbin and Koopman, 2001, Section 3.2). This appendix outlines the univariate and multivariate model with trigonometric seasonal factors and examines the results of the relative efficiency for a few designs from the study in Chapter 4.

B.2 Univariate BSM

The univariate basic structural model is described in detail in Section 3.2. By replacing the dummy seasonal component, S_t, with a trigonometric seasonal component, $\tilde{S}_t$, Y_t may be written as:

$$Y_t = L_t + \tilde{S}_t + \varepsilon_{U,\,t}, \qquad \varepsilon_{U,\,t} \sim N(0, \sigma^2_{U,\,\varepsilon}). \tag{B.1}$$

If $\tilde{S}_t$ represents a trigonometric seasonal component, then the $s/2$ trigonometric terms are usually denoted in the texts (such as in Durbin and Koopman, 2001, Section 3.2) by γ_{jt}. Here, they will be denoted by ${}_j\beta_t$, thereby putting the j subscript

on the left, to avoid confusion later with the k subscript relating to the individual series in the multivariate BSM.

$$\tilde{S}_t = \sum_{j=1}^{[s/2]} {}_j\beta_t \tag{B.2}$$

where the ${}_j\beta_t$ evolve over time according to

$$\begin{aligned} {}_j\beta_{t+1} &= {}_j\beta_t \cos(\lambda_j) + {}_j\beta_t^* \sin(\lambda_j) + {}_j\tilde{\omega}_{U,\,t} \\ {}_j\beta_{t+1}^* &= -{}_j\beta_t \sin(\lambda_j) + {}_j\beta_t^* \cos(\lambda_j) + {}_j\varpi_{U,\,t} \end{aligned} \tag{B.3}$$

and where ${}_j\tilde{\omega}_{U,\,t}$ and ${}_j\varpi_{U,\,t}$ are independent $N(0, \sigma^2_{U,\,\tilde{\omega}})$ variables. (The U subscript is a label for the univariate model as previously denoted). Also

$$\lambda_j = \tfrac{2\pi j}{s}$$

$$j = 1, \ldots, [s/2], \text{ and } [s/2] = \begin{cases} s/2 & \text{if } s \text{ is even} \\ (s-1)/2 & \text{if } s \text{ is odd} \end{cases}$$

For quarterly data, $s = 4$, and hence λ_j is given by $\lambda_j = \frac{\pi j}{2}$ for $j = 1, 2$, resulting in the seasonal frequencies of $\frac{\pi}{2}$ and π. The equations defining the trigonometric model above (B.3) simplify to give four elements:

$$\begin{aligned} {}_1\beta_{t+1} &= {}_1\beta_t^* + {}_1\tilde{\omega}_{U,\,t} \\ {}_1\beta_{t+1}^* &= -{}_1\beta_t + {}_1\varpi_{U,\,t} \\ {}_2\beta_{t+1} &= -{}_2\beta_t + {}_2\tilde{\omega}_{U,\,t} \\ {}_2\beta_{t+1}^* &= -{}_2\beta_t^* + {}_2\varpi_{U,\,t} \end{aligned} \tag{B.4}$$

and then the seasonal component, $\tilde{S}_t$, becomes the sum of two elements:

$$\tilde{S}_t = {}_1\beta_t + {}_2\beta_t. \tag{B.5}$$

Similarly to the dummy seasonal variable, (B.2) still allows the sum constraint of the seasonal effects to be zero over any s time periods. Here, it can be shown that by substituting the appropriate terms from (B.4) into (B.5), and by taking expectations, the following constraint holds.

$$E\left(\sum_{i=0}^{s-1} \tilde{S}_{t+1-i}\right) = 0 \tag{B.6}$$

This constraint means that the trigonometric model only requires $s-1$ seasonal elements to be defined, which is the same as in the dummy seasonal model. For quarterly data, there are three seasonal elements for the trigonometric form, namely ${}_1\beta_t,\ {}_1\beta_t^*,\ {}_2\beta_t$, whereas for the dummy seasonal form they are $S_t,\ S_{t-1},\ S_{t-2}$.

B.2.1 Univariate State Space Form

In state space form, the state vector and system matrices for a seasonal local level model (without slope) were given in Section 4.2.1 in equations (4.6 to 4.8) for the dummy seasonal model. These are replaced by the following:

$$
\begin{aligned}
\tilde{\alpha}_t &= [L_t,\ {}_1\beta_t,\ {}_1\beta_t^*,\ {}_2\beta_t]', && \tilde{\alpha}_1 \sim \mathbf{N}(\mathbf{a}_1, \mathbf{P}_1)\\
\tilde{\gamma}_t &= [\eta_{U,\,t},\ {}_1\tilde{\omega}_{U,\,t},\ {}_1\varpi_{U,\,t},\ {}_2\tilde{\omega}_{U,\,t}]', && \tilde{\gamma}_t \sim \mathbf{N}(0, \tilde{\mathbf{Q}})\\
\tilde{\mathbf{Z}} &= \begin{pmatrix} 1 & 1 & 0 & 1 \end{pmatrix},\\
\tilde{\mathbf{G}} &= \mathbf{I}_4, && \varepsilon_{U,\,t} \sim \mathbf{N}(0, H), \qquad H = \sigma^2_{U,\,\varepsilon},\\
\tilde{\mathbf{T}} &= \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 0 & -1 \end{pmatrix}
\end{aligned}
$$

$$
\mathrm{Var}\left(\tilde{\mathbf{G}}\tilde{\gamma}_t\right) = \tilde{\mathbf{G}}\tilde{\mathbf{Q}}\tilde{\mathbf{G}}' = \begin{pmatrix} \sigma^2_{U,\,\eta} & 0 & 0 & 0\\ 0 & \sigma^2_{U,\,\tilde{\omega}} & 0 & 0\\ 0 & 0 & \sigma^2_{U,\,\tilde{\omega}} & 0\\ 0 & 0 & 0 & \sigma^2_{U,\,\tilde{\omega}} \end{pmatrix} \tag{B.7}
$$

where $\mathbf{I}_4$ is a 4×4 identity matrix.

The main difference is that each trigonometric seasonal element $({}_1\beta_t,\ {}_1\beta_t^*,\ {}_2\beta_t)$ has a disturbance term associated with it, namely ${}_1\tilde{\omega}_{U,\,t},\ {}_1\varpi_{U,\,t},\ {}_2\tilde{\omega}_{U,\,t}$ respectively. The dummy seasonal model has only one disturbance term, $(\omega_{U,\,t})$, which is associated with the S_t element of the state vector (α_t) in (4.8).

B.3 Multivariate BSM

If the dummy seasonal component, described in (4.1), is replaced by the trigonometric seasonal component, then an observation at time t for series k is given by Y_{kt}:

$$Y_{kt} = L_{kt} + \tilde{S}_{kt} + \varepsilon_t + \varepsilon^*_{kt}. \tag{B.8}$$

The equation for the seasonal component for series k is:

$$\tilde{S}_{kt} = \sum_{j=1}^{[s/2]} {}_j\beta_{k,\,t} \tag{B.9}$$

where the trigonometric terms ${}_j\beta_{k,\,t}$ evolve over time by

$$\begin{aligned} {}_j\beta_{k,\,t+1} &= {}_j\beta_{kt}\cos(\lambda_j) + {}_j\beta^*_{kt}\sin(\lambda_j) + {}_j\tilde{\omega}_t + {}_j\tilde{\omega}^*_{kt} \\ {}_j\beta^*_{k,\,t+1} &= -{}_j\beta_{kt}\sin(\lambda_j) + {}_j\beta^*_{kt}\cos(\lambda_j) + {}_j\varpi_t + {}_j\varpi^*_{kt}. \end{aligned}$$

The ${}_j\tilde{\omega}_t$ and ${}_j\varpi_t$ are the common effects which are assumed to be independent Normal random variables, each with zero mean and variance $\sigma^2_{\tilde{\omega}}$. The time-unit specific effects are ${}_j\tilde{\omega}^*_{kt}$ and ${}_j\varpi^*_{kt}$ similar to those described for the dummy seasonal model in Section 3.3.1, and are assumed to be independent Normal random variables each with zero mean and variance of $\sigma^2_{\tilde{\omega}^*}$.

B.3.1 Multivariate State Space Form

The state vector and system matrices for the multivariate local level seasonal model (without slope) are given in Section 4.2.1 in equations (4.9 to 4.11) for the dummy seasonal model. When $K = 2$, the transformed state vector, $\alpha_{(M),\,t}$ and the vector of disturbance terms, $\gamma_{(M),\,t}$ are replaced by the following:

$$\begin{aligned}\tilde{\alpha}_{(M),\,t} &= \left[L_{tot,\,t},\ L_{1t},\ {}_1\beta_{tot,\,t},\ {}_1\beta_{1t},\ {}_1\beta^*_{tot,\,t},\ {}_1\beta^*_{1t},\ {}_2\beta_{tot,\,t},\ {}_2\beta_{1t},\ \varepsilon_{tot,\,t},\ (\varepsilon_t + \varepsilon^*_{1t})\right]' \\ \tilde{\gamma}_{(M),\,t} &= [\ \eta_{tot,\,t},\ (\eta_t + \eta^*_{1t}),\ {}_1\tilde{\omega}_{tot,\,t},\ ({}_1\tilde{\omega}_t + {}_1\tilde{\omega}^*_{1t}),\ {}_1\varpi_{tot,\,t},\ ({}_1\varpi_t + {}_1\varpi^*_{1t}), \\ &\qquad {}_2\tilde{\omega}_{tot,\,t},\ ({}_2\tilde{\omega}_t + {}_2\tilde{\omega}^*_{1t}),\ \varepsilon_{tot,\,t+1},\ (\varepsilon_{t+1} + \varepsilon^*_{1,\,t+1})\ ]'\end{aligned} \tag{B.10}$$

The system matrices $\mathbf{Z}$, $\mathbf{G}$ and $\mathbf{T}$ described in (4.11) are replaced by:

$$\tilde{\mathbf{Z}} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \end{pmatrix} \qquad \tilde{\mathbf{G}} = \mathbf{I}_5 \qquad \tilde{\mathbf{T}} = \begin{pmatrix} 1 & 0\,0 & 0\,0 \\ 0 & 0\,1 & 0\,0 \\ 0 & -1\,0 & 0\,0 \\ 0 & 0\,0 & -1\,0 \\ 0 & 0\,0 & 0\,0 \end{pmatrix}$$

$$\mathrm{Var}\left((\tilde{\mathbf{G}} \otimes \mathbf{I}_2)\tilde{\gamma}_{(M),\,t}\right) = \begin{pmatrix} \mathbf{\Sigma}_{(M),\,\eta} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} \\ \mathbf{0}_{(2\times2)} & \mathbf{\Sigma}_{(M),\,\tilde{\omega}} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} \\ \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{\Sigma}_{(M),\,\tilde{\omega}} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} \\ \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{\Sigma}_{(M),\,\tilde{\omega}} & \mathbf{0}_{(2\times2)} \\ \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{0}_{(2\times2)} & \mathbf{\Sigma}_{(M),\,\varepsilon} \end{pmatrix}. \tag{B.11}$$

The transformed 2×2 covariance matrix, $\mathbf{\Sigma}_{(M),\,\eta}$, given in (4.12) remains unchanged, and similarly for $\mathbf{\Sigma}_{(M),\,\varepsilon}$. The covariance matrix for each of the seasonal components, $\mathbf{\Sigma}_{(M),\,\tilde{\omega}}$, is given by the following:

$$\mathbf{\Sigma}_{(M),\,\tilde{\omega}} = \begin{pmatrix} \sigma^2_{tot,\,\tilde{\omega}} & 2\sigma^2_{\tilde{\omega}} + \sigma^2_{1\tilde{\omega}^*} \\ 2\sigma^2_{\tilde{\omega}} + \sigma^2_{1\tilde{\omega}^*} & \sigma^2_{\tilde{\omega}} + \sigma^2_{1\tilde{\omega}^*} \end{pmatrix}. \tag{B.12}$$

B.4 Calculation of Relative Efficiency

With a trigonometric seasonal component for quarterly data, the seasonal component, $\tilde{S}_t$, is the sum of two elements, as given in (B.5). To obtain the variance of the seasonally adjusted series, the variance of the seasonal component, $\tilde{S}_t$, is determined.

$$\text{Var}(\tilde{S}_t) = \text{Var}({}_1\beta_t) + \text{Var}({}_2\beta_t) + 2\text{Cov}({}_1\beta_t,\ {}_2\beta_t). \tag{B.13}$$

Both $\text{Var}({}_1\beta_t)$ and $\text{Var}({}_2\beta_t)$ are diagonal elements of the $\mathbf{P}_{t|t}$ (3.48) matrix. These values are included in the standard `S+FinMetrics` output from the Kalman filter for each element of the state vector and for each t. However, the covariance term is not given directly.

Without using `S+FinMetrics`, the Kalman filter may be applied directly using the filtering equations, and then for each t, the relevant elements of the $\mathbf{P}_{t|t}$ matrix may be extracted for use in the formula in (B.13). This method has been used to obtain the MSE of the seasonal component for the univariate and the multivariate methods. The calculations have been carried out using MAPLE software and details are included on the CD enclosed with this thesis. Hence, the relative efficiency is given by

$$RE_t(M) = \frac{\text{MSE}(\hat{\tilde{S}}^U_{t|t})}{\text{MSE}(\hat{\tilde{S}}^M_{t|t})}, \qquad t = 1, \ldots, T \tag{B.14}$$

Alternatively, it is possible to obtain the covariance term in (B.13) from `S+FinMetrics` by using the output for $\mathbf{P}_{t|t}$, $\mathbf{P}_{t-1|t-1}$, $\mathbf{a}_{t|t}$, and $\mathbf{a}_{t-1|t-1}$. Details can be found on the CD enclosed with this thesis.

B.5 Relative Efficiency Results

To demonstrate the similarity of the relative efficiency using dummy seasonal and trigonometric seasonal components, three designs have been chosen from the study carried out in Chapter 4. They are *A1a14*, *A1a21*, and *A1b42*. For comparison purposes, the parameters for the aggregated series are chosen to be

$$\sigma^2_{U,\,\eta} = 0.01, \quad \sigma^2_{U,\,\tilde{\omega}} = 0.25, \quad \sigma^2_{U,\,\varepsilon} = 1. \tag{B.15}$$

For each design, the exact parameters for the multivariate model were calculated using the constraints in Section 4.3.2 according to the c-ratios and correlation settings for the design (refer to Section 4.3.1). The results of $RE_t(M)$ (B.14) at $T = 40$ are shown in Table B.1.

Table B.1: Results for $RE_{40}(M)$ for trigonometric seasonal component.

Design	c_ω	c_η and c_ε	$RE_{40}(M)$
A1a14	1	20	1.1623
A1a21	5	1	1.0935
A1b42	20	0.2	1.9020

The results in Table B.1 may be compared with those obtained using the dummy seasonal factor, given in Table 4.4 for designs *A1a14*, *A1a21* and Table 4.5 for design *A1b42*. It can be seen that the results are equivalent if rounded to two decimal places. Therefore, for the purposes of this thesis, the choice of the dummy or trigonometric seasonal factor in the basic structural model was not significant.

Appendix C

Derivations for Chapter 5

C.1 Quasi-Likelihood Result for the Error Variance Matrix of the Seasonal Components

In Section 5.4.1, equation (5.56) gives an expression for the error variance matrix of the seasonal components for the $K = 2$ multivariate case. The derivation of that expression is presented here.

The formula for the error variance matrix associated with the quasi-likelihood multivariate state vector estimator is given in (5.53), and restated here:

$$\begin{aligned}
\mathbf{P}^{(q)}_{(M),\,t} &= \mathrm{Var}_{t-1}\left(\alpha^{(q)}_{(M),\,t} - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,1}\right) \\
&= \left[\mathbf{I}_6 - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,a}\right]\mathrm{Var}_{t-1}\left(\alpha^{(q)}_{(M),\,t}\right)\left[\mathbf{I}_6 - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,a}\right]' \\
&\quad + \mathcal{A}_{(M)}\mathbf{X}_{(M),\,b}\,\mathrm{Var}_{t-1}\left(\mathcal{E}_{(M),\,t}\right)\left(\mathcal{A}_{(M)}\mathbf{X}_{(M),\,b}\right)', \qquad \text{(C.1)}
\end{aligned}$$

where $\mathbf{I}_6$ is the 6×6 identity matrix. Similarly, $\mathbf{I}_r$ will denote the $r \times r$ identity matrix in what follows. From (5.55), the matrix $\mathcal{A}_{(M)}$ is partitioned as

$$\mathcal{A}_{(M)} = \begin{pmatrix} \mathcal{A}_{11}\ \mathcal{A}_{12}\ \mathcal{A}_{13}\ \mathcal{A}_{14} \\ \mathcal{A}_{21}\ \mathcal{A}_{22}\ \mathcal{A}_{23}\ \mathcal{A}_{24} \\ \mathcal{A}_{31}\ \mathcal{A}_{32}\ \mathcal{A}_{33}\ \mathcal{A}_{34} \end{pmatrix}. \qquad \text{(C.2)}$$

From (5.52), the partitioned vector $\mathbf{X}_{(M),\,1}$ was given as

$$\begin{aligned}\mathbf{X}_{(M),\,1} = &\left[\begin{pmatrix}\mathbf{ZU}'\\ \mathbf{0}_{(3\times 3)}\end{pmatrix}\otimes\mathbf{I}_2\right]\alpha^{(q)}_{(M),\,t} + \left[\begin{pmatrix}1\\ \mathbf{0}_{(3\times 1)}\end{pmatrix}\otimes\mathbf{I}_2\right]\mathcal{E}_{(M),\,t}\\ &+\left[\begin{pmatrix}\mathbf{0}_{(1\times 5)}\\ -\mathbf{UT}\end{pmatrix}\otimes\mathbf{I}_2\right]\alpha_{(M),\,t-1}\\ &= (\mathbf{X}_a\otimes\mathbf{I}_2)\,\alpha^{(q)}_{(M),\,t} + (\mathbf{X}_b\otimes\mathbf{I}_2)\,\mathcal{E}_{(M),\,t} + (\mathbf{X}_c\otimes\mathbf{I}_2)\,\alpha_{(M),\,t-1}\\ &= \mathbf{X}_{(M),\,a}\;\alpha^{(q)}_{(M),\,t} + \mathbf{X}_{(M),\,b}\;\mathcal{E}_{(M),\,t} + \mathbf{X}_{(M),\,c}\;\alpha_{(M),\,t-1}. \qquad \text{(C.3)}\end{aligned}$$

The expression for $\mathbf{X}_{(M),\,a}$ in (C.3) may be written in terms of 2×2 matrices as

$$\mathbf{X}_{(M),\,a} = \begin{pmatrix}\mathbf{ZU}'\\ \mathbf{0}_{(3\times 3)}\end{pmatrix}\otimes\mathbf{I}_2 = \begin{pmatrix}\mathbf{I}_2 & \mathbf{0}_{(2\times 2)} & \mathbf{I}_2\\ \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)}\\ \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)}\\ \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)}\end{pmatrix}. \qquad \text{(C.4)}$$

Similarly, the expression for $\mathbf{X}_{(M),\,b}$ may be written as

$$\mathbf{X}_{(M),\,b} = \begin{pmatrix}1\\ \mathbf{0}_{(3\times 1)}\end{pmatrix}\otimes\mathbf{I}_2 = \begin{pmatrix}\mathbf{I}_2\\ \mathbf{0}_{(2\times 2)}\\ \mathbf{0}_{(2\times 2)}\\ \mathbf{0}_{(2\times 2)}\end{pmatrix}. \qquad \text{(C.5)}$$

Therefore, the expression $\mathbf{I}_6 - \mathcal{A}_{(M)}\mathbf{X}_{(M)\,a}$ in (C.1) may be re-expressed as

$$\begin{aligned}&\mathbf{I}_6 - \mathcal{A}_{(M)}\mathbf{X}_{(M)\,a}\\ &= \begin{pmatrix}\mathbf{I}_2 & \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)}\\ \mathbf{0}_{(2\times 2)} & \mathbf{I}_2 & \mathbf{0}_{(2\times 2)}\\ \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} & \mathbf{I}_2\end{pmatrix} - \begin{pmatrix}\mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{A}_{13} & \mathcal{A}_{14}\\ \mathcal{A}_{21} & \mathcal{A}_{22} & \mathcal{A}_{23} & \mathcal{A}_{24}\\ \mathcal{A}_{31} & \mathcal{A}_{32} & \mathcal{A}_{33} & \mathcal{A}_{34}\end{pmatrix}\begin{pmatrix}\mathbf{I}_2 & \mathbf{0}_{(2\times 2)} & \mathbf{I}_2\\ \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)}\\ \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)}\\ \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)}\end{pmatrix}\\ &= \begin{pmatrix}\mathbf{I}_2 - \mathcal{A}_{11} & \mathbf{0}_{(2\times 2)} & -\mathcal{A}_{11}\\ -\mathcal{A}_{21} & \mathbf{I}_2 & -\mathcal{A}_{21}\\ -\mathcal{A}_{31} & \mathbf{0}_{(2\times 2)} & -\mathcal{A}_{31}\end{pmatrix}. \qquad \text{(C.6)}\end{aligned}$$

Similarly,

$$\mathcal{A}_{(M)}\mathbf{X}_{(M)\,b} = \begin{pmatrix} \mathcal{A}_{11} & \mathcal{A}_{12} & \mathcal{A}_{13} & \mathcal{A}_{14} \\ \mathcal{A}_{21} & \mathcal{A}_{22} & \mathcal{A}_{23} & \mathcal{A}_{24} \\ \mathcal{A}_{31} & \mathcal{A}_{32} & \mathcal{A}_{33} & \mathcal{A}_{34} \end{pmatrix} \begin{pmatrix} \mathbf{I}_2 \\ \mathbf{0}_{(2\times 2)} \\ \mathbf{0}_{(2\times 2)} \\ \mathbf{0}_{(2\times 2)} \end{pmatrix} = \begin{pmatrix} \mathcal{A}_{11} \\ \mathcal{A}_{21} \\ \mathcal{A}_{31} \end{pmatrix}. \tag{C.7}$$

The first term of (C.1) becomes:

$$\begin{aligned} &\left[\mathbf{I}_6 - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,a}\right] \mathrm{Var}_{t-1}\left(\alpha^{(q)}_{(M),\,t}\right) \left[\mathbf{I}_6 - \mathcal{A}_{(M)}\mathbf{X}_{(M),\,a}\right]' \\ &\quad = \begin{pmatrix} \mathbf{I}_2 - \mathcal{A}_{11} & \mathbf{0}_{(2\times 2)} & -\mathcal{A}_{11} \\ -\mathcal{A}_{21} & \mathbf{I}_2 & -\mathcal{A}_{21} \\ -\mathcal{A}_{31} & \mathbf{0}_{(2\times 2)} & -\mathcal{A}_{31} \end{pmatrix} \begin{pmatrix} \mathbf{\Sigma}_{(M),\,\eta} & \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} \\ \mathbf{0}_{(2\times 2)} & \mathbf{\Sigma}_{(M),\,\zeta} & \mathbf{0}_{(2\times 2)} \\ \mathbf{0}_{(2\times 2)} & \mathbf{0}_{(2\times 2)} & \mathbf{\Sigma}_{(M),\,\omega} \end{pmatrix} \\ &\quad \times \begin{pmatrix} \mathbf{I}_2 - \mathcal{A}_{11} & -\mathcal{A}_{21} & -\mathcal{A}_{31} \\ \mathbf{0}_{(2\times 2)} & \mathbf{I}_2 & \mathbf{0}_{(2\times 2)} \\ -\mathcal{A}_{11} & -\mathcal{A}_{21} & -\mathcal{A}_{31} \end{pmatrix}. \end{aligned} \tag{C.8}$$

Denote the resulting matrix from (C.8) as V_a, and then the element in row r and column c as $V_a[r, c]$. The elements of the resulting 3×3 matrix are given below.

$$\begin{aligned} V_a[1,1] &= (\mathbf{I}_2 - \mathcal{A}_{11})\,\mathbf{\Sigma}_{(M),\,\eta}\,(\mathbf{I}_2 - \mathcal{A}_{11})' + \mathcal{A}_{11}\mathbf{\Sigma}_{(M),\,\omega}\mathcal{A}_{11}' \\ V_a[1,2] &= -(\mathbf{I}_2 - \mathcal{A}_{11})\,\mathbf{\Sigma}_{(M),\,\eta}\mathcal{A}_{21}' + \mathcal{A}_{11}\mathbf{\Sigma}_{(M),\,\omega}\mathcal{A}_{21}' \\ V_a[1,3] &= -(\mathbf{I}_2 - \mathcal{A}_{11})\,\mathbf{\Sigma}_{(M),\,\eta}\mathcal{A}_{31}' - \mathcal{A}_{11}\mathbf{\Sigma}_{(M),\,\omega}\,(\mathbf{I}_2 - \mathcal{A}_{31})' \\ V_a[2,1] &= -\mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\eta}\,(\mathbf{I}_2 - \mathcal{A}_{11})' + \mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\omega}\mathcal{A}_{11}' \\ V_a[2,2] &= \mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\eta}\mathcal{A}_{21}' + \mathbf{\Sigma}_{(M),\,\zeta} + \mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\omega}\mathcal{A}_{21}' \\ V_a[2,3] &= -\mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\eta}\mathcal{A}_{31}' - \mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\omega}\,(\mathbf{I}_2 - \mathcal{A}_{31})' \\ V_a[3,1] &= -\mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\eta}\,(\mathbf{I}_2 - \mathcal{A}_{11})' - (\mathbf{I}_2 - \mathcal{A}_{31})\,\mathbf{\Sigma}_{(M),\,\omega}\mathcal{A}_{11}' \\ V_a[3,2] &= \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\eta}\mathcal{A}_{21}' - (\mathbf{I}_2 - \mathcal{A}_{31})\,\mathbf{\Sigma}_{(M),\,\omega}\mathcal{A}_{21}' \\ V_a[3,3] &= \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\eta}\mathcal{A}_{31}' + (\mathbf{I}_2 - \mathcal{A}_{31})\,\mathbf{\Sigma}_{(M),\,\omega}\,(\mathbf{I}_2 - \mathcal{A}_{31})' \end{aligned} \tag{C.9}$$

The second term of (C.1) becomes

$$\begin{aligned}
&\mathcal{A}_{(M)}\mathbf{X}_{(M),\,b}\mathrm{Var}_{t-1}\left(\mathcal{E}_{(M),\,t}\right)\left(\mathcal{A}_{(M)}\mathbf{X}_{(M),\,b}\right)' \\
&\qquad = \begin{pmatrix} \mathcal{A}_{11} \\ \mathcal{A}_{21} \\ \mathcal{A}_{31} \end{pmatrix} \mathbf{\Sigma}_{(M),\,\varepsilon} \begin{pmatrix} \mathcal{A}_{11}' & \mathcal{A}_{21}' & \mathcal{A}_{31}' \end{pmatrix} \\
&\qquad = \begin{pmatrix} \mathcal{A}_{11}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{11}' & \mathcal{A}_{11}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{21}' & \mathcal{A}_{11}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{31}' \\ \mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{11}' & \mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{21}' & \mathcal{A}_{21}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{31}' \\ \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{11}' & \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{21}' & \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{31}' \end{pmatrix}. \qquad \text{(C.10)}
\end{aligned}$$

The solution to (C.1) can now be determined by adding together the two matrices, (C.9) and (C.10). The third diagonal element which is the 2×2 matrix representing $\mathrm{Var}_{t-1}\left(S_{tot,\,t}^{(q)} - \hat{S}_{tot,\,t}^{(q)},\ S_{1t}^{(q)} - \hat{S}_{1t}^{(q)}\right)$, is given by:

$$\begin{aligned}
&\mathrm{Var}_{t-1}\left(S_{tot,\,t}^{(q)} - \hat{S}_{tot,\,t}^{(q)},\ S_{1t}^{(q)} - \hat{S}_{1t}^{(q)}\right) \\
&\qquad = V_a[3,3] + \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{31}' \\
&\qquad = \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\eta}\mathcal{A}_{31}' + \left(\mathbf{I}_2 - \mathcal{A}_{31}\right)\mathbf{\Sigma}_{(M),\,\omega}\left(\mathbf{I}_2 - \mathcal{A}_{31}\right)' + \mathcal{A}_{31}\mathbf{\Sigma}_{(M),\,\varepsilon}\mathcal{A}_{31}' \\
&\qquad = \mathcal{A}_{31}\left[\mathbf{\Sigma}_{(M),\,\eta} + \mathbf{\Sigma}_{(M),\,\varepsilon}\right]\mathcal{A}_{31}' + \left(\mathbf{I}_2 - \mathcal{A}_{31}\right)\mathbf{\Sigma}_{(M),\,\omega}\left(\mathbf{I}_2 - \mathcal{A}_{31}\right)'. \qquad \text{(C.11)}
\end{aligned}$$

Hence, the element in the first row and first column of the 2×2 matrix (C.11) is the estimator of $\mathrm{Var}_{t-1}\left(S_{tot,\,t}^{(q)} - \hat{S}_{tot,\,t}^{(q)}\right)$.

C.2 Denominator of the Quasi-Likelihood Indicator of Relative Efficiency

In Chapter 5, equation (5.58) is the factorised version of the result for the error variance of the seasonal component for the aggregate series obtained with the multivariate method. This Appendix presents the derivation of the factorised result in terms of the parameter values.

The element which relates to the aggregate series is given by the element in the first row and first column of the 2×2 matrix given in Appendix C.1:

$$\begin{aligned}\mathrm{Var}_{t-1}&\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t},\ S^{(q)}_{1t} - \hat{S}^{(q)}_{1t}\right) \\ &= \mathcal{A}_{31}\left[\mathbf{\Sigma}_{(M),\,\eta} + \mathbf{\Sigma}_{(M),\,\varepsilon}\right]\mathcal{A}_{31}^{'} + \left(\mathbf{I}_2 - \mathcal{A}_{31}\right)\mathbf{\Sigma}_{(M),\,\omega}\left(\mathbf{I}_2 - \mathcal{A}_{31}\right)^{'} \qquad \text{(C.12)}\end{aligned}$$

To obtain the result in (5.58), the abbreviations from (5.57) are required and are stated again here for ease of reference.

$$\begin{aligned}\sigma^2_\eta + \sigma^2_{1\eta^*} &= \mathrm{n}_1, & \sigma^2_\omega + \sigma^2_{1\omega^*} &= \mathrm{w}_1, & \sigma^2_\varepsilon + \sigma^2_{1\varepsilon^*} &= \mathrm{e}_1 \\ \sigma^2_\eta + \sigma^2_{2\eta^*} &= \mathrm{n}_2, & \sigma^2_\omega + \sigma^2_{2\omega^*} &= \mathrm{w}_2, & \sigma^2_\varepsilon + \sigma^2_{2\varepsilon^*} &= \mathrm{e}_2 \\ \sigma^2_{tot,\,\eta} &= \mathrm{N}, & \sigma^2_{tot,\,\omega} &= \mathrm{W}, & \sigma^2_{tot,\,\varepsilon} &= \mathrm{E}\end{aligned} \qquad \text{(C.13)}$$

Using the `Maple` software, the elements of the 2×2 matrix, $\mathcal{A}_{31}$, are calculated. Element $[1,1]$ is

$$\begin{aligned}\mathcal{A}_{31}[1,1] &= \frac{1}{D}\left(\sigma^2_\varepsilon\sigma^2_\omega + \sigma^2_\varepsilon\mathrm{w}_1 + \sigma^4_\omega - \mathrm{w}_1\mathrm{w}_2 - \sigma^2_\omega\mathrm{n}_1 - \mathrm{e}_1\mathrm{w}_2\right. \\ &\qquad \left. -\mathrm{e}_1\sigma^2_\omega + \sigma^2_\omega\sigma^2_\eta - \mathrm{w}_2\mathrm{n}_1 + \mathrm{w}_1\sigma^2_\eta\right) \\ &= \frac{1}{D}\left[\sigma^4_\omega - \mathrm{w}_1\mathrm{w}_2 + (\sigma^2_\omega + \mathrm{w}_1)(\sigma^2_\eta + \sigma^2_\varepsilon) - (\sigma^2_\omega + \mathrm{w}_2)(\mathrm{n}_1 + \mathrm{e}_1)\right] \qquad \text{(C.14)}\end{aligned}$$

where D is given by:

$$\begin{aligned}D &= \sigma^4_\omega + 2\sigma^2_\varepsilon\sigma^2_\omega + 2\sigma^2_\omega\sigma^2_\eta - \mathrm{w}_1\mathrm{w}_2 - \mathrm{n}_1\mathrm{n}_2 + \sigma^4_\eta - \mathrm{e}_1\mathrm{e}_2 - \mathrm{w}_2\mathrm{n}_1 \\ &\quad - \mathrm{e}_1\mathrm{n}_2 - \mathrm{w}_1\mathrm{e}_2 + 2\sigma^2_\varepsilon\sigma^2_\eta - \mathrm{w}_1\mathrm{n}_2 - \mathrm{e}_1\mathrm{w}_2 + \sigma^4_\varepsilon - \mathrm{n}_1\mathrm{e}_2 \\ &= \sigma^4_\omega + \sigma^4_\eta + \sigma^4_\varepsilon + 2\sigma^2_\varepsilon\sigma^2_\omega + 2\sigma^2_\omega\sigma^2_\eta + 2\sigma^2_\varepsilon\sigma^2_\eta - \mathrm{w}_1\mathrm{w}_2 - \mathrm{n}_1\mathrm{n}_2 \\ &\quad -\mathrm{e}_1\mathrm{e}_2 - \mathrm{w}_2\mathrm{n}_1 - \mathrm{e}_1\mathrm{n}_2 - \mathrm{w}_1\mathrm{e}_2 - \mathrm{w}_1\mathrm{n}_2 - \mathrm{e}_1\mathrm{w}_2 - \mathrm{n}_1\mathrm{e}_2\end{aligned} \qquad \text{(C.15)}$$

and can be factorised as

$$D = (\sigma_\omega^2 + \sigma_\eta^2 + \sigma_\varepsilon^2)^2 - (\mathrm{w}_1 + \mathrm{n}_1 + \mathrm{e}_1)(\mathrm{w}_2 + \mathrm{n}_2 + \mathrm{e}_2). \tag{C.16}$$

The remaining elements of $\mathcal{A}_{31}$ are given by:

$$\begin{aligned}
\mathcal{A}_{31}[1,2] &= -\frac{1}{D}\big(\sigma_\varepsilon^2\mathrm{w}_1 - \sigma_\varepsilon^2\mathrm{w}_2 - \mathrm{e}_1\mathrm{w}_2 - \mathrm{w}_2\sigma_\eta^2 + \sigma_\omega^2\mathrm{n}_2 - \mathrm{w}_2\mathrm{n}_1 \\
&\qquad -\sigma_\omega^2\mathrm{n}_1 - \mathrm{e}_1\sigma_\omega^2 + \mathrm{w}_1\mathrm{n}_2 + \mathrm{e}_2\sigma_\omega^2 + \mathrm{w}_1\sigma_\eta^2 + \mathrm{w}_1\mathrm{e}_2\big) \\
&= \frac{1}{D}\big[(\sigma_\omega^2 + \mathrm{w}_2)(\mathrm{n}_1 + \mathrm{e}_1) - (\sigma_\omega^2 + \mathrm{w}_1)(\mathrm{n}_2 + \mathrm{e}_2) \\
&\qquad -(\sigma_\eta^2 + \sigma_\varepsilon^2)(\mathrm{w}_1 - \mathrm{w}_2)\big] \qquad \text{(C.17)} \\
\mathcal{A}_{31}[2,1] &= \frac{1}{D}\big(\sigma_\varepsilon^2\mathrm{w}_1 - \sigma_\omega^2\mathrm{n}_1 - \mathrm{e}_1\sigma_\omega^2 + \mathrm{w}_1\sigma_\eta^2\big) \\
&= \frac{1}{D}\big[\mathrm{w}_1(\sigma_\eta^2 + \sigma_\varepsilon^2) - \sigma_\omega^2(\mathrm{n}_1 + \mathrm{e}_1)\big] \qquad \text{(C.18)}
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_{31}[2,2] &= \frac{1}{D}\big(\sigma_\varepsilon^2\sigma_\omega^2 - \sigma_\varepsilon^2\mathrm{w}_1 - \mathrm{w}_1\mathrm{w}_2 - \mathrm{w}_1\mathrm{e}_2 \\
&\qquad + \sigma_\omega^4 - \mathrm{w}_1\sigma_\eta^2 + \sigma_\omega^2\sigma_\eta^2 + \sigma_\omega^2\mathrm{n}_1 + \mathrm{e}_1\sigma_\omega^2 - \mathrm{w}_1\mathrm{n}_2\big) \\
&= \frac{1}{D}\big[\sigma_\omega^4 - \mathrm{w}_1\mathrm{w}_2 + (\sigma_\eta^2 + \sigma_\varepsilon^2)(\sigma_\omega^2 - \mathrm{w}_1) + \sigma_\omega^2(\mathrm{n}_1 + \mathrm{e}_1) \\
&\qquad -\mathrm{w}_1(\mathrm{n}_2 + \mathrm{e}_2)\big]. \qquad \text{(C.19)}
\end{aligned}$$

From (5.45), the covariance matrices can be rewritten using the abbreviated notation in (C.13). Hence, taking the sum of the level and error covariance matrices as required in the first term of (C.12), the result is

$$\begin{aligned}
\big[\boldsymbol{\Sigma}_{(M),\,\eta} + \boldsymbol{\Sigma}_{(M),\,\varepsilon}\big] &= \begin{pmatrix} \mathrm{N} & \sigma_\eta^2 + \mathrm{n}_1 \\ \sigma_\eta^2 + \mathrm{n}_1 & \mathrm{n}_1 \end{pmatrix} + \begin{pmatrix} \mathrm{E} & \sigma_\varepsilon^2 + \mathrm{e}_1 \\ \sigma_\varepsilon^2 + \mathrm{e}_1 & \mathrm{e}_1 \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{N} + \mathrm{E} & \sigma_\eta^2 + \sigma_\varepsilon^2 + \mathrm{n}_1 + \mathrm{e}_1 \\ \sigma_\eta^2 + \sigma_\varepsilon^2 + \mathrm{n}_1 + \mathrm{e}_1 & \mathrm{n}_1 + \mathrm{e}_1 \end{pmatrix}. \qquad \text{(C.20)}
\end{aligned}$$

So, if $\mathcal{A}_{31}$ is the matrix

$$\mathcal{A}_{31} = \begin{pmatrix} \mathcal{A}_{31}[1,1] & \mathcal{A}_{31}[1,2] \\ \mathcal{A}_{31}[2,1] & \mathcal{A}_{31}[1,2] \end{pmatrix}, \tag{C.21}$$

then (C.12) may be written in matrix form as

$$\begin{aligned}\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t},\ S^{(q)}_{1t} - \hat{S}^{(q)}_{1t}\right) \\ = \begin{pmatrix} \mathrm{a} & \mathrm{b} \\ \mathrm{c} & \mathrm{d} \end{pmatrix} \begin{pmatrix} \mathrm{N}+\mathrm{E} & \sigma^2_\eta + \sigma^2_\varepsilon + \mathrm{n}_1 + \mathrm{e}_1 \\ \sigma^2_\eta + \sigma^2_\varepsilon + \mathrm{n}_1 + \mathrm{e}_1 & \mathrm{n}_1 + \mathrm{e}_1 \end{pmatrix} \begin{pmatrix} \mathrm{a} & \mathrm{c} \\ \mathrm{b} & \mathrm{d} \end{pmatrix} \\ + \begin{pmatrix} 1-\mathrm{a} & \mathrm{b} \\ \mathrm{c} & 1-\mathrm{d} \end{pmatrix} \begin{pmatrix} \mathrm{W} & \sigma^2_\omega + \mathrm{w}_1 \\ \sigma^2_\omega + \mathrm{w}_1 & \mathrm{w}_1 \end{pmatrix} \begin{pmatrix} 1-\mathrm{a} & \mathrm{c} \\ \mathrm{b} & 1-\mathrm{d} \end{pmatrix},\end{aligned} \tag{C.22}$$

where, for convenience, the elements of $\mathcal{A}_{31}$ are denoted

$$\begin{aligned}\mathcal{A}_{31}[1,1] = \mathrm{a}, \quad \mathcal{A}_{31}[1,2] = \mathrm{b}, \\ \mathcal{A}_{31}[2,1] = \mathrm{c}, \quad \mathcal{A}_{31}[2,2] = \mathrm{d}.\end{aligned} \tag{C.23}$$

Element [1,1] in (C.22) is given by:

$$\begin{aligned}\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}\right) = \mathrm{a}^2(\mathrm{N}+\mathrm{E}) + 2\mathrm{ab}(\sigma^2_\eta + \sigma^2_\varepsilon + \mathrm{n}_1 + \mathrm{e}_1) + \mathrm{b}^2(\mathrm{n}_1 + \mathrm{e}_1) \\ + (1-\mathrm{a}^2)\mathrm{W} + \mathrm{b}^2\mathrm{w}_1 + 2\mathrm{b}(1-\mathrm{a})(\sigma^2_\omega + \mathrm{w}_1).\end{aligned} \tag{C.24}$$

Using `Maple`, the expressions for a, b, c, and d, are substituted into (C.24) and the following result is obtained:

$$\begin{aligned}\mathrm{Var}_{t-1}\left(S^{(q)}_{tot,\,t} - \hat{S}^{(q)}_{tot,\,t}\right) \\ = \frac{\mathrm{W}\left(\left[\sigma^2_\eta + \sigma^2_\varepsilon\right]^2 - \left[\mathrm{n}_1 + \mathrm{e}_1\right]\left[\mathrm{n}_2 + \mathrm{e}_2\right]\right) + (\mathrm{N}+\mathrm{E})\left(\sigma^4_\omega - \mathrm{w}_1\mathrm{w}_2\right)}{\left(\sigma^2_\eta + \sigma^2_\varepsilon + \sigma^2_\omega\right)^2 - \left(\mathrm{n}_1 + \mathrm{e}_1 + \mathrm{w}_1\right)\left(\mathrm{n}_2 + \mathrm{e}_2 + \mathrm{w}_2\right)}\end{aligned} \tag{C.25}$$

This expression becomes the denominator of Q. It is the error variance of the seasonal component for the aggregate series obtained with the multivariate QL method.

C.3 Stirling Numbers of the Second Kind

The 18th century Scottish mathematician, James Stirling (1692 - 1770), determined the number of ways, $S(n,\ m)$, of partitioning a set of n elements into m non-empty sets. For an annotated translation of the original book (written in Latin) by Stirling, refer to Tweddle (2003).

The $S(n,\ m)$ which are known as 'Stirling numbers of the second kind', can be computed as the sum

$$S(n,\ m) = \frac{1}{m!}\sum_{i=0}^{m}(-1)^i \begin{pmatrix} m \\ i \end{pmatrix} (m-i)^n. \tag{C.26}$$

Alternatively, the following recursion can be used:

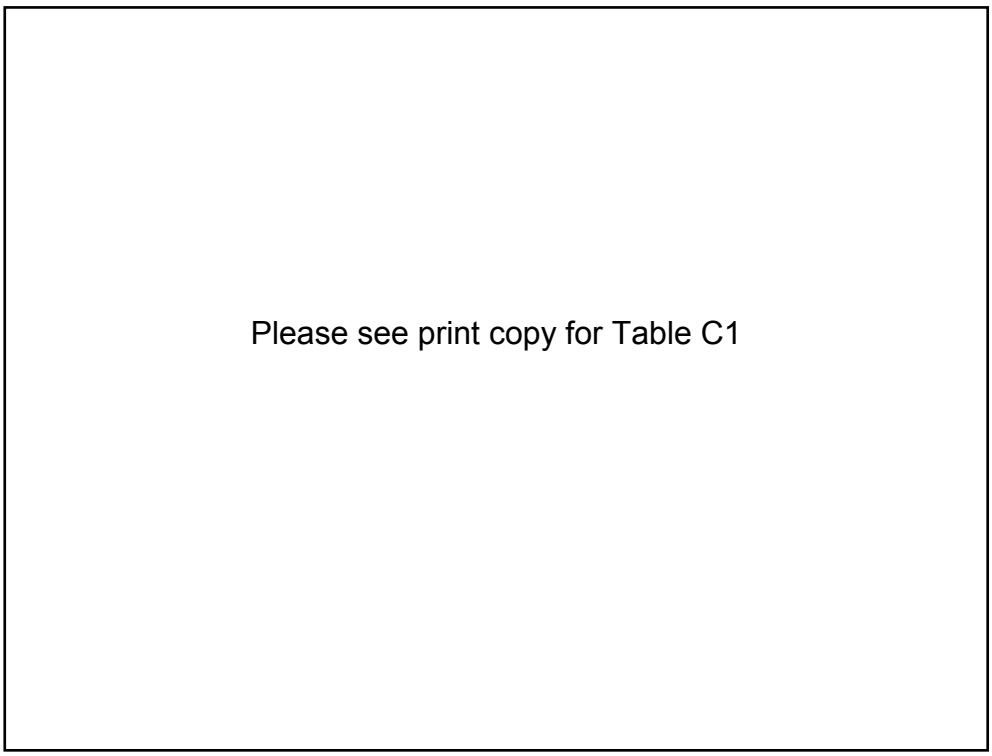
$$S(n,m) = mS(n-1,m) + S(n-1,m-1) \quad \text{for} \quad 1 \leq m < n, \tag{C.27}$$

subject to the initial conditions,

$$S(n,n) = S(n,1) = 1. \tag{C.28}$$

An excerpt from the table of values found in Abramowitz and Stegun (1965, p835) is given in Table C.1. From row 4 ($n = 4$) of this table, there are seven ways of grouping four elements into 2 sets, six ways of grouping four elements into 3 sets, and one way of grouping four elements into 4 sets. Thus, in total, there are 14 possible ways of grouping 4 elements. A property of the table that closely parallels that for Pascal's triangle of Binomial coefficients is best given by example. For instance, the number 7 in the column for $m = 2$ and row $n = 4$, is given by $7 = 1 + (2 \times 3)$, where the number 1 is the number above and to the left of 7, 3 is the number above 7, and 2 is the column number (m) (PlanetMath, 2007).

Table C.1: Stirling numbers of the second kind, $S(n, m)$ for $n = 0, \ldots, 9$, $m = 0, \ldots, 9$.



Appendix D

Derivations for Chapter 6

D.1 Diffuse Log-likelihood

In order to define w_t in the second term of the diffuse log-likelihood, given in (6.1):

$$\ln \mathrm{L}_d(Y;\psi) = -\frac{TK}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{d} w_t - \frac{1}{2}\sum_{t=d+1}^{T}\left(\ln|\mathbf{F}_t| + \nu_t'\mathbf{F}_t^{-1}\nu_t\right), \quad \text{(D.1)}$$

it is also necessary to write down the recursions for the exact initial Kalman filter.

In Durbin and Koopman (2001, Section 7.2.2), w_t is defined as the following:

$$w_t = \begin{cases} \ln|\mathbf{F}_{\infty,t}|, & \text{if } \mathbf{F}_{\infty,t} \text{ is positive definite,} \\ \ln|\mathbf{F}_{*,t}| + \nu_t^{(0)'}\mathbf{F}_{*,t}^{-1}\nu_t^{(0)}, & \text{if } \mathbf{F}_{\infty,t} = \mathbf{0}, \end{cases} \quad \text{(D.2)}$$

where the formula for $\mathbf{F}_{\infty,t}$ and $\mathbf{F}_{*,t}$ are given by the recursions for the exact initial Kalman filter.

D.1.1 Exact Initial Kalman Filter

In Section 3.5.1, the initial mean squared error matrix $\mathbf{P}_1$ was given in equation (3.50). Similarly, the matrix $\mathbf{P}_t$ has the decomposition:

$$\mathbf{P}_t = \kappa\mathbf{P}_{\infty,t} + \mathbf{P}_{*,t} + O(\kappa^{-1}), \qquad t = 2,\ldots,T, \quad \text{(D.3)}$$

where $\mathbf{P}_{\infty,t}$ and $\mathbf{P}_{*,t}$ do not depend on κ. The term $O(\kappa^{-1})$ denotes a function $f(\kappa)$ of κ such that the limit of $\kappa f(\kappa)$ as $\kappa \to \infty$ is finite. For $t = 1\ldots d$, $P_{\infty,t} \neq \mathbf{0}$ but

for $t > d$, $P_{\infty,t} = \mathbf{0}$. For $t = d+1, \ldots, T$, the Kalman filter as defined in Sections 3.5.1 and 3.5.2 applies with $\mathbf{P}_t = \mathbf{P}_{t|t-1} = \mathbf{P}_{*,t}$.

The decomposition (D.3) is defined by the exact initial Kalman filter, for which details can be found in Durbin and Koopman (2001, Section 5.2.1). The relevant formula are repeated here for convenience. Analogously to (D.3), decomposition formula may be written for $\mathbf{F}_t$ and $\mathbf{M}_t$, with

$$\begin{aligned} \mathbf{F}_{\infty,t} &= \mathbf{Z}\mathbf{P}_{\infty,t}\mathbf{Z}', & \mathbf{M}_{\infty,t} &= \mathbf{P}_{\infty,t}\mathbf{Z}', \\ \mathbf{F}_{*,t} &= \mathbf{Z}\mathbf{P}_{*,t}\mathbf{Z}' + \mathbf{H}, & \mathbf{M}_{*,t} &= \mathbf{P}_{*,t}\mathbf{Z}', \end{aligned} \tag{D.4}$$

where $t = 1 \ldots d$. Other required formula for the calculation of w_t are:

$$\begin{aligned} \nu_t^{(0)} &= y_t - \mathbf{Z}\mathbf{a}_t^{(0)} & \mathbf{a}_1^{(0)} &= \mathbf{a}, \\ \nu_t^{(1)} &= -\mathbf{Z}\mathbf{a}_t^{(1)}, & \mathbf{a}_1^{(1)} &= \mathbf{0} \end{aligned} \tag{D.5}$$

The updating equations for $\mathbf{P}_{\infty,t}$, $\mathbf{P}_{*,t}$ and $\mathbf{a}_{t+1}^{(0)}$ are:

$$\begin{aligned} \mathbf{P}_{\infty,t+1} &= \mathbf{T}\mathbf{P}_{\infty,t}\mathbf{L}_t^{(0)\prime}, \\ \mathbf{P}_{*,t+1} &= \mathbf{T}\mathbf{P}_{\infty,t}\mathbf{L}_t^{(1)\prime} + \mathbf{T}\mathbf{P}_{*,t}\mathbf{L}_t^{(0)\prime} + \mathbf{G}\mathbf{Q}\mathbf{G}', \\ \mathbf{a}_{t+1}^{(0)} &= \mathbf{T}\mathbf{a}_t^{(0)} + \mathbf{K}_t^{(0)}\nu_t^{(0)}, \end{aligned} \tag{D.6}$$

where

$$\begin{aligned} \mathbf{L}_t^{(0)} &= \mathbf{T} - \mathbf{K}_t^{(0)}\mathbf{Z}, & \mathbf{L}_t^{(1)} &= -\mathbf{K}_t^{(1)}\mathbf{Z}, \\ \mathbf{K}_t^{(0)} &= \mathbf{T}\mathbf{M}_{\infty,t}\mathbf{F}_t^{(1)}, & \mathbf{K}_t^{(1)} &= \mathbf{T}\mathbf{M}_{*,t}\mathbf{F}_t^{(1)} + \mathbf{T}\mathbf{M}_{\infty,t}\mathbf{F}_t^{(2)}, \\ \mathbf{F}_t^{(1)} &= \mathbf{F}_{\infty,t}^{-1}, & \mathbf{F}_t^{(2)} &= -\mathbf{F}_{\infty,t}^{-1}\mathbf{F}_{*,t}\mathbf{F}_{\infty,t}^{-1}. \end{aligned} \tag{D.7}$$

D.2 Derivations for Section 6.2.1

Using formula in Section D.1, log-likelihood equation (6.2) for the univariate local level seasonal model is derived in this section. Since $\mathbf{F}_{\infty,t}$ and $\mathbf{F}_t$ are scalar quantities here, and since $K = 1$, equation (D.1) may be rewritten as:

$$\begin{aligned} \ln \mathrm{L}_d(Y;\psi) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{d} w_t - \frac{1}{2}\sum_{t=d+1}^{T}\left(\ln F_t + \nu_t' F_t^{-1}\nu_t\right) \\ &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{d} w_t - \frac{1}{2}\sum_{t=d+1}^{T}\ln F_t - \frac{1}{2}\sum_{t=d+1}^{T}\frac{\nu_t^2}{F_t}. \end{aligned} \tag{D.8}$$

D.2.1 Calculation of w_t

The system matrices $\mathbf{Z}$, $\mathbf{T}$ and $\mathbf{G}$ for the local level seasonal model are given in (4.8). Here, $\mathbf{P}_{\infty,1} = \mathbf{I}_4$ and $\mathbf{P}_{*,1} = \mathbf{0}_{4\times 4}$. Applying the recursions given in the previous section, it is found that for $t = 1 \ldots 4$, $\mathbf{P}_{\infty,t} \neq \mathbf{0}_{4\times 4}$. This implies that $d = 4$, and hence the values of w_t for $t = 1 \ldots 4$ are required.

The calculations of the exact initial Kalman filter have been carried out using MAPLE. The MAPLE code has been included in Appendix E (the CD enclosed with this thesis). The following results were found:

$$F_{\infty,1} = 2, \quad F_{\infty,2} = 4, \quad F_{\infty,3} = \frac{3}{2}, \quad F_{\infty,4} = \frac{4}{3}. \tag{D.9}$$

Since $\ln F_{\infty,t} \neq 0$ for $t = 1 \ldots 4$, the constant terms $w_1, \ldots w_4$ are calculated by $w_t = \ln F_{\infty,t}$ and hence

$$\begin{aligned} \sum_{t=1}^{4} w_t &= \ln 2 + \ln 4 + \ln\frac{3}{2} + \ln\frac{4}{3} \\ &= \ln 16 \\ &= 4\ln 2. \end{aligned}$$

Substituting into (D.8) and noting $d = 4$, $\ln \mathrm{L}_d(Y;\psi)$ becomes

$$\ln \mathrm{L}_d(Y;\psi) = -\frac{T}{2}\ln(2\pi) - 2\ln 2 - \frac{1}{2}\sum_{t=5}^{T}\ln F_t - \frac{1}{2}\sum_{t=5}^{T}\frac{\nu_t^2}{F_t}, \tag{D.10}$$

which is given as equation (6.2).

D.2.2 Concentrated Diffuse Log-likelihood

When $F_t = \sigma^2_{U,\,\varepsilon} F^c_t$ (6.5), is substituted into (D.10) above, the concentrated diffuse log-likelihood function for the univariate model is derived. Detailed steps are provided here:

$$\begin{aligned}
\ln \mathrm{L}_{dc} &= -\frac{T}{2}\ln(2\pi) - 2\ln 2 - \frac{1}{2}\sum_{t=5}^{T} \ln\left(\sigma^2_{U,\,\varepsilon} F^c_t\right) - \frac{1}{2}\sum_{t=5}^{T} \frac{\nu_t^2}{\sigma^2_{U,\,\varepsilon} F^c_t} \\
&= -\frac{T}{2}\ln(2\pi) - 2\ln 2 - \frac{1}{2}\sum_{t=5}^{T} \left(\ln\sigma^2_{U,\,\varepsilon} + \ln F^c_t\right) - \frac{1}{2\sigma^2_{U,\,\varepsilon}}\sum_{t=5}^{T} \frac{\nu_t^2}{F^c_t} \\
&= -\frac{T}{2}\ln(2\pi) - 2\ln 2 - \frac{(T-4)}{2}\ln(\sigma^2_{U,\,\varepsilon}) - \frac{1}{2}\sum_{t=5}^{T} \ln F^c_t - \frac{1}{2\sigma^2_{U,\,\varepsilon}}\sum_{t=5}^{T} \frac{\nu_t^2}{F^c_t}.
\end{aligned} \tag{D.11}$$

This result is given in (6.6).

D.3 Derivations for Section 6.2.2

In this section the result for $\ln \mathrm{L}_{(m),c}$ given in (6.15) is derived using formula in Section D.1 for the multivariate local level seasonal model.

D.3.1 Calculation of w_t

The system matrices $\mathbf{Z}$, $\mathbf{T}$ and $\mathbf{G}$ for the multivariate local level seasonal model are given in (4.11). Here, the $\mathbf{P}_{\infty,1}$ matrix (with dimensions 10×10) is such that the lower right 2×2 block diagonal of a 10×10 identity matrix is replaced by a 2×2 zero matrix. The $\mathbf{P}_{*,1}$ matrix, which holds the variance of the stationary part of α_1, is no longer a zero matrix, but now includes the $\mathbf{\Sigma}_{(m),\,\varepsilon}$ covariance matrix (6.11) in the lower right 2×2 block diagonal. All other elements of the $\mathbf{P}_{*,1}$ matrix are zero.

Applying the recursions given in Section D.1, and recalling that $\mathbf{H} = \mathbf{0}_{2\times 2}$ here since the measurement error is placed within the state vector, it is found that for $t = 1 \ldots 4$, $\mathbf{P}_{\infty,t} \neq \mathbf{0}_{10\times 10}$. This implies that $d = 4$, and hence the values $w_1, \ldots w_4$ are required.

The calculations of the exact initial Kalman filter have been carried out using MAPLE. The MAPLE code has been included in Appendix E. The following results were found:

$$\mathbf{F}_{\infty,1} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \mathbf{F}_{\infty,2} = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}, \quad \mathbf{F}_{\infty,3} = \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}, \quad \mathbf{F}_{\infty,4} = \begin{pmatrix} \frac{4}{3} & 0 \\ 0 & \frac{4}{3} \end{pmatrix}. \tag{D.12}$$

Since $\ln\mathbf{F}_{\infty,t} \neq \mathbf{0}_{2\times 2}$ for $t = 1 \ldots 4$, the constant terms $w_1, \ldots w_4$ are calculated by $w_t = \ln|\mathbf{F}_{\infty,t}|$. Note that since each of the matrices in (D.12) are diagonal, the determinant of each is easily calculated as the product of the diagonal elements. Hence,

$$\begin{aligned} \sum_{t=1}^{d} w_t &= \ln 4 + \ln 16 + \ln\frac{9}{4} + \ln\frac{16}{9} \\ &= \ln\left(4 \times 16 \times \frac{9}{4} \times \frac{16}{9}\right) \\ &= 2\ln 16. \\ &= 8\ln 2. \end{aligned}$$

Substituting into (D.1) and noting that $d = 4$ and $K = 2$, the diffuse log-likelihood for the untransformed multivariate model is denoted by $\ln \mathrm{L}_{(m),d}(Y;\psi)$. It is given by

$$\ln \mathrm{L}_{(m),d}(Y;\psi) = -T\ln(2\pi) - 4\ln 2 - \frac{1}{2}\sum_{t=5}^{T}\left(\ln|\mathbf{F}_{(m),t}| + \nu_t'\mathbf{F}_{(m),t}^{-1}\nu_t\right). \tag{D.13}$$

D.3.2 Concentrated Diffuse Log-likelihood

When $\mathbf{F}_{(m),t} = \sigma_c^2 \mathbf{F}^c_{(m),t}$ given in (6.14), is substituted into (D.13) above, the concentrated diffuse log-likelihood function for the univariate model is derived. Detailed steps are provided here:

$$
\begin{aligned}
\ln \mathrm{L}_{(m),dc} &= -T\ln(2\pi) - 4\ln 2 - \frac{1}{2}\sum_{t=5}^{T}\left(\ln|\sigma_c^2\mathbf{F}^c_{(m),t}| + \nu_t^{'}\left[\sigma_c^2\mathbf{F}^c_{(m),t}\right]^{-1}\nu_t\right) \\
&= -T\ln(2\pi) - 4\ln 2 - \frac{1}{2}\sum_{t=5}^{T}\left(\ln|\sigma_c^2\mathbf{I}_2\mathbf{F}^c_{(m),t}|\right) \\
&\quad -\frac{1}{2}\sum_{t=5}^{T}\left(\frac{1}{\sigma_c^2}\nu_t^{'}\left[\mathbf{F}^c_{(m),t}\right]^{-1}\nu_t\right) \\
&= -T\ln(2\pi) - 4\ln 2 - \frac{1}{2}\sum_{t=5}^{T}\left(\ln|\sigma_c^2\mathbf{I}_2|\right) - \frac{1}{2}\sum_{t=5}^{T}\left(\ln|\mathbf{F}^c_{(m),t}|\right) \\
&\quad -\frac{1}{2\sigma_c^2}\sum_{t=5}^{T}\nu_t^{'}\left[\mathbf{F}^c_{(m),t}\right]^{-1}\nu_t \\
&= -T\ln(2\pi) - 4\ln 2 - \frac{1}{2}(T-4)\ln\left[\left(\sigma_c^2\right)^2\right] - \frac{1}{2}\sum_{t=5}^{T}\left(\ln|\mathbf{F}^c_{(m),t}|\right) \\
&\quad -\frac{1}{2\sigma_c^2}\sum_{t=5}^{T}\nu_t^{'}\left[\mathbf{F}^c_{(m),t}\right]^{-1}\nu_t \\
&= -T\ln(2\pi) - 4\ln 2 - (T-4)\ln\left(\sigma_c^2\right) - \frac{1}{2}\sum_{t=5}^{T}\left(\ln|\mathbf{F}^c_{(m),t}|\right) \\
&\quad -\frac{1}{2\sigma_c^2}\sum_{t=5}^{T}\nu_t^{'}\left[\mathbf{F}^c_{(m),t}\right]^{-1}\nu_t. \qquad \text{(D.14)}
\end{aligned}
$$

This result is given in (6.15).

Appendix E

Program Files

The enclosed CD contains the source code of the programs used in the calculation of theoretical and simulation results. The files have been organised into folders which are named with the relevant chapter and/or section of the thesis. If the program output provides results for a particular table and/or figure in the thesis, the table and figure are referred to by number. Typically, there is a main (run) program which calls other secondary programs such as model specifications. If the program is a secondary program, the main program is cited. Some of these secondary programs are used to obtain results in more than one chapter or section of a chapter. These have been listed each time so that the program list for each chapter stands alone.

The program files have been created using the following software: `S-PLUS` (Version 7.0 for Windows with `S+FinMetrics` module), `Maple` (Version 9.01) and Microsoft Excel (Microsoft Office Professional Edition 2003). The `S-PLUS` programs (.ssc) may be opened with a text editor such as Wordpad or Notepad. The `Maple` files have been saved as both a `Maple` file (.mw) and a text file (.txt) without output. The corresponding text file is not listed here, although it appears on the CD. Parameter input files are saved in Microsoft Excel spreadsheets.

Please see print copy for Appendix E: Program files

Bibliography

Abramowitz, M. and Stegun, I. (eds.) (1965) *Handbook of mathematical functions with formulas, graphs amd mathematical tables*. United States Department of Commerce.

ABS (2007a) Business Indicators, Australia - Explanatory Notes, Cat.no. 5676.0. *Tech. rep.*, Canberra, accessed 10/9/2007, http://www.abs.gov.au.

— (2007b) Sales of Australian Wine and Brandy by Winemakers, Cat.no. 8504.0. *Tech. rep.*, Canberra, accessed 2/2/2007, http://www.abs.gov.au.

Anderson, B. and Moore, J. (1979) *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, N.J.

Ansley, C. and Kohn, R. (1986) Prediction mean squared error for state space models with estimated parameters. *Biometrika*, **73**, 467–473.

Bell, W. R. (1984) Signal extraction for non-stationary time series. *The Annals of Statistics*, **12**, 646–664.

Bell, W. R. and Hillmer, S. C. (1984) Issues involved with the seasonal adjustment of economic time series. In *Modelling Seasonality* (ed. S. Hylleberg), 83–138. Oxford University Press.

Box, G. E. P. and Jenkins, G. M. (1970) *Time series analysis : forecasting and control*. San Francisco: Holden-Day.

Burman, J. P. (1980) Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society, Series A*, **143**, 321–337.

Burridge, P. and Wallis, K. (1984) Unobserved-components models for seasonal adjustment filters. *Journal of Business and Economic Statistics*, **2**, 350–359.

— (1985) Calculating the variance of seasonally adjusted series. *Journal of the American Statistical Association*, **80**, 541–552.

Dagum, E. (1988) The X-11-ARIMA/88 seasonal adjustment method, foundations and user's manual. *Tech. rep.*, Time Series Research and Analysis Division, Statistics Canada.

Dagum, E. and Quenneville, B. (1993) Dynamic linear models for time series components. *Journal of Econometrics*, **55**, 333–351.

Durbin, J. and Koopman, S. (2000) Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perpectives. *Journal of the Royal Statistical Society, Series B(Statistical Methodology)*, **62**, 3–56.

— (2001) *Time Series Analysis by State Space Methods*, vol. 24 of *Oxford Statistical Science Series*. New York: Oxford University Press.

Engle, R. F. (1978) Estimating structural models of seasonality. In *Seasonal Analysis of Economic Time Series* (ed. A. Zellner), 281–295. Washington DC: US Department of Commerce, Bureau of Census.

Feder, M. (2001) Time series analysis of repeated surveys: the state-space approach. *Statistica Neerlandica*, **55**, 182–199.

Feder, M., Nathan, G. and Pfeffermann, D. (2000) Multilevel modelling of complex survey longitudinal data with time varying random effects. *Survey Methodology*, **26**, 53–65.

Fernandez, F. and Harvey, A. (1990) Seemingly unrelated time series equations and a test for homogeneity. *Journal of Business and Economic Statistics*, **8**, 71–81.

Findley, D., Monsell, B., Bell, W., Otto, M. and Chen, B. (1998) New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics*, **16**, 127–152.

Gatto, R. (2006) Series revision and seasonal adjustment of short time series in presence of a major methodological break. Paper given at the *Conference on seasonality, seasonal adjustment and their implications for short term analysis and forecasting*, Luxembourg, May 2006.

Geweke, J. (1978) The temporal and sectoral aggregation of seasonally adjusted time series. In *Seasonal Analysis of Economic Time Series* (ed. A. Zellner), 411–427. U.S. Government Printing Office.

Ghysels, E. (1997) Seasonal adjustment and other data transformations. *Journal of Business and Economic Statistics*, **15**, 410–418.

Godambe, V. and Heyde, C. (1987) Quasi-likelihood and optimal estimation. *International Statistical Review*, **55**, 231–244.

Gomez, V. (1999) Three equivalent methods for filtering finite nonstationary time series. *Journal of Business and Economic Statistics*, **17**, 109–116.

Gomez, V. and Maravall, A. (1996) Programs TRAMO and SEATS. Instructions for the user (with some updates). *Working Paper 9628*, Servicio de Estudios, Banco de Espana.

Hamilton, J. (1986) A standard error for the estimated state vector of a state space model. *Journal of Econometrics*, **33**, 387–397.

Harrison, J. and Stevens, C. (1976) Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society*, **38**, 205–247.

Harvey, A. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Harvey, A. and Chung, C.-H. (2000) Estimating the underlying change in unemployment in the UK. *Journal of the Royal Statistical Society*, **163**, 303–339.

Harvey, A. and Durbin, J. (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*, **149**, 187–227.

Harvey, A. and Scott, A. (1994) Seasonality in dynamic regression models. *The Economic Journal*, **104**, 1324–1345.

Harvey, A. and Shephard, N. (1993) Structural time series models. In *Handbook of Statistics, Volume 11* (eds. G. Maddala, C. Rao and H. Vinod), chap. 10, 261–302. Amsterdam: Elsevier Science Publishers.

Harvey, A. and Todd, P. (1983) Forecasting economic time series with structural and Box-Jenkins models: A case study. In *Modelling Seasonality* (ed. S. Hylleberg), 341–358. Oxford University Press.

Harville, D. A. (1997) *Matrix algebra from a statistician's persective.* New York: Springer-Verlag.

Hausman, J. and Watson, M. (1985) Errors in variables and seasonal adjustment procedures. *Journal of the American Statistical Association*, **80**, 531–540.

Heyde, C. (1997) *Quasi-Likelihood and Its Application. A general approach to optimal parameter estimation.* New York: Springer-Verlag.

Hillmer, S. and Tiao, G. (1982) An ARIMA model-based approach to seasonal adjustment. In *Modelling Seasonality* (ed. S. Hylleberg), 321–339. Oxford University Press.

Holy Bible, New International Version (1985) *The NIV Study Bible.* Zondervan.

Hood, C. and Findley, D. F. (2003) Comparing direct and indirect seasonal adjustments of aggregate series. In *Seasonal Adjustment* (eds. M. Manna and R. Peronaci), chap. 1, 9–21. Frankfurt am Main, Germany: European Central Bank.

Jain, R. K. (2001) A state space model-based method of seasonal adjustment. *Monthly Labour Review*, **July**, 37–45.

de Jong, P. (1991) The diffuse Kalman filter. *The Annals of Statistics*, **19**, 1073–1083.

Kaiser, R. and Maravall, A. (2003) Seasonal outliers in time series. *Estadistica, Journal of the Inter-American Statistcial Institute*, **15**, 101–142. Special Issue on Time Series.

Kalman, R. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions, ASMA, Series D*, **82**, 35–45.

Koopman, S. (1997) Exact initial Kalman filtering and smoothing for non-stationary time series models. *Journal of the American Statistical Association*, **92**, 1630–1638.

Koopman, S. and Durbin, J. (2000) Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, **21**, 281–296.

Koopman, S., Shephard, N. and Doornik, J. (1999) Statistical algorithms for models in state space using SsfPack 2.2. *The Econometrics Journal*, **2**, 113–166.

Ladiray, D. and Mazzi, G. (2003) Seasonal adjustment of European aggregates: direct versus indirect approach. In *Seasonal Adjustment* (eds. M. Manna and R. Peronaci), 37–65. Frankfurt am Main, Germany: European Central Bank.

Ladiray, D. and Quenneville, B. (2001) *Seasonal adjustment with the X-11 method*. New York: Springer-Verlag.

Lin, Y.-X. (2007) An alternative derivation of the Kalman filter using the quasi-likelihood method. *Journal of Statistical Planning and Inference*, **137**, 1627–1633.

Maravall, A. (1985) On structural time series models and the characterization of components. *Journal of Business and Economic Statistics*, **3**, 350–355.

— (1993) Stochastic linear trends. *Journal of Econometrics*, **56**, 5–37.

— (1995) Unobserved components in economic time series. In *The Handbook of Applied Econometrics* (eds. M. Pesaran and M. Wickens), 12–72. Basil Blackwell, Oxford.

— (2006) An appliction of the TRAMO-SEATS automatic procedure; direct versus indirect adjustment. *Computational Statistics and Data Analysis*, **50**, 2167–2190.

Marshall, P. (1990) *Analysis of a cross-section of time series using structural time series models*. Ph.D. thesis, London School of Economics and Political Science, University of London.

— (1992) Estimating time-dependent means in dynamic models for cross-sections of time series. *Empirical Economics*, **17**, 25–33.

Mazzi, G. and Savio, G. (2003) Seasonal adjustment of short series. In *Proceedings of Statistics Canada Symposium*, no. 11-522-X1E in Statistics Canada International Symposium Series.

Mir, A. and Rondonotti, V. (2003) The performance of X-12 in the seasonal adjustment of short time series. In *Seasonal Adjustment* (eds. M. Manna and R. Peronaci), 149–160. Frankfurt am Main, Germany: European Central Bank.

Moosa, I. and Lenten, L. (2000) In defence of model-based seasonal adjustment: An illustration using Australian data. *Australian Economic Papers*, **39**, 372–392.

Otranto, E. and Triacca, U. (2002) Measures to evaluate the discrepancy between direct and indirect model-based seasonal adjustment. *Journal of Official Statistics*, **18**, 511–530.

Pena, D., Tsay, R. and Tiao, G. C. (eds.) (2001) *A Course in Time Series Analysis*. Chichester: John Wiley and Sons.

Penzer, J. (2006) Diagnosing seasonal shifts in time series using state space models. *Statistical Methodology*, **3**, 193–210.

Pfeffermann, D., Feder, M. and Signorelli, D. (1998) Estimation of autocorrelations of survey errors with application to trend estimation in small areas. *Journal of Business and Economic Statistics*, **16**, 339–348.

Pfeffermann, D. and Tiller, R. (2003) State-space modelling with correlated measurements with application to small area estimation under benchmark constraints. *Working Paper M03/11*, Southampton Statistical Sciences Research Institute, University of Southampton.

— (2005) Bootstrap approximation to prediction MSE for state space models with estimated parameters. *Journal of Time Series Analysis*, **26**, 893–916.

Pierce, D. A. (1980) Data revisions with moving average seasonal adjustment procedures. *Journal of Econometrics*, **14**, 95–114.

Planas, C. and Campolongo, F. (2001) The seasonal adjustment of contemporaneously aggregated series. *Working Paper 8221011/1-Lot3*, EUROSTAT, Luxembourg.

Planas, C. and Depoutot, R. (2002) Controlling revisions in ARIMA-model-based seasonal adjustment. *Journal of Time Series Analysis*, **23**, 193–213.

Planas, C. and Rossi, A. (2004) Can inflation data improve the real-time reliability of output gap estimates? *Journal of Applied Econometrics*, **19**, 121–133.

PlanetMath (2007) Stirling numbers of the second kind, accessed 6/9/2007, http://planetmath.org/?op=getobj& from=objects& id=2805.

Pollock, D. (2002) A review of TSW: the Windows version of the TRAMO-SEATS program. *Journal of Applied Econometrics*, **17**, 291–299.

Proietti, T. (2000) Comparing seasonal components for structural time series models. *International Journal of Forecasting*, **16**, 247–260.

Quenneville, B. and Singh, A. (2005) Bayesian prediction mean squared error for state space models with estimated parameters. *Journal of Time Series Analysis*, **21**, 219–236.

Riani, M. (1998) Weights and robustness of model-based seasonal adjustment. *Journal of Forecasting*, **17**, 19–34.

Shiskin, J., Young, A. and Musgrave, J. (1967) The X-11 variant of the census method II seasonal adjustment program. *Technical Paper 15*, Bureau of the Census, U.S. Department of Commerce, Washington D.C.

Sridharan, S., Vujic, S. and Koopman, S. (2003) Intervention time series analysis of crime rates. *Timbergen Institute Discussion Paper*.

Taylor, J. B. (1978) Comments on 'The temporal and sectorial aggregation of seasonally adjusted time series' by John Geweke. In *Seasonal Analysis of Economic Time Series* (ed. A. Zellner), 431–432. U.S. Government Printing Office.

Tweddle, I. (2003) *James Stirling's Methodus Differentialis: An annotated translation of Stirling's text*. London: Springer.

Whittle, P. (1963) *Prediction and Regulation using Least Squares Methods*. London: English Universities Press.

Zivot, E. and Wang, J. (2006) *Modeling Financial Time Series with S-PLUS*. New York: Springer Science+Business Media Inc, 2nd edn.

Zivot, E., Wang, J. and Koopman, S. (2004) State space modelling in macroeconomics and finance using SsfPack in S+Finmetrics. In *State Space and Unobserved Component Models* (eds. A. Harvey, S. Koopman and N. Shephard), chap. 13, 284–335. Cambridge University Press.