Modelling Joint Autoregressive Moving Average Processes

Ross S Bowden

This thesis is presented for the degree of Doctor of Philosophy of
Murdoch University
2017

Abstract

This thesis explores Joint Autoregressive Moving-Average (JARMA) models for independent replicated univariate time series with common ARMA coefficients whose innovations variances are either in common, unique to each series or vary with the series mean. The constraint of a common variance is also applied to vector ARMA processes. Interleaving is shown to represent replicated series with a common variance as one series from the same process. The time and frequency domain properties of interleaved replicated stationary and invertible processes are established. As an aid to identification, hypothesis tests for comparing series are reviewed and several new tests are presented and explored along with a graphical method for identification. Unconditional maximum likelihood estimates of the parameters of various JARMA processes are derived using the methods of joint likelihood and interleaving. The properties of the estimators are examined using simulation and asymptotics. Finally JARMA models are fitted to over 60 years of daily univariate and bivariate temperature data to estimate differences in level due to location and climate change.

"He who loves practice without theory is like the sailor who boards ship without a rudder and compass. He does not know where he may be cast." - Leonardo da Vinci

I declare that this thesis is my own account of my research except where other sources are fully acknowledged. It contains, as its main content, work which has not previously been submitted for a degree at any tertiary education institution.

. . . . . .
Ross Bowden

Acknowledgements

Firstly I would very much like to thank Dr Brenton Clarke for his supervision during my PhD. His encouragement and enthusiasm kept me going during some difficult times and his review of my research was invaluable. I am also very grateful to my co-supervisor, Dr Nicola Armstrong, who provided a wealth of suggestions for improving my thesis.

I thank my family, especially my wife, Gail, for supporting me during the past eight years. Gail has been my wonderful partner for 34 years and I couldn't have completed my thesis without her. I also appreciate the help of my daughter, Clare, who diligently checked the bibliography. I thank my mother, Aileen Bowden, for her determination that I should attend university and become the first in my extended family to do so. I thank my friends, Dr Harry Suehrcke and Mr Scott Davis, for their unwavering support. Finally I thank the authors of the computer packages, R and Lyx, for producing such extraordinary software.

This research was undertaken on the land of the Whadjuk people of the Noongar nation. I wish to pay my respects to their elders - past, present and future - and to acknowledge the important role all Aboriginal and Torres Strait Islander people play in Australian society.

Eight years ago I remember quoting Robert Frost, "The woods are lovely, dark and deep. But I have promises to keep and miles to go before I sleep". So now I can sleep.

To the memory of my mother and father, Aileen and Arthur Bowden.

Contents

Chapter 1

Introduction

This thesis addresses the specification, identification, estimation and application of joint independent stationary and invertible autoregressive moving average (ARMA) and vector ARMA (VARMA) time series processes where some or all of the parameters are in common.

In analysing ARMA time series, it is typically assumed that only one realization is available for model fitting. However, there are many circumstances where multiple independent realisations of the same ARMA process are available. Examples can be found in replicated experiments on evolving chemical processes, repeated measures on "individuals" recorded at fixed time intervals (in panel and longitudinal studies) and daily weather data for the same season over many years.

A common ARMA generating mechanism is likely to apply to each realization (possibly differing in mean level) with one model to be fitted to all realizations simultaneously. This could apply equally well to multivariate time series such as daily maximum and minimum temperatures (see Chapter 7).

It is also possible that the generating process has the same ARMA parameters for each series but the innovations variances may vary reflecting differing levels of variation for each series. Furthermore the variances could be proportional to each series mean as is often encountered in observed data (for example, in quarterly Gross Domestic Product values by country). Models with differences in the variance are candidates for a generalised model to simulate daily solar radiation readings across a range of sites. These solar models have been explored in a preliminary sense

by Graham et al. [1988] and could be extended to encompass the relationship to sunshine fraction as investigated by the author of this dissertation in Suehrcke, Bowden, and Hollands [2013][1].

This thesis develops and compares the methods of joint likelihood and interleaving (Bowden and Clarke [2012]) to simultaneously model time series that have the same ARMA parameters but whose innovations variances are either the same, unique to each series or proportional to the series mean. The time and frequency domain properties of interleaved stationary and invertible processes are established to better understand the effects of interleaving.

Models for replicated independent time series are discussed in Chapter 2 along with interleaving. To aid in identification, in Chapter 3, hypothesis tests for comparing time series are reviewed and several new tests are presented. The size and power of the tests are compared and a graphical method for identifying the appropriate model for joint processes is proposed.

Unconditional maximum likelihood estimates of the parameters of joint ARMA processes are derived and explored in Chapter 4. The estimation makes minimal use of bespoke programming by incorporating existing software in deriving joint and interleaved likelihood functions. The properties of the estimators are examined using both asymptotic methods and simulation (for the latter, see Chapter 5). Two published papers co-authored by the writer of this thesis (Bowden and Clarke [2012] and Bowden and Clarke [2017]) are included (Chapters 6 and 7) that fit ARMA and VARMA models to (replicated) daily maximum and minimum temperature data for Perth, Western Australia. The analysis is used to measure changes in mean temperature due to climate change and movements in recording site. Finally Chapter 8 contains conclusions and recommendations.

In this thesis, the main reference texts used are Priestley [1981] and Brockwell and Davis [1991] along with Jenkins and Watts [1968], Chatfield [2003] and Shumway and Stoffer [2011]. All vectors are column vectors and all matrices and vectors are shown in bold font. In selected parts of this thesis (see, for example, Section 2.4) the time series designation, $\{..\}$, say $\{y_t\}$, is used as a form of $\{..\}_{t=1}^{n}$, that is, of $\{y_t\}_{t=1}^{n}$ but where t ranges across all integers. The term "likelihood" herein refers to unconditional likelihood.

[1] Awarded "2013 Best Paper" by the journal, Solar Energy.

The statistical package used for data analysis was chosen in order that it be well-known, programmable, provides excellent graphical capabilities, creates reasonably fast code, is easily documented, is associated with an extensive library of time series routines and has available an efficient IDE (Integrated Development Environment). R (Version 3.3.1) (R Development Core Team [2010]) is the candidate of choice in this regard especially given its time series capabilities (see McLeod et al. [2012]). All the fitting routines used the "BFGS" (Broyden, Fletcher, Goldfarb and Shanno) quasi-Newton optimisation method (see the `optim` function in R).

The next chapter defines various joint time series models and explores the properties of interleaved ARMA and VARMA processes.

Chapter 2

Joint ARMA Processes

A Joint ARMA process (that is, a JARMA process) is defined in this thesis as a set of independent stationary and invertible ARMA processes with common ARMA coefficients.

The simplest JARMA model assumes repeated independent realisations of the same univariate stationary and invertible ARMA process (that is, with common ARMA coefficients *and* innovations variances). To generalise this model, the component processes, although having the same ARMA parameters, may differ in their innovations variances. This could be the case for example in simultaneous modelling of the outputs from independent physical processes. These could have the same autocorrelation structure but the processes may have varying levels of random disturbance represented by different innovations variances. In this the mean level could also vary and the variance may be proportional to the mean level of each series presenting another form of JARMA process.

Formal definitions of the these models, which are the focus of this thesis, are presented in this chapter.

It is also possible that the ARMA coefficients as well as the innovations variances may vary between series reflecting processes with independent likelihood functions. This is shown in Section 2.5 to be equivalent, after interleaving, to a single-series periodic ARMA (that is, PARMA) process with parameter constraints (that is, some parameters equal to zero). Finally, parameters between the independent replicated series may be related by some functional form. These processes and the PARMA

processes just described will not be estimated in this thesis but can be fitted using a generalisation of the joint likelihood approach (see, for example, Section 4.1).

2.1 Univariate Joint ARMA Process

Let the m stationary and invertible independent univariate time series, $\{y_{i,t}\}_{t=1}^{n_i}$, $i = 1, \ldots, m$, be generated by

$$\phi(B)(y_{i,t} - \boldsymbol{\beta}^\top \mathbf{z}_{i,t}) \;=\; \theta(B)\epsilon_{i,t}, \tag{2.1.1}$$

where each element of the series has a linear mean, $\boldsymbol{\beta}^\top \mathbf{z}_{i,t}$, with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ and $\boldsymbol{z}_{i,t} = (z_{1,i,t}, \ldots, z_{k,i,t})^\top$ and where the $\phi(B)$ and $\theta(B)$ are polynomials in B, the backshift operator, of orders p and q, that is, $\phi(B) = 1+\phi_1 B+\ldots+\phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$. Also $\{\epsilon_{i,t}\}_{t=1}^{n_i}$ is the i^{th} independent identically distributed time series with $E(\epsilon_{i,t}) = 0$, $\mathrm{Var}(\epsilon_{i,t}) = \sigma^2_{\epsilon_i}$ and $E(\epsilon_{i,t}\epsilon_{i,t-\kappa}) = 0$ for all $\kappa \neq 0$. The regression coefficients can be assumed to be unique to each series and then combined with the use of additional dummy regressor variables into one common $\boldsymbol{\beta}$ for specification purposes.

This process comprising m independent time series is defined here as a JARMA process and either,

1. there is one common innovations variance, $\sigma^2_{\epsilon_i} = \sigma^2_\epsilon$ for all i. This is called a Replicated ARMA (RARMA) process of order (p, q, m), that is, a RARMA(p,q,m) process. The standard single-series ARMA model corresponds to the RARMA model with $m = 1$. It has been shown (see Theorem 1 and Bowden and Clarke [2012] (see Chapter 6)) that a RARMA process can be represented by one interleaved ARMA series with certain ARMA coefficients set to zero.

2. there are unique innovations variances per series, $\left\{\sigma^2_{\epsilon_i}\right\}_{i=1}^{m}$. This process, comprising m independent time series with the same ARMA parameters but differing variances, is called an Almost Identical ARMA (AIARMA) process.

3. as a special case of an AIARMA process, the innovations standard deviations are proportional to each series mean, that is, $\sigma_{\epsilon_i} = c\mu_i$ where $E(y_{i,t}) = \mu_i$. This is called a Conditional Almost Identical ARMA (CAIARMA) Process.

As an example of a JARMA process, consider two independent identically distributed time series ($m = 2$) of length 100 ($n_1 = n_2 = 100$) which are each generated as an ARMA(1,1) process. The generating mechanisms share the same ARMA coefficients ($\phi_1 = -0.4$ and $\theta_1 = 0.3$). However the first series has mean, $\mu_1 = 1$, and innovations variance, $\sigma^2_{\epsilon_1} = 1$, whilst the second has mean, $\mu_2 = 2$, and innovations variance, $\sigma^2_{\epsilon_2} = 4$. This is an AIARMA process but can also be represented as a CAIARMA process with $c = 1$.

Hence the CAIARMA generating model for the two independent identically distributed series, $\{y_{i,t}\}_{t=1}^{100}, i = 1, 2$, is,

$$(1 - 0.4B)(y_{i,t} - \boldsymbol{\beta}^\top \mathbf{z}_{i,t}) \;=\; (1 + 0.3B)\epsilon_{i,t},$$

where $\boldsymbol{\beta} = (1, 2)^\top$ and $c = 1$, with $\mathbf{z}_{1,t} = (1, 0)^\top$and $\mathbf{z}_{2,t} = (0, 1)^\top$ for all t.

The next section reviews the literature on univariate joint ARMA models.

2.1.1 Literature Review for Joint Univariate Time Series

An original reference on simultaneous estimation in a repeated measures environment is Yates [1960]. He presents a correction for first order autocorrelation in the analysis of repeated measures in sample surveys and this is further explored in Scott and Smith [1974].

Azzalini [1981] examines the fitting of a single model to replicated series from an autoregressive process of order one or two. The emphasis is on asymptotic efficiency where the number of replications (as opposed to the length of each series) tends to infinity. The conditional and unconditional maximum likelihood results are derived and compared. The latter are shown to be a substantially superior result especially near the stationarity boundary. Azzalini [1984] enhances the modelling to incorporate what is effectively time series modelling in a random-effects two-way ANOVA setting. The results are applied to the plasma citrate concentration of n=10 subjects measured at 14 equal time points to detect changes in the plasma readings during the day.

Wong et al. [2002], in an extension of Wong and Miller [1990], model repeated realisations of ARMA processes where the error variance and the number of realisations are allowed to vary over time. Also each repeated realisation is assumed to

be a combination of an underlying ARIMA process and an additional independent error term (ARIMAN process)[1]. The model differs from those considered in this thesis because (a) the current models do not employ the additional noise term and (b) Wong et al. [2002] constrain the underlying ARIMA process (i.e. without noise) to not only have the same parameters but the same underlying ARIMA realisation. Without this constraint the model with noise wouldn't be identifiable from the autocorrelation function. Wong et al. [2002] fit the ARIMAN model using maximum likelihood in a state space representation.

Based on extensions of repeated measures models, Diggle et al. [2002] look at the consequences of autocorrelation amongst the errors of these processes but limited their analysis to AR(1) models which are fitted using least squares. Shi and Chaganty [2004], in a similar context, compare maximum likelihood, Yule-Walker and quasi-least squares estimates for autoregressive models for errors within regression models. Browne and Zhang [2007] discuss fitting modified univariate AR models of order p to observations taken over time on a number of individuals. An independent error term for each observation on each individual is added to the standard AR model which also incorporates an initialisation of the process for the first p observations using a so-called initial state vector. If specified *a priori*, this effectively transforms the system into a conditional model. A comparison of the maximum likelihood fit of conditional and unconditional models in the context of differences between individuals indicates that the latter is a superior result.

Peiris et al. [2003] working with short time series (of, say, medical observations on individuals) discuss the maximum likelihood fits to a replicated AR(1) process, each realisation being of equal length. They derive the conditional and exact likelihood function for the AR(1) model, provide the asymptotic variance matrix of the estimates and run a simulation study to assess bias and efficiency. They conclude that the conditional and exact maximum likelihood estimates are unbiased and efficient.

[1]ARMA processes with added noise are a particular form of aggregated ARMA process as discussed by Box and Jenkins [1970], Anderson [1975] and Granger and Morris [1976]. See also Ansley et al. [1976], Engel [1984] and Granger [1988]. The sum of ARMA processes are themselves ARMA processes of higher but finite order which is proved in an elegant manner by Anderson [1975] (for MA processes) and Granger and Morris [1976] (for ARMA processes). This result is used by Anderson [1976] to show how ARMA processes of higher order can eventuate when it may be difficult to find a clear generating mechanism using, say, economic theory.

Quinn [2006] discusses the maximum likelihood estimation of common AR models in the context of testing for spectrum change. Quinn's approach allows the innovations variance to vary between the series.

For additional references on replicated time series, see also Ledolter and Chang-Soo [1993], Nandram and Petruccelli [1997] and Cipra [1999] who adopt state space and Bayesian approaches to fitting time series models to replicated series.

Camacho et al. [1987a] and Camacho et al. [1987b] explore a form of parallel univariate time series model, the so-called Contemporaneous ARMA model, where the ARMA coefficients for each series are unique to the series but the contemporaneous error terms are correlated. This is in comparison to the Replicated ARMA models defined in Section 2.1 where the series are independent but share their ARMA coefficients and innovations variances. The likelihood function of the CARMA process is derived and the asymptotic distribution of the maximum likelihood estimators is obtained. The models are applied to hydrological data.

A natural area for the use of Joint ARMA models is the modelling of panel or longitudinal data where data is collected on individuals or groups over time. Cameron and Trivedi [2005] provide an overview of econometrics in general and its application to panel data and time series.

There are two main uses of time series in panel data, ARMA models for error terms and so-called dynamic models. The latter, over and above the use of ARMA error models, applies what is effectively an autoregressive model to the observed dependent variable. This could be said to multiplicatively add a polynomial in B to the error term's autoregressive polynomial. The standard econometric approach to dynamic model fitting is the use of the Arellano-Bond estimator (see Cameron and Trivedi [2005] p. 765) which is a generalised method of moments approach based on first differences.

Baltagi [2005] provide a more targeted overview of panel data analysis in general and of time series in particular (p. 84-104). Estimation using generalised least squares is presented for various AR and MA models of the error term in panel data models. The estimation routines are adjusted to take advantage of the particular form of the autocovariance matrix. Quasi, approximate and limited information maximum likelihood estimation methods are addressed in MaCurdy [1982], Chen [2006] and Moral-Benito et al. [2017] respectively.

2.2 Multivariate Joint ARMA Process

The univariate JARMA process definition can be extended to the multivariate case. Let the m stationary and invertible independent multivariate time series, $\{\mathbf{y}_{i,t}\}_{t=1}^{n_i}$, $i = 1, \ldots, m$, of dimension T with linear mean vector, $\boldsymbol{\beta}^\top \mathbf{z}_{i,t}$, be generated by

$$\boldsymbol{\phi}(B)(\mathbf{y}_{i,t} - \boldsymbol{\beta}^\top \mathbf{z}_{i,t}) \;=\; \boldsymbol{\theta}(B)\boldsymbol{\epsilon}_{i,t}, \tag{2.2.1}$$

where the $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are matrix polynomials in B of order p and q, that is, $\boldsymbol{\phi}(B) = \boldsymbol{I} + \boldsymbol{\phi}_1 B + \ldots + \boldsymbol{\phi}_p B^p$ and $\boldsymbol{\theta}(B) = \boldsymbol{I} + \boldsymbol{\theta}_1 B + \ldots + \boldsymbol{\theta}_q B^q$. The process, $\{\boldsymbol{\epsilon}_{i,t}\}_{t=1}^{n}$, is the i^{th} independent T-dimensional identically-distributed multivariate time series with $E(\boldsymbol{\epsilon}_{i,t}) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{\epsilon}_{i,t}) = \sum_{\boldsymbol{\epsilon}_i}$ and $E(\boldsymbol{\epsilon}_{i,t}\boldsymbol{\epsilon}_{i,t-\kappa}) = \mathbf{0}$ for all $\kappa \neq 0$. The vector, $\boldsymbol{\beta}^\top \mathbf{z}_{i,t}$, is defined with $\boldsymbol{\beta}$ as a $k \times T$ matrix of regression coefficients and $\mathbf{z}_{i,t}$ as the vector from (2.1.1). Only the Replicated VARMA (RVARMA) process with a constant error variance matrix, $\sum_{\boldsymbol{\epsilon}_i} = \sum_{\boldsymbol{\epsilon}}$ for all i (that is, a RVARMA(p,q,m) process), is modeled in this thesis, being a natural generalisation of the univariate RARMA process.

As an example of a RVARMA process, consider two ($m = 2$) independent realisations of the same bivariate VAR(1) (vector autoregressive of order one) time series process each of length 100 ($T = 2$ and $n_1 = n_2 = n = 100$) with common autoregressive coefficient matrix, $\boldsymbol{\phi}_1 = \begin{bmatrix} 0.1 & 0.4 \\ 0.3 & 0.2 \end{bmatrix}$, mean vectors, $\boldsymbol{\mu}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$, and common innovations variance matrix, $\sum_{\epsilon} = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 1.2 \end{bmatrix}$.

Hence the RVARMA generating model for the two independent identically distributed bivariate series, $\left\{\boldsymbol{y}_{i,t}\right\}_{t=1}^{100}, i = 1, 2$, (with $\boldsymbol{y}_{i,t} = (y_{1,i,t,}, y_{2,i,t,})^\top$) is,

$$\left(\boldsymbol{I} + \begin{bmatrix} 0.1 & 0.4 \\ 0.3 & 0.2 \end{bmatrix} B\right)\left(\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} - \boldsymbol{\beta}^\top \mathbf{z}_{i,t}\right) \;=\; \begin{bmatrix} \epsilon_{1,i,t} \\ \epsilon_{2,i,t} \end{bmatrix},$$

where $\boldsymbol{\beta} = \begin{bmatrix} 1 & 2 \\ 4 & 6 \end{bmatrix}$ and $\sum_{\epsilon} = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 1.2 \end{bmatrix}$ with $\boldsymbol{z}_{1,t} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{z}_{2,t} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ for all t.

The next section reviews the (rather scarce) literature on multivariate joint ARMA models.

2.2.1 Literature Review for Joint Multivariate Time Series

Anderson [1978] considered the case of first order vector AR models where the available time series are short in length but have multiple realisations. In that paper, maximum likelihood parameter estimates are derived where the parameters are either constant over time or allowed to vary by time interval and by process (i.e. treatment). The latter makes use of the multiple observations on each time interval.

Browne and Nesselroade [2005] discuss the modelling of replicated VARMA processes from a psychometric perspective. The authors initially use a generalised mean and covariance matrix for the moments of the time series vector. This is then reduced to a restricted set of parameters using a VARMA model and a result from du Toit and Browne [2007]. In the latter the covariance matrix of the vector time series is represented as a closed-form function of the autoregressive matrices, of the variance of the initial (unobserved) system state and of the covariance matrix of the one-step ahead forecasting errors. This then facilitates estimation of the model parameters using maximum likelihood. The mean of the process is modelled using Gompertz trend curves.

Feder [2001] modelled repeated multivariate time series data from sample surveys. The author employs conventional state space models and incorporates sample survey error (see also Beck [2001] and Abraham and Vijayan [1992]).

2.3 The Interleaved Representation of a Replicated Time Series

This section will present theorems setting out the representation of replicated independent univariate and multivariate series by one interleaved (V)ARMA series with constraints on the (V)ARMA coefficients. It provides a somewhat counter intuitive but equivalent representation of several independent (stationary and invertible) time series as one (stationary and invertible) series. Significantly this

permits the use of standard time series routines to find simultaneous maximum likelihood parameter estimates. It also leads to a simple spectral interpretation of a RARMA (and RVARMA) process (see Section 2.4).

2.3.1 The Interleaved Representation of a RARMA Process

The following Univariate Interleaving Theorem and proof are from Bowden and Clarke [2012] (see Chapter 6).

Theorem 1. Let $\{y_{i,t}\}_{t=1}^{n}$, $i = 1, \ldots, m$, be a RARMA(p,q,m) process from Section 2.1, and let,

$$\begin{aligned} x_{m(t-1)+i} &= y_{i,t} \\ \mathbf{w}_{m(t-1)+i} &= \mathbf{z}_{i,t} \text{ and} \\ e_{m(t-1)+i} &= \epsilon_{i,t} \, . \end{aligned}$$

Then,
$$\phi(B^m)(x_s - \boldsymbol{\beta}^\top \mathbf{w}_s) = \theta(B^m)e_s, \tag{2.3.1}$$

where $E(e_s) = 0$, $\mathrm{Var}(e_s) = \sigma_\epsilon^2$ and $E(e_s e_r) = 0$, $s \neq r$. That is, the interleaved series, $\{x_s\}_{s=1}^{mn}$, is an ARMA process of order (mp, mq).

Proof. Given $\{x_s\}$, $\{\boldsymbol{w}_s\}$ and $\{e_s\}$ from (2.3.1) and for a constant i from $(1, ..., m)$, let $s = m(t-1)+i$, $t = 1, ..., n$,. The difference equation (2.3.1) can then be expressed as,

$$\phi(B)(y_{i,t} - \boldsymbol{\beta}^\top \mathbf{z}_{i,t}) = \theta(B)\epsilon_{i,t} \, ,$$

where $t = 1, ..., n$. It is also known that, for $s = i$, $m + i$, $2m + i, ...$, $E(e_s) = E(\epsilon_{i,t}) = 0$, and $\mathrm{Var}(e_s) = \mathrm{Var}(\epsilon_{i,t}) = \sigma_\epsilon^2$. Also, for $r = i$, $m + i$, $2m + i, ...$, where $r \neq s$, then $E(e_s e_r) = E(\epsilon_{i,t}\epsilon_{i,u}) = 0$, $t \neq u$. Finally , for $s = i$, $m+i$, $2m+i, ...$ and $r^* = j$, $m + j$, $2m + j, ...$,[2] and $(s \mid m) \neq (r^* \mid m)$,[3] then $E(e_s e_{r^*}) = E(\epsilon_{i,t}\epsilon_{j,u}) = 0$, $i \neq j$, that is, that the replicated univariate time series are independent, Hence the interleaved process (2.3.1) leads to the RARMA process.

The above formulation permits m independent ARMA(p,q) processes, each

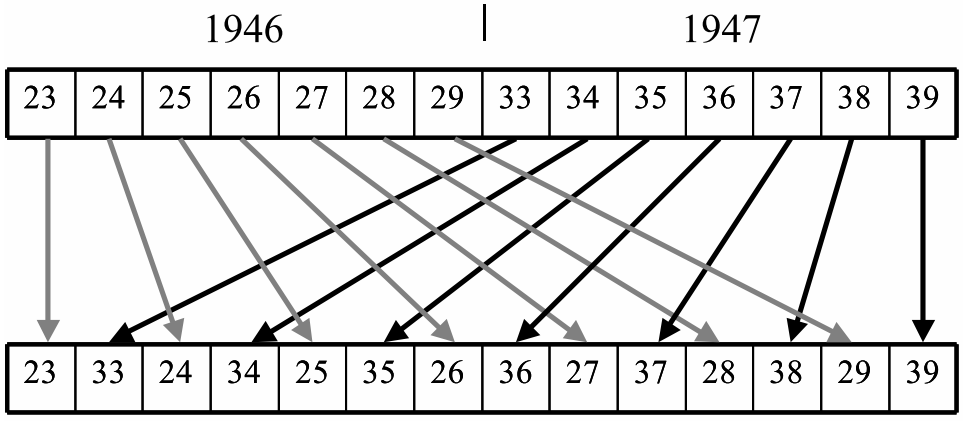
[2] When used in a subscript, j is an integer. Otherwise $j = \sqrt{-1}$.

[3] "|" is the modulus operator.

of length n, with identical parameters to be modelled as a single (univariate) ARMA(pm, qm) process of length, nm. Given that the component series are assumed to be stationary and invertible, so too will be the interleaved series, that is, if the roots of $\phi(B) = 0$ and of $\theta(B) = 0$ all lie outside the unit circle, then so too will those of $\phi(B^m) = 0$ and of $\theta(B^m) = 0$.

The interleaving used in Theorem 1 is illustrated in Figure 2.3.1. In this example, it is assumed that there are, say, seven daily maximum temperature values of sequential integers available for the first week in the year for two years, 1946 and 1947. The two series are interleaved to create a final single series of length fourteen.

Figure 2.3.1 – Univariate Interleaving of Two Artificial Bivariate Series ("1946" and "1947") of Length Seven Containing Sequential Integers to Produce One Interleaved Univariate Series of Length Fourteen.



As an example of a RARMA process, let there be $m = 10$ realisations of a zero-mean ARMA(2,1) time series process with AR and MA polynomials in B of $\phi(B) = (1 + \phi_1 B + \phi_2 B^2)$ and $\theta(B) = (1 + \theta_1 B)$ and each series is of length 25. After interleaving the ten series to create a new zero-mean series of length 250, the interleaved model is of the form, ARMA(20,10), with AR and MA polynomials in B as follows,

$$\phi(B^{10}) = (1 + \phi_1 B^{10} + \phi_2 B^{20}) \text{ and } \theta(B^{10}) = (1 + \theta_1 B^{10}) .$$

The interleaving representation of a RARMA process permits the use of readily available software routines in say R, SAS or GAUSS to estimate the parameters of RARMA processes. The routine must permit the fixing of selected coefficients to zero. In the example above, the autoregressive lags of 1 to 9 and 11 to 19 are fixed at zero as are the moving average lags of 1 to 9. Also to accommodate non-uniform series lengths, $\{n_i\}_{i=1}^{m}$, the routine must be able to accept missing values, added to make all series the same length.

It is possible to combine the replicated series in ways other than interleaving such as simple concatenation to create one series for joint modelling. However Bowden and Clarke [2012] show that these alternatives result in biased or inefficient parameter estimates.

The interleaving approach is adopted in Chapter 6 to model daily maximum temperature readings for Perth, Western Australia, by week-in-the-year from 1943 to 2009.

2.3.2 The Interleaved Representation of a RVARMA Process

This section will now state and prove a multivariate extension of the Univariate Interleaving Theorem (called the Multivariate Interleaving Theorem) that reduces the apparent dimensionality of a RVARMA process with m vector series (see Section 2.2) by a factor of m.

Theorem 2. *Let the replicated T-dimensional series $\{\boldsymbol{y}_{i,t}\}_{t=1}^{n}, i = 1, ..., m$, be a RVARMA(*p,q,m*) process, and let,*

$$\begin{aligned} \boldsymbol{x}_{m(t-1)+i} &= \boldsymbol{y}_{i,t}, \\ \boldsymbol{w}_{m(t-1)+i} &= \boldsymbol{z}_{i,t} \text{ and} \\ \boldsymbol{e}_{m(t-1)+i} &= \boldsymbol{\epsilon}_{i,t}\,. \end{aligned} \tag{2.3.2}$$

Then,

$$\boldsymbol{\phi}(B^m)(\boldsymbol{x}_s - \boldsymbol{\beta}^\top \mathbf{w}_s) = \boldsymbol{\theta}(B^m)\boldsymbol{e}_s, \tag{2.3.3}$$

where $E(\boldsymbol{e}_s) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{e}_s) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ and $E(\boldsymbol{e}_s\boldsymbol{e}_r^\top) = \mathbf{0}$, $s \neq r$. That is, the interleaved

series, $\{\boldsymbol{x}_s\}_{s=1}^{mn}$*, is a* T*-dimensional VARMA process of order* (mp, mq)*.*

Proof. Given $\{\boldsymbol{x}_s\}$, $\{\boldsymbol{w}_s\}$ and $\{\boldsymbol{e}_s\}$ from (2.3.2) and for a constant i from $(1, ..., m)$, let $s = m(t-1)+i$, $t = 1, ..., n$. The difference equation (2.3.3) can then be expressed as,

$$\boldsymbol{\phi}(B)(\boldsymbol{y}_{i,t} - \boldsymbol{\beta}^\top \mathbf{z}_{i,t}) = \boldsymbol{\theta}(B)\boldsymbol{\epsilon}_{i,t}\,,$$

where $t = 1, ..., n$. It is also known that, for $s = i,\ m+i,\ 2m+i, ...,\ E(\boldsymbol{e}_s) = E(\boldsymbol{\epsilon}_{i,t}) = \mathbf{0}$, and $\mathrm{Var}(\boldsymbol{e}_s) = \mathrm{Var}(\boldsymbol{\epsilon}_{i,t}) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. Also, for $r = i,\ m+i,\ 2m+i, ...,$ where $r \neq s$, then $E(\boldsymbol{e}_s \boldsymbol{e}_r^\top) = E(\boldsymbol{\epsilon}_{i,t}\boldsymbol{\epsilon}_{i,u}^\top) = \mathbf{0}\,,\ t \neq u$. Finally, for $s = i,\ m+i,\ 2m+i, ...$ and $r^* = j,\ m+j,\ 2m+j, ...,$ and $(s \mid m) \neq (r^* \mid m)\,$, then $E(\boldsymbol{e}_s \boldsymbol{e}_{r^*}^\top) = E(\boldsymbol{\epsilon}_{i,t}\boldsymbol{\epsilon}_{j,u}^\top) = \mathbf{0}$, $i \neq j$, that is, that the replicated multivariate time series are independent, Hence the interleaved process (2.3.3) leads to the RVARMA process.

To paraphrase Theorem 2, any m replicated independent T-dimensional VARMA(p, q) time series, each of length n, can be represented by one T-dimensional VARMA(mp, mq) process of length mn. This equivalence is achieved by interleaving the m series and by ensuring that AR and MA parameters are only non-zero at orders that are multiples of m. The interleaving is illustrated in Figure 2.3.2 for two artificial bivariate series (being sequential integers), each of length seven.

As in the univariate case, given that the component series are assumed to be stationary and invertible, so too will be the interleaved series. That is, if the roots of $\det(\boldsymbol{\phi}(B)) = 0$ and of $\det(\boldsymbol{\theta}(B)) = 0$ all lie outside the unit circle, then so too will those of $\det(\boldsymbol{\phi}(B^m)) = 0$ and of $\det(\boldsymbol{\theta}(B^m)) = 0$.

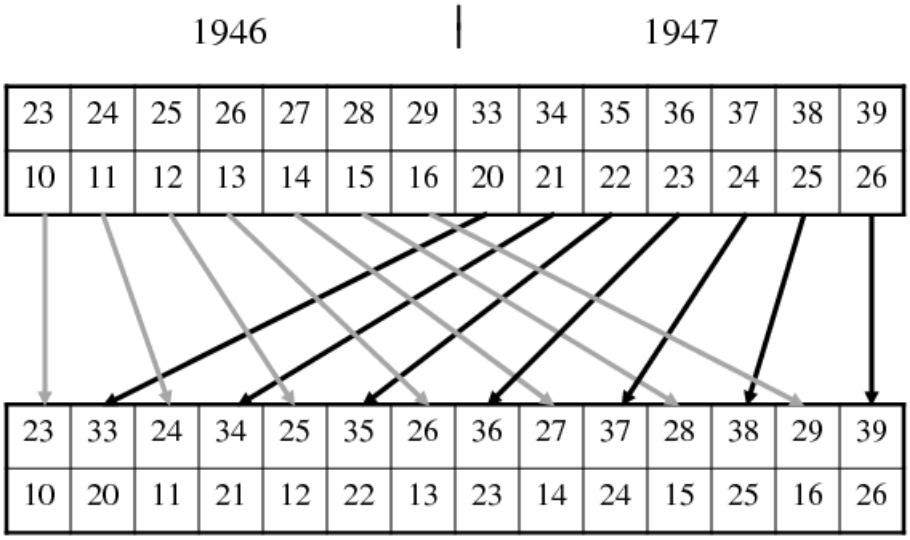
As an extension of the univariate example from Section 2.3.1, let there be $m = 10$ realisations of a bivariate VARMA(2,1) time series process, each of length 25. After interleaving the series, a single bivariate series of length 250 is created and the interleaved model to be fitted is of the form, VARMA(20,10), with,

$$\boldsymbol{\phi}(B^{10}) = (1 + \boldsymbol{\phi}_1 B^{10} + \boldsymbol{\phi}_2 B^{20}) \text{ and } \boldsymbol{\theta}(B^{10}) \ = (1 + \boldsymbol{\theta}_1 B^{10})\,.$$

The interleaving (to create one multivariate series) and the associated constraints on the VARMA model allows the use of existing software to estimate the parameters of RVARMA models. In R, a RVARMA model can be fitted via maximum likelihood to the interleaved VARMA series using the package, `dse` (Gilbert [2006]), which allows for the relevant AR and MA matrices to be fixed to zero. The

RVARMA estimation is discussed further in Chapter 7 and follows a similar process to the RARMA interleaved estimation described in Section 4.2.

Figure 2.3.2 – Multivariate Interleaving of Two Artificial Bivariate Series ("1946" and "1947") of Length Seven Containing Sequential Integers to Produce One Interleaved Bivariate Series of Length Fourteen.



2.4 Properties of an Interleaved Replicated Time Series Process

This section will explore the time domain and spectral properties of interleaved replicated stationary and invertible univariate and multivariate time series processes (that is, RARMA and RVARMA processes) to better understand the effects of interleaving. The influence of interleaving on the periodogram and of compressing the spectrum of a continuous process have also been derived but, to limit the length of this thesis, will be explored in a future paper.

As defined in Section 2.3, an interleaved process is created by the merging of two or more univariate or multivariate series through sequentially interspersing the scalar or vector elements of the respective series.

2.4.1 Spectral Theory

To recall some spectral theory (see Priestley [1981] p. 225) in the context of a set of replicated time series, if $\{y_{i,t}\}$[4] is the i^{th} independent replicated real zero-mean (without loss of generality) stationary and invertible time series process ($i = 1, 2, \ldots, m$), each with common autocovariance function, $\{R_y(r) = E(y_{i,t}y_{i,t-r}), r = 0, 1, 2, \ldots\}$, then the common non-normalised spectral density function (for each series) is,

$$h_y(\omega) = \frac{1}{2\pi}\sum_{r=-\infty}^{\infty} R_y(r)e^{-jr\omega},\ -\pi \leq \omega \leq \pi \tag{2.4.1}$$

$$= \frac{\sigma_y^2}{2\pi} + \frac{1}{\pi}\sum_{r=1}^{\infty} R_y(r)\cos r\omega,\ -\pi \leq \omega \leq \pi, \tag{2.4.2}$$

where $j = \sqrt{-1}$ and σ_y^2 is the variance of $\{y_{i,t}\}$.

Given $h_y(\omega)$ and $R_y(r)$ are real (because $\{y_{i,t}\}$ is real), then

$$R_y(r) = \int_{-\pi}^{\pi} e^{j\omega r}h_y(\omega)d\omega,\ r \in \mathbb{N} \tag{2.4.3}$$

$$= 2\int_{0}^{\pi} \cos(\omega r)\, h_y(\omega)d\omega,\ r \in \mathbb{N} \tag{2.4.4}$$

In the multivariate case, let $\{\mathbf{y}_{i,t},\ t = 0, \pm1, \pm2, \ldots\}$ be the i^{th} replicated (and independent) stationary and invertible real vector time series process ($i = 1, 2, \ldots, m$) of dimension, T, that is, $\mathbf{y}_{i,t} = (y_{1,i,t}, \ldots, y_{T,i,t})^\top$, each with autocovariance matrix, $\mathbf{R}_{\mathrm{y}}(s),\ s = 0, \pm1, \pm2, \ldots$, then its common (non-normalised) spectral density function is (see Priestley [1981] p. 667),

$$\mathbf{h}_y(w) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} \mathbf{R}_{\mathrm{y}}(s)e^{-js\omega}. \tag{2.4.5}$$

[4]See the end of Chapter 1 for the explanation of this series representation.

2.4.2 Compressed Spectral Density from an Interleaved Process

Here the effect of interleaving on the univariate and multivariate autocovariance and spectral density functions is derived.

Univariate Interleaving. The interleaved univariate stationary and invertible discrete parameter time series, $\{x_s\}$, is derived from the m stationary and invertible time series, $\{y_{i,t}\}$, $i = 1, 2, \ldots, m$, as follows,

$$x_{(m-1)t+i} = y_{i,t},\ t = 0, \pm 1, \pm 2, \ldots \text{ and } i = 1, \ldots, m.$$

Given that each of the series, $\{y_{i,t}\}$, are independent of each other, then the autocovariance function of $\{x_s\}$ is ,

$$R_x(r) = \begin{cases} R_y(\frac{r}{m}) & r|m = 0 \\ 0 & r|m \neq 0 \end{cases} \tag{2.4.6}$$

and $r \in \mathbb{N}$. In this thesis, "critical lags" are defined as the lags which are multiples of the number of interleaved series. Hence $R_x(r)$ will be zero at all non-critical lags.

Now using the formula (2.4.2) for the spectral density of $\{x_s\}$, the following is derived,

$$\begin{aligned} h_x(\omega) &= \frac{\sigma_x^2}{2\pi} + \frac{1}{\pi}\sum_{r=1}^{\infty} R_x(r)\cos r\omega,\ -\pi \leq \omega \leq \pi \\ &= \frac{\sigma_x^2}{2\pi} + \frac{1}{\pi}\sum_{r=m\,by\,m}^{\infty} R_x(r)\cos r\omega,\ -\pi \leq \omega \leq \pi \\ &= \frac{\sigma_y^2}{2\pi} + \frac{1}{\pi}\sum_{s=1}^{\infty} R_y(s)\cos ms\omega,\ -\pi \leq \omega \leq \pi \\ &= h_y(m\omega), \end{aligned}$$

where "$r = m\,by\,m$" reads as "r from m in multiples of m", $\sigma_x^2 = \sigma_y^2$ and $h_x(\omega) = h_x(\omega + {}^{2\pi}/_{m})$ because $h_y(\omega) = h_y(\omega + 2\pi)$.

Hence the spectral density of the interleaved process is of the same form as for the component series but with the frequencies "compressed" by a factor, m. The new spectral density is periodic with period, $2\pi/m$. This result can be seen

heuristically by noting that the length of each of the interleaved series is effectively extended by a factor of m. Hence the periods of the associated spectral density are increased by the same factor and the frequencies are reduced (or compressed) by a factor of $1/m$.

For an MA(1) process with $\theta_1 = 0.5$ and $\sigma_\epsilon^2 = 1$ the spectral density is the solid line in Figure 2.4.1 (that is, $h_y(\omega) = (2\pi)^{-1}(1.25 + \cos\omega)$). Of course, because the series is real, the spectrum is symmetric around zero. If, say $m = 2$, that is, there are two replicated interleaved MA(1) series, the spectral density is given by the dashed line (that is, $h_x(\omega) = (2\pi)^{-1}(1.25 + \cos 2\omega)$). If the series is complex-valued, the only change is that the initial spectral density is not symmetric around zero. So, interleaving creates the phenomenon of a compressed spectral density, periodic with period, $\frac{2\pi}{2} = \pi$.

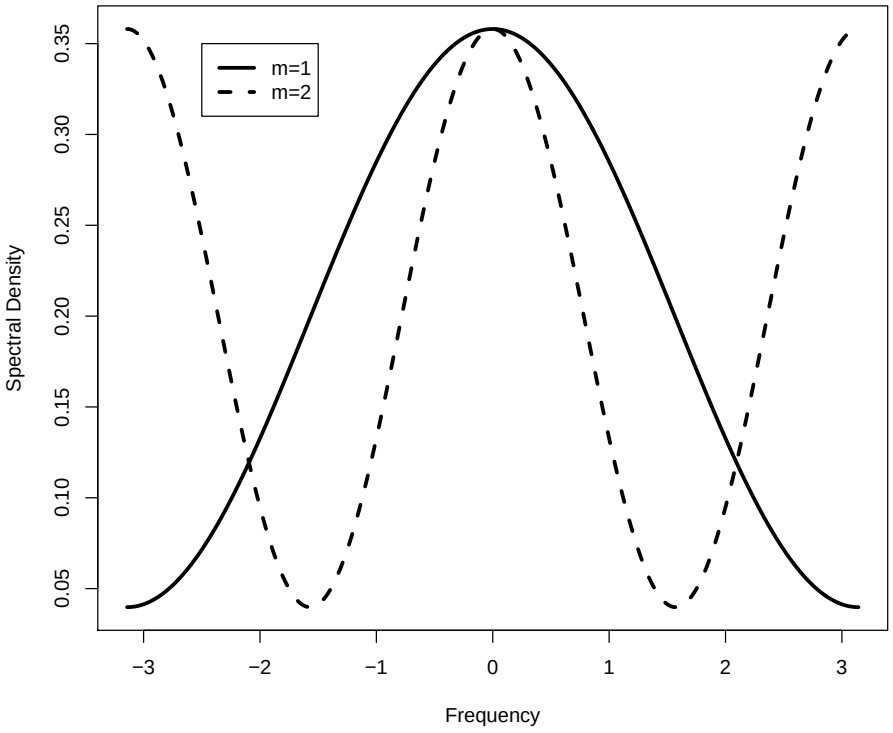
To explore the reverse relationship, that is, the time series that is associated with a compressed spectral density, the autocovariance function from a compressed spectral density is first derived. The compressed spectral density is the spectrum of $\{x_s\}$ defined from the spectrum of $\{y_t\}$ as follows,

$$h_x(\omega) \;=\; h_y(m\omega),\ -\pi \le \omega \le \pi, \tag{2.4.7}$$

where m is a positive integer.

The formula for deriving the autocovariance function from the spectral density (that is, (2.4.4)) is,

Figure 2.4.1 – Spectral Density for an Interleaved First Order Real Moving Average Process ($\theta_1 = 0.5$). The solid line is the original spectral density ($m = 1$) and the dashed line is after interleaving two replicated series ($m = 2$).



$$
\begin{aligned}
R_x(r) &= 2\int_0^{\pi} \cos(\omega r)\, h_x(\omega) d\omega,\ r \in \mathbb{N} \\
&= 2\int_0^{\pi} \cos(\omega r)\, h_y(m\omega) d\omega,\ r \in \mathbb{N} \\
&= 2\int_0^{\pi} \cos(\omega r) \left\{ \frac{\sigma_y^2}{2\pi} + \frac{1}{\pi}\sum_{s=1}^{\infty} R_y(s)\cos(ms\omega) \right\} d\omega,\ r \in \mathbb{N} \\
&= 2\int_0^{\pi} \cos(\omega r)\, \frac{\sigma_y^2}{2\pi} d\omega \\
&\quad + \frac{2}{\pi}\int_0^{\pi} \cos(\omega r) \sum_{s=1}^{\infty} R_y(s)\cos(ms\omega) d\omega,\ r \in \mathbb{N} \\
&= \frac{\sigma_y^2}{\pi}\int_0^{\pi} \cos(\omega r)\, d\omega \\
&\quad + \frac{2}{\pi}\sum_{s=1}^{\infty} R_y(s) \int_0^{\pi} \cos(\omega r)\cos(ms\omega) d\omega,\ r \in \mathbb{N}
\end{aligned}
$$

Now, on the right hand side of the equation, the first integral is zero and the second integral is only non-zero (that is, it only equals $\frac{\pi}{2}$) when $r = ms$ (remembering that r, m and s are all integers). Hence $R_x(r)$ is also only non-zero (that is, it equals $R_y(\frac{r}{m})$) when r is a multiple of m (that is, when r is a critical lag) and therefore (2.4.6) is the autocovariance of the interleaved process. Accordingly if (2.4.7) is assumed then the sparse autocovariance function (2.4.6) is the result.

Hence a compressed periodic spectral density with period $2\pi/m$ has an autocovariance function that has zero values at all non-critical lags (that is, that are not multiples of m).

Multivariate Interleaving As in the univariate case, the m stationary and invertible (vector) time series, $\{\boldsymbol{y}_{i,t}\}$, $i = 1, 2, \ldots, m$, are interleaved to create the stationary and invertible (vector) series $\{\mathbf{x}_s\}$ using,

$$\mathbf{x}_{(m-1)t+i} = \mathbf{y}_{i,t},\ t = 0, \pm 1, \pm 2, \ldots,\ i = 1, \ldots, m,$$

where $\mathbf{x}_s = (x_{1,s}, \ldots, x_{T,s})^\top$. Immediately it is noted that,

$$\mathbf{R}_{\mathbf{x}}(s) = \begin{cases} \mathbf{R}_{\mathbf{y}}(\frac{r}{m}) & r|m = 0 \\ 0 & r|m \neq 0. \end{cases}$$

Hence, from (2.4.5),

$$\begin{aligned} \mathbf{h}_{\mathbf{x}}(w) &= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \mathbf{R}_{\mathbf{y}}(s) e^{-jms\omega} \\ &= \mathbf{h}_{\mathbf{y}}(mw). \end{aligned} \tag{2.4.8}$$

This is a compressed version of the spectral density of each of the time series, $\{\mathbf{y}_{i,t}\}$. The marginal spectral densities of each of the scalar time series components,$\{x_{j,s}\}$, of $\{\boldsymbol{x}_s\}$ are the same as shown in (2.4.7) and the cross-spectral density between $\{x_{i,s}\}$ and $\{x_{j,s}\}$ is (see Priestley [1981] p. 657),

$$\begin{aligned} h_{x_i x_j}(\omega) &= \frac{\sigma_{x_i x_j}}{2\pi} + \frac{1}{\pi}\sum_{r=1}^{\infty} R_{x_i x_j}(r)\cos r\omega,\ -\pi \leq \omega \leq \pi \\ &= \frac{\sigma_{y_i y_j}}{2\pi} + \frac{1}{\pi}\sum_{s=1}^{\infty} R_{y_i y_j}(s)\cos ms\omega,\ -\pi \leq \omega \leq \pi \\ &= h_{y_i y_j}(m\omega), \end{aligned}$$

where $R_{uv}(s)$ is the cross-covariance between $\{u_t\}$ and $\{v_t\}$ at lag s and σ_{uv} is the covariance (at lag zero) between $\{u_t\}$ and $\{v_t\}$. In other words, the cross-spectral elements of the interleaved multivariate time series are the compressed cross-spectral elements of the original multivariate series.

2.5 Joint ARMA Processes with Unique Coefficients and Innovations Variances

Let $\{y_{i,t}\}_{t=1}^{n_i}, i = 1, \ldots, m$, be a set of m independent stationary and invertible univariate time series, each with *unique* ARMA orders and coefficients, unique levels of the mean, $\{\mu_i\}_{i=1}^{m}$, and unique innovations variances, $\left\{\sigma_{\epsilon_i}^2\right\}_{i=1}^{m}$ where

$$\phi_i(B)(y_{i,t} - \mu_i) = \theta_i(B)\epsilon_{i,t} \tag{2.5.1}$$

with $E(\epsilon_{i,t}) = 0$, $Var(\epsilon_{i,t}) = \sigma_{\epsilon_i}^2$, $E(\epsilon_{i,t}\epsilon_{i,t-\kappa}) = 0$ for all $\kappa \neq 0$, $E(y_{i,t}) = \mu_i$ and $E(\epsilon_{i,t}\epsilon_{j,t-\kappa}) = 0$ for all $i \neq j$, and all κ. This is a generalisation of the Joint RARMA process. An alternative specification for these processes is now derived.

Using interleaving, let

$$x_{(i-1)m+t} = y_{i,t},\ i = 1, \ldots, m$$

and

$$\varepsilon_{(i-1)m+t} = \epsilon_{i,t},\ i = 1, \ldots, m.$$

Therefore

$$\phi_{((s-1)|m)+1}(B^m)(x_s - \eta_{((s-1)|m)+1}) = \theta_{((s-1)|m)+1}(B^m)\varepsilon_s,$$

where $\{\varepsilon_s\}$ is independent identically distributed with $\text{Var}(\varepsilon_s) = \sigma^2_{\epsilon_{((s-1)|m)+1}}$ and $\eta_{((s-1)|m)+1} = E(x_s) = \mu_{((s-1)|m)+1}$. The series, $\{x_s\}$, is a periodic ARMA (PARMA) process (see Parzen and Pagano [1979]) of period m with non-critical ARMA parameters set to zero.

Hence m independent ARMA processes, each with unique ARMA coefficients and innovations variances, can be represented after interleaving by a single PARMA process of period m with coefficient constraints. The proof of this follows directly from Theorem 1. This model has been presented for completeness but will not be pursued further in this thesis.

Chapter 3, to follow, will allow for RARMA, AIARMA and CAIARMA process identification by reviewing hypothesis tests of process equivalence from the literature, presenting some new tests and exploring their size and power using simulation.

In all the hypothesis testing and estimation in the following chapters it is assumed that the innovations are normally distributed, a common assumption in time series modelling.

Chapter 3

Empirical Identification

The purpose of this chapter is to explore hypothesis tests that could be used to determine whether two stationary and invertible time series of equal length may be represented by the Joint ARMA (univariate) models from Chapter 2. The current chapter reviews the literature on testing whether two time series are generated by the same process and presents several new tests. It introduces a graphical approach to testing and identifying the underlying processes.

The testing can involve a number of null and alternative hypotheses. These include whether the two processes have the same spectral density shape and, if so, whether they have the same spectral density. This is equivalent to deciding whether the stationary and invertible series have generating mechanisms with a common set of ARMA coefficients and, if so, whether they also share a common innovations variance.

Seven new tests are proposed which either may improve on an existing test or are arguably simpler than existing tests but with comparable power. The analysis will focus on comparing two series of equal length but it is possible to extend this to several series of unequal length (see Chapter 8).

3.1 Background

It is assumed that the first time series, $\{y_{1,t}\}_{t=1}^{n}$, has been been generated by a stationary and invertible ARMA model of the form, $\phi_1(B)y_{1,t} = \theta_1(B)\epsilon_{1,t}$ with

spectral density,

$$h_1(\omega) = \frac{|\theta_1(e^{-j\omega})|^2}{|\phi_1(e^{-j\omega})|^2}\sigma^2_{\epsilon_1},$$

where $\omega \in [-\pi, \pi]$ is the frequency and $\sigma^2_{\varepsilon_1}$ is the variance of the normal identically-distributed independent innovations series, $\{\epsilon_{1,t}\}_{t=1}^n$. Similarly, $\{y_{2,t}\}_{t=1}^n$ is an independent stationary and invertible time series with normal identically-distributed independent innovations, $\{\epsilon_{2,t}\}_{t=1}^n$, innovations variance, $\sigma^2_{\epsilon_2}$, generating process, $\phi_2(B)y_{2,t} = \theta_2(B)\epsilon_{2,t}$ and spectral density,

$$h_2(\omega) = \frac{|\theta_2(e^{-j\omega})|^2}{|\phi_2(e^{-j\omega})|^2}\sigma^2_{\epsilon_2}.$$

The periodogram ordinates of $\{y_{k,t}\}_{t=1}^n$, $k = 1, 2$, are $\left\{\hat{h}_k(\omega_i)\right\}_{i=1}^N$ where n is assumed to be even (for convenience), $N = \frac{n}{2}$ and $\omega_i = \pi i/N$ (see Priestley [1981] p. 395). The ratio of the two spectral densities at each frequency, ω_i, $i = 1, \ldots, N$, is

$$\alpha_i = \frac{h_2(\omega_i)}{h_1(\omega_i)}$$

and

$$\hat{\alpha}_i = \frac{\hat{h}_2(\omega_i)}{\hat{h}_1(\omega_i)}$$

is an estimate of each α_i using the ratio of the periodogram ordinates for the series $\{y_{1,t}\}_{t=1}^n$ and $\{y_{2,t}\}_{t=1}^n$. In the following the ratio can be inverted without loss of generality. A review of distributional theory for periodograms and ratios of periodogram ordinates is presented in Appendix B.

The literature on hypothesis tests for comparing the generating processes for two (or, in some cases, more) time series is reviewed in Section 3.2, followed by a summary of the published size and power studies. The current chapter then continues with the introduction of several new tests (Section 3.3), presents a simulation-based size and power analysis of the new and already published tests (Section 3.4) and finally suggests a graphical approach to identification of AIARMA processes (Section 3.5).

3.2 Tests of Process Equivalence from the Literature

This section reviews the main tests from the literature and compares their size and power from published simulation studies.

For the two independent stationary and invertible series, $\{y_{1,t}\}_{t=1}^{n}$ and $\{y_{2,t}\}_{t=1}^{n}$, the five hypotheses that are typically present in the literature are defined below,

1. $H_{B_1}: \alpha_i = 1$ for all i (The two series have identical generating processes),
2. $H_{B_2}: \alpha_i = \alpha \neq 1$ for all i (The two series have the same spectral shape but different innovations variances),
3. $H_B: \alpha_i = \alpha$ for all i (The two series have the same spectral shape),
4. $H_C: \alpha_i \neq \alpha$ for some i (The two series have different spectral shapes (versus H_B)) and
5. $H_D: \alpha_i \neq 1$ for some i (The two series have different spectral densities (versus H_{B_1})).

The null versus alternative hypotheses that are most commonly tested in the literature are:

1. H_{B_1} v H_{B_2}, that is, the series have the same spectral densities versus the same spectral shapes but different variances,
2. H_B v H_C, that is, the series have the same versus different spectral shapes, and
3. H_{B_1} v H_D, that is, the series have the same versus different spectral densities.

The same spectral shape implies $\alpha_i = \alpha$ for all i (that is, the ratio of the spectral densities is constant across all frequencies) because,

$$\alpha_i = \frac{h_2(\omega_i)}{h_1(\omega_i)} = \frac{|\theta(-j\omega_i)|^2/|\phi(-j\omega_i)|^2\sigma^2_{\varepsilon_2}}{|\theta(-j\omega_i)|^2/|\phi(-j\omega_i)|^2\sigma^2_{\epsilon_1}} = \frac{\sigma^2_{\varepsilon_2}}{\sigma^2_{\varepsilon_1}} = \alpha. \tag{3.2.1}$$

Further to what hypotheses are involved, the tests in the literature can be categorised according to whether the periodogram, autocovariance function or other

statistic is involved. The periodogram tests can be subdivided into whether they are regression-based, use a single-parameter likelihood ratio test or a distance-based measure. Furthermore some tests can readily use either the non-normalised or normalised periodogram or the sample autocovariance or autocorrelation function. Effectively these address the scale or shape of the spectral density.

Priestley [1981] p. 479 summarises a number of tests for single spectral densities including those that test for white noise. An important test here which is used in the process comparison literature (see Coates and Diggle [1986]) employs the fact that, for a normally distributed time series under the white noise assumption, the periodogram ordinates are exponentially distributed and the cumulative periodogram ordinates divided by the sum of the periodogram ordinates are distributed as an ordered uniform random variable on $[0, 1]$. This permits the use of the Kolmogorov-Smirnov statistic to test for white noise, as the ordinates of the re-scaled cumulative periodogram behave as if they were an empirical distribution function.

A table summarising every test identified in the literature is included as Appendix A (see also Grant [2015], Chapter 2) along with a table detailing the design of the associated simulation studies.

3.2.1 Regression-Based Likelihood Ratio Tests using Periodograms

The tests in this group all employ some form of regression model explaining the logarithms of the ratio of the non-normalized periodogram ordinates in terms of the associated frequencies (versus an hypothesis of a common α for all frequencies). The models are fitted using maximum likelihood and then employ the likelihood ratio test.

Coates and Diggle [1986] suggest a quadratic function which is of the form $\lambda_1 + \lambda_2\omega_i + \lambda_3\omega_i^2$ and they test $\lambda_1 = \lambda_2 = \lambda_3 = 0$ versus $\lambda_1 \neq 0,\ \lambda_2 \neq 0\ , \lambda_3 \neq 0$, and $\lambda_1 = \lambda_2 = \lambda_3 = 0$ versus $\lambda_1 \neq 0,\ \lambda_2 \neq 0,\ \lambda_3 = 0$. These correspond to testing H_{B_1} v H_C. Jin [2011], Vassiliadis and Rigas [2009], Fokianos and Savvides [2008] and Lu and Li [2013] use Lagrangian polynomials, generalised linear models, cosine terms and Fourier models respectively, all with similar constraints and tests

to Coates and Diggle [1986].

3.2.2 Single-Parameter Likelihood Ratio Tests Using Periodograms

Both Lund et al. [2009] and Tugnait [2013] employ a pooled estimate of the periodogram ordinates by frequency. The distribution of the test statistic (a likelihood ratio value but distinct from that in Section 3.3.4) is obtained using the Central Limit Theorem and likelihood ratio theory respectively. Both tests examine H_{B_1} v H_D.

3.2.3 Distance-Based Tests for Periodograms and Other Periodogram Tests

In the literature there are several distance-based tests for comparing periodograms. These all involve some aggregate measure of the gap between the periodogram ordinates and all except one compare H_{B_1} v H_D. Chik [2002], Diggle and Fisher [1991], Caiado et al. [2006] and Luengo et al. [2006] employ distance measures which can be used with either non-normalised or normalised periodograms and hence allow expansion for testing of either H_{B_1} v H_D or H_B v H_C. Coates and Diggle [1986] (with two tests), Lund et al. [2009], Jentsch and Pauly [2012] (two tests), Hidalgo and Souza [2014], and Jentsch and Pauly [2015] all test only H_{B_1} v H_D. Also Coates and Diggle [1986] contain an additional test that uses only the range of the log of the ratios of the periodogram ordinates to compare H_B v H_C.

In addition to the above periodogram tests, there is also Lund et al. [2009] based on a count of the number of ordinate ratios outside certain bounds and Dette et al. [2011] which uses the sum of squares and cross-products of the ordinates. Both these test H_{B_1} v H_D .

3.2.4 Autocovariance-Based Tests

There are a number of tests based on the sample autocovariances and autocorrelations. Caiado et al. [2006], Alonso and Maharaj [2005], and Maharaj [2000] employ distances between each series' sample autocovariances and test H_{B_1} v H_D.

This approach can be modified using sample autocorrelations to also test $H_B \, v \, H_C$ (see Section 3.3.2). Lund et al. [2009] employ Bartlett's formulae for the asymptotic mean, variance/covariance structure and distribution of the sample autocovariances to derive a χ^2 test statistic of the distance between the two processes. Jin and Wang [2016] use a test statistic that is the maximum over all lags of a linear function of the sample autocorrelations (in order to ensure a consistent test).

3.2.5 Other Tests

Quinn [2006] uses a likelihood test to compare an autoregressive model of the same order fitted to each series and allows for a mixed spectral process. Tunno [2015] compare various measures of the length of the line segments connecting the points of each series (including the sum of square of the first differences) and Decowski and Li [2015] compares wavelet models fitted to each series.

Cox and Solomon [1988], Peiris and Rao [2004] and Perera et al. [2008] develop tests for the first lag autocorrelation from time series panel data and these results could be adapted to testing for differences in the autocorrelations between series.

3.2.6 Summary of Power Studies

The shortage of simulation studies and the lack of design consistency makes it difficult to conclude which of the existing tests have the highest power. Tests with relatively high power are Coates and Diggle [1986] polynomial and Jin [2011] Lagrangian regression tests. Several randomisation and bootstrapping tests (Alonso and Maharaj [2005], Maharaj [2000] and Jin and Wang [2016]) also show good power. On the other hand some tests show especially low power including Lund et al. [2009] likelihood ratio and Central Limit Theorem tests and Coates and Diggle [1986] range and extrema tests.

3.3 New Tests of Process Equivalence

This section will introduce seven new approaches for testing the equivalence of time series processes, four for H_B v H_C, the spectral "shape decision", and then, given

H_B, three for H_{B_1} v H_{B_2}, the spectral "scale decision". Also the likelihood ratio test of scale from Coates and Diggle [1986] is further explored but is based here on the untransformed ratios, not their logs. The transformation doesn't, of course, change the maximum likelihood estimate but the interpretation of the estimation algorithm is different (see Section 3.3.4).

The testing method proceeds in sequence (see Section 3.4.2) to firstly test the spectral density for shape equivalence and then, if a common shape is accepted, for scale equivalence. This spectral testing nomenclature can be said to apply even to testing (of "shape") using the sample autocorrelations (see Section 3.3.2) because of the symmetry between time- and frequency-domain models of time series.

As mentioned previously, in some circumstances there will be a need to compare several time series simultaneously to determine whether at least one has a different spectral shape to the others, and then, if they are shown to all have the same shape, whether they all have the same innovations variance. There is also the circumstance of differing series lengths. These issues are discussed in Section 8.2.

The following sections will introduce the differences-of-logs (two tests), autocorrelation and variance shape tests, explore the likelihood ratio scale test and introduce the Wald, mean log and Central Limit Theorem scale tests.

3.3.1 Goodness-of-Fit Test using the Differences of the Logarithms of the Periodogram Ratios (Shape Test)

The relatively simple differencing test for shape introduced in this section removes the dependence of the test statistic on the population value of $\log \alpha$ (under the null hypothesis, H_B) by differencing the sample ratios. This also permits the use of more powerful test statistics than the Kolmogorov-Smirnov statistic employed by Coates and Diggle [1986] (see the discussion later in this section) who remove the dependence on $\log \alpha$ by using the range of the sample ratios as the test statistic.

It is known that each $\{\log(\hat{\alpha}_i)\}_{i=1}^{N}$ is asymptotically distributed as an independent identically distributed logistic random variable with parameters, $\log(\alpha_i)$ and one. That is, each has the density function,

$$p(\log(\hat{\alpha}_i)) = \frac{e^{-(\log(\hat{\alpha}_i)-\log(\alpha_i))}}{(1+e^{-(\log(\hat{\alpha}_i)-\log(\alpha_i))})^2} \text{ for } \log(\hat{\alpha}_i) \in (-\infty, \infty). \tag{3.3.1}$$

Now if the differences of each pair of values are taken, k lags apart, that is,

$$d_i^{(k)} = \log(\hat{\alpha}_i) - \log(\hat{\alpha}_{i-k}) \, , i = k+1, \ldots, N$$

then $d_i^{(k)}$=$\log(\alpha_i\zeta_i) - \log(\alpha_{i-k}\zeta_{i-k})$ where asymptotically $\{\zeta_i\}_{i=1}^N$ are independent identically distributed $F_{2,2}$-distributed random variables. Under H_B, $\alpha_i = \alpha$ for all ω_i so $d_i^{(k)} = \log(\alpha\zeta_i) - \log(\alpha\zeta_{i-k}) = \log(\zeta_i) - \log(\zeta_{i-k})$. So under H_B the differences of the logarithms of the ratios are asymptotically distributed as the differences between two independent logistic-distributed random variables with mean zero and unit scale.

Now the logistic distributed random variables, $\{\log(\zeta_i)\}_{i=1}^N$, can be represented as the difference between the logarithms of independent logged χ^2-distributed random variables with two degrees of freedom. Hence $d_i^{(k)} = (u_3 - u_4) - (u_1 - u_2)$ where $u_j \sim \log v_j$, $v_j \sim \chi_2^2$. Re-arranging the difference gives $d_i^{(k)} = (u_2 - u_1) + (u_3 - u_4)$ and the result becomes the sum of the difference between two independent χ^2-distributed random variables with two degrees of freedom. This is equivalent to the sum of two independent logistic random variables of zero mean and unit scale (see (3.3.2) for the associated distribution function from del Castillo [2016] p. 112 and Wästlund [2006] p. 35). Note that in general any random variable that is the difference of the difference of independent identically distributed random variables has the same distribution as sum of the difference of similar independent identically distributed random variables.

This sequence of sums, $d_i^{(k)}$, $i = k+1, \ldots, N$, across all $\{\hat{\alpha}_i\}_{i=1}^N$ is a set of identically-distributed random variables with marginal distribution function,

$$F_d(d_i^{(k)} = d) = \frac{e^d\left(e^d - 1 - d\right)}{(1 - e^d)^2} \tag{3.3.2}$$

(again see del Castillo [2016] p. 112 and Wästlund [2006] p. 35). However the sums (that is, the differences) are not in general independent. Nevertheless asymptotic independence for a subset of the differences can be ensured by retaining only those differences that have no common $\{\log(\hat{\alpha}_i)\}_{i=1}^N$ in the calculation of the $\left\{d_i^{(k)}\right\}$. The distribution of this set of asymptotically-independent identically-distributed differences is then suitable for testing against (3.3.2).

For a differencing lag of ℵ, let $\xi = \lceil N/(2\aleph) \rceil$ where $\lceil x \rceil$ is the largest integer less than or equal to x. Then the number of asymptotically independent differences is at least $\aleph\xi$ illustrated by Figure 3.3.1 which details the outcome for ℵ= 3, 4 and 5 and $N = 32$. Using Figure 3.3.1 it is clear that the ratios, $\hat{\alpha}_1$ to $\hat{\alpha}_{2\aleph}$, represent a "cycle" which contains ℵ asymptotically independent differences and there are $\lceil N/(2\aleph) \rceil$ of these in the dataset. So the total number of independent differences is at least $\aleph\lceil N/(2\aleph) \rceil$.

In general if two comparative differencing lags, $\aleph_1$ and $\aleph_2$, are both factors of N (and N is even for convenience) then the number of asymptotically independent differences is $\aleph_i\lceil N/(2\aleph_i) \rceil = N/2$ for both lags. To express this another way, in this circumstance, each ratio is used once only in a pair of differenced ratios to create the set of asymptotically independent differences. Of course if the chosen lag isn't a factor then there is some "spillage" of differences at the end of the series as shown in Figure 3.3.1.

The use of one of these potential differencing choices, the alternate first differences ($\aleph = 1$), will now be explored for testing H_B v H_C.

First Proposed Difference Test: Alternate First Differences The alternate first differences are $\left\{ d_{2k}^{(1)} \right\}_{k=1}^{\lceil N/2 \rceil}$ where, again, $\lceil h \rceil$ is the largest integer less than or equal to h. These differences are asymptotically independent identically-distributed under H_B with a fully specified distribution (that is, (3.3.2)). They can be used in a classical goodness-of-fit test of the difference of the logarithms of the ratio of the periodogram ordinates against the assumed distribution (under H_B) of the differences of a logistic random variable (with mean zero and unit scale). The Kolmogorov-Smirnov, Cramér-Von Mises or Anderson-Darling statistics can be used although Razali and Wah [2011] show that in general the Anderson-Darling test is more powerful than the Kolmogorov-Smirnov in testing distributional fit and is employed here.

Second Proposed Difference Test: Half Length Differences The above first difference test could exhibit low power if the spectral densities of the two processes change slowly with increasing frequency as would be the case for typical low order autoregressive, moving average or ARMA processes. The first differences

Figure 3.3.1 – Retained Differences for Asymptotically Independent Logged Periodogram Ratios by Differencing Lags of 3, 4 and 5 for N=32 (that is, $n = 64$). The "x" symbols indicate the indices of the ratios employed in the differencing whose interval is printed above each table. The tables which would normally extend out to the right are a compact representation of the span of the differences.

3

1	x		
2	I	x	
3	I	I	x
4	x	I	I
5		x	I
6			x
7	x		
8	I	x	
9	I	I	x
10	x	I	I
11		x	I
12			x
13	x		
14	I	x	
15	I	I	x
16	x	I	I
17		x	I
18			x
19	x		
20	I	x	
21	I	I	x
22	x	I	I
23		x	I
24			x
25	x		
26	I	x	
27	I	I	x
28	x	I	I
29		x	I
30			x
31			
32			

4

1	x			
2	I	x		
3	I	I	x	
4	I	I	I	x
5	x	I	I	I
6		x	I	I
7			x	I
8				x
9	x			
10	I	x		
11	I	I	x	
12	I	I	I	x
13	x	I	I	I
14		x	I	I
15			x	I
16				x
17	x			
18	I	x		
19	I	I	x	
20	I	I	I	x
21	x	I	I	I
22		x	I	I
23			x	I
24				x
25	x			
26	I	x		
27	I	I	x	
28	I	I	I	x
29	x	I	I	I
30		x	I	I
31			x	I
32				x

5

1	x				
2	I	x			
3	I	I	x		
4	I	I	I	x	
5	I	I	I	I	x
6	x	I	I	I	I
7		x	I	I	I
8			x	I	I
9				x	I
10					x
11	x				
12	I	x			
13	I	I	x		
14	I	I	I	x	
15	I	I	I	I	x
16	x	I	I	I	I
17		x	I	I	I
18			x	I	I
19				x	I
20					x
21	x				
22	I	x			
23	I	I	x		
24	I	I	I	x	
25	I	I	I	I	x
26	x	I	I	I	I
27		x	I	I	I
28			x	I	I
29				x	I
30					x
31					
32					

of the logged ratio of the respective spectral densities show only relatively minor aberrations from zero. In this circumstance it would be difficult to detect changes in the distribution of the first differences, compared to the distribution under the null hypothesis.

If the lag in the differences is extended to a larger number of frequencies in the periodogram the level of the differences under these typical models become much more pronounced. Hence the "Alternative First Differences" test statistic can arguably be improved by taking differences across a larger gap than just one time point whilst still avoiding any differences with shared ratios of periodogram ordinates. A possible candidate is to use a gap equivalent to half the number of periodogram ordinates (that is, $\aleph = \lceil N/2 \rceil$). Assuming $N/2$ is a whole number (and using $(N-1)/2$ if not),

$$d_i^{(N/2)} = \log(\hat{\alpha}_i) - \log(\hat{\alpha}_{i-(N/2)}), i = N/2+1, \ldots, N.$$

This again can be compared using the Anderson-Darling test against the distribution function, (3.3.2).

3.3.2 Test using the Sum of Squares of the Differences between the Sample Autocorrelations (Shape Test)

Using sample moments from the time domain, the shape test introduced here is based on the sampling distribution of the weighted sum of the squared differences between the sample autocorrelations. This is an extension of Lund et al. [2009] who use sample autocovariances but the proposed test allows a more specific test of H_B v H_C (compared to H_{B_1} v H_D).

Let $\hat{\rho}_u(k)$ be the sample autocorrelation function for the stationary and invertible series, $\{u_t\}_{t=1}^n$, with autocorrelation function, $\rho_u(k)$. For normally distributed data, it is known that, to the first order,

$$
\begin{aligned}
cov(\hat{\rho}_u(r), \hat{\rho}_u(r+\nu)) \;\approx\; \frac{1}{n} \sum_{m=-\infty}^{\infty} & [\rho_u(m)\rho_u(m+\nu) \\
& + \rho_u(m+r+\nu)\rho_u(m-r) \\
& + 2\rho_u(r)\rho_u(r+\nu)\rho_u^2(m) \\
& - 2\rho_u(r)\rho_u(m)\rho_u(m-r-\nu) \\
& - 2\rho_u(r+\nu)\rho_u(m)\rho_u(m-r)] \,. \qquad (3.3.3)
\end{aligned}
$$

(see Priestley [1981] p. 332).

Letting $\boldsymbol{\rho}_u(L) = (\rho_u(1), \ldots, \rho_u(L))^\top$ and $\hat{\boldsymbol{\rho}}_u(L) = (\hat{\rho}_u(1), \ldots, \hat{\rho}_u(L))^\top$ for a given maximum lag, L, then it is also known that

$$\hat{\boldsymbol{\rho}}_u(L) \xrightarrow{d} N(\boldsymbol{\rho}_u(L), \tfrac{1}{n}\mathbf{W})$$

(Priestley [1981] p. 339) where $\mathbf{W}$ is an $L \times L$ matrix with elements, $w_{ij} = n\, cov(\hat{\rho}_u(i), \hat{\rho}_u(j))$. It therefore follows that, under H_B,

$$\triangle\hat{\boldsymbol{\rho}}_{y_2y_1}(L) \;= \left(\hat{\boldsymbol{\rho}}_{y_2}(L) - \hat{\boldsymbol{\rho}}_{y_1}(L)\right) \xrightarrow{d} N(0, \frac{2}{n}\mathbf{W}),$$

being the difference between two independent multivariate normal random variables with common distribution, $N(\boldsymbol{\rho}_u(L), \frac{1}{n}\mathbf{W})$.

Hence, under H_B,

$$\frac{n}{2}\left(\triangle\hat{\boldsymbol{\rho}}^\top_{y_2y_1}(L)\right)\mathbf{W}^{-1}\left(\triangle\hat{\boldsymbol{\rho}}_{y_2y_1}(L)\right) \xrightarrow{d} \chi^2_L,$$

where χ^2_L denotes a chi-squared random variable with L degrees of freedom. In this (under the null common-shape hypothesis, H_B) for some suitably large $k_{max} > L$, the average of $\{\hat{\rho}_{y_1}(k)\}_{k=1}^{k_{max}}$ and $\{\hat{\rho}_{y_2}(k)\}_{k=1}^{k_{max}}$ by lag, k, are used as an approximation for the population autocorrelations in (3.3.3). That is,

$$\rho_y(k) \;\simeq\; \frac{1}{2}\left(\hat{\rho}_{y_1}(k) + \hat{\rho}_{y_2}(k)\right),\; k = 1, \ldots k_{max}.$$

The statistic,

$$D = \tfrac{n}{2}\left(\triangle\hat{\boldsymbol{\rho}}^\top_{y_2y_1}(L)\right)\mathbf{W}^{-1}\left(\triangle\hat{\boldsymbol{\rho}}_{y_2y_1}(L)\right)$$

provides a test value to be compared to the critical values of the chi-squared (χ^2_L) distribution. If the population autocorrelations differ between series then the difference between the sample autocorrelations will exhibit increased variation and hence D will tend to increase leading to a (one-sided) test of H_B v H_C.

3.3.3 Variance Test (Shape Test)

The shape test of H_B v H_C described here involves the distribution of the sample variance of the logged ratios and is simple to calculate. Knowing that, under H_B, $\{\log \hat{\alpha}_i\}_{i=1}^N$ are asymptotically distributed as independent identically-distributed logistic random variables with mean, $\log(\alpha)$, and scale, 1, define the sample variance of the log of the periodogram ordinate ratios, $\{\log \hat{\alpha}_i\}_{i=1}^N$ as,

$$g_N = \frac{1}{N-1}\sum_{i=1}^{N}\left(\log(\hat{\alpha}_i) - \overline{\log(\hat{\alpha}_i)}\right)^2, \tag{3.3.4}$$

where $\overline{\log(\hat{\alpha}_i)} = \frac{1}{N}\sum_{i=1}^{N}\log(\hat{\alpha}_i)$.

Now, under H_B, $(\log(\hat{\alpha}_i) - \overline{\log(\hat{\alpha}_i)})$ has the same asymptotic distribution as $\left(\hat{z}_i - \overline{\hat{z}}\right)$ where $\hat{z}_i = \log(\hat{\alpha}_i) - \log\alpha$ and $\overline{\hat{z}} = \frac{1}{N}\sum_{i=1}^{N}\hat{z}_i$. Hence for this ratio, if $\log(\hat{\alpha}_i)$ is asymptotically distributed as a logistic random variable with mean, $\log\alpha$, and scale, 1, then for the purposes of (3.3.4), without loss of generality, it can also be said to be asymptotically distributed as a logistic random variable with mean, 0, and scale, 1. This provides a measure of the asymptotic distribution of the sample ratio under H_B, independent of α. That is, it has the asymptotic distribution of the sample variance of an independent identically-distributed set of logistic random variables with zero mean and unit scale.

Given the alternative hypothesis, H_C, that is, that α_i is not constant across all i, then the log of the sample periodogram ratios will diverge from their assumed distribution and will show greater variation (and g_N) providing a one-sided test statistic. The critical values of g_N can be determined by simulation noting that, under H_B, the distribution of the variance of the logged periodogram ratios is independent of α. The critical values have been tabulated for an hypothesis test (of H_B versus H_C) for a range of test sizes and sample counts in Table 3.1 (10,000 repetitions).

Table 3.1 – Critical Values for the Variance Test of Shape. The series lengths (per series, that is, n) include the lengths for the series used in the simulations.

n	Size			
	10%	5%	2.5%	1%
25	5.541	6.557	7.611	9.102
50	4.867	5.500	6.094	6.891
64	4.674	5.222	5.730	6.396
75	4.572	5.067	5.556	6.165
100	4.401	4.803	5.189	5.681
150	4.191	4.504	4.804	5.165
200	4.067	4.337	4.580	4.904
256	3.974	4.211	4.428	4.684
300	3.921	4.127	4.319	4.554
500	3.779	3.937	4.079	4.244
1000	3.634	3.740	3.834	3.946
1024	3.628	3.732	3.826	3.938

3.3.4 Likelihood Ratio Test of the Ratio of the Periodogram Ordinates (Scale Test)

A maximum likelihood estimate of α allows for a likelihood ratio test of H_{B_1} v H_{B_2} (scale test). In this section the maximum likelihood estimate is derived following but expanding on Coates and Diggle [1986] who use the maximum likelihood estimate of $\log \alpha$ in a test of H_{B_1} v H_C (and ultimately H_{B_1} v H_{B_2} and H_B v H_C). For the first time, the uniqueness of the estimate is proven thereby simplifying its interpretation.

Maximum Likelihood Estimate of α Consider the ratio of the ordinates of the periodograms of the two series,

$$\hat{\alpha}_i = \frac{\hat{h}_2(\omega_i)}{\hat{h}_1(\omega_i)},\ i = 1, \ldots, N.$$

Under H_B, $\{\hat{\alpha}_i\}_{i=1}^N$ have the asymptotic probability density function (see Appendix B) given by,

$$p(\hat{\alpha}_i) = \frac{1}{\alpha_i} f(\hat{\alpha}_i/\alpha, 2, 2),$$

where $f(x, n_1, n_2)$ is the probability density function for x being an $F_{(n_1,n_2)}$ distributed random variable.

Hence, the joint likelihood under H_B is,

$$L_1 = \frac{1}{\alpha^N} \prod_{i=1}^{N} \left(1 + \frac{\hat{\alpha}_i}{\alpha}\right)^{-2}.$$

So, the log likelihood, LL_1, is,

$$LL_1 = \log L_1 = -N \log \alpha - 2 \sum_{i=1}^{N} \log \left(1 + \frac{\hat{\alpha}_i}{\alpha}\right).$$

Now, to find the maximum likelihood estimate, $\tilde{\alpha}$ [1], the first derivative with respect to α is derived as,

$$\frac{\partial LL_1}{\partial \alpha} = -\frac{N}{\alpha} - 2 \sum_{i=1}^{N} \frac{1}{(1 + \frac{\hat{\alpha}_i}{\alpha})} (-\hat{\alpha}_i \alpha^{-2})$$

and solved for $\tilde{\alpha}$ when set equal to zero,

$$-\frac{N}{\tilde{\alpha}} - 2 \sum_{i=1}^{N} \frac{1}{(1 + \frac{\hat{\alpha}_i}{\tilde{\alpha}})} (-\hat{\alpha}_i \tilde{\alpha}^{-2}) = 0.$$

So,

$$\begin{aligned}
\sum_{i=1}^{N} \left[\frac{1}{(\frac{\tilde{\alpha}}{\hat{\alpha}_i} + 1)} - \frac{1}{2}\right] &= 0 \\
\sum_{i=1}^{N} \left[\frac{\hat{\alpha}_i}{(\tilde{\alpha} + \hat{\alpha}_i)} - \frac{1}{2}\right] &= 0 \\
\sum_{i=1}^{N} \left[\frac{(\hat{\alpha}_i - \tilde{\alpha})}{(\hat{\alpha}_i + \tilde{\alpha})}\right] &= 0 \\
\frac{\sum_{i=1}^{N} (\hat{\alpha}_i - \tilde{\alpha}) \prod_{j=1 \neq i}^{N} (\hat{\alpha}_j + \tilde{\alpha})}{\prod_{i=1}^{M} (\hat{\alpha}_i + \tilde{\alpha})} &= 0 \\
\sum_{i=1}^{N} (\hat{\alpha}_i - \tilde{\alpha}) \prod_{j=1 \neq i}^{N} (\hat{\alpha}_j + \tilde{\alpha}) &= 0. \qquad (3.3.5)
\end{aligned}$$

That is, the maximum likelihood estimate, $\tilde{\alpha}$, will endeavour to be as close as

[1] Note that $\tilde{\alpha}$ is distinct from $\{\hat{\alpha}_i\}_{i=1}^{N}$ which are the ratios of the periodogram ordinates.

possible to every $\hat{\alpha}_i$ to make the (weighted) summation equal to zero.

Equation 3.3.5 is a polynomial in $\tilde{\alpha}$ of order, N, which potentially has N roots. This arguably presents issues of selection because of the multiple solutions. However it can be shown (see Theorem C.1 in Appendix C) that, knowing all $\{\hat{\alpha}_i > 0\}_{i=1}^{N}$, the polynomial in (3.3.5) has coefficients that, when sorted in order of their descending orders of $\tilde{\alpha}$, have only one change in their signs. Hence by Descartes Rule of Signs only one of the solutions is positive and hence lies in the feasible region. This makes the calculation and interpretation of a maximum likelihood estimate more straightforward by establishing that there is only one solution in the (positive) feasible region.

To derive the asymptotic variance of $\tilde{\alpha}$, the second derivative of the log likelihood is,

$$\begin{aligned}
\frac{\partial^2 LL_1}{\partial^2 \alpha} &= \frac{N}{\alpha^2} - 2\sum_{i=1}^{N} \partial \left[\frac{1}{(1+\frac{\hat{\alpha}_i}{\alpha})} (-\hat{\alpha}_i \alpha^{-2}) \right] / \partial \alpha \\
&= \frac{N}{\alpha^2} + 2\sum_{i=1}^{N} \partial \left[\frac{\hat{\alpha}_i}{(\alpha^2 + \alpha\hat{\alpha}_i)} \right] / \partial \alpha \\
&= \frac{N}{\alpha^2} - 2\sum_{i=1}^{N} \left[\frac{\hat{\alpha}_i (2\alpha + \hat{\alpha}_i)}{(\alpha^2 + \alpha\hat{\alpha}_i)^2} \right].
\end{aligned}$$

The expected value of the negative of the Hessian is,

$$\begin{aligned}
E\left(-\frac{\partial^2 LL_1}{\partial^2 \alpha} \right) &= -E\left(\frac{N}{\alpha^2} - 2\sum_{i=1}^{N} \left[\frac{\hat{\alpha}_i (2\alpha + \hat{\alpha}_i)}{(\alpha^2 + \alpha\hat{\alpha}_i)^2} \right] \right) \\
&= -\frac{N}{\alpha^2} + 2E\left(\sum_{i=1}^{N} \left[\frac{\hat{\alpha}_i (2\alpha + \hat{\alpha}_i)}{(\alpha^2 + \alpha\hat{\alpha}_i)^2} \right] \right) \\
&= -\frac{N}{\alpha^2} + 2\sum_{i=1}^{N} E\left[\frac{\hat{\alpha}_i (2\alpha + \hat{\alpha}_i)}{(\alpha^2 + \alpha\hat{\alpha}_i)^2} \right].
\end{aligned}$$

Hence the asymptotic variance of $\tilde{\alpha}$ is,

$$Var(\tilde{\alpha}) = E\left(-\tfrac{\partial^2 LL_1}{\partial^2 \alpha} \right)^{-1} = \left(-\frac{N}{\alpha^2} + 2\sum_{i=1}^{N} E\left[\frac{\hat{\alpha}_i (2\alpha + \hat{\alpha}_i)}{(\alpha^2 + \alpha\hat{\alpha}_i)^2} \right] \right)^{-1}. \quad (3.3.6)$$

This can be calculated setting $\alpha \approx \tilde{\alpha}$ in (3.3.6) and dropping the expectation (see Section 4.1). Alternatively the negative inverse of the gradient can be numerically derived from the log likelihood at $\alpha = \tilde{\alpha}$.

Likelihood Ratio Tests To prepare a likelihood ratio (scale) test of H_{B_1} v H_{B_2}, under $H_{B_1}$the log likelihood is,

$$LL_2 = \sum_{i=1}^{N} -2\log\left(1+\hat{\alpha}_i\right).$$

The statistic for the likelihood ratio test of H_{B_2} versus H_{B_1} is,

$$LL_{12} = 2(LL_1 - LL_2),$$

where LL_1 is evaluated at $\alpha = \tilde{\alpha}$. This statistic will be asymptotically distributed as a χ^2 random variable with one degree of freedom.

To test H_B versus H_C (a shape test) the likelihood under H_C requires estimation of a unique ratio for each frequency, that is, $\{\alpha_i\}_{i=1}^N$. Hence the likelihood is,

$$L_0 = \prod_{j=1}^{N} \frac{1}{\alpha_i}\left(1+\frac{\hat{\alpha}_i}{\alpha_i}\right)^{-2}$$

and the log likelihood becomes,

$$LL_0 = \sum_{i=1}^{N}\left[-\log(\alpha_i) - 2\log\left(1+\frac{\hat{\alpha}_i}{\alpha_i}\right)\right].$$

The maximum likelihood estimates of $\{\alpha_i\}_{i=1}^N$ are simply $\{\hat{\alpha}_i\}_{i=1}^N$. Hence the likelihood ratio statistic is,

$$\begin{aligned} LL_{01} &= 2(LL_0 - LL_1) \\ &= 2\left(\sum_{i=1}^{N}\left[-\log(\hat{\alpha}_i) - 2\log 2\right] - LL_1\right) \\ &= 2\left(\sum_{i=1}^{N}\left[-\log(\hat{\alpha}_i)\right] - N\log 4 - LL_1\right) \end{aligned}$$

which will be asymptotically distributed as a χ^2 random variable with $N-1$ degrees of freedom.

This test is unlikely to have high power given the large critical values of LL_{01} associated with the high $N-1$ degrees of freedom and is not pursued further in this thesis. However it is possible to approximate a unique ratio for each frequency by representing the log of the ratios by say a quadratic in the frequencies as suggested by Coates and Diggle [1986]. This is undertaken in the size and power studies reported in Section 3.4 as the reference shape test from the literature. None of the multiple sets of 10,000 simulations (for either the model of constant $\{\alpha_i = \alpha\}_{i=1}^{N}$ or the quadratic model) showed any convergence issues (using R's `optim` function).

3.3.5 Wald Test using the Maximum Likelihood Estimate of the Ratio of the Spectral Densities (Scale Test)

Assuming that the shape test, that is, H_B v H_C, has concluded that the two processes have the same spectral shape and that the maximum likelihood value of α has been derived, the Wald test can then be used to test whether the ratio of the two spectral densities (that is, the ratio of the innovations variances) equals one, that is H_{B_1} v H_{B_2}. This allows a simple scale test to be formed from the outcome of the maximum likelihood estimation.

In the present circumstance the Wald test statistic, based on the maximum likelihood estimate for α, $\tilde{\alpha}$, is,

$$W = \frac{(\tilde{\alpha} - 1)}{\hat{\sigma}_{\tilde{\alpha}}},$$

where $\hat{\sigma}_{\tilde{\alpha}}$ is the square root of the estimated variance of $\tilde{\alpha}$ (using the Hessian). Under H_{B_1}, W is asymptotically distributed as a normal random variable with mean zero and standard deviation of one. This suggests the use of W in a two-sided scale test.

3.3.6 Test for the Mean Log Ratio (Scale Tests)

In line with the simplicity and ease of calculation of the variance test (Section 3.3.3), the scale test, H_{B_1} v H_{B_2}, can be undertaken by using the mean logged ratio in the

statistic,

$$Q_N = (\overline{\log \hat{\alpha}} - \log \alpha), \tag{3.3.7}$$

where $\overline{\log \hat{\alpha}}$ is the sample mean of the N logged periodogram ratios. Under H_{B_1} the $\log \hat{\alpha}_i$ are asymptotically distributed as independent identically distributed logistic random variables with mean zero ($\log \alpha = \log 1 = 0$) and scale of one, so Q_N has the form,

$$Q_N = \overline{\log \hat{\alpha}} \quad . \tag{3.3.8}$$

Q_N is asymptotically distributed as $\bar{\hat{z}}_N$ where $\bar{\hat{z}}_N$ is the sample mean of N (asymptotically) independent identically distributed logistic random variables with mean zero and scale of one.

The sample mean of the untransformed $\{\hat{\alpha}_i\}_{i=1}^{N}$is not used because asymptotically they are distributed as $F_{2,2}$ independent random variables which have an undefined population mean.

This Mean-Log-Ratio scale test is a two-sided test for equivalence of the expected value of the log of the ratios of the periodograms to 0. If α is not equal to one, then Q_N will tend to diverge from its expected distribution with values that are too high or too low. Q_N can be employed as a two-sided test statistic using finite sample critical values from Table 3.2. These were derived using simulation employing the mean of independent random variables from a logistic distribution with mean zero and unit scale (10,000 simulations).

An analytical result exists for the mean (that is, the sum) of independent identically distributed logistic random variables (see George and Mudholkar [1983] and Ojo [2003]). However the formula is complex and involves the summation of an infinite number of derivatives of increasing degree.

3.3.7 Central Limit Theorem Test (Scale Test)

Under H_{B_1} and using the Central Limit Theorem, Q_N will asymptotically converge in distribution to the zero-mean normal with variance $\frac{\pi^2}{3N}$ (see Appendix B) with associated (two-sided) normal-distribution critical values. The critical values are very similar to those for the Mean-Log-Ratio test.

The approximate power of this mean log ratio test can be expressed analytically.

Table 3.2 – Critical Values for the Mean Log Ratio Test (Simulated) and Central Limit Theorem Test of Scale (Two-Sided). The series lengths (per series, that is, n) include the lengths for the series used in the simulations.

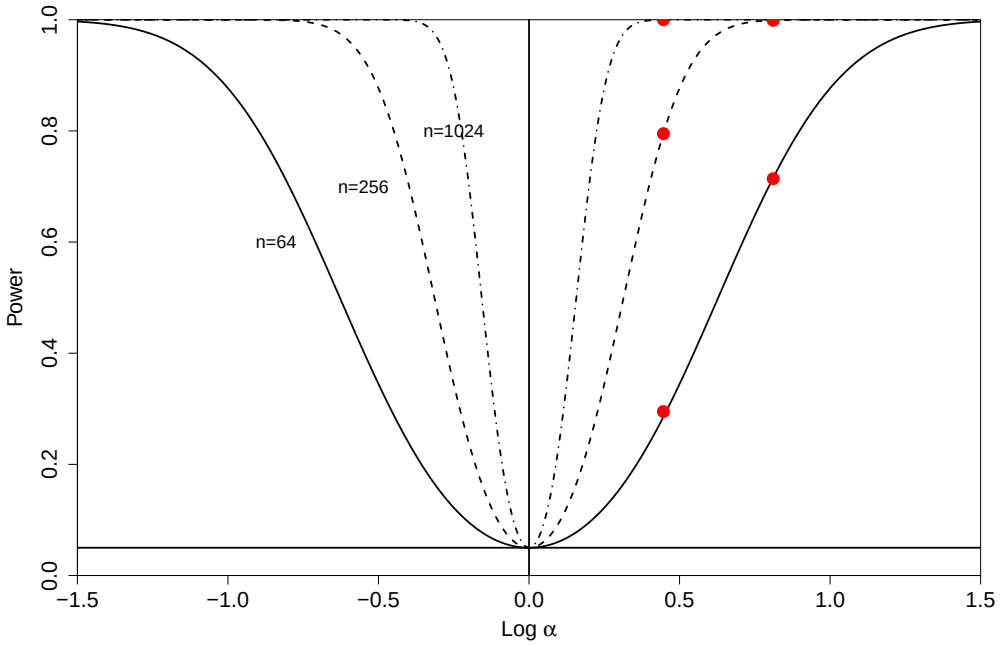
n	Size											
	99.5%		97.5%		95%		5%		2.5%		0.5%	
	CLT	Sim.	*CLT*	Sim.	*CLT*	Sim.	*CLT*	Sim.	*CLT*	Sim.	*CLT*	Sim.
25	*-1.321*	-1.382	*-1.006*	-1.032	*-0.844*	-0.862	*0.844*	0.862	*1.006*	1.032	*1.321*	1.382
50	*-0.934*	-0.938	*-0.711*	-0.711	*-0.597*	-0.595	*0.597*	0.595	*0.711*	0.711	*0.934*	0.938
64	*-0.826*	-0.839	*-0.628*	-0.630	*-0.527*	-0.528	*0.527*	0.528	*0.628*	0.630	*0.826*	0.839
75	*-0.763*	-0.774	*-0.581*	-0.585	*-0.487*	-0.489	*0.487*	0.489	*0.581*	0.585	*0.763*	0.774
100	*-0.661*	-0.668	*-0.503*	-0.501	*-0.422*	-0.420	*0.422*	0.420	*0.503*	0.501	*0.661*	0.668
150	*-0.539*	-0.541	*-0.410*	-0.409	*-0.344*	-0.343	*0.344*	0.343	*0.410*	0.409	*0.539*	0.541
200	*-0.467*	-0.467	*-0.355*	-0.356	*-0.298*	-0.298	*0.298*	0.298	*0.355*	0.356	*0.467*	0.467
256	*-0.413*	-0.416	*-0.314*	-0.315	*-0.264*	-0.264	*0.264*	0.264	*0.314*	0.315	*0.413*	0.416
300	*-0.381*	-0.380	*-0.290*	-0.289	*-0.244*	-0.243	*0.244*	0.243	*0.290*	0.289	*0.381*	0.380
500	*-0.295*	-0.291	*-0.225*	-0.224	*-0.189*	-0.189	*0.189*	0.189	*0.225*	0.224	*0.295*	0.291
1000	*-0.209*	-0.210	*-0.159*	-0.160	*-0.133*	-0.134	*0.133*	0.134	*0.159*	0.160	*0.209*	0.210
1024	*-0.206*	-0.207	*-0.157*	-0.157	*-0.132*	-0.132	*0.132*	0.132	*0.157*	0.157	*0.206*	0.207

If, under H_{B_2}, the actual population logged ratio is ℓ, then the asymptotic power of the test assuming size, υ, becomes $\beta = P(|z| > q_N(\mu = 0, \sigma^2 = \frac{\pi^2}{3N}, 1 - \frac{\upsilon}{2}))$ where $q_N(\mu, \sigma^2, 1 - \frac{\upsilon}{2})$ is the $1 - \frac{\upsilon}{2}$ quantile of a normal random variable with mean, μ, and variance, σ^2, and z is distributed as a normal random variable with mean ℓ and variance, $\frac{\pi^2}{3N}$. That is,

$$\begin{aligned} \beta &= 1 - N(x = q_N(\mu = 0, \sigma^2 = \frac{\pi^2}{3N}, 1 - \frac{\upsilon}{2}), \mu = \ell, \sigma^2 = \frac{\pi^2}{3N}) \\ &\quad + N(x = -q_N(\mu = 0, \sigma^2 = \frac{\pi^2}{3N}, 1 - \frac{\upsilon}{2}), \mu = \ell, \sigma^2 = \frac{\pi^2}{3N}), \end{aligned} \tag{3.3.9}$$

where $N(x, \mu, \sigma^2)$ is the normal distribution function for random variable, x, with mean μ and variance σ^2. The power of the test for a range of ℓ for three sample sizes is plotted in Figure 3.3.2 with the results of the simulation of power from the Central Limit Theorem scale test (from Table 3.8) plotted as large red dots with a horizontal reference line at 0.05.

Figure 3.3.2 – Power of the Central Limit Theorem Scale Test for 5% Significance (Horizontal Reference Line) with the Simulated Power Plotted as Red Dots.



3.4 Size and Power Analysis

In this section, simulation methods are used to explore the size (that is, the probability of a Type I error) and power (that is, the probability of avoiding a Type II error) of the shape and scale tests proposed in Section 3.3 for comparing two series.

From the discussion of testing in the introduction to Section 3.3, it is logical to consider firstly whether two series have a common generating mechanism but possibly different innovations variance. That is, do they have the same shape in their spectral densities (the shape decision)? If so, do they have the same innovations variance (the scale decision)? This is reflected in Diggle and Fisher [1991] and effectively encompasses all the hypotheses in the literature although some tests examine in one step whether the two series have the same shape *and* scale. This unfortunately puts aside the possibility of more detailed conclusions.

Hence the approach used in the current simulation studies will firstly test whether two series have the same spectral shape (or equivalently the same au-

tocorrelation function) and then, if accepted, if they have the same scale (that is, the same innovations variance).

When deciding on the models and model parameters to be used in the simulation work, there are some characteristics of the shape and scale testing which suggest the choice of the underlying time series processes. In investigating the shape tests for size, the probability is explored of rejecting the null hypothesis that the series have the same spectral shape given this is true. However the shape tests introduced here are all invariant under a constant multiplicative transformation of either series and hence are invariant to scale. Hence without loss of generality the size review of the shape tests can use two series with the same innovations variance (say, unity), as well as the same spectral shape as required under the null hypothesis.

The scale tests don't exhibit a similar invariance with respect to spectral shape. This is because the ratios of the periodogram ordinates of two series are not invariant if a common filter is applied to both series, as opposed to the effect of the filtering on the ratios of the population spectral densities. This invariance can be demonstrated with a single simulation (not shown). However, given that the scale tests are only used if a common spectral shape is accepted then both series will be generated using the same process mechanism. Hence when investigating the scale tests for size, they will be explored under the null hypothesis of the same scale (that is, say, common unit innovations variance) and using various common ARMA models.

As a representative of existing tests, the likelihood ratio shape test of Coates and Diggle [1986] (based on a quadratic approximation to the shape of the log of the ratios against the frequencies) was included along with an extension of their likelihood ratio test as a scale test.

The shape tests considered in the current size and power study are (along with their abbreviations):

1. LRT 2: Likelihood ratio test using a second order polynomial in the logged ratios (from Coates and Diggle [1986]).

2. Δ^1: Alternate first differences of the logged ratios.

3. $\Delta^{\frac{N}{2}}$: Differences with a lag equal to half the number of ratios.

4. ACF: Weighted sum of the squared differences between the sample autocorrelations.

5. Var.: Variance-of-ratios test.

The scale tests are:

1. LRT 0: Likelihood ratio test using the untransformed ratios (after Coates and Diggle [1986] but they employ the logged ratios).

2. Wald: Wald test using the maximum likelihood estimate of the ratio.

3. $\overline{\text{Log}}$: Test using the sample mean of the logged ratios.

4. CLT: Test employing the Central Limit Theorem to obtain the asymptotic distribution of the sample mean of the logged ratios.

3.4.1 Simulation Design

The simulation design process defined in Paxton et al. [2001] is followed here.

As mentioned earlier, the measures used to assess the performance of each of the tests for a given pair of ARMA models, ratios of spectral densities and set of series lengths, are size (the probability of rejecting the null hypothesis when it's true) and power (the probability of rejecting the null hypothesis when it's false).

Candidate Representative Models Processes selected for shape and scale comparison are simple models deemed to be most useful in applied work. AR(1) and MA(1) processes are selected for shape and scale size exploration with the parameter set at 0.5 and -0.5 (see Table 3.5). White noise is also used. For shape power analysis, white noise is tested against a range of AR(1) and MA(1) processes (see Table 3.7); a comparison is also undertaken of AR(1) versus AR(1) series. For scale power analysis, $\sqrt{\alpha}$ values from 1.00 to 3.00 are used along with the same AR and MA processes employed in the shape and scale size analysis (see Table 3.8).

Simulations Simulations are undertaken for series lengths which most closely matched those used in the literature (see Table A.4), being $n = 64$, 256 and 1024. These also present a substantial range of sample sizes. Ten thousand simulations are undertaken for each scenario.

The simulation results have an associated approximate confidence interval of $\pm 1.96\sqrt{\frac{p_{sim}(1-p_{sim})}{n_{sim}}}$ where n_{sim} is the number of simulation cycles and p_{sim} is the associated test size or power. For the nominal 5% significance level and 10,000 simulations, the confidence intervals for the simulated size results are approximately ± 0.004. For the power studies for 10,000 simulations, the confidence intervals are at most ± 0.010.

The analysis will firstly address the verification of test size followed by power studies. However a discussion of the sequential nature of the testing will be undertaken beforehand.

3.4.2 Shape and Scale Hypothesis Tests in Sequence

Independent size and power analyses don't address the fact that the shape and scale tests are designed to be undertaken in sequence. A scale test is only undertaken if the hypothesis of a common shape has been accepted. The nature of this relationship is shown in Tables 3.3 and 3.4. Table 3.3 relates the shape test outcome to the true shape hypothesis and illustrates that the testing only proceeds to the scale test after the shape null hypothesis has been accepted. Table 3.4 shows the veracity of the combined shape and scale testing outcome given the true hypothesis and further illustrates the dependent nature of the size and power outcomes for the scale test. This is discussed below.

Table 3.3 – Shape Hypothesis Testing in Sequence

		Test Outcome	
		Accept H_{Sh}	Reject H_{Sh}
True Shape Hypothesis	H_{Sh}	Correct $\Rightarrow$ Test H_{Sc}	Incorrect$\Rightarrow$ Halt
	Not H_{Sh}	Incorrect$\Rightarrow$ Test H_{Sc}	Correct$\Rightarrow$ Halt

Table 3.4 – Shape and Scale Hypothesis Testing Outcomes

		Tests Outcome		
		Reject H_{Sh}	Accept H_{Sh} and Accept H_{Sc}	Accept H_{Sh} and Reject H_{Sc}
True Hypothesis	Not H_{Sh}	Correct	Incorrect	Incorrect
	H_{Sh} and H_{Sc}	Incorrect	Correct	Incorrect
	H_{Sh} and Not H_{Sc}	Incorrect	Incorrect	Correct

Size of the Shape and Scale Tests For linked shape and scale tests, the size for any of the shape tests is unaffected by sequential testing given its the first test. The actual size of the shape test is therefore simply the nominal size of the shape test.

However, for the scale test, the size of the test is the probability that the scale test rejects the null hypothesis of a common scale given that there is actually a common scale. This test also assumes that there is a common shape (otherwise testing for scale without a common shape is nonsensical) whether there is a common scale or not. Moreover it would typically only follow a shape test that had accepted the null hypothesis of a common shape.

So for the scale test there is an assumption for the size calculation that the two series share both a common shape and a common scale. However the scale test is only undertaken if the shape test accepts the null hypothesis test of a common shape. This has a chance of being falsely rejected which would then also reject the scale test by not leading to it.

Hence the final size of the scale test is the probability of incorrectly rejecting a common shape given there is a common shape plus the probability of correctly accepting a common shape given a common shape times the probability of rejecting a common scale given a common shape and scale (see (3.4.1)). The events involved here correspond to the outcomes in the second row (that is, "H_{Sh} and H_{Sc}") in Table 3.4.

Note that, under the assumption of a common shape, the test outcome of ac-

cepting a common shape has no effect on the probability of a common scale other than shown in (3.4.1) because, under the assumption of a common shape, the scale of the series doesn't affect the shape tests used here. Hence the outcome of accepting a common shape doesn't affect the probability of accepting (or rejecting) a common scale.

Following the logic above, the overall size of the scale test following the shape test can be expressed as,

$$\upsilon_{Sc} \quad = \quad \upsilon_{Sh} + (1 - \upsilon_{Sh}) * \upsilon_{Sc|Sh}, \tag{3.4.1}$$

where υ_{Sc} is the resultant final size of the scale test, υ_{Sh} is the size of the shape test, and $\upsilon_{Sc|Sh}$ is the size of the scale test (after the shape test has correctly concluded a common shape). Typically then, if the objective is $\upsilon_{Sc} = 0.05$ and letting $\upsilon_{Sh} = \upsilon_{Sc|Sh}$, then $\upsilon_{Sc|Sh} = 1 - \sqrt{1 - \upsilon_{Sc}} \approx \frac{\upsilon_{Sc}}{2}$ for small υ_{Sc}. Hence for an overall 5% significance level for the scale test, this requires approximately a notional 2.5% level for both the shape test and the subsequent scale test. For this reason the studies in this thesis on size and power include a 2.5% significance level.

Power of the Shape and Scale Tests The powers of the shape and scale tests reflect a similar issue to the size of the tests. The power of any test is the probability of rejecting the null hypothesis when the alternative hypothesis is true. Given that it's undertaken first, the shape test has a power that's unaffected in sequential testing.

However the power of the scale test following the shape test is affected. The power definition used here is the probability that the scale and shape test in sequence will produce a decision in favour of a difference in scale after a decision in favour of a common shape, given that there is a common shape but difference in scale (any other shape assumption for the power derivation would be nonsensical as with the size assessment). This is the outcome of choosing a common shape (given a common shape) and of then choosing a difference in scale (given a difference in scale and a common shape). If a common shape is not chosen, then the scale test can't be executed to reject the null hypothesis and hence the power of the scale test is lower.

Hence the power outcome for the scale test becomes,

$$\beta_{Sc} \;=\; (1-\upsilon_{Sh})\beta_{Sc|Sh},$$

where β_{Sc} is the overall power of the scale test, υ_{Sh} is the size of the shape test and $\beta_{Sc|Sh}$ is the power of the scale test after the shape test correctly concludes a common shape. Given that say $\upsilon_{Sh} = 0.025$ (from the size discussion) then the resultant effect on the power of the scale test will be quite modest (that is, $0.975\beta_{Sc|Sh}$).

3.4.3 Size Verification

Tables 3.5 and 3.6 show the simulated sizes for the shape (first five) and scale (last four) tests for nominal sizes of 10% and 5% and 2.5% (to accommodate testing in sequence from Section 3.4.2) and 1% respectively. Figure 3.4.1 plots the results for sizes of 10% (blue), 5% (red) and 1% (green) by the generating process and sample sizes (64, 256 and 1024 per series) for each of the two series. "WN" refers to white noise.

In general the simulated sizes are somewhat larger than the nominal size but typically converge to the nominal values as $n \to \infty$. The simulated sizes are relatively consistent between generating processes for the same test and sample size.

Amongst the shape tests, the likelihood ratio and variance tests show actual sizes that are roughly 2-3% above the nominal values for small n. The 1^{st} and $\frac{N}{2}^{th}$ difference tests are consistently close to the nominal value. The size for the test based on the difference in the autocorrelations varies between generating processes.

For the scale tests, the likelihood ratio, mean log and Central Limit Theorem tests are close to the nominal size for all n whilst the Wald test shows somewhat higher sizes than the nominal values for small n.

3.4.4 Power of the Shape Tests

Table 3.7 and Figure 3.4.2 present the power of the two-series shape tests at 5% significance level for sample sizes of 64, 256 and 1024 per series. The tests are

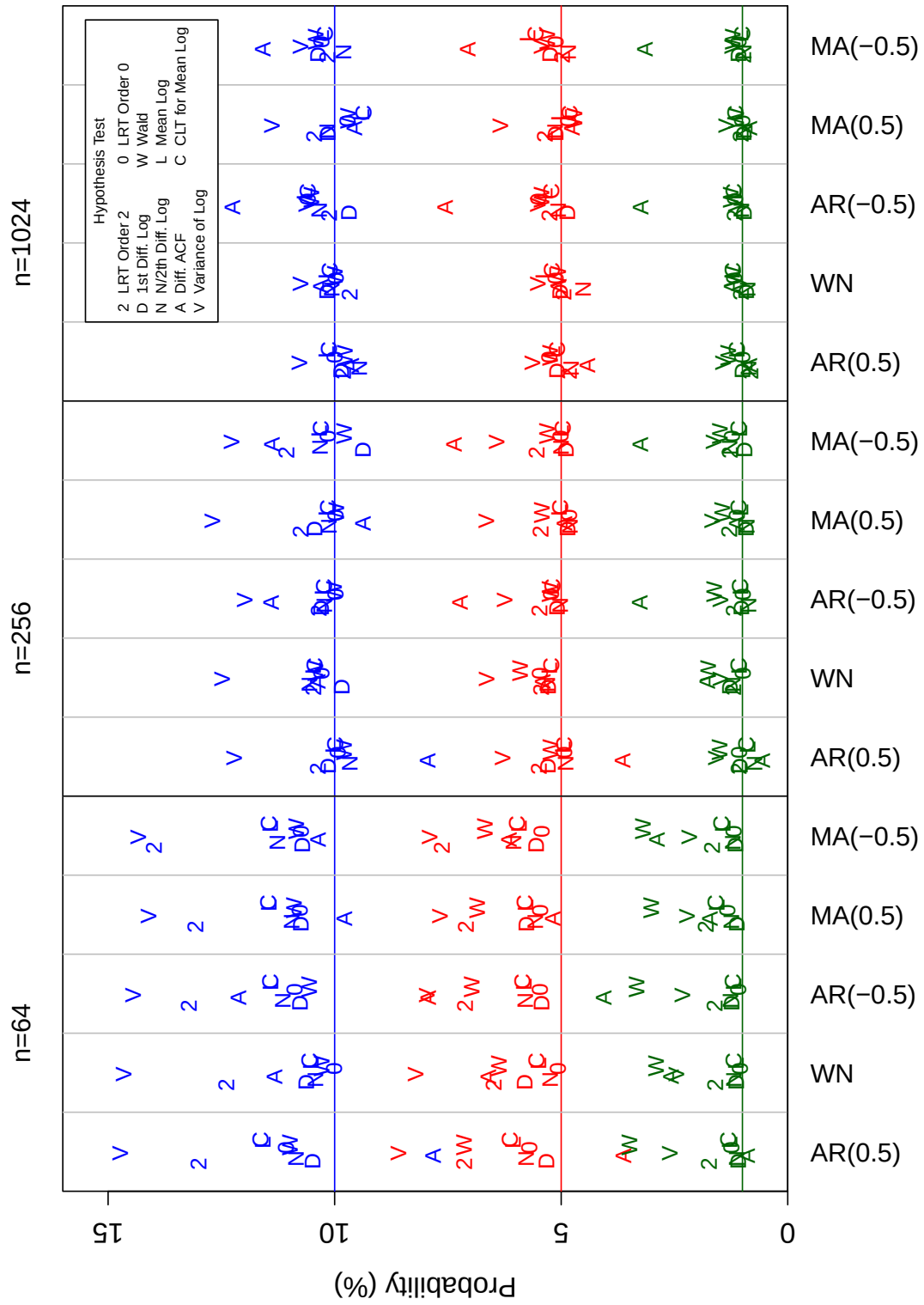
Table 3.5 – Verification of Hypothesis Test Size for Two Series (Nominal Sizes of 0.10 and 0.05) for Shape and Scale Tests

Size	n	Process	LRT 2	Δ^1	$\Delta^{\frac{N}{2}}$	ACF	Var.	LRT 0	Wald	$\overline{\text{Log}}$	CLT
0.10	64	AR(0.5)	0.130	0.105	0.109	0.078	0.147	0.111	0.110	0.116	0.116
0.10	64	WN	0.124	0.106	0.104	0.113	0.147	0.100	0.103	0.105	0.106
0.10	64	AR(-0.5)	0.132	0.108	0.112	0.121	0.144	0.109	0.106	0.114	0.114
0.10	64	MA(0.5)	0.131	0.108	0.110	0.098	0.141	0.108	0.109	0.115	0.115
0.10	64	MA(-0.5)	0.140	0.107	0.113	0.104	0.143	0.107	0.108	0.114	0.115
0.10	256	AR(0.5)	0.104	0.102	0.097	0.080	0.122	0.099	0.098	0.100	0.100
0.10	256	WN	0.105	0.098	0.106	0.104	0.125	0.103	0.104	0.105	0.104
0.10	256	AR(-0.5)	0.104	0.103	0.102	0.114	0.120	0.100	0.100	0.103	0.102
0.10	256	MA(0.5)	0.108	0.104	0.101	0.094	0.127	0.100	0.100	0.102	0.102
0.10	256	MA(-0.5)	0.111	0.094	0.103	0.114	0.123	0.102	0.098	0.104	0.103
0.10	1024	AR(0.5)	0.098	0.099	0.095	0.096	0.108	0.100	0.098	0.102	0.101
0.10	1024	WN	0.097	0.102	0.101	0.103	0.108	0.100	0.100	0.102	0.101
0.10	1024	AR(-0.5)	0.101	0.097	0.104	0.123	0.106	0.106	0.105	0.107	0.106
0.10	1024	MA(0.5)	0.104	0.102	0.102	0.096	0.114	0.097	0.097	0.094	0.094
0.10	1024	MA(-0.5)	0.102	0.104	0.098	0.116	0.108	0.103	0.104	0.103	0.102
0.05	64	AR(0.5)	0.071	0.053	0.058	0.036	0.086	0.057	0.072	0.061	0.061
0.05	64	WN	0.065	0.058	0.052	0.066	0.082	0.051	0.064	0.055	0.056
0.05	64	AR(-0.5)	0.071	0.054	0.058	0.079	0.080	0.055	0.070	0.058	0.059
0.05	64	MA(0.5)	0.071	0.058	0.056	0.052	0.077	0.055	0.069	0.058	0.058
0.05	64	MA(-0.5)	0.076	0.056	0.061	0.062	0.079	0.054	0.067	0.059	0.060
0.05	256	AR(0.5)	0.055	0.053	0.049	0.036	0.063	0.049	0.052	0.050	0.049
0.05	256	WN	0.054	0.053	0.053	0.054	0.066	0.055	0.059	0.053	0.052
0.05	256	AR(-0.5)	0.055	0.051	0.050	0.072	0.062	0.052	0.053	0.052	0.052
0.05	256	MA(0.5)	0.054	0.049	0.048	0.049	0.067	0.048	0.054	0.050	0.050
0.05	256	MA(-0.5)	0.055	0.049	0.050	0.074	0.064	0.050	0.053	0.050	0.050
0.05	1024	AR(0.5)	0.048	0.051	0.048	0.044	0.056	0.053	0.052	0.052	0.051
0.05	1024	WN	0.050	0.050	0.045	0.051	0.055	0.051	0.051	0.054	0.052
0.05	1024	AR(-0.5)	0.053	0.049	0.051	0.076	0.055	0.055	0.054	0.054	0.052
0.05	1024	MA(0.5)	0.054	0.051	0.051	0.048	0.064	0.048	0.047	0.050	0.048
0.05	1024	MA(-0.5)	0.050	0.052	0.048	0.071	0.054	0.051	0.053	0.058	0.056

Table 3.6 – Verification of Hypothesis Test Size for Two Series (Nominal Sizes of 0.025 and 0.01) for Shape and Scale Tests

Size	n	Process	LRT 2	Δ^1	$\Delta^{\frac{N}{2}}$	ACF	Var.	LRT 0	Wald	$\overline{\text{Log}}$	CLT
0.025	64	AR(0.5)	0.040	0.027	0.028	0.019	0.050	0.028	0.051	0.033	0.032
0.025	64	WN	0.035	0.029	0.027	0.041	0.046	0.024	0.042	0.029	0.029
0.025	64	AR(-0.5)	0.038	0.027	0.029	0.056	0.046	0.027	0.049	0.029	0.029
0.025	64	MA(0.5)	0.039	0.028	0.030	0.031	0.044	0.030	0.047	0.032	0.032
0.025	64	MA(-0.5)	0.039	0.029	0.030	0.044	0.045	0.026	0.045	0.032	0.032
0.025	256	AR(0.5)	0.029	0.028	0.024	0.016	0.033	0.025	0.029	0.026	0.026
0.025	256	WN	0.029	0.027	0.028	0.032	0.034	0.026	0.035	0.025	0.025
0.025	256	AR(-0.5)	0.028	0.024	0.023	0.048	0.033	0.025	0.030	0.025	0.026
0.025	256	MA(0.5)	0.028	0.024	0.023	0.025	0.037	0.026	0.029	0.026	0.026
0.025	256	MA(-0.5)	0.031	0.025	0.028	0.050	0.033	0.027	0.030	0.026	0.027
0.025	1024	AR(0.5)	0.023	0.027	0.023	0.021	0.032	0.027	0.030	0.028	0.027
0.025	1024	WN	0.026	0.024	0.023	0.028	0.026	0.026	0.027	0.029	0.028
0.025	1024	AR(-0.5)	0.027	0.023	0.027	0.050	0.029	0.027	0.028	0.028	0.027
0.025	1024	MA(0.5)	0.026	0.025	0.025	0.024	0.034	0.024	0.025	0.026	0.025
0.025	1024	MA(-0.5)	0.026	0.027	0.023	0.047	0.028	0.026	0.028	0.028	0.027
0.01	64	AR(0.5)	0.017	0.011	0.011	0.009	0.026	0.012	0.035	0.013	0.013
0.01	64	WN	0.016	0.011	0.012	0.026	0.025	0.010	0.029	0.012	0.012
0.01	64	AR(-0.5)	0.016	0.012	0.013	0.041	0.023	0.011	0.033	0.012	0.012
0.01	64	MA(0.5)	0.018	0.011	0.012	0.017	0.022	0.013	0.030	0.016	0.016
0.01	64	MA(-0.5)	0.017	0.012	0.012	0.029	0.022	0.012	0.032	0.014	0.015
0.01	256	AR(0.5)	0.011	0.011	0.007	0.006	0.016	0.011	0.015	0.009	0.009
0.01	256	WN	0.012	0.013	0.012	0.018	0.015	0.010	0.018	0.011	0.011
0.01	256	AR(-0.5)	0.012	0.010	0.009	0.033	0.015	0.010	0.016	0.010	0.011
0.01	256	MA(0.5)	0.013	0.009	0.009	0.011	0.017	0.011	0.014	0.010	0.011
0.01	256	MA(-0.5)	0.013	0.010	0.013	0.033	0.016	0.012	0.015	0.011	0.011
0.01	1024	AR(0.5)	0.008	0.010	0.009	0.008	0.014	0.010	0.013	0.011	0.011
0.01	1024	WN	0.010	0.009	0.009	0.012	0.011	0.012	0.012	0.012	0.012
0.01	1024	AR(-0.5)	0.012	0.010	0.010	0.032	0.012	0.012	0.013	0.012	0.012
0.01	1024	MA(0.5)	0.010	0.010	0.010	0.008	0.014	0.010	0.011	0.012	0.012
0.01	1024	MA(-0.5)	0.010	0.011	0.010	0.032	0.012	0.011	0.012	0.010	0.010

Figure 3.4.1 – Verification of Hypothesis Test Size (Scale and Shape Tests) for Two Series (Significance Levels of 10% (blue), 5% (red) and 1% (green)). The vertical lines differentiate the sample size values and ARMA processes. The x-axis labels show the type of ARMA model and parameterisation employed.



examined for power using white noise (WN) versus lag-one autoregressive process with ranges of parameters from -0.75 to 0.75. The modelling also compares two autoregressive processes and WN versus moving average processes. Some of these power results effectively show the size of the test by testing shape for processes with the same ARMA coefficients. This is to demonstrate the continuum of results for comparison purposes.

Whilst noting the discussion of sequential testing from Section 3.4.2, for ease of comparison against other power studies and to retain reader familiarity, the studies use a test size of 5%.

Across the process comparisons and sample sizes for shape tests, the likelihood ratio and autocorrelation tests have comparable power which is also the highest of all the shape tests. The first difference test (see column "$\triangle^1$") shows low power in all circumstances reflecting the slow and smooth change in the population spectral densities of the two processes. However if the lag in the differences is extended to half the number of frequencies in the periodogram (see column "$\triangle^{\frac{N}{2}}$") the power becomes much larger and is very close to that of the likelihood ratio and autocorrelation tests especially for moderate to large sample size ($n \geq 256$); it is also arguably simpler to use. It could even be said to be superior to the likelihood ratio test given that the $N/2^{th}$ difference test has a closer match to its nominal size than the likelihood ratio test (see Table 3.5). The latter has a higher actual size.

The variance test has quite low power for all sample sizes and processes. For all tests, for the autoregressive versus white noise processes, the power increases as the absolute value of the autoregressive coefficient increases towards one.

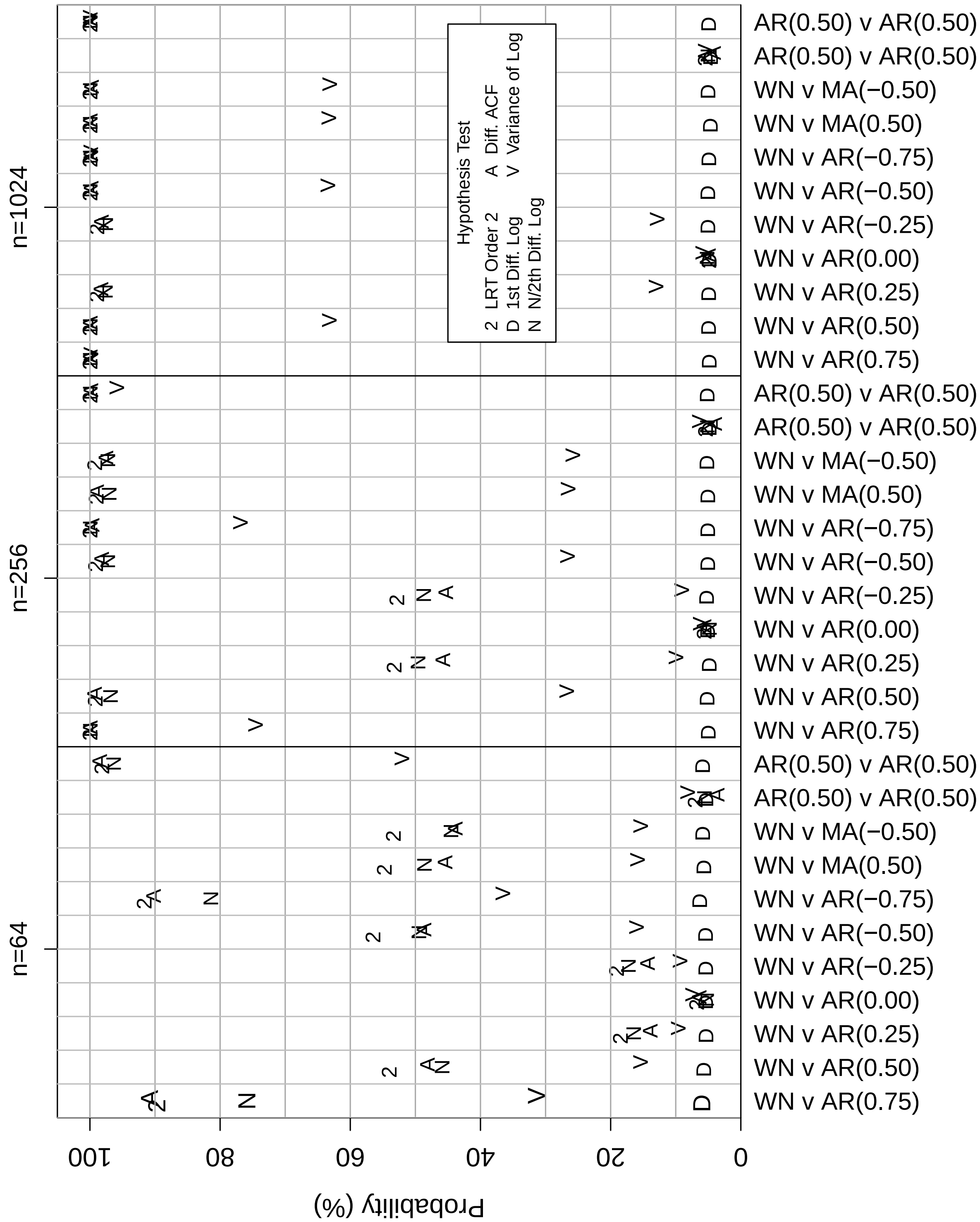
3.4.5 Power of the Scale Tests

The power of the two-series scale tests at 5% significance level for sample sizes of 64 and 256 per series are presented in Table 3.8 and Figure 3.4.3. The results for $n = 1024$ are not shown as the power for $\alpha > 1$ was always 1.000. Ratios of 1.00 versus 1.00, 1.25, 1.50 and 3.00 (being $\sqrt{\alpha}$) have been employed for various common AR and MA generating processes. Again as with the shape tests, some of the power results show the size of the tests by testing the scale for processes with the same innovations variance. As with the shape tests, for ease of comparison against other

Table 3.7 – Power of Hypothesis Tests for Shape for Two Series (5% Test Size)

n	1st	2nd	LRT 2	Δ^1	$\Delta^{\frac{N}{2}}$	ACF	Var.
64	WN	AR(0.75)	0.898	0.061	0.760	0.909	0.315
64	WN	AR(0.50)	0.540	0.058	0.459	0.482	0.155
64	WN	AR(0.25)	0.185	0.054	0.165	0.140	0.097
64	WN	AR(0.00)	0.068	0.054	0.053	0.064	0.075
64	WN	AR(-0.25)	0.191	0.054	0.173	0.144	0.094
64	WN	AR(-0.50)	0.565	0.055	0.495	0.487	0.161
64	WN	AR(-0.75)	0.917	0.064	0.815	0.903	0.366
64	WN	MA(0.50)	0.548	0.057	0.487	0.455	0.159
64	WN	MA(-0.50)	0.534	0.059	0.446	0.439	0.154
64	AR(0.50)	AR(0.50)	0.070	0.054	0.056	0.037	0.082
64	AR(0.50)	AR(-0.50)	0.982	0.059	0.964	0.986	0.521
256	WN	AR(0.75)	1.000	0.050	1.000	1.000	0.746
256	WN	AR(0.50)	0.992	0.052	0.969	0.993	0.268
256	WN	AR(0.25)	0.533	0.049	0.496	0.458	0.100
256	WN	AR(0.00)	0.056	0.052	0.049	0.057	0.063
256	WN	AR(-0.25)	0.529	0.053	0.488	0.454	0.091
256	WN	AR(-0.50)	0.991	0.051	0.973	0.982	0.267
256	WN	AR(-0.75)	1.000	0.051	1.000	0.997	0.769
256	WN	MA(0.50)	0.990	0.051	0.971	0.990	0.266
256	WN	MA(-0.50)	0.993	0.053	0.973	0.976	0.259
256	AR(0.50)	AR(0.50)	0.054	0.049	0.052	0.041	0.064
256	AR(0.50)	AR(-0.50)	1.000	0.052	1.000	0.999	0.959
1024	WN	AR(0.75)	1.000	0.049	1.000	1.000	0.999
1024	WN	AR(0.50)	1.000	0.050	1.000	1.000	0.633
1024	WN	AR(0.25)	0.990	0.050	0.977	0.984	0.131
1024	WN	AR(0.00)	0.049	0.049	0.052	0.050	0.059
1024	WN	AR(-0.25)	0.989	0.051	0.976	0.983	0.129
1024	WN	AR(-0.50)	1.000	0.051	1.000	0.999	0.635
1024	WN	AR(-0.75)	1.000	0.050	1.000	0.999	0.999
1024	WN	MA(0.50)	1.000	0.047	1.000	1.000	0.634
1024	WN	MA(-0.50)	1.000	0.051	1.000	0.998	0.632
1024	AR(0.50)	AR(0.50)	0.055	0.047	0.051	0.042	0.055
1024	AR(0.50)	AR(-0.50)	1.000	0.050	1.000	1.000	1.000

Figure 3.4.2 – Power Results for Shape Tests for Two Series (5% Test Size). The vertical lines differentiate the sample size values and ARMA processes. The x-axis labels show the pairs of ARMA models and parameterisations employed.



power studies and to retain reader familiarity, the studies use a significance level of 5%.

The power for all tests except the Wald test is at or over 80% for sample sizes greater than or equal to 256. The mean log test is comparable to the likelihood ratio test in power across all processes, sample sizes and ratios of innovations variances and is much simpler to implement. The Central Limit Theorem test has almost exactly the same power as both of the tests and is even simpler to use. Figure 3.3.2 shows that the power of the Central Limit Theorem test agrees very closely with the analytical approximation from (3.3.9). The Wald test shows substantially reduced power compared to the other tests (note the very low power of <10% for $\sqrt{\alpha} = 1.25$ and $n = 64$) but the gap closes with increased sample size. In general all tests show relatively good power for sample sizes greater than or equal to 256.

A phenomenon evident in the tables for all the scale test powers is that, for the same α value and sample size, the power for all scenarios is roughly the same irrespective of the common ARMA process.

3.4.6 Conclusions

The simulated sizes for all shape and scale tests for all generating processes are typically slightly greater than the nominal sizes but reduce to the nominal values with increasing series length.

When testing shape, the likelihood ratio, autocorrelation and $\frac{N}{2}^{th}$ difference tests show the highest power with the latter having the most consistent test size. The literature review in this thesis (see Section 3.2) suggests that the likelihood ratio test is amongst the best performing of the existing tests and hence the new shape tests show promise especially given their simplicity.

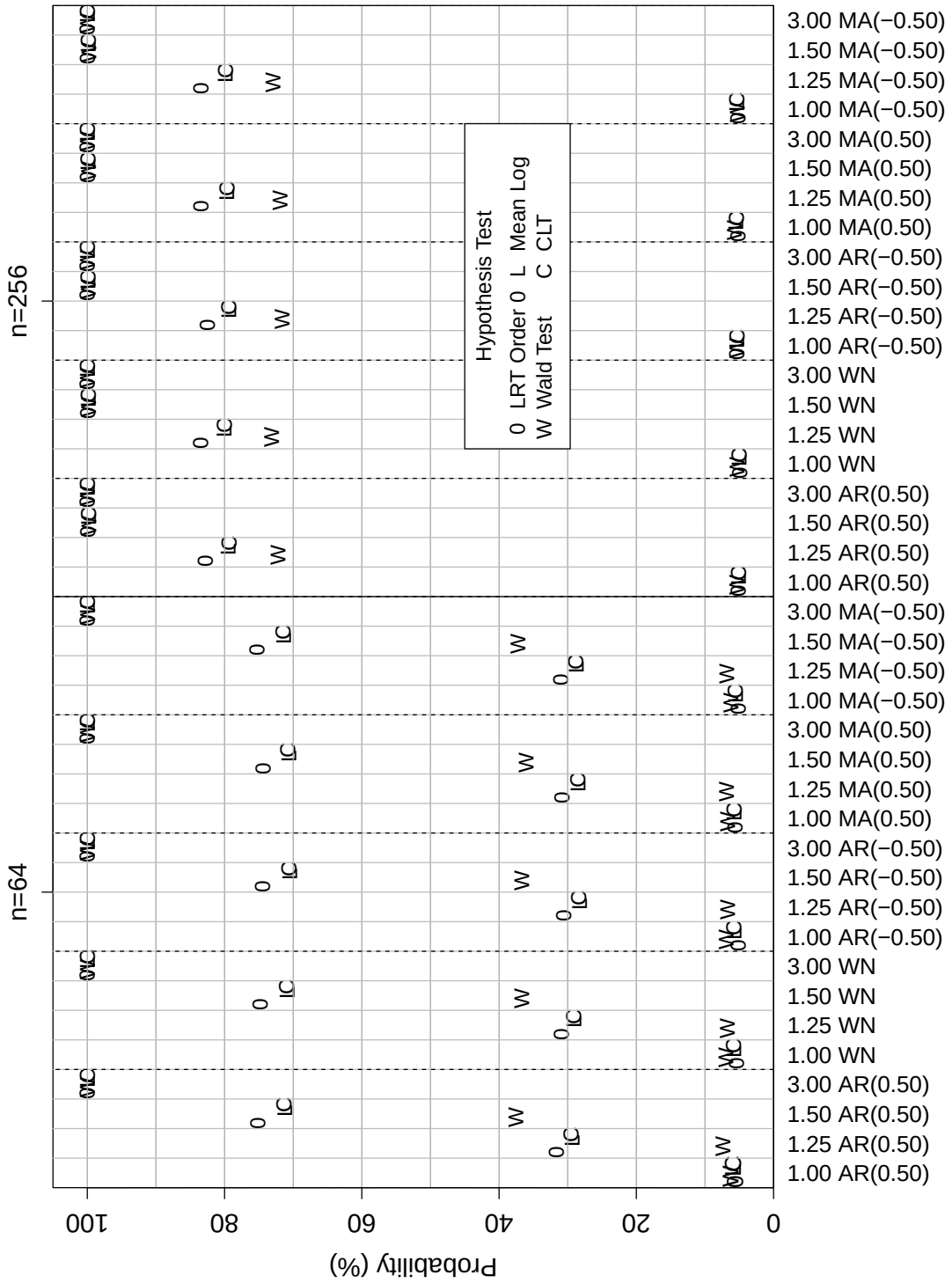
For the scale tests, the mean log test even when incorporating the Central Limit Theorem is at least as powerful as the other tests and is very easy to deploy.

As well as their use in generally comparing two processes, these shape and scale tests can form the basis of a graphical tool for comparing and identifying processes as presented in the next section.

Table 3.8 – Power of Hypothesis Tests for Scale for Two Series with $n = 64$ and 256 per Series where α is the Ratio of the Innovations Variances (5% Test Size). For $n = 1024$, all powers were 100% except at $\alpha = 1$.

		n=64				n=256			
Process	$\sqrt{\alpha}$	LRT 0	Wald	$\overline{\text{Log}}$	CLT	LRT 0	Wald	$\overline{\text{Log}}$	CLT
AR(0.5)	1.00	0.056	0.063	0.059	0.060	0.051	0.054	0.052	0.051
AR(0.5)	1.25	0.316	0.074	0.294	0.295	0.828	0.722	0.795	0.794
AR(0.5)	1.50	0.752	0.376	0.713	0.714	0.999	0.998	0.999	0.999
AR(0.5)	3.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
WN	1.00	0.054	0.070	0.058	0.058	0.049	0.054	0.051	0.050
WN	1.25	0.309	0.068	0.291	0.292	0.835	0.732	0.801	0.801
WN	1.50	0.748	0.367	0.709	0.710	0.999	0.998	0.999	0.999
WN	3.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AR(-0.5)	1.00	0.051	0.069	0.058	0.058	0.053	0.054	0.054	0.054
AR(-0.5)	1.25	0.306	0.067	0.283	0.284	0.825	0.716	0.794	0.794
AR(-0.5)	1.50	0.744	0.367	0.705	0.706	1.000	0.999	0.999	0.999
AR(-0.5)	3.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MA(0.5)	1.00	0.056	0.067	0.058	0.058	0.052	0.058	0.055	0.055
MA(0.5)	1.25	0.308	0.069	0.285	0.286	0.834	0.719	0.797	0.797
MA(0.5)	1.50	0.744	0.361	0.706	0.708	1.000	0.999	0.999	0.999
MA(0.5)	3.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MA(-0.5)	1.00	0.051	0.062	0.056	0.056	0.052	0.053	0.054	0.054
MA(-0.5)	1.25	0.310	0.068	0.288	0.289	0.834	0.730	0.799	0.799
MA(-0.5)	1.50	0.752	0.373	0.714	0.715	1.000	0.998	0.999	0.999
MA(-0.5)	3.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 3.4.3 – Power Results for Scale Tests for Two Series (5% Test Size). The vertical lines differentiate the sample size values, ARMA processes and population spectral ratios. The x-axis labels show the square root of the population spectral ratio and the common ARMA process. For $n = 1024$, all powers were 100% except at $\alpha = 1$.



3.5 Graphical Identification Method

The objective of this section is to propose a graphical method to explore whether two processes have the same spectral shape. If they do, then the method tests whether they have the same innovations variance and determines the order of their common ARMA representation.

The approach prepares plots of the logged ratios of the periodogram ordinates against frequency and plots of the merged sample autocorrelations and partial autocorrelations. This therefore presents both a frequency and time domain image of the data.

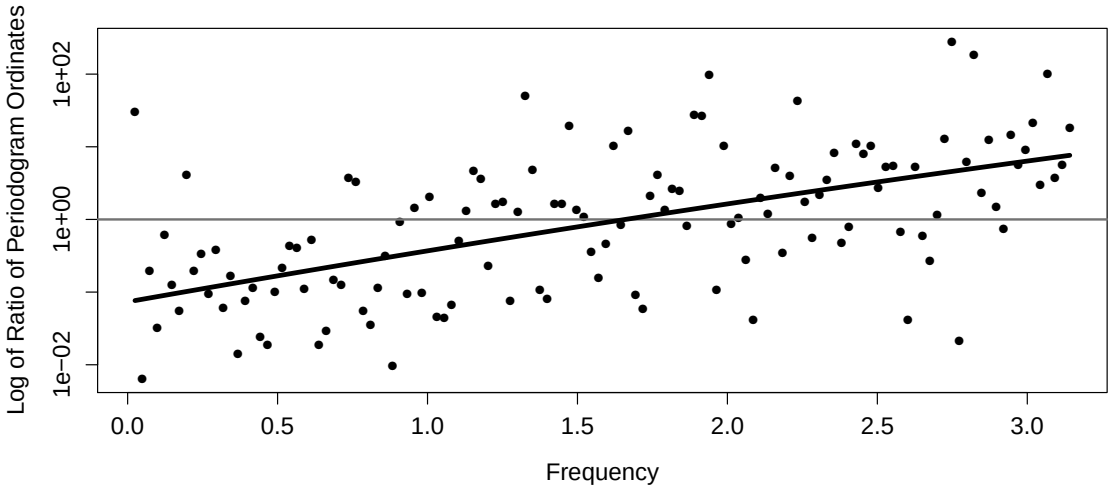
A shape test based on the count of individual ratio values outside critical limits which could be readily implemented graphically was initially investigated for size and power. However the test proved to have low power (which is perhaps not surprising given its similarity to a non-parametric test) and was deemed unsuitable. As a substitute the likelihood ratio test from Coates and Diggle [1986] using a quadratic approximation to the spectral density is undertaken before plotting the fitted model (quadratic or a horizontal line) overlayed on the plot of logged ratios.

If the null hypothesis of no difference in spectral shape is rejected using Coates and Diggle [1986], then the resulting quadratic is plotted. Figure 3.5.1 demonstrates this outcome for two autoregressive processes of length 256 with parameters $\phi_1 = 0.5$ and $\phi_1 = -0.5$ respectively. The shape of the curve (and the position of the plotted individual ratios) suggests the form of the relationship between the spectra of the two processes. The graphing proceeds no further.

If the processes are shown to have the same spectral shape via the likelihood ratio test (that is, the null hypothesis is not rejected), then, instead of the quadratic, a line of the simple mean of the logged ratios is overlayed on the ratios plot and a graphical test of significance of $\alpha = 1$ is constructed. This employs parallel horizontal lines which are a confidence interval around the sample mean. The Central Limit Theorem is used to devise the confidence limits (see Section 3.3.7). Figure 3.5.2 demonstrates this approach for two simulated AR(2) processes with common parameter ($\phi_1 = 0.25$) and innovations variances of 1 and 2.25. The plot was produced because the test from Coates and Diggle [1986] didn't reject the null hypothesis of a common spectral shape. If the confidence limits do not encompass

one as is the case with Figure 3.5.2 , then it is concluded that $\alpha \neq 1$.

Figure 3.5.1 – Plot of the Logged Ratios for the Shape Test. The processes employed (AR($\phi_1 = 0.5$ v $\phi_1 = -0.5$)) reflecting $\alpha = 1$ and $n = 256$ with the plot showing a fitted quadratic (after rejecting the null hypothesis of a common shape using the likelihood ratio test from Coates and Diggle [1986]). The reference line is at $\alpha = 1$.

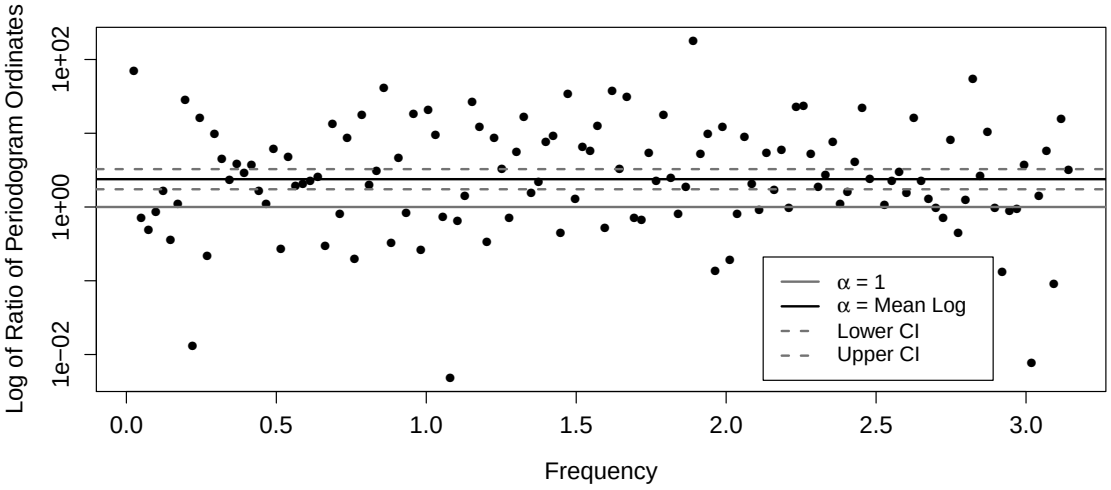


The use of the quadratic test from Coates and Diggle [1986] is not to suggest that the variance, log difference and autocorrelation tests proposed in this thesis do not have merit. However the quadratic likelihood ratio test is arguably more suited to a graphical presentation by indicating the form of the relationship between the two spectral densities.

Finally if the processes are shown to have at least the same spectral shape a sample autocorrelation and partial autocorrelation plot of the combined data can be constructed (Figure 3.5.3). The construction involves finding the sample autocorrelations and partial autocorrelations for each series separately and then averaging each of them by their respective lags. Given that the two series are assumed independent the standard errors of the averaged correlations under the assumption of two white noise processes can be readily calculated using $\frac{1}{2}\sqrt{\frac{1}{n} + \frac{1}{n}} = \sqrt{\frac{1}{2n}}$ (for identification purposes, white noise is assumed for the confidence limits). Empirical identification of the order of the ARMA process proceeds as with the Box-Jenkins methodology.

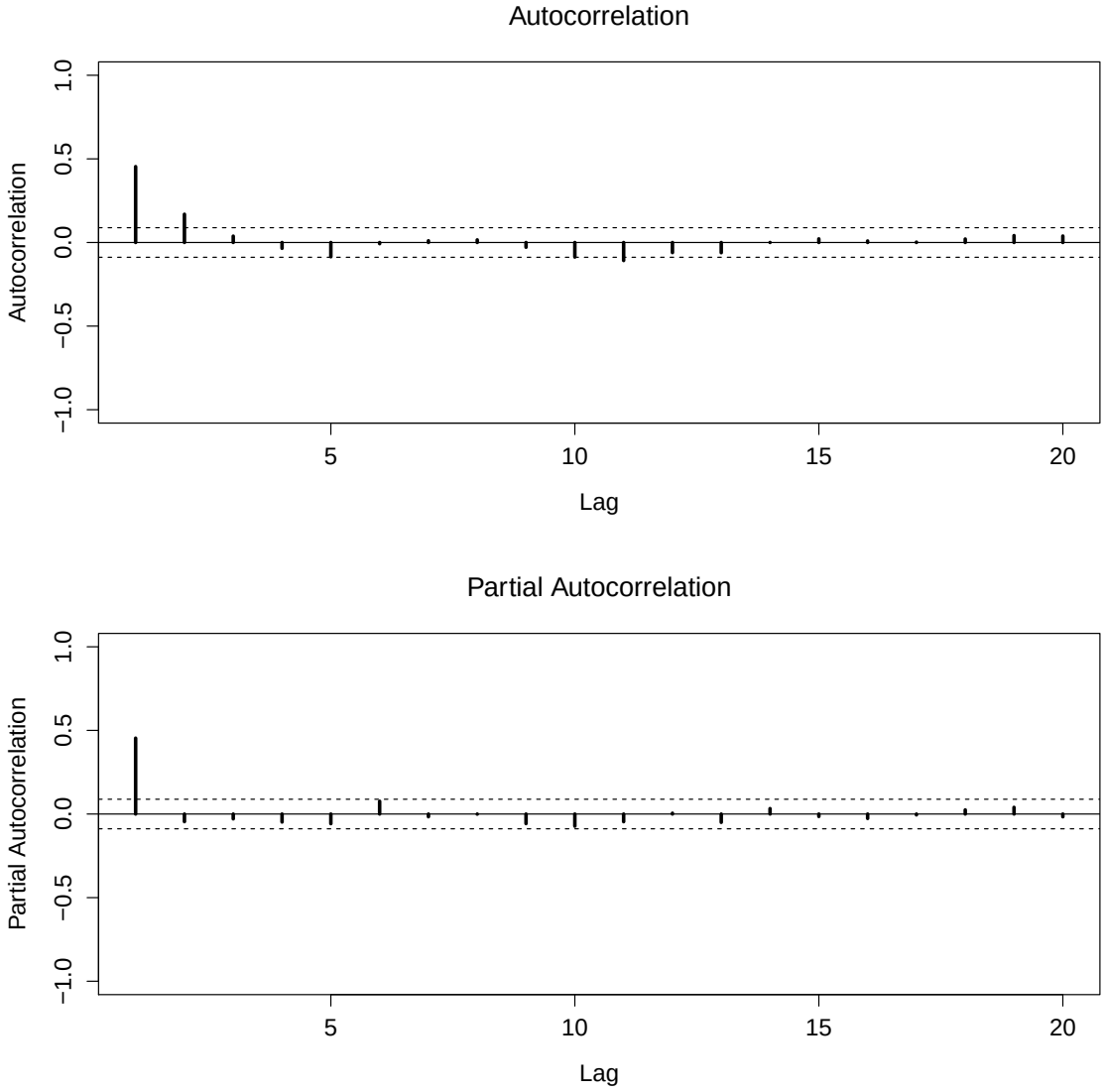
The correlation plots could also be used for shape testing. In the shape testing

Figure 3.5.2 – Plot of the Logged Ratios for the Graphical Scale Test. Each process with $\alpha = 2.25$ and $n = 256$ used a common AR($\phi_1 = 0.5$) model (after accepting the null hypothesis of a common shape using the likelihood ratio test from Coates and Diggle [1986]). There are reference lines at $\alpha = 1$, at the mean logged ratio and at the 95% confidence limits.



the two sets (that is, from the two series) of autocorrelation and partial autocorrelation values are plotted by lag. They have confidence intervals constructed for each value and overlayed on the plots such that the confidence intervals only overlap if the respective autocorrelation and partial autocorrelation are different. This will require some development of the theory of the lag-by-lag confidence intervals beyond the material in this thesis but would present a similar graphical style to that used for ARMA identification.

Figure 3.5.3 – Merged Sample Autocorrelation and Partial Autocorrelation Functions for (AR($\phi_1 = 0.5$ v $\phi_1 = 0.5$)) Processes with $\alpha = 2.25$, $n = 256$ and White Noise 95% Confidence Intervals.



Chapter 4

Maximum Likelihood Estimation

In this chapter the unconditional maximum likelihood estimation of the parameters of single-series ARMA (for reference and comparison), RARMA, AIARMA and CAIARMA processes, all with extraneous variables for maximum model flexibility, is explored. Unconditional maximum likelihood estimation is chosen because of the relative ease of implementation with current computer programs (see Section 4.1.2) and because of the typically superior properties of its estimates (see Chatfield [2003] p. 65) under standard conditions.

Three methods of calculating the maximum likelihood estimates are presented: two joint likelihood algorithms (that is, with and without stationarity and invertibility constraints) and an interleaving approach. The comparison allows assessment of the stability of the three estimation methods. The maximum likelihood estimates and their associated asymptotic distributions are derived.

4.1 Single-Series ARMA Process

Firstly some known results for single-series ARMA model estimation (with extraneous variables) will be stated that will be used later in this chapter. In addition the presentation of these single-series results allows immediate comparison with new results for RARMA, AIARMA and CAIARMA processes.

Let $\boldsymbol{y} = \{y_t\}_{t=1}^{n}$ be a single-series ARMA process, whose t^{th} entry has expected value, $\boldsymbol{\beta}^\top \boldsymbol{z}_t$ (where $\boldsymbol{z}_t = (z_{i,t}, \ldots, z_{k,t})^\top$ is the vector of extraneous variables at

time, t), with ARMA parameters, $(\boldsymbol{\phi}, \boldsymbol{\theta})$, and with innovations variance, σ_ϵ^2. That is, the process has ARMA and regression coefficients,

$$\boldsymbol{\Lambda} = (\boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\beta}^\top)^\top = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \beta_1, \ldots, \beta_k)^\top,$$

and the vector of all parameters is $\boldsymbol{\varsigma} = \left(\boldsymbol{\Lambda}^\top, \sigma_\epsilon^2\right)^\top$. The unconditional (full) likelihood for $\boldsymbol{y}$ given $\boldsymbol{z} = \{\boldsymbol{z}_t\}_{t=1}^n$, is ,

$$\begin{aligned} L_{ARMA}(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z}, \sigma_\epsilon^2) &= (2\pi\sigma_\epsilon^2)^{-n/2} \left[\prod_{t=1}^n r_t(\boldsymbol{\phi}, \boldsymbol{\theta})\right]^{-1/2} \\ &\quad \times \exp\left[-\frac{S(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z})}{2\sigma_\epsilon^2}\right] \end{aligned} \tag{4.1.1}$$

(see Shumway and Stoffer [2011] p. 127[1]) where $r_t(\boldsymbol{\phi}, \boldsymbol{\theta})$ are functions of $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ only, and

$$S(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z}) = \sum_{t=1}^n \left[\frac{((y_t - \boldsymbol{\beta}^\top \mathbf{z}_t) - (y_t - \boldsymbol{\beta}^\top \mathbf{z}_t)^{t-1})^2}{r_t(\boldsymbol{\phi}, \boldsymbol{\theta})}\right],$$

where $\left(y_t - \boldsymbol{\beta}^\top \mathbf{z}_t\right)^{t-1}$ is the conditional mean of $(y_t - \boldsymbol{\beta}^\top \mathbf{z}_t)$ given $((y_{t-1} - \boldsymbol{\beta}^\top \mathbf{z}_{t-1}), \ldots, (y_1 - \boldsymbol{\beta}^\top \mathbf{z}_1))$ and is a linear function of $\left\{y_s - \boldsymbol{\beta}^\top \mathbf{z}_s\right\}_{s=1}^{t-1}$ only (with no additional intercept term) and does not involve σ_ϵ^2. So the log likelihood is,

$$\begin{aligned} LL_{ARMA}(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z}, \sigma_\epsilon^2) &= -\frac{n}{2}\log(2\pi\sigma_\epsilon^2) - \frac{1}{2}\sum_{t=1}^n \log\, r_t(\boldsymbol{\phi}, \boldsymbol{\theta}) \\ &\quad - \frac{S(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z})}{2\sigma_\epsilon^2}. \end{aligned} \tag{4.1.2}$$

The unconditional maximum likelihood estimate of σ_ϵ^2 is

$$\hat{\sigma}_\epsilon^2 = \frac{S(\boldsymbol{y}, \hat{\boldsymbol{\Lambda}}, \boldsymbol{z})}{n}, \tag{4.1.3}$$

where $\hat{\boldsymbol{\Lambda}}$ is the unconditional maximum likelihood estimate of $\boldsymbol{\Lambda}$. Given this, equation (4.1.1) becomes the "concentrated" likelihood,

[1] When those authors state that the $r_t(\boldsymbol{\phi}, \boldsymbol{\theta})$ are "functions only of the regression parameters", they are actually referring to any parameters that are neither the innovations variance nor the parameters for the extraneous variables, as shown in their expressions for the $r_t(\boldsymbol{\phi}, \boldsymbol{\theta})$.

$$CL_{ARMA}(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z}) = (2\pi \frac{S(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z})}{n})^{-n/2} \left[\prod_{t=1}^{n} r_t(\boldsymbol{\phi}, \boldsymbol{\theta})\right]^{-1/2} \exp\left[-\frac{n}{2}\right]$$

The concentrated log likelihood is then,

$$\begin{aligned} CLL_{ARMA}(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z}) &= -\frac{n}{2}(\log 2\pi + 1) - \frac{n}{2} \log \frac{S(\boldsymbol{y}, \boldsymbol{\Lambda}, \boldsymbol{z})}{n} \\ &\quad -\frac{1}{2} \sum_{t=1}^{n} \log r_t(\boldsymbol{\phi}, \boldsymbol{\theta}). \end{aligned} \tag{4.1.4}$$

This equation can be differentiated with respect to Λ and set equal to zero to find (numerically) the maximum likelihood estimates. Alternatively it can be maximised directly using a purely numerical approach which is adopted here. The maximum likelihood estimate of σ_ϵ^2 is then given by (4.1.3). Estimated asymptotic standard errors of $\hat{\boldsymbol{\Lambda}}$ and $\hat{\sigma}_\epsilon^2$ can be obtained in the usual way from the square root of the diagonals of the inverse of the negative Hessian matrix of the full log likelihood at $\hat{\boldsymbol{\Lambda}}$ and $\hat{\sigma}_\epsilon^2$, evaluated either analytically or numerically (following Efron and Hinkley [1978]).

4.1.1 Re-Parameterisation to Ensure Stationarity and Invertibility

Ideally the parameters of the ARMA models returned from the fitting routines should be constrained to be stationary and invertible given that this constraint forms part of the model definitions for all models in this thesis. To ensure stationarity the AR coefficients can be reparameterised before optimisation using Jones [1980] (see also Hamilton [1994] pp. 146-148, Kitagawa [2010] p. 88 and 154, and Prado and West [2010] pp. 75-76). The moving average coefficients are similarly constrained to lie in their invertibility region and the general constraining approach is set out below.

For convenience in optimisation the reparameterisation for stationarity in this thesis uses partial autoregressive autocorrelation coefficients (see Kitagawa [2010] p. 88). For a stationary process the partial autocorrelation coefficients must all be (independently) between -1 and 1 thereby providing a set of simple unrelated

constraints for each partial autocorrelation coefficient. The latter are then functionally related to the autoregressive coefficients by the standard relationships (see Kitagawa [2010] p. 88). To completely remove the constraints on the range of optimising variables, the p partial autocorrelation coefficients $\{\pi_i\}_{i=1}^{p}$ can be derived from p unconstrained real variables, $\{\chi_i\}_{i=1}^{p}$ using say,

$$\pi_i \quad = \quad \frac{e^{\chi_i} - 1}{e^{\chi_i} + 1}.$$

This reparameterisation can be applied separately to the seasonal and nonseasonal components of the AR (and MA - see Jones [1980], p.393) coefficients.

4.1.2 Use of R's `arima` Function

R's `arima` function can be employed to derive the log likelihood and concentrated log likelihood of a single-series ARMA process with and without constrained (to lie in the stationary and invertible region) estimates obviating the need to create much in the way of bespoke computer code. This approach is set out in Section 4.2.2 as an application of `arima` to model multiple independent sets of univariate stationary and invertible time series from the same process (that is, a RARMA process).

4.1.3 Asymptotic Distribution of the Maximum Likelihood Estimates

The asymptotic distribution of parameter estimates can be very useful in understanding the performance of estimators especially if the distribution is shown to hold for smaller sample sizes. The asymptotic distribution of the unconditional maximum likelihood estimates, $\hat{\boldsymbol{\varsigma}}$, of the single-series ARMA parameters, $\boldsymbol{\varsigma}$, is,

$$(\hat{\boldsymbol{\varsigma}} - \boldsymbol{\varsigma}) \quad \sim \quad N(\mathbf{0}, \frac{\sigma_\epsilon^2}{n}\boldsymbol{\Omega}_{ARMA})$$

(see Pierce [1971], Hannan [1973], Brockwell and Davis [1991], p.258, Baillie [Forthcoming] and Professor R. Baillie, pers. comm., 23rd January 2017) where

$$\boldsymbol{\Omega}_{ARMA} = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \boldsymbol{C} & \boldsymbol{D} \\ \boldsymbol{D}^\top & \boldsymbol{F} \end{bmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \frac{1}{2}\sigma_\epsilon^{-2} \end{bmatrix}^{-1} \tag{4.1.5}$$

and $\boldsymbol{C} = \{c_{l,j}\} = \left\{\gamma^u_{l-j}\right\}_{l,j}$ and $\boldsymbol{F} = \{f_{l,j}\} = \left\{\gamma^v_{l-j}\right\}_{l,j}$ are the $p \times p$ and $q \times q$ autocovariance matrices for $\{u_t\}$ and $\{v_t\}$ with $\phi(B)u_t = \omega_t$ and $\theta(B)v_t = \omega_t$. Also $\{\omega_t\}$ is an independent identically-distributed time series with variance, σ^2_ϵ (both the $\phi(B)$ and $\theta(B)$ filters are applied to the left-hand side of the difference equation). Finally $\boldsymbol{D} = \{d_{l,j}\} = \left\{\gamma^{uv}_{l-j}\right\}_{l,j}$ is the $p \times q$ cross-covariance matrix between $\{u_t\}$ and $\{v_t\}$ and

$$\boldsymbol{G} = \{g_{l,j}\} = \left\{\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \left(\vartheta_{l,t}\vartheta_{j,t}\right),\ l,j = 1,\ldots,k\right\},$$

where $\vartheta_{l,t} = \theta^{-1}(B)\phi(B)z_{l,t}$ [2], $l = 1,\ldots,k$. Here the limit is used instead of expectation because strictly the elements of the vector series, $\boldsymbol{z} = \{\boldsymbol{z}_t\}_{t=1}^n, l = 1,\ldots,k$, are not random variables.

It is well known that

$$\begin{bmatrix} \boldsymbol{C} & \boldsymbol{D} \\ \boldsymbol{D}^\top & \boldsymbol{F} \end{bmatrix} \sigma_\epsilon^{-2} = \begin{bmatrix} \boldsymbol{C}_* & \boldsymbol{D}_* \\ \boldsymbol{D}_*^\top & \boldsymbol{F}_* \end{bmatrix} \tag{4.1.6}$$

is a matrix in $(\boldsymbol{\phi}, \boldsymbol{\theta})$ only and does not depend on σ^2_ϵ. That is,

$$\begin{aligned} \frac{\sigma^2_\epsilon}{n}\boldsymbol{\Omega}_{ARMA} &= \frac{1}{n}\begin{bmatrix} \boldsymbol{G}/\sigma_\epsilon^{-2} & 0 & 0 \\ 0 & \begin{bmatrix} \boldsymbol{C}_* & \boldsymbol{D}_* \\ \boldsymbol{D}_*^\top & \boldsymbol{F}_* \end{bmatrix} & 0 \\ 0 & 0 & \frac{1}{2}\sigma_\epsilon^{-4} \end{bmatrix}^{-1} \\ &= \frac{1}{n}\begin{bmatrix} \boldsymbol{G}^{-1}\sigma^2_\epsilon & 0 & 0 \\ 0 & \begin{bmatrix} \boldsymbol{C}_* & \boldsymbol{D}_* \\ \boldsymbol{D}_*^\top & \boldsymbol{F}_* \end{bmatrix}^{-1} & 0 \\ 0 & 0 & 2\sigma^4_\epsilon \end{bmatrix}. \end{aligned} \tag{4.1.7}$$

[2] Note that $\phi(B)\theta^{-1}(B) = \theta^{-1}(B)\phi(B)$.

In this thesis the variance of the asymptotic distribution will be referred to as the asymptotic variance.

Therefore (for later use) the expected value of the negative of the Hessian matrix (being the inverse of the asymptotic variance matrix) is,

$$E(-\boldsymbol{H}_{ARMA}) = \frac{n}{\sigma_\epsilon^2}\boldsymbol{\Omega}_{ARMA}^{-1}. \tag{4.1.8}$$

In this thesis, a sequence of random variables, (X_n, Y_n), are said to be "asymptotically independent" if (X_n, Y_n) are possibly dependent for all finite n, but converge in distribution to some random variables (X, Y) which are independent. If (X_n, Y_n) converge in distribution to bivariate normal random variables, (X, Y), with covariance of zero then this implies that (X, Y) are independent and hence (X_n, Y_n) are asymptotically independent. Hence, the fact that, in $\boldsymbol{\Omega}_{ARMA}$ and except for the diagonal matrices and scalars, the last rows and columns are all zero implies that $\hat{\sigma}_\epsilon^2$ is asymptotically independent of $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ which are also asymptotically independent of $\hat{\boldsymbol{\beta}}$.

4.2 Replicated ARMA (RARMA) Process

In this thesis, for RARMA time series, there are three approaches to unconditional maximum likelihood model fitting, joint likelihood with and without stationarity and invertibility constraints (see Section 4.1.1) and interleaving. This section explores these three approaches and derives the asymptotic distribution of the maximum likelihood parameter estimates.

4.2.1 Joint Likelihood Approach

The unconditional likelihood for a RARMA process with m series, $\boldsymbol{Y} = (\boldsymbol{y}^{(1)} \cup \ldots \cup \boldsymbol{y}^{(m)})$ (where "$\cup$" is the union operator) with sample sizes, $\{n_i\}_{i=1}^m$, given m extraneous vector time series, $\boldsymbol{Z} = \left\{\boldsymbol{z}^{(1)} \cup \ldots \cup \boldsymbol{z}^{(m)}\right\}$, can be expressed as the following joint likelihood of all the series (see 4.1.1),

$$
\begin{aligned}
L_{RARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \sigma_\epsilon^2) &= \prod_{i=1}^{m} L_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}, \sigma_\epsilon^2) && (4.2.1)\\
&= \prod_{i=1}^{m} \left\{ (2\pi\sigma_\epsilon^2)^{-n_i/2} \left[\prod_{t=1}^{n_i} r_t(\boldsymbol{\phi}, \boldsymbol{\theta}) \right]^{-1/2} \exp\left[-\frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}))}{2\sigma_\epsilon^2} \right] \right\}\\
&= (2\pi\sigma_\epsilon^2)^{-\sum_{i=1}^{m} n_i/2} \prod_{i=1}^{m} \left(\left[\prod_{t=1}^{n_i} r_t(\boldsymbol{\phi}, \boldsymbol{\theta}) \right]^{-1/2} \right)\\
&\qquad \times \exp\left[-\left(\frac{\sum_{i=1}^{m} S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}))}{2\sigma_\epsilon^2} \right) \right]. && (4.2.2)
\end{aligned}
$$

So the joint log likelihood is

$$
\begin{aligned}
LL_{RARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \sigma_\epsilon^2) &= -\left(\frac{\sum_{i=1}^{m} n_i}{2} \right) \log(2\pi) - \left(\frac{\sum_{i=1}^{m} n_i}{2} \right) \log\left(\sigma_\epsilon^2\right)\\
&\quad -\frac{1}{2} \sum_{i=1}^{m} \sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta}))\\
&\quad - \left(\frac{\sum_{i=1}^{m} S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{2\sigma_\epsilon^2} \right). && (4.2.3)
\end{aligned}
$$

If the first derivative with respect to σ_ϵ^2 is taken and the result set to zero this gives,

$$
\hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^{m} S(\boldsymbol{y}^{(i)}, \hat{\boldsymbol{\Lambda}}, \boldsymbol{z}^{(i)})}{\sum_{i=1}^{m} n_i}, \qquad (4.2.4)
$$

where $\hat{\boldsymbol{\Lambda}}$ is the unconditional maximum likelihood estimate of $\boldsymbol{\Lambda}$. This parallels the result for one series (see (4.1.3)). Substituting this equality into equation (4.2.3) gives the concentrated log likelihood,

$$
\begin{aligned}
CLL_{RARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}) &= -\frac{\sum_{i=1}^{m} n_i}{2} \log(2\pi) - \frac{\sum_{i=1}^{m} n_i}{2} \log\left(\frac{\sum_{i=1}^{m} S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{\sum_{i=1}^{m} n_i}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})) - \left(\frac{\sum_{i=1}^{m} n_i}{2}\right) \\
&= -\frac{\sum_{i=1}^{m} n_i}{2} (\log(2\pi) + 1) \\
&\quad -\frac{\sum_{i=1}^{m} n_i}{2} \log\left(\frac{\sum_{i=1}^{m} S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{\sum_{i=1}^{m} n_i}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})). \qquad (4.2.5)
\end{aligned}
$$

This can be numerically maximised to derive the maximum likelihood values, $\hat{\boldsymbol{\Lambda}}$. The maximum likelihood estimate of σ_ϵ^2 is obtained from (4.2.4). Estimated asymptotic standard errors of $\hat{\boldsymbol{\Lambda}}$ and $\hat{\sigma}_\epsilon^2$ can be derived from the Hessian matrix of the full log likelihood at $\hat{\boldsymbol{\Lambda}}$ and $\hat{\sigma}_\epsilon^2$.

4.2.2 Use of R's `arima` Function

The above process can be facilitated by use of the `arima` function in R rather than requiring the writing of fully bespoke code. The single-series modelling of the ARMA process using this approach can be obtained by setting $m = 1$ in the following.

If all values of $\boldsymbol{\Lambda}$ are fixed in calling `arima` then the function returns the associated concentrated unconditional log likelihood value (see (4.1.4)) along with the innovations variance evaluated at the specified $\boldsymbol{\Lambda}$ values (see (4.1.3)). Using (4.1.3) and (4.1.4) the components of this result can then be extracted and used to optimise the RARMA concentrated joint likelihood (see below), with respect to $\boldsymbol{\Lambda}$ leading to the unconditional maximum likelihood estimate of the innovations variances. After optimisation these resultant estimates can be employed in `arima` to derive the surface of the joint log likelihood around the maximum likelihood estimates for evaluating the Hessian matrix and the estimated variance matrix.

Specifically, the summation of the m concentrated log likelihoods from the m calls to `arima` with fixed $\boldsymbol{\Lambda}$, (designated as $CLL_{RARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z})^*$), gives (see

(4.1.4)),

$$CLL_{RARMA}(\boldsymbol{Y},\boldsymbol{\Lambda},\boldsymbol{Z})^* = -\frac{\sum_{i=1}^m n_i}{2}\left(\log(2\pi)+1\right) - \sum_{i=1}^m\left(\frac{n_i}{2}\log\left(\frac{S(\boldsymbol{y}^{(i)},\boldsymbol{\Lambda},\boldsymbol{z}^{(i)})}{n_i}\right)\right) - \frac{1}{2}\sum_{i=1}^m\sum_{t=1}^{n_i}\log(r_t(\boldsymbol{\phi},\boldsymbol{\theta})). \quad (4.2.6)$$

Hence, the concentrated log likelihood for a RARMA process can be derived via,

$$\begin{aligned} CLL_{RARMA}(\boldsymbol{Y},\boldsymbol{\Lambda},\boldsymbol{Z}) &= CLL_{RARMA}(\boldsymbol{Y},\boldsymbol{\Lambda},\boldsymbol{Z})^* \\ &\quad + \sum_{i=1}^m\left(\frac{n_i}{2}\log\left(\frac{S(\boldsymbol{y}^{(i)},\boldsymbol{\Lambda},\boldsymbol{z}^{(i)})}{n_i}\right)\right) \\ &\quad - \left(\frac{\sum_{i=1}^m n_i}{2}\right)\log\left(\frac{\sum_{i=1}^m S(\boldsymbol{y}^{(i)},\boldsymbol{\Lambda},\boldsymbol{z}^{(i)})}{\sum_{i=1}^m n_i}\right). \end{aligned}$$

Accordingly this correction can be applied to the sum of the m concentrated log likelihood results from the m calls to `arima` to derive the concentrated log likelihood for the RARMA process, that is, with a common innovations variance. This likelihood can then be optimised with respect to $\boldsymbol{\Lambda}$ whereby `arima` provides the values of $\check{\sigma}_\epsilon^2(i) = S(\boldsymbol{y}^{(i)},\hat{\boldsymbol{\Lambda}},\boldsymbol{z}^{(i)})/n_i$, i=1,..,m, for each of the m series (see (4.1.3)) given the optimal $\hat{\boldsymbol{\Lambda}}$. The $\left\{S(\boldsymbol{y}^{(i)},\hat{\boldsymbol{\Lambda}},\boldsymbol{z}^{(i)})\right\}_{i=1}^m$ values can be obtained via the outputted $\{\check{\sigma}_\epsilon^2(i)\}_{i=1}^m$estimates. The RARMA maximum likelihood estimate of $\hat{\sigma}_\epsilon^2$ is

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^m n_i\check{\sigma}_\epsilon^2(i)}{\sum_{i=1}^m n_i} \quad (4.2.7)$$

using (4.2.4), that is, it is the weighted sum of the maximum likelihood estimates by series assuming a common $\boldsymbol{\Lambda}$.

Finally repeated calls to `arima` across all series centred at the maximum likelihood values allows numerical estimation of the Hessian matrix using as the log likelihood,

$$\begin{aligned} LL_{RARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \sigma_\epsilon^2) &= CLL_{RARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}) + \frac{\sum_{i=1}^m n_i}{2} \\ &\quad + \frac{\sum_{i=1}^m n_i}{2} \log\left(\frac{\sum_{i=1}^m S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{\sum_{i=1}^m n_i}\right) \\ &\quad - \left(\frac{\sum_{i=1}^m S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{2\sigma_\epsilon^2}\right) - \frac{1}{2}\left(\sum_{i=1}^m n_i\right) \log \sigma_\epsilon^2. \end{aligned}$$

Employing the methods detailed in Section 4.1.1, the AR and MA parameters can be constrained to ensure stationarity and invertibility, if required.

4.2.3 Estimation using an Interleaved Process

A conceptually simpler but equivalent approach to the same maximum likelihood estimation is to use the method of interleaving. Section 2.3.1 shows that two or more RARMA processes of the same length can be represented as a single univariate ARMA process with certain parameters set equal to zero.

If the series are not of equal length then missing values can be added at the end of each series to ensure that all series are of the same length. If the missing values are incorporated in the estimation via maximum likelihood methods (see Jones [1980]) the resulting estimates, $\hat{\boldsymbol{\Lambda}}$ and $\hat{\sigma}_\epsilon^2$, will be maximum likelihood values. The aim here is not to estimate the missing values but to simply account for them in the maximum likelihood estimation, that is, in the likelihood function.

The R function `arima` allows for missing values in unconditional maximum likelihood estimation using the approach of Jones [1980]. In fitting using state space models, Jones [1980] employs an updating of the "state" for each additional time series value and missing values simply do not update the state. The updating system moves on to consider the next series value.

The variance matrix of the estimates can be derived directly from the Hessian of the log likelihood of the interleaved series.

4.2.4 Asymptotic Distribution of the Maximum Likelihood Estimates

For a RARMA process with m series and using $\boldsymbol{\varsigma}$ from Section 4.1 to represent the vector of all parameters, $\boldsymbol{\varsigma} = (\boldsymbol{\Lambda}^\top, \sigma_\epsilon^2)^\top$, the unconditional log likelihood is (from (4.2.1)),

$$LL_{RARMA}(\boldsymbol{Y}, \boldsymbol{\varsigma}, \boldsymbol{Z},) \;=\; \sum_{i=1}^{m} LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\varsigma}, \boldsymbol{z}^{(i)}).$$

Now the asymptotic variance matrix of maximum likelihood estimates is the inverse of the expected value of the negative of the Hessian matrix of the log likelihood (in the current situation, the term "asymptotic" implies "$n_i \to \infty$ for all i"). The Hessian matrix of the above log likelihood is the sum of the Hessian matrices of the components of the aggregate log likelihood. That is, using $\boldsymbol{\varsigma} = (\varsigma_1, \ldots, \varsigma_{k+p+q+1})^\top$ and given

$$\begin{aligned}
\boldsymbol{H}_{RARMA} &= \left\{ \partial^2 \left[\sum_{i=1}^{m} LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\varsigma}, \boldsymbol{z}^{(i)}) \right] / \partial\varsigma_l \partial\varsigma_j \right\} \\
&= \left\{ \sum_{i=1}^{m} \left[\partial^2 LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\varsigma}, \boldsymbol{z}^{(i)}) / \partial\varsigma_l \partial\varsigma_j \right] \right\} \\
&= \sum_{i=1}^{m} \boldsymbol{H}_{ARMA_i},
\end{aligned}$$

where $\boldsymbol{H}_{ARMA_i}$ is the Hessian for the log likelihood for the i^{th} series assuming common parameters with the other $m-1$ series, then,

$$E(-\boldsymbol{H}_{RARMA}) \;=\; E(-\sum_{i=1}^{m} \boldsymbol{H}_{ARMA_i}).$$

From (4.1.8),

$$E\left(-\boldsymbol{H}_{ARMA_i}\right) \;=\; \frac{n_i}{\sigma_\epsilon^2} \boldsymbol{\Omega}_{ARMA_i}^{-1},$$

where $\boldsymbol{\Omega}_{ARMA_i}^{-1}$ is from (4.1.5) and where $\boldsymbol{G}_i$, substituting for $\boldsymbol{G}$ in $\boldsymbol{\Omega}_{ARMA_i}$ (see

(4.1.5)), is evaluated for each set of extraneous time series, $\boldsymbol{z}^{(i)}$, (that is, $\boldsymbol{G}_i$ has elements $\lim_{n_i\to\infty} n_i^{-1}\sum_{i=1}^{n_i}\left(u_{l,i,t}u_{j,i,t}\right), l,j=1,\ldots,k$ where $u_{l,i,t}=\theta^{-1}(B)\phi(B)z_{l,i,t}$ and $z_{l,i,t}$ is the l^{th} variable at the t^{th} time point of the extraneous input vector, $\boldsymbol{z}^{(i)}$).

Hence

$$\begin{aligned}
E(-\boldsymbol{H}_{RARMA}) &= \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{m} n_i\boldsymbol{\Omega}_{ARMA_i}^{-1} \\
&= \frac{1}{\sigma_\epsilon^2}\begin{bmatrix} \sum_{i=1}^m n_i\boldsymbol{G}_i & 0 & 0 \\ 0 & \sum_{i=1}^m n_i\begin{bmatrix}\boldsymbol{C} & \boldsymbol{D}\\ \boldsymbol{D}^\top & \boldsymbol{F}\end{bmatrix} & 0 \\ 0 & 0 & \frac{\sum_{i=1}^m n_i}{2}\sigma_\epsilon^{-2}\end{bmatrix} \\
&= \frac{\sum_{i=1}^m n_i}{\sigma_\epsilon^2}\begin{bmatrix} \sum_{i=1}^m \frac{n_i}{\sum_{i=1}^m n_i}\boldsymbol{G}_i & 0 & 0 \\ 0 & \begin{bmatrix}\boldsymbol{C} & \boldsymbol{D}\\ \boldsymbol{D}^\top & \boldsymbol{F}\end{bmatrix} & 0 \\ 0 & 0 & \frac{1}{2}\sigma_\epsilon^{-2}\end{bmatrix}.
\end{aligned}$$

Subsequently, the asymptotic variance of $\hat{\boldsymbol{\varsigma}}$ is,

$$\begin{aligned}
\text{Var}(\hat{\boldsymbol{\varsigma}}) &= E(-\boldsymbol{H}_{RARMA})^{-1} \\
&= \frac{\sigma_\epsilon^2}{\sum_{i=1}^m n_i}\begin{bmatrix} \left(\sum_{i=1}^m \frac{n_i}{\sum_{i=1}^m n_i}\boldsymbol{G}_i\right)^{-1} & 0 & 0 \\ 0 & \begin{bmatrix}\boldsymbol{C} & \boldsymbol{D}\\ \boldsymbol{D}^\top & \boldsymbol{F}\end{bmatrix}^{-1} & 0 \\ 0 & 0 & 2\sigma_\epsilon^2\end{bmatrix} \\
&= \frac{1}{\sum_{i=1}^m n_i}\begin{bmatrix} \sigma_\epsilon^2\left(\sum_{i=1}^m \frac{n_i}{\sum_{i=1}^m n_i}\boldsymbol{G}_i\right)^{-1} & 0 & 0 \\ 0 & \begin{bmatrix}\boldsymbol{C}_* & \boldsymbol{D}_*\\ \boldsymbol{D}_*^\top & \boldsymbol{F}_*\end{bmatrix}^{-1} & 0 \\ 0 & 0 & 2\sigma_\epsilon^4\end{bmatrix}
\end{aligned}$$

using (4.1.6).

Therefore, noting that the asymptotic distribution of an unconditional maximum likelihood estimate is normal with mean equal to the population parameter, the asymptotic distribution of the maximum likelihood estimates of the parameters

of a RARMA process is,

$$\lim_{n_i \to \infty \forall i} (\hat{\boldsymbol{\varsigma}} - \boldsymbol{\varsigma}) \quad \sim \quad N(\mathbf{0}, Var(\hat{\boldsymbol{\varsigma}})). \tag{4.2.8}$$

Note that the asymptotic condition of $n_i \to \infty$ for all i includes both finite m and $m \to \infty$. However the asymptotic distribution for $m \to \infty$ and finite n_i is not encompassed here and will be pursued in future research.

The above result implies that $\hat{\boldsymbol{\beta}}$ is asymptotically independent of $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ which is asymptotically independent of $\hat{\sigma}_\epsilon^2$.

Also, (4.2.8) implies that the estimates from a RARMA process with m series are asymptotically equivalent to fitting the same model to one series of the same length as the sum of the lengths of the m series. This suggests that the determining factor in the variance of the estimators, at least asymptotically, is the total length of all the series, rather than being primarily determined by either each series's length or by the number of series.

Finally the variance matrix of the regression coefficients of the extraneous vector time series is the inverse of the weighted mean of the cross-product matrices of the (filtered) extraneous time series for each series.

4.3 Almost Identical ARMA (AIARMA) Process

As with the RARMA modelling, there are three approaches to AIARMA model fitting employing the constrained versus unconstrained joint likelihood approaches and the interleaving method. This section shows how to use these procedures and determines the asymptotic distribution of the maximum likelihood AIARMA estimates.

4.3.1 Joint Likelihood Approach

The likelihood for an AIARMA process with m series and sample sizes, $\{n_i\}_{i=1}^m$, can be expressed as the following joint likelihood (see (4.1.1)),

$$
\begin{aligned}
L_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \{\sigma_{\epsilon_i}\}_{i=1}^m) &= \prod_{i=1}^m L_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}, \sigma_{\epsilon_i}) \qquad (4.3.1)\\
&= \prod_{i=1}^m \left[(2\pi\sigma_{\epsilon_i}^2)^{-n_i/2} \left[\prod_{t=1}^{n_i} r_t(\boldsymbol{\phi}, \boldsymbol{\theta}) \right]^{-1/2} \exp\left[-\frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{2\sigma_{\epsilon_i}^2} \right] \right]\\
&= \prod_{i=1}^m \left[(2\pi\sigma_{\epsilon_i}^2)^{-n_i/2} \left[\prod_{t=1}^{n_i} r_t(\boldsymbol{\phi}, \boldsymbol{\theta}) \right]^{-1/2} \right]\\
&\quad \times \left[\exp\left[-\sum_{i=1}^m \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{2\sigma_{\epsilon_i}^2} \right] \right].
\end{aligned}
$$

So the joint log likelihood is

$$
\begin{aligned}
LL_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \{\sigma_{\epsilon_i}\}_{i=1}^m) &= -\left(\frac{\sum_{i=1}^m n_i}{2} \right) \log(2\pi) - \sum_{i=1}^m \left(\frac{n_i}{2} \log\left(\sigma_{\epsilon_i}^2\right) \right)\\
&\quad -\frac{1}{2} \sum_{i=1}^m \sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta}))\\
&\quad -\sum_{i=1}^m \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{2\sigma_{\epsilon_i}^2}. \qquad (4.3.2)
\end{aligned}
$$

If the first derivatives with respect to σ_i^2, $i = 1, \ldots, m$, are taken and the results set to zero this gives,

$$
\hat{\sigma}_i^2 = \frac{S(\boldsymbol{y}^{(i)}, \hat{\boldsymbol{\Lambda}}, \boldsymbol{z}^{(i)})}{n_i},\ i = 1, \ldots, m, \qquad (4.3.3)
$$

where $\hat{\boldsymbol{\Lambda}}$ are the joint maximum likelihood estimates of the elements of $\boldsymbol{\Lambda}$. This again parallels the result for one series (see (4.1.3)), noting that a common set of $\hat{\boldsymbol{\Lambda}}$ is used in the estimation of all $\{\hat{\sigma}_i^2\}_{i=1}^m$. Substituting these equalities into equation (4.3.2) gives the following concentrated log likelihood,

$$
\begin{aligned}
CLL_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}) &= -\frac{\sum_{i=1}^{m} n_i}{2} \log(2\pi) - \sum_{i=1}^{m} \left(\frac{n_i}{2} \log \left(\frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{n_i} \right) \right) \\
&\quad - \frac{1}{2} \sum_{i=1}^{m} \sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})) - \frac{\sum_{i=1}^{m} n_i}{2} \\
&= -\frac{\sum_{i=1}^{m} n_i}{2} (\log(2\pi) + 1) - \sum_{i=1}^{m} \left(\frac{n_i}{2} \log \left(\frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{n_i} \right) \right) \\
&\quad - \frac{1}{2} \sum_{i=1}^{m} \sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})). \qquad (4.3.4)
\end{aligned}
$$

The maximum of (4.3.4) can be found using a numerical optimisation routine. After finding $\hat{\boldsymbol{\Lambda}}$, $\{\hat{\sigma}_i^2\}_{i=1}^m$ can be calculated using (4.3.3). Estimated asymptotic standard errors of $\hat{\boldsymbol{\Lambda}}$ and $\{\hat{\sigma}_i^2\}_{i=1}^m$ can be obtained using the Hessian matrix of the (full) log likelihood (4.3.2) at $\hat{\boldsymbol{\Lambda}}$ and $\{\hat{\sigma}_i^2\}_{i=1}^m$.

4.3.2 Use of R's `arima` Function

The above process can be facilitated by use of the `arima` function in R as employed in RARMA modelling (see Section 4.2.2). If all values of $\boldsymbol{\Lambda}$ are fixed in calling the function for one series then the function returns the concentrated log likelihood value (4.1.4) along with the calculated innovations variance using (4.1.3). This can be used to optimise the joint concentrated log likelihood, (4.3.4), with respect to $\boldsymbol{\Lambda}$ thereby also providing the maximum likelihood estimates of $\{\sigma_i^2\}_{i=1}^m$. These estimates can then be employed in `arima` as shown below to derive the surface of the joint log likelihood around the maximum likelihood estimates for evaluating the Hessian matrix and hence the estimated variance matrix of the estimates.

The AIARMA concentrated log likelihood is simply the sum of the concentrated log likelihoods for each series with common $\boldsymbol{\Lambda}$, that is,

$$
CLL_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}) = \sum_{i=1}^{m} CLL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}).
$$

To evaluate the Hessian, the full log likelihood, $LL_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \{\sigma_{\epsilon_i}\}_{i=1}^m)$, is related to the concentrated log likelihood with common $\boldsymbol{\Lambda}$ (that is, $CLL_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z})$

from (4.3.4)) resulting from m calls to `arima`) using,

$$\begin{aligned} LL_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \left\{\sigma^2_{\epsilon_i}\right\}_{i=1}^m) &= CLL_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}) + \frac{1}{2}\sum_{i=1}^m n_i \\ &\quad + \sum_{i=1}^m \frac{n_i}{2} \log\left(\frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{n_i}\right) \\ &\quad - \sum_{i=1}^m \left(\frac{n_i}{2}\log\sigma^2_{\epsilon_i}\right) - \sum_{i=1}^m \left(\frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{2\sigma^2_{\epsilon_i}}\right). \end{aligned}$$

As with RARMA model fitting, the methods detailed in Section 4.1.1 can be applied to the AR and MA parameters to ensure stationarity and invertibility.

4.3.3 Estimation using a Transformed Interleaved Process

From 4.3.1, the joint log likelihood of AIARMA processes with m series can be expressed as,

$$\begin{aligned} LL_{RARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \left\{\sigma^2_{\epsilon_i}\right\}_{i=1}^m) &= \sum_{i=1}^m LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}, \sigma^2_{\epsilon_i}) \\ &= \sum_{i=1}^m LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}, \tau_i^2\sigma^2_{\epsilon_1}) \\ &= -\frac{n_i}{2}\log(2\pi\tau_i^2\sigma^2_{\epsilon_1}) - \frac{1}{2}\left[\sum_{t=1}^{n_i}\log r_t(\boldsymbol{\phi}, \boldsymbol{\theta})\right] \\ &\quad - \sum_{t=1}^m \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})}{2\tau_i^2\sigma^2_{\epsilon_1}}, \end{aligned}$$

where $\tau_i^2 = \sigma^2_{\epsilon_i}/\sigma^2_{\epsilon_1}$, $i = 2, \ldots, m$ and $\tau_1 = 1$. Now, noting that $S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})$ is a linear function in the squared and cross-product elements of $\left(\boldsymbol{y}^{(i)} - \boldsymbol{\beta}^\top \boldsymbol{z}^{(i)}\right)$ with no intercept (see Section 4.1), the log likelihood for the i^{th} series, $LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}, \tau_i^2\sigma^2_{\epsilon_1}) = LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}, \tau_i^2, \sigma^2_{\epsilon_1})$, can be expressed as,

$$\begin{aligned} LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}, \tau_i^2, \sigma_{\epsilon_1}^2) &= -\frac{n_i}{2}\log(2\pi) - \frac{n_i}{2}\log\tau_i^2 - \frac{n_i}{2}\log\sigma_{\epsilon_1}^2 \\ &\quad -\frac{1}{2}\left[\sum_{t=1}^{n_i}\log r_t(\boldsymbol{\phi},\boldsymbol{\theta})\right] - \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)})/\tau_i^2}{2\sigma_{\epsilon_1}^2} \\ &= -\frac{n_i}{2}\log(2\pi) - \frac{n_i}{2}\log\tau_i^2 - \frac{n_i}{2}\log\sigma_{\epsilon_1}^2 \\ &\quad -\frac{1}{2}\left[\sum_{t=1}^{n_i}\log r_t(\boldsymbol{\phi},\boldsymbol{\theta})\right] - \frac{S(\boldsymbol{y}^{(i)}/\tau_i, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}/\tau_i)}{2\sigma_{\epsilon_1}^2} \\ &= LL_{ARMA}(\boldsymbol{y}^{(i)}/\tau_i, \boldsymbol{\Lambda}, z^{(i)}/\tau_i, \sigma_{\epsilon_1}^2) - n_i\,\log\tau_i. \quad (4.3.5) \end{aligned}$$

This gives,

$$\begin{aligned} LL_{RARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \left\{\sigma_{\epsilon_i}^2\right\}_{i=1}^m) &= \sum_{i=1}^{n_i}[LL_{ARMA}\left(\boldsymbol{y}^{(i)}/\tau_i, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}/\tau_i, \sigma_{\epsilon_1}^2\right) \\ &\quad -n_i\log(\tau_i)] \\ &= LL_{ARMA}(\boldsymbol{U}, \boldsymbol{\Lambda}, \boldsymbol{W}, \sigma_{\epsilon_1}^2) \\ &\quad -\sum_{i=2}^{m}\left(n_i\,\log\tau_i\right), \qquad (4.3.6) \end{aligned}$$

where $\boldsymbol{u}^{(i)} = \boldsymbol{y}^{(i)}/\tau_i$, $\boldsymbol{w}^{(i)} = \boldsymbol{z}^{(i)}/\tau_i$, $i = 1, \ldots, m$, $\boldsymbol{U} = (\boldsymbol{u}^{(1)} \cup \ldots \cup \boldsymbol{u}^{(m)})$ and $\boldsymbol{W} = (\boldsymbol{w}^{(1)} \cup \ldots \cup \boldsymbol{w}^{(m)})$.

This shows that the joint log likelihood of the m AIARMA series (that is, with different innovations variances) can be represented as the sum of the log likelihoods of m replicated series from a RARMA process minus the sum of the length of each series times the log of the ratio of the innovations variance. The last term in (4.3.6) only involves parameters, $\{\tau_i\}_{i=2}^m$, so the maximisation of the joint log likelihood for the transformed $\boldsymbol{U}$ and $\mathbf{W}$ with respect to $\boldsymbol{\Lambda}$ and σ_1^2 conditional on $\{\tau_i\}_{i=2}^m$ only involves the term, $LL_{ARMA}(\boldsymbol{U}, \boldsymbol{\Lambda}, \boldsymbol{W}, \sigma_{\epsilon_1}^2)$, in (4.3.6) (which, as noted, is the joint log likelihood of a RARMA process). Moreover m time series assumed to be from a RARMA process can be represented as an interleaved series.

Hence to estimate the parameters using maximum likelihood via interleaving, the following algorithm is used:

- re-scale the last $(m-1)$ series by dividing the i^{th} series and the i^{th} set of

extraneous variables by the respective elements of $\{\tilde{\tau}_i\}_{i=2}^m$, an initial guess of the maximum likelihood values, $\{\hat{\tau}_i\}_{i=2}^m$,

- interleave the m series to create one series after padding each series with missing values to the same length as the longest (say, the first) series,
- find the maximum likelihood estimates of the ARMA coefficients $(\boldsymbol{\phi}, \boldsymbol{\theta})$, of the regression coefficients $(\boldsymbol{\beta})$ and of σ_{ϵ_1} with the resultant log likelihood, given $\{\tilde{\tau}_i\}_{i=2}^m$,
- correct the log likelihood by subtracting $\sum_{i=2}^m n_i \log \tilde{\tau}_i$, and
- optimise the resultant log likelihood with respect to $\{\tilde{\tau}_i\}_{i=2}^m$ (using say R's `optim` function) which will include redoing the four steps above.

Hence after rescaling, interleaving and fitting via maximum likelihood (with the constraints imposed by interleaving) the resulting parameters will be maximum likelihood values with respect to $(\boldsymbol{\Lambda}, \sigma_1^2, \{\tau_i\}_{i=2}^m))$.

The standard errors of the estimates of $\boldsymbol{\Lambda}$ and $\{\tau_i\}_{i=2}^m$ can be obtained by using the Hessian matrix of the joint log likelihood.

In summary, the above method can be used to determine the unconditional maximum likelihood estimates of all coefficients of the AIARMA model including $\{\tau_i\}_{i=2}^m$. The variance matrix of the parameter estimates can be determined from a numerical approximation to the slope of the log likelihood around the maximum likelihood estimates.

4.3.4 Asymptotic Distribution of the Maximum Likelihood Estimates

The unconditional joint log likelihood for an AIARMA process with m series is,

$$LL_{AIARMA}(\boldsymbol{Y}, \boldsymbol{\Lambda}, \boldsymbol{Z}, \left\{\sigma_{\epsilon_i}^2\right\}_{i=1}^m) \;=\; \sum_{i=1}^m LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\Lambda}, \boldsymbol{z}^{(i)}, \sigma_{\epsilon_i}^2). \quad (4.3.7)$$

The asymptotic variance matrix of the maximum likelihood estimates is the inverse of the expected value of the negative of the Hessian matrix of the log likelihood. The expected value, in the case of the above log likelihood, becomes the sum

of the expected values of the negative of the Hessian matrices of the components of the aggregate log likelihood. That is, letting $\boldsymbol{\varphi} = (\boldsymbol{\Lambda}^\top, \{\sigma^2_{\epsilon_i}\}_{i=1}^m)^\top = \{\varphi_j\}$, the Hessian matrix for the AIARMA process can be expressed as,

$$\begin{aligned}
\boldsymbol{H}_{AIARMA}(\boldsymbol{\varphi}) &= \left\{ \partial^2 \left[\sum_{i=1}^{m} LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\varphi}, \boldsymbol{z}^{(i)}) \right] / \partial\varphi_l \partial\varphi_j \right\} \\
&= \left\{ \sum_{i=1}^{m} \partial^2 LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\varphi}, \boldsymbol{z}^{(i)}) / \partial\varphi_l \partial\varphi_j \right\} \\
&= \sum_{i=1}^{m} \boldsymbol{H}_{ARMA_i}(\boldsymbol{\varphi}),
\end{aligned}$$

where here $\boldsymbol{H}_{ARMA_i}(\boldsymbol{\varphi}) = \left\{ \partial^2 LL_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\varphi}, \boldsymbol{z}^{(i)}) / \partial\varphi_l \partial\varphi_j \right\}$ is the Hessian matrix for the ARMA log likelihood for the i^{th} series with unique innovations variance. The partial derivatives for each of the m Hessians are with respect to all the parameters, $\left\{\sigma^2_{\epsilon_i}\right\}_{i=1}^m$, even if $m-1$ of the parameters are not actually present in the respective log likelihood. The respective elements of each Hessian are therefore zero. The expected negative AIARMA Hessian becomes,

$$E(-\boldsymbol{H}_{AIARMA}(\boldsymbol{\varphi})) = \sum_{i=1}^{m} E(-\boldsymbol{H}_{ARMA_i}(\boldsymbol{\varphi})). \tag{4.3.8}$$

For the i^{th} series, the expectation of the negative of the ARMA Hessian matrix from (4.1.8) is,

$$E(-\boldsymbol{H}_{ARMA_i}(\boldsymbol{\varphi})) = \frac{n_i}{\sigma_i^2} \begin{pmatrix} \boldsymbol{G}_i & 0 & 0 \\ 0 & \begin{bmatrix} \boldsymbol{C}_i & \boldsymbol{D}_i \\ \boldsymbol{D}_i^\top & \boldsymbol{F}_i \end{bmatrix} & 0 \\ 0 & 0 & \frac{1}{2} \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \sigma_{\epsilon_i}^{-2} & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \end{pmatrix},$$

where $\boldsymbol{C}_i = \left\{c^i_{l,j}\right\} = \left\{\gamma^{u_i}_{(l-j)}\right\}_{l,j}$ and $\boldsymbol{F}_i = \left\{f^i_{l,j}\right\} = \left\{\gamma^{v_i}_{(l-j)}\right\}_{l,j}$ are the $p \times p$ and $q \times q$

autocovariance matrices for $\{u_{i,t}\}$ and $\{v_{i,t}\}$ with $\phi(B)u_{i,t} = \omega_{i,t}$, $\theta(B)v_{i,t} = \omega_{i,t}$ and $\{\omega_{i,t}\}$ is an independent identically-distributed time series with variance, $\sigma^2_{\epsilon_i}$. Also $\boldsymbol{D}_i = \left\{d^i_{l,j}\right\} = \left\{\gamma^{u_i v_i}_{(l-j)}\right\}_{l,j}$ is the $p \times q$ cross-covariance matrix between $\{u_{i,t}\}$ and $\{v_{i,t}\}$ and $\boldsymbol{G}_i$ has the (l,j) element $\lim_{n_i\to\infty} n_i^{-1} \sum_{i=1}^{n_i} \left(\vartheta_{l,i,t}\vartheta_{j,i,t}\right)$, $l, j = 1, \ldots, k$ where $\vartheta_{i,l,t} = \theta^{-1}(B)\phi(B)z_{l,i,t}$. Again $z_{l,i,t}$ is the l^{th} variable at the t^{th} time point of the extraneous input vector, $\boldsymbol{z}^{(i)}$ (that is, for the i^{th} series). Finally the lower right-hand matrix has the single non-zero value, $\frac{\sigma^{-2}_{\epsilon_i}}{2}$, at location, (i,i).

So, including *all* elements of $\boldsymbol{\varphi}$ including $\left\{\sigma^2_{\epsilon_i}\right\}_{i=1}^m$ in each of the the additive components of the AIARMA log likelihood (whether they are all used in that component or not) and hence in each of the additive components of the associated expected negative Hessian, the overall expected negative Hessian, $E(-\boldsymbol{H}_{AIARMA}(\boldsymbol{\varphi}))$, from (4.3.8) becomes,

$$\begin{pmatrix} \sum_{i=1}^m \left(\frac{n_i}{\sigma_i^2}\boldsymbol{G}_i\right) & 0 & 0 \\ 0 & \sum_{i=1}^m \left(\frac{n_i}{\sigma_i^2}\begin{bmatrix} \boldsymbol{C}_i & \boldsymbol{D}_i \\ \boldsymbol{D}_i^\top & \boldsymbol{F}_i \end{bmatrix}\right) & 0 \\ 0 & 0 & \begin{bmatrix} \frac{n_1}{2\sigma^4_{\epsilon_1}} & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{n_m}{2\sigma^4_{\epsilon_m}} \end{bmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^m \left(\frac{n_i}{\sigma_i^2}\boldsymbol{G}_i\right) & 0 & 0 \\ 0 & \left(\sum_{i=1}^m n_i\right)\begin{bmatrix} \boldsymbol{C}_* & \boldsymbol{D}_* \\ \boldsymbol{D}_*^\top & \boldsymbol{F}_* \end{bmatrix} & 0 \\ 0 & 0 & \begin{bmatrix} \frac{n_1}{2\sigma^4_{\epsilon_1}} & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{n_m}{2\sigma^4_{\epsilon_m}} \end{bmatrix} \end{pmatrix}. \quad (4.3.9)$$

In the above, $\boldsymbol{C}_* = \boldsymbol{C}_i/\sigma^2_{\epsilon_i}$, $\boldsymbol{D}_* = \boldsymbol{D}_i/\sigma^2_{\epsilon_i}$ and $\boldsymbol{F}_* = \boldsymbol{F}_i/\sigma^2_{\epsilon_i}$ are constant across all m series and are not dependent on $\sigma^2_{\epsilon_i}$, $i = 1, \ldots m$ because each process only differs in its innovations variance (and its mean level).

Hence the asymptotic variance of the maximum likelihood estimates, $\hat{\boldsymbol{\varphi}}_{AIARMA}$,

of the parameters of an AIARMA process with m series is,

$$
\begin{aligned}
&\mathrm{Var}(\hat{\boldsymbol{\varphi}}_{AIARMA}) \\
&= E(-\boldsymbol{H}_{AIARMA}(\boldsymbol{\varphi}))^{-1} \\
&= \begin{pmatrix}
\left(\sum_{i=1}^{m} \frac{n_i}{\sigma^2_{\epsilon_i}} \boldsymbol{G}_i\right) & 0 & 0 \\
 & \left(\sum_{i=1}^{m} n_i\right) \begin{bmatrix} \boldsymbol{C}_* & \boldsymbol{D}_* \\ \boldsymbol{D}_*^\top & \boldsymbol{F}_* \end{bmatrix} & 0 \\
0 & 0 & \begin{bmatrix} \frac{n_1}{2\sigma^4_{\epsilon_1}} & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{n_m}{2\sigma^4_{\epsilon_m}} \end{bmatrix}
\end{pmatrix}^{-1} \\
&= \begin{pmatrix}
\left(\sum_{i=1}^{m} \frac{n_i}{\sigma^2_{\epsilon_i}} \mathbf{G}_i\right)^{-1} & 0 & 0 \\
 & \frac{1}{\sum_{i=1}^{m} n_i} \begin{bmatrix} \boldsymbol{C}_* & \boldsymbol{D}_* \\ \boldsymbol{D}_*^\top & \boldsymbol{F}_* \end{bmatrix}^{-1} & 0 \\
0 & 0 & 2\begin{bmatrix} \frac{\sigma^4_{\epsilon_1}}{n_1} & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\sigma^4_{\epsilon_m}}{n_m} \end{bmatrix}
\end{pmatrix}.
\end{aligned}
$$

Hence asymptotically,

$$(\hat{\boldsymbol{\varphi}}_{AIARMA} - \boldsymbol{\varphi}_{AIARMA}) \;\sim\; N\left(\mathbf{0}, \mathrm{Var}(\hat{\boldsymbol{\varphi}}_{AIARMA})\right).$$

As with maximum likelihood estimates for a RARMA process, $\hat{\boldsymbol{\beta}}$ is asymptotically independent of $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ which is asymptotically independent of the elements of $\left(\hat{\sigma}^2_{\epsilon_1}, \ldots, \hat{\sigma}^2_{\epsilon_m}\right)$ which in turn are asymptotically independent of each other.

4.3.5 Asymptotic Distribution of the Estimates of the Parameters of the First Order AIARMA Process with a Unique Mean for Each Series

As an example, the asymptotic distribution of the parameters of the AIARMA model will be derived which reflects a common AR(1), MA(1) or ARMA(1,1) process with unique mean and innovations variance by series.

Using the results of Section 4.3.4, the complete variance matrix of the AIARMA parameter estimates is,

$$\mathrm{Var}(\hat{\boldsymbol{\varphi}}_{AIARMA}) \;=\; \begin{bmatrix} \boldsymbol{V}^{R}_{AIARMA} & 0 & 0 \\ 0 & \boldsymbol{V}^{C}_{AIARMA} & 0 \\ 0 & 0 & \boldsymbol{V}^{E}_{AIARMA} \end{bmatrix},$$

where $\boldsymbol{V}^{R}_{AIARMA}$ is the variance sub-matrix for the regression parameters, $\boldsymbol{V}^{C}_{AIARMA}$ is for the ARMA coefficients and $\boldsymbol{V}^{E}_{AIARMA}$ is for the innovation variances.

The asymptotic distribution of the maximum likelihood estimates is,

$$(\hat{\boldsymbol{\varphi}}_{AIARMA} - \boldsymbol{\varphi}_{AIARMA}) \;\sim\; N(\mathbf{0}, Var(\hat{\boldsymbol{\varphi}}_{AIARMA}),$$

where the components are derived below.

Let the j$^{\text{th}}$ series of $\boldsymbol{z}^{(i)}$, for i, $j = 1, \ldots, m$, be defined as,

$$z_{j,i,t} \;=\; \begin{cases} 1 & j = 1 \\ 1 & j > 1 \,\text{and}\, i = j \\ 0 & j > 1 \,\text{and}\, i \neq j \end{cases}$$

(that is, following (2.1.1), β_1 is the mean level for series 1 and β_i, $i > 1$, is the difference between the means for the i^{th} and first series) and using the results from Section 4.3.4, the variance matrix of the estimates of the regression parameters is,

$$\boldsymbol{V}^{R}_{AIARMA} \;=\; \left(\sum_{i=1}^{m} \left(n_i / \sigma^2_{\epsilon_i} \right) \boldsymbol{G}_i \right)^{-1}.$$

The design matrix of extraneous variables used here is commonly employed

in regression modelling and it is straightforward to show that the $m \times m$ matrix of cross-covariances of the filtered extraneous variables, $\theta^{-1}(B)\phi(B)z_{j,i,t}$ for each series, $i = 1, \ldots, m$, is

$$\boldsymbol{G}_i = \begin{cases} (1+\theta_1)^{-2}(1+\phi_1)^2 \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \mathbf{0} & \vdots \\ \vdots & \mathbf{0} & \ddots & \\ 0 & \ldots & & 0 \end{bmatrix} & i = 1 \\ (1+\theta_1)^{-2}(1+\phi_1)^2 \begin{bmatrix} 1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ 0 & & & & 0 & & & \vdots \\ \vdots & & \mathbf{0} & & \vdots & & \mathbf{0} & \\ 0 & & & & 0 & & & \\ 1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ 0 & & & & 0 & & & \vdots \\ \vdots & & \mathbf{0} & & \vdots & & \mathbf{0} & \\ 0 & \ldots & & & 0 & & & 0 \end{bmatrix} & i \geqslant 2, \end{cases}$$

where the "1"s are in the i^{th} row and i^{th} columns only. Hence,

$$\begin{aligned} \boldsymbol{V}^R_{AIARMA} &= (1+\theta_1)^2(1+\phi_1)^{-2} \begin{bmatrix} \sum_{j=1}^m \frac{n_j}{\sigma^2_{\epsilon_j}} & \frac{n_2}{\sigma^2_{\epsilon_2}} & \cdots & \frac{n_m}{\sigma^2_{\epsilon_m}} \\ \frac{n_2}{\sigma^2_{\epsilon_2}} & \frac{n_2}{\sigma^2_{\epsilon_2}} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ \frac{n_m}{\sigma^2_{\epsilon_m}} & \ldots & 0 & \frac{n_m}{\sigma^2_{\epsilon_m}} \end{bmatrix}^{-1} \\ &= (1+\theta_1)^2(1+\phi_1)^{-2} \begin{bmatrix} \frac{\sigma^2_{\epsilon_1}}{n_1} & \ldots & -\frac{\sigma^2_{\epsilon_1}}{n_1} & \ldots & -\frac{\sigma^2_{\epsilon_1}}{n_1} \\ -\frac{\sigma^2_{\epsilon_1}}{n_1} & 0 & \left(\frac{\sigma^2_{\epsilon_1}}{n_1} + \frac{\sigma^2_{\epsilon_2}}{n_2}\right) & 0 & \vdots \\ \vdots & 0 & 0 & \ddots & 0 \\ -\frac{\sigma^2_{\epsilon_1}}{n_1} & 0 & \ldots & 0 & \left(\frac{\sigma^2_{\epsilon_1}}{n_1} + \frac{\sigma^2_{\epsilon_m}}{n_m}\right) \end{bmatrix}. \end{aligned}$$

When fitting AR(1), MA(1) or ARMA(1,1) models, the variance matrix of the

estimates of the ARMA coefficients is,

$$\boldsymbol{V}^{C}_{AIARMA} = \left(\sum_{i=1}^{m} n_i\right)^{-1} \begin{cases} (1-\phi_1^2) & p=1,\, q=0 \\ (1-\theta_1^2) & p=0,\, q=1 \\ \begin{bmatrix} (1-\phi_1^2)^{-1} & (1-\phi_1\theta_1)^{-1} \\ (1-\phi_1\theta_1)^{-1} & (1-\theta_1^2)^{-1} \end{bmatrix}^{-1} & p=1,\, q=1 \end{cases}$$

(see Shumway and Stoffer [2011] p. 134 allowing for the difference in defining the signs of the AR coefficients). The result for the last outcome expands to,

$$\left(\sum_{i=1}^{m} n_i\right)^{-1} \left(\left(1-\phi_1^2\right)^{-1}\left(1-\theta_1^2\right)^{-1} - \left(1-\phi_1\theta_1\right)^{-2}\right)^{-1} \begin{bmatrix} \left(1-\theta_1^2\right)^{-1} & -\left(1-\phi_1\theta_1\right)^{-1} \\ -\left(1-\phi_1\theta_1\right)^{-1} & \left(1-\phi_1^2\right)^{-1} \end{bmatrix}.$$

Finally the variance of the estimates of the innovations variances is,

$$\boldsymbol{V}^{E}_{AIARMA} = \begin{bmatrix} \frac{2\sigma_{\epsilon_1}^4}{n_1} & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \frac{2\sigma_{\epsilon_m}^4}{n_m} \end{bmatrix}.$$

4.4 Conditional AIARMA (CAIARMA) Process

Again, as with the RARMA and AIARMA modelling, there are three approaches to CAIARMA model fitting employing the constrained and unconstrained joint likelihood versus interleaving methods.

4.4.1 Joint Likelihood Approach

The unconditional likelihood for a CAIARMA process with m series and series lengths $\{n_i\}_{i=1}^{m}$ can be expressed as the following joint likelihood where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)^\top$, representing a simple mean per series rather than a full linear model. This is consistent with the definition of CAIARMA processes from Section 2.1 where the specification of a constant mean per series is required to facilitate the assumption of an innovations variance proportional to each series mean.

The likelihood is (following (4.3.1)),

$$\begin{aligned} L_{CAIARMA}(\boldsymbol{Y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\mu}, c) &= \prod_{i=1}^{m} L_{ARMA}(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mu_i, \sigma_{\epsilon_i} = c\mu_i) \qquad (4.4.1) \\ &= \prod_{i=1}^{m} \left((2\pi(c\mu_i)^2)^{-n_i/2} \left[\prod_{t=1}^{n_i} r_t(\boldsymbol{\phi}, \boldsymbol{\theta}) \right]^{-1/2} \exp\left[-\frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mu_i)}{2(c\mu_i)^2} \right] \right) \\ &= \prod_{i=1}^{m} \left((2\pi(c\mu_i)^2)^{-n_i/2} \left[\prod_{t=1}^{n_i} r_t(\boldsymbol{\phi}, \boldsymbol{\theta}) \right]^{-1/2} \right) \\ &\quad \times \exp\left[-\sum_{i=1}^{m} \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mu_i)}{2(c\mu_i)^2} \right]. \qquad (4.4.2) \end{aligned}$$

So the joint log likelihood is

$$\begin{aligned} LL_{CAIARMA}(\boldsymbol{Y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\mu}, c) &= -\frac{\sum_{i=1}^{m} n_i}{2} \log(2\pi) - \sum_{i=1}^{m} n_i \left[\log c + \log \mu_i \right] \\ &\quad -\frac{1}{2} \sum_{i=1}^{m} \sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})) \\ &\quad -\sum_{i=1}^{m} \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mu_i)}{2(c\mu_i)^2}. \qquad (4.4.3) \end{aligned}$$

If the first derivative with respect to c is taken and the result set to zero this gives,

$$0 = -\sum_{i=1}^{m} \frac{n_i}{c} + \frac{1}{c^3} \left(\sum_{i=1}^{m} \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \hat{\mu}_i)}{\hat{\mu}_i{}^2} \right). \qquad (4.4.4)$$

That is,

$$\hat{c} = \sqrt{\frac{\sum_{i=1}^{m} \frac{S(\boldsymbol{y}^{(i)}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \hat{\mu}_i)}{\hat{\mu}_i{}^2}}{\sum_{i=1}^{m} n_i}} \quad , \qquad (4.4.5)$$

where $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ and $\{\hat{\mu}_i\}_{i=1}^{m}$ are the unconditional maximum likelihood estimates of $(\boldsymbol{\phi}, \boldsymbol{\theta})$ and $\{\mu_i\}_{i=1}^{m}$. Note that if

$$\hat{c}_i^2 = \frac{S(\boldsymbol{y}^{(i)}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \hat{\mu}_i)}{n_i \hat{\mu}_i{}^2}$$

then

$$\hat{c} = \sqrt{\frac{\sum_{i=1}^{m} n_i \hat{c}_i^2}{\sum_{i=1}^{m} n_i}} \quad ,$$

that is, $\hat{c}$ is the square root of the weighted mean of the $\{\hat{c}_i^2\}_{i=1}^m$, weighted according to the relative sample sizes. Substituting (4.4.5) into equation (4.4.3) gives the concentrated log likelihood,

$$\begin{aligned} CLL_{CAIARMA}(\boldsymbol{Y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \{\mu_i\}_{i=1}^m) &= -\frac{\sum_{i=1}^m n_i}{2} \log(2\pi) - \sum_{i=1}^m n_i \log \mu_i \\ &\quad -\frac{1}{2}\left(\sum_{i=1}^m n_i\right) \log\left(\frac{\sum_{i=1}^m \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \hat{\mu}_i)}{{\mu_i}^2}}{\sum_{i=1}^m n_i}\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^m \sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})) - \left(\frac{\sum_{i=1}^m n_i}{2}\right) \\ &= -\frac{\sum_{i=1}^m n_i}{2}(\log(2\pi) + 1) - \sum_{i=1}^m n_i \log \mu_i \\ &\quad -\frac{1}{2}\left(\sum_{i=1}^m n_i\right) \log\left(\frac{\sum_{i=1}^m \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \hat{\mu}_i)}{{\mu_i}^2}}{\sum_{i=1}^m n_i}\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^m \sum_{t=1}^{n_i} \log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})). \end{aligned} \tag{4.4.6}$$

This can be maximised numerically to derive the maximum likelihood values, $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}, \{\hat{\mu}_i\}_{i=1}^m)$. The maximum likelihood estimate of c is then calculated using (4.4.5). The standard errors of $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$, $\hat{c}$ and $\{\hat{\mu}_i\}_{i=1}^m$ can be obtained using the Hessian matrix of the full log likelihood.

4.4.2 Use of R's `arima` Function

As with RARMA and AIARMA modelling, the above process can be facilitated by use of the `arima` function in R. If all values of $(\boldsymbol{\phi}, \boldsymbol{\theta}, \{\mu_i\}_{i=1}^m)$ are fixed in calling the function (m times, one for each series) then the function calls return the concentrated log likelihood value (separately) for each series along with the maximum

likelihood values by series of the m innovations variances. The resultant summated concentrated log likelihoods can then be adjusted as follows to derive the joint concentrated log likelihood under the CAIARMA model,

$$\begin{aligned}\sum_{i=1}^{m} CLL_{ARMA}(\boldsymbol{Y},\boldsymbol{\phi},\boldsymbol{\theta},\{\mu_i\}_{i=1}^{m}) &= -\frac{\sum_{i=1}^{m} n_i}{2}\log(2\pi+1) - \frac{1}{2}\sum_{i=1}^{m} n_i \log\left(\frac{S(\boldsymbol{y}^{(i)},\boldsymbol{\phi},\boldsymbol{\theta},\mu_i)}{n_i}\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{n_i}\log(r_t(\boldsymbol{\phi},\boldsymbol{\theta})) \\ &= CLL_{CAIARMA}(\boldsymbol{Y},\boldsymbol{\phi},\boldsymbol{\theta},\{\mu_i\}_{i=1}^{m}) + \sum_{i=1}^{m} n_i \log\mu_i \\ &\quad +\frac{1}{2}\left(\sum_{i=1}^{m} n_i\right)\log\left(\frac{\sum_{i=1}^{m}\frac{S(\boldsymbol{y}^{(i)},\boldsymbol{\phi},\boldsymbol{\theta},\mu_i)}{{\mu_i}^2}}{\sum_{i=1}^{m} n_i}\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{m} n_i \log\left(\frac{S(\boldsymbol{y}^{(i)},\boldsymbol{\phi},\boldsymbol{\theta},\mu_i)}{n_i}\right). \end{aligned} \tag{4.4.7}$$

Hence,

$$\begin{aligned} CLL_{CAIARMA}(\boldsymbol{Y},\boldsymbol{\phi},\boldsymbol{\theta},\{\mu_i\}_{i=1}^{m}) &= \sum_{i=1}^{m} CLL_{ARMA}(\boldsymbol{Y},\boldsymbol{\phi},\boldsymbol{\theta},\{\mu_i\}_{i=1}^{m}) - \sum_{i=1}^{m} n_i\log\mu_i \\ &\quad -\frac{1}{2}\left(\sum_{i=1}^{m} n_i\right)\log\left(\frac{\sum_{i=1}^{m}\frac{S(\boldsymbol{y}^{(i)},\boldsymbol{\phi},\boldsymbol{\theta},\mu_i)}{{\mu_i}^2}}{\sum_{i=1}^{m} n_i}\right) \\ &\quad +\frac{1}{2}\sum_{i=1}^{m} n_i \log\left(\frac{S(\boldsymbol{y}^{(i)},\boldsymbol{\phi},\boldsymbol{\theta},\mu_i)}{n_i}\right). \end{aligned} \tag{4.4.8}$$

This concentrated log likelihood, $CLL_{CAIARMA}(\boldsymbol{Y},\boldsymbol{\phi},\boldsymbol{\theta},\{\mu_i\}_{i=1}^{m})$, can then be optimised with respect to $(\boldsymbol{\phi},\boldsymbol{\theta},\{\mu_i\}_{i=1}^{m})$ leading to the maximum likelihood estimate of c using (4.4.5). This is then employed in a final call to `arima` to derive the full (that is, not concentrated) log likelihood surface after an appropriate transformation of the output (see (4.4.3), (4.4.6) and (4.4.8)), that is,

$$\begin{aligned} LL_{CAIARMA}(\boldsymbol{Y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\mu}, c) &= CLL_{CAIARMA}(\boldsymbol{Y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\mu}) \\ &\quad + \frac{\sum_{i=1}^{m} n_i}{2} \\ &\quad + \left(\frac{\sum_{i=1}^{m} n_i}{2}\right) \log\left(\frac{\sum_{i=1}^{m} S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mu_i)}{\sum_{i=1}^{m} n_i}\right) \\ &\quad - \left(\sum_{i=1}^{m} n_i\right) \log(c) - \sum_{i=1}^{m} \frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mu_i)}{2\left(c\mu_i\right)^2}. \end{aligned}$$

This is used to evaluate the Hessian matrix and derive the estimated standard errors of the estimates.

As with RARMA and AIARMA model fitting, the methods detailed in Section 4.1.1 can be applied to the AR and MA parameters to ensure stationarity and invertibility.

4.4.3 Estimation using a Re-Scaled Zero-Mean Interleaved Process

In order to use interleaving to fit a CAIARMA process, the log likelihood of the CAIARMA model (4.4.3) can be expressed as,

$$
\begin{aligned}
LL_{CAIARMA}(\mathbf{Y}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\mu}, c) &= -\frac{\log(2\pi)}{2}\sum_{i=1}^{m} n_i - \sum_{i=1}^{m}\left(\frac{1}{2}n_i \log c^2 + \frac{1}{2}n_i \log \mu_i^2\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{n_i}\log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})) \\
&\quad -\left(\sum_{i=1}^{m}\frac{S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, E(\boldsymbol{y}^{(i)}) = \mu_i)/\mu_i^2}{2c^2}\right) \\
&= -\frac{\log(2\pi)}{2}\sum_{i=1}^{m} n_i - \sum_{i=1}^{m}\left(\frac{1}{2}n_i \log c^2 + \frac{1}{2}n_i \log \mu_i^2\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{n_i}\log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})) \\
&\quad -\left(\sum_{i=1}^{m}\frac{S(\boldsymbol{y}^{(i)}/\mu_i, \boldsymbol{\phi}, \boldsymbol{\theta}, E(\boldsymbol{y}^{(i)}/\mu_i) = \frac{\mu_i}{\mu_i} = 1)}{2c^2}\right) \\
&= -\frac{\log(2\pi)}{2}\sum_{i=1}^{m} n_i - \frac{\log c^2}{2}\sum_{i=1}^{m} n_i \\
&\quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{t=1}^{n_i}\log(r_t(\boldsymbol{\phi}, \boldsymbol{\theta})) \\
&\quad -\left(\sum_{i=1}^{m}\frac{S(\boldsymbol{y}^{(i)}/\mu_i, \boldsymbol{\phi}, \boldsymbol{\theta},\ E(\boldsymbol{y}^{(i)}/\mu_i) = 1)}{2c^2}\right) \\
&\quad -\frac{1}{2}\sum_{i=1}^{m} n_i \log \mu_i^2 \\
&= \sum_{i=1}^{m} LL_{ARMA}(\boldsymbol{u}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_\epsilon^2 = c^2,\ E(\boldsymbol{u}^{(i)}) = 0) \\
&\quad -\sum_{i=1}^{m} n_i \log \mu_i, \qquad (4.4.9)
\end{aligned}
$$

where $\boldsymbol{u}^{(i)} = \boldsymbol{y}^{(i)}/\mu_i - 1$ and noting as mentioned previously that $S(\boldsymbol{y}^{(i)}, \boldsymbol{\phi}, \boldsymbol{\theta}, \mu_i)$ is a linear function of squared and cross-product values of the elements of $(\boldsymbol{y}^{(i)} - \mu_i)$ (with no intercept). Hence the above transformation can be used to create series with the same innovations variance (and zero mean). These can be modeled as a RARMA process with m series using interleaving (and missing values as required to equalise the series lengths) and then employing the appropriate log likelihood

modification as shown in (4.4.9).

Hence the maximum likelihood estimates of all the parameters can be achieved by

- rescaling each series by initial estimates, $\{\tilde{\mu}_i\}_{i=1}^m$, of $\{\hat{\mu}_i\}_{i=1}^m$,
- subtracting one from each rescaled series to give a population mean of zero,
- interleaving the re-scaled and re-located series padded with missing values to equalise the series's lengths,
- estimating the coefficients of the resulting interleaved process with zero mean (conditional on $\{\tilde{\mu}_i\}_{i=1}^m$),
- correcting the resultant log likelihood to derive the final log likelihood by subtracting $\sum_{i=1}^m n_i \log \tilde{\mu}_i$ and
- converging on the maximised log likelihood with respect to $\{\tilde{\mu}_i\}_{i=1}^m$.

The maximum likelihood estimates of $(\boldsymbol{\phi}, \boldsymbol{\theta})$ and c (being what is now the square root of the innovations variance of the transformed series) are available from the fourth step above and, of $\{\mu_i\}_{i=1}^m$, from the result of the final step. The standard errors of the coefficients and mean estimates and of $\hat{c}$ can be got from the inverse of the negative of the Hessian matrix of the full log likelihood at the maximum likelihood values using empirical methods.

4.4.4 Asymptotic Distribution of the Maximum Likelihood Estimates

In this section the asymptotic distribution of the unconditional maximum likelihood estimates of the parameters of a CAIARMA process are derived.

The expected value of the negative of the Hessian matrix for the log likelihood of a single-series ARMA model with a constant mean, μ, is (see (4.1.8)),

$$E(-\boldsymbol{H}_{ARMA}) = \begin{bmatrix} \frac{n}{\sigma_\epsilon{}^2}\theta(1)^{-2}\phi(1)^2 & 0 & 0 \\ 0 & n\boldsymbol{P} & 0 \\ 0 & 0 & \frac{n}{2\sigma_\epsilon{}^4} \end{bmatrix}, \tag{4.4.10}$$

where $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{C}_* & \boldsymbol{D}_* \\ \boldsymbol{D}_*^\top & \boldsymbol{F}_* \end{bmatrix}$ from (4.1.6). Also note that, following Section 4.1.3, the filter $\theta^{-1}(B)\phi(B)$ applied to the series of all 1's (representing the mean) results in the series of equal values, $\theta(1)^{-1}\phi(1)$, the mean of the square of which is $\theta(1)^{-2}\phi(1)^2$.

The parameter transformations, $\mu = \mu$, $\{\phi_i = \phi_i\}_{i=1}^p$, $\{\theta_i = \theta_i\}_{i=1}^q$ and $c = \sqrt{\sigma_\epsilon^2}/\mu$ are now undertaken. If the original parameters are $\{\kappa_i\}_{i=1}^{1+p+q+1}$ and the transformed parameters as a function of $\{\kappa_i\}_{i=1}^{1+p+q+1}$ are $\left\{\delta_j\left(\{\kappa_i\}_{i=1}^{1+p+q+1}\right)\right\}_{j=1}^{1+p+q+1}$, the Jacobian matrix, $\boldsymbol{J}$, of the transformation of the parameters with elements, $\{J_{i,j} = \partial\delta_i/\partial\kappa_j\}_{i,j=1}^{1+p+q+1}$, is,

$$\boldsymbol{J} = \begin{bmatrix} 1 & \boldsymbol{0} & 0 \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ -\frac{\sqrt{\sigma_\epsilon^2}}{\mu^2} & \boldsymbol{0} & \frac{1}{2\sqrt{\sigma_\epsilon^2}\mu} \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{0} & 0 \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ -\frac{c}{\mu} & \boldsymbol{0} & \frac{1}{2c\mu^2} \end{bmatrix}.$$

This definition of the Jacobian matrix uses rows for the partial derivatives of each function (whilst some authors use columns). Also, substituting into (4.4.10),

$$E(-\boldsymbol{H}_{ARMA}) = \begin{bmatrix} \frac{n}{c^2\mu^2}\theta(1)^{-2}\phi(1)^2 & 0 & 0 \\ 0 & n\boldsymbol{P} & 0 \\ 0 & 0 & \frac{n}{2c^4\mu^4} \end{bmatrix}. \tag{4.4.11}$$

Now it is well known that, if a set of multivariate parameters are functionally transformed, then the maximum likelihood estimates of the transformed parameters are the original maximum likelihood estimates transformed in the same way (given certain regularity conditions). Hence the expected value of the negative Hessian of the log likelihood of the new parameters (that is, $E(-\boldsymbol{H}^*_{ARMA})$) can be derived as follows using the Multivariate Delta Method (see Taboga [2010]),

$$\boldsymbol{\Xi}^* = \boldsymbol{J}\,\boldsymbol{\Xi}\,\boldsymbol{J}^\top,$$

where $\boldsymbol{\Xi}$ is the variance matrix of the estimates of the original parameters and $\boldsymbol{\Xi}^*$ is the variance matrix of the estimates of the transformed parameters. That is,

$$
\begin{aligned}
E(-\boldsymbol{H}^*_{ARMA})^{-1} &= \boldsymbol{J}.E(-\boldsymbol{H}_{ARMA})^{-1}\boldsymbol{J}^\top. \\
\text{Hence, } E(-\boldsymbol{H}^*_{ARMA}) &= \left[\boldsymbol{J}\, E(-\boldsymbol{H}_{ARMA})^{-1}\, \boldsymbol{J}^\top\right]^{-1} \\
&= (\boldsymbol{J}^{-1})^\top E(-\boldsymbol{H}_{ARMA})\, \boldsymbol{J}^{-1} \\
&= \begin{bmatrix} 1 & \mathbf{0} & 2c^2\mu \\ \mathbf{0} & \boldsymbol{I} & \mathbf{0} \\ 0 & \mathbf{0} & 2c\mu^2 \end{bmatrix} E(-\boldsymbol{H}_{ARMA}) \begin{bmatrix} 1 & \mathbf{0} & 0 \\ \mathbf{0} & \boldsymbol{I} & \mathbf{0} \\ 2c^2\mu & \mathbf{0} & 2c\mu^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{(2c^2+a)n}{c^2\mu^2} & \mathbf{0} & \frac{2n}{c\mu} \\ \mathbf{0} & n\boldsymbol{P} & \mathbf{0} \\ \frac{2n}{c\mu} & \mathbf{0} & \frac{2n}{c^2} \end{bmatrix},
\end{aligned}
$$

where $a = \theta(1)^{-2}\phi(1)^2$.

This result is for one series (with one mean). It can be readily extend to m series with m unique means and a common c following the approach in Section 4.2.4 which reflects m unique means and a common σ^2_ϵ. The i^{th} expected negative Hessian, that is, for the i^{th} series, (with respect to all of the m means, $\{\mu_i\}_{i=1}^m$, c and the ARMA coefficients) has the following form,

$$
E(-\boldsymbol{H}^*_{ARMA_i}) = \begin{bmatrix} 0 & \cdots & 0 & \cdots & \mathbf{0} & 0 \\ \vdots & \ddots & \vdots & 0 & \vdots & \vdots \\ 0 & 0 & 0 & \mathbf{0} & \vdots & 0 \\ \vdots & 0 & \frac{(2c^2+a)n_i}{c^2{\mu_i}^2} & 0 & \mathbf{0} & \frac{2n_i}{c\mu_i} \\ 0 & \cdots & 0 & \ddots & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 & n_i\boldsymbol{P} & \mathbf{0} \\ 0 & \cdots & \frac{2n_i}{c\mu_i} & \cdots & \mathbf{0} & \frac{2n_i}{c^2} \end{bmatrix},
$$

where the upper left hand $m \times m$ matrix only has a non-zero element at (i, i). Note also that any elements involving partial derivatives with respect to $\{\mu_j\}_{j=1\neq i}^m$are zero values as these $m-1$ parameters are not included in the associated log likelihood for the i^{th} series.

When combined as the sum of the expected negative Hessians, the following expected negative Hessian of the CAIARMA process results,

$$
\begin{aligned}
E(-\boldsymbol{H}^*_{CAIARMA}) &= \sum_{i=1}^{m} E(-\boldsymbol{H}^*_{ARMA_i}) \\
&= \begin{bmatrix}
\frac{(2c^2+a)n_1}{c^2{\mu_1}^2} & 0 & 0 & 0 & \frac{2n_1}{c\mu_1} \\
0 & \ddots & \vdots & \vdots & \vdots \\
\vdots & 0 & \frac{(2c^2+a)n_m}{c^2{\mu_m}^2} & 0 & \frac{2n_m}{c\mu_m} \\
0 & 0 & 0 & (\sum_{i=1}^{m} n_i)\,\boldsymbol{P} & 0 \\
\frac{2n_1}{c\mu_1} & \cdots & \frac{2n_m}{c\mu_m} & 0 & \frac{2\sum_{i=1}^{m} n_i}{c^2}
\end{bmatrix}.
\end{aligned}
$$

Setting $n_* = \sum_{j=1}^{m} n_j$, the inverse of this expected negative Hessian (that is, $E(-\boldsymbol{H}^*_{CAIARMA})^{-1}$) provides the variance, $Var(\hat{\boldsymbol{\eta}}_{CAIARMA})$, of the unconditional maximum likelihood estimates of the CAIARMA parameters, $\boldsymbol{\eta}_{CAIARMA} = (\eta_1, \ldots, \eta_{m+p+q+1})^\top = (\{\mu_i\}_{i=1}^{m}, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top, c)^\top$,

$$
\mathrm{Var}(\hat{\boldsymbol{\eta}}_{CAIARMA}) =
$$

$$
\begin{bmatrix}
\frac{(an_*+2\,n_1c^2)c^2{\mu_1}^2}{(an_*)n_1(2\,c^2+a)} & \frac{2c^4\mu_1\mu_2}{(an_*)(2\,c^2+a)} & \cdots & \frac{2c^4\mu_1\mu_m}{(an_*)(2\,c^2+a)} & 0 & -\frac{c^3\mu_1}{an_*} \\
\frac{2c^4\mu_2\mu_1}{(an_*)(2\,c^2+a)} & \ddots & & \frac{2c^4\mu_2\mu_m}{(an_*)(2\,c^2+a)} & \vdots & \vdots \\
\vdots & \ddots & & \vdots & \vdots & \vdots \\
\frac{2c^4\mu_m\mu_1}{(an_*)(2\,c^2+a)} & \frac{2c^4\mu_m\mu_2}{(an_*)(2\,c^2+a)} & \cdots & \frac{(an_*+2\,n_mc^2)c^2{\mu_m}^2}{(an_*)n_m(2\,c^2+a)} & 0 & -\frac{c^3\mu_m}{an_*} \\
0 & \cdots & \cdots & 0 & n_*^{-1}\boldsymbol{P}^{-1} & 0 \\
-\frac{c^3\mu_1}{an_*} & \cdots & \cdots & -\frac{c^3\mu_m}{an_*} & 0 & \frac{(2\,c^2+a)c^2}{2an_*}
\end{bmatrix}. \tag{4.4.12}
$$

Note that the asymptotic variance of c is not dependent on the series means but the asymptotic variance of the estimated means is dependent on c and their own population value.

Hence, the asymptotic distribution of the maximum likelihood estimates, $\hat{\boldsymbol{\eta}}_i$, is,

$$
(\hat{\boldsymbol{\eta}}_{CAIARMA} - \boldsymbol{\eta}_{CAIARMA}) \;\sim\; N(\mathbf{0}, \mathrm{Var}(\hat{\boldsymbol{\eta}}_{CAIARMA})).
$$

As an example of this asymptotic result, for a two-series CAIARMA process with first order autoregressive generating mechanism, the matrix, $n_*^{-1}\boldsymbol{P}^{-1} = \boldsymbol{V}^C_{AIARMA}$, from Section 4.3.5 can be substituted into (4.4.12) giving the asymptotic variance

matrix,

$$\begin{bmatrix} \frac{(an_*+2\,n_1c^2)c^2{\mu_1}^2}{(an_*)n_1(2\,c^2+a)} & \frac{2c^4\mu_1\mu_2}{(an_*)(2\,c^2+a)} & 0 & -\frac{c^3\mu_1}{an_*} \\ \frac{2c^4\mu_2\mu_1}{(an_*)(2\,c^2+a)} & \frac{(an_*+2\,n_2c^2)c^2{\mu_2}^2}{(an_*)n_2(2\,c^2+a)} & 0 & -\frac{c^3\mu_2}{an_*} \\ 0 & 0 & \frac{(1-\phi_1^2)}{n_*} & 0 \\ -\frac{c^3\mu_1}{an_*} & -\frac{c^3\mu_2}{an_*} & 0 & \frac{(2\,c^2+a)c^2}{2an_*} \end{bmatrix}. \tag{4.4.13}$$

Note that the asymptotic variance of the ARMA parameter doesn't depend on c. In general, for any single-series ARMA, RARMA, AIARMA or CAIRMA process, the variances of the ARMA parameters don't depend on the innovations variance either directly or through c.

Chapter 5

Simulation Studies of the Estimates

The purpose of this chapter is to explore the finite sample properties of the maximum likelihood estimators from Chapter 4. The simulation method used to assess the small sample properties of the unconditional maximum likelihood estimators (MLEs) is described. This is then applied in the first instance to single-series ARMA processes (for reference and comparison purposes) and then extended to RARMA, AIARMA and CAIARMA processes.

5.1 Simulation Design

As with the hypothesis testing in Chapter 3, the simulation design process defined in Paxton et al. [2001] is followed. The measures to assess the performance of the maximum likelihood estimator, say, $\hat{\varsigma}$, of a population parameter, ς, estimated for a given model and set of series lengths, are:

1. Finite-sample bias, that is, $\text{Bias}(\hat{\varsigma}) = E(\hat{\varsigma} - \varsigma)$. Here the estimated bias is defined as $\widehat{\text{Bias}}(\hat{\varsigma}) = \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} (\hat{\varsigma}_j - \varsigma)$ where $\hat{\varsigma}_j$ is the j^{th} simulated estimate out of n_{sim} simulations.

2. The finite-sample standard error, that is, $SE(\hat{\varsigma}) = \sqrt{Var(\hat{\varsigma})} = \sqrt{E[(\hat{\varsigma} - E(\hat{\varsigma}))^2]}$. The estimated standard error is defined here as,

$$\widehat{SE}(\hat{\varsigma}) = \sqrt{\frac{1}{n_{sim}-1}\sum_{j=1}^{n_{sim}}\left(\hat{\varsigma}_j - \frac{1}{n_{sim}}\sum_{j=1}^{n_{sim}}\hat{\varsigma}_j\right)^2}.$$

The Mean Square Error (MSE) (that is, $E[(\hat{\varsigma}-\varsigma)^2]$) will not be addressed as it can be derived from the combination of bias and standard error.

3. Coverage of the confidence interval which uses the estimated asymptotic standard errors, that is, $P\left(\left(\hat{\varsigma} - 1.96\sqrt{\widehat{Var}(\hat{\varsigma})}, \hat{\varsigma} + 1.96\sqrt{\widehat{Var}(\hat{\varsigma})}\right) \ni \varsigma\right)$ where $\widehat{Var}(\hat{\varsigma})$ is the estimate of the variance of $\hat{\varsigma}$ derived from the inverse of the negative of the Hessian matrix of the realised log likelihood function. This measures the ability of the derived standard errors for each estimator to accurately provide confidence intervals and ultimately to act as inputs to Wald hypothesis testing. Coverage is estimated using,

$$\frac{1}{n_{sim}}\sum_{i=1}^{n_{sim}}\left[\left(\hat{\varsigma}_j - 1.96\sqrt{\widehat{Var}(\hat{\varsigma}_j)},\ \hat{\varsigma}_j + 1.96\sqrt{\widehat{Var}(\hat{\varsigma}_j)}\right) \ni \varsigma\right],$$

where "[...]" is the Iverson bracket.

For each model this chapter will also report on the alignment of the simulations with the asymptotic distribution which involves a combined assessment based on finite sample bias, variance and coverage. Alignment is useful in at least two ways. Firstly it suggests the use of the asymptotic distribution in understanding the finite sample properties of the estimates. Secondly it helps validate the conventional confidence intervals in assessing the variability of the estimates.

The simulations also investigate:

- Stability of the estimation process. The convergence behaviour of the estimation routines is reported for each type of Joint ARMA model. This provides some indication of the numerical stability of the model fitting.

- Agreement of results with published studies on properties of estimators of standard single-series ARMA models. This helps verify the simulations on those models and the extension to other models.

- Intervention ARMA (ARMAX) models as a commonly used extension to standard ARMA models. This permits the introduction of external explanatory

variables.

- Models close to the invertibility region (overdifferencing) and stationarity region (requires differencing). This explores model fitting in circumstances often encountered in econometrics and other disciplines (for example, see Hamilton [1994] Chapter 15).

Mis-specified models and non-normal data were not considered as the current work is focused on the performance of the estimation routines under standard conditions.

5.1.1 Candidate Representative Models

Models were chosen for simulations that most closely resemble those used in practice or are of most interest to researchers. This is inevitably a judgement decision. The following processes were deemed the most appropriate from a casual sampling of the applied literature and from the author's own experience:

1. The single-series equivalents of Joint ARMA models with the same total series length. This allowed comparisons between estimation performance for several series (in the current circumstance, two) and one series of the same total length as the (two) replicated series.

2. AR(1), MA(1) and ARMA(1,1) processes. These all appear regularly in the applied literature (see, for example, Jenkins [1979], Alfares and Nazeeruddin [2002] and Stevenson [2003]), although perhaps ARMA less so.

3. Models with intervention variables representing differences in the series means from the first series. This is arguably the simplest and most widely used form of intervention model.

4. Innovations variances which are not equal between series but differ by only a moderately-sized factor. This is likely to be the case in most practical circumstances.

Seasonal ARMA models were not included but the results can be easily extended to these processes.

Table 5.1 – ARMA Models and Parameters Used in Simulations

Model	ϕ_1	θ_1	Model	ϕ_1	θ_1	Model	ϕ_1	θ_1
1	-0.9	0.0	7	0.0	-0.9	13	-0.4	-0.9
2	-0.5	0.0	8	0.0	-0.5	14	-0.4	-0.5
3	-0.1	0.0	9	0.0	-0.1	15	-0.4	-0.1
4	0.1	0.0	10	0.0	0.1	16	-0.4	0.1
5	0.5	0.0	11	0.0	0.5	17	-0.4	0.5
6	0.9	0.0	12	0.0	0.9	18	-0.4	0.9

5.1.2 Specific Experimental Conditions

The circumstance of having a large number of series, m, but a relatively modest series length, n, is a common problem with panel data in econometrics and with longitudinal data in biometrics and this scenario is examined in this thesis at least asymptotically. There is also a need to investigate the opposite, to examine the relative efficiency of using a small number of "long" series. It is not immediately clear whether this is less or more efficient than the reverse. Scenarios are also examined where the lengths of the replicated series differ. In the current case for half the scenarios, the second series length is half the first.

5.1.3 Population Parameters

The chosen models include those with AR and MA parameters close to the unit circle. Also a range of positive and negative AR and MA parameters were selected. For the ARMA processes, to restrict the number of runs and volume of reporting, the AR parameter is fixed at -0.4 and the MA parameter is varied (being typically the more difficult to estimate). The parameters chosen are shown in Table 5.1.

The innovations variances use $\sigma_\epsilon^2 = 1$ for RARMA processes and, for AIARMA, $\sigma_{\epsilon_1}^2 = 1$ and $\sigma_{\epsilon_2}^2 = 4$. For CAIARMA processes, $c = 0.2$. Intervention terms involved are $\mu_1 = 5$ (first series mean) and $\mu_2 = 10$ (second series mean).

5.1.4 The Simulations

It was decided to run scenarios featuring each of the following attributes:

Table 5.2 – Sample Sizes Used in the Simulations

n_1	n_2	$n = n_1 + n_2$	Designation
64	32	96	Small
64	64	128	Small
256	128	384	Moderate
256	256	512	Large
1024	512	1536	Large
1024	1024	2048	Large

1. Number of independent series. The simplest scenario of two series is used to minimise the total number of simulation runs. For 1,000 simulations, the CAIARMA runs took just under 19 hours elapsed time on a dedicated Windows 10 PC with an Intel i7 2.1GHz Dual Core processor having 8GB of memory and using seven parallel processors. The simulation runs for the single-series ARMA, RARMA and AIARMA models recorded similar execution times.

2. Length of each of the series. This parallels the lengths of the series used in the hypothesis testing of Section 3. A further set of scenarios is run with the second series half the length of the first to show how the routines can handle series of unequal length. The series lengths used in this thesis are shown in Table 5.2 with a description of the series length, being "Small", etc. For ARMA models with only one series, the separate sample sizes are still relevant reflecting the point of change in series means.

The simulations are undertaken for:

- Unconstrained joint likelihood versus constrained (for stationarity and invertibility) joint likelihood methods versus (unconstrained) interleaved fitting. The latter isn't used for the single-series ARMA modelling.

- ARMA v RARMA v AIARMA v CAIARMA models.

- 1,000 simulations to provide a reasonable level of accuracy consistent with (recent) published studies.

5.2 Single-Series ARMA Process

Relatively few studies have been published using simulation to understand the finite sample properties of unconditional maximum likelihood estimators of the parameters of single-series ARMA processes. The main results are Dent and Min [1978], Watson and Nicholls [1992], Hauser [1999] and Krone et al. [2017] (AR(1) only) and align with results of the current simulations. The exploration here will follow the level of detail used for the RARMA, AIARMA and CAIARMA models both for comparison and to enhance the existing literature.

5.2.1 Comparison of the Two Estimation Routines

The results for the estimated AR parameter are closely aligned between the constrained and unconstrained joint ARMA estimates (not shown but see Figure 5.2.1 for MA(1) results). Note that only two estimation methods are compared here because, as mentioned previously, interleaving is not required for a single-series process.

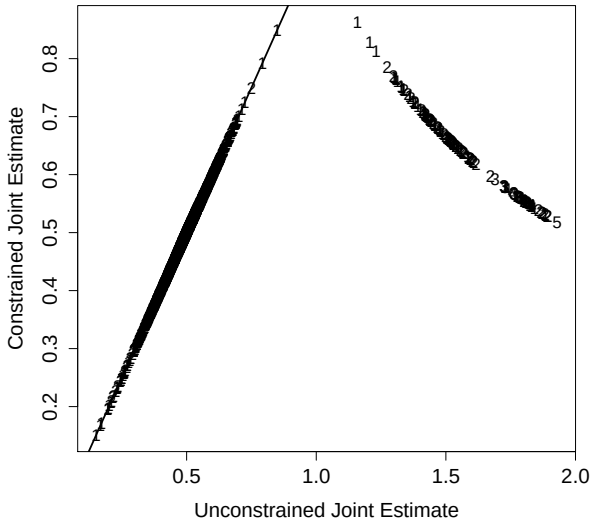
For ARMA models the AR coefficient results for the unconstrained versus constrained joint likelihood estimates for the ARMA model with $\phi_1 = -0.4$ and $\theta_1 = -0.5$ diverge in a minority of cases (but stay in the stationary region), likely because of the near cancellation of the factors in B (not shown).

For the MA parameter, the results for the unconstrained method reflect the phenomenon whereby the same likelihood applies for sets of MA parameters equidistant from the unit circle. It can be readily shown that this circumstance results in the same autocorrelation function such as in the MA(1) case with $\theta(B) = (1 + \theta_1 B)$ and $\theta(B) = (1 + \theta_1^{-1} B)$. As an illustration, Figure 5.2.1 compares the simulation outcomes for the constrained versus unconstrained methods for $\hat{\theta}_1$ from the MA(1) model with $\theta_1 = 0.5$.

The points are designated such that "1" is for $n = 64 + 32$, "2" for $n = 64 + 64$, "3" for $n = 256 + 128$, "4" for $n = 256 + 256$, "5" for $n = 1,024 + 512$ and "6" for $n = 1,024 + 1,024$ where n is the total series length being the sum of the length with the first and second means; the line is a reference 1:1 indicator.

The estimated mean level for the first series, the estimated difference for the second series and the estimated innovations variance show more alignment between

Figure 5.2.1 – Comparison of Simulated θ_1 Estimates (Single-Series ARMA Models - MA(1), $\theta_1 = 0.5$). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024) with a 1:1 reference line.



the two estimation methods than for the AR or MA parameters (not shown).

In summary, the optimisation routines for the two systems show no substantial convergence issues. Given the requirement of stationarity and invertibility and the consistency of the two results, only the second constrained joint likelihood approach is discussed below.

5.2.2 Bias

The average bias for $\hat{\phi}_1$ and $\hat{\theta}_1$ from the simulations are shown in Figure 5.2.2. The x-axis labels indicate the (ϕ_1, θ_1) parameterisation and the coding of points follows Figure 5.2.1.

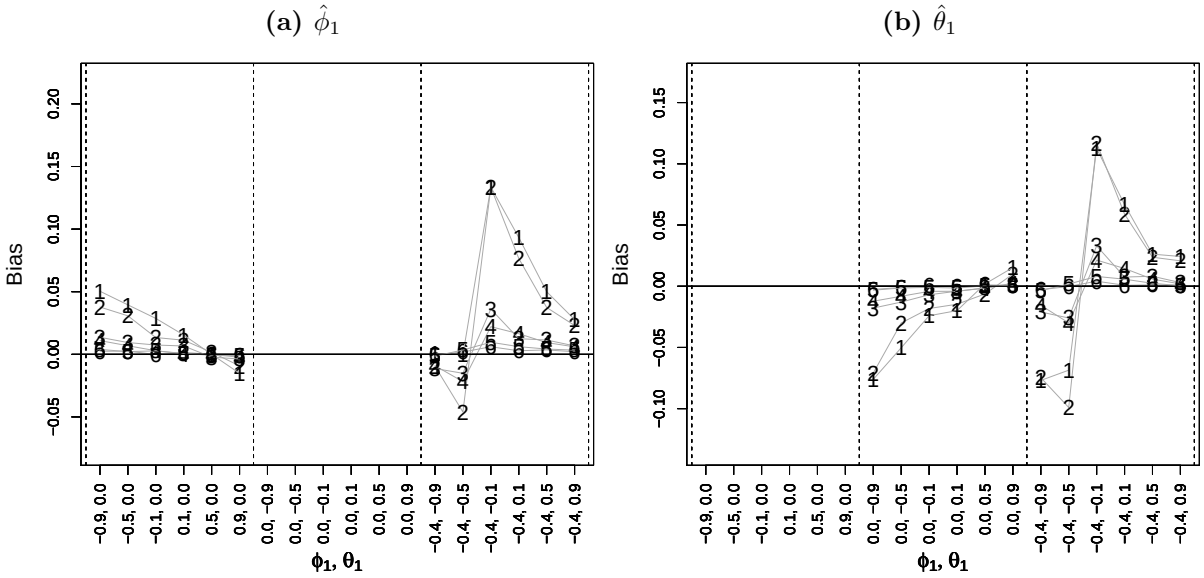
AR Parameter. For the estimate of the AR parameter, for a pure AR(1) process, the typically positive bias decreases as the sample size increases and as the autogressive parameter increases from -0.9 to 0.9 (it is zero at approximately $\phi_1 = 0.3$). This is in line with Cordeiro and Klein [1994] where the first order approximation to the bias is shown to be $(1 - 3\phi_1)/n$ for an AR(1) process with an estimated constant mean. For ARMA processes the typically positive bias in the AR parameter

estimate is large (up to 0.14) at $\phi_1 = -0.4$, $\theta_1 = -0.1$ and $\theta_1 = 0.1$ and $n \leq 128$.

MA Parameter. For the MA(1) parameter estimates, the bias is greatest (negative) at $\theta_1 = -0.9$ where, for small sample sizes, it is -0.07 declining to zero at $\theta_1 \approx 0.5$. This approximately agrees with Cordeiro and Klein [1994] who show that the first order approximation to the bias is $(2\theta_1 - 1)/n$ for an MA(1) process with a estimated constant mean. The MA result for ARMA processes shows typically large ($\sim \pm 0.10$) biases for small sample sizes ($n \leq 128$).

Other Parameters. There is no apparent bias in the estimated mean for the first series or the estimated difference for the second series (not shown). This again reflects the results of Cordeiro and Klein [1994] who derive a first-order zero bias in this case. The estimate of the innovations variance shows a negative bias (~ -0.03 for n$\leq$ 128) which declines towards zero as n increases (again not shown in this thesis).

Figure 5.2.2 – Simulated Bias of MLE Parameter Estimates (Single-Series ARMA Models - $\hat{\phi}_1$ and $\hat{\theta}_1$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).

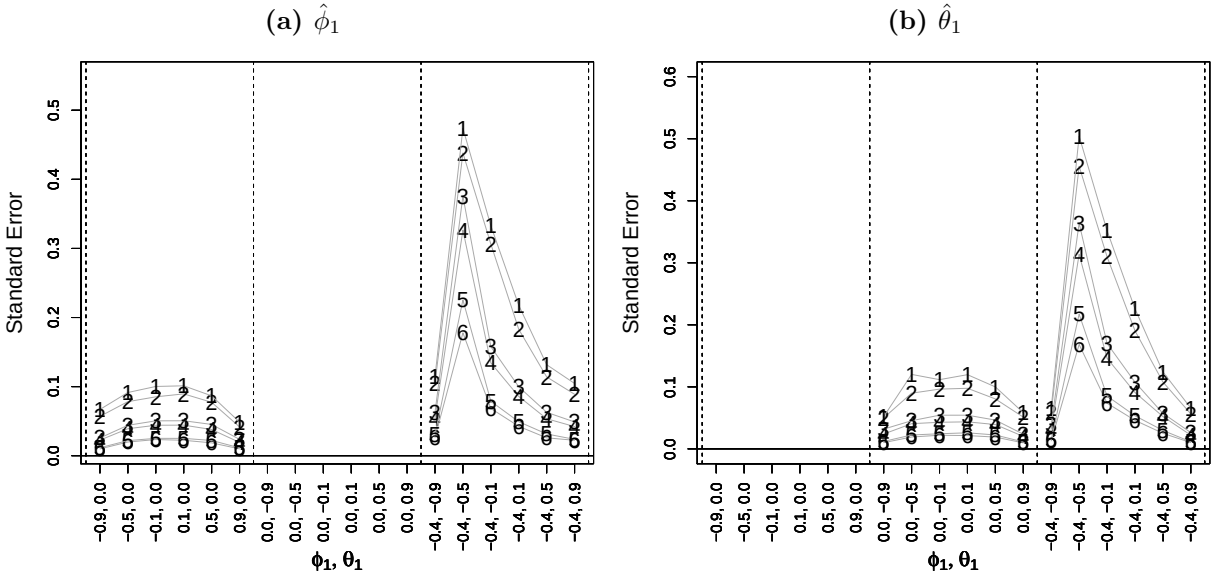


5.2.3 Standard Error

The simulated standard errors for $\hat{\phi}_1$ and $\hat{\theta}_1$ from the simulated estimates are presented in Figure 5.2.3. For a pure AR(1) process the standard error of the AR parameter estimate is minimised at the extremes of parameter values and is at a maximum near zero. The standard errors are much higher for the ARMA model and reach a peak at $\phi_1 = -0.4$ and $\theta_1 = -0.5$ likely reflecting again the near cancellation of the AR and MA difference equation factors in B.

The results for the estimated MA parameter are very similar to those for the AR parameter. The standard errors of the estimated mean level of the first series and of the difference for the second series decrease with increasing ϕ_1 and increase with increasing θ_1 (not shown). The standard errors of the estimated innovations variance are approximately constant across all models and parameterisations (again not shown).

Figure 5.2.3 – Simulated Standard Error of MLE Parameter Estimates (Single-Series ARMA Models - $\hat{\phi}_1$ and $\hat{\theta}_1$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).



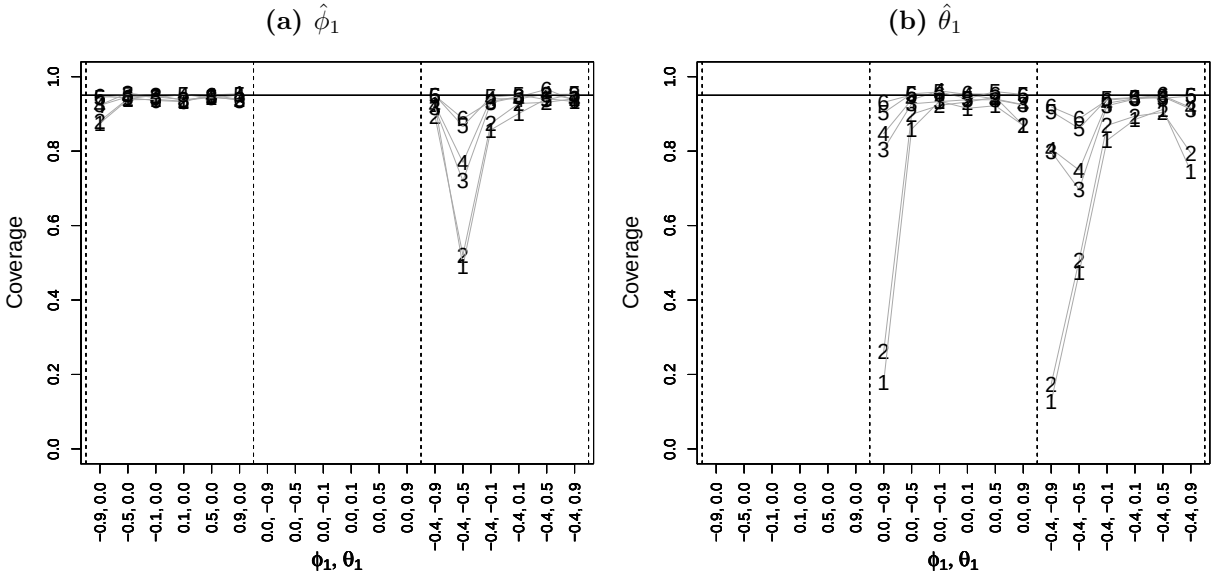
5.2.4 Coverage

The coverage for $\hat{\phi}_1$ and $\hat{\theta}_1$ from the simulations are plotted in Figure 5.2.4. For the AR parameter in a pure AR process, the coverage is very close to the notional 0.95. However for the mixed process the coverage is typically lower than 0.95 and drops to between 0.85 to 0.5 for $\phi_1 = -0.4$ and $\theta_1 = -0.5$ for even large sample sizes of $n = 1536$. For the MA(1) θ_1 parameter, coverage is typically below 0.95 sometimes substantially below. At $\theta_1 = -0.9$ the coverage drops to approximately 0.25 for sample sizes below 128. This also happens for the mixed model.

The coverage of the estimated mean of the first series and for the difference for the second series (not shown) is close to 0.95 except for a substantial shortfall for $n \leq 512$ for $\phi_1 \leq -0.5$ and $\theta_1 \leq -0.5$. The coverage of the estimated innovations variance confidence intervals are typically between 0.90 and 0.95 except for small samples.

These results suggest that the empirical confidence intervals for maximum likelihood estimates for AR, MA and ARMA models are generally accurate.

Figure 5.2.4 – Simulated Coverage of MLE Parameter Estimates (Single-Series ARMA Models - $\hat{\phi}_1$ and $\hat{\theta}_1$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024). The horizontal reference line shows the 95% nominal value.



5.2.5 Alignment to Asymptotic Distribution

Across all parameters, the standard errors from simulation (see Figure 5.2.3 and Section 4.1.3) generally agree with the asymptotic results (not shown) albeit having a slightly higher value (as expected). On a proportional basis the asymptotic standard errors are typically within 10% of the simulated values except for simulations at small to moderate sample sizes (≤ 384) with large positive and negative AR and MA coefficients and for ARMA models with factors in B that nearly cancel.

Given the reduction in bias with sample size, the agreement with the asymptotic standard errors and the reasonably accurate coverage, it is concluded that the asymptotic distribution can be assumed to be applicable for moderate to large samples for single-series ARMA processes. Moreover this further suggests that the Hessian-based empirical confidence intervals are reliable.

5.2.6 Conclusions

These results for the single-series ARMA models suggest that, except for parameters near the unit circle and those with nearly cancelling factors in B, the MLEs are "well-behaved" reflecting relatively low bias, stable standard errors and accurate empirical confidence intervals. The constrained likelihood approach showed the most stable MLE values.

5.3 RARMA Processes

As with single-series ARMA modelling, the following comments refer to the typical estimator performance measures, being bias, standard error, coverage and alignment to the asymptotic distribution. The discussion begins with the stability of the optimisation process from the three estimation methods including the interleaving method.

5.3.1 Comparison of the Three Estimation Routines

For the AR parameter, the outcomes from the three estimation systems are very similar for AR and ARMA processes and are typically closely aligned. There is

some variation between the three estimation systems when the AR and MA factors in B almost cancel. For the MA parameter, results were similar to those for single-series ARMA processes whereby the same likelihood is realised for MA parameters equidistant from the unit circle.

Figure 5.3.1 compares the three estimation results for $\hat{\theta}_1$ using the MA(1) model with $\theta_1 = 0.5$ where the line is a reference 1:1 indicator. The points are coded as used previously, that is, the codes indicated the series lengths but, in this case, for the first versus second series. This point designation applies also for the AIARMA and CAIARMA results to come (see Sections 5.4.1 and 5.5.1).

The estimated mean level for the first series and the estimated difference for the second are more stable than for the ARMA coefficients. The estimated innovations variance varies for a substantial number of runs when the parameters are close to the unit circle or the factors for the AR and MA difference equations in B are close to cancelling. The optimisation routines for all three systems for "well-behaved" parameterisations shows no substantial convergence issues.

Given the consistency of the three results and the requirement to ensure stationarity and invertibility, the constrained joint likelihood approach is chosen for reporting the simulation analysis.

5.3.2 Bias

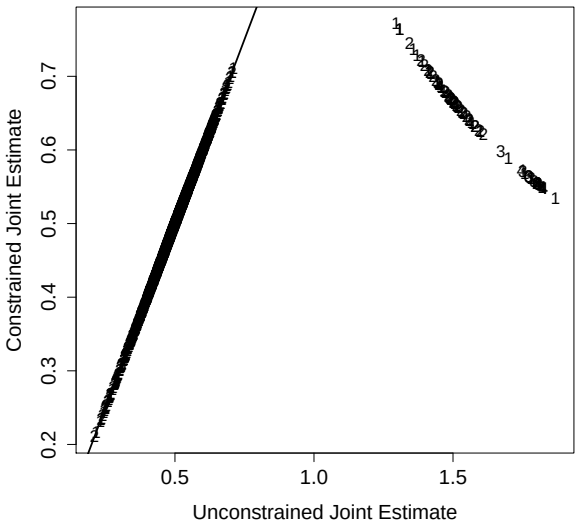
The mean bias for $\hat{\phi}_1$ and $\hat{\theta}_1$ for the simulations undertaken in this thesis are shown in Figure 5.3.2 by process, parameterisation and sample size. The x-axis labels indicate the (ϕ_1, θ_1) parameterisation as with the single-series ARMA plots.

For the AR parameter, the results are very similar to those from the single-series modelling (see Figure 5.2.2a) with the largest (positive) bias for AR(1) processes at $\phi_1 = -0.9$, and for ARMA(1,1) processes, at $\phi_1 = -0.4$ and $\theta_1 = -0.1$. For the MA parameter, again the results are very close to those from single-series modelling (see Figure 5.2.2b) reflecting a typically negative bias for the MA(1) process and showing maximum bias for small total samples (≤ 128) at $\theta_1 = -0.9$ and at ($\phi_1 = -0.4$, $\theta_1 \leq -0.1$).

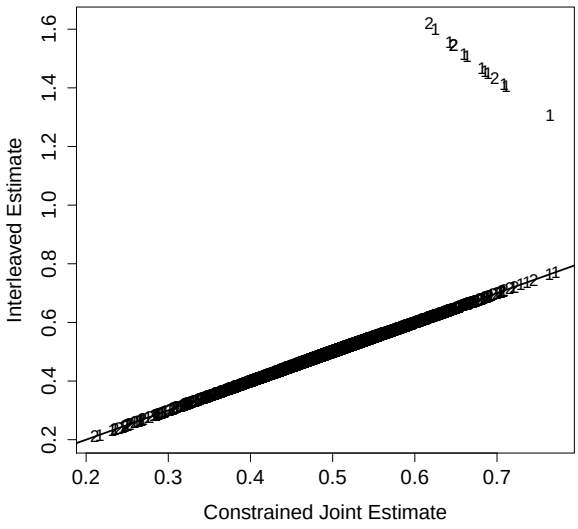
There is no apparent bias in the estimated mean for the first series or the difference for the second (not shown). The estimate of the innovations variance

Figure 5.3.1 – Comparison of Simulated θ_1 Estimates (RARMA Models - MA(1), $\theta_1 = 0.5$). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024) with 1:1 reference line.

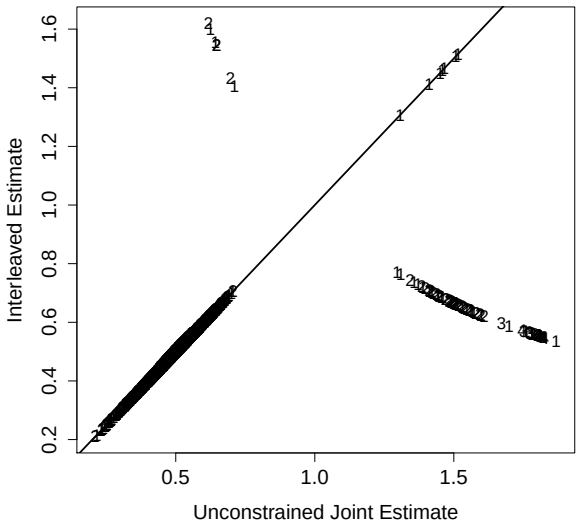
(a) Unconstrained versus Constrained Joint Likelihood



(b) Constrained Joint Likelihood versus Interleaved

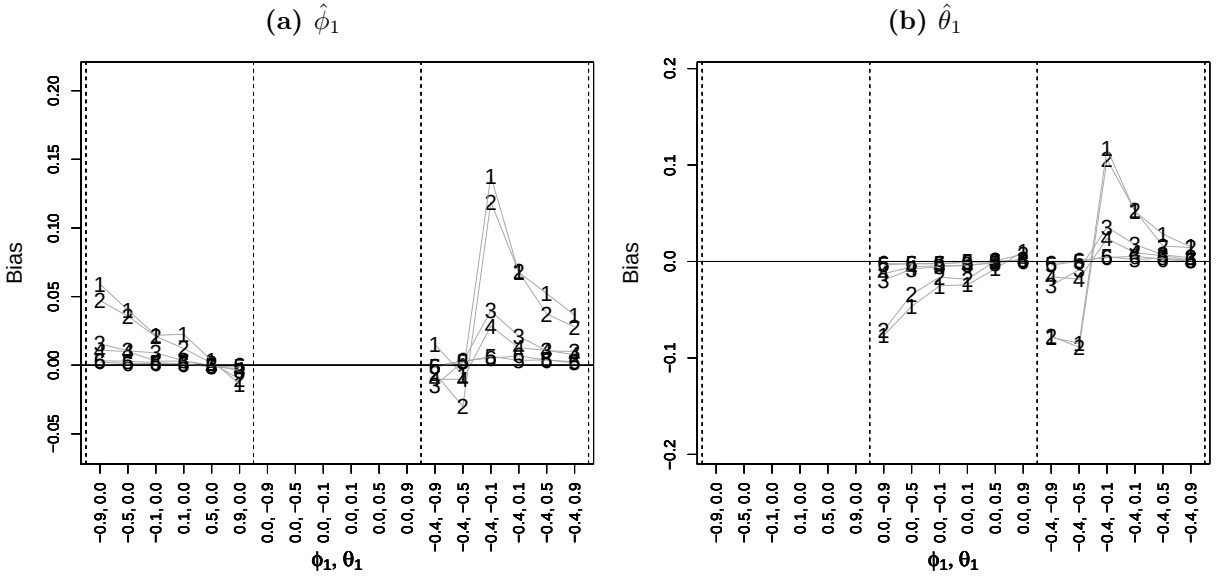


(c) Unconstrained Joint Likelihood versus Interleaved



(also not shown) exhibits a small negative bias (-0.08) which declines towards zero as n increases.

Figure 5.3.2 – Simulated Bias of MLE Parameter Estimates (RARMA Models - $\hat{\phi}_1$ and $\hat{\theta}_1$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).



5.3.3 Standard Error

The standard errors for the simulations are presented in Figure 5.3.3 for $\hat{\phi}_1$ and $\hat{\theta}_1$.

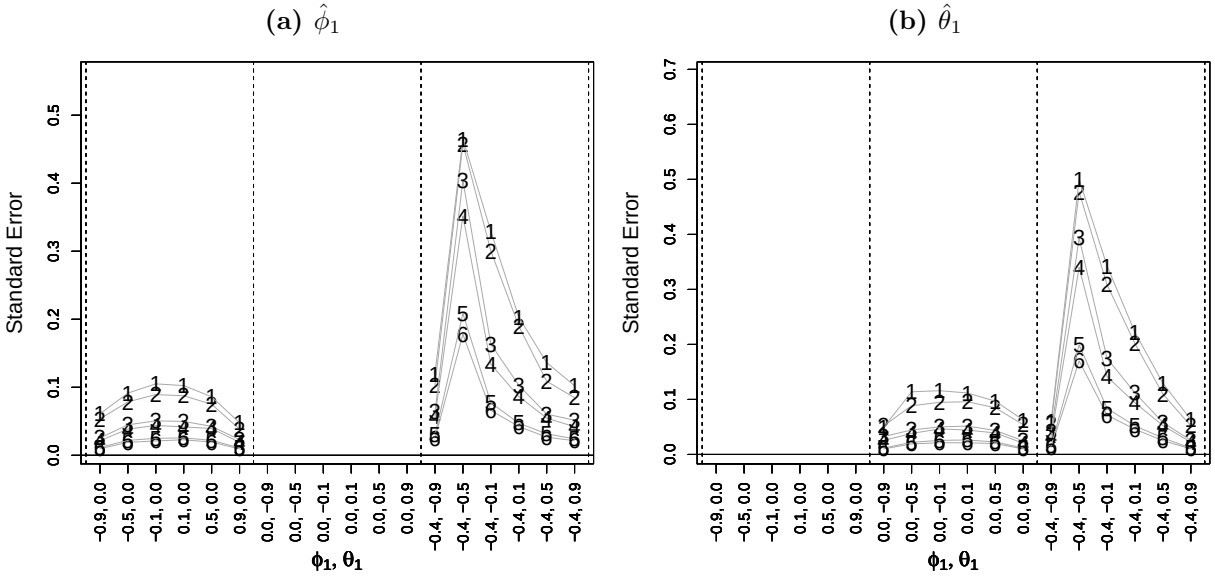
The results for the AR parameter are again very similar to those for single-series models as suggested by the alignment of the asymptotic standard errors (that is, comparing the analytical results from Sections 4.1.3 and 4.2.4). For a pure AR(1) process the standard error of the AR parameter estimate is at a minimum at the extremes of parameter values and is at a maximum near zero. The standard errors are much higher for the ARMA model and reach a peak at $\phi_1 = -0.4$ and $\theta_1 = -0.5$ likely reflecting again the near cancellation of the AR and MA factors in B.

The results for the MA parameter are similar to those for the AR parameter and very similar to those for the single-series modelling.

The standard errors of the estimated mean level of the first series and for the

difference for the second series are very close to those for single-series models (not shown). The standard errors of the estimated innovations variance are consistent with those for the single-series models and are approximately constant across all models and parameterisations again as suggested by the asymptotic variances.

Figure 5.3.3 – Simulated Standard Error of MLE Parameter Estimates (RARMA Models - $\hat{\phi}_1$ and $\hat{\theta}_1$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).



5.3.4 Coverage

The average coverage from the simulations for 95% confidence intervals for $\hat{\phi}_1$ and $\hat{\theta}_1$ are plotted in Figure 5.3.4. The outcome for the AR parameter closely follows the single-series results where the coverage is close to 0.95 except for the almost cancelling factors in B. For the MA parameter, the result is also close to that for the single-series models with coverage substantially below 0.95 for large negative θ_1 in the MA(1) and ARMA(1,1) processes.

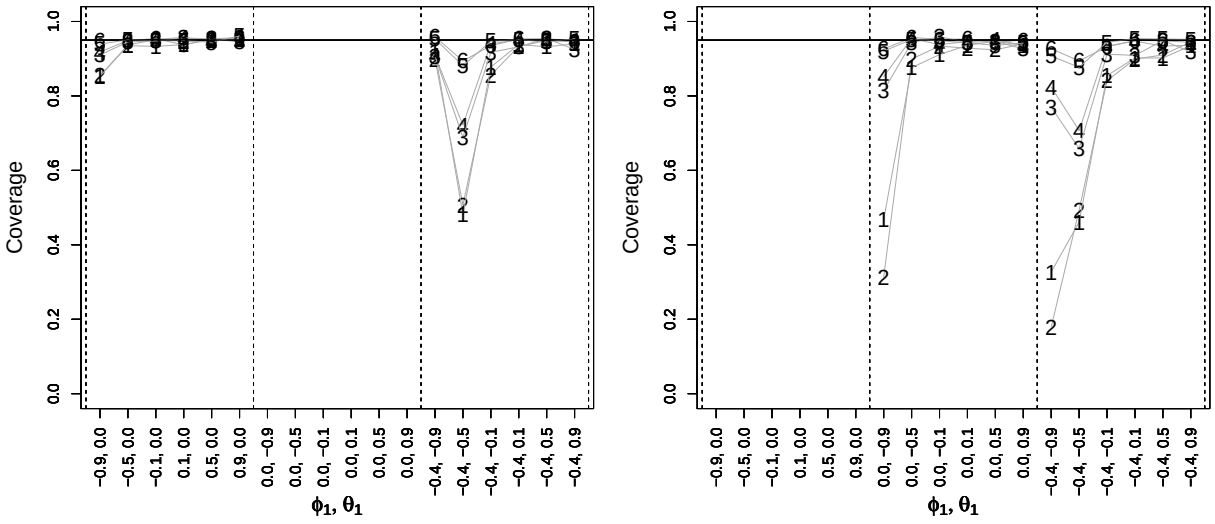
The coverage of the estimated mean of the first series and of the estimated difference for the second series (not shown) is typically marginally below 95% with more substantial reductions for small samples at $\phi_1 = -0.9$ and $\theta_1 = -0.9$ for AR(1)

and MA(1) processes and for ARMA(1,1) processes at $\phi_1 = -0.4$ and $\theta_1 = -0.5$ or -0.9.

The above suggest that the unconditional maximum likelihood confidence intervals for RARMA processes are reasonably accurate for moderate to large sample sizes ($n > 128$. See Table 5.2 for sample size classifications).

Figure 5.3.4 – Simulated Coverage of MLE Parameter Estimate (RARMA Models - $\hat{\phi}_1$ and $\hat{\theta}_1$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024). The horizontal reference line shows the 95% nominal value.

(a) $\hat{\phi}_1$ **(b)** $\hat{\theta}_1$



5.3.5 Alignment to Asymptotic Distribution

As with the single-series results, on a proportional basis the asymptotic standard errors are typically within 0.10 of the simulated values (not shown) except for small sample sizes, for large negative AR coefficients (-0.9) and for ARMA models with nearly cancelling factors in B. Again as with single-series processes, the finite sample bias in the maximum likelihood estimates appear to converge to zero as $n \to \infty$ (though it is not clear what type of convergence is reflected here) and the empirical confidence intervals and asymptotic standard errors are accurate for finite samples.

Hence it is concluded that the asymptotic distribution of the RARMA estimates can be used as a reliable indication of estimate behaviour for moderate to large sample behaviour. Moreover, as with single-series ARMA MLEs, this further suggests that the Hessian-based empirical confidence intervals are reliable.

5.3.6 Comparative Efficiency of Joint versus Single-Series Estimation

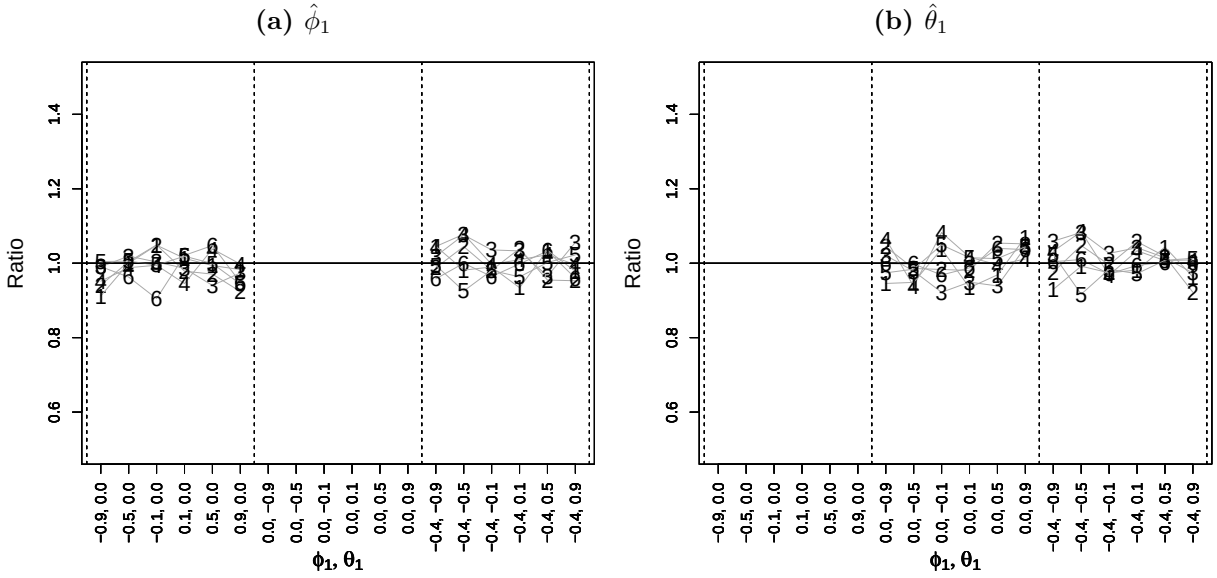
A question of some interest is whether multiple stationary and invertible series result in joint unconditional maximum likelihood estimates which have similar finite sample standard errors to those from one series of the same total length and same generating mechanism.

Figure 5.3.5 shows the ratios of the simulation standard errors for $\hat{\phi}_1$ and $\hat{\theta}_1$ from Figures 5.2.3 and 5.3.3, that is, from one ARMA process versus two ARMA processes with the same generating mechanism and the same total length. There is no clear trend away from unity and this is also reflected in similar plots (not shown) for the variance of the estimated mean of the first series, the difference in mean for the second and the innovations variance. The asymptotic standard errors (see (4.1.7) and (4.2.8)) suggest a similar outcome (for two or more component series).

The Interleaving Theorem (see Section 2.3.1) states that multiple independent (stationary and invertible) series with the same generating mechanism can be represented by one (stationary and invertible) series with certain ARMA parameters set to zero and a length equal to the total length of all of the original replicated series. This further suggests that the unconditional maximum likelihood estimates of the parameters of the single ARMA and the RARMA process can be considered equivalent in efficiency for the same total series length.

Hence the simulation outcomes, the Interleaving Theorem and the asymptotic results all lead to the same conclusion. The joint use of many series from the same stationary and invertible process versus one series of equivalent total length with the same generating mechanism results in unconditional maximum likelihood parameter estimates with equivalent sampling properties, at least for moderate series lengths.

Figure 5.3.5 – Ratio of Simulated Standard Errors of Unconditional Maximum Likelihood Estimates of Two RARMA Processes to a Single-Series ARMA Process ($\hat{\phi}_1$ and $\hat{\theta}_1$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).



5.3.7 Conclusions

As argued in Section 5.2.6 for single-series ARMA models, these simulation results for the RARMA models suggest that, except when ARMA coefficients are near the unit circle or reflect nearly cancelling factors in B, the MLE's are "well-behaved" reflecting acceptable bias, stable standard errors and accurate empirical confidence intervals. The constrained joint likelihood approach derives the most stable MLE values. Finally for the same underlying process the joint use of multiple series versus analysis with one series of the same total length produces maximum likelihood estimates with very similar properties.

5.4 AIARMA Processes

As with the ARMA and RARMA modelling, the following comments refer to the estimator performance measures listed in Section 5.1, being bias, standard error, coverage and alignment to the asymptotic distribution as well as the stability of

the optimisation process with a comparison of the three estimation methods. The parameters whose results are plotted are the AR and MA coefficients and the difference between the innovations variance for the first and second series. The estimate of the difference between the innovations variances is designated $\widehat{\Delta\sigma^2} = \widehat{\sigma^2_{\epsilon_2} - \sigma^2_{\epsilon_1}}$.

5.4.1 Comparison of the Three Estimation Routines

There is a confluence of results between the three estimation systems except at the extremes of the MA parameterisation and where the AR and MA factors in B almost cancel. Also again there is evidence that the same likelihood is realised for MA parameters equidistant from the unit circle. As an illustration, Figure 5.4.1 compares the three estimation results for $\hat{\theta}_1$ using the MA(1) model with $\theta_1 = 0.5$.

Again the optimisation routines for all three systems for "well-behaved" parameterisations show convergence issues for only a small fraction of the runs. The constrained joint likelihood approach is chosen for the reporting the AIARMA simulations.

5.4.2 Bias

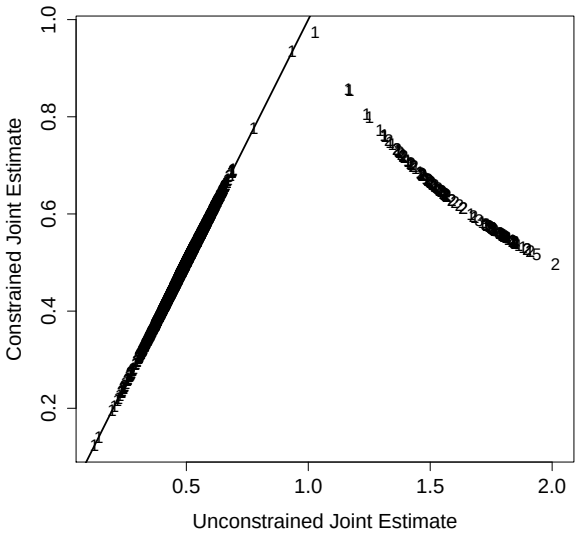
The mean bias for $\hat{\phi}_1$, $\hat{\theta}_1$ and $\widehat{\Delta\sigma^2}$ from the AIARMA simulations are shown in Figure 5.4.2 by process, sample size and parameterisation. For the AR parameter for the AR(1) model, bias decreases from ~ 0.06 at $\phi_1 = -0.9$ at low sample sizes through zero at $\phi_1 \approx 0.5$ closely mirroring the RARMA and single-series ARMA results.

The estimated MA parameter follows a similar pattern again as with the single-series ARMA and RARMA modelling.

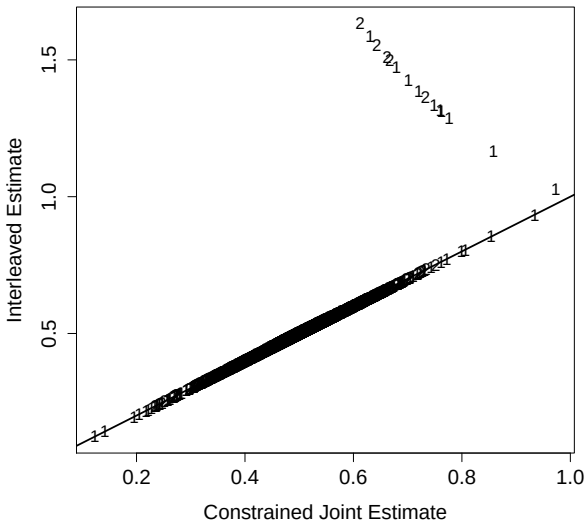
There is no apparent bias in the estimated mean for the first series or the difference for the second (not shown). The estimates of the innovations variances show a negative bias which declines to zero as n increases. Again this follows the previous model results. The bias in the difference in innovations variance (see Figure 5.4.2c) is typically modestly negative for most generating processes (bias is less than -0.1 for $n_1 + n_2 > 384$) and decreases rapidly with increasing sample size.

Figure 5.4.1 – Comparison of Simulated θ_1 Estimates (AIARMA Models - MA(1), $\theta_1 = 0.5$). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024) with 1:1 reference line.

(a) Unconstrained versus Constrained Joint Likelihood



(b) Constrained Joint Likelihood versus Interleaved



(c) Unconstrained Joint Likelihood versus Interleaved

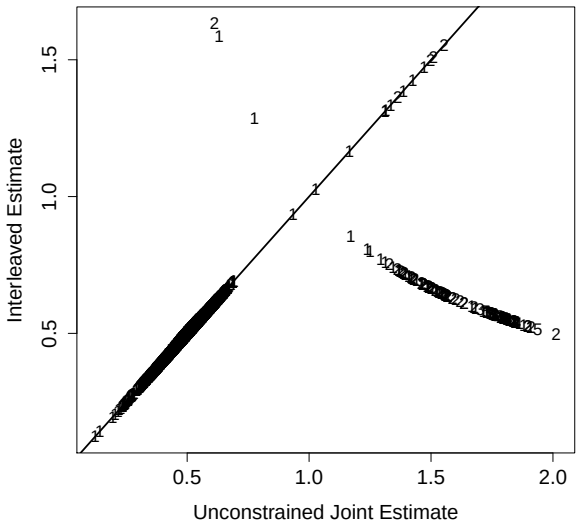
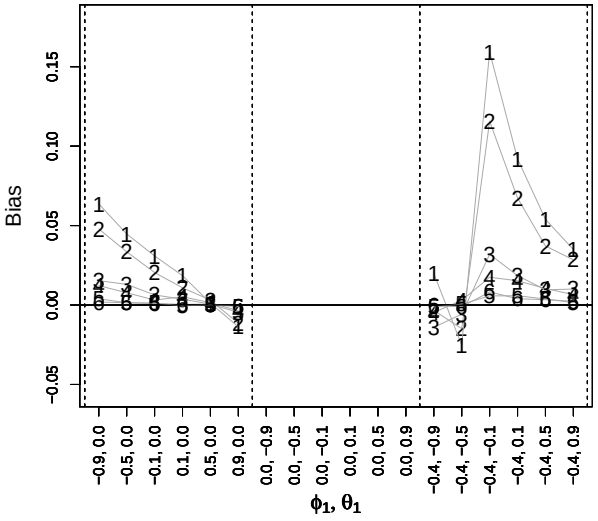
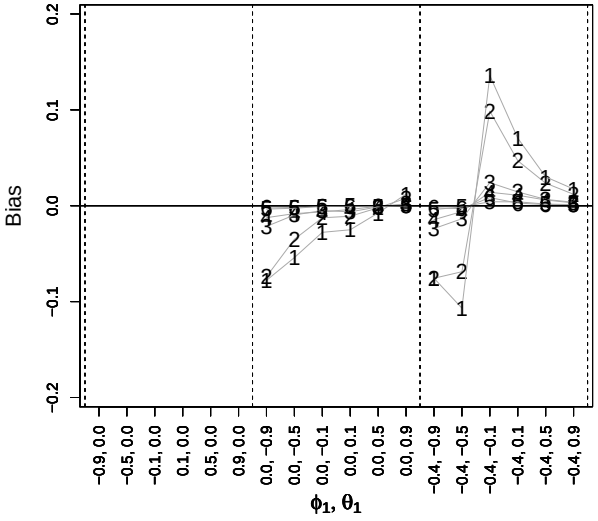


Figure 5.4.2 – Simulated Bias of MLE Parameter Estimates (AIARMA Models - $\hat{\phi}_1$, $\hat{\theta}_1$ and $\widehat{\Delta\sigma^2}$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).

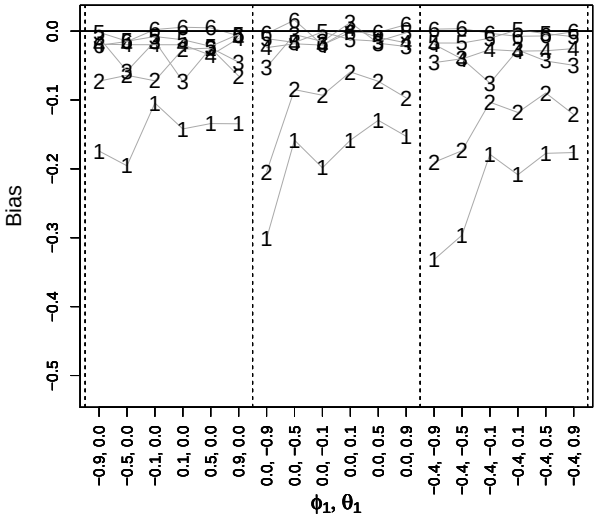
(a) $\hat{\phi}_1$



(b) $\hat{\theta}_1$



(c) $\widehat{\Delta\sigma^2}$



5.4.3 Standard Error

The standard errors for the simulations are presented in Figure 5.4.3 for $\hat{\phi}_1$, $\hat{\theta}_1$ and $\widehat{\Delta\sigma^2}$.

The standard errors of the estimated AR and MA parameters align closely to the standard errors from the RARMA and single-series ARMA models. The standard errors of the differences in the innovations variances are relatively constant across ARMA models and parametrisations.

5.4.4 Coverage

The average coverage for $\hat{\phi}_1$, $\hat{\theta}_1$ and $\widehat{\Delta\sigma^2}$ from the simulations for 95% confidence intervals are plotted in Figure 5.4.4. The AR and MA parameter results follow the coverage for the RARMA and single series ARMA models with substantial lack of coverage when $\theta_1 = -0.9$ and when the factors in B almost cancel. The coverage for the estimated difference between the innovations variances are typically close to 95% except for $n_1 = 64$ and $n_2 = 32$ where it is closer to 90%.

The coverage of the estimated mean of the first series and of the estimated difference for the second series is typically marginally below 95% (as with the RARMA and single-series ARMA processes) with more substantial reductions for small samples at $\theta_1 = -0.9$ for AR processes and for ARMA processes at $\phi_1 = -0.4$ and $\theta_1 = -0.5$ (not shown).

5.4.5 Alignment to Asymptotic Distribution

From the asymptotic variances of the estimated innovations variance in Section 4.3.4, the asymptotic variance of the estimated difference between the second and first innovations variances, $\widehat{\Delta\sigma^2}$, is readily derived as,

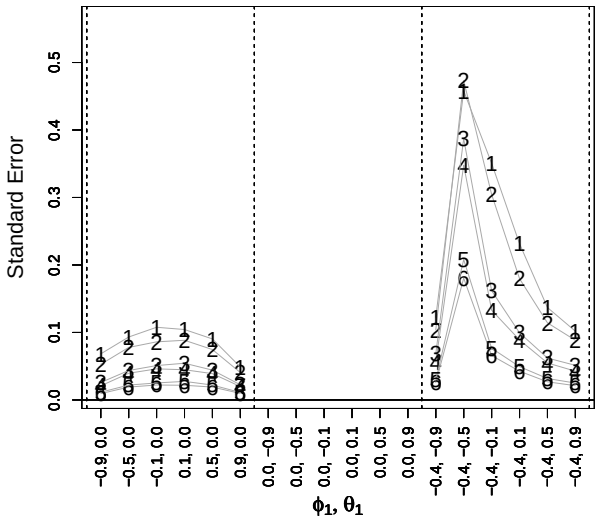
$$\text{Var}\left(\widehat{\Delta\sigma^2}\right) \;=\; \frac{2\sigma_{\epsilon_2}^4}{n_2} + \frac{2\sigma_{\epsilon_1}^4}{n_1}.$$

This is used as the asymptotic reference value to compare to the simulation results plotted in Figure 5.4.3c.

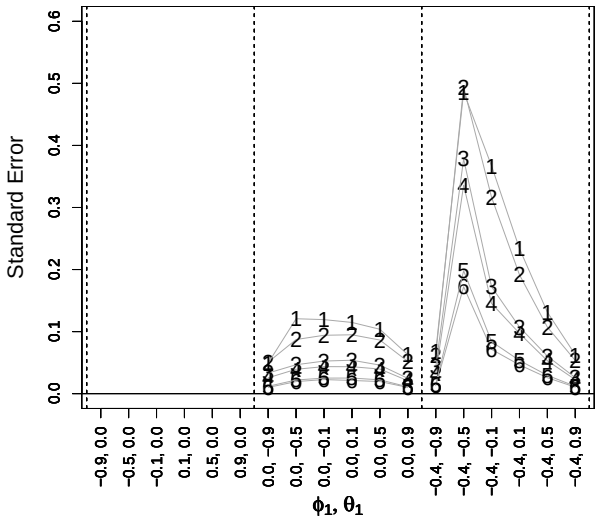
As with the RARMA and single-series ARMA results, on a proportional basis,

Figure 5.4.3 – Simulated Standard Error of MLE Parameter Estimates (AIARMA Models - $\hat{\phi}_1$, $\hat{\theta}_1$ and $\widehat{\Delta\sigma^2}$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).

(a) $\hat{\phi}_1$



(b) $\hat{\theta}_1$



(c) $\widehat{\Delta\sigma^2}$

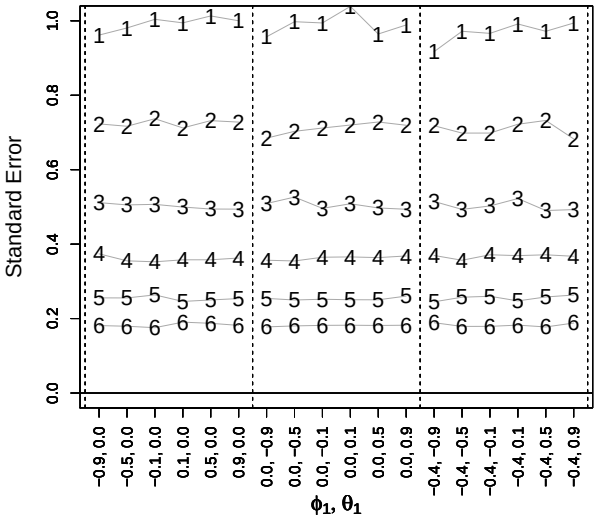
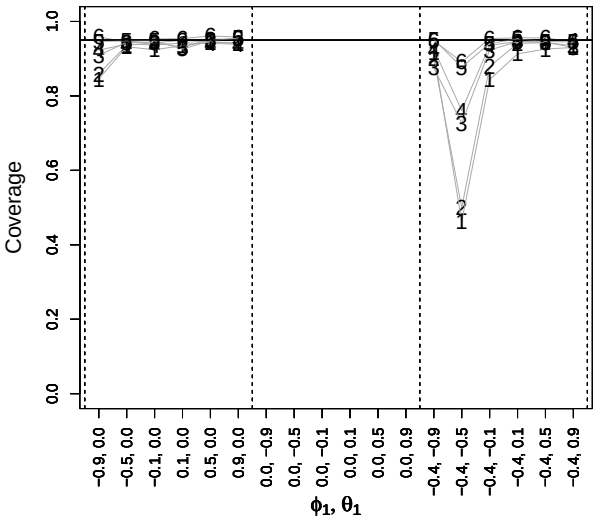
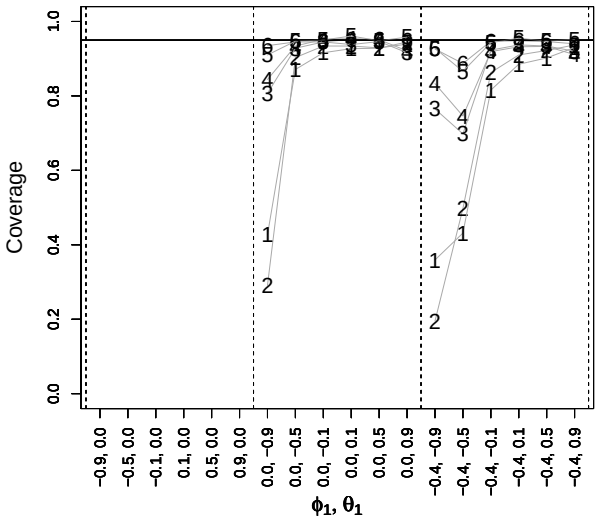


Figure 5.4.4 – Simulated Coverage of MLE Parameter Estimates (AIARMA Models - $\hat{\phi}_1$, $\hat{\theta}_1$ and $\widehat{\Delta\sigma^2}$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024). The horizontal reference line shows the 95% nominal value.

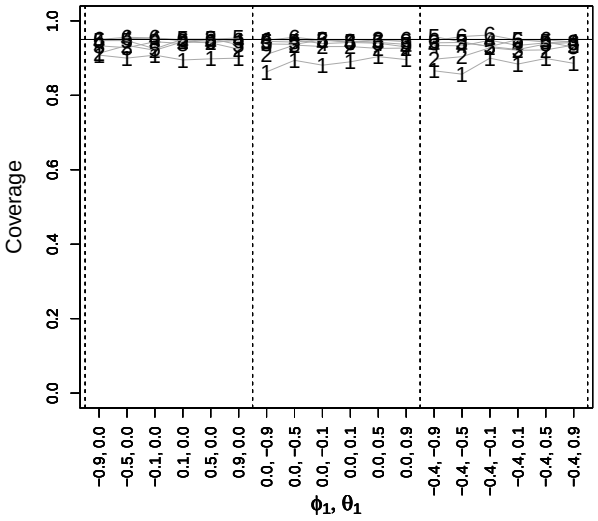
(a) $\hat{\phi}_1$



(b) $\hat{\theta}_1$



(c) $\widehat{\Delta\sigma^2}$



the asymptotic standard errors are typically within 0.10 of the simulated values except at small sample sizes, that is, $n_1 + n_2 \leq 128$, for large AR and MA coefficients (± 0.9) and for ARMA models with nearly cancelling factors in B. Again as with RARMA and single-series ARMA processes, AIARMA maximum likelihood estimates appear to approach their population values, have variance similar to their asymptotic values and have accurate empirical confidence intervals.

The asymptotic distribution therefore can be used as a reliable indicator of finite moderate to large sample behaviour. Similarly the empirical Hessian-based confidence intervals are also reliable.

5.4.6 Conclusions

As argued for RARMA and single-series ARMA models, these simulation results for the AIARMA models suggest that, except when ARMA coefficients are near the unit circle or reflect nearly cancelling factors in B, the MLE's are "well-behaved" showing only moderate bias, stable standard errors and representative empirical confidence intervals. Furthermore the empirical Hessian-based confidence intervals can be used with confidence and the constrained joint likelihood approach again produced the most stable MLE values.

As with the RARMA analysis, the simulation results (comparing Figures 5.2.3 and 5.4.3) and the asymptotic outcomes (see Section 4.3.4) suggest that AIARMA versus single-series ARMA processes of the same total length produce (for shared parameters and with the same underlying ARMA filter) maximum likelihood estimates with very similar sampling properties, at least for moderate series lengths.

5.5 CAIARMA Processes

As with ARMA, RARMA and AIARMA modelling, the following comments refer to the typical estimator performance measures, being bias, standard error, coverage and alignment to the asymptotic distribution as well as the stability of the optimisation process with a comparison of the three estimation methods.

5.5.1 Comparison of the Three Estimation Routines

The CAIARMA results generally reflect those for the RARMA, AIARMA and single-series ARMA processes. There is a confluence of outcomes between unconstrained and constrained joint likelihood and interleaved methods except at the extremes of the MA parameterisation and where the AR and MA factors in B almost cancel. Also the same likelihood is reflected for MA parameters equidistant from the unit circle. As an illustration, Figure 5.5.1 compares the results for $\hat{\theta}_1$ from the MA(1) process with $\theta_1 = 0.5$ where the line is a reference 1:1 indicator.

The optimisation routines show some convergence issues for the interleaving method. Given that the constrained joint likelihood approach has no such convergence issues, again it is chosen for the reporting the CAIARMA simulations.

5.5.2 Bias

The mean bias for $\hat{\phi}_1$, $\hat{\theta}_1$ and $\hat{c}$ from the simulations are shown in Figure 5.5.2 by process, sample size and parameterisation. Overall the bias results are very similar to those for $\hat{\phi}_1$, $\hat{\theta}_1$, $\hat{\mu}_1$ and $\widehat{\triangle\mu} = \widehat{(\mu_2 - \mu_1)}$ for single-series ARMA, RARMA and AIARMA modelling (the results for the latter two estimated parameters are not shown). For $\hat{c}$, for small sample sizes ($n_1 + n_2 \leq 128$) there is a small negative bias of up to -0.01.

5.5.3 Standard Error

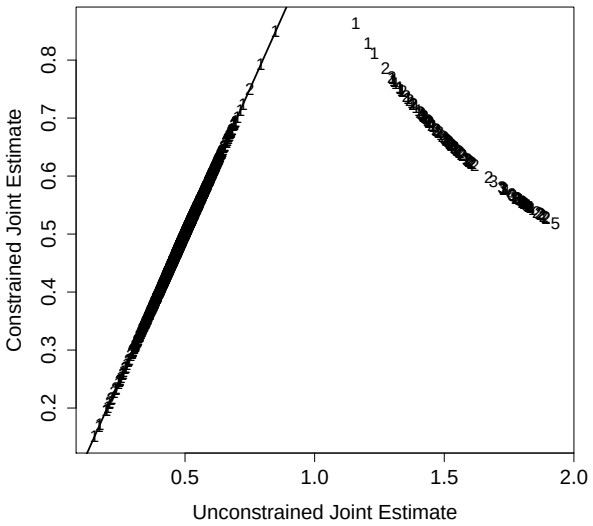
The standard errors for the simulations are presented in Figure 5.5.3 for $\hat{\phi}_1$, $\hat{\theta}_1$ and $\hat{c}$. For $\hat{\phi}_1$ and $\hat{\theta}_1$ the results closely follow those for RARMA, AIARMA and single-series ARMA processes. The standard errors of the estimated c parameter are relatively constant by sample size but show a larger value for the AR(1) process at $\phi_1 = -0.9$.

5.5.4 Coverage

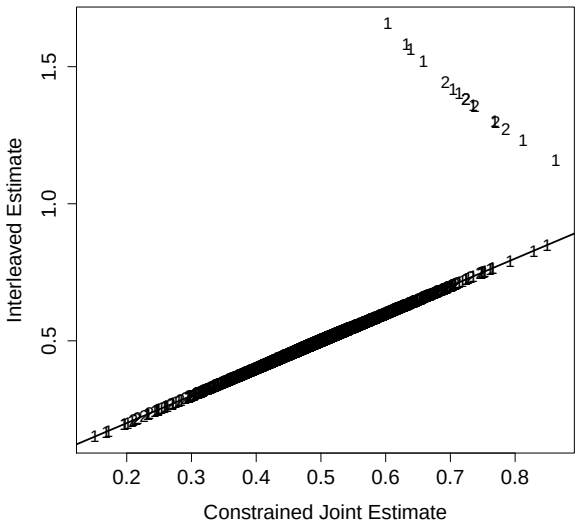
The average coverage for $\hat{\phi}_1$, $\hat{\theta}_1$ and $\hat{c}$ from the simulations for 95% confidence intervals are plotted in Figures 5.5.4. The coverage outcome is aligned with those

Figure 5.5.1 – Comparison of Simulated θ_1 Estimates (CAIARMA Models - MA(1), $\theta_1 = 0.5$). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024) with 1:1 reference line.

(a) Unconstrained versus Constrained Joint Likelihood



(b) Constrained Joint Likelihood versus Interleaved



(c) Unconstrained Joint Likelihood versus Interleaved

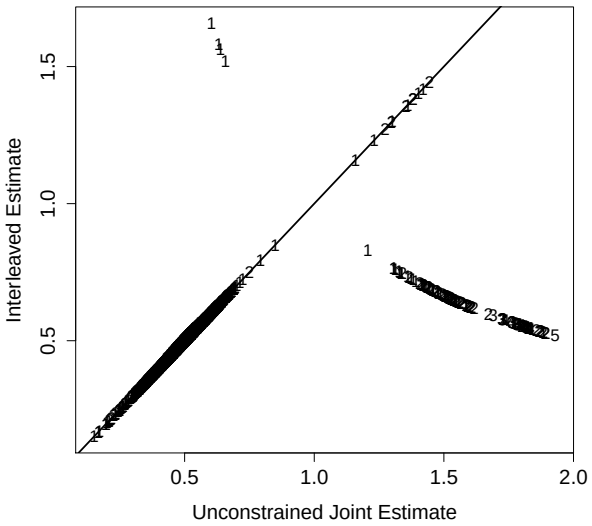
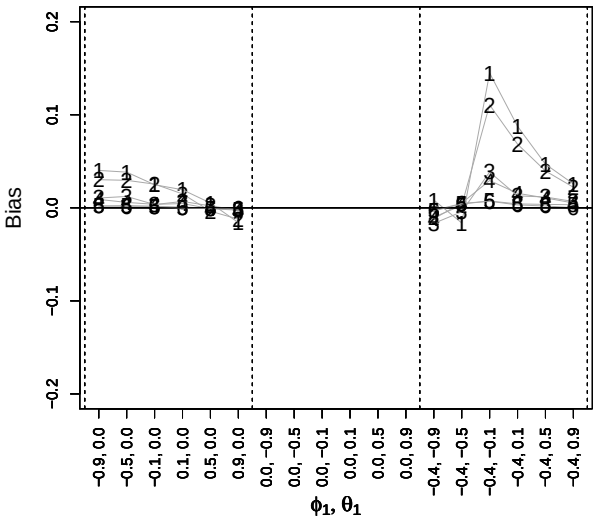
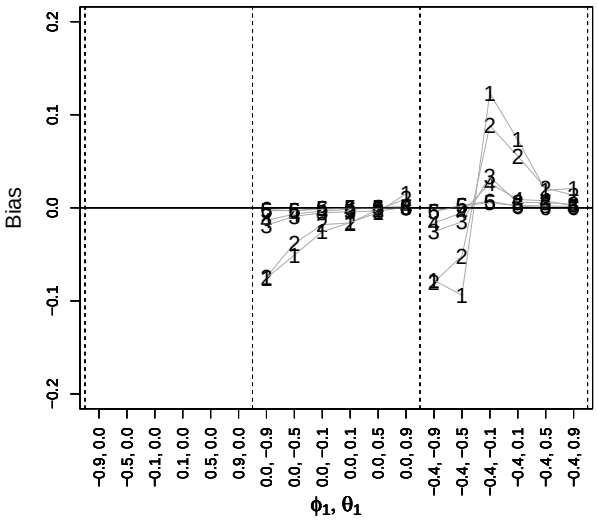


Figure 5.5.2 – Simulated Bias of MLE Parameter Estimates (CAIARMA Models - $\hat{\phi}_1$, $\hat{\theta}_1$ and $\hat{c}$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).

(a) $\hat{\phi}_1$



(b) $\hat{\theta}_1$



(c) $\hat{c}$

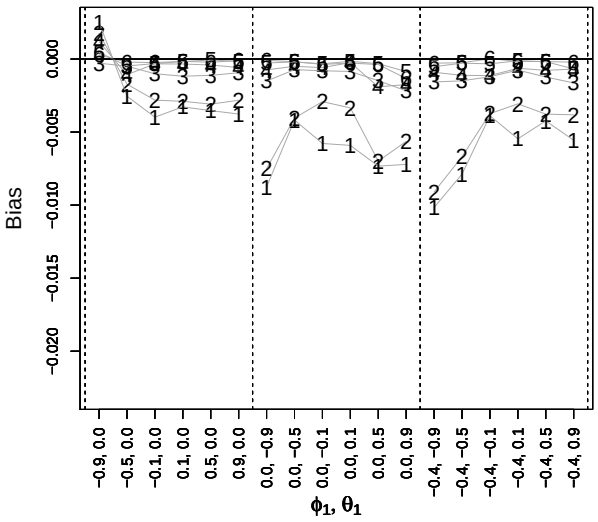
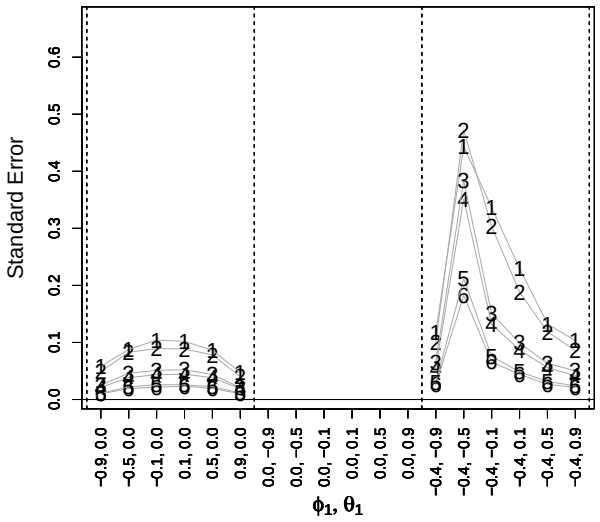
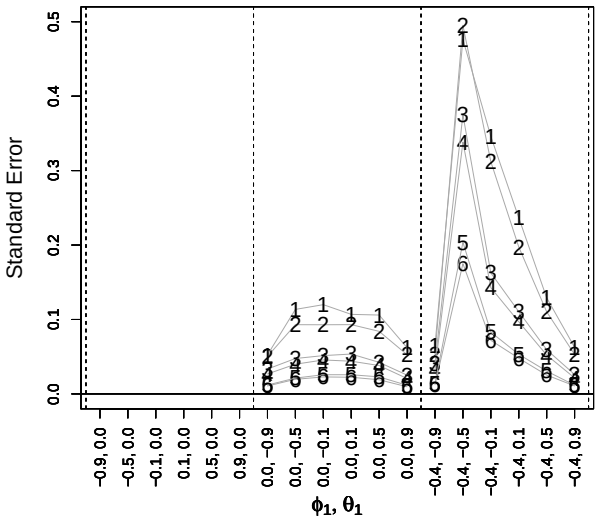


Figure 5.5.3 – Simulated Standard Error of MLE Parameter Estimates (CAIARMA Models - $\hat{\phi}_1$, $\hat{\theta}_1$ and $\hat{c}$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024).

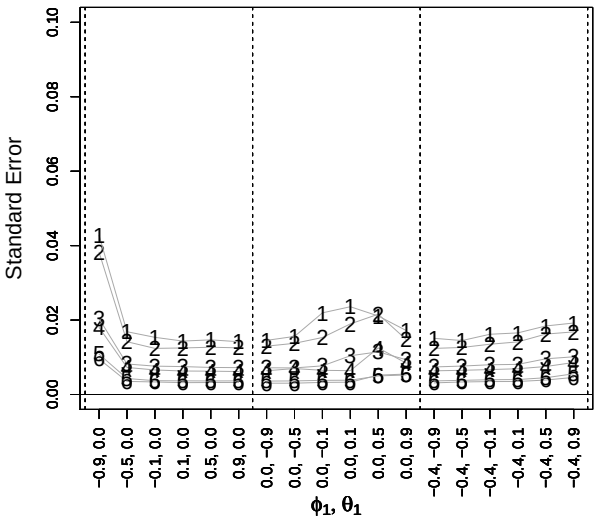
(a) $\hat{\phi}_1$



(b) $\hat{\theta}_1$



(c) $\hat{c}$



reported previously in this thesis. For $\hat{c}$, coverage is typically over 90% except when $n_1 + n_2 \leq 128$ where it drops to 86%.

5.5.5 Alignment to Asymptotic Distribution

The standard errors from the simulations are close to the asymptotic results for all parameters for most models (not shown). The only circumstances where there is substantial deviation is where the AR and MA factors in B almost cancel with $n_1 + n_2 \leq 128$, and where $|\theta_1|$ and $|\phi_1|$ are close to 0.9. As expected the simulated standard errors appear to converge to the asymtotic results as $n \to \infty$. Also there is some deviation for $\hat{c}$ for MA processes but this reflects issues with convergence of the θ_1 estimates.

Given the results on bias, coverage and asymptotic variance, the asymptotic distribution of the MLE CAIARMA estimates appear to be an accurate approximation, at least for sample sizes above 128. This also suggests that for most purposes the empirical Hessian-derived empirical confidence intervals are reliable.

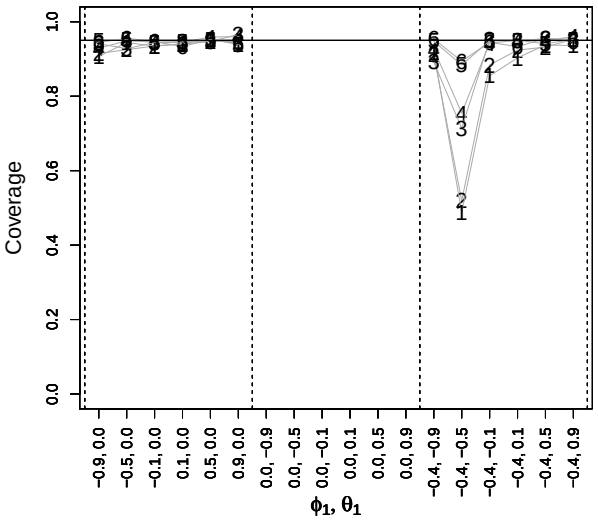
5.5.6 Conclusions

The asymptotic distribution of the MLE CAIARMA estimates appear to be a close approximation for moderate to large samples and the empirical Hessian-derived empirical confidence intervals can be used with confidence. Moreover as with the single-series ARMA, RARMA and AIARMA modelling, the constrained joint likelihood approach result in the most stable MLE values.

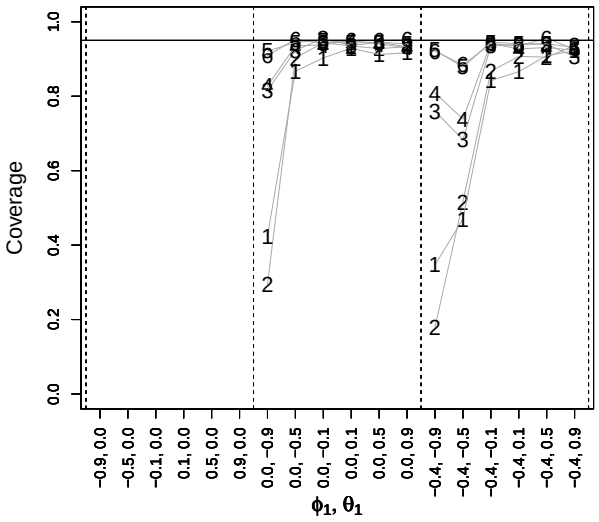
As with the RARMA and AIARMA modelling, for shared parameters and with the same underlying ARMA filter, the simulation results (comparing Figures 5.2.3 and 5.5.3) and the asymptotic outcomes (see Section 4.4.4) suggest that CAIARMA versus single-series ARMA processes of the same total length produce maximum likelihood estimates with very similar sampling properties, at least for moderate to large series lengths.

Figure 5.5.4 – Simulated Coverage of MLE Parameter Estimates (CAIARMA Models - $\hat{\phi}_1$, $\hat{\theta}_1$ and $\hat{c}$ for AR v MA v ARMA Processes). Sample size is coded as 1=(64+32) 2=(64+64) 3=(256+128) 4=(256+256) 5=(1,024+512) 6=(1,024+1,024). The horizontal reference line shows the 95% nominal value.

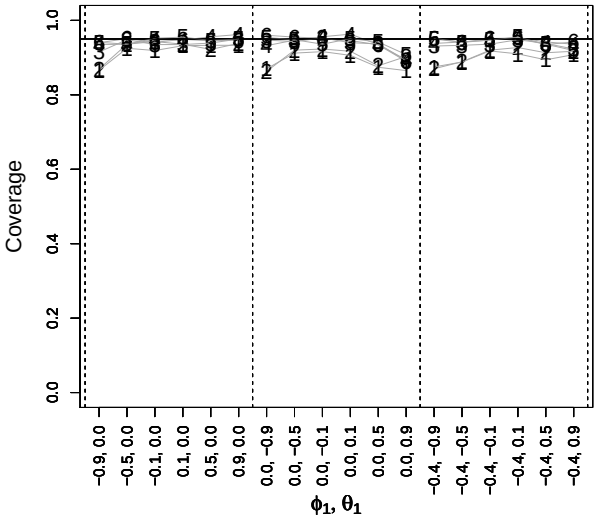
(a) $\hat{\phi}_1$



(b) $\hat{\theta}_1$



(c) $\hat{c}$



Chapter 6

Application to Daily Maximum Temperatures

This chapter contains the paper by Bowden and Clarke [2012] from the Journal of Time Series Analysis co-authored by the author of this thesis and his supervisor, Dr Brenton Clarke. It introduced the Univariate Interleaving Theorem (from Section 2.3.1) and applied it to daily maximum temperature data for Perth, Western Australia. The paper uses interleaving to capture the effect of climate and location change on over sixty years of temperature readings by week-in-the-year. It also illustrates the shortcomings of alternative methods of modelling replicated time series. Note that the paper shown here incorporates a slight correction to the original (see the strike out on the second page). Also this paper uses $\{x_t\}$ to refer to the original series and $\{y_t\}$ for the interleaved series whereas this thesis uses a reversed designation.

Original Article

Journal of
Time Series Analysis

First version received July 2011

Published online in Wiley Online Library: 29 November 2011

(wileyonlinelibrary.com) DOI: 10.1111/j.1467-9892.2011.00766.x

A single series representation of multiple independent ARMA processes

Ross S. Bowden[a] and Brenton R. Clarke[a,*,†]

This article shows that multiple independent time series from the same ARMA process can be represented by a single univariate ARMA time series through an interleaving of the original series. Using this result, existing univariate modelling software can be used to fit a single ARMA time series model simultaneously to multiple independent realizations of the same ARMA process. The interleaving approach and its properties will be presented and compared with alternative estimation options. It will be applied to the modelling of 66 years of daily maximum temperatures for Perth, Western Australia and to other time series models.

Keywords: Univariate ARMA; interleaving; simultaneous estimation; multiple time series.

1. INTRODUCTION

In analysing ARMA time series, it is typically assumed that only one realization is available for model fitting (see Anderson, 1976, p. 1). However, there are many circumstances where multiple realizations of the same ARMA process are available. Examples can be found in recordings from replicated experiments on evolving chemical processes, repeated measures on individuals recorded at fixed time intervals, daily weather data for the same season over many years and annual economic growth rates for various industry sectors over several years. A common ARMA generating mechanism is likely to apply to each realization (possibly differing in mean level) with one model to be fitted to all realizations simultaneously.

Repeated series realizations present particular issues with ARMA model fitting. Standard time series software only allows for one realization of the process (see PROC ARIMA in SAS/ETS (SAS, 2004), function 'arima' in R's 'stats' package (R Development Core Team, 2010) and the 'Forecasting' module in SPSS (SPSS, 2008)). There are a number of alternative fitting options that could be applied including concatenation of the series end-to-end, representation as a multiple time series (vector ARMA) process, and averaging across each time index. As discussed in Section 3.2, these typically involve either less than fully efficient estimation, result in biased estimation or require parameter constraints that are not commonly available in current software.

To permit fully efficient estimation, the 'interleaving method' is introduced. This is based on an equivalent representation of multiple independent (identical) ARMA processes as a single univariate ARMA process. It uses readily available software (as in R and SAS) and makes full provision for the data in all the series. It can be immediately extended to fitting one model to repeated realisations of multiple ARMA processes. It can also incorporate intervention and other extraneous variables to model the level of the replicated series, within and between series.

This article reviews the literature on simultaneous ARMA time series estimation in Section 2. In Section 3, the interleaving method will be presented and compared to other estimation options. In Section 4 the method will be applied to 66 years of daily maximum temperature readings which will be represented using AR(2) models applied separately to each week in the year. The article concludes in Section 5 with comments on extending the method to other time series models.

2. REVIEW OF THE LITERATURE

There is a relative dearth of literature on fitting a single time series model to multiple realizations of the same ARMA process and the articles reviewed here are from a wide variety of sources.

An original reference on simultaneous estimation in a repeated measures environment is Yates (1960). He presents a correction for first order autocorrelation in the analysis of repeated measures in sample surveys and this is further explored in Scott and Smith (1974).

Anderson (1978) considered the case of first-order vector AR models where the available time series are short in length but have multiple realizations. Maximum likelihood parameter estimates are derived where the parameters are either constant over time or

[a]*Murdoch University*
**Correspondence to: Brenton R. Clarke, Mathematics and Statistics, School of Chemical and Mathematical Sciences, Faculty of Science, Engineering and Sustainability, Murdoch University, Murdoch, Australia, 6150.*
[†]*E-mail: b.clarke@murdoch.edu.au*

allowed to vary by time interval and by process (i.e. treatment). The latter makes use of the multiple observations on each time interval.

Azzalini (1981) examines the fitting of a single model to replicated series from an autoregressive process of order one or two. The emphasis is on asymptotic efficiency where the number of replications (as opposed to the length of each series) tends to infinity. The conditional and unconditional maximum likelihood results are derived and compared. The later are shown to be a substantially superior result especially near the stationarity boundary. Azzalini (1984) enhances the modelling to incorporate what is effectively 'measurement-with-error' modelling in a random-effects two-way ANOVA setting. The results are applied to the plasma citrate concentration of $n = 10$ subjects measured at 14 equal time points to detect changes in the plasma readings during the day.

Wong *et al.* (2002) in an extension of Wong and Miller (1990) model repeated realisations of ARMA processes where the error variance and the number of realizations are allowed to vary over time. Also each repeated realization is assumed to be a combination of a underlying ARIMA process and an additional independent error term (ARIMAN process).[1] The model differs from those considered in this article because (i) the current models do not employ the additional noise term and (ii) Wong *et al.* (2002) constrain the underlying ARIMA process (i.e. without noise) to not only have the same parameters but the same underlying ARIMA realization. Without this constraint the model with noise wouldn't be identifiable from the autocorrelation function. Wong *et al.* (2002) fit the ARIMAN model using maximum likelihood in a state space representation.

Based on extensions of repeated measures models, Diggle *et al.* (2002) looked at the consequences of autocorrelation amongst the errors of these processes but limited their analysis to AR(1) models which are fitted using least squares. Shi and Chaganty (2004) in a similar context compare maximum likelihood, Yule–Walker and quasi-least squares estimates for autoregressive models for errors within regression models. Browne and Zhang (2007) discuss fitting modified univariate AR models of order p to observations taken over time on a number of individuals. An independent error term for each observation on each individual is added to the standard AR model which also incorporates an initialization of the process for the first p observations using a so-called initial state vector. If specified a priori, this effectively transforms the system into a conditional model. A comparison of the maximum likelihood fit of conditional and unconditional models in the context of differences between individuals indicates that the later is a superior result.

Peiris *et al.* (2003) working with short time series of medical observations on individuals discuss the maximum likelihood fit to a replicated AR(1) process, each realization being of equal length ~~but incorporating random contamination.~~ They derive the conditional and exact likelihood function for the AR(1) ~~plus error term~~ model and run a simulation study to assess bias and efficiency. They conclude that the conditional and exact maximum likelihood estimates are unbiased. However it is likely that the results are only asymptotically unbiased given the large sample size ($n = 100$) and the fact that it is well-known that maximum likelihood parameter estimates for AR models are (substantially) biased for even moderate sample sizes (see Shaman and Stine (1988)).

Quinn (2006) discusses the maximum likelihood estimation of common AR models in the context of testing for spectrum change. Quinn's approach differs from the models in the current article because Quinn allows the innovations variance to vary between the series. This suggests an area of further research, into the unconditional likelihood of a replicated ARIMA process.

For additional references on replicated time series, see also Ledolter and Chang-Soo (1993), Nandram and Petruccelli (1997) and Cipra (1999) who adopt a Bayesian approach to fitting first order autoregressive models to replicated series.

3. PARAMETER ESTIMATION

Here, we shall establish a univariate representation of multiple independent ARMA processes. This method is more convenient for model fitting and is compared to alternative model fitting approaches.

3.1. Interleaving method

Let the ith repeated series ($i = 1, \ldots, m$) over the time span, $t = 1, \ldots, n$, be $\{x_{i,t}\}_{t=1}^{n}$ and assume each series is generated by the following ARMAX(p, q) process,

$$\boldsymbol{\phi}(B)(x_{i,t} - \mu(\mathbf{z}_{i,t})) = \theta(B)a_{i,t}, \tag{1}$$

where $\{a_{i,t}\}_{i=1}^{n}$ is a series of i.i.d. random errors with constant variance, σ_a^2, $E(a_{i,t}) = 0$ for all i and t, $E(a_{i,\,t}a_{j,u}) = 0$ for all t, u and $i \neq j$ and $E(a_{i,t}\,a_{i,u}) = 0$ for all i and $t \neq u$. Also $\phi(B)$ and $\theta(B)$ are polynomials in B, the backshift operator, of order p and q, respectively, and

$$\mu(\mathbf{z}_{i,t}) = E(x_{i,t} \,|\, \mathbf{z}_{i,t}) = \sum_{k=1}^{l} \psi_k z_{k,i,t},$$

where $\{\mathbf{z}_{i,t}\}_{t=1}^{n}$, $i = 1, \ldots, m$, are m series of explanatory vectors with kth element, $z_{k,i,t}$.

We will call this a RARMA (replicated ARMA) process i.e. RARMA(p,q,m). It has a mean which can vary with the series realization but otherwise maintains a consistent generating process between realizations. In fact in general the mean can be any linear combination of the extraneous vector variables, $\mathbf{z}_{i,t}$.

We now state and prove a theorem of equivalence between multiple and univariate representations of this process.

THEOREM 1. *Let $x_{i,t}$ be generated by the above RARMA(p,q,m) process, and let,*

$$y_{m(t-1)+i} = x_{i,t}$$
$$\mathbf{w}_{m(t-1)+i} = \mathbf{z}_{i,t} \text{ and}$$
$$\epsilon_{m(t-1)+i} = a_{i,t}.$$

Then,

$$\phi(B^m)(y_s - \mu(\mathbf{w}_s)) = \theta(B^m)\epsilon_s, \tag{2}$$

where $E(\epsilon_s) = 0, \mathrm{E}(\epsilon_s^2) = \sigma_a^2$ and $E(\epsilon_s\epsilon_r) = 0, s \neq r$.

PROOF. Consider the autoregressive and moving average difference equation, eqn (2) and select all s such that $s \mid m$ (i.e. s *modulo* m) equals some constant i. We then have $y_s = x_{i,t}$, $\epsilon_s = a_{i,t}$, and $\mathbf{w}_s = \mathbf{z}_{i,t}$, $t = 1, \ldots, n$, and, setting $D = B^m$,

$$\phi(D)(x_{i,t} - \mu(\mathbf{w}_t)) = \theta(D)a_{i,t}, \tag{3}$$

where D is equivalent to a one-lag backshift operator on $x_{i,t}$ i.e. $D(x_{i,t}) = x_{i,(t-1)}$. We note that, from the specification of the interleaved model (2) alone, $E(\epsilon_s) = 0$ i.e $E(a_{i,t}) = 0$, and $E(\epsilon_s^2) = \sigma_a^2$ i.e. $E(a_{i,t}^2) = \sigma_a^2$. If we now choose r such that $r \mid m = j \neq i$ then $E(\epsilon_s\epsilon_r) = 0$ i.e. $E(a_{i,t}a_{j,u}) = 0$ for all $i \neq j$ and all t,u. This completes the proof of equivalence between representations, eqns (2) and (1).

The above formulation permits m multiple independent ARMA(p,q) processes with identical parameters – but different realisations – to be modelled as a single univariate ARMA(pm,qm) process. The interleaving used in Theorem 1 is illustrated in Figure 1. In this example, it is assumed that there are daily maximum temperature values available for the first week in the year (with seven artificial readings) for 2 years, 1946 and 1947. The two series are interleaved to create a final single series of length fourteen.[2]

The interleaving formulation is equivalent to eqn (1) but provides significant advantages for fitting a single ARMA model to multiple series realizations. An ARMA(pm,qm) model can now be fitted to the data using standard time series software where the data is interleaved placing the tth data points from each realization adjacent to each other in a consistent order. All AR and MA parameters are set to zero for all orders apart from multiples of m, the number of realizations. The error term (as in most packages assumed to be normally distributed) retains its original ARMA definition as do the parameter estimates.

Most ARMA packages such as R's 'arima' in the 'stats' package allow specification of which parameters are non-zero and hence the model expression suggested above can be presented in this form. Furthermore, standard tests of significance and standard diagnostics can be applied to the results using lags which are multiples of m.

As an example, let there be m=10 realizations of a time series process, each of length 25 (although see Section 5), and the model to be fitted is ARMA(2,1). After interleaving the series to create a new series of length 250, the model to be fitted would be ARMA(20,10) with,

$$\phi(B^{10}) = (1 + \phi_1 B^{10} + \phi_2 B^{20}) \text{ and } \theta(B^{10}) = (1 + \theta_1 B^{10}).$$

In the estimation software, the autoregressive lags of 1 – 9 and 11 – 19 are fixed at zero as are the moving average lags of 1–9.

Given the possibly large lags involved in the interleaving method, it is worth exploring whether there may be computational issues when fitting ARMA models. The R 'arima' function was accessed in this regard using simulations. In the 'arima' code, the estimated interleaved model is tested for stationarity using the function 'polyroot'. For, say, an interleaved AR(1) model the polynomial, $(1 - \phi_1 B^m)$, in B is checked for stationary by ensuring that all the moduli of the roots of the polynomial are greater than one.

The exact moduli are all equal to $\frac{1}{\sqrt[m]{\phi_1}}$ and will be greater than one for all $|\phi_1|$ less than one. For lags less than approximately 65 the 'polyroot' moduli converged towards one from above. However, thereafter the roots and hence moduli become unstable and some moduli drop below one, even for values of ϕ_1 close to zero, for example, $\phi_1 = 0.05$. In the current modelling this was remedied by substituting the following function for 'polyroot' which collapses the polynomial in B^m to one in say D involving much smaller orders (e.g., $(1 - \phi_1 D^1)$),

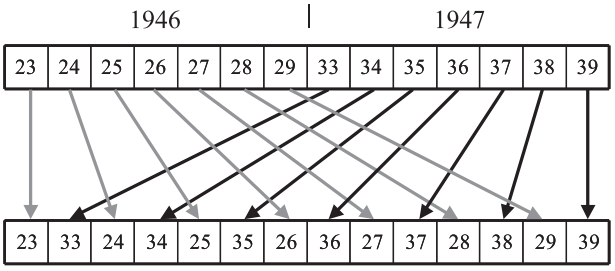


Figure 1. Interleaving of an artificial series

```
polyroot <- function(x) {
   indices <- which(x[-1] !=0)
   if (all (indices%%indices[1] ==0))
   return(base::polyroot(c(x[1],x[-1][ indices] ))) else
   return(base::polyroot(x))
   }
```

Over and above the false identification of non-stationarity for large m, simulations show that the R 'arima' package when used with the interleaved method ($m = 60$ and 120) converges to the true parameter values for a range of AR(2), MA(2) and ARMA(2,2) parameterisations. Hence convergence doesn't appear to be a major problem (this is reinforced by the results of the model fitting in Section 4) although some memory management issues developed for very large lags ($m = 200$).

In Section 4, the interleaving method will be used to fit one ARMA model to multiple realisations of daily weather data from 66 years of recordings. However, before proceeding we shall examine alternative methods of fitting parameters for model (1).

3.2. Alternative estimation methods

An alternative representation of a RARMA process is as a constrained VARMA model (Quenouille, 1957 and Lutkepohl, 1991) with constraints on the parameter and error covariance matrices. To frame (1) as a multiple time series process, we form an m-dimensional vector of 'replicated' observations, $\upsilon_t = (x_{1,t}, \ldots, x_{m,t})$. This doesn't imply that these observations are concurrent in real time but rather are assigned to the same time index, t, for estimation purposes. We assume that $\{\upsilon_t; t = 1, \ldots, n\}$ follows a standard VARMA time series process generated by:

$$\boldsymbol{\phi}(B)(\boldsymbol{\upsilon}_t - \boldsymbol{\pi}(\mathbf{z}_t)) = \boldsymbol{\theta}(B)\mathbf{a}_t,$$

where $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are matrix polynomials in B, the backshift operator, $\{\mathbf{a_t}\}_{t=1}^{n}$ is an m-dimensional white noise process with covariance matrix, $\boldsymbol{\Sigma}$, and $\boldsymbol{\pi}(\mathbf{z}_t) = E(\upsilon_t \mid \mathbf{z}_t) = \boldsymbol{\psi}\mathbf{z}_t$ where $\boldsymbol{\psi}$ is a $m \times r$ matrix of coefficients and $\mathbf{z}_t$ is a vector of extraneous variables of length r. To ensure unique identification of $\boldsymbol{\phi}(B)$, $\boldsymbol{\theta}(B)$ and $\boldsymbol{\Sigma}$ from the covariance matrix of υ_t, the zero'th order matrices of $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are $m \times m$ identity matrices.

To model a set of replicated time series from the same ARMA process, we now constrain $\boldsymbol{\phi}(B)$, $\boldsymbol{\theta}(B)$ and $\boldsymbol{\Sigma}$ to be diagonal matrices. This results in a set of m independent univariate ARMA processes. We further constrain each diagonal element of $\boldsymbol{\phi}(B)$ to equal $\phi(B)$, each diagonal element of $\boldsymbol{\theta}(B)$ to equal $\theta(B)$ and the diagonal elements of $\boldsymbol{\Sigma}$ to equal σ_a^2. This completes the representation of the $\{\boldsymbol{\upsilon}_t\}_{t=1}^{n}$ as a set of independent time series generated using eqn (1).

This model could be fitted using appropriate VARMA modelling software if such software were available. However the ability to constrain the VARMA model terms as required is not commonly to hand. SAS/VARMAX (the suffix 'X' indicates that the modelling can incorporate extraneous explanatory variables) allows constraining of parameter values to constants but doesn't permit specification of linear relationships (e.g. equivalence) between parameters (SAS (2004)). SSATS in GAUSS (GAUSS, 2010) allows estimation of VARMA models via state space models but doesn't have facilities for parameter constraints. The package GROCER for Scilab (Scilab, 2010) provides VARMA modelling capability (which can be used within R via R/SCILAB) but with similar shortcomings to SSATS. R fits VARMA models using the DSE package which utilizes a state space representation. Parameter constraints cannot be imposed as required above.

Using existing univariate software, there are a number of possibilities for fitting a single time series model to replicated ARMA time series. Perhaps the most obvious is to join the time series head-to-tail to create one long series and fit the required ARMA model to the result (herein called the 'concatenation method'). However, at the joins of the time series there will be a disruption to the autocorrelation structure in the data which is not represented in the ARMA model. The fitted model's parameters will be biased towards zero.

Alternatively, the elements of each time series could be averaged across each time point to produce one series for analysis. That is, a new series, $\{\hat{x}_t\}_{t=1}^{n}$, could be formed where

$$\hat{x}_t = \frac{1}{m}\sum_{i=1}^{m} x_{i,t}, t = 1, \ldots, n.$$

Given that each series, $\{x_{i,t}\}_{t=1}^{n}, i = 1, \ldots, m$, is assumed to be generated by the same ARMA(p,q) process, their sum will also be generated by an ARMA(p,q) process with the same parameters except that the error variance will be $\frac{1}{m}$ times[3] the original error variance, σ_a^2. The suggestion here is that the reduced error variance will lead to similar reductions in the standard errors of the parameter estimates. However, it is well known that the standard errors of the least squares and maximum likelihood parameter estimates of an ARMA process are independent of the error variance (Anderson, 1976 and Godolphin, 1984). Moreover the bias of the common parameter estimates are also independent of the error variance (see Shaman and Stine, 1988). Hence there is no advantage gained by averaging the series given that this results in no reduction in either the standard errors or biases of the parameter estimates fitted to the individual replicated series.

A final alternative method of estimation is to fit the required ARMA models independently to each replicated series in turn and then average the results with a commensurate reduction in the standard errors of the estimates. The parameter estimates are likely to be biased for each estimation due to the typically small sample sizes for each replication. Averaging will not correct this.

Accordingly, the interleaving method from Section 3.1 is the only approach that allows the use of readily available software to produce fully efficient parameter estimates for a replicated independent ARMA process. The remainder of this article will apply the interleaving method to daily maximum temperature data, compare the results to the concatenation method and discuss the application of the new method to other time series models.

4. AN AR MODEL FOR SIXTY-SIX YEARS OF DAILY MAXIMUM TEMPERATURES IN PERTH, WESTERN AUSTRALIA

It is proposed to model the daily maximum temperature readings for Perth, Western Australia using the RARMA model. The time series modelling of daily weather data can be used to better understand a number of weather-dependent activities (for example, daily maximum electricity demand). However, such a time series process is likely to change over the year. Accordingly, it would be useful to focus on daily data for individual weeks in deriving a model where the first week extends from the 1st to the 7th January and succeeding weeks continue in 7 day contiguous intervals.

An historical series of daily weather readings was extracted for Perth, Western Australia, from 1944 to 2009. The interleaving method allows the use of all years of this data simultaneously to fit a unique ARMA model for each week in the year and to derive a set of fully efficient weekly parameter estimates. These are not likely to be substantially affected by estimation bias because of the large combined sample size of each interleaved series ($n = 7 \times 66 = 462$).

Data after 1963 were possibly affected by a change in recording site which will be investigated using the interleaving method and intervention analysis. Initially only the data for the first week in the year will be modelled but the interleaving analysis will then be extended to every full week of data.

We use the terminology and formulation from eqn (2) where $m = 66$. To assess the order of the AR and MA models the daily readings for the first week were interleaved ($n = 462$). A logarithmic transformation was considered but the distribution of the temperatures did not show sufficient skewness. The sample autocorrelation and partial autocorrelation functions are shown in Figure 2 and have been produced for lags up to 198 (which is effectively a 3 day lag). ACFs and PACFs at lags other than multiples of 66 can be assumed to be zero and ignored because they represent the relationship between the lagged daily temperatures many years apart. The standard sample PACF for lag p was derived by fitting successively higher-order AR models and extracting the pth order autoregressive parameter. However, to strictly represent the nature of the interleaved time series, the sample PACF should only control for AR orders corresponding to multiples of the number of series as opposed to successively fitting AR models involving all lags up to p. Hence, the PACF for lag 132 should only use an AR model of lags 66 and 132. In general, the PACF for all lags other than multiples of 66 should be set to zero. The sample ACF and PACF results assuming interleaving are shown in Figure 2 as circled markers.

These diagnostics indicate that the underlying ARMA process is second-order autoregressive only. A similar analysis of data for each week in the year indicates that an AR(2) model is appropriate, that is, in (2), $\theta(B^{66}) = 1$ and,

$$\phi(B^{66}) = (1 + \phi_1 B^{66} + \phi_2 B^{132}).$$

Hence, we now fit a separate RAR model of order two with $m = 66$ for each of the 52 weeks (resulting in fifty-two RARMA(2,0,66) models) and the parameter estimates are shown in Figure 3. Diagnostic tests (again accommodating the interleaved nature of the residuals) suggest that this model is adequate. The grey lines are 95% confidence limits.

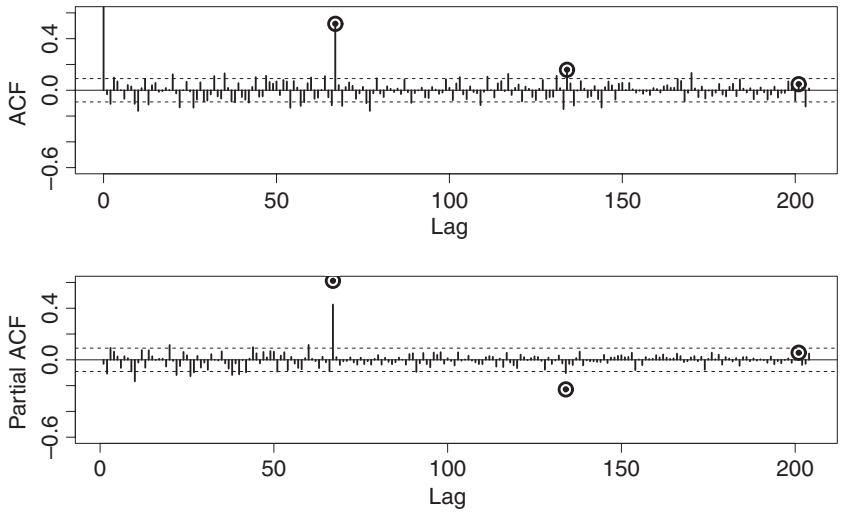


Figure 2. Sample ACF and PACF of the interleaved daily maximum temperatures for the first week. The circled results are the estimates using the interleaving method

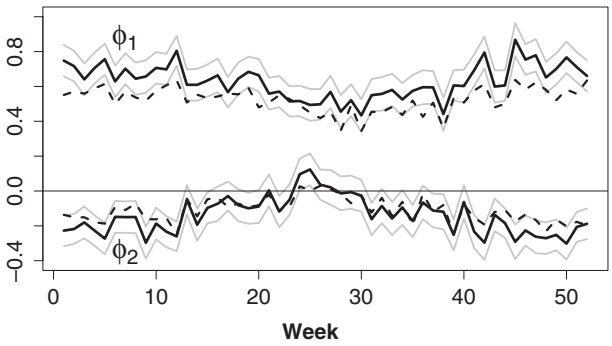


Figure 3. Estimated AR parameters by week for daily maximum temperature data (Perth, Western Australia, 1944–2009). The grey lines are 95% confidence limits and the dashed lines are the estimates from the concatenation method

Table 1. Change of location for Perth's temperature recording device

Location	Last date
King's Park	August 1963
Old Hale School	June 1967
Wellington St	May 1992
Perth Airport	November 1993
Mt Lawley	Current location

The results indicate that the lagged relationships are significantly stronger in summer than in winter. Also the standard deviation of the error term for summer months is higher than for winter reflecting the greater variability of summer temperatures (not shown). In general, the fitted models suggest that the relationship between successive daily maximum temperatures persist for only 2 days at the most and this persistence is strongest in summer.

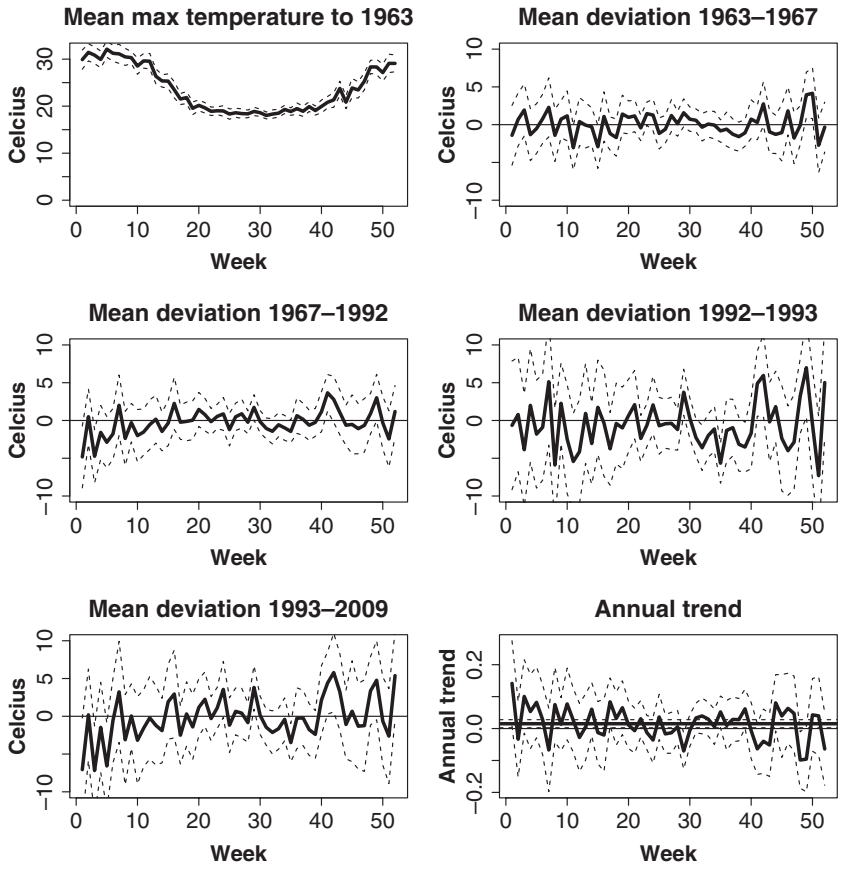


Figure 4. Estimates of effect by week of change of location and of trend over time (with average annual trend)

Figure 3 also shows the estimates using the concatenation method (dashed line) which as expected are biased towards zero.

Unfortunately, the location of the recording device for daily maximum temperatures has changed over time (see Table 1). The effect of this on the level (i.e. the mean) of daily maximum temperatures was modelled using intervention analysis with the interleaving method and the results are shown in Figure 4 with 95% confidence intervals. The top left plot shows the mean daily maximum temperature by week prior to 1963. The next four plots show the estimated change by week from that base for the four shifts in location. Although the weekly estimates in the plot have relatively high standard errors it appears that the change of location to both Wellington St (in the CBD) and to Mt Lawley (in the suburbs) resulted in lower daily maximum temperatures in the first few weeks in the year compared to the initial King's Park location. This analysis could possibly be improved using a reference site for temperature readings such as Perth Airport where the recording device has remained in a relatively unchanged location.

As a final demonstration of the potential of the interleaving method, a constant trend was fitted through the 66 years of daily data by week to model the effect of climate change. The annual trend estimates by week are plotted in the bottom right plot in Figure 4 with 95% confidence intervals. The average trend value shown as a reference line is +0.015°C per year (with 95% confidence bounds of ±0.013°C) or +0.15°C per decade. Over the 66 years of data, this equates to a 0.99°C increase in daily maximum temperatures.

5. CONCLUSIONS AND FURTHER APPLICATIONS

The model specification contained in this article and its application to daily maximum temperatures by week suggest that the interleaving technique is effective in fitting a single standard ARMA model to multiple independent realizations of a time series process. It is simple to use and can be implemented using standard ARMA software packages. The only requirement is that the package allows ARMA parameters to be set to zero. A problem was noted in R in checking stationarity with the 'polyroot' function but this was remedied by some simple code substitution.

The interleaving technique has been applied to AR models in Section 5 but can be used with MA and ARMA models (and extraneous variables, if required). Furthermore the technique readily extends to differenced series, seasonal ARMA models, VARMA models and GARCH models. It can be used with fractional differencing but, instead of the formulation $(1 - B)^{\lambda}$ for each series, the interleaved series must be modelled as $(1 - B^m)^{\lambda}$ (so called 'seasonal fractional differencing').

A constraint on the interleaving technique is that the repeated series must be of the equal length. However, each series can be extended to the same length by adding missing values. The models can then be fitted as above (most ARMA software packages such as R's 'stats' can accept missing values within the series).

Acknowledgement

The authors acknowledge the assistance of Professor John Henstridge who was Mr Bowden's Masters supervisor in the 1980s.

NOTES

1. ARMA processes with added noise are a particular form of aggregated ARMA process as discussed by Box and Jenkins (1976), Anderson (1975) and Granger and Morris [1979]. See also Ansley *et al.* (1976), Engel (1984) and Granger (1988). The sum of ARMA processes are themselves ARMA processes of higher but finite order which is proved in an elegant manner by Anderson (1975) (for MA processes) and Granger and Morris (1979) (for ARMA processes). This result is used by Anderson (1976) to show how ARMA processes of higher order can eventuate when it may be difficult to find a clear generating mechanism using, say, economic theory.
2. If the model's AR or MA parameters do vary with the series realisation then this can be represented as a periodic ARMA process (see Parzen and Pagano (1979) and Bowden (1989)) which can in turn be represented as a VARMA model.
3. That is,

$$\frac{1}{m}\sum_{i=1}^{m}\phi(B)(x_{i,t}-\mu(\mathbf{z}_{i,t}))=\frac{1}{m}\sum_{i=1}^{m}\theta(B)a_{i,t}$$
$$\phi(B)(\hat{x}_t-\mu(\hat{\mathbf{z}}_t))=\theta(B)\hat{a}_t,$$

where $\mu(.)$ is a linear function, $\hat{\mathbf{z}}_t = \frac{1}{m}\sum_{i=1}^{m}\mathbf{z}_{i,t}$ and $\hat{a}_t = \frac{1}{m}\sum_{i=1}^{m}a_{i,t}$ with $\mathrm{Var}(\hat{a}_t) = \frac{\sigma_a^2}{m}$ and $E(\hat{a}_t\hat{a}_u) = 0,\ t \neq u$.

REFERENCES

Anderson, O. D. (1975) On a lemma associated with Box, Jenkins and Granger. *Journal of Econometrics* **3**, 151–6.
Anderson, O. D. (1976) *Time Series Analysis and Forecasting The Box-Jenkins Approach*. London: Butterworths.
Anderson, T. W. (1978) Repeated measurements on autoregressive processes. *Journal of American Statistical Association* **73**, 371–8.
Ansley, C. F., Spivey, W. A. and Wrobleski, W. J. (1976) *On the Structure of Moving Average Processes*. Technical Report, Graduate School of Business Administration, The University of Michigan.
Azzalini, A. (1981) Replicated observations of low order autoregressive time series. *Journal of Time Series Analysis* **2**, 63–70.
Azzalini, A. (1984) Estimation and hypothesis testing for collections of autoregressive time series. *Biometrika* **71**, 85–90.

Bowden, R. (1989) *Forecasting Electricity Demand*. Master's thesis, University of Western Australia.
Box, G. E. P. and Jenkins, G. M. (1976) *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
Browne, M. W. and Zhang, G. (2007) Repeated time series models for learning data. In *Data Analytic Techniques for Dynamical Systems in the Social and Behavioural Sciences* (eds S. M. Boker and M. J. Wenger), Lawrence Erlbaum Associates, Mahwah, NJ, pp. 25–46.
Cipra, T. (1999) Many short time series with outliers and missing observations. *Proceedings of the 52nd Session of the International Statistical Institute*, Helsinki.
Diggle, P. J., Heargerty, P. J., Liang, K-Y. and Zeger, S. L. (2002) *Analysis of Longitudinal Data*, 2edn. Oxford University Press, Oxford.
Engel, E. M. R. A. (1984) A unified approach to the study of sums, products, time-aggregation and other functions of arma processes. *Journal of Time Series Analysis* **5**, 159–71.
GAUSS (2010) *GAUSS Version 10.* Aptech systems, Inc., Black Diamond, WA, USA.
Godolphin, E. J. (1984) A direct representation for the large-sample maximum likelihood estimator of a gaussian autoregressive-moving average process. *Biometrika* **71**, 281–9.
Granger, C. W. J. (1988) Aggregation of time series variables – a survey. *Technical Report*, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis.
Granger, C. W. J. and Morris, M. J. (1979) Time series modelling interpretation. *Journal of the Royal Statistical Society Series A* **139**, 246–57.
Ledolter, J. and Chang-Soo, L. (1993) Analysis of many short time sequences: forecast improvements achieved by shrinkage. *Journal of Forecasting* **12**, 1–11.
Lutkepohl, H. (1991) *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.
Nandram, B. and Petruccelli, J. D. (1997) A bayesian analysis of autoregressive time series panel data. *Journal of Business and Economic Statistics* **15**, 328–34.
Parzen, E. and Pagano, M. (1979) An approach to modelling seasonally stationary time series. *Journal of Econometrics* **9**, 137–53.
Peiris, S., Mellor, R. and Ainkaran, P. (2003) Maximum likelihood estimation for short time series with replicated observations : a simulation study. *Interstat* **9**, 1–15.
Quenouille, M. H. (1957) *The Analysis of Multiple Time-Series*. Charles Griffon and Co. Ltd, London.
Quinn, B. G. (2006) Statistical methods of spectrum change detection. *Digital Signal Processing* **16**, 588–96.
R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org.
SAS (2004) SAS/ETS 9.1 *User's Guide, Volumes 1–4*, Cary, NC, USA.
SCILAB (2010) *Scilab version 5.2*, http:www.scilab.org/products/scilab (last accessed date October 2, 2011).
Scott, A. J. and Smith, T. M. F. (1974) Analysis of repeated surveys using time series methods. *Journal of the American Statistical Association* **69**, 674–8.
Shaman, P. and Stine, R. A. (1988) The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association* **83** (403), 842–8.
Shi, G. and Chaganty, N. R. (2004) Application of quasi-least squares to analyse replicated autoregressive time series regression models. *Journal of Applied Statistics* **31** (10), 1147–56.
SPSS (2008) *SPSS Statistics 17.0*, Cary, NC, USA.
Wong, W-K. and Miller, R. B. (1990) Repeated time series analysis of ARIMA-Noise models. *Journal of Business and Economic Statistics* **8**, 243–50.
Wong, W-K., Miller, R. B. and Shrestha, K. (2002) Maximum likelihood estimation of ARMA model with error processes for replicated observations. Working Paper No. 0217. *Technical Report,* Department of Economics, National University of Singapore.
Yates, F. (1960) *Sampling Methods for Censuses and Surveys*, 3edn. Charles Griffon and Co. Ltd, London.

Chapter 7

Application to Daily Maximum and Minimum Temperatures

This chapter contains the paper by Bowden and Clarke [2017] accepted for publication in the Australian and New Zealand Journal of Statistics which extends the Univariate Interleaving Theorem from Section 2.3 to the multivariate case. Again it was co-authored by the author of this thesis and his supervisor, Dr Brenton Clarke. The multivariate theorem is applied to daily maximum and minimum temperature data for Perth, Western Australia. As with Bowden and Clarke [2012] (see Chapter 6), the purpose of the modelling is to estimate the effect of climate and location change on over sixty years of maximum and minimum temperature readings. The marginal univariate processes corresponding to the multivariate models fitted in this chapter are discussed in Appendix D. As with the paper in Chapter 6, this paper uses $\{\boldsymbol{x}_t\}$ to refer to the original series and $\{\boldsymbol{y}_t\}$ to the interleaved series whereas this thesis uses a reversed designation. In Bowden and Clarke [2012] it was incorrectly reported that sixty-six years of data were analysed whereas sixty-seven years were actually modelled.

Aust. N. Z. J. Stat. 2017 doi: 10.1111/j.1467-842X.XXX

Using multivariate time series methods to estimate location and climate change effects on temperature readings employed in electricity demand simulation

Ross S. Bowden[1] and Brenton R. Clarke[2*]

Mathematics and Statistics, School of Engineering and Information Technology, Murdoch University

Summary

Long-term historical daily temperatures are used in electricity forecasting to simulate the probability distribution of future demand but can be affected by changes in recording site and climate. This paper presents a method of adjusting for the effect of these changes on daily maximum and minimum temperatures. The adjustment technique accommodates the autocorrelated and bivariate nature of the temperature data which has not previously been taken into account. The data are from Perth, Western Australia, the main electricity demand centre for the South-West of Western Australia. The statistical modelling involves a multivariate extension of the univariate time series "interleaving method", which allows fully efficient simultaneous estimation of the parameters of replicated Vector Autoregressive Moving Average processes. Temperatures at the most recent weather recording location in Perth are shown to be significantly lower compared to previous sites. There is also evidence of long-term heating due to climate change especially for minimum temperatures.

Key words: VARMA; replicated process; data correction; forecasting; maximum likelihood

1. Introduction

This paper uses time series methods to estimate and adjust for the effect of changes of location and climate on daily maximum and minimum temperature data for Perth, Western Australia, as used in electricity demand simulation for the South-West of Western Australia. In undertaking the estimation, this paper also explores the temperature data's stochastic generating mechanism which is one of the primary determinants of daily electricity demand. This mechanism could be used for temperature simulation but it is not the main purpose of this paper. Nevertheless the

*Author to whom correspondence should be addressed.
[1] Mathematics and Statistics, School of Engineering and Information Technology, Murdoch University, Murdoch, Western Australia, 6150 (E-mail: ross.bowden@iinet.net.au)
[2] Mathematics and Statistics, School of Engineering and Information Technology, Murdoch University, Murdoch, Western Australia, 6150 (E-mail: B.Clarke@murdoch.edu.au)

Acknowledgment. The authors would like to acknowledge the work and advice of Dr Paul Gilbert, the author of the R multiple time series package, `dse`, used in the final model fitting. We also acknowledge the assistance of Mr Éric Dubois, the author of the Scilab econometrics package, `Grocer`, and Professor Ruey Tsay, the author of the R package, `MTS`, both used at an earlier stage of the current research. We thank Professor Adrian Bowman for his insightful comments on a draft of the paper. Finally we acknowledge the contribution from the Technical Editor, Dr Rolf Turner, and two anonymous referees whose insightful comments significantly improved the paper.

Prepared using anzsauth.cls *[Version: 2017/10/03 Version 7]*

corrected temperature record from the current work could be used in bootstrapped temperature simulations.

The daily maximum and minimum temperatures are modelled as a bivariate time series while accommodating and estimating the effect of location shifts and climate change on the daily values. The model used also incorporates the annually replicated nature of the time series process. To accommodate this replication, the current research analyses the daily temperature data by week-in-the-year with fully efficient use of all available historical data, simultaneously for all years.

To date, none of the published temperature adjustment methods (as used for example in climate change studies) appear to have employed time series or multivariate models (see Section 1.2). In particular, the authors could identify no specific published work on the adjustment of the historical temperature record used for estimating future electricity demand.

Electricity systems are required to supply the daily and (resulting) annual maximum demand on the power system (as well as providing for aggregate energy needs). Hence there is a major focus in power supply companies on servicing the annual peak demand from customers. To this end, $479 US billion was invested per year on average by the world's electricity companies from 2011 to 2013 (Wilkinson 2014).

These investment decisions are critically determined by long-term (5 to 20 year) forecasts of electricity maximum demand. Any improvement in the accuracy of demand forecasting can result in substantial savings in capital expenditure. Moreover an appreciation of the probability of certain future demand outcomes allows the electricity planning staff to more exactly match the risk of plant deficit with the variability of customer demand. This results in an improved balance of capital expenditure with system reliability.

The underlying growth in annual electricity maximum demand is primarily driven by economic and social factors. However the annual maximum demand is simply the largest of the daily maximum demands which are further determined by season of the year, day-of-the-week and public holiday effects, daily prevailing as well as lagged weather variables, plus autocorrelated (but otherwise unexplained) random influences.

The primary weather variables which influence daily maximum demand and which are readily available for most locations are the daily maximum and minimum temperatures. As a point of reference, high temperatures in Western Australia in summer have strong effects on peak demand whereas low temperatures in winter have a less pronounced effect; in countries in higher latitudes the influence of winter temperatures is stronger. Either way, stable historical series of daily maximum and minimum temperatures are required for demand simulation and these depend on adjustments that make allowance for location shift and climate change in recordings of past data. Methods for implementing such adjustments are discussed below.

1.1. Electricity demand forecasting and weather recordings

The methods for predicting long-term maximum demand growth typically encompass a wide variety of approaches. These include regression and econometric models, time series analysis (including exponential smoothing), neural networks,

support vector machines and knowledge-based expert systems (see Alfares & Nazeeruddin 2002, Hahn, Meyer-Nieberg & Pickl 2009 and Feinberg & Genethliou 2005).

To some extent all these models use an estimate of the pattern of weather on the day of the annual maximum demand. This is typically the mean of the relevant weather variables on the maximum demand day even though there is usually little actual data on which to base a direct estimate (with only one value per year of strictly relevant historical daily temperatures) . Also the weather variables of interest may include lagged and transformed values of readily available weather readings. For example variables can be constructed representing the effect of runs of days of relatively unchanged (but hot) weather which can see a gradual increase in electricity demand due to the so-called heat bank effect in homes and businesses. Typically some form of running total of weighted daily degree-days over say 35 degrees Celsius is employed here. Hence not only will there typically be a very limited number of days of annual maximum demand on which to base an assessment of the typical peak day weather but there are also subtle relationships with the weather occurring on peak and past days.

One approach to addressing these issues is illustrated in Figure 1. This is similar to the work of Hyndman & Fan (2010) although they use daily half-hourly demand data rather than daily maximum demands. They also employ the full daily profile of temperatures (from two sites) rather than just the daily minimum and maximum temperatures. The current model uses daily minimum and maximum temperatures because the associated historical record extends back to 1943 providing over sixty years of data. Additionally it could be argued that it is not absolutely necessary to use the full daily temperature and demand profile to capture the main determinants of daily and annual maximum demand alone. Nevertheless it is recognised that the analysis of daily profile data provides additional useful information. Moreover the multivariate interleaving method introduced in this paper could be used to correct the daily vector time series of historical temperatures.

The method illustrated in Figure 1 forecasts both the mean and the distribution of annual maximum demands in any future year by using a detailed and sophisticated regression model that relates historical daily maximum electricity demand to weather conditions (as well as other variables discussed earlier such as day-of-the-week). The daily regression model is typically fitted using daily demand and weather data for the past 5 to 10 years. This approach effectively allows the estimation of the (daily) maximum demand on any day given the predictors referred to above. The model is then used with a forecast of growth (discussed previously) and say sixty years of year-by-year historical daily weather data (as well as simulated daily error terms) to repeatedly generate daily electricity demands for a future year. This then results in replications of the forecast annual maximum demand for that same future year. Empirical distributions of those simulated future annual electricity demands (which, of course, take as given the underlying forecast of growth) can be created.

This is particularly useful for modeling the balance of possible electricity demand versus possible available power plant capacity because it creates an empirical distribution of forecast annual maximum demand. Amongst other metrics, this allows

for the estimation of the so-called once-in-ten-year demand forecast (which, given an accurate forecast of growth, has a 50% chance of being exceeded once every ten years).

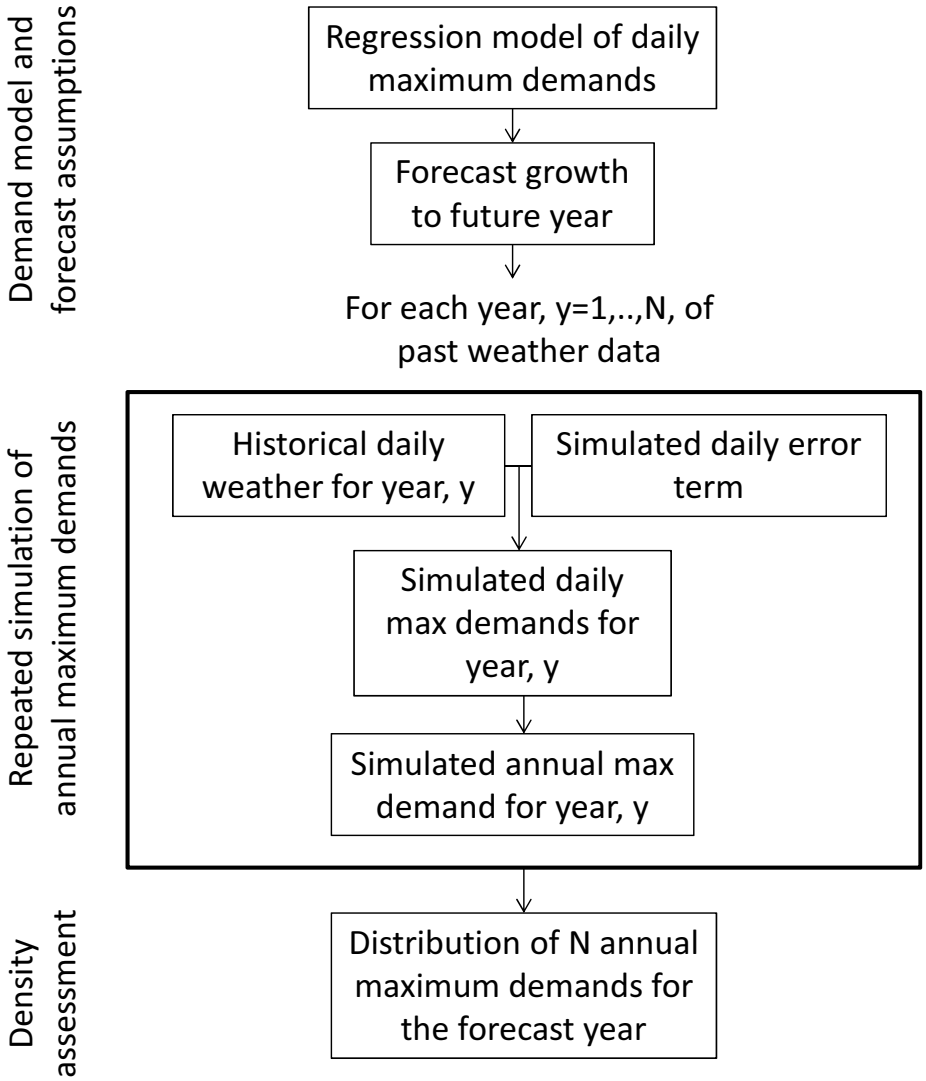


Figure 1. Forecasting the distribution of annual electricity maximum demand using simulation.

A critical assumption here is that the historical weather record is stable, at least reflecting current patterns. However this is unlikely to be true for two reasons. It is now well established that global temperatures are increasing due to climate change. The Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) conclude that "Australia's climate has warmed by 0.9°C since 1910" (CSIRO & BOM 2014) with similar changes for Perth, Western Australia, of between 1.0°C and 1.5°C. This will likely have influenced the historical temperature record used in demand simulation. Also recording stations often change location over time for various reasons such as alterations in the use of the site. It is possible that this has also led to changes in the level of recorded temperatures (see the next section).

A linear trend term was used as it was felt that the relatively subtle temperature change for essentially only one site (Perth) would not support a more intricate trend model. Other possible trend models will be further explored in future work.

Both the location and climate change effects will influence the long-term historical weather record. It is estimated that a one degree Celsius increase in daily maximum temperatures in summer can increase maximum electricity demand in the South-West Interconnected System (SWIS) by up to 2% (Bowden & Gamble 1995) (The annual system maximum demand in the SWIS occurs in summer when the main drivers of demand are airconditioning and refrigeration). This temperature relationship (with appropriate seasonal variation) is employed in electricity demand simulation.

Moreover the time series generating mechanism of the daily temperature data should be consistent from year-to-year within the expected seasonal cycles. Such consistency would be expected from the seasonally recurring nature of the local climate system.

Therefore, whilst accounting for the replicated bivariate autocorrelated nature of the data, the historical readings for Perth should be adjusted for location shift and climate change. This paper estimates the effect of these changes on daily temperature readings for Perth, Western Australia, which is the main centre of electricity demand for the SWIS.

1.2. Adjustment of historical weather records for exogenous effects

The long-term daily weather data available from Europe and the USA are reviewed in Wijngaard, Klein Tank & Können (2003) and Menne, Williams Jr. & Palecki (2010) respectively. They conclude that adjustment of the data is required because of the number of sites where disturbance of the historical record has resulted in substantial data contamination. A similar conclusion for a group of US sites is reached by Pielke Sr. et al. (2007).

To adjust for these issues, the World Meteorological Organization (WMO) has published guidelines for undertaking so-called homogenisation of weather records for (amongst other effects) location shifts (Aguilar et al. 2003). The WMO guidelines focus initially on identifying so-called break-points which are suspected of initiating a change in (at least) the level of the time series process but are of unknown date. Once these dates are identified (which is already the case with the Perth data), the time series is then adjusted for these interventions to match the level of the most recent recording station so that on-going adjustment is not required (at least in the medium term). The adjustment methods (which include averaging, site differencing and regression) typically do not account for the autocorrelated multivariate nature of the data.

The published methods up to 1998 are summarised in Peterson et al. (1998). As with the WMO guidelines the authors firstly discuss detecting a change in homogeneity and then review adjustment methods. The authors consider approaches by country and region. In common with the WMO methods, there is little accommodation for the autocorrelated or multivariate nature of the daily readings. Reviews of methods for identifying change points with respect to homogeneity and for subsequently adjusting the data series are contained in Rhoades & Salinger (1993) and

Guttman (1998). The adjustment methods focus on simple average differences with respect to a relatively stable reference site. The actual site adjustments for a New Zealand dataset from Rhoades & Salinger (1993) range from 0.02 to 1.58°C. Similar methods for estimating the effects of urbanisation and shift of location of recording stations around Beijing were employed by Yan et al. (2010) with site effects varying between 0.43 and 0.95°C.

1.3. Statistical modelling of replicated bivariate temperature time series

The time series modelling of daily temperatures used in this paper employs the interleaving method (see Bowden & Clarke 2012) extended herein to the multivariate case to model replicated realisations of a multiple time series process (specifically the Perth daily maximum and minimum temperature records since 1943). This extension allows fully efficient estimation of locational and climate change effects as well as of the time series coefficients. The time series analysis uses Vector Autoregressive Moving Average (VARMA) models which are a multivariate extension of the univariate ARMA models (see Section 2).

Section 2 in this paper provides a brief overview of VARMA models and Section 3 extends the univariate interleaving method to multivariate processes. Section 4 uses the interleaving methodology to estimate the effect of location shift and climate change by week and season on over sixty years of daily maximum and minimum temperatures for Perth. It also explores the time series structure which is the generating mechanism behind one of the major day-to-day determinants of electricity demand. The weather data used in this paper and the modelling results are contained (as an R dataset) in the Murdoch University Research Repository at `http://researchrepository.murdoch.edu.au/27282/`.

In this paper the terms VARMA and VARMAX (VARMA with exogenous inputs) are used interchangeably. The term VARMA is employed in general to refer to multivariate extensions of autoregressive moving average models. However, when considered necessary to be explicit concerning exogenous inputs, the term VARMAX will be employed.

2. VARMAX models

A Vector Autoregressive Moving Average process of order p and q with exogenous inputs (VARMAX(p,q)), $\{\boldsymbol{x}_t\}_{t=1}^n$, is a k-dimensional multiple time series generated by the model,

$$\boldsymbol{\phi}(B)(\boldsymbol{x}_t - \boldsymbol{\mu}_{\boldsymbol{x}_t}(\boldsymbol{z}_t)) = \boldsymbol{\theta}(B)\boldsymbol{a}_t \tag{1}$$

where $\{\boldsymbol{a}_t\}_{t=1}^n$ is a series of k-dimensional independent identically distributed random error vectors with constant covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{a}}$, $\mathrm{E}(\boldsymbol{a}_t) = \mathbf{0}$ for all t and $\mathrm{E}(\boldsymbol{a}_t\boldsymbol{a}_u^\top) = \mathbf{0}$ for all $t \neq u$. Also $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are matrix polynomials in B, the backshift operator, of degree p and q respectively. The roots of $\det(\boldsymbol{\phi}(B)) = 0$ and $\det(\boldsymbol{\theta}(B)) = 0$ all lie outside the unit circle ensuring stationarity and invertibility respectively.

Typically the zeroth order coefficient matrix in the polynomial, $\boldsymbol{\phi}(B)$, is an identity matrix and similarly for $\boldsymbol{\theta}(B)$. In this case, $\boldsymbol{\Sigma_a}$ is of general symmetric positive-definite form and this specification results in a canonical formulation for the VARMAX model which allows for unique identification. It is often assumed that the innovations are multivariate normal (This is not required for Theorem 1 in Section 3 although it is needed for our model fitting).

We will assume that $\boldsymbol{\mu_{x_t}}(\boldsymbol{z}_t)\ (= \mathrm{E}(\boldsymbol{x}_t|\boldsymbol{z}_t)) = \boldsymbol{\psi}\boldsymbol{z}_t$ where $\{\boldsymbol{z}_t\}_{t=1}^{n}$ is a series of explanatory (input) vectors and $\boldsymbol{\psi}$ is the matrix of regression parameters.

The foregoing is one form of standard VARMAX specification. The VARMAX process also could be described as being generated by a seemingly-unrelated regression model (see Johnston & Dinardo 1996) with VARMA errors in that the series mean vector corrects the series mean level to zero before application of the VARMA filter. However if the exogenous variables are introduced on the right hand-side of (1) their influence on the time series vector can only be assessed with knowledge of the VAR filter. This alternative expression for the VARMAX model is,

$$\boldsymbol{\phi}(B)\boldsymbol{x}_t = \boldsymbol{\Upsilon}\boldsymbol{z}_t + \boldsymbol{\theta}(B)\boldsymbol{a}_t, \tag{2}$$

where $\boldsymbol{\Upsilon} = \boldsymbol{\phi}(1)\boldsymbol{\psi}$, that is,

$$\boldsymbol{\psi} = \boldsymbol{\phi}(1)^{-1}\boldsymbol{\Upsilon}. \tag{3}$$

This specification (2) is used by the software employed in this paper (the R package, `dse` (Gilbert 2006)) to fit the replicated VARMAX model to daily temperatures.

An alternative specification of $\boldsymbol{\phi}(B)$, $\boldsymbol{\theta}(B)$ and $\boldsymbol{\Sigma_a}$ is possible which allows unique identification (See Lütkepohl 2005, pp. 447 ff.). This uses the unique Cholesky LDL decomposition of the innovations covariance matrix, that is, $\boldsymbol{\Sigma_a} = \mathbf{L}\mathbf{D}\mathbf{L}^{\top}$ where $\mathbf{L}$ is upper triangular with a unit diagonal (so-called "unitriangular") and $\mathbf{D}$ is a diagonal matrix.

This alternative specification to (1) is,

$$\begin{aligned}(\mathbf{L}^{-1} + \mathbf{L}^{-1}\boldsymbol{\phi_1}B + \mathbf{L}^{-1}\boldsymbol{\phi}_2B^2 + ... + \mathbf{L}^{-1}\boldsymbol{\phi}_pB^p)(\boldsymbol{x}_t - \boldsymbol{\mu_{x_t}}(\boldsymbol{z}_t)) \\ = (\mathbf{I} + \mathbf{L}^{-1}\boldsymbol{\theta}_1\mathbf{L}B + \mathbf{L}^{-1}\boldsymbol{\theta}_2\mathbf{L}B^2 + ... + \mathbf{L}^{-1}\boldsymbol{\theta}_q\mathbf{L}B^q)\boldsymbol{u}_t.\end{aligned} \tag{4}$$

where $\boldsymbol{u}_t = \mathbf{L}^{-1}\boldsymbol{a}_t$ and hence $\mathrm{var}(\boldsymbol{u}_t) = \mathbf{D}$.

This now provides a representation with a diagonal innovations covariance matrix but where the zeroth order MA matrix is an identity matrix and the zeroth order AR matrix is upper unitriangular because $\mathbf{L}^{-1}$ is upper unitriangular. This AR formulation explicitly makes the first entry of $\boldsymbol{x}_t$ (that is, $x_{1,t}$) a linear function of elements $(x_{2,t}, ..., x_{k,t})$ as well as other elements of $\boldsymbol{x}_t$ at non-zero lags. Similarly $x_{2,t}$ is a linear function of $(x_{3,t}, ..., x_{k,t})$ as well as other elements of $\boldsymbol{x}_t$ at non-zero lags, and so on.

In the context of the bivariate application in Section 4, this formulation, which is used in this paper, appears more natural inasmuch as the $x_{1,t}$ and $x_{2,t}$ are recorded sequentially on the same day and is similar to that of a periodically correlated ARMA model (see Parzen & Pagano 1979).

3. The interleaving method

In this section, we prove that replicated independent VARMAX processes can be represented as a single VARMAX process with the same dimension as each of the replicated series. This result allows model fitting using existing VARMAX software.

3.1. Replicated VARMAX process

Let the i^{th} replicated k-dimensional vector series over the time span, $t = 1,...,n$, be $\{\boldsymbol{x}_{i,t}\}_{t=1}^{n}, i = 1, ..., m$ and assume each series is generated by the following VARMAX(p,q) model,

$$\boldsymbol{\phi}(B)(\boldsymbol{x}_{i,t} - \boldsymbol{\mu}_{\boldsymbol{x}_{i,t}}(\boldsymbol{z}_{i,t})) = \boldsymbol{\theta}(B)\boldsymbol{a}_{i,t} \tag{5}$$

where $\{\boldsymbol{a}_{i,t}\}_{i=1}^{n}$ is a series of independent zero-mean identically-distributed random error vectors with $\mathrm{E}(\boldsymbol{a}_{i,t}\boldsymbol{a}_{j,u}^{\top}) = \boldsymbol{\Sigma}_{\boldsymbol{a}}$ for $i = j$ and $t = u$ and $\mathbf{0}$ otherwise. Hence the error vectors have a covariance matrix of general form but otherwise the vectors are assumed to be independent between realisations at all lags and within realisations at all non-zero lags. Also $\boldsymbol{\phi}(B)$ and $\boldsymbol{\theta}(B)$ are matrix polynomials in B, the backshift operator, of degree p and q respectively.

The (conditional) mean of $\boldsymbol{x}_{i,t}$ is,

$$\mathrm{E}(\boldsymbol{x}_{i,t}|\boldsymbol{z}_{i,t}) = \boldsymbol{\mu}_{\boldsymbol{x}_{i,t}}(\boldsymbol{z}_{i,t}) = \boldsymbol{\psi}\boldsymbol{z}_{i,t} \ ,$$

where $\{\boldsymbol{z}_{i,t}\}_{t=1}^{n}$, i=1,...,m, are m series of explanatory vectors. It is possible to effectively have a unique parameterisation of $\boldsymbol{\psi}$ (say $\boldsymbol{\psi}_i$) for each realisation, i, through simply expanding the dimension of the input vector, $\boldsymbol{z}_{i,t}$, by a factor of m.

We will call the time series (5) an RVARMA (replicated VARMA) process, that is, RVARMA(p,q,m). It has a mean which can vary with each series realisation but it otherwise maintains a consistent generating mechanism between realisations. In fact the mean can be any linear combination of the exogenous vector variables, $\boldsymbol{z}_{i,t}$ (and, in general, can be a non-linear function of the $\boldsymbol{z}_{i,t}$).

3.2. Equivalent replicated VARMAX representation

We now state a theorem that reduces the apparent dimensionality of the replicated process by a factor of m. The proof is provided in the Appendix.

Theorem 1. *Let the replicated k-dimensional series* $\{\boldsymbol{x}_{i,t}\}_{t=1}^{n}, i = 1, ..., m,$ *be generated by the above RVARMA(*p,q,m*) process (see (5)), and let*

$$\begin{aligned} \boldsymbol{y}_{m(t-1)+i} &= \boldsymbol{x}_{i,t}, \\ \boldsymbol{w}_{m(t-1)+i} &= \boldsymbol{z}_{i,t} \text{ and} \\ \boldsymbol{\epsilon}_{m(t-1)+i} &= \boldsymbol{a}_{i,t} \, . \end{aligned} \tag{6}$$

Then

$$\boldsymbol{\phi}(B^m)(\boldsymbol{y}_s - \boldsymbol{\mu}_{\boldsymbol{y}_s}(\boldsymbol{w}_s)) = \boldsymbol{\theta}(B^m)\boldsymbol{\epsilon}_s \tag{7}$$

where $\mathrm{E}(\boldsymbol{\epsilon}_s) = \mathbf{0}$, $\mathrm{var}(\boldsymbol{\epsilon}_s) = \boldsymbol{\Sigma}_{\boldsymbol{a}}$ *and* $\mathrm{E}(\boldsymbol{\epsilon}_s \boldsymbol{\epsilon}_r^{\top}) = \mathbf{0}$, $s \neq r$. *That is, the interleaved series,* $\{\boldsymbol{y}_s\}_{s=1}^{mn}$, *is a k-dimensional VARMA process of order* (mp, mq).

To paraphrase Theorem 1 (henceforth called the Multivariate Interleaving Theorem), any m replicated independent k-dimensional VARMA(p, q) time series, each of length n, can be represented by one k-dimensional VARMA(mp, mq) process of length mn. This equivalence is achieved by interleaving the m series and by ensuring that AR and MA parameters are only non-zero at orders that are multiples of m. The equivalence uses an interleaving which is illustrated in Figure 2 for two artificial bivariate series, each of length seven.

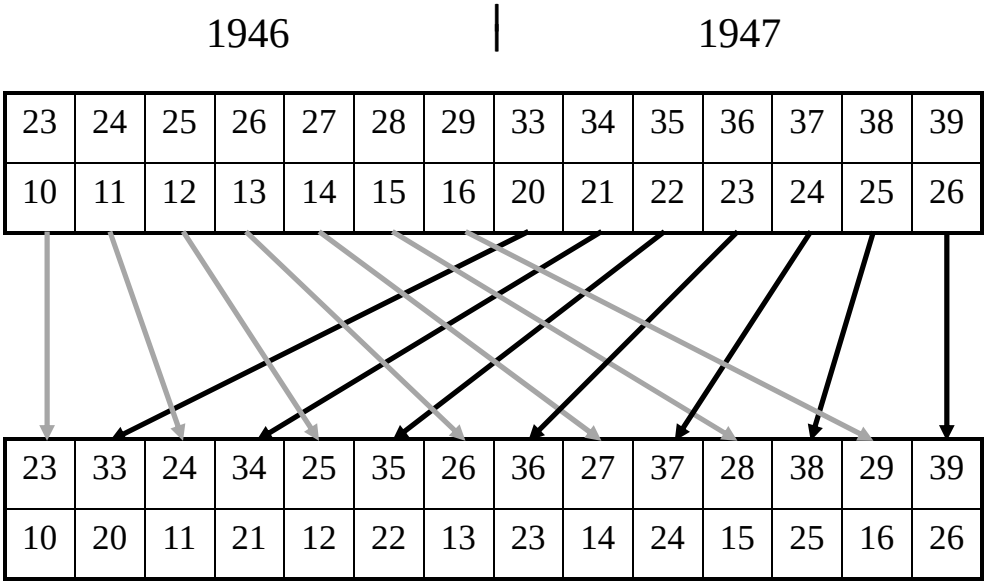


Figure 2. Multivariate interleaving of an artificial bivariate series.

Interleaving allows replicated independent stationary vector time series to be modelled as one stationary series without "end of series" discontinuities, as would be the case if say the series were merely concatenated together. Bowden & Clarke (2012) contains a discussion of these discontinuity effects for a univariate series.

By using VARMA software such as R's `dse` (Gilbert 2006) and `MTS` (Tsay 2015) packages, Scilab's `Grocer` (Dubois & Michaux 2016) and Gauss's Time Series `MT` (GAUSS 2012) (which all allow subsets of the VAR and VMA matrix parameters to be set to zero) the interleaving method can be employed in RVARMA model fitting without preparing purpose-built computer programs. These packages use maximum likelihood estimation. Of course, the interleaving method can also be applied to estimation approaches other than those employed in this paper including robust methods and least squares.

4. Effects of location and climate change on daily maximum and minimum temperatures for Perth, Australia

This section deals with estimating the effect of location shift and climate change on the daily maximum and minimum temperature readings from 1943 to 2009 for Perth, Western Australia. The results can be used to adjust the historical record employed in electricity demand simulation. As mentioned previously, the `dse` package from R is used to fit the associated RVARMA model via (conditional) maximum likelihood. The modelling also provides an understanding of the relationship between maximum and minimum temperatures which is informative in forecasting daily electricity demand up to a week ahead.

Figure 3 displays a plot of the daily maximum and minimum values for three years and it is clear that there is a strong relationship between values on the same day. The expected seasonal cycle is also evident as is an increase in the variability of the maximum temperatures over summer.

Given the increased variability in summer and the known changes in weather patterns between summer and winter it is likely that VARMA models of the bivariate daily temperature data vary over the year. However, from year to year, the models are likely to be unchanged for any particular part of the year. Hence the modelling in Sections 4.1 to 4.3 was undertaken separately for each week-in-the-year of daily data but simultaneously for all years. This provides estimates of the effect of location shift (using binary intervention variables which are similar to dummy variables in regression analysis (see Chatfield 2003, p. 259)) and climate change (via a trend term) by week-in-the-year and ultimately by month and season. The five locations where the temperature data were collected are listed in Table 1.

The use of multivariate interleaving allows one stationary vector series to represent a week of data for all sixty-seven years without any "end of week" effects. Moreover "end of year" effects are managed by analysing the sixty-seven years of data for week 52 as an interleaved series separate from that for week 1.

4.1. Model identification with interleaving

To begin the RVARMA model fitting, an interleaved bivariate series by each week-in-the-year was created using the daily maximum and minimum temperatures from 1943 to 2009. The RVARMA model order was determined by examining the raw sample cross-correlations (see Chatfield 2003), the prewhitened sample cross-correlation function (see Jenkins & Alavi 1981 and Granger & Newbold 1986, pp. 237

ff.) and the sample partial lag autocorrelation function (Wei 2006, pp. 408 ff.) after correction for all intervention effects (henceforth termed detrending) and adjusting for interleaving. For the first week of interleaved data ($n = 67 \times 7$) these correlations are plotted in Figures 4, 5 and 6 respectively with the 95 percent confidence intervals delimited by dashed lines. The full (and hollow - see below) points indicate the interleaved lags $(-3 \times 67, ..., 3 \times 67)$ that correspond to lags -3 to 3 in the original series.

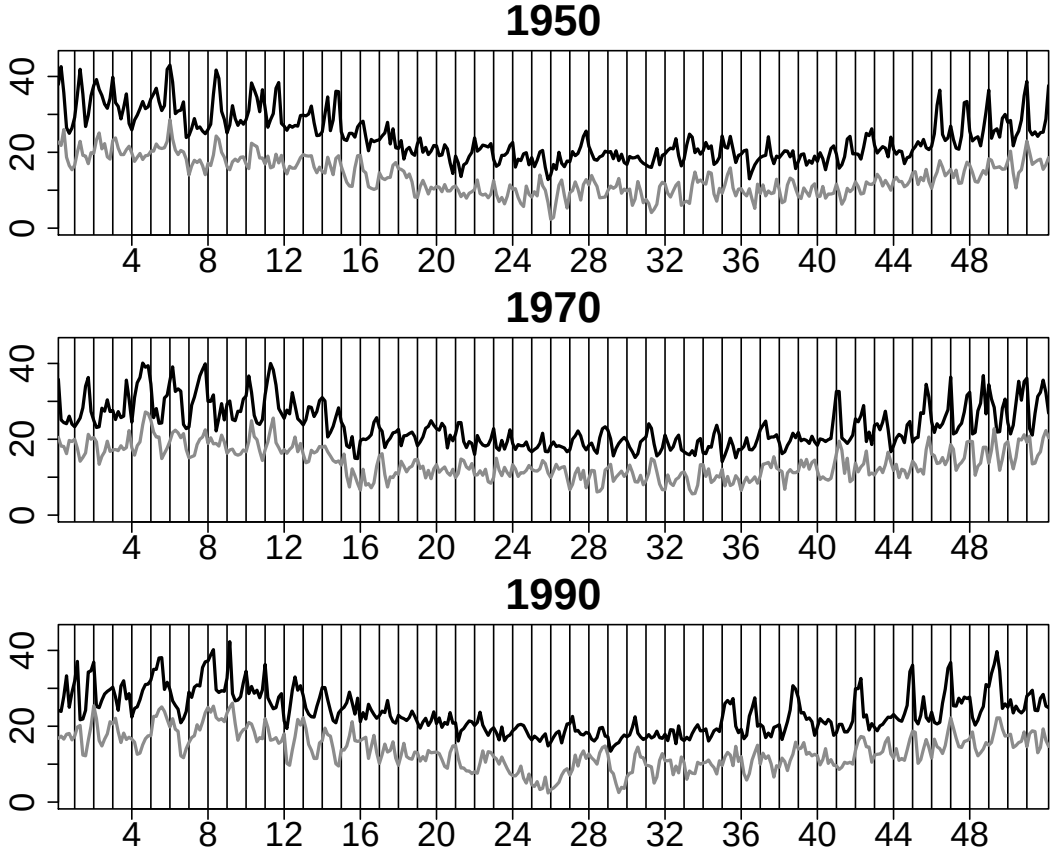


Figure 3. Three years of daily maximum and minimum temperatures (in degrees Celsius). The x-axis values are the week-in-the-year. The plot reveals the seasonality in mean and variance.

Detrending involved identifying and fitting a univariate interleaved AR(2) model (Bowden & Clarke 2012) separately to both daily maximum and minimum temperatures with associated intervention terms for the change in location and for the long-term trend in temperature. The resulting intervention terms (without the autoregressive filters) were then used to adjust the two univariate time series for the change in location and for the long-term increase in temperatures. The resultant zero-mean series were then prewhitened.

Prewhitening consisted of applying the AR(2) filter from the AR detrending model for daily minimum temperatures to both the detrended daily maximum and minimum temperatures. The sample cross-correlations were then calculated for the two filtered series. This prewhitening took account of the constraints imposed by

interleaving, that is, only parameters for AR orders corresponding to multiples of the number of replicated series (years) were non-zero.

As mentioned above, the sample partial lag correlation matrix was also employed in model selection (Wei 2006). The sample partial lag correlation matrix is the sample cross correlation matrix at each lag of k time intervals after removing the (linear) influence of the intervening lags. For an AR(p) process the correlations cut off at lag p as is the case with the multivariate partial autoregression matrix.

To accommodate the constraints imposed by interleaving, the partial lag correlations were derived by only fitting autoregressions to multiples of the number of replicated series (that is, of years). In the approach of Wei (2006) for estimating the partial lag correlations, this implies fixing the (detrended) sample cross-correlations at zero for the other lags before using them in the estimation routine.

4.2. Identification results

In Figure 4, the sample detrended cross-correlations between maximum and minimum temperatures for week one (as indicated by the black dots at lags that are multiples of 67) include a number of statistically significant values. However the sample prewhitened detrended cross-correlations (in Figure 5) have significant values at lags 0 and -1 only (that is, lags 0 and -67 with interleaving). This demonstrates the ability of prewhitening to substantially simplify the model selection process within an interleaving paradigm.

Sample partial lag correlations between maximum and (lagged) minimum temperatures are plotted in Figure 6. As with the cross-correlations the values corresponding to lags -3 to 3 in the original replicated series are marked by full dots. The values reveal significant correlations at lags ± 1 and arguably at ± 2. The hollow dots indicate partial lag correlations calculated without setting relevant intermediate correlations to zero.

Given that these correlation results were similar for all weeks, we decided to fit an RVAR model of order two (that is, an RVAR(2,0,67) model) for each week. This was undertaken using the interleaving method from Section 3 (with m=67, that is, a total sample per week-in-the-year of $67 \times 7 = 469$) and employing the method of conditional maximum likelihood via the R package, `dse`. The residuals from the model fits showed no consistent autocorrelation or cross-correlation structure.

The `dse` package uses the VARMAX representation (2) but the results in this paper employ the representation (4), derived by applying the Cholesky LDL transformation from Section 2 and the transformation of the process mean from (3). Hence the VAR(2) models in this paper utilise an intervention vector that is the (conditional) mean of the process, a zeroth order VAR matrix that is upper unitriangular and a covariance matrix of the innovations vector that is diagonal.

This formulation permits what is arguably a simpler interpretation of the estimates whereby the mean correction due to the intervention terms is applied directly to the vector time series before application of the AR filter, the daily maximum temperature is related to the (earlier) minimum temperature on the same day and the elements of the innovation vector are independent.

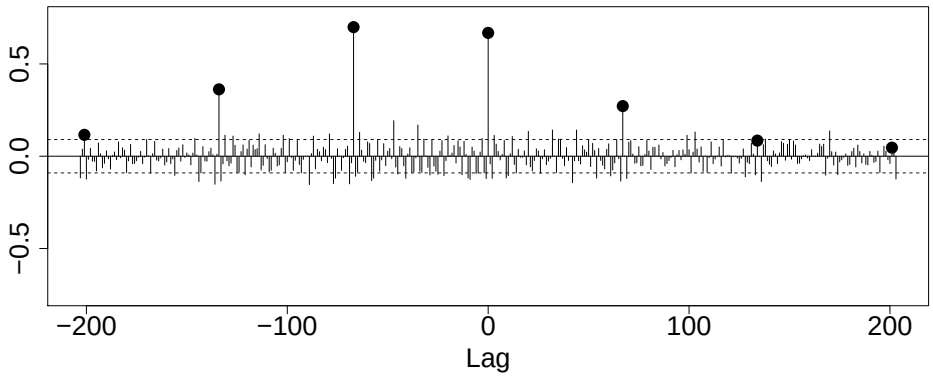


Figure 4. Detrended cross-correlations using interleaving (week 1) with 95% confidence intervals. The black dots indicate the sample correlations at lags -3 to 3 in the original series. The lack of prewhitening makes order selection difficult.

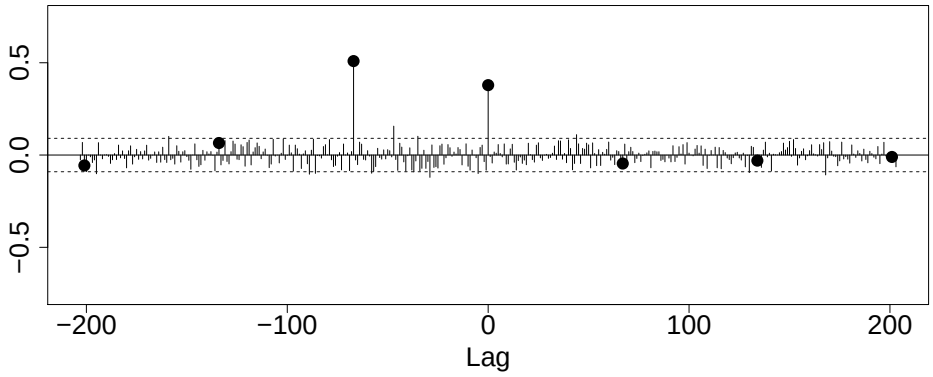


Figure 5. Cross-correlations of the series which were detrended and prewhitened using interleaving (week 1) with 95% confidence intervals. The black dots indicate the sample correlations at lags -3 to 3 in the original series. Prewhitening is effective in simplifying the cross-correlations.

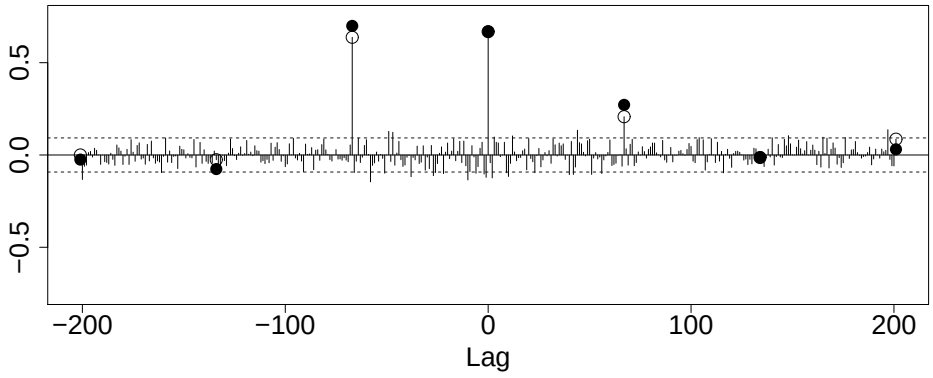


Figure 6. Wei's partial lag detrended correlations using interleaving (week 1) with 95% confidence intervals. The black dots indicate the sample correlations at lags -3 to 3 in the original series. The circles are the same values uncorrected for assumed zero correlations at lags that are not multiples of the number of replicated series (67 in this case). The significant correlations (across all weeks-in-the-year) suggest an AR(2) process.

The standard errors of the sample parameter estimates for model (4) were not immediately available from the model fits and had to be derived after transformation

from the fitted model (2) using simulation (10,000 simulations). In order to carry out this procedure, we assumed the estimated innovations covariance matrix for (2) is independent of the other sample parameter estimates for (2). Repeated realisations of the sample covariance matrix of the innovations from the fitted model (2) were simulated using bootstrapping on the model's residual vectors. The other parameter estimates from (2) were simulated using an assumption of multivariate normality where the mean vector consisted of the VAR and intervention parameter estimates (in (2)) and the covariance matrix was the associated Hessian-derived covariance matrix.

Each set of combined simulated parameter estimates for (2) was transformed using the LDL transformation from Section 2 used in model (4) and the mean transformation (3). The empirical distribution of the resulting simulated parameters was then used to calculate the sample covariance matrix of the parameter estimates for (4). The square roots of the diagonal elements of this matrix were used as the standard errors of the transformed parameter estimates shown in Figures 7, 8, 9 and 10.

4.3. Estimation of the location shift and climate change effects and of the VARMA generating mechanism

The VAR parameters are plotted by week in Figure 7 where, for example, the top left panel displays a plot against week-of-the-year of the parameter for the relationship between the maximum versus the minimum temperatures at lag zero (that is, on the same day). Similarly the bottom right panel displays a plot against week-of-the-year of the autoregressive parameter for minimum temperatures at a lag of two days. Note that all these plotted values are the negatives of the VAR parameters from representation (4) which helps to reflect the use of the values in a predictive formulation.

It is clear that the parameters change substantially and relatively smoothly over the year with the strongest relationship between maximum temperatures and past maximum and minimum temperatures occurring in summer and with there being little relationship in winter. The minimum temperatures show a much weaker set of relationships although the VAR parameters are now generally strongest in winter.

The variances of the (independent) innovations by week are shown in Figure 8. The panels of this figure indicate that the variation of the maximum temperatures changes substantially over the year with the greatest variance in summer. The minimum temperatures display a relatively unchanging variance.

The RVARMA modelling by week produced estimates of the effect of location shift and climate change. However the week-by-week results were noisier than the VAR parameters already discussed which made application of these results difficult. Hence the results were averaged by month (that is, by 28 days) with associated calculation of standard errors (see below) and these were used in the effect plots that are displayed in this paper.

Given the autocorrelated nature of the data, it is likely that the weekly intervention estimates are autocorrelated. Hence the standard errors of the monthly mean effect estimates were derived using the well-known result for the variance of the mean of correlated random variables (See Cressie 1993, pp. 15 ff.). In calculating the

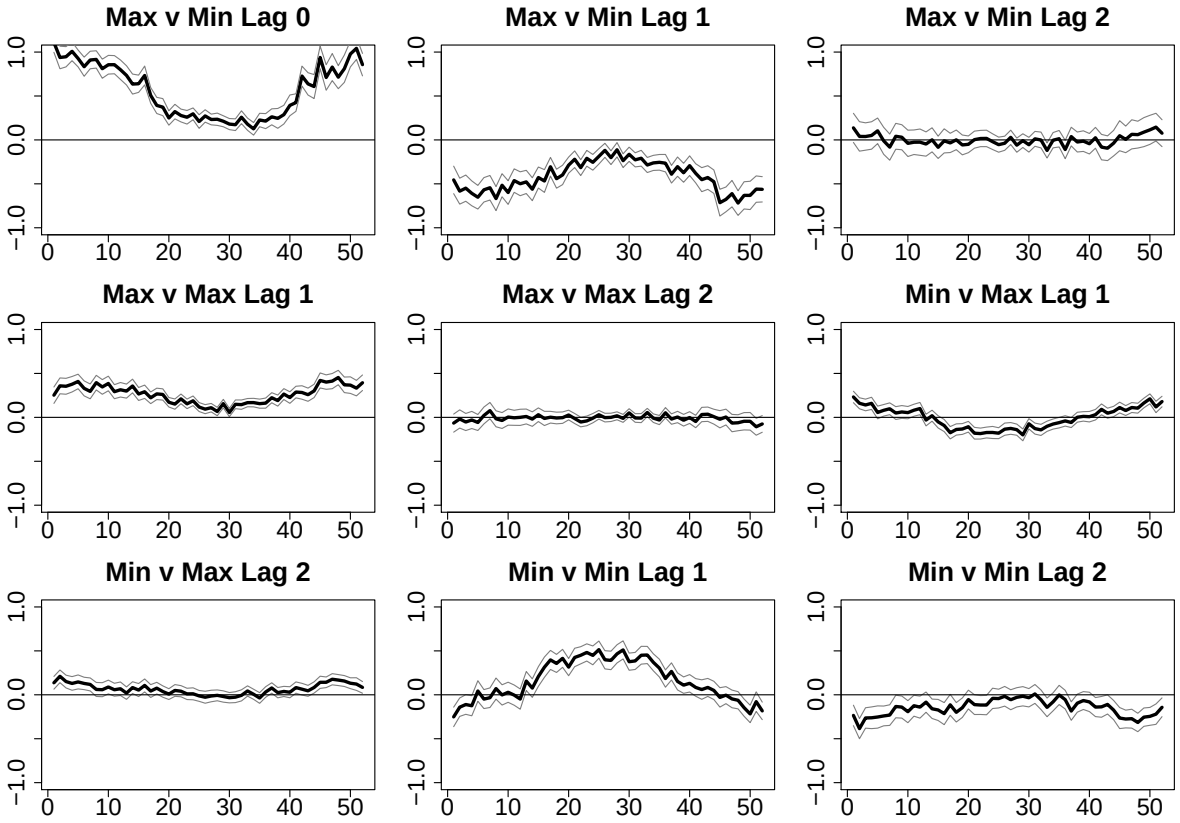


Figure 7. VAR(2) parameters by week with 95% confidence intervals. There is evidence of seasonality in the AR parameters for the maximum and, to a lesser extent, minimum temperature models.

standard errors we assumed that only the (auto-)correlation between adjacent week's intervention terms is non-zero. This correlation was estimated using a simulation procedure which employed the estimated VARMAX models to repeatedly generate new bivariate input series. The models used in this paper were then fitted to the simulated data and the sample week-to-week correlation of the resulting parameter estimates was derived from the simulated estimates. These correlations increased the standard errors of the estimated effects by approximately 10 percent compared to an independence model.

The top left plot in Figure 9 shows the mean daily maximum temperatures by month for the site used to 1963 and the values exhibit the expected seasonal cycle. The next four plots show the mean difference by month between the data recorded at each subsequent site and the initial site. The sixth plot shows the annual trend in maximum temperatures (by month) to 2009.

Over and above climate change, the results suggest colder temperatures in summer for the Wellington St site (1967-92) and the current Mt Lawley site (1993-09). This implies that maximum temperatures recorded at Mt Lawley in summer are likely to underestimate the true maximum temperature compared to the historical record from King's Park (pre-1963).

To accommodate the apparent variation in effects by season, we calculated the mean maximum temperature effects by season (summer = months 1 to 3, autumn

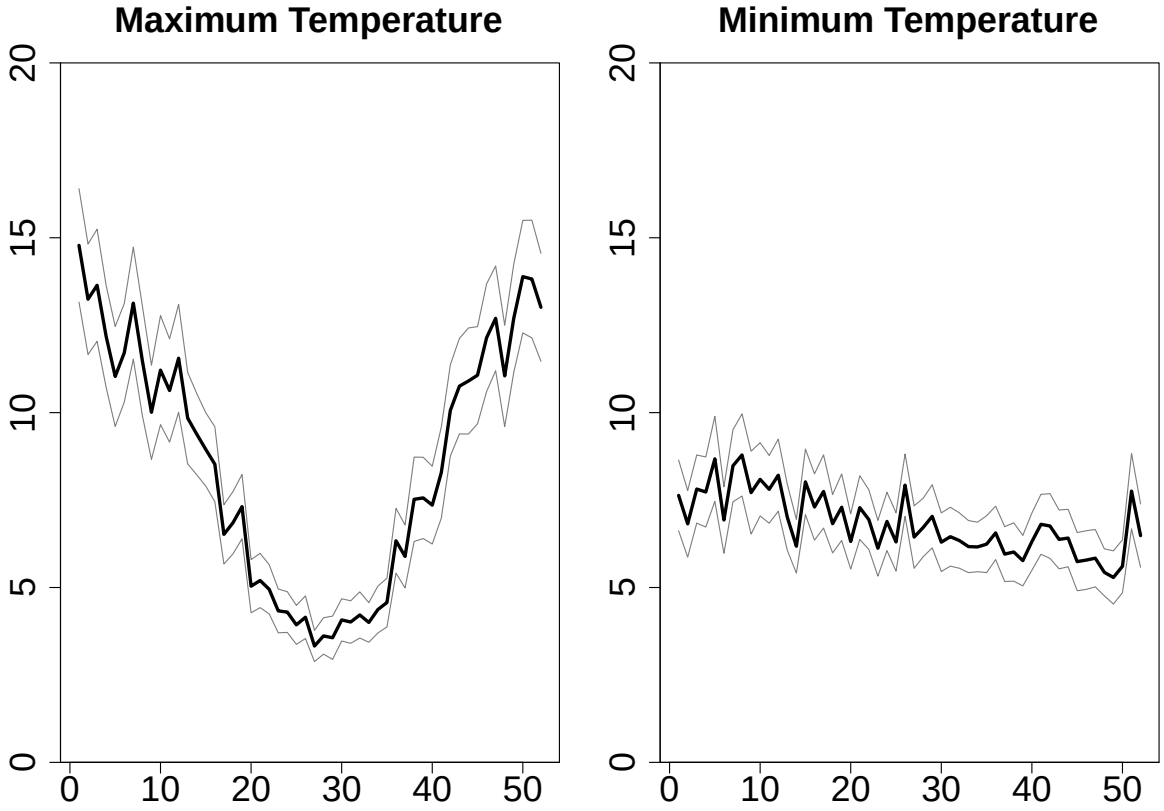


Figure 8. Innovation variance by week with 95% confidence intervals. Again there is evidence of seasonality, at least for maximum temperature innovations.

= 4 to 5, winter = 6 to 8, spring = 9 to 13). These results were chosen so as to group together months showing similar VARMA parameter estimates and mean temperatures. Standard errors and t statistics were also calculated. The results are displayed in Table 2. Significant effects are preceded by "*". The seasonal standard errors were calculated by means of the same approach as for the monthly standard errors.

The estimated annual (positive) trend in maximum temperatures from combining the fifty-two weeks' results is 0.0147 °C (±0.0108 being 95% confidence limits). This equates to 0.98 °C (±0.72) as the total increase over 67 years. There appears to be a higher rate of increase for summer compared to other seasons.

For minimum temperatures (see Figure 10) the monthly results for locality are similar to those for maximum temperatures although there is additional evidence of lower minimum temperatures over the whole year at the current Mt Lawley site. To again accommodate the apparent variation in effects by season, we display in Table 2 the mean effects by season. Over the fifty-two weeks the mean difference for the Mt Lawley site compared to the King's Park site is −1.84 °C (±0.40). The (positive) annual trend in minimum temperatures is 0.0179 °C (±0.0078) or 1.20 °C (±0.52) as a total over 67 years. Again, as with the maximum temperatures, there appears to be a higher rate of increase for summer compared to other seasons.

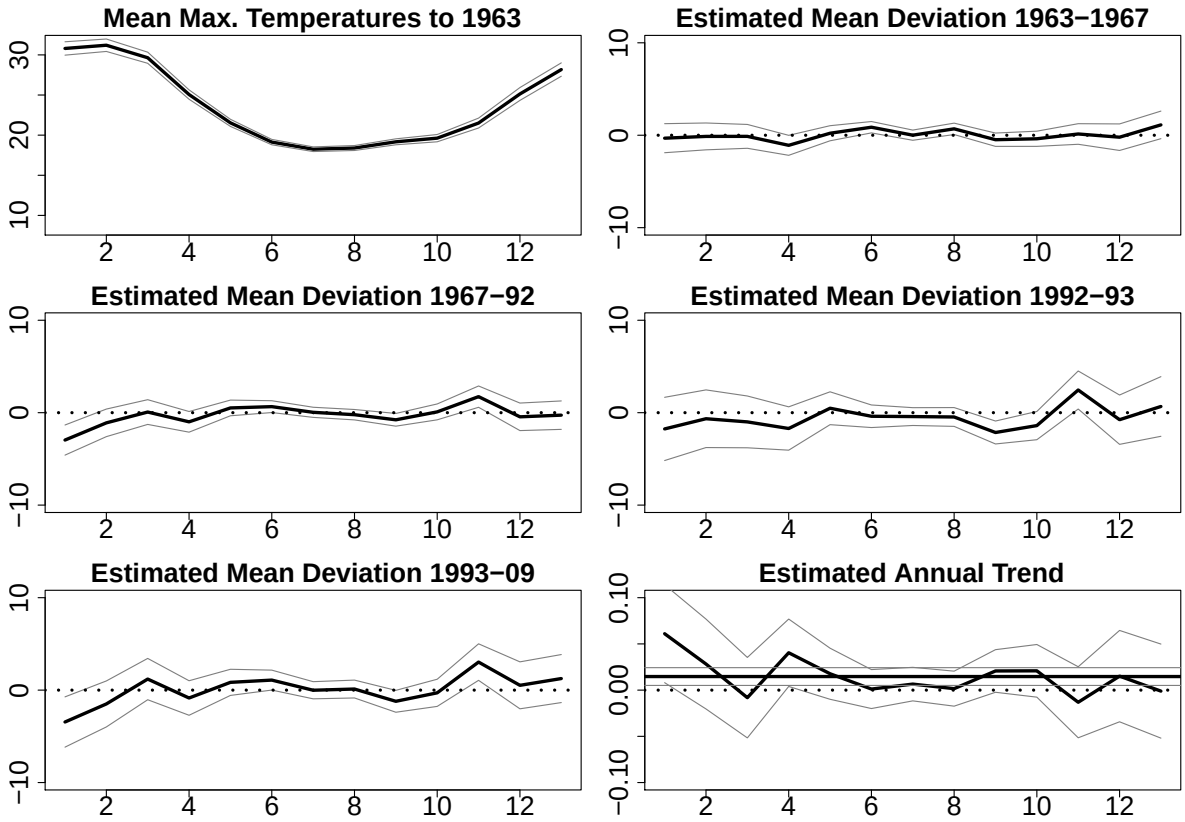


Figure 9. Estimates by month for maximum temperatures of the effect of change of location and of trend over time with 95% confidence intervals. The plot of annual trend also shows the overall mean annual change as a horizontal full line. The results suggest colder temperatures in summer for the Wellington St site (1967-92) and the current Mt Lawley site (1993-09). Also there is a total increase over 67 years of 0.98 °C (±0.72).

5. Conclusions

In this paper the effects of location shift and climate change on historical temperature recordings for Perth, Western Australia, (as employed in electricity demand simulation) are estimated. The analysis uses multivariate time series models and incorporates a multivariate extension of the univariate interleaving method (Bowden & Clarke 2012). The interleaving approach allows replicated realisations of the same VARMA process to be modelled as a single VARMA series of the same dimension as each of the original series but with extended length. The analysis shows that the current weather recording site for Perth has colder temperatures than the original location and there was also evidence of rising temperatures due to climate change.

Accordingly, in order to prepare a stable time series of daily readings for electricity demand simulation, it is recommended that the daily historical values for maximum and minimum temperatures be adjusted for location effects using the mean seasonal effects described in detail above. The effects of climate change should be incorporated by using the associated annual growth factors indexed by season. Allowing for the effects of climate and site change alters the temperature readings for certain locations

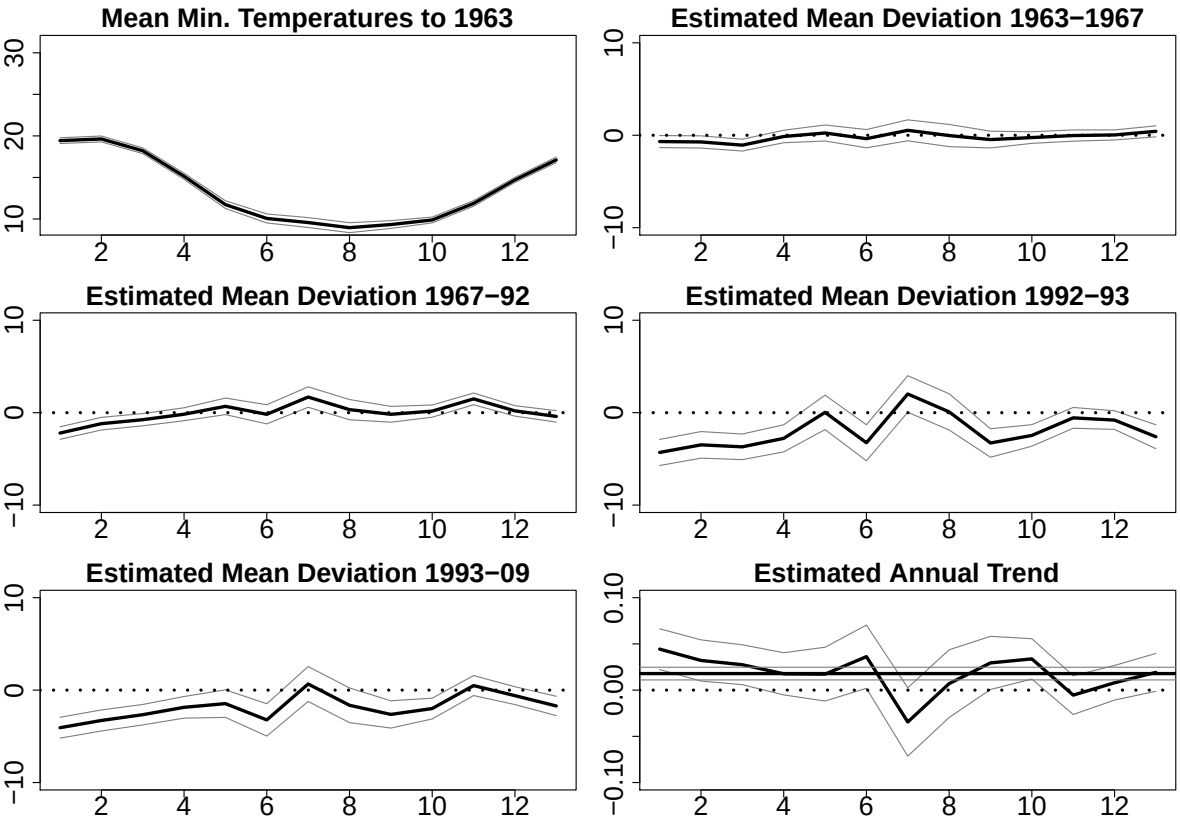


Figure 10. Estimates by month for minimum temperatures of the effect of change of location and of trend over time with 95% confidence intervals. The plot of annual trend also shows the overall mean annual change as a horizontal full line. The total increase over 67 years is 1.20 °C (±0.52). There is evidence of lower temperatures over the whole year at the current Mt Lawley site (The mean difference compared to the King's Park site is −1.84 °C (±0.40)).

by over two degrees Celsius. Hence the associated simulated electricity demand forecasts could change by up to 4% (see Section 1).

Appendix

Proof of the Multivariate Interleaving Theorem. Given $\{\boldsymbol{y}_s\}$, $\{\boldsymbol{w}_s\}$ and $\{\boldsymbol{\epsilon}_s\}$ from (6), a constant i from $(1, ..., m)$, and $s = m(t-1)+i$, $t = 1, ..., n$, the difference equation (7) can then be expressed as

$$\boldsymbol{\phi}(B)(\boldsymbol{x}_{i,t} - \boldsymbol{\mu}_{x_{i,t}}(\boldsymbol{z}_{i,t})) = \boldsymbol{\theta}(B)\boldsymbol{a}_{i,t}\ .$$

where $t = 1, ..., n$. It is known that for $s = i,\ m+i,\ 2m+i, ..., (n-1)m+i$, $E(\boldsymbol{\epsilon}_s) = E(\boldsymbol{a}_{i,t}) = \mathbf{0}$, and $V(\boldsymbol{\epsilon}_s) = V(\boldsymbol{a}_{i,t}) = \boldsymbol{\Sigma}_{\boldsymbol{a}}$. Also, for $r = i,\ m+i,\ 2m+i, ..., (n-1)m+i$, where $r \neq s$, it is the case that $E(\boldsymbol{\epsilon}_s\boldsymbol{\epsilon}_r^\top) = E(\boldsymbol{a}_{i,t}\boldsymbol{a}_{i,u}^\top) = \mathbf{0},\ t \neq u$. We finally note that for $s = i,\ m+i,\ 2m+i, ..., (n-1)m+i$ and $r^* = j,\ m+j,\ 2m+j, ..., (n-1)m+j$, and $s \mid m \neq r^* \mid m$, it is the case that $E(\boldsymbol{\epsilon}_s\boldsymbol{\epsilon}_{r^*}^\top) = E(\boldsymbol{a}_{i,t}\boldsymbol{a}_{j,u}^\top) = \mathbf{0},\ i \neq j$.

That is, the replicated multivariate time series are independent. Consequently the interleaved process (7) is a RVARMA process.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C. & Wieringa, J. (2003). Guidelines on climate metadata and homogenization. *Technical report*, World Meteorological Organization.

Alfares, H.K. & Nazeeruddin, M. (2002). Electricity load forecasting: Literature survey and classification of methods. *International Journal of Systems Science* **33**, 23–34.

Bowden, R.S. & Clarke, B.R. (2012). A single series representation of multiple independent ARMA processes. *Journal of Time Series Analysis* **33**, 304–311.

Bowden, R.S. & Gamble, D.K. (1995). Western Power's peak loads (CS 04/05). *Technical report*, Western Power Corporation.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*. Chapman and Hall, 6th edn.

Cressie, N.A.C. (1993). *Statistics for Spatial Data*. John Wiley and Sons, New York.

CSIRO & BOM (2014). State of the climate 2014. *Technical report*, Commonwealth Scientific, Industry and Research Organisation and the Australian Bureau of Meteorology.

Dubois, E. & Michaux, E. (2016). *Grocer 1.72: an econometric toolbox for Scilab*. URL `http://dubois.ensae.net/grocer.html`.

Feinberg, E.A. & Genethliou, D. (2005). Load forecasting. In *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence*. Springer Science+Business Media Inc.

GAUSS (2012). Time Series MT Version 1.0. URL `http://www.aptech.com/wp-content/uploads/2012/11/tsmt1.0.pdf`.

Gilbert, P. (2006). *Brief user's guide: dynamic systems estimation distributed with the dse package*. Bank of Canada, Bank of Canada. URL `www.cran.r-project.org/web/packages/dse/vignettes/Guide.pdf`.

Granger, C.W.J. & Newbold, P. (1986). *Forecasting Economic Time Series*. Academic Press.

Guttman, N.B. (1998). Homogeneity, data adjustments and climatic normals. *Technical report*, National Climatic Data Center, Asheville, USA.

Hahn, H., Meyer-Nieberg, S. & Pickl, S.W. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research* **199**, 902–907.

Hyndman, R.J. & Fan, S. (2010). Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems* **25**, 1142–1153.

Jenkins, G.M. & Alavi, A.S. (1981). Some aspects of modelling and forecasting multivariate time series. *Journal of Time Series Analysis* **2**, 1–47.

Johnston, J. & Dinardo, J. (1996). *Econometric Methods*. McGraw-Hill/Irwin.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

Menne, M.J., Williams Jr., C.N. & Palecki, M.A. (2010). On the reliability of the U.S. surface temperature record. *Journal of Geophysical Research: Atmospheres* **115**, 1–9.

Parzen, E. & Pagano, M. (1979). An approach to modelling seasonally stationary time series. *Journal of Econometrics* **9**, 137–153.

Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Farland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P. & Parker, D. (1998). Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology* **18**, 1493–1517.

Pielke Sr., R., Nielsen-Gammon, J., Davey, C., Angel, J., Bliss, O., Doesken, N., Cai, M., Fall, S., Niyogi, D., Gallo, K., Hale, R., Hubbard, K.G., Lin, X., Li, H. & Raman, S. (2007). Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. *Bulletin of the American Meteorological Society* **88**, 913–928.

Rhoades, D.A. & Salinger, M.J. (1993). Adjustment of temperature and rainfall records for site changes. *International Journal of Climatology* **13**, 899–913.

Tsay, R.S. (2015). *MTS: all-purpose toolkit for analyzing multivariate time series (MTS) and estimating multivariate volatility models*. URL `www.CRAN.R-project.org/package=MTS`.

Wei, W.S.W. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley.

Wijngaard, J.B., Klein Tank, A.M.G. & Können, G.P. (2003). Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* **23**, 679–692.

Wilkinson, D. (2014). World energy investment outlook 2014 factsheet. *Technical report*, International Energy Agency.

Yan, Z., Li, Z., Qingxiang li & Jones, P. (2010). Effects of site change and urbanisation in the Beijing temperature series 1977-2006. *International Journal of Climatology* **30**, 1226–1234.

Table 1. Change of location for Perth's temperature recording device. The readings for King's Park begin in the current analysis on 1st January 1943.

Location	Last Recording Date
King's Park	August 1963
Old Hale School	June 1967
Wellington St	May 1992
Perth Airport	November 1993
Mt Lawley	December 2009

Table 2. Seasonal effect of location shift and of climate change (summer = months 1 to 3, autumn = 4 to 5, winter = 6 to 8, spring = 9 to 13). Values that are statistically significant at the 0.05 level are preceded by "*".

	Season	Old Hale School	Wellington Street	Perth Airport	Mt Lawley	Annual Trend
			Maximum Temperature			
Mean Effect	Summer	-0.19	-1.33	-1.14	-1.25	0.0270
	Autumn	-0.44	-0.24	-0.62	-0.01	0.0291
	Winter	0.52	0.16	-0.42	0.40	0.0031
	Spring	0.04	0.07	-0.24	0.66	0.0085
SE	Summer	0.43	0.45	0.94	0.75	0.0146
	Autumn	0.35	0.36	0.76	0.60	0.0118
	Winter	0.18	0.18	0.32	0.30	0.0059
	Spring	0.27	0.28	0.53	0.47	0.0092
t Value	Summer	-0.44	*-2.97	-1.21	-1.67	1.85
	Autumn	-1.26	-0.66	-0.81	-0.01	*2.47
	Winter	*2.96	0.90	-1.31	1.32	0.53
	Spring	0.13	0.24	-0.45	1.39	0.92
			Minimum Temperature			
Mean Effect	Summer	-0.83	-1.38	-3.83	-3.34	0.0346
	Autumn	0.05	0.26	-1.37	-1.66	0.0175
	Winter	0.04	0.61	-0.38	-1.40	0.0029
	Spring	-0.06	0.26	-1.94	-1.29	0.0170
SE	Summer	0.20	0.20	0.43	0.34	0.0066
	Autumn	0.28	0.29	0.61	0.49	0.0095
	Winter	0.34	0.33	0.59	0.56	0.0109
	Spring	0.16	0.16	0.29	0.27	0.0052
t Value	Summer	*-4.21	*-6.79	*-8.97	*-9.81	*5.22
	Autumn	0.19	0.88	*-2.25	*-3.41	1.84
	Winter	0.12	1.88	-0.64	*-2.51	0.27
	Spring	-0.38	1.63	*-6.74	*-4.81	*3.26

Chapter 8

Conclusions

The aim of this thesis was to define, identify, fit and apply models for replicated stationary and invertible time series processes. A literature review on replicated processes revealed that they are encountered in a wide range of areas including panel and longitudinal studies in econometrics and medical statistics. However much of the existing work uses AR(1) models only and the parameter estimation typically employs least squares or method-of-moments approaches rather than maximum likelihood (which is commonly used in standard time series analysis (see Chatfield [2003] p. 65)).

8.1 Joint ARMA Processes

Replicated stationary and invertible time series are represented in this thesis by so-called Joint ARMA or JARMA processes. These are processes that have shared ARMA coefficients with innovations variances that are either in common (Replicated ARMA or RARMA processes), vary with the series (Almost Identical ARMA or AIARMA processes) or are proportional to the series mean (Conditional AIARMA or CAIARMA processes). For a shared innovations variance matrix, the extension to Joint VARMA (Replicated VARMA) processes is undertaken. All models include extraneous variables.

In this thesis it is proven that a RARMA process can be represented by one stationary and invertible univariate series using interleaving and this is shown to

also apply to Replicated VARMA processes. The interleaving of the replicated series allows ARMA and VARMA models to be fitted to all series simultaneously using readily available time series software. The spectral density of an interleaved Replicated ARMA (and VARMA) process is shown to be a compressed version of the shared spectral density of each of the replicated series. Conversely the autocorrelation function is an expanded version of the shared autocorrelation function. These properties of an interleaved process provide a simple interpretation of the effects of interleaving.

The JARMA models could be further extended by adding an intercept term to the variance expression for the CAIARMA model. The asymptotics of the associated maximum likelihood estimates can be derived using the methods in this treatise and simulation studies similar to those in Chapter 5 could be undertaken. An intercept would allow the use of this extended CAIARMA model to be employed in likelihood ratio testing against the original CAIARMA model and the RARMA model.

8.2 Empirical Identification

In order to identify JARMA processes, tests for comparing time series were investigated. A comprehensive review of the literature on series testing was undertaken. The existing tests typically involve regression models of the logged ratio of the periodogram ordinates, single-parameter likelihood ratio tests or distance-based measures. There are also tests that employ the autocovariance function. The tests generally assume two series of equal length (but this is discussed further at the end of this section). Tests with relatively high power from published simulation studies are the regression tests and randomisation and bootstrapping tests.

In this treatise, the system of testing for process equivalence begins with testing for the same spectral shape (or, equivalently, the same autocorrelation function). If accepting of the same shape, testing is undertaken for the same scale (that is, the same spectral density or equivalently the same innovations variance). The sequential nature of the shape and scale tests has not, to the author's knowledge, been previously addressed and requires adjustments to the size and power of the scale test.

To attempt to improve on the power of the existing tests and/or develop tests that are simpler to implement, seven new tests are proposed. These tests include four new shape tests using differences across frequencies in the logged ratio of the periodogram ordinates (with two alternatives for the differencing), the distance between the sample autocorrelation functions and the calculation of the sample variance of the logged ratios. A shape test employing a count of periodogram ratios outside critical levels was investigated but showed low power. The three new scale tests employ the Wald test, the mean logged ratios and the Central Limit Theorem applied to the sample mean of the logged ratios.

A likelihood ratio shape test using a quadratic regression model of the logged ratios against frequency and a likelihood ratio scale test using the un-transformed ratios (see Coates and Diggle [1986]) are also employed as reference tests being amongst the most powerful from the literature. It is proven for the first time that the maximum likelihood estimate associated with the likelihood ratio scale test has only one feasible (that is, positive) solution thereby aiding its interpretation.

Size and power simulation studies of the new and existing tests were undertaken. The simulations indicate that the sizes for all tests are slightly higher than their nominal values. The most powerful shape tests for a range of alternative hypotheses are the autocorrelation test, the $N/2^{th}$ difference test and the likelihood ratio regression test. Testing using the variance of the logged ratio and the first difference of the logged ratios show disappointing power. Of the scale tests, the mean logged ratio and the Central Limit Theorem tests show comparative power with the likelihood ratio test.

On the basis of the simulations it is recommended that, for sequential time-series testing, the $N/2^{th}$ difference shape test and the Central Limit Theorem scale test be employed, being simple to implement and relatively powerful. However the $N/2^{th}$ difference test is not so suitable as a graphical test and an alternative is recommended below.

Given the popularity of the Box-Jenkins identification method, a graphical procedure has been developed that is a visual approach to exploring whether two series have the same spectral shape and scale and to determine their common ARMA order. The method uses the Coates and Diggle [1986] shape test, the Central Limit Theorem scale test and a merging of the autocorrelation functions. As mentioned

above, a shape test using a count of periodogram ratios outside critical levels proved unsuitable. However the convenience of this form of count-based test suggests that further research to improve the power is worthwhile. This is particularly the case given that the use of the Coates and Diggle [1986] shape test is not altogether satisfactory as a graphical tool as it results in a test outcome that is not visually determined from the plot. However the logged ratio plot and the overlayed regression model outcome does provide valuable information on the relationship between the two series's processes.

In some hypothesis testing there will be a need to compare more than two time series but this situation is not addressed in this thesis. One possible approach could be to use the maximum likelihood estimation for RARMA, AIARMA and single-series ARMA models in a multiple-series likelihood ratio test of shape and of scale. The same tests may also be used for series of differing lengths which is not considered here. The variances for the autocorrelation shape test can also be readily modified to accommodate varying series length.

8.3 Maximum Likelihood Estimation

Unconditional maximum likelihood estimates of RARMA, AIARMA and CAIARMA models were derived analytically and numerically using unconstrained and constrained (for stationarity and invertibility) joint likelihood and interleaving approaches. In the simulation estimation studies in this thesis, these three methods all use the R function, `arima`, which is employed as part of the optimisation process with some adjustment to the inputs and outputs. This avoids the preparation of extensive bespoke coding making the modelling of replicated series much easier to implement.

For the ARMA (single-series), RARMA, AIARMA and CAIARMA simulation modelling, the constrained joint likelihood approach resulted in the most stable MLE values.

The asymptotic distribution of the maximum likelihood estimates were analytically derived and show a natural progression from the known results for single-series ARMA models. Extensive simulation studies were conducted into the properties of the maximum likelihood estimates of the ARMA, RARMA, AIARMA and CA-

IARMA model parameters.

Across most models, the estimates have low bias even for small samples ($n \leq 96$), show accurate coverage for the associated confidence intervals and follow the asymptotic standard errors. Hence the asymptotic distribution of the estimates can be used as a reliable indicator of estimate behaviour. Moreover, the Hessian-based empirical confidence intervals appear to be reliable. The estimates do not perform well when the ARMA coefficients are near the unit circle nor when the AR and MA polynomials in B are close to cancelling.

An area of some interest in the literature is the contrast in estimator efficiency between using a small number of long series or a large number of short series. This was not fully explored in the current work, given that the simulations were limited to two component series and restricted to maximum likelihood methods. However, for RARMA, AIARMA and CAIARMA versus single-series ARMA processes, the asymptotic and simulation results as well as the Interleaving Theorem indicate that it is the total length of all series that is the determining factor when estimating common parameters, at least when using joint maximum likelihood methods. This issue could be pursued by expanding on the analytical asymptotic results presented here (for a large number of series of finite length) and by further simulations.

The use of the models in this thesis for panel and longitudinal data could also be further explored. This type of data commonly arises in economic, medical and biological studies and often involves a suspected autocorrelated structure given the sequential nature of the readings. The current Joint ARMA and VARMA models could also be extended to incorporate random effects which are often included in panel and longitudinal models.

The joint likelihood and interleaving methods lend themselves to use with factional differencing, non-normal innovations distributions and robust estimation methods (for example, see trimmed likelihood in Bednarski and Clarke [1993]). They could also be extended to the multivariate equivalent of the AIARMA and CAIARMA models.

The process of identification and estimation of the Joint ARMA models could be improved by pursuing research into residual checks for the adequacy of the models. In this regard the Ljung-Box test for white noise residuals could be extended to the case of joint time series. It would use the interleaved residuals and would be

modified to accommodate the assumed independence between the series. With the latter, certain population autocorrelations of the interleaved series would be assumed to be zero.

Finally, much R code has been written for this thesis. It is the author's intention to develop this code into an R package for the identification and fitting of JARMA models.

8.4 Application

The interleaving method for both univariate and multivariate time series was applied to over sixty years of daily maximum and minimum temperatures for Perth, Western Australia (Bowden and Clarke [2012, 2017]). This shows that there have been changes in mean temperatures due to location and climate change. The modelling of interleaved daily data by day-in-the-week returns univariate and multivariate results which concur with known climate change effects. The location change estimates suggest that the current recording site, 5km north of Perth, is 1.8°C colder overnight than the original site at King's Park near the city. The estimated movements in temperature due to climate and location change can be used in correcting historical weather records employed in the simulation of annual electricity maximum demand. The demand simulations employ an estimated model (not included here) linking daily demand to temperatures combined with a long-term daily temperature record and a forecast of underlying demand growth.

Over and above changes in the mean, work by Vasseur et al. [2014] and others suggest that there is evidence of movements in the distribution of temperatures due to climate change. The current work on Perth temperatures using the models in this thesis could be extended to investigate this phenomenon.

The Joint ARMA models can be employed in exploring time series models of daily total solar radiation readings; these radiation models are likely to vary by location and season. Past research suggests some common generating mechanisms involving similar models to those developed in this thesis. Such work will likely require further enhancements to the Joint ARMA models, allowing the coefficients to vary with the series mean. This will build on the author's previous research in Suehrcke, Bowden, and Hollands [2013] which studied monthly solar radiation and

sunshine fraction.

8.5 Personal Reflections

Given the commentary above, there are clearly many prospects for future research into replicated time series. I hope that my own work in this area has followed the advice of Gwilym Jenkins in his excellent 1979 book, "Practical Experiences with Modelling and Forecasting Time Series" where he writes:

> "As a profession, we, as statisticians, have a great deal to contribute....To be effective in a practical context, we need to be concerned first with the solution of problems and second with the relevance of techniques, statistical or otherwise."

I trust that this treatise has contributed to the research community's knowledge of both the applied and theoretical aspects of time series analysis.

Appendix A

Literature Review of Equivalence Tests

The tests of process equivalence referenced in this literature review are detailed in the tables below and can be classified according to the type of test statistic involved. The first three tables (sorted by publication date) detail the form and performance of the tests and the final table shows the design of any associated simulation studies. A summary is contained in Section 3.2.

The capital letter codes described below are used in the column "T" of the tables (with the counts of tests in brackets shown below). For the designations in columns, H_0 (the null hyypothesis), and, H_1(the alternative hypothesis), see Section 3.2.

The first group of tests involves use of the periodogram and these tests can be further sub-grouped according to :

1. R : Regression-based modelling of the log of the ratio of the periodogram ordinates (as a function of frequency) (9)
2. L : Other likelihood ratio tests (LRT) using the periodogram (2)
3. D : Other measures of distance between periodogram ordinates (12) and
4. P : Other periodogram methods (2)

There are also methods that use sample (normalised or non-normalised) autocovariances (C) (5) and other approaches (O) (3) such as the difference in the lengths of the line segments connecting the data points.

A number of the tests below can employ either the normalised or non-normalised estimated second-order moments (that is, the periodogram or autocovariance/autocorrelation values) with little or no adjustment of the methodology. This typically allows testing of either H_B versus H_C or H_{B_1} versus H_D respectively. These tests are indicated by an asterisk in the "T" column. Columns "S" and "P" provide a summary of the size and power outcomes for A(1) process ($\phi_1 = 0.5$) versus white noise for $n = 64$ (when available, at least approximately). An integer below the decimal size or power indicates a different sample size.

Table A.1 – Hypothesis Tests of Process Equality (Periodogram-Based Regression Tests and Other Periodogram-Based Single Parameter Likelihood Ratio Tests) with Size (S) and Power (P) for AR(1) ($\phi_1 = 0.5$) v White Noise, n=64. For column (T) see earlier and for H_0 and H_1 see Section 3.2.

Reference	Description	H_0	H_1	T	S	P
Coates and Diggle [1986]	This employs the LRT using the logs of the ratio of the non-normalised paired periodogram ordinates represented as a quadratic in $i = \lambda_1 + \lambda_2 i + \lambda_3 i^2$ where $\lambda_1 = \lambda_2 = \lambda_3 = 0$ versus $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_3 \neq 0$	H_B	H_C	R	-	~ 0.63, ($\phi_1 = -0.5$), 10%
Coates and Diggle [1986]	This employs LRT using the logs of the ratio of the non-normalised paired periodogram ordinates represented as a quadratic in $i = \lambda_1 + \lambda_2 i + \lambda_3 i^2$ where $\lambda_1 = \lambda_2 = \lambda_3 = 0$ versus $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_3 = 0$.	H_{B_1}	H_{B_2}	R	-	-
Fokianos and Savvides [2008]	This employs the LRT of model of the log of the ratio of the non-normalised periodogram ordinates using cosine terms (only the intercept is non-zero versus all parameters are non-zero).	H_B	H_C	R	-	-
Fokianos and Savvides [2008]	As above (all parameters equal zero versus intercept only is non-zero).	H_{B_1}	H_{B_2}	R	-	-
Vassiliadis and Rigas [2009]	This models the log of the ratio of the paired non-normalised periodogram ordinates as a function of the frequency via a generalised linear model fitted using maximum likelihood (only the intercept is non-zero versus all parameters are non-zero).	H_B	H_C	R	-	-
Vassiliadis and Rigas [2009]	As above (all parameters equal zero versus intercept only is non-zero).	H_{B_1}	H_{B_2}	R	-	-
Lund et al. [2009]	This uses a LRT that the spectral density functions are the same versus unique to each series. The LRT statistic via the CLT and non-normalised periodogram theory leads to a distribution of the statistic.	H_{B_1}	H_D	L	0.055 1,024	~ 0.2 1,024
Jin [2011]	This models the log of the ratio of the non-normalised paired periodogram ordinates using orthogonal Lagrangian polynomials and a χ^2 statistic from a least square fit with an Akaike-like criterion for order selection (only the intercept is non-zero versus all parameters are non-zero).	H_B	H_C	R	0.055	0.59
Jin [2011]	As above (all parameters equal zero versus intercept only is non-zero).	H_{B_1}	H_{B_2}	R	-	-
Tugnait [2013]	LRT of the smoothed non-normalised periodograms using the Daniel window. The LRT compares averaged versus unique periodograms finding the associated LRT asymptotic distribution.	H_{B_1}	H_D	L	-	-
Lu and Li [2013]	The test statistic (adaptive Neyman testing) is the squared coefficients of a Fourier model fitted to the logged ratio of non-normalised periodogram ordinates.	H_{B_1}	H_D	R	0.05 256	0.2-1.0 256

Table A.2 – Hypothesis Tests of Process Equality (Periodogram-Based Distance Tests) with Size (S) and Power (P) for AR(1) ($\phi_1 = 0.5$), n=64. For column (T) see earlier and for H_0 and H_1 see Section 3.2.

Reference	Description	H_0	H_1	T	S	P
Coates and Diggle [1986]	It uses the extrema of the log of the non-normalised periodogram ratios with periodogram distribution theory.	H_{B_1}	H_D	D	-	<0.13 10%
Coates and Diggle [1986]	It employs the range of the log periodogram ratios using periodogram distribution theory. There is no difference between the use of normalised or non-normalised statistics.	H_B	H_C	D	-	<0.16 10%
Coates and Diggle [1986]	This exact K-S test uses $c_i = \sum_{i=1}^{N} \log\left\{1 + \hat{\alpha_i}^{-1}\right\}$ and c_i/c_N with non-normalised periodogram.	H_{B_1}	H_D	D	-	<0.57 10%
Diggle and Fisher [1991]	A K-S and CVM distance between two *non-normalised* periodograms uses randomisation for probability density function under H_{B_1}.	H_{B_1}	H_D	D*	0.05	0.65
Chik [2002]	A randomisation procedure derives the pdf of the K-S measure of distance between two periodograms.	H_{B_1}	H_D	D*	-	0.64
Caiado et al. [2006]	Various distance measures between periodograms are used for classification purposes.	H_{B_1}	H_D	D*	-	-
Luengo et al. [2006]	A bootstrapped test is based on the average squared distance between the non-normalised periodogram ordinates of two groups of series.	H_{B_1}	H_D	D*	-	-
Lund et al. [2009]	This uses the CLT with the mean of the absolute values of the log ratios of the non-normalised periodogram ordinates which under H_0 is a mean of iid log-logistic random vars.	H_{B_1}	H_D	D	0.059 1,024	~ 0.2 1,024
Jentsch and Pauly [2012]	This employs a weighted (c_i) sum of squares of the difference between non-normalised periodogram ordinates distributed as weighted sum of iid double exponential distributed random vars.	H_{B_1}	H_D	D	0.09	0.42
Jentsch and Pauly [2012]	This utilises a weighted (d_i) sum of squares of the squared log of the ratio of the non-normalised periodogram ordinates distributed as weighted sum of squared iid standard exponential distributed random vars.	H_{B_1}	H_D	D	0.06	<0.10
Hidalgo and Souza [2014]	For a large number of series, the sum of squares of the ratio of the non-normalised periodogram ordinates and the averaged ordinates at each frequency is employed. The asymptotic distribution is derived.	H_{B_1}	H_D	D	-	-
Jentsch and Pauly [2015]	The test statistic is a randomised L2-measure of the difference between the non-normalised periodogram of a multivariate process. The asymptotic distribution is derived.	H_{B_1}	H_D	D	-	-

Table A.3 – Hypothesis Tests of Process Equality (Other Periodogram-Based Tests, Autocovariance-Based Tests and Other Tests) with Size(S) and Power(P) for AR(1) ($\phi_1 = 0.5$), n=64. For column (T) see earlier and for H_0 and H_1 see Section 3.2.

Reference	Description	H_0	H_1	T	S	P
Maharaj [2000] See Alonso et al.	The moving block time series bootstrap method derives the empirical probability density function of the K-S and CVM distances between the sample autocorrelations and partial autocorrelations.	H_B	H_C	C*	>0.05 200	0.9 200
Alonso and Maharaj [2005]	This employs Politis and Romano's time series bootstrapping method to estimate the pdf of the sum of squared differences between the sample autocorrelations.	H_B	H_C	C*	~ 0.08 256	~ 0.87 256
Caiado et al. [2006]	Various distance measures between sample autocorrelations, inverse autocorrelations and partial autocorrelations are employed for classification purposes.	H_B	H_C	C*	-	-
Quinn [2006]	A LRT test is utilised involving the difference between estimated AR parameters of the two series (as a mixed spectral model).	H_B	H_C	O	-	-
Lund et al. [2009]	This test uses a count of number of ratios of non-normalised periodogram ordinates outside binomial critical values	H_{B_1}	H_D	P	-	-
Lund et al. [2009]	This uses the known asymptotic variances, covariances and distribution of the sample autocovariances in a distance measure between the sample autocovariance of the two series.	H_{B_1}	H_D	C	0.047 1,024	1.0 1,024
Dette et al. [2011]	The test statistic here is essentially the sum of squares of the non-normalised periodogram ordinates.	H_{B_1}	H_D	P	~ 0.08 128	-
Jin and Wang [2016]	The test statistic is the maximum over all lags of a linear function of the differences in sample autocorrelations. The asymptotic distribution is derived.	H_B	H_C	C	~ 0.06 40	0.60 32/64
Tunno [2015]	This test is based on the the asymptotic probability density function of 1) the difference in total length of the line segment connecting the non-normalised points, 2) of the squared first differences and 3) of the area to the x-axis.	H_{B_1}	H_D	O	-	-
Decowski and Li [2015]	This tests for differences in a wavelet model of each of the non-normalised periodogram ordinates.	H_{B_1}	H_D	O	-	-

Table A.4 – Processes and Parameters used in Simulation

Reference	Type	Processes	n	Sizes	#Sim.
Coates and Diggle [1986]	Power	WN v AR(1)[0.2, 0.4, 0.6, 0.8] WN v AR(2)[0 and above ϕ's] WN v AR(3)[-ϕ's, 0, ϕ's] AR(1)[0.5] v AR(1)[0.1...0.9 by 0.1] WN with var ratio of 1.1 and 1.2	64 1024	0.10 0.05	50 100 (Scale)
Diggle and Fisher [1991]	Size	AR(1)[0, 0.1, 0.5, 0.9] MA(1)[0, 0.1, 0.5, 0.9]	64 256 1,024	0.10 0.05 0.01	1,000
	Power	WN v AR(1)[0.2, 0.4, 0.6, 0.8] WN v AR(2)[0 and above ϕ's] WN v AR(3)[-ϕ's, 0, ϕ's] AR(1)[0.5] v AR(1)[0.1...0.9 by 0.1] WN with var ratio of 1.1 to 1.5 by 0.1	64 1,024	0.10 0.05 0.01	1,000
Lund et al. [2009]	Size	AR(1) [-0.75 to 0.75 by 0.25] MA(1) [-2.0 to 2.0 by 0.5]	1,024	0.05	10,000
	Power	AR(1) v MA(1) [-0.75 to 0.75 by 0.25] MA(1) [1.0] v MA(2)=[1.0]+ [0.10, 0.25, 0.50]	256 512 1,024	0.05	10,000
Dette et al. [2011]	Size	WN MA(1)[0.2]	128 256 512 1,024	0.05 0.10 0.15	1,000
	Power	WN v MA(2)[θ_1s.t.$\rho = [0.1, 0.5]$ given $\theta_2 = [0, 0.5]$]	128 256 512 1,024	0.05 0.10 0.15	1,000
Lu and Li [2013]	Size	AR(1)[-0.75 to 0.75 by 0.25] MA(1)[-0.75 to 0.75 by 0.25]	1,024	0.05	10,000
	Power	AR(1)[-0.75 to 0.75 by 0.25] v MA(1)[$\theta = \sqrt{\phi^2(1-\phi)^2}$. MA(1)= [0.1, 1.0], MA(2)=[0.1,1.0]+[0.1,0.25,0.5] WN v Ann. Seas. AR[0.1,0.25,0.5]	1,024	0.05	10,000
Tunno [2015]	Size	MA but not invertible	1,000	0.05	10,000
Jentsch and Pauly [2015]	Size	AR(1)[0.5] Unclear	50/75 200/300 800/1,200	0.01 0.05 0.10	?
	Power	Unclear but possibly WN v AR(1)	50/75 200/300 800/1,200	0.01 0.05 0.10	?
Jin and Wang [2016]	Size	ARMA(1,1)[(0,0), (0.8,0.0), (0.0,1.0), (0.5,0.1)]	40 120 360	0.10 0.05	1,000
	Power	WN v AR(1) [0.0,0.1,0.2,0.4,0.6] AR(1) [-0.75 to 0.75 by 0.25] v MA(1) [$\theta_1 = sign(\phi_1)\sqrt{\frac{\phi_1}{1-\phi_1}}$]	32 64 128 256 $n_1 \neq n_2$	0.05	1,000

Appendix B

Distributional Theory for Periodograms

This appendix summaries general distributional outcomes and periodogram distribution results that are used in Chapter 3.

Firstly the following probability density functions for various distributions and the associated distributional relationships are listed:

1. **Exponential distribution with parameter** λ (which is designated as $\text{Exp}(\lambda)$). $f_X(x,\lambda)= \lambda e^{-\lambda x},\ x \geq 0$.

2. **Chi-squared distribution with two degrees of freedom**. $f_X(x,2)= \frac{1}{2}e^{-\frac{1}{2}x},\ x \geq 0$, that is, $X \sim \text{Exp}(\frac{1}{2})$.

3. **F distribution with degrees of freedom (2,2)** (that is, $F_{2,2}$). $f_X(x)= (1+x)^{-2},\ x \geq 0$. It has the same distribution as the ratio of two independent chi-squared random variables with two degrees of freedom (that is, the ratio of two independent $\text{Exp}(\frac{1}{2})$ random variables). The natural logarithm of one plus an $F_{2,2}$ random variable is an exponential random variable with parameter one (Coates and Diggle [1986] p. 10).

4. **Logistic distribution with parameters** μ **and** s. $f_X(x)= \frac{e^{-(x-\mu)/s}}{s(1+e^{-(x-\mu)/s})^2}, x \in (-\infty,\infty)$. The natural logarithm of an $F_{2,2}$ random variable is a logistic random variable with a mean 0 and a scale of 1 (see Coates and Diggle [1986] p. 8). The logistic random variable with scale of one has a variance of $\frac{\pi^2}{3}$.

The distributions associated with various functions of the periodogram ordinates is now discussed recalling that, from Section 3, $\alpha_i = \frac{h_2(\omega_i)}{h_1(\omega_i)}$, that is, α_i is the ratio of the spectral densities of the two independent series at frequency, ω_i, and $\{\hat{\alpha}_i\}$ is the estimate of the $\{\alpha_i\}$ from the ratio of the periodogram ordinates for the independent series $\{y_{1,t}\}_{t=1}^n$ and $\{y_{2,t}\}_{t=1}^n$, both from a stationary and invertible normally-distributed process.

The non-normalised periodogram ordinates of $\{y_{1,t}\}_{t=1}^n$ [1] at frequencies $\{\omega_i\}_{i=1}^N$, that is, $\left\{\hat{h}_1(\omega_i)\right\}_{i=1}^N$, with $N = n/2$, are asymptotically independently distributed as $h_1(\omega_i)$ times a chi-squared random variable with two degrees of freedom, that is, an Exp($\frac{1}{2}$) random variable times $h_1(\omega_i)$. If the series $\{y_{1,t}\}_{t=1}^n$ and $\{y_{2,t}\}_{t=1}^n$ are two independent stationary and invertible series (not necessarily from the same process), then $\hat{\alpha}_i = \hat{h}_1(\omega_i)/\hat{h}_2(\omega_i)$ is asymptotically distributed as α_i times an $F_{2,2}$ random variable.

Let $\hat{\delta}_i = \log(\hat{\alpha}_i)$. So $\hat{\delta}_i$ is therefore asymptotically distributed as a logistic random variable with mean, $\log(\alpha_i)$, and scale of one. This derives from $\hat{\delta}_i = \log(\alpha_i) + \log(y)$ where $\log(y)$ is asymptotically distributed as a logistic random variable with mean zero and scale of one. Letting $z_i = \log(1 + \hat{\alpha}_i)$ and assuming $\alpha_i = 1$ then z_i is asymptotically distributed as an Exp(1) random variable.

[1] As in the main text we assume n is even for convenience sake.

Appendix C

Proof of One Positive Root

Theorem 1. *Let,*

$$\sum_{j=1}^{N}(x-a_j)\prod_{k=1\neq j}^{N}(x+a_k)=0, \tag{C.0.1}$$

where the constants $a_j>0$ for all $j=1,\ldots,N$. *Of this polynomial's* N *roots, only one is a positive real number.*

Proof. Consider the polynomial in (C.0.1) but, in the first summation (that is, in $\sum_{j=1}^{N}(x-a_j)\ldots$), let $-a_j \Longrightarrow a_j$ for all $j=1,\ldots,N$. Hence (C.0.1) becomes N additive identical polynomials, each of the form,

$$\prod_{k=1}^{N}(x+a_k)\,. \tag{C.0.2}$$

Let this "sub-polynomial" be expanded in powers of x. For the term, x^i, the associated positive coefficient is the sum of the products of all combinations of $(N-i)$ of the all-positive $\{a_j\}_{j=1}^{N}$. So, for each x^i, the coefficient is the sum of $\binom{N}{i}$ positive additive constants, each of which are one of the unique (combinatorial) products of $N-i$ of the N a_j's . These $\binom{N}{i}$ products are summed across all the sub-polynomials to create the coefficient for x^i in the polynomial (C.0.3) below, which corresponds to N times (C.0.2) (after the $-a_j \Longrightarrow a_j \forall j=1,\ldots,N$ transformation), that is,

$$N\prod_{k=1}^{N}(x+a_k)\,. \tag{C.0.3}$$

The transformation is now reversed, letting $a_j \Longrightarrow -a_j$ in the j^{th} sub-polynomial, $j = 1, \ldots, N$. For each of the additive sets of the N identical combination products of unique form this change of sign affects only $N - i$ of the products (see Table C.1 from the example below). Hence, of each of the N identical (positive) products, $N - i$ become negative. Hence the total of these N products is only negative when $N - i > i$, that is, $i < {}^{N}/_{2}$. This applies equally to all the unique additive products and hence to the sum of all of them, being the coefficient of x^i in (C.0.1)).

Accordingly the coefficient of x^i in (C.0.1) is only negative when $i < {}^{N}/_{2}$. Similarly it is only zero when $i = {}^{N}/_{2}$ (that is, in general, zero coefficients only exist when N is even) and only positive when $i > {}^{N}/_{2}$. This implies that there is only one change of sign of the coefficients of (C.0.1) when sorted in order the powers of x. Hence, by Descartes Rule of Signs, there can be only one positive (real) root.■

As an illustration of part of the detail of this proof, let $N = 4$ so,

$$\begin{aligned}
&(x - a_1)(x + a_2)(x + a_3)(x + a_4) \quad + \\
&(x + a_1)(x - a_2)(x + a_3)(x + a_4) \quad + \\
&(x + a_1)(x + a_2)(x - a_3)(x + a_4) \quad + \\
&(x + a_1)(x + a_2)(x + a_3)(x - a_4) \qquad = 0.
\end{aligned}$$

If $-a_j \Longrightarrow a_j \, \forall \, j = 1, \ldots, 4$, (that is, by each sub-polynomial, $j = 1, \ldots, 4$) then this gives 4 identical sub-polynomials, each of the form,

$$(x + a_1)(x + a_2)(x + a_3)(x + a_4) \quad .$$

Expanding this gives 4 identical sub-polynomials, each of the form,

$$\begin{aligned}
x^4 + (a_1 + \ldots + a_4)x^3 + (a_1a_2 + a_2a_4 + a_2a_3 + a_1a_4 + a_1a_3 + a_3a_4)x^2 + \\
(a_1a_2a_3 + a_1a_2a_4 + a_1a_3a_4 + a_2a_3a_4)x^1 + \\
a_1a_2a_3a_4 \qquad .
\end{aligned}$$

Taking the coefficients of, say, x^1 (that is, $a_1a_2a_3 + a_1a_2a_4 + a_1a_3a_4 + a_2a_3a_4$) and setting in turn, by sub-polynomial, $a_j \Longrightarrow -a_j \; \forall \; j = 1, \ldots, 4$, the following sub-polynomial coefficients and the aggregate (that is, "Sum") coefficient for x^1

result,

Table C.1 – Sub-Polynomials for x^1

Sub-Polynomial	Sub-Polynomial Coefficients
1	$-a_1a_2a_3 - a_1a_2a_4 - a_1a_3a_4 + a_2a_3a_4$
2	$-a_1a_2a_3 - a_1a_2a_4 + a_1a_3a_4 - a_2a_3a_4$
3	$-a_1a_2a_3 + a_1a_2a_4 - a_1a_3a_4 - a_2a_3a_4$
4	$+a_1a_2a_3 - a_1a_2a_4 - a_1a_3a_4 - a_2a_3a_4$
Sum	$-2a_1a_2a_3 - 2a_1a_2a_4 - 2a_1a_3a_4 - 2a_2a_3a_4$

In this case the negatives predominate which, given that this is the aggregate coefficient for x^1, is shown to satisfy $i = 1 < \frac{4}{2} = 2$.

To further illustrate this theorem, the following table shows the outcome regarding positive, zero and negative coefficients associated with powers of x for polynomials of up to order six. The "Cutpoint" specifies the power of x above which the associated coefficient is positive, at which it is zero and below which it is negative. All these polynomials show only one change in the sign of the coefficients when sorted in order of the powers of x.

Table C.2 – Signs of the Coefficients of x^i

Polynomial Order (N)	Sign of Coefficient of Powers of x			Cutpoint ($N/2$)
	Positive	Zero	Negative	
2	2	1	-	1
3	3 2	-	1 0	1.5
4	4 3	2	1 0	2
5	5 4 3	-	2 1 0	2.5
6	6 5 4	3	2 1 0	3

Appendix D

Marginal ARMA Models from Vector ARMA Processes

It is informative to compare the form of the temperature model's VAR filter in Bowden and Clarke [2017] (see Chapter 7) to the univariate ARMA filter from Bowden and Clarke [2012] (see Chapter 6). Corollary 11.1.1 from Lütkepohl [2005] can be used to show that the univariate (marginal) time series from a multivariate VAR(2) process are ARMA(4,2).

To demonstrate this, in the current bivariate case we define a vector, $\boldsymbol{F} = [1, 0]$, and hence $M = 1$ and $K = 2$ as used in Lütkepohl [2005] (note that Corollary 11.1.2 is not applicable as $M \not> 1$). So the (univariate) process, $u_t = \boldsymbol{F}\boldsymbol{x}_t$, is an ARMA($\tilde{p}, \tilde{q}$) process where $\tilde{p} \leq Kp = 2p$ and $\tilde{q} \leq (K-1)p+q = (2-1)p+q = p+q$ (The inequalities accommodate the potential cancellation of AR and MA terms). Hence with $p = 2$ and $q = 0$ (from our VARMAX(2,0) models) the marginal ARMA models are of order at most $(4, 2)$, that is, ARMA(4,2). This compares to the AR(2) models identified and fitted in Bowden and Clarke [2012].

Although not undertaken here it is possible to calculate the parameters of the univariate marginal process from the VARMA parameters by equating the auto-covariances of the marginal ARMA models to the diagonal elements of the cross-covariance matrix of the VARMA process.

Bibliography

B. Abraham and K. Vijayan. Time series analysis for repeated surveys. *Communications in Statistics - Simulation and Computation*, 21:893–908, 1992.

H. K. Alfares and M. Nazeeruddin. Electricity load forecasting: Literature survey and classification of methods. *International Journal of Systems Science*, 33:23–34, 2002.

A. M. Alonso and E. A. Maharaj. On the comparison of time series using subsampling. Technical report, Department of Statistics, Universidad Carlos III de Madrid, 2005.

O. D. Anderson. On a lemma associated with Box, Jenkins and Granger. *Journal of Econometrics*, 3:151–156, 1975.

O. D. Anderson. *Time Series Analysis and Forecasting The Box-Jenkins Approach.* Butterworth, London and Boston, 1976.

T. W. Anderson. Repeated measurements on autoregressive processes. *Journal of the American Statistical Association*, 73:371–378, 1978.

C. F. Ansley, W. A. Spivey, and W. J. Wrobleski. On the structure of moving average processes. Technical report, Graduate School of Business Administration, The University of Michigan, 1976.

A. Azzalini. Replicated observations of low order autoregressive time series. *Journal of Time Series Analysis*, 2:63–70, 1981.

A. Azzalini. Estimation and hypothesis testing for collections of autoregressive time series. *Biometrika*, 71:85–90, 1984.

R. Baillie. *Time Series Econometrics*. Forthcoming.

B. H. Baltagi. *Econometric Analysis of Panel Data*. John Wliey and Sons, 2005.

N. Beck. Time-series-cross-section data. *Statistica Neerlandica*, 55:111–133, 2001.

T. Bednarski and B. R. Clarke. Trimmed likelihood estimation of location and scale of the normal distribution. *Australian Journal of Statistics*, 35:141–153, 1993.

R. S. Bowden and B. R. Clarke. A single series representation of multiple independent ARMA processes. *Journal of Time Series Analysis*, 33:304–311, 2012.

R. S. Bowden and B. R. Clarke. Using multivariate time series to estimate location and climate change effects on temperatures employed in future electricity demand simulation. *Australian and New Zealand Journal of Statistics*, 2017. Accepted for publication.

G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag New York, 1991.

M. W. Browne and J. R. Nesselroade. Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. In *Contemporary Psychometrics: A Festschrift for Roderick P. McDonald*. Lawrance Erlbaum Associates Inc., 2005.

M. W. Browne and G. Zhang. Repeated time series models for learning data. In *Data Analytical Techniques for Dynamical Systems in the Social and Behavioral Sciences*. Mahwah, NJ, 2007.

J. Caiado, N. Crato, and D. Peña. A periodogram-based metric for time series classification. *Computational Statistics and Data Analysis*, 50:2668–2684, 2006.

F. Camacho, A. I. McLeod, and K. W. Hipel. Contemporaneous bivariate time series. *Biometrika*, 74:103–113, 1987a.

F. Camacho, A. I. McLeod, and K. W. Hipel. Multivariate contemporaneous ARMA model with hydrological applications. *Stochastic Hydrology and Hydraulics*, 1: 141–154, 1987b.

A. C. Cameron and P. K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005.

C. Chatfield. *The Analysis of Time Series: An Introduction*. Chapman and Hall, 6th edition, 2003.

W-D. Chen. An approximate likelihood function for panel data with a mixed ARMA(p,q) remainder disturbance model. *Journal of Time Series Analysis*, 27:911–921, 2006.

Z. Chik. A periodogram-based test method for comparing stationary stochastic signals. *Pakistan Journal of Information and Technology*, 1:208–212, 2002.

T. Cipra. Many short time series with outliers and missing observations. In *Proceedings of the 52nd Session of the International Statistical Institute*, 1999.

D. S. Coates and P. J. Diggle. Tests for comparing two estimated spectral densities. *Journal of Time Series Analysis*, 7:7–20, 1986.

G. M. Cordeiro and R. Klein. Bias correction in ARMA models. *Statistics and Probability Letters*, 19(3):169–176, 1994.

D. R. Cox and P. J. Solomon. On testing for serial correlation in large numbers of small samples. *Biometrika*, 75:145–148, 1988.

J. Decowski and L. Li. Wavelet-based tests for comparing two time series with unequal lengths. *Journal of Time Series Analysis*, 36:189–208, 2015.

J. M. del Castillo. Slash distributions of the sum of independent logistic random variables. *Statistics and Probability Letters*, 110:111–118, 2016.

W. Dent and A. S. Min. A Monte Carlo study of autoregressive integreated moving average processes. *Journal of Econometrics*, 7:23–55, 1978.

H. Dette, T. Kinsvater, and M. Vetter. Testing nonparametric hypotheses for stationary processes by estimating minimal distances. *Journal of Time Series Analysis*, 32:447–461, 2011.

P. J. Diggle and N. I. Fisher. Nonparametric comparison of cumulative periodograms. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 40:423–434, 1991.

P. J. Diggle, P. Heargerty, K-Y. Liang, and S. L. Zeger. *Analysis of Longitudinal Data*. Oxford University Press, 2nd edition, 2002.

S. H. C. du Toit and M. W. Browne. Structural equation modeling of multivariate time series. *Multivariate Behavioural Research*, 42:67–101, 2007.

B. Efron and D. V. Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65:457–482, 1978.

E. M. R. A. Engel. A unified approach to the study of sums, products, time-aggregation and other functions of ARMA processes. *Journal of Time Series Analysis*, 5:159–171, 1984.

M. Feder. Time series analysis of repeated surveys: The state-space approach. *Statistica Neerlandica*, 55:182–199, 2001.

K. Fokianos and A. Savvides. On comparing several spectral densities. *Technometrics*, 50:317–331, 2008.

E. O. George and G. S. Mudholkar. On the convolution of logistic random variables. *Metrika*, 30:1–13, 1983.

P. Gilbert. *Brief user's guide: dynamic systems estimation distributed with the dse package*. Bank of Canada, Bank of Canada, 2006. URL `www.cran.r-project.org/web/packages/dse/vignettes/Guide.pdf`.

V. A. Graham, K. G. T. Hollands, and T. E. Unny. A time series model for K_t with application to global synthetic weather generation. *Solar Energy*, 40:83–92, 1988.

C. W. J. Granger. Aggregation of time series variables - a survey. Technical report, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, 1988.

C. W. J. Granger and M. J. Morris. Time series modelling and interpretation. *Journal of the Royal Statistical Society Series A*, 139:246–257, 1976.

A. Grant. Discriminant analysis of time series. Master's thesis, Department of Statistics, Faculty of Science and Engineering, Macquarie University, 2015.

J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

E. J. Hannan. The asymptotic theory of linear time-series models. *Journal of Applied Probability*, 10:130–145, 1973.

M. A. Hauser. Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study. *Journal of Statistical Planning and Inference*, 80:229–255, 1999.

J. Hidalgo and P. C. L. Souza. *Topics in Nonparametric Statistics*, chapter Testing for equality of an increasing number of spectral density functions, pages 137–154. Springer Science+Business Media New York, 2014.

G. M. Jenkins. *Practical Experiences with Modelling and Forecasting Time Series*. Gwilym Jenkins & Partners (Overseas), 1979.

G. M. Jenkins and D. G. Watts. *Spectral Analysis and its Applications*. Holden-Day Inc., 1968.

C. Jentsch and M. Pauly. A note on using periodogram-based distances for comparing spectral densities. *Statistics and Probability Letters*, 82:158–164, 2012.

C. Jentsch and M. Pauly. Testing equality of spectral densities using randomization techniques. *Bernoulli*, 21:697–739, 2015.

L. Jin. A data-driven test to compare two or multiple time series. *Computational Statistics and Data Analysis*, 55:2183–2196, 2011.

L. Jin and S. Wang. A new test for checking the equality of the correlation structures of two time series. *Journal of Time Series Analysis*, 37:355–368, 2016.

R. H. Jones. Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, 22:389–395, 1980.

G. Kitagawa. *Introduction to Time Series Modelling*. CRC Press, 2010.

T. Krone, C. J. Albers, and M. E. Timmerman. A comparative simulation study of AR(1) estimators in short time series. *Quality and Quantity*, 51:1–21, 2017.

J. Ledolter and L. Chang-Soo. Analysis of many short time sequences: Forecast improvements achieved by shrinkage. *Journal of Forecasting*, 12:1–11, 1993.

K. Lu and L. Li. On Fan's adaptive Neyman tests for comparing two spectral densities. *Journal of Statistical Computation and Simulation*, 83:1585–1601, 2013.

I. Luengo, C. N. Hernández, and P. Saavedra. Test to compare two population logspectra. *Computational Statistics*, 21:91–101, 2006.

R. Lund, H. Bassily, and B. Vidakovic. Testing equality of stationary autocovariances. *Journal of Time Series Analysis*, 30:332–348, 2009.

H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, 2005.

T. E. MaCurdy. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics*, 18:83–114, 1982.

E. A. Maharaj. Comparison of stationary time series using distribution-free methods. In *COMPSTAT*, 2000.

A. I. McLeod, H. Yu, and E. Mahdi. *Time Series Analysis: Methods and Applications*, volume 30 of *Handbook of Statistics*, chapter Time series analysis with R, pages 661–712. North-Holland, 2012.

E. Moral-Benito, P. Allison, and R. Williams. Dynamic panel data modelling using maximum likelihood: An alternative to Arellano-Bond. Technical Report 1703, Banco de Espãna and Working Paper, 2017.

B. Nandram and J. D. Petruccelli. A Bayesian analysis of autoregressive time series panel data. *Journal of Business and Economic Statistics*, 15:328–334, 1997.

M. O. Ojo. A remark on the convolution of the generalised logistic random variable. *Interstat*, 9:1–5, 2003.

E. Parzen and M. Pagano. An approach to modelling seasonally stationary time series. *Journal of Econometrics*, 9:137–153, 1979.

P. Paxton, P. J. Curran, K. A. Bollen, J. Kirby, and F. Chen. Monte Carlo experiments: Design and implementation. *Structural Equation Modelling: A Multidisciplinary Journal*, 8:287–312, 2001.

M. S. Peiris and C. R. Rao. A note on testing for serial correlation in large numbers of small samples using tail probability approximations. *Communications in Statistics - Theory and Methods*, 33:1767–1777, 2004.

M. S. Peiris, R. Mellor, and P. Ainkaran. Maximum likelihood estimation for short time series with replicated observations: A simulation study. *Interstat*, 9:1–16, 2003.

D. I. Perera, M. S. Peiris, J. Robinson, and N. C. Weber. The empirical saddlepoint method applied to testing for serial correlation in panel time series data. *Statistics and Probability Letters*, 78:2876–2882, 2008.

D. A. Pierce. Least squares estimation in the regression model with autoregressive-moving average errors. *Biometrika*, 58:299–312, 1971.

R. Prado and M. West. *Time Series Modelling, Inference and Forecasting*. Chapman and Hall/CRC, 2010.

M. B. Priestley. *Spectral Analysis and Time Series, Volume I and II*. Academic Press, 1981.

B. G. Quinn. Statistical methods of spectrum change detection. *Digital Signal Processing*, 16:588–596, 2006.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL `http://www.R-project.org`.

N. M. Razali and Y. B. Wah. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2:21–33, 2011.

A. J. Scott and T. M. F. Smith. Analysis of repeated surveys using time series methods. *Journal of the American Statistical Association*, 69:674–678, 1974.

G. Shi and N. R. Chaganty. Application of quasi-least squares to analyse replicated autoregressive time series regression models. *Journal of Applied Statistics*, 31 (10):1147–1156, 2004.

R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications with R Examples*. Springer, 2011.

S. Stevenson. A comparison of the forecasting abilities of ARIMA models. In *Proceedings of the Pacific-Rim Real Estate Society Annual Conference, Brisbane, Australia, January 19-22*, 2003.

H. Suehrcke, R. S. Bowden, and K. G. T. Hollands. Relationship between sunshine duration and solar radiation. *Solar Energy*, 92:160–171, 2013.

M. Taboga. *Lectures on probability and statistics*. http://www.statlect.com, 2010.

J. K. Tugnait. Wireless user authentication via comparison of power spectral densities. *IEEE Journal on Selected Areas in Communications*, 31:1791–1802, 2013.

F. Tunno. Bounded area tests for comparing the dynamics between ARMA processes. *Communications in Statistics - Theory and Methods*, 44:3921–3941, 2015.

D. A. Vasseur, J. P. DeLong, B. Gilbert, H. S. Greig, C. D. G. Harley, K. S. McCann, V. Savage, T. D. Tunney, and M. I. O'Connor. Increased temperature

variation poses a greater risk to species than climate warming. *Proceedings of the Royal Society of London B: Biological Sciences*, 281, 2014.

V. G. Vassiliadis and A. G. Rigas. A semi-parametric test based on a quasi-likelihood approach for comparing the estimated spectra of two neural spike trains. *Neurocomputing*, 72:3212–3219, 2009.

J. Wästlund. Random assignment and shortest path problems. In *Fourth Colloquium on Mathematics and Computer Science Algorithms, Nancy, France, 2006*, pages 31–36, 2006.

P. Watson and S. M. A. Nicholls. ARIMA modelling in short data sets: Some Monte Carlo results. *Social and Economic Studies*, 41:53–75, 1992.

W-K. Wong and R. B. Miller. Repeated time series analysis of ARIMA-noise models. *Journal of Business and Economic Statistics*, 8:243–250, 1990.

W-K. Wong, R. B. Miller, and K. Shrestha. Maximum likelihood estimation of ARMA model with error processes for replicated observations, Working Paper No. 0217. Technical report, Department of Economics, National University of Singapore, 2002.

F. Yates. *Sampling Methods for Censuses and Surveys*. Charles Griffon and Co. Ltd, 1960.