ESSAYS ON
LONG MEMORY TIME SERIES
AND
FRACTIONAL COINTEGRATION

Submitted by

AMR SABER IBRAHIM ALGARHI

to the **University of Exeter**

as a thesis for the degree of

Doctor of Philosophy in

Economics

April 2013

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this dissertation which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

....

© Copyright 2013- Amr Saber Algarhi
All Rights Reserved

Table of Contents

To my Family

إلي عائلتي

Acknowledgements

First and foremost, I am profoundly grateful to my dissertation advisors, Professor James E. H. Davidson and Dr Andreea Halunga, for providing intellectual guidance and support throughout my studies, and for always being willing to read and comment on several drafts of this dissertation. During my studies in University of Exeter, I have worked with Dr Carlos Cortinhas, Dr Paulo M.D.C. Parente and Juliette Stephenson in teaching undergraduate tutorials. I am grateful for having had this opportunity to collaborate. I would like to thank the Business School Staff and the students in Exeter for generating a very pleasant environment.

In addition, my gratitude goes to all those who taught me Economics. In particular, I thank the following:

Abla El-Khawaga, Cairo University
Heba Nassar, Cairo University
Mahmoud Mohieldin, World Bank
Tarek Abdelfattah Moursi, Cairo University
Alaa El-Shazly, Cairo University
Ahmed Kamaly, American University in Cairo
Tarek Selim, American University in Cairo
Rodrigo Seda, USAID
John W. Salevurakis, American University in Cairo
Simon Wren-Lewis, Oxford University
Christos Kotsogiannis, University of Exeter
Tatiana Kirsanova, University of Glasgow

Moreover, special thanks go to Mohamed Ibrahim Saleh, Shady Elhady, Mohamed Khairy, Mahmoud Marei Abdelaziz, Mustafa Aboeldahab, Mohamed Shoukry and Panagiotis Anastasiou for their loyal friendship and all the unforgettable times we have spent together; thank you for giving me the honour of being your friend. Very special thanks go to, Ahmed Zaki and Alaa Zaki, my cousins and friends.

Last but not least, I would like, of course, to thank my family: my sisters Aliaa and May- only when you are far away you realise what family really means, my mother who made it possible for me to concentrate on my studies while still feeling that our family remained family and my father, my first teacher, who made it possible for me to follow the right path in life and who is, and always has been my rock. My parents supported me all my life. I am very lucky to have parents like them.

AMR SABER ALGARHI
Exeter, United Kingdom, April 2013

The viva voce examination of this dissertation was held on 20 June 2013. I would like to take this opportunity to thank the members of the board of examiners, Professor Javier Hidalgo (London School of Economics) and Dr Paulo M.D.C. Parente (University of Exeter) for their constructive and helpful comments and suggestions.

AMR SABER ALGARHI
Exeter, United Kingdom, August 2013

List of Tables

List of Figures

Introductory Chapter

In the modelling of macroeconomic and financial data, high persistence over sufficiently long periods of time and dependence structure across time play an important role. Such data (series) is reported to be characterised by distinct, but non periodic, cyclical patterns and behaves such that current values are not only influenced by immediate past values but values from pervious time periods. Increasing interest has been shown in capturing such phenomenon referred to as ‘long memory’. A long memory property can be expressed in many ways (discussed in chapter 1). It can be defined by the series autocovariances with a hyperbolic (slow) decay and hence not summable, unlike in the short memory time series where autocovariances decay exponentially (rapidly); or by a spectral density that is unbounded and characterised by a pole at the origin, whereas in the short memory series the spectral density is bounded and very smooth. Moreover, long memory time series include both stationary and nonstationary where in the latter the dependence falls off even more slowly over time. The first attempt in economics to parameterise the long memory feature, represented in a memory parameter d that govern the long run dynamics of the series, was developed by Granger (1980) and Granger and Joyeux (1980) where the concept of long memory in a time series was coupled with ‘fractional integration’ to provide a theoretical explanation for the hyperbolic (slow) decay of sample correlograms in certain empirical contexts.

The term fractional came up to refer to a generalised operation of a non integer order, where the conventional and usual order of the operation has to be an integer. For instance, some series levels in macroeconomics change with time in a smooth way, such smooth movements were explained either by assuming the series to be stationary $I(0)$ around a deterministic trend or by the means of stochastic unit roots $I(1)$ models. At the beginning, the interest in long memory feature was through the focus on the unit root $I(1)$ models as a special case of a more general case of nonstationary fractional series where the memory parameter is greater and equal to half. Despite the simplicity and usefulness of the unit root $I(1)$ models by assuming a known degree of memory which can be reduced by differencing to a stationary and invertible short memory time series, the relaxation of such assumption to a more flexible and fractional memory measure that is unknown and estimable from the data would lead to a valid statistical inference, even in large samples. Such fractional models can also form a convenient bridge from short memory stationary to nonstationary models. Although the immense interest was in the fractional non stationary models, the main development and focus in literature was on estimating the memory parameter in the stationary case where the memory parameter is between zero and half.

Consider a stochastic process, x_t is written in a simple generalization of a random walk,

$$x_t = (1-L)^{-d}\varepsilon_t,$$

and the process $\varepsilon_t \in I(0)$, i.e. an *i.i.d.* zero mean and finite variance time series. For $d = 0$, $x_t = \varepsilon_t$ and thus x_t is weakly autocorrelated. When $d > 0$, x_t is strongly dependent and referred to as persistent, because of the strong association between observations widely separated in time, while when $d < 0$, x_t is negatively dependent and referred to as anti-persistent. The term $(1-L)^{-d}$ can be defined by the binomial expansion and only exists for $-\frac{1}{2} < d < \frac{1}{2}$,

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j,$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the gamma function and L ($Lx_t = x_{t-1}$) is the lag operator. The autocorrelation function (ACF) of x_t is expressed using Stirling's formula,

$$\rho_k \sim C k^{2d-1} \text{ as } k \to \infty$$

where C is a strictly positive constant. Table 1 shows a summary of time series characteristics related to the long memory parameter d.

Table 1: Long memory characteristics

	Series	Memory	ACF
$d \in (-0.5, 0)$	stationary fractional	anti-persistent	hyperbolic
$d = 0$	stationary	short memory	exponential
$d \in (0, 0.5)$	stationary fractional	long memory	hyperbolic
$d \in [0.5, 1)$	non-stationary fractional	long memory	hyperbolic
$d = 1$	unit root	infinite memory	linear

If two or more time series exhibit long memory, then cointegrating relationships among them, if existed, may also exhibit long memory, implying relatively slow adjustments to re-establish equilibrium, hence the term fractional cointegration was developed. Engle and Granger (1987) defined the fractional cointegration along with the standard cointegration. A collection of two or more time series is fractionally cointegrated if their linear combination exists and less persistent (measured in terms of the memory parameters) than any of the individual series. In standard cointegration analysis, all variables are either $I(1)$ and the linear combinations are $I(0)$ process or alternatively they are $I(0)$. In other words, standard cointegration allows only integer values for the memory parameter and the equilibrium errors to be a $I(0)$ process which are rather restrictive. However, there is no integer constraint and no assumption on the memory parameters of the individual series or their linear combination in the fractional cointegration analysis. In this dissertation, a selective survey of the literature on representation and estimation in fractional cointegration is presented in chapters 2 and 3 for the stationary case and in chapter 4 for the nonstationary case.

The dissertation considers an indirect approach for the estimation of the cointegrating parameters, in the sense that the estimators are jointly constructed along with estimating other nuisance parameters. This approach was proposed by Robinson (2008) where a bivariate local Whittle estimator was developed to jointly estimate a cointegrating parameter along with the memory parameters and the phase parameters (discussed in chapter 2). The main contributions of this dissertation is to establish, similar to Robinson (2008), a joint estimation of the memory, cointegrating and phase parameters in stationary and nonstationary fractionally cointegrated models in a multivariate framework. In order to accomplish such task, a general shape of the spectral density matrix, first noted in Davidson and Hashimzade (2008), is utilised to cover multivariate jointly dependent stationary long memory time series allowing more than one cointegrating relation (discussed in chapter 3). Consequently, the notion of the extended discrete Fourier transform is adopted based on the work of Phillips (1999) to allow for the multivariate estimation to cover the non stationary region (explained in chapter 4). Overall, the estimation methods adopted in this dissertation follows the semiparametric approach, in that the spectral density is only specified in a neighbourhood of zero frequency.

The dissertation is organised in four self-contained chapters that are connected to each other, in additional to this introductory chapter:

- Chapter 1 discusses the univariate long memory time series analysis covering different definitions, models and estimation methods. Consequently, parametric and semiparametric estimation methods were applied to a univariate series of the daily Egyptian stock returns to examine the presence of long memory properties. The results show strong and significant evidence of long memory in the Egyptian stock market which refutes the hypothesis of market efficiency.
- Chapter 2 expands the analysis in the first chapter using a bivariate framework first introduced by Robinson (2008) for long memory time series in stationary system. The bivariate model presents four unknown parameters, including two memory parameters, a phase parameter and a cointegration parameter, which are jointly estimated. The estimation analysis is applied to a bivariate framework includes the US and Canada inflation rates where a linear combination between the US and Canada inflation rates that has a long memory less than the two individual series has been detected.
- Chapter 3 introduces a semiparametric local Whittle (LW) estimator for a general multivariate stationary fractional cointegration using a general shape of the spectral density matrix first introduced by Davidson and Hashimzade (2008). The proposed estimator is used to jointly estimate the memory parameters along with the cointegrating and phase parameters. The consistency and asymptotic normality of the proposed estimator is proved. In addition, a Monte Carlo study is conducted to examine the performance of the new proposed estimator for different sample sizes. The multivariate local whittle estimation analysis is applied to three different relevant examples to examine the presence of fractional cointegration relationships.
- In the first three chapters, the estimation procedures focused on the stationary case where the memory parameter is between zero and half. On the other hand, the analysis in chapter 4, which is a natural progress to that in chapter 3, adjusts the estimation procedures in order to cover the non-stationary values of the memory parameters. Chapter 4 expands the analysis in chapter 3 using the

extended discrete Fourier transform and periodogram to extend the local Whittle estimation to non stationary multivariate systems. As a result, the new extended local Whittle (XLW) estimator can be applied throughout the stationary and non stationary zones. The XLW estimator is identical to the LW estimator in the stationary region, introduced in chapter 3. Application to a trivariate series of US money aggregates is employed.

CHAPTER 1

LONG MEMORY IN THE EGYPTIAN STOCK MARKET RETURNS

AMR ALGARHI*
UNIVERSITY OF EXETER

Abstract

This chapter examines the presence of long memory in the daily returns of the Egyptian stock market, using parametric and semiparametric methods. Both techniques have their merits and demerits. Accordingly, the Exact Maximum Likelihood (EML) estimation is employed to estimate the ARFIMA model in the time domain; while two main semiparametric techniques, log periodogram (LP) and local Whittle (LW), were applied to estimate the memory parameter in the frequency domain. Unlike the findings for developed equity markets, the results show strong and significant evidence of long memory in the Egyptian stock returns, which refutes the hypothesis of market efficiency. As a result the Egyptian stock returns can be predicted using historical information. The findings of this paper are helpful to regulators, financial managers and investors dealing in the Egyptian stock market.

JEL Classification: C14, C22.

Keywords: ARFIMA; Egyptian stock market; Exact maximum likelihood; Local Whittle estimation; Log-periodogram regression; Long memory; Semiparametric estimation.

---

* Amr Algarhi: a.s.i.algarhi@exeter.ac.uk
I would like to thank Professor James Davidson at University of Exeter for supplying the TSM software. Empirical applications were performed in OxMetrics 6 and TSM 4.32. All remaining errors are mine.

1.1 Introduction

This chapter uses parametric and semiparametric methods to estimate the long memory parameter for the Egyptian stock market returns. A time series has a long memory, whenever the dependence between apart events diminishes very slowly as the number of lags increases. The presence of long memory properties in asset returns has important implications for asset pricing models. Such features can be used to construct a profitable trading strategy. In another words, long memory entails that perfect arbitrage is impossible and contradicts standard derivative pricing models based on Brownian and martingale assumptions (Mandelbrot, 1963).

In the last two decades, it has been of great importance, theoretically and empirically, to study the properties of long memory in financial asset returns which is used as a proxy for analysing market efficiency. The stock market returns are said to exhibit long memory properties, if there is a significant autocorrelation (dependence) between observations widely separate in time. This dependence between apart observations can be utilised to predict future returns, leading to the possibility of consistent speculative profits. Consequently, the existence of long memory in the return series refutes the weak form of the market efficiency hypothesis. The price of an asset determined in an efficient market should follow a martingale process in which each price change is unaffected by its predecessor and has no memory. Therefore, if the returns series display significant autocorrelation between distant observations then past returns can help to predict future returns, thus violating the market efficiency hypothesis which states that, asset prices incorporate all relevant information, where future asset returns are unpredictable, conditioning on past returns.

The development of statistical long memory processes was inspired by Hurst (1951) who was the first to introduce a method for the quantifying of the long memory called rescaled range analysis (R/S). This method involves parameter estimation to capture the scaling behaviour of the range of partial sums of the variable under consideration. Using the rescaled range analysis, Mandelbrot (1971) has found evidence of long memory in the stock returns. However, Lo (1991) pointed out the lack of robustness of the statistical R/S test in the presence of short term memory and heteroskedasticity. Lo (1991) suggested a modified R/S test and tested for long memory in daily US stock market indices and found no evidence of long range dependence. Mills (1993) found weak evidence of long memory in a sample of monthly UK stock returns. Cheung and Lai (1995) provided little evidence of long memory in the Morgan Stanley Capital International stock index data. Ding and Granger (1996) reported evidence of long memory for S&P 500 returns, while Lobato and Savin (1997) saw no evidence of long memory in daily S&P 500 returns over the period July 1962 to December 1994. The majority of the above studies have employed either parametric or semiparametric methods to test and estimate the long memory property. For the parametric method, a complete parametric model to express the autocovariance function as a parametric function of the parameters, d, is built, such as ARFIMA model. In contrast, the semiparametric method is only interested in the memory parameter d and does not require the modelling of a complete set of the autocovariances. In the main, both parametric and semiparametric techniques have their merits and demerits. Estimation of fully parametric long memory models is computationally expensive and is subject to misspecification; hence the correct choice of the model is important. On the other hand, the semiparametric estimation considers d as the main parameter of interest. The SPE derives robust estimators since it avoids difficulties over the specification of the short run ARMA parameters; however, the idea of explaining the entire autocorrelation structure by a single parameter d is highly restrictive. This study employs parametric and semiparametric methods to estimate the long range dependence.

Compared to the world's well-developed financial markets (the U.S. markets), the presence of long memory in emerging capital markets in developing economies has received little attention. Nevertheless, there are various conditions and reasons that contribute to a different dynamics regarding returns in emerging stock markets. Emerging markets are typically much smaller, less liquid, and more volatile than well known world financial markets. Emerging markets may be less informationally efficient. This could be due to several factors such as poor-quality (low precision) information, high trading costs, and less competition due to international investment barriers. Furthermore, the industrial organization found in emerging economies is often quite different from that in developed economies. As a result, it is very important to study the emerging securities markets and the complete characterization of the dynamic behaviour underlying stock returns in these developing economies, in order to attract investors and investment funds seeking to diversify their assets.

The objective of this chapter is to examine the presence of long memory in the Egyptian stock returns using parametric and semiparametric methods. The Daily EGX30 price index is considered as a proxy for the Egyptian stock market. There has been very limited research on the behaviour of stocks traded on the Egyptian stock exchange, although the capital market in Egypt is apt to exhibit different characteristics from those observed in developed capital markets. Biases due to market thinness and nonsynchronous, trading should be expected to be more severe in the case of the Egyptian stock market. The Egyptian stock market is not expected to be highly efficient in terms of the speed of information reaching traders compared to the developed capital markets. Furthermore, traders and investors in the Egyptian stock market tend to react slowly and gradually to new information. The existence of long memory will have significant implications in the Egyptian stock market, where future returns can be predicted from past returns, thus violating the market efficiency hypothesis.

This chapter is organised as follows. The next section provides an overview of the theoretical and relevant literature review of long memory. Section 1.3 covers the parametric and semiparametrics methods used to estimate the long memory parameter. Section 1.4 describes the data and reports the results of the empirical application to the daily Egyptian stock market and finally section 1.5 offers some concluding remarks.

1.2 Literature Review

1.2.1 Background

Long memory, or **LM**, processes were initially documented in non-economic literature, with interest starting from the empirical examination of data in physical science since at least 1950s. The famous British hydrologist Harold Edwin Hurst (1951), during the engineering of the high Aswan dam, developed an analysis to determine if the yearly flows and inflows into reservoirs of the Nile were random or clustered from year to year using long reliable historical data for the years 622- 1281 recorded at the Roda gauge in Cairo. He determined they were not random and found evidence of dependence over long intervals of time, with stretches when floods are high tending to be above the mean and others when they are low tending to be below the mean. As a result, the data was found to show several cycles; however, these cycles did not exhibit periodicity. Mandelbrot and Wallis (1968) called this behaviour the *Joseph effect*[1] in reference to the biblical seven good years of abundance and seven bad years of famine[2]. Hurst (1951, 1956 and 1957) also examined 900 records of other natural phenomena (for example, annual river levels, rainfall, temperature and pressure records, tree rings, and sunspot activity) finding non-random positive correlations in most of them (Baillie, 1996).

Since then, LM processes have been investigated by many researchers from very different fields; and because of the diversity of its applications, its literature is generally spread over a large number of journals including Agronomy, Astronomy, Chemistry, Climatology, Engineering[3], Geo-science, Hydrology, Mathematics, Physics and Statistics. Examples of these are presented in, *inter alia*, Hurst (1951, 1956 and 1957), Lawrance and Kottegoda (1977) and Painter (1998) in geophysical data, Mandelbrot and Wallis (1968), Mandelbrot (1972), Hipel and McLeod (1978a, 1978b and 1978c), Bloomfield (1992), Seater (1993) and Kirk-Davidoff and Varotsos (2006) in climatology.

As the data in natural sciences demonstrate a preference towards LM and the source of uncertainty in Economics can be considered as natural phenomena, then we can expect LM to be found in economic data. The importance of LM in economic data was recognised in Mandelbrot (1963), Adelman (1965) and particularly, in Granger (1966), who noticed that for economic time series, the typical shape of the spectral density is a function with a pole at the origin that then decays monotonically at high frequencies. It was not until 1980 that LM models were used by Econometricians and

[1] The *Joseph effect* involves long stretches of time when the process tends to be above the mean, and long stretches of time when the process tends to be below the mean.
[2] In the Bible (Genesis 41, 29-30): “Seven years of great abundance are coming throughout the land of Egypt, but seven years of famine will follow them”, and the same phenomenon was also mentioned in the Koran (Joseph 12, 47-48): “He [Joseph] said, What you cultivate during the next seven years, when the time of harvest comes, leave the grains in their spikes, except for what you eat. After that, seven years of drought will come; this will consume most of what you stored for them”. However, there are no records of the water level of the Nile from those times.
[3] In signal and image processing.

by Financial Researchers *circa* 1995. Granger and Joyeux (1980) propose the use of the fractional differencing to model LM which is related to earlier work by Mandelbrot and Van Ness (1968) describing fractional Brownian motion.

There is substantial evidence that LM models can be successfully applied to time series data in both Macroeconomics and Financial Economics data; for example, real national output measures, inflation rates, exchange rates, interest rate differentials, stock prices, commodity prices, market indices and forward premiums. These time series show evidence of being neither $I(0)$ nor $I(1)$. When first differenced, those series appear as being over-differenced. This feature is typical of long memory processes. LM processes has also been used in modelling the volatility of asset prices and power transformations of returns. Investigations for LM in real output measures were first studied in Diebold and Rudebusch (1989) and Haubrich and Lo (1989). Baillie, Chung, and Tieslau (1992) and Hassler and Wolters (1995) apply fractionally integrated $ARMA$, or $ARFIMA$, models to describe the fluctuations of the inflation rates. They provided empirical evidence in favour of LM models. Baillie et al. (1992) examined the relationship between the mean and the variability of inflation rates by means of $ARFIMA - GARCH$ models for 10 countries using monthly observations from 1948 to 1990. Baillie, Bollerslev and Mikkelson (1996) find LM in the volatility of the Deutsche Mark- U.S. Dollar (DM-USD) exchange rate. Long range dependence, in asset price series, was reviewed by Brock and De Lima (1996); yet, LM seems much more likely in asset volatility than in asset returns themselves. LM was found in the Deutshcer Aktien IndeX (DAX, the German stock index) by Lux (1996). In addition, some research demonstrates the existence of LM in smaller and less developed markets. Tolvi (2003) examined the Finnish stock market; Madhusoodanan (1998) provides evidence on the individual stocks in the Indian Stock Exchange. Barkoulas and Baum (2000) give similar evidence on the Greek financial market; while Cavalcante et al (2002) demonstrate LM in Brazil stock market.

Finally, the monographs by Beran (1994), Robinson (2003), Palma (2007) and Samorodnitsky (2007) provide an excellent introduction to LM processes. Additionally, several survey-type articles on LM have been written, for example

Taqqu (1986), Hampel (1987), Beran (1992), Robinson (1994), Baillie (1996), Guégan (2005) and Banerjee and Urga (2005). Recently, research on LM is growing significantly leaving some of these surveys very out-of-date.

1.2.2 Defining Long Memory

The terms long memory, long-range dependence, strong dependence or persistence can be used interchangeably. LM can be defined in several ways. Traditionally, LM has been specified in the time domain in terms of long lag autocovariance, or in the frequency domain in terms of explosion of low frequency spectra. Given a stationary time series process $\{y_t\}$ with an autocovariance function $\gamma(j) = Cov(y_t, y_{t+j})$ at lag j that does not depend on t, then the process has LM if,

$$\gamma(j) \sim c_\gamma j^{2d-1} \text{ , as } j \to \infty \qquad (1.2.2.1)$$

for $0 < d < \frac{1}{2}$, where d is the memory (differencing) parameter, or the fractional difference parameter. The constant c_γ is finite positive $(0 < c_\rho < \infty)$, and the notation "~" means that the ratio of the left and right sides tend to one for large j. The intuition interpretation for this definition is that the dependence between apart events diminishes very slowly as the number of lags increases (tends to infinity) often called a hyperbolic decay. On the contrary, short-range dependence is characterised by quickly decaying correlations at an exponential rate to zero (e.g. $ARMA$ and Markov processes). The asymptotic behaviour in (1.2.2.1) indicates that the autocovariance decreases very slowly with long lags, or in other words the autocovariances are not summable so that,

$$\lim_{n\to\infty} \sum_{j=-n}^{n} \gamma(j) = \infty \qquad (1.2.2.2)$$

On the other hand, LM can be described in the frequency domain using the spectral density structure. It is interesting to see how long-range dependence, or LRD, translates from the time domain to the frequency domain. Suppose that $\{y_t\}$ has absolutely continuous spectral density function, then it has a spectral density $f(\lambda)$ that is

$$f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-i\lambda j} \text{ , } -\pi \le \lambda \le \pi \qquad (1.2.2.3)$$

where $f(\lambda)$ is a non-negative, even function, periodic of period 2π when extended beyond the Nyqvist[4] range $[-\pi, \pi]$. LM in the time domain is expected to be translated into the behaviour of the spectral density around the origin because low frequencies (frequencies around the origin) account for big lags in the time domain. A process with spectral density f is defined to exhibit long memory if,

$$f(\lambda) \sim c_f |\lambda|^{-2d}, \text{ as } \lambda \to 0 \tag{1.2.2.4}$$

Both definitions, in the time and frequency domain, are not equivalent but connected (Beran, 1994a and Taqqu, 1986). The spectral density in (1.2.2.4) implies that the spectral density will be unbounded at low frequencies. Hence long range dependence corresponds to the blow-up of the spectral density $f(\lambda)$ at the origin so that it has a pole at frequency zero,

$$f(0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) = \infty \tag{1.2.2.5}$$

Thus if $\gamma(j)$ behaves like a power function at infinity, so does $f(\lambda)$ at zero and the relation can be remembered by the "add one, change sign" rule, where the exponent $2d - 1$, is the asymptotic behaviour of $\gamma(j)$.

1.2.3 Long Memory Models

During the World War II, a massive momentum in time series research has evolved as a result of advances in many engineering applications, including spectral analysis and radio signals. Afterwards, a flexible group of models, known as $ARMA$, also called short-range dependent models involving correlation functions that decrease exponentially fast over time, was developed in the time domain. Although short-memory models were used widely, by economists, these models had a number of shortcomings and could not be applied to all fields. Some data seemed to require models, whose correlation functions would decay much less quickly.

Kolmogorov (1940) discovered the fractional Brownian motion, or fBm, which was used along with its increments by Mandelbrot to generate long-range dependence. The characteristic of LRD in economic and financial data is lately described by a number of models. This includes the fractional differencing model, the autoregressive fractionally integrated moving average models ($ARFIMA$) and fractional

[4] Nyqvist range is named after the Swedish-American Engineer Harry Nyqvist (1889-1976).

cointegration models. Among these models, the focus would be on the $ARFIMA(p,d,q)$ models introduced by Granger and Joyeux (1980).

In 1971, Box and Jenkins introduced the $ARIMA(p,d,q)$ model,

$$\Phi(L)(1-L)^d y_t = \Psi(L)\varepsilon_t \qquad (1.2.3.1)$$

where d is an integer, $\Phi(L)$ and $\Psi(L)$ are the polynomials $\Phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j$ and $\Psi(L) = 1 + \sum_{j=1}^{q} \psi_j L^j$ involving autoregressive and moving average coefficients of order p and q respectively and ε_t is a white noise process. To ensure the stationarity and inverstibility conditions, the roots of $\Phi(L)$ and $\Psi(L)$ must lie outside the unit circle. Granger and Joyeux (1980) managed to extend the set of $ARIMA$ models by considering instead fractional $d \in (-0.5, 0.5)$ in (1.2.3.1) which introduces a fractional autoregressive integrated moving average model orders p, d and q, or $ARFIMA(p,d,q)$ or $FARIMA(p,d,q)$. It has spectral density,

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|1-e^{i\lambda}\right|^{-2d}\left|\frac{\phi(e^{i\lambda})}{\psi(e^{i\lambda})}\right|^2, \; -\pi \leq \lambda \leq \pi \qquad (1.2.3.2)$$

A fractional white noise process is a particular case which is equivalent to an $ARFIMA(0,d,0)$ process. $ARFIMA$ processes are covariance stationary for $-0.5 < d < 0.5$, mean reverting for $d < 1$ and weakly correlated for $d = 0$. For $d > 0.5$ these processes have infinite variance. For $d \geq 0.5$ the processes have infinite variance but in the literature it is more usual to impose initial value conditions so that y_t has changing, but finite, variance. Granger and Joyeux (1980) and Hosking (1981) considered $ARFIMA(0,d,0)$ and $ARFIMA(1,d,0)$ respectively, which based on Adenstedt's (1974) model. Further information on $ARFIMA(p,d,q)$ models was give by Sowell (1992), Chung (1994) and others.

1.2.4 Estimation Methods

One of the main interests in the literature of long memory is to estimate the unknown parameter d that describes the long memory properties or the low frequency behaviour of the spectral density function $f(\lambda)$. There are two main groups of estimation methods used to test for LM: the parametric estimation ($\boldsymbol{PE}$) and the semi-parametric estimation ($\boldsymbol{SPE}$). For the parametric estimation, a complete parametric model that expresses $\gamma(j)$ for all j, or the spectral density function $f(\lambda)$ for all λ, as a parameteric function of the parameters, d and unknown scale factors, is built, such as $ARFIMA$ model. In contrast, the semi-parametric estimation is only interested in the memory parameter d and do not require the modelling of a complete set of the autocovariances.

In the main, each method has its merits and demerits. Estimation of fully parametric long memory models is computationally expensive, especially in the time domain. Additionally, parametric methods are subject to misspecification. Under- or over-specification of the autoregressive and moving average orders p and q, which describes the short range dependent component of y_t, can lead to invalidation of the statistical properties and can dangerously bias the estimation of d. On the other hand, SPE considers d as the main parameter of interest. *SPE* derives robust estimators since it avoids difficulties over the specification of the short run *ARMA* parameters; however, the idea of explaining the entire autocorrelation structure by a single parameter d is highly restrictive.

1.**2.4.1 Parametric Estimation**

Several joint estimation methods of the unknown parameters in the $ARFIMA(p, d, q)$ model in equation (1.2.3.1) were considered. If y_t is assumed Gaussian process, then the log-likelihood function is,

$$\mathcal{L}(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Omega| - \frac{1}{2}y'\Omega^{-1}y \qquad (1.2.4.1.1)$$

The Gaussian maximum likelihood estimate, or MLE, is obtained by maximising $\mathcal{L}(\theta)$ and might be expected to have optimal asymptotic statistical properties. The log likelihood function requires the calculation of the determinant and the inverse of the variance-covariance matrix Ω. These calculations can be done by means of several procedures, for example, Cholesky decomposition method, Durbin-Levinson algorithm and state space techniques. Furthermore, Sowell (1990, 1992) derives the exact MLE of the $ARFIMA$ process with unconditional normally distributed error terms. Sowell's estimator performs poorly if the model is misspecified, like all maximum likelihood. Baillie and Chung (1993) developed a conditional sum of squares estimator in the time domain and show that it performs similarly to Sowell's estimator for the $ARFIMA(0, d, 0)$ model. Computer programs for the exact MLE were developed by Doornik and Ooms (2003, 2004), who showed, by building on the work by Sowell (1992), that the exact MLE can be efficiently estimated with storage of order n and computation of order n^2. Their ML approach is applied in the Arfima package and therefore also in PcGive (See, Doornik and Hendry (2001)).

The calculation of the exact MLE is complicated and computationally demanding. As a result, alternative procedures have been considered to replace the exact MLE. The use of approximations to Gaussian ML was developed to speed up the calculation of parameter estimates, without affecting the first order limit distributional behaviour. Estimates maximising such approximations are called Whittle estimates due to Whittle (1951) described in detail in Beran (1994a). Whittle estimates are all $\sqrt{n}$ −consistent and have the same limit normal distribution as the Gaussian MLE. One type of Whittle estimates is the discrete form described in the frequency domain. Suppose the parametric spectral density is $f(\lambda;\theta,\sigma^2)$, where θ is an r −dimensional unknown parameter vector and σ^2 is a scalar. Under $ARFIMA(p,d,p)$ specification, the vector θ is an estimate of the autoregressive, moving average and long memory coefficients. Under a fractional white noise specification, θ reduces to the long memory parameter d. Now suppose the periodogram,

$$I(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t e^{it\lambda}\right|^2 \tag{1.2.4.1.2}$$

where $\lambda_j = \frac{2\pi j}{n}$ are the Fourier frequencies. The discrete frequency Whittle estimate, mentioned in Hannan (1973), minimised the Whittle objective function,

$$\mathcal{L}_W(\theta) = \sum_{j=1}^{n-1}\left[\log f(\lambda_j;\theta) + \frac{I(\lambda)}{f(\lambda_j;\theta)}\right] \tag{1.2.4.1.3}$$

This form of Whittle estimation has many advantages. One of these is that it is based on the rapid calculation of the periodogram by means of the fast Fourier transform (FFT), even when n is large. Another advantage of Whittle estimate is that their limit distribution is unchanged by many departures from Gaussianity, which means that the same rules of statistical inference can be used without worrying about Gaussianity. Thus the same relatively convenient rules of statistical inference can be used without worrying too much about the question of Gausiannity.

Alternative parametric methods are available but they are less efficient than Whittle estimation when y_t is Gaussian. For example; generalised method of moments, or GMM, has been used to estimate LM models, in both time and frequency domains. However, GMM estimates are not only less efficient in the Gaussian case than the

Whittle estimates, but GMM is also computationally less attractive than the Whittle estimation in (1.2.4.1.3). Beran (1994b) proposed M-estimators for Gaussian long-memory models, while Pai and Ravishanker (1998) use Bayesian analysis to detect changing parameters in $ARIMA$ processes. Hauser (1998) has compared the various maximum likelihood estimators on samples of size 100 using Monte Carlo methods. He concludes that the Whittle estimator with tapered data is most reliable. Hauser, Pötscher and Reschenhofer (1999) are critical of $ARFIMA$ models for estimating persistence in aggregate output. They show that over-parameterisation of an $ARMA$ model may bias the estimates of persistence downwards. In general all the above methods were applied on the stationary LM models, the case $0 < d < 0.5$.

1.2.4.2 Semiparametric Estimation, or SPE

Semiparametric estimation methods were developed to overcome some of the difficulties found in the parametric methods. These methods include the log-periodogram (LP) regression and the local Whittle (LW) estimation, also known as the Gaussian semiparametric estimation. The LP regression is longer established and was the most widely used, but it is less efficient than the LW estimate. The semiparametric estimators of the LM parameter assume the spectral density model,

$$f(\lambda) \sim |\lambda|^{-2d} g(h), \text{ as } \lambda \to 0 \qquad (1.2.4.2.1)$$

where $g(\cdot)$ is an even function on the Nyqvist range $[-\pi, \pi]$ that determines the short run dynamics of the stationary process y_t and satisfying $0 < g(0) < \infty$.

LP regression estimator was first proposed by Geweke and Porter-Hudak (1983) and also known as GPH estimator. It is considered the first semiparametric estimation of the LM parameter, d in the frequency domain. GPH is based on the characteristic pattern of the periodogram around zero frequencies, which is first estimated from the series, and its logarithm is regressed on the logarithm of a trigonometric function of frequency. This method of estimation has been used comprehensively in macroeconomic and financial time series application because it is easy to implement even before development of any satisfactory theoretical analysis of its asymptotic distributional properties. The performance of this estimator, however, has several drawbacks; one of which concerns the number of values of the periodogram to be used in the regression. Geweke and Porter-Hudak (1983) proposed a heuristic based on the length of the time series. Agiakloglou, Newbold and Wohar (1992) shows that

GPH is biased in the presence of strongly autoregressive short memory and in addition does not possess satisfactory asymptotic properties.

Robinson (1995a) has further refined the GPH log-periodogram regression. Using the same notation as GPH, the estimator is based on the least-square regression using spectral ordinates $\lambda_1, \lambda_2, \ldots, \lambda_m$ from the periodogram of $y_t, I_y(\lambda_j)$, and $j = 1, 2, \ldots, m$, where m, a bandwidth or smoothing number, is less than n but is regarded as increasing slowly with n in asymptotic theory.

$$\log[I_y(\lambda_j)] = a + b\log(\lambda_j) + v_j \qquad (1.2.4.2.2)$$

where v_j is assumed to be $i.i.d.$ The least square estimator $\hat{b}$, which gives $\hat{d} = -\frac{1}{2}\hat{b}$, is asymptotically normal and the corresponding theoretical standard error is $\pi(24m)^{-\frac{1}{2}}$. This version is easier to use for actual computation. The value of the estimator $\hat{d}$ depends on the choice of truncation parameter m. Diebold and Inoue (2001) showed that the choice of a large value for m would result in reducing standard error at the expense of biasness in the estimator, as the relationship that the GPH regression is based on holds only at low frequencies. On the other hand, consistency requires that m grows with sample size, but at a slower rate. They adapt the rule of thumb of $m = \sqrt{T}$, where T is the number of observations. Additionally, Wright (2000) develops log-periodogram estimators with conditional heavy tails, while Henry (2001) introduced a periodogram spectral estimation for the case of long memory conditional heteroscedasticity.

There are a plethora of new *SPEs* of the long memory parameter that are more efficient and robust, for example, Kunsch (1987), Robinson (1994b), Lobato and Robinson (1996), Moulines and Soulier (1999), Phillips and Shimotsu (2006) and Phillips (2007). Nevertheless, the most widely used and preferred SPE is the local Whittle estimation proposed by Robinson (1995b), and was further investigated by Dalla, Giraitis, and Hidalgo (2004) and Phillips and Shimotsu (2006), where the objective function is a discrete form of an approximate frequency domain Gaussian likelihood, averaged over a neighbourhood of zero frequency,

$$L(d, C) = \sum_{j=1}^{m} \left[\log\left(C\lambda_j^{-2d}\right) + \frac{I(\lambda_j)}{C\lambda_j^{-2d}}\right] \qquad (1.2.4.2.3)$$

where m is a bandwidth (see, Kunsch (1987)). The argument requires m to be of smaller order than n. It is inadvisable to choose m too large as bias can then result. However the longer the series length n, the larger we can choose m, so that in very long series the extra robustness gained by the semiparameteric approach may be worthwhile. LW estimator is shown to be asymptotically normal and more efficient than previous estimators. SPE also includes data differing and data tapering methods. Phillips and Shimotsu (2004) propose variant of the local Whittle estimation procedure that does not rely on differencing or tapering and they further extend the range where the estimator of d has standard asymptotic.

1.3 Methodology

This section introduces the methodology used to estimate the long memory parameter using parametric and semiparametric methods. The parametric estimation, analysed in the time domain, is based on the likelihood function as mentioned before in section 1.2.4. The parametric estimation used in this paper is the Exact Maximum Likelihood (EML) estimation method developed by Sowell (1992). On the other hand, the semiparametric estimation for the memory parameter is based on frequency domain. Two semiparametric estimators will be considered in this section, the GPH log-periodogram regression and the local Whittle estimator.

1.3.1 The Exact Maximum Likelihood (EML)

Consider the following $ARFIMA(p,d,q)$ process,

$$\Phi(L)(1-L)^d y_t = \Psi(L)\varepsilon_t$$

where $\Phi(L)$ and $\Psi(L)$ are the polynomials

$$\Phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j$$

and

$$\Psi(L) = 1 + \sum_{j=1}^{q} \psi_j L^j$$

involving autoregressive and moving average coefficients of order p and q respectively and ε_t is a white noise process. Now assume $Y = (Y_1, \dots, Y_T)'$ follows a normal distribution with $Y \sim N(0, \Sigma)$. The EML procedure allows for simultaneous estimation of both the long memory parameter and ARMA parameters. The maximum likelihood objective function is expressed as,

$$l_E(\Phi, \Psi, d; Y) = -\frac{T}{2}\log|\Sigma| - \frac{1}{2}Y'\Sigma^{-1}Y \quad (1.3.1.1)$$

As a result, the EML estimator of d can be derived as,

$$\hat{d}_{EML} = \arg\max\left[-\frac{T}{2}\log|\Sigma| - \frac{1}{2}Y'\Sigma^{-1}Y\right] \quad (1.3.1.2)$$

This estimator can be inconsistent if the AR and MA orders of the ARFIMA model are misspecified, like all maximum likelihood. The ARFIMA model's EML estimate in the OxMetrics 6 package was used to estimate the long memory parameter (see Doornik and Ooms, 2003).

1.3.2 GPH Log-periodogram Regression

This and the next subsections investigate the main semiparametric methods applied to the Egyptian stock market to estimate the long memory parameter d in the frequency domain. The semiparametric estimation used in this paper is carried out in the Time Series Modelling (TSM) 4.32. These methods are not a recommended substitute for maximum likelihood estimation of an $ARFIMA(p, d, q)$ model if there is confidence that the ARMA components are correctly specified, but they impose fewer assumptions about the short-run. The assumption is that the spectrum of the process takes the form

$$f(\lambda) = \left|1 - e^{-i\lambda}\right|^{-2d} f^*(\lambda) \quad (1.3.2.1)$$

where f^* represents the spectral density of an $ARMA(p, q)$; and hence, the short-range component of the dependence. This is assumed smooth in the neighbourhood of the origin, with $f^*(0)' = 0$. Note the alternative representation

$$f(\lambda) = \lambda^{-2d} g(\lambda) \quad (1.3.2.2)$$

where g is likewise assumed smooth at the origin with $g(0)' = 0$.

Equation (1.3.2.1) is a semiparametric model, where the long memory parameter, d, is parametrically specified in the frequency domain; on the other hand, the short memory component represented in $f^*(\lambda)$ is not required to obey any parametric model. The two semiparametric estimators discussed in this chapter are the GPH and LW estimators. The log-periodogram estimator (i.e. GPH) minimise some distance between the periodogram and the spectral density function at low frequencies represented by the first m Fourier frequencies, $\lambda_j = \frac{2\pi j}{T}, j = 1, \dots, m \ll [\frac{T}{2}]$. Estimation is usually between a set frequency band $(0,m]$ to capture the long run component $f(\lambda) = \lambda^{-2d} g(\lambda)$ whilst the remainder of the frequencies capture the local variations.

This method is based on the periodogram of the time series defined by

$$I(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} e^{it\lambda} (y_t - \bar{y}) \right|^2 \quad (1.3.2.3)$$

A series with long memory has a spectral density proportional to λ^{-2d} close to the origin (2.2.4). Assuming the fact that the spectral density of a stationary process can be formulated as $f(\lambda) = f_0(\lambda) 4 \sin^2(\frac{\lambda}{2})$ we may consider a regression of the logarithm should give a coefficient of $-2d$. The GPH estimator is based on the log linearization of the periodogram as follows,

$$\log\{I(\lambda_j)\} = C - d \log\left\{4 \sin^2\left(\frac{\lambda_j}{2}\right)\right\} + \varepsilon \quad (1.3.2.4)$$

The memory parameter is estimated

$$\hat{d}_{GPH} = -\frac{\sum_{j=1}^{m} (y_j - \bar{y}) \log\{I(\lambda_j)\}}{2 \sum_{j=1}^{m} (y_j - \bar{y})} \quad (1.3.2.5)$$

We consider only harmonic frequencies $\lambda_j = \frac{2\pi j}{T}$, $j \in (l, m]$, where l is a trimming parameter discarding the lowest frequencies and m is a bandwidth parameter. A necessary condition for consistency which depends on the bandwidth is that $\frac{m}{T} \to 0$ as $T \to \infty$.

1.3.3 Local Whittle Estimation

Kunsch (1987) proposed a local Whittle (LW) estimator and then developed by Robinson (1995). This estimator represents approximately a MLE in the frequency domain, since for larger T

$$I(\lambda_j) \sim e^{f(\lambda_j)^{-1}} \quad (1.3.3.1)$$

As a result, the likelihood function is,

$$L\{I(\lambda_j), \dots, I(\lambda_m), \theta\} = \prod_{j=1}^{m} \frac{1}{f_\theta(\lambda_j)} e^{-I(\lambda_j) f(\lambda_j)^{-1}} \quad (1.3.3.2)$$

where $\theta = (C, d)$ is the parameter vector. The log-likelihood function becomes,

$$l(\theta) = \sum_{j=1}^{m} \left[-\log f_\theta(\lambda_j) - \frac{I(\lambda_j)}{f_\theta(\lambda_j)} \right] \quad (1.3.3.3)$$

In the neighbourhood of zero frequency we obtain,

$$L(d, C) = \sum_{j=1}^{m} \left[\log C - 2d \log(\lambda_j) + \frac{I(\lambda_j)}{C\lambda_j^{-2d}} \right] \quad (1.3.3.4)$$

$$\frac{\partial l(C,d)}{\partial C} = \sum_{j=1}^{m} \left[\frac{1}{C} + \frac{I(\lambda_j)}{C^2 \lambda_j^{-2d}} \right] \quad (1.3.3.5)$$

yielding

$$\hat{C} = m^{-1} \sum_{j=1}^{m} \left[\frac{I(\lambda_j)}{\lambda_j^{-2d}} \right] \quad (1.3.3.6)$$

Inserting $\hat{C}$ for C in (1.3.3.4) and by minimisation, the local Whittle estimator can be written as,

$$\hat{d}_{LW} = \arg\min \left(\log \left[m^{-1} \sum_{j=1}^{m} \left[\frac{I(\lambda_j)}{\lambda_j^{-2d}} \right] \right] - 2dm^{-1} \sum_{j=1}^{m} \log(\lambda_j) \right) \quad (1.3.3.7)$$

Robinson (1995) showed the LW estimator is consistent for $d \in (-0.5, 0.5)$. However, its consistency depends on the bandwidth m, which satisfy $\frac{1}{m}+\frac{m}{T} \to 0$ as $T \to \infty$. The LW estimator is more attractive due to its nice asymptotic properties, the mild assumptions underlying it and the likelihood interpretation. Robinson (1995) also showed that

$$\sqrt{m}(\hat{d}_{LW} - d) \to N(0, \tfrac{1}{4})$$

1.4 Data and Empirical Results

To analyse the Egyptian stock market, the daily EGX30 stock index traded on Cairo Stock Exchange has been used in this chapter. The data covers the period from the first transaction, 01 January 1998 to 09 May 2010 for a total of 3,050 observations. The EGX30 Price Index includes the top 30 companies in terms of liquidity and activity in Egypt. It is weighted by market capitalisation adjusted by the free float. This stock index can be considered as a proxy for the Egyptian stock market. The period under analysis is of major importance because it starts with the revival of the stock market after major changes in political and economic reforms in 1990s and before the January Revolution in 2011. All subsequent analysis is done on the daily return series (see Figure 4 for the daily stock returns) by taking the natural logarithmic first-difference on EGX30 price index (see Figure 1),

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln P_t - \ln P_{t-1}$$

where P_t denotes the stock index in day t.

Figure 1: The EGX30 daily stock index

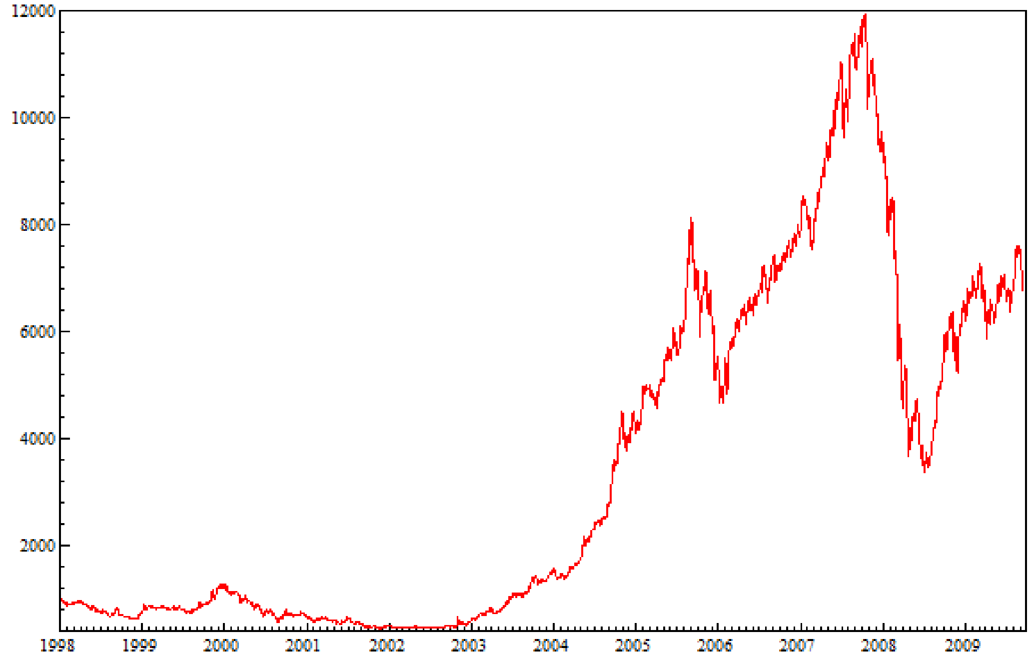


Figure 2: The periodogram of the daily EGX30 index

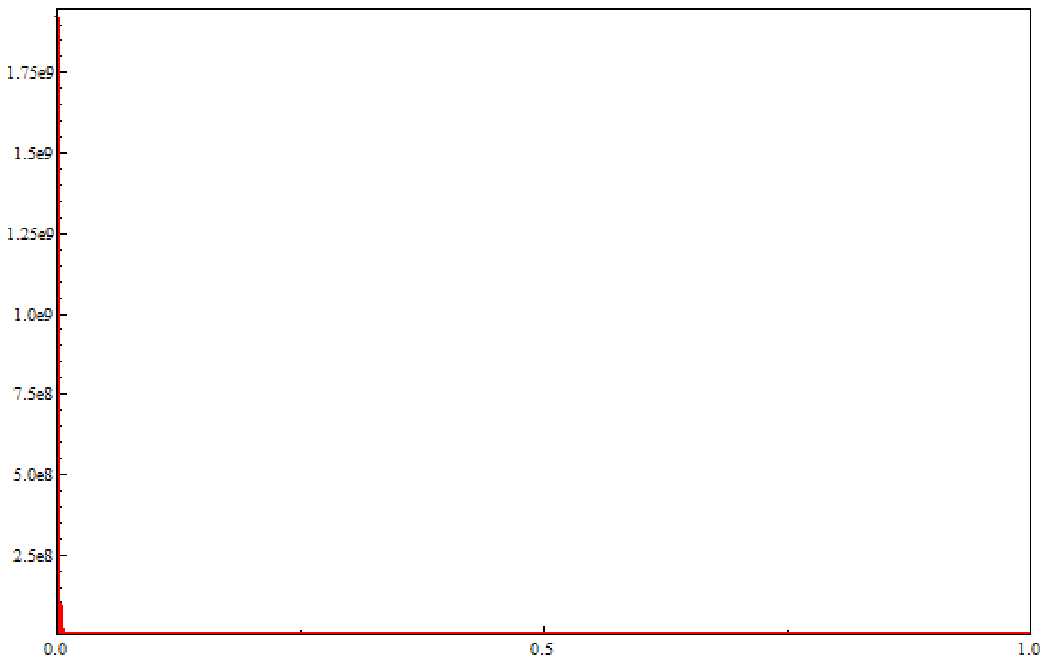


Figure 3: The correlogram of the daily EGX30 index

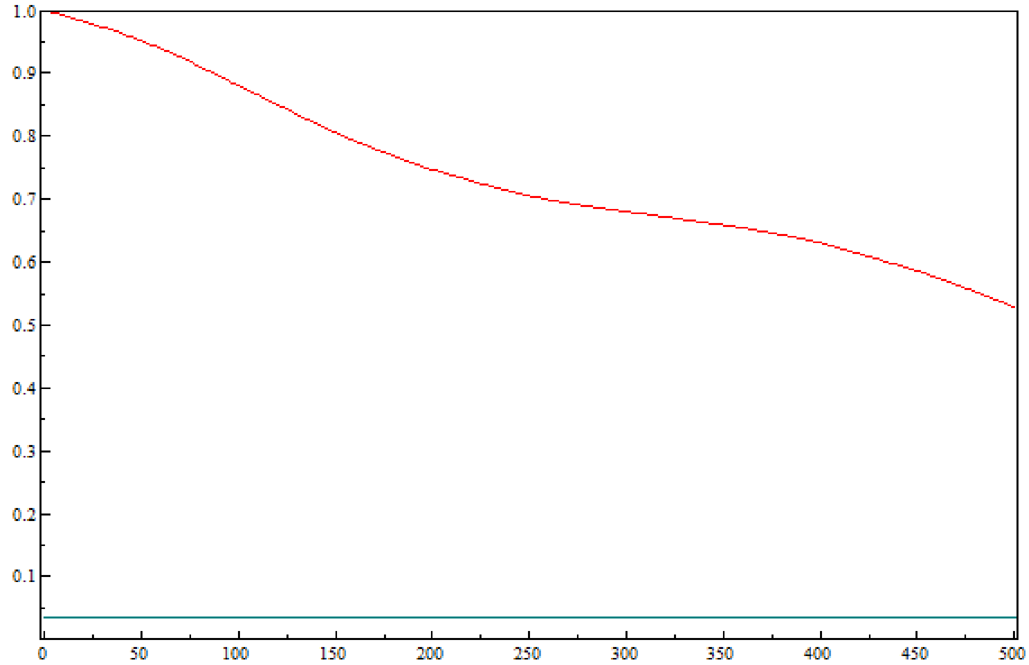


Figure 4: The EGX30 daily returns

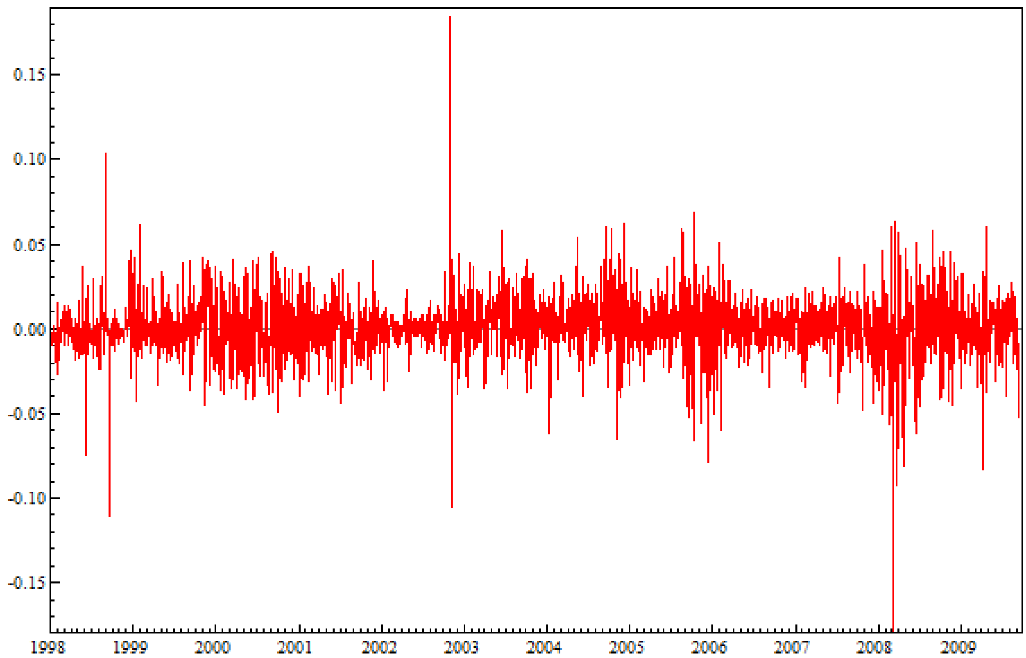


Figure 5: The periodogram of the daily EGX30 returns series

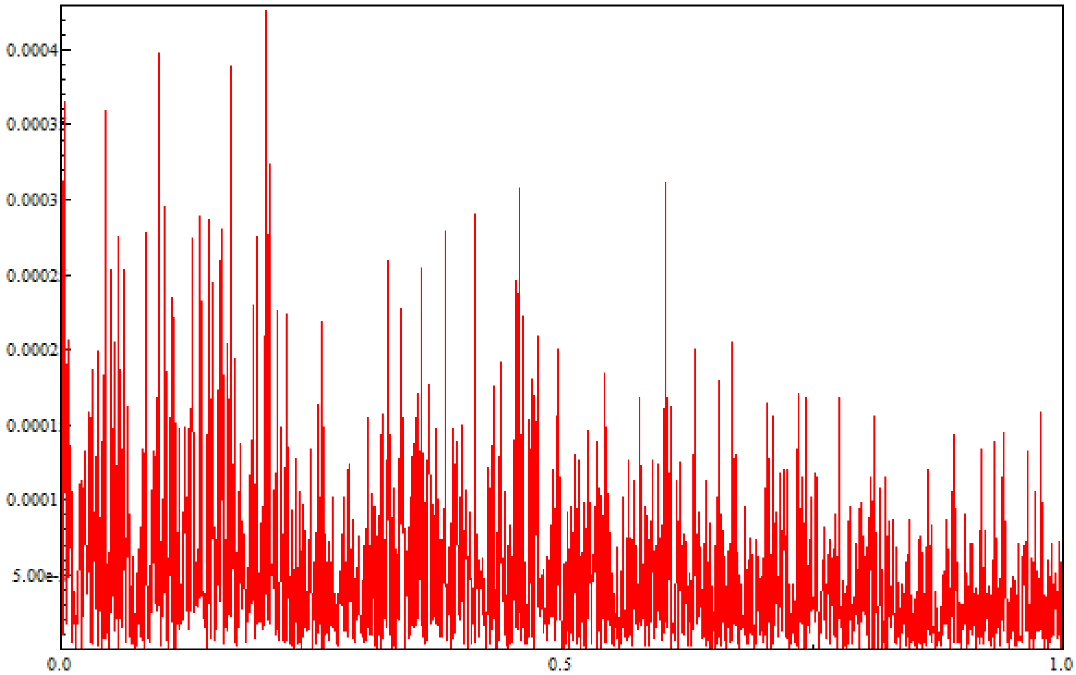


Figure 6: The correlogram of the daily EGX30 returns series

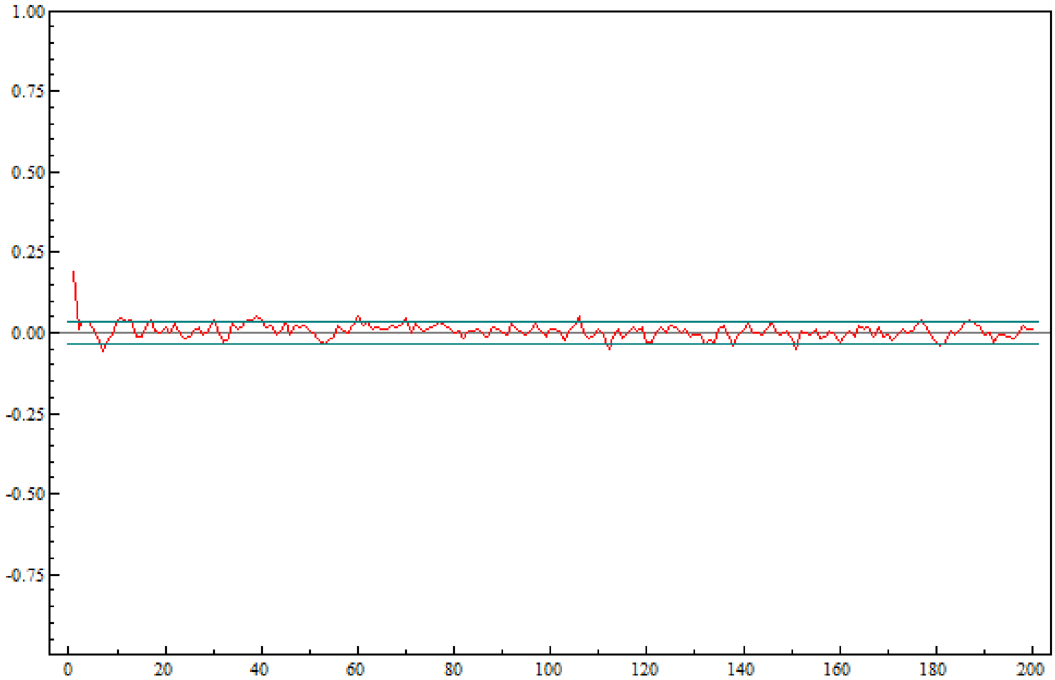


Figure 1 displays plots of the original EGX30 price index series where a nonstationary appearance can be observed. This can be also confirmed through its corresponding periodogram where large values (a large peak) are observed around the zero frequency and also across the correlogram with values decaying very slowly. Plots of the EGX30 daily return series data, with its corresponding periodogram and correlogram are displayed in Figures 4, 5 and 6. The return series may now be stationary (see figure 4). Based on the shapes of the corresponding periodogram and correlogram, the series may be over-differenced suggesting a presence of long memory. Dominant peak areas occurs around low frequencies (figure 5) and the correlogram declines steadily but very slowly and remains positive for many lags, indicating the presence of stationary long memory component (figure 6).

Table 2: Descriptive Statistics of EGX30 daily returns (1 Jan. 1989- 9 May 2010)

Obs.	Mean	S.D.	Skewness	Kurtosis
3049	0.0006	0.0179	-0.248	12.337
Min.	Max.	Jarque-Bera	ADF	KPSS
-0.179	0.183	11107	-21.25	0.3845

Note: The critical values of ADF unit root tests are -2.54, -1.95, -1.61 at 1%, 5%, 10% levels of significance.

Table 2 displays the descriptive statistics for the EGY30 daily returns over the full sample. The sample mean return is positive and very close to zero. There are significant departures from normality as the returns series is negatively skewed possibly due to the large negative returns associated with the financial crisis of 2007-2009. The unconditional distribution is peaked with fat tails. The data also display a high degree of kurtosis. Such skewness and kurtosis are common features in asset return distributions, which are repeatedly found to be leptokurtic. The data also fail to satisfy the null hypothesis of normality of the Bera-Jarque at the 1% level. Table 2 also includes the implementation of ADF and the KPSS tests. The ADF test shows evidence of non-stationary. The results of the ADF unit root test indicate that the return series are stationary by rejecting the null hypothesis of $I(1)$ at 1% level. For the KPSS test, the critical values are 0.739, 0.463 and 0.347 corresponding to the 1%, 5% and 10% level respectively. The null hypothesis of $I(0)$ against long memory alternatives is rejected (KPSS = 0.38) at the 10% level suggesting that the long memory process can be appropriate representation for the return series.

Figure 7: The distribution of the daily EGX30 returns series

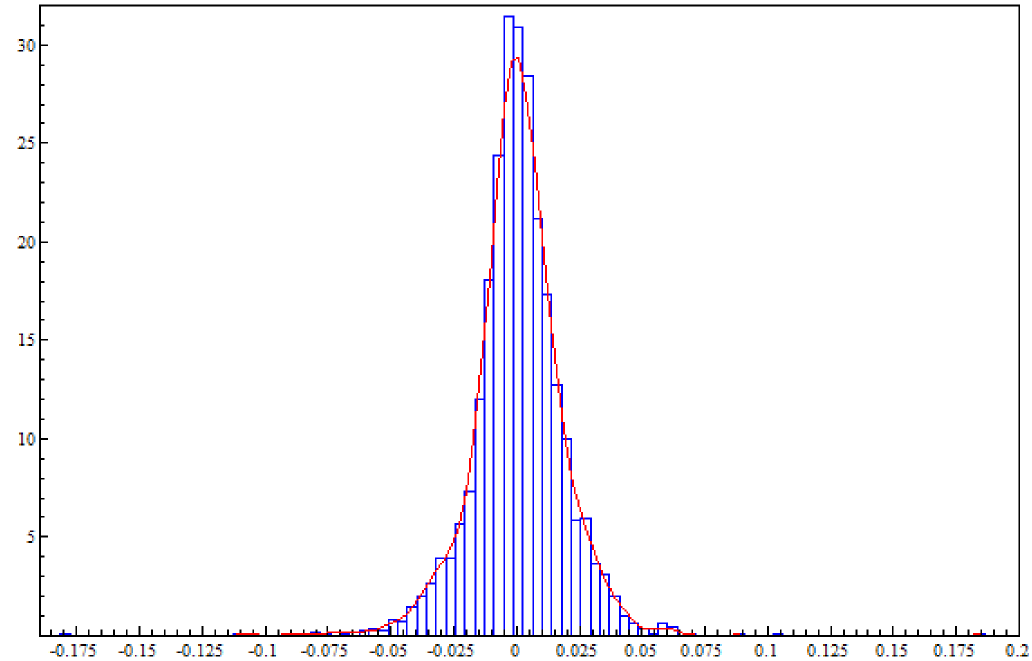


Figure 8: The QQ plot of the daily EGX30 returns series

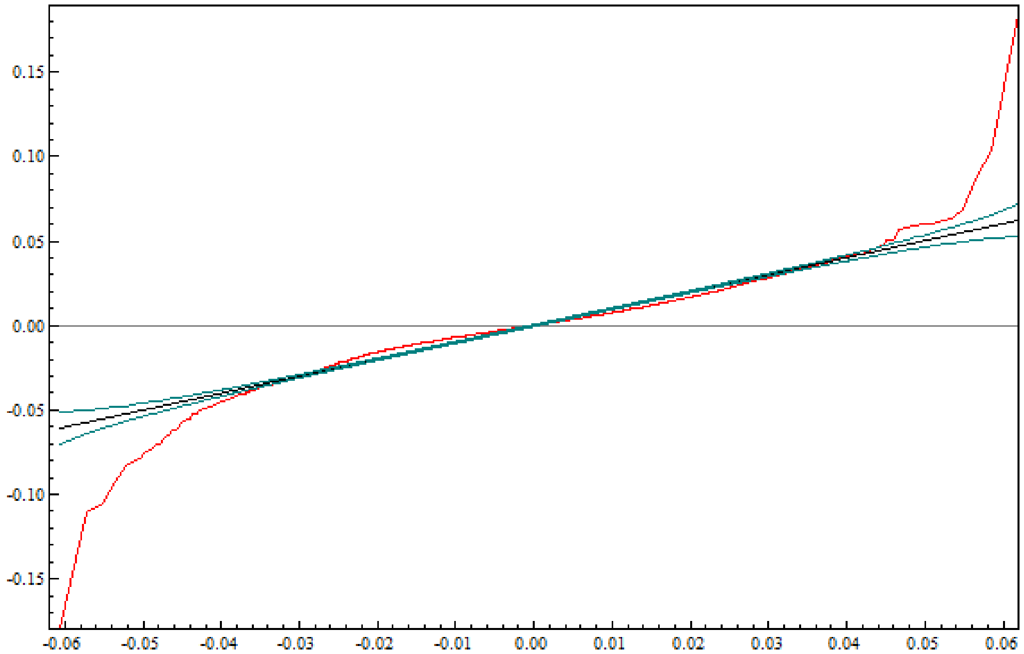


Table 3: ML estimation of ARFIMA models using EXG30 returns

p	q	$\hat{d}$	S.E.	AIC
0	1	0.035	0.019	-5.232
1	0	0.031	0.023	-5.231
1	1	0.055	0.023	-5.232
0	2	0.061	0.026	-5.233
1	2	0.058	0.027	-5.232
2	0	0.068	0.028	-5.232
2	1	0.070	0.029	-5.231
2	2	0.063	00.30	-5.231
3	0	0.051	0.032	-5.232
3	1	0.046	0.034	-5.231
3	2	0.275**	0.109	-5.232
0	3	0.056	0.030	-5.232
1	3	0.049	0.035	-5.231
2	3	0.259**	0.107	-5.232
3	3	0.031	0.032	-5.235
0	4	0.043	0.033	-5.231
1	4	0.255**	0.108	-5.232
2	4	0.032	0.029	-5.235
3	4	0.174**	0.078	-5.232
4	0	0.044	0.035	-5.231
4	1	0.258**	0.123	-5.234
4	2	0.054	0.027	-5.231
4	3	0.186	0.081	-5.232
4	4	0.416*	0.062	-5.236

Note: * and ** indicate statistical significance at the 1% and 5% levels respectively.

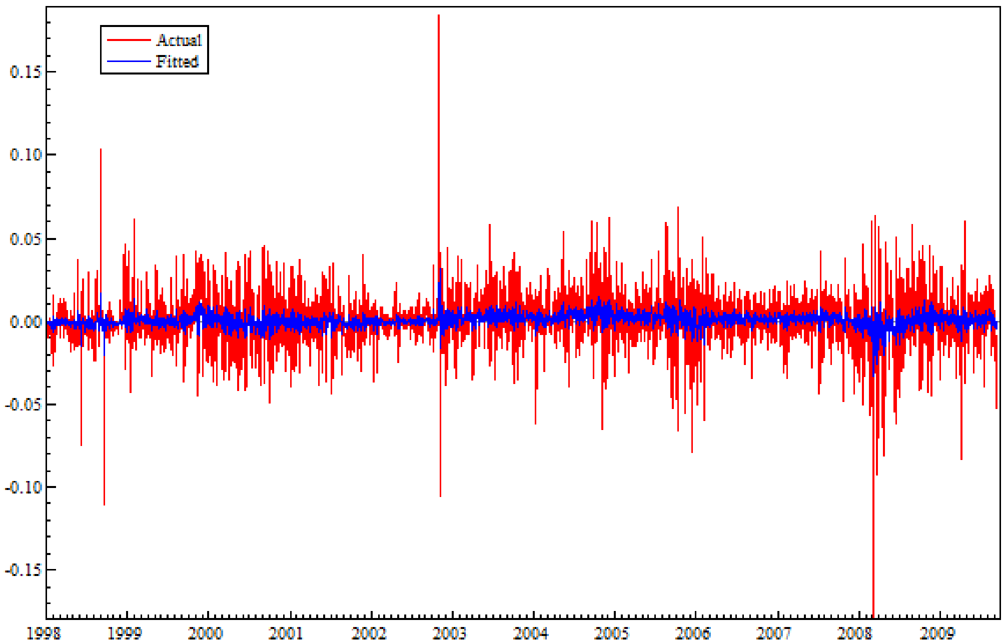
Table 4: ML estimation of ARFIMA$(4, d, 4)$ model

	Coefficient	S.E.	t-value	p-value
$\hat{\mu}$	0.0003	0.035	0.09	0.925
$\hat{d}$	0.461	0.062	6.67	0.000
$\hat{\phi}_1$	0.721	0.055	9.76	0.000
$\hat{\phi}_2$	0.168	0.061	3.10	0.002
$\hat{\phi}_3$	-0.933	0.058	-15.9	0.000
$\hat{\phi}_4$	0.616	0.068	9.65	0.000
$\hat{\psi}_1$	-0.965	0.070	-15.3	0.000
$\hat{\psi}_2$	-0.159	0.063	-2.92	0.004
$\hat{\psi}_3$	0.101	0.051	16.2	0.000
$\hat{\psi}_4$	-0.785	0.046	-15.6	0.000

Note: All estimators are statistically significant at the 1% level.

The parametric estimation for the returns series were derived by means of the Exact MLE of the OxMetrics 6 ARFIMA package, while the TSM modelling was used to obtain the long memory estimates via semiparametric methods. The ARFIMA model's Exact MLE (Maximum Likelihood Estimate) in the OxMetrics 6 package was used (see Doornik and Ooms, 2003). The models with different orders are estimated for ARFIMA (p, d, q). Table 3 show the results from various ARFIMA models with different specifications where $p + q$ equals and less than 4. The model is selected based on the Akaike's information criterion (AIC) and log likelihood values. The selected ARFIMA model is ARFIMA$(4, d, 4)$[5]. The estimated results show that the memory parameter is 0.41. The evidence of long memory property can be found in the model estimation where the long memory parameter is statistically significant at 1% level (see table 4). Hence, the EGX30 returns series exhibit long memory features. ARFIMA $(4, 0.41, 4)$ model is fitted to the data to capture the long memory characteristics of the returns series as in Figure 9.

Figure 9: The fitted ARFIMA$(4, 0.41, 4)$ model



[5] If the lag polynomials for AR and MA have common roots, a more economical ARMA $(p-1, q-1)$ model suffices and hence written as a lower-order process. Unique roots were found and are either real or in complex conjugate pairs. The ϕ 's roots are outside the unit circle, while the ψ's roots are inside the unit circle. So, it is an ARMA (4, 4). Alternatively, a purely autoregressive process can be considered which may typically require a higher number of parameters.

Moreover, the presence of the long memory properties in the Egyptian stock market suggests that ARFIMA models can improve forecasting performance by providing very reliable out-of-sample forecasts for both the long memory and the short run dynamic properties of the return series. The Egyptian stock returns series is forecast by using the ARFIMA model fitted to the EGX30 returns series according to the AIC. This forecast should significantly outperform any others using standard linear models. The period 9 April 2010 to 9 May 2010 is used for out of-sample forecasting. Table 5 reports the ex-ant forecasting performance for the EGX30 returns series (see figure 10).

Table 5: Out-of-sample Forecasting Performance for the daily EGX30 Returns

Forecasting Horizon	1	5	10	15	30
RMSE	0.0622	0.1145	0.1899	0.2536	0.4019
MAD	0.0597	0.0871	0.1540	0.2321	0.3358

Note: The out-of-sample period is from 9 April 2010 to 9 May 2010. The forecasting horizon is reported in *k*-steps ahead. The RMSE stands for the root mean square error, while the MAD is the mean absolute deviation.

Figure 10: The ARFIMA(4, 0.41, 4) model forecast

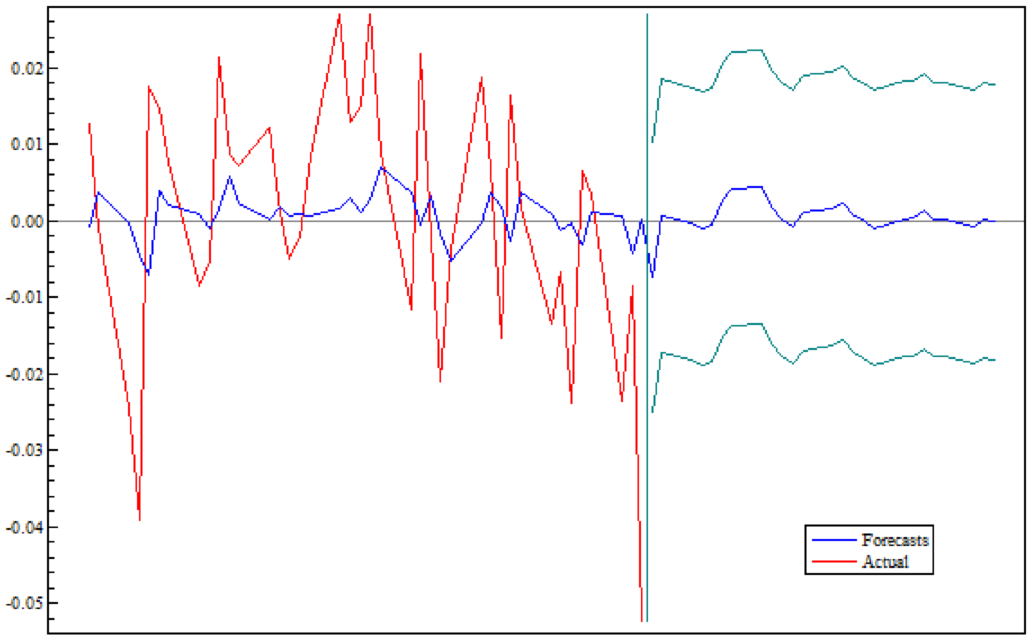


Table 6: Semiparametric estimates of d for returns

	Bandwidth			
	$m = n^{0.45}$	$m = n^{0.5}$	$m = n^{0.55}$	$m = n^{0.6}$
$\hat{d}_{GPH}$	0.2644 (0.1212)	0.1981 (0.0954)	0.1979 (0.0787)	0.1838 (0.0631)
$\hat{d}_{LW}$	0.2536 (0.0833)	0.1758 (0.0680)	0.1346 (0.0559)	0.1290 (0.0458)

Note: The standard errors are provided in parentheses.

Table 6 reports the semiparametric estimates of the long memory parameter d for the two estimators GPH and LW. The conventional setting of the bandwidth to be equal to the square root of the same size ($m = n^{0.5}$) was adopted. Moreover, d estimates were reported for different bandwidths $m = n^{0.45}$, $n^{0.55}$ and $n^{0.6}$ in order to evaluate the sensitivity of the results to the choice of the bandwidth. The results are not too sensitive to the bandwidth. Looking at the returns series, both estimators present similar results which show the existence of long memory features. The estimated d values range between 0.1 and 0.3, which is the property of stationary long memory processes. All the estimates of d are significantly positive at the 1% level. This result, due to semiparametric techniques, confirms the presence of long memory in the Egyptian stock returns as that of the parametric method.

1.5 Concluding Remarks

This chapter applied the parameter and semiparametric techniques to examine the long memory property in the daily Egyptian stock market returns. The exact maximum likelihood estimation was employed as a parametric method in the time domain to estimate the ARFIMA model, while two semiparametric methods were used to estimate the memory parameter in the frequency domain. The results from the ARFIMA model show evidence of long memory in the EGX30 returns. The results were also confirmed using the semiparametric methods. Both techniques provide strong evidence of long range dependence in the EGX30 returns. This implies that price movements in the Egyptian stock market appear to be related and affected by past and remote observations. The paper's findings suggest that long memory plays an important role in the structure and the dynamic behaviour of the Egyptian stock market returns and hence, influence the investment strategies involving multinational equities portfolios. Moreover, the presence of long memory in the Egyptian stock market may suggest constructing nonlinear econometric models, such as ARFIMA, for improved and more efficient price forecasting performance.

Furthermore, long memory in returns is not consistent with market efficiency. This market inefficiency in the Egyptian stock market can be attributed to the high persistence of risk factors in the market or due to the lack of liquidity. Accordingly, investors can exploit such inefficiency to earn excess returns. In addition, regulators should analyse the sources of the persistence in the Egyptian stock market that takes the form of long memory in order to improve its efficiency.

CHAPTER 2

LOCAL WHITTLE ESTIMATION IN STATIONARY SYSTEMS: EVIDENCE FROM US AND CANADA INFLATION

AMR ALGARHI*
UNIVERSITY OF EXETER

Abstract

This chapter uses the bivariate framework introduced by Robinson (2008) to analyse the long run relationship between the monthly inflation rates in the US and Canada. For two stationary long memory time series driven by a common stochastic trend, there may exist a linear combination of the two series with smaller memory parameter. The bivariate model introduces four unknown parameters (two memory parameters, a phase parameter and a cointegration parameter) to be jointly estimated by optimising a local Whittle function. The results indicate the existence of a linear combination between the US and Canada inflation rates that has a long memory less than the two individual series.

JEL Classification: C14, C32.

Keywords: Frequency domain; Local Whittle estimation; Long memory; Semiparametric estimation; US and Canada inflation.

---

* Amr Algarhi: a.s.i.algarhi@exeter.ac.uk
I would like to thank Professor James Davidson at University of Exeter for supplying the TSM software. Codes and empirical applications were performed in Ox and TSM. All remaining errors are mine.

2.1 Introduction

In the analysis of a multivariate framework of long memory time series, two main features emerge: the possibility of cointegration and a phase shift that does not need to be zero. This chapter concerns with the joint estimation of the memory parameters along with the cointegrating and phase parameters in the bivariate framework developed by Robinson (2008). This procedure is applied to monthly US and Canada inflation rates to examine the long-run equilibrium relationship which consequently has its implications on the interdependence of their monetary policies. The local Whittle estimation is employed where four unknown parameters (two memory parameters, a phase parameter and a cointegration parameter) are introduced. Robinson (2008) introduces an additional parameter to model the phase (γ) between the linear combination between the two series y_t and x_t, ($y_t - \beta x_t$), and x_t flexibly. Moreover, Robinson (2008) derived the consistency and established the joint asymptotic normality of the estimates under the assumption that the memory parameters lie between zero and $\frac{1}{2}$ and indicated how his results to be applied in statistical inference.

Stationarity of time series was usually associated with the Box-Jenkins modelling methodology with inherent short-memory properties of a series; while lack of statistical evidence for existence of long memory in economic time series made research restricted to the intuitively conventional $I(1)/I(0)$ case. However, recent research in long memory time series modelling has provided enough tools to explore the idea of fractional cointegration empirically. In addition, the term long memory time series includes both stationary and nonstationary series. Since 1990, the mainstream econometric time series literature shows considerable interest in long memory by focusing on unit roots time series. Unit roots series can be perceived as special cases of nonstationary fractional series. The analysis in this chapter only covers stationary long memory series. Stationary long memory time series displays a statistically significant dependence between distant observations. This dependence can be formalised by assuming that the autocorrelations decay very slowly, hyperbolically, to zero as a function of the time lag or spectral density displaying a pole at zero frequency. In addition, this dependence structure across time played a vital role in the modelling of macroeconomic and financial data.

As in economic literature, relationships between variables comes in pairs; and hence it is a natural starting point to focus on analysing bivariate relations between stationary long memory time series. One of the first estimation methods in the bivariate framework was the semiparameter narrow-band ordinary least squares (NBLS) regression in the frequency domain, developed in a series of papers by Robinson (1994) and Robinson and Marinucci (2001). The NBLS estimator is an OLS estimator in a spectral regression with a degenerating frequency band around the origin. Robinson (1994) proved the consistency of this estimator in the stationary case. The NBLS estimator reduces the bias in comparison to the OLS estimator, by reducing the effect of correlation between cointegration errors, while the convergence rate of the NBLS estimator depends on the values of the long memory and the bandwidth used in estimation. On the other hand, Lobato (1999) derived a semiparametric two-step estimator of parameters that characterise long memory for a time series vector in the frequency domain. Asymptotic normality of his estimator was established, but did not include Gaussianity condition. The two main methods described above were combined by Marinucci and Robinson (2001) and Christensen and Nielsen (2004), who suggested conducting a fractional cointegration analysis in several steps. First, the memory parameters of the original series are seperately estimated by local Whittle QMLE. Secondly, the narrow band FDLS estimator for the cointegrating vector is calculated, and finally the persistence of the residuals is estimated assuming that the approach is equally valid for residuals. In addition, Velasco (2003) and Hassler et al. (2006) sought to estimate the memory parameter of the equilibrium error by applying semiparametric estimators to the residuals from cointegrating regressions. Nielsen (2007) considered joint estimation of the memory parameters and the cointegrating vector for stationary long memory series in a multivariate framework, but derived its asymptotic distribution only under the long-run exogeneity between the stochastic trend and equilibrium error. Nielsen and Frederiksen (2008) considered a fully modified narrowband least squares estimator that corrects the endogeneity bias of the NBLS, and analysed the estimation of the memory parameters from modified NBLS regression. Shimotsu (2007) also developed a semiparametric estimator for the multivariate stationary framework. He used a more general local form of the spectral density. In general, a joint estimation method for the memory parameters and the cointegrating vector is more preferable. The estimators for the cointegrating parameter considered above are mostly direct in the sense that they do not require estimation of memory parameters. An alternative approach, first introduced by Robinson (2008) in a context of stationary bivariate system, jointly estimate the cointegrating parameters along with the memory parameters or other nuisance parameters which is adopted in this chapter.

Many previous empirical studies in economic literature have examined the characteristics of aggregate US and Canada inflation rates. Klein (1976) and Nelson and Schwert (1977) imposed a unit root on the inflation process; while Ball and Cecchetti (1990) and Kim (1993) modelled inflation as a transitory and a permanent component, which is represented as a random walk. On the other hand, Barsky (1987) and Brunner and Hess (1993) argued that inflation was covariance stationary. Hassler and Wolters (1995) found evidence in favour of long memory properties. Furthermore, Doornik and Ooms (2004) used ARFIMA models with different estimation methods in order to model and forecast the long memory characteristics in inflation. This question, regarding the examination and modelling the long memory features in inflation rate series, should take on a new investigation of whether inflation rates are related across countries. Examining such relation has very important implications on the interdependence of domestic monetary policies and the validity of purchasing power parity. As a result, the purpose of this chapter is to examine and investigate the interdependence between the inflation rates in US and Canada by studying their long memory properties in a bivariate framework which avoids mathematical complexity of any general multivariate structure. However, unlike the conventional classical approaches discussed in literature, this model allows for the possibility of cointegration and phase shifts.

The rest of the chapter is structured as follows. The next section describes the methodology used by presenting Robinson's (2008) bivariate system and demonstrates the local Whittle (or Narrow-Band) estimation. Section 2.3 reports some simulation results. Section 2.4 offers an empirical application to analyse the long run equilibrium between the inflation rates in the US and Canada, and finally section 2.5 concludes.

2.2 Methodology

Suppose a bivariate jointly covariance stationary process $u_t = (u_{1t}, u_{2t})'$ has a spectral density matrix,

$$f_u(\lambda) \sim \begin{bmatrix} \lambda^{-d_1} & 0 \\ 0 & \lambda^{-d_2} e^{i\gamma \text{sgn}(\lambda)} \end{bmatrix} \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix} \begin{bmatrix} \lambda^{-d_1} & 0 \\ 0 & \lambda^{-d_2} e^{-i\gamma \text{sign}(\lambda)} \end{bmatrix} \text{ as } \lambda \to 0 \quad (2.2.1)$$

For simplicity, this can be written as,

$$f_u(\lambda) \sim \Lambda^{-1} \Omega \overline{\Lambda}^{-1} \text{ as } \lambda \to 0 \quad (2.2.2)$$

The parameters d_1, d_2 and γ are unknown real valued and will be collected in vector $\alpha = (d_1, d_2, \gamma)'$, where d_1 and d_2 are the memory parameters and lay in the interval $[0, \frac{1}{2})$ and γ is the phase parameter between u_{1t} and u_{2t} at zero frequency and lies in the interval $\gamma \in (-\pi, \pi]$. The term $\text{sign}(\lambda) = 1$ if $\lambda \geq 0$. The symbol "~" means that for each element, the ratio of real/imaginary parts of the left and right sides tend to 1. In (2.2.2), the over bar denotes the complex conjugate and the parameters and Ω is a 2×2 positive definite matrix.

The spectral density matrix in (2.2.1) can be written as,

$$f_u(\lambda) \sim \begin{bmatrix} \omega_{11} |\lambda|^{-2d_1} & \omega_{12} \lambda^{-d_1 - d_2} e^{-i\gamma \text{sign}(\lambda)} \\ \omega_{21} \lambda^{-d_1 - d_2} e^{i\gamma \text{sign}(\lambda)} & \omega_{22} |\lambda|^{-2d_2} \end{bmatrix}. \quad (2.2.3)$$

From the main diagonal element, it can be deduced that the bivariate series has the memory parameter d_1 and d_2 respectively. On the other hand, the off diagonal elements represent the cross spectrum between the bivariate series. It takes a real value only if ω_{12}, ω_{21} and/or $\gamma = 0$.

For any two time series to be cointegrated and shape a long run equilibrium relationship, they need to share a common stochastic trend with a specific memory parameter. The long run equilibrium relationship is represented in the linear combination that becomes less persistent. Intuitively, most studies focused on the conventional $I(1)/I(0)$ where persistence is reduced from 1 to zero. However, this model is developed where persistence takes values between 0 and $\frac{1}{2}$. Now consider the model that includes the bivariate series $(y_t, x_t)'$,

$$\begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \quad (2.2.4)$$

When $d_1 \neq d_2$ and $\beta = 0$, then y_t and x_t have unequal memories d_1 and d_2 respectively. When $d_1 < d_2$ and $\beta \neq 0$, the bivariate series are said to be cointegrated and the unobservable linear combination $u_{1t} = y_t - \beta x_t$ has a memory of d_1 which is less than the memory for the bivariate series. Robinson's (2008) local Whittle (or narrow-band) estimate $\theta = (d_1, d_2, \gamma, \beta)'$ is considered in this paper where $0 \leq d_1 < d_2 < \frac{1}{2}$ and $\beta \neq 0$.

A local Whittle estimation is considered which employs Fourier frequencies in the neighbourhood of the origin. To begin with, the discrete Fourier transform (dFt) and the periodogram of a time series W_t are defined and evaluated at frequency λ as

$$w_j(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} W_t e^{-it\lambda_j} \tag{2.2.5}$$

$$I_w(\lambda) = w_j(\lambda) w_j^*(\lambda) = n^{-1}\left(\sum_{t=1}^{n} W_t e^{-it\lambda}\right)\left(\sum_{t=1}^{n} W_t e^{-it\lambda}\right)' \tag{2.2.6}$$

where $w_j^*(\lambda)$ is the conjugate transpose of $w_j(\lambda)$. The Whittle function, $Q(\theta, \Omega)$, is approximated to the (negative) log-likelihood function is

$$Q(\theta, \Omega) = \frac{1}{m} \sum_{j=1}^{m} \left(\log\left|\Lambda^{-1}\Omega\bar{\Lambda}^{-1}\right| + tr\left[\Omega^{-1} Re\left(\Lambda I(\lambda_j)\bar{\Lambda}\right)\right]\right) \tag{2.2.7}$$

To find the local Whittle estimator, function (2.2.7) is minimised with respect to the unknown parameters θ and Ω. The first step is to concentrate (2.2.7) with respect to the parameter Ω solving the resulting first order condition for Ω and then substituting the result back into (2.2.7). The solution of the first order condition with respect to Ω gives

$$\hat{\Omega}(\theta) = \frac{1}{m} \sum_{j=1}^{m} \Lambda_j Re\left(I(\lambda_j)\right) \bar{\Lambda}_j$$

By substituting $\hat{\Omega}(\theta)$ into $Q(\theta, \Omega)$, this yields the concentrated likelihood function $R(\theta)$ in terms of the four parameters,

$$R(\theta) = \log\det\left\{\hat{\Omega}(\theta)\right\} - 2\left(\sum_{s=1}^{2} d_s\right) \frac{1}{m} \sum_{j=1}^{m} \log\left|\lambda_j\right| \tag{2.2.8}$$

The local Whittle estimator of the parameter of interest, θ, can then be defined in terms of the concentrated likelihood

$$\hat{\theta} = \arg\min_{\theta \in \Theta} R(\theta) \tag{2.2.9}$$

The space of the true parameter θ is the compact set $\Theta \in \mathbb{R}^4$. The consistency and asymptotic properties of the local Whittle estimator $\hat{\theta}$ was also established in Robinson (2008).

2.3 Finite sample simulations

In this section, the finite sample behaviour of LW estimator is investigated by conducting a Monte Carlo study. The following four generating mechanisms for u_{1t} and u_{2t} are considered.

Model A:

$$u_{1t} = (1-L)^{-d_1}\varepsilon_{1t} \qquad u_{2t} = (1-L)^{-d_2}\varepsilon_{2t}$$

Model B:

$$u_{1t} = (1-L)^{-d_1}\eta_{1t} \qquad u_{2t} = (1-L)^{-d_2}\varepsilon_{2t}$$

$$\eta_{1t} = 0.5\eta_{1,t-1} + \varepsilon_{1t}$$

Model C:

$$u_{1t} = (1-L)^{-d_1}\varepsilon_{1t} \qquad u_{2t} = (1-L)^{-d_2}\eta_{2t}$$

$$\eta_{2t} = 0.5\eta_{2,t-1} + \varepsilon_{2t}$$

Model D:

$$diag\{(1-L)^{d_1}, (1-L)^{d_2}\}(1-0.5L)u_t = \sqrt{R}\varepsilon_t$$

where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is bivariate independently and identically distributed with mean zero and unit variance, ρ is the correlation between ε_{1t} and ε_{2t}, and

$$R = \begin{bmatrix} 1 & 2\rho \\ 2\rho & 4 \end{bmatrix}.$$

Based on the above generating mechanisms, the process y_t in (2.2.4) is generated with $\beta = 1$. The data is generated (for all simulations) with two sets of memory parameters. Firstly the memory parameters used are $(d_1, d_2) = (0.05, 0.4)$ which is close to many practical situations and supported by the empirical application reported in the next section, and then $(d_1, d_2) = (0.2, 0.3)$ which indicates a weaker form of fractional cointegration where the two memory parameters are very close. Model A has no short-run dynamics, unlike Models B and C where short-run dynamics are introduced to u_{1t} and u_{2t} respectively. Model D satisfies the spectral density function adopted in this chapter in (2.2.1) to (2.2.3). The elements of the main diagonal for R are 1 and 4, while the off-diagonal elements is 2ρ, thus the phase parameter is set as $\gamma = (d_2 - d_1)\frac{\pi}{2}$. For the Monte Carlo study, 10000 replications for sample sizes n are used where $n = 128$ and $n = 512$ are chosen. The former sample size is chosen to be close to the application in the next section. The bandwidth parameters chosen are $m = n^{0.5}$ and $m = n^{0.4}$ to examine the robustness of the LW estimator due to changes in the bandwidth. The Monte Carlo bias and root mean squared error (RMSE) results of the local Whittle estimator for all above models are reported in Tables 7 and 8. Simulations are performed using Ox 6.0 and TSM 4.35.

Table 7: Simulation Results for bias and RMSE where $\rho = 0$

	Model A						Model B						Model C						Model D					
	Bias			RMSE			Bias			RMSE			Bias			RMSE			Bias			RMSE		
	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β
$d_1 = 0.05, d_2 = 0.4$																								
$n = 128$																								
$m = n^{0.4}$	0.19	0.13	0.22	0.26	0.23	0.31	0.34	0.27	0.43	0.39	0.35	0.44	0.15	0.12	0.17	0.14	0.11	0.16	0.06	0.04	0.09	0.18	0.12	0.15
$m = n^{0.5}$	0.14	0.09	0.17	0.24	0.18	0.25	0.26	0.24	0.31	0.28	0.31	0.42	0.11	0.10	0.14	0.16	0.10	0.10	0.03	0.02	0.07	0.09	0.07	0.12
$n = 512$																								
$m = n^{0.4}$	0.16	0.11	0.18	0.14	0.11	0.25	0.28	0.21	0.33	0.36	0.35	0.38	0.10	0.09	0.12	0.11	0.09	0.12	0.04	0.03	0.05	0.13	0.09	0.08
$m = n^{0.5}$	0.12	0.08	0.15	0.15	0.14	0.21	0.25	0.20	0.27	0.32	0.25	0.34	0.08	0.08	0.10	0.07	0.06	0.09	0.00	0.00	0.02	0.06	0.05	0.03
$d_1 = 0.2, d_2 = 0.3$																								
$n = 128$																								
$m = n^{0.4}$	0.65	0.42	0.35	0.74	0.55	0.49	0.86	0.81	0.76	0.84	0.78	0.59	0.44	0.35	0.40	0.33	0.25	0.31	0.18	0.15	0.23	0.31	0.27	0.29
$m = n^{0.5}$	0.61	0.43	0.31	0.67	0.50	0.45	0.83	0.84	0.73	0.79	0.77	0.54	0.41	0.30	0.36	0.32	0.27	0.35	0.16	0.13	0.25	0.28	0.25	0.20
$n = 512$																								
$m = n^{0.4}$	0.53	0.36	0.27	0.63	0.51	046	0.77	0.69	0.71	0.80	0.73	0.52	0.39	0.32	0.37	0.35	0.24	0.25	0.17	0.16	0.19	0.25	0.21	0.25
$m = n^{0.5}$	0.49	0.33	0.24	0.61	0.42	0.40	0.70	0.67	0.65	0.73	0.76	0.48	0.46	0.34	0.39	0.28	0.20	0.22	0.10	0.11	0.21	0.24	0.19	0.15

Table 8: Simulation Results for bias and RMSE where $\rho = 0.5$

	Model A						Model B						Model C						Model D					
	Bias			RMSE			Bias			RMSE			Bias			RMSE			Bias			RMSE		
	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β
$d_1 = 0.05, d_2 = 0.4$																								
$n = 128$																								
$m = n^{0.4}$	-0.25	-0.19	0.27	0.34	0.26	0.33	-056	-034	0.46	0.43	0.38	0.46	0.20	0.15	0.21	0.16	0.17	0.24	0.14	0.06	0.15	0.21	0.15	0.16
$m = n^{0.5}$	-0.17	-0.12	0.24	0.28	0.16	0.29	-0.43	-0.31	0.37	0.35	0.37	0.43	0.15	0.10	0.15	0.22	0.13	0.18	0.09	0.05	0.09	0.14	0.09	0.11
$n = 512$																								
$m = n^{0.4}$	-0.21	-0.17	0.25	0.18	0.17	0.29	-0.35	-0.26	0.37	0.41	0.32	0.34	0.16	0.11	0.19	0.15	0.14	0.16	0.06	0.04	0.11	0.15	0.07	0.09
$m = n^{0.5}$	-0.17	-0.13	0.19	0.18	0.20	0.28	-0.29	-0.29	0.32	0.36	0.27	0.32	0.16	0.13	0.15	0.11	0.08	0.12	0.02	0.01	0.05	0.08	0.09	0.08
$d_1 = 0.2, d_2 = 0.3$																								
$n = 128$																								
$m = n^{0.4}$	-0.60	-0.46	0.42	0.68	0.57	0.54	-0.74	0.78	0.81	0.81	0.80	0.68	0.36	0.43	0.41	0.54	0.46	0.38	0.31	0.26	0.20	0.25	0.26	0.32
$m = n^{0.5}$	-0.44	-0.51	0.37	0.62	0.52	0.47	-0.65	0.73	0.76	0.84	0.75	0.65	0.44	0.25	0.28	0.41	0.43	0.42	0.27	0.17	0.21	0.15	0.29	0.24
$n = 512$																								
$m = n^{0.4}$	-0.47	-0.52	0.31	0.57	0.46	0.47	-0.68	0.75	0.74	0.76	0.75	0.67	0.42	0.35	0.35	0.45	0.40	0.35	0.19	0.15	0.14	0.18	0.16	0.21
$m = n^{0.5}$	-0.56	-0.45	0.29	0.55	0.45	0.48	-0.59	0.66	0.66	0.70	0.71	0.70	0.39	0.29	0.22	0.38	0.36	0.39	0.15	0.16	0.15	0.21	0.16	0.23

Table 9: Simulation Results for median bias and MAD where $\rho = 0$

	Model A						Model B						Model C						Model D					
	Bias			MAD			Bias			MAD			Bias			MAD			Bias			MAD		
	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β
$d_1 = 0.05, d_2 = 0.4$																								
$n = 128$																								
$m = n^{0.4}$	0.048	0.046	0.050	0.05	0.046	0.06	0.067	0.054	0.07	0.061	0.051	0.08	0.042	0.043	0.03	0.041	0.044	0.04	0.029	0.032	0.03	0.045	0.044	0.04
$m = n^{0.5}$	0.041	0.027	0.041	0.06	0.037	0.05	0.053	0.049	0.07	0.049	0.054	0.07	0.036	0.041	0.05	0.045	0.038	0.04	0.024	0.034	0.03	0.048	0.036	0.03
$n = 512$																								
$m = n^{0.4}$	0.045	0.042	0.043	0.04	0.045	0.05	0.057	0.049	0.06	0.066	0.061	0.06	0.042	0.029	0.04	0.042	0.036	0.04	0.034	0.061	0.03	0.049	0.028	0.03
$m = n^{0.5}$	0.040	0.037	0.045	0.04	0.043	0.04	0.055	0.051	0.05	0.064	0,057	0.06	0.037	0.031	0.04	0.035	0.035	0.03	0.026	0.019	0.02	0.034	0.024	0.04
$d_1 = 0.2, d_2 = 0.3$																								
$n = 128$																								
$m = n^{0.4}$	0.092	0.074	0.065	0.10	0.083	0.07	0.102	0.103	0.09	0.107	0.103	0.08	0.073	0.068	0.07	0.065	0.053	0.06	0.046	0.041	0.03	0.062	0.058	0.06
$m = n^{0.5}$	0.090	0.076	0.060	0.09	0.081	0.08	0.105	0.102	0.09	0.093	0.112	0.09	0.072	0.064	0.07	0.063	0.052	0.07	0.049	0.047	0.05	0.055	0.057	0.05
$n = 512$																								
$m = n^{0.4}$	0.086	0.065	0.054	0.09	0.073	0.07	0.095	0.086	0.09	0.109	0.094	0.07	0.069	0.057	0.06	0.064	0.051	0.04	0.038	0.045	0.03	0.056	0.051	0.06
$m = n^{0.5}$	0.082	0.062	0.047	0.09	0.075	0.08	0.094	0.095	0.08	0.098	0.097	0.06	0.075	0.066	0.07	0.054	0.054	0.05	0.041	0.042	0.05	0.054	0.047	0.04

Table 10: Simulation Results for median bias and MAD where $\rho = 0.5$

	Model A						Model B						Model C						Model D					
	Bias			MAD			Bias			MAD			Bias			MAD			Bias			MAD		
	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β	d_1	d_2	β
$d_1 = 0.05, d_2 = 0.4$																								
$n = 128$																								
$m = n^{0.4}$	-0.051	-0.04	0.057	0.06	0.056	0.06	-0.08	-0.065	0.07	0.064	0.071	0.07	0.053	0.042	0.05	0.045	0.048	0.05	0.045	0.003	0.04	0.055	0.042	0.04
$m = n^{0.5}$	-0.046	-0.04	0.054	0.05	0.048	0.05	-0.07	-0.063	0.05	0.075	0.068	0.07	0.042	0.041	0.04	0.051	0.045	0.05	0.038	0.008	0.04	0.043	0.038	0.03
$n = 512$																								
$m = n^{0.4}$	-0.053	-0.03	0.057	0.04	0.046	0.05	-0.07	-0.26	0.06	0.071	0.062	0.06	0.047	0.040	0.05	0.045	0.044	0.04	0.037	0.012	0.05	0.063	0.046	0.03
$m = n^{0.5}$	-0.046	-0.04	0.048	0.04	0.052	0.05	-0.06	-0.29	0.06	0.067	0.052	0.07	0.049	0.038	0.04	0.041	0.037	0.04	0.032	0.004	0.08	0.019	0.014	0.02
$d_1 = 0.2, d_2 = 0.3$																								
$n = 128$																								
$m = n^{0.4}$	-0.093	0.07	0.074	0.09	0.086	0.08	-0.09	0.101	0.08	0.121	0.095	0.08	0.064	0.074	0.07	0.085	0.073	0.06	0.061	0.055	0.05	0.054	0.046	0.06
$m = n^{0.5}$	-0.075	0.07	0.066	0.09	0.083	0.07	-0.08	0.098	0.09	0.104	0.099	0.08	0.072	0.053	0.05	0.082	0.075	0.07	0.059	0.049	0.04	0.052	0.052	0.05
$n = 512$																								
$m = n^{0.4}$	-0.078	-0.08	0.063	0.08	0.074	0.08	-0.06	0.093	0.09	0.091	0.097	0.09	0.073	0.063	0.06	0.078	0.073	0.07	0.048	0.045	0.04	0.046	0.047	0.05
$m = n^{0.5}$	-0.087	-0.09	0.057	0.08	0.075	0.06	-0.07	0.089	0.11	0.090	0.092	0.09	0.068	0.059	0.06	0.079	0.065	0.07	0.043	0.042	0.05	0.042	0.050	0.05

For Model A, the values of the bias are high for almost all the specifications. The RMSE decreases for all the parameters for a larger bandwidth. The bias and RMSE of d_1 are higher than those of d_2. In Model B, the biases and RMSE are found to be larger when there is no short run dynamics. However, both bias and RMSE decreases for larger bandwidth and sample size chosen. For Model C, the LW estimator appears to perform better than Model A, as the bias and RMSE are lower. Finally, the simulation for Model D works very well and produces unbiased estimates with very low bias and RMSE compared to the other models. In general, for all models, when the memory parameters are closer $(d_1, d_2) = (0.2, 0.3)$, even for larger n, the bias is more severe; however matters improve for larger bandwidth. On the other hand, for $(d_1, d_2) = (0.05, 0.4)$ the sizes of bias and RMSE are better on average. In tables 7 and 8, the values of the mean bias are very high, which might indicate that the first moment of the estimator does not exist. As a result, the median bias and the median absolute deviation (MAD) are reported in tables 9 and 10 instead of the mean bias and the root mean square error (RMSE). The median bias for Model A is very low. In Model B, the biases and RMSE are found to be larger when there is no short run dynamics. In addition, the LW estimator appears to perform better in Model C than in Model A, as the median bias and MAD are lower over the different bandwidths. Finally, the simulation for Model D works very well and produces unbiased estimates with very low bias and RMSE compared to the other models. In general, both median bias and MAD decreases for all the parameters as the bandwidth increases. The median bias and MAD of d_1 are higher than those of d_2. Overall, it seems difficult to draw exact conclusions about the effect of ρ, as it only causes the bias to change sign but does not change the size of bias or RMSE and it has no significant effect on the performance of the LW estimator. On the other hand, relatively larger bandwidth appears to be preferable as the LWE works best.

2.4 Application to the US and Canada inflation rates

Consumer price indices of the United States and Canada are originally examined. Monthly inflation rates (116 observations) are calculated based on the CPI measure of the US and Canada. This data measures the inflation rate for each month as the percentage increase from the same month of the previous year. The empirical analysis

Figure 11: The inflation rate, the correlogram (ACF) and the periodogram of USA and Canada respectively

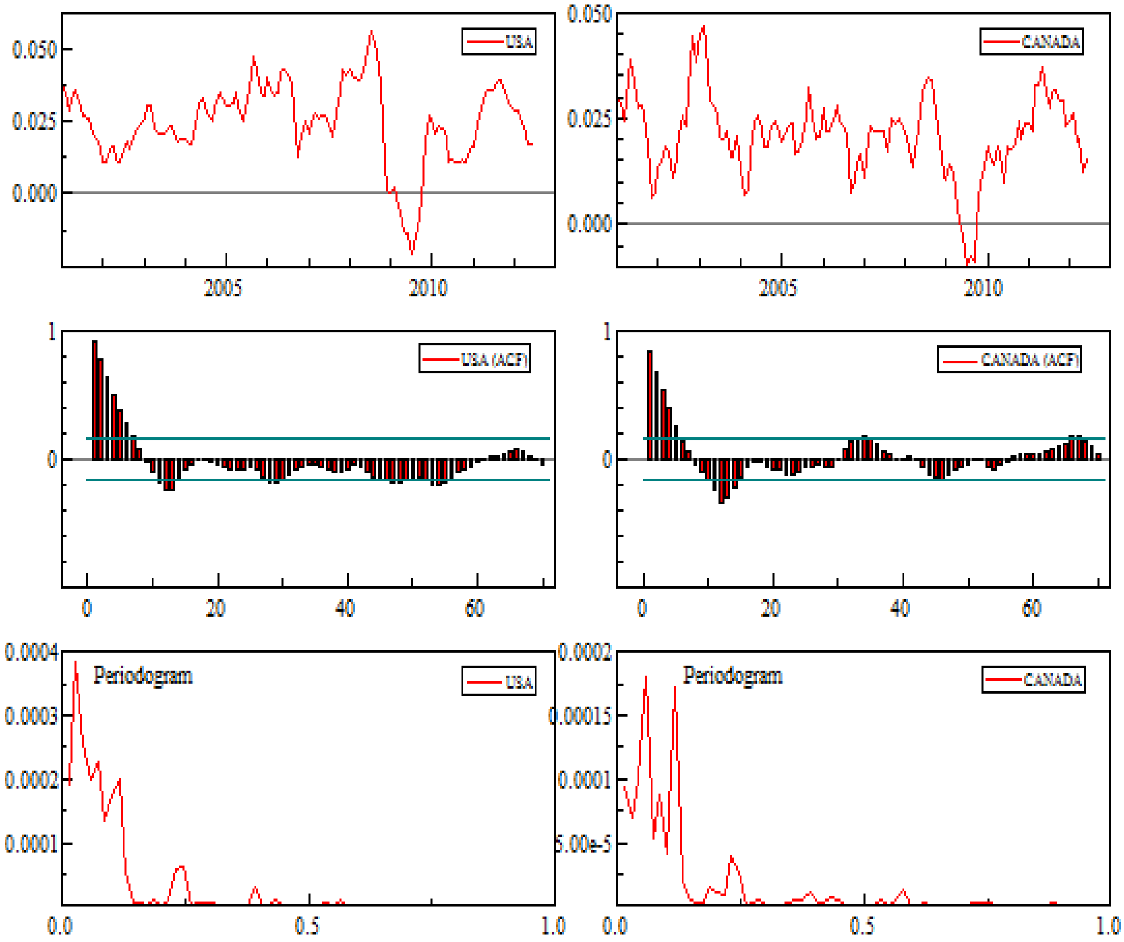


Table 11: Descriptive Statistics and Unit Root Tests

USA					
	Obs.	Mean	S.D.	Min.	Max.
	116	0.024371	0.014422	-0.021	0.056
	Skewness	Kurtosis	J.B.	ADF	KPSS
	-0.76105	4.0196	16.223*	-3.817	0.483
Canada					
	Obs.	Mean	S.D.	Min.	Max.
	116	0.020147	0.010036	-0.009	0.047
	Skewness	Kurtosis	J.B.	ADF	KPSS
	0.23046	4.1825	7.7850**	-5.912	0.365

Note: * and ** denote statistical significance of J.B. at the 1% and 5% levels respectively. The critical values of ADF unit root tests are -2.54, -1.94, -1.61 at 1%, 5%, 10% levels of significance. The critical values for KPSS test are 0.784, 0.521 and 0.437 at 1%, 5%, 10% levels of significance.

has been carried out using the monthly US and Canada inflation rates for the time period of January 2001 to August 2010. The series was obtained from the USA Federal Reserve Bank and the Bank of Canada respectively. Figure 1 provides graphs of the inflation rates, correlograms and periodograms respectively. The inflation rates in United States and Canada show the same trend movements which increased steadily with some oscillations to mid 2008 where the trend sharply declined. This similarity in the patterns between the US and Canada inflation rates, though not the levels, can lead to a potential cointegration relation between the two series. The corresponding correlograms exhibit the typical hyperbolic decline associated with long memory processes, while the periodograms in figure 11 confirm the presence of long memory features in the inflation series.

Table 12: The estimates of the LM parameters

	USA		Canada	
$m = n^{0.45}$				
	$\hat{d}_{GPH}$	$\hat{d}_{LW}$	$\hat{d}_{GPH}$	$\hat{d}_{LW}$
	0.473	0.488	0.318	0.274
	(0.147)	(0.112)	(0.098)	(0.073)
$m = n^{0.5}$				
	$\hat{d}_{GPH}$	$\hat{d}_{LW}$	$\hat{d}_{GPH}$	$\hat{d}_{LW}$
	0.448	0.426	0.291	0.236
	(0.172)	(0.124)	(0.109)	(0.080)

Note: The numbers in the parenthesis are standard errors.

Table 11 reports several descriptive statistics along with two unit root tests, including mean, standard deviation, skewness, kurtosis, Jarque-Bera statistic, ADF and KPSS. The US inflation rate averaged 2.4%, while the Canada inflation rate averaged 2%[6]. The values in the table give some information about the distribution of the US and Canada inflation rates. Both skewness and kurtosis statistics indicate that distributions are not normal. According to JB statistic, it is very clear that there are significant departures from normality. The next step of the analysis is to examine the unit root properties of the inflation rates using Augmented Dickey-Fuller (ADF) and Kwiatkowski, Phillips, Schmidt and Shin (KPSS) unit root tests. The results are presented in Table 11 and suggest that both series can be represented by stationary

[6] The Bank of Canada aims to keep inflation rate at the 2% midpoint of an inflation-control target range of 1-3 %.

long memory processes. However, these unit root tests, especially the ADF, do not take account of the possible long memory properties of the series. Therefore, two semiparametric methods are employed to examine the long memory properties of the data. The GPH and LW estimators for long memory parameters are reported in Table 12 for different bandwidths $m = n^{0.45}$ and $n^{0.5}$ where all the estimates can be seen to be statistically significant. The results show that the memory estimates are not sensitive to the bandwidth choice, although they decrease as the bandwidth increases (the memory estimates vary from 0.42 to 0.48 and 0.23 to 0.31 for US and Canada respectively) and that inflation rates exhibits stationary long memory properties. Consequently, if there exists a stable relationship between the inflation rates, a stationary fractional cointegration would be expected.

Now, consider the bivariate model in (2.2.4),

$$\pi_{USA,t} - \beta\pi_{CAN,t} = u_{1t}$$
$$\pi_{CAN,t} = u_{2t}$$

when $\beta \neq 0$, the two series $\pi_{USA,t}$ and $\pi_{CAN,t}$ are said to be cointegrated where the linear combination u_{1t} has a memory of d_1 which less than the memory of the original two series.

Table 13: Application to the US and Canada inflation rates

	$\hat{d}_1$	$\hat{d}_2$	$\hat{\beta}$	$\hat{\gamma}$
$m = n^{0.45}$	0.072	0.356	1.165	0.206
	(0.033)	(0.1138)	(0.298)	(0.081)
$m = n^{0.5}$	0.056	0.328	1.149	0.281
	(0.029)	(0.156)	(0.352)	(0.110)

Note: Standard errors are reported in the parentheses.

Table 13 reports the joint local Whittle estimation including the estimates of the four unknown (two memory, phase and cointegration) parameters, while the standard errors are represented in parentheses. The results indicate that all the coefficients are statistically significant for both bandwidths $m = n^{0.45}$ and $n^{0.5}$, respectively. The estimate of memory parameter d_2 indicates that the inflation rates can be described as stationary long memory series confirming the results in table 12. In addition, the estimate of d_1 for the unknown linear combination appears to have less memory than d_2. Moreover, the estimate of the cointegrating parameter β is close to unity reflecting a cointegration relationship between the US and Canada inflation rates. In particular, the LW estimates of the cointegration coefficient are significantly higher than unity for bandwidths $m = n^{0.45}$ and $n^{0.5}$, implying that the long-run rate of inflation in the US is higher than that in Canada.

2.5 Conclusion

One contribution of this paper is to apply the theoretical framework in Robinson (2008) to inflation rates which allows for a new parameter, the phase shift, to the bivariate model. The possibility of existence of long memory features in the inflation rates was initially examined, then the relationship between the monthly US and Canada inflation rates was analysed using the analysis in Robinson (2008). This approach is preferable to other conventional methods as it allows for the possibility of phase shifts along with cointegration. In addition, the four unknown parameters were jointly estimated using local Whittle estimation. The main finding is that the monthly US and Canada inflation rates exhibit the properties of stationary long memory series confirming the presence of long memory in macroeconomic time series which is consistent with the results reported in Hassler and Wolters (1995) and Doornik and Oooms (2004). The local Whittle estimate gives evidence to a fractional cointegration relationship between the US and Canada inflation rates, with the estimate of the cointegrating parameter, β, is higher than unity. This implies that the long-run rate of inflation in the US is higher than that in Canada. Furthermore, this link between inflation rates in the US and Canada has its vital implications on the interdependence of monetary policies in both countries and the validity of purchasing power parity. As the US and Canada have differing rates of inflation, and the relative price of goods is linked to the exchange rate through the purchasing power parity theory. The relative prices of goods should change and the value of US dollar may decline against the Canadian dollar.

CHAPTER 3

LOCAL WHITTLE ESTIMATION FOR MULTIVARIATE STATIONARY FRACTIONALLY COINTEGRATED SYSTEMS

AMR ALGARHI*

UNIVERSITY OF EXETER

Abstract

This chapter proposes a semiparametric estimator for multivariate fractionally cointegrated systems where the values of the memory parameter (d) lie between 0 and ½ by optimising a local Whittle function in the frequency domain. The proposed local Whittle estimator (LWE) is used to jointly estimate the memory, cointegrating and phase parameters. To derive this estimator, a general shape of the spectral density matrix first noted in Davidson and Hashimzade (2008) is utilised to cover multivariate jointly dependent long memory time series. A Monte Carlo study exhibits the performance of the LWE for different sample sizes. Finally, three different empirically relevant examples are presented to examine the existence of stationary fractional cointegration relationships.

JEL Classification: C14, C32.

Keywords: Fractional cointegration; Frequency domain; Long memory; Semiparametric estimation; Whittle likelihood.

---

* Amr Algarhi: a.s.i.algarhi@exeter.ac.uk

I would like to thank Professor James Davidson at University of Exeter for his constant guidance and suggestions, and audiences at the 2010 Economics PGR Conference at University of Exeter, especially Professor Gareth Myles, Professor Tatiana Kirsanova, Andreea Halunga and Joao Madeira for the useful comments. Simulations and empirical applications were performed in Ox and TSM. All remaining errors are mine.

3.1 Introduction

Robinson (2008) proposed an alternative version of the semiparametric approach to jointly estimate four unknown parameters (two memory parameters, a cointegrating parameter and a phase parameter) by optimising a local Whittle function. Robinson (2008) introduced a semiparametric local Whittle estimator and derived its consistency and asymptotic normality. Unlike previous studies, in the literature, that directly estimated the cointegrating parameter without the requirement of estimating the memory parameter or other nuisance parameters. Robinson analysis was in the context of a stationary bivariate system for long memory time series which raise the attention to two main issues. The first issue to emerge is the possibility of fractional cointegration among the two individual series and the second issue raised was that the phase does not need to be zero as in the short memory series. In literature, the former issue has been developed recently; unlike the latter which attracted very little attention.

This chapter establishes, similar to Robinson (2008), a joint estimation of the memory, cointegrating and phase parameters in stationary fractionally cointegrated models. However, the bivariate framework is extended to consider a more general multivariate case. In order to extend Robinson (2008) analysis and to consider a general multivariate process with more than one cointegrating relation, a general shape of the spectral density matrix, first noted in Davidson and Hashimzade (2008), is utilised to cover multivariate jointly dependent stationary long memory time series. The estimation method adopted in this chapter follows the semiparametric approach, in that the spectral density is only specified in a neighbourhood of zero frequency.

It has been commonly thought of cointegration as a stationary relationship between non-stationary variables. However, cointegration can be present between stationary processes with long memory where their linear combination is another stationary process with less memory. Throughout this chapter the main focus will be on the covariance stationary observable series with long memory where the memory parameter is between zero and ½, hence the notion of stationary fractionally cointegrated systems. The term fractional refers to a generalised operation of a non-integer order and the stationary region lies in [0, ½). This interval is relevant for many applications in macroeconomics and finance. Moreover, the most common value of the phase of the spectral density at the origin adopted in the literature is equal to zero. The spectral density that overlooks the information in the phase parameter leads to less efficient estimates of the cointegrating parameter. This chapter employs a more general form of the spectral density, as in Davidson and Hashimzade (2008), which allows for values of the phase parameters that is different from zero.

A Monte Carlo study is presented to illustrate the properties of the proposed local Whittle estimator in finite samples. The stationary fractional cointegration model has many potential applications. The model developed in this chapter is applied to three different empirically relevant applications. The first example is to examine for inflation rate harmonization in Spain and France. Secondly, the fractional cointegration among the volatilities of three stock market indices is inspected. Finally, the model developed is applied on another trivariate series of the daily US Treasury rate, at constant maturities of 2 years, 3years and 7 years, in order to analyse the long range dependence. An evidence of a weak form of fractional cointegration was reported in the three empirical examples.

The remainder of the chapter is organised as follows. The next section covers the relevant literature review for stationary fractionally cointegrated models. Section 3.3 introduces the multivariate fractional cointegration model. Section 3.4 sets up the local Whittle likelihood. Sections 3.5 and 3.6 derive the consistency and the asymptotic normality of the local Whittle estimator for the multivariate case respectively. Section 3.7 reports and discusses the results of a Monte Carlo study, illustrating the finite sample behaviour of the LWE. Section 3.8 demonstrates the empirical application to three different relevant examples and finally section 3.9 offers some concluding remarks.

3.2 Literature Review

In this section, a selective survey of the literature on stationary fractional cointegration model is presented. This section starts with the definition of cointegration which is the existence of a linear combination among two or more time series. Cointegrated time series form a long-run equilibrium relationship where they share a common stochastic trend with a memory parameter d_0. On the other hand, their linear combination is less persistent than any of the individual series which is measured in terms of a smaller memory parameter d_u. Standard cointegration, which is considered so far the most studied special case of cointegration, reduces the memory parameter from $d_0 = 1$ to $d_u = 0$ by the linear combination. Tests for standard cointegration mainly depend on unit root theory. Alternatively, fractional cointegration generalizes the conventional $I(1)/I(0)$ standard cointegration by allowing the memory parameter to be any real number, making it more flexible structure for analysing long run relationships between economic time series and enables more proper modelling of interdependence between them. Although both types of cointegration, standard and fractional, were concurrently defined and discussed in the seminal paper of Engle and Granger (1987); in the literature they have been developed separately. The literature on cointegration under autoregressive unit roots has exceeded that under long memory, until recently when theoretical studies of fractional cointegration have been rapidly expanding in several directions. In this section the focus is on the literature review that covers a stationary fractionally cointegrated systems where $0 \leq d_0 < d_u < \frac{1}{2}$.

In literature, the most used simple model for stationary fractional cointegration can be presented as follows. Suppose the p-vector $z_t = (y_t, x_t')'$ is observed, which is integrated of order $d \in (0, 1/2)$, where $z_t \in I(d)$, In other words, $z_t \in I(d)$ if

$$(1 - L)^d z_t = \varepsilon_t, \tag{3.2.1}$$

where $\varepsilon_t \in I(0)$, a process is $I(0)$ if it is covariance stationary and has spectral density that is bounded and bounded away from zero at the origin, and $(1 - L)^d$ is defined by its binomial expansion

$$(1 - L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j,$$

where $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$ is the gamma function and L $(Lz_t = z_{t-1})$ is the lag operator.

A scalar-valued stochastic process generated by (3.2.1) has spectral density

$$f(\lambda) \sim \omega \lambda^{-2d} \quad \text{as } \lambda \to 0^+, \tag{3.2.2}$$

where ω is a constant and the symbol "~" means that the ratio of the left- and right-hand sides tends to one in the limit. Such a process is said to possess strong dependence (or long range dependence), since the autocorrelations decay at a hyperbolic rate in contrast to the much faster exponential rate in the weak dependence case. The parameter d determines the memory of the process. If $d > -\frac{1}{2}$, z_t is invertible and admits a linear representation, and if $d < \frac{1}{2}$ it is covariance stationary. If $d = 0$, the spectral density (3.2.2) is bounded at the origin, and the process has only weak dependence. Sometimes, z_t is said to have intermediate memory, short memory, and long memory when $d < 0$, $d = 0$, and $d > 0$, respectively.

Moreover, suppose that $z_t = (y_t, x_t')'$ satisfies the regression model

$$y_t = \beta' x_t + u_t, \qquad (3.2.3)$$

where the error term is integrated of a smaller order $d_u < d$, $u_t \in I(d_u)$. When there is no integer constraint on d or d_u, the model is called a fractional cointegration model. In addition, assuming $0 \leq d_u < d < \frac{1}{2}$ the model is called a stationary fractionally cointegrated model, since it is concerned with the long-run linear co-movement between two or more stationary fractionally integrated processes. The properties of the model in the fractional cointegration framework have been examined only recently in Robinson and Yajima (2002).

Many estimators of the memory parameter d have been suggested in the literature; however, a selective survey on a semiparametric approach is presented in this section. Different attempts have been made to develop a semiparametric estimator of fractionally cointegrated systems. The two main methods of semiparametric estimation of the memory parameter discussed in empirical studies are log-periodogram (LP) regression and local Whittle (LW) estimation. In the stationary case, the LW estimator was found to be more efficient than LP regression.

Geweke and Porter-Hudak (1983) was the first to develop a semiparametric approach to estimate the memory parameter assuming only the model (3.2.2) for the spectral density and use a degenerating part of the periodogram around the origin to estimate the model. This approach has the advantage of being invariant to any short and medium term dynamics (as well as mean terms since the zero frequency is usually left out). Künsch (1987) firstly proposed the univariate local Whittle estimator. He developed a local Whittle quasi maximum likelihood estimator (QMLE) based on the maximization of a local Whittle approximation to the likelihood to estimate the memory parameter of the univariate stationary fractionally integrated time series. Robinson (1995a) worked on the univariate local Whittle estimator by showing its consistency and asymptotic normality for $d \in (-\frac{1}{2}, \frac{1}{2})$ and called it a Gaussian semiparametric estimator. Velasco (1999) showed that the estimator is consistent for $d \in (-\frac{1}{2}, 1)$ and asymptotic normally for $d \in (-\frac{1}{2}, \frac{3}{4})$.

In the multivariate framework, Robinson (1994a) proposed a semiparametric narrow least squares (NBLS) estimator in the frequency domain that assumes only a multivariate generalization of (3.2.3), and essentially performs OLS on a degenerating band of frequencies around the origin. He also proved the consistency of this estimator in the stationary case. In addition, Christensen & Nielsen (2004) showed that its asymptotic distribution is normal when the collective memory of the regressors and the error term is less than ½. Nielsen and Frederiksen (2008) considered a fully modified narrowband least squares (NBLS) estimator that corrects the endogeneity bias of the NBLS, and analysed the estimation of the memory parameters from modified NBLS regression. Lobato (1999) derived a semiparametric two-step estimator in a multivariate stationary long memory model. Velasco (2003) and Hassler et al. (2006) sought to estimate the memory parameter of the equilibrium error, d_u, by applying semiparametric estimators to the residuals from cointegrating regressions. Both Velasco (2003) and Hassler et al. (2006) required $d_0 - d_u > 1/2$, because, when $d_0 - d_u$ is small, the cointegrating vector estimate converges at a too slow rate to validate the subsequent residual-based analysis. Nielsen (2007) considered joint estimation of d_u, d_0 and the cointegrating vector under the assumption $0 \leq d_u < d_0 < 1/2$, but derived its asymptotic distribution only under the long-run exogeneity between the stochastic trend and equilibrium error. Shimotsu (2007) introduced a Gaussian semiparametric estimator of multivariate stationary fractionally integrated processes. He used a more general local form of the spectral density and from this he derived a semiparametric estimator of multivariate fractionally integrated processes. The class of spectral densities included in

Shimotsu's specification includes those of multivariate fractionally integrated processes which includes the information in phase shifts and which will lead to more efficient estimates of the integration orders. Robinson (2008) introduced a Gaussian semiparametric estimator of bivariate stationary fractionally integrated processes by extending the work by Robinson (1995a) and derives its consistency and asymptotic normality under the assumption $0 \leq d_u < d_0 < \frac{1}{2}$.

The methods described above are combined by Marinucci and Robinson (2001) and Christensen and Nielsen (2006), who suggest conducting a fractional cointegration analysis in several steps. First, the integration order of the raw data is estimated by local Whittle QMLE. Secondly, the narrow band FDLS estimator for the cointegrating vector is calculated, and finally the integration order of the residuals is estimated assuming that the approach is equally valid for residuals. Hypothesis testing is then conducted on d_u as if u_t were observed, and on β as if d_u were known. Although this is indeed a valid course of action, a joint estimation method for the integration orders and the cointegration vector is preferable.

Furthermore, the different estimators developed, as mentioned above, in the literature can also be classified according to the assumptions on the phase parameter. The simplest special case, as assumed by Christensen and Nielsen (2006), is that all the phase parameters are equal to zero. This assumption is plausible, when the spectral density is real. However, neglecting the information in the phase parameter in the spectral density can lead to less efficient estimates of the cointegrating and memory parameter. Another case is considered in the literature where the phase parameter is equal to $(\pi/2)d$ as in Lobato (1999), Robinson and Yajima (2002) and Shimotsu (2007). Finally, Robinson (2008) introduces a general case where the phase parameter in a bivariate framework is equal to $(\pi/2 - c)d$ where c is a real number satisfying the constraints that the phase parameter is in the Nyqvist range.

This chapter's contribution to the literature is to estimate the multivariate framework for a stationary fractional cointegration model by introducing a joint estimator of the cointegrating parameters together with the memory and phase parameters. Moreover, this chapter uses a more general form of the spectral density that allows for phase shifts as in Davidson and Hashimzade (2008). The next two sections introduce the multivariate model and develop the semiparametric local Whittle estimator.

3.3 Multivariate Stationary Fractional Cointegrated Model

Consider a real-valued stationary fractional process Y_t that is generated by the model,

$$\begin{bmatrix} Y_{1t} \\ \vdots \\ Y_{qt} \end{bmatrix} = \begin{bmatrix} g_1(\cdot) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & g_q(\cdot) \end{bmatrix} \begin{bmatrix} u_{1t} \\ \vdots \\ u_{qt} \end{bmatrix} \tag{3.3.1}$$

where

$$g_s(\cdot) = g(L; d_s, \kappa_s, a_s) \,, \qquad \text{where } s = 1, \dots, q$$

$$g_s(\cdot) = \kappa_s (1-L)^{a_s - d_s}(1-L^{-1})^{-a_s} + (1-\kappa_s)(1-L)^{-a_s}(1-L^{-1})^{a_s - d_s}$$

The parameters d_s, κ_s, a_s are real valued, $0 < d_s < \frac{1}{2}$, $0 \leq a_s \leq \frac{d_s}{2}$ and $0 \leq \kappa_s \leq 1$. These parameters are collected in $\alpha = (d', \kappa', a')'$ for $d' = (d_1, \dots, d_q)'$, $\kappa' = (\kappa_1, \dots, \kappa_q)'$ and $a' = (a_1, \dots, a_q)'$. The error vector, $u_t = (u_{1t}, \dots, u_{qt})'$, is serially independent Gaussian process with $(0, \sigma^2)$ whose spectral density $f_u(\lambda)$ is bounded and bounded away from zero at the zero frequency $\lambda = 0$. The memory parameters, d_s, govern the long-run dynamics of the process Y_{st} and the behaviour of its spectral density representation around the origin. Therefore, if empirical interest lies in the long-run dynamics of the process, it is useful to specify the spectral density only locally in the vicinity of the origin and avoid specifying the short-run dynamics of u_t explicitly. The parameters a_s present different degrees of forward (lead) and backward (lag) memories that the model can exhibit and different choices of the parameters κ_s influence the amount of short run memory relative to long run either symmetrically or asymmetrically (see Davidson and Hashimzade (2008)).

The spectral density representation of the process (3.3.1) is

$$f(\lambda) \sim \Lambda^{-1} \Omega \bar{\Lambda}^{-1} \,, \text{ as } \lambda \to 0 \tag{3.3.2}$$

where

$$\Lambda^{-1} = diag\{\Lambda_s^{-1}(d, \kappa, a)\}$$

$$\Lambda_s^{-1}(d, \kappa, a) = |\lambda|^{-d_s}[\kappa_s e^{-i \operatorname{sng}(\lambda) \gamma_s} + (1-\kappa_s) e^{i \operatorname{sng}(\lambda) \gamma_s}]$$

The over bar denotes the complex conjugate and the parameters $\gamma_s = (2a_s - d_s)\frac{\pi}{2}$. The term $\operatorname{sng}(\lambda) = 1$ if $\lambda \geq 0$ and -1 otherwise. The matrix Ω is a $q \times q$ real, symmetric, finite and positive definite. The spectral density $f(\lambda)$ has non-zero complex part. The notation "~" in equation means that for each element, the ratio of real/imaginary parts of the left and right sides tend to 1. We can deduce the cross-spectrum $f_{jk}(\lambda)$, where $j \neq k$, (off the diagonal elements of $f(\lambda)$),

$$f_{jk}(\lambda) = \omega_{jk}|\lambda|^{-d_j-d_k} \sum_{l=1}^{4} \zeta_l \, e^{-i\text{sng}(\lambda)\xi_l} \tag{3.3.3}$$

where

$$\begin{aligned} \zeta_1 &= \kappa_j \kappa_k & \xi_1 &= \gamma_j + \gamma_k \\ \zeta_2 &= \kappa_j(1-\kappa_k) & \xi_2 &= \gamma_j - \gamma_k \\ \zeta_3 &= (1-\kappa_j)\kappa_k & \xi_3 &= -(\gamma_j - \gamma_k) \\ \zeta_4 &= (1-\kappa_j)(1-\kappa_k) & \xi_4 &= -(\gamma_j + \gamma_k) \end{aligned}$$

3.4 Local Whittle Estimation

A semiparametric local Whittle estimation is considered which utilises only Fourier frequencies in the neighbourhood of the origin and therefore is nonparametric with respect to the short run dynamics of the data. To begin with, the discrete Fourier transform (d.F.t.) and the periodogram of a time series X_t are defined and evaluated at frequency λ as

$$x_j(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{it\lambda_j} \tag{3.4.1}$$

$$I_w(\lambda) = x_j(\lambda) x_j^*(\lambda) = n^{-1}\left(\sum_{t=1}^{n} X_t e^{it\lambda}\right)\left(\sum_{t=1}^{n} X_t e^{-it\lambda}\right)' \tag{3.4.2}$$

where $x_j^*(\lambda)$ is the conjugate transpose of $x_j(\lambda)$. The Whittle approximation to the (negative) log-likelihood function is

$$W(\theta, \Omega) = \frac{1}{m}\sum_{j=1}^{m}\left(\log\left|f(\lambda_j)\right| + tr[f^{-1}(\lambda_j)Re(I(\lambda_j))]\right)$$

$$W(\theta, \Omega) = \frac{1}{m}\sum_{j=1}^{m}\left(\log\left|\Lambda_j^{-1}\Omega\overline{\Lambda}_j^{-1}\right| + tr[(\Lambda_j^{-1}\Omega\overline{\Lambda}_j^{-1})^{-1}Re(I(\lambda_j))]\right)$$

$$W(\theta, \Omega) = \frac{1}{m}\sum_{j=1}^{m}\left(\log\left|\Lambda_j^{-1}\Omega\overline{\Lambda}_j^{-1}\right| + tr[\Omega^{-1}Re(\Lambda_j I(\lambda_j)\overline{\Lambda}_j)]\right) \tag{3.4.3}$$

To find the LW estimator, the minimization of (3.4.3) is needed with respect to the unknown parameters $\theta = (\alpha', \beta')'$ and Ω, where $\beta = (1, \beta_{12}, \dots, \beta_{1q})$ is the cointegrating parameters. The first step in minimising (3.4.3) is to concentrate it with respect to the parameter Ω. This means differentiating (3.4.3) with respect to Ω, solving the resulting first order condition for Ω and then substituting the result back into (3.4.3) as follows,

$$\frac{\partial W(\theta,\Omega)}{\partial \Omega}=\sum_{j=1}^{m}[\Omega^{-1}-\Omega^{-1}Re(\Lambda_j I(\lambda_j)\bar{\Lambda}_j)\Omega^{-1}]=0$$

The solution of the first order condition with respect to Ω gives

$$\hat{\Omega}(\theta)=\frac{1}{m}\sum_{j=1}^{m}\Lambda_j Re[I(\lambda_j)]\bar{\Lambda}_j$$

By substituting $\widehat{\Omega}(\theta)$ into $W(\theta,\Omega)$, this yields the concentrated likelihood function,

$$R(\theta)=W\left(\theta,\widehat{\Omega}(\theta)\right)=\log\left|\hat{\Omega}(\theta)\right|-2\left(\sum_{s=1}^{q}d_s\right)\frac{1}{m}\sum_{j=1}^{m}\log\left|\lambda_j\right|$$
$$+\sum_{s=1}^{q}\log(1-4\kappa_s(1-\kappa_s)\sin^2(-\gamma_s)) \qquad (3.4.4)$$

The local Whittle estimator of the parameter of interest, θ, can then be defined in terms of the concentrated likelihood

$$\hat{\theta}=\arg\min_{\theta\in\Theta}R(\theta) \qquad (3.4.5)$$

The space of the true parameter θ_0 is the compact set $\Theta=\Theta_d\times\Theta_\beta\times\Theta_\gamma$ which is defined in the next section.

3.5 Consistency

This section introduces assumptions on the spectral density and the bandwidth that is needed to establish the consistency of the LW estimator. These assumptions are mainly multivariate extensions of those of Robinson (1995). They are similar to the assumptions imposed by Lobato (1999), Shimotsu (2007) and Robinson (2008). The assumptions imposed are as follows,

Assumption 3.1: *The cross spectral density between f_{jt} and f_{kt} , as $\lambda\to 0$, satisfies*

$$\left|f_{jk}(\lambda)-\omega_{jk}|\lambda|^{-d_j-d_k}\sum_{l=1}^{4}\varsigma_l\exp[i\xi_l\,\mathrm{sgn}(\lambda)]\right|=o(\lambda^{-d_j-d_k})\,,\ \ j,k=1,\ldots,q$$

Assumption 3.2: $u_t=\sum_{j=0}^{\infty}C_j\varepsilon_{t-j},\sum_{j=0}^{\infty}\left\|C_j\right\|^2<\infty$

The innovations satisfy $E(\varepsilon_t|\mathcal{F}_{t-1})=0$ *and* $E(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1})=I_q$ *a.s.*

Assumption 3.3: $\theta\in int(\Theta)$, *where the compact set $\Theta = \Theta_d \times \Theta_\beta \times \Theta_\gamma$, where $\Theta_d = \{d: -\eta_1 \leq d \leq \frac{1}{2} - \eta_2\}$, Θ_β is an arbitrary large interval which includes {0} and $\Theta_\gamma = \{\gamma: -\eta_3 - \frac{\pi}{2} \leq \gamma \leq \frac{\pi}{2} - \eta_4\}$ and η_i are arbitrarily small positive numbers.*

Assumption 3.4: $\frac{1}{m} + \frac{m}{n} = o(1)$, *as* $n \to \infty$.

Assumption 3.5: $0 < \omega_{jk} < \sqrt{\omega_{kk}\omega_{jj}}$

Assumption 3.1 is a smoothness condition that is typically imposed in spectral analysis. A rate of convergence of $f_{jk}(\lambda)$ is imposed to $\lambda^{-d_j-d_k}$. Assumption 3.2 follow that of Robinson (1995) and Lobato (1999) in presenting the innovations in the Wold representation to be a square summable integrable martingale difference sequence that need not be strictly stationary, but satisfies a mild homogeneity restriction where the symbol $\|\cdot\|$ denotes the supremum (Euclidean) norm and $\mathcal{F}_{t-1}$ is the σ-field generated by ε_s, $s < t$. Assumption 3.2 also implies the existence of $f_u(\lambda)$. Assumption 3.3 states that the unknown parameter θ is an interior point to the compact set Θ. Assumption 3.4 restricts the rate of expansion of the bandwidth parameter m. The bandwidth must tend to infinity for consistency; however it must do so at slower rate compared to n in order to remain in the neighbourhood of the origin. Assumption 3.5 implies that Ω is positive definite and that the phase shift is identifiable. Under these assumptions, the following theorem can be set up which delivers the consistency of the LW estimator.

Theorem 1: *Assume that assumptions 3.1 through 3.5 hold, then*

$$\hat{\alpha} \xrightarrow{p} \alpha \qquad as\ n \to \infty$$

$$\hat{\beta} - \beta = O_p(\lambda^{-d_j-d_k}) \qquad as\ n \to \infty$$

Proof of Theorem 1: Consistency is derived by showing that the objective function converges uniformly in the parameter space to a limit which identifies all parameters and can thus be uniquely optimised. This proof follows mainly those of Robinson (1995), Lobato (1999), Shimotsu (2007) and Robinson (2008).

For any $c > 0$ and $\varepsilon > 0$, define the neighbourhoods which depend on s,

$N_\beta(c) = \{\beta_s: |b_s| < c\}$ where $b_s = \beta_s - \beta_{0s}$

$N_\gamma(c) = \{\gamma_s: |v_s| < c\}$ where $\gamma_s - \gamma_{0s}$

$N_d(c) = \{d_s: |\tau_s| < c\}$ where $\tau_s = d_s - d_{0s}$

$\bar{N}_\beta(c) = \Theta_\beta \backslash N_\beta(c)$

$\bar{N}_\gamma(c) = \Theta_\gamma \backslash N_\gamma(c)$

$\bar{N}_d(c) = \Theta_d \backslash N_d(c)$

$\Theta_\alpha = \Theta_\gamma \times \Theta_d$

$\bar{N}_\alpha(c) = \{\bar{N}_\gamma(c) \times \Theta_d\} \cup \{\Theta_\gamma \times \bar{N}_d(c)\}$

$N(\varepsilon) = N_\beta(\varepsilon^{-1}\left(\frac{m}{n}\right)^{d_j - d_k}) \times N_\gamma(\varepsilon) \times N_d(\varepsilon)$

$\bar{N}(\varepsilon) = \Theta \backslash N(\varepsilon)$

We have,

$$\Pr\left[\hat{\theta} \in \bar{N}(\varepsilon)\right] = \Pr\,[\inf_{\bar{N}(\varepsilon)} R(\theta) \leq \inf_{N(\varepsilon)} R(\theta)] \leq \Pr\,[\inf_{\bar{N}(\varepsilon)} S(\theta) \leq 0]$$

where $S(\theta) = R(\theta) - R(\theta_0)$

We need to show that,

$$\Pr\,[\inf_{\bar{N}(\varepsilon)} S(\theta) \leq 0] \to 0 \tag{3.5.1}$$

Then,

$$S(\theta) = \log\left|\widehat{\Omega}(\theta)\widehat{\Omega}(\theta)^{-1}\right| - 2\textstyle\sum_{s=1}^{q} \tau_s \sum_{j=1}^{m} \log \lambda_j$$

$$S(\theta) = \log\left|\Upsilon(d)\Xi(\theta)\widehat{\Omega}(\theta)\Xi(\theta)\Upsilon(d)\widehat{\Omega}(\theta)^{-1}\right| + u(d) \tag{3.5.2}$$

where

$$\Upsilon(d) = diag\{(2\tau_s + 1)^{\frac{1}{2}}\}$$

$$\Xi(\theta) = diag\{(\frac{2\pi m}{n})^{\tau_s}\}$$

$$u(d) = -2\sum_{s=1}^{q} \tau_s \left(\frac{1}{m}\sum_{j=1}^{m} \log j - \log m\right) - \sum_{s=1}^{q} \log(2\tau_s + 1)$$

$$= \sum_{s=1}^{q} \{2\tau_s - \log(2\tau_s + 1)\} - 2\tau_s(\frac{1}{m}\sum_{j=1}^{m} \log j - \log m + 1)\}$$

$$= \sum_{s=1}^{q} \{2\tau_s - \log(2\tau_s + 1)\} + O(\frac{\log m}{m})$$

since $\frac{1}{m}\sum_{j=1}^{m}\log j - \log m + 1 = O(\frac{\log m}{m})$ and $\tau_s = d_s - d_{0s}$

because $x - \log(x+1)$ achieves a unique global minimum on $(-1,\infty)$ at $x = 0$ and $x - \log(x+1) > \frac{x^2}{6}$ for $0 \le x < 1$, for all sufficiently large n,

$$\min_{\bar{N}_d(\varepsilon)} u(d) \ge \frac{d}{8}$$

(see, Robinson, 1995, p.1635).

The first term in (3.5.2) can be written as $\log\left|\Upsilon(d)\widehat{\Omega}^*(\theta)\Upsilon(d)\widehat{\Omega}(\theta)^{-1}\right|$ where

$$\widehat{\Omega}^*(\theta) = \Xi(\theta)\widehat{\Omega}(\theta)\Xi(\theta) = \frac{1}{m}\Xi(\theta)\Lambda_j B\sum_{j=1}^{m} Re(I(\lambda_j))\,B'\bar{\Lambda}_j\Xi(\theta)$$

$$H_j = \Lambda_j(I(\lambda_j))\bar{\Lambda}_j$$

As in Robinson (2008, p.23), Rearranging

$$\widehat{\Omega}^*(\theta) = \hat{G}^{(1)}(\alpha) + b_n(\beta)\hat{G}^{(2)}(\alpha) + b_n^2\hat{G}^{(3)}(\alpha) \tag{3.5.3}$$

$$b_n(\beta) = \left(\frac{2\pi m}{n}\right)^{\tau_s}(\beta_0 - \beta)$$

$$\hat{G}^{(i)}(\alpha) = \left(\hat{g}_{uv}^{(i)}\right)$$

Then the first term in (3.5.2) can be defined as $U_\alpha(\alpha) + U_\beta(\theta)$ where

$$U_\alpha(\alpha) = \Upsilon(d)\hat{G}^{(1)}(\alpha)\Upsilon(d)\hat{G}^{(1)}(\alpha)^{-1} \tag{3.5.4}$$

$$U_\beta(\theta) = \widehat{\Omega}^*(\theta)\hat{G}^{(1)}(\alpha)^{-1} \tag{3.5.5}$$

It suffices to show that as $n \to \infty$

$$\Pr\,[\inf_{\bar{N}_\alpha(\varepsilon)} U_\alpha(\alpha) \le 0] \to 0 \tag{3.5.6}$$

$$\Pr\,[\inf_{\bar{N}_\alpha\left(\varepsilon^{-1}(\frac{m}{n})^{d_j - d_k}\right)} U_\beta(\alpha) \le 0] \to 0 \tag{3.5.7}$$

To prove (3.5.6), it suffices to show following Lobato (1999) and Robinson (2008)

$$\sup_{\Theta_\alpha}\left\|\Upsilon(d)\{G^{(1)}(\alpha) - G^{(1)}(\alpha)\}\Upsilon(d)\right\| \xrightarrow{P} 0 \tag{3.5.8}$$

$$\sup_{\Theta_\alpha}\left\|[\Upsilon(d)G^{(1)}(\alpha)\Upsilon(d)]^{-1}\right\| < \infty \tag{3.5.9}$$

$$\inf_{\bar{N}_\gamma(\varepsilon)\times\Theta_d}\log\left|\Upsilon(d)G^{(1)}(\alpha)\Upsilon(d)G^{(1)}(\alpha)^{-1}\right| > 0 \tag{3.5.10}$$

Now to prove (3.5.7), $U_\beta(\alpha) = \log Q(b_n(\beta))$ where $Q(y) = 1 + \widehat{\aleph}_1 y + \widehat{\aleph}_2 y^2$, $\widehat{\aleph}_1 = \hat{g}_{uu}^{(i)} \hat{g}_{vv}^{(i)} - 2\hat{g}_{uv}^{(i)} \hat{g}_{uv}^{(i)} / \det\{G^{(1)}(\alpha)\}$ and $\widehat{\aleph}_2 = \hat{g}_{uu}^{(i)} \hat{g}_{vv}^{(i)} - \hat{g}_{uv}^{(i)^2} / \det\{G^{(1)}(\alpha)\}$ for all $\widehat{\aleph}_1$ and $\widehat{\aleph}_2 \geq 0$. Since $\widehat{\Omega}^*(\theta)$ and $\hat{G}^{(1)}(\alpha)$ are non-negative definite, then $Q(y)$ is non-negative for all real y, thus the probability on the left side of (3.5.7) is bounded. The proof is complete. □

3.6 Asymptotic Normality

Some further assumptions which need in deriving the asymptotic normality are listed. These assumptions are analogous to the assumptions in Lobato (1999), Shimotsu (2007) and Robinson (2008).

Assumption 3.6: *The cross spectral density between f_{jt} and f_{kt} , as$\lambda \to 0$, satisfies*

$$\left| f_{jk}(\lambda) - \omega_{jk} |\lambda|^{-d_j - d_k} \sum_{l=1}^{4} \varsigma_l \exp[\, i\xi_l \operatorname{sgn}(\lambda)] \right| = O(\lambda^{-d_j - d_k + \tau})$$

where $j, k = 1, \ldots, q$ *and* $\tau \in (0,1]$.

Assumption 3.7: *Assumption 3.2 holds with also the elements of ε_t having a.s. finite third and fourth moments and cross-moments, conditional on $\mathcal{F}_{t-1}$.*

Assumption 3.8: *Assumption 3.3 holds*

Assumption 3.9: $\frac{m^{1+2\tau}(\log m)^2}{n^{2\tau}} + \frac{\log n}{m} = o(1)$, *as* $n \to \infty$.

Assumption 3.10: *Assumption 3.5 holds*

Assumption 3.6 is the smoothness conditions used in the asymptotic theory of power spectral density estimates. This assumption imposes a rate of convergence on cross spectrum analogous to the one used in Robinson (1995) and Shimotsu (2007). Assumption 3.6 does not hold for $\tau > 1$. Assumption 3.7 implied that the innovations have third and fourth finite moments. Assumption 3.7 establishes that the process is linear with finite fourth moment. Assumption 3.8 is the same as assumption 3.3 (in section 3.5) that states that the unknown parameter θ is an interior point to the compact set Θ. Assumption 3.9 allows the bandwidth m to increase arbitrarily slowly with n, but it also imposes an upper bound on the rate of increase of m with n. It is similar to assumption 3.9 but slightly stronger.

Theorem 2. *Assume that conditions 3.6 through 3.10 hold, then*

$$\sqrt{m}(\hat{\theta} - \theta) \xrightarrow{d} N(0, E^{-1})$$

$\hat{\theta}$ is asymptotic normal with zero mean and asymptotic variance, E^{-1}, where

$$E = 2\left[\Omega_0 \odot (\Omega_0)^{-1} + I_q + \frac{\pi^2}{4}\left(\Omega_0 \odot (\Omega_0)^{-1} - I_q\right)\right]$$ [7]

Proof of Theorem2: To prove the asymptotic normality of the LW estimator given the assumptions 3.6 to 3.10 and from theorem 1 in section 3.5, $\hat{\alpha} \xrightarrow{p} \alpha$ implies that $x(\lambda_j, \hat{\alpha}) - x(\lambda_j, \alpha) = o_p(1)$ or tends to zero in probability as $n \to \infty$; this enables us to use the same proofs of asymptotic normality developed by Shimotsu (2007), Lobato(1999) and Robinson (2008).

The theorem established if, for any $q \times 1$ vector η, as $n \to \infty$,

$$\eta' \sqrt{m} \frac{\partial R(\theta)}{\partial \theta} \xrightarrow{d} N(0, \eta' E \eta) \tag{3.6.1}$$

$$\frac{\partial^2 R(\theta)}{\partial \theta \partial \theta'} \xrightarrow{p} E \tag{3.6.2}$$

The proof to show (3.6.1) is similar to that of Lobato (1999) and Shimotsu (2007). Consider,

$$\sqrt{m} \frac{\partial R(\theta)}{\partial \theta_u} = -\frac{2}{m} \sum_{j=1}^{m} \lambda_j + \operatorname{tr}\left[\hat{\Omega}(\theta)^{-1} \sqrt{m} \frac{\partial \hat{\Omega}(\theta)}{\partial \theta_u}\right] \tag{3.6.3}$$

Let i_u be a $q \times q$ matrix whose uth diagonal element is one and all other elements are zero, and let Λ_j denote $\Lambda_j(\theta)$.

As

$$\Lambda_j^{-1} = \operatorname{diag}\left\{\left|\lambda_j\right|^{-d_u} [\kappa_u e^{-i\operatorname{sng}(\lambda)\gamma_u} + (1 - \kappa_u) e^{i\operatorname{sng}(\lambda)\gamma_u}\right\}$$

and

$$\operatorname{Re}[(a + bi)(c + di)] = ac - bd$$

[7] The symbol ⊙ denotes for the hadamard product.

then

$$\sqrt{m}\frac{\partial\widehat{\Omega}(\theta)}{\partial\theta_u}=\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\log\lambda_j\,\mathrm{Re}\left[\left(\Lambda_j^0\right)^{-1}\left(i_u I_j+I_j i_u\right)\left(\overline{\Lambda}_j^0\right)^{-1}\right]$$

$$+\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}\mathrm{Im}\left[\left(\Lambda_j^0\right)^{-1}\left(-i_u I_j+I_j i_u\right)\left(\overline{\Lambda}_j^0\right)^{-1}\right]=H_{1a}+H_{2a}$$

Therefore (3.6.1) can be written as

$$\sum_{u=1}^{q}\eta_u\left\{-\frac{2}{m}\sum\log\lambda_j+\mathrm{tr}\left[\widehat{\Omega}(\theta_0)^{-1}H_{1a}\right]\right\}+\sum_{u=1}^{q}\eta_u\,\mathrm{tr}\left[\widehat{\Omega}(\theta_0)^{-1}H_{2a}\right]=M_1+M_2 \tag{3.6.4}$$

We follow the same procedures in Shimotsu (2007, p.296) to find an approximation of M_1 and M_2.

$$M_1=\sum\varepsilon_t'\sum\Psi_{t-s}\varepsilon_s+o_p(1) \tag{3.6.5}$$

$$\Psi_s=\frac{1}{\pi\sqrt{mn}}\sum_{j=1}^{m}\left(\log\lambda_j-m^{-1}\sum_{j=1}^{m}\log\lambda_j\right)\mathrm{Re}\left[E_j+E_j'\right]\cos(s\lambda_j)$$

$$M_2=\sum\varepsilon_t'\sum\widetilde{\Psi}_{t-s}\varepsilon_s+o_p(1) \tag{3.6.6}$$

$$\widetilde{\Psi}_s=\frac{1}{2\sqrt{mn}}\sum_{j=1}^{m}\left(\log\lambda_j-m^{-1}\sum_{j=1}^{m}\log\lambda_j\right)\mathrm{Re}\left[E_j+E_j'\right]\sin(s\lambda_j)$$

It follows from (3.6.5) and (3.6.6) that, with $z_t=0$,

$$\sum_{u=1}^{q}\eta_u\sqrt{m}\frac{\partial R(\theta_0)}{\partial\theta_u}=\sum_{t=1}^{n}z_t+o_p(1)$$

$$z_t=\varepsilon_t'\sum_{s=1}^{t-1}\left[\Psi_{t-s}\widetilde{\Psi}_{t-s}\right]\varepsilon_s$$

By standard martingale CLT, (3.6.1) follows if

$$\sum_{t=1}E(z_t^2|\mathcal{F}_{t-1})-\sum_{u=1}^{q}\sum_{v=1}^{q}\eta_u\eta_v E_{uv}\overset{P}{\to}0 \tag{3.6.7}$$

$$\sum_{t=1}E(z_t^2 I(|z_t|>\delta))\overset{P}{\to}0,\text{ for all }\delta>0 \tag{3.6.8}$$

(3.6.7) and (3.6.8) are proved in Robinson (1995, p.1645), Lobato (1999, pp.141-143) and Shimotsu (2007, pp.298-299). By establishing (3.6.7) and (3.6.8), (3.6.1) is proved.

To prove (3.6.2) which is similar to that in Lobato (1999) and Shimotsu (2007), observe that

$$\frac{\partial^2 R(\theta)}{\partial\theta_u \partial\theta_v'} = \operatorname{tr}\left[-\widehat{\Omega}(\theta)^{-1}\frac{\partial\widehat{\Omega}(\theta)}{\partial\theta_u}\widehat{\Omega}(\theta)^{-1}\frac{\partial\widehat{\Omega}(\theta)}{\partial\theta_v} + \widehat{\Omega}(\theta)^{-1}\frac{\partial^2\widehat{\Omega}(\theta)}{\partial\theta_u\partial\theta_v'}\right] \qquad (3.6.9)$$

where the derivatives of $\widehat{\Omega}(\theta)$ are given by

$$\frac{\partial\widehat{\Omega}(\theta)}{\partial\theta_u} = \frac{1}{m}\sum_{j=1}^{m}\operatorname{Re}\left[\left(\log\lambda_j + \frac{\lambda_j - \pi}{2}i\right) i_u (\Lambda_j)^{-1} I_j (\overline{\Lambda}_j)^{-1}\right]$$
$$+\frac{1}{m}\sum_{j=1}^{m}\operatorname{Re}\left[\left(\log\lambda_j + \frac{\lambda_j - \pi}{2}i\right)(\Lambda_j)^{-1} I_j (\overline{\Lambda}_j)^{-1} i_u\right]$$

and

$$\frac{\partial^2\widehat{\Omega}(\theta)}{\partial\theta_u\partial\theta_v'} = \frac{1}{m}\sum_{j=1}^{m}\operatorname{Re}\left[\left(\log\lambda_j + \frac{\lambda_j - \pi}{2}i\right)^2 i_u i_v (\Lambda_j)^{-1} I_j (\overline{\Lambda}_j)^{-1}\right]$$
$$+\frac{1}{m}\sum_{j=1}^{m}\operatorname{Re}\left[\left|\log\lambda_j + \frac{\lambda_j - \pi}{2}i\right|^2 i_u (\Lambda_j)^{-1} I_j (\overline{\Lambda}_j)^{-1} i_v\right]$$
$$+\frac{1}{m}\sum_{j=1}^{m}\operatorname{Re}\left[\left|\log\lambda_j + \frac{\lambda_j - \pi}{2}i\right|^2 i_v (\Lambda_j)^{-1} I_j (\overline{\Lambda}_j)^{-1} i_u\right]$$
$$+\frac{1}{m}\sum_{j=1}^{m}\operatorname{Re}\left[\left(\log\lambda_j + \frac{\lambda_j - \pi}{2}i\right)^2 (\Lambda_j)^{-1} I_j (\overline{\Lambda}_j)^{-1} i_u i_v\right]$$

Define for $h = 0, 1, 2$

$$\widehat{\Omega}_h(\theta) = \frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h \operatorname{Re}\left[\left(\Lambda_j\right)^{-1} I_j \left(\overline{\Lambda}_j\right)^{-1}\right] \qquad (3.6.10)$$

$$\widehat{\Omega}_h^*(\theta) = \frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h \operatorname{Im}\left[\left(\Lambda_j\right)^{-1} I_j \left(\overline{\Lambda}_j\right)^{-1}\right] \qquad (3.6.11)$$

then it follows that

$$\frac{\partial\widehat{\Omega}(\theta)}{\partial\theta_u} = i_u\widehat{\Omega}_1(\theta) + \widehat{\Omega}_1(\theta) i_u + \frac{\pi}{2} i_u \Omega_0^*(\theta) - \frac{\pi}{2} i_u \Omega_0^*(\theta) i_u + o_p[(\log n)^{-1}]$$

$$\frac{\partial^2\widehat{\Omega}(\theta)}{\partial\theta_u\partial\theta_v'} = i_u i_v \widehat{\Omega}_2(\theta) + i_u\widehat{\Omega}_2(\theta) i_v + i_v\widehat{\Omega}_2(\theta) i_u + \widehat{\Omega}_2(\theta) i_u i_v$$
$$+\frac{\pi}{4}\left[-i_u i_v \widehat{\Omega}_0(\theta) + i_u\widehat{\Omega}_0(\theta) i_v + i_v\widehat{\Omega}_0(\theta) i_u - \widehat{\Omega}_0(\theta) i_u i_v\right]$$
$$+\pi i_u i_v \widehat{\Omega}_1^*(\theta) - \pi\widehat{\Omega}_1^*(\theta) i_u i_v + o_p(1)$$

We can write (3.6.10) and (3.6.11) as follows uniformly in θ

$$\widehat{\Omega}_h(\theta) = \widehat{\Omega}_0 \frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h + o_p[(\log n)^{h-2}], \qquad \widehat{\Omega}_h^*(\theta) = o_p[(\log n)^{h-2}] \quad (3.6.12)$$

The second term in assumption 3.9 is necessary in showing (3.6.12) because the terms with $\widehat{\Omega}_1^*(\theta)$ do not cancel out even if we take the trace of $\widehat{\Omega}(\theta)^{-1}\frac{\partial^2\widehat{\Omega}(\theta)}{\partial\theta_u\partial\theta_v'}$. Define

$$\Omega_{01u} = i_u\Omega_0 + \Omega_0 i_u$$
$$\Omega_{02uv} = i_u i_v\Omega_0 + i_u\Omega_0 i_v + i_v\Omega_0 i_u - \Omega_0 i_u i_v$$
$$\Omega_{03uv} = -i_u i_v\Omega_0 + i_u\Omega_0 i_v + i_v\Omega_0 i_u - \Omega_0 i_u i_v$$

It follows from (3.6.12) as Shimotsu (2007, p.301), that we can obtain

$$\frac{\partial^2 R(\theta)}{\partial\theta_u\partial\theta_v'} = \mathrm{tr}\,\Omega_0^{-1}\,\Omega_{02uv} + \frac{\pi}{4}\Omega_0^{-1}\Omega_{03uv} + o_p(1)$$

and (3.6.2) follows. Since $\widehat{\Omega}(\theta)\xrightarrow{p}\Omega_0$ follows from (3.6.12). It remains to show (3.6.12) as in Shimotsu (2007, pp.301-302). Define

$$F_h(\theta) = \frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h\,\Lambda_j(\rho)^{-1}\Omega_0\overline{\Lambda}_j(\rho)^{-1}$$

where $\rho = \theta - \theta_0$, then (3.6.12) follows if

$$\sup_{\Theta}(\theta)\left\|\frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h\,\Lambda_j(\theta)^{-1}I_j\overline{\Lambda}_j(\theta)^{-1} - F_h(\theta)\right\| = o_p[(\log n)^{h-2}] \quad (3.6.13)$$

$$\sup_{\Theta}(\theta)\left\|F_h(\theta) - \Omega_0\frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h\right\| = o_p[(\log n)^{h-2}] \quad (3.6.14)$$

(3.6.14) can be written as

$$\sup_{\Theta}(\theta)\left\|\frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h\,\Lambda_j(\theta)^{-1}\left[\Lambda_j(\theta)^{-1}I_j\overline{\Lambda}_j(\theta)^{-1} - \Omega_0\right]\overline{\Lambda}_j(\theta)^{-1}\right\| \quad (3.6.15)$$

where the (u,v) term inside $\sup_{\Theta}(\theta)$ in (3.6.15) is equal to

$$\frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h\,|\lambda|^{-d_u-d_v}\sum_{l=1}^{4}\zeta_l\,e^{-i\mathrm{sng}(\lambda)\xi_l}\left[|\lambda|^{-d_u-d_v}\sum_{l=1}^{4}\zeta_l\,e^{-i\mathrm{sng}(\lambda)\xi_l}x(\lambda_j,\alpha)x^*(\lambda_j,\alpha) - \Omega_{0uv}\right]$$

and the (u,v) term inside $\sup_{\Theta}(\theta)$ in (3.6.14) is equal to

$$\frac{1}{m}\sum_{j=1}^{m}(\log\lambda_j)^h\left[|\lambda|^{-d_u-d_v}\sum_{l=1}^{4}\zeta_l\,e^{-i\mathrm{sng}(\lambda)\xi_l}x(\lambda_j,\alpha)x^*(\lambda_j,\alpha) - 1\right]\Omega_{0uv} \quad (3.6.16)$$

and hence (3.6.13), (3.6.14) and (3.6.16) are $o_p[(\log n)^{h-2}]$ as in Shimotsu (2007, p.302), then (3.6.12) is shown. The proof is complete. □

3.7 Simulation

This section reports some simulations that were conducted to examine the finite sample behaviour of the developed local Whittle estimator (or LWE). The following generating mechanism (a two sided moving average process), introduced in Davidson and Hashimzade (2008), is adopted in this chapter to generate fractionally cointegrated systems,

$$x_t = g(L; d, \kappa, a)u_t = \sum_{j=-\infty}^{\infty} b_j u_{t-j}$$

where

$$g(L; d, \kappa, a) = \kappa(1-L)^{a-d}(1-L^{-1})^{-a} + (1-\kappa)(1-L)^{-a}(1-L^{-1})^{a-d}$$

The lag structure is a convolution of the binomial series associated with fractional integration as follows where $j > 0$,

$$b_0 = \frac{1}{\Gamma(a)\Gamma(d-a)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(d-a+k)}{\Gamma(k+1)^2}$$

$$b_j = \frac{1}{\Gamma(a)\Gamma(d-a)} \sum_{k=0}^{\infty} \left(\frac{\kappa\Gamma(a+k)\Gamma(d-a+k+j)+(1-\kappa)\Gamma(a+k+j)\Gamma(d-a+k)}{\Gamma(k+1)\Gamma(k+1+j)} \right)$$

$$b_{-j} = \frac{1}{\Gamma(a)\Gamma(d-a)} \sum_{k=0}^{\infty} \left(\frac{(1-\kappa)\Gamma(a+k)\Gamma(d-a+k+j)+\kappa\Gamma(a+k+j)\Gamma(d-a+k)}{\Gamma(k+1)\Gamma(k+1+j)} \right)$$

Using Stirling's formula, it is easy to show that both,

$$\sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(d-a+k+j)}{\Gamma(k+1)\Gamma(k+1+j)} \sim Cj^{d-1}$$

and

$$\sum_{k=0}^{\infty} \frac{\Gamma(a+k+j)\Gamma(d-a+k)}{\Gamma(k+1)\Gamma(k+1+j)} \sim Cj^{d-1}$$

where C is a strictly positive constant. The relationship between the value of the memory parameter and the persistence of a shock can be understood in terms of the coefficient in the binomial expansion. Given the long memory parameter d, different choices of κ and a can affect the amount of short-run memory relative to long run, either symmetrically forwards and backwards or asymmetrically. The case $a = 0$, $\kappa = 1$ gives $g(L; d) = (1-L)^{-d}$ which corresponds to the one-sided model (lags but no leads). Another case is $a = 0$, $\kappa = 0$, which yields $g(L; d) = (1-L^{-1})^{-d}$ (leads but no lags). Moreover, when $a = \frac{d}{2}$, the model does not depend on κ. The symmetry property also holds for the case $\kappa = \frac{1}{2}$ (see Davidson and Hashimzade (2008)).

Accordingly, to satisfy the process (3.3.1) the following fractionally cointegrated system is generated for simulation,

$$\boldsymbol{B}\boldsymbol{z}_t = \begin{bmatrix} g(L; d_1, \kappa_1, a_1) & 0 \\ 0 & g(L; d_2, \kappa_2, a_2) \end{bmatrix} \boldsymbol{u_t I} \quad (3.7.1)$$

where

$$\boldsymbol{B} = \begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{z}_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{u_t} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

u_t is independently distributed and generated by $iidN(0, \Omega)$, where the diagonal elements of Ω were fixed to 1 and the off-diagonal elements of Ω, ρ, were selected to be (0.0, 0.4, 0.8). ρ is the correlation between u_{1t} and u_{2t}. The fractional parameters of interest (d_1, d_2) are set as (0.05, 0.4) and (0.2, 0.3). The former cases (0.05, 0.4) are close to what is expected in many practical situations; while, the latter describes a weaker form of cointegration where the memory parameters are closer. Sample size n is chosen to be $n = 512$ and $n = 1024$, and the bandwidth parameters chosen are $m = n^{0.5}$ and $m = n^{0.6}$ to check the robustness of the estimator due to changes in m. The Monte Carlo bias and root mean squared error (RMSE) of the local Whittle estimator are computed using 10,000 replications. The performance of the LW estimator is studied for different values of κ and a with fixed cointegrating parameter $\beta = 1$ in (3.7.1). The simulation results based on $a = 0$ and $\kappa = 1$, $a = 0$ and $\kappa = 0$ and $a = 0$ and $\kappa = 0.5$ are reported in the following tables (other values of κ and a were tried, but are not reported here). Simulations are done in Ox6.0 and TSM4.3.

Table 14: Simulation results for bias and RMSE where $a = 0$ and $\kappa = 1$

d_1	d_2	ρ	d_1		d_2		β		d_1		d_2		β	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
			$n = 512$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.05	0.4	0	-0.005	0.086	-0.008	0.093	0.007	0.085	-0.003	0.080	-0.006	0.088	0.005	0.079
0.05	0.4	0.4	0.003	0.081	0.005	0.082	0.005	0.080	0.002	0.078	0.004	0.082	0.005	0.077
0.05	0.4	0.8	0.002	0.079	0.002	0.078	0.005	0.081	0.002	0.077	0.003	0.080	0.004	0.075
0.2	0.3	0	-0.012	0.095	-0.011	0.096	0.008	0.086	-0.010	0.092	-0.011	0.098	0.006	0.081
0.2	0.3	0.4	0.009	0.090	0.010	0.092	0.007	0.084	0.006	0.085	0.008	0.094	0.007	0.084
0.2	0.3	0.8	0.009	0.091	0.008	0.087	0.005	0.080	0.007	0.089	0.008	0.093	0.008	0.083
			$n = 1024$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.05	0.4	0	-0.003	0.081	-0.004	0.082	0.005	0.085	-0.002	0.076	-0.002	0.077	0.003	0.076
0.05	0.4	0.4	0.002	0.074	0.002	0.075	0.004	0.082	0.000	0.066	0.001	0.069	0.004	0.079
0.05	0.4	0.8	0.000	0.065	0.001	0.071	0.004	0.080	0.000	0.064	0.000	0.063	0.002	0.070
0.2	0.3	0	-0.009	0.089	-0.010	0.092	0.006	0.087	-0.009	0.093	-0.008	0.095	0.004	0.081
0.2	0.3	0.4	0.008	0.085	0.008	0.087	0.008	0.085	0.006	0.088	0.007	0.093	0.006	0.091
0.2	0.3	0.8	0.006	0.084	0.009	0.089	0.007	0.082	0.004	0.085	0.005	0.084	0.006	0.090

Table 15: Simulation results for bias and RMSE where $a = 0$ and $\kappa = 0$

d_1	d_2	ρ	d_1		d_2		β		d_1		d_2		β	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
			$n = 512$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.05	0.4	0	-0.014	0.089	-0.013	0.085	0.009	0.090	-0.009	0.095	-0.007	0.094	0.007	0.094
0.05	0.4	0.4	0.010	0.090	0.011	0.080	0.008	0.084	0.008	0.092	0.006	0.091	0.008	0.096
0.05	0.4	0.8	0.016	0.091	0.015	0.081	0.007	0.087	0.011	0.096	0.006	0.097	0.005	0.082
0.2	0.3	0	-0.016	0.090	-0.018	0.095	0.011	0.094	-0.012	0.099	-0.012	0.095	0.005	0.083
0.2	0.3	0.4	0.011	0.089	0.020	0.093	0.009	0.089	0.011	0.084	0.011	0.098	0.006	0.089
0.2	0.3	0.8	0.020	0.097	0.021	0.098	0.006	0.084	0.015	0.106	0.013	0.102	0.007	0.086
			$n = 1024$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.05	0.4	0	-0.010	0.084	-0.009	0.080	0.007	0.088	-0.008	0.094	-0.005	0.091	0.005	0.092
0.05	0.4	0.4	0.007	0.088	0.009	0.078	0.006	0.086	0.006	0.098	0.004	0.095	0.006	0.090
0.05	0.4	0.8	0.012	0.087	0.012	0.082	0.005	0.087	0.009	0.103	0.003	0.098	0.008	0.091
0.2	0.3	0	-0.013	0.090	-0.016	0.089	0.008	0.091	-0.010	0.096	-0.009	0.101	0.007	0.089
0.2	0.3	0.4	0.010	0.079	0.015	0.087	0.009	0.087	0.008	0.090	0.008	0.095	0.009	0.085
0.2	0.3	0.8	0.014	0.094	0.016	0.092	0.004	0.080	0.011	0.109	0.010	0.099	0.008	0.083

Table 16: Simulation results for bias and RMSE where $a = 0$ and $\kappa = \frac{1}{2}$

d_1	d_2	ρ	d_1		d_2		β		d_1		d_2		β	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
			$n = 512$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.05	0.4	0	-0.013	0.089	-0.012	0.087	0.011	0.091	-0.008	0.096	-0.009	0.091	0.005	0.093
0.05	0.4	0.4	0.012	0.083	0.014	0.079	0.007	0.085	0.009	0.090	0.007	0.090	0.007	0.090
0.05	0.4	0.8	0.014	0.094	0.013	0.082	0.009	0.084	0.010	0.094	0.008	0.095	0.006	0.085
0.2	0.3	0	-0.013	0.095	-0.015	0.093	0.010	0.095	-0.011	0.103	-0.010	0.092	0.007	0.080
0.2	0.3	0.4	0.010	0.087	0.021	0.094	0.011	0.091	0.010	0.085	0.010	0.095	0.004	0.085
0.2	0.3	0.8	0.018	0.099	0.019	0.095	0.005	0.083	0.013	0.094	0.011	0.100	0.008	0.089
			$n = 1024$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.05	0.4	0	-0.009	0.080	-0.011	0.084	0.010	0.094	-0.007	0.090	-0.007	0.094	0.006	0.095
0.05	0.4	0.4	0.008	0.089	0.010	0.079	0.008	0.089	0.007	0.096	0.006	0.097	0.008	0.097
0.05	0.4	0.8	0.010	0.085	0.011	0.076	0.007	0.091	0.008	0.097	0.005	0.094	0.005	0.087
0.2	0.3	0	-0.011	0.084	-0.014	0.085	0.010	0.095	-0.011	0.099	-0.007	0.095	0.007	0.086
0.2	0.3	0.4	0.009	0.075	0.017	0.093	0.008	0.085	0.009	0.094	0.009	0.098	0.007	0.089
0.2	0.3	0.8	0.012	0.090	0.014	0.087	0.005	0.083	0.010	0.103	0.008	0.095	0.009	0.085

The results are summarised in tables 14-16. Table 14 reports the simulation results for $a = 0$ and $\kappa = 1$ where (3.7.1) which corresponds to lags-no leads model. The simulation results are reported in table 14 for bandwidth $m = n^{0.5}$ and $m = n^{0.6}$ for sample sizes $n = 512$ and $n = 1024$ and the correlation between u_{1t} and u_{2t} are set to different values 0, 0.4 and 0.8. The estimates of the memory parameters are d_1 and d_2 are discussed along with the cointegrating parameter β. The bias in the estimates d_1 and d_2 is very low and effectively diminishes (significantly decreases) when considering a larger sample. For most of the specifications, the bias is negative and it is uniformly lower than 0.012, in absolute value. The RMSE decreases in the bandwidth parameter which suggests that larger bandwidths are preferable. The RMSE shows that the memory parameters are estimated quite accurately. Looking at the case where correlation is 0.4 and 0.8, the same magnitude of bias on average and a bit lower RMSE is observed. Hence, the performance of the estimator improves with the presence of endogeneity for $\rho = 0.4$ and $\rho = 0.8$. Moreover, the estimates of β works well over different bandwidth and sample size, even for $(d_1, d_2) = (0.2,0.3)$ where $d_2 - d_1$ is small.

Similar results can be deduced from tables 15 and 16. Table 15 reports the simulation results for $a = 0$ and $\kappa = 0$ in which the data generating model (3.7.1) corresponds to leads-no lags. On the other hand, table 16 reports the simulation results for $a = 0$ and $\kappa = 0.5$ that corresponds to a symmetric two sided model. In (3.7.1), the cointegrating parameter is fixed to $\beta = 1$ for both cases. The estimates d_1 and d_2 suffer from severe bias, however it decreases for larger sample $n = 1024$. On the other hand, the RMSE reflects the bias and substantially increases in the bandwidth. In addition, results reports worse estimates for higher values of correlation $\rho = 0.4$ and $\rho = 0.8$. The estimates of beta in tables 15 and 16 are imprecise and have almost identical poor performance. In general, the simulation results for the LWE, tables 15 and 16, are significantly worse than that in table 14. Hence, the one-sided model for (3.7.1) is a better option for LWE.

Overall, the one sided model where $a = 0$ and $\kappa = 1$ dominates in terms of both bias and RMSE for all values of bandwidth and sample size and performs fairly stable especially where $(d_1, d_2) = (0.05, 0.4)$. In addition, the LW estimator improves and performs best with the presence of endogeneity for $\rho = 0.4$ and $\rho = 0.8$.

Furthermore, the Wald test with the above described simulation designs for the null hypotheses $\beta = 0, \kappa_1 = 1, a_1 = 0$ and $d_1 = 0$ is discussed in terms of size and power[8]. Table 17 reports the size for the hypotheses $\kappa_1 = 1$ and $a_1 = 0$ and the power for the hypotheses $\beta = 0$ and $d_1 = 0$ of Wald test for the one-sided model where $a_1 = 0$ and $\kappa_1 = 1$. The test on κ_1 and a_1 is over-sized, however it improves with the sample size and larger m. In addition, the sizes are better when $(d_1, d_2) = (0.05, 0.4)$. For the test on d_1, the power is very poor, but it seems that it increases as the bandwidth increases. On contrast, the power for testing β is very high and it increases as β moves away from the null. Overall, the Wald test is oversized in nearly all the cases, the LW estimator appears to work best with larger bandwidth and sample.

Table 17: Rejection frequency of Wald test with 5% level

					$n = 512$					$n = 1024$		
d_1	d_2	ρ	m	d_1	a_1	κ_1	β	m	d_1	a_1	κ_1	β
0.05	0.4	0	22	0.40	0.06	0.08	1.00	32	0.46	0.03	0.05	1.00
		0	42	0.63	0.04	0.05	1.00	64	0.65	0.02	0.03	1.00
		0.4	22	0.51	0.08	0.09	1.00	32	0.58	0.06	0.07	1.00
		0.4	42	0.78	0.06	0.08	1.00	64	0.85	0.03	0.03	1.00
		0.8	22	0.49	0.08	0.09	0.98	32	0.55	0.04	0.04	1.00
		0.8	42	0.75	0.06	0.06	1.00	64	0.79	0.02	0.03	1.00
0.2	0.3	0	22	0.38	0.08	0.13	0.79	32	0.45	0.05	0.07	0.91
		0	42	0.47	0.08	0.10	0.84	64	0.59	0.04	0.05	0.98
		0.4	22	0.43	0.10	0.12	0.88	32	0.64	0.09	0.05	0.89
		0.4	42	0.67	0.09	0.09	0.95	64	0.82	0.09	0.07	0.98
		0.8	22	0.45	0.12	0.10	0.85	32	0.65	0.10	0.09	0.94
		0.8	42	0.72	0.10	0.07	0.98	64	0.89	0.07	0.08	1.00

3.8 Empirical Examples

In this section, the LW estimation developed in the preceding sections is applied to three different empirically relevant examples. The examples considered are a bivariate

[8] The null hypothesis of a linear set of q restrictions $W = (\hat{\theta} - \theta)'\hat{E}(\hat{\theta} - \theta) \xrightarrow{d} \chi_q^2$, where $(\hat{\theta} - \theta)$ is a $q \times 1$ vector and the matrix $\hat{E}$ is a $q \times q$ covariance matrix obtained by replacing Ω_0 by the estimate $\hat{\Omega}(\hat{\theta})$ as $\hat{\Omega}(\hat{\theta}) \xrightarrow{p} \Omega_0$ (see theorem 2).

series and two trivariate series denoting by x_{1t} the variable chosen to be dependent. Suppose the following model that accommodates the trivariate series x_t,

$$Bx_t = u_t$$

where

$$B = \begin{bmatrix} 1 & -\beta_{12} & -\beta_{13} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} \text{ and } u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

When $d_1 \neq d_2 \neq d_3$, and $B = 0$, then x_{1t} , x_{2t} and x_{3t} have unequal memories d_1, d_2 and d_3 respectively. When $d_1 < d_2$ and/or d_3 and $B \neq 0$ (i.e. $\beta_{12} \neq 0$ or $\beta_{13} \neq 0$); then the unobservable linear combination $u_{1t} = x_{1t} - \beta_{12}x_{2t} - \beta_{13}x_{3t}$ has less memory of d_1 than the returns series, andx_{1t} , x_{2t} and x_{3t} are said to be fractionally cointegrated.

3.8.1 Inflation Rate Harmonization in the European Union

The consumer price indices (CPIs) of Spain and France are studied in this subsection. Hence, methods of calculating the CPI differs from one country to another, the harmonized index for consumer prices (HICP) developed in the European Union are used to examine for inflation rate harmonization in Spain and France. It is expected to find a stationary fractional cointegration between the inflation rates where evidence of long memory has been previously found, see Doornik and Ooms (2004) and Nielsen and Frederiksen (2010). A bivariate series of 159 observations on monthly inflation rates based on HICP of Spain and France is applied- $\pi_{SP,t}$ and $\pi_{FR,t}$ respectively. This data was obtained from Eurostat and spans from the January 1992 to April 2005.

Table 18: LM estimates under the assumption of no cointegration

Bandwidth	Spain $\pi_{SP,t}$	France $\pi_{FR,t}$
$m = n^{0.5}$	0.3136 (0.1382)	0.2057 (0.1438)
$m = n^{0.6}$	0.1080 (0.0987)	0.0912 (0.0975)
$m = n^{0.7}$	0.0790 (0.0721)	0.0643 (0.0706)

Note: The table shows long memory estimates for the LWEs with standard errors in the parentheses.

The estimates of the memory parameters are reported in Table 18 under the assumption of no cointegration. The results suggest that the presence of long memory properties in the inflation rates, however, the memory estimates decrease as the bandwidth chosen increases. Across the different bandwidth, it can be easily concluded that the inflation rates in Spain and France are stationary series with long memory. As the results so far are consistent in a sense that the inflation rates in Spain and France can be described as stationary series with long memory parameters, now let's proceed to conduct the multivariate local Whittle estimation on the fractionally cointegration system to estimate the memory and cointegrating parameters (see table 19).

Table 19: Application to inflation rates relation where $\kappa_i = 1$ and $a_i = 0$

Bandwidth	Parameters	Spain $x_{1t} = \pi_{SP,t}$	France $x_{2t} = \pi_{FR,t}$	Wald test $W_{d_1=0,\beta_{12}=1}$
$m = n^{0.5} = 12$	$\hat{d}_i$	0.0736	0.2098	2.3463
		(0.1023)	(0.1853)	
	$\hat{\beta}_{12}$		1.407	
			(0.1136)	
$m = n^{0.6} = 20$	$\hat{d}_i$	0.0960	0.2681	7.852*
		(0.0725)	(0.0816)	
	$\hat{\beta}_{12}$		1.029	
			(0.0945)	

Note: The table shows long memory and cointegrating estimates for the LWEs with standard errors in the parentheses. For the Wald tests, one or two asterisks denote significance at 5% or 1% respectively.

Table 19 report the results for the joint local Whittle estimation where $\kappa_i = 1$ and $a_i = 0$. Two different values for the bandwidth ($m = n^{0.5} = 12$ and $m = n^{0.6} = 20$) were used. The estimates of the memory parameters d_1 and d_2 along with cointegrating parameter β_{12} are reported, while the standard errors are reported in parentheses. The LW estimates specify that the inflation rate series can still be described as stationary long memory processes, and hence a stationary fractional cointegration is expected to be observed. The estimate of d_1 is smaller than d_2, and increases for larger bandwidth. The estimate of the cointegration coefficient β_{12} represents the existence of the fractional cointegration between the inflation rates. For the two different bandwidths 12 and 20, the estimates of the cointegration coefficient β_{12} are 1.407 and 1.029 respectively, which is higher than unity. Consequently, this implies that the long run inflation rate in Spain is much higher than that in France. The estimate of β_{12} decreases as the bandwidth increases.

The above results shows that the inflation rates relationship can be explained by a fractionally cointegrated system, as they tend to move together where the errors have less memory. Moreover, the last column in Table 19 reports the Wald test for the joint hypothesis $d_1 = 0$ and $\beta_{12} = 1$. The test does not reject the null hypothesis at the 5% level for $m = n^{0.5} = 12$. On the other hand, the null hypothesis is rejected for $m = n^{0.6} = 20$. Nevertheless, at 1% level the test does not reject the null hypothesis for both bandwidths $m = 12$ and 20. These results suggest the presence of long memory property in inflation rates with a unit cointegrating coefficient. In other words, the results imply that we cannot reject that the inflation series are stationary fractionally cointegrated, where d_1 is less that d_2. Finally, the estimates of a_i were captured (but not reported here) to reflect the different degree of lag and lead memory. The phase shifts among the inflation rates depends on the estimates of the parameters d_i and a_i. Overall, the results show an evidence of fraction cointegration among the inflation rates in Spain and France.

3.8.2 Volatility Relations in the Middle East

This subsection analyses the relation among the volatilities of the Morgan Stanley Capital International (MSCI) indices in the Middle East. A trivariate series is used of MSCI Jordan, MSCI Israel and MSCI Egypt covering three different markets in the Middle East, i.e. the frontier, developed and emerging markets respectively. The data is obtained from the Thomson One Banker database. The sample covers the period from March 2002 to February 2012 with 121 observations. The analysis is implemented on the volatilities of MSCI Jordan, Israel and Egypt denoted as $\sigma^2_{JRD,t}$, $\sigma^2_{ISR,t}$ and $\sigma^2_{EGY,t}$ respectively.

Table 20: LM estimates under the assumption of no cointegration

Bandwidth	MSCI Jordan	MSCI Israel	MSCI Egypt
	$\sigma^2_{JRD,t}$	$\sigma^2_{ISR,t}$	$\sigma^2_{EGY,t}$
$m = n^{0.5}$	0.4330	0.3546	0.3683
	(0.0867)	(0.0801)	(0.0832)
$m = n^{0.6}$	0.4582	0.4011	0.4130
	(0.0755)	(0.0712)	(0.0772)
$m = n^{0.7}$	0.3943	0.4266	0.4322
	(0.0534)	(0.0549)	(0.0583)

Note: The table shows long memory estimates for the LWEs with standard errors in the parentheses.

Table 20 shows that the memory estimates of the three volatilities which are very similar and seem to be stable across the different bandwidths chosen. The memory estimates suggest that the volatilities exhibit stationary long memory properties.

Table 21: Application to volatility relations where $\kappa_i = 1$ and $a_i = 0$

Bandwidth	Parameters	MSCI Jor. $x_{1t} = \sigma^2_{JRD,t}$	MSCI Isr. $x_{2t} = \sigma^2_{ISR,t}$	MSCI Egy. $x_{3t} = \sigma^2_{EGY,t}$	Wald test $W_{d_1=0,\beta_{12}=1}$
$m = n^{0.5} = 11$	$\hat{d}_i$	0.2016	0.4775	0.4299	2.0917
		(0.0624)	(0.0785)	(0.0881)	
	$\hat{\beta}_{1i}$		1.2190	0.3525	
			(0.1143)	(0.0727)	
$m = n^{0.6} = 17$	$\hat{d}_i$	0.2642	0.4899	0.3945	2.8644
		(0.0713)	(0.0880)	(0.0726)	
	$\hat{\beta}_{1i}$		1.0587	0.2759	
			(0.0955)	(0.0781)	

Note: The table shows long memory and cointegrating vector estimates for the LWEs with standard errors in the parentheses. For the Wald tests, one or two asterisks denote significance at 5% or 1% respectively.

Table 21 report the results for the joint local Whittle estimation for the bandwidth ($m = n^{0.5} = 11$ and $m = n^{0.6} = 17$), where $\kappa_i = 1$ and $a_i = 0$. The estimates of the memory and cointegrating parameters are reported, while the standard errors are reported in parentheses. The LW estimates show that the volatilities can be described as stationary long memory processes. The volatility of MSCI Jordan was chosen to be the dependent variable as it is expected to depend on the volatilities in both Israel and Egypt stock markets. The estimates of the memory parameters d_1 and d_2 along with cointegrating parameters β_{1j} (where $j = 2, 3$) are reported, while the standard errors are reported in parentheses. The estimate of d_1 increases as bandwidth increases on average and is smaller than d_2 and d_2. The stationary fractional cointegration relation between the volatilities of MSCI Jordan with that of MSCI Israel and Egypt indices is represented in the estimates of β_{1j}. The estimates of β_{12}, corresponding to that of the MSCI Israel, are 1.22 and 1.06; while, the estimates of β_{13}, corresponding to the MSCI Egypt, are 0.35 and 0.28. The cointegrating estimates between Jordan and Israel are close to 1. Accordingly, the results indicate that the volatility of MSCI Jordan follows and depends mostly on that of MSCI Israel.

The above results show that the volatilities in both MSCI Jordan and Israel can be explained by a fractionally cointegrated system. In Table 21, the Wald test are also reported for the joint hypothesis $d_1 = 0$ and $\beta_{12} = 1$. The test does not reject the null the hypothesis at the 5% level for both bandwidths $m = n^{0.5} = 11$ and $m = n^{0.6} = 17$. These results suggest that we cannot reject that the volatility relations are stationary fractionally cointegrated, where d_1 is less that d_2.

3.8.3 Returns on the US Treasury Rates

Finally, a potential long run relationship among the returns on the US treasury rates is examined. The model developed is applied to a trivariate series of 5,503 observations on the daily U.S. Treasury rate, at constant maturities of 2 years, 3 years and 7 years (Y2, Y3 and Y7 respectively). These series were obtained from the U.S. Department of the Treasury and the data spans from 02/01/1990 to 23/12/2011. The returns (differences of the log rates) of Y2, Y3 and Y7 are calculated and analysed.

Table 22: LM estimates under the assumption of no cointegration

Bandwidth	Y2	Y3	Y7
$m = n^{0.5}$	0.2092	0.2843	0.3237
	(0.0968)	(0.0984)	(0.0989)
$m = n^{0.6}$	0.2216	0.3075	0.3429
	(0.0876)	(0.0885)	(0.0891)
$m = n^{0.7}$	0.2591	0.3357	0.3722
	(0.0690)	(0.0693)	(0.0711)

Note: The table shows long memory estimates for the LWEs with standard errors in the parentheses.

Table 22 reports the estimates of the memory parameters for Y2, Y3 and Y5 under the assumption of no cointegration. The standard errors are reported in parentheses. The memory estimates suggests that all the returns are stationary series with long memories as d_i is between 0 and $\frac{1}{2}$ over different bandwidth chosen. As the memory parameters reflect the degree of the persistence of the series, the results indicate that the degree of persistence increases with maturity.

Table 23: Application to the US treasury returns where $\kappa_i = 1$ and $a_i = 0$

Bandwidth	Parameters	Y2	Y3	Y7	Wald test
		$x_{1t} = Y2_t$	$x_{2t} = Y3_t$	$x_{3t} = Y7_t$	$W_{d_1=0\ ,\beta_{12}=1}$
$m = n^{0.5} = 74$	$\hat{d}_i$	0.1724	0.3799	0.3285	16.2946*
		(0.6540)	(0.852)	(0.0833)	
	$\hat{\beta}_{1i}$		0.8193	0.3567	
			(1.0934)	(0.0908)	
$m = n^{0.6} = 175$	$\hat{d}_i$	0.2583	0.3162	0.3937	18.2445*
			(0.7958)	(0.0889)	
	$\hat{\beta}_{1i}$		0.7411	0.4486	
			(0.9766)	(0.0914)	

Note: The table shows long memory and cointegrating vector estimates for the LWEs with standard errors in the parentheses. For the Wald tests, one or two asterisks denote significance at 5% or 1% respectively.

Table 23 reports the results for the joint local Whittle estimation where $\kappa_i = 1$ and $a_i = 0$ with large bandwidths ($m = n^{0.5} = 74$) and ($m = n^{0.6} = 175$). Similar to the last two examples, the estimates of the memory and cointegrating parameters are reported (the standard errors are reported in parentheses). The estimate of d_1 is smaller than d_2 and d_3, and seems to be larger when bandwidth increases. On the other hand, the estimates of β_{12} are 0.82 and 0.74 and the estimates of β_{13} are 0.36 and 0.45 for the bandwidth $n^{0.5}$ and $n^{0.6}$ respectively. However, the estimates of β_{12}, corresponding to Y3, are very close to unity, unlike the estimates of β_{13} that corresponding to Y7. According to this results, we can expect strong cointegration between Y2 and Y3 compared to the one between Y2 and Y7 which concludes that the shorter time interval between the maturities, the stronger the evidence of fraction cointegration. In addition, the Wald tests for the joint hypothesis $d_1 = 0$ and $\beta_{12} = 1$ are reported. The test rejects the null the hypothesis at the 5% level for $m = 74$ and 175 which implies that we can reject that the US treasury returns are stationary fractionally cointegrated between Y2 and Y7.

3.9 Conclusion

This chapter proposes a semiparametric local Whittle estimator for multivariate stationary fractionally cointegrated systems. The multivariate framework is based on a spectral density that has both a real and complex part even at the origin, which accommodates for phase shifts. The phase of the proposed spectral density at the origin depends on the long memory parameters d and the parameter a which exhibits the degree of forward or backward memory.

Furthermore, the proposed multivariate semiparametric estimator was applied on three different empirically relevant examples to examine the presence of stationary fractional cointegration relationships. The analysis covered both bivariate and trivariate framework and can accommodate for a multivariate framework. The first example applied to a bivariate series of the inflation rates in Spain and France where a strong evidence of a fractional cointegration was found. Secondly, a trivariate series of the volatilities of three stock market indices in the Middle East (MSCI Jordan, MSCI Israel and MSCI Egypt) was examined and the results showed that the volatility of MSCI Jordan strongly follows that of MSCI Israel. Finally, the model developed is applied on another trivariate series of the daily US Treasury rate at constant maturities of 2 years, 3 years and 7 years and an evidence of a stationary fractional cointegration was reported between the returns with short time interval between the maturities.

CHAPTER 4

EXTENDED LOCAL WHITTLE ESTIMATION FOR STATIONARY FRACTIONALLY COINTEGRATED SYSTEMS

AMR ALGARHI*
UNIVERSITY OF EXETER

Abstract
This chapter introduces a semiparametric extended local Whittle (XLW) estimator to be applied throughout the stationary and non stationary regions of the memory parameters for fractionally cointegrated systems based on the analysis in chapter 3. The extended local Whittle estimator utilises the idea of extended discrete Fourier transform and periodogram as in Phillips (1999) and Abadir et al. (2007). The model employed in this chapter covers the multivariate framework. A Monte Carlo study exhibits the performance of the XLWE in non-stationary region. Finally, an empirical analysis of a trivariate series of US money aggregates (LTD_t, $M2_t$ and $M3_t$) is presented. It is found that there is strong evidence of fractional cointegration between the growth rates ($m2_t$ and $m3_t$) concluding that $m2_t$ can be considered as one of the main long-term contribution to the persistence properties in $m3_t$.

JEL Classification: C14, C32.

Keywords: Extended discrete Fourier transform; Fractional cointegration; Local Whittle; Money aggregates; Nonstationary; Semiparametric estimation.

* Amr Algarhi: a.s.i.algarhi@exeter.ac.uk
I would like to thank Professor James Davidson at University of Exeter for his constant guidance and suggestions, and audiences at the 2011 Economics PGR Conference at University of Exeter. Simulations and empirical applications were performed in Ox and TSM. All remaining errors are mine.

4.1 Introduction

This chapter develops a semiparametric local Whittle (LW) estimator to be applied throughout the stationary and non stationary zones in a multivariate framework. In chapter 3, a general LW estimator for multivariate stationary fractionally cointegrated system was introduced for the stationary case when the long memory parameter d is between 0 and $\frac{1}{2}$, by extending the work by Robinson (2008) on the bivariate LW estimator. Since the importance of larger values of d in economics as evidence from the number of papers in economics that have derived estimators that are robust to non-stationary values of d, it is a natural progress to modify the estimation procedures in chapter 3 to cover a wider range of d. The notion of the extended discrete Fourier transform (dFt) is utilised in this chapter to extend the multivariate LW estimator in chapter 3 to cover non stationary values of d.

In the univariate framework, the LW estimator is a common semiparametric estimator. Robinson (1995) proved its consistency and asymptotic normality for the stationary region where $d \in (-\frac{1}{2}, \frac{1}{2})$. Velasco (1999) extended Robinson's (1995) results to show that the estimator is consistent for $d \in (-\frac{1}{2}, 1)$ and asymptotically normally distributed for $d \in (-\frac{1}{2}, \frac{3}{4})$. Phillips and Shimotsu (2004) showed that the LW estimator has a non-normal limit distribution for $d \in (\frac{3}{4}, 1)$, and a mixed normal limit distribution for $d = 1$. When $d > 1$ the LW estimator converges in probability to unity as shown in many simulations studies and consequently is inconsistent. Therefore, the LW estimator is not a good general purpose estimator when d takes on values in the non-stationary region beyond ¾. The asymptotic theory is discontinuous at $d \in \{\frac{3}{4}, 1\}$ and the estimator is not consistent for $d > 1$. For the multivariate framework, Lobato (1999) derived a semiparametric two-step estimator for a long memory process for stationary values of d. Lobato and Velasco (2000) extended the results of Lobato (1999) by using tapering, and thereby allowing for non-stationary values of d and potential trends in the data generating process. Shimotsu (2007) derived a semiparametric estimator of multivariate fractionally integrated processes using a different form of the spectral density than that in Lobato (1999). The class of spectral densities included in Shimotsu's (2007) specification includes those of multivariate fractionally integrated processes, whereas the specification used in Lobato (1999) is an alternative local form of the spectral density that neglects the information in phase shifts and which will lead to less efficient estimates of the integration orders. Shimotsu (2007) showed that the estimator of Lobato (1999) is consistent given the more precise spectral density representation, but the limiting distribution is more evolved. Therefore, the estimator of Shimotsu (2007) has a smaller limiting distribution than the two-step estimator of Lobato (1999).

The contribution of this chapter is to develop a semiparametric LW estimator to be applied throughout the stationary and non stationary zones in a multivariate framework. Outside the stationary region, LW estimator is discontinuous at $d = \frac{3}{4}$ and hence the asymptotic theory is awkward to use because of non normal limit theory. It is obvious that the LW estimator is not a good general-purpose estimator when the value of d may take on values in the non stationary region. Several methods were introduced in the literature to avoid the problems when entering the non stationary region e.g. fractional differencing (the series is differenced before using the semiparametric estimator and then add the fractional differencing to the estimate) and tapering (the multiplication of an observed series by a sequence of constants -the taper- prior to Fourier transformation). However, each of these methods used has drawbacks. For data differencing, prior information is needed on the appropriate order of differencing, and data tapering has the disadvantage that the filter distorts the trajectory of the data and inflates the asymptotic variance.

In this chapter, the notion of the extended discrete Fourier transform is used to allow for the multivariate LW estimation to cover the non stationary region. The concept of the extending the dFt to the non stationary case is based on the work of Phillips (1999), Lahiri (2003), Dalla et al. (2006) and Abadir et al. (2007). Abadir et al. (2007) showed that when the $I(d)$ series is generated by a linear sequence the extended dFt and periodogram have the same asymptotic behaviour for $d \in (-\frac{3}{2}, \infty)$. A new extended local Whittle (XLW) estimator will be introduced to cover both stationary and non stationary case. The XLW estimator is identical to the LW estimator in the stationary region. Furthermore the model developed will be applied to three US money aggregates.

The remainder of the chapter is organised as follows. The next section introduces the extended local Whittle estimator. Section 4.3 and 4.4 establish the consistency and asymptotic normality for the XLWE. Section 4.5 reports and discusses the results of a Monte Carlo study, illustrating the finite sample behaviour of the XLWE. Section 4.6 demonstrates the empirical application to three US money aggregates and finally section 4.7 concludes.

4.2 The Extended Multivariate Local Whittle Estimator

In this section, the framework in the previous chapter is expanded to cover the non stationary region for the memory parameters, similar to that in Abadir et al. (2007). Firstly, assume a real-valued fractional process $\boldsymbol{X}_t = (X_{1,} \dots, X_q)'$ that is generated by the following model,

$$\mathbf{X}_t = \mathbf{G}_s(.)\mathbf{u}_t \qquad (4.2.1)$$

where,

$$\mathbf{G}_s(.) = diag\{\kappa_s(1-L)^{a_s-d_s}(1-L^{-1})^{-a_s} + (1-\kappa_s)(1-L)^{-a_s}(1-L^{-1})^{a_s-d_s}\}$$

and $u_t = (u_{1t}, \dots, u_{qt})'$ is serially independent Gaussian process with $(0, \sigma^2)$ whose spectral density is bounded and finite. The parameters d_s, κ_s, a_s are real valued, $0 < d_s < \frac{1}{2}$, $0 \le a_s \le \frac{d_s}{2}$ and $0 \le \kappa_s \le 1$ and $s = 1, \dots, q$. These parameters, similar to that in the previous chapter can be collected in $\alpha = (d', \kappa', a')'$ for $d' = (d_1, \dots, d_q)'$, $\kappa' = (\kappa_1, \dots, \kappa_q)'$ and $a' = (a_1, \dots, a_q)'$. The memory parameters, d_s, govern the long-run dynamics of the process and the behaviour of $f(\lambda)$ around the origin. Therefore, if empirical interest lies in the long-run dynamics of the process, it is useful to specify the spectral density only locally in the vicinity of the origin and avoid specifying the short-run dynamics of u_t explicitly. The spectral representation of the process (4.2.1) is

$$\mathbf{f}(\lambda) \sim \boldsymbol{\Lambda}_s^{-1}(d,\kappa,a)\boldsymbol{\Omega}\overline{\boldsymbol{\Lambda}}_s^{-1}(d,\kappa,a) \text{ , as } \lambda \to 0 \qquad (4.2.2)$$

$$\boldsymbol{\Lambda}_s^{-1}(d,\kappa,a) = diag\{|\lambda|^{-d_s}[\kappa_s e^{-i\text{sng}(\lambda)\gamma_s} + (1-\kappa_s)e^{i\text{sng}(\lambda)\gamma_s}]\}$$

where $\gamma_s = (a_s - \frac{d_s}{2})\pi$, $\text{sng}(\lambda) = 1$ if $\lambda \ge 0$ and -1 otherwise, $\boldsymbol{\Omega}$ is a $q \times q$ real, symmetric, finite and positive definite matrix, $\mathbf{f}(\lambda)$ has non-zero complex part. The notation "~" in equation (4.2.2) means that for each element, the ratio of real/imaginary parts of the left and right sides tend to 1. In addition, the over bar in

(4.2.2) denotes the complex conjugate. As in Davidson and Hashimzade (2008) and in the previous chapter, the cross-spectrum $f_{jk}(\lambda)$ in (4.2.2), where $j \neq k$, (off the diagonal elements of $f(\lambda)$), can be written as,

$$f_{jk}(\lambda) = \omega_{jk}|\lambda|^{-d_j-d_k} \sum_{l=1}^{4} \zeta_l \, e^{-i\mathrm{sng}(\lambda)\xi_l}$$

where

$$\zeta_1 = \kappa_j \kappa_k \qquad \xi_1 = \gamma_j + \gamma_k$$
$$\zeta_2 = \kappa_j(1-\kappa_k) \qquad \xi_2 = \gamma_j - \gamma_k$$
$$\zeta_3 = (1-\kappa_j)\kappa_k \qquad \xi_3 = -(\gamma_j - \gamma_k)$$
$$\zeta_4 = (1-\kappa_j)(1-\kappa_k) \qquad \xi_4 = -(\gamma_j + \gamma_k)$$

The memory parameters, d_s, emerge in both the power decay and the phase shift. The shift spectrum of X_{jt} and X_{kt} is nonzero and depends on d_j, d_k, $a_j$and a_k even at the zero frequency. To cover non-stationary values of d as well as stationary values, the framework of Abadir et al. (2007) is used in order to expand the set up in the previous chapter. Denote the Fourier frequencies by $\lambda_j = 2\pi j/n$, where $j = 1, \dots, n$. The extended periodogram is defined by

$$I_n(\lambda_j; d) = |w(\lambda_j, d)|^2 \tag{4.2.3}$$

where the extended dFt is

$$w(\lambda_j; d) = w_X(\lambda_j) + c(\lambda_j; d) \tag{4.2.4}$$

and has the property,

$$w(\lambda_j; d) = (1 - e^{i\lambda_j})^{-p} w_X(\lambda_j).$$

The extended dFt consists of the usual dFt ($w_X(\lambda_j)$),

$$w_X(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{it\lambda_j} \tag{4.2.5}$$

and a new correction term $c(\lambda_j; d)$ is a step function that takes constant values on the intervals $d \in \left[p - \frac{1}{2}, p + \frac{1}{2}\right)$, $p = -1, 0, 1, 2, \dots$ and is defined by

$$c(\lambda_j; d) = 0 \qquad \text{if } d \in \left[-\tfrac{1}{2}, -\tfrac{1}{2}\right)$$

$$c(\lambda_j; d) = e^{i\lambda_j} \textstyle\sum_{r=1}^{p} (1 - e^{i\lambda_j})^{-r} Z_r \quad \text{if } d \in \left[p - \tfrac{1}{2}, p + \tfrac{1}{2}\right), p = 1, 2, \dots$$

where

$$Z_r = \frac{1}{\sqrt{2\pi n}} ((1 - \mathrm{L})^{\mathrm{r}-1} X_t - (1 - \mathrm{L})^{\mathrm{r}-1} X_0), \; r = 1, \dots, p$$

Since the correction term is a step function, this feature facilitates the estimation procedures. Accordingly, the notion of the extended dFt allows estimating the usual

LW estimator in the context of non-stationary values of the memory parameter d. The XLW estimator is based on the behaviour of,

$$\eta_j = \frac{I(\lambda_j)}{f(\lambda_j)}, \text{ for } 1 \leq j \leq m \tag{4.2.6}$$

Following Lahiri (2003), Shimotsu (2007) and Abadir et al. (2007), the random variables in (4.2.6) satisfy,

$$E[\eta_j] = 1 + O(j^{-1}\log(j)), \text{ as } n \to \infty \tag{4.2.7}$$

$$Var[\eta_j] \leq K \tag{4.2.8}$$

$$Cov[\eta_j, \eta_k] \to 0, \text{ where } j \neq k \tag{4.2.9}$$

Equations (4.2.7), (4.2.8) and (4.2.9) imply that η_j satisfies the weak law of large numbers (WLLN) as $n \to \infty$

$$\frac{1}{m}\sum_{j=1}^{m} \eta_j \xrightarrow{p} 1$$

The extended local Whittle estimator, $\hat{\theta}$, can be derived by minimising the extended local Whittle objective function, $R(\theta, \Omega)$, over the admissible parameter space as follows,

$$R(\theta) = \log\left|\frac{1}{m}\sum_{j=1}^{m} \Lambda_j Re(I(\lambda_j))\bar{\Lambda}_j\right| - 2\left(\sum_{s=1}^{q} d_s\right)\frac{1}{m}\sum_{j=1}^{m}\log|\lambda_j| + \sum_{s=1}^{q}\log(1 - 4\kappa_s(1-\kappa_s)\sin^2(-\gamma_s))$$

$$\hat{\theta} = \arg\min_{\theta\in\Theta} R(\theta)$$

where $\theta = (\alpha', \beta')'$ and where $\beta = (1, \beta_{12}, \ldots, \beta_{1q})$. If the process is stationary the XLW estimator is identical to the LW estimator in chapter 3.

4.3 Consistency

Assumptions on the spectral density and the bandwidth are introduced to establish the consistency of the XLW estimator. These assumptions are mainly multivariate extensions and similar to those imposed by Robinson (2005) and Abadir et al. (2007). This section and the following one will focus on the memory parameter d_s, assuming the parameters $\kappa_s = 1$ and $a_s = 0$. Relaxing this condition can be a subject for further studies. The assumptions imposed are as follows,

Assumption 4.1: *The cross spectral density between f_{jt} and f_{kt}, as $\lambda \to 0$, satisfies*

$$\left|f_{jk}(\lambda) - \omega_{jk}|\lambda|^{-d_j-d_k}\sum_{l=1}^{4}\varsigma_l \exp[i\xi_l \operatorname{sgn}(\lambda)]\right| = o(\lambda^{-d_j-d_k}), \ \ j,k = 1, \ldots, q$$

Assumption 4.2: $d \in \Delta$, *where the compact set* $\Delta = [\frac{1}{2}, \frac{3}{2}]$.

Assumption 4.3: *The bandwidth parameter satisfies* $\frac{1}{m} + \frac{m}{n} = o(1)$, *as* $n \to \infty$.

Assumption 4.4: η_j *satisfies the weak law of large numbers (WLLN)* as $n \to \infty$

$$\frac{1}{m}\sum_{j=1}^{m} \eta_j \xrightarrow{p} 1$$

Assumption 4.5: $\{u_t\}$ is linear where, $u_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$, $\sum_{j=0}^{\infty} \left\| C_j \right\|^2 < \infty$
The innovations are martingale difference that satisfy $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ *and* $E(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = I_q$ *a.s.*

Assumption 4.1 is the usual smoothness condition that was also used in chapter 3. A rate of convergence of $f_{jk}(\lambda)$ is imposed to $\lambda^{-d_j - d_k}$. Assumption 4.2 states that the unknown parameter d is an interior point to the compact set Δ; however, the restriction $d \neq \frac{1}{2}$ is applied. The chosen interval is very important for economic data. Assumption 4.3 restricts the rate of expansion of the bandwidth parameter m. This assumption is used in chapter 3 for the LW estimator in the stationary region. Assumption 4.4 was discussed in the previous section. And finally assumption 4.5 follow that of Abadir et al. (2007) where the innovations are relaxed and assumed to be a martingale difference sequence. The following theorem can be set up which delivers the consistency of the XLW estimator.

Theorem 3: *Assume that assumptions 4.1 through 4.4 hold, then*

$$\hat{d} \xrightarrow{p} d \qquad\qquad as\ n \to \infty$$

Proof of Theorem 3: This proof follows mainly that of Abadir et al. (2007). For any $c > 0$, define the neighbourhood $N_d(c) = \{d: |d - d^0| \geq c\}$. It suffices to prove consistency we need to show,

$$\Pr\left[\inf_{N_d(c)} S(\theta) \geq \delta\right] \to 1$$

where, $S(\theta) = R(\theta) - R(\theta^0)$, θ^0 is an interior point in the parameter space Θ, and $\delta > 0$. Since $\frac{1}{m}\sum_{j=1}^{m} \log j + 1 = O(\frac{\log m}{m})$ then,

$$S(\theta) = \log \Xi(d) - \log \Xi(d^0) + 2\sum_{s=1}^{q} \theta_s + o(1)$$

where

$$\Xi(d) = \frac{1}{m}\sum_{j=1}^{m} Re[\Upsilon_j(d)\hat{\Omega}(\theta)I(\lambda_j)\bar{\Upsilon}_j(d)]$$

$$\Upsilon_{js}(d) = diag\{\left(\frac{j}{m}\right)^{d_s} [\kappa_s e^{-i\mathrm{sng}(\lambda)\gamma_s} + (1-\kappa_s)e^{i\mathrm{sng}(\lambda)\gamma_s}]\}$$

Hence by definition the extended periodogram matrix is

$$I(\lambda_j, d) = \left|w(\lambda_j, d)\right|^2 = w(\lambda_j, d)w^*(\lambda_j, d)$$

$$I(\lambda_j, d) = \left(w_X(\lambda_j) + c(\lambda_j; d)\right)\left(w_X(\lambda_j) + c(\lambda_j; d)\right)^*$$

From Abadir et al. (2007, p.1368), it follows that

$$w(\lambda_j; d^0) = \frac{w_X(\lambda_j)}{(1 - e^{i\lambda_j})}$$

and,

$$\tau(\lambda_j; d) = c(\lambda_j; d) - c(\lambda_j; d^0)$$

Consequently

$$I(\lambda_j, d) = \left|w(\lambda_j; d^0) + \tau(\lambda_j; d)\right|^2$$

Then $\Xi(d)$ can be written as,

$$\Xi(d) = \frac{1}{m}\sum_{j=1}^{m} Re[\Upsilon_j(d)\hat{\Omega}(\theta)\left(w(\lambda_j; d^0) + \tau(\lambda_j; d)\right)\left(w(\lambda_j; d^0) + \tau(\lambda_j; d)\right)^* \bar{\Upsilon}_j(d)]$$

Now let's set $\tau(\lambda_j; d) = 0$ as in Abadir et al. (2007, p.1368),

$$U(d) = \frac{1}{m}\sum_{j=1}^{m} Re[\Upsilon_j(d)\hat{\Omega}(\theta)w(\lambda_j; d^0)w^*(\lambda_j; d^0)\bar{\Upsilon}_j(d)]$$

Because of the uniform behaviour of $U(d)$, now let any $\varepsilon > 0$ and $d^{\varepsilon} = d^0 - \frac{1-\varepsilon}{2}$ where the s[th] element of ε is denoted by $\varepsilon_s > 0$ and very small. Then as proved using section 4.2 lemmas in Abadir et (2007)- particularly lemmas 4.2, 4.3 and 4.5 , we can deduce that,

$$\Xi(d) \geq (1 + o(1)) \odot U(\max\{d, d^{\varepsilon}\}),$$

where

$$U(\max\{d, d^{\varepsilon}\}) = \Lambda_j^{-1}[E(\theta^{\varepsilon}) \odot F(\theta^{\varepsilon})] \bar{\Lambda}_j^{-1}$$

where $\odot$ is the hadamard product. The matrices $E(\theta^{\varepsilon})$ and $F(\theta^{\varepsilon})$ have the elements $\sum_{l=1}^{4} \zeta_l \, e^{-i\xi_l}$ and $\left(1 + \gamma_j^{\varepsilon} + \gamma_k^{\varepsilon}\right)^{-1}$ respectively and hence for $\varepsilon > 0$,

$$S(\theta) \geq \log\left|\frac{U(\max\{d, d^{\varepsilon}\})}{U(d^0)}\right| + 2\sum_{s=1}^{q} \theta_s + o(1) \qquad (4.3.2)$$

$$= 2\sum_{s=1}^{q} \theta_s - \log|E(\theta^{\varepsilon}) \odot F(\theta^{\varepsilon})| + o(1) \qquad (4.3.3)$$

then,

$$2(\theta_j + \theta_k) - \log\left|Re\left[\sum_{l=1}^{4} \zeta_l \, e^{-i\xi_l}\left(1 + \gamma_j^{\varepsilon} + \gamma_k^{\varepsilon}\right)^{-1}\right]\right| \qquad (4.3.4)$$

$$\geq 2(\theta_j + \theta_k) - \log\left|Re\left[\sum_{l=1}^{4} \zeta_l \, e^{-i\varepsilon_l}\left(1 + \varepsilon_j + \varepsilon_k\right)^{-1}\right]\right| \qquad (4.3.5)$$

Therefore,

$$\inf_{N_d(c), d^{\varepsilon} \leq d} 2(\theta_j + \theta_k) - \log\left|Re\left[\sum_{l=1}^{4} \zeta_l \, e^{-i\xi_l}\left(1 + \gamma_j^{\varepsilon} + \gamma_k^{\varepsilon}\right)^{-1}\right]\right| \geq \delta > 0 \qquad (4.3.6)$$

Then from (4.3.2), (4.3.4) and (4.3.6), it is shown that (4.3.1) holds and the proof is complete. □

4.4 Asymptotic Normality

Some further assumptions which need in deriving the asymptotic normality are listed. These assumptions are analogous to the assumptions in Abadir et al (2007) and Shimotsu (2007).

Assumption 4.6: *The cross spectral density between f_{jt} and f_{kt} , as $\lambda \to 0$, satisfies*

$$\left| f_{jk}(\lambda) - \omega_{jk}|\lambda|^{-d_j - d_k} \sum_{l=1}^{4} \varsigma_l \exp[\, i\xi_l \operatorname{sgn}(\lambda)] \right| = O(\lambda^{-d_j - d_k + \tau})$$

where $j, k = 1, \ldots, q$ and $\tau \in (0,2]$.

Assumption 4.7: *Assumption 4.2 holds*

Assumption 4.8: $\frac{m^{1+2\tau}(\log m)^2}{n^{2\tau}} + \frac{\log n}{m} = o(1)$, *as* $n \to \infty$.

Assumption 4.9: *Assumption 4.5 holds with also the elements of* ε_t *having a.s. finite third and fourth moments and cross-moments, conditional on* $\mathcal{F}_{t-1}$.

Assumption 4.6 is the smoothness conditions. Assumption 4.7 is the same as assumption 4.2. Assumption 4.8 imposes an upper bound on the rate of increase of m with n which is similar to assumption 4.3 but slightly stronger. Assumption 4.9 implied that the innovations have third and fourth finite moments.

Theorem 4: *Assume that conditions 4.6 through 4.9 hold, then*

$$\sqrt{m}(\hat{d} - d) \xrightarrow{d} N(0, G^{-1}), \text{ where } G = 2\left[\Omega_0 \odot (\Omega_0)^{-1} + I_q + \frac{\pi^2}{4}\left(\Omega_0 \odot (\Omega_0)^{-1} - I_q\right)\right]$$

Proof of Theorem 4: Given the above assumptions, theorem 3, and the extended dFt in (4.2.4), the following result can be deducted as $n \to \infty$,

$$w(\lambda, \hat{d}) - w(\lambda, d^0) = o_p(1),$$

where d^0 is an interior point in Δ. As a result, Shimotsu (2007) proof of asymptotic normality can be used here as in section 3.6. Therefore, the same argument discussed in the proof of theorem 2 in the previous chapter (pp. 72-75) can applied here to prove that $\hat{d}$ is asymptotic normal with zero mean and asymptotic variance G^{-1}, where for any $q \times 1$ vector η, as $n \to \infty$,

$$\eta' \sqrt{m} \frac{\partial R(d)}{\partial d} \xrightarrow{d} N(0, \eta' G \eta) \tag{4.4.1}$$

$$\frac{\partial^2 R(\theta)}{\partial d \partial d'} \xrightarrow{p} G \tag{4.4.2}$$

Using the score approximation arguments in Shimotsu (2007, pp.295-299) to prove (4.4.1) and the Hessian approximation arguments (see Shimotsu 2007, pp.299-302) to prove (4.4.2). The proof is complete. □

4.5 Simulation

A Monte Carlo study is carried out to study the finite sample behaviour of the extended local Whittle estimator. The following bivariate framework is used to generate a two sided moving average processes $\mathbf{x}_t = (x_{1t}, x_{2t})'$,

$$\begin{bmatrix} 1 & -\beta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} g(L; d_1, \kappa_1, a_1) & 0 \\ 0 & g(L; d_2, \kappa_2, a_2) \end{bmatrix} \boldsymbol{u_t I} \tag{4.5.1}$$

and

$$\boldsymbol{u_t} \sim iidN(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}) \tag{4.5.2}$$

where

$$g(L; d, \kappa, a) = \kappa(1-L)^{a-d}(1-L^{-1})^{-a} + (1-\kappa)(1-L)^{-a}(1-L^{-1})^{a-d}$$

The experiment is carried out with the following specifications: the correlation ρ was set equal to 0, 0.4, 0.8. The memory parameters of interest (d_1, d_2) are set as (0.6, 0.6) and (0.6, 0.8). The stationary case is not reported here as the performance of the XLW estimator is expected to be equal to that of the LW estimator in chapter 3 for the stationary region. Series of length $n = 512$ and 1024 are generated, and the bandwidth parameters m chosen are $m = n^{0.5}$ and $m = n^{0.6}$ to check the robustness of the estimator due to changes in m. Simulations are done in Ox 6.0 and TSM 4.35. Throughout the simulation exercise, all the results are based on 10,000 replications. The performance of the LW estimator is examined for different values of κ and a (simulations are reported for the cases $a = 0$ and $\kappa = 1$, $a = 0$ and $\kappa = 0$ and $a = 0$ and $\kappa = 0.5$) with fixed cointegrating parameter $\beta = 1$.

Table 24: Simulation results for bias and RMSE where $a = 0$ and $\kappa = 1$

d_1	d_2	ρ	d_1		d_2		β		d_1		d_2		β	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
			$n = 512$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.6	0.6	0	0.004	0.064	0.006	0.069	0.005	0.071	0.004	0.061	0.005	0.067	0.003	0.067
0.6	0.6	0.4	0.006	0.067	0.008	0.081	0.006	0.075	0.005	0.062	0.008	0.075	0.005	0.071
0.6	0.6	0.8	0.007	0.070	0.008	0.079	0.005	0.073	0.006	0.065	0.006	0.071	0.005	0.068
0.6	0.8	0	0.006	0.069	0.006	0.072	0.007	0.072	0.007	0.079	0.007	0.081	0.005	0.074
0.6	0.8	0.4	0.009	0.073	0.007	0.075	0.008	0.078	0.007	0.066	0.006	0.070	0.006	0.074
0.6	0.8	0.8	0.011	0.075	0.009	0.078	0.010	0.081	0.008	0.072	0.009	0.075	0.009	0.078
			$n = 1024$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.6	0.6	0	0.002	0.060	0.004	0.066	0.004	0.065	0.003	0.058	0.002	0.061	0.001	0.054
0.6	0.6	0.4	0.004	0.062	0.005	0.074	0.005	0.072	0.005	0.060	0.003	0.067	0.003	0.063
0.6	0.6	0.8	0.005	0.066	0.007	0.073	0.007	0.078	0.007	0.064	0.004	0.065	0.005	0.064
0.6	0.8	0	0.004	0.067	0.004	0.066	0.006	0.070	0.005	0.076	0.005	0.071	0.003	0.065
0.6	0.8	0.4	0.007	0.068	0.006	0.074	0.005	0.075	0.007	0.061	0.007	0.076	0.005	0.070
0.6	0.8	0.8	0.010	0.074	0.009	0.077	0.009	0.085	0.009	0.076	0.011	0.083	0.005	0.067

Table 25: Simulation results for bias and RMSE where $a = 0$ and $\kappa = 0$

d_1	d_2	ρ	d_1		d_2		β		d_1		d_2		β	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
			$n = 512$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.6	0.6	0	0.007	0.069	0.008	0.073	0.006	0.075	0.006	0.073	0.005	0.079	0.006	0.081
0.6	0.6	0.4	0.007	0.070	0.008	0.084	0.008	0.078	0.005	0.076	0.007	0.086	0.006	0.084
0.6	0.6	0.8	0.009	0.073	0.009	0.083	0.007	0.076	0.008	0.075	0.008	0.089	0.005	0.078
0.6	0.8	0	0.011	0.089	0.010	0.084	0.010	0.085	0.010	0.094	0.008	0.087	0.012	0.086
0.6	0.8	0.4	0.013	0.091	0.009	0.087	0.011	0.092	0.009	0.095	0.007	0.092	0.009	0.099
0.6	0.8	0.8	0.015	0.098	0.013	0.085	0.012	0.093	0.012	0.096	0.012	0.087	0.010	0.095
			$n = 1024$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.6	0.6	0	0.004	0.061	0.006	0.072	0.004	0.074	0.004	0.064	0.006	0.074	0.006	0.078
0.6	0.6	0.4	0.005	0.064	0.007	0.077	0.007	0.070	0.004	0.069	0.005	0.072	0.006	0.076
0.6	0.6	0.8	0.006	0.070	0.008	0.076	0.007	0.071	0.005	0.072	0.007	0.079	0.009	0.083
0.6	0.8	0	0.009	0.085	0.008	0.082	0.008	0.073	0.008	0.086	0.007	0.089	0.010	0.086
0.6	0.8	0.4	0.011	0.086	0.007	0.081	0.010	0.085	0.010	0.087	0.006	0.086	0.008	0.082
0.6	0.8	0.8	0.013	0.094	0.011	0.079	0.013	0.089	0.012	0.093	0.008	0.083	0.010	0.095

Table 26: Simulation results for bias and RMSE where $a = 0$ and $\kappa = 0.5$

d_1	d_2	ρ	d_1		d_2		β		d_1		d_2		β	
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
			$n = 512$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.6	0.6	0	0.008	0.072	0.006	0.070	0.005	0.073	0.008	0.075	0.004	0.076	0.005	0.086
0.6	0.6	0.4	0.007	0.071	0.007	0.081	0.006	0.075	0.006	0.079	0.008	0.089	0.007	0.087
0.6	0.6	0.8	0.008	0.070	0.010	0.085	0.006	0.075	0.008	0.077	0.007	0.085	0.004	0.081
0.6	0.8	0	0.009	0.084	0.011	0.085	0.009	0.084	0.009	0.097	0.009	0.088	0.010	0.084
0.6	0.8	0.4	0.011	0.086	0.010	0.085	0.010	0.089	0.011	0.098	0.008	0.096	0.007	0.095
0.6	0.8	0.8	0.013	0.092	0.012	0.086	0.011	0.089	0.014	0.105	0.010	0.083	0.012	0.099
			$n = 1024$											
			$m = n^{0.5}$						$m = n^{0.6}$					
0.6	0.6	0	0.005	0.067	0.005	0.071	0.004	0.071	0.003	0.069	0.003	0.074	0.003	0.073
0.6	0.6	0.4	0.004	0.063	0.006	0.074	0.007	0.073	0.002	0.060	0.004	0.076	0.008	0.078
0.6	0.6	0.8	0.007	0.071	0.008	0.078	0.005	0.070	0.004	0.065	0.007	0.075	0.006	0.071
0.6	0.8	0	0.008	0.080	0.009	0.081	0.007	0.082	0.006	0.077	0.007	0.083	0.004	0.079
0.6	0.8	0.4	0.009	0.083	0.011	0.084	0.009	0.085	0.008	0.085	0.010	0.081	0.006	0.078
0.6	0.8	0.8	0.011	0.085	0.010	0.083	0.013	0.087	0.010	0.088	0.007	0.085	0.010	0.081

Table 24 reports the simulation results for the case for $a = 0$ and $\kappa = 1$ where bandwidth $m = n^{0.5}$ and $m = n^{0.6}$ for sample sizes $n = 512$ and $n = 1024$ are chosen and the correlation between u_{1t} and u_{2t} are set to different values 0, 0.4 and 0.8. The estimates of the memory parameters are d_1 and d_2 are discussed along with the parameter β. The extended LW estimator has little bias for the different values of d and significantly decreases when the sample size increases. The RMSE diminishes as bandwidth increases. The RMSE shows that the memory parameters are estimated quite accurately.

Tables 25 and 26 report the simulation results for the cases ($a = 0$ and $\kappa = 0$) and ($a = 0$ and $\kappa = 0.5$) respectively. The estimates d_1, d_2 and β have larger bias compared to that in table 24, however bias decreases for larger sample. The RMSE is substantially large, but decreases with larger bandwidth. Overall, the quality of the estimator does not seem to be affected by much for $\rho = 0.4$ and $\rho = 0.8$. The bias is slightly larger but the RMSE is higher than those when $\rho = 0$. The estimator becomes positively biased, and the value of the bias seems roughly constant across the different d. Regardless of the region studied, the estimator is unbiased and RMSE indicates that the memory parameter of interest is estimated accurately. In summary, there seems to be little doubt from these results that the XLW estimator is the best general purpose estimator over both stationary and non-stationary regions.

4.6 Empirical Example

The model, developed in section 4.2, were coded and performed in Ox 6.0 and TSM 4.35 where can take the values in the interval (0.5, 1.5) by assuming $p = 1$ in the correction term. In this case, the correction term in the extended dFt can be expressed as,

$$c(\lambda_j; d) = \frac{1}{\sqrt{2\pi n}} e^{i\lambda_j} \left(1 - e^{i\lambda_j}\right)^{-1} (X_t - X_0)$$

The time series data analysed, in this section, is the three US monetary aggregates. The data consists of 361 monthly observations from January 1976 to February 2006 and is obtained from the St Louis Federal Reserve Bank for three observed series, large-denomination time deposits (LTD), $M2$ and $M3$. The increment rates of the original series are calculated, ltd_t, $m2_t$ and $m3_t$ where $ltd_t = \Delta \log LTD$, $m2_t = \Delta \log M2$ and $m3_t = \Delta \log M3$. The larger aggregate $M3$ contains both $M2$ and LTD.

Figure 12: The increment rates of LTD, M2 and M3are calculated as the differences of the logarithm of the original data

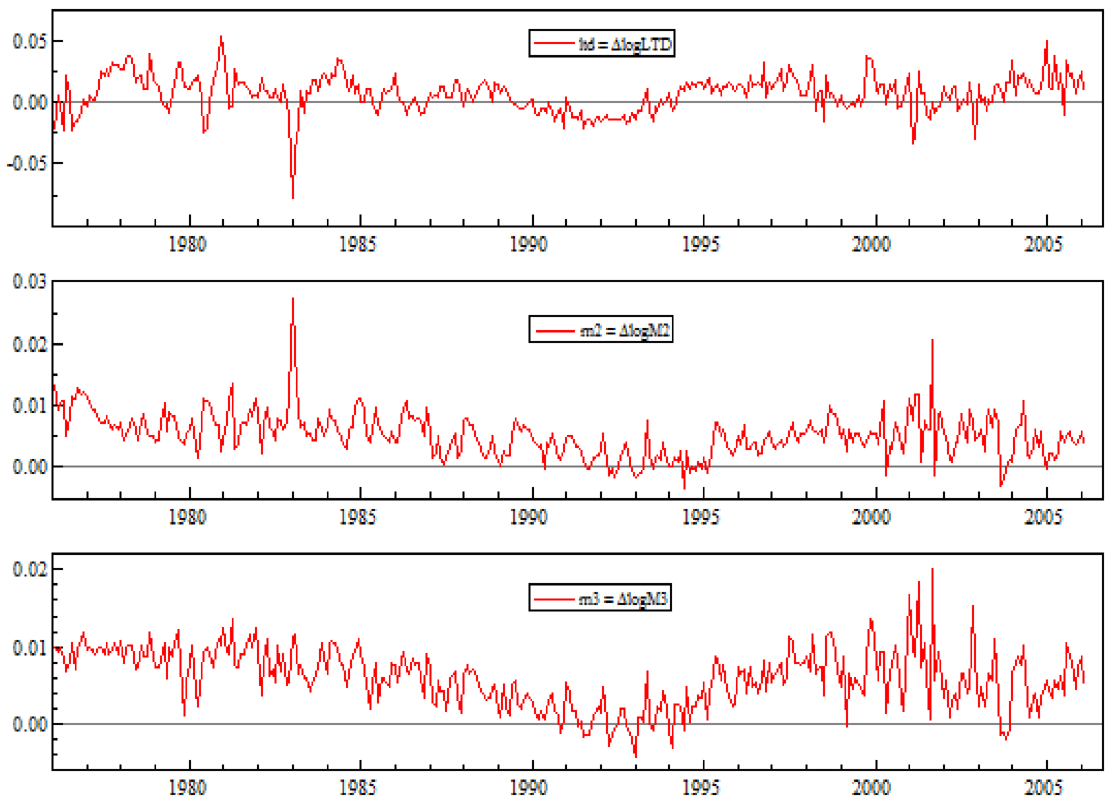


Table 27: LM estimates under the assumption of no cointegration

Bandwidth	ltd_t	$m2_t$	$m3_t$
$m = n^{0.5}$	0.3922 (0.0425)	0.7164 (0.0637)	0.6893 (0.0561)
$m = n^{0.6}$	0.4066 (0.0537)	0.6940 (0.0611)	0.6548 (0.0514)
$m = n^{0.7}$	0.3819 (0.0348)	0.6802 (0.0587)	0.6375 (0.0566)

Note: The table shows long memory estimates for the XLWEs with standard errors in the parentheses.

Figure 12 shows a plot of the trivariate series. Visual inspection suggests that the three series exhibit similar trend movements, particularly between $m2_t$ and $m3_t$ (downward trend until the mid of the 90s, when the trend is reverted). As a first step, the memory parameters d are estimated assuming no cointegration among the trivariate series. Table 27 reports the estimates of d over different bandwidth choices. It can be seen that the estimates of d are found to be decreasing as the bandwidth chosen increases in the three cases. The memory estimates suggest that ltd_t, $m2_t$ and $m3_t$ exhibit long memory properties. The results in Table 27 also demonstrate that ltd_t is stationary series, with less memory than $m2_t$ and $m3_t$. On the other hand, $m2_t$ and $m3_t$ are non stationary and have very similar memory levels. These results may show that the two series $m2_t$ and $m3_t$ can be cointegrated. Accordingly, the case of cointegration is considered next by relaxing the assumption $\beta_{1j} = 0$. The following model is adopted to accommodate the cointegration case,

$$\begin{aligned} m3_t - \beta_{12} ltd_t - \beta_{13} m2_t &= u_{1t} \\ ltd_t &= u_{2t} \\ m2_t &= u_{3t} \end{aligned}$$

Table 28: Application to the US money aggregates where $\kappa_i = 1$ and $a_i = 0$

Bandwidth	Parameters	$m3_t$	ltd_t	$m2_t$
$m = n^{0.5} = 19$	$\hat{d}_i$	0.3782	0.4019	0.8105
		(0.0891)	(0.0922)	(0.1630)
	$\hat{\beta}_{1j}$		0.5375	0.9286
			(0.0779)	(0.1749)
$m = n^{0.6} = 34$	$\hat{d}_i$	0.4669	0.5629	0.8991
		(0.1033)	(0.1352)	(0.1951)
	$\hat{\beta}_{1j}$		0.4599	0.8922
			(0.0718)	(0.1685)

Note: The table shows long memory and cointegrating estimates for the XLWEs with standard errors in the parentheses.

Table 28 demonstrates the results for the joint XLW estimation for $\kappa_i = 1$ and $a_i = 0$ with different values for bandwidth ($m = n^{0.5} = 19$ and $m = n^{0.6} = 34$). The estimates of the memory d_i and cointegrating parameters β_{1i} are reported, while the standard errors are reported in parentheses. The estimate of the cointegration coefficient β_{1j} represents the existence of the cointegration relation among the US monetary aggregates, where the estimates of β_{12}, corresponding to the cointegration relation between $m3_t$ and ltd_t, are 0.53 and 0.46 for $m = 19$ and 34 respectively. And the estimates of β_{13} that correspond to the cointegration relation between $m3_t$ and $m2_t$, take the values of 0.93 and 0.89 for the same bandwidths respectively. This indicates that the growth rates in $M3$ can be explained mainly by the increment rates in $M2$. This result is expected as $M2$ is a main component of the larger aggregate $M3$ and hence it would contribute to the persistence properties of its growth rate $m3_t$. In general, it seems that the memory parameters representing the degree of persistence in the US monetary aggregates have decreased as the chosen bandwidths increase. It can also be concluded that $m2_t$ is the main long-term contribution to $m3_t$, and their cointegration relationship might be cointegrated with ltd_t which can be considered as a second-order long term contribution to $m3_t$.

4.7 Conclusion

In this chapter, an extension of the multivariate LW estimator introduced in the previous chapter was proposed to cover potentially non-stationary multivariate fractional cointegrated system using the notion of the extended dFt and periodogram. The multivariate framework is based on a general form of the spectral density as introduced in Davidson and Hashimzade (2008), where the long memory parameters appear both in the slope and the phase of the spectral density at the origin. A simulation study confirms the asymptotic results. Finally, the proposed multivariate semiparametric estimation method is applied to the analysis of the US monetary aggregates.

Appendix

This appendix presents excerpts of the codes for the methods developed in this dissertation covering chapters 2, 3 and 4.

```
MakeB(const vParam)
{
	decl bee = zeros(2, 2);

	for(decl j = 0; j < 2; j++)
	for(decl k = 0; k < 2; k++)
		if (k != j)
		{
			bee[j][k] = vParam[3];
		}
		else bee[j][k] = 1;

	return bee;
}

BigPsi(const lamdaj, const vParam)
{
	decl bigpsi = new array[2];
	bigpsi[0] = bigpsi[1] = zeros(2, 2);

	decl lilpsi = fabs(lamdaj);


	decl lamdasgn;
	if(lamdaj < 0) lamdasgn = -1;
	else lamdasgn = 1;

	for (decl k = 0; k<2; k++)
	{
		decl d = vParam[k];
		decl gamma = vParam[2];
		decl Q = 1;
		if (k == 1) Q = gamma*lamdasgn;
		decl cosq = cos(Q);
		decl sinq = -sin(Q);

		bigpsi[0][k][k] = (lilpsi^d)*cosq;
		bigpsi[1][k][k]= (lilpsi^d)*sinq;
	}

	return bigpsi;
}
```

```
MakeB(const vParam, const cM)
{
	decl parpos = 3*cM;
	decl bee = zeros(cM, cM);

	for (decl j = 0; j<cM; j++)
```

```
	for (decl k = 0; k<cM; k++)
		if (k != j)
		{
			bee[j][k] = vParam[parpos];
			parpos++;
		}
		else bee[j][k] = 1;

	return bee;
}

LilPsi(const lamdaj)
{
	return fabs(lamdaj);
}

BigPsi(const lamdaj, const vParam, const cM)
{
	decl bigpsi = new array[2];
	decl lamsign;

	if (lamdaj < 0) lamsign = -1;
	else lamsign = 1;

	decl lilpsi =  LilPsi(lamdaj);
	bigpsi[0] = bigpsi[1] = zeros(cM, cM);

	for (decl k = 0; k<cM; k++)
	{
		decl d = vParam[3*k];
		decl kappa = vParam[3*k + 1];
		decl astar = vParam[3*k + 2];
		decl Q = M_PI*d*(1-astar)*lamsign/2;
		decl cosq = cos(Q);
		decl sinq = -(1 - 2*kappa)*sin(Q);
		decl mod = sqrt(cosq^2 + sinq^2);
		bigpsi[0][k][k] = (lilpsi^d)*cosq/mod;
		bigpsi[1][k][k] = (lilpsi^d)*sinq/mod;
	}

	return bigpsi;
}

AMat(const lamdaj, const perdg, const vParam)
{
	decl m = rows(perdg[0]);
	decl amat = 0;
	decl P0, P1;
	[P0, P1] = BigPsi(lamdaj, vParam, m);
	decl mB = MakeB(vParam, m);
	decl X0 = mB*perdg[0]*mB';
	decl X1 = mB*perdg[1]*mB';
	decl P0X0 = P0*X0;
	decl P0X1 = P0*X1;
	decl P1X0 = P1*X0;
	decl P1X1 = P1*X1;
	amat = P0X0*P0 + P1X0*P1 + P0X1*P1 - P1X1*P0;
	return amat;
}
```

```
EvalQfunc(const vParam, const mcX, const bMode)
{
	decl qfunc = 0;
	decl nobs = rows(mcX);
	decl neq = columns(mcX);
	decl cm = floor(sqrt(nobs));
	decl omegahat = 0, lilpsiterm = zeros(cm,1);
	decl amat = new array[cm];
	decl bigpsi = new array[cm];
	decl perdg;
	decl store = UserRetrieve();

	if (!bMode)
	{	perdg = new array[cm];
		for (decl j=1; j<cm; j++)

		{
			decl lamdaj = M_2PI*j/nobs;
			decl coslamt = cos(range(1, nobs)'*lamdaj);
			decl sinlamt = sin(range(1, nobs)'*lamdaj);
			decl w1star = sumc(mcX.*coslamt);
			decl w2star = sumc(mcX.*sinlamt);
			perdg[j] = new array[2];
			perdg[j][0] = (w1star'w1star + w2star'w2star);
			perdg[j][1] = (w2star'w1star - w1star'w2star);
		}
		store[0]= perdg;
		UserStore(store);
	}
	else perdg = UserRetrieve();

	decl sumds = 0;

	for (decl j=0; j<neq; j++) sumds += vParam[3*j];

	for (decl j=1; j<cm; j++)

	{	decl lamdaj	= M_2PI*j/nobs;
		omegahat += AMat(lamdaj, perdg[j], vParam);
		lilpsiterm[j-1] = 2*sumds*log(LilPsi(lamdaj));
	}

	omegahat /= cm;
	decl loglikcontr = -log(determinant(omegahat)) + lilpsiterm;
	return loglikcontr;
}

UserLikelihood(const mcDataset, const cStart, const cEnd, const vParam,
	const aName, const bMode)
{
	aName[0] = "Multifractional";
	decl x = <>, datalocs = VarNum(SERIES);
	for (decl j=0; j<columns(datalocs); j++)
		x ~= mcDataset[cStart:cEnd][datalocs[j]];
	return EvalQfunc(vParam, x, bMode);
}

MakeGam(const delta, const tee)
{
	decl h = zeros(tee+1, 1);
```

```
	h[0]=1;

	for (decl j=1; j<=tee; j++)
		h[j] = h[j-1]*(j - delta - 1)/j;
	return h;
}

UserGenerate(const mcX, const cStart, const cEnd, const vP,
	const aName, const bMode)
{
	for (decl k = 0; k<2; k++)
	{
		decl d = vP[3*k];
		decl kappa = vP[3*k + 1];
		decl a = vP[3*k + 2];
	}

	decl d, kappa, a;
	decl tee = cEnd - cStart + 1;
	decl r = 3*tee;
	decl shocks = zeros (r, 2);

	decl u1 = rann(r, 1)*aa;
	decl u2 = u1.*b + rann(r, 1).*c;
	shocks = u1 ~ u2;

	decl h, hh, store = UserRetrieve();

	if (!isarray(store)) store = new array[3];
	if (bMode) {	h = store[1];	}
	else
	{
		decl h_a = MakeGam( -a, tee)|zeros(tee,1);
		decl h_ad = MakeGam( a -d, tee)|zeros(tee,1);

		decl h_af = h_a;
		decl h_adf = h_ad;
		decl h_ab = h_a;
		decl h_adb = h_ad;

		decl h = zeros(1, 2*tee+1);
		h[tee+1] = h_a'h_ad;

		for (decl j=0; j<tee; j++)
		{
			h_af = 0 | h_af[:rows(h_af)-2];

			h_adf = 0 | h_adf[:rows(h_adf)-2];
			h[tee+j+1] = kappa*h_ad'h_af + (1-kappa)*h_a'h_adf;

		}
		for (decl j=0; j<tee; j++)
		{
			h_ab = 0|h_ab[:rows(h_ab)-2];
			h_adb = 0|h_adb[:rows(h_adb)-2];
			h[tee-j-1] = kappa*h_a'h_adb + (1-kappa)*h_ad'h_ab;
		}
		store[1] = h;
	}
```

```
	if (!isarray(store)) store = new array[3];
	if (bMode) {	hh = store[2];	}
	else
	{
		decl h_a = MakeGam( -a, tee)|zeros(tee,1);
		decl h_ad = MakeGam( a -d, tee)|zeros(tee,1);

		decl h_af = h_a;
		decl h_adf = h_ad;
		decl h_ab = h_a;
		decl h_adb = h_ad;

		decl hh = zeros(1, 2*tee+1);
		hh[tee+1] = h_a'h_ad;

		for (decl j=0; j<tee; j++)
		{
			h_af = 0 | h_af[:rows(h_af)-2];

			h_adf = 0 | h_adf[:rows(h_adf)-2];
			hh[tee+j+1] = kappa*h_ad'h_af + (1-kappa)*h_a'h_adf;

		}
		for (decl j=0; j<tee; j++)
		{
			h_ab = 0|h_ab[:rows(h_ab)-2];
			h_adb = 0|h_adb[:rows(h_adb)-2];
			hh[tee-j-1] = kappa*h_a'h_adb + (1-kappa)*h_ad'h_ab;

		}
		store[2] = hh;
	}

	decl x;
	for (decl j=tee; j<2*tee; j++)
	{
		decl x;

		x[j-tee][0] = h*shocks[j-tee: j+tee][0];
		x[j-tee][1] = hh*shocks[j-tee: j+tee][1];
		return x;
	}
	return x;
}
```

```
LWES(const y, const x, const gamma, const m0, const m1, const m2)
{
	decl k=columns(x);
	decl T=rows(x);
	decl Ixxr=zeros(k^2,T);
	decl Ixxi=zeros(k^2,T);
	decl Ixyr=zeros(k,T);
	decl lam=range(0,M_2PI-M_2PI/T,M_2PI/T);
	decl ic,jc,temp;
	for (jc=0; jc<k; jc++)
		{
		Ixyr[jc][]=PER(xdat[][jc]',y')[0][];
		for (ic=0; ic<k; ic++)
			{
```

```
                    temp=PER(xdat[][ic]',x[][jc]');
                    Ixxr[ic+jc*k][]=temp[0][];
                    Ixxi[ic+jc*k][]=temp[1][];
                    }
            }

        decl Fhatxx=shape(sumr(Ixxr[][1:m0]),k,k);
        decl Fhatxy=sumr(Ixyr[][1:m0]);
        decl lw0=invert(Omhatxx)*Fhatxy;
        Fhatxx=shape(sumr(Ixxr[][1:m2]),k,k);
        Fhatxy=sumr(Ixyr[][1:m2]);
        decl lw2=invert(Fhatxx)*Fhatxy;

        decl u=y-x*lw0;
        decl Ixur=zeros(k,T);
        decl Ixui=zeros(k,T);
        for (ic=0; ic<k; ic++)
            {
            temp=PER(x[][ic]',u');
            Ixur[ic][]=temp[0][];
            Ixui[ic][]=temp[1][];
            }

        decl dh=zeros(k+1,1);
        for (ic=0; ic<k; ic++) dh[ic]=EstLW(x[][ic],1,m1,FALSE)[0];
        dh[k]=EstLW(u,1,m1,FALSE)[0];


            decl data=x~u;
            decl perr=zeros((k+1)^2,T);
            decl peri=zeros((k+1)^2,T);
            decl OMEGA=zeros(k,k);
            decl K=zeros(k,k);
            decl H=zeros(k,1);
            decl J=zeros(k,k);
            decl Vhat=zeros(k,k);
            for (jc=0; jc<k+1; jc++)
                for (ic=0; ic<k+1; ic++)
                    {
                    temp=PER(data[][ic]',data[][jc]');
                    perr[ic+jc*(k+1)][]=temp[0][];
                    peri[ic+jc*(k+1)][]=temp[1][];
                    }
            for (jc=0; jc<k+1; jc++)
                for (ic=0; ic<k+1; ic++)
                    {
                    perr[ic+jc*(k+1)][]=perr[ic+jc*(k+1)][].*cos((lam-M_PI).*(dh[ic]-
dh[jc])/2)-peri[ic+jc*(k+1)][].*sin((lam-M_PI).*(dh[ic]-dh[jc])/2);
                    peri[ic+jc*(k+1)][]=perr[ic+jc*(k+1)][].*(lam.^(dh[ic]+dh[jc]));
                    }
            OMEGA=shape(meanr(perr[][1:m2]),k+1,k+1);
            for (jc=0; jc<k; jc++)
                {
                H[jc]=OMEGA[jc][k]*cos(M_PI_2*(dh[jc]-dh[k]))/(1-dh[jc]-dh[k]);
                for (ic=0; ic<k; ic++)
                    {
                    K[ic][jc]=OMEGA[ic][jc]*cos(M_PI_2*(dh[ic]-dh[jc]))/(1-dh[ic]-
dh[jc]);
```

```
	J[ic][jc]=OMEGA[ic][k]*OMEGA[k][jc]*cos(M_PI_2*(dh[ic]+dh[jc]-2*dh[k]))/(2-2*dh[ic]-2*dh[ic]-4*dh[k]);
				J[ic][jc]=J[ic][jc]+
OMEGA[ic][jc]*OMEGA[k][k]*cos(M_PI_2*(dh[ic]-dh[jc]))/(2-2*dh[ic]-2*dh[ic]-4*dh[k]);
				}
			}
		Vhat=invert(K)*J*invert(K)/m0;
		Vhat=Vhat.*lam[m0]^(-2*dhat[k]);
		Vhat=diag(lam[m0].^dhat[0:k-1])*Vhat*diag(lam[m0].^dh[0:k-1]);
	decl lwes=lw0|vec(Vhat);


	decl c,xd,yd,ud;
	if (gamma>=0) c=gamma;
	if (gamma==-1) c=dhat[k];
	if (c!=0)
		{
		xd=diffpow(x,c);
		yd=diffpow(y,c);
		for (jc=0; jc<k; jc++)
			{
			Ixyr[jc][]=PER(xd[][jc]',yd')[0][];
			for (ic=0; ic<k; ic++)
				{
				temp=PER(xd[][ic]',xd[][jc]');
				Ixxr[ic+jc*k][]=temp[0][];
				Ixxi[ic+jc*k][]=temp[1][];
				}
			}
		Fhatxx=shape(sumr(Ixxr[][1:m0]),k,k);
		Fhatxy=sumr(IxyR[][1:m0]);
		lwe0=invert(Fhatxx)*Fhatxy;
		Fhatxx=shape(sumr(Ixxr[][1:m2]),k,k);
		Fhatxy=sumr(Ixyr[][1:m2]);
		lwe2=invert(Fhatxx)*Fhatxy;
		ud=yd-xd*nbls0;
		for (ic=0; ic<k; ic++)
			{
			temp=PER(xd[][ic]',ud');
			Ixur[ic][]=temp[0][];
			Ixui[ic][]=temp[1][];
			}
		}

	decl Fhatxu=sumr(Ixur[][m0+1:m2]);
	Fhatxx=shape(sumr(Ixxr[][m0+1:m2]),k,k);
	decl OMEGAhat=invert(Fhatxx)*Fhatxu;
	decl lammat1=diag((ones(k,1)*lam[m1]).^dh[:k-1]);
	decl lammat2=diag((ones(k,1)*lam[m2]).^dh[:k-1]);
	decl lweshat=lwe2-((m1/m2)^dhat[k])*lammat2*invert(lammat1)*OMEGAhat;


	u=y-x*lweshat;
	decl du=EstLW(u,1,m1,FALSE)[0];
	if (gamma==-1) c=du;
	data=diffpow(xdat~udat,c);
	perR=zeros((k+1)^2,T);
	perI=zeros((k+1)^2,T);
	OMEGA=zeros(k,k);
```

```
	K=zeros(k,k);
	H=zeros(k,1);
	J=zeros(k,k);
	Vhat=zeros(k,k);
	for (jc=0; jc<k+1; jc++)
		for (ic=0; ic<k+1; ic++)
			{
			temp=PER(data[][ic]',data[][jc]');
			perr[ic+jc*(k+1)][]=temp[0][];
			peri[ic+jc*(k+1)][]=temp[1][];
			}
	for (jc=0; jc<k+1; jc++)
		for (ic=0; ic<k+1; ic++)
			{
			perr[ic+jc*(k+1)][]=perr[ic+jc*(k+1)][].*cos((lam-M_PI).*(dh[ic]-
dh[jc])/2)-peri[ic+jc*(k+1)][].*sin((lam-M_PI).*(dh[ic]-dh[jc])/2);
			perr[ic+jc*(k+1)][]=perr[ic+jc*(k+1)][].*(lam.^(dh[ic]+dh[jc]-2*dh[k]));
			}
	OMEGA=shape(meanr(perr[][1:m2]),k+1,k+1);
	for (jc=0; jc<k; jc++)
		{
		H[jc]=OMEGA[jc][k]*cos(M_PI_2*(dh[jc]-du))/(1-dh[jc]+du);
		for (ic=0; ic<k; ic++)
			{
			K[ic][jc]=OMEGA[ic][jc]*cos(M_PI_2*(dh[ic]-dh[jc]))/(1-dh[ic]-
dh[jc]+2*c);
			J[ic][jc]=OMEGA[ic][k]*OMEGA[k][jc]*cos(M_PI_2*(dh[ic]+dh[jc]-
2*dh[k]))/(2-2*dh[ic]-2*dh[ic]+4*c);
			J[ic][jc]=J[ic][jc]+ OMEGA[ic][jc]*OMEGA[k][k]*cos(M_PI_2*(dh[ic]-
dh[jc]))/(2-2*dh[ic]-2*dh[ic]+4*c);
			}
		}
	Vhat=invert(K)*J*invert(K)/m0;
	Vhat=Vhat.*lam[m0]^(-2*dhat[k]);
	Vhat=diag(lam[m0].^dh[0:k-1])*Vhat*diag(lam[m0].^dh[0:k-1]);
	lweshat=lweshat|vec(Vhat);


	return lwes~lweshat;
}
```

```
XLWfunc(const vParam, const mcX, const bMode)
{
	decl qfunc = 0;
	decl nobs = rows(mcX);
	decl neq = columns(mcX);
	decl cm = floor(sqrt(nobs));
	decl omegahat = 0, lilpsiterm = zeros(cm,1);
	decl amat = new array[cm];
	decl bigpsi = new array[cm];
	decl perdg;
	decl store = UserRetrieve();

	if (!bMode)
	{	perdg = new array[cm];
		for (decl j=1; j<cm; j++)

		{
			decl lamdaj = M_2PI*j/nobs;
```

```
				decl coslamt = cos(range(1, nobs)'*lamdaj);
				decl sinlamt = sin(range(1, nobs)'*lamdaj);
				decl coslam = cos(lamdaj);
				decl sinlam = sin(lamdaj);
				decl w1star = sumc(mcX.*coslamt)+ sumc(mcX.*coslam);
				decl w2star = sumc(mcX.*sinlamt)+ sumc(mcX.*sinlam);
				perdg[j] = new array[2];
				perdg[j][0] = (w1star'w1star + w2star'w2star);
				perdg[j][1] = (w2star'w1star - w1star'w2star);
			}
			store = perdg;
			UserStore(store);
		}
		else perdg = UserRetrieve();

		decl sumds = 0;

		for (decl j=0; j<neq; j++) sumds += vParam[3*j];

		for (decl j=1; j<cm; j++)

		{	decl lamdaj		= M_2PI*j/nobs;
			omegahat += AMat(lamdaj, perdg[j], vParam);
			lilpsiterm[j-1] = 2*sumds*log(LilPsi(lamdaj));
		}

		omegahat /= cm;
		decl loglikcontr = -log(determinant(omegahat)) + lilpsiterm;
		return loglikcontr;
}
```

Bibliography

ABADIR, K.M., W. DISTASO and L. GIRAITIS (2007), "Nonstationarity extended Local Whittle Estimation", *Journal of Econometrics* **141**, 1353-1384.

ADELMAN, I. (1965), "Long Cycles: Fact or Artifact?", *American Economic Review* **55**, 444-463.

ADENSTEDT, R., K. (1974), "On Large-sample Estimation for the Mean of a Stationary Random Sequence", *Annals of Statistics*, **2**, 1095-1107.

AGIAKLOGLOU, C., P. NEWBOLD and M. WOHAR (1992), "Bias in an Estimator of the Fractional Differencing Parameter", *Journal of Time Series Analysis*, **14**, 235-246.

BALL, L. and S.G. CECCHETTI (1990), "Inflation and Uncertainty at Short and Long Horizons", *Brookings Papers on Economic Activity*, 215-254.

BAILLIE, R.T. (1996), "Long Memory Processes and Fractional Integration in Econometrics", *Journal of Econometrics* **73**, 5–59.

BAILLIE, R.T., T. BOLLERSLEV and H.O. MIKKELSON (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* **74**, 3-30.

BAILLIE, R.T. and C.F. CHUNG (1993), "Small Sample Bias in Conditional Sum of Squares Estimators of Fractionally Integrated *ARMA* Models", *Empirical Economics* **18**, 791–806.

BAILLIE, R.T., C.F. CHUNG, and M.A. TIESLAU, (1992), "The Long Memory and Variability of Inflation: A Reappraisal of the Friedman Hypothesis," *Discussion Paper* **9246**, Tilburg University, Center.

BANERJEE A. and G. URGA (2005), "Modelling Structural Breaks, Long Memory and Stock Market Volatility: An Overview", *Journal of Econometrics* **129**, 1-34.

BARKOULAS. J.T and C.F. BAUM (2000), "Long Memory in the Greek Stock Market", *Applied Financial Economics* **10**, 177-184.

BARSKY, R.B. (1987), "The Fisher hypothesis and the Forecastability and Persistence of Inflation", *Journal of Monetary Economics* **19**, 3-24.

BERAN, R. (1992), "Statistical Methods for Data with Long-Range", *Statistical Science* **7**, 404-427.

BERAN, J. (1994a), *Statistics for Long Memory Processes*, Chapman and Hall, New York.

BERAN, J. (1994b), "On a class of M-estimators for Gaussian Long-Memory Models", *Biometrika* **81**, 755–766.

BLOOMFIELD, P. (1992), “Trends in Global Temperature”, *Climatic Change* **21**, 1-16.

BOX, G.E.P. and G.M. JENKINS (1971), *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day.

BROCK, W. and P. DE LIMA (1996), "Nonlinear Time Series, Complexity Theory and Finance", in MADDALA, G.S. and C.R. RAO (Ed.), *Handbook of Statistics **14**, Finance,* Amsterdam: North Holland.

BRUNNER, A.D. and G.D. HESS (1993), “Are Higher Levels of Inflation Less Predictable? A State-dependent Conditional Heteroskedasticity Approach”, *Journal of Business and Economic Statistics* **11**,187-197.

CAVALCANTE, J. and A. ASSAF (2002), “Long Range Dependence in the Returns and Volatility of the Brazilian Stock Market”, Manuscript, Rio de Janeiro.

CHEUNG, Y. and K. LAI (1995), "A Search for Long Memory in International Stock Market Returns", *Journal of International Money and Finance* **14**, 597-615.

CHRISTENSEN, B.J. and M.O.NIELSEN (2004),"Asymptotic Normality of Narrow band Least Squares in the Stationary Fractional Cointegration Model and Volatility Forecasting", *Journal of Econometrics* **133**, 343–371.

CHUNG, C.F. (1994), “A Note on Calculating the Autocovariance of the Fractionally Integrated *ARMA* Models”, *Economics Letters* **45**, 293-297.

DAVIDSON, J.E.H. and N. HASHIMZADE (2008), “Alternative Frequency and Time Domain Versions of Fractional Brownian Motion”, *Econometric Theory* **24**, No. 1, 256-293.

DALLA, V., L. GIRAITIS, and J. HIDALGO (2004), “Consistent Estimation of the Long Memory Parameter for Nonlinear Time Series”, *J. Time Series Analysis* **27**, 211-251.

DIEBOLD, F.X. and A. INOUE (2001), “Long Memory and Regime Switching”, *Journal of Econometrics* **105**, 131–159.

DIEBOLD, F.X. and G.D. RUDEBUSCH (1989), "Long Memory and Persistence in Aggregate Output", *Journal of Monetary Economics* **24**, 189-209.

DING, Z. and C.W.J. GRANGER (1996), "Modelling Volatility Persistence of Speculative Returns: A New Approach", *Journal of Econometrics* **73**, 185-215.

DOORNIK, J.A. and D.F. HENDRY (2001), *Econometric Modelling Using PcGive*, Volume III, Timberlake Consultants Press, London.

DOORNIK, J.A. and M. OOMS (2003), “Computational Aspects of Maximum Likelihood Estimation of Autoregressive Fractionally Integrated Moving Average Models”, *Computational Statistics and Data Analysis* **41**, 333–348.

DOORNIK, J.A. and M. OOMS (2004), “Inference and forecasting for ARFIMA models with an application to US and UK inflation”, *Studies in Nonlinear Dynamics and Econometrics* **8**, No. 2, Article 14.

ENGLE, R.F. and C.W.J. GRANGER (1987), “Co-integration and Error Correction: Representation, Estimation, and Testing”, *Econometrica* **55** (**2**), 251-276.

GEWEKE, J. and S. PORTER-HUDAK (1983), “The Estimation and Application of Long Memory Time Series Models”, *Journal of Time Series Analysis*, **4**, 221-238.

GRANGER, C.W.J. (1966), “The Typical Spectral Shape of an Economic Variable”, *Econometrica* **34 (1)**, 150–61.

GRANGER, C.W.J. and R. JOYEUX (1980), “An introduction to Long Memory Time Series Models and Fractional Differencing”, *Journal of Time Series Analysis* **I**, 15-39.

GUÉGAN, D. (2005), “How can we define the concept of Long Memory? An Econometric survey”, *Econometric Reviews* **24**, 113–149.

HAMPEL, F. (1987), “Data Analysis and Self-Similar Processes”, invited paper, ISI, Tokyo.

HANNAN, E.J. (1973), “The asymptotic theory of linear time series models”, *Journal of Applied Probability* **10**, 130-145.

HASSLER, U. and J. WOLTERS (1994), “On the power of Unit Roots against Fractional Alternatives”, *Economics Letters* **45**, 1–5.

HASSLER, U. and J. WOLTERS (1995), “Long Memory in Inflation Rates: International Evidence”, *Journal of Business and Economic Statistics* **13** (**1**), 37-45.

HASSLER, U., F. MARMOL and C. VELASCO (2006), "Residual Log-periodogram Inference for Long-run Relationships", *Journal of Econometrics* **130**, 165-207.

HAUBRICH, J.G. and A.W. LO (1989), "The Sources and Nature of Long- Term Memory in the Business Cycle", *NBER Working Paper* **2951**, Cambridge, MA.

HAUSER, M. A. (1998), "Maximum Likelihood Estimators for ARMA and ARFIMA Models: A Monte Carlo Study", Working paper, University of Economics and Business Administration, Department of Statistics, Vienna, forthcoming in *Journal of Statistical Planning and Inference*.

HAUSER, M.A., B.M. PÖTSCHER and E. RESCHENHOFER (1999), “Measuring Persistence in Aggregate Output: ARMA Models, Fractionally Integrated ARMA Models and Nonparametric Procedures”, *Empirical Economics* **22**, 243-269.

HENRY, M. (2001), “Averaged Periodogram Spectral Estimation with Long Memory Conditional Heteroscedasticity”, *Journal of Time Series Analysis* **22**, 431–459.

HIPEL, K. W. and A. I. MCLEOD (1978a), “Preservation of the Rescaled Adjusted Range, 1: A Reassessment of the Hurst phenomenon”, *Water Resources Research* **14**, 491-508.

HIPEL, K. W. and A. I. MCLEOD (1978b), “Preservation of the Rescaled Adjusted Range, 3: Fractional Gaussian Noise Algorithms”, *Water Resources Research* **14**, 517-518.

HIPEL, K. W. and A. I. MCLEOD (1978c), “Preservation of the Rescaled Adjusted Range, 2: Simulation studies using Box-Jenkins models”, *Water Resources Research* **14**, 509-516.

HOSKING, J.R.M. (1981), “Fractional Differencing”, *Biometrika* **68**, 165-176.

HURST, H.E. (1951), “Long-term Storage Capacity of Reservoirs”, *Transactions of the American Society of Civil Engineers* **116**, 770-799.

HURST, H.E. (1956), “Methods of Using Long Term Storage in Reservoirs”, *Proceedings of the Institute of Civil Engineers* **1**, 519-543.

HURST, H.E. (1957), “A suggested Statistical Model of some Time Series that occur in Nature”, *Nature* **180**, 494.

KANG, S.H. and S.M. YOON (2008), "Long Memory Features in the High Frequency Data of the Korean Stock Market", *Physical Review* **387**, 5189-5196.

KIM, C.J. (1993), "Unobserved component time series models with Markov switching heteroskedasticity: change in regime and the link between inflation rates and inflation uncertainty", *Journal of Business and Economic Statistics* **11**, 341-349.

KIRK-DAVIDOFF, D. and C. VAROTSOS (2006), “Long Memory Processes in Global Ozone and Temperature Variations at the Region 60º S– 60º N,” *Atmospheric Chemistry and Physics* **6**, 4093–4100.

KLEIN, B. (1976), "The Social Costs of the recent Inflation: the Mirage of Steady Anticipated Inflation", Carnegie-Rochester Conference Serieson Public Policy 3, 185-212.

KOLMOGOROV, A.N. (1940), “Wienersche Spiralen und einige andere interessante Kurven im Hilberschen raum”, *Comptes Rendus (Doklady) de l’ Academie des Sciences de l’URSS (N.S.)* **26**, 115-118.

KUNSCH, H.R. (1987), “Statistical Aspects of Self-similar Processes”, in Y. Prokhorov and V.V. Sazanov (eds), *Proceedings of the First World Congress of the Bernoulli Society*, VNU Science Press, Utrecht, 67-74.

KWIATKOWSKI, D., P.C.B. PHILLIPS, P. SCHMIDT AND Y. SHIN (1992), “Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure are We That Economic Time Series Have a Unit Root?”, *Journal of Econometrics* **54**, 159-178.

LAHIRI, S.N. (2003), “A Necessary and Sufficient Condition for Asymptotic Independence of Discrete Fourier Transforms under Short- and Long-range Dependence”, *Annals of Statistics* **31**,613-641.

LAWRANCE, A.J. and N.T. KOTTEGODA (1977), “Stochastic Modelling of River Flow Time Series”, *Journal of the Royal Statistical Society* **A 140**, 1-47.

LO, A.W. (1991), “Long Term Memory in Stock Market Prices”, *Econometrica* **59**, 1279–1313.

LOBATO, I.N. (1999), “A Semiparametric Two-step Estimator in a Multivariate Long Memory Model”, *Journal of Econometrics* **90**, 129-153.

LOBATO, I.N. and P. ROBINSON (1996), “Averaged Periodogram Estimation of Long Memory”, *Journal of Econometrics* **73**, 303-324.

LOBATO, I.N. and N.E. SAVIN (1998),"Real and Spurious Long-Memory Properties of Stock Market Data", *Journal of Business and Economics Statistics* **16**, 261-268.

LOBATO, I.N. and C. VELASCO (2000), “Long Memory in Stock Market Trading Volume”, *Journal of Business and Economic Statistics* **18**, 570-576.

LUX, T. (1996), “Long-term Stochastic Dependence in Financial Prices: Evidence from the German Stock Market”, *Applied Economics Letters* **3**, 701-706.

MADHUSOODANAN, T. P (1998), “Long-Term Dependence in the Indian Stock Market”, *Journal of Financial Studies* **5**, 33-53.

MANDELBROT, B.B. (1963), “The Variation of Certain Speculative Prices”, *Journal of Business* **36**, 394-419.

MANDELBROT, B.B. (1971), "When can price be arbitraged efficiently? A limit to the Validity of the Random Walk and Martingale Models", *Review of Economics and Statistics* **53**, 225-236.

MANDELBROT, B.B. (1972), “Statistical methodology for non periodic cycles: From the covariance to R/S analysis”, *Annals of Economic and Social Measurement I*, 259-290.

MANDELBROT, B.B. and J. VAN NESS (1968), "Fractional Brownian Motion, Fractional Noises and Application”, *SIAM Review* **10**, 422-437.

MANDELBROT, B.B. and J. WALLIS (1968), “Noah, Joseph and Operational Hydrology”, *Water Resources Research* **4**, 909-918.

MARINUCCI D. and P.M. ROBINSON (2001), “Narrow-bandAnalysisofNonstationary Processes”, *Annals of Statistics* **29**, 947-86.

MILLS, T.C. (1993), “Is There Long-Term Memory in the UK Stock Returns?”, *Applied Financial Economics* **3**, 303-306.

MOULINES, E. and P. SOULIER (1999), "Broadband Log Periodogram Regression of Time Series with Long Range Dependence", *Annals of Statistics* **27**, 1415-1439.

NIELSON, M.O. (2007), "Local Whittle Analysis of Stationary Fractional Cointegration and the Implied-realized Volatility Relation", *Journal of Business Economics and Statistics* **25**, 427-446.

NIELSEN, M.O. and P. FREDERIKSEN (2008), "Fully Modified Narrow-Band Least Squares Estimation of Stationary Fractional Cointegration", Queen's Economics Department, Working Paper No. 1171.

PAI, J. and S.N. RAVISHANKER (1998), "Bayesian Analysis of Autoregressive Fractionally Integrated Moving Average Processes", *Journal of Time Series Analysis* **19**, 99–112.

PAINTER, S. (1998), "Long-range dependence in the subsurface: Empirical evidence and simulation methods", Invited paper at the American Geophysical Union 1998 Fall Meeting.

PALMA, Wilfredo (2007), *Long Memory Times Series: Theory and Methods*, John Wiley & Sons Inc. Hoboken, New Jersey. .

PHILLIPS, P.C.B. (1999), "Discrete Fourier Transformation of Fractional Processes", *Cowles Foundation Discussion Papers* **1243**, Cowles Foundation, Yale University.

PHILLIPS, P.C.B. (2007), "Unit Root Periodogram", *Journal of Econometrics* **138**, 104-124.

PHILLIPS, P.C.B. and K. SHIMOTSU (2004), "Local Whittle Estimation in Non-stationary and Unit Root Cases", *Annals of Statistics* **32**, 656–692.

PHILLIPS, P.C.B. and K. SHIMOTSU (2006), "Local Whittle Estimation of Fractional Integration and Some of its Variants", *Journal of Econometrics,* **130**, 209-233.

ROBINSON, P.M. (1994a), "Time Series with Strong Dependence", In: C.A. Sims (Ed.), *Advances in Econometrics: Sixth World Congress* **1**, Cambridge, Cambridge University Press, 47-95.

ROBINSON, P.M. (1994b), "Semiparametric Analysis of Long Memory Time Series", *The Annals of Statistics* **22**, 515–539.

ROBINSON, P.M. (1995a), "Log-periodogram Regression of Time Series with Long Range Dependence", *Annals of Statistics*, **23**, 1048-1072.

ROBINSON, P.M. (1995b), "Gaussian Semiparametric Estimation of Long Range Dependence", *Annals of Statistics*, **23**, 1630-1661.

ROBINSON, P.M. (Ed.) (2003), *Time Series with Long Memory*. Advanced Texts in Econometrics. Oxford University Press, Oxford.

ROBINSON, P.M. (2008), "Multiple Local Whittle Estimation in Stationary Systems", *Annals of Statistics* **36**, 2508-2530.

ROBINSON, P.M. and Y. YAJIMA (2002), "Determination of Cointegrating Rank in Fractional Systems", *Journal of Econometrics* **106**, 217—241.

SAMORODNISTSKY, G. (2007), *Long Range Dependence*, Now Publishers Inc.

SEATER, J.J. (1993), "World Temperature-trend Uncertainties and their Implications for Economic Policy", *Journal of Business and Economic Statistics* **1I,** 265-277.

SHIMOTSU, K. (2007), "Gaussian Semiparametric Estimation of Multivariate Fractionally Integrated Processes", *Journal of Econometrics* **137**, 277-310.

SOWELL, F.B. (1990), "The Fractional Unit Root Distribution", *Econometrica* **58**, 495–505.

SOWELL, F.B. (1992), "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models", *Journal of Econometrics* **53**, 165–188.

TAQQU, M.S. (1986), "A Bibliographical Guide to Self-Similar Processes and Long-Range Dependence", In: E. Eberlein and M. Taqqu (Ed.), *Dependence in Probability and Statistics*, Boston, Birkhauser, 137-162.

TOLVI, J. (2003), "Long-Term Memory in a Small Stock Market", *Economics Bulletin* **7**, 1-13.

VELASCO, C. (1999), "Gaussian Semiparametric Estimation of Non-stationary Time Series", *Journal of Time Series Analysis* **20**, 87-127.

VELASCO, C. (2003), "Gaussian Semiparametric Estimation of Fractional Cointegration", *Journal Time Series Analysis* **24**, 345-378.

WHITTLE, P. (1951), *Hypothesis Testing in Time Series Analysis*, Almquist and Wiksells, Uppsala, Sweden.

WRIGHT, J.H., (2000), "Log-periodogram Estimation of Long Memory Volatility Dependencies with Conditionally Heavy Tailed Returns" *FRB International Finance Discussion Paper* **685**.