Bias Corrected Minimum Distance Estimator for Short and Long Memory Models

Gustavo de Carvalho Lana
Orientadora: Glaura da Conceição Franco
Co-Orientador: Sokol Ndreca

Departamento de Estatística
Universidade Federal de Minas Gerais

Belo Horizonte, Abril de 2016

Resumo

Este trabalho propõe um novo estimador de mínima distância (MDE) para os parâmetros de modelos de memória curta e longa. Este estimador de mínima distância com correção de vício (BCMDE) considera uma correção para o MDE usual de modo a levar em consideração o vício da função de autocorrelação amostral quando a média é desconhecida. Provamos a concistência fraca do BCMDE para modelo ARFIMA(p, q, d) geral e derivamos sua distribuição assintótica no caso dos modelos ARFIMA$(0, d, 0)$, AR(1) e MA(1). Estudos de simulação mostram que para tamanhos amostrais finitos o BCMDE frequentemente possui erro quadrático médio menor que o estimador Whittle (para memória longa) e estimador de máxima verossimilhança (para memória curta). Ademais, quando a média do processo não é constante no tempo, o BCMDE é também menos viciado que o estimador Whittle.

Palavras-chave: funções de autocorrelação e autocovariância, modelos ARMA e ARFIMA, média desconhecida, estimador Whittle.

Abstract

This work proposes a new minimum distance estimator (MDE) for the parameters of short and long memory models. This bias corrected minimum distance estimator (BCMDE) considers a correction in the usual MDE to account for the bias of the sample autocorrelation function when the mean is unknown. We prove the weak consistency of the BCMDE for the general ARFIMA(p, d, q) model and derive its asymptotic distribution in the case of the ARFIMA$(0, d, 0)$, AR(1) and MA(1) models. Simulation studies show that for finite sample sizes, the BCMDE often has a lower mean squared error compared to the commonly used Whittle estimator (for long memory) and the maximum likelihood estimator (for short memory). Additionaly, when the mean of the process is non constant in time, the BCMDE is also less biased than the Whittle estimator.

Keywords: sample autocorrelation and autocovariance functions, ARMA and ARFIMA models, unknown mean, Whittle estimator.

Contents

Chapter 1

Introduction

A stochastic process is a family of random variables indexed in time and a time series can be defined as one of its realizations. The most used model for time series is the so-called ARMA(p, q) model (Box and Jenkins, 1976), ARMA standing for autoregressive moving average. In a series that follows an ARMA model, each value linearly depends on the last p values of the series and on the last q values of a white noise process which are the random errors of the model. The popularity of the ARMA model stems from its simplicity and the variety of autocorrelation function forms it can assume. A peculiarity of the ARMA model is that the autocorrelation function always has an exponential decay. That is, $\rho_k \sim c^k$, where ρ_k is the autocorrelation function at lag k and $|c| < 1$.

The autoregressive integrated moving average (ARIMA) model is a generalization of the ARMA model for non-stationary time series. A non-stationary process follows the ARIMA model if an integer number of differentiations of the process leads to a process following the ARMA model.

The fact that the ARMA model has autocorrelations functions with fast decay created the need for models that allow for slow decay. The ARFIMA model (Hosking, 1981, Granger and Joyeux, 1980) is a generalization of the ARIMA model for time series with the long memory property. Besides the usual p autoregressive and q moving average parameters , the ARFIMA model includes a so-called memory parameter (usually denoted by d) which regulates the long term behavior of the autocorrelation. In the ARFIMA model, the autocorrelation function has hyperbolic decay, that is, $\rho_k \sim k^c$. As such, the autocorrelation function is not absolutely summable: $\sum_{i=1}^{\infty} |\rho_i| = \infty$. Autocorrelation functions with that behaviour are called slowly-decaying functions.

The SARFIMA (Porter-Hudak, 1990) model is an extension to the ARFIMA model for seasonal time series. It can be also seen as an extension to the SARIMA model (seasonal ARIMA) for series with long memory. Series which follows the SARFIMA model may present short memory seasonality as well as long memory seasonality. Its non-seasonal components may also present long memory or short memory. For some properties of invertibility and stationarity on SARFIMA models see Bisognin and Lopes (2009).

The most common estimator for the memory parameter of both ARFIMA and SARFIMA models is the Whittle estimator (Whittle, 1951, Fox and Taqqu, 1986), an approximation of the maximum likelihood estimator. The Whittle estimator is based on the periodogram, an estimator of the spectral function consisting in a Fourier transform of the autocorrelation function. Many other estimators have arised in the literature, for example, the estimator of Haslett-Raftery (Haslett and Raftery, 1989), the GPH (Geweke and Porter-Hudak, 1983) and the SPR (Reisen, 1994). The Haslett-Raftery estimator is an approximate minimum square error estimator with truncation. The GPH estimator is based on a linear regression on the lower frequencies of the periodogram. The SPR is also based on a linear regression, but on the lower frequencies of a smoothed periodogram. Simulation studies show that the Whittle estimator possesses a good combination of accuracy and computational simplicity (Rea et al. 2013, Palma, 2007).

Recently, many estimators have appeared in the literature based on the sample autocorrelations which have some intuitive appeals. For example, sample correlations are consistent at each lag, even for long-memory processes (Hosking, 1996). Back in 1986, Andel suggested using the first sample autocorrelation to estimate d for the ARFIMA$(0, d, 0)$ model. Tieslau et al. (1996) introduced the minimum distance estimator (MDE), which allows for the use of more than one lag of sample autocorrelations for the ARFIMA(p, d, q) model and derived its asymptotic distribution for $d \in (-0.5, 0.25)$. The MDE minimizes the distance between the sample autocorrelations and the respective theoretical autocorrelations. Zevallos and Palma (2013) proposed a filtered version of the MDE estimator in order to obtain an asymptotic distribution also for $d \in [0.25, 0.5)$, henceforth called the MDEFF. Another minimum distance estimator is the one of Mayoral (2007) which minimizes the distance between the sample and theoretical autocorrelations of the residuals. The theoretical autocorrelations of the residuals, evidently, are supposed to be zero for any lag different from zero. As far as the author knows, the MDE, MDEFF and Mayoral estimators were never tested in the literature to SARFIMA models, nor were the asymptotic properties in such cases studied.

When the mean of the process is unknown, a very common situation in practice, sample autocovariances and autocorrelations are biased. There are many studies in the literature about this subject going as far as Marriott and Pope (1954). For instance, Hassani et al. (2012) discuss how this bias may affect the identification of long memory processes. Arnau and Bono (2001) and Huitema and McKean (1994) suggest alternative autocorrelation estimators with lower bias. Although these alternative estimators do tend to reduce the bias, they do not take into account the fact that the bias is not a function of the sample size alone. The values of autocorrelation in other lags can affect the bias too.

If the sample autocorrelations are biased, the same could be expected to happen to the minimum distance estimators, as they rely on these statistics. Thus, the main objective of this work is to propose a minimum distance estimator that takes these sample autocorrelation and autocovariance bias into account, instead of trying to correct them. This can be done by minimizing the distance between the sample autocorrelations and their expectations, given a set of parameters and sample size. It is very complicated to calculate the exact expectation of sample autocorrelations as they are not a linear combination of the sample, unlike the sample autocovariance. But instead we use a reasonable approximation for it. We call this new estimator the bias corrected minimum distance estimator (BCMDE).

Several models were contemplated in this work. Besides the ARFIMA and SARFIMA models, we have also considered the short memory ARMA model. Finally, models in which the mean is not a constant function of time are also presented. For instance, models with structural break or models in which the mean is a linear function of time.

We show that the BCMDE is weakly consistent and we obtain its asymptotic distribution in the case of ARFIMA$(0, d, 0)$, AR(1) and MA(1) models. We have also performed a large simulation study to assess the small sample properties of the BCMDE and make a comparison with some other estimators in the literature. The simulation studies showed that the BCMDE is more precise on many instances. In particular, for ARFIMA$(0, d, 0)$ with constant mean or not, ARFIMA(1,d,0) and AR(1) models. The BCMDE also reduced the bias (compared to the Whittle) in the simulations for models AR(1) and ARFIMA$(0, d, 0)$ with non constant mean.

It should be clear here that the expression "bias corrected" in the BCMDE name refers to a correction for the bias of the sample autocorrelation. It is a statement about how it is constructed, not necessarily about its properties.

This work is organized in the following way. In Chapter 2 we review basic definitions of stationarity as some properties of the autocorrelation and autocovariance functions and their estimation. In Chapter 3 the ARFIMA and SARFIMA models are defined and the main methods of estimation of their parameters are reviewed. In Chapter 4 we introduce the main contribution of this work, the Bias Corrected Minimum Distance Estimator. Chapter 5 presents simulation results in order to compare the BCMDE with other estimators in the literature. In Chapter 6 a simple application to a real time series is presented. Finally, Chapter 7 includes conclusion and future works.

Chapter 2

Stationary time series and estimation of the autocovariance and autocorrelation functions

We deal in this work with a class of stochastic process in which time is discrete. Henceforth, any reference to a stochastic process here should be understood as a reference to a stochastic process in discrete time. Likewise, whenever the term 'time series' is used here, it should also be understood in the same way.

The autocovariance function of a stochastic process, $\{X_t\}, t \in \mathbb{Z}$ is defined as

$$\gamma_{t_1,t_2} = Cov(X_{t_1}, X_{t_2}).$$

Note that $\gamma_{t_1,t_2} = \gamma_{t_2,t_1}$. A stochastic process is called weakly stationary if

1. $E(X_t) = \mu$, for any $t \in \mathbb{Z}$, where $\mu \in \mathbb{R}$.

2. $\text{Var}(X_t) = \sigma^2$, for any $t \in \mathbb{Z}$, where $\sigma^2 < \infty$.

3. $\text{Cov}(X_{t_1}, X_{t_1+k}) = \text{Cov}(X_{t_2}, X_{t_2+k})$ for any $t_1, t_2, k \in \mathbb{Z}$.

From now on, for simplicity, we will call a weakly stationary stochastic process as a stationary stochastic process. In a stationary time series, the autocovariance function depends only on the difference between the indexes, therefore the autocovariance function is represented as a function of a single argument:

$$\gamma_k = \mathrm{Cov}(X_t, X_{t+k}), \quad t, k \in \mathbb{Z}.$$

Naturally, $\gamma_k = \gamma_{-k}$. The autocorrelation function of a stationary time series is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad k \in \mathbb{Z}.$$

Obvious consequences of this definition are that $\rho_0 = 1$ and $|\rho_k| \leq 1$ for all $k \in \mathbb{Z}$.

The spectral function of a stationary time series with autocovariance function γ is defined as

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j}, \quad \omega \in [-\pi, \pi].$$

Another equivalent representation of the spectral function of stationary series that dispenses the use of complex numbers is

$$f(\omega) = \frac{\gamma_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \gamma_k \cos(\omega k), \quad \omega \in [-\pi, \pi].$$

The spectral function of a stationary process is an even function, so it is common to represent it only on the interval $(0, \pi]$.

2.1 Estimation of γ_k in the presence of a constant mean

Let $\{X_t\}$ be a stationary stochastic process. For a realization of size T of a stochastic process one possible estimator for the autocovariance, γ_k, is

$$\tilde{\gamma}_k = \frac{\sum_{j=1}^{T-k} (X_j - \bar{X})(X_{j+k} - \bar{X})}{T}, \quad k = 0, ..., T-1. \tag{2.1}$$

Another possible estimator is

$$\hat{\gamma}_k = \frac{\sum_{j=1}^{T-k}(X_j - \bar{X})(X_{j+k} - \bar{X})}{T-k}, \quad k = 0, ..., T-1. \tag{2.2}$$

The estimator $\tilde{\gamma}_k$ is more commonly used in the literature (see, for example, Brockwell and Davis, 1991), but here we will use both $\tilde{\gamma}_k$ and $\hat{\gamma}_k$. For example, we will use $\tilde{\gamma}_k$ to calculate the periodogram, while $\hat{\gamma}_k$ will be employed to calculate minimum distance estimators, which are defined in Section 3.4.2 and Chapter 4.

If the true mean of the process is known, it can replace the sample mean in (2.1) and (2.2), and the following estimators can be built:

$$\tilde{\gamma}_k^{'} = \frac{\sum_{j=1}^{T-k}(X_j - \mu)(X_{j+k} - \mu)}{T}, \quad k = 0, ..., T-1,$$

and

$$\hat{\gamma}_k^{'} = \frac{\sum_{j=1}^{T-k}(X_j - \mu)(X_{j+k} - \mu)}{T-k}, \quad k = 0, ..., T-1.$$

It is easy to verify that $E(\tilde{\gamma}_k^{'}) = \gamma_k(T-k)/T$ and $E(\hat{\gamma}_k^{'}) = \gamma_k$. The unbiased estimator under known mean, $\hat{\gamma}_k^{'}$, has some drawbacks. For example, when k is big, the variance of this estimator is excessively large. This can result in an estimation of the spectral function that is too inaccurate. Details on the estimation of the spectral function can be found on Section 2.3. The sample autocovariance is also impaired by the high variance at the higher lags, with the autocovariance function assuming unusual values at these lags (this fact will be illustrated in Section 4.1 for the ARFIMA process). This happens because when k is close to T, $\hat{\gamma}_k^{'}$ is based on fewer sums of the quantity $(X_j - \mu)(X_{j+k} - \mu)$. This does not happen for $\tilde{\gamma}_k^{'}$ because the denominator is T, instead of $T-k$. For minimum distance estimators, though, this is not an issue. These estimators tend to use the smaller lags of the autocovariance function, so the large variance of $\hat{\gamma}_k$ at higher lags will not cause any damage on them.

In practice, though, a known mean is a rare situation. Thus, the estimation of the autocovariance function is usually made through $\tilde{\gamma}_k$ and $\hat{\gamma}_k$.

The issues regarding the variance at higher lags discussed in the previous paragraph are also present in the estimators. Furthermore, both $\tilde{\gamma}_k$ and $\hat{\gamma}_k$ are biased estimators for the autocovariance function, a fact which is already proven in the literature (see Shkolnisky et al, 2008, Priestley, 1981).

In Shkolnisky et al. (2008) the bias of $E(\hat{\gamma}_k)$ is given by an equation with number of operations of order T^2. The expectation of $\hat{\gamma}_k$ is given in the following proposition, with number of operations of order T. To obtain $E(\tilde{\gamma}_k)$, one just needs to multiply $E(\hat{\gamma}_k)$ by $(T-k)/T$.

Proposition 1: The expectation of $\hat{\gamma}_k$ can be written as

$$E(\hat{\gamma}_k) = \gamma_k - \frac{T+k}{T-k}\left[\frac{T\gamma_0 + \sum_{i=1}^{T-1} 2(T-i)\gamma_i}{T^2}\right] + 2\frac{\sum_{i=1}^{k}\sum_{j=1}^{T}\gamma_{|i-j|}}{T(T-k)}, \quad (2.3)$$

which is an equation with number of operations of order of magnitude T. *Proof in Appendix A.*

The bias of $\hat{\gamma}_k$ originates from the fact that $\bar{X}$ was used to estimate μ. We should note that $Var(\bar{X})$ appears in the expectation of $\hat{\gamma}_k$, in Equation (2.3), as

$$Var(\bar{X}) = \frac{T\gamma_0 + \sum_{i=1}^{T-1} 2(T-i)\gamma_i}{T^2}.$$

Throughout this work, we will use the following notations

$$B^{\gamma}_{T,k} = -\frac{T+k}{T-k}\left[\frac{T\gamma_0 + \sum_{i=1}^{T-1} 2(T-i)\gamma_i}{T^2}\right] + 2\frac{\sum_{i=1}^{k}\sum_{j=1}^{T}\gamma_{|i-j|}}{T(T-k)}, \quad (2.4)$$

$$B^{\rho}_{T,k} = -\frac{T+k}{T-k}\left[\frac{T\rho_0 + \sum_{i=1}^{T-1} 2(T-i)\rho_i}{T^2}\right] + 2\frac{\sum_{i=1}^{k}\sum_{j=1}^{T}\rho_{|i-j|}}{T(T-k)}. \quad (2.5)$$

Note that $B^{\rho}_{T,k} = B^{\gamma}_{T,k}/\gamma_0$. Equation (2.4) provides the bias of the sample autocovariance in the context of a model with a single constant mean which is estimated through $\bar{X} = \sum_{i=1}^{T} X_i/T$. Other types of models, like the ones in Section 2.2, may have a different formula for the bias. Both $B^{\gamma}_{T,k}$ and $B^{\rho}_{T,k}$ converge to zero as $T \to \infty$. A detailed proof of that can be seen in the proof of Proposition 6, more specifally on Lemma 2. In a broad way, it is suffices to note that $B^{\gamma}_{T,k}$ is a function of weighted means of the autocovariance function from lag 0 to $T-1$ and that the autocovariance function goes to zero as $T \to \infty$ (and therefore so does the weighted means).

For any estimator of the autocovariance function, the autocorrelation at lag k can be estimated as the ratio between the sample autocovariance at lag k and the sample autocovariance at lag 0.

2.2 Estimation of γ_k in the presence of a non-constant mean

Time series models may be generalized to cases where the mean is not a constant. In what follows, let μ_t be the mean of the process at time t. In the previous section, it was shown that the mean estimation in a model with constant mean causes bias in the autocovariance estimators. Intuitively, the same could be expected to happen if the mean is non-constant. In this work, three specific cases of time series with time varying mean will be considered:

- Simple structural break: $\mu_t = \alpha$ for $t \leq T_0$, $\mu_t = \beta$ for $t > T_0$.
- Simple linear regression: $\mu_t = \alpha + \beta z_t$, $t = 1, ..., T$ where $z_1, ..., z_T$ are non-stochastic regressing variables.
- Non-stochastic seasonality: $\mu_t = \alpha_{t-s\lfloor (t-1)/s \rfloor}$, where s is the period and $\lfloor . \rfloor$ is the floor function.

Under these conditions, the interest is not in estimating the autocovariance function of the series X_t (which is non-stationary), but that of the stationary series $X_t - \mu_t$. As typically the values of μ_t are not known, it is not possible to calculate the sample autocovariances through $X_t - \mu_t$. But it is possible to estimate μ_t and calculate

$$\hat{\gamma}_k = \frac{\sum_{t=1}^{T-k} (X_t - \hat{\mu}_t)(X_{t+k} - \hat{\mu}_{t+k})}{T-k}, \tag{2.6}$$

where $\hat{\mu}_t$ is an unbiased estimator of μ_t. Recall that we will use $\hat{\gamma}_k$ instead of $\tilde{\gamma}_k$ in order to calculate the sample autocovariances when using the minimum distance estimators. Thus in this section the results will be obtained only for $\hat{\gamma}_k$.

The following subsections show the calculation of $\hat{\mu}_t$ for each of the three scenarios listed above and the consequences on the sample autocovariance expectations.

2.2.1 Simple Structural Break

If the series X_t presents a simple structural break at T_0, the mean, before and after the structural break, can be estimated as

$$\hat{\alpha} = \frac{\sum_{i=1}^{T_0} X_i}{T_\alpha},$$

$$\hat{\beta} = \frac{\sum_{i=T_0+1}^{T} X_i}{T_\beta},$$

where $T_\alpha = T_0$ and $T_\beta = T - T_0$ represent, respectively, the time series size before and after the break. The value of μ_t is then estimated as

$$\hat{\mu}_t = \begin{cases} \hat{\alpha}, & t \leq T_0 \\ \hat{\beta}, & t > T_0 \end{cases}.$$

Proposition 2: In the case of simple structural break, and when $T_0 - k > 0, T - T_0 - k > 0$, the expectation of $\hat{\gamma}_k$ is given by

$$E(\hat{\gamma}_k) = \gamma_k - \frac{2\sum_{t=1}^{T_0-k} f_{\hat{\alpha}}(t) + 2\sum_{t=T_0+1}^{T-k} f_{\hat{\beta}}(t)}{T-k} - \frac{\sum_{t=T_0-k+1}^{T_0}[f_{\hat{\beta}}(t) + f_{\hat{\alpha}}(t+k)]}{T-k} + \frac{(T_\alpha - k)f_{\hat{\alpha}^2} + (T_\beta - k)f_{\hat{\beta}^2} + kf_{\hat{\alpha},\hat{\beta}}}{T-k},$$

where

$$f_{\hat{\alpha}}(t) = \frac{\sum_{i=1}^{T_0} \gamma_{|t-i|}}{T_\alpha},$$

$$f_{\hat{\beta}}(t) = \frac{\sum_{i=T_0+1}^{T} \gamma_{|t+k+i|}}{T_\beta},$$

$$f_{\hat{\alpha}^2} = \frac{\gamma_0 + 2\sum_{i=1}^{T_\alpha - 1}(T_\alpha - i)\gamma_i}{T_\alpha^2},$$

$$f_{\hat{\beta}^2} = \frac{\gamma_0 + 2\sum_{i=1}^{T_\beta - 1}(T_\beta - i)\gamma_i}{T_\beta^2},$$

$$f_{\hat{\alpha},\hat{\beta}} = \frac{\sum_{i=1}^{T_0}\sum_{j=T_0+1}^{T}\gamma_{|i-j|}}{T_\alpha T_\beta},$$

Proof in Appendix B.

2.2.2 Simple linear regression

In the case of a simple linear regression, given the independent variable, $z_1, ..., z_T$, the parameters α and β in the equation $\mu_t = \alpha + \beta z_t$ can be estimated through the ordinary least squares method:

$$\hat{\beta} = \frac{\sum_{t=1}^{T}\tilde{z}_t X_t}{\sum_{t=1}^{T}\tilde{z}_t^2}$$

$$\hat{\alpha} = \bar{X} - \hat{\beta}\bar{z},$$

where $\tilde{z}_t = z_t - \bar{z}$, and $\bar{z}$ is the mean of the dependent variable. Then the estimator of μ_t is given by

$$\begin{aligned}\hat{\mu}_t &= \hat{\alpha} + \hat{\beta} z_t \\ &= \bar{X} + \hat{\beta}\tilde{z}_t.\end{aligned}$$

We can also note that:

$$\mu_t = \alpha + \beta(z_t - \bar{z} + \bar{z})$$

$$\mu_t = \alpha + \beta\bar{z} + \beta\tilde{z}_t.$$

We will call the unknown constant $\alpha + \beta\bar{z}$ as $\mu_{1:T}$.

Proposition 3: In the case of a simple linear regression, the expectation of $\hat{\gamma}_k$ is given by

$$E(\hat{\gamma}_k) = \gamma_k - \frac{2\sum_{t=1}^{T-k}\sum_{i=1}^{T}\gamma_{|t-i|}}{(T-k)T} - \frac{\sum_{t=1}^{T-k}\sum_{i=1}^{T}(\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|} + \tilde{z}_t\tilde{z}_i\gamma_{|t+k-i|})}{(T-k)\sum_{i=1}^{T}\tilde{z}_i^2}$$
$$+ \frac{[\sum_{t=1}^{T-k}(\tilde{z}_t + \tilde{z}_{t+k})]\sum_{i=1}^{T}\sum_{j=1}^{T}\tilde{z}_j\gamma_{|i-j|}}{(T-k)T\sum_{i=1}^{T}\tilde{z}_i^2}$$
$$+ \frac{[\sum_{t=1}^{T-k}\tilde{z}_t\tilde{z}_{t+k}]\sum_{i=1}^{T}\sum_{j=1}^{T}\tilde{z}_i\tilde{z}_j\gamma_{|i-j|}}{(T-k)(\sum_{i=1}^{T}\tilde{z}_i^2)^2} + \frac{\sum_{i=1}^{T}\sum_{j=1}^{T}\gamma_{|i-j|}}{T^2}. \tag{2.7}$$

Proof in Appendix C.

In the particular case where $z_t = z_{t-1} + 1$, $t = 2, ..., T$, as when z_t is the time, the formula of the expectation of $\hat{\gamma}_k$ can be simplified, as it is shown in Proposition 4.

Proposition 4: In the case of a simple linear regression, if the independent variable, z_t, satisfies $z_t = z_{t-1} + 1$, the expectation of $\hat{\gamma}_k$ is given by

$$E(\hat{\gamma}_k) = \gamma_k - \frac{2\sum_{t=1}^{T-k}\sum_{i=1}^{T}\gamma_{|t-i|}}{(T-k)T} - \frac{24\sum_{t=1}^{T-k}\sum_{i=1}^{T}z_{t+k}z_i\gamma_{|t-i|}}{(T-k)(T^3-T)}$$
$$+\frac{12[(T-k)^3 - (T-k)(3k^2+1)]\sum_{i=1}^{T}\sum_{j=1}^{T}z_iz_j\gamma_{|i-j|}}{(T-k)(T^3-T)^2} + \frac{\sum_{i=1}^{T}\sum_{j=1}^{T}\gamma_{|i-j|}}{T^2}.$$

Proof in Appendix D.

2.2.3 Non-stochastic seasonality

In the case of non-stochastic seasonality the mean at time t is given by $\mu_t = \alpha_{t-s\lfloor(t-1)/s\rfloor}$. That is, $\mu_1 = \alpha_1$, $\mu_2 = \alpha_2$, $\mu_s = \alpha_s$, $\mu_{s+1} = \alpha_1$, $\mu_{s+2} = \alpha_2$, etc. Each constant term α_l is estimated as

$$\hat{\alpha}_l = \frac{\sum_{i=0}^{\lfloor(T-l)/s\rfloor}X_{l+si}}{\lfloor(T-l)/s\rfloor + 1}, \quad l = 1, ..., s$$

and the estimate of μ_t is given by

$$\hat{\mu}_t = \hat{\alpha}_{t-s\lfloor(t-1)/s\rfloor}.$$

In order to ease the notation, define the functions

$$s_1(t) = t - s\lfloor (t-1)/s \rfloor$$

and

$$s_2(t) = \lfloor (T - s_1(t))/s \rfloor + 1.$$

The function s_1 gives, for each time t, a value between 1 and s in a way such that if $\mu_t = \alpha_r$, then $s_1(t) = r$. The function s_2 gives, for each time t, the number of times with mean $\alpha_{s_1(t)}$.

Proposition 5: In the case of non-stochastic seasonality, the expectation of $\hat{\gamma}_k$ is given by

$$E(\hat{\gamma}_k) = \gamma_k - \sum_{t=1}^{T-k} \frac{\sum_{i=0}^{s_2(t+k)-1} \gamma_{|t-s_1(t+k)-si|}}{(T-k)s_2(t+k)} - \sum_{t=1}^{T-k} \frac{\sum_{i=0}^{s_2(t)-1} \gamma_{|t+k-s_1(t)-si|}}{s_2(t)}$$

$$+ \sum_{t=1}^{s} \left(\left\lfloor \frac{T-k-t}{s} \right\rfloor + 1 \right) \frac{\sum_{i=0}^{s_2(t)-1} \sum_{j=0}^{s_2(t+k)-1} \gamma_{|s_1(t)+Si-s_1(t+k)-sj|}}{(T-k)s_2(t)s_2(t+k)}.$$

Proof in Appendix E.

2.3 Estimation of the spectral function

The most common estimator of the spectral function is the periodogram, which is given by

$$I(\omega) = \frac{\tilde{\gamma}_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{T-1} \tilde{\gamma}_k \cos(\omega k), \quad \omega \in \mathbb{R}.$$

The periodogram can be estimated at any frequency but in practice it is usually calculated at the Fourier frequencies: $\omega_j = 2\pi j/T$, $j = 1, 2, ..., \lfloor T/2 \rfloor$. Note that the formula of $I(\omega)$ is dependent on $\tilde{\gamma}_0, ..., \tilde{\gamma}_{T-1}$ and the weight of each lag is dependent on the frequency. This is why a high variance at higher lags of the sample autocorrelation function can be so detrimental to the periodogram. In Chapter 4 an empirical example of this fact will be presented for the ARFIMA process.

For any frequencies $0 < \omega_1 < ... < \omega_m < \pi$, the periodogram at these points converges in distribution to independent exponential random variables with mean $f(\omega_1), .., f(\omega_m)$ (see Brillinger, 1975).

Chapter 3

Definition and Estimation of ARIMA, ARFIMA and SARFIMA processes

This chapter presents the methodology for ARIMA and ARFIMA processes and their seasonal generalization, the SARFIMA processes. The basic properties of such processes are presented, including the autocorrelation, autocovariance and spectral function and their estimators. Additionally, estimation in short and long memory models is discussed.

3.1 ARIMA Process

The ARIMA(p, d, q) process (Box and Jenkins, 1976) is the most commonly used model for time series. It satisfies the equation

$$\phi(B)(X_t - \mu) = (1 - B)^{-d}\theta(B)a_t, \quad t \in \mathbb{Z},$$

where $\phi(B) = (1 - \phi_1 B - ... - \phi_p B^p)$, $\theta(B) = (1 + \theta_1 B + ... + \theta_q B^q)$, B is the backward shift operator such that $B^k X_t = X_{t-k}$, $\phi_1, ..., \phi_p, \theta_1, ..., \theta_q$ are real numbers, μ is the mean of the process, a_t, is a zero-mean white noise process with $Var(a_t) = \sigma^2 < \infty$ and d is a non-negative integer. The parameters $\phi_1, ..., \phi_p$ are called the autoregressive parameters, $\theta_1, ..., \theta_q$ are called the moving average parameters and d is the differentiation parameter.

When $d = 0$, it is called an ARMA(p, q) process and satisfies $(X_t - \mu) = \phi_1(X_{t-1} - \mu) + ... + \phi_p(X_{t-p} - \mu) + \theta_1 a_{t-1} + ... + \theta_q at - q + a_t$. If X_t is an ARIMA(p, d, q) process and $d = 1$, then $X_t - X_{t-1}$ is an ARMA (p, q) process. The ARMA process is stationary if all roots of $\phi(B)$ lay outside the unit circle in the complex plane. The process is invertible if $d > -0.5$ and $\theta(B)$ has all its roots outside the unitary circle of the complex plane. A process in invertible if it can be written at time t as a linear combination of all past values plus the error at time t and the weights of such linear combination are absolutely summable.

If $p > 0$ and $q = 0$ the process is called an autoregressive process (AR(p)) and if $p = 1$, the autocorrelation function is given by $\rho_k = \phi^k$. For larger values of p the autocorrelation function has a more complicate formula, but it is known to satisfy the recursion $\rho_k = \phi_1\rho_{k-1} + ... + \phi_p\rho_{k-p}$. If $p = 0$ and $q > 0$ the process is called a moving average process (MA(q)) and its autocorrelation function is given by $\rho_k = \sum_{i=0}^{q} \theta_i\theta_{i+k} / \sum_{i=0}^{q} \theta_i^2$, considering that $\theta_0 = 1$ and $\theta_i = 0$ for $i > q$. The autocorrelation function for the general ARMA(p,q) model can be achieved through the splitting method that is described in details in Section 3.2. A common characteristic of the autocorrelation functions of ARMA processes is that all of them are absolutely summable.

The spectral density of the ARMA process is given by

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}, \quad \omega \in \mathbb{R}.$$

3.2 ARFIMA Process

A stochastic process $\{X_t\}$ is an ARFIMA(p, d, q) process (Hosking, 1981, Granger and Joyeux, 1980) if it satisfies

$$X_t - \mu = (1 - B)^{-d} U_t, \quad t \in \mathbb{Z}, \tag{3.1}$$

where $U_t = \frac{\theta(B)}{\phi(B)} a_t$ is an ARMA(p,q) where $d \in \mathbb{R}$ is the memory parameter and

$$(1-B)^{-d} = 1 + \sum_{k=1}^{\infty} \frac{d(1+d)...(k-1+d)}{k!} B^k. \tag{3.2}$$

For $d \neq 0, -1, -2, ...$, Equation (3.2) can be written as

$$(1-B)^{-d} = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} B^k. \tag{3.3}$$

An ARFIMA process described as in (3.2) is a generalization of ARIMA processes for cases in which the parameter d may assume other values besides non-negative integers. It will be stationary if $d < 0.5$ and if all roots of $\phi(B)$ lay outside the unit circle in the complex plane. The process is called invertible if $d > -0.5$ and if $\theta(B)$ has all its roots outside the unitary circle of the complex plane. If $d \in (0, 0.5)$ the process has the property of long memory, characterized by an autocorrelation function that is not absolutely summable. For any ARFIMA process, the autocorrelation and autocovariance functions decay asymptotically proportionally to c^{2d-1} (Hosking, 1996), where c is any positive constant.

Even though the ARFIMA model is often called invertible in the literature when $d \in (-0.5, 0)$, its infinite autoregressive representation does not have absolutely summable coefficients in this interval. To show this, let φ_k, $k = 1, 2, ...$, be the coefficients of its infinite autoregressive form $X_t - \mu = \sum_{k=1}^{\infty} \varphi_k (X_{t-k} - \mu) + a_t$. From equation (3.2), it is easy to show that $\varphi_k = \varphi_{k-1}(k-1-d)/k$, $k \geq 1$. Therefore, for $d < 0$, the sequence φ_k, $k = 1, 2, ...$, decays more slowly than the harmonic sequence, which satisfies, for any term b_k, $k \geq 2$, $b_k = b_{k-1}(k-1)/k$. Accordingly, because the terms of the infinite moving average representation are equal to the terms of the infinite autoregressive representation for the opposite values of d, it follows that for $d \in (0, 0.5)$, the infinite moving average representation is not absolutely summable.

When $p, q = 0$ the process is called a fractional white noise. In this case, for $d < 0.5$, $Var(X_t) = \gamma_0 = (-2d)!/(-d)!^2$ (Hosking, 1981). Its autocorrelation function, for $d < 0.5$, is given by

$$\rho_k = \prod_{i=1}^{k} \frac{i-1+d}{i-d}, \quad k \in \mathbb{Z}. \tag{3.4}$$

If $p > 0$ or $q > 0$ the autocorrelation function is more difficult to be obtained, but it can be accurately calculated through the splitting method (Brockwell and Davis, 1991, Bertelli and Caporin, 2002). Following this method, if $\gamma_k^{(1)}$ is the autocovariance function of the ARMA component and $\gamma_k^{(2)}$ the autocovariance function of the fractional white noise component, then the autocovariance of the ARFIMA process, γ_k, can be decomposed as

$$\gamma_k = \sigma^{-2} \sum_{i=-\infty}^{\infty} \gamma_i^{(1)} \gamma_{i-k}^{(2)}, \quad k \in \mathbb{Z}, \tag{3.5}$$

and the autocorrelation function can be calculated as

$$\rho_k = \frac{\sum_{i=-\infty}^{\infty} \gamma_i^{(1)} \gamma_{i-k}^{(2)}}{\sum_{i=-\infty}^{\infty} \gamma_i^{(1)} \gamma_i^{(2)}}, \quad k \in \mathbb{Z}. \tag{3.6}$$

The splitting method is valid even if the infinite autoregressive representations of the ARFIMA model for $d < 0$ are not absolutely summable. The origin of the splitting method is Proposition 3.1.2 in Brockwell and Davis (1991), stablishing sufficient conditions in order for the process $Y_t = A(B)Z_t$ be a stationary process, where $A(B)$ is a polynomial in B and Z_t is a stationary process. Nevertheless, if $Y_t = A(B)Z_t$ is stationary, the splitting method is still valid for $A(B)$ non-absolutely summable, as it can be easily seen in the proof of the proposition, though such fact is not mentioned there, nor in Bertelli and Caporin (2002).

Let ϱ be a vector of autocorrelations of an ARFIMA process with $d \in (-0.5, 0.25)$ and $\hat{\varrho}$ the vector of sample autocorrelations of ϱ. Hosking (1996) shows that

$$\sqrt{T}(\hat{\varrho} - \varrho) \xrightarrow{D} N(0, C),$$

where C is a matrix whose element C_{ij} is given by

$$C_{ij} = \sum_{l=1}^{\infty} (\rho_{l-i} + \rho_{l+i} - 2\rho_i \rho_l)(\rho_{l-j} + \rho_{l+j} - 2\rho_j \rho_l). \tag{3.7}$$

The spectral density of the ARFIMA process is given by

$$f(\omega) = \frac{\sigma^2}{2\pi}\left(2\sin\frac{\omega}{2}\right)^{-2d}\frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}, \quad \omega \in \mathbb{R}.$$

Let f_U be the spectral density of the differentiated series, $\{U_t\}$, an ARMA(p,q) model with same parameters of the short memory components of the original ARFIMA(p, d, q) model. Note that it is possible to write the spectral function of $\{X_t\}$ as

$$f(\omega) = \left(2\sin\frac{\omega}{2}\right)^{-2d} f_U(\omega).$$

For $d > 0$, the spectral function satisfies $\lim_{\omega\to 0} f(\omega) = \infty$.

3.3 SARFIMA process

The SARFIMA (Porter-Hudak, 1990) process is a generalization of both the ARFIMA and the SARIMA processes (seasonal ARIMA) to account for seasonality and long memory (which might be seasonal or not). The process $\{X_t\}$ is a SARFIMA$(p, d, q) \times (p_s, d_s, q_s)_s$ process if it satisfies the equation

$$\phi(B)\Phi(B^s)X_t = (1-B)^{-d}(1-B^s)^{-d_s}\theta(B)\Theta(B^s)a_t, \quad t \in \mathbb{Z},$$

where $\Phi(B^s) = (1-\Phi_1 B^s - ... - \Phi_{p_s} B^{sp_s})$, $\Theta(B^s) = (1+\Theta_1 B^s + ... + \Theta_{q_s} B^{sq_s})$, $\phi(B)$ and $\theta(B)$ are given as in Section 3.1, $d_s \in \mathbb{R}$ is the seasonal memory parameter, $d \in \mathbb{R}$ is the non-seasonal memory parameter and s is the seasonal period. There are SARFIMA models with more than one period, but this work will focus on models with one period. The stationarity and invertibility conditions are given by Bisognin and Lopes (2009). It will be stationary if $d + d_s < 0.5$, $d_s < 0.5$ and $\phi(B)\Phi(B^s)$ has all roots outside the unit circle and invertible if $d < 0.5$, $d_s < 0.5$ and $\theta(B)\Theta(B^s)$ has all roots outside the unit circle.

Let $\{X_t\}$ be a SARFIMA$(0, 0, 0) \times (p_s, d_s, q_s)_s$ process. The autocovariance and autocorrelation functions of $\{X_t\}$ satisfy $\gamma_{X,k} = \gamma_{Y,k/s}$ and $\rho_{X,k} = \rho_{Y,k/s}$ for $k = 0, \pm s, \pm 2s, ...$ and $\gamma_{X,k} = 0$ and $\rho_{X,k} = 0$ otherwise, where γ_Y and ρ_Y are the autocovariance and autocorrelation functions of an

ARFIMA (p_s,d_s,q_s) process. The general form of the autocovariance function and hence the autocorrelation is rather complicated but it can also be calculated through the splitting method, after obtaining the autocovariances of the seasonal and non-seasonal components. It should be noted, though, that the convolution of two slowly-decreasing functions may take time to converge. Take, as an example, a SARFIMA$(0, 0.2, 0) \times (0, 0.2, 0)_{12}$ process. Let m be the number of sums actually calculated in (3.6). Setting a large value for m (say $m = 5 \times 10^4$) in order to obtain approximations to the true autocorrelations of order 1 and 12, leads to $\rho_1 \approx 0.3388$ and $\rho_{12} \approx 0.3746$, respectively. Assuming these values to be the truth, it is necessary, so that the error due to the truncation be less than 2%, to use $m = 8639$ and $m = 6786$ for the first and twelfth autocorrelations respectively.

The spectral function of a SARFIMA process is given by

$$f(\omega) = \frac{\sigma^2}{2\pi}\left(2\sin\frac{\omega}{2}\right)^{-2d}\frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}\left(2\sin\frac{\omega s}{2}\right)^{-2d_s}\frac{|\Theta(e^{-i\omega s})|^2}{|\Phi(e^{-i\omega s})|^2}, \quad \omega \in \mathbb{R}.$$

As in the case of an ARFIMA process, for $d > 0$, we have $\lim_{\omega\to 0} f(\omega) = \infty$. Furthermore, for $d_s > 0$, $\lim_{\omega\to 2k\pi/s} f(\omega) = \infty$ for any $k \in 0, 1, 2, ...$ At these points the spectral function is undefined.

3.4 Parameter estimation

In this section we review some of the common estimators for ARIMA and ARFIMA models in the literature. The mean, μ, can be estimated separately by $\hat{\mu} = \bar{X}$. The methods here described will be applied to $Z_t = X_t - \bar{X}$, $t = 1, ..., T$.

3.4.1 Estimation in ARIMA processes

This subsection reviews two of the most common methods of obtaining estimators for the ARMA model: conditional sum of squares (CSS) and the maximum likelihood (ML). We are interested in estimating the parameter vector $\delta = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)'$. Note that here the differentiation parameter,

d, is not included in δ. Usually the value of d is chosen in a subjective way. For example, through an analysis of the sample autocorrelation function.

The CSS estimator searchs for the parameter vector δ that minimizes

$$\text{CSS}(\delta) = \sum_{t=p+1}^{T} [Z_t - \hat{Z}_t(\delta, t-1, ..., 1)]^2,$$

where $\hat{Z}_t(\delta, t-1, ..., 1)$ is calculated as

$$\hat{Z}_t(\delta, t-1, ..., 1) = \sum_{i=1}^{\infty} \pi_i(\delta) Z_{t-i},$$

where $\pi_i(\delta)$, $i = 1, 2, ...,$ are the coefficients of the pure autoregressive form of the ARMA process given the parameters δ. For $t < 0$, Z_t is set to be zero.

Let Σ_δ be the autocovariance matrix for $Z = (Z_1, ..., Z_T)$ given δ. Under the assumption of Gaussian errors, the maximum likelihood (ML) estimator maximizes

$$l(\delta, \sigma^2) = -\frac{1}{2T} \log(\sigma^{-2} \det \Sigma_\delta) - \frac{1}{2T\sigma^2} Z'(\Sigma_\delta)^{-1} Z.$$

In order to evaluate the likelihood function, the necessity of calculating the determinant and the inverse of a $T \times T$ matrix can make this procedure rather unpractical. Fortunately, for ARMA processes this is not necessary, as to maximize $L(\delta, \sigma^2)$ is the same as to maximize

$$L(\delta, \sigma^2)^* = (v_1 ... v_T)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} [Z_t - \hat{Z}_t^*(\delta, t-1, ..., 1)]^2 / v_t \right\},$$

where v_t, $t = 1, ..., T$ are the mean squared errors of $Z_t - \hat{Z}_t^*(\delta, t-1, ..., 1)$ as predictors of Z_t and $Z_t - \hat{Z}_t^*(\delta, t-1, ..., 1)$ is the best predictor under δ of Z_t given $Z_1, ..., Z_{t-1}$. See Brockwell and Davis (1991) to learn in details how $v_1, ..., v_T$ and $\hat{Z}_{1|\delta}^*, ..., \hat{Z}_{T|\delta}^*$ are calculated.

Define the autoregressive processes $\mathcal{U}_t$ and $\mathcal{V}_t$ that satisfy the equation $\phi(B)\mathcal{U}_t = a_t$, $\theta(B)\mathcal{V}_t = a_t$. Here again a_t is a white noise process. Finally define the vectors $\mathcal{U} = (\mathcal{U}_t, ..., \mathcal{U}_{t-1+p})'$ and $\mathcal{V} = (\mathcal{V}_t, ..., \mathcal{V}_{t-1+p})'$. The matrix of asymptotic covariances of the ML estimator is given by (Brockwell and Davis, 1981)

$$Var(\hat{\delta}) = \begin{bmatrix} E(\mathcal{U}\mathcal{U}') & E(\mathcal{U}\mathcal{V}') \\ E(\mathcal{V}\mathcal{U}') & E(\mathcal{V}\mathcal{V}') \end{bmatrix}^{-1}.$$

The CSS estimator has the same asymptotic variance than the maximum likelihood estimator (Brockwell and Davis, 1981).

3.5 Estimation in ARFIMA and SARFIMA processes

This section reviews some estimators for the parameters for ARFIMA and SARFIMA models, which are the maximum likelihood estimator, the Whittle (Whittle, 1951, Fox and Taqqu, 1986) estimator, the minimum distance estimator (Tieslau et al., 1996) and the minimum distance estimator of the filtered series (Zevallos and Palma, 2013). The first three can be used for both ARFIMA and SARFIMA models, but the last one only for ARFIMA models. The aim is to estimate the parameter vector $\lambda = (\phi_1, ..., \phi_p, d, \theta_1, ..., \theta_q)'$ or $\lambda_s = (\phi_1, ..., \phi_p, d, \theta_1, ..., \theta_q, \Phi_1 \ , ..., \Phi_{p_s}, d_s, \ \Theta_1, ..., \Theta_{q_s})'$, if the process is an ARFIMA or a SARFIMA model, respectively.

The most common estimator for λ in the literature is the Whittle estimator (Whittle, 1951, Fox and Taqqu, 1986). The Whittle estimator is based on minimizing an approximation of the log-likelihood function given by

$$l_w(\lambda, \sigma^2) = \sum_{j=1}^{\lfloor T/2 \rfloor} \left[\log f_{\lambda,\sigma^2}(\omega_j) + \frac{I(\omega_j)}{f_{\lambda,\sigma^2}(\omega_j)} \right] \tag{3.8}$$

where $\omega_j = 2\pi j/T$, $j = 1, 2, ..., \lfloor T/2 \rfloor$, are the Fourier frequencies, and f_{λ,σ^2} is the spectral function given λ and σ^2. Numeric procedures are necessary to find the values of λ and σ^2 that minimize (3.8).

The asymptotic distribution of both the exact maximum likelihood and the Whittle ($\hat{\lambda}_W$) estimators are the same. Let λ_i, $i = 1, ..., k$ be the elements of the k-sized parameter vector λ, then $\sqrt{T}(\hat{\lambda}_W - \lambda) \xrightarrow{D} N(0, V^{-1})$, where V is a matrix with elements V_{ij} given by

$$V_{ij} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\frac{\partial \log f_{\lambda,\sigma^2}(\omega)}{\partial \lambda_i} \right] \left[\frac{\partial \log f_{\lambda,\sigma^2}(\omega)}{\partial \lambda_j} \right] d\omega.$$

For SARFIMA processes with $d_s > 0$, some of the Fourier frequencies may coincide with points in which the spectral function is undefined. One possible approach under these conditions is to restrict the sum in (3.8) to values where the spectral function is well defined. The Whittle estimator combines relatively little computational complexity and good accuracy (Palma, 2007, Rea et al, 2013).

The minimum distance estimator (MDE) for ARFIMA processes was proposed by Tieslau et al. (1996). The idea is to minimize the difference between the theoretical autocorrelations and the sample autocorrelations. Define $\hat{\varrho}$ as a vector of sample autocorrelations and $\varrho(\lambda)$ as the vector of corresponding theoretical autocorrelations, given the parameter vector λ. The minimum distance estimator is the one that minimizes

$$S(\lambda) = [\hat{\varrho} - \varrho(\lambda)]' W [\hat{\varrho} - \varrho(\lambda)], \tag{3.9}$$

where W, the weighting matrix, is a symmetric, positive-definite matrix. The asymptotically optimal W matrix is $W = C^{-1}$, where C is the asymptotic autocovariance matrix of the sample autocorrelations (Tieslau et al., 1996) whose elements are given in (3.7). It should be noted, though, that if the parameters are unknown, so is C. Tieslau et al. (1996) show that for $d \in (-0.5, 0.25)$,

$$\sqrt{T}(\hat{\lambda}_{\text{mde}} - \lambda) \xrightarrow{D} N(0, (D'WD)^{-1} D'WCWD(D'WD)^{-1})$$

when $T \to \infty$, where D is the matrix of derivatives of $\rho(\lambda)$ with respect to λ. The inverval $(-0.5, 0.25)$ is the inverval in which $\hat{\rho}_k$ has an asymptotic variance that decays as T^{-1}. For $d = 0.25$, Hosking (1996) shows that the asymptotic variance of $\hat{\rho}_k$ decays as $T^{-1}\log T$, while for $d \in (0.25, 0.5)$, $\hat{\rho}_k$ decays as $T^{2(1-2d)}$. Tieslau does not calculate the asymptotic variance of $\hat{\lambda}$

in these cases, but it is reasonable to conjecture that it will be similar to the asymptotic variances of $\hat{\rho}_k$ also for $d \geq 0.25$.

If $W = C^1$ the variance matrix of the asymptotic distribution simplifies to $D'C^{-1}D$. Meanwhile if W is the identity matrix, the variance matrix of the asymptotic distribution becomes $(D'D)^{-1}D'CD(D'D)^{-1}$. Tieslau et al. calculated the asymptotic variance of the MDE estimator using $W = C^{-1}$ using autocorrelations in the lag $1, ..., k$. Higher efficiency was obtained the higher was the value of k.

Zevallos and Palma (2013) try to overcome the problem of the limiting distribution for $d \in [0.25, 0.5)$ by applying a fractional filtering. Instead of estimating the autocorrelations for the original series, they use the sample autocorrelations of the filtered series

$$Y_t = (1 - B)^{1/2}(X_t - \bar{X}), \quad t \in \mathbb{Z}, \tag{3.10}$$

which will be approximately an ARFIMA$(p, d - 0.5, q)$ process, if $\{X_t\}$ is an ARFIMA(p, d, q) process. This estimator was called the MDEFF estimator by the authors. The asymptotic distribution of $\hat{\lambda}_{\text{mdeff}}$ for an ARFIMA(p, d, q) process is the same one of an ARFIMA$(p, d - 0.5, q)$ process for the MDE.

The weighting matrix W in the MDE and the MDEFF is formally defined as a fixed constant. Nevertheless, there is intuitive appeal to use a weighting matrix that depends on the parameter vector: $W(\lambda)$. Considering that the asymptotic optimal is $W = C^{-1}$, one could consider definying $W(\lambda) = C(\lambda)^{-1}$, where $C(\lambda)$ is the matrix of asymptotic variances of ϱ. It is important to note, though, that the proofs in the literature about consistence and asymptotic distribution of the estimators are for fixed W.

So far, minimum distance estimators have not been used to estimate the parameters of SARFIMA processes. In the case of the MDE, the extension is straightforward. The MDE estimator for SARFIMA may be defined in a way such that it minimizes an equation like Equation (3.9), replacing λ by λ_s. But there is not a straightforward intuitive generalization of the MDEFF for SARFIMA processes. The MDEFF filters the original series in order to generate a new series without long memory. Applying a filter like the one in (3.10) in a SARFIMA process will not necessarily generate a series without the long memory property. Evidently, a generalization of the MDEFF for SARFIMA processes can be thought about, but this is not in the main scope of this work.

Chapter 4

Bias Corrected Minimum Distance Estimator

This chapter focus on the main proposal of this work, the bias corrected minimum distance estimator (BCMDE). We begin by discussing the behaviour of the bias of autocovariance and autocorrelation estimators. Minimum distance estimators depend on estimators of the autocorrelation, therefore, their bias are likely to cause a negative impact on the MDE. This is the main motivation of this work. The MDEFF is not much affected by this problem because the filtered series does not have the property of long memory. Filtering a series generates a series which follows an ARFIMA model with d equal to the original series minus 0.5. If the original series is stationary, the filtered series will have a value of d not greater than 0. After the problems with the sample autocorrelation and autocovariance are analysed, this chapter closes defining a minimum distance estimator that tries to take this bias into account.

4.1 Empiral analysis of the bias in the sample autocovariance and autocorrelation functions

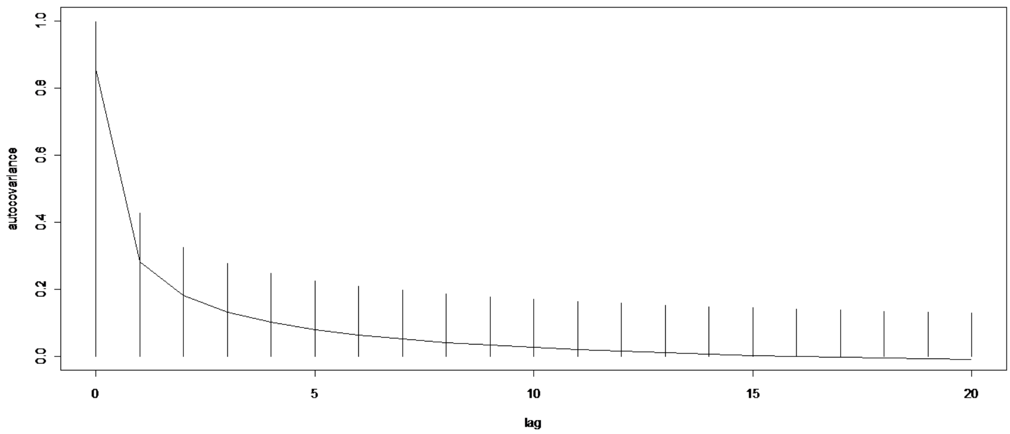
In Chapter 2 we derived the exact values of the bias of the sample autocovariance function as a function of the values of the autocovariance function at the lags $0, ..., T-1$. Calculation for the bias of the sample autocorrelation is far more complicated. In this section we provide some examples of how the sample autocovariance function can be affected by the bias in its estimation through a comparison with the theoretical function. In the case of the autocorrelation function, we discuss an approximation for calculating the expectation of its estimator.

Estimation of the autocovariance, autocorrelation and spectral functions are important not only because they help to identify the correct model, but also because they are frequently used in the estimation of model parameters. Therefore, it is essential to understand their behaviour and how an under (or over) estimation of these functions may affect parameter estimation.

In what follows, we will perform some empirical examples using the ARFIMA process. Regarding autocovariance and periodogram functions, we compare the behavior of the expectation of their estimators with respect to the theoretical autocovariance and spectral function, respectively. Concerning the estimator of the autocorrelation function, as its expectation is difficult to calculate, we make some simulations to assess the behavior of an approximation for the expectation of the sample autocorrelation.

Figure 4.1 shows, for an ARFIMA$(0, 0.3, 0)$ process with $T = 100$, the difference in the behavior of the expectation of the autocovariance estimator $\hat{\gamma}_k$ compared to the theoretical autocovariance function. By analysing Figure 4.1, some interesting conclusions can be reached. The first one is that all lags of the sample autocovariance in the figure are different from its expectation. The second one is that the expectation of the sample autocovariance appears to behave more like the autocovariance functions of short memory process,

Figure 4.1: Autocovariance function (bars) and expectation (full line) of the estimator $\hat{\gamma}_k$ for ARFIMA(0,0.3,0) with $T = 100$.

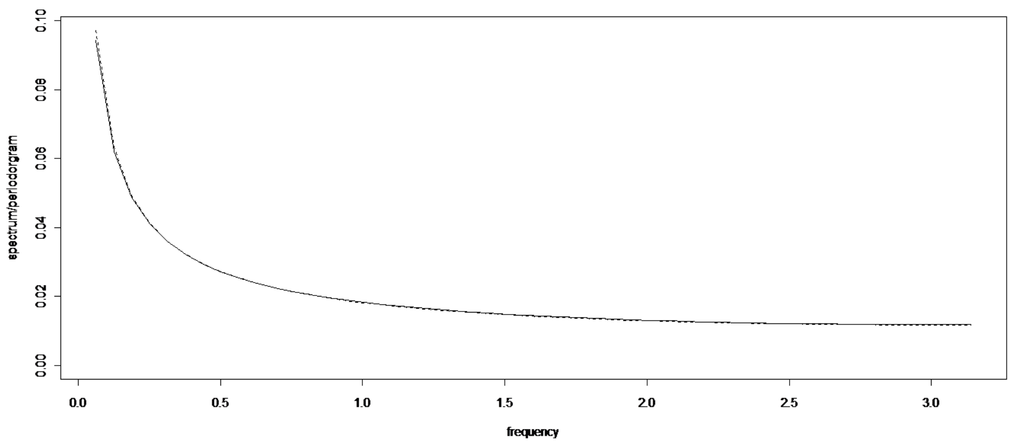


possessing a rapid decay.

As a function of sample autocovariances, which may be strongly biased, it could be thought that the same problem would affect the periodogram. But empirical analysis indicates this is not a problem. The bias affects each frequency in a way such that the expectation of the periodogram tends to have a shape similar to the spectrum, as it can be seen in Figure 4.2 for an ARFIMA$(0, 0.3, 0)$ with $T = 100$.

For any estimator of the autocovariance function, the autocorrelation at lag k can be estimated as $\hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$. A derivation of the bias for the sample autocorrelations is more complicated, and it is likely distribution-dependent, as $\hat{\gamma}_0$ enters in the denominator of $\hat{\rho}_k$. One approach is to approximate $E(\hat{\rho}_k)$ by $E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$. This approach makes sense asymptotically, as $E(\hat{\gamma}_k) \to \gamma_k$ and $E(\hat{\gamma}_0) \to \gamma_0$ and therefore $E(\hat{\gamma}_k)/E(\hat{\gamma}_0) \to \rho_k$. To verify if it works reasonably well also for small samples, a Monte Carlo with 1000 replications study was performed for ARFIMA(0,d,0) processes with $T = 100$. Table 4.1 shows the results of this study for $d = 0.2, 0.3, 0.4$ and $k = 1, 2, 3$. In this simulation we compared the approximation $E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$ with $\widehat{E(\hat{\rho}_k)}$, which is the mean in the simulation of the sample autocorrelations (an em-

Figure 4.2: Standardized spectrum (solid line) and expectation of the periodogram (dashed line) for ARFIMA(0,0.3,0) with $T = 100$. The lines are indistinguishable at naked eye.



pirical estimation of $E(\hat{\rho}_k)$). From the results it seems the approximation is adequate.

With the results on Table 1 in mind, we can further investigate the approximation $E(\hat{\rho}_k) \approx E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$. The exact value of $E(\hat{\rho}_k)$ is given by

$$E(\hat{\rho}_k) = E(\hat{\gamma}_k)E(\hat{\gamma}_0^{-1}) + \text{Cov}(\hat{\gamma}_k, \hat{\gamma}_0^{-1}).$$

The quantities $E(\hat{\gamma}_0^{-1})$ and $\text{Cov}(\hat{\gamma}_k, \hat{\gamma}_0^{-1})$ are both unknown. It is possible to say, though, that $E(\hat{\gamma}_0^{-1}) \geq E(\hat{\gamma}_0)^{-1}$, and therefore $E(\hat{\gamma}_k)E(\hat{\gamma}_0^{-1}) \geq E(\hat{\gamma}_k)E(\hat{\gamma}_0)^{-1}$. This is because $\hat{\gamma}_0$ is a positive random variable and $f(x) = 1/x$ is a convex function on $\mathbb{R}^+$ (Jensen's inequality). But in all simulations we encounter $\widehat{E(\hat{\rho}_k)} > E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$. Therefore, this underestimation of the bias caused by the approximation (and overestimation of the mean) in these simulations is a resulf of the negative autocovariance between $\hat{\gamma}_k$ and $\hat{\gamma}_0^{-1}$.

We now proceed to compare theoretical autocorrelations with the approximation for the expectation of the sample autocorrelations seen in the last paragraph. This can be seen in Figure 4.3 for an ARFIMA$(0, 0.3, 0)$ process with $T = 100$. We can see observations similar to those that were

Table 4.1: Monte Carlo simulation for ARFIMA$(0, d, 0)$ with $d = 0.2, 0.3, 0.4$ and $T = 100$. $E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$ are the ratios of the means of autocovariance estimators, and $\widehat{E(\hat{\rho}_k)}$ is the estimated mean of the autocorrelation estimator.

		ρ_1	ρ_2	ρ_3
$d = 0.2$	$E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$	0.2108	0.1231	0.0855
	$\widehat{E(\hat{\rho}_k)}$	0.1982	0.1115	0.0762
$d = 0.3$	$E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$	0.3374	0.2205	0.1642
	$\widehat{E(\hat{\rho}_k)}$	0.3205	0.2062	0.1496
$d = 0.4$	$E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$	0.4727	0.3408	0.2698
	$\widehat{E(\hat{\rho}_k)}$	0.4511	0.3156	0.2452

made for Figure 4.1: strong bias and behaviour similar to short memory processes. Naturally there are caveats about the fact that we are dealing with approximations in Figure 4.3.

We now show empirically in Figure 4.4 how the use of $\hat{\gamma}_k$ or $\tilde{\gamma}_k$ may affect the periodogram. Using $\hat{\gamma}_k$ as an estimator for the autocovariance instead of $\tilde{\gamma}_k$ leads to less accuracy in the estimation of the spectrum. Figure 4.4 was built using 999 Monte Carlo simulated periodograms from an ARFIMA$(0, 0.3, 0)$ model.

4.2 The bias corrected MDE

As shown in Sections 2.1, 2.2 and 4.1, the autocovariance estimators are biased when the mean must be estimated. The sample autocorrelations also show strong signs of bias, though they can not be calculated analytically. These biases can cause many negative effects on the MDE, specially in the case of small samples. The idea of the bias corrected MDE (BCMDE) is to minimize the distance, not between sample and theoretical autocorrelations, but between sample autocorrelation and one approximation of its expecta-

Figure 4.3: Autocorrelation function (bars) and approximation for the expectation (full line) of the estimator $\hat{\rho}_k$ for ARFIMA(0,0.3,0) with $T = 100$.

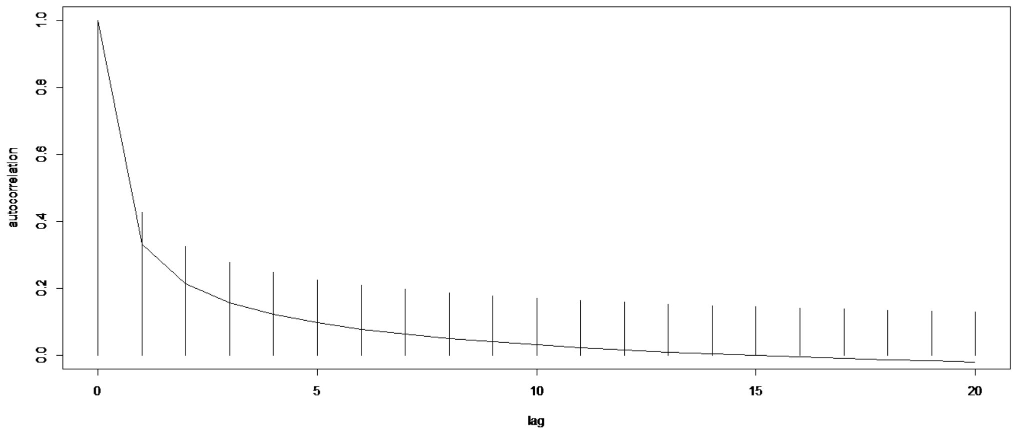
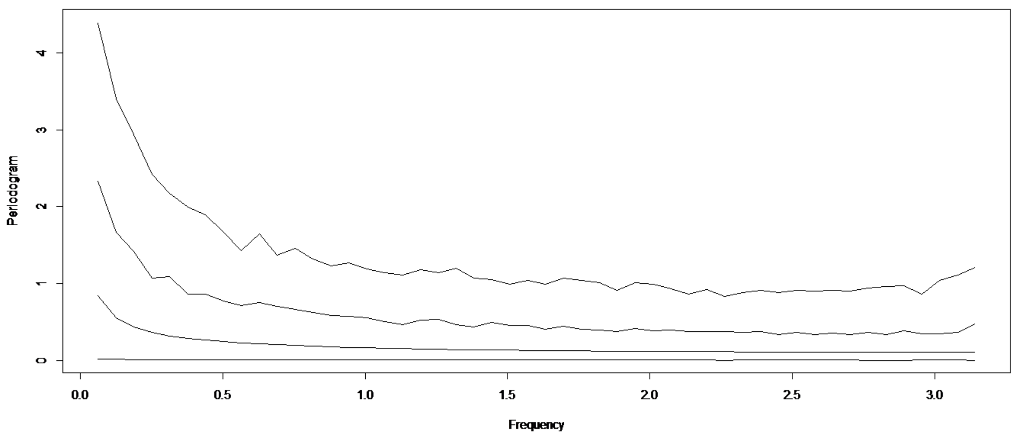


Figure 4.4: Empirical confidence intervals for the periodogram of a ARFIMA$(0, 0.3, 0)$ model. The lines from top to bottom are the upper bound of the 95% confidence interval using $\hat{\gamma}_k$, the upper bound of the 95% confidence interval using $\tilde{\gamma}_k$, the theoretical spectrum and the lower bounds of the confidence intervals (which are indistinguishable in this figure).



tion, using the approximation given in Section 4.1 of the sample autocorrelations given the parameters, that is $E(\hat{\rho}_k) \approx E(\hat{\gamma}_k)/E(\hat{\gamma}_0)$.

Let $\rho_{T,k}$ be the ratio between $E(\hat{\gamma}_k)$ and $E(\hat{\gamma}_0)$ whose formulas was given in Proposition 1. Thus $\rho_{T,k}$ can be written as

$$\rho_{T,k} = \frac{\rho_k + B^{\rho}_{T,k}}{1 + B^{\rho}_{T,0}}, \tag{4.1}$$

where $B^{\rho}_{T,k}$ is given in (2.5). Note that the equation above is written as function of autocorrelations instead of autocovariances. That is because γ_0 can be isolated either in the numerator and the denominator. Furthermore, ρ_0 is always equal to one. Let $\hat{\varrho}$ be a vector of sample autocorrelations and ϱ_T be the vector corresponding to $\rho_{T,k}$. The BCMDE minimizes

$$S(\lambda) = (\hat{\varrho} - \varrho_T(\lambda))' W (\hat{\varrho} - \varrho_T(\lambda)). \tag{4.2}$$

The weighting matrix W in (4.2) can be any symmetric positive definite matrix in order that $S(\lambda)$ becomes a measure of distance (not in the strict mathematical sense) between $\hat{\varrho}$ and ϱ_T. It can be the matrix of asymptotic covariances of the sample autocorrelations or the identity matrix (specially in cases where the asymptotic covariances are unknown). If only one lag of autocorrelation is being used, the choice of W, a single number in this case, is irrelevant. What minimizes $(\hat{\varrho} - \varrho_T(\lambda))^2/w_1$, minimizes $(\hat{\varrho} - \varrho_T(\lambda))^2/w_1$, $w_1, w_2 \in \mathbb{R}$. We will see that in the BCMDE the choice of W impacts in the variance of the asymptotic distribution, as it was the case for the MDE. What would be a choice of W that reduces uniformly the asymptotic variance, if it even exists, could be a subject of future research.

An obvious question regarding the vector of sample autocorrelation, $\hat{\varrho}$, is which lags should be chosen to be part of $\hat{\varrho}$. For an ARFIMA model, the most intuitive choice for the lags in the vectors $\hat{\varrho}$ and ϱ is $1, ..., m$, $m \in \mathbb{N}$. Tieslau et al. (1996) showed that for an ARFIMA$(0, d, 0)$ process this choice reduces the asymptotic variance of the MDE estimator compared to the choice of the lags $k, ..., k+m$, for any $k \geq 2$. For a SARFIMA model it might be a better idea to include the first lags of order multiple of the seasonal period.

We will proceed now to prove the weak consistency of the BCMDE estimator. For the proof to be valid, it is necessary first to establish the following proposition:

Proposition 6: Let $\{X_t\}$ be an ARFIMA(p,d,q) process, $d \in (-1, 0.5)$, in a compact parametric space with constant mean in which the mean is estimated as $\bar{X} = \sum_{t=1}^{T} X_t/T$. Assume that the autoregressive parameters are such that $\sum_{i=1}^{p} |\phi_i| \leq K < 1$, where K is the maximum value in the parametric space of the sum of the absolute value of the autoregressive parameters. Then $\rho_{T,k}$ defined in (4.1) converges uniformly to ρ_k as $T \to \infty$. *Proof in Appendix F.*

Proposition 6 combined with the following theorem proves the weak consistency of the BCMDE in the case of ARFIMA processes with constant mean.

Theorem 1: Let $\{X_t\}$ be an ARFIMA(p,d,q) process, $d \in (-1, 0.5)$, for which $\rho_{T,k}$ converges uniformly to ρ_k in a compact parametric space Λ. In addition, let the vector of theoretical autocovariances $\varrho = (\rho_{k_1}, ..., \rho_{k_m})$, $k_1, ..., k_m \in \mathbb{N}$, be such that $\varrho : \Lambda \to \mathbb{R}^m$ is injective. Then, $\hat{\lambda}$ converges in probability to λ_0, the real parameter values, as $T \to \infty$. *Proof in Appendix G.*

In some cases it is easy to check and guarantee that ϱ is injective. For example, in the case of ARFIMA$(0,d,0)$ processes, the first theoretical autocorrelation is monotonous as a function of d. Therefore, if the first lag of the autocorrelation is present, ϱ is injective. It is even more trivial to guarantee the injectivity of AR(1) or MA(1) processes.

The following result establishes the asymptotic distribution of the BCMDE in the case of an ARFIMA$(0,d,0)$ process.

Theorem 2: Let Λ be a compact parametric space. Let $\varrho = (\rho_{k_1}, ..., \rho_{k_m})$, $k_1, ..., k_m \in \mathbb{N}$, be such that $\varrho : \Lambda \to \mathbb{R}^m$ is injective.If $\{X_t\}$ is an ARFIMA$(0,d,0)$ with $d < 0.25$ and constant mean, then as $T \to \infty$,

$$\sqrt{T}(\hat{\lambda} - \lambda) \xrightarrow{D} N(0, (D'WD)^{-1}D'WCWD(D'WD)^{-1}),$$

where D is the matrix of derivatives of ϱ with respect to the parameters.

Proof in Appendix H.

Theorem 2 is also valid for AR(1) and MA(1) processes. Nevertheless, it is worth mentioning some particularities. The bias of the sample autocovariance of both the AR(1) and the MA(1) processes decay faster than $\sqrt{T}$. The derivative of the autocorrelation function of the AR(1) process is given by $\rho_k'(\phi) = k\phi^{k-1}$. In a compact parametric space, $\rho_k'(\phi)$ clearly converges uniformly to zero. The derivative of the autocorrelation function of the MA(1) process is zero after the first lag. Finally, as in the case of the ARFIMA$(0, d, 0)$ process, a non-zero derivative is guaranteed adding the first lag to ϱ.

Chapter 5

Monte Carlo simulations

In order to compare the small sample properties of the different estimators for the different models, Monte Carlo simulations were performed. When generating the Monte Carlo series a burn-in of size 1000 was used. In each instance, 1000 Monte Carlo replications were used.

5.1 ARMA model

In this section we compare the performance of the conditional sum of squares estimator (CSSE), maximum likelihood estimator (MLE), MDE and BCMDE for the ARMA model with constant and unknown mean. The simulations were performed for AR(1) and MA(1) models, with parameter values fixed at 0.4 and 0.8.

In each case, the sample sizes were $T = 25$ and $T = 100$ and the errors were generated from a standard normal distribution. In all models only the first sample autocorrelation was used in the BCMDE. For every estimator, mean, standard deviation and square root of the mean squared error (RMSE) were calculated.

Table 5.1 shows the results for the autoregressive model. The MDE was not used in this case as it is in this case almost identical to the CSS. The BCMDE has a significantly better performance particularly in terms of bias,

Table 5.1: Mean, standard deviation and RMSE for AR(1)

	CSS	MLE	BCMDE
$T = 25$, $\phi = 0.4$			
Mean	0.31011	0.31063	**0.37297**
SD	**0.19154**	0.19163	0.19358
RMSE	0.21149	0.21136	**0.19536**
$T = 100$, $\phi = 0.4$			
Mean	0.37975	0.37962	**0.39495**
SD	**0.08913**	0.08916	0.09012
RMSE	0.09136	0.09142	**0.09021**
	CSS	MLE	BCMDE
$T = 25$, $\phi = 0.8$			
Mean	0.66003	0.66549	**0.74346**
SD	0.16515	**0.16135**	0.18487
RMSE	0.21643	0.21000	**0.19324**
$T = 100$, $\phi = 0.8$			
Mean	0.76736	0.76854	**0.78625**
SD	0.06842	**0.06697**	0.07047
RMSE	0.07578	0.07396	**0.07176**

Obs.: In bold are the means closest to the real value of the parameter and the smallest SD and RMSE

but also in terms of the RMSE. This is true for both parameter values and sample sizes used.

Table 5.2 shows the results for the moving average model. Contrary to the case of the AR(1) model seen in Table 5.1, the estimators ML and CSS are significantly better than the BCMDE both in terms of bias and RMSE, regardless the sample size and parameter value. The MDE does have smaller RMSE for $\theta = 0.4$ and $T = 25$, but this result is not replicated for other parameter values and sample sizes.

The contrast between the performance of the estimators in the case AR(1)

Table 5.2: Mean, standard deviation and RMSE for MA(1)

	CSSE	MLE	MDE	BCMDE
$T = 25$, $\theta = 0.4$				
Mean	**0.39558**	0.39105	0.27558	0.43643
SD	0.27751	0.25771	**0.16770**	0.30192
RMSE	0.27741	0.25774	**0.20874**	0.30396
$T = 100$, $\theta = 0.4$				
Mean	0.38980	0.39096	0.38888	**0.40870**
SD	**0.10107**	0.10231	0.15907	0.16647
RMSE	**0.10153**	0.10266	0.15938	0.16661

	CSSE	MLE	MDE	BCMDE
$T = 25$, $\theta = 0.8$				
Mean	0.81376	**0.80827**	0.62133	0.69885
SD	0.26854	**0.19653**	0.31137	0.29998
RMSE	0.26876	**0.19661**	0.35886	0.31643
$T = 100$, $\theta = 0.8$				
Mean	**0.79335**	0.80902	0.75738	0.78187
SD	**0.07597**	0.07646	0.23142	0.22398
RMSE	**0.07622**	0.07695	0.23520	0.22460

Obs.: In bold are the means closest to the real value of the parameter and the smallest SD and RMSE and in the case MA(1) deserves some comments. The bias of the autocovariance estimators are linear combinations of the autocovariance function. In the MA(1) model, the autocovariance function is zero for lags greater than one, causing the bias of sample autocovariances to be irrelevant. Furthermore, sample autocovariances and autocorrelations seem simply not to be good identifiers of the MA(1) model.

5.2 ARFIMA model with constant mean

In this section we compare the performance of Whittle, MDE, MDEFF and BCMDE estimators for the constant mean ARFIMA model, through Monte Carlo simulations. Although in Section 4.2 we have only proved the consistency of BCMDE for the ARFIMA(0,d,0), will also implement some simulations for ARFIMA(1,d,0) and SARFIMA models, in order to verify the empirical behavior of the BCMDE in these cases. The simulations were performed for an ARFIMA$(0, 0.3, 0)$ model with known and unknown mean, ARFIMA$(1, 0.3, 0)$ models with $\phi = 0.3$, SARFIMA$(0, 0.3, 0) \times (0, 0.1, 0)_{12}$ and SARFIMA$(0, 0.1, 0) \times (0, 0.3, 0)_{12}$ with unknown mean.

In each case, the sample sizes were $T = 100$ and $T = 500$ and the errors were generated from a standard normal distribution. For every estimator, mean, standard deviation and square root of the mean squared error (RMSE) were calculated.

The identity matrix was used as the weighting matrix for the MDE, the BCMDE and the MDEFF. We also ran simulations for the MDEFF with a non fixed weighting matrix, $W(\lambda)$, $W(\lambda)$ being the inverse of the asymptotic distribution of ϱ given the parameters. The MDEFF estimated this way will be called MDEFF* in this section. For the ARFIMA$(0, d, 0)$ model, only the first sample autocorrelations were used for the MDE, MDEFF (fixed W) and the BCMDE. For the MDEFF* the first ten sample autocorrelations were used. These choices on the number of sample autocorrelations were based on the results of preliminary simulations. For the ARFIMA$(1, d, 0)$ model, the first and second sample autocorrelation were used for all the minimum distance estimators. For the SARFIMA$(0, d, 0) \times (0, d_{12}, 0)_{12}$ model, the first and the thirteenth sample autocorrelations.

Table 5.3 shows the result of the simulations for the ARFIMA$(0, d, 0)$ model. When the mean is known, the MDE and BCMDE are identical due to the sample autocovariance being unbiased. Even though the MDE is slightly more biased than the other estimators, it compensates for it by being more precise, with a RMSE 22.5% smaller than the one of the Whittle estimator and 17.5% smaller than the one of the MDEFF* for $T = 100$. For $T = 500$, the RMSE for the MDE is still the smallest one among the assessed estimators, but the gap towards the other estimators shortens. The MDEFF presents the largest RMSE.

Comparing the case of an ARFIMA$(0, 0.3, 0)$ model with known and unknown mean, still in Table 5.3, it is possible to see that the knowledge of the mean poses little effect on the Whittle and MDEFF estimators. The MDE, on the other hand, is heavily affected by it. When the mean is unknown, the MDE is much more biased than the remaining estimators. The BCDME is not affected that much by bias when the mean is unknown, but its RMSE is grater than the RMSE of the MDE with known mean, reflecting the difficulty of estimating the autocovariance function under this condition. Notwithstanding, the BCMDE presents the smallest RMSE for $T = 100$ and $T = 500$. The MDEFF has very small bias, for known or unknown mean, but a very large standard deviation, resulting in higher values of RMSE. The MDEFF* also has very small bias and its RMSE is much smaller than the one of the MDEFF, though not beating the BCMDE in this aspect.

In the simulation of Table 5.3 with known mean and $T = 100$, we measure the times taken by each estimator. The average estimation time in seconds for each estimator was 0.00392 for the Whittle, 0.00017 for the MDE, 0.00687 for the BCMDE, 0.00123 for the MDEFF and 0.08314 for the MDEFF*.

Table 5.4 shows the results of the simulations for $ARFIMA(0, 0.3, 0)$, mean known and $T = 2500$. Even for this large sample size, the BCMDE

Table 5.3: Mean, standard deviation and RMSE for ARFIMA$(0, 0.3, 0)$

	Whittle	MDE	MDEFF	MDEFF*	BCMDE
T=100, known μ					
Mean	0.29305	0.27954	**0.29370**	0.31093	0.27954
SD	0.09363	**0.06982**	0.13503	0.08760	**0.06982**
RMSE	0.09389	**0.07275**	0.13517	0.08827	**0.07275**
T=500, known μ					
Mean	0.29901	0.29094	0.29820	**0.30133**	0.29094
SD	0.03876	**0.03445**	0.05918	0.03957	**0.03445**
RMSE	0.03877	**0.03562**	0.05921	0.03959	**0.03562**
	Whittle	MDE	MDEFF	MDEFF*	BCMDE
T=100, unknown μ					
Mean	0.29261	0.23825	0.29264	**0.30251**	0.29007
SD	0.09319	**0.06773**	0.13365	0.09064	0.08822
RMSE	0.09348	0.09165	0.13384	0.09067	**0.08877**
T=500, unknown μ					
Mean	**0.29942**	0.27346	0.29935	0.30124	0.29735
SD	0.03807	**0.03042**	0.05898	0.03920	0.03689
RMSE	0.03807	0.04037	0.05898	0.03922	**0.03698**

Table 5.4: Mean, standard deviation and RMSE for ARFIMA$(0, 0.3, 0)$, $T = 2500$

	Whittle	MDE	MDEFF	BCMDE
Mean	**0.29967**	0.28683	0.29982	0.29852
SD	0.01648	**0.01436**	0.02665	0.01631
RMSE	0.01648	0.01948	0.02664	**0.01637**

is still competitive with the Whittle in RMSE and bias. Remember that Whittle is an asymptotically efficient estimator and that we were not able to proof the asymptotic distribution of the BCMDE for $d > 0.25$.

Table 5.5 shows the results for the ARFIMA$(1, 0.3, 0)$. It is possible to see that the MDEFF presents less bias, except in the case of the estimation of ϕ for $T = 100$, when the BCMDE presented the smaller bias. The MDE again showed the largest bias for both parameters. The behavior of the bias of the Whittle estimator and the MDE are very similar, with a tendency towards underestimating d and overestimating ϕ. In the case of the estimation of ϕ for $T = 100$, though, the bias of the BCMDE was significantly smaller. The BCMDE is, nevertheless, more precise in terms of RMSE than the Whittle estimator and that is particularly clear for $T = 100$. Besides having a bigger bias, the BCMDE is also more precise than the MDEFF.

Table 5.6 shows the results for the SARFIMA$(0, d, 0) \times (0, d_s, 0)$ model. Some observations are similar to those made for ARFIMA models. The MDE is again heavily biased, for both d and d_{12}, regardless their true values. With respect to the estimation of d, the BCMDE is again slightly more biased than the Whittle estimator. In terms of RMSE, though, the BCMDE outperforms the Whittle estimator for $d = 0.3$, as it happened in the ARFIMA(0,0.3,0) model. The Whittle behaves better compared to the BCMDE when d is smaller. For $d = 0.1$, the Whittle estimator is better in terms of bias and RMSE. Some similar observations can be made regarding the estimation of d_{12}. In particular, the Whittle estimator tends to behave comparatively

Table 5.5: Mean, standard deviation and RMSE for ARFIMA$(1, d, 0)$.

	Whittle	MDE	MDEFF	BCMDE
$T = 100$		$d = 0.3$		
Mean	0.21462	0.13020	**0.29737**	0.22770
SD	0.25472	**0.15545**	0.21848	0.19680
RMSE	0.26853	0.23016	0.21839	**0.20957**
		$\phi = 0.3$		
Mean	0.35427	0.39579	0.26823	**0.32358**
SD	0.24946	**0.17899**	0.20822	0.20243
RMSE	0.25517	**0.20293**	0.21052	0.20370
	Whittle	MDE	MDEFF	BCMDE
$T = 500$		$d = 0.3$		
Mean	0.27497	0.22617	**0.29632**	0.27223
SD	0.12231	**0.09416**	0.15569	0.10948
RMSE	0.12478	0.11962	0.15566	**0.11289**
		$\phi = 0.3$		
Mean	0.32084	0.35553	**0.29209**	0.31948
SD	0.13197	**0.12124**	0.15037	0.13113
RMSE	0.13354	0.13330	0.15050	**0.13250**

better for smaller values of d_{12}.

5.3 ARFIMA model with non-constant mean

In this section we compare the performance of the Whittle estimator, MDE, MDEFF and BCMDE for ARFIMA models with non-constant mean through Monte Carlo simulations. Once again, we aim here to verify the empirical behavior of the BCMDE, as the proofs of consistency were only made for models with constant mean. The models with non-constant mean are those described in Section 2.2: structural break, simple linear regression and non-stochastic seasonality. The simulations were performed for the ARFIMA$(0, d, 0)$ model with structural break, linear regression and non-stochastic seasonality.

For each model, except in the case of structural break, where $T = 100$, the sample sizes were $T = 100$ and $T = 500$ and the errors were generated from a standard normal distribution. Only the first autocorrelation was used for the MDE, BCMDE and MDEFF. For every estimator, mean, standard deviation and square root of the mean squared error (RMSE) were calculated.

In the case of the structural break model, the break was set up at $T_0 = 40$. In the case of the linear regression model, time was the independent variable, and the slope coefficient was equal to one. In the case of the non-stochastic seasonality model, the period $s = 12$ was used.

Tables 5.7 shows the results for structural break. In the presence of structural break, the Whittle estimator becomes significantly more biased when compared to a model with constant mean. In this case, not only the BCMDE is less biased but also it has much smaller RMSE for the two values of d.

Table 5.8 show the results for the model in which the mean is a linear function of time for $d = 0.2, 0.4$ and $T = 100, 500$. Once again, the Whittle estimator is severely affected by the estimation of the mean, specially when $T = 100$. The bias is smaller when $T = 500$, but still bigger than the bias

Table 5.6: Mean, standard deviation and RMSE for SARFIMA$(0, d, 0) \times (0, d_{12}, 0)_{12}$,

	Whittle	MDE	BCMDE	W	MDE	BCMDE
T=100		d=0.3			$d_{12} = 0.1$	
Mean	**0.29382**	0.23359	0.28093	**0.09199**	-0.00385	0.07297
SD	0.09612	**0.07991**	0.08665	0.11915	**0.10967**	0.11283
RMSE	0.09627	0.10388	**0.08868**	0.11935	0.15100	**0.11597**
T=500						
Mean	**0.30167**	0.26860	0.29687	**0.09816**	0.05044	0.09167
SD	0.03975	**0.03327**	0.03786	**0.04043**	0.04777	0.05131
RMSE	0.03976	0.04573	**0.03797**	**0.04045**	0.06882	0.05196
	Whittle	MDE	BCMDE	W	MDE	BCMDE
T=100		d=0.1			$d_{12} = 0.3$	
Mean	**0.08666**	0.03901	0.08288	0.34628	0.23069	**0.27195**
SD	0.09970	**0.09645**	0.11241	0.12521	**0.06489**	0.06494
RMSE	**0.10054**	0.11407	0.11365	0.13343	0.09492	**0.07071**
T=500						
Mean	0.09739	**0.06660**	0.09538	0.31525	0.26903	**0.28433**
SD	**0.03877**	0.04761	0.05653	0.04324	0.03086	**0.03046**
RMSE	**0.03883**	0.05814	0.05670	0.04583	0.04370	**0.03424**

Table 5.7: Structural break, $T = 100$, $d = 0.2, 0.4$.

$d = 0.2$	Mean	RSME	SD
Whittle	0.15847	0.10801	0.09976
MDE	0.13767	0.10385	**0.08311**
MDEFF	**0.19720**	0.12994	0.12998
BCMDE	0.18743	**0.10305**	0.10233
$d = 0.4$	Mean	RSME	SD
Whittle	0.35532	0.10276	0.09259
MDE	0.28376	0.13198	**0.06253**
MDEFF	0.38210	0.12383	0.12259
BCMDE	**0.38407**	**0.08915**	0.08775

of the BCMDE or the MDEFF. In all cases the BCMDE has the smallest RMSE. Although the MDEFF is the less biased, the bias of the BCMDE is only slightly superior to the former.

Table 5.9 shows the results for the models with non-stochastic seasonality for $d = 0.2, 0.4$ and $T = 100, 500$. Contrary to the other models with non-constant mean, in this case the Whittle estimator is not so biased. Even for $T = 100$, the Whittle estimator tends to have a smaller RMSE compared to the BCMDE. For $T = 500$, Both the Whittle and the BCMDE have very similar performances, both in terms of bias and RMSE.

Table 5.8: Linear regression, $T = 100, 500$, $d = 0.2, 0.4$.

$T = 100$, $d = 0.2$	Mean	RSME	SD
Whittle	0.16878	0.10210	0.09725
MDE	0.14336	**0.09879**	**0.08098**
MDEFF	**0.19954**	0.12992	0.12998
BCMDE	0.19358	0.09996	0.09981
$T = 500$, $d = 0.2$	Mean	RSME	SD
Whittle	0.19022	0.03993	0.03874
MDE	0.18041	0.04024	**0.03516**
MDEFF	**0.19751**	0.06686	0.06685
BCMDE	0.19750	**0.03954**	0.03948
$T = 100$, $d = 0.4$	Mean	RSME	SD
Whittle	0.34751	0.10622	0.09239
MDE	0.27668	0.13828	**0.06259**
MDEFF	**0.37896**	0.12271	0.12096
BCMDE	0.37291	**0.09187**	0.08783
$T = 500$, $d = 0.4$	Mean	RSME	SD
Whittle	0.38220	0.041694	0.03772
MDE	0.33488	0.07025	**0.02635**
MDEFF	**0.39369**	0.05377	0.05342
BCMDE	0.39357	**0.03893**	0.03840

Table 5.9: Non-stochastic seasonality, $T = 100, 500$, $d = 0.2, 0.4$.

$T = 100$, $d = 0.2$	Mean	RSME	SD
Whittle	0.21115	0.09880	0.09822
MDE	0.16674	**0.09038**	**0.08408**
MDEFF	**0.19856**	0.13337	0.13343
BCMDE	0.18673	0.09902	0.09817
$T = 500$, $d = 0.2$	Mean	RSME	SD
Whittle	0.20381	0.03686	0.03668
MDE	0.19073	**0.03498**	**0.03375**
MDEFF	0.20103	0.06712	0.06715
BCMDE	**0.19935**	0.03679	0.03681
$T = 100$, $d = 0.4$	Mean	RSME	SD
Whittle	**0.40428**	**0.08250**	0.08243
MDE	0.31507	0.10431	**0.06059**
MDEFF	0.39045	0.12783	0.12754
BCMDE	0.37980	0.08658	0.08423
$T = 500$, $d = 0.4$	Mean	RSME	SD
Whittle	0.40435	0.04158	0.04137
MDE	0.34985	0.05717	**0.02746**
MDEFF	**0.39910**	0.05274	0.05276
BCMDE	0.39508	**0.04000**	0.03971

Chapter 6

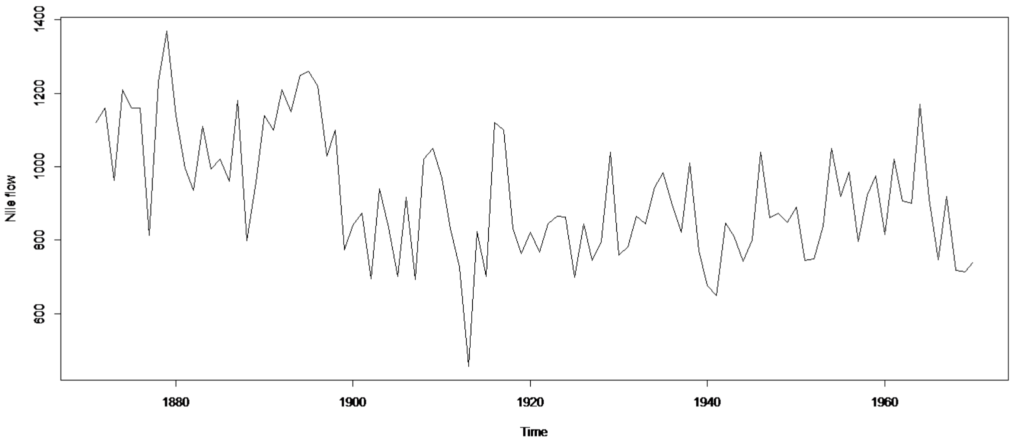
Application to a real time series

Figure 6.1 shows the annual flow discharge of the Nile river at Aswan between 1871 and 1970, measured in $10^8\ m^3$. This specific series is part of a larger set of Nile databases that were extensively analysed in the statistical literature. The series in this work can be accessed in newer versions of the R software by simply typing "Nile". It was originally used by Cobb (1978). In Cobb's work it is supposed, for the sake of giving a practical example to his theory, that there is a structural break around 1898 and that before and after that break, the annual flow follows independent normal distributions. Cobb ponders that the break could be the result of the construction of a dam, but he himself refutes that possibility because the break can also be observed in rainfall series in the tropical regions near the upper Nile.

The most common Nile series is the one of Nile overflow between 622 and 1470 that was famously studied by Hurst (1951, 1957). It is a seminal example of the presence of long memory in nature. In general, long memory models have been extensively used in the literature to model hydrological phenomena.

In this particular short times series that we will study in this work, an indisputable evidence of long memory which eliminates any reasonable doubt on contrary can not be presented. This is often the case for short time series, in part for the reasons we have already discussed here: problems of bias in

Figure 6.1: Annual flow of the River Nile at Ashwan between 1871 and 1970.



the estimation of the sample autocorrelation. But knowledge that this kind of phenomena do usually carry long memory together with lack of evidence on the contrary for this particular time series (as we will see in the next paragraphs) makes the choice of a long memory model adequate.

The sample autocorrelation function of this series with respect to the sample mean can be seen in Figure 6.2. Its periodogram is given in Figure 6.3. An analysis of Figures 6.2 and 6.3 suggests the existency of long memory in the series: The sample autocorrelation function has a slow decay and the periodogram peaks are in the lower frequencies, apparently tending to infinity as the frequency goes to zero.

We fitted an ARFIMA$(0, d, 0)$ model to the series from 1871 to 1960. Using the Whittle estimator we found $d = 0.3758$ while using BCMDE we found $d = 0.4216$. Figure 6.4 shows the sample autocorrelation of the residuals of the fitted model using the BCMDE and Figure 6.5 shows the periodogram of these residuals. Visual analysis of Figures 6.4 and 6.5 suggest that the chosen model was well fitted to the series. Similar conclusions can be reached using the Whittle estimator.

Figure 6.6 shows the predictions 10 steps ahead using the Whittle estimator and the BCMDE. As a result of estimating a higher value of d, the

Figure 6.2: Sample autocorrelation function of the Nile flow series.

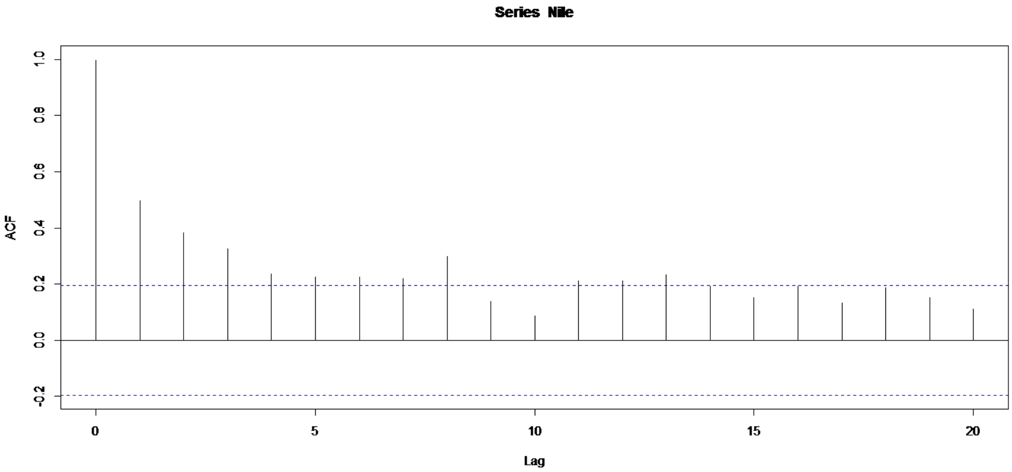


Figure 6.3: Periodogram of the Nile flow series.

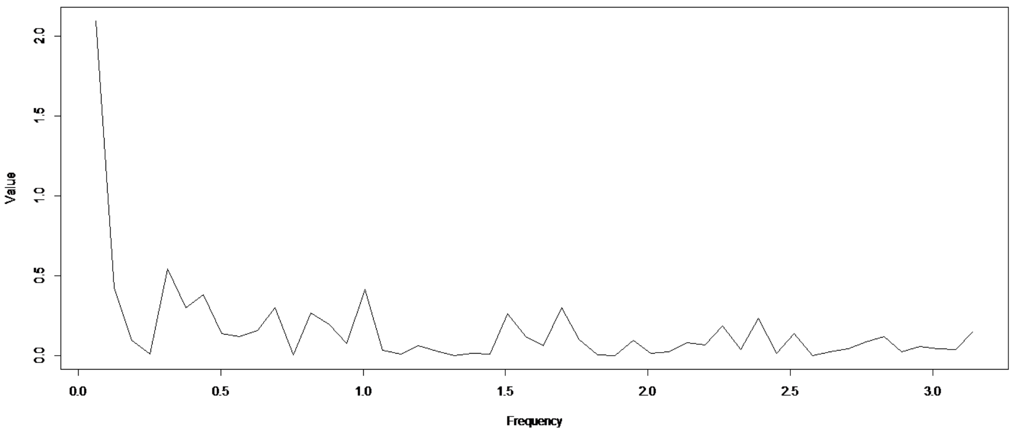


Figure 6.4: Autocorrelation function of the residuals of the fitted model.

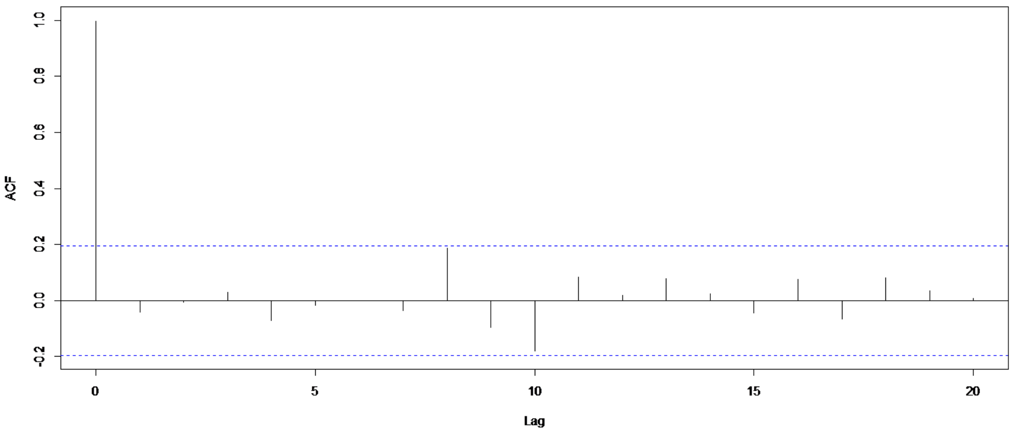


Figure 6.5: Periodogram of the residuals of the fitted model.

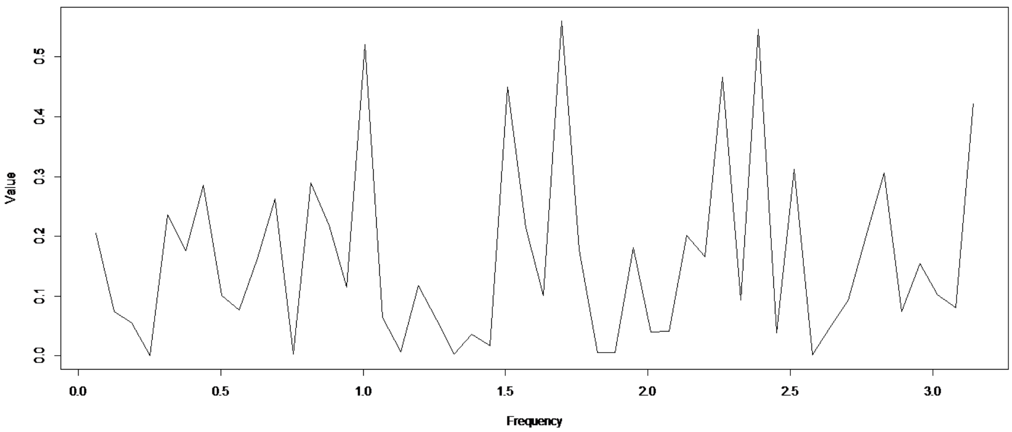
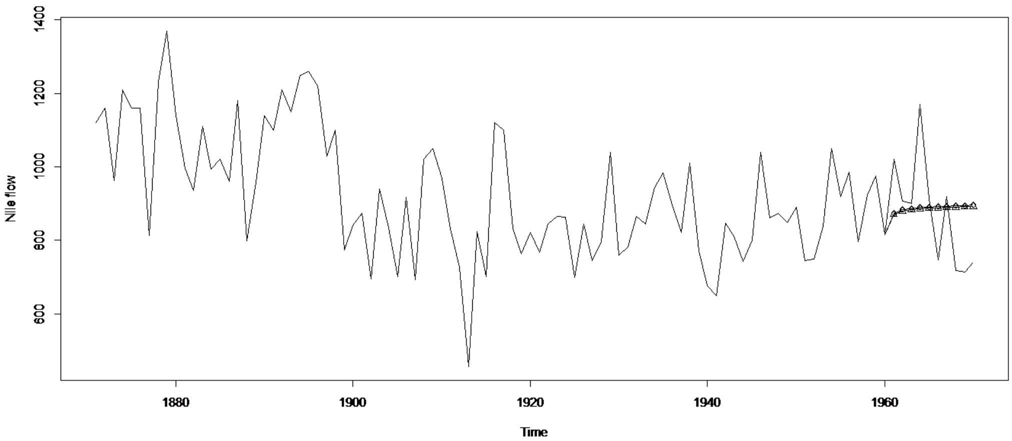


Figure 6.6: Prediction of the Nile flow series 10 steps ahead, using the Whittle estimator (circle) and the BCMDE (triangle).



predictions of the BCMDE are slightly further away from the sample mean than the predictions of the Whittle estimator. The squared prediction error of the BCMDE was 21025.3 while for the Whittle estimator it was 21079.7.

Chapter 7

Conclusion and future works

This work proposed a new estimator for short and long memory models, the BCMDE. This estimator belongs to the class of minimum distance estimators which are based on the sample autocorrelation function.

Previous minimum distance estimators in the literature find the parameter values that minimize the distance between the sample autocorrelations and the theoretical autocorrelations. A problem with this approach is that the expectation of the sample autocorrelations may differ substantially from the theoretical autocorrelations. We have shown the exact formula for the bias of the sample autocovariance in different scenarios: constant mean, structural break, regression and non-stochastic seasonality.

It has become clear that the bias caused by the necessity to estimate the mean can not be neglected. There is enough empirical and intuitive reasons to believe that bias also affect the sample autocorrelation, but the exact expectation of the sample autocorrelation is very difficult to be derived. The central idea of the BCMDE is to find the parameter values that minimize the distance between the sample autocorrelation and an approximation of its expectation. The approximation we have chosen is the ratio of the expectation of the sample autocovariance at lags k and 0. Simulation studies show that this is a good approximation, both in the sense of the BCMDE being a good estimator (according to criteria like the RMSE) as well as in the sense of comparing this approximation with the average of sample autocorrelations in a Monte Carlo study. This approximation can also be viewed as some kind of penalizing function.

We have proved the weak consistency of the BCMDE in the case of a constant mean and we have also derived its asymptotic distribution for the ARFIMA$(0, d, 0)$ ($d < 0.25$), AR(1) and MA(1) models. In these circumstances, both the BCMDE and the MDE have the same asymptotic distribution.

Similarly to the case of the MDE, the BCMDE is not more efficient than the Whittle estimator for the ARFIMA model (Tieslau et al. 1996). Nevertheless, simulation studies have been performed to evaluate the behavior of the BCMDE in small samples and the results were encouraging, as the BCMDE has beaten its competitors in autoregressive models and long memory models in terms of the RMSE. This performance was observed through simulations for AR(1), ARFIMA$(1, d, 0)$ and ARFIMA$(0, d, 0)$ models with constant mean plus ARFIMA$(0, d, 0)$ with structural break or with the mean as a linear function of time. It is interesting to note that in many of these cases, the BCMDE perform better even for samples of size 500. The BCMDE also reduced the bias (compared to the Whittle)in the simulations for AR(1) and ARFIMA$(0, d, 0)$ models with structural break or mean as a linear function of the time.

Future works encompass the search for a better approximation for the expectation of the sample autocorrelation and the proof of some asymptotic properties that were not covered in this work, such as the asymptotic distribution of the estimators when more than one parameter must be estimated. Some of these aymptotic properties due to technical difficulties, some of them due to time constraints. The rationale behind the BCMDE could also be expanded to realms outside the area of time series, for instance, comparing a sample statistic with its expectation, given the parameters. It would also be interesting to perform simulations to a wider variety of parameters and sample sizes.

Appendix A

Proof of Proposition 1

Taking expectation in (2.2) we have,

$$
\begin{aligned}
E(\hat{\gamma}_k) &= \frac{\sum_{i=1}^{T-k} E((X_i - \bar{X})(X_{i+k} - \bar{X}))}{T-k} \\
&= \frac{\sum_{i=1}^{T-k} E((X_i - \mu)(X_{i+k} - \mu))}{T-k} - \frac{\sum_{i=1}^{T-k} E((X_i - \mu)(\bar{X} - \mu))}{T-k} \\
&- \frac{\sum_{i=1}^{T-k} E((X_{i+k} - \mu)(\bar{X} - \mu))}{T-k} + E((\bar{X} - \mu)^2).
\end{aligned} \tag{A.1}
$$

The first term in the right hand side of(A.1) is the autocorrelation of lag k, γ_k. For the second and third terms we have that

$$
E((X_i - \mu)(\bar{X} - \mu)) = \frac{\sum_{j=1}^{T} E((X_i - \mu)(X_j - \mu))}{T} = \frac{\sum_{j=1}^{T} \gamma_{|i-j|}}{T}.
$$

For the fourth term,

$$
\begin{aligned}
E((\bar{X} - \mu)^2) &= E\left(\left(\frac{\sum_{j=1}^{T}(X_j - \mu)}{T}\right)^2\right) \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} E((X_i - \mu)(X_j - \mu))}{T^2} \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \gamma_{|i-j|}}{T^2} = \frac{T\gamma_0 + 2\sum_{i=1}^{T-1}(T-i)\gamma_i}{T^2}.
\end{aligned}
$$

Thus, using these equalities on Equation (A.1), we find that

$$\begin{aligned} E(\hat{\gamma}_k) = \gamma_k - \frac{\sum_{i=1}^{T-k}\sum_{j=1}^{T}\gamma_{|i-j|}}{T(T-k)} - \frac{\sum_{i=1}^{T-k}\sum_{j=1}^{T}\gamma_{|i+k-j|}}{T(T-k)} \\ + \frac{T\gamma_0 + 2\sum_{i=1}^{T-1}(T-i)\gamma_i}{T^2}, \end{aligned} \tag{A.2}$$

which is an equation with number of operations of order of magnitude T^2. Now note that $\sum_{i=1}^{T}\gamma_{|i-j|} = \sum_{i=1}^{T}\gamma_{|i-(T-j+1)|}$. As a result $\sum_{i=1}^{T-k}\sum_{j=1}^{T}\gamma_{|i-j|} = \sum_{i=1}^{T-k}\sum_{j=1}^{T}\gamma_{|i+k-j|}$. Therefore, Equation (A.2) becomes

$$E(\hat{\gamma}_k) = \gamma_k + \frac{T\gamma_0 + 2\sum_{i=1}^{T-1}(T-i)\gamma_i}{T^2} - 2\frac{\sum_{i=1}^{T-k}\sum_{j=1}^{T}\gamma_{|i-j|}}{T(T-k)}.$$

Besides $\sum_{i=1}^{T-k}\sum_{j=1}^{T}\gamma_{|i-j|} = \sum_{i=1}^{T}\sum_{j=1}^{T}\gamma_{|i-j|} - \sum_{i=1}^{k}\sum_{j=1}^{T}\gamma_{|i-j|}$. Therefore,

$$\begin{aligned} E(\hat{\gamma}_k) &= \gamma_k + \left(\frac{1}{T^2} - \frac{2}{T(T-k)}\right)\left(T\gamma_0 + 2\sum_{i=1}^{T-1}(T-i)\gamma_i\right) + 2\left[\frac{\sum_{i=1}^{k}\sum_{j=1}^{T}\gamma_{|i-j|}}{T(T-k)}\right] \\ &= \gamma_k - \frac{(T+k)}{(T-k)}\left[\frac{T\gamma_0 + \sum_{i=1}^{T-1}2(T-i)\gamma_i}{T^2}\right] + 2\left[\frac{\sum_{i=1}^{k}\sum_{j=1}^{T}\gamma_{|i-j|}}{T(T-k)}\right], \end{aligned}$$

which is an equation with number of operations of order of magnitude T. □

Appendix B

Proof of Proposition 2

Applying the expectation to the formula of $\hat{\gamma}_k$ in (2.6), we get

$$E(\hat{\gamma}_k) = \frac{\sum_{t=1}^{T-k}\big(E(X_tX_{t+k}) - E(X_t\hat{\mu}_{t+k}) - E(\hat{\mu}_tX_{t+k}) + E(\hat{\mu}_t\hat{\mu}_{t+k})\big)}{T-k}, \tag{B.1}$$

with $\hat{\mu}_t = \sum_{i=1}^{T_0} X_i/T_\alpha$ for $t \leq T_0$ or $\hat{\mu}_t = \sum_{i=T_0+1}^{T} X_i/T_\beta$ for $t > T_0$. The summation in (B.1) can be partitioned in three parts: $t = 1, ..., T_0 - k$, $t = T_0 - k + 1, ..., T_0$ and $t = T_0 + 1, ..., T - k$. These parts have size $T_\alpha - k$, k and $T_\beta - k$ respectively.

If $t = 1, ..., T_0 - k$, then:

$$E(X_tX_{t+k}) = \gamma_k + \alpha^2$$

$$\begin{aligned} E(X_t\hat{\mu}_{t+k}) &= E(X_t\hat{\alpha}) = E\Big(X_t\Big(\frac{\sum_{i=1}^{T_0} X_i}{T_\alpha}\Big)\Big) \\ &= \frac{\sum_{i=1}^{T_0}\gamma_{|t-i|}}{T_\alpha} + \alpha^2 \end{aligned}$$

$$E(\hat{\mu}_tX_{t+k}) = E(\hat{\alpha}X_{t+k}) = \frac{\sum_{i=1}^{T_0}\gamma_{|t+k-i|}}{T_\alpha} + \alpha^2$$

$$E(\hat{\mu}_t\hat{\mu}_{t+k}) = E(\hat{\alpha}^2) = \frac{\sum_{i=1}^{T_0}\sum_{j=1}^{T_0}\gamma_{|i-j|}}{T_\alpha^2} + \alpha^2 = \frac{\gamma_0 + 2\sum_{i=1}^{T_\alpha-1}(T_\alpha - 1)\gamma_i}{T_\alpha^2} + \alpha^2.$$

Analogous results can be found when $t = T_0 + 1, ..., T - k$:

$$E(X_tX_{t+k}) = \gamma_k + \beta^2$$

$$E(X_t\hat{\mu}_{t+k}) = E(X_t\hat{\beta}) = \frac{\sum_{i=T_0+1}^{T} \gamma_{|t-i|}}{T_\beta} + \beta^2$$

$$E(\hat{\mu}_t X_{t+k}) = E(\hat{\beta} X_{t+k}) = \frac{\sum_{i=T_0+1}^{T} \gamma_{|t+k-i|}}{T_\beta} + \beta^2$$

$$E(\hat{\mu}_t\hat{\mu}_{t+k}) = E(\hat{\beta}^2) = \frac{\gamma_0 + 2\sum_{i=1}^{T_\beta-1}(T_\beta - 1)\gamma_i}{T_\beta^2} + \beta^2$$

Finally, when $t = T_0 - k + 1, ..., T_0$:

$$E(X_t X_{t+k}) = \gamma_k + \alpha\beta$$

$$E(X_t\hat{\mu}_{t+k}) = E(X_t\hat{\beta}) = \frac{\sum_{i=T_0+1}^{T} \gamma_{|t-i|}}{T_\beta} + \alpha\beta$$

$$E(\hat{\mu}_t X_{t+k}) = E(\hat{\alpha} X_{t+k}) = \frac{\sum_{i=1}^{T_0} \gamma_{|t+k-i|}}{T_\alpha} + \alpha\beta$$

$$E(\hat{\mu}_t\hat{\mu}_{t+k}) = E(\hat{\alpha}\hat{\beta}) = \frac{\sum_{i=1}^{T_0}\sum_{j=T_0+1}^{T} \gamma_{|i-j|}}{T_\alpha T_\beta} + \alpha\beta$$

Now note that in each term of the summation in (B.1), the expressions that depend on parameters, α^2, β^2 and $\alpha\beta$ cancel themselves as they appear twice with positive sign and twice with negative sign.

The second thing we should note is that $E(X_1\hat{\mu}_{1+k}) = E(\hat{\mu}_{T_0-k}X_{T_0})$, $E(X_2\hat{\mu}_{2+k}) = E(\hat{\mu}_{T_0-k-1}X_{T_0-1})$, and so on.

More generally, $E(X_t\hat{\mu}_{t+k}) = E(\hat{\mu}_{T_0-k-(t-1)}X_{T_0-(t-1)})$, $t = 1, ..., T_0 - k$.

In a similar way, we can see that $E(X_{T_0+1}\hat{\mu}_{T_0+1+k}) = E(\hat{\mu}_{T-k}X_T)$, $E(X_{T_0+2}\hat{\mu}_{T_0+2+k}) = E(\hat{\mu}_{T-k-1}X_{T-1})$, and so on.

More generally, $E(X_t\hat{\mu}_{t+k}) = E(\hat{\mu}_{T-k-(t-(T_0+1))}X_{T-(t-(T_0+1))})$, $t = T_0 + 1, ..., T - k$.

Because of these remarks, $\sum_{t=1}^{T_0-k} E(X_t\hat{\mu}_{t+k}) = \sum_{t=1}^{T_0-k} E(X_{t+k}\hat{\mu}_t)$ and $\sum_{t=T_0+1}^{T-k} E(X_t\hat{\mu}_{t+k}) = \sum_{t=T_0+1}^{T-k} E(X_{t+k}\hat{\mu}_t)$.

Therefore:

$$\begin{aligned} E(\hat{\gamma}_k) = \gamma_k &- \frac{2\sum_{t=1}^{T_0-k} f_{\hat{\alpha}}(t) + 2\sum_{t=T_0+1}^{T-k} f_{\hat{\beta}}(t)}{T-k} \\ &- \frac{\sum_{t=T_0-k+1}^{T_0}[f_{\hat{\beta}}(t) + f_{\hat{\alpha}}(t+k)]}{T-k} \\ &+ \frac{(T_\alpha - k)f_{\hat{\alpha}^2} + (T_\beta - k)f_{\hat{\beta}^2} + kf_{\hat{\alpha},\hat{\beta}}}{T-k}, \end{aligned}$$

where $f_{\hat{\alpha}}(t)$, $f_{\hat{\beta}}(t)$, $f_{\hat{\alpha^2}}(t)$, $f_{\hat{\beta^2}}(t)$ and $f_{\hat{\alpha},\hat{\beta}}$ are as defined in Proposition 2. □

Appendix C

Proof of Proposition 3

In this case, we have again

$$E(\hat{\gamma}_k) = \frac{\sum_{t=1}^{T-k}(E(X_t X_{t+k}) - E(X_t \hat{\mu}_{t+k}) - E(\hat{\mu}_t X_{t+k}) + E(\hat{\mu}_t \hat{\mu}_{t+k}))}{T-k},$$

with $\hat{\mu}_t = \bar{X} + \hat{\beta}\tilde{z}_t$. The value of $E(X_t X_{t+k})$ is given by:

$$\begin{aligned} E(X_t X_{t+k}) &= \gamma_k + \mu_t \mu_{t+k} \\ &= \gamma_k + (\mu_{1:T} + \beta\tilde{z}_t)(\mu_{1:T} + \beta\tilde{z}_{t+k}) \\ &= \gamma_k + \mu_{1:T}^2 + \mu_{1:T}\beta(\tilde{z}_t + \tilde{z}_{t+k}) + \beta^2 \tilde{z}_t \tilde{z}_{t+k}. \end{aligned}$$

The value of $E(X_t \hat{\mu}_{t+k})$ is given by:

$$\begin{aligned} E(X_t \hat{\mu}_{t+k}) &= E(X_t(\bar{X} + \hat{\beta}\tilde{z}_{t+k})) \\ &= E(X_t \bar{X}) + E(X_t \hat{\beta})\tilde{z}_{t+k}. \end{aligned}$$

The first term can be calculated as:

$$\begin{aligned} E(X_t \bar{X}) &= E\left(X_t \left(\frac{\sum_{i=1}^{T} X_i}{T}\right)\right) = E\left(\frac{\sum_{i=1}^{T} X_t X_i}{T}\right) = \frac{\sum_{i=1}^{T} E(X_t X_i)}{T} \\ &= \frac{\sum_{i=1}^{T}(\gamma_{|t-i|} + \mu_t \mu_i)}{T} = \frac{\sum_{i=1}^{T}\gamma_{|t-i|}}{T} + \frac{\sum_{i=1}^{T}(\mu_{1:T} + \beta\tilde{z}_t)(\mu_{1:T} + \beta\tilde{z}_i)}{T} \\ &= \frac{\sum_{i=1}^{T}\gamma_{|t-i|}}{T} + \frac{\sum_{i=1}^{T}(\mu_{1:T}^2 + \mu_{1:T}\beta\tilde{z}_i + \mu_{1:T}\beta\tilde{z}_t + \beta^2 \tilde{z}_t \tilde{z}_i)}{T} \\ &= \frac{\sum_{i=1}^{T}\gamma_{|t-i|}}{T} + \frac{T\mu_{1:T}^2 + T\mu_{1:T}\beta\tilde{z}_T}{T} = \frac{\sum_{i=1}^{T}\gamma_{|t-i|}}{T} + \mu_{1:T}(\mu_{1:T} + \beta\tilde{z}_t). \end{aligned}$$

And the second term as

$$
\begin{aligned}
E(X_t\hat{\beta}) &= E\left(X_t\left(\frac{\sum_{i=1}^T \tilde{z}_i X_i}{\sum_{i=1}^T \tilde{z}_i^2}\right)\right) = \frac{\sum_{i=1}^T \tilde{z}_i E(X_t X_i)}{\sum_{i=1}^T \tilde{z}_i^2} \\
&= \frac{\sum_{i=1}^T \tilde{z}_i(\gamma_{|t-i|} + \mu_t\mu_i)}{\sum_{i=1}^T \tilde{z}_i^2} = \frac{\sum_{i=1}^T \tilde{z}_i\gamma_{|t-i|}}{\sum_{i=1}^T \tilde{z}_i^2} + \frac{\sum_{i=1}^T \tilde{z}_i(\mu_{1:T} + \beta\tilde{z}_i)(\mu_{1:T} + \beta\tilde{z}_i)}{\sum_{i=1}^T \tilde{z}_i^2} \\
&= \frac{\sum_{i=1}^T \tilde{z}_i\gamma_{|t-i|}}{\sum_{i=1}^T \tilde{z}_i^2} + \frac{\sum_{i=1}^T \tilde{z}_i(\mu_{1:T}^2 + \mu_{1:T}\beta\tilde{z}_i + \mu_{1:T}\beta\tilde{z}_t + \beta^2\tilde{z}_t\tilde{z}_i)}{\sum_{i=1}^T \tilde{z}_i^2} \\
&= \frac{\sum_{i=1}^T \tilde{z}_i\gamma_{|t-i|}}{\sum_{i=1}^T \tilde{z}_i^2} + \frac{\mu_{1:T}\beta\sum_{i=1}^T \tilde{z}_i^2 + \beta^2\tilde{z}_t\sum_{i=1}^T \tilde{z}_i^2}{\sum_{i=1}^T \tilde{z}_i^2} = \frac{\sum_{i=1}^T \tilde{z}_i\gamma_{|t-i|}}{\sum_{i=1}^T \tilde{z}_i^2} + \mu_{1:T}\beta + \beta^2\tilde{z}_t.
\end{aligned}
$$

Analogous results can be found to $E(\hat{\mu}_t X_{t+k})$.

The value of $E(\hat{\mu}_t\hat{\mu}_{t+k})$ is given by:

$$
\begin{aligned}
E(\hat{\mu}_t\hat{\mu}_{t+k}) &= E((\bar{X} + \hat{\beta}\tilde{z}_t)(\bar{X} + \hat{\beta}\tilde{z}_{t+k})) \\
&= E(\bar{X}^2) + E(\bar{X}\hat{\beta})(\tilde{z}_t + \tilde{z}_{t+k}) + E(\hat{\beta}^2)\tilde{z}_t\tilde{z}_{t+k}.
\end{aligned}
$$

Where:

$$
\begin{aligned}
E(\bar{X}^2) &= \frac{\sum_{i=1}^T\sum_{j=1}^T E(X_iX_j)}{T^2} = \frac{\sum_{i=1}^T\sum_{j=1}^T(\gamma_{|i-j|} + \mu_i\mu_j)}{T^2} \\
&= \frac{\sum_{i=1}^T\sum_{j=1}^T \gamma_{|i-j|}}{T^2} + \frac{\sum_{i=1}^T\sum_{j=1}^T \mu_i\mu_j}{T^2} \\
&= \frac{\sum_{i=1}^T\sum_{j=1}^T \gamma_{|i-j|}}{T^2} + \frac{\sum_{i=1}^T\sum_{j=1}^T(\mu_{1:T} + \beta\tilde{z}_i)(\mu_{1:T} + \beta\tilde{z}_j)}{T^2} \\
&= \frac{\sum_{i=1}^T\sum_{j=1}^T \gamma_{|i-j|}}{T^2} + \frac{\sum_{i=1}^T\sum_{j=1}^T(\mu_{1:T}^2 + \beta\tilde{z}_i + \beta\tilde{z}_j + \beta^2\tilde{z}_i\tilde{z}_j)}{T^2} \\
&= \frac{\sum_{i=1}^T\sum_{j=1}^T \gamma_{|i-j|}}{T^2} + \frac{T^2\mu_{1:T}^2}{T^2} = \frac{\sum_{i=1}^T\sum_{j=1}^T \gamma_{|i-j|}}{T^2} + \mu_{1:T}^2.
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{X}\hat{\beta}) &= E\left(\left(\frac{\sum_{i=1}^{T} X_i}{T}\right)\left(\frac{\sum_{i=1}^{T} \tilde{z}_i X_i}{\sum_{i=1}^{T} \tilde{z}_i^2}\right)\right) = E\left(\frac{\sum_{i=1}^{T}\sum_{j=1}^{T} X_i X_j}{T\sum_{i=1}^{T} \tilde{z}_i^2}\right) \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_j E(X_i X_j)}{T\sum_{i=1}^{T} \tilde{z}_i^2} = \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_j (\gamma_{|i-j|} + \mu_i\mu_j)}{T\sum_{i=1}^{T} \tilde{z}_i^2} \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_j \gamma_{|i-j|}}{T\sum_{i=1}^{T} \tilde{z}_i^2} + \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_j \mu_i\mu_j}{T\sum_{i=1}^{T} \tilde{z}_i^2} \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_j \gamma_{|i-j|}}{T\sum_{i=1}^{T} \tilde{z}_i^2} + \left(\frac{\sum_{i=1}^{T} \mu_i}{T}\right)\left(\frac{\sum_{i=1}^{T} \mu_i \tilde{z}_i \mu_i}{\sum_{i=1}^{T} \tilde{z}_i^2}\right) \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_j \gamma_{|i-j|}}{T\sum_{i=1}^{T} \tilde{z}_i^2} + \left(\frac{\sum_{i=1}^{T} (\mu_{1:T} + \beta\tilde{z}_i)}{T}\right)\left(\frac{\sum_{i=1}^{T} \tilde{z}_i(\mu_{1:T} + \beta\tilde{z}_i)}{\sum_{i=1}^{T} \tilde{z}_i^2}\right) \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_j \gamma_{|i-j|}}{T\sum_{i=1}^{T} \tilde{z}_i^2} + \left(\frac{T\mu_{1:T}}{T}\right)\left(\frac{\beta\sum_{i=1}^{T} \tilde{z}_i^2}{\sum_{i=1}^{T} \tilde{z}_i^2}\right) \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_j \gamma_{|i-j|}}{T\sum_{i=1}^{T} \tilde{z}_i^2} + \mu_{1:T}\beta.
\end{aligned}
$$

$$
\begin{aligned}
E(\hat{\beta}^2) &= E\left(\left(\frac{\sum_{i=1}^{T} \tilde{z}_i X_i}{\sum_{i=1}^{T} \tilde{z}_i^2}\right)^2\right) = E\left(\frac{\sum_{i=1}^{T}\sum_{j=1}^{T} \tilde{z}_i\tilde{z}_j X_i X_j}{(\sum_{i=1}^{T} \tilde{z}_i)^2}\right) \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} z_i z_j E(X_i X_j)}{(\sum_{i=1}^{T} \tilde{z}_i)^2} = \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} z_i z_j (\gamma_{|i-j|} + \mu_i\mu_j)}{(\sum_{i=1}^{T} \tilde{z}_i)^2} \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} z_i z_j \gamma_{|i-j|}}{(\sum_{i=1}^{T} \tilde{z}_i)^2} + \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} z_i z_j \mu_i\mu_j}{(\sum_{i=1}^{T} \tilde{z}_i)^2} \\
&= \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} z_i z_j \gamma_{|i-j|}}{(\sum_{i=1}^{T} \tilde{z}_i)^2} + \left(\frac{\sum_{i=1}^{T} \tilde{z}_i\mu_i}{\sum_{i=1}^{T} \tilde{z}_i^2}\right)^2 = \frac{\sum_{i=1}^{T}\sum_{j=1}^{T} z_i z_j \gamma_{|i-j|}}{(\sum_{i=1}^{T} \tilde{z}_i)^2} + \beta^2.
\end{aligned}
$$

Joining together all the equation shown above we can simplify the calculations. Note that:

$$
E[(X_t - \hat{\mu}_t)(X_{t+k} - \hat{\mu}_{t+k})] =
$$

$$= \gamma_k + \mu_{1:T}^2 + \mu_{1:T}\beta(\tilde{z}_t + \tilde{z}_{t+k}) + \beta^2 \tilde{z}_t \tilde{z}_{t+k} - \frac{\sum_{i=1}^T \gamma_{|t-i|}}{T} - \mu_{1:T}^2 - \mu_{1:T}\beta\tilde{z}_t$$
$$- \tilde{z}_{t+k}\frac{\sum_{i=1}^T \tilde{z}_i\gamma_{|t-i|}}{\sum_{i=1}^T \tilde{z}_i^2} - \mu_{1:T}\beta\tilde{z}_{t+k} - \beta^2\tilde{z}_t\tilde{z}_{t+k} - \frac{\sum_{i=1}^T \gamma_{|t+k-i|}}{T} - \mu_{1:T}^2 - \mu_{1:T}\beta\tilde{z}_{t+k}$$
$$- \tilde{z}_t\frac{\sum_{i=1}^T \tilde{z}_i\gamma_{|t-i|}}{\sum_{i=1}^T \tilde{z}_i^2} - \mu_{1:T}\beta\tilde{z}_t - \beta^2\tilde{z}_t\tilde{z}_{t+k} + \frac{\sum_{i=1}^T \gamma_{|i-j|}}{T} + \mu_{1:T}^2$$
$$+ \tilde{z}_t\frac{\sum_{i=1}^T\sum_{j=1}^T \tilde{z}_j\gamma_{|t-i|}}{T\sum_{i=1}^T \tilde{z}_i^2}(\tilde{z}_t + \tilde{z}_{t+k}) + \mu_{1:T}\beta(\tilde{z}_t + \tilde{z}_{t+k})$$
$$+ \tilde{z}_t\tilde{z}_{t+k}\frac{\sum_{i=1}^T\sum_{j=1}^T \tilde{z}_i\tilde{z}_j\gamma_{|i-j|}}{(\sum_{i=1}^T \tilde{z}_i^2)^2} + \beta^2\tilde{z}_t\tilde{z}_{t+k}$$
$$= \gamma_k - \frac{\sum_{i=1}^T \gamma_{|t-i|}}{T} - \tilde{z}_{t+k}\frac{\sum_{i=1}^T \tilde{z}_i\gamma_{|t-i|}}{\sum_{i=1}^T \tilde{z}_i^2} - \frac{\sum_{i=1}^T \gamma_{|t+k-i|}}{T} - \tilde{z}_t\frac{\sum_{i=1}^T \tilde{z}_i\gamma_{|t+k-i|}}{\sum_{i=1}^T \tilde{z}_i^2}$$
$$+ \frac{\sum_{i=1}^T\sum_{j=1}^T \gamma_{|i-j|}}{T^2} + \frac{\sum_{i=1}^T\sum_{j=1}^T \tilde{z}_j\gamma_{|i-j|}}{T\sum_{i=1}^T \tilde{z}_i^2}(\tilde{z}_t + \tilde{z}_{t+k}) + \tilde{z}_t\tilde{z}_{t+k}\frac{\sum_{i=1}^T\sum_{j=1}^T \tilde{z}_i\tilde{z}_j\gamma_{|i-j|}}{(\sum_{i=1}^T \tilde{z}_i^2)^2}.$$

At last:

$$E(\hat{\gamma}_k) = \gamma_k - \frac{\sum_{t=1}^{T-k}\sum_{i=1}^T \gamma_{|t-i|}}{(T-k)T} - \frac{\sum_{t=1}^{T-k}\tilde{z}_{t+k}\sum_{i=1}^T \tilde{z}_i\gamma_{|t-i|}}{(T-k)\sum_{i=1}^T \tilde{z}_i^2} - \frac{\sum_{t=1}^{T-k}\sum_{i=1}^T \gamma_{|t+k-i|}}{(T-k)T}$$
$$- \frac{\sum_{t=1}^{T-k}\tilde{z}_t\sum_{i=1}^T \tilde{z}_i\gamma_{|t+k-i|}}{(T-k)\sum_{i=1}^T \tilde{z}_i^2} + \frac{[\sum_{t=1}^{T-k}(\tilde{z}_t\tilde{z}_{t+k})]\sum_{i=1}^T\sum_{j=1}^T \tilde{z}_i\gamma_{|i-j|}}{(T-k)T\sum_{i=1}^T \tilde{z}_i^2}$$
$$+ \frac{[\tilde{z}_t\tilde{z}_{t+k}]\sum_{i=1}^T\sum_{j=1}^T \tilde{z}_i\tilde{z}_j\gamma_{|i-j|}}{(T-k)(\sum_{i=1}^T \tilde{z}_i^2)^2} + \frac{\sum_{i=1}^T\sum_{j=1}^T \gamma_{|i-j|}}{T^2}.$$

As $\sum_{i=1}^T \gamma_{|1-i|} = \sum_{i=1}^T \gamma_{|T-i|}$, $\sum_{i=1}^T \gamma_{|2-i|} = \sum_{i=1}^T \gamma_{|T-1-i|}$, and so on, thus $\sum_{t=1}^{T-k}\sum_{i=1}^T \gamma_{|t-i|} = \sum_{t=1}^{T-k}\sum_{i=1}^T \gamma_{|t+k-i|}$. Therefore, the equation above becomes:

$$E(\hat{\gamma}_k) = \gamma_k - \frac{2\sum_{t=1}^{T-k}\sum_{i=1}^T \gamma_{|t-i|}}{(T-k)T} - \frac{\sum_{t=1}^{T-k}\sum_{i=1}^T(\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|} + \tilde{z}_t\tilde{z}_i\gamma_{|t+k-i|})}{(T-k)\sum_{i=1}^T \tilde{z}_i^2}$$
$$+ \frac{[\sum_{t=1}^{T-k}(\tilde{z}_t + \tilde{z}_{t+k})]\sum_{i=1}^T\sum_{j=1}^T \tilde{z}_j\gamma_|i-j|}{(T-k)T\sum_{i=1}^T \tilde{z}_i^2}$$
$$+ \frac{[\sum_{t=1}^{T-k}\tilde{z}_t\tilde{z}_{t+k}]\sum_{i=1}^T\sum_{j=1}^T \tilde{z}_i\tilde{z}_j\gamma_|i-j|}{(T-k)(\sum_{i=1}^T \tilde{z}_i^2)^2} + \frac{\sum_{i=1}^T\sum_{j=1}^T \gamma_{|i-j|}}{T^2}.$$

□

Appendix D

Proof of Proposition 4

The following lemma will be used in the proof of Proposition 4:

Lemma 1: Define $\tilde{z}_1, ..., \tilde{z}_T$ as $\tilde{z}_i = i - (T+1)/2$, $i = 1, ..., T$. Then

$$\sum_{t=1}^{T-k} \tilde{z}_t \tilde{z}_{t+k} = \frac{(T-k)^3 - (T-k)(3k^2+1)}{12}.$$

Proof:

$$\begin{aligned}
\sum_{t=1}^{T-k} \tilde{z}_t \tilde{z}_{t+k} &= \sum_{t=1}^{T-k} \left(t - \frac{T+1}{2}\right)\left(t + k - \frac{T+1}{2}\right) \\
&= \sum_{t=1}^{T-k} \left(t^2 + tk - t(T+1) - k\left(\frac{T+1}{2}\right) + \left(\frac{T+1}{2}\right)^2\right) \\
&= \frac{2(T-k)^3 + 3(T-k)^2 + T - k}{6} + \frac{k(T-k)(T-k+1)}{2} \\
&- \frac{2(T+1)(T-k)(T-k+1)}{4} - \frac{k(T-k)(T+1)}{2} + \frac{(T-k)(T+1)^2}{4}.
\end{aligned}$$

The results in the above equation can be achieved using the fact that $\sum_{t=1}^{T} t = T(T+1)/2$ and $\sum_{t=1}^{T} t^2 = (2T^3 + 3T^2 + T)/6$. The above equation can be written as $\sum_{t=1}^{T-k} \tilde{z}_t \tilde{z}_{t+k} = (*)(T-k)/2$ where:

$$
\begin{aligned}
(*) = {} & \frac{2(T-k)^2 + 3(T-k) + 1}{3} + k(T-k+1)2 - (T+1)(T-k+1) - k(T+1) \\
& + \frac{(T+1)^2}{2}. \\
= {} & \frac{2(T^2 - 2Tk + k^2) + 3T - 3k + 1}{3} + Tk - k^2 + k - (T^2 - Tk + T + T - k + 1) - Tk \\
& - k + \frac{T^2 + 2T + 1}{2} \\
= {} & \frac{2T^2 - 4Tk + 2k^2 + 3T - 3k + 1}{3} - T^2 + Tk - 2T - k^2 + k - 1 + \frac{T^2 + 2T + 1}{2} \\
= {} & \frac{4T^2 - 8Tk + 6T + 4k^2 - 6k + 2 - 6T^2 + 6Tk - 12T - 6k^2 + 6k - 6 + 3T^2 + 6T + 3}{6} \\
= {} & \frac{T^2 - 2Tk - 2k^2 - 1}{6} \\
= {} & \frac{(T-k)^2 - (3k^2 + 1)}{6}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\sum_{t=1}^{T-k} \tilde{z}_t \tilde{z}_{t+k} &= \frac{(T-k)^2 - (3k^2+1)}{6} \frac{(T-k)}{2} \\
&= \frac{((T-k)^3 - (T-k)(3k^2+1))}{12}.
\end{aligned}
$$

□

Proof of Proposition 4: Under the conditions of Proposition 4, it is easy to show that $\tilde{z}_i = i - (T+1)/2$. In the case of the numerator of the third term on the right side of Equation (2.7):

$$\begin{aligned}
\sum_{t=1}^{T-k}\sum_{i=1}^{T}(\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|}+\tilde{z}_t\tilde{z}_i\gamma_{|t+k-i|}) &= \sum_{t=1}^{T-k}\sum_{i=1}^{T}\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|}+\sum_{t=1}^{T-k}\sum_{i=1}^{T}\tilde{z}_t\tilde{z}_i\gamma_{|t+k-i|}\\
&= \sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|}+\sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_t\tilde{z}_i\gamma_{|t+k-i|}\\
&= \sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|}+\sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_t\tilde{z}_{T+1-i}\gamma_{|t+k-(T+1-i)|}\\
&= \sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|}+\sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_{T-k+1-t}\tilde{z}_{T+1-i}\gamma_{|T+1-t-(T+1-i)|}\\
&= \sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|}+\sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_{T-k+1-t}\tilde{z}_{T+1-i}\gamma_{|t-i|}\\
&= \sum_{i=1}^{T}\sum_{t=1}^{T-k}(\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|}+\tilde{z}_{T-k+1-t}\tilde{z}_{T+1-i}\gamma_{|t-i|})
\end{aligned}$$

As $\tilde{z}_{t+k}=-\tilde{z}_{T-k+1-t}$ and $\tilde{z}_i=-\tilde{z}_{T+1-i}$, that becomes:

$$=2\sum_{i=1}^{T}\sum_{t=1}^{T-k}\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|}.$$

Furthermore, note that

$$\begin{aligned}
\sum_{t=1}^{T-k}(\tilde{z}_t+\tilde{z}_{t+k}) &= \sum_{t=1}^{T-k}\tilde{z}_t+\sum_{t=1}^{T-k}\tilde{z}_{t+k}\\
&= \sum_{t=1}^{T-k}\tilde{z}_t+\sum_{t=1}^{T-k}\tilde{z}_{T+1-t}\\
&= \sum_{t=1}^{T-k}(\tilde{z}+\tilde{z}_{T+1-t})=0.
\end{aligned}$$

Finally, it is possible to show that:

$$\begin{aligned}
\sum_{i=1}^{T}\tilde{z}_i^2 &= \frac{T^3-T}{12},\\
\sum_{j=1}^{T-k}\tilde{z}_j\tilde{z}_{j+k} &= \frac{(T-k)^3-(T-k)(3k^2+1)}{12}.
\end{aligned}$$

The first equation above is trivial, given the well known variance of an uniform distribution. The proof of the second equation can be found on Lemma 1. Applying these results to Equation (2.7) we find:

$$E(\hat{\gamma}_k) = \gamma_k - \frac{2\sum_{t=1}^{T-k}\sum_{i=1}^{T}\gamma_{|t-i|}}{(T-k)T} - \frac{\sum_{t=1}^{T-k}\sum_{i=1}^{T}(\tilde{z}_{t+k}\tilde{z}_i\gamma_{|t-i|} + \tilde{z}_t\tilde{z}_i\gamma_{|t+k-i|})}{(T-k)\sum_{i=1}^{T}\tilde{z}_i^2}$$
$$+ \frac{[\sum_{t=1}^{T-k}(\tilde{z}_t + \tilde{z}_{t+k})]\sum_{i=1}^{T}\sum_{j=1}^{T}\tilde{z}_j\gamma_|i-j|}{(T-k)T\sum_{i=1}^{T}\tilde{z}_i^2}$$
$$+ \frac{[\sum_{t=1}^{T-k}\tilde{z}_t\tilde{z}_{t+k}]\sum_{i=1}^{T}\sum_{j=1}^{T}\tilde{z}_i\tilde{z}_j\gamma_|i-j|}{(T-k)(\sum_{i=1}^{T}\tilde{z}_i^2)^2} + \frac{\sum_{i=1}^{T}\sum_{j=1}^{T}\gamma_{|i-j|}}{T^2}.$$

□

Appendix E

Proof of Proposition 5

$$E(\hat{\gamma}_k) = \frac{\sum_{t=1}^{T-k}(E(X_t X_{t+k}) + E(X_t \hat{\mu}_{t+k}) + E(\hat{\mu}_t X_{t+k}) + E(\hat{\mu}_t \hat{\mu}_{t+k}))}{T-k},$$

where:

$$E(X_t X_{t+k}) = \gamma_k + \alpha_{s_1(t)} \alpha_{s_1(t+k)},$$

$$\begin{aligned} E(X_t \hat{\mu}_{t+k}) &= \frac{\sum_{i=0}^{s_2(t+k)-1} E(X_t X_{s_1(t+k)+si})}{s_2(t+k)}, \\ &= \frac{\sum_{i=0}^{s_2(t+k)-1} \gamma_{s_2(t+k)}}{s_2(t+k)} + \alpha_{s_1(t)} \alpha_{s_1(t+k)} \end{aligned}$$

$$\begin{aligned} E(\hat{\mu}_t X_{t+k}) &= \frac{\sum_{i=0}^{s_2(t)-1} \gamma_{|t+k-s_1(t)-si|}}{s_2(t)} + \alpha_{s_1(t)} \alpha_{s_1(t+k)}, \\ E(\hat{\mu}_t \hat{\mu}_{t+k}) &= \frac{\sum_{i=0}^{s_2(t)-1} \sum_{j=0}^{s_2(t+k)-1} \gamma_{|s_1(t)+si-s_2(t+k)-sj|}}{s_2(t) s_2(t+k)}. \end{aligned}$$

Therefore:

$$\begin{aligned} E(\hat{\gamma}_k) = \gamma_k &- \sum_{t=1}^{T-k} \left(\frac{\sum_{i=0}^{s_2(t+k)-1} \gamma_{|t-s_1(t+k)-si|}}{(T-k) s_2(t+k)} \right) - \sum_{t=1}^{T-k} \left(\frac{\sum_{i=0}^{s_2(t)-1} \gamma_{|t+k-s_1(t)-si|}}{(T-k) s_2(t)} \right) \\ &+ \sum_{t=1}^{T-k} \left(\frac{\sum_{i=0}^{s_2(t)-1} \sum_{j=0}^{s_2(t+k)-1} \gamma_{|s_1(t)+si-s_1(t+k)-sj|}}{(T-k) s_2(t) s_2(t+k)} \right). \end{aligned}$$

In the third sum of the equation above, for $t_1, t_2 \in \{1, ..., T\}$ such that $s_1(t_1) = s_1(t_2)$ (both belong to the same season), the corresponding terms of the sun are equal. Note the terms of the sum are not dependent on t given $s_1(t)$ and $s_2(t)$. Therefore, we can rewrite this sum as:

$$\sum_{t=1}^{s} \left(\left(\lfloor \frac{T-k-t}{S} \rfloor + 1 \right) \frac{\sum_{i=0}^{s_2(t)-1} \sum_{j=0}^{(} s_2(t+k)-1) \gamma_{|s_1(t)+si-s_1(t+k)-sj|}}{s_2(t)s_2(t+k)} \right).$$

This allows for $E(\hat{\gamma}_k)$ to be written in a way that calculations are of order T^2 instead of T^3:

$$E(\hat{\gamma}_k) = \gamma_k - \sum_{t=1}^{T-k} \left(\frac{\sum_{i=0}^{s_2(t+k)-1} \gamma_{|t-s_1(t+k)-si|}}{(T-k)s_2(t+k)} \right) - \sum_{t=1}^{T-k} \left(\frac{\sum_{i=0}^{s_2(t)-1} \gamma_{|t+k-s_1(t)-si|}}{(T-k)s_2(t)} \right)$$
$$+ \sum_{t=1}^{s} \left[\left(\left\lfloor \frac{T-k-t}{s} \right\rfloor + 1 \right) \frac{\sum_{i=0}^{s_2(t)-1} \sum_{j=0}^{s_2(t+k)-1} \gamma_{|s_1(t)+si-s_1(t+k)-sj|}}{s_2(t)s_2(t+k)} \right].$$

□

Appendix F

Proof of Proposition 6

To proof this proposition, the following lemma will be needed to reach this goal.

Lemma 2: Consider the function:

$$f_T = \sum_{i=0}^{\infty} w_{T,i} a_i, \quad T \in \mathbb{N}$$

where a_i and $w_{T,i}$ are sequences of real numbers satisfying the following conditions:

1. $a_i \to 0$ as $i \to \infty$.
2. There exists a real number $U > 0$ such that $\sum_{i=1}^{\infty} |w_{T,i}| \leq U$ for a sufficiently large T.
3. For a fixed i, $w_{T,i} \to 0$ as $T \to \infty$.

Then the function of f_T goes to zero as $T \to \infty$. $\square$

Proof: Let $\epsilon > 0$. Define i_0 such that $a_i < \epsilon$ for $i > i_0$. Note that $f_T = \sum_{i=0}^{i_0} w_{T,i} a_i + \sum_{i=i_0+1}^{\infty} w_{T,i} a_i$. The sum $\sum_{i=0}^{i_0} w_{T,i} a_i$ goes to zero as $T \to \infty$ due to the fact that $w_{T,i} \to 0$ when $T \to \infty$. The second sum is bounded by ϵU. $\square$

Now consider $w_{T,i}$ as the weight of the i-th autocorrelation of $B^{\rho}_{T,k}(\lambda)$ (given in Equation (2.5)). We can determine an upper bound for the sum of the weights:

$$\sum_{i=0}^{T-1} |w_{T,i}| \leq \frac{T+k}{T-k} + \frac{Tk}{T(T-k)},$$

which converges to one as $T \to \infty$ when k is fixed. We also have that, for any i:

$$|w_{T,i}| \leq \frac{(T+k)}{(T-k)} \frac{2(T-1)}{T^2} + \frac{4k}{(T-k)}, \tag{F.1}$$

which goes to zero as $T \to \infty$.

To complete the proof it is sufficient to show that the autocorrelation function of an ARFIMA process under the proposition assumptions converges uniformly to zero as $T \to \infty$. In order to do that we will employ the splitting method.

For a pure MA(q) model, the autocovariance function is given by:

$$\gamma_k^{(ma)} = \sigma^2 \sum_{j=0}^{q} \theta_k \theta_{k+j}, \quad k = 0, ..., q$$

where $\theta_0 = 1$. For $k > q$, $\gamma_k^{(ma)} = 0$. Clearly, such autocovariance function converges uniformly to zero as $k \to \infty$.

For a pure AR(p) model, the autocovariance function obtained satisfies the recursive relation $\gamma_k = \phi_1 \gamma_{k-1} + ... + \phi_p \gamma_{k-p}$ for $k > 0$. For any values of $\phi_1, ..., \phi_p$ in the parametric space, the following generic inequality can be established:

$$|\gamma_k^{(\mathrm{ar})}| \leq K^{\lceil k/p \rceil} \gamma_0^{(\mathrm{ar})}.$$

We can also determine a bound for the sum of the autocovariance function, a result that will be important in the remaining of the proof:

$$\left| \sum_{k=0}^{\infty} \gamma_k^{(\mathrm{ar})} \right| \leq \gamma_0^{(\mathrm{ar})} + \gamma_0^{(\mathrm{ar})} \sum_{k=1}^{\infty} K^{\lceil k/p \rceil} \leq \gamma_0^{(\mathrm{ar})} + \gamma_0^{(\mathrm{ar})} p \sum_{k=1}^{\infty} K^k \leq \gamma_0^{(\mathrm{ar})} + \frac{\gamma_0^{(\mathrm{ar})} pK}{1-K}.$$

For an ARFIMA(0,d,0), the autocorrelation function, $\rho_k^{(\mathrm{afm})}$, is given in (3.4). The value of $\rho_k^{(\mathrm{afm})}(d)$ is positive for $d > 0$, negative for $d < 0$ and zero for $d = 0$. Therefore, for $d \geq 0$,

$$\rho_k^{(\mathrm{afm})}(d) \leq \rho_k^{(\mathrm{afm})}(d^{\max})$$

where $d^{\max}$ stands for the possible values of d in the parametric space. For any stationary value of d, $\rho_k^{(afm)}(d^{\max}) \to 0$ as $k \to \infty$. For $d < 0$, it is easy to see that $|d/(1-d)|$ is decreasing while $|(k-1+d)/(k-d)|$ for $k > 1$ is increasing. Therefore, for $d < 0$

$$|\rho_k^{(\text{afm})}(d)| \leq \frac{1}{2}\prod_{i=2}^{k}\frac{i-1}{i} = \frac{1}{2k}.$$

That is, both the lower and upper bounds for $\rho_k^{(afm)}(d)$ converges to zero as $k \to \infty$.

To show the uniform convergence of ρ_k in the ARFIMA(p,d,q) model, it suffices to show the uniform convergence of the numerator in (3.6), as the denominator is inferiorly bounded by σ^2 as a function of the parameters, regardless their values.

We begin the proof with the ARFIMA$(p,d,0)$ model. The numerator in (3.6) can be written as

$$\gamma_k = \sigma^{-2}\gamma_0^{(1)}\gamma_0^{(2)}\Big[\rho_0^{(1)}\rho_{-k}^{(2)} + \sum_{i=1}^{\infty}\rho_i^{(1)}\big(\rho_{i-k}^{(2)} + \rho_{i+k}^{(2)}\big)\Big], \quad k \geq 0. \tag{F.2}$$

Now consider $\gamma_0^{(1)}$ and $\rho_0^{(1)}$ to be the autocovariance and autocorrelation functions of an ARFIMA$(0,d,0)$ process and $\gamma_0^{(2)}$ and $\rho_0^{(2)}$ to be the autocovariance and autocorrelation functions of an AR(p) process. The value of $\gamma_0^{(1)}$ is given by $\Gamma(-2d+1)/\Gamma(-d+1)^2$. For $d \in (-1, 1/2)$, $\gamma_0^{(1)}$ as a function of d is well defined and continuous. As the parametric space is compact, $\gamma_0^{(1)}$ is clearly bounded. The value of $\gamma_0^{(2)}$ is given by $\gamma_0^{(2)} = \sigma^2/(1 - \phi_1\rho_1 - ... - \phi_p\rho_p)$. Therefore, as the absolute values of the autocorrelations are smaller than 1:

$$\begin{aligned}\gamma_0^{(2)} &\leq \frac{\sigma^2}{1 - |\phi_1\rho_1 + ... + \phi_p\rho_p|} \\ &\leq \frac{\sigma^2}{1 - (|\phi_1| + ... + |\phi_p|)} \leq \frac{\sigma^2}{1-K}.\end{aligned}$$

The expression inside the brackets in Equation (F.2) has upper bound

$$\left|\rho_0^{(1)}\rho_{-k}^{(2)} + \sum_{i=1}^{\infty}\rho_i^{(1)}(\rho_{i-k}^{(2)} + \rho_{i+k}^{(2)})\right| \leq$$

$$\rho_0^{(1)}(d^{\max})K^{\lceil k/p\rceil} + \sum_{i=1}^{\infty} \rho_i^{(1)}(d^{\max}) \left(K^{\lceil |i-k|/p\rceil} + K^{\lceil |i+k|/p\rceil}\right).$$

In the above equation, if $k \to \infty$, clearly $\rho_0^{(1)}(d^{\max})K^{\lceil k/p\rceil} \to 0$. Additionally, $\rho_i^{(1)}(d^{\max}) \to 0$ as $i \to \infty$ and $K^{\lceil |i-k|/p\rceil} + K^{\lceil |i+k|/p\rceil} \to 0$ as $k \to \infty$. Furthermore, note that

$$\sum_{i=1}^{\infty} K^{\lceil |i+k|/p\rceil} \leq \sum_{i=0}^{\infty} K^{\lceil i/p\rceil} = 1 + \frac{pK}{1-K} \leq 2 + \frac{2pK}{1-K}.$$

And,

$$\sum_{i=1}^{\infty} K^{\lceil |i-k|/p\rceil} \leq \sum_{i=1}^{\infty} K^{\lceil |i|/p\rceil} + \sum_{i=0}^{\infty} K^{\lceil i/p\rceil} = 2\sum_{i=0}^{\infty} K^{\lceil i/p\rceil} = 2 + \frac{2pK}{1-K}.$$

Therefore, if we set $\rho_i^{(1)} = a_i$ and $K^{\lceil |i-k|/p\rceil} + K^{\lceil |i+k|/p\rceil} = w_{k,i}$, the criterias of Lemma 2 are satisfied, proving the uniform convergence to zero of the autocovariance function of the ARFIMA$(p, d, 0)$ model.

The generalization of this conclusion to a general ARFIMA(p, d, q) model is easy using the splitting method combining the autocovariance of the ARFIMA$(p, q, 0)$ model and the MA(q) model, both of which are now known to converge uniformly to zero. Because only a finite number of lags in the autocovariance function of a MA(q) model is nonzero, the sum in (3.6) becomes:

$$\rho_k = \frac{\sum_{i=-q}^{q} \gamma_i^{(1)} \gamma_{i-k}^{(2)}}{\sum_{i=-q}^{q} \gamma_i^{(1)} \gamma_i^{(2)}}.$$

Because $\gamma_i^{(1)}$ and $\gamma_{i-k}^{(2)}$ are uniformly convergent when $k \to \infty$, so are $\gamma_i^{(1)}\gamma_{i-k}^{(2)}$ and $\sum_{i=-q}^{q} \gamma_i^{(1)}\gamma_{i-k}^{(2)}$. Therefore, the numerator of the above equation converges uniformly to zero, proving the uniform convergence of the autocorrelation function of the general ARFIMA(p, d, q) model.

Using Lemma 2, with the weights given in (F.1), in conjunction with the uniform convergence of ρ_k results in $B_{T,k}^{\rho}$ converging uniformly to zero as $T \to \infty$ for a fixed k. Thus, $\rho_{T,k}$ converges uniformly to ρ_k as $T \to \infty$. □

Appendix G

Proof of Theorem 1

The BCMDE searches for the λ that minimizes $S(\lambda)$, given in Equation (3.9). We will define the function $f : \mathbb{R}^{2K} \to \mathbb{R}^{+}$, $f(a,b) = (a-b)'W(a-b)$, where W is any positive definite matrix. Note that $S(\lambda) = f(\hat{\varrho}, \varrho_T(\lambda))$.

Part 1: For any three vectors with same dimension, a, b, c, $f(a,b)/2 \leq f(a,c) + f(b,c)$.

Using the definition of f it is easily seen that $f(a,c) = f(a-c,0)$ and $f(a,b) = f(a-c, b-c)$. Therefore we just have to prove that $f(a-c,b-c)/2 \leq f(a-c,0) + f(b-c,0)$. Define the vectors $x = a-c$ and $y = b-c$. As W is positive definite,:

$$\begin{aligned} (x+y)'W(x+y) &\geq 0, \\ x'Wx + y'Wy + x'Wy + y'Wx &\geq 0, \\ (-x'Wx - y'Wy - x'Wy - y'Wx)/2 &\leq 0, \\ (x'Wx + y'Wy - x'Wy - y'Wx)/2 &\leq x'Wx + y'Wy, \\ (x-y)'W(x-y)/2 &\leq x'Wx + y'Wy, \\ f(x,y)/2 &\leq f(x,0) + f(y,0), \end{aligned}$$

Part 2: Let λ_0 be the true parameter value, then $f(\hat{\varrho}, \varrho_T(\lambda_0)) \xrightarrow{P} 0$ as $T \to \infty$.

Define the vector $B_T(\lambda) = \varrho_T(\lambda) - \varrho(\lambda)$. From Proposition 2, $\varrho_T(\lambda)$ converges uniformly to $\varrho(\lambda)$, so $B_T(\lambda)$ converges uniformly to zero. Thus,

$$\begin{aligned} f(\hat{\varrho}, \varrho_T(\lambda_0)) &= (\hat{\varrho} - \varrho(\lambda_0) - B_T(\lambda_0))' W (\hat{\varrho} - \varrho(\lambda_0) - B_T(\lambda_0)) \\ &= (\hat{\varrho} - \varrho(\lambda_0))' W (\hat{\varrho} - \varrho(\lambda_0)) - (\hat{\varrho} - \varrho(\lambda_0))' W B_T(\lambda_0) \\ &- B_T(\lambda_0) W (\hat{\varrho} - \varrho(\lambda_0)) + B_T(\lambda_0)' W B_T(\lambda_0). \end{aligned}$$

Due to the fact that $\hat{\varrho} - \varrho(\lambda_0) \to 0$ in probability as $T \to \infty$, so does $(\hat{\varrho} - \varrho(\lambda_0))' W (\hat{\varrho} - \varrho(\lambda_0))$, $(\hat{\varrho} - \varrho(\lambda_0))' W B_T(\lambda_0)$ and $B_T(\lambda_0) W (\hat{\varrho} - \varrho(\lambda_0))$. Which implies that $B_T(\lambda_0)' W B_T(\lambda_0)$ also converges to zero.

Part 3: For $\lambda \neq \lambda_0$, $f(\hat{\varrho}, \varrho_T(\lambda)) \xrightarrow{P} c$, $c > 0$, as $T \to \infty$.

We have already seen that, if $T \to \infty$, then $B_T(\lambda) \to 0$ and $\hat{\varrho} - \varrho(\lambda)$ converges in probability to a non-zero vector (if the injectivity assumption of ϱ is satisfied). Therefore, $\hat{\varrho} - \varrho(\lambda) - B_T(\lambda)$ converges in probability to a non-zero vector and $f(\hat{\varrho}, \varrho_T(\lambda)) = (\hat{\varrho} - \varrho_T(\lambda))' W (\hat{\varrho} - \varrho_T(\lambda)) \to c$, $c > 0$, using the assumption that W is positive definite.

Part 4: For any neighborhood $V(\lambda_0)$ around λ_0, there exists a $L_2 > 0$ such that $f(\varrho_T(\lambda_0), \varrho_T(\lambda)) \geq L_2$ if $\lambda \notin V(\lambda_0)$, for T large enough.

$f(\varrho_T(\lambda), \varrho_T(\lambda_0))$ can be written as

$$\begin{aligned} f(\varrho_T(\lambda_0), \varrho_T(\lambda)) &= (\varrho(\lambda_0) - \varrho(\lambda))' W (\varrho(\lambda_0) - \varrho(\lambda)) \\ &+ (\varrho(\lambda_0) - \varrho(\lambda))' W (B_T(\lambda_0) - B_T(\lambda)) \\ &+ (B_T(\lambda_0) - B_T(\lambda))' W (\varrho(\lambda_0) - \varrho(\lambda)) \\ &+ (B_T(\lambda_0) - B_T(\lambda))' W (B_T(\lambda_0) - B_T(\lambda)). \end{aligned}$$

As $\varrho(\lambda)$ is a continuos injective function and Λ is a compact set, than there exists a $L > 0$ such that, for $\lambda \notin V(\lambda_0)$, $(\varrho(\lambda_0) - \varrho(\lambda))' W (\varrho(\lambda_0) - \varrho(\lambda)) > L$. Furthermore, because $B_T(\lambda)$ and converges uniformly to zero as $T \to \infty$, then $(\varrho(\lambda_0) - \varrho(\lambda))' W (B_T(\lambda_0) - B_T(\lambda)) \to 0$, $(B_T(\lambda_0) - B_T(\lambda))' W (\varrho(\lambda_0) - \varrho(\lambda)) \to 0$ and $(B_T(\lambda_0) - B_T(\lambda))' W (B_T(\lambda_0) - B_T(\lambda)) \to 0$, both uniformly. Therefore, for T large enough, $f(\varrho_T(\lambda_0), \varrho_T(\lambda)) \geq L_2$, for L_2 satisfying $0 < L_2 < L$.

Part 5: For any neighborhood of λ_0, $V(\lambda_0)$, $P(\hat{d} \in V(\lambda_0)) \to 1$, where $\hat{\lambda}$ is the BCMDE of λ.

In Part 2 it was shown that $f(\hat{\varrho}, \varrho_T(\lambda_0)) \xrightarrow{P} 0$. Therefore for any $L_2 > 0$, $P(f(\hat{\varrho}, \varrho_T(d_0)) < L_2/4) \to 1$. It was also shown, in Part 1, that for any three vectors a, b, c, $f(a,b)/2 \leq f(a,c) + f(b,c)$. Therefore:

$$f(\hat{\varrho}, \varrho_T(\lambda)) \geq f(\varrho_T(\lambda), \varrho_T(\lambda_0))/2 - f(\hat{\varrho}, \varrho_T(\lambda_0)).$$

For T large enough and $\lambda \notin V(\lambda_0)$, the first term in the right hand side of the above equation is greater or equal to $L_2/2$, while the probability that the second term is less than $L_2/4$ goes to 1. When both conditions are satisfied, $f(\hat{\varrho}, \varrho_T(\lambda)) \geq L_2/4$. In other words, the probability that the minimum of f can be found at $V(\lambda_0)$, instead of outside $V(\lambda_0)$, goes to 1. The proof of the convergence in probability is completed. □

Appendix H

Proof of Theorem 2

To evaluate the asymptotic distribution of the BCMDE, we should first guarantee that the conditions of Theorem 3.2 of Newey and McFadden (1994) are satisfied. Define $\hat{g}_T(\lambda) = \hat{\varrho} - \varrho_T(\lambda)$ and let λ_0 be an interior point of Λ. We need to fulfill the following conditions:

1. $\hat{g}_T(\lambda)$ is continuously differentiable in a neighborhood $V(\lambda_0)$ of λ_0;

2. $\sqrt{T}\hat{g}_T(\lambda_0) \xrightarrow{D} N(0, \Omega)$, Ω being a covariance matrix;

3. There exists a $G(\lambda)$ that is continuos at λ_0 and $\sup_{\lambda \in \mathcal{N}} ||\nabla_\lambda \hat{g}_T(\lambda) - G(\lambda)|| \to 0$;

4. For $G = G(\lambda_0)$, $G'WG$ is non-singular.

If all these requirements are satisfied, then $\sqrt{T}(\hat{\lambda} - \lambda) \to N(0, \Omega)$.

1. In this case the vector $\varrho_T(d)$ is given by $\varrho_T(d) = (\rho_{T,k_1}, ..., \rho_{T,k_m})$ where $\rho_{T,k} = \frac{\rho_k + B^\rho_{T,k}}{1 + B^\rho_{T,0}}$. That is, $\rho_{T,k}$ is the ratio between two linear combinations of autocorrelations. It is already known that the theoretical autocorrelations of a ARFIMA$(0, d, 0)$ model are continuously differentiable. Therefore, both the numerator and the denominator of $\rho_{T,k}$ are continuously differentiable. Moreover the denominator does not vanish because it is the expectation of a

non-negative random variable. Thus, as the derivative of $\hat{\rho}$ is equal to zero, $\hat{g}_T(d)$ is continuously differentiable in a neighborhood $V(d_0)$ of d.

2. We have that

$$\begin{aligned}\sqrt{T}\hat{g}_T(d_0) &= \sqrt{T}(\hat{\varrho} - \varrho_T(d_0)) \\ &= \sqrt{T}(\hat{\varrho} - \varrho(d_0) - B_T(d_0))\end{aligned}$$

$$= \sqrt{T}(\hat{\varrho} - \varrho(d_0)) - \sqrt{T}B_T(d_0). \tag{H.1}$$

The fist term in Equation (H.1) converges to $N(0, C)$ when $d < 0.25$ as show in Hosking (1996), C being the asymptotic covariance matrix of $\hat{\varrho}$. For the second term, $B_T(d_0)$, we will first write each element of vector $B_T(d_0)$ as

$$\begin{aligned}\sqrt{T}B_T(d_0)_k &= \sqrt{T}\rho_k(d_0) - \sqrt{T}\frac{\rho_k(d_0) + B^{\rho}_{T,k}}{1 + B^{\rho}_{T,0}} \\ &= \frac{\rho_k(d_0)\sqrt{T}B^{\rho}_{T,0} - \sqrt{T}B^{\rho}_{T,k}}{1 + B^{\rho}_{T,0}}.\end{aligned} \tag{H.2}$$

We have already seen that the denominator in (H.2) converges to one. Regarding the numerator, $B^{\rho}_{T,0}$ and $B^{\rho}_{T,k}$ stand for the bias of the autocovariance (for $\gamma_0 = 1$) for lags 0 and k. Hosking (1996) shows that the bias of the sample autocovariance decays at the rate T^{2d-1}, (if $E(a_t^4) < \infty$). Therefore, for $d < 0.25$, $\rho_k(d_0)\sqrt{T}B^{\rho}_{T,0} - \sqrt{T}B^{\rho}_{T,k} \to 0$.

3. The derivative of $\hat{g}_T(d)$ is given by

$$\begin{aligned}\nabla_d\hat{g}_T(d) &= \nabla_d\hat{\varrho} - \nabla_d\varrho_T(d) \\ &= 0 - \nabla_d\varrho_T(d).\end{aligned}$$

The elements of vector $\varrho_T(d)$ are approximations of the expectancies of $\hat{\varrho}$ given in (4.1). And therefore,

$$\nabla_d\rho_{T,k}(d) = \frac{[\nabla_d\rho_k(d) - \nabla_d B^{\rho}_{T,k}(d)][1 - B^{\rho}_{T,0}(d)] - [\rho_k(d) - B^{\rho}_{T,k}(d)][-\nabla_d B^{\rho}_{T,0}(d)]}{(1 - B^{\rho}_{T,0}(d))^2}. \tag{H.3}$$

It has already been shown in the proof of Proposition 2 that $B^{\rho}_{T,k}(d)$ converges uniformly to 0. Thus we only need to analyse the behaviour of the derivatives of $B_{T,k}(d)$. First of all, we investigate the behaviour of $\nabla_d \rho_k(d)$. We can write $\nabla_d \rho_k(d)$ as

$$\nabla_d \rho_k(d) = \nabla_d(\log|\rho_k(d)|)\rho_k(d)$$

for any $d \neq 0$, where $\rho_k(d)$ is given in (3.6). Thus,

$$\rho_k'(d) = \frac{1}{(1-d)^2} \prod_{i=2}^{k} \frac{i-1+d}{i-d}.$$

At $d = 0$, $\nabla_d \rho_k(d) = 1/k$, which obviously converges to zero as $k \to \infty$. For $d \neq 0$ note that

$$\log|\rho_k(d)| = \log|d| + \sum_{i=2}^{k} \log(i-1+d) - \sum_{i=1}^{k} \log(i-d).$$

Consequently,

$$(\log|\rho_k(d)|)' = \frac{\text{sgn}(d)}{|d|} + \sum_{i=2}^{k} \frac{1}{i-1+d} + \sum_{i=1}^{k} \frac{1}{i-d},$$

where sgn(d) is the sign of d. Therefore,

$$\begin{aligned}
\rho_k'(d) &= \left(\frac{\text{sgn}(d)}{|d|} + \sum_{i=2}^{k} \frac{1}{i-1+d} + \sum_{i=1}^{k} \frac{1}{i-d} \right) \prod_{i=1}^{k} \frac{i-1+d}{i-d} \\
&= \frac{1}{1-d} \prod_{i=2}^{k} \frac{i-1+d}{i-d} + \left(\sum_{i=2}^{k} \frac{1}{i-1+d} + \sum_{i=1}^{k} \frac{1}{i-d} \right) \prod_{i=1}^{k} \frac{i-1+d}{i-d}.
\end{aligned}$$

Let d_{sup} and d_{inf} be respectively the supremum and infimum of Λ, the parameter space. Then, an upper bound for the absolute value of $\rho_k'(d)$ is given by

$$\text{UB}(\rho_k'(d)) = \frac{1}{1-d_{sup}} \prod_{i=2}^{k} \frac{i-1+d_{sup}}{i-d_{sup}} + \left(\sum_{i=2}^{k} \frac{1}{i-1+d_{inf}} + \sum_{i=1}^{k} \frac{1}{i-d_{sup}} \right) \prod_{i=1}^{k} \frac{i-1+d_{sup}}{i-d_{sup}}.$$

Note that:

$$\prod_{i=2}^{k} \frac{i-1+d_{sup}}{i-d_{sup}} = O(k^{2d_{sup}-1}),$$

$$\sum_{i=2}^{k} \frac{1}{i-1+d_{inf}} = O(\log(k)),$$

$$\sum_{i=1}^{k} \frac{1}{i-d_{sup}} = O(\log(k)),$$

$$\prod_{i=1}^{k} \frac{i-1+d_{sup}}{i-d_{sup}} = O(k^{2d_{sup}-1}),$$

therefore UB($\rho'_k(d)$) $\to 0$ as $k \to \infty$. That is, $\rho'_k(d)$ converges uniformly to zero. If $\rho'_k(d)$ converges uniformly to zero, so does $B'_{T,k}(d)$, as it is a linear combination of $\rho'_1(d), ..., \rho'_{T-1}(d)$ that satisfies the properties of Lemma 2.

4. As the matrix W is by definition positive definite, it suffices to show that at least one element of $G(d_0) = \nabla_d \rho_k(d_0)$ is nonzero. One way to guarantee that is to use the first lag, $\rho_1(d)$ in vector ϱ as $\rho'_1(d) > 0$ for all values of d in the invertibility and stationarity regions. $\square$

Bibliography

Andel, J. (1986). Long memory time series models. *Kybernetika*, 22, 105-123.

Arnau, J., Bono, R. (2001). Autocorrelation and bias in short time-series: an alternative estimator. *Quality & Quantity*, 35, 365-387.

Bertelli, S., Caporin, M. (2002). A note on calculating autocovariances of long memory processes. *Journal of Time Series Analysis*, 23, 503-508.

Bisognin, C., Lopes, S. R. C. (2009). Properties of seasonal long memory processes. *Mathematical and Computer Modelling*, 49, 1837-1851.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, Inc. New York.

Box, G. E. P., Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, 2nd ed. San Francisco: Holden Day.

Brockwell, P. J., Davis, R. A. (1991). *Time series: Theory and Methods*, 2nd ed. New York: Springer-Verlag.

Cobb, G. W. (1978). The problem of the Nile: conditional solution to a change-point problem. *Biometrika*, 65, 243-251.

Fox, R., Taqqu, M., S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, 14, 2, 517-532.

Geweke, J., Porter-Hudak, S. (1983). The estimation and application of long memory time series model. *Journal of Time Series Analysis*, 4, 221-238.

Granger, C.M.G., Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1, 15-29.

Haslett, J., Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of Applied Statistics*, 38, 1-50.

Hassani, H., Leonenko, N., Patterson, K. (2012). The sample autocorrelation function and the detection of long-memory processes. *Physica A*, 391, 6367-6379.

Hosking, J. R. M. (1981). Fractional Differencing. *Biometrika*, 68, 1, 165-176.

Hosking, J. R. M. (1996). Asymptotic distribution of the sample mean, autocovariance and autocorrelation of long-memory time series. *Journal of Econometrics*, 73, 261-284.

Huitema, B. E., McKean, J. W. (1994). Two reduced-bias autocorrelation estimator: r_{F1} and r_{F2}. *Perceptual and Motor Skills*, 78, 323-330.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 16, 770-799.

Hurst, H. E. (1957). A suggested statistical model of time series that occur in nature. *Nature*, 180, 494.

Marriott, F. H. C., Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika*, 41, 390-402.

Mayoral, L. (2007). Minimum distance estimation of stationary and non-stationary ARFIMA processes. *Econometrics Journal*, 10, 124-148.

Newey, W. K., McFadden D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics IV*, North Holland, 2113-2241.

Palma, W. (2007). *Long-Memory Time Series.* Hoboken, New Jersey: Wiley.

Porter-Hudak, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of the Americal Statistical Association*, 84, 410: 338-344.

Priestley, M. B. (1981). *Spectral Analysis and Time Series.* London: Elsevier.

Rea, W., Oxley, L., Reale, M. and Brown, J. (2013). Not all estimators are born equal: The empirical properties of some estimators of long memory. *Mathematics and Computers in Simulation*, 93, 29-42.

Reisen, V. A. (1994). Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram. *Journal of Time Series Analysis*, 15, 335-350.

Shkolnisky, Y., Sigworth, F., Singer, A. (2008). A note on estimating autocovariance from short-time observations. *Technical Report.*

Tieslau, M. A., Schmidt, P. and Baillie, R. T. (1996). A minimum distance estimator for long-memory processes. *Journal of Econometrics*, 71, 249-264.

Zevallos, M., Palma, W. (2013). Minimum distance estimation of ARFIMA processes. *Computational Statistics and Data Analysis*, 58, 242-256.

Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis.* Hafner, New York.