Long memory and fractional cointegration with deterministic trends

thesis submitted by Fabrizio Iacone,

registered at the London School of Economics,

for the degree of Doctor of Philosophy

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

We discuss the estimation of the order of integration of a fractional process that may be contaminated by a time-varying deterministic component, or subject to a break in the dynamics of the zero-mean stochastic component, and the estimation of the cointegrating parameter in a bivariate system generated by fractionally integrated processes and by additive polynomial trends. In Chapter 1 we review the theoretical literature on fractional integration and cointegration, and we analyse a situation in which a fractional model reconciles two apparently conflicting economic theories. In Chapter 2 we consider local Whittle estimation of the order of integration when the process is contaminated by a deterministic trend or by a break in the mean. We propose a simple condition to assess whether the asymptotic properties of the estimate are unaffected by the time-varying mean, and a test, with asymptotically normal test statistic under the null, to detect if that condition is met. In Chapter 3 we discuss local Whittle estimation when the zero-mean stochastic component is subject to a break: we show that the estimate is robust to instability in the short term dynamics, while in presence of a break in the long term dynamics only the highest order of integration is consistently estimated. We propose a test to detect that break: the limit distribution of the test statistic under the null is not standard, but it is well known in the literature. We also propose a procedure to estimate the location of a break when it is present. In Chapter 4 we consider a cointegrating relation in which a nonstationary, bivariate process is augmented by a deterministic trend. We derive the limit properties of the Ordinary Least Squares and Generalised Least Squares estimates: these depend on the comparison between the deterministic and the stochastic components.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Peter M. Robinson for his constant presence, participation and encouragement during these years. His guidance has been truly invaluable. I will be always be grateful to him and I will remember him as a unique example and source of inspiration.

I would also like to thank many other people. Thanks to Vassilis Hajivassiliou, Javier Hidalgo, Liudas Giraitis, Domenico Marinucci, Marcia Shafgans, who taught me during my time at the LSE, and to Sue Kirkbirde; thanks to my fellow LSE students and to all my friends, and thanks to my flatmates Fernando and Alex, who had to endure me most of the time.

A special thanks to Javier Hualde and Stepana Lazarova, who shared with me the experience of being a PhD student, the passion for this topic, and many endless conversations.

Financial support, through the Dennis Sargan Memorial Fund, the Ente per gli studi monetari, bancari e finanziari "Luigi Einaudi" and ESRC Grant R000239936, is also gratefully acknowledged.

Finally, thanks to my parents, Giovanni and Rita, to whom I dedicate this work.

Declaration

- No part of this doctoral dissertation has been presented to any University for any degree.

- Chapter 4 was undertaken as joint work with professor Peter M. Robinson.

Chapter 1

Overview

1.1 Introduction

Fractional integration is a very popular way to model strong autocorrelation. It is a parsimonious model, because dependence at long lags can be summarised by a single parameter, often referred to as "order of integration" or "memory parameter", and a satisfactory one, because the dependence it prescribes for the data in the long term often matches the one observed in reality: indeed, the low frequency spectral shape that can be associated with the autocorrelation at long lags was acknowledged by Granger (1966) as "typical" for economic variables.

Cointegration is a non-trivial extension of the concept of integration to multivariate processes. In that case the long term dynamics of two or more integrated processes are driven by the same stochastic trends, and there is at least one linear combination of the variables which has a lower order of integration. Each group of weights that combines the variables so that the order of integration is reduced is known as "cointegrating vector", and the combination itself is often regarded as a long run equilibrium.

However in many applications the assumption that the observations are generated by a fractionally integrated process may be too restrictive, and we then propose to extend it by either considering a time-varying mean, such as a deterministic trend or a sudden shift in the mean, or a break in the (zero-mean) fractionally integrated process.

In order to address these issues, we first review the current relevant literature.

In Section 1.2 we introduce the concepts of fractional integration and cointegration and in Section 1.3 we describe several techniques to estimate the memory parameter and the cointegrating vectors.

In Section 1.4 we present an application of some of these techniques: the example we have chosen, a dynamic model for the term structure of interest rates, is motivated by the fact that alternative groups of economic theories prescribe conflicting orders of integration if only integers are considered, but this incompatibility may be resolved by introducing intermediate, "fractional" orders.

In Section 1.5 we introduce the topics that we intend to analyse in the thesis and we discuss how they are related to the current literature.

In this thesis we will use the following notation: the operator $\sim$ indicates that the ratio between left- and right-sides tends to 1 (when applied to matrices, it refers to each element of the matrix), and $\Longrightarrow$ indicates weak convergence of the associated probability measure. The "prime", ($'$), operator denotes transpose of a matrix or of a vector, the "star", (*), the complex conjugate. The lag operator L is such that $Lx_t = x_{t-1}$, and $\Delta = (1 - L)$ is the (first) difference operator. We use the operator $\|.\|$ to refer to the (spectral) norm of a vector or of a matrix, $|.|$ to refer to the absolute value of a number or to the determinant of a matrix, $[.]$ to refer to the integer part of a number,

and $\mathrm{Re}(.)$ to the real part of a number or of a matrix. The function $1(.)$ is the indicator function, while 1_p is the $p \times 1$ vector $(1, ..., 1)'$. For a group of observations $x_1, ..., x_n$, we indicate the sample mean as $\overline{x}$. We introduce C and K, such that each one of them may be a positive, finite constant or a positive definite matrix with finite norm, not necessarily always the same: the difference between the two is that C is used to set upper bounds in identities, equations or limit approximations, and K is introduced when we intend that the identities, equations or limit approximations hold exactly. By O_p we indicate a stochastic order of magnitude: for a stochastic sequence S_n and a deterministic one b_n this is defined by saying that $S_n = O_p(b_n)$ if for any $\varepsilon > 0$ there is C and n_0 such that $P(|S_n/b_n| > C) < \varepsilon$ for any $n > n_0$; if $S_n/b_n \to_p 0$ then we say that $S_n = o_p(b_n)$. By O_e we indicate an exact order of magnitude: for deterministic sequences d_n, b_n this is defined by saying that $d_n = O_e(b_n)$ if $|d_n/b_n| \longrightarrow K$ as $n \longrightarrow \infty$, while for stochastic sequences S_n it is defined by saying that $S_n = O_e(b_n)$ if S_n/b_n converges (in distribution) to a random variable with positive and finite variance or to a non-zero constant. The sets $Z = \{0, \pm 1, ...\}$ and $\mathbb{R}$ are composed of the integer and real numbers respectively. We also introduce the following abbreviations: for a process that is independent and identically distributed with mean μ, variance σ^2, we say that it is $i.i.d.(\mu, \sigma^2)$, if λ tends to zero from above, we say $\lambda \to 0^+$.

1.2 Long memory, fractional integration and cointegration

1.2.1 Long memory and fractional integration

For a generic column vector stochastic process ξ_t, $t \in Z$, such that ξ_t is zero-mean, weakly stationary and invertible, we introduce the autocovariance γ_ξ

$$E(\xi_t \xi'_{t+j}) = \gamma_\xi(j) \tag{1.1}$$

where $j \in Z$, and we assume that ξ_t has spectral density $f_\xi(\lambda)$ such that

$$\gamma_\xi(j) = \int_{-\pi}^{\pi} f_\xi(\lambda) e^{i\lambda j} d\lambda. \tag{1.2}$$

Definition 1.1. Strong autocorrelation, weak autocorrelation and antipersistence. *Let the scalar ξ_t, $t \in Z$, be a zero-mean, weakly stationary and invertible stochastic process with spectral density $f_\xi(\lambda)$. Then*

(i) *ξ_t is strongly autocorrelated (has long memory, has long range dependence) if*

$$f_\xi(0) = \infty; \tag{1.3}$$

(ii) *ξ_t is weakly autocorrelated (has short memory, has short range dependence) if*

$$0 < f_\xi(0) < \infty; \tag{1.4}$$

(iii) *ξ_t is antipersistent if*

$$f_\xi(0) = 0. \tag{1.5}$$

For any non-integer d, we introduce the binomial expansion,

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \psi_j(d) L^j, \text{ where } \psi_j(d) = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \tag{1.6}$$

and $\Gamma(.)$ is the Gamma function. Using the Stirling approximation for the Gamma function, $\psi_j(d)$ in (1.6) can be approximated as

$$\psi_j(d) \sim \frac{j^{d-1}}{\Gamma(d)} \text{ as } j \to \infty. \tag{1.7}$$

Definition 1.2. "Type I" (scalar) fractionally integrated process. *For a scalar process u_t, $t \in Z$, with positive and finite spectral density $f_u(\lambda)$ for any $\lambda \in [0, 2\pi)$, for an integer k and any real number δ so that $-1/2 < \delta - k < 1/2$, let*

$$\eta_t = \Delta^{-(\delta-k)} u_t. \tag{1.8}$$

Then the process

$$\phi_t = \begin{cases} \Delta^{-k}\eta_t\,, & k \leq 0 \\ \Delta^{-k}\{\eta_t 1(t>0)\}\,, & k > 0 \end{cases} \tag{1.9}$$

is a "Type I" integrated of order δ process, and we write as $\phi_t \in I_1(\delta)$.

Definition 1.3. "Type II" (scalar) fractionally integrated process. *For a scalar process u_t, $t \in Z$, with positive and finite spectral density $f_u(\lambda)$ for any $\lambda \in [0, 2\pi)$, for any real number δ, then the process*

$$\varphi_t = \Delta^{-\delta}\{u_t 1(t>0)\}, \tag{1.10}$$

is a "Type II" integrated of order δ process, and we write as $\varphi_t \in I_2(\delta)$.

Both Type I and Type II are fractionally integrated processes, and for both

of them the notation $I(\delta)$ is used in the literature: in the rest of the thesis however we will only use $I(\delta)$ as an abbreviation for $I_2(\delta)$. The parameter δ is also known as the memory parameter or as the order of integration. It is particularly important, because, as we are going to show, under regularity conditions it summarises properties of the long term dynamics and it characterises the rate of convergence and the limit distribution of estimates such as the sample mean or the OLS regression coefficient.

We presented the three definitions at the same time in order to discuss their differences and similarities.

In order to compare them, we first remark that due to the truncation in (1.10) and, when $k > 0$, in (1.9), only Type I fractionally integrated processes with $\delta < 1/2$ may be stationary, and that invertibility requires $\delta > -1/2$. For Type II processes, notice that even $\varphi_t \in I(0)$ is nonstationary: we prefer this notation to keep it consistent with Robinson and Hualde (2003) and related works, but for practical purposes the difference seems to be negligible.

Following Velasco (1999a) we generalise the spectral density for the cases in which it is not defined, and introduce the "pseudo-spectrum", for $\phi_t \in I_1(\delta)$,

$$f_\phi(\lambda) = \left|1 - e^{-i\lambda}\right|^{-2\delta} f_u(\lambda) \tag{1.11}$$

(clearly, when $\delta < 1/2$ this is actually a spectrum). The pseudo-spectrum still maintains several properties that were defined for the proper spectral density, as we also discuss in Section 1.3.

For $\varphi_t \in I(\delta)$ Robinson and Marinucci (2001) defined a "time-varying spectral density"

$$f_\varphi^{(n)}(\lambda) = |\psi_n(\lambda;\delta)|^2 f_u(\lambda) \tag{1.12}$$

where

$$\psi_t(\lambda;\delta) = \sum_{s=0}^{t-1} \psi_s(\delta)\, e^{i\lambda s} \tag{1.13}$$

and $\psi_s(\delta)$ is defined as in (1.6). Since t is finite, $f_\varphi^{(t)}(\lambda)$ has no pole at $\lambda = 0$ even when $\delta > 0$, but for $\phi_t \in I_1(\delta)$, $\varphi_t \in I(\delta)$, $\lambda \in (0, 2\pi)$,

$$f_\varphi^{(n)}(\lambda) \to f_\varphi^{(\infty)}(\lambda) = f_\phi(\lambda) \text{ as } n \to \infty. \tag{1.14}$$

In order to simplify the notation, we will, in the rest of the thesis, drop (∞) and simply write $f_\varphi(\lambda)$ for $f_\varphi^{(\infty)}(\lambda)$ for $\varphi_t \in I(\delta)$. For $\xi_t \in I_1(\delta)$ or $\xi_t \in I(\delta)$, making use of the approximation $\left|1 - e^{-i\lambda}\right| \sim |\lambda|$ as $\lambda \to 0$,

$$f_\xi(\lambda) \sim G_\xi |\lambda|^{-2\delta} \text{ as } \lambda \to 0 \tag{1.15}$$

where $0 < G_\xi < \infty$.

Type I fractionally integrated processes with $\delta \in (0, 1/2)$ are strongly autocorrelated according to Definition 1.1, but Type I processes with $\delta \geq 1/2$ and Type II processes are not stationary and then cannot be classified according to that definition. Indeed, even the popular "unit root" model, which corresponds to $\delta = 1$, is not included in Definition 1.1, although it is generally acknowledged that the autocorrelation is particularly strong in that case. It seems fair then to generalise Definition 1.1 at least to allow for nonstationary processes by also considering the pseudo-spectrum and the limit of the time-varying spectral densities.

On the other hand, Definition 1.1 is very general because, at least if we restrict our attention to stationary processes, the three cases cover all the possible outcomes for $f_\xi(0)$; fractional integration imposes on the spectral density a parametric structure at low frequencies, which may be considered a restriction on long memory because (1.3) and (1.5) may also be generated

by functions that do not meet (1.15); the same remark applies of course if we extend Definition 1.1 to nonstationary fractionally integrated processes.

Unfortunately though, the fact that Definition 1.1 is very general is also the reason why it is of little practical use: additional assumptions are necessary in order to derive the limit dynamics and the asymptotic distributions of transformations of long memory processes.

One reason for this interest in fractional integration is precisely in the fact that, given little further regularity conditions, this model includes enough information to make the derivation of those limit dynamics and asymptotic properties possible. For example, for $\varphi_t \in I(\delta)$, $\delta > 0$, following Marinucci and Robinson (2000) (and regularity conditions therein):

$$\frac{1}{n^{\delta+1/2}\sqrt{2\pi f_u(0)}} \sum_{t=1}^{r} \varphi_t \Longrightarrow B(r;\delta) \text{ as } n \to \infty \tag{1.16}$$

where

$$B(r;\delta) = \int_0^r \frac{(r-s)^{\delta-1}}{\Gamma(\delta)} dB(s) \tag{1.17}$$

and $B(s)$ is a standardised Brownian motion (that is to say, $B(s)$ is such that $E\left(B(s)^2\right) = s$ when $s \geq 0$). The functional $B(r;\delta)$ in (1.17) is a Type II fractional Brownian motion: this is described by Mandelbrot and Van Ness (1968) and by Marinucci and Robinson (1999, 2000). The same authors also discuss an alternative form of Brownian motion, that Marinucci and Robinson (1999) called "Type I fractional Brownian motion": this is obtained by summing Type I fractionally integrated observations instead. We do not present the two fractional Brownian motions in greater detail because we only use the Type II in Chapter 4 and even then only following Robinson and Hualde's (2003) approach, and we refer to Marinucci and Robinson (1999) instead. We point out, however, that the two Brownian motions are different, being associ-

ated with different autocovariance structures, although these autocovariances become equal at least at long lags. It is also worth noticing that the Type I process is only defined for a limited range of values for δ.

The two types of processes are, anyway, very similar in many other respects: Marinucci and Robinson (2001) showed that, when $\delta < 1/2$, $\phi_t \in I_1(\delta)$, $\varphi_t \in I(\delta)$,

$$E(\phi_t \phi_{t+j}) - E(\varphi_t \varphi_{t+j}) = O\left(t^{\delta-1/2}\right) \text{ as } t \to \infty \tag{1.18}$$

uniformly for any $j \geq 0$, while Robinson (2005b) showed that the asymptotic distribution of the Whittle estimate of δ does not change according to whether ϕ_t or φ_t is used (we discuss Whittle estimation in Section 1.3), and on the basis of his work, it is also fair to conjecture that the same irrelevance holds for other estimation techniques as well. This is an important result because it means that although it is not possible to distinguish between the two Types in the empirical analysis, this difference has no asymptotic impact on the estimate.

We refer to Robinson and Marinucci (2001) and to Robinson (2005b) for a more detailed comparison of Type I and II processes.

Fractionally integrated processes are also characterised by two other properties: that

$$\gamma_\xi(j) \sim c_\xi j^{2\delta-1} \text{ as } j \to \infty \tag{1.19}$$

for a non-zero c_ξ, and

$$Var\left(\bar{\xi}\right) = O_e\left(n^{2\delta-1}\right) \tag{1.20}$$

(notice that (1.19) is restricted by (1.1) to stationarity, that is, to Type I processes; (1.18) however provides a clear indication of the applicability of the same concept to Type II processes as well). From (1.19) we then see that the dependence at long range can be satisfactorily summarised by the order of integration only, c_ξ being just a scaling factor, and that the dependence is

stronger the higher δ. The power law approximation for the spectrum conveys the same piece of information in the frequency domain.

Semiparametric techniques to estimate δ have been realised using any of (1.15), (1.19) and (1.20). Notice anyway that (1.19) and (1.20) can be observed also for processes that are not fractionally integrated, and for which (1.16) does not hold (either for the Type I or for the Type II fractional Brownian motion): Diebold and Inoue (2001) for example showed several models that are not fractionally integrated and nevertheless have sample mean of order $n^{\delta-1/2}$ as in (1.20). Their examples were mainly particular types of Markov-switching models, but a sample mean of order $n^{\delta-1/2}$ may also be generated by a neglected deterministic component, as we actually consider in Chapter 2 and in Chapter 4: when

$$x_t = \xi_t + \mu t^{\delta-1/2}, \tag{1.21}$$

with $\xi_t \in I_1(0)$, $\delta > 0$ and for some nonzero μ, then

$$\overline{x} = O_e\left(n^{\delta-1/2}\right). \tag{1.22}$$

In other cases, only some of the properties (1.15), (1.19) and (1.20) are present. Some cyclical models, for example, may generate slowly decaying autocovariances as in (1.19) (but with a further cosine factor that induces a cyclical path in the autocorrelation function in the long run), and yet have spectral density bounded at zero and possibly with a pole at another frequency: we refer to Baillie (1999) or Robinson (2003) for a discussion of one of these examples, the Gegenbauer process, and to Yong (1974) for a discussion of the conditions under which (1.19) and (1.15) are equivalent.

We also mention that our definitions of fractional integration may be generalised replacing $\Delta^{-(\delta-k)}$ in (1.8) and $\Delta^{-\delta}$ in (1.10) by any weighting struc-

ture $\sum_{j=0}^{\infty}\psi_j(\delta-k)L^j$ or $\sum_{j=0}^{\infty}\psi_j(\delta)L^j$ provided that (1.7) and that $|\psi_j(\delta) - \psi_{j+1}(\delta)| = O\left(|\psi_j(\delta)|\, j^{-1}\right)$ are met, as indeed Robinson and Marinucci (2001) did for the Type II fractionally integrated process. This may be a sensible generalisation, because the relevant properties, in terms of low frequencies approximation of the spectral density, high lags approximation of the autocorrelation function, order of magnitude of the sample variance and weak convergence of partial sums of the observables to the appropriate (Type I or Type II) fractional Brownian motion, are still met. Definitions 1.2 and 1.3 are actually so restrictive that they do not even include the fractional noise, the process having autocovariance structure

$$\gamma_{fgn}(j) = \frac{1}{2}\left(|j+1|^{2\delta+1} - 2|j|^{2\delta+1} + |j-1|^{2\delta+1}\right) \tag{1.23}$$

for $\delta \in (-1/2, 1/2)$, introduced by Mandelbrot and Van Ness (1968).

We then prefer Definitions 1.2 and 1.3 because they are more intuitive and because they are more familiar to many readers, especially when δ is restricted to integers. However, we acknowledge that the results we are presenting apply to a broader class of processes.

We conclude the subsection with a discussion of a parametric model that generates a fractionally integrated process: the ARFIMA(p, δ, q) model.

ARFIMA is an acronym for AutoRegressive Fractionally Integrated Moving Average; p and q indicate the number of lags in the AR and MA components respectively, and δ the order of fractional integration. This model was introduced by Adenstedt (1974), who set $k = 0$ in a Type I integrated process and considered

$$a_t = \sum_{j=0}^{\infty}\psi_j(\delta)\, u_{t-j}, \tag{1.24}$$

further assuming that the innovations u_t were $i.i.d.(0, \sigma^2)$. This model was

later referred to as ARFIMA$(0, \delta, 0)$ by Granger and Joyeux (1980) and Hosking (1981).

In the ARFIMA$(0, \delta, 0)$, the weights (1.7) describe the responses to past impulses, which then vanish only at hyperbolic rate when $\delta > 0$, much more slowly than the exponential decay of the innovation of any stationary and invertible ARMA(p, q). Notice, also, that the weights with which the past innovations still affect the current observation increase the larger δ. The autocovariance function is $\gamma_a(j) = \sigma^2 \frac{(-1)^j \Gamma(1-2\delta)}{\Gamma(j-\delta+1)\Gamma(1-j-\delta)}$, and the spectral density $f_a(\lambda) = \frac{\sigma^2}{2\pi} \left|1 - e^{-i\lambda}\right|^{-2\delta}$.

The ARFIMA$(0, \delta, 0)$ model provides some flexibility in the description of the long term properties of a process, but the short term dynamics are constrained because the structure of the covariances is already set. To satisfactorily model the short term dynamics as well, Granger and Joyeux (1980) and Hosking (1981) suggested treating the ARFIMA$(0, \delta, 0)$ as the building block of a more general structure, and introduced the ARFIMA(p, δ, q). This is obtained by passing an ARFIMA$(0, \delta, 0)$ process through a (stationary and invertible) ARMA(p, q) filter. Hosking (1981) discussed several ARFIMA(p, δ, q) structures, showing that the long range dependence is still dictated by (1.19), and that (1.15) and (1.20) still hold too (thus justifying why δ is often the main parameter of interest), but more flexible short run dynamics are possible.

Of course, Type I ARFIMA(p, δ, q) for $\delta \geq 1/2$ or Type II ARFIMA(p, δ, q) may be easily defined following the lines set out in Definitions 1.2 and 1.3.

1.2.2 Fractional cointegration

When all the elements of a vector are integrated processes, we say that the whole vector is a multivariate integrated process.

For a $p \times 1$ vector $z_t = \{z_{1,t}, ..., z_{p,t}\}'$ such that $z_{1,t} \in I(\delta_1)$, ... , $z_{p,t} \in$

$I(\delta_p)$, we write $z_t \in I(\delta_1, ..., \delta_p)$. In most of the literature, especially when $\delta_1 = ... = \delta_p$, as it is typically when the order of integration is either 0 or 1, the notation $z_t \in I(\delta)$, for a scalar δ, is used instead. We on the other hand set $\delta_+ = \{\delta_1, ..., \delta_p\}'$ and then use $z_t \in I(\delta_+)$.

In the same way, we define a Type I multivariate integrated process, and we introduce the notation $z_t \in I_1(\delta_1, ..., \delta_p)$ and the abbreviation $z_t \in I_1(\delta_+)$.

When z_t is stationary, let $f_z(\lambda)$ be the spectral density matrix, as from (1.1) and (1.2): the approximation (1.15) can be generalised to

$$f_z(\lambda) \sim \Lambda(\lambda) G_z \Lambda(\lambda)^* \text{ as } \lambda \to 0^+ \tag{1.25}$$

where

$$\Lambda(\lambda) = diag\{e^{i\pi\delta_1/2}\lambda^{-\delta_1}, ..., e^{i\pi\delta_p/2}\lambda^{-\delta_p}\} \tag{1.26}$$

and G_z is positive semidefinite.

In order to define cointegration, we follow Robinson and Yajima (2002) and assume that the elements of z_t are all ordered so that

$$\delta_1 = ... = \delta_{k_1} > \delta_{k_1+1} = ... = \delta_{k_2} > ... > \delta_{k_{s-1}+1} = ...\delta_{k_s}. \tag{1.27}$$

We then partition z_t following (1.27)

$$z_t = \left(z_t^{(1)\prime}, ..., z_t^{(s)\prime}\right)' \tag{1.28}$$

so $z_t^{(l)} = \left(z_{k_{l-1}+1,t}, ...z_{k_l,t}\right)'$ for $1 \leq l \leq s$, and introduce a p-dimensional vector α, which is also partitioned in the same way:

$$\alpha = \left(\alpha(1)', \alpha(2)', ..., \alpha(s)'\right)'. \tag{1.29}$$

Definition 1.4. Cointegration and cointegrating rank (Robinson and Yajima (2002)). *If there exists a non-null vector* $\alpha(l)$*, such that* $\alpha(l)' z_t^{(l)} \in I(\gamma_h)$ *with* $\gamma_h < \delta_{k_l}$*, then we say* z_t *is cointegrated with cointegrating vector* $\alpha = (0, ..., 0, \alpha(l)', 0, .., 0)'$*. The number of such linearly independent* $\alpha(l)$ *is* r_l*, and the cointegrating rank of* z_t *is* $r = \sum_{l=1}^{s} r_l$*.*

The processes $\alpha' z_t$ are referred to as cointegrating errors.

Notice that different combinations of different subsets of $z_t^{(l)}$ may generate cointegrating errors with different orders of integration.

We arrange the orders of integration of the cointegrating errors in the vector $\gamma_+ = \{\gamma_1, ..., \gamma_r\}'$.

Robinson and Yajima (2002) proposed the Definition 1.4 to generalise the original definition of Engle and Granger (1987) by allowing for alternative levels of integration (although not for linear combinations involving variables and cointegrating errors).

Intuitively, cointegration means that there is at least one non trivial linear combination of the elements of a multivariate vector having order of integration lower than the order of the components of the given vector: this definition applies this principle to groups of variables in z_t that have the same order of integration. Robinson and Yajima (2002) also compared this definition to several others already present in the literature, and showed with some examples that it is closer to the intuitive concept of cointegration.

The time domain description has a correspondence in the frequency domain and, as we have already seen for integration, cointegration too is a phenomenon that may be better observed at low frequencies: the matrix G_z in (1.25) in fact is positive semidefinite with rank $p - r$ (which also implies that G_z has full rank when the process is not cointegrated).

For a given vector $z_t^{(l)}$ and for a given $\alpha(l)$ such that $\alpha(l)' z_t^{(l)} \in I(\gamma_h)$

with $\gamma_h < \delta_{k_l}$ as in Definition 1.4, we also introduce $\beta_h = \delta_{k_l} - \gamma_h$ and $\beta_+ = \{\beta_1, ..., \beta_r\}'$. For a p-dimensional $z_t \in I(\delta_+)$ with cointegrating rank r and cointegrating errors of order γ then we introduce the notation $z_t \in CI(\delta_1, ..., \delta_p, \beta_1, ..., \beta_r)$ that we also abbreviate to $z_t \in CI(\delta_+, \beta_+)$. When in particular only unit root observables and short memory cointegrating errors are considered, we should then write $z_t \in CI(1_p, 1_r)$: since anyway this case is known in the literature as $CI(1,1)$, we prefer to drop the references to the dimensions p and r and use the notation that is more familiar to the reader.

Letting $v_t = \{v_{1,t}, ..., v_{r,t}\}'$, $x_t = \{x_{1,t}, ..., x_{p-r,t}\}'$, Definition 1.4 means that there is a non-trivial $r \times (p-r)$ matrix ν such that

$$y_t = \nu x_t + v_t. \tag{1.30}$$

The elements in the vector $(y_t', x_t')'$ are the same as those in z_t, but they may be ordered in a different way, because the elements in z_t are arranged according to (1.27) and (1.28).

For each row h, $1 \leq h \leq r$

$$y_{h,t} = \nu_h' x_t + v_{h,t}, \tag{1.31}$$

where ν_h' is the h^{th} row of ν; we can introduce $\delta_{\{h\}}$ as the order of integration of $y_{h,t}$, so $y_{h,t} \in I(\delta_{\{h\}})$, and γ_h as the order of integration of $v_{h,t}$, so $v_{h,t} \in I(\gamma_h)$: Definition 1.4 then also means that $\delta_{\{h\}} > \gamma_h \geq 0$. It is also worth noticing that for each k, $1 \leq k \leq p-r$, then $\nu_{hk} \neq 0$ implies that $x_{k,t} \in I(\delta_{\{h\}})$ as well: this means that the elements of x_t that actually appear with a nonzero coefficient in the h^{th} equation share the same order of integration.

In many theoretical and applied papers, each equation (1.31) is considered to describe an equilibrium relation between $y_{h,t}$ and x_t: deviations from it,

represented by the cointegrating error $v_{h,t}$, are temporary and, upon taking a long enough time span, the variables $y_{h,t}$ and x_t move together. This actually imposes the additional requirement that the cointegrating error $v_{h,t}$ is mean-reverting, which corresponds to $\gamma_h < 1$: although we agree that this is usually the most interesting case due to the importance that the concept of equilibrium often has in economic theory, we notice that this is not necessarily imposed from the definition of cointegration.

Fractional integration adds a lot to the $CI(1,1)$ design because $\gamma_h > 0$ allows a much slower return to the equilibrium; it also makes mean reversion (or trend reversion) possible for y_t and x_t as well, a property that in many practical applications may be required by the economic theory. Finally, it also provides the researcher with another measure, β_h, that indicates how strong the cointegration itself is.

1.3 Estimation and testing

1.3.1 Estimation of the memory parameter

The memory parameter may be estimated either individually, typically using the limited information in (1.15), (1.19) or (1.20), or jointly with other parameters, when a complete model, such as an ARFIMA(p, δ, q) for example, is assumed. The first approach is called "semiparametric", because it does not require the specification of a whole model but only of some of its properties, while the other one is "parametric".

There is a large number of estimates in the literature: however, we only discuss those that we are going to use in our applications, or that are of relevance to the models we discuss in Chapters 2 to 4.

We begin by discussing the case in which ξ_t is a scalar.

An early estimation procedure was based on the "rescaled range" R/S statistic proposed by Hurst (1951). Given the observations $\xi_1, ..., \xi_n$, the R/S statistic is

$$RS = \frac{\max_{1\leq k\leq n} \sum_{t=1}^{k} \left(\xi_t - \overline{\xi}\right) - \min_{1\leq k\leq n} \sum_{t=1}^{k} \left(\xi_t - \overline{\xi}\right)}{\left\{\frac{1}{n} \sum_{t=1}^{n} \left(\xi_t - \overline{\xi}\right)^2\right\}^{1/2}}. \tag{1.32}$$

The ratio $\ln(RS) / \ln n$ converges to $1/2$ when ξ_t is a short memory process, and to $\delta + 1/2$ when it is fractionally integrated of order δ. Regularity conditions are very mild: consistent estimation of δ is obtained even when the second moment is not finite, which may occur for example for distributions with very heavy tails, as sometimes is the case for a financial time series.

The R/S statistic can also be used to test for the presence of fractional integration. Under the null of short memory and regularity conditions (these, however, included the existence of finite second moments), by setting

$$c_L^2 = \frac{f_\xi(0)}{\gamma_\xi(0) / (2\pi)}, \tag{1.33}$$

then

$$\frac{1}{\sqrt{n} c_L} RS \Longrightarrow V, \tag{1.34}$$

where V is the range of a Brownian bridge. We refer to Lo (1991) for further details on V, including critical values, and for a discussion about nonparametric estimation of c_L.

When the second moments of the process are defined, consistent estimation of δ may also be based on the low frequency approximation of the spectral density by the power law (1.15). The estimates we describe are appealing because they are intuitive and, given regularity conditions, they are also characterised

by limit normal distribution, a distinct advantage over the nonstandard asymptotics in (1.34).

The building block for these estimates is the periodogram. We define it by introducing, for the observations $\xi_1, ..., \xi_n$, the discrete Fourier transform

$$F_\xi(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \xi_t e^{-i\lambda t}. \tag{1.35}$$

The periodogram is then

$$I_\xi(\lambda) = F_\xi(\lambda) F_\xi(-\lambda). \tag{1.36}$$

Although (1.35) and (1.36) can be computed for any λ, the frequencies

$$\lambda_j = \frac{2\pi j}{n}, \text{ for } j \in Z_n \tag{1.37}$$

where

$$Z_n = \{Z \cap [0, n-1]\} \tag{1.38}$$

are particularly important and are referred to as Fourier frequencies: in the rest of the thesis then when we use the notation λ_j we also assume that $j \in Z_n$. Notice that with this definition j cannot take the value n, nor any multiple of it.

At the Fourier frequencies the periodogram provides a decomposition of the sum of squares of ξ_t

$$\sum_{t=1}^{n} \xi_t^2 = 2\pi \sum_{j=0}^{n-1} I_\xi(\lambda_j) \tag{1.39}$$

and

$$I_\xi(0) = \frac{n}{2\pi} \left(\bar{\xi}\right)^2. \tag{1.40}$$

Since

$$\sum_{t=1}^{n} e^{i\lambda_j t} = 0 \text{ when } j \neq 0 \tag{1.41}$$

it can also be noticed that the presence of a constant, non-zero mean does not affect the periodogram for $\lambda_j \neq 0$.

At Fourier frequencies different from $\lambda_j = 0$, the definition (1.36) is equivalent to

$$I_\xi(\lambda_j) = \frac{1}{2\pi} \sum_{s=-n+1}^{n-1} \widehat{\gamma}_\xi(s) \cos(\lambda_j s), \tag{1.42}$$

where

$$\widehat{\gamma}_\xi(s) = \frac{1}{n} \sum_{t=1}^{n-s} (\xi_t - \overline{\xi})(\xi_{t+s} - \overline{\xi}) \tag{1.43}$$

(the mean-correction is irrelevant for the definition of the periodogram, but we retained it because it is included in the definition of the sample autocovariance). The periodogram is then an estimate of the spectral density.

If ξ_t is a zero-mean, weakly autocorrelated process with $0 < f_\xi(\lambda) < \infty$ at any λ (and given other mild regularity conditions, including continuity of $f_\xi(\lambda)$), the periodogram is asymptotically unbiased,

$$E(I_\xi(\lambda)) = f_\xi(\lambda) + o(1), \tag{1.44}$$

and for any two Fourier frequencies λ_j and λ_k, with $j \neq k$, and $j \neq 0$, $k \neq 0$,

$$Cov(I_\xi(\lambda_j), I_\xi(\lambda_k)) = o(1), \tag{1.45}$$

and

$$Var(I_\xi(\lambda_j)) = f_\xi^2(\lambda_j) + o(1). \tag{1.46}$$

The periodograms at different Fourier frequencies are then asymptotically uncorrelated, and the potential weak temporal dependence of ξ_t is transformed

into heteroscedasticity of $I_\xi(\lambda_j)$. We refer to Brockwell and Davies (1987) for more details on the properties of the periodogram of a weakly autocorrelated process.

Asymptotic unbiasedness and absence of correlation between different Fourier frequencies cannot be extended to long memory time series: for given λ_j, Künsch (1986) noticed this for a process ξ_t having spectral density that can be approximated around 0 by the power law (1.15) and $\delta \in (0, 1/2)$, and Robinson (1995a) extended the result to $\delta \in (-1/2, 0)$.

But Robinson (1995a) also showed that the bias of the periodogram and the correlation of the discrete Fourier transforms at different frequencies can be bounded, and that the bound decreases with the distance from the origin, so for some sequences $j(n)$ the asymptotic unbiasedness and uncorrelation still hold: when $\delta \in (-1/2, 1/2)$, for any positive integer j such that $j/n \to 0$ as $n \to \infty$, then

$$E\left(f_\xi^{-1}(\lambda_j) I_\xi(\lambda_j)\right) = 1 + O\left(\frac{\ln j}{j}\right), \tag{1.47}$$

and for any positive integer $k < j$ (and j defined as before)

$$E\left(\lambda_j^\delta F_\xi(\lambda_j) \lambda_k^\delta F_\xi(-\lambda_k)\right) = O\left(\frac{\ln j}{k}\right). \tag{1.48}$$

The assumptions of Robinson (1995a) were very general, requiring only stationarity and a certain degree of smoothness of $f_\xi(\lambda)$ as $\lambda \to 0$.

If ξ_t is observable, the low frequency approximation (1.15) may be rearranged as

$$\ln\left(I_\xi(\lambda_j)\right) \sim c - 2\delta \ln \lambda_j + u_j \text{ as } \lambda_j \to 0^+, \tag{1.49}$$

where $u_j = \ln\left(I_\xi(\lambda_j) / f_\xi(\lambda_j)\right)$: due to the logarithmic transformation, (1.49) is usually known as a "log-periodogram regression" model.

The condition $\lambda_j \to 0^+$ is met by running the regression only for the Fourier

frequencies $0 < j < m$, where m is such that $m/n \to 0$ as $n \to \infty$. When the process ξ_t is stationary and invertible, the OLS regression estimate of δ in (1.49), $\widehat{\delta}_{LP}$, is consistent if $m \to \infty$; under the additional condition that $m = o\left(n^{4/5}\right)$ (when ξ_t is an ARFIMA, this rate depends on the smoothness of $f_\xi(\lambda)$ as $\lambda \to 0$ and may be smaller for other processes), the estimate is also asymptotically normal, with limit distribution

$$\sqrt{m}\left(\widehat{\delta}_{LP} - \delta\right) \to_d N\left(0, \frac{\pi^2}{24}\right) \text{ as } n \to \infty. \tag{1.50}$$

The idea to treat (1.49) as a regression model can be traced to a comment by Granger and Joyeux (1980), but the estimate was first addressed by Geweke and Porter-Hudak (1983). However, a rigorous proof was only supplied by Robinson (1995a). He discussed the regression over the Fourier frequencies associated to $j = l, ..., m$ with $1/l \to 0$ as $n \to \infty$, and under the additional assumption of Gaussianity, but later Hurvich, Deo and Brodsky (1998) obtained (1.50) under alternative conditions that would also allow for $l = 1$, while Velasco (2000) proved (1.50) for non-Gaussian ξ_t as well.

Another estimate of δ was discussed by Robinson (1995b) following a remark by Künsch (1987): he suggested using the Whittle approximation of the Gaussian likelihood in the frequency domain, but to estimate δ on a band that degenerates to zero asymptotically. This means computing

$$\left\{\widehat{\delta}_{LW}, \widehat{G}\right\} = \arg \min_{d \in \Theta, G \in S_G} \frac{1}{m} \sum_{j=1}^{m} \left(G^{-1} \lambda_j^{2d} I_\xi(\lambda_j) + \ln G \lambda_j^{-2d}\right) \tag{1.51}$$

where Θ and S_G are compact sets such that $\Theta \subset (-1/2, 1/2)$, $S_G \subset (0, \infty)$. After concentrating

$$\widehat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d} I_\xi(\lambda_j), \tag{1.52}$$

the local Whittle estimate of δ, $\widehat{\delta}_{LW}$, is

$$\widehat{\delta}_{LW} = \arg\min_{d\in\Theta} \ln \widehat{G}(d) - 2\frac{d}{m}\sum_{j=1}^{m} \ln \lambda_j. \tag{1.53}$$

Robinson (1995b) established consistency as $m \to \infty$, $m/n \to 0$, and limit normality when $m = o\left(n^{4/5}\right)$ (as for $\widehat{\delta}_{LP}$, slower rates may be necessary according to the smoothness of $f_\xi(\lambda)$ as $\lambda \to 0$). The asymptotic distribution is

$$\sqrt{m}\left(\widehat{\delta}_{LW} - \delta\right) \to_d N\left(0, \frac{1}{4}\right) \text{ as } n \to \infty. \tag{1.54}$$

Both $\widehat{\delta}_{LW}$ and $\widehat{\delta}_{LP}$ are subject to a lower order bias which increases in m, so the choice of the bandwidth is very important because, on the other hand, the larger m is the smaller the dispersion is. Optimal (in Mean Squared Error sense) bandwidths are discussed by Henry and Robinson (1996) for $\widehat{\delta}_{LW}$ and by Hurvich, Deo and Brodsky (1998) for $\widehat{\delta}_{LP}$: in both cases the choice depends on the precision of the approximation (1.15) and on the smoothness and on the steepness of $\lambda^{2\delta} f_\xi(\lambda)$ as $\lambda \to 0^+$. The most favourable situation is when $f_\xi(\lambda)$ is sufficiently smooth, as it actually occurs for many parametric models, including the stationary and invertible ARFIMA(p,δ,q).

Nonstationarity ($\delta \geq 1/2$) can be addressed by differencing the data, but this requires a certain preliminary knowledge at least of the range in which δ lies. Velasco (1999a, 1999b) showed that (1.50) and (1.54) also hold for $1/2 \leq \delta < 3/4$, and consistency even for $\delta < 1$. Key to this result is the computation of the bound for the expected periodogram when $1/2 \leq \delta < 1$: for positive integer j such that $j/n \to 0$ as $n \to \infty$,

$$E\left(f_\xi^{-1}(\lambda_j) I_\xi(\lambda_j)\right) = 1 + O\left(j^{2(\delta-1)} \ln(j+1)\right) \tag{1.55}$$

and, for positive integer $k < j$,

$$E\left(\lambda_j^{\delta} F_{\xi}\left(\lambda_j\right) \lambda_k^{\delta} F_{\xi}\left(-\lambda_k\right)\right) = O\left((jk)^{\delta-1} \ln\left(k+1\right)\right). \tag{1.56}$$

Velasco (1999a) also showed that when the data is weighted by a suitable filter ("tapering"), the bias of the expectation of the tapered periodogram can still be bounded, and in a way such that the bound may be made negligible for some sequences $j\left(n\right)$ such that $j/n \to 0$ as $n \to \infty$, even for larger δ or for $\delta \leq -1/2$. We refer to Velasco (1999a) for a discussion of the properties required for the taper; we only mention that the tapers may be classified according to the maximum δ for which they can successfully eliminate, at least for some sequences $j\left(n\right)$, the bias, and that the requirements get stronger the larger δ (or, the smaller δ when $\delta \leq -1/2$).

This however is acquired at the cost of higher correlation across neighbouring frequencies (and the higher the order of the taper, the higher the correlation induced), so Velasco (1999a, 1999b) modified the definitions of log-periodogram regression and of local Whittle estimates, and imposed more distance between the points used in the estimation by skipping frequencies. Since less points are used in the optimisation for any given m, the variances are comparatively bigger than in (1.50) and (1.54): tapering then is only advisable if no preliminary information on the range in which δ lies is available.

Semiparametric estimates have the advantage of not requiring any specification of the spectral density for the remaining frequencies. Yet if the whole parametric model is known, even if only up to a known function of a vector of unknown parameters, δ may be estimated more efficiently by using all the Fourier frequencies, rather than just a degenerating narrow band.

Suppose that the spectral density of ξ_t is a known function of the parameters σ^2, θ, δ, and indicate this as $f_{\xi}\left(\lambda; \sigma^2, \theta, \delta\right)$, and that there is $g_{\xi}\left(\lambda; h, d\right)$

such that $f_\xi(\lambda; s^2, h, d) = s^2/(2\pi)\, g_\xi(\lambda; h, d)$ for all the admissible values of $(s^2, h', d)'$. Then, if s^2 varies freely from h, d and $\int_{-\pi}^{\pi} \ln g_\xi(\lambda; h, d)\, d\lambda = 0$, the parametrisation is refereed to as "standard", for example, by Robinson (2003). Then the Whittle estimation of $(\theta', \delta)'$ can be obtained by minimising

$$\frac{1}{n}\sum_{j=1}^{n-1} g_\xi(\lambda_j; h, d)^{-1} I_\xi(\lambda_j) \tag{1.57}$$

with respect to $(h', d)'$ over a compact set.

Consistency of the Whittle estimates for stationary and invertible ξ_t follows already from the argument of Hannan (1973), but his proof could not be directly extended to establish the limit distribution; this was treated by Fox and Taqqu (1986) for Gaussian processes, and by Giraitis and Surgalis (1990) for possibly non-normal ones too. Both Fox and Taqqu (1986) and Giraitis and Surgalis (1990) had a slightly different loss function, the summation being replaced by an integral (Hannan considered this case too). Velasco and Robinson (2000) on the other hand used (1.57): they extended the results to nonstationary processes, proving consistency for $\delta < 1$ and root-n limit normality for $\delta < 3/4$. Velasco and Robinson (2000) also replaced the raw periodogram in (1.57) with a tapered one, but as in Velasco (1999b) the loss function had to be modified by discarding neighbouring frequencies: with these modifications to (1.57) they established consistency and root-n limit normality for even higher δ. In all these cases, additional regularity conditions were required: we do not discuss this in details but mention that they include a certain smoothness of the spectrum (or of the pseudo-spectrum).

Knowledge of the parametric model (possibly up to a vector of unknown parameters) can also be exploited in the time domain, to design a procedure that delivers a consistent and root-n asymptotically normal estimate by minimising a conditional sum of squares.

Variations on this procedure are fairly common, and they can all be treated as a special case of "Whittle" in the sense that the limit distribution of the estimates is the same.

Introduce the notation

$$\xi_t(d) = \Delta^d \xi_t^{\#},\ \xi_t^{\#} = \xi_t 1(t > 0), \tag{1.58}$$

and

$$\widehat{\varepsilon}_t\left(\widehat{\theta}(d)\right) = \xi_t(d) - \sum_{j=1}^{k} B_j\left(\widehat{\theta}(d)\right)\xi_{t-j}(d) \tag{1.59}$$

where $\widehat{\theta}(d)$ is an estimate of the parameters characterising the autoregressive structure B_j for a given d (for example, $B_j\left(\widehat{\theta}(d)\right)$ could be estimated with a finite order autoregression of $\xi_t(d)$ on some lagged values). The parameters $(\theta', \delta)'$ can then be estimated by minimising

$$\frac{1}{n}\sum_{t=1}^{n}\left(\widehat{\varepsilon}_t\left(\widehat{\theta}(d)\right)\right)^2 \tag{1.60}$$

with respect to some values of d defined in a certain compact set. A time domain procedure based on a (slightly different) conditional sum of squares approach was advocated by Beran (1995), although the proof was not complete.

Sowell (1992) claimed that exact maximum likelihood should be preferred, arguing that it would be more precise in finite samples.

In all these cases (exact maximum likelihood, conditional sum of square in the time domain and Whittle approximation without tapering in the frequency domain), the estimates are root-n consistent and have the same limit distribution.

It may be worth noticing that the proofs of Robinson (1995a, 1995b), Velasco (1999a, 1999b) and Velasco and Robinson (2000) were formulated for Type I processes only, and they do not immediately accommodate Type II

processes. Robinson (2005b) addressed the issue by discussing the difference between the Fourier transforms at Fourier frequencies of two processes of the different types, and showed that the difference can be bounded, although he also found that it gets larger with δ; for $\delta > 1/2$ he compared the tapered Fourier transforms instead. He then showed that the Whittle estimate is robust to the type of process used in the estimation, and that the limit distribution does not change.

When a vector process is analysed, simultaneous estimation may be preferred, because the correlation between the different elements composing the vector can be taken into account and the efficiency of the estimates is improved with respect to the case in which the parameters are estimated separately.

For the p-dimensional $z_t = (z_{1,t}, ..., z_{p,t})'$, introduce the $p \times p$ periodogram $I_z(\lambda) = F_z(\lambda) F_z(-\lambda)'$, with $F_z(\lambda) = \left(F_{z_1}(\lambda), ..., F_{z_p}(\lambda)\right)'$.

A multivariate generalisation of the local Whittle loss function in (1.53) can be presented following Lobato (1999). Letting $d_+ = (d_1, ..., d_p)'$, introduce

$$\Lambda(\lambda; d_+) = diag(\lambda^{-d_1}, ..., \lambda^{-d_p}), \tag{1.61}$$

$$\widehat{G}_z(d_+) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}(\Lambda(\lambda_j; d_+)^{-1} I_z(\lambda_j)(\Lambda(\lambda_j; d_+)^{-1})), \tag{1.62}$$

$$L(d_+) = \left\{ \ln\left|\widehat{G}_z(d_+)\right| - \frac{2}{m}(d_1 + ... + d_p)\sum_{j=1}^{m} \ln(\lambda_j) \right\}, \tag{1.63}$$

the local Whittle estimates are

$$\widehat{\delta}_{+,LW} = \arg\min_{d_+ \in \Theta} L(d_+), \text{ where } 1/m + m/n \to 0 \text{ as } n \to \infty \tag{1.64}$$

where $\widehat{\delta}_+ = \left\{\widehat{\delta}_1, ..., \widehat{\delta}_p\right\}'$ and Θ is a compact subset of $(-1/2, 1/2) \times ... \times$

$(-1/2, 1/2)$ (Lobato (1999) actually went further, because he showed that the minimisation of (1.63) can be replaced by a two-step procedure that generates the same asymptotic properties).

If each element of z_t is fractionally integrated with spectral density as in (1.26), (1.25), and the matrix G_z in (1.25) is positive definite, then the local Whittle estimates are root$-m$ consistent (the usual $m = o\left(n^{4/5}\right)$ or less applies, for the limit normality of the estimates, according to the smoothness of the spectrum) and more efficient than in the univariate case: in the bivariate case for example the increase of efficiency is $c_V^2/8$, where $c_V^2 = G_{12}^2/\left(G_{11}G_{22}\right)$ and G_{ab} is the element in the position (a, b) in the matrix G_z.

Multivariate parametric estimation can also be considered, again resulting in more efficient estimates provided that G_z is not singular. Since the generalisations of the procedures for univariate series are rather straightforward, we omit them.

1.3.2 Estimation of the cointegrating vectors

In this subsection and in the following one we discuss the estimation of the cointegrating vectors ν and of the cointegration rank r in the cointegrated model (1.30) for a p-dimensional process z_t.

We assume the model

$$y_t = \nu x_t + u_{y,t}(-\gamma_+) \tag{1.65}$$

$$x_t = u_{x,t}(-\delta_+), \tag{1.66}$$

where $u_{y,t}(-\gamma_+) = (u_{1,t}(-\gamma_1), ..., u_{r,t}(-\gamma_r))'$ is a $r \times 1$ process, $u_{x,t}(-\delta_+) = (u_{r+1,t}(-\delta_1), ..., u_{p,t}(-\delta_{p-r}))'$ is $(p-r) \times 1$, and $u_t = (u_{1,t}, ..., u_{p,t})'$ is a $p \times 1$, $I_1(0)$ process with spectral density $f_u(\lambda)$.

The scalar processes $u_{1,t}(-\gamma_1)$, ..., $u_{p,t}(-\delta_{p-r})$ may of course be fractionally integrated, and, since they are generated by using the notation (1.10), they are of Type II. This has the advantage of not limiting the order of integration only to the range for which the Type I fractional Brownian motion is defined.

We further assume that the process u_t is a linear transformation $u_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$ of an i.i.d. vector ε_t with covariance matrix $E(\varepsilon_t \varepsilon_t') = \Omega$. Regularity conditions for the weights A_j include the normalisation $A_0 = I_p$ and $\sum_{j=0}^{\infty} j \|A_j\| < \infty$, $\sum_{j=0}^{\infty} j \|A_j\|^2 < \infty$: this is a fairly general specification, and other details on the design are in Chapter 4.

This structure is sufficient to derive the limit behaviour of some semi-parametric estimates like OLS (and narrow band least squares, introduced later in this subsection); when we specify a fully parametric model we also assume that $E(\varepsilon_t \varepsilon_t') = \Omega(\theta)$ and that u_t admits an autoregressive representation $B(L;\theta) u_t = \varepsilon_t$, and both $\Omega(\theta)$ and $B(L;\theta)$ are known up to a set of parameters θ.

We introduced a p-dimensional system because in that case the determination of the cointegration rank is not trivial, and we can then present the techniques we use to address that problem in the application that we discuss in the next section, where we have $p = 4$. However, in the remaining part of this subsection and in Chapter 4 we only intend to describe and discuss some techniques for the estimation of ν, and in these parts we focus on $p = 2$, $r = 1$, thereby avoiding the discussion of identifiability of some parameters in presence of differing orders of integration in the explanatory variables, as Robinson and Hualde (2003) also noticed; moreover, in (1.30), $y_t \in I(\delta)$ and $x_t \in I(\delta)$, $v_t \in I(\gamma)$ and $\beta = \delta - \gamma$, and we can then describe the theoretical literature and, in Chapter 4, our results, with a simpler notation.

Engle and Granger (1987) estimated ν via OLS (we refer to this estimate

as $\overline{\nu}_{OLS}$ in the rest of the Chapter, in order to keep the notation consistent with $\overline{\nu}$ in Chapter 4). Engle and Granger (1987) assumed $\delta = 1$ and $\gamma = 0$: in that case the OLS estimate is consistent and converges to the true value with the rate $\overline{\nu}_{OLS} = \nu + O_e\left(n^{-1}\right)$, which is faster than in a regression model with $I(0)$ regressors. That result holds for a rather general specification of the cointegrating errors, including forms of (short memory) autocorrelation or even heteroscedasticity; potential correlation between $u_{1,t}$ and $u_{2,t}$, which would make OLS inconsistent if $\delta = 0$, only generates a lower order bias when $\delta = 1$.

When δ and γ are no longer restricted to integers, though, in general rates depend on both γ and δ: Robinson (1994b) showed that when $\delta < 1/2$ OLS is inconsistent if the regressor x_t is correlated with the cointegrating error v_t, and Robinson and Marinucci (2001) discussed the case $\delta > 1/2$, finding that $\overline{\nu}_{OLS} = \nu + O_p\left(n^{1-2\delta}\right)$ when $\delta + \gamma < 1$, $\overline{\nu}_{OLS} = \nu + O_p\left(n^{1-2\delta} \ln n\right)$ when $\delta + \gamma = 1$ and $\delta < 1$, and $\overline{\nu}_{OLS} = \nu + O_e\left(n^{\gamma-\delta}\right)$ when $\delta + \gamma > 1$ (notice the use of an upper bound for the orders of magnitude, rather than the exact rate, when $\delta + \gamma \leq 1$: this is because the rate of convergence may be faster, when x_t and v_t are not correlated at any lag).

Since the inconsistency of OLS when $\delta < 1/2$ and the suboptimal rate of convergence when $\delta \leq 1$ are caused by the correlation between x_t and the cointegrating errors v_t in (1.30), Robinson (1994b) suggested to focus on the lowest frequencies, where the "noise" due to the correlation with v_t should be of a lower order when compared to the "signal" in x_t. The idea of a regression on selected frequencies with the purpose to minimise the bias induced by the noise in the extraction of a signal is due to Hannan (1963), who discussed time series with continuous spectra. Robinson (1994b) exploited the power law approximation (1.15) to reduce the bias: he suggested to run the regression on low frequencies only, and he introduced the crucial assumption that the

band (the set $\{\lambda_0, ..., \lambda_m\}$ or $\{\lambda_1, ..., \lambda_m\}$, where $m \leq n$ is the bandwidth) degenerates towards zero as in the log-periodogram regression or in the local Whittle estimation.

The narrow band least square estimate (NBLS), $\overline{\nu}_{NBLS}$, is defined as

$$\overline{\nu}_{NBLS} = \frac{\sum_{j=l}^{m} \mathrm{Re}\left(I_{xy}\left(\lambda_j\right)\right)}{\sum_{j=l}^{m} I_x\left(\lambda_j\right)}, \tag{1.67}$$

where

$$I_{xy}\left(\lambda_j\right) = F_x\left(\lambda\right) F_y\left(-\lambda\right) \tag{1.68}$$

is the cross periodogram of x_t and y_t, and m is such that $\frac{1}{m} + \frac{m}{n} \to 0$ as $n \to \infty$; the choice of l depends on the presence of an intercept in (1.65): l is set to 1 if an unknown intercept is included in the model, and to 0 otherwise.

Robinson and Marinucci (2003) showed that the NBLS estimate is consistent even when $\delta < 1/2$, in which case $\overline{\nu}_{NBLS} = \nu + O_p\left((n/m)^{\gamma-\delta}\right)$ (and they conjectured that the rate is sharp), while when $\delta > 1/2$, $\overline{\nu}_{NBLS} = \nu + O_e\left(n^{\gamma-\delta} m^{1-\gamma-\delta}\right)$ when $\gamma + \delta < 1$, $\overline{\nu}_{NBLS} = \nu + O_e\left(n^{1-2\delta} \ln m\right)$ when $\gamma + \delta = 1$ and $\delta < 1$, and $\overline{\nu}_{NBLS} = \nu + O_e\left(n^{\gamma-\delta}\right)$ when $\gamma = 0$, $\delta = 1$ or when $\gamma + \delta > 1$. The rates of convergence are then improved with respect to OLS when the joint memory is relatively small, i.e. $\gamma + \delta \leq 1$: the only exception is for $\delta = 1$ and $\gamma = 0$, but even in that situation NBLS can be preferred, because it succeeds in eliminating the lower order bias.

An even faster rate of convergence was attained by Chen and Hurvich (2003) for some combinations of δ and γ: they kept m fixed and used a combination of tapering and differencing of the data.

The OLS and NBLS estimates have the advantages of being very simple to compute, and of not requiring any preliminary knowledge of the distribution of the cointegrating errors (in this sense, they can both be considered semi-

parametric). On the other hand the limit distribution is not standard, and Wald type statistics are not asymptotically χ^2, so a test on ν based on $\overline{\nu}_{OLS}$ or $\overline{\nu}_{NBLS}$ is not really practical because it requires critical values that depend on the model of $u_{1,t}$ and $u_{2,t}$.

Maximum likelihood, or pseudo-maximum likelihood if Gaussianity is not assumed, may overcome these shortcomings.

Given the model $B(L;\theta)u_t = \varepsilon_t$, $E(\varepsilon_t\varepsilon_t') = \Omega(\theta)$, when the parameters γ, δ and θ are known a closed form estimate based in the time domain is

$$\widetilde{\nu}(\gamma,\delta,\theta) = \frac{\sum_{t=1}^{n}\left(b(L;\theta)x_t(\gamma)\right)'\Omega(\theta)^{-1}\left(B(L;\theta)\left(y_t(\gamma),x_t(\delta)\right)'\right)}{\sum_{t=1}^{n}\left(b(L;\theta)x_t(\gamma)\right)'\Omega(\theta)^{-1}\left(b(L;\theta)x_t(\gamma)\right)} \tag{1.69}$$

where $b(L;\theta)$ is the first column of $B(L;\theta)$. In the same set-up, a closed form estimate based in the frequency domain is

$$\widehat{\nu}(\gamma,\delta,\theta) = \frac{\sum_{j=1}^{n}p(\lambda_j;\theta)F_{x(\gamma)}(\lambda_j)\left(F_{y(\gamma)}(-\lambda_j),F_{x(\delta)}(-\lambda_j)\right)'}{\sum_{j=1}^{n}q(\lambda_j;\theta)I_{x(\gamma)}(\lambda_j)}. \tag{1.70}$$

where $p(\lambda;\theta) = (1,0)f_u^{-1}(\lambda;\theta)$ and $q(\lambda;\theta) = (1,0)f_u^{-1}(\lambda;\theta)(1,0)'$.

Since they both have the same asymptotic properties at least when $\beta > 1/2$, as Robinson and Hualde (2003) showed, in the rest of the subsection we will comment on $\widehat{\nu}(\gamma,\delta,\theta)$ only.

Phillips (1991) showed that when $\delta = 1$, $\gamma = 0$, and θ is known, under regularity conditions $\widehat{\nu}(0,1,\theta)$ is asymptotically mixed normal, with $\widehat{\nu}(0,1,\theta) = \nu + O_e(n^{-1})$, and the Wald test on ν has limit χ_q^2 distribution, where q is the number of restrictions tested.

When any of θ, γ or δ are unknown, maximum likelihood requires joint estimation with ν.

When γ and δ are known, efficient estimation of ν and θ can be realised with one single regression, even though possibly a non-linear one. By rewriting

(1.65) as

$$\Delta^{\gamma} y_t = \nu \Delta^{\gamma} x_t + (\omega_{12}/\omega_{22})\, \Delta^{\delta} x_t + \varepsilon_{1,t} - (\omega_{12}/\omega_{22})\, \varepsilon_{2,t}, \qquad (1.71)$$

ν can be estimated by maximum likelihood as the coefficient of $\Delta^{\gamma} x_t$ in the regression of $\Delta^{\gamma} y_t$ on $\Delta^{\gamma} x_t$ and $\Delta^{\delta} x_t$: Phillips (1991) anticipated it for $\delta = 1$, $\gamma = 0$ and under the assumption of independence for u_t, and Robinson and Hualde (2003) generalised it to fractional orders (provided that $\beta > 1/2$); Phillips and Loretan (1991) discussed the extension to an autoregressive structure in u_t, augmenting the model (1.71) by leads and lags of $\Delta^{\delta} x_t$ and by lags of $\Delta^{\gamma} v_t$. Phillips and Loretan (1991) only considered $\delta = 1$, $\gamma = 0$, and even in that case the regression imposes non-linear constraints, so a two-step procedure may actually be faster. For that case, Phillips (1991) showed that the limit distribution of $\widehat{\nu}(0, 1, \theta)$ does not change if θ is replaced by a consistent estimate $\widehat{\theta}$.

On the other hand, the assumption of preliminary knowledge of the unit root was essential: indeed, if ρ was estimated rather than imposed as $\rho = 1$ in

$$x_t = \rho x_{t-1} + u_{2,t} \qquad (1.72)$$

(notice that when $\rho = 1$ then $\delta = 1$ so this is another way of formulating (1.66)), then Phillips (1991) showed that the limit distribution of the estimate of ν is in general contaminated by the unit root distribution of the estimate of ρ.

Robinson and Hualde (2003) showed that this difference in the limit distribution of ν depended on the restriction to integer orders only for δ and γ, and on the estimation of δ via the regression of x_t on x_{t-1}. They compared $\widehat{\nu}(\gamma, \delta, \theta)$, $\widehat{\nu}\left(\gamma, \delta, \widehat{\theta}\right)$ and $\widehat{\nu}\left(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}\right)$, where $\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}$ are consistent estimates of

γ, δ, θ, and found that the three estimates of ν all have the same limit distribution (regularity conditions included $\beta > 1/2$ and minimal rates of convergence for $\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}$). They also derived the limit distribution of these estimates of ν, finding that $\widehat{\nu}(\gamma, \delta, \theta) = \nu + O_e\left(n^{\gamma-\delta}\right)$ with mixed normal asymptotics and that a Wald test on ν has limit χ_1^2 distribution (χ_1^2 because they only considered a scalar ν).

The situation $\gamma = 0$, $\delta = 1$ did not require a particular treatment or discussion, except possibly the remark that then the results were the same as those of Phillips (1991) when $\delta = 1$, $\gamma = 0$ are known in advance. The fact that preliminary estimation of δ (or of γ, for that matter) does not affect the limit distribution of the estimate of the cointegrating parameter, is a remarkable difference with respect to the result of Phillips (1991), and it seems to confirm that estimating a possibly fractional memory parameter rather than imposing the alternative between short range dependence or unit root is a more natural approach.

The case in which $\beta < 1/2$ was called by Hualde and Robinson (2002) "weak cointegration".

It is fair to conjecture that under regularity conditions the maximum likelihood estimates are root-n consistent and Gauss-Markov efficient. The regression estimate, on the other hand, may generate nonstandard asymptotics, and possibly a lower rate of convergence as well, as discussed by Robinson (1994a), or even inconsistency.

Assuming no correlation between x_t and v_t, Hidalgo and Robinson (2002) proposed adaptive GLS estimation, showing that the resulting estimate is root-n consistent and Gauss-Markov efficient. Allowing for potential correlation, Robinson and Marinucci (2003) discussed NBLS under stationarity: the estimate is consistent but (they conjectured) converges at a slower rate. Hualde

and Robinson (2002) on the other hand proposed a simple two-step procedure that delivers root-n consistent estimation of the cointegrating parameter even when x_t and v_t are correlated, although the estimate may be less efficient than the maximum likelihood one.

1.3.3 Testing for cointegration

Engle and Granger (1987) also considered the problem of testing for a cointegrating relation. They proposed to run an OLS regression and then to test for a unit root in the residuals with the augmented version of the Dickey-Fuller test. This approach was further discussed by Hansen (1992), who noticed that in that case the critical values of Said and Dickey (1984) for that unit root test are not valid, and that the correct ones depend on the number and on the nature (deterministic or stochastic) of the regressors.

Yet if the cointegration rank is unknown, detecting the cointegrating relations through testing the order of the residuals may be problematic, because all possible combinations should be considered. Several different procedures then have been developed to estimate the cointegration rank.

Phillips and Ouliaris (1988) proposed to estimate the cointegration rank by looking at the rank of G_z in (1.25). Since that matrix is unknown, they considered a nonparametric estimate of $f_{\Delta z}(0)$ (they assumed a $CI(1,1)$ model): the rank of G_z is then estimated by testing how many eigenvalues of that estimate are significantly different than zero. Unfortunately, the limit distribution theory only covers the case in which the eigenvalues are not zero, but Phillips and Ouliaris (1988) proposed to use this procedure and that limit distribution anyway, arguing that it would help at least to spot the situations in which the eigenvalues are far away from 0.

A third procedure to estimate the cointegration rank, still under the $CI(1,1)$

assumption, was proposed by Johansen (1991) in a maximum likelihood framework. Johansen (1991) derived the limit distribution of the likelihood ratio test of the hypothesis that the cointegration rank is r against $r+1$: it is not a standard χ^2, but he showed that it only depends on r and p and on the type of deterministic component in the model, and it can be tabulated.

Testing for cointegration by looking at the order of integration of the residuals of a regression is also popular when fractional cointegration is analysed. In many applied works this had been done by a semiparametric estimate, a procedure that actually seems more appropriate, given that the residuals also had been estimated semiparametrically (usually by OLS or NBLS).

An early example was provided by Cheung and Lai (1993), who used log-periodogram regression on OLS residuals to investigate a potential Purchasing Power Parity (PPP) relation in the long run. Cheung and Lai (1993) also argued that when the estimate of γ is based on estimated residuals then its limit distribution is not necessarily normal, supporting their claim with a Monte Carlo exercise. Indeed Hassler, Marmol and Velasco (2006) found that the necessary conditions for limit normality are rather strong if OLS residuals are used: $\beta > 1/2$ is required and the lowest frequencies have to be trimmed and excluded from the regression, and an even larger β is necessary if $\delta + \gamma < 1$.

Robinson (2005b) obtained a root-n consistent estimate of γ under the milder condition that $\delta > \gamma + 1/2$ and no trimming: he obtained this much stronger result by using more information, because he discussed Whittle estimation, which is parametric and uses the whole range $\left\{\frac{2\pi}{n}, .., \frac{2\pi(n-1)}{n}\right\}$, and by employing a better estimate of the cointegrating parameter in the first step, because he considered the residuals of a NBLS regression (OLS residuals can also be used if $\delta + \gamma \geq 1$).

However, in the fractional setting, the simple estimation of the order γ

only gives a heuristic piece of evidence rather than to provide a proper test of cointegration, because the order δ is unknown as well.

Marinucci and Robinson (2001) thus proposed an alternative test, based on the remark that the rank of the matrix G_z in (1.25) is reduced under cointegration. They considered a variation of (1.63), augmented by the additional assumption that the order of integration is the same for all the cointegrated processes:

$$\widehat{\delta}_{*,LW} = \arg \min_{d_* \in \Theta_*} L\left(d_* 1_p\right) \tag{1.73}$$

where $d_* 1_p$ is a p-dimensional vector in which each element is d_* (scalar), and Θ_* is a compact subset in $(-1/2, 1/2)$.

When $p = 2$ and the rank of G_z is full, under $m = o\left(n^{4/5}\right)$ (or less, depending as usual on the smoothness of $f_z(\lambda)$ as $\lambda \to 0$) and other regularity conditions

$$\sqrt{8m}\left(\widehat{\delta}_{*,LW} - \delta_*\right) \to_d N(0,1) \text{ as } n \to \infty. \tag{1.74}$$

When the restriction to a common order of integration is correct, estimating the memory parameter using two processes jointly is more efficient than estimating it using just one of the series: indeed Marinucci and Robinson (2001) considered

$$\widetilde{H}_k = 8m(\widehat{\delta}_{*,LW} - \widehat{\delta}_{k,LW})^2 \tag{1.75}$$

where $\widehat{\delta}_{k,LW}$, the local Whittle estimate of the memory parameter of the kth element of bivariate vector z_t, is computed by minimising the loss function for a scalar process as in (1.53), and showed that

$$\widetilde{H}_k \to_d \chi_1^2 \text{ as } n \to \infty. \tag{1.76}$$

If on the other hand G_z is singular, Marinucci and Robinson (2001) argued

that $\widehat{\delta}_{*,LW}$ is inconsistent and the statistic in (1.75) diverges. Marinucci and Robinson (2001) then suggested to test for cointegration by comparing $\widehat{\delta}_{*,LW}$ and $\widehat{\delta}_{k,LW}$, rejecting cointegration if (1.75) is below a critical value (given the similarity with the Hausman test, Marinucci and Robinson referred to this as a "Hausman-type" test).

Robinson and Yajima (2002) proposed a variation of the test of Phillips and Ouliaris (1988) in which the δ_* difference is taken by applying the filter $\Lambda(\lambda; \delta_* 1_p)$ as in (1.61) to the periodogram $I_z(\lambda)$, thus estimating G_z as $\widehat{G}_z(\delta_* 1_p)$ in (1.62), provided that $\delta_* \in (0, 1/2)$. Since δ_* is unknown, and cannot be estimated using the multivariate approach (1.73) because G_z is not of full rank under cointegration, they proposed

$$\overline{\delta}_{*,LW} = (\widehat{\delta}_{1,LW} + ... + \widehat{\delta}_{p,LW})/p. \tag{1.77}$$

Robinson and Yajima (2002) also observed that if the same bandwidth is used for $\overline{\delta}_{*,LW}$ and $\widehat{G}_z(\overline{\delta}_{*,LW} 1_p)$, then these are perfectly correlated, so they suggested to compute $\overline{\delta}_{*,LW}$ using another bandwidth m_1 that increases sufficiently fast to remove that effect. Let $\lambda_1(\delta_* 1_p)$, ..., $\lambda_p(\delta_* 1_p)$ be the ordered eigenvalues of $G_z(\delta_* 1_p)$ and let $\widehat{\lambda}_1(\overline{\delta}_{*,LW} 1_p)$, ..., $\widehat{\lambda}_p(\overline{\delta}_{*,LW} 1_p)$ be the ordered eigenvalues of $\widehat{G}_z(\overline{\delta}_{*,LW} 1_p)$, and drop $(\delta_* 1_p)$ and $(\overline{\delta}_{*,LW} 1_p)$ in order to make the notation lighter: Robinson and Yajima (2002) defined

$$\sigma_{k,l}^{(i)} = \sum_{a=k}^{l} \lambda_a^i, \; \widehat{\sigma}_{k,l}^{(i)} = \sum_{a=k}^{l} \widehat{\lambda}_a^i \tag{1.78}$$

$$\pi_j = \frac{\sigma_{p-j+1,p}^{(1)}}{\sigma_{1,p}^{(1)}}, \; \widehat{\pi}_j = \frac{\widehat{\sigma}_{p-j+1,p}^{(1)}}{\widehat{\sigma}_{1,p}^{(1)}} \text{ for } j = 1, ..., p-1 \tag{1.79}$$

$$s_j = \frac{\left(\widehat{\sigma}^{(1)2}_{p-j+1,p}\widehat{\sigma}^{(2)}_{1,p-j} + \widehat{\sigma}^{(1)2}_{1,p-j}\widehat{\sigma}^{(2)2}_{p-j+1,p}\right)^{1/2}}{\widehat{\sigma}^{(1)2}_{1,p}} \tag{1.80}$$

and showed that when the rank of G_z is full

$$m^{1/2}(\widehat{\pi}_j - \pi_j)/s_j \to_d N(0,1) \text{ as } n \to \infty. \tag{1.81}$$

To test for the cointegration rank r they considered

$$\widehat{t}_r = \widehat{\pi}_r + cv_a s_r / m^{1/2} \tag{1.82}$$

(where cv_a is the critical value for the size a): evidence of cointegration is found if the computed value of the expression in (1.82) is below a pre-specified threshold (they suggested $0.1/p$). They also proposed, as an alternative, to confront $\widehat{\pi}_r$ with another pre-specified threshold (they suggested $0.01/p$).

The last test we present was proposed by Breitung and Hassler (2002) and is based on the extension to the fractional case of the maximum likelihood approach. They considered (1.65) and (1.66) with the additional assumptions that a representation $B(L;\theta)\, u_t = \varepsilon_t$ is feasible, and in particular that $u_{x,t} = (u_{r+1,t}, ..., u_{p,t})'$ has VAR(k) structure, that the orders of the observations and of the cointegrating errors are always the same, so $\delta_1 = ... = \delta_p = \delta_*$ and $\gamma_1 = ...\gamma_r = \gamma_*$ (although they stated that this restriction was only introduced to ease the notation) and that there are no parametric restrictions across the elements of Ω and $\nu, \delta_*, \gamma_*, \theta$. Finally, they assumed that $\beta_* = \delta_* - \gamma_* > 1/2$. Introducing

$$e_t = \Delta^{\delta_*} z_t,\ \widehat{\varepsilon}^{\dagger}_t = z_t - B_1(\theta)\, e_{t-1} - ... - B_k(\theta)\, e_{t-k} \tag{1.83}$$

where $B_1, ..., B_k$ are the first k elements of the VAR representation for u_t, and

$$\widehat{\varepsilon}^*_{t-1} = \sum_{j=1}^{t-k-1} j^{-1}\widehat{\varepsilon}^{\dagger}_{t-1}, \quad \widetilde{\varepsilon}^*_{t-1} = \left(\widehat{\varepsilon}^{*\prime}_{t-1}, e'_{t-1}, ..., e'_{t-k}\right)' \tag{1.84}$$

$$\widetilde{\Omega} = 1/n \sum_{t=1}^{n} \widehat{\varepsilon}^{\dagger}_t \left(\widehat{\varepsilon}^{\dagger}_t\right)', \; \widetilde{S}_{10} = \sum_{t=1}^{n} \widehat{\varepsilon}^*_{t-1} \left(\widehat{\varepsilon}^{\dagger}_t\right)', \; \widetilde{S}_{11} = \sum_{t=1}^{n} \widehat{\varepsilon}^*_{t-1}\widehat{\varepsilon}^{*\prime}_{t-1}, \tag{1.85}$$

following Johansen (1991), Breitung and Hassler (2002) tested the hypothesis of r cointegrating vectors by computing $\widetilde{\lambda}_1, ... , \widetilde{\lambda}_p$, the solutions of the equation in λ

$$\left|\lambda\widetilde{\Omega} - \widetilde{S}'_{10}\left(\widetilde{S}_{11}\right)^{-1}\widetilde{S}_{10}\right| = 0. \tag{1.86}$$

Ordering these eigenvalues as $\widetilde{\lambda}_1 \leq ... \leq \widetilde{\lambda}_p$, Breitung and Hassler then showed that the trace statistic $\sum_{a=1}^{p-r}\widetilde{\lambda}_a$ is asymptotically $\chi^2_{(p-r)^2}$. Since in practice δ_* and $B_1(\theta), ..., B_k(\theta)$ are unknown, Breitung and Hassler suggested to replace them by consistent estimates.

Cointegration requires that the order of integration of the processes y_t and x_t is the same. In the $CI(1,1)$ case this is usually verified by testing for a unit root for both the series. In the fractional model though there is no reason to expect any particular value for δ_*, so the test has to be different.

Robinson and Yajima (2002) proposed to test the hypothesis

$$H_0 : \{\delta_k = \delta_l\} \tag{1.87}$$

where l and k refer to the position in the p-dimensional vector z_t, against the alternative that the two parameters are different, by using the statistic

$$\widetilde{T}_{k,l} = \frac{m^{1/2}(\widehat{\delta}_{k,LW} - \widehat{\delta}_{l,LW})}{\{\frac{1}{2}(1 - \widehat{G}^2_{kl}/(\widehat{G}_k\widehat{G}_l)\}^{1/2}}, \tag{1.88}$$

where G_l and G_k are the elements in the main diagonal of the $p \times p$ matrix

G_z, and G_{kl} are the elements in the position (k,l), and $\widehat{G}_z$ is estimated as in (1.62). When the process is not cointegrated,

$$\widetilde{T}_{k,l} \rightarrow_d N(0,1) \text{ as } n \rightarrow \infty, \tag{1.89}$$

but otherwise the test statistic is not well defined. Robinson and Yajima (2002) then also suggested the alternative statistic

$$\widehat{T}_{k,l} = \frac{m^{1/2}(\widehat{\delta}_{k,LW} - \widehat{\delta}_{l,LW})}{\{\frac{1}{2}(1 - \widehat{G}_{kl}^2/(\widehat{G}_k\widehat{G}_l)\}^{1/2} + h\,(n)}, \tag{1.90}$$

where $h(n)$ is a sequence that tends to zero at an appropriate speed, because under cointegration

$$\widehat{T}_{k,l} \rightarrow_p 0 \text{ as } n \rightarrow \infty, \tag{1.91}$$

while otherwise the same limit distribution of $\widetilde{T}_{k,l}$ holds. They remarked that $\left|\widehat{T}_{k,l}\right| < \left|\widetilde{T}_{k,l}\right|$ so a non rejection of the hypothesis from $\widetilde{T}_{k,l}$ would be made with even greater confidence from $\widehat{T}_{k,l}$. Robinson and Yajima also generalised the statistic $\widehat{T}_{k,l}$ to test several hypotheses simultaneously by a χ^2 test.

1.4 An applied example: a fractional cointegration analysis of the term structure of interest rates

1.4.1 Introduction

We apply the techniques described in Section 1.3 to estimate the memory parameter, to determine the cointegrating rank of a vector, and to estimate a cointegrating matrix, for a vector of four interest rates.

The application of the fractional model to a vector of interest rates is of

particular importance because it helps to reconcile two apparently conflicting groups of theories about the order of integration of the interest rates.

By plotting together two comparable interest rates with different maturities, it is fair to conjecture that the long term dynamics of the two series is largely driven by a highly persistent common factor, while the persistence of the difference between those two rates (the "spread") is much less. This of course matches the intuitive description of the concept of cointegration, and it is not surprising then that an application for two rates with different maturities had been proposed already by Engle and Granger (1987).

Campbell and Shiller (1987) provided a theoretical model linking rates with different maturities: they showed that if two different rates are $I(1)$, then under rational expectations they are cointegrated with $\nu = 1$; if a vector of p rates is considered, then the cointegration rank is $r = p - 1$. Campbell and Shiller (1987) proposed a $CI(1,1)$ model, but fractional integration can be introduced in their design without any modification: applying rational expectations (and then adding an $I(0)$, *i.i.d.* disturbance term) to their equation (4), the cointegrating errors should be of order $\max\{\delta - 1, 0\}$.

If indeed the interest rates are integrated, then a cointegrated model is also necessary for a successful transmission of monetary policy. The central bank operates by supplying liquidity on the interbank market by open market operations and discount window loans, so that the short term rate is tightly managed by the monetary authority. Yet the relevant macroeconomic indicators are only affected by the rates of contracts with much longer maturities: the demand of money from the individual agents depends on the return of the alternative asset, which is more likely to be the Treasury bills or Treasury bonds rate, while the economic activity depends on other rates, like the bank loan rate or the commercial paper rate. The transmission of monetary policy then requires that an impulse originated in a market for a very short

term contract is transmitted to contracts with much longer maturity, and for a successful policy it is also necessary that the transmission is fast and reliable.

If we are restricted to integer orders only, the evidence seems to be in favour of a $CI(1,1)$ model, and it is quite robust with respect to changes in the pairs of rates used (Treasury bills and bonds, eurodeposits, synthetic rates generated by fitting splines...) and to the sample period. When more than two rates were analysed jointly, though, the hypothesis of $p-1$ independent cointegrating vectors was not always met: Hall, Anderson and Granger (1992), Engsted and Tangaard (1994) and Lanne (2000) found more than one stochastic trend, but explained the result allowing for a structural break in 1979 (Hall *et al.* also introduced a break in 1982); Domingues and Novales (2000) too found that the results were sensitive to whether the sample started in 1979 or a few years later. The potential breaks were explicitly tested by Hansen (2003), who found that the two changes in monetary policy in 1979 and in 1982 altered the short run dynamics of the rates.

Most researchers assumed integer orders only, and in many cases applied the maximum likelihood approach introduced by Johansen (1991). Arguing that the $I(1)$ representation contrasts the experience that rates do not take negative values, Lanne (2000) proposed near unit roots instead.

An $I(1)$ interest rate is also difficult to motivate because it is likely to imply an $I(1)$ model for inflation also: this is the case, for example, if the Fisher equation holds, or if the central bank sets the interest rate using a linear reaction function like the ones described by Taylor (1993) or by Svensson (1997). Such a strong persistence in inflation is hardly acceptable in monetary theory, because it implies that the central bank does not stabilise inflation around a constant value, as the inflation targeting commitment requires. Indeed, Clarida, Gali and Gertler (2000), for example, adopted the $I(0)$ representation instead, and dismissed the results of the Dickey-Fuller test arguing that they

were due to the low power rather than to a unit root.

1.4.2 Empirical analysis

We applied the fractional model (1.65) and (1.66) to the US interest rates with maturities of 1, 3, 6 and 12 months. We augmented the equations by non-zero constant means, but imposed no trends because they would imply explosive rates. By indicating the four rates as $i1_t$, $i3_t$, $i6_t$, 12_t respectively, and under the assumption that $r = 3$, the model prescribed by the theory is

$$\begin{cases} i1_t = \mu_1 + u_{1,t}\left(-\delta_*\right) \\ i3_t = \mu_3 + \nu_3 i1_t + u_{3,t}\left(-\gamma_3\right) \\ i6_t = \mu_6 + \nu_6 i1_t + u_{6,t}\left(-\gamma_6\right) \\ i12_t = \mu_{12} + \nu_{12} i1_t + u_{12,t}\left(-\gamma_{12}\right). \end{cases} \tag{1.92}$$

Notice that we do not assume $r = 3$, but test for it instead; for each rate we assume the univariate model

$$ij_t = \mu_j + u_{j,t}\left(-\delta_j\right),\ j \in \{1, 3, 6, 12\}. \tag{1.93}$$

The contract is the London interbank deposit in US$ (monthly averages of the offer rate) over the period 10/1979 to 01/2002 (inclusive), corresponding to the DataStream identification codes USI60LDC, USI60LDD, USI60LDE, USI60LDF. The period was selected because Clarida Gali and Gertler (2000) suggested that, with the appointment of Volker as chairman, the Fed took a more aggressive attitude towards inflation, and indeed a break at that point was suspected in many applied analyses. The London InterBank Offer Rate (LIBOR) is a typical measure of the cost of funds in the US, because it is not affected by any regulation imposed by the central bank; in fact for the same

argument the LIBOR had been in the past a good measure of the effective cost of funds for several European countries too.

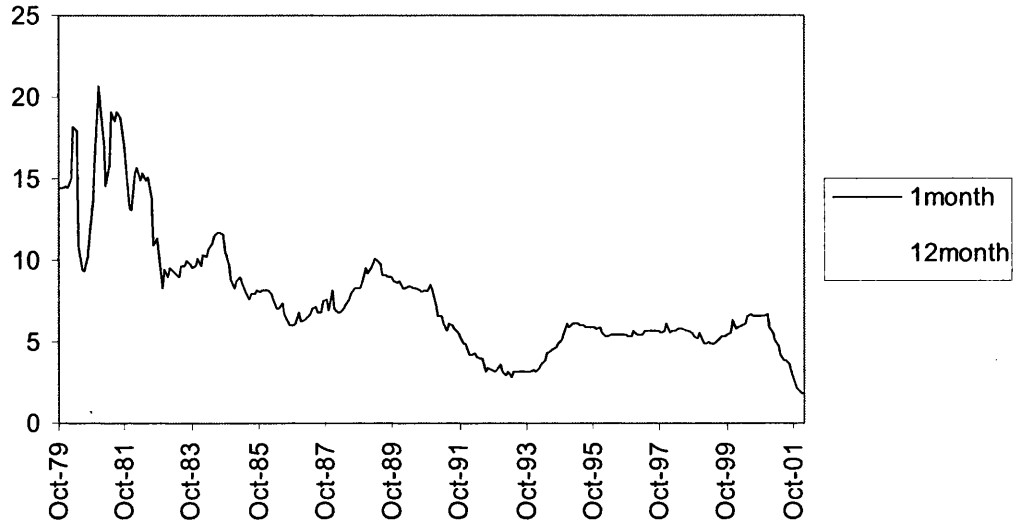
In the rest of the section we only use some of the procedures described in the previous part, so we can shorten the notation by using $\widetilde{\delta}$, $\widetilde{\gamma}$ for $\widehat{\delta}_{LW}$, $\widehat{\gamma}_{LW}$ as in (1.53) and $\widetilde{\nu}$ for $\overline{\nu}_{NBLS}$ as in (1.67); we also drop the subscripts LW in $\overline{\delta}_{*,LW}$ (1.77) and OLS in $\overline{\nu}_{OLS}$; finally we use $\widehat{\delta}$, $\widehat{\gamma}$, $\widehat{\nu}$ for the maximum likelihood estimates of the corresponding parameters (we compute these by minimising a conditional sum of square like in (1.60) but applying the approach to a multivariate cointegrated process; more details are to be found further on in this subsection, see (1.100)). In all the tests, the size is 5% unless specified otherwise.

We run the analysis in two parts: a preliminary, semiparametric treatment of the data, and then a fully parametric one. We begin with the semiparametric analysis in order to obtain robust evidence. For this reason, we prefer a conservative approach in the choice of the bandwidths: for the local Whittle procedure and similar ones, we set $m = 25$ (unless the theory required otherwise, as in the computation of $\overline{\delta}_*$ in the test of Robinson and Yajima (2002)): this is optimal for an AR(1) structure with autoregressive coefficient of 0.4 for the short memory component, but of course we did not assume such a structure. Following Marinucci and Robinson (2001), we also set a much smaller m to estimate the cointegration parameter.

The plot of $i1$ and $i12$ is in Figure 1.1.

A naive inspection of the plot suggested that if the data are reverting to a potential mean, they are only doing it very slowly. The downward trends in the first part of the period could be a movement towards the long run equilibrium after a particularly large perturbation, if $\delta_j < 1$, but we did not rule out the possibility that $\delta_j = 1$, as in the mainstream cointegration literature. The

Figure 1.1: 1 month and 12 months rates



plot is also informative of a potential common stochastic trend: indeed the two rates seem to move together, the differences vanishing rather quickly.

Since it is widely accepted that $\delta_j \leq 1$, we estimated the memory parameters both for the levels and for the first differences. We estimated δ_j by the local Whittle procedure (we initialised the optimisation by the log-periodogram regression estimate). The results are presented in Table 1.1: the label δ_i refers to the memory parameters for data in levels, $\delta_{\Delta i}$ to the first differences; a.s.e. indicates the asymptotic standard errors as per (1.54).

Table 1.1: local Whittle estimates of the memory parameters

	$i1$	$i3$	$i6$	$i12$
$\widetilde{\delta}_i$	0.75	0.76	0.78	0.81
$\widetilde{\delta}_{\Delta i}$	-0.14	-0.11	-0.08	-0.05
a.s.e.	0.1	0.1	0.1	0.1

The estimates of δ_j were indeed below 1, but we never rejected the null hypothesis in the four tests

$$H_0 : \{\delta_{\Delta i,j} = 0\} \text{ v.s. } H_1 : \{\delta_{\Delta i,j} < 0\} \tag{1.94}$$

on the basis of a t test.

The estimates were, in any case, very distant from 0, and testing

$$H_0 : \{\delta_{i,j} = 0\} \text{ v.s. } H_1 : \{\delta_{i,j} > 0\} \tag{1.95}$$

the null hypothesis was always rejected on the basis of a t test.

In order to justify the $I(0)$ structure despite the extensive evidence in the literature against it based on the Dickey and Fuller test, it was often argued that the power of the test is very low. Having estimated the orders of integration directly, we treated both the $I(0)$ and $I(1)$ specifications in the same way, and the evidence against the $I(0)$ model was then far more convincing. Indeed, considering our estimated orders of integration, it is not surprising that the conventional unit root tests were in favour of the $I(1)$ model: even assuming that the data were neither $I(0)$ nor $I(1)$, they appeared to be far closer to being $I(1)$, and the Dickey and Fuller test just reflected this fact.

We next tested the hypothesis that the memory was the same for all the series: the squares of the pairwise statistics $\widetilde{T}_{k,l}$ (1.88) are in Table 1.2. In accordance with our previous findings, we run the tests on first differences of the data. We never rejected the null hypothesis of a common order of integration.

This result was confirmed by the joint test that the four rates are the same: the realisation of test statistics was 2.74, well below 7.81, the 5% critical value of a χ^2_3 distribution.

Table 1.2: Semiparametric tests for the equality of the orders

pairs:	$\Delta i1, \Delta i3$	$\Delta i1, \Delta i6$	$\Delta i1, \Delta i12$	$\Delta i3, \Delta i6$	$\Delta i3, \Delta i12$	$\Delta i6, \Delta i12$
$\widetilde{T}^2_{k,l}$	2.31	1.99	2.18	1.44	1.76	1.67

To estimate the memory of the data more efficiently we then pooled the four individual estimates, and we computed $\overline{\delta}_{*,\Delta i} + 1 = 0.91$.

Having obtained evidence in favour of a common order of integration, we tested for cointegration. The obvious candidate cointegration rank was $p - 1$ (3 then in our case), since this was the one required under the expectations theory and it was also used in a wide number of empirical analyses. We begun by testing if pairs of interest rates were cointegrated by the Hausman-type of test (1.75) proposed by Marinucci and Robinson (2001). The results of the tests are presented in Table 1.3: $\widetilde{H}_x$ refers to the test statistic (1.75) computed using the interest rate with shorter maturity, $\widetilde{H}_y$ to the other one. We rejected the null of no cointegration in 9 combinations out of 12. Moreover, in all the combinations we always found that $\widetilde{\delta}_*$ from (1.73) was lower than both the individual estimates of the orders of the two series, so we suspect that the failure to reject in the remaining cases may be due to a type II error.

Table 1.3: Marinucci and Robinson (2001) test for (no) cointegration

pairs:	$\Delta i1, \Delta i3$	$\Delta i1, \Delta i6$	$\Delta i1, \Delta i12$	$\Delta i3, \Delta i6$	$\Delta i3, \Delta i12$	$\Delta i6, \Delta i12$
$\widetilde{\delta}_{*,\Delta i}$	-0.32	-0.31	-0.26	-0.31	-0.24	-0.20
$\widetilde{H}_x$	6.64	6.21	2.87	8.19	3.29	2.51
$\widetilde{H}_y$	8.78	10.57	8.69	10.46	7.20	4.32

We also analysed the data using the test proposed by Robinson and Yajima (2002) (of course we intend the definition to be valid for Type II fractionally integrated processes as well, and we used the vector of the four rates instead of z_t). Notice that Robinson and Yajima (2002) only formulated the test for variables having $\delta_* < 0.5$: on the basis of Velasco (1999b) we may conjecture that the same procedure holds for $\delta_* < 0.75$ but this is unlikely to be the case anyway, so we considered two alternatives. In the first procedure we took δ_* differences of the data in the time domain, setting all the observations

before the first one to zero, as is common practice. We then estimated G_z in (1.25) simply by averaging the periodograms over the first m frequencies, as if computing $\widehat{G}_{\Delta^{\delta_* i}}(0 \times 1_p)$ as in (1.62): this is exactly the procedure of Phillips and Ouliaris (1988) allowing for fractional integration as well. In the second procedure we computed the first differences for the data and then estimated G_z by $\widehat{G}_{\Delta i}((\delta_* - 1)1_p)$ as in (1.62). We refer to these two approaches as "time domain" based and "frequency domain" based respectively. In both the procedures, δ_* was actually unknown, so we estimated it by computing $\overline{\delta}_{*,\Delta i}$ and then by adding 1 back; notice that we used 27 rather than 25 frequencies as suggested by Robinson and Yajima (2002).

In Table 1.4 we present the test statistics $\widehat{\pi}_r$, $\widehat{t}_r$ for the joint four dimensional vector. The hypothesis of interest in that case was $r = 3$, as prescribed by the economic theory. With the time domain approach, the maximum of the rescaled sum of the eigenvalues, $\widehat{\pi}_3$, was about 0.009, larger than the suggested threshold $0.01/p$=0.0025; the statistic $\widehat{t}_3$ on the other hand was well within the threshold $0.1/p$=0.025. With the frequency domain approach, on the other hand, both these indicators rejected the null hypothesis.

Table 1.4: Robinson and Yajima (2002) cointegration tests: joint test

rank tested:	$r = 3$	$r = 2$	$r = 1$
	time domain		
$\widehat{\pi}_r$	0.00888	0.00034	0.00005
$\widehat{t}_r$	0.01178	0.00046	0.00007
	frequency domain		
$\widehat{\pi}_r$	0.02749	0.00110	0.00017
$\widehat{t}_r$	0.03628	0.00145	0.00023

We also considered testing only pairs of rates: in that case the null hypothesis was $r = 1$. These results are in Table 1.5. With the time domain approach,

the evidence was mixed for the $\widehat{\pi}_1$ statistics, cointegration not having been rejected for rates that were "contiguous" in terms of maturities ($i1_t$ and $i3_t$, $i3_t$ and $i6_t$, $i6_t$ and $i12_t$) but not otherwise; no rejections took place when $\widehat{t}_1$ was used. Once again, the frequency domain approach yielded less rejections of the null hypothesis: 1 out of 6 for the $\widehat{\pi}_1$ statistics and 5 out of 6 for $\widehat{t}_1$.

Table 1.5: Robinson and Yajima (2002) cointegration tests: pairwise tests

pairs:	$\Delta i1, \Delta i3$	$\Delta i1, \Delta i6$	$\Delta i1, \Delta i12$	$\Delta i3, \Delta i6$	$\Delta i3, \Delta i12$	$\Delta i6, \Delta i12$
			time domain			
$\widehat{\pi}_1$	0.00140	0.00616	0.01599	0.00195	0.00700	0.00286
$\widehat{t}_1$	0.00187	0.00818	0.02064	0.00259	0.00950	0.00382
			frequency domain			
$\widehat{\pi}_1$	0.00141	0.01813	0.04717	0.00616	0.02784	0.00912
$\widehat{t}_1$	0.00187	0.02398	0.06195	0.00817	0.03674	0.01210

The group of tests of the rank of G_z gave then a less clear indication of three independent cointegrating vectors, especially when the frequency domain approach was considered.

In order to obtain additional evidence, we set $r = 3$ and proceeded to the semiparametric estimation of the cointegrating parameters ν_3, ν_6, ν_{12}, and then of the orders of integration of the cointegrating errors γ_3, γ_6, γ_{12}, as defined in (1.92). We already noticed that this does not constitute a formal test but it still provides a further piece of information about the cointegration rank, because if $r = 3$ we should observe in the residuals a sensible reduction in the order of integration.

Since we assumed $\delta_* \leq 1$, OLS may be subject to a rather large bias in small samples, which can be reduced by using NBLS. Marinucci and Robinson (2001) showed by a Monte Carlo exercise that only a very small number of frequencies should be used. Since in our data approximately 80% of the total variation was

concentrated in the first five frequencies, we set $m = 5$ for the computation of the NBLS estimate. We report both groups of estimates in Table 1.6: the correction of NBLS on OLS was very small, as if either δ was rather close to 1 or the correlation between innovations and explanatory variables was small. The estimates were very close to 1, as predicted by the expectations theory, and decreased slightly with the increase of the difference in the maturities, the minimum of 0.945 corresponding to the relation between $i1_t$ and $i12_t$.

Table 1.6: NBLS and OLS estimates of the cointegrating parameters

	series:	$i1, i3$	$i1, i6$	$i1, i12$
$NBLS:$	$\widetilde{\nu}$	1.008	0.998	0.945
$OLS:$	$\overline{\nu}$	1.005	0.990	0.931

We then moved on to estimate the memory parameters of the three series of residuals: this is often of interest in its own right, but in our case it was also important because it could provide yet another piece of evidence that indeed the order of integration was three, as prescribed by the expectations theory and acknowledged in most of the applied analyses on similar data under the restriction of integer integration.

We intended to discuss primarily the NBLS residuals, because the rate of convergence is likely to be faster than for OLS, but in practice the results were extremely similar.

The local Whittle estimates of the orders of integration of the residuals are in Table 1.7. The estimates based on the residuals of the NBLS regression ranged between 0.22 and 0.44, the reduction of the order of integration being larger for the pair $i1_t$, $i3_t$, and smaller for the pair $i1_t$, $i12_t$: if we interpret this result as an indication of a potentially stronger cointegrating relationship between the rates with closer maturities, it may be important to notice that it mirrored the outcome of the cointegration tests. We do not show the estimates

for the OLS-based residuals because the potential rate of convergence should be smaller; the estimates anyway were extremely similar to the other ones.

In the second part of Table 1.7 we present the estimates for the memory parameters of the spreads $(i12-i1)$, $(i6-i1)$, $(i3-i1)$, because according to the expectations theory each pair of interest rates is cointegrated with long term coefficient $\nu_j = 1$ and the series of the spreads should be weakly autocorrelated. The outcome was essentially the same as if NBLS residuals were used instead, therefore supporting the hypothesis that indeed the cointegration rank was $p-1$ and also that $\nu_3 = 1$, $\nu_6 = 1$ and $\nu_{12} = 1$. But the order of integration of the spreads was always larger than that prescribed for it by the expectations theory, and testing

$$H_0 : \{\gamma_{i,j} = 0\} \text{ v.s. } H_1 : \{\gamma_{i,j} > 0\} \tag{1.96}$$

the null hypothesis was always rejected.

Table 1.7: LW estimates of the memory parameters of the cointegrating errors

pairs:	$i3_t, i1_t$	$i6_t, i1_t$	$i12_t, i1_t$
residuals of NBLS regression			
$\widetilde{\gamma}$	0.22	0.30	0.44
spreads (structural assumption)			
$\widetilde{\gamma}$	0.25	0.30	0.42

The high persistence signalled by the larger than zero γ_j may be interpreted as too slow an adjustment of the long rates to current and expected future shocks in the short term rate.

Evidence of failure of the expectations theory is common in the literature, but it was mainly observed through particular reparameterisations of the short run dynamics: the $CI(1,1)$ model with $\nu_j = 1$ was either taken for granted or found to be broadly compatible with the data. We on the contrary found that the expectations theory failed in describing even the long run dynamics,

because according to our estimates the reaction to shocks and the approach to the long term equilibrium is much slower than predicted by the theory.

The rejection of the expectations theory means that long term rates do not anticipate future short term rates precisely.

As Campbell and Shiller (1987) pointed out, the failure to observe some strong implications of the expectations theory does not necessarily mean that the long term maturities are not informative at all, about future short term rates dynamics: a sudden increase in the long term rates for example may be a prelude to a tightening of the monetary policy even if we cannot rely on the expectations theory to quantify the exact extent of the future intervention. We addressed this issue by estimating the complete parametric model and then by analysing the impulse response function.

We first tested for cointegration again, by using the parametric approach proposed by Breitung and Hassler (2002).

Preliminary knowledge of δ_* and of the order k of the VAR representation of $\Delta^{\delta_*} z_t^{\#}$ are necessary. Breitung and Hassler (2002) remarked that it can be replaced by a consistent estimate, so we could simply use the semiparametric estimate $\overline{\delta}_{*,\Delta i} + 1$ we computed before, but, in line with the spirit of the parametric model, we fitted an ARFIMA(2,d,0) to the first differences of each rate, averaged the estimates of each memory parameter and added back 1 (we obtained 0.86, very close to $\overline{\delta}_{*,\Delta i} + 1$, which is what we should expect).

Once again we found rank 2 using the whole vector and the 5% test, but notice that the realisation of the test statistic was very close to the critical value, and rank 3 would have followed had we taken a size of 10%. On the other hand, when testing only pairs of rates, we always found evidence of cointegration, thus pointing at rank 3 in the whole vector. These results are shown in Tables 1.8 and 1.9.

Table 1.8: Breitung and Hassler (2002) cointegration tests: joint test

rank tested:	$r=3$	$r=2$	$r=1$	$r=0$
$\widetilde{\lambda}_1$	0.01	8.84	29.91	49.39
$\sum_{j=1}^{n-r_0}\widetilde{\lambda}_j$	0.01	8.85	38.05	87.45
$\chi^2_{0.95,(n-r_0)^2}$	3.84	9.49	16.92	26.30

Table 1.9: Breitung and Hassler (2002) cointegration tests: pairwise tests

pairs:	$\Delta i1, \Delta i3$	$\Delta i1, \Delta i6$	$\Delta i1, \Delta i12$	$\Delta i3, \Delta i6$	$\Delta i3, \Delta i12$	$\Delta i6, \Delta i12$
$\widetilde{\lambda}_1$	0.02	0.03	0.00	0.19	0.02	0.00
$\widetilde{\lambda}_2$	22.82	28.61	22.36	22.64	17.67	14.61

We imposed a cointegration rank $r = 3$, and moved on to formulate and estimate a VAR(k) for $(u_{1t}, u_{3t}, u_{6t}, u_{12t})'$.

Introduce the notation

$$g = (d_*, c_3, c_6, c_{12}, b_3, b_6, b_{12})' \tag{1.97}$$

$$\xi_t(g) = \begin{bmatrix} \Delta^{d_*}\left[\left(i1_t - \overline{i1}\right) 1(t>0)\right] \\ \Delta^{c_3}\left[\left(i3_t - \overline{i3} - b_3(i1_t - \overline{i1})\right)1(t>0)\right] \\ \Delta^{c_6}\left[\left(i6_t - \overline{i6} - b_6(i1_t - \overline{i1})\right)1(t>0)\right] \\ \Delta^{c_{12}}\left[\left(i12_t - \overline{i12} - b_{12}(i1_t - \overline{i1})\right)1(t>0)\right] \end{bmatrix} \tag{1.98}$$

and

$$\widehat{\varepsilon}_t\left(\widehat{\theta}(g)\right) = \xi_t(g) - \sum_{j=1}^{k} B_j\left(\widehat{\theta}(g)\right)\xi_{t-j}(g) \tag{1.99}$$

where $B_j\left(\widehat{\theta}(g)\right)$ are estimated by regressing $\xi_t(g)$ for k lags.

Letting $\widetilde{g} = \left(\overline{\delta}_{*,\Delta i} + 1, \widetilde{\gamma}_3, \widetilde{\gamma}_6, \widetilde{\gamma}_{12}, \widetilde{\nu}_3, \widetilde{\nu}_6, \widetilde{\nu}_{12}\right)'$, then $\xi_t(\widetilde{g})$ and $\widehat{\varepsilon}_t\left(\widehat{\theta}(\widetilde{g})\right)$ are semiparametric estimates of the series of original innovations u_t and ε_t respectively. To determine the order of the VAR, we first discarded the first three $\xi_t(\widetilde{g})$ (because only a few observations were available to compute the fractional difference of the data, and the effect of the truncation can be very

strong in those situations), and then applied the Schwartz and Hannan and Quinn information criteria to the remaining observations. In both cases a VAR(2) was selected.

This two-step approach can also be used to show that the likelihood is a function of g, so the maximum likelihood estimation of the whole parametric model follows as

$$\widehat{g} = \arg_{g\in\Theta} \min \ln \left| \frac{1}{n-l} \sum_{t=l}^{n} \widehat{\varepsilon}_t \left(\widehat{\theta}(g)\right) \widehat{\varepsilon}_t \left(\widehat{\theta}(g)\right)' \right| \tag{1.100}$$

on a compact set Θ.

The maximum likelihood estimates of the elements of $B(L)$ and of Ω are $\widehat{B}_j = B_j\left(\widehat{\theta}(\widehat{g})\right)$, $\widehat{\Omega} = \frac{1}{n-l}\sum_{t=l}^{n} \widehat{\varepsilon}_t \left(\widehat{\theta}(\widehat{g})\right) \widehat{\varepsilon}_t \left(\widehat{\theta}(\widehat{g})\right)'$.

We presented the maximum likelihood estimates of the long run parameters g in Table 1.10, those of $B(L)$ in Table 1.11 and those of Ω in Table 1.12.

Table 1.10: ML estimates of the long run parameters

$\widehat{\nu}_3$	$\widehat{\nu}_6$	$\widehat{\nu}_{12}$	$\widehat{\gamma}_3$	$\widehat{\gamma}_6$	$\widehat{\gamma}_{12}$	$\widehat{\delta}_*$
1.01	1.01	0.98	0.19	0.21	0.25	0.86

Table 1.11: ML estimate of the autoregressive parameters

$\widehat{B}_1$				$\widehat{B}_2$			
0.52	1.08	-0.57	0.41	-0.17	0.15	-0.95	0.45
0.00	-0.72	0.36	0.26	0.02	-0.58	0.73	-0.38
-0.06	-1.35	0.51	0.64	0.06	-0.19	0.45	-0.40
-0.09	-1.51	0.18	1.10	0.06	0.07	0.18	-0.30

Table 1.12: ML estimate of the covariance matrix of the innovations

$10000\widehat{\Omega}$				Correlation structure			
0.025	0.03	0.04	-0.04	1	0.90	0.78	-0.34
0.03	0.06	0.08	-0.07	0.90	1	0.93	-0.42
0.04	0.08	0.11	-0.11	0.78	0.93	1	-0.50
-0.04	-0.07	-0.11	0.44	-0.34	-0.42	-0.50	1

The estimates of the long run parameters resembled those from the semi-parametric analysis: $\widehat{\nu}_{12}$ and $\widehat{\gamma}_6$ were slightly closer to the parameters prescribed from the expectations theory, and a larger correction took place for $\widehat{\gamma}_{12}$. The estimated cointegrating parameters were very close to 1, indeed even closer than the NBLS estimates, and the effect of cointegration was quite relevant, the gaps $\widehat{\delta}_* - \widehat{\gamma}_3$, ..., $\widehat{\delta}_* - \widehat{\gamma}_{12}$ ranging between 0.61 and 0.67. Since Robinson and Hualde (2003) showed that when $\delta_* - \gamma_j > 1/2$ a simple two-step GLS estimate would be as efficient as maximum likelihood, we first tested

$$\begin{aligned} &H_0 : \{\delta_* - \gamma_3 = 0.49\} \text{ v.s. } H_1 : \{\delta_* - \gamma_3 \neq 0.49\}, \\ &H_0 : \{\delta_* - \gamma_6 = 0.49\} \text{ v.s. } H_1 : \{\delta_* - \gamma_6 \neq 0.49\}, \\ &H_0 : \{\delta_* - \gamma_{12} = 0.49\} \text{ v.s. } H_1 : \{\delta_* - \gamma_{12} \neq 0.49\} \end{aligned} \tag{1.101}$$

with three likelihood ratio tests. Each test statistic has a χ^2_1 distribution (under the null), and the realisations were 0.93, 0.68 and 0.38 respectively: despite the fact that the point estimates yielded gaps $\delta_* - \gamma_3$, ..., $\delta_* - \gamma_{12}$ larger than 1/2, the differences were then not statistically significant, so with these data the simultaneous estimation might be safer.

We next tested the structural hypothesis that all the cointegrating parameters were 1, as it is assumed in the expectations theory:

$$H_0 : \{\nu_3 = 1, \nu_6 = 1, \nu_{12} = 1\} \text{ v.s. } H_1 : \{\nu_3 \neq 1 \text{ \&/or } \nu_6 = 1 \text{ \&/or } \nu_{12} = 1\}, \tag{1.102}$$

The corresponding test statistic took the value 7.5, just below the critical value 7.8, so the hypothesis was not rejected. The two other relevant hypothesis

concerned the order of integration of the data: the expectations theory required

$$H_0 : \{\gamma_3 = 0,\ \gamma_6 = 0,\ \gamma_{12} = 0\},\ \text{v.s.}\ H_1 : \{\gamma_3 \neq 0\ \&/\text{or}\ \gamma_6 = 0\ \&/\text{or}\ \gamma_{12} = 0\}, \tag{1.103}$$

but the hypothesis was rejected (the test statistic took the value 8.3); on the other hand, we again failed to provide any convincing statistical evidence against the hypothesis of a unit root:

$$H_0 : \{\delta_* = 1\}\ \text{v.s.}\ H_1 : \{\delta_* \neq 1\} \tag{1.104}$$

had a realised test statistic of 1.7.

As in the semiparametric analysis, the estimates of the orders of the residuals maintained the property that the closer the maturities were, the faster was the adjustment, but this feature was not statistically significant: testing

$$H_0 : \{\gamma_3 = \gamma_6 = \gamma_{12}\}\ \text{v.s.}\ H_1 : \{\gamma_3 \neq \gamma_6\ \&/\text{or}\ \gamma_6 \neq \gamma_{12}\}, \tag{1.105}$$

the computed statistic was 1.3, far below the critical value of 6.0: the estimate of the memory parameter of the residuals under this restriction was 0.2.

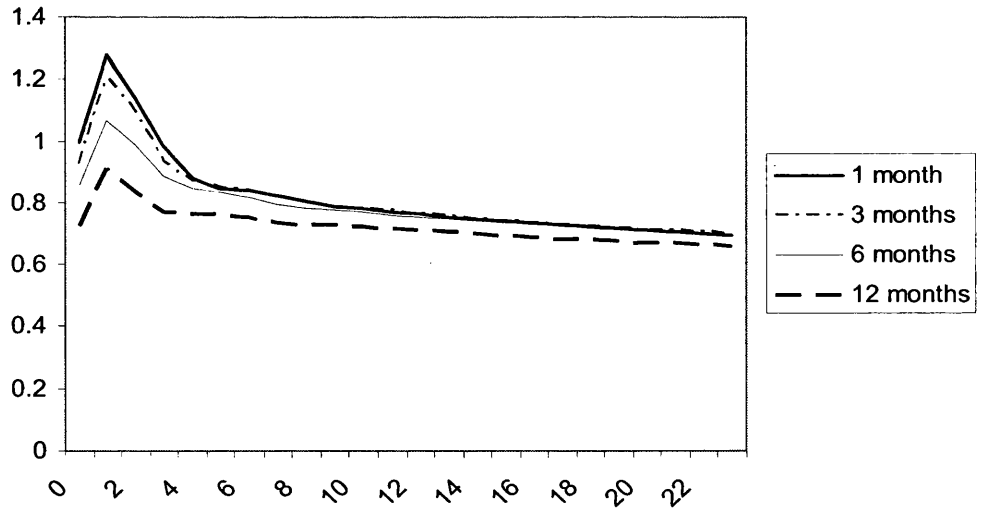
We analysed the short term dynamics via a structuralised impulse response function.

For the structural identification of contemporaneous shocks, we assumed that the contemporaneous correlation moved from the shortest to the longest maturity. We already noticed that the central bank operates by supplying liquidity on the interbank market, with the aim of controlling a very short term rate (indeed, the overnight has maturity of only one day): with our assumption, an innovation to the rate with the shortest maturity is interpreted as driven by monetary policy decisions, whereas innovations to rates with longer maturities

may depend on other factors as well, there including the possibility that long term rates do still roughly anticipate the movements of future short term ones even if with less precision than the expectations theory hypothesis prescribes.

We plotted in Figure 1.2 the estimated reaction of the three rates with respect to an innovation to the short term rate. The estimated effect of a 100 basis points (b.p.) innovation was temporary: the peak was reached after two periods and then the plot reverted to 0, yet after 24 periods the one month rate was still estimated to be 69 b.p. above the starting value, so the reversion to 0 was indeed very slow.

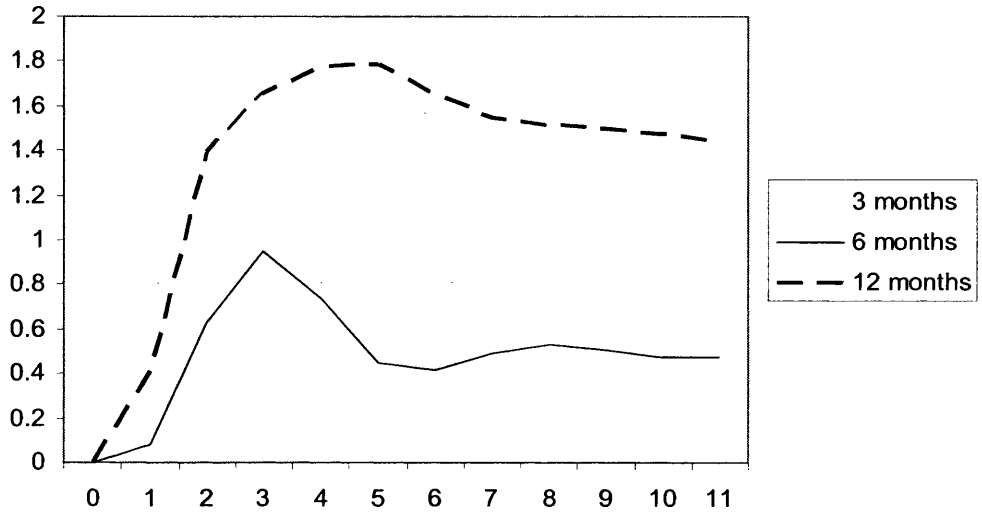
Figure 1.2: Reaction of the rates to a 100 b.p. increase in the 1 month rate



The estimated contemporaneous reaction of the longer rates was very strong and indeed nearly one to one: a 100 b.p. increase in the short rate was met by a 93, 86 and 72 b.p. increase in the 3, 6, 12 months rates respectively. They also followed the 1 month rate in the subsequent periods, maintaining the characteristic that the 3 months rate was the closest one to the 1 month, the 12 month the furthest away. Overall though these differences among the plots were very little: the whole term structure drifted, peaked and then reverted to

the mean in quite the same way.

Figure 1.3: Reaction of 1 month rate to a 100 b.p. increase in the longer rates



In Figure 1.3 we plotted the reaction of the 1 month rates to a 100 b.p. innovation in the other three rates: the three months rate anticipated the future dynamics better in the short run, but in the medium and long run the informative content of the long term rate was clearly superior.

1.4.3 Conclusions

Fractional integration and cointegration allow a more flexible description of the characteristics of an economic time series. We considered an example in which by restricting the attention to $I(1)$ and $I(0)$ models only we would implicitly exclude some properties that are required by the economic theory. Fractional models made it possible to reconcile the apparently alternative theories. We presented a multivariate model for the US interest rates for different maturities, to study the implications of fractional integration and cointegration on the expectations theory for interest rates and on the transmission mechanism of monetary policy.

The first, semiparametric, analysis was sufficient to rule out the expectation theory, because the high persistence of the residuals implies a reaction much slower than what the theory predicted. But it still provided evidence in favour of the existence of a long run relation as required for the transmission of monetary policy, possibly taking the form of the interest rate spread. The fully parametric analysis confirmed this conclusion, and indicated that the spreads were informative with respect to the future rates. Evidence of fractional integration of the data, obtained in the semiparametric analysis, was confirmed by the parametric specification: point estimates indicated a slow mean-reverting dynamic for the interest rates, although we were not able to reject the hypothesis of an $I(1)$ process.

1.5 Estimation when the process is contaminated by unobserved deterministic components or subject to breaks in the stochastic ones.

1.5.1 Memory estimation in the presence of deterministic components of various kinds

We now consider a time series x_t, observed at times $t = 1, ..., n$, which is composed of two unobservable parts: a deterministic sequence s_t and a zero-mean stochastic process ξ_t

$$x_t = \xi_t + s_t. \tag{1.106}$$

In many economic time series it seems that the deterministic component changes over time: it could include a time trend, for example, or a mean subject to a break. Modelling these terms may sometimes be difficult: in some applications a linear time trend can be confused with a shift in the mean, or

the location of a certain break can be disputed; more often some features may be neglected altogether.

The practical consequences of incorrectly modelling the deterministic component were greatly exemplified by Perron (1991). He considered the same thirteen time series for which Nelson and Plosser (1982) did not reject the hypothesis of a unit root using a Dickey-Fuller test, and showed that the conclusion could be reversed in ten of them if a break in the intercept or in the slope was allowed for in 1929 (in terms of (1.106), Perron considered for x_t the residuals of a regression of the data on a linear trend).

He also showed that the estimate of the autoregressive coefficient in a Dickey-Fuller type of regression with $I(0)$ observables can be inconsistent if the deterministic component is not correctly specified. When a shift in the intercept is not accounted for, the true value of the autoregressive coefficient is overestimated and the limit distribution of the Dickey-Fuller test statistic is different from the one specified by the asymptotic theory; when the trend is modelled in an incorrect way, the estimate of the autoregressive coefficient converges to 1, then giving spurious evidence of a unit root.

This can lead to the application of inappropriate limit theory, and it can also have important implications for economic interpretation of the results, for example because the spurious strong autocorrelation could be regarded as a slow response to shocks by the policymaker or by the agents.

The two deterministic components discussed by Perron (1991) may be associated with a dimension based on the Euclidean norm, and the different results may be classified according to that dimension. It is of course a rather coarse classification, because only two deterministic components and two orders of integration are considered, but it is worth noticing that the spurious evidence of a unit root occurs when the order of magnitude of s_t is bigger than the order of ξ_t, and that only the limit distribution is affected otherwise.

Perron's empirical findings were mitigated by several authors, who argued that the choice of the point of the break was driven by the data, and the critical value should have been modified to take that endogeneity into account.

Allowing for a potential random break in the intercept, Zivot and Andrews (1992) reversed Perron's results (so again they failed to reject the hypothesis of a unit root) in five cases even with critical values computed without breaks under the null, and argued that even more reversions would occur with a critical value generated by introducing a break under the null. But this does not make the example of Perron less important: Zivot and Andrews' remark simply means that Perron was exposed to the same criticism he raised, because the way in which he proposed the evidence depended on the model as well, the Dickey-Fuller test requiring a specification of the deterministic component.

We referred to the examples of Perron (1991) because of their popularity, but we generalise the integer powers and the unit root to a fractional set-up,

$$\xi_t \in I(\delta), \quad \delta \in (-1/2, 1/2), \tag{1.107}$$

$$s_t = \mu t^{\phi - 1/2}, \ 0 < \phi < 1/2, \tag{1.108}$$

for some finite, nonzero μ. Undoubtedly the trends of interest in practical applications have integer powers, but these are only special cases of (1.108) if $\phi = 1/2$, $\phi = 3/2$,... (these are not in (1.108) but we nonetheless refer to them in Chapter 2). By using fractional powers on the other hand we can provide a much more refined classification. Moreover, trends with non-integer powers are not necessarily unrealistic: if for example d (fractional) differences are taken from a time series with a linear trend, the resulting time series has a time trend with fractional power $1 - d$.

Bhattacharya, Gupta and Waymire (1983) showed that, assuming (1.106)

- (1.108), the R/S statistic computed using x_t indicates the presence of the Hurst effect even when $\delta = 0$, the determinist trend being then mistaken for a stochastic one of order $\delta = \phi$. Teverovsky and Taqqu (1997) considered not only the fractional trend (1.108) but also the break in the mean: they found that both the deterministic components induce spurious evidence of fractional integration in a variance-type estimate as well, although they showed that the effect can be removed by differencing the variance (they also acknowledged that the original estimate is rather imprecise, though, and that the robust one is even worse). Giraitis, Kokoszka and Leipus (2001) generalised the class of deterministic components for which the R/S and the V/S statistics do not detect spurious evidence of the Hurst effect to

$$\limsup_{t=1,\ldots,n} \left(\sum_{t=1}^{n} (s_t)^2 \right)^{1/2} < C < \infty. \tag{1.109}$$

This spurious evidence depended partially on the adoption of R/S-type statistics: with $|s_t| \leq Ct^{\phi-1/2}$ Heyde and Dai (1996) claimed that the asymptotic distribution of the Whittle estimate in a parametric model for data having $\delta = 0$ is not affected by the time trend if $\phi < 1/4$, which means that the Whittle estimation can distinguish better between a stochastic and a deterministic component. For a larger δ, they suggested the condition

$$\phi < \min(1/4, 1/2 - \delta). \tag{1.110}$$

Setting s_t to be a break in the mean, several Monte Carlo exercises were proposed to illustrate the conjecture that it too may induce spurious evidence of long memory even though $\delta = 0$. Indeed, as Lobato and Savin (1988) showed, the autocorrelation function of a time series subject to a neglected shift in the mean does not approach 0 at all, so it is certainly not summable.

Breaks in the mean were also discussed by Granger and Hyung (2004) and Diebold and Inoue (2001), and an introductory discussion of the corresponding periodogram is in Mikosch and Stărică (1999).

Having seen that often a deterministic component cannot be neglected without consequences, detecting its presence is very important. The difference between the two estimates of Teverovsky and Taqqu (1997) can at most be a preliminary indication, and actually they did not even provide a limit distribution for it. Hidalgo and Robinson (1996) addressed the detection of the shift in the mean at least when the location of the break is known. They proposed a version of the Chow test that is robust to strong autocorrelation: the test is semiparametric in the sense that it does not require specification of the short memory dynamics of the disturbance process, but it is very model-specific because it is only designed for one particular type of deterministic component, and it even requires knowledge of the location of the break.

In Chapter 2 we investigate the consequences of applying the local Whittle estimation procedure to a stochastic process which is contaminated by deterministic terms of various kinds. We also show that, by modifying the loss function, the estimate can be made robust to the presence of a much wider class of deterministic components. We then propose a test to detect the presence of deterministic components that may affect the properties of the estimates.

1.5.2 Memory estimation in the presence of a structural break in the stochastic component

As Lucas (1976) remarked, the assumption that the data generating process remains stable over time is often questionable: the economy is subject to structural shifts and to changes in policy regimes that may alter the dynamics of the target variable or the structure of a macroeconomic relationship.

In Chapter 2 we discuss the effects of changes over time of the deterministic component s_t in a univariate process (1.106); in Chapter 3 we consider instability in the zero-mean, stochastic part of (1.106), ξ_t. We then consider the model $\xi_t = \xi_{1,t}1(t \leq \tau n) + \xi_{2,t}1(t > \tau n)$, $\tau \in [0,1]$: because of the break, ξ_t is no longer stationary, but both $\xi_{1,t}$ and $\xi_{2,t}$ may be so, and indeed in Chapter 3 we assume that they are stationary and fractionally integrated. We distinguish between changes in the long term dynamics, that we associate with the lowest frequencies and summarise with the order of integration δ, and changes in the short term dynamics, that we associate with the remaining frequencies.

Changes in δ are often important in policy evaluation because *ceteris paribus* they indicate a tighter (when δ decreases, provided that $\delta < 1$ after the change) or weaker (otherwise) control of the variable of interest, the return to the targeted mean being faster the lower δ.

Potential changes in persistence and in long term dynamics had often been considered in the applied literature, but the evidence is largely anecdotal and restricted to integer δ only.

A formal approach was proposed by Kim (2000) who introduced a ratio-based statistic to test the null that $\delta = 0$ in the whole process against the alternative that a shift between $\delta = 0$ and $\delta = 1$ took place. Kim, Belaire-Franch and Badilli-Amador (2002) and Busetti and Taylor (2004) proposed some corrections and further developments, but they did not alter the original structure. Harvey, Leybourne and Taylor (2004) remarked that with that design the case $\delta = 1$ and no breaks can be confused with the presence of a break, so those tests are not very informative. Harvey *et al.* (2004) then proposed a modification of the test statistic to make it such that the critical values would be the same (although the limit distribution would still be different) regardless of whether $\delta = 0$ or $\delta = 1$.

We have already argued that the restriction to integer δ seems too strong because it leaves no alternative between fast reversion to the mean and no reversion at all: this is much more the case when a potential change in integer δ is discussed, because with that restriction it is implicitly assumed that the process jumps between the two extreme situations avoiding all the intermediate ones, while important variations in the long term dynamics may be represented with relatively small changes in (fractional) δ.

A second drawback, more specifically related to the approach introduced by Kim (2000), is the sensitivity of the test to instability in the short term dynamics of the process: even simple changes like the shift of the variance in an independent sequence can be detected by the test and confused with a change in δ. This seems quite unappealing, because it requires an assumption, the stability of the short term dynamics, that is not directly related to the object of the analysis, and it is particularly unfortunate in this case because in practice it is at least doubtful that such an assumption can be imposed when the long term stability is being tested. Indeed, in applied work the reverse is more often assumed: Kim and Nelson (1999) for example discussed the change in volatility of the GDP, while Hansen (2003) found instability in the short term dynamics of the interest rates. Neither the Dickey-Fuller nor similar tests can provide a reliable indication in this situation because, as Hamori and Tokihisa (1997) showed in the particular case of a volatility shift, it may be sensitive to short term instability.

In Chapter 3 we propose to address these two issues simultaneously, using the local Whittle estimation procedure. This removes the constraint to integer δ by allowing for a fractional model for the order of integration, and it also avoids the sensitivity to the short term dynamics because it only uses a band of frequencies degenerating towards 0.

1.5.3 Cointegration in presence of deterministic trends

In Section 1.2 we introduced fractional cointegration as the application of the concept of cointegration to fractionally integrated processes as well. Yet, as we already noticed, economic time series are rarely zero-mean (or constant-mean) stochastic processes, and they are more often characterised by a time-varying deterministic component as well.

The large $I(1)/I(0)$ literature focuses on many features of economic time series, in particular recognising empirical evidence that the stochastic unit root trend frequently needs to be supplemented by a deterministic trend, such as one increasing linearly with time (see e.g. West (1988), Stock and Watson (1988), Park and Phillips (1988), Johansen (1991), Hansen (1992), Perron and Campbell (1993), Campos, Ericsson and Hendry (1996)). For empirical applications on the analysis of demand for money, see Hoffman and Rasche (1991), Stock and Watson (1993), and on the PPP/UIP relations see Johansen and Juselius (1992). A review of applications of cointegrated models with a background in economic theory is in Söderlind and Vredin (1996), and a particular treatment of cotrending, deterministic and stochastic cointegration is in Ogaki and Park (1997), who modelled the allocation of income in consumption of durable and non-durable goods.

On the other hand, the fractional cointegration literature has mostly not allowed for deterministic trends. An exception is the discussion in Robinson and Marinucci (2000) of the properties of OLS and NBLS, but they only considered a particular combination of deterministic and stochastic trends.

In Chapter 4 we develop properties of the OLS and GLS estimates of the cointegrating coefficient in a bivariate model that either ignores or takes account of additive deterministic trends.

A model of fractional integration and cointegration with fractional deter-

ministic trends is an important extension of the standard $I(0)/I(1)$ and linear trend case that is more often discussed: we can give a much more precise classification of the conditions under which the stochastic or the deterministic component define the properties of the estimate.

A cointegration model is quite a change of perspective with respect to the analysis we run in Chapter 2. In the estimation of the memory parameter we treat the deterministic component as a nuisance that may obscure the signal originating from the zero-mean stochastic term, so we only deal with it because we suspect we are unable to model it properly, in order to filter it and remove it from the data. In the cointegration framework, the deterministic component is part of the model: it may contribute, for example, to determinate the long term dynamics of the explanatory or of the dependent variables, and it may even increase the rate of convergence of the estimate of the cointegrating parameter, so it should not be removed from the data, even if we have precise knowledge of its structure.

Chapter 2

Local Whittle estimation of the memory parameter in presence of a deterministic component

2.1 Introduction

In Chapter 1 we presented several methods for the estimation of the memory parameter of a (constant-mean) fractionally integrated process when the process itself is observable.

We also noticed though that often the economic time series cannot be modelled as a process with constant mean. Since in several empirical cases there is little agreement on the nature of the deterministic component, we motivated our interest in the problem with the conjecture, rather widespread in the applied literature, that neglecting it or modelling it in an incorrect way may compromise the limit distribution of the estimate of the memory parameter, or even its consistency.

In this chapter we define the class of deterministic components which can be safely neglected or misspecified for the local Whittle estimate, and propose

a test to detect the presence of relevant deterministic terms: although these two purposes have already been addressed before (but for different estimates), we propose to do it simultaneously. Key to our approach is the computation of the periodogram of the deterministic component of interest at the relevant Fourier frequencies, which gives us the possibility to exploit the differences between that periodogram and the spectral density of a stochastic process. Frequency domain estimates can then be made robust to even more potential deterministic terms: we show, for example, that the estimate may be robust even to a break in the mean.

We also generalise the previous studies in two other ways: we explicitly discuss the break in the mean by computing its periodogram and showing that it can be treated as a particular fractional trend, and we allow for a wider range of deterministic trends.

Since we intend to propose an automatic testing procedure that could be considered as part of the preliminary analysis of the data, a semiparametric estimate has the advantage of not requiring the specification of the short term dynamics: this makes the test robust and fast to implement. Although some theoretical work has already been done for R/S-type statistics, we prefer to consider a different class of estimates: R/S-type procedures are rather *ad hoc* and the estimates are characterised by a nonstandard limit distribution, whereas other estimates, like the log-periodogram regression or the local Whittle ones, are very intuitive and their limit distributions are asymptotically normal and parameter free, a great advantage if we are also interested in designing a test that is fast and easy to implement. Moreover, on the basis of other published works, we anticipate that these frequency domain based estimates are less prone than R/S-type statistics to be affected by neglected deterministic components. We choose the local Whittle estimate for its smaller variance, but we think that the results derived here are also of interest because

they are a reliable anticipation of the properties of the log-periodogram regression estimate, and indeed also for the full spectrum Whittle estimate in the same situation.

In Section 2.2 we present the asymptotic theory, in Section 2.3 we analyse the small sample properties with a Monte Carlo exercise and in Section 2.4 we propose two empirical studies: the S&P500 and three inflation rates. We conclude in Section 2.5, summarizing the results and discussing some potential extensions. The proofs of the theorems are in the Appendix.

2.2 Local Whittle estimation with neglected deterministic terms

We consider a process x_t observed at times $t = 1, ..., n$ such that

$$x_t = \xi_t + s_t, \tag{2.1}$$

where s_t is a deterministic sequence and ξ_t is a stochastic process, that we assume to be zero-mean, weakly stationary and invertible, with spectral density $f_\xi(\lambda)$ such that

$$f_\xi(\lambda) \sim G\,|\lambda|^{-2\delta} \text{ as } \lambda \to 0, \tag{2.2}$$

where $G > 0$. Notice that stationarity requires $\delta < 1/2$ and invertibility $\delta > -1/2$.

We indicate by $F_x(\lambda)$, $F_\xi(\lambda)$, $F_s(\lambda)$ the discrete Fourier transforms of x_t, ξ_t and s_t respectively, and by $I_x(\lambda)$, $I_\xi(\lambda)$, $I_s(\lambda)$ the corresponding periodograms, and by $I_{s\xi}(\lambda)$ the crossperiodogram between s_t and ξ_t.

2.2.1 The periodograms of the stochastic and of the deterministic components

Since the local Whittle loss function is a weighted average of periodograms, the asymptotic properties of the estimate depend on whether enough elements in the summation are dominated by the stochastic rather than by the deterministic component.

To appreciate the different contributions, we analyse the order of magnitude of the periodograms I_ξ and I_s at the Fourier frequencies used in the local Whittle estimation.

We consider three models for the deterministic component: the shift in the mean, the deterministic trend and the single impulse (we also refer to them as $s(\mu)$, $s\left(t^{\phi-1/2}\right)$ and $\Delta s(\mu)$ respectively in the rest of the thesis). These are defined as:

shift in the mean:

$$s(\mu) = \begin{cases} s_t = \mu_1 \text{ for } t \leq [\tau n] \\ \\ s_t = \mu_2 \text{ for } t > [\tau n] \end{cases} \tag{2.3}$$

where $\tau \in (0,1)$, $|\mu_1| < \infty$, $|\mu_2| < \infty$ *and* $\mu_1 \neq \mu_2$;

deterministic trend:

$$s\left(t^{\phi-1/2}\right) \sim \mu_3 t^{\phi-1/2} \text{ as } t \to \infty, \tag{2.4}$$

where $0 < |\mu_3| < \infty$, $0 < \phi < \infty$;

single impulse:

$$\Delta s\left(\mu\right) = \begin{cases} s_t = \mu_4 \text{ for } t = [\tau n] \\ \\ s_t = 0 \text{ for } t \neq [\tau n] \end{cases} \tag{2.5}$$

where $0 < |\mu_4| < \infty$.

Shifts in the mean as in (2.3) are often considered in applied analysis, while deterministic trends as in (2.4) are important to provide a general classification and to compare our results with the rest of the literature (notice that the trends may well have non-integer powers); single impulses as in (2.5) have had less theoretical and empirical importance, but we consider them explicitly nevertheless because this structure emerges when first differences of a shift in the mean are taken, a procedure that is very common when $\delta > 1/2$.

The periodogram of a deterministic fractional trend was first discussed by Künsch (1986), who also advocated trimming to remove the potential effects of that term on the estimate; a more general discussion is in Robinson and Marinucci (2000). A reference to the exact order of magnitude of the periodogram of the shift in the mean is in Mikosch and Stărică (1999) eq. (3.4) - (3.6), although they did not provide a proof and required the condition $n\lambda_j^2 \to 0$, which at the Fourier frequencies corresponds to $j^2/n \to 0$. Notice that this condition, if necessary, would reduce the frequencies available for the computation of the loss function: Robinson (1995b) showed that, when the other regularity conditions are met, the local Whittle estimate is consistent when $m/n \to 0$, where m is the largest frequency used in the loss function, but $j^2/n \to 0$ for all $j < m$ only holds for the stronger condition $m^2/n \to 0$.

We summarise these results and fill in a few gaps for the three models of interest (we recall that K is a positive, finite constant, not necessarily the same).

Theorem 2.1. *(i) Shift in the mean. If* $s_t \in s(\mu)$, *then*

$$\left| \sum_{t=1}^{v} e^{i\lambda t} s_t \right| \leq C |\lambda|^{-1} \text{ for } v \geq 1,\ 0 < |\lambda| < \pi \tag{2.6}$$

and, for $j > 0$,

$$I_s(\lambda_j) \leq C |\lambda_j|^{-1} j^{-1} \tag{2.7}$$

and

$$n^{-1} I_s(\lambda_j) \sim K j^{-2} \sin^2 \tau\pi j \text{ as } j/n \to 0; \tag{2.8}$$

(ii) Fractional trend. If $s_t \in s\left(t^{\phi-1/2}\right)$, $\phi \in (-1/2, 1/2)$ *then, for* $j > 0$,

$$I_s(\lambda_j) \leq C |\lambda_j|^{-2\phi} j^{-1} \tag{2.9}$$

and

$$n^{-2\phi} I_s(\lambda_j) \sim K j^{-2\phi-1} \text{ as } j/n \to 0; \tag{2.10}$$

if $\phi = 1/2$ *then, for* $j > 0$,

$$I_s(\lambda_j) = 0; \tag{2.11}$$

if $\phi \in (1/2, 3/2)$ *then, for* $j > 0$,

$$I_s(\lambda_j) \leq C |\lambda_j|^{-2\phi} j^{-2(1-\phi)} \tag{2.12}$$

and

$$n^{-2\phi} I_s(\lambda_j) \sim K j^{-2} \text{ as } j/n \to 0. \tag{2.13}$$

(iii) Single impulse. If $s_t \in \Delta s(\mu)$, *then*

$$n I_s(\lambda) = K. \tag{2.14}$$

The periodograms of these deterministic components can then be indexed

by ϕ, breaks in the mean and single impulses having $\phi = 1/2$ and $\phi = -1/2$ respectively.

More general situations are implicitly dealt with in Theorem 2.1: polynomial trends with different orders, for example, or mixed situations with trends and breaks. In general, the order of magnitude is only determined by the largest ϕ, and trends of lower orders can be ignored in the analysis.

The periodograms of the deterministic components are similar to the spectral density of a long memory stochastic process since they too have a pole at frequency 0. Notice, however, that because of the damping factor j^{-1} (or $j^{-2(1-\phi)}$ if $\phi \geq 1/2$), they do not meet the condition (2.2) for any δ so they cannot be confused with the spectrum of a fractionally integrated time series.

In order to compare these periodograms with that of ξ_t, we recall that, as we already mentioned in Chapter 1, Robinson (1995a) showed that although $I_\xi(\lambda_j)$ is asymptotically a biased estimate of the spectral density, the bias can be bounded and it becomes less and less relevant the more distant λ_j is from $\lambda = 0$, and the average of the upper bound of the bias becomes negligible when enough Fourier frequencies are used.

Loosely speaking, then, the comparison of the (possibly stochastic) orders of magnitude of the two periodograms is a comparison between $(j/n)^{-2\delta}$ for ξ_t and $(j/n)^{-2\phi} j^{-1}$ for s_t (or $(j/n)^{-2\phi} j^{2\phi-2}$ for $\phi > 1/2$).

When $\phi < \delta$ the order of magnitude of the periodogram of the deterministic component is clearly smaller. Yet even for some $\phi \geq \delta$ consistent and root-m, zero-mean, asymptotically normal estimation of δ is still possible, because the damping factor j^{-1} may be enough to make the periodogram $I_s(\lambda_j)$ irrelevant: for φ such that $2(\phi - \delta) \,/\, [2(\phi - \delta) + 1] < \varphi < 1$, the spectral density (2.2) still dominates for the frequencies having $j > c_\varphi n^\varphi$ for some positive, finite c_φ (if $\phi \leq 1/2$, the condition is slightly different otherwise; notice that c_φ can be arbitrarily close to 0, and of course it cannot be too large because $j \leq n/2$

must still hold).

In comparing the orders of magnitude it is also possible to see that trimming may improve the quality of the estimate, because most of the power of the periodograms of the given deterministic components is concentrated in the few lowest frequencies, exactly those that are going to be removed.

To summarise, if we regard the local Whittle estimation of δ as the extraction of the signal from a "dirty" time series, as indeed (2.1) might suggest, it is clear that s_t is a very peculiar type of contamination, different from a weakly dependent and indeed even from a fractionally integrated "noise". The treatment is then different as well, because in case of a stochastic contamination the highest frequencies should be trimmed, while with an unobserved time-varying deterministic component the strategy is reversed.

2.2.2 Robust estimation of the memory parameter

The local Whittle estimate $\widehat{\delta}$ is obtained by minimising, with respect to $d \in [\Delta_1, \Delta_2] \subset (-1/2, 1/2)$, the expression

$$R(d) = \ln \left\{ \frac{1}{m-l+1} \sum_{j=l}^{m} \lambda_j^{2d} I_x(\lambda_j) \right\} - 2d \frac{1}{m-l+1} \sum_{j=l}^{m} \ln(\lambda_j). \qquad (2.15)$$

This is a slight generalisation of the function originally considered by Robinson (1995b), who set $l = 1$: when $l > 1$, one or more of the lowest frequencies are trimmed.

The loss function (2.15) was considered also by Giraitis and Robinson (2003), although for a different purpose, because they were interested in deriving an Edgeworth expansion for $\widehat{\delta}$ and trimming was only required when the tapered, rather than the raw, periodograms were used. Giraitis and Robinson, though, were only interested in the bias generated by the low frequency

approximation of the spectral density $f_\xi(\lambda) \sim G\lambda^{-2\delta}(1 + O(\lambda^{\beta_\xi}))$ for some $\beta_\xi \in (0, 2]$. That bias is generated at the highest available frequencies, where λ^{β_ξ} is larger: indeed, while $1/m + m/n \to 0$ was sufficient for Robinson (1995b) to show consistency of the local Whittle estimate of δ, the stronger $1/m + m^{1+2\beta_\xi} \ln^2 m/n^{2\beta_\xi} \to 0$ was required to make the bias due to λ^{β_ξ} small enough to obtain root-m consistency as well.

We discuss consistency of $\widehat{\delta}$ in Theorem 2.2 and limit distribution in Theorem 2.3 for some cases in which the deterministic component is not a simple constant.

To prove consistency, we introduce the following assumptions.

Assumption A.1. *Let $m = c_\kappa n^\kappa$, where $c_\kappa \in (0, \infty)$.*

Assumption A.2. *Let $l = c_\upsilon n^\upsilon$, where $c_\upsilon \in (0, \infty)$.*

Assumption A.3. *The deterministic component s_t is such that, for $j > 0$,*

$$\left(|\lambda_j|^{-2\phi} j^{-1}\right)^{-1} I_s(\lambda_j) \sim c_1 + c_2 \sin^2 \tau\pi j \text{ as } j/n \to 0 \tag{2.16}$$

where $0 \leq c_1 < \infty$, $0 \leq c_2 < \infty$, $c_1 + c_2 > 0$, $\tau \in (0, 1)$.

Assumption 2.1. *As $\lambda \to 0^+$,*

$$f_\xi(\lambda) \sim G\lambda^{-2\delta} \tag{2.17}$$

where $G \in (0, \infty)$ and $\delta \in [\Delta_1, \Delta_2] \subset (-1/2, 1/2)$.

Assumption 2.2. *In a neighbourhood $(0, \iota)$ of the origin, $f_\xi(\lambda)$ is differentiable and*

$$\frac{d}{d\lambda} \ln f_\xi(\lambda) = O(\lambda^{-1}) \text{ as } \lambda \to 0^+. \tag{2.18}$$

Assumption 2.3. *The sequence* ξ_t *is such that*

$$\xi_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \ \sum_{j=0}^{\infty} \alpha_j^2 < \infty \tag{2.19}$$

where

$$E(\varepsilon_t | F_{t-1}) = 0, \ E(\varepsilon_t^2 | F_{t-1}) = 1, \ a.s., \ t = 0, \pm 1, ... \tag{2.20}$$

in which F_t *is the* $\sigma-$*field generated by* ε_s, $s \leq t$, *and there exists a random variable* ϵ *such that* $E(\epsilon) < \infty$ *and for all* $\eta > 0$ *and some* $C > 0$, $P(|\varepsilon_t| > \eta) \leq CP(|\epsilon| > \eta)$.

Assumption 2.4. *Assumptions A.1 and A.2 hold and*

$$0 \leq \upsilon < \kappa < 1. \tag{2.21}$$

Assumption 2.5. *Assumptions A.3, 2.1 and 2.4 hold and*

$$\phi < \delta + \frac{1}{2} \frac{\kappa}{1 - \upsilon}. \tag{2.22}$$

We use a different notation to distinguish between Assumptions A.1 to A.3 and Assumptions 2.1 to 2.5 because those in the first group define some characteristics of the model (A.3) or of the loss function (A.1 and A.2), and are to remain unchanged both in the proof of consistency of $\widehat{\delta}$, and in the derivation of its limit distribution or of its lower order bias, whereas those in the second group are modified according to the problem. Assumption A.1 and A.2 define the bandwidth m and the trimming point l as proportional to n^{κ} and to n^{υ} respectively: notice that both c_{κ} and c_{υ} may be arbitrarily close to 0 but in practice a relatively large c_{κ} may be preferred in order to minimise the MSE of the estimate (see Henry and Robinson (1996) for a more detailed discussion on the choice of c_{κ}).

Assumptions 2.1, 2.2 and 2.3 were introduced by Robinson (1995b) to characterise the stochastic component in his original work and are discussed therein. These are semiparametric in the sense that Assumptions 2.1 and 2.2 are only defined for $\lambda \to 0^+$. Assumption 2.3 is a very general specification for the stochastic component: the linear structure is ensured by the Wold representation theorem so the assumption is only about the second moment of the martingale difference sequence of innovations ε_t.

We have modified Assumption 2.4 of Robinson (1995b) slightly: consistency would follow for any l, m with

$$l/m + m/n \to 0, \tag{2.23}$$

but with Assumptions A.1 and A.2 we restricted m and l to be proportional to n^κ and n^υ respectively, because it allows a simple computation of the orders of magnitude of the weighted averages of $I_s(\lambda_j)$ and of $I_{s\xi}(\lambda_j)$. We think that this is only a very mild restriction, because it still leaves a wide range of rates of divergence for m and l, and we also justify it by noticing that in applied works the bandwidth is often chosen according to this practice anyway.

We characterise the deterministic component in Assumption A.3 and 2.5. It is based on the approximations computed in Theorem 2.1, although it does not actually require knowledge of s_t, but only of the order of magnitude of its periodogram. As we saw, this is more general because it can be generated also by other deterministic components not considered in Theorem 2.1; it is also "semiparametric" in the sense that it does not require knowledge of the location of the break, if we included that case in s_t. On the other hand, Assumption 2.5 is apparently rather restrictive in that it requires the presence of that type of deterministic component, so for instance even the case $s_t = 0$ is not included. This is due to the structure of the proof, which requires the

calculation of the loss function over the whole parameter space: indeed, notice that Heyde and Day (1996), discussing a similar problem (they considered the full spectrum Whittle estimate, rather than focusing on the lowest frequencies), proposed the more general $|s_t| \leq Ct^{\phi-1/2}$, but they did not actually prove the consistency of the estimate, and rather assumed it and went to discuss the potential lower order bias in the limit distribution. Assumption A.3 could be relaxed, for example to $|s_t| \leq Ct^{\phi-1/2}$ when $\phi \leq \delta$, but other details should be given to deal with the case in which $\phi \leq \delta$ does not hold but Assumption 2.5 is still met, if we want to follow the proof of Robinson (1995b). In any case, for practical purposes we conjecture that the order of magnitude in Assumption A.3 could be treated as an upper bound instead.

The condition (2.22) indicates which deterministic components are irrelevant. Higher trends can be ignored the stronger the autocorrelation is, as we already conjectured when comparing the periodograms of the deterministic and of the stochastic terms. Higher trends can also be neglected the larger κ and υ are, because high κ means including more frequencies in which the stochastic rather than the deterministic component dominates the order of magnitude of the periodogram of x_t (due to the damping factor j^{-1}); higher υ is similar, because it means that less periodograms in which the deterministic component may be relevant are used in the estimation. It also indicates that trimming is not necessary when $\phi \leq \delta$, because $\frac{\kappa}{1-\upsilon}$ can only be positive.

Since in practice ϕ and δ are unknown, we suggest using (2.22) to choose υ if at least we have some preliminary information on $\phi - \delta$: rewriting that condition as

$$\upsilon > \frac{2(\phi - \delta) - \kappa}{2(\phi - \delta)}, \tag{2.24}$$

if, for example, we expect $\phi = 1/2$ and $\delta \geq 0$, and we set $\kappa = 0.79$, the minimal trimming has $\upsilon > 0.21$.

Theorem 2.2. *Under Assumptions 2.2, 2.3, and 2.5,*

$$\widehat{\delta} \to_p \delta \text{ as } n \to \infty. \tag{2.25}$$

Notice that we do not mention Assumptions 2.1 and 2.4 explicitly because they are already included in Assumption 2.5.

Under a stronger set of conditions, Robinson (1995b) also derived the limit distribution of the estimate $\widehat{\delta}$. We repeat these below, updating them in order to take the deterministic component into account as well.

Assumption 2.1'. *For some* $\beta_\xi \in (0, 2]$

$$f_\xi(\lambda) \sim G\lambda^{-2\delta}(1 + O(\lambda^{\beta_\xi})) \text{ as } \lambda \to 0^+, \tag{2.26}$$

where $G \in (0, \infty)$ *and* $\delta \in [\Delta_1, \Delta_2] \subset (-1/2, 1/2)$.

Assumption 2.2'. *In a neighbourhood* $(0, \iota)$ *of the origin,*

$$\alpha(\lambda) = \sum_{s=0}^{\infty} \alpha_s e^{i\lambda s} \tag{2.27}$$

is differentiable and

$$\frac{d}{d\lambda}\alpha(\lambda) = O\left(\frac{|\alpha(\lambda)|}{\lambda}\right) \text{ as } \lambda \to 0^+. \tag{2.28}$$

Assumption 2.3'. *Assumption 2.3 holds and also*

$$E(\varepsilon_t^3 | F_{t-1}) = c_3,\ E(\varepsilon_t^4 | F_{t-1}) = c_4,\ a.s.,\ t = 0, \pm 1, \ldots \tag{2.29}$$

for some finite constants c_3 *and* c_4.

Assumption 2.4'. *Assumptions A.1 and A.2 hold and*

$$0 \leq \upsilon < \kappa < 2\beta_\xi / \left(1 + 2\beta_\xi\right). \tag{2.30}$$

Assumption 2.5'. *Assumptions A.3, 2.1' and 2.4' hold and*

$$\phi < \delta + \frac{1}{4}\frac{\kappa}{1-\upsilon}, \tag{2.31}$$

Assumptions 2.1' to 2.3' are those originally proposed for the stochastic component. The information on the shape in (2.26), that was not already provided in Assumption 2.1, is necessary to define whether the approximation of the density with $G\lambda^{-2\delta}$ may generate a lower order bias: since that approximation is less precise at high frequencies, the highest ones must be removed, as the assumption on $\kappa < 2\beta_\xi / \left(1 + 2\beta_\xi\right)$ also indicates. The weakest upper bound is for $\beta_\xi = 2$, a class that also includes the case in which ξ_t is an ARFIMA process.

Assumption 2.5' replicates Assumption 2.5 but the condition (2.31) on ϕ is stronger than the one in (2.22): in fact in this case consistency is not enough, and it is also necessary that the bias is of order smaller than $1/\sqrt{m}$. Intuitively, when $\phi > \delta$ consistency is still possible (given the regularity conditions of Theorem 2.2) because the stochastic component dominates the periodogram of the deterministic term on enough frequencies. But the fact that the deterministic term prevails in some frequencies may induce a positive lower order bias, because on the lowest frequencies $I_s\left(\lambda_j\right)$ is markedly steeper than $f_\xi\left(\lambda_j\right)$ in that case. The condition (2.31) then ensures that the contribution from the deterministic component dominates in $I_x\left(\lambda_j\right)$ on such a little range of frequencies that this effect is negligible. As we did for (2.22), we suggest reading

(2.31) as a condition on the trimming as well, in this case being

$$v > \frac{2(\phi - \delta) - \kappa/2}{2(\phi - \delta)}. \tag{2.32}$$

In the example above, $\phi = 1/2$, $\delta \geq 0$ and $\kappa = 0.79$, the minimal trimming has $v > 0.605$.

Since these assumptions are sufficient to confirm the limit distribution given by Robinson (1995b) for the case of no deterministic component, as we discuss in Theorem 2.3 below, we think that this is a very strong result, because it means that even a break in the mean can be dealt with.

Theorem 2.3. *Under Assumptions 2.2', 2.3', and 2.5',*

$$\sqrt{m}(\widehat{\delta} - \delta) \to_d N(0, \frac{1}{4}) \text{ as } n \to \infty. \tag{2.33}$$

Theorem 2.3 seems to offer a free lunch: trimming may help to reduce the distortionary effect of the deterministic component without even inflating the variance.

When Assumption 2.4 is met but 2.4' is not, then the lower order bias prevents reaching the limit distribution of Theorem 2.3. We show the nature of the lower order bias in the following theorem.

Theorem 2.4. *Under Assumptions 2.1', 2.2', 2.3', 2.5, and*

$$0 < 2(\phi - \delta)(1 - v) < \kappa < \min\left\{4(\phi - \delta)(1 - v), 2\beta_\xi / (1 + 2\beta_\xi)\right\}, \tag{2.34}$$

then

$$mn^{2(\delta-\phi)} l^{2(\phi-\delta)} \ln^{-1} m \left(\widehat{\delta} - \delta\right) \to_p K \text{ as } n \to \infty. \tag{2.35}$$

The limit (2.35) gives the lower order bias under the given assumptions and (2.34). Notice, first, that the estimate is indeed consistent, because the

conditions of Theorem 2.2 are met, so this is really only a lower order bias.

We need Assumption 2.1' and $\kappa < 2\beta_\xi / \left(1 + 2\beta_\xi\right)$ to deal with another potential lower order bias, due to the approximation of the spectral density at intermediate frequencies: these two assumptions ensure that it is $o\left(1/\sqrt{m}\right)$.

The rest of condition (2.34) specifies when the lower order bias due to the neglected deterministic component can be relevant: we already argued that the problem should only emerge when $\phi > \delta$, and indeed in (2.34) this appears in $0 < 2\left(\phi - \delta\right)\left(1 - \upsilon\right)$. The condition $\kappa < 4\left(\phi - \delta\right)\left(1 - \upsilon\right)$ ensures that Assumption 2.5' is not met, because $m^{-1}n^{2(\phi-\delta)}l^{2(\delta-\phi)}$ would be of order smaller than $m^{-1/2}$ if $-\kappa + 2\left(\phi - \delta\right) - 2\upsilon\left(\phi - \delta\right) < -\kappa/2$, so if $\kappa > 4\left(\phi - \delta\right)\left(1 - \upsilon\right)$, as we can see simply replacing m with $c_\kappa n^\kappa$, l with $c_\upsilon n^\upsilon$ and then comparing the exponents.

Finally, Theorem 2.4 also confirms two other conjectures we stated before: that the bias, when it exists, is positive, and that it is smaller the larger the trim (notice there that the order of magnitude depends on $l^{2(\delta-\phi)}$, so it is smaller the larger υ is, because $\phi > \delta$), while it is bigger the larger $\phi - \delta$ is for given m, l. We will take these results into account in the design of a test to detect relevant deterministic components.

2.2.3 A test to detect deterministic components

Theorems 2.2 and 2.3 gave the combinations of δ, ϕ that are sufficient for consistency and $\sqrt{m}$, zero-mean limit normality of the estimate.

This requires knowledge of $\delta - \phi$, a piece of information that is not usually available. There are, however, cases in which the researcher has preliminary information on the highest possible ϕ, and this, combined with the results of the estimation, can be enough: if for example we are only concerned about a shift in the mean ($\phi = 1/2$), and we intend to estimate δ with $\kappa = 0.80 - \varepsilon$

(for a very little positive ε), $l = 1$, we then know that consistency requires $\delta > 0.1$ and the limit distribution (2.33) also requires $\delta > 0.3$. If we estimated $\widehat{\delta} = 0.4$, we could conclude that the potential shift of the mean is not a relevant problem.

Yet most of the cases are less simple to handle: $\widehat{\delta} = 0.2$ in the example above could be evidence of the lower order bias of Theorem 2.4, or simply indicate the proper order of integration because in fact the feared deterministic component is not present. Of course when there is no preliminary information about the deterministic component at all, it is not possible to conclude whether our estimate is consistent simply by looking at it.

We then propose a simple test to detect the presence of a relevant deterministic component. For this purpose, we introduce the notation $\widehat{\delta}^{(l)}$ to refer to the local Whittle estimate when trimming is actively used (that is, $l \to \infty$),

$$\widehat{\delta}^{(l)} = \arg \min_{d \in [\Delta_1, \Delta_2] \subset (-1/2, 1/2)} R(d) \ \textit{with} \ 1/l \to 0 \ \textit{when} \ 1/n \to 0, \qquad (2.36)$$

and the new estimate

$$\widetilde{\delta}^{\dagger} = \arg \min_{d \in [\Delta_1, \Delta_2] \subset (-1/2, 1/2)} \ln \left\{ \frac{1}{m/2} \sum_{j=0}^{m/2-1} \lambda_{2j+1}^{2d} I_x(\lambda_{2j+1}) \right\} - 2d \frac{1}{m/2} \sum_{j=0}^{m/2-1} \ln(\lambda_{2j+1}) \qquad (2.37)$$

for some even m. The estimate $\widetilde{\delta}^{\dagger}$ is still of local Whittle type, so we anticipate that the results stated for $\widehat{\delta}$ hold for $\widetilde{\delta}^{\dagger}$ too, the only difference being that the variance is doubled because only half of the frequencies are used in the estimation. Since we used $2j$ rather than j, the optimization is still done on the Fourier frequencies spanning the same subset of $(0, 2\pi)$ used for $\widehat{\delta}$, so the approximation of the spectral density in (2.2) is roughly the same for both the estimates: any relevant difference between $\widetilde{\delta}^{\dagger}$ and $\widehat{\delta}^{(l)}$ should then depend on the deterministic component.

In order to derive the asymptotic properties of $\widetilde{\delta}^{\dagger}$, we modify Assumptions 2.4 and 2.5 (and then 2.4' and 2.5' as well) to take into account the fact there is no trimming: these are replaced by Assumption 2.6 and 2.6', that we introduce below.

Assumption 2.6. *Assumptions A.1, A.3 and 2.1 hold and*

$$0 < \kappa < 1 \tag{2.38}$$

and

$$\phi < \delta + \frac{\kappa}{2}. \tag{2.39}$$

Assumption 2.6'. *Assumptions A.1, A.3 and 2.1' hold and*

$$0 < \kappa < 2\beta_{\xi}/\left(1 + 2\beta_{\xi}\right) \tag{2.40}$$

and

$$\phi < \delta + \frac{\kappa}{4}. \tag{2.41}$$

The asymptotic properties of $\widetilde{\delta}^{\dagger}$ are then summarised in the theorem below.

Theorem 2.5. *(i) under Assumptions 2.2, 2.3, 2.6,*

$$\widetilde{\delta}^{\dagger} \rightarrow_{p} \delta \text{ as } n \rightarrow \infty; \tag{2.42}$$

(ii) under Assumptions 2.2', 2.3', 2.6',

$$\sqrt{m}(\widetilde{\delta}^{\dagger} - \delta) \rightarrow_{d} N(0, \frac{1}{2}) \text{ as } n \rightarrow \infty; \tag{2.43}$$

(iii) under Assumptions 2.1', 2.2', 2.3', 2.6, and

$$0 < 2\left(\phi - \delta\right) < \kappa < \min\left\{4\left(\phi - \delta\right), 2\beta_{\xi}/\left(1 + 2\beta_{\xi}\right)\right\}, \tag{2.44}$$

then

$$mn^{2(\delta-\phi)} \ln^{-1} m \left(\widetilde{\delta}^{\dagger} - \delta\right) \rightarrow_p K \text{ as } n \rightarrow \infty. \tag{2.45}$$

The test is based on the following result:

Theorem 2.6. *(i) under Assumptions 2.2', 2.3', 2.5', 2.6',*

$$\sqrt{m}(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}) \rightarrow_d N\left(0, \frac{1}{4}\right) \text{ as } n \rightarrow \infty, \tag{2.46}$$

and (ii) under Assumptions 2.1', 2.2', 2.3', 2.5, 2.6, and

$$0 < 2(\phi - \delta) < \kappa < \min\left\{4(\phi - \delta), 2\beta_\xi / (1 + 2\beta_\xi)\right\}, \tag{2.47}$$

then

$$\sqrt{m}\left(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}\right) \rightarrow \infty \text{ as } n \rightarrow \infty. \tag{2.48}$$

The test we propose is then based on a comparison of the estimate with trimming, $\widehat{\delta}^{(l)}$, and without trimming, $\widetilde{\delta}^{\dagger}$: under the given assumptions, a large value (when compared to a critical value) of the test statistic $\sqrt{m}\left(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}\right)$ is evidence of the presence of a bias of order bigger than $m^{-1/2}$ in $\widetilde{\delta}^{\dagger}$. Notice that the test does not require root-m consistent estimation of $\widehat{\delta}^{(l)}$: from Theorem 2.4, even if the latter too is subject to a lower order bias, the dimension of that bias is smaller than the one of the bias of $\widetilde{\delta}^{\dagger}$, so the test statistic still detects the presence of the deterministic component.

Since the bias, if it exists, can only be positive, we suggest taking the critical value for a test for a one sided alternative.

For a formal definition, we introduce

$$K_0 = Plim\left(\min\left\{m^{1/2}, mn^{2(\delta-\phi)}\right\} \ln m^{-1}\left(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}\right)\right) \tag{2.49}$$

under Assumptions *2.1', 2.2', 2.3', 2.4', 2.6,* $v > 0$: we suggest then testing

$$H_0 : \{K_0 = 0\} \quad \text{vs} \quad H_1 : \{K_0 > 0\} \tag{2.50}$$

by comparing the realisation of the test statistic $\sqrt{m}\left(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}\right)$ with a critical value from a standard normal.

If the null hypothesis is not rejected, we can presume that the estimate $\widehat{\delta}$ is consistent and the limit distribution of Theorem 2.3 holds. When the null hypothesis is rejected, though, we can only conclude that $\widetilde{\delta}^{\dagger}$ is at least subject to a lower order bias.

It should be noticed that those deterministic components for which consistency is not proved are not included in the assumptions above: our object of interest is the root-m consistent estimation of δ rather than the discussion, for example, of the cases having $\phi > \delta + \kappa/(2(1-v))$. We anyway conjecture that the deterministic trend always has a stronger impact on $\widetilde{\delta}^{\dagger}$ than on $\widehat{\delta}^{(l)}$ due to the trimming, so these cases too should be detected by the test.

We conclude by explaining why only a fraction of the available frequencies are used in $\widetilde{\delta}^{\dagger}$. It may indeed seem more obvious to use

$$\widehat{\delta}^{(1)} = \arg \min_{d \in [\Delta_1, \Delta_2] \subset (-1/2, 1/2)} R(d) \ \textit{when} \ l = 1, \tag{2.51}$$

which is the original local Whittle estimate and has smaller asymptotic variance than $\widetilde{\delta}^{\dagger}$.

Theorem 2.7. *Under Assumptions 2.1', 2.2', 2.3', 2.4' with* $v > 0$, $s_t = 0$,

$$\left(m/\sqrt{l}\right)\left(\widehat{\delta}^{(l)} - \widehat{\delta}^{(1)}\right) \to_d N\left(0, \frac{1}{4}\right) \ \textit{as} \ n \to \infty. \tag{2.52}$$

Unfortunately then $\widehat{\delta}^{(1)}$ and $\widehat{\delta}^{(l)}$ are asymptotically too similar when the deter-

ministic term is irrelevant, so their difference is of a smaller order of magnitude: a test statistic based on $\left(m/\sqrt{l}\right)\left(\widehat{\delta}^{(l)} - \widehat{\delta}^{(1)}\right)$ would also detect deterministic components which do not affect the limit distribution stated in Theorem 2.3.

2.3 Monte Carlo evidence

In order to investigate how reliable a guide the asymptotic theory is in moderate-sized samples, a small Monte Carlo study was carried out.

We considered three deterministic structures and two stochastic components; in each situation we compared the estimate with and without trimming.

The case of no deterministic structure, $s_t = 0$, was our benchmark. Following Bhattacharya *et al.* (1983) and other works in the literature, we allowed for a fractional trend and set $s_t = 2t^{-1/4}$, corresponding to $\phi = 1/4$. The last deterministic structure we considered is the shift in the mean, posing it in the middle of the sample, so $s_t = 0$ for $t \leq n/2$ and $s_t = 1$ for $t > n/2$: the possible bias induced by this component is a serious concern in the applied literature so we think it was important to observe the performance of the estimate with trimming in this case.

For the stochastic component, we set $\delta = 0$ and $\delta = 0.4$: since it is the difference $\delta - \phi$ that really matters, we considered in this way quite a wide range of situations. A large δ was also important to analyse a case in which the condition stated by Heyde and Dai (1996) is not met.

The data were generated as a sequence of independent standard normals for $\delta = 0$, and using the Davies and Harte (1987) simulator for $\delta = 0.4$.

We set the bandwidth and the trimming parameter as $m = 0.8n^{0.79}$, $l = 0.2n^{0.62}$, and employed $n = 64$, 128, 256, 512, 1024, with 1000 replications.

For each combination we computed the local Whittle estimates with and without trimming the lowest frequencies, and the statistic $\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}$. Since

the local Whittle estimate does not have a closed form formulation, we used the log-periodogram regression estimate, trimming the lowest frequencies, as a starting value in the numerical optimization.

In the rest of the section and in the Tables, we refer to the three deterministic models as $s(-\infty)$, $s(1/4)$ and $s(1/2)$ respectively, while for the stochastic model we use $d(0)$ and $d(0.4)$; l and 1 distinguish the case in which trimming was applied or not.

Making use of Theorems 2.2, 2.3 and 2.4, all the combinations yield consistent and asymptotically normal estimates under that rather aggressive trimming; without it, root-m convergence fails for $(1, d(0), s(1/4))$, and consistency could fail altogether for $(1, d(0), s(1/2))$.

In Tables 2.3 and 2.4 we report for $\widehat{\delta}^{(1)}$ and $\widehat{\delta}^{(l)}$ the average of the deviations of the estimates from δ (*bias*), the sample standard deviation (*s.d.*) of the estimates and the one prescribed by the asymptotic theory (*a.s.d.*). Notice that two measures are presented there: under the column $\widehat{\delta}^{(1)}$ we report $1/\sqrt{4m}$, while under the column $\widehat{\delta}^{(l)}$ we propose as an alternative reference $1/\sqrt{4(m-l+1)k_{l,m}}$ where

$$k_{l,m} = \frac{1}{m-l+1}\sum_{j=l}^{m}\nu_j^2 \text{ where } \nu_j = \ln j - \frac{1}{m-l+1}\sum_{j=l}^{m}\ln j. \tag{2.53}$$

The choice of this factor depends on its presence in the calculation of the limit normality and of the variance in Theorem 2.3: of course, the statement in the theorem is only asymptotic, in which case the correction is irrelevant; moreover, many other terms are involved in the approximation, so we do not attempt to propose this as a rigorous correction for the small samples, but we mention it because it worked well at least in our Monte Carlo exercise.

We summarise the results looking at the root of the sample mean squared error ($rMSE$): these are in the first two columns of Tables 2.5 and 2.6. In the

rest of the two tables we present the empirical sizes of some tests of interest: in the columns $t_{\widehat{\delta}^{(1)}}$ and $t_{\widehat{\delta}^{(l)}}$ we reported 100 times the percentage in which the standardized t statistic $2\sqrt{m}(\widehat{\delta} - \delta_0)$ to test H_0:$\{\delta = \delta_0\}$ *vs* H_1:$\{\delta > \delta_0\}$, where δ_0 is 0 or 0.4 according to the situation, exceeded the critical value of a 5% significance test; in the last column, $t_{\widehat{\delta}^{(l)} - \widetilde{\delta}^{\dagger}}$, we analyse the reliability of the test to detect the deterministic component by looking at 100 times the percentage in which $2\sqrt{m}\left(\widehat{\delta}^{(l)} - \widetilde{\delta}^{\dagger}\right)$ exceeded the 5% threshold with a one sided alternative.

Despite the smallness of the samples, the results were broadly in line with the theory, at least if we only consider the main features. We found that the bias was always quite small but for the case $(1, d(0), s(1/2))$, the only one not covered by the theory, where it was about 0.25. Not surprisingly, $(1, d(0), s(1/4))$ was the only other one exhibiting a certain systematic deviation from the true value (approximately 0.06; it did not decrease much, if at all, with the increase of the dimension of the sample).

Given that the periodogram of the deterministic component may still dominate in the frequencies closer to 0, a minimal residual bias, which should vanish at a rate faster than root-m, can still appear in small samples even when the conditions for Theorem 2.3 are met. This was the case in a few combinations: since the bias depends on the gap $\delta - \phi$ and on the trimming, it was larger in the case $(1, d(0.4), s(1/2))$, where it was 0.04 for $n = 64$ and 0.02 for $n = 1024$. On the other hand the reduction of the bias realised trimming was complete: even in the most unfavourable situation, $(1, d(0), s(1/2))$, it quickly dropped below 0.02.

The bias generated by the deterministic component did not affect the dispersion at least if the hypotheses for Theorem 2.2 are met.

Trimming on the other hand had a strong effect on the dispersion, despite the fact that it should not, according to the asymptotic theory: the standard

deviation of the estimates with trimming was in some cases nearly 60% more, and the approximation $2\sqrt{m}$ did not seem very close to it even in the largest sample. We said before that Theorem 2.3 seemed to offer a free lunch, but we saw here that in practice this is not the case. This was not a surprise: commenting on trimming for the log-periodogram regression estimate, Hurvich, Deo and Brodsky (1998) noticed that the removal of the few lowest frequencies resulted in a marked increase in the dispersion of the estimates in the simulation, and a poorer approximation of the variance indicated by the asymptotic theory. Notice however that the correction by the factor (2.53) would greatly improve the precision of the approximation.

Despite the potential lower order bias, the dispersion clearly dominated the rMSE: as a consequence, $\widehat{\delta}^{(1)}$ was always superior in the cases in which the conditions for Theorem 2.3 were met, and it was roughly equivalent to $\widehat{\delta}^{(l)}$ when at least consistency was achieved, thus trimming was only superior in a rMSE sense when the gap $\phi - \delta$ was very large.

Turning to the approximation stated in Theorem 2.3, first notice that in the case with no trend and no trim, the test statistic replicated the theoretical size of the t test for $H_0 : \{\delta = \delta_0\}$ vs. $H_1 : \{\delta > \delta_0\}$ very effectively.

The lower order bias had a certain impact on the distribution as a whole: the sizes $t_{\widehat{\delta}^{(1)}}$ and $t_{\widehat{\delta}^{(l)}}$ increased with the gap $\phi - \delta$ even when Theorem 2.3 still held, if the gap was relatively large. In general, the discrepancy in size with respect to the case having the same δ and the same l, and no trend, was smaller the larger the sample. The sizes $t_{\widehat{\delta}^{(1)}}$ computed for $(1, d(0), s(1/4))$ and $(1, d(0), s(1/2))$ were on the other hand quite large, confirming that the limit distribution stated in Theorem 2.3 did not follow when the deterministic component was too strong.

Trimming shifted the size above 5% as well, although here too the approximation improved with the dimension of the sample.

If trimming is not necessary, it is advisable to set $l = 1$, especially when the sample is small. The test that we proposed to detect a deterministic component can also help choosing whether to trim the lowest frequencies. We analysed it looking at the 5% size for the test statistic $\sqrt{4m}(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)})$. This was a little too large: about 20% already in the case without deterministic component, and it only improved slowly as the sample increased. The situation was even worse at least in one case with a mildly strong deterministic component: for $d\,(0.4)$, $s\,(1/2)$ the combination of the residual bias in $\widetilde{\delta}^{\dagger}$ and of the excessive dispersion of $\widehat{\delta}^{(l)}$ caused the rejection of the hypothesis of no relevant deterministic component in 30% to 40% of the cases.

We would like to conclude by saying that when there is no additional information on ϕ and the null hypothesis of the test is rejected, then trimming is a safer strategy: even if we ignore if the estimate is consistent, we may at least expect that the bias is sensibly reduced. If additional information on ϕ is available, we suggest not to follow the result of the test blindly, but rather to decide on a case by case approach.

2.4 Two empirical applications

We illustrate these results by means of two empirical examples.

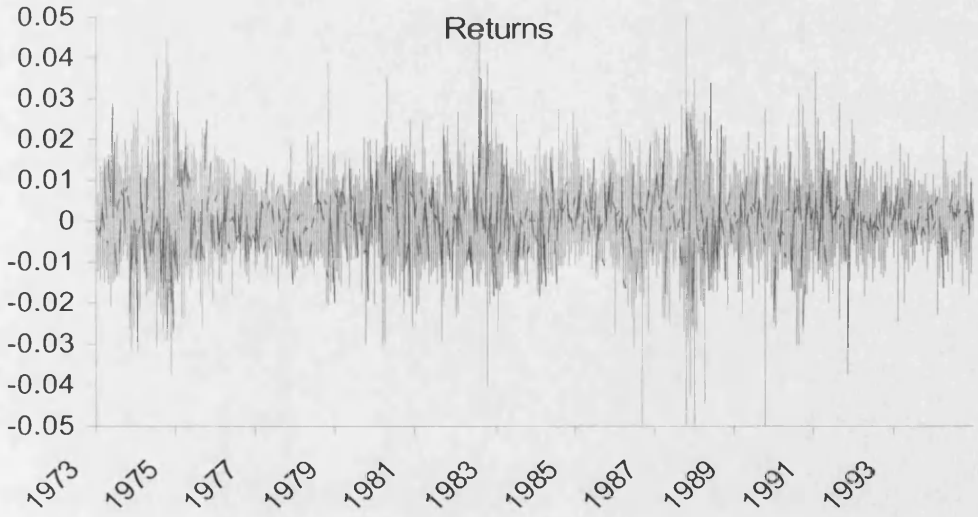
First, we discuss the daily S&P500 Index.

Lobato and Savin (1998) analysed the returns, r_t, their absolute values $|r_t|$ and their squares r_t^2 for each day over the period July 1962 - December 1994. Since they suspected that the oil shock in 1973 and the stock market crash in 1987 caused shifts in the mean of r_t^2 in the second part of the sample, thus inducing spurious evidence of fractional integration, they split the sample in 1973. We then analysed the subset 1973 - 1994, to assess the presence of the instability that Lobato and Savin expected.

This period was also analysed by Granger and Hyung (2004). Using log-periodogram regression, they compared a fractional model with no break with the case in which the number of breaks is unknown and endogenously estimated. They took subsamples of the period 1928 - 2002 and concluded that both the models describe $|r_t|$ equally well. Yet notice that their estimate when no breaks are allowed for was remarkably higher in the period 1973 - 1979: this seems to indirectly provide evidence in favour of one or more breaks in those years.

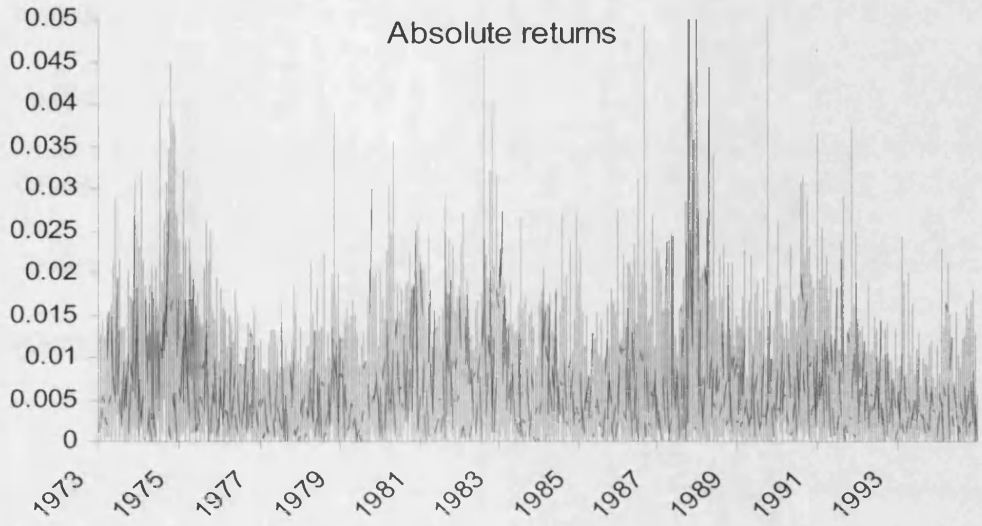
Our data were collected from Datastream and have code S&PCOMP(PI): this is a price index, and the returns were computed by taking first differences of the logarithms. The plots of r_t, $|r_t|$ and r_t^2 are presented in Figure 2.1 to 2.3 respectively.

Figure 2.1: S&P500 index returns



The period following the 1973 oil shock and the one following the 1987 stock market crash seem to be characterised by higher volatility, as Lobato and Savin (1998) warned, so a break in the mean can be anticipated both for $|r_t|$ and r_t^2. The raw returns r_t on the other hand do not seem to exhibit any deterministic component.

Figure 2.2: S&P500 index absolute returns

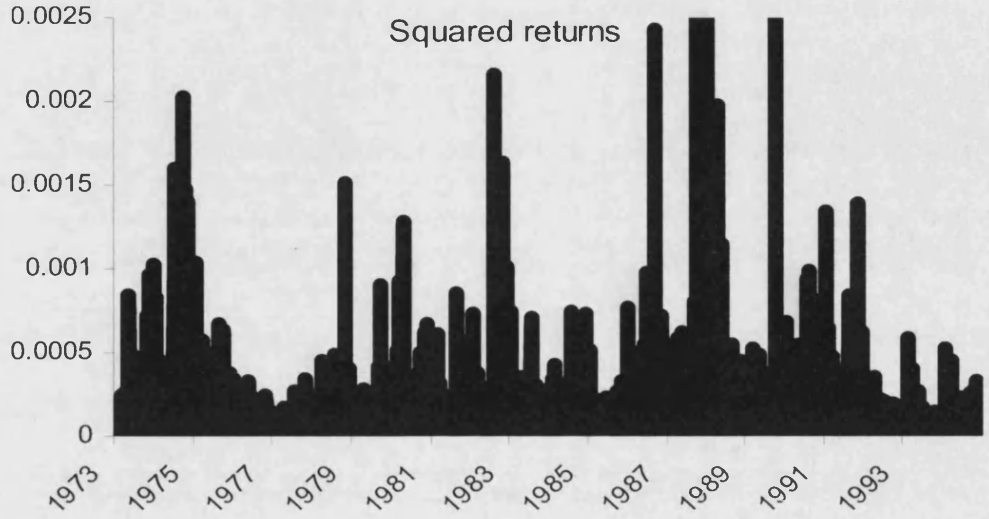


There are 5740 observations, but one is lost to first differencing, so $n = 5739$. Lobato and Savin (1998) considered bandwidths ranging between 30 and 100: assuming $m = c_\kappa n^{0.79}$, their approach was extremely conservative, and would not leave many periodograms for the optimization when the lowest frequencies are trimmed in the loss function. We kept $m = 100$ for comparison, but also took $m = 0.2n^{0.79} = 186$ and $m = 0.25n^{0.79} = 233$, which are still very conservative but left a reasonable number of periodograms in the optimization even allowing for trimming; we set $l = 40$ in the case $m = 100$, but otherwise considered $l = 0.2n^{0.62} = 42$ and $l = 0.25n^{0.62} = 53$.

The results are summarised in Table 2.1.

The estimates $\widehat{\delta}^{(1)}$ for $m = 100$ were very close to those in Lobato and Savin (1998), who found that the returns r_t did not exhibit strong autocorrelation, while the absolute $|r_t|$ and the squared returns r_t^2 appeared to do so. Increasing the bandwidth resulted in slightly smaller estimates for $|r_t|$ and for r_t^2 and had no effects for r_t: summarising, the estimated value of the memory parameter of $|r_t|$ was approximately 0.4 while the one of r_t^2 was smaller, it being less than 0.15.

Figure 2.3: S&P500 index squared returns



Yet, we found that when trimming the lowest frequencies, the instability that Lobato and Savin suspected may have really increased the estimated values for the memory parameters of r_t^2 and $|r_t|$. In all the cases with trimming, the squared returns appeared to have short memory; the evidence of short memory is less clear for the absolute returns, but there too we found a strong reduction in the estimated values; in all the cases the test to detect relevant deterministic components rejected the null hypothesis so trimming should be preferred.

For the second example, we discuss the first difference in the logged quarterly price indices of Boston, New York and Philadelphia from 1950 (first quarter) through 2003 (third quarter), so $n = 214$ (again, one observation is lost because we used first differences of the logarithm of the price index). The data were collected from Datastream and have codes USCPBOMAF, USCPNYMAF, USCPPHMAF; the sampling frequency was intended to be monthly, but for several years the data were only collected every second or third month both for Boston and Philadelphia: for each city we produced quarterly data by averaging.

Table 2.1: Estimates of the memory parameter S&P500 index

		r_t	$\lvert r_t \rvert$	$(r_t)^2$
$m = 100$, $l = 40$	$\widehat{\delta}^{(1)}$	0.002	0.440	0.131
	$\widehat{\delta}^{(l)}$	0.168	0.270	-0.061
	$\sqrt{4m}\left(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}\right)$	-3.324	4.949	4.511
$m = 186$, $l = 42$	$\widehat{\delta}^{(1)}$	-0.016	0.358	0.108
$(m = 0.2n^{0.79}, l = 0.2n^{0.62})$	$\widehat{\delta}^{(l)}$	-0.019	0.113	0.014
	$\sqrt{4m}\left(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}\right)$	0.141	7.010	3.003
$m = 186$, $l = 53$	$\widehat{\delta}^{(1)}$	-0.016	0.358	0.108
$(m = 0.2n^{0.79}, l = 0.25n^{0.62})$	$\widehat{\delta}^{(l)}$	-0.053	-0.017	0.002
	$\sqrt{4m}\left(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}\right)$	1.082	10.543	3.314
$m = 233$, $l = 53$	$\widehat{\delta}^{(1)}$	-0.003	0.342	0.106
$(m = 0.25n^{0.79}, l = 0.25n^{0.62})$	$\widehat{\delta}^{(l)}$	-0.026	0.034	0.040
	$\sqrt{4m}\left(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}\right)$	-1.154	9.424	2.339

Interest in inflation is justified by the fact that central banks are committed to some forms of inflation stabilisation, thus a mean-reverting dynamics should be anticipated and the memory parameter can be treated as an indicator of how quickly inflation shocks are absorbed by the economic system, under the action of the monetary authority. Sudden phases of high inflation, such as those taking place after the oil shocks, may be regarded as temporary shifts in the mean due to external, exogenous phenomena, rather than periods of careless or inappropriate monetary management. With our analysis, we can identify if they affected the estimation of the memory parameter, and eventually remove their effect.

The data are plotted in Figure 2.4.

Commentators often identify at least two phases for inflation: an initial period of relatively low inflation, then a sudden increase associated to the two oil shocks, and then a slow return to the original lower level. The intermediate "high" inflation could then be grossly associated with a shift in the mean,

although in this case even some polynomial trends may help to describe the relatively slow transition.

Figure 2.4: Annualised quarterly inflation: BY, Bo, Ph

We considered here the two pairs $m = 16$, $l = 3$ and $m = 50$, $l = 6$. The bandwidth $m = 16$ is the MSE - optimal when the process is a ARFIMA(1,δ,0) with autoregressive coefficient 0.5 (see Henry and Robinson (1996); we did not assume that model, we merely referred to it since it is intermediate in a range of possible short term autoregressive structures); $m = 50$ is quite a large bandwidth, corresponding to $m = 0.72n^{0.79}$, but it still does not include the frequencies involved in a potential seasonal cycle, the peak of seasonality being around $m = 53$. The lowest frequency l was decided considering $l = 0.2n^{0.62} = 5$ as a reference, but allowing for a certain flexibility around it.

The estimated values are presented in Table 2.2.

The estimates were mainly in the range 0.4 - 0.5, albeit on a few points they passed that threshold: we think that the estimates were reliable anyway, though, because Velasco (1999b) showed that the limit distribution of Theorem 2.3 holds even for stochastic processes having $\delta \in [0.5, 0.75)$. Admittedly he did

Table 2.2: Estimates of the memory parameter for quarterly growth of prices

		NY	Bo	Ph
$m = 16$, $l = 3$	$\widehat{\delta}^{(1)}$	0.530	0.454	0.397
	$\widetilde{\delta}^{(l)}$	0.422	0.321	0.101
	$\sqrt{4m}\left(\widetilde{\delta}^{\dagger} - \widetilde{\delta}^{(l)}\right)$	0.783	0.319	2.464
$m = 50$, $l = 6$	$\widehat{\delta}^{(1)}$	0.511	0.439	0.459
	$\widetilde{\delta}^{(l)}$	0.509	0.449	0.488
	$\sqrt{4m}\left(\widetilde{\delta}^{\dagger} - \widetilde{\delta}^{(l)}\right)$	0.586	0.924	0.915

not consider trimming and excluded the generic deterministic component (at least in the case with no taper), but it is fair to conjecture that the extension of the arguments of Velasco can be done on the same lines we gave for Robinson (1995b), so we conclude that if indeed $\delta \geq 0.5$ then a potential shift in the mean is irrelevant. Yet even if $\delta < 0.5$ we found that the estimates were so large that the signal of a shift in the mean should be covered by the stochastic component in the periodogram. The only potential exception was Philadelphia for $m = 16$, $l = 3$, which had $\widetilde{\delta}^{(l)} = 0.10$, but notice that when $m = 50$, $l = 6$ then the estimated value was again 0.49, so we concluded that a break in the mean, if present, did not affect the estimate of the persistence.

Finally, we looked at the test statistics $\sqrt{4m}\left(\widetilde{\delta}^{\dagger} - \widetilde{\delta}^{(l)}\right)$. Since we already ruled out shifts in the mean and indeed any deterministic component having $\phi < 1/2$, a strict interpretation of the result of this test may only reveal the presence of a trend with $\phi > 1/2$: we did not treat this case explicitly, but this can be conjectured on the basis of Theorems 2.1 and 2.4. Even in that case we found that the estimation was not affected by any deterministic component, since we did not reject that hypothesis in 5 cases out of 6.

2.5 Discussion

We have studied the local Whittle estimate of the memory parameter in presence of a time-varying deterministic component.

We have found that the local Whittle estimate is less prone than the R/S and related statistics to be fooled into confounding long memory and deterministic components. By studying the periodograms of the deterministic trends and of the shift in the mean, we have also found that they concentrate much more power in the lowest frequencies, and their effect can then be easily removed by trimming those. We have shown that whether the deterministic or the stochastic component prevails, depends on the difference $\phi - \delta$, and that high ϕ can be neglected if the order of integration of the data is high (thus reversing the finding of Heyde and Dai (1996)). Finally, we have proposed a test to detect relevant deterministic components.

We conclude by discussing some conjectures which we also derived from our results and some potential extensions.

1. We only discussed the local Whittle estimate, but we think that the same results apply for the log-periodogram regression estimate and, setting $\kappa = 1$, for the Whittle estimate.

2. We discussed a Type I integrated process only, but we expect that all the results carry through if a Type II is considered instead. Also, we focused on the range of δ that is more often considered in the literature, but we think that a wider range for δ could be treated, following Velasco (1999b); notice anyway that, at least in the univariate analysis, restricting to $\delta \in (-1/2, 1/2)$ is a very common practice, because if the order of integration is higher it is still possible to recover a stationary and invertible process differencing the original observations enough times.

3. It would be of interest to characterise the deterministic component in Theorem 2.2 and 2.3 in the most general way. We think that a condition of practical use is

$$\frac{1}{m-l+1}\sum_{j=l}^{m}\lambda_j^{-2\delta}I_s(\lambda_j)=o(1) \tag{2.54}$$

for consistency and

$$\frac{\sqrt{m-l+1}}{m-l+1}\ln m\sum_{j=l}^{m}\lambda_j^{-2\delta}I_s(\lambda_j)=o(1) \tag{2.55}$$

for root-m, zero-mean limit normality.

4. We considered a relatively small range of ϕ (except in Theorem 2.1), mainly in order to keep the proofs simple. For $\phi > 1/2$, by using the order of magnitude in Theorem 2.1 and (2.54) and (2.55), the condition for consistency would be

$$v > \frac{2(\phi-\delta)-\kappa}{1-2\delta}, \tag{2.56}$$

while for the limit distribution of Theorem 2.3,

$$v > \frac{2(\phi-\delta)-\kappa/2}{2(1-\delta)}. \tag{2.57}$$

5. We argued that when (2.22) in Assumption 2.5 is not met, then the estimate is inconsistent but we did not formally prove it. We did not pursue this because our object of interest was the consistent estimation of δ rather than the test, but we nonetheless think the discussion of the case in which consistency fails might be an interesting topic for future research.

6. When the hypothesis of no relevant deterministic component is rejected, no conclusion can be made for $\widehat{\delta}$ unless we have some preliminary information on $\phi - \delta$. Of course, by introducing trimming in the definition of $\widetilde{\delta}^{\dagger}$ it would also be possible to test if a certain trim is sufficient to eliminate the effect of the unobserved deterministic component.

7. Our Monte Carlo exercise confirmed the remark of Hurvich *et al.* (1998) that trimming increases the variance above the measure indicated by the asymptotic theory. This in turn inflates the size of the tests, in some cases quite above the level desired by the researcher. It could be interesting to see if bootstrapping the critical value improves the small sample performance.

8. We did not consider tapering, despite its explicit treatment given by Velasco (1999b). Unfortunately, the tapers he considered only remove particular trends such as t, t^2, ... so they would not be very interesting in the more general framework that we intend to discuss. A combination of trimming and tapering may nonetheless be helpful: consider for example the cosine bell taper

$$h_t = \frac{1}{2}(1 - 2\cos\frac{2\pi t}{n}), \tag{2.58}$$

with $\sum_{t=1}^{n} h_t^2 = (3/8)n$: in this case the tapered Fourier transform can be written as

$$F_s^T(\lambda_j) = \frac{1}{\sqrt{6}}(-F_s(\lambda_{j-1}) + 2F_s(\lambda_j) - F_s(\lambda_{j+1})) \tag{2.59}$$

and, using the mean value theorem twice, the tapered periodogram of

$s(t^{\phi-1/2})$, $I_s^T(\lambda_j)$, can be approximated, for $j > 0$, as $j/n \to \infty$, as

$$I_s^T(\lambda_j) \sim K\lambda_j^{-2\phi} j^{-5} \tag{2.60}$$

when $\phi < 1/2$ and as

$$I_s^T(\lambda_j) \sim K\lambda_j^{-2\phi} j^{-6+2\phi} \tag{2.61}$$

for larger ϕ, so much less trimming should be required.

2.6 Appendix to Chapter 2

We present the proofs of the theorems in the first subsection; some technical lemmas which we used in the arguments are discussed in the second subsection.

2.6.1 Proofs of the theorems

Proof of Theorem 2.1. Part *(i)*, shifts in the mean.

To prove (2.6) rewrite $\sum_{t=1}^n s_t e^{i\lambda t}$ when $|\lambda| \in (0, \pi)$ as

$$\sum_{t=1}^{n} s_t e^{i\lambda t} = \mu_1 \sum_{s=1}^{[\tau n]} e^{i\lambda s} + \mu_2 \sum_{s=[\tau n]+1}^{n} e^{i\lambda s} = (\mu_1 - \mu_2) \sum_{s=1}^{[\tau n]} e^{i\lambda s} + \mu_2 \sum_{s=1}^{n} e^{i\lambda s} = O(\lambda^{-1}) \tag{2.62}$$

where we used $\sum_{r=s}^{t} e^{i\lambda r} \leq \frac{C(t-s)}{1+(t-s)\lambda}$.

The periodogram on $|\lambda| \in (0, \pi)$ is then bounded as $I_s(\lambda) \leq \frac{C}{n} |\lambda|^{-2}$, and (2.7) follows replacing λ with $2\pi j/n$.

Next, we approximate, for $j > 0$,

$$\left(\frac{2\pi}{n}\right) \sum_{t=1}^{[\tau n]} \cos(\lambda_j t) \to \int_0^{\tau 2\pi} \cos jx dx \tag{2.63}$$

as $j/n \to 0$, and similarly $\left(\frac{2\pi}{n}\right)\sum_{t=1}^{[\tau n]}\sin(\lambda_j t) \to \int_0^{\tau 2\pi}\sin jxdx$. Then, as $n \to \infty$

$$
\begin{aligned}
& \frac{1}{(\mu_2-\mu_1)^2}(2\pi)^3 n^{-1} I_s(\lambda_j) \\
\to & \left(\int_0^{\tau 2\pi}\cos jxdx - i\int_0^{\tau 2\pi}\sin jxdx\right)\left(\int_0^{\tau 2\pi}\cos jxdx + i\int_0^{\tau 2\pi}\sin jxdx\right) \\
= & \left(\int_0^{\tau 2\pi}\cos jxdx\right)^2 + \left(\int_0^{\tau 2\pi}\sin jxdx\right)^2 \\
= & \left(\left[j^{-1}\sin jx\right]_0^{\tau 2\pi}\right)^2 + \left(\left[-j^{-1}\cos jx\right]_0^{\tau 2\pi}\right)^2 \\
= & j^{-2}\left[\sin^2 j\tau 2\pi + (1-\cos j\tau 2\pi)^2\right] \\
= & j^{-2}(2-2\cos j\tau 2\pi) = j^{-2}4\sin^2 j\tau\pi. \qquad (2.64)
\end{aligned}
$$

Part *(ii)*, fractional trend.

Consider the case $\phi \in (-1/2, 1/2)$ first.

The bound (2.9) follows replacing $2\pi j/n$ (with $j > 0$) in λ in $I_s(\lambda) \leq C/n\,|\lambda|^{-2\phi-1}$ in theorem 1 of Robinson and Marinucci (2000).

To prove (2.10) we use

$$\lim_{n\to\infty}\sum_{t=1}^{n} t^{\phi-1/2}\cos(\lambda t) \sim \Gamma(\phi+1/2)\sin\frac{(\phi+1/2)\pi}{2}\lambda^{-(\phi+1/2)} \text{ as } \lambda \to 0^+ \qquad (2.65)$$

and similarly

$$\lim_{n\to\infty}\sum_{t=1}^{n} t^{\phi-1/2}\sin(\lambda t) \sim \Gamma(\phi+1/2)\cos\frac{(\phi+1/2)\pi}{2}\lambda^{-(\phi+1/2)} \text{ as } \lambda \to 0^+ \qquad (2.66)$$

(see for example Zygmund, 1988, p. 70). Combining the two, and computing the periodogram at λ_j, for $j > 0$,

$$n^{-2\phi} I_s(\lambda_j) \to \mu_3\left(\Gamma(\phi+1/2)\right)^2 (2\pi)^{1-2\phi} j^{-2\phi-1} \text{ as } j/n \to 0. \qquad (2.67)$$

Next, $\phi = 1/2$, for which it is very well known that $\sum_{t=1}^{n} e^{-i\lambda_j t} = 0$ for $j \neq 0, n$.

Next, $\phi \in (1/2, 3/2)$. The bound (2.12) follows using the same remark we made for (2.9). For the remaining bound, using summation by parts,

$$\sum_{t=1}^{n} t^{\phi-1/2}\cos(\lambda_j t) = \sum_{t=1}^{n-1}\left\{t^{\phi-1/2} - (t+1)^{\phi-1/2}\right\}\sum_{s=1}^{t}\cos(\lambda_j s) + n^{\phi-1/2}\sum_{t=1}^{n}\cos(\lambda_j t). \tag{2.68}$$

Since

$$\left(\frac{2\pi}{n}\right)\sum_{t=1}^{n}\cos(\lambda_j t) \to \int_0^{2\pi}\cos jx dx = \left|j^{-1}\sin jx\right|_0^{2\pi} = 0 \text{ as } n \to \infty, \tag{2.69}$$

then $n^{\phi-1/2}\sum_{t=1}^{n}\cos(\lambda_j t) = o\left(n^{\phi+1/2}\right)$, while

$$\left(\frac{2\pi}{n}\right)\sum_{s=1}^{t}\cos(\lambda_j s) = j^{-1}\sin j\frac{2\pi t}{n} + o(1) \tag{2.70}$$

using integral approximation; also, using a second order expansion,

$$\sum_{t=1}^{n-1}\left\{\left(\frac{t}{n}\right)^{\phi-1/2} - \left(\frac{t+1}{n}\right)^{\phi-1/2}\right\} = \left\{\sum_{t=1}^{n-1}\frac{1}{n}\left(\frac{t}{n}\right)^{\phi-3/2} + \left(\frac{1}{n}\right)^2\left(\frac{t_{m_t}}{n}\right)^{\phi-3/2}\right\} \tag{2.71}$$

where $t_{m_t} \in [t, t+1]$ follows from the application of the mean value theorem and it may be different for each t. Then,

$$\begin{aligned} & n^{\phi-1/2}\sum_{t=1}^{n-1}\frac{1}{n}\left(\frac{t}{n}\right)^{\phi-3/2}\sum_{s=1}^{t}\cos(\lambda_j s) \\ = \; & \frac{n}{2\pi}\sum_{t=1}^{n-1} t^{\phi-3/2}\left(j^{-1}\sin(\lambda_j t) + o(1)\right) = O_e\left(n^{\phi+1/2}j^{-\phi-1/2}\right) \end{aligned} \tag{2.72}$$

making use of (2.65). Finally, for the remainder,

$$n^{\phi-1/2}\sum_{t=1}^{n-1}\left(\frac{1}{n}\right)^2\left(\frac{t_{m_t}}{n}\right)^{\phi-5/2}\sum_{s=1}^{t}\cos(\lambda_j s) \leq C\sum_{s=1}^{t} t^{\phi-5/2}t = O(1), \tag{2.73}$$

so the order is lower.

For the complex part,

$$\sum_{t=1}^{n} t^{\phi-1/2}\sin(\lambda_j t) = \sum_{t=1}^{n-1}\left\{t^{\phi-1/2} - (t+1)^{\phi-1/2}\right\}\sum_{s=1}^{t}\sin(\lambda_j s) + n^{\phi-1/2}\sum_{t=1}^{n}\sin(\lambda_j t) \tag{2.74}$$

and $n^{\phi-1/2}\sum_{t=1}^{n}\sin(\lambda_j t) = o\left(n^{\phi+1/2}\right)$ using the same argument as in (2.69). For the first term

$$\left(\frac{2\pi}{n}\right)\sum_{s=1}^{t}\sin(\lambda_j s) = -j^{-1}\cos j\frac{2\pi t}{n} + j^{-1} + o(1), \tag{2.75}$$

and

$$n^{\phi-1/2}\sum_{t=1}^{n-1}\left(t^{\phi-1/2} - (t+1)^{\phi-1/2}\right)\left(-j^{-1}\cos(\lambda_j t)\right) = O\left(n^{\phi+1/2}j^{-\phi-1/2}\right) \tag{2.76}$$

as (2.72), but the other term has a different order: using the expansion (2.71) again,

$$\begin{aligned} & n^{\phi-1/2}\sum_{t=1}^{n-1}\left(\left(\frac{t}{n}\right)^{\phi-1/2} - \left(\frac{t+1}{n}\right)^{\phi-1/2}\right)\sum_{s=1}^{t} j^{-1} \\ = \; & n^{\phi-1/2}\sum_{t=1}^{n-1}\frac{1}{n}\left(\frac{t}{n}\right)^{\phi-3/2} j^{-1}t + O\left(j^{-1}\right) = O_e\left(j^{-1}n^{\phi+1/2}\right). \end{aligned} \tag{2.77}$$

The result then follows directly from applying (2.77) in the formula of the periodogram.

Next, $\phi = 3/2$: using integral approximation again,

$$\left(\frac{2\pi}{n}\right)^2\sum_{t=1}^{n} t\cos(\lambda_j t) = o(1) \tag{2.78}$$

and, for $j > 0$,as $n \to \infty$,

$$\left(\frac{2\pi}{n}\right)^2 \sum_{t=1}^{n} t \sin(\lambda_j t) \to \int_0^{2\pi} x \sin jx dx$$

$$= -\left|j^{-1} x \cos x\right|_0^{2\pi} - j^{-1} \int_0^{2\pi} \cos jx dx = j^{-1} \tag{2.79}$$

so again the result follows from the application of the formula of the periodogram.

Part *(iii)*, single impulses. In that case the results follows from

$$\sum_{t=1}^{n} \Delta s_t(\mu) \cos(\lambda t) = \mu_4 \cos(\lambda[\tau n]) \tag{2.80}$$

and

$$\sum_{t=1}^{n} \Delta s_t(\mu) \sin(\lambda t) = \mu_4 \sin(\lambda[\tau n]), \tag{2.81}$$

so the periodogram is $\frac{(\mu_4)^2}{2\pi n}$ at all the frequencies.

Proof of Theorem 2.2. In this and in the following proofs of this chapter we replace the scaling factor $m-l+1$ in the loss function with m: since $\frac{m}{m-l+1} \to 1$ as $n \to \infty$, this replacement does not affect the asymptotic properties but it saves space in the presentation. Also notice that, because $l/m \to 0$, dropping the frequencies $\lambda_1,...,\lambda_{l-1}$ does not affect the proofs of Robinson, so we can refer to them even if in the original paper only $l = 1$ was considered.

We follow the same argument of Robinson (1995b), replacing $H = d + 1/2$, $H_0 = \delta + 1/2$. Let $\Theta_1 = \{d : \Delta \leq d \leq \Delta_2\}$ where $\Delta = \Delta_1$ when $\delta < 1/2 + \Delta_1$ and $\delta \geq \Delta > \delta - 1/2$ otherwise; when $\Delta > \Delta_1$, define $\Theta_2 = \{d : \Delta_1 \leq d < \Delta\}$, and otherwise take Θ_2 to be empty. Robinson showed that

$$P\left(\left|\widehat{\delta} - \delta\right| \geq \iota\right) \leq P\left(\inf_{\overline{N}_\iota \cap \Theta_1} S(d) \leq 0\right) + P\left(\inf_{\Theta_2} S(d) \leq 0\right) \tag{2.82}$$

where $\overline{N}_\iota = (-\infty, \infty) - N_\iota$ and $N_\iota = (d : |d - \delta| < \iota)$, and $S(d) = R(d) - R(\delta)$, choosing ι so that $\sup_{\varepsilon \in (0,\iota)} \frac{1}{m} \sum_{j=l}^{m} \lambda_j^{-2\delta+\varepsilon} I_s(\lambda_j) = o(1)$ (notice that this exists, as we show in Lemma 2.B.2 *(i)*).

Next rewrite

$$S(d) = \ln\left(\frac{\widehat{G}(d)}{G(d)} \frac{G(d)}{G} \frac{G}{\widehat{G}(\delta)}\right) - 2(d - \delta) \frac{1}{m} \sum_{j=l}^{m} \ln \lambda_j \tag{2.83}$$

where

$$G(d) = \frac{1}{m} \sum_{j=l}^{m} \lambda_j^{2d} g(\lambda_j) \text{ with } g(\lambda_j) = G\lambda_j^{-2\delta} + I_s(\lambda_j) \tag{2.84}$$

(notice the difference in the definition of $g(\lambda_j)$ with respect to Robinson, in order to take the deterministic component into account too).

Following Robinson, for $d \in N_\iota$,

$$\begin{aligned} & \ln\left(\frac{\frac{1}{m}\sum_{j=l}^{m} \lambda_j^{2d} g(\lambda_j)}{G}\right) - 2(d - \delta) \frac{1}{m} \sum_{j=l}^{m} \ln \lambda_j \\ = \; & 2(d - \delta) - \ln(2(d - \delta) + 1) + o(1), \end{aligned} \tag{2.85}$$

where we used Lemma 2.B.2 *(i)* again; since

$$\ln\left(\frac{\frac{1}{m}\sum_{j=l}^{m} \lambda_j^{2d} g(\lambda_j)}{G}\right) \geq \ln\left(\frac{\frac{1}{m}\sum_{j=l}^{m} \lambda_j^{2d} G\lambda_j^{-2\delta}}{G}\right) \tag{2.86}$$

it also follows that, in general,

$$\begin{aligned} & \ln\left(\frac{\frac{1}{m}\sum_{j=l}^{m} \lambda_j^{2d} g(\lambda_j)}{G}\right) - 2(d - \delta) \frac{1}{m} \sum_{j=l}^{m} \ln \lambda_j \\ \geq \; & 2(d - \delta) - \ln(2(d - \delta) + 1) + o(1), \end{aligned} \tag{2.87}$$

and

$$2(d - \delta) - \ln(2(d - \delta) + 1) > 0 \text{ for } d \in \overline{N}_\iota. \tag{2.88}$$

Consistency then follows from showing

$$\sup_{d\in\Theta_1}\left|\frac{\widehat{G}(d)-G(d)}{G(d)}\right|=o_p(1) \tag{2.89}$$

and

$$\left|\frac{\widehat{G}(\delta)-G}{G}\right|=o_p(1). \tag{2.90}$$

The limit (2.89) follows from

$$\begin{aligned}&\sup_{d\in\Theta_1}\left|\frac{\widehat{G}(d)-G(d)}{G(d)}\right|\\ \leq\ &\sup_{d\in\Theta_1}\left|\frac{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}\left(I_\xi(\lambda_j)-G\lambda_j^{-2\delta}\right)}{G(d)}\right|+2\sup_{d\in\Theta_1}\left|\frac{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}\operatorname{Re}\left(I_{s\xi}(\lambda_j)\right)}{G(d)}\right|\end{aligned} \tag{2.91}$$

where we used $I_x(\lambda_j)=I_\xi(\lambda_j)+I_{\xi s}(\lambda_j)+I_{s\xi}(\lambda_j)+I_s(\lambda_j)$. For the first term in the upper bound (2.91),

$$\sup_{d\in\Theta_1}\left|\frac{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}\left(I_\xi(\lambda_j)-G\lambda_j^{-2\delta}\right)}{G(d)}\right|\leq\sup_{d\in\Theta_1}\left|\frac{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}\left(I_\xi(\lambda_j)-G\lambda_j^{-2\delta}\right)}{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}G\lambda_j^{-2\delta}}\right|, \tag{2.92}$$

which is $o_p(1)$ following the same argument of Robinson. The second term in the upper bound (2.91) is $o_p(1)$ from Lemma 2.B.2 *(ii)*.

To show (2.90),

$$\begin{aligned}\left|\frac{\widehat{G}(\delta)-G}{G}\right|\leq&\left|\frac{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2\delta}I_\xi(\lambda_j)-G}{G}\right|\\ &+\left|\frac{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2\delta}I_s(\lambda_j)}{G}\right|+2\left|\frac{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2\delta}\operatorname{Re}\left(I_{s\xi}(\lambda_j)\right)}{G}\right|.\end{aligned} \tag{2.93}$$

Following Robinson, it is immediate to show that the first term is $o_p(1)$. The

second term is $o(1)$ too from Lemma 2.B.2 *(i)*, while the third one is $o_p(1)$ from Lemma 2.B.2 *(ii)*.

To discuss the set Θ_2 in case $\Delta > \Delta_1$, we rearrange (3.21) of Robinson as

$$P\left(\inf_{\Theta_2} S(d) \leq 0\right) \leq P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j - 1)G^{-1}\lambda_j^{2\delta}I_x(\lambda_j) \leq 0\right) \tag{2.94}$$

with

$$a_j = \begin{cases} \left(\frac{j}{p}\right)^{2(\Delta-\delta)} & l \leq j \leq p \\ \left(\frac{j}{p}\right)^{2(\Delta_1-\delta)} & p < j \leq m \end{cases} \tag{2.95}$$

and

$$p = \exp(\frac{1}{m}\sum_{j=l}^{m}\ln j) \text{ so that } p \sim m/e \text{ as } m \to \infty. \tag{2.96}$$

Following Robinson, $\frac{1}{m}\sum_{j=l}^{p} a_j \sim \frac{m/e}{2(\Delta-\delta)+1}$ as $m \to \infty$ (the fact that the summation starts in l rather than in 1 does not matter as long as $l/p \to 0$, and this is indeed the case because $p/m \sim 1/e$ as $m \to \infty$), so

$$\frac{1}{m}\sum_{j=l}^{m}(a_j - 1) \geq \frac{1}{e\left(2\left(\Delta - \delta\right) + 1\right)} - 1 > 1 \tag{2.97}$$

choosing $\Delta < \delta - 1/2 + 1/(4e)$, there is $\iota > 0$ such that

$$\frac{1}{m}\sum_{j=l}^{m}(a_j - 1) \geq 1 + \iota, \tag{2.98}$$

thus strengthening slightly the original result.

We then rewrite the bound in (2.94) as

$$P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j - 1)G^{-1}\lambda_j^{2\delta}I_x(\lambda_j) \leq 0,\right.$$
$$\left.\frac{1}{m}\sum_{j=l}^{m}(a_j - 1)G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j) + 2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right) < -\iota\right) \tag{2.99}$$

$$+P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j-1)G^{-1}\lambda_j^{2\delta}I_x(\lambda_j)\leq 0,\right.$$
$$\left.\frac{1}{m}\sum_{j=l}^{m}(a_j-1)G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right)\geq -\iota\right). \quad (2.100)$$

Clearly, (2.99) can be bounded by

$$P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j-1)G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right)<-\iota\right), \quad (2.101)$$

and, taking $\varepsilon<\iota$, this is

$$P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j-1)G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right)<-\iota,\right.$$
$$\left.\left|\frac{1}{m}\sum_{j=l}^{m}G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right)\right|\leq\varepsilon\right) \quad (2.102)$$

$$+P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j-1)G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right)<-\iota,\right.$$
$$\left.\left|\frac{1}{m}\sum_{j=l}^{m}G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right)\right|>\varepsilon\right) \quad (2.103)$$

$$\leq P\left(\frac{1}{m}\sum_{j=l}^{m}(a_jG^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right)<\varepsilon-\iota\right) \quad (2.104)$$

$$+P\left(\left|\frac{1}{m}\sum_{j=l}^{m}G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right)\right|>\varepsilon\right) \quad (2.105)$$

where (2.105) goes to zero because

$$\begin{aligned} & \left| \frac{1}{m} \sum_{j=l}^{m} G^{-1} \lambda_j^{2\delta} \left(I_s(\lambda_j) + 2 \operatorname{Re} \left(I_{s\xi} \left(\lambda_j \right) \right) \right) \right| \\ \leq & \left| \frac{1}{m} \sum_{j=l}^{m} G^{-1} \lambda_j^{2\delta} I_s(\lambda_j) \right| + \left| \frac{2}{m} \sum_{j=l}^{m} G^{-1} \lambda_j^{2\delta} \operatorname{Re} \left(I_{s\xi} \left(\lambda_j \right) \right) \right|. \end{aligned} \tag{2.106}$$

and each term in (2.106) is smaller than $\varepsilon/2$ for n large enough applying Lemma 2.B.2 *(i)* and *(ii)*. We discuss (2.104) rearranging the argument as

$$\frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} 2 \operatorname{Re} I_{s\xi} \left(\lambda_j \right) < \varepsilon - \iota - \frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} I_s(\lambda_j) \tag{2.107}$$

and then as

$$\frac{\frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} 2 \operatorname{Re} I_{s\xi} \left(\lambda_j \right)}{1 + \frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} I_s(\lambda_j)} < \frac{\varepsilon - \iota - \frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} I_s(\lambda_j)}{1 + \frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} I_s(\lambda_j)}. \tag{2.108}$$

Since $\frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} I_s(\lambda_j) \geq 0$, (2.104) can then be bounded as

$$P \left(\frac{\frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} 2 \operatorname{Re} I_{s\xi} \left(\lambda_j \right)}{1 + \frac{1}{m} \sum_{j=l}^{m} a_j G^{-1} \lambda_j^{2\delta} I_s(\lambda_j)} < \varepsilon - \iota \right) \tag{2.109}$$

which goes to zero using Lemma 2.B.2 *(iii)*.

We characterise (2.100) noticing that

$$\frac{1}{m} \sum_{j=l}^{m} (a_j - 1) G^{-1} \lambda_j^{2\delta} I_x(\lambda_j) \leq 0 \tag{2.110}$$

is equivalent to

$$\frac{1}{m}\sum_{j=l}^{m}(a_j-1)G^{-1}\lambda_j^{2\delta}I_\xi(\lambda_j) \leq -\frac{1}{m}\sum_{j=l}^{m}(a_j-1)G^{-1}\lambda_j^{2\delta}\left(I_s(\lambda_j)+2\,\mathrm{Re}\left(I_{s\xi}\left(\lambda_j\right)\right)\right) \tag{2.111}$$

so (2.100) can be bounded by

$$P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j-1)G^{-1}\lambda_j^{2\delta}I_\xi(\lambda_j) \leq \iota\right). \tag{2.112}$$

Rearranging the original argument of Robinson:

$$\begin{aligned} & P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j-1)\left(G^{-1}\lambda_j^{2\delta}I_\xi(\lambda_j)-1+1\right) \leq \iota\right) \\ = \; & P\left(\frac{1}{m}\sum_{j=l}^{m}(a_j-1)\left(G^{-1}\lambda_j^{2\delta}I_\xi(\lambda_j)-1\right)+\frac{1}{m}\sum_{j=l}^{m}(a_j-1) \leq \iota\right) \\ \leq \; & P\left(\left|\frac{1}{m}\sum_{j=l}^{m}(a_j-1)\left(G^{-1}\lambda_j^{2\delta}I_\xi(\lambda_j)-1\right)\right| \geq 1\right) \to 0. \end{aligned} \tag{2.113}$$

Proof of Theorem 2.3. As for Theorem 2.2, the proof follows the one in Robinson (1995b). Using the mean value theorem,

$$0 = \left.\frac{dR(d)}{d\,d}\right|_{\widehat{\delta}} = \left.\frac{dR(d)}{d\,d}\right|_{\delta} + \left.\frac{d^2R(d)}{d\,d^2}\right|_{\delta_m}\left(\widehat{\delta}-\delta\right) \tag{2.114}$$

where δ_m is such that $|\delta_m-\delta| \leq \left|\widehat{\delta}-\delta\right|$. The proof that $\left.\frac{d^2R(d)}{d\,d^2}\right|_{\delta_m} \to_p 4$ follows the same argument of Robinson, once again replacing $g\left(\lambda_j\right) = G\lambda^{-2\delta} + I_s\left(\lambda_j\right)$ as in (2.84). The proof proceeds as in the original paper: equation (4.6) of Robinson requires

$$\sup_{\Theta_1\cap N_\iota}\left|\frac{\widehat{G}\left(d\right)-G\left(d\right)}{G\left(d\right)}\right| = o_p\left(\ln m^{-6}\right) \tag{2.115}$$

and

$$\left|\frac{\widehat{G}(\delta)-G}{G}\right| = o_p\left(\ln m^{-6}\right). \tag{2.116}$$

Using the same arguments as in Theorem 2.2, (2.115) is bounded by

$$\sup_{\Theta_1\cap N_\iota}\left|\frac{\sum_{j=l}^m \lambda_j^{2d} I_\xi(\lambda_j)}{G\sum_{j=l}^m \lambda_j^{2d}\lambda_j^{-2\delta}} - 1\right| + 2\sup_{\Theta_1\cap N_\iota}\left|\frac{\sum_{j=l}^m \lambda_j^{2d}\,\mathrm{Re}\left(I_{s\xi}(\lambda_j)\right)}{\sum_{j=l}^m \lambda_j^{2d} g(\lambda_j)}\right| : \tag{2.117}$$

Robinson already showed that the first element of (2.117) is $o_p\left(\ln m^{-6}\right)$, while the second one is $o_p\left(\ln m^{-6}\right)$ according to Lemma 2.B.3 *(i)*. To show (2.116) use the upper bound of the expression (2.93) again: the first element of the bound is $o_p\left(\ln m^{-6}\right)$ following Robinson; the second bound is $o\left(\ln m^{-6}\right)$ using Lemma 2.B.3 *(ii)* while the last one is $o_p\left(\ln m^{-6}\right)$ noticing that noticing that $\sum_{j=l}^m \lambda_j^{2\delta} I_s(\lambda_j) = o(1)$ and then applying Lemma 2.B.3 *(i)*: this holds for the supremum for $\varepsilon \in (-\iota, \iota)$, so it also holds for $\varepsilon = 0$ in particular.

Finally, using $\frac{1}{m}\sum_{j=l}^m \lambda_j^{-2\delta} I_x(\lambda_j) \to_p G$, we consider a normalization of $\left.\frac{dR(d)}{d\,d}\right|_\delta$,

$$\begin{aligned} m^{-1/2}\left.\frac{dR(d)}{d\,d}\right|_\delta &= 2m^{-1/2}\sum_{j=l}^m \nu_j \frac{\lambda_j^{2\delta} I_x(\lambda_j)}{G+o_p(1)} \\ &= 2m^{-1/2}\sum_{j=l}^m \nu_j \frac{\lambda_j^{2\delta} I_\xi(\lambda_j)}{G+o_p(1)} \\ &\quad +4m^{-1/2}\sum_{j=l}^m \nu_j \frac{\lambda_j^{2\delta}\,\mathrm{Re}\left(I_{s\xi}(\lambda_j)\right)}{G+o_p(1)} \\ &\quad +2m^{-1/2}\sum_{j=l}^m \nu_j \frac{\lambda_j^{2\delta} I_s(\lambda_j)}{G+o_p(1)}. \end{aligned} \tag{2.118}$$

Robinson showed that $2m^{-1/2}\sum_{j=l}^m \nu_j \frac{\lambda_j^{2\delta} I_\xi(\lambda_j)}{G+o_p(1)} \to_d N(0,4)$; Lemma 2.B.3 *(iii)* and Lemma 2.B.3 *(iv)* are sufficient to prove that the remainder is negligible.

Proof of Theorem 2.4. The result follows from computing the first order

expansion for $\widehat{\delta}-\delta$ based on the mean value theorem as in Theorem 2.3. Recall that $\left.\frac{d^2R(d)}{d\,d^2}\right|_{\delta_m} \to_p 4$; to find out the term with largest order of magnitude, we discuss $-\left.\frac{dR(d)}{d\,d}\right|_{\delta}$: assume for simplicity that $c_1 > 0$, $c_2 = 0$ in Assumption A.3: as $n \to \infty$,

$$\begin{aligned} & -2m^{-1}\sum_{j=l}^{m}\nu_j\frac{\lambda_j^{2\delta}I_s(\lambda_j)}{G+o_p(1)} \\ \sim & -\frac{2c_1(2\pi)^{2(\delta-\phi)}}{G}m^{-1}n^{2(\phi-\delta)}\left(\sum_{j=l}^{m}\ln j\; j^{2(\delta-\phi)-1}-(\ln m-1)\,j^{2(\delta-\phi)-1}\right) \\ \sim & \frac{2c_1(2\pi)^{2(\delta-\phi)}}{G}m^{-1}n^{2(\phi-\delta)}\frac{l^{2(\delta-\phi)}}{2(\phi-\delta)}\ln m, \end{aligned} \tag{2.119}$$

which is positive recalling that $\delta-\phi<0$. The term $\frac{l^{2(\delta-\phi)}}{2(\phi-\delta)}$ is replaced by $\sum_{j=1}^{\infty}j^{2(\delta-\phi)-1}$ when $l=1$. The bias for the other combinations of c_1, c_2 can be treated in the same way.

Proof of Theorem 2.5. All the results follows as in Theorems 2.2 to 2.4. Just notice, for the limit distribution, that $m/2$ frequencies are used, so $\sqrt{m/2}\left(\widetilde{\delta}^{\dagger}-\delta\right)\to_d N(0,1/4)$ and then $\sqrt{m}\left(\widetilde{\delta}^{\dagger}-\delta\right)\to_d N(0,1/2)$.

Proof of Theorem 2.6. The limit normality follows from the fact both $\widehat{\delta}^{(l)}$ and $\widetilde{\delta}^{\dagger}$ are asymptotically normally distributed. We also already have

$$\lim_{n\to\infty} m\;Var(\widehat{\delta}^{(l)})=1/4,\;\lim_{n\to\infty} m\;Var(\widetilde{\delta}^{\dagger})=1/2. \tag{2.120}$$

For the asymptotic covariance, notice that

$$
\begin{aligned}
& m\left(\widehat{\delta}^{(l)}-\delta\right)\left(\widehat{\delta}^{\dagger}-\delta\right) \\
= & \left(\frac{1}{4+o_p(1)}2\frac{1}{\sqrt{m}}\sum_{j=l}^{m}\nu_j\frac{\lambda_j^{-2\delta}I_\xi(\lambda_j)}{G(1+o_p(1))}\right) \\
& \times\left(2\frac{1}{\sqrt{m/2}}\sum_{j=0}^{m/2-1}\nu_{2j+1}\frac{\lambda_{2j+1}^{-2\delta}I_\xi(\lambda_{2j+1})}{G(1+o_p(1))}\frac{1}{4+o_p(1)}\right) \\
= & \left(\frac{1}{2}\frac{1}{\sqrt{2}\sqrt{m/2}}\sum_{j=l}^{m}\nu_j\frac{\lambda_j^{-2\delta}I_\xi(\lambda_j)}{G(1+o_p(1))}\right)\left(\frac{\sqrt{2}}{\sqrt{m/2}}\sum_{j=0}^{m/2-1}\nu_{2j+1}\frac{\lambda_{2j+1}^{-2\delta}I_\xi(\lambda_{2j+1})}{G(1+o_p(1))}\frac{1}{2}\right).
\end{aligned}
\tag{2.121}
$$

Rewriting (for l even)

$$
\frac{1}{\sqrt{m/2}}\sum_{j=l}^{m}\nu_j\frac{\lambda_j^{-2\delta}I_\xi(\lambda_j)}{G(1+o_p(1))} \tag{2.122}
$$

$$
= \frac{1}{\sqrt{m/2}}\sum_{j=l/2}^{m/2}\nu_{2j}\frac{\lambda_{2j}^{-2\delta}I_\xi(\lambda_{2j})}{G(1+o_p(1))} \tag{2.123}
$$

$$
+\frac{1}{\sqrt{m/2}}\sum_{j=l/2}^{m/2-1}\nu_{2j+1}\frac{\lambda_{2j+1}^{-2\delta}I_\xi(\lambda_{2j+1})}{G(1+o_p(1))} \tag{2.124}
$$

(2.121) is

$$
\left(\frac{1}{2}\frac{1}{\sqrt{m/2}}\sum_{j=l/2}^{m/2}\nu_{2j}\frac{\lambda_{2j}^{-2\delta}I_\xi(\lambda_{2j})}{G(1+o_p(1))}\right)\left(\frac{1}{\sqrt{m/2}}\sum_{j=1/2}^{m/2}\nu_{2j}\frac{\lambda_{2j}^{-2\delta}I_\xi(\lambda_{2j})}{G(1+o_p(1))}\frac{1}{2}\right) \tag{2.125}
$$

$$
+\left(\frac{1}{2}\frac{1}{\sqrt{m/2}}\sum_{j=l/2}^{m/2-1}\nu_{2j+1}\frac{\lambda_{2j+1}^{-2\delta}I_\xi(\lambda_{2j+1})}{G(1+o_p(1))}\right)\left(\frac{1}{\sqrt{m/2}}\sum_{j=1/2}^{m/2}\nu_{2j}\frac{\lambda_{2j}^{-2\delta}I_\xi(\lambda_{2j})}{G(1+o_p(1))}\frac{1}{2}\right) \tag{2.126}
$$

and since

$$
\frac{1}{\sqrt{m/2}}\sum_{j=l/2}^{m/2-1}\nu_{2j+1}\frac{\lambda_{2j+1}^{-2\delta}I_\xi(\lambda_{2j+1})}{G(1+o_p(1))}\rightarrow_d N(0,1) \tag{2.127}
$$

the product of the two summations in (2.126) converges in distribution to $\frac{1}{4}\chi_1^2$, which is a random variable with variance $1/4$.

On the other hand the two summations in (2.125) converge to two independent normals. We can see this noticing that the expression in (2.122) converges to a $N(0,2)$, and since both (2.123) and (2.124) have asymptotically variance 1, then they are asymptotically uncorrelated. Therefore,

$$\lim_{n\to\infty} m\ Var(\widetilde{\delta}^{\dagger} - \widehat{\delta}^{(l)}) = 1/4 + 1/2 - 2(1/4) = 1/4. \tag{2.128}$$

Proof of Theorem 2.7. Consider the second order expansion

$$0 = \left.\frac{dR(d)}{d\,d}\right|_{\widehat{\delta}} = \left.\frac{dR(d)}{d\,d}\right|_{\delta} + \left.\frac{d^2R(d)}{d\,d^2}\right|_{\delta}\left(\widehat{\delta}-\delta\right) + \frac{1}{2}\left.\frac{d^3R(d)}{d\,d^3}\right|_{\delta_m}\left(\widehat{\delta}-\delta\right)^2 \tag{2.129}$$

where δ_m is such that $|\delta_m - \delta| \leq \left|\widehat{\delta}-\delta\right|$. Taking another derivative in equation (4.3) of Robinson,

$$\frac{d^3R(d)}{d\,d^3} = \frac{4\left[\widehat{F}_3(d)\,\widehat{F}_0^3(d) - 3\widehat{F}_2(d)\,\widehat{F}_1(d)\,\widehat{F}_0^2(d) + 2\widehat{F}_0(d)\,\widehat{F}_1^3(d)\right]}{\widehat{F}_0^4(d)} \tag{2.130}$$

where

$$\widehat{F}_k(d) = \frac{1}{m}\sum_{j=l}^{m}(\ln j)^k\,\lambda_j^{2d} I_\xi(\lambda_j) \tag{2.131}$$

and noticing that

$$\left|\widehat{F}_k(d)\right| \leq (\ln m)^k \frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} I_\xi(\lambda_j) = (\ln m)^k\,\widehat{F}_0(d) \tag{2.132}$$

then

$$\begin{aligned}\left|\frac{d^3R(d)}{d\,d^3}\right| &\leq \frac{4\left[\widehat{F}_3(d)\,\widehat{F}_0^3(d)+3\widehat{F}_2(d)\,\widehat{F}_1(d)\,\widehat{F}_0^2(d)+2\widehat{F}_0(d)\,\widehat{F}_1^3(d)\right]}{\widehat{F}_0^4(d)}\\ &\leq \frac{4\left[\ln^3 m\widehat{F}_0^4(d)+3\ln^3 m\;\widehat{F}_0^4(d)+2\ln^3 m\widehat{F}_0^4(d)\right]}{\widehat{F}_0^4(d)}\\ &\leq 24\ln^3 m = O_p\left(\ln^3 m\right). \end{aligned} \tag{2.133}$$

Notice then that this holds for any d: actually, using the fact that $\delta_m \rightarrow_p \delta$ a sharper bound could be obtained for $\frac{d^3R(d)}{d\,d^3}\Big|_{\delta_m}$, but the one in (2.133) is sufficient for our purpose so we do not discuss the case in more details. The remainder in (2.129) is bounded as

$$\frac{d^3R(d)}{d\,d^3}\Big|_{\delta_m}\left(\widehat{\delta}-\delta\right)^2 = O_p\left(\frac{\ln^3 m}{m}\right). \tag{2.134}$$

Introducing then the notation

$$R^{(l)}(d) \text{ for } R(d) \text{ when } 1/l \rightarrow 0,\; R^{(1)}(d) \text{ for } R(d) \text{ when } l=1, \tag{2.135}$$

then

$$\begin{aligned}&\widehat{\delta}^{(1)}-\widehat{\delta}^{(l)} = \widehat{\delta}^{(1)}-\delta+\delta-\widehat{\delta}^{(l)}\\ =\;& -\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(1)}(d)}{d\,d}\Big|_{\delta}-\frac{1}{2}\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\Big|_{\delta}\frac{d^3R^{(1)}(d)}{d\,d^3}\Big|_{\delta_m^{(1)}}\left(\widehat{\delta}^{(1)}-\delta\right)^2\\ &+\left(\frac{d^2R^{(l)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(l)}(d)}{d\,d}\Big|_{\delta}+\frac{1}{2}\left(\frac{d^2R^{(l)}(d)}{d\,d^2}\right)^{-1}\Big|_{\delta}\frac{d^3R^{(l)}(d)}{d\,d^3}\Big|_{\delta_m^{(l)}}\left(\delta-\widehat{\delta}^{(l)}\right)^2\\ =\;& -\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(1)}(d)}{d\,d}\Big|_{\delta}+\left(\frac{d^2R^{(l)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(l)}(d)}{d\,d}\Big|_{\delta}+O_p\left(\frac{\ln^3 m}{m}\right)\end{aligned} \tag{2.136}$$

where the expansion for $\widehat{\delta}^{(1)}-\delta$ and the one for $\delta-\widehat{\delta}^{(l)}$ are computed in two

different points $\delta_m^{(1)}$ and $\delta_m^{(l)}$ such that $\left|\delta_m^{(1)}-\delta\right| \leq \left|\widehat{\delta}^{(1)}-\delta\right|$ and $\left|\delta_m^{(l)}-\delta\right| \leq \left|\widehat{\delta}^{(l)}-\delta\right|$, but in both the cases (2.134) holds; using the fact that $\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)\Big|_\delta \to_p 4$, $\left(\frac{d^2R^{(l)}(d)}{d\,d^2}\right)\Big|_\delta \to_p 4$, the remainder is of the order stated in (2.136). Adding and subtracting $\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(l)}(d)}{d\,d}\Big|_\delta$, (2.136) is

$$-\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(1)}(d)}{d\,d}\Big|_\delta+\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(l)}(d)}{d\,d}\Big|_\delta \tag{2.137}$$

$$-\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(l)}(d)}{d\,d}\Big|_\delta+\left(\frac{d^2R^{(l)}(d)}{d\,d^2}\right)^{-1}\frac{dR^{(l)}(d)}{d\,d}\Big|_\delta \tag{2.138}$$

$$+O_p\left(\frac{\ln^3 m}{m}\right). \tag{2.139}$$

The term in (2.137) is

$$-\frac{1}{2+o_p(1)}\frac{1}{m}\sum_{j=1}^{l}\nu_j\frac{\lambda_j^{-2\delta}I_\xi(\lambda_j)}{G(1+o_p(1))}, \tag{2.140}$$

so,

$$\frac{m}{\sqrt{l}}\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\left(\frac{dR^{(l)}(d)}{d\,d}-\frac{dR^{(1)}(d)}{d\,d}\right)\Big|_\delta \to_d N\left(0,\frac{1}{4}\right). \tag{2.141}$$

To discuss (2.138), we introduce the following notation:

$$\widehat{H}_k(l)=\frac{1}{m}\sum_{j=l}^{m}(\ln j)^k\lambda_j^{2\delta}I_\xi(\lambda_j) \tag{2.142}$$

(this is the same as (2.131), but it is formulated for one given d only, and it is a function of l). The factor

$$\left(\frac{d^2R^{(l)}(d)}{d\,d^2}\right)^{-1}-\left(\frac{d^2R^{(1)}(d)}{d\,d^2}\right)^{-1}\Big|_\delta \tag{2.143}$$

is then

$$\frac{1}{4}\left(\frac{\widehat{H}_0^2(l)}{\widehat{H}_2(l)\widehat{H}_0(l)-\widehat{H}_1^2(l)}-\frac{\widehat{H}_0^2(1)}{\widehat{H}_2(1)\widehat{H}_0(1)-\widehat{H}_1^2(1)}\right). \tag{2.144}$$

Robinson already showed, in equations (4.3) to (4.10), that $\frac{\widehat{H}_2(1)\widehat{H}_0(1)-\widehat{H}_1^2(1)}{\widehat{H}_0^2(1)} = 1+o_p(1)$, and in the same way, it holds that $\frac{\widehat{H}_2(l)\widehat{H}_0(l)-\widehat{H}_1^2(l)}{\widehat{H}_0^2(l)} = 1+o_p(1)$. Since $\widehat{H}_0^2(1) = G^2(1+o_p(1))$, $\widehat{H}_0^2(l) = G^2(1+o_p(1))$, then $\widehat{H}_2(1)\widehat{H}_0(1)-\widehat{H}_1^2(1) = G^2(1+o_p(1))$ and $\widehat{H}_2(l)\widehat{H}_0(l)-\widehat{H}_1^2(l) = G^2(1+o_p(1))$. Summing the two terms in (2.144), the denominator converges to G^2, so the order of magnitude depends on the numerator. This is

$$\left(\widehat{H}_2(l)\widehat{H}_0(l)-\widehat{H}_1^2(l)\right)\widehat{H}_0^2(1)-\widehat{H}_0^2(l)\left(\widehat{H}_2(1)\widehat{H}_0(1)-\widehat{H}_1^2(1)\right). \tag{2.145}$$

Introducing $\widehat{h}_k$ such that

$$\widehat{h}_k = \widehat{H}_k(1)-\widehat{H}_k(l), \tag{2.146}$$

notice that

$$\widehat{h}_k = \frac{1}{m}\sum_{j=1}^{l-1}(\ln j)^k \lambda_j^{-2\delta} I_\xi = O_p\left(\frac{l}{m}\ln^k l\right) \text{ and } \widehat{H}_k(l) = O_p\left(\ln^k m\right). \tag{2.147}$$

Replacing $\widehat{H}_k(1)$ with $\widehat{h}_k+\widehat{H}_k(l)$ in (2.145) and simplifying terms, (2.143) can be bounded by $O_p\left(\frac{l}{m}\ln^2 m\right)$, so, since $\left.\frac{dR^{(l)}(d)}{d\,d}\right|_\delta = O\left(m^{-1/2}\right)$, the term (2.138) is $O_p\left(\frac{l}{m^{3/2}}\ln^2 m\right)$.

Taking $m = c_\kappa n^\kappa$ and $l = c_\upsilon n^\upsilon$ with $\upsilon < \kappa$ then both (2.138) and (2.139) have a smaller order of magnitude than (2.137).

2.6.2 Technical lemmas

Lemma 2.B.1.

(i) under Assumptions 2.1, 2.2 and 2.3, for $d \in [\Delta, \Delta_2]$, *where* Δ *is defined as in Theorem 2.2, for* $l > 0$,

$$\frac{1}{r}\sum_{j=l}^{r} \lambda_j^{2d} G \lambda_j^{-2\delta} = O_e\left(\left(\frac{r}{n}\right)^{2(d-\delta)}\right) \text{ as } l/r \to 0; \tag{2.148}$$

(ii) under Assumption A.3 and $l > 0$,

$$\frac{1}{r}\sum_{j=l}^{r} \lambda_j^{2d} I_s(\lambda_j) = \begin{cases} O_e\left(\frac{1}{r}\left(\frac{l}{n}\right)^{2(d-\phi)}\right) & \text{if } d < \phi \\ O_e\left(\frac{\ln r}{r}\right) & \text{if } d = \phi \\ O_e\left(\frac{1}{r}\left(\frac{r}{n}\right)^{2(d-\phi)}\right) & \text{if } d > \phi \end{cases} \tag{2.149}$$

as $l/r \to 0$;

(iii) under Assumptions A.3, 2.1, 2.2 and 2.3, for $d \in [\Delta, \Delta_2]$, *where* Δ *is defined as in Theorem 2.2,* $j > 0$,

$$|I_{s\xi}(\lambda_j)| = O_p\left(\left(\frac{j}{n}\right)^{-\phi-\delta} j^{-1/2}\right) \text{ as } j/n \to 0, \tag{2.150}$$

and, for $l > 0$,

$$\left|\frac{1}{r}\sum_{j=l}^{r} \lambda_j^{\delta+\phi} j^{1/2} I_{s\xi}(\lambda_j)\right| = O_p\left(r^{-1/2}\ln r\right) \text{ as } l/r \to 0. \tag{2.151}$$

Proof. The orders of magnitude in (2.148) and in (2.149) can be computed directly; a little remark is only needed in (2.149) when the deterministic components includes a relevant the shift in the mean. Since $\sin^2 \tau\pi j \leq 1$, it is clear that the quantity in (2.149) is an upper bound. To see that the bound is exact, notice that $\sin^2 \tau\pi j > 0$ unless τj is an integer, but can only be an

integer at most every second j (when $\tau = 1/2$), leaving still $m/2$ non-zero elements in summation.

To show (2.150), just notice that

$$|I_{s\xi}(\lambda_j)| = O_p\left(\left(E\left|I_{s\xi}(\lambda_j)\right|^2\right)^{1/2}\right) = O_p\left(\left(F_s(\lambda_j)\, E\left(F_\xi(-\lambda_j)\, F_\xi(\lambda_j)\right) F_s(-\lambda_j)\right)^{1/2}\right) \tag{2.152}$$

and the conclusion follows using Assumption A.3 and Theorem 2 of Robinson (1995a). Although the bound in (2.150) is sometimes enough for our proofs, the sharper bound in (2.151) can be derived.

Since

$$\left|\frac{1}{r}\sum_{j=l}^{r}\lambda_j^{\delta+\phi} j^{1/2} I_{s\xi}(\lambda_j)\right| = O_p\left(\left[E\left|\frac{1}{r}\sum_{j=l}^{r}\lambda_j^{\delta+\phi} j^{1/2} I_{s\xi}(\lambda_j)\right|^2\right]^{1/2}\right) \tag{2.153}$$

we start considering

$$E\left|\sum_{j=l}^{r}\lambda_j^{\delta+\phi} j^{1/2} I_{s\xi}(\lambda_j)\right|^2 = \tag{2.154}$$

$$= E\left(\sum_{j=l}^{r}\lambda_j^{2(\delta+\phi)} j F_s(\lambda_j) F_\xi(-\lambda_j) F_\xi(\lambda_j) F_s(-\lambda_j)\right) \tag{2.155}$$

$$+2E\left(\sum_{k=l}^{r}\sum_{j=l}^{k-1}\lambda_j^{\delta+\phi} j^{1/2}\lambda_k^{\delta+\phi} k^{1/2} F_s(\lambda_j) F_\xi(-\lambda_j) F_\xi(\lambda_k) F_s(-\lambda_k)\right). \tag{2.156}$$

The expectation in (2.155) is

$$\left(\sum_{j=l}^{r}\lambda_j^{2(\delta+\phi)} j F_s(\lambda_j) E\left(F_\xi(-\lambda_j) F_\xi(\lambda_j)\right) F_s(-\lambda_j)\right)$$

$$= O\left(\sum_{j=l}^{r}\left(\frac{j}{n}\right)^{2(\delta+\phi)} j\left(\frac{j}{n}\right)^{-2\delta-2\phi} j^{-1}\right) = O(r), \tag{2.157}$$

while, using Assumption A.3 to derive the bound $F_s(\lambda_j) = O\left(\lambda_j^{-\phi} j^{-1}\right)$ and Theorem 2 of Robinson (1995a) for $E\left(\lambda_j^{\delta} F_\xi(-\lambda_j) F_\xi(\lambda_k)\lambda_k^{\delta}\right)$, the expectation

in (2.156) has order of magnitude

$$O\left(\sum_{k=l}^{r}\sum_{j=l}^{k-1}\left(\frac{j}{n}\right)^{\delta+\phi} j^{1/2}\left(\frac{k}{n}\right)^{\delta+\phi} k^{1/2}\left(\frac{j}{n}\right)^{-(\delta+\phi)} j^{-1/2}\left(\frac{k}{n}\right)^{-(\delta+\phi)} k^{-1/2}\frac{\ln k}{j}\right)$$
$$= O\left(r\ln^2 r\right). \tag{2.158}$$

The orders of magnitude in (2.150) and (2.151) also holds for shifts in the mean, since $\sin^2(\tau\pi j) \leq 1$.

Lemma 2.B.2. *Under Assumptions 2.2, 2.3, and 2.5, as $n \to \infty$,*

(i) there is $\iota > 0$ such that

$$sup_{\varepsilon\in(-\iota,\iota)}\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2(\delta+\varepsilon)} I_s(\lambda_j)\right| = o(1); \tag{2.159}$$

Let $\Delta = \Delta_1$ if $\delta - \Delta_1 < 1/2$, and Δ such that $\delta - 1/2 < \Delta < \delta - 1/2 + 1/(4e)$ if $\delta - \Delta_1 \geq 1/2$. Then

(ii)

$$\sup_{d\in[\Delta,\,\Delta_2]}\left|\left(\sum_{j=l}^{m}\lambda_j^{2d}\left(G\lambda_j^{-2\delta} + I_s(\lambda_j)\right)\right)^{-1}\sum_{j=1}^{m}\lambda_j^{2d} I_{s\xi}(\lambda_j)\right| = o_p(1); \tag{2.160}$$

if $\delta - \Delta_1 \geq 1/2$, also define

$$a_j = \left(\frac{j}{p}\right)^{2(\Delta-\delta)} \quad when \quad l \leq j \leq p, \tag{2.161}$$

$$a_j = \left(\frac{j}{p}\right)^{2(\Delta_1-\delta)} \quad when \quad p < j \leq m, \tag{2.162}$$

$$where\ p = \exp\left(\frac{1}{m}\sum_{j=l}^{m}\ln j\right). \tag{2.163}$$

Then,

(iii)

$$\left|\left(1+\frac{1}{m}\sum_{j=l}^{m}a_j\lambda_j^{2\delta}I_s(\lambda_j)\right)^{-1}\frac{1}{m}\sum_{j=l}^{m}a_j\lambda_j^{2\delta}I_{s\xi}(\lambda_j)\right| = o_p(1). \tag{2.164}$$

Proof. Since $j/n < 1$,

$$sup_{\varepsilon\in(-\iota,\iota)}\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2(\delta+\varepsilon)}I_s(\lambda_j)\right| = \left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2(\delta-\iota)}I_s(\lambda_j)\right|. \tag{2.165}$$

From (2.149),

$$\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2(\delta-\iota)}I_s(\lambda_j) = \begin{cases} O_e\left(\frac{1}{m}\left(\frac{l}{n}\right)^{2(\delta-\iota-\phi)}\right) & \text{if } \delta-\iota<\phi \\ O_e\left(\frac{\ln m}{m}\right) & \text{if } \delta-\iota=\phi \\ O_e\left(\frac{1}{m}\left(\frac{m}{n}\right)^{2(\delta-\iota-\phi)}\right) & \text{if } \delta-\iota>\phi. \end{cases} \tag{2.166}$$

The result then follows from $m/n \to 0$ when $\delta \geq \phi + \iota$; the condition

$$\phi < (\delta-\iota) + \frac{1}{2}\frac{\kappa}{1-\upsilon} \tag{2.167}$$

is sufficient when $\delta - \iota < \phi$, and it can be derived from (2.22) choosing ι small enough for given κ, υ.

For 2.B.2 *(ii)* we first find the order of magnitude of the numerator:

$$\begin{aligned} &\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}I_{s\xi}(\lambda_j)\right| \\ &= O_p\left(\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}\lambda_j^{-(\delta+\phi)}j^{-1/2}\lambda_j^{\delta+\phi}j^{1/2}I_{s\xi}(\lambda_j)\right|\right) \end{aligned} \tag{2.168}$$

$$= O_p\left(n^{(\phi+\delta)-2d}m^{-1}\sum_{j=l}^{m-1}\left|j^{2d-(\delta+\phi)-1/2}-(j+1)^{2d-(\delta+\phi)-1/2}\right|\left|\sum_{r=l}^{j-1}\lambda_r^{\delta+\phi}r^{1/2}I_{s\xi}(\lambda_r)\right|\right) \tag{2.169}$$

$$+O_p\left(n^{(\phi+\delta)-2d}m^{-1}m^{2d-(\delta+\phi)-1/2}\left|\sum_{r=l}^{m}\lambda_r^{\delta+\phi}r^{1/2}I_{s\xi}(\lambda_r)\right|\right) \tag{2.170}$$

applying summation by parts.

Using an argument based on the mean value theorem, the order of magnitude in (2.169) is at most

$$O_p\left(n^{(\phi+\delta)-2d}m^{-1}\sum_{j=l}^{m-1}j^{2d-(\delta+\phi)-3/2}\left|\sum_{r=l}^{j-1}\lambda_r^{\delta+\phi}r^{1/2}I_{s\xi}(\lambda_r)\right|\right) \tag{2.171}$$

$$= \begin{cases} O_p\left(n^{(\phi+\delta)-2d}m^{-1}l^{2d-(\delta+\phi)}\ln l\right) & \text{if } 2d-(\delta+\phi)<0 \\ O_p\left(m^{-1}\ln^2 m\right) & \text{if } 2d-(\delta+\phi)=0 \\ O_p\left(n^{(\phi+\delta)-2d}m^{-1}m^{2d-(\delta+\phi)}\ln m\right) & \text{if } 2d-(\delta+\phi)>0 \end{cases} \tag{2.172}$$

while (2.170) is

$$O_p\left(n^{(\phi+\delta)-2d}m^{-1}m^{2d-(\delta+\phi)}\ln m\right). \tag{2.173}$$

The two orders of magnitude are then the same when $2d-(\delta+\phi)>0$, but (2.169) prevails in the other two cases: when $2d-(\delta+\phi)<0$ the ratio of the orders of magnitude of the two expressions is $(m/l)^{2d-(\delta+\phi)}\ (\ln m/\ln l)$ which is $o(1)$ in that situation, while when $2d-(\delta+\phi)=0$ the same ratio is $1/\ln m$. We then consider (2.169) only in the rest of the proof.

Next, we discuss the ratio

$$\frac{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}I_s(\lambda_j)}{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d}\lambda_j^{-2\delta}G} \tag{2.174}$$

which has order of magnitude

$$\begin{cases} n^{2(\phi-\delta)}m^{2(\delta-d)-1}l^{2(d-\phi)} & \text{if } d<\phi \\ n^{2(\phi-\delta)}m^{2(\delta-\phi)-1}\ln m & \text{if } d=\phi \\ n^{2(\phi-\delta)}m^{2(\delta-\phi)-1} & \text{if } d>\phi. \end{cases} \tag{2.175}$$

The assumption on υ and κ is sufficient to show that

$$n^{2(\phi-\delta)}m^{2(\delta-\phi)-1} = o(1). \tag{2.176}$$

This is immediate when $\delta \geq \phi$, while if $\delta < \phi$

$$n^{2(\phi-\delta)}m^{2(\delta-\phi)-1} < n^{2(\phi-\delta)}l^{2(\delta-\phi)}m^{-1}, \tag{2.177}$$

which is $o(1)$ using (2.22).

The ratio (2.174) is then $o(1)$ when $d \geq \phi$. If $\phi > d$, it is still $o(1)$ if

$$n^{2(\phi-\delta)}m^{2(\delta-d)-1}l^{2(d-\phi)} = o(1), \tag{2.178}$$

which corresponds to $d > \Delta(\kappa, \upsilon)$, where

$$\Delta(\kappa,\upsilon) = \delta + \frac{1-\upsilon}{\kappa-\upsilon}(\phi-\delta) - \frac{\kappa}{2(\kappa-\upsilon)}. \tag{2.179}$$

Otherwise, if $d = \Delta(\kappa, \upsilon)$ the ratio in (2.174) converges to a constant, and if $d < \Delta(\kappa, \upsilon)$, it diverges to ∞. Notice that for any eligible combination κ,υ, we have that

$$\delta > \Delta(\kappa,\upsilon) \text{ and } \phi > \Delta(\kappa,\upsilon). \tag{2.180}$$

We then introduce the bound

$$\sup_{d\in[\Delta,\Delta_2]} \frac{\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} I_{s\xi}(\lambda_j)\right|}{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} g(\lambda_j)} \tag{2.181}$$

$$\leq \max\left\{ \sup_{d\in[\Delta,\Delta_2],\, d\in[\Delta(\kappa,\upsilon),\Delta_2]} \frac{\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} I_{s\xi}(\lambda_j)\right|}{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} g(\lambda_j)}, \tag{2.182}$$

$$\sup_{d\in[\Delta,\Delta_2],\, d\in[\Delta,\Delta(\kappa,\upsilon)]} \frac{\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} I_{s\xi}(\lambda_j)\right|}{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} g(\lambda_j)} \right\}, \tag{2.183}$$

and (2.183) is set to 0 when $\Delta(\kappa,\upsilon) < \Delta$. Letting

$$\Delta_a = \max\{\Delta, \Delta(\kappa,\upsilon)\} \tag{2.184}$$

(2.182) can be bounded as

$$\sup_{d\in[\Delta_a,\Delta_2]} \frac{\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} I_{s\xi}(\lambda_j)\right|}{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} g(\lambda_j)} \leq \sup_{d\in[\Delta_a,\Delta_2]} \frac{\left|\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} I_{s\xi}(\lambda_j)\right|}{\frac{1}{m}\sum_{j=l}^{m}\lambda_j^{2d} G\lambda^{-2\delta}} \tag{2.185}$$

$$\leq \sup_{d\in[\Delta_a,\Delta_2]} \frac{Cn^{2(d-\delta)}\frac{2(d-\delta)+1}{m}\left(\frac{1}{m}\right)^{2(d-\delta)}\left|\sum_{j=l}^{m}\lambda_j^{2d} I_{s\xi}(\lambda_j)\right|}{\frac{2(d-\delta)+1}{m}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2(d-\delta)}} \tag{2.186}$$

$$\leq \frac{\sup_{d\in[\Delta_a,\Delta_2]} Cn^{2(d-\delta)}\frac{2(d-\delta)+1}{m}\left(\frac{1}{m}\right)^{2(d-\delta)}\left|\sum_{j=l}^{m}\lambda_j^{2d} I_{s\xi}(\lambda_j)\right|}{\inf_{d\in[\Delta,\Delta_2]}\frac{2(d-\delta)+1}{m}\sum_{j=l}^{m}\left(\frac{j}{m}\right)^{2(d-\delta)}}. \tag{2.187}$$

Robinson showed that $\inf_{d\in[\Delta,\Delta_2]}\frac{2(d-\delta)+1}{m}\sum_{j=1}^{m}\left(\frac{j}{m}\right)^{2(d-\delta)} \geq 1/2$ for m large enough, so we only have to discuss the numerator of (2.187). Using (2.169)-

(2.172), this is

$$= \begin{cases} O_p\left(n^{\phi-\delta} m^{2(\delta-\Delta_a)-1} l^{2\Delta_a-(\delta+\phi)} \ln l\right) \text{ if } 2\Delta_a - (\delta+\phi) < 0 \\ O_p\left(n^{\phi-\delta} m^{2(\delta-\Delta_a)-1} \ln^2 m\right) \text{ if } 2\Delta_a - (\delta+\phi) = 0 \\ O_p\left(n^{\phi-\delta} m^{\delta-\phi-1} \ln m\right) \text{ if } 2\Delta_a - (\delta+\phi) > 0. \end{cases} \tag{2.188}$$

When $2\Delta_a - (\delta+\phi) < 0$, we rewrite the bound as

$$\left(n^{\phi-\delta} m^{\delta-\Delta_a-1/2} l^{\Delta_a-\phi}\right)\left(m^{\delta-\Delta_a-1/2} \ln l\right)\left(l^{\Delta_a-\delta}\right) \tag{2.189}$$

and notice that, while the first factor is $O(1)$ because $\Delta_a \geq \Delta(\kappa, \upsilon)$, and the third one is $O(1)$ because $\delta \geq \Delta_a$, the second factor is $o(1)$ and so the whole sequence converges to zero. For the remaining terms, $2\Delta_a - (\delta+\phi) = 0$ is only possible if $\Delta_a \geq \phi$ (when $\Delta_a = \Delta$), because $\delta \geq \Delta_a$, so $\delta \geq \phi$. Rewriting then the bound as

$$\left(n^{\phi-\delta} m^{\delta-\phi}\right)\left(m^{-1} \ln^2 m\right) \tag{2.190}$$

the first factor is $O(1)$ while the second one is $o(1)$; with a similar argument $2\Delta_a - (\delta+\phi) > 0$ implies $\Delta_a > \phi$ and then $\delta > \phi$, so both the factors in the bound

$$\left(n^{\phi-\delta} m^{\delta-\phi}\right)\left(m^{-1} \ln m\right) \tag{2.191}$$

are $o(1)$.

When $\Delta(\kappa, \upsilon) \geq \Delta$, (2.183) has to be taken into account explicitly: we bound it as

$$\sup_{d \in [\Delta, \Delta(\kappa,\upsilon)]} \frac{\left|\frac{1}{m}\sum_{j=l}^{m} \lambda_j^{2d} I_{s\xi}(\lambda_j)\right|}{\frac{1}{m}\sum_{j=l}^{m} \lambda_j^{2d} I_s(\lambda_j)} \tag{2.192}$$

$$\leq \sup_{d\in[\Delta,\Delta(\kappa,\upsilon)]} \frac{(n/l)^{2(d-\phi)}}{(n/l)^{2(d-\phi)}} \frac{C\frac{1}{m}\left|\sum_{j=l}^{m}\lambda_j^{2d}I_{s\xi}(\lambda_j)\right|}{n^{2(\phi-d)}\frac{1}{m}\sum_{j=l}^{m}j^{2(d-\phi)-1}} \tag{2.193}$$

$$\leq \frac{\sup_{d\in[\Delta,\Delta(\kappa,\upsilon)]} Cn^{2(d-\phi)}l^{2(\phi-d)}\left|\sum_{j=l}^{m}\lambda_j^{2d}I_{s\xi}(\lambda_j)\right|}{\inf_{d\in[\Delta,\Delta(\kappa,\upsilon)]} l^{2(\phi-d)}\sum_{j=l}^{m}j^{2(d-\phi)-1}}. \tag{2.194}$$

Since $\phi > \Delta(\kappa,\upsilon) \geq \Delta$, then

$$\inf_{d\in[\Delta,\Delta(\kappa,\upsilon)]} l^{2(\phi-d)}\sum_{j=l}^{m}j^{2(d-\phi)-1} = l^{2(\phi-\Delta)}\sum_{j=l}^{m}j^{2(\Delta-\phi)-1} \to \frac{1}{2(\phi-\Delta)} > 0, \tag{2.195}$$

and we only have to discuss the numerator of (2.194).

Using (2.169)-(2.172), this is

$$O_p\left(n^{\delta-\phi}l^{\phi-\delta}\ln l\right) \text{ if } 2\Delta(\kappa,\upsilon)-(\delta+\phi)<0 \tag{2.196}$$

(and notice that this is the only possible outcome, given (2.180)).

Reversing the argument used to show (2.178),

$$\text{for } d\in[\Delta,\Delta(\kappa,\upsilon)],\ n^{2(\delta-\phi)}m^{2(d-\delta)+1}l^{2(\phi-d)} = O(1). \tag{2.197}$$

We then rewrite

$$n^{\delta-\phi}l^{\phi-\delta}\ln l = \left(n^{\delta-\phi}m^{d-\delta+1/2}l^{\phi-d}\right)\left(m^{\delta-d-1/2}\ln l\right)\left(l^{d-\delta}\right) \tag{2.198}$$

for $d\in[\Delta,\Delta(\kappa,\upsilon)]$: while the first and third factors are $O(1)$ (recall (2.180) for the last term), the second one is $o(1)$.

This proves Lemma 2.B.2 *(ii)*.

Finally, we prove Lemma 2.B.2 *(iii)* considering the sets $l\leq j\leq p$ and $p< j\leq m$ separately (notice that, for m large enough, $l<p$ because $p=O_e(m)$).

For $l \leq j \leq p$, using (2.151),

$$\begin{aligned}
&\frac{1}{m}\sum_{j=l}^{p}(j/p)^{2(\Delta-\delta)}(j/n)^{2\delta}I_{s\xi}(\lambda_j) \\
&= \frac{1}{m}\sum_{j=l}^{p}(j/p)^{2(\Delta-\delta)}(j/n)^{2\delta}(j/n)^{-\delta}(j/n)^{-\phi}j^{-1/2}(j/n)^{\delta}(j/n)^{\phi}j^{1/2}I_{s\xi}(\lambda_j) \\
&= \begin{cases} O_p\left(m^{2(\delta-\Delta)-1}n^{\phi-\delta}l^{2\Delta-(\delta+\phi)}\ln l\right) \text{ if } 2\Delta-(\delta+\phi)<0 \\ \\ O_p\left(m^{\delta-\phi-1}n^{\phi-\delta}\ln^2 m\right) \text{ if } 2\Delta-(\delta+\phi)=0 \\ \\ O_p\left(m^{\delta-\phi-1}n^{\phi-\delta}\ln m\right) \text{ if } 2\Delta-(\delta+\phi)>0. \end{cases}
\end{aligned} \tag{2.199}$$

Clearly the third bound is $\left(m^{\delta-\phi-1/2}n^{\phi-\delta}\right)\left(m^{-1/2}\ln m\right)$, which is $o(1)$, using (2.176) for the first factor; the second bound can be dealt with in a similar way.

Rewriting the first bound in (2.199) as

$$\left(m^{2(\delta-\Delta)-1}\ln l\right)\left(n^{\phi-\delta}\right)\left(l^{2\Delta-(\delta+\phi)}\right), \tag{2.200}$$

when $\phi \leq \delta$ then this too is $o(1)$ because the first factor is, and the other two are $O(1)$. If $\phi > \delta$, we look at

$$\frac{1}{m}\sum_{j=l}^{p}(j/p)^{2(\Delta-\delta)}(j/n)^{2\delta}I_s(\lambda_j) = O_e\left(m^{2(\delta-\Delta)-1}n^{2(\phi-\delta)}l^{2(\Delta-\phi)}\right) \tag{2.201}$$

instead (notice that $\phi > \delta$ implies $\phi > \Delta$). The ratio that we have to consider is

$$\frac{m^{2(\delta-\Delta)-1}n^{\phi-\delta}l^{2\Delta-(\delta+\phi)}\ln l}{m^{2(\delta-\Delta)-1}n^{2(\phi-\delta)}l^{2(\Delta-\phi)}} = n^{\delta-\phi}l^{\phi-\delta}\ln l = o(1). \tag{2.202}$$

For $p+1 \leq j \leq m$, using (2.150),

$$\frac{1}{m}\sum_{j=p+1}^{m}(j/p)^{2(\Delta_1-\delta)}(j/n)^{2\delta}I_{s\xi}(\lambda_j) = O_p\left(n^{\phi-\delta}m^{\delta-\phi-1/2}\right) \tag{2.203}$$

which is $o_p(1)$ for any eligible κ.

This concludes the proof of Lemma 2.B.2.

Lemma 2.B.3. *Under Assumptions 2.2', 2.3', and 2.5', as* $n \to \infty$,

(i) there is $\iota > 0$ *such that*

$$\sup_{\varepsilon \in (-\iota, \iota)} \left| \ln m^6 \left(\sum_{j=l}^{m} \lambda_j^{2(\delta+\varepsilon)} \left(G\lambda_j^{-2\delta} + I_s(\lambda_j) \right) \right)^{-1} \sum_{j=l}^{m} \lambda_j^{2(\delta+\varepsilon)} I_{s\xi}(\lambda_j) \right| = o_p(1); \tag{2.204}$$

(ii)

$$(\ln m)^6 \sum_{j=l}^{m} \lambda_j^{2\delta} I_s(\lambda_j) = o_p(1); \tag{2.205}$$

(iii)

$$\left| \frac{\sqrt{m}}{m} \sum_{j=l}^{m} \nu_j \lambda_j^{2\delta} I_{s\xi}(\lambda_j) \right| = o_p(1), \text{ where } \nu_j = \ln j - \frac{1}{m} \sum_{j=l}^{m} \ln j; \tag{2.206}$$

(iv)

$$\left| \frac{\sqrt{m}}{m} \sum_{j=l}^{m} \nu_j \lambda_j^{2\delta} I_s(\lambda_j) \right| = o(1). \tag{2.207}$$

Proof. Parts *(i)* and *(ii)* follow from the same proofs for Lemma 2.B.2 *(ii)* and *(i)* respectively, since the extra factor $\ln m^6$ is irrelevant when m, l are chosen as required in Assumptions A.1 and A.2.

To show part *(iii)* and *(iv)*, we bound

$$\begin{aligned}
&\left|\frac{\sqrt{m}}{m}\sum_{j=l}^{m}\nu_j\lambda_j^{2\delta}I_{s\xi}(\lambda_j)\right| \\
=&\left|\frac{\sqrt{m}}{m}\sum_{j=l}^{m}\left(\ln j-\frac{1}{m}\sum_{j=l}^{m}\ln j\right)\lambda_j^{2\delta}I_{s\xi}(\lambda_j)\right| \\
=&\frac{\sqrt{m}}{m}\left|\sum_{j=l}^{m}(\ln j)\,\lambda_j^{2\delta}I_{s\xi}(\lambda_j)-\sum_{j=l}^{m}\left(\frac{1}{m}\sum_{j=l}^{m}\ln j\right)\lambda_j^{2\delta}I_{s\xi}(\lambda_j)\right| \\
\leq&\frac{\sqrt{m}}{m}\left|\sum_{j=l}^{m}(\ln j)\,\lambda_j^{2\delta}I_{s\xi}(\lambda_j)\right|+\frac{\sqrt{m}}{m}\left|\sum_{j=l}^{m}\left(\frac{1}{m}\sum_{j=l}^{m}\ln j\right)\lambda_j^{2\delta}I_{s\xi}(\lambda_j)\right| \\
\leq&2\frac{\sqrt{m}}{m}\ln m\left|\sum_{j=l}^{m}\lambda_j^{2\delta}I_{s\xi}(\lambda_j)\right| \qquad (2.208)
\end{aligned}$$

so

$$= \begin{cases} O_p\left(m^{-1/2}n^{\phi-\delta}l^{\delta-\phi}\ln l\ln m\right) \text{ if } \delta-\phi<0 \\ O_p\left(m^{-1/2}\ln^3 m\right) \text{ if } \delta-\phi=0 \\ O_p\left(n^{\phi-\delta}m^{\delta-\phi-1/2}\ln^2 m\right) \text{ if } \delta-\phi>0. \end{cases} \qquad (2.209)$$

when $\phi>\delta$ the bound is $\left(m^{-1/4}\right)\left(n^{\phi-\delta}m^{-1/4}l^{\delta-\phi}\ln l\ln m\right)$, the second factor being $o(1)$ using (2.32); clearly, for $\phi=\delta$, $m^{-1/2}\ln^3 m=o(1)$; when $\delta>\phi$, the bound is $(n/m)^{\phi-\delta}\left(m^{-1/2}\ln^2 m\right)=o(1)$.

Finally, for part *(iv)*, using

$$|\nu_j|\leq 2\ln m, \qquad (2.210)$$

$$\left|\frac{\sqrt{m}}{m}\sum_{j=l}^{m}\nu_j\lambda_j^{2\delta}I_s(\lambda_j)\right|\leq 2\frac{\sqrt{m}}{m}\ln m\sum_{j=l}^{m}\lambda_j^{2\delta}I_s(\lambda_j). \qquad (2.211)$$

Replacing d with δ in (2.149), clearly this is $o(1)$ if $\delta\geq\phi$; if $\delta<\phi$, (2.211) is $=O\left(n^{2(\phi-\delta)}l^{2(\delta-\phi)}m^{-1/2}\ln m\right)$, which is $o(1)$ using (2.32).

Table 2.3: Monte Carlo bias and standard deviation, $\delta = 0$

$\delta(0)$		*bias*		*s.d.*		*a.s.d*	
	n	$\widehat{\delta}^{(1)}$	$\widehat{\delta}^{(l)}$	$\widehat{\delta}^{(1)}$	$\widehat{\delta}^{(l)}$	$\widehat{\delta}^{(1)}$	$\widehat{\delta}^{(l)}$
$s(-\infty)$	64	-0.020	-0.020	0.144	0.225	0.109	0.172
	128	-0.015	-0.020	0.101	0.171	0.083	0.146
	256	-0.011	-0.014	0.073	0.118	0.063	0.107
	512	-0.007	-0.012	0.053	0.083	0.048	0.078
	1028	-0.005	-0.006	0.038	0.060	0.036	0.057
$s(1/4)$	64	0.056	0.015	0.141	0.225	0.109	0.172
	128	0.062	0.008	0.095	0.173	0.083	0.146
	256	0.065	0.008	0.066	0.118	0.063	0.107
	512	0.065	0.005	0.047	0.083	0.048	0.078
	1028	0.062	0.008	0.035	0.060	0.036	0.057
$s(1/2)$	64	0.234	0.063	0.106	0.216	0.109	0.172
	128	0.246	0.031	0.067	0.164	0.083	0.146
	256	0.247	0.027	0.043	0.118	0.063	0.107
	512	0.249	0.014	0.028	0.082	0.048	0.078
	1028	0.252	0.017	0.019	0.060	0.036	0.057

Table 2.4: Monte Carlo bias and standard deviation, $\delta = 0.4$

$\delta(0.4)$		*bias*		*s.d.*		*a.s.d.*	
	n	$\widehat{\delta}^{(1)}$	$\widehat{\delta}^{(l)}$	$\widehat{\delta}^{(1)}$	$\widehat{\delta}^{(l)}$	$\widehat{\delta}^{(1)}$	$\widehat{\delta}^{(l)}$
$s(-\infty)$	64	-0.029	-0.042	0.148	0.143	0.109	0.172
	128	-0.019	-0.034	0.108	0.172	0.083	0.146
	256	-0.012	-0.022	0.074	0.140	0.063	0.107
	512	-0.007	-0.015	0.053	0.118	0.048	0.078
	1028	-0.004	-0.011	0.040	0.109	0.036	0.027
$s(1/4)$	64	-0.016	-0.029	0.143	0.160	0.109	0.172
	128	-0.011	-0.028	0.107	0.184	0.083	0.146
	256	-0.006	-0.016	0.075	0.156	0.063	0.107
	512	-0.002	-0.012	0.054	0.136	0.048	0.078
	1028	0.000	-0.009	0.041	0.116	0.036	0.057
$s(1/2)$	64	0.039	-0.014	0.147	0.191	0.109	0.172
	128	0.036	-0.016	0.107	0.202	0.083	0.146
	256	0.029	-0.014	0.075	0.163	0.063	0.107
	512	0.024	-0.010	0.054	0.126	0.048	0.078
	1028	0.022	-0.006	0.039	0.120	0.036	0.057

Table 2.5: Monte Carlo root-MSE and size of 5% t tests, $\delta = 0$

$\delta(0)$		$rMSE$				
	n	$\widehat{\delta}^{(1)}$	$\widehat{\delta}^{(l)}$	$t_{\widehat{\delta}^{(1)}}$	$t_{\widehat{\delta}^{(l)}}$	$t_{\widehat{\delta}^{(1)}-\widetilde{\delta}^{\dagger}}$
$s(-\infty)$	64	0.145	0.226	0.071	0.182	0.199
	128	0.102	0.172	0.056	0.171	0.192
	256	0.074	0.119	0.049	0.165	0.177
	512	0.053	0.084	0.047	0.143	0.172
	1028	0.038	0.061	0.043	0.141	0.154
$s(1/4)$	64	0.152	0.226	0.182	0.226	0.277
	128	0.114	0.173	0.209	0.219	0.333
	256	0.092	0.118	0.294	0.210	0.362
	512	0.080	0.083	0.390	0.184	0.457
	1028	0.072	0.060	0.532	0.193	0.506
$s(1/2)$	64	0.257	0.225	0.689	0.315	0.641
	128	0.255	0.167	0.943	0.256	0.863
	256	0.251	0.121	1.000	0.259	0.975
	512	0.251	0.083	1.000	0.233	0.998
	1028	0.252	0.062	1.000	0.240	1.000

Table 2.6: Monte Carlo root-MSE and size of 5% t tests, $\delta = 0.4$

$\delta(0.4)$		$rMSE$				
	n	$\widehat{\delta}^{(1)}$	$\widehat{\delta}^{(l)}$	$t_{\widehat{\delta}^{(1)}}$	$t_{\widehat{\delta}^{(l)}}$	$t_{\widehat{\delta}^{(l)}-\widetilde{\delta}^{\dagger}}$
$s(-\infty)$	64	0.151	0.223	0.062	0.130	0.206
	128	0.109	0.176	0.069	0.162	0.233
	256	0.075	0.121	0.056	0.129	0.211
	512	0.054	0.084	0.052	0.115	0.198
	1028	0.040	0.058	0.046	0.103	0.190
$s(1/4)$	64	0.144	0.212	0.079	0.150	0.217
	128	0.108	0.176	0.076	0.172	0.253
	256	0.075	0.120	0.068	0.142	0.202
	512	0.054	0.084	0.076	0.130	0.202
	1028	0.041	0.058	0.058	0.108	0.189
$s(1/2)$	64	0.153	0.221	0.158	0.181	0.313
	128	0.113	0.173	0.177	0.190	0.370
	256	0.080	0.122	0.149	0.148	0.376
	512	0.059	0.082	0.146	0.120	0.374
	1028	0.045	0.057	0.160	0.113	0.408

Chapter 3

Local Whittle estimation of the memory parameter for processes subject to a break

3.1 Introduction

We discussed in Chapter 2 the estimation of the memory parameter when the process is characterised by an unstable mean. We consider here a complementary form of instability: the one related to the (mean-corrected) stochastic component itself, either in the order of integration or in the short term dynamics.

We have already motivated our interest in the memory parameter interpreting it as an indicator of persistence over time and, for a policy variable, of the attitude towards stabilisation from the authority that is controlling or targeting it (assuming of course a stable framework in the rest of the economy). More in general, the stability of the memory parameter may also be important when the focus is on long term dynamics, including cointegrating relationships.

As we did in Chapter 2, we complement the analysis with a test that may be applied to detect evidence of such a shift, and also with a procedure to estimate its location.

As before, we discuss the local Whittle estimate for its small variance and because the theoretical treatment is simpler than that for the log-periodogram regression, but we conjecture that these results also provide reliable guidelines for the other case.

In Section 3.2 we present the asymptotic theory, in Section 3.3 we analyse the small sample properties with a Monte Carlo exercise and in Section 3.4 we analyse inflation in the euro-area. We conclude in Section 3.5; the proofs of the theorems are to be found in the Appendix.

3.2 Local Whittle estimation in presence of a potential break

We formalise our model by introducing the process x_t, observed at $t = 1, ..., n$, which we describe as the sum of two unobservable processes x_{1t} and x_{2t},

$$x_t = x_{1t} + x_{2t} \tag{3.1}$$

such that

$$x_{1t} = \begin{cases} \xi_{1t} \text{ if } t \leq [\tau_0 n] \\ \\ 0 \text{ otherwise} \end{cases}, x_{2t} = \begin{cases} \xi_{2t} \text{ if } t > [\tau_0 n] \\ \\ 0 \text{ otherwise} \end{cases} \tag{3.2}$$

for a constant $\tau_0 \in (0, 1)$. The process ξ_{1t} is stationary and invertible and has autocovariance $\gamma_{\xi 1}(s)$ and spectral density $f_{\xi 1}(\lambda)$; in a similar way ξ_{2t}

has autocovariance $\gamma_{\xi 2}(s)$ and spectral density $f_{\xi 2}(\lambda)$ and, introducing the covariance $\gamma_{\xi 12}(s) = E(\xi_{1t}\xi_{2,t+s})$, we indicate the cross-spectrum with $f_{\xi 12}(\lambda)$.

The processes ξ_{1t}, ξ_{2t} are characterised by spectral densities having

$$f_{\xi 1}(\lambda) \sim G_{\xi 1}\lambda^{-2\delta_1}, \quad f_{\xi 2}(\lambda) \sim G_{\xi 2}\lambda^{-2\delta_2} \text{ when } \lambda \longrightarrow 0^+, \tag{3.3}$$

so they can be fractionally integrated.

We do not make any other assumption on ξ_{1t} and on ξ_{2t}, thus encompassing several cases: for example they may be independently distributed, but they may also be actually the same process (in which case there is no break).

We indicate with $F_x(\lambda)$, $F_{x1}(\lambda)$ and $F_{x2}(\lambda)$ the discrete Fourier transform of x_t, x_{1t} and x_{2t} respectively, and with $I_x(\lambda)$, $I_{x1}(\lambda)$ and $I_{x2}(\lambda)$ the corresponding periodograms; finally, for the crossperiodogram between $x1$ and $x2$, we use $I_{x12}(\lambda)$.

3.2.1 The periodogram of a process potentially subject to a break

The processes x_{1t}, x_{2t} are not stationary, and the bound for the expectation of the periodogram provided by Robinson (1995a) cannot be directly applied. Yet we find that the same result can be quickly derived: introduce

$$\tau^{\triangle} = [\tau n]/n, \tag{3.4}$$

and the following assumptions

Assumption B.1. *For* $a \in \{1,2\}$ *there exists* $G_{\xi a} \in (0,\infty)$, $\delta_a \in (-1/2, 1/2)$, *and* $\alpha \in (0,2]$ *such that*

$$f_{\xi a}(\lambda) = G_{\xi a}\lambda^{-2\delta_a} + O\left(\lambda^{\alpha-2\delta_a}\right) \text{ as } \lambda \to 0^+. \tag{3.5}$$

Assumption B.2. *For $a \in \{1,2\}$, in a neighbourhood $(0,\varepsilon)$ of the origin $f_{\xi a}(\lambda)$, $f_{\xi 12}(\lambda)$ are differentiable and*

$$\left|\frac{\partial f_{\xi a}(\lambda)}{\partial \lambda}\right| = O\left(\lambda^{1-2\delta_a}\right), \left|\frac{\partial f_{\xi 12}(\lambda)}{\partial \lambda}\right| = O\left(\lambda^{1-\delta_1-\delta_2}\right) \text{ as } \lambda \to 0^+. \tag{3.6}$$

Assumption B.3. *Letting $R_{12}(\lambda) = f_{\xi 12}(\lambda)/\sqrt{f_{\xi 1}(\lambda) f_{\xi 2}(\lambda)}$, then for some $\beta_g \in (0,2]$,*

$$|R_{12}(\lambda) - R_{12}(0)| = O\left(\lambda^{\beta_g}\right) \text{ as } \lambda \to 0^+. \tag{3.7}$$

Assumptions B.1 to B.3 were introduced by Robinson (1995a) and are also discussed therein. In accordance with the semiparametric approach to the problem, all the assumptions are local to 0. Assumption B.1 imposes a rate of convergence to the approximation of $f_{\xi a}(\lambda)/G_{\xi a}\lambda^{-2\delta_a}$ to 1 and it was introduced because Robinson formulated his result for $I_{\xi_a}(\lambda_j)/G_{\xi a}\lambda_j^{-2\delta_a}$ so the additional approximation of $f_{\xi a}(\lambda)$ by $G_{\xi a}\lambda^{-2\delta_a}$ had to be taken into account; it also imposes stationarity, the extension to the nonstationary process having been discussed by Velasco (1999a). Assumption B.2 is a common smoothness condition and it is also present, for example, in the Whittle estimation of a fully parametric model; Assumption B.3 is automatically met when ξ_{1t} and ξ_{2t} are observed from the same process, while the situation in which the spectral density matrix for $(\xi_{1t},\xi_{2t})'$ is not singular is discussed by Robinson (1995a), where he also mentions that these conditions are met, for example, from standard ARFIMA processes.

We then have the following

Theorem 3.1. *Let Assumptions B.1, B.2 and B.3 hold and introduce*

$$v_{x1}(\lambda) = \frac{F_{x1}(\lambda)}{\sqrt{\tau_0^{\triangle} G_{\xi 1} \lambda^{-2\delta_1}}}, \; v_{x2}(\lambda) = \frac{F_{x2}(\lambda)}{\sqrt{\left(1-\tau_0^{\triangle}\right) G_{\xi 2} \lambda^{-2\delta_2}}} \tag{3.8}$$

For $a, b \in \{1,2\}$, for any positive integer sequence $j(n)$ with $j/n \to 0$,

$$E\left(v_a(\lambda_j)\, v_b(-\lambda_j)\right) = 1(a=b) + O\left(\frac{\ln j}{j} + \left(\frac{j}{n}\right)^{\alpha}\right), \tag{3.9}$$

and, with $j > k$, k positive integer,

$$E\left(v_a(\lambda_j)\, v_b(-\lambda_k)\right) = O\left(\frac{\ln j}{k}\right). \tag{3.10}$$

Since

$$I_x(\lambda) = I_{x1}(\lambda) + 2\,\mathrm{Re}\, I_{x12}(\lambda) + I_{x2}(\lambda), \tag{3.11}$$

it follows from Theorem 3.1 that, for $j > 0$,

$$\begin{aligned} & E\left(I_x(\lambda_j)\right) \\ = \; & \tau_0^{\triangle} f_{\xi 1}(\lambda_j) + \left(1-\tau_0^{\triangle}\right) f_{\xi 2}(\lambda_j) + O\left(\left(f_{\xi 1}(\lambda_j) + \mathrm{Re}\, f_{\xi 12}(\lambda_j) + f_{\xi 2}(\lambda_j)\right)\left(j^{-1}\ln j\right)\right) \end{aligned} \tag{3.12}$$

as $j/n \to 0$. When the process x_t is not subject to any break, then, clearly, $f_{\xi 1}(\lambda) = f_{\xi 2}(\lambda) = f_{\xi 12}(\lambda)$, and the orders of magnitude in (3.12) are the same as those given in Robinson (1995a).

If a break in δ took place, so that, for example,

$$\delta_1 > \delta_2 \tag{3.13}$$

(which we can assume without loss of generality), the process x_t behaves like

a long memory process with parameter δ_1, but subject to an unobservable disturbance with memory of order δ_2, and the stochastic order of magnitude of the periodogram, for $j > 0$, is

$$E\left(I_x\left(\lambda_j\right)\right) = \tau_0^{\triangle} G_{\xi 1} \lambda_j^{-2\delta_1} + O\left(\lambda_j^{-2\delta_1} \frac{\ln j}{j} + \lambda_j^{-2\delta_1} \left(\frac{j}{n}\right)^{\alpha} + \lambda_j^{-2\delta_2}\right) \quad (3.14)$$

as $j/n \to 0$.

The term $\lambda_j^{-2\delta_2}$ in the bound in (3.14) is not necessarily negligible: indeed, considering j proportional to n^{φ} for a certain φ, it is of order bigger than $\lambda_j^{-2\delta_1} \frac{\ln j}{j}$ for φ large enough ($\varphi > \left(1 + 2\left(\delta_1 - \delta_2\right)\right)^{-1} 2\left(\delta_1 - \delta_2\right)$).

It is interesting to compare (3.14) to the expected value of the periodogram of a process in which the stochastic component is corrupted by a deterministic component that acts as a noise and obscures the signal, as discussed in Chapter 2. The situation is clearly reversed: while with a time-varying deterministic component the signal is mainly obscured at the lowest frequencies, from (3.14) we see that here those are the frequencies in which the signal is more clear. Breaks in the mean and in the memory parameter have then the opposite effect when analysed in the frequency domain.

The expectation in (3.12) is also useful to observe the effect of a break in the short term dynamics. In that case,

$$\delta_1 = \delta_2 = \delta, \quad (3.15)$$

so the expectation becomes, for $j > 0$,

$$E\left(I_x\left(\lambda_j\right)\right) = \left(\tau_0^{\triangle} G_{\xi_1} + \left(1 - \tau_0^{\triangle}\right) G_{\xi 2}\right) \lambda_j^{-2\delta} \left(1 + \frac{\ln j}{j} + \left(\frac{j}{n}\right)^{\alpha}\right) \quad (3.16)$$

as $j/n \to 0$, and there is then no effect on the slope of the expectation of the

periodogram at low frequencies. Since the term $\left(\tau_0^{\triangle} G_{\xi_1} + \left(1 - \tau_0^{\triangle}\right) G_{\xi 2}\right)$ is only a scaling factor, we can already anticipate that the local Whittle estimate is robust to changes in the short term dynamics. This is hardly a surprising result, given that the local Whittle estimate does not actually require any specification of the short term dynamics at all, but it is important to state it explicitly because it provides a strong argument in favour of semiparametric estimates, such as the local Whittle or the log-periodogram regression ones, when these are compared to other estimates that may even be more efficient (in the sense of having a faster rate of convergence, as for example is the case of the fully parametric Whittle estimate), but are sensitive to other breaks as well.

We conclude the subsection by discussing

$$z_t = x_t,\ z_{1t}(\tau) = \begin{cases} x_t \text{ if } t \leq [\tau n] \\ \\ 0 \text{ otherwise} \end{cases},\ z_{2t}(\tau) = \begin{cases} x_t \text{ if } t > [\tau n] \\ \\ 0 \text{ otherwise.} \end{cases} \tag{3.17}$$

In z_{1t} the second block of observations is supplemented with zeros, while in z_{2t} the design is reversed: z_{1t} and z_{2t} are the processes that are going to be used in the recursive Chow test.

From Theorem 3.1 we already have a bound for $E(I_z(\lambda_j))$; for $E(I_{z1}(\lambda_j))$ and $E(I_{z2}(\lambda_j))$ however, we only have it when $\tau^{\triangle} = \tau_0^{\triangle}$. Yet using the same arguments it is also immediate to show that, under (3.13) and under the

assumptions stated for Theorem 3.1, $j > 0$,

$$
\begin{aligned}
& if\ \tau^{\triangle} < \tau_0^{\triangle}, \\
E\left(I_{z1}\left(\lambda_j\right)\right) &= \tau^{\triangle} G_{\xi 1} \lambda_j^{-2\delta_1} + O\left(\lambda_j^{-2\delta_1}\left[\frac{\ln j}{j} + \left(\frac{j}{n}\right)^{\alpha}\right]\right) && (3.18) \\
E\left(I_{z2}\left(\lambda_j\right)\right) &= \left(\tau_0^{\triangle} - \tau^{\triangle}\right) G_{\xi 1} \lambda_j^{-2\delta_1} + \left(1 - \tau_0^{\triangle}\right) G\left(\frac{j}{n}\right)^{\alpha}\Big] + O\left(\lambda_j^{-2\delta_1}\left[\frac{\ln j}{j} + \left(\frac{j}{n}\right)^{\alpha}\right]\right) && (3.19) \\
and\ if\ \tau^{\triangle} &> \tau_0^{\triangle}, \\
E\left(I_{z1}\left(\lambda_j\right)\right) &= \tau_0^{\triangle} G_{\xi 1} \lambda_j^{-2\delta_1} + \left(\tau^{\triangle} - \tau_0^{\triangle}\right) G_{\xi 2} \lambda_j^{-2\delta_2} + O\left(\lambda_j^{-2\delta_1}\left[\frac{\ln j}{j} + \left(\frac{j}{n}\right)^{\alpha}\right]\right) && (3.20) \\
E\left(I_{z2}\left(\lambda_j\right)\right) &= \left(1 - \tau^{\triangle}\right) G_{\xi 2} \lambda_j^{-2\delta_2} + O\left(\lambda_j^{-2\delta_2}\left[\frac{\ln j}{j} + \left(\frac{j}{n}\right)^{\alpha}\right]\right) && (3.21)
\end{aligned}
$$

as $j/n \to 0$.

When $\delta_1 > \delta_2$, the order of magnitude of the expectation of the periodogram of z_{1t} always depends on δ_1; for the periodogram of z_{2t} on the other hand, the conclusion depends on the position of τ with respect to the break τ_0: when all the observations with the highest memory are removed, the dominating term depends on δ_2, otherwise on δ_1. Intuitively then, the slopes of two periodograms are very different when τ is set large enough, and a test can be realised simply by comparing them.

3.2.2 The estimate when a break in the memory parameter is present

We analyse the effects of the break using the Local Whittle estimate as in Robinson (1995b): this is computed by minimising the loss function

$$R(d; m, I) = \ln\left\{\frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2d} I(\lambda_j)\right\} - 2d\frac{1}{m}\sum_{j=1}^{m} \ln(\lambda_j). \qquad (3.22)$$

This is also described in Chapter 1 and it is repeated here because we introduced the notation $R(d; m, I)$: this explicitly states that the result depends on the dataset, of which $I(\lambda_j)$ is the periodogram, and on m, a user chosen bandwidth parameter. We introduce

$$\widehat{\delta} = \arg \min_{d \in \Theta \subset (-1/2,1/2)} R(d; m, I_z) \tag{3.23}$$

$$\widehat{\delta}_1(\tau) = \arg \min_{d \in \Theta \subset (-1/2,1/2)} R(d; m, I_{z1}) \tag{3.24}$$

$$\widehat{\delta}_2(\tau) = \arg \min_{d \in \Theta \subset (-1/2,1/2)} R(d; m, I_{z2}): \tag{3.25}$$

$\widehat{\delta}$ is the estimate when the whole dataset is used, and we are interested in it because we can then see what happens when the data are subject to a break in the memory parameter; $\widehat{\delta}_1(\tau)$ and $\widehat{\delta}_2(\tau)$ are the estimates computed using only the first or the second part of the sample for given τ, and are used to detect a change in δ with the Chow test.

In order to get their limits in probability we introduce a set of assumptions

Assumption 3.1. *For $a \in \{1, 2\}$, as $\lambda \to 0^+$,*

$$f_{\xi a}(\lambda) \sim G_{\xi a} \lambda^{-2\delta_a} \tag{3.26}$$

where $G_{\xi a} \in (0, \infty)$ and $\delta_a \in [\Delta_1, \Delta_2]$, where $-1/2 < \Delta_1 < \Delta_2 < 1/2$.

Assumption 3.2. *For $a \in \{1, 2\}$, in a neighbourhood $(0, \iota)$ of the origin, $f_{\xi a}(\lambda)$, $f_{\xi 12}(\lambda)$ are differentiable and*

$$\frac{d}{d\lambda} \ln f_{\xi a}(\lambda) = O(\lambda^{-1}), \ \frac{d}{d\lambda} \ln f_{\xi 12}(\lambda) = O(\lambda^{-1}) \text{ as } \lambda \to 0^+. \tag{3.27}$$

Assumption 3.3. *For $a \in \{1, 2\}$, the sequence $\xi_{a,t}$ is such that*

$$\xi_{a,t} = \sum_{j=0}^{\infty} \alpha_{a,j} \varepsilon_{a,t-j}, \ \sum_{j=0}^{\infty} \alpha_{a,j}^2 < \infty \tag{3.28}$$

where

$$E(\varepsilon_{a,t}|F_{a,t-1}) = 0,\ E(\varepsilon_{a,t}^2|F_{a,t-1}) = 1,\ a.s.,\ t = 0, \pm 1, ... \tag{3.29}$$

in which $F_{a,t}$ *is the* $\sigma-$*field generated by* $\varepsilon_{a,s}$, $s \leq t$, *and there exists a random variable* ϵ *such that* $E(\epsilon) < \infty$ *and for all* $\eta > 0$ *and some* $C > 0$, $P(|\varepsilon_{a,t}| > \eta) \leq CP(|\epsilon| > \eta)$.

Assumption 3.4. *As* $n \to \infty$,

$$\frac{1}{m} + \frac{m}{n} \to 0. \tag{3.30}$$

These are the same assumptions as those of Robinson (1995b), augmented to take $f_{\xi 12}(\lambda)$ into account as well. Assumptions 3.1 to 3.3 have appeared in Chapter 2, to which we refer for a detailed discussion. Notice that we do not require $\delta_1 = \delta_2$, so we can discuss the consequences of a break in δ.

Following Robinson (1995b) we can then show:

Theorem 3.2. *Under (3.13) and Assumptions 3.1, 3.2, 3.3, 3.4, then, as* $n \to \infty$,

$$\widehat{\delta} \to_p \delta_1 \tag{3.31}$$

and

$$if\ \tau < \tau_0\text{: } \widehat{\delta}_1(\tau) \to_p \delta_1,\ \widehat{\delta}_2(\tau) \to_p \delta_1; \tag{3.32}$$

$$if\ \tau = \tau_0\text{: } \widehat{\delta}_1(\tau) \to_p \delta_1,\ \widehat{\delta}_2(\tau) \to_p \delta_2; \tag{3.33}$$

$$if\ \tau > \tau_0\text{: } \widehat{\delta}_1(\tau) \to_p \delta_1,\ \widehat{\delta}_2(\tau) \to_p \delta_2. \tag{3.34}$$

Notice that here and after we formulate the theorem for τ because $\tau^{\triangle}/\tau \to 1$ as $n \to \infty$.

The Local Whittle estimate $\widehat{\delta}$ then converges in probability to the largest of

the two memory parameters, confirming the intuition from Theorem 3.1 that the process with lower order of integration acts like a disturbance: neither the value of the lower order of integration, nor the fraction of observations with lower memory as opposed to those with higher memory, matter. This result depends on the fact that the estimate is semiparametric and it only uses the frequencies where the features of the spectral density are dominated by the long term component: had all the frequencies been used, as in the Whittle estimate, it would converge to a point intermediate between δ_1 and δ_2, the exact location depending also on the fraction of observations with lower memory.

The same considerations apply to $\widehat{\delta}_1(\tau)$ and $\widehat{\delta}_2(\tau)$, but for some τ it may happen that only observations with lower memory are used, and in that case the estimate converges to that (lower) level instead.

Robinson (1995b) also showed that the Local Whittle estimate is root-m consistent and the limit distribution is zero-mean, asymptotically normal. In case of a break, root-m consistent estimation of the largest memory parameter depends on how large the gap

$$\vartheta = |\delta_1 - \delta_2| \tag{3.35}$$

is. Introduce the following assumptions:

Assumption 3.1'. *For* $a \in \{1,2\}$ *and some* $\beta_\xi \in (0,2]$

$$f_{\xi a}(\lambda) \sim G_{\xi a}\lambda^{-2\delta_a}(1 + O(\lambda^{\beta_\xi})) \quad as\ \lambda \to 0^+, \tag{3.36}$$

where $G_{\xi a} \in (0,\infty)$ *and* $\delta_a \in [\Delta_1, \Delta_2]$, *where* $-1/2 < \Delta_1 < \Delta_2 < 1/2$.

Assumption 3.2'. *For* $a \in \{1,2\}$, *in a neighbourhood* $(0,\iota)$ *of the origin,*

$\alpha_a(\lambda)$ *is differentiable and*

$$\frac{d}{d\lambda}\alpha_a(\lambda) = O\left(\frac{|\alpha_a(\lambda)|}{\lambda}\right) \text{ as } \lambda \to 0^+ \tag{3.37}$$

where $\alpha_a(\lambda) = \sum_{l=0}^{\infty} \alpha_{a,l} e^{i\lambda l}$.

Assumption 3.3'. *Assumption 3.3 holds and also*

$$E(\varepsilon_{a,t}^3 | F_{a,t-1}) = c_{a,1},\ E(\varepsilon_{a,t}^4 | F_{a,t-1}) = c_{a,2},\ a.s.,\ t = 0, \pm 1, ... \tag{3.38}$$

for some finite constants $c_{a,1}$ *and* $c_{a,2}$.

Assumption 3.4'. *As* $n \to \infty$,

$$\frac{1}{m} + \frac{m^{1+2\beta_\xi} \ln^2 m}{n^{2\beta_\xi}} \to 0. \tag{3.39}$$

Assumption 3.5'. *As* $n \to \infty$, *if* $\delta_2 \neq \delta_1$, *letting* $\vartheta = |\delta_1 - \delta_2|$,

$$\frac{m^{2\vartheta+1/2}}{n^{2\vartheta}} \to 0. \tag{3.40}$$

Assumptions 3.1' to 3.4' replicate those in Robinson (1995b) and are discussed in Chapter 2 as well.

Assumption 3.5' on the other hand has been introduced precisely to treat the case in which a change in δ took place. We postpone the discussion to after Theorem 3.4, where the consequences of not meeting it are presented. It is sufficient here to notice that it removes the highest frequencies, and that it is stronger the smaller ϑ is. Whether it is more or less restrictive than Assumption 3.4' depends on ϑ and on β_ξ: for the popular case $\beta_\xi = 2$, and indeed for any $\beta_\xi > 1$, Assumption 3.5' is stronger than 3.4', the reverse happening when β_ξ is very close to 0.

Theorem 3.3. *(i) Under Assumptions 3.1', 3.2', 3.3', 3.4', as* $n \to \infty$,

$$if\ \tau \leq \tau_0\text{: } 2\sqrt{m\tau}(\widehat{\delta}_1(\tau) - \delta_1) \to_d N(0,1) \tag{3.41}$$

$$if\ \tau \geq \tau_0\text{: } 2\sqrt{m(1-\tau)}(\widehat{\delta}_2(\tau) - \delta_2) \to_d N(0,1); \tag{3.42}$$

(ii) Under (3.13) and Assumptions 3.1', 3.2', 3.3', 3.4', 3.5', as $n \to \infty$,

$$2\sqrt{m\tau_0}(\widehat{\delta} - \delta_1) \to_d N(0,1) \tag{3.43}$$

$$if\ \tau > \tau_0\text{: } 2\sqrt{m\tau_0}(\widehat{\delta}_1(\tau) - \delta_1) \to_d N(0,1) \tag{3.44}$$

$$if\ \tau < \tau_0\text{: } 2\sqrt{m(\tau-\tau_0)}(\widehat{\delta}_2(\tau) - \delta_1) \to_d N(0,1). \tag{3.45}$$

This is the same result as the original paper of Robinson (1995b): we simply have to replace m by $m\tau$ (or by $m(1-\tau)$, $m\tau_0$, $m(\tau-\tau_0)$ according to the situation), to take into account that only a fraction of observations has a higher memory (this may well be δ_2, when no observations with δ_1 are present, as in (3.42)).

The statement of the theorem is divided into two parts according to whether the break in the memory parameter is included in the sample or not. In the first part it is not: the two time series are both composed of a stationary process padded with 0, and we verified that in such a case the results of Robinson (1995b) did not change.

In the second part of the Theorem on the other hand the data have been subject to the shock, and we need Assumption 3.5' as well, to make its effect negligible.

When Assumption 3.5' is not met, consider instead

Assumption 3.6'. *As* $n \to \infty$, *if* $\delta_2 \neq \delta_1$, *letting* $\vartheta = |\delta_1 - \delta_2|$,

$$\frac{m^{2\vartheta+1/2}}{n^{2\vartheta}} \to \infty. \tag{3.46}$$

Theorem 3.4. *Under (3.13), Assumptions 3.1', 3.2', 3.3', 3.4' and 3.6',* *as* $n \to \infty$,

$$\left(n^{2\vartheta}m^{-2\vartheta}\right)\left(\widehat{\delta} - \delta_1\right) \to_p -\frac{1}{2}\frac{1-\tau_0}{\tau_0}\frac{G_{\xi 2}}{G_{\xi 1}}(2\pi)^{2\vartheta}\frac{2\vartheta}{(1+2\vartheta)^2} \tag{3.47}$$

and

$$if\ \tau < \tau_0\text{:}\quad \left(n^{2\vartheta}m^{-2\vartheta}\right)\left(\widehat{\delta}_2(\tau) - \delta_1\right) \to_p -\frac{1}{2}\frac{1-\tau_0}{\tau_0-\tau}\frac{G_{\xi 2}}{G_{\xi 1}}(2\pi)^{2\vartheta}\frac{2\vartheta}{(1+2\vartheta)^2}; \tag{3.48}$$

$$if\ \tau > \tau_0\text{:}\quad \left(n^{2\vartheta}m^{-2\vartheta}\right)\left(\widehat{\delta}_1(\tau) - \delta_1\right) \to_p -\frac{1}{2}\frac{\tau-\tau_0}{\tau_0}\frac{G_{\xi 2}}{G_{\xi 1}}(2\pi)^{2\vartheta}\frac{2\vartheta}{(1+2\vartheta)^2}. \tag{3.49}$$

When Assumption 3.5' is not met but it is replaced by Assumption 3.6', the estimate is subject to a lower order bias. Intuitively this happens because the term λ_j^{2d} in the loss function has to accommodate both $I_{x1}(\lambda_j)$ and $I_{x2}(\lambda_j)$: at the lowest frequencies the latter is irrelevant, but for higher frequencies the stochastic orders of magnitude of the two periodograms get closer and the estimate is then contaminated by δ_2.

In Theorem 3.2 then, Assumption 3.4 ensures that such a contamination is always of a lower order, and in Theorem 3.3, Assumption 3.5' is sufficient to have that order smaller than $m^{-1/2}$, so root-m consistent estimation of δ_1 still follows, but here we see that when Assumption 3.5' is not met then the root-m consistency fails.

Notice that Theorem 3.4, by requiring Assumption 3.6', implicitly sets a

limit on β_ξ in Assumption 3.1': indeed if Assumption 3.4' is more restrictive than 3.5' (as it happens for very small β_ξ) then the assumptions of Theorem 3.4 cannot be met.

The "jump" ϑ affects the limit in two ways: the value of the potential bias increases with ϑ, but its order of magnitude gets smaller the larger ϑ. Larger ϑ should overall deliver smaller bias because the order of magnitude should prevail, but it is still possible that in small samples the reverse happens.

The fact that the bias from δ_1 gets smaller the more distant δ_2 is, may seem counter-intuitive: if, for example, we used fully parametric Whittle estimation, we would expect that the estimate converges to a point intermediate between δ_1 and δ_2, and that it would be further away from δ_1 the bigger ϑ. The result we obtained here depends crucially on $m/n \to 0$, so $I_{x2}(\lambda_j)$ is always dominated by $I_{x1}(\lambda_j)$, and the gap in the stochastic orders is bigger the bigger ϑ. Indeed, setting m proportional to n, the bias induced by the break (disregarding then the bias due to the approximation of the spectral density as in Assumption 3.1'), would then be bigger the larger ϑ, as we conjectured before for the fully parametric Whittle estimate.

When m is set as

$$m = c_\kappa n^\kappa \tag{3.50}$$

for some $\kappa \in (0, 4/5)$ and $c_k > 0$, Assumption 3.5' requires

$$\vartheta > \frac{1}{4}\frac{\kappa}{1-\kappa}: \tag{3.51}$$

we can see then that if $\kappa = 0.8 - \varepsilon$ (for some $\varepsilon > 0$) is set, then no combination of δ_1, δ_2 in a closed subset of $(-1/2, 1/2)$ is available to eliminate the bias, confirming that when $\beta_\xi = 2$ in Assumption 3.1' then Assumption 3.5' requires the elimination of more of the intermediate frequencies.

3.2.3 Tests for parameter instability and estimation of the breakpoint

Theorem 3.4 is also very interesting for its potential application to detect the presence of a break. Since the bias increases with the bandwidth m, a test can be realised by comparing two estimates for two different bandwidths, if these are properly chosen, because the estimate computed with the smaller bandwidth should be less subject to the lower order bias. More formally, defining

$$\widehat{\delta}_{(1)} = \arg\min_{\Theta} R\left(d; m_1; I_z\left(\lambda_j\right)\right) \tag{3.52}$$

$$\widehat{\delta}_{(2)} = \arg\min_{\Theta} R\left(d; m_2; I_z\left(\lambda_j\right)\right) \tag{3.53}$$

with

$$m_2/m_1 \to 0 \text{ as } n \to \infty, \tag{3.54}$$

and introducing

$$\widetilde{t} = \sqrt{4m_2}\left(\widehat{\delta}_{(2)} - \widehat{\delta}_{(1)}\right), \tag{3.55}$$

we have the following theorem.

Theorem 3.5. *Let $m = m_1$ in Assumptions 3.1', 3.2', 3.3', 3.4', 3.6' and $1/m_2 + m_2/m_1 \to 0$ as $n \to \infty$. Then as $n \to \infty$,*

$$\begin{cases} \widetilde{t} \to_d N(0,1) \text{ if } \vartheta = 0 \\ \\ \widetilde{t} \to \infty \text{ if } \vartheta > 0. \end{cases} \tag{3.56}$$

The statistic $\widetilde{t}$ can then detect the presence of a break even when ϑ is so small that both $\widehat{\delta}_{(2)}$ and $\widehat{\delta}_{(1)}$ are subject to the lower order bias, because the bias is smaller in the first case.

Since if there is a break then $\widetilde{t}$ should be positive, a one sided alternative should be preferred in order to have more power. We then suggest to compute $\widetilde{t}$ and then to reject $H_0 : \{\vartheta = 0\}$ if the realisation of the test statistic exceeds the critical value 1.65 (assuming of course a 5% size).

Notice that this test requires Assumption 3.6', that is, a certain smoothness on β_ξ in Assumption 3.1' as we saw when discussing Theorem 3.4: the test then cannot be used for very small β_ξ, but it is still valid for the β_ξ associated to the more popular models (including the ARFIMA).

A more important comment has to do with the power of the test. This may be very small, for two reasons: because the slower rate of convergence $m_2^{1/2}$ has to be used, and because the test is only consistent because of the lower order bias. Therefore the performance of the test in small samples may be less than satisfactory.

For a more powerful test, we introduce

$$\widehat{t}(\tau) = \sqrt{4\tau(1-\tau)m}\left(\widehat{\delta}_1(\tau) - \widehat{\delta}_2(\tau)\right). \tag{3.57}$$

Theorem 3.6. *(i) Under Assumptions 3.1', 3.2', 3.3', 3.4', $\delta_1 = \delta_2$, as $n \to \infty$,*

$$2\sqrt{\frac{\tau_1(1-\tau_2)}{1+\tau_1-\tau_2}m}\left(\widehat{\delta}_1(\tau_1) - \widehat{\delta}_2(\tau_2)\right) \to_d N(0,1), \ \tau_2 \geq \tau_1 \tag{3.58}$$

and

$$\widehat{t}(\tau)^2 \to_d \chi_1^2 \tag{3.59}$$

for given τ;

(ii) under Assumptions 3.1', 3.2', 3.3', 3.4', $\delta_1 > \delta_2$ and $\tau \geq \tau_0$, as

$n \to \infty$,

$$\widehat{t}(\tau)^2 \to \infty; \tag{3.60}$$

(iii) under Assumptions 3.1', 3.2', 3.3', 3.4', 3.6', $\delta_1 > \delta_2$ and $\tau < \tau_0$, as $n \to \infty$,

$$\widehat{t}(\tau)^2 \to \infty. \tag{3.61}$$

Of course the same result holds, replacing $\delta_1 > \delta_2$ and $\tau \geq \tau_0$ with $\delta_1 < \delta_2$ and $\tau \leq \tau_0$ in (ii) and replacing $\delta_1 > \delta_2$ and $\tau < \tau_0$ with $\delta_1 < \delta_2$ and $\tau > \tau_0$ in (iii), so $\widehat{t}(\tau)^2$ can be used to detect any break.

In part (i) of Theorem 3.6 we derive the limit distribution under the null: the estimates $\widehat{\delta}_1(\tau_1)$ and $\widehat{\delta}_2(\tau_2)$ when $\tau_2 \geq \tau_1$ are asymptotically independent.

In parts (ii) and (iii) we verify that the test statistic diverges under the break. We then suggest to compute $\widehat{t}(\tau)^2$ and eventually to reject $H_0 : \{\vartheta = 0\}$ if the realisation of the test statistic exceeds the critical value 3.84 (assuming of course a 5% size). This is a simple Wald test, and it only requires the estimation of $\widehat{\delta}_1(\tau)$ and $\widehat{\delta}_2(\tau)$.

Notice that if Assumption 3.6' is met then the test is consistent even if we compute $\widehat{t}(\tau)^2$ in points different than τ_0; otherwise, it requires knowledge of the location of τ_0. Yet in (iii) the test is only consistent because of the lower order bias, so the power may be rather low when computed for $\tau < \tau_0$, and in (ii) with $\tau \neq \tau_0$ $\widehat{\delta}_1(\tau)$ is still subject to a lower order bias, so the power should be higher in $\tau = \tau_0$.

Given the potential sensitivity of the power to the distance from τ_0, $\widehat{t}(\tau)^2$ seems to be particularly useful when preliminary knowledge of τ_0 is given, as it happens when analysing a shift in persistence originated by a change of policy.

If the location of the break is unknown, the test statistic should be analysed

in any potential τ in a closed subset of $(0,1)$:

$$\widehat{t}^2 = \sup_{\tau \in [\tau_l, \tau_h] \subset (0,1)} \widehat{t}(\tau)^2 . \tag{3.62}$$

Recalling the notation $B(\tau)$, $\tau \in [0,1]$, for a standardised Brownian motion on $[0,1]$, then

Theorem 3.7. *Under Assumptions 3.1', 3.2', 3.3', 3.4', 3.6', as* $n \to \infty$,

$$\begin{cases} \widehat{t}^2 \Rightarrow \sup_{\tau \in [\tau_l, \tau_h] \subset (0,1)} \frac{(B(\tau) - \tau B(1))^2}{4\tau(1-\tau)} \text{ if } \vartheta = 0 \\ \\ \widehat{t}^2 \to \infty \text{ if } \vartheta > 0. \end{cases} \tag{3.63}$$

The limit process $\sup_{\tau \in [\tau_l, \tau_h] \subset (0,1)} \frac{(B(\tau) - \tau B(1))^2}{4\tau(1-\tau)}$ is the supremum on $[\tau_l, \tau_h] \subset (0,1)$ of the square of a standardised tied down Bessel process and references for that are already in Andrews' (1993) work, where he also discussed what happens when $[\tau_l, \tau_h] = [0,1]$. Critical values can be tabulated, and indeed Andrews (1993) provided them: the critical value for the 5% test is 8.85 when $[\tau_l, \tau_h] = [0.15, 0.85]$, and 9.31 when $[\tau_l, \tau_h] = [0.1, 0.9]$; we used 9.01 when $[\tau_l, \tau_h] = [1/8, 7/8]$, interpolating as described in the original paper.

We conclude by proposing an estimate of the location of the break when there is one: letting

$$\widehat{\tau} = \arg \min_{\tau \in [\tau_l, \tau_h] \subset (0,1)} \widehat{Q}_n(\tau) = \tau \widehat{\delta}_1(\tau) + (1-\tau)\widehat{\delta}_2(\tau) \tag{3.64}$$

Theorem 3.8. *Under Assumptions 3.1, 3.2, 3.3, 3.4 with* $\delta_1 \neq \delta_2$, $\tau_0 \in [\tau_l, \tau_h]$

$$\widehat{\tau} \to_p \tau_0. \tag{3.65}$$

3.3 A Monte Carlo exercise

We investigate the validity in small samples of the theoretical results with a little Monte Carlo exercise. We considered the models:

Model 1 (M1): no break in δ, $x_t \in I(0.4)$;

Model 2 (M2): no break in δ, the variance doubles in the second part of the sample, $x_t \in I(0.4)$;

Model 3 (M3): break $\tau_0 = 1/2$: $x_{1t} \in I(0.4)$, $x_{2t} \in I(0)$;

Model 4 (M4): break $\tau_0 = 1/3$: $x_{1t} \in I(0.4)$, $x_{2t} \in I(0)$;

Model 5 (M5): break $\tau_0 = 1/2$: $x_{1t} \in I(0.2)$, $x_{2t} \in I(0)$.

Model 1 is the standard design, and we use it as a benchmark. It also provides us with a reference for $\widehat{\delta}_1(\tau)$ and $\widehat{\delta}_2(\tau)$, which are not discussed in Robinson's original paper, and for the statistics based on them, including the ones used to test for the presence of the break.

Model 2 is included to verify that changes in the short term component do not affect the quality of the estimation in small samples either. We decided to model the break in the short term dynamics with a change in the variance because, as we reviewed in the introduction, the problem received some applied and theoretical attention. We can then observe here how sensitive our techniques are to that change.

Models 3 to 5 are the ones with a break. We intend to evaluate the precision of the estimates $\widehat{\delta}$ and $\widehat{\delta}_{(2)}$ (we set $m_1 = m$, so $\widehat{\delta}_{(1)} = \widehat{\delta}$ in our design) thus appreciating the sensitivity of the bias to the bandwidths m and m_2, to the location of the break τ_0 and to the difference ϑ. We also compare the performance of the tests to detect the presence of a shift in δ, $\widetilde{t}$ and $\widehat{t}^2$, when τ_0 is unknown, and discuss the reliability of $\widehat{\tau}$ as an estimate of the location of the break.

We generated 64, 128, 256, 512 and 1024 observations, using the Davies and Harte (1987) algorithm; for each cell we simulated 1000 runs. We set $m = 0.75n^{0.79}$ and $m_2 = 0.75n^{0.49}$ and estimated $\widehat{\delta}_{(2)}$, $\widehat{\delta}$, $\widehat{\delta}_1(1/4)$, $\widehat{\delta}_2(1/4)$, $\widehat{\delta}_1(3/4)$, $\widehat{\delta}_2(3/4)$. According to the design, three outcomes are possible: root-m consistent, asymptotically normal estimation of δ_1, root-m consistent, asymptotically normal estimation of δ_2, consistent estimation of δ_1 with lower order bias. These are summarised in Table 3.1.

Table 3.1: Limit properties of the estimates used in the Monte Carlo exercise

	$\widehat{\delta}_{(2)}$	$\widehat{\delta}$	$\widehat{\delta}_1(1/4)$	$\widehat{\delta}_2(1/4)$	$\widehat{\delta}_1(3/4)$	$\widehat{\delta}_2(3/4)$
M1	A	A	A	A	A	A
M2	A	A	A	A	A	A
M3	A	X	A	X	X	B
M4	A	X	A	X	X	B
M5	X	X	A	X	X	B

A: root-m consistent and asymptotically normal estimation of δ_1;
B: root-m consistent and asymptotically normal estimation of δ_2;
X: consistent estimation of δ_1, lower order bias.

For each estimate we computed the average of the difference between the estimates and the theoretical limit value, indicating it as "bias" in Table 3.4. We also computed the sample standard deviation as a measure of the dispersion, presenting it in Table 3.5: for comparison, in Table 3.6 we report the standard deviation prescribed by the asymptotic theory as from Theorem 3.3.

Comparing the bias and the standard deviation gives a preliminary indication of the reliability of the limit normal approximation, but we also analyse it by counting the number of occurrences in which the standardised t statistic exceeded the critical value of a two sided 5% test: these are in Table 3.7. These t statistics are infeasible in the case of a break, because its location is actually unknown, but here we are only interested in appreciating the precision of the

approximation as stated in Theorem 3.3, to see when the effect of a change in δ is really negligible. In the same way, in Table 3.8 we analyse the limit normal approximation of $\widehat{\delta}_1(1/4) - \widehat{\delta}_2(1/4)$, $\widehat{\delta}_1(1/4) - \widehat{\delta}_2(3/4)$, $\widehat{\delta}_1(3/4) - \widehat{\delta}_2(3/4)$: when there is no break, root-m consistency and limit normality follow from Theorem 3.6 part (i).

Under a break, however, the same test statistics should diverge, as from Theorem 3.6 part (ii). We observe this effect in the second part of Table 3.8. We also discuss the detection of a shift in δ using the test statistics $\widetilde{t}$ (this one using a one sided alternative) and $\widehat{t}^2$ (for which we considered $\tau \in [1/8, 7/8]$): the size and power of the tests are in Table 3.9. Finally, in the last two columns of that table we present bias, dispersion and selected quantiles of the estimate of the break $\widehat{\tau}$.

In the two models without breaks in δ, the standard local Whittle estimate $\widehat{\delta}$ was more precise than the other estimates, having similar bias and smaller sample standard deviation. This is consistent with the asymptotic theory, because more information (either in terms of more frequencies or of more observations) is used.

In all the cases without breaks in δ the bias was negligible; the dispersion broadly accorded with the asymptotic one, with a couple of exceptions: $\widehat{\delta}_1(1/4)$ and $\widehat{\delta}_2(3/4)$ and, even more, $\widehat{\delta}_{(2)}$, especially when n was very little. It is possible that in these cases the samples or the number of frequencies used were so small that the asymptotic approximation was very poor: indeed with $n = 64$, $\widehat{\delta}_1(1/4)$ and $\widehat{\delta}_2(3/4)$ used only 16 observations, while for $\widehat{\delta}_{(2)}$ the bandwidth was $m_2 = 3$. In all the cases, the dispersion got closer to the theoretical value in larger samples. The empirical size of the 5% test, which we regard as the best indicator to summarize the properties of the estimates, was often slightly higher than predicted by the theory, but the only case in

which the limit approximation was too poor was for $\widehat{\delta}_{(2)}$ with few observations ($n = 64$ and $n = 128$), where it was above 20%. We conclude this part of the discussion by commenting on the fact that the performances were very similar both in Model 1 and in Model 2, and any difference appeared to be randomly generated: this is very important because it confirmed that the semiparametric procedure is robust to short term dynamics instability, which is often thought to be more frequent in real cases.

In Models 3 to 5, only $\widehat{\delta}_1\,(1/4)$ and $\widehat{\delta}_2\,(3/4)$ are unaffected by the break, the former estimating $\delta_1 = 0.4$ or $\delta_1 = 0.2$, the latter always $\delta_2 = 0$. Not surprisingly then their performances were in every respect comparable to their counterparts under Models 1 and 2.

In the other cases, Assumption 3.5' is only met by $\widehat{\delta}_{(2)}$, and only when $\delta_1 = 0.4$. The asymptotic theory (eq. (3.43)) prescribes a bigger dispersion for $\widehat{\delta}_{(2)}$ under Model 4, but the same rate of convergence: we indeed found a slightly bigger variance in the second case, but we also found a certain bias, at least for $\tau_0 = 1/3$ in the smallest samples. Although not accounted for in the asymptotic theory, we can hardly consider it as unexpected: the result stated in Theorem 3.3 rests on the presence of some observations with higher memory, but it is possible that in the smallest sample their number was just negligible, τ_0 being so small. Notice anyway that this bias disappeared in moderately sized samples, like those where $n = 512$ or above.

In any case, the approximation of the limit distribution was not hampered by the break, as indeed predicted by Theorem 3.3: the size was still slightly too large, but not more than under Models 1 and 2, and actually, possibly thanks to a more precise approximation of the variance, it was even marginally better.

When on the other hand Assumption 3.5' was not met, the lower order bias was a dominant feature. Consider, in particular, $\widehat{\delta}_{(2)}$ under Models 3 and 5:

the two cases only differ for ϑ, but have the same τ_0 and m_2, and we found a much bigger bias when Assumption 3.5' was not met, regardless of the fact that δ_1 was much closer to δ_2 in Model 5 than in Model 3.

The effect of a change in ϑ was less clear when Assumption 3.5' was not met anyway: for $\widehat{\delta}$, $\widehat{\delta}_2\,(1/4)$ and $\widehat{\delta}_1\,(3/4)$ the bias happened in our exercise to be roughly the same both in Model 3 and in Model 5.

Contrary to the case of $\widehat{\delta}_{(2)}$, the different τ_0 between Models 3 and 4 should give a bigger bias in the second case, as from Theorem 3.4, and this proved indeed to be the case.

One can also appreciate this effect within each model, by comparing $\widehat{\delta}_2\,(1/4)$, $\widehat{\delta}$ and $\widehat{\delta}_1\,(3/4)$: the lowest proportion of observations with high δ was in the first estimate, the biggest in the last one, and the bias is ranked accordingly.

It is worth noticing that when Assumption 3.5' was not met, then the lower order bias still had a strong effect even with 1024 observations, its reduction proceeding only rather slowly as n increased.

Finally, we compare the bias between $\widehat{\delta}_{(2)}$ and $\widehat{\delta}$: according to the theory, given that Assumption 3.5' is not met for $\widehat{\delta}$, it should always have a bigger bias. This, anyway, only appeared in Models 3 and 4, while no clear ranking emerged for Model 5.

Overall, the lower order bias was more important the smaller the sample and the larger the percentage of frequencies with δ_2 compared to δ_1; in most of the cases it was also larger the bigger the bandwidth, while only in a few cases the difference ϑ mattered.

In all these cases it was also sufficient to cause a clear failure of the normal approximation of the t statistic, the effective size depending on the comparison between the bias and the dispersion; this was more clear for $\widehat{\delta}$, where the bias was very large with respect to the standard deviation, and much less for $\widehat{\delta}_{(2)}$ when $\delta = 0.2$, where the variance was so large anyway that no size distortion

at all could be appreciated.

In the last block of results we deal with detecting and estimating a break in δ.

We begin again discussing the case in which there is no break in δ. The limit distribution of $\widehat{\delta}_1(\tau_1) - \widehat{\delta}_2(\tau_2)$ should then be normal, as per Theorem 3.6 (i), and indeed in all the three combinations it resulted in being compatible with the theory. The empirical size only exceeded 5% slightly, if at all: it reached a maximum of 11%, and the approximation improved quickly with the dimension of the sample. A similar pattern emerged for the approximation of the limit distributions of the two statistics $\widetilde{t}$, $\widehat{t}^2$: the latter anyway was more precise, the empirical size being closer to the theoretical 5%. The worse performance in terms of size of $\widetilde{t}$ probably depended on the poor approximation of the variance of $\widehat{\delta}_{(2)}$, which had too big a dispersion. Notice, again, that neither of these statistics was sensitive to breaks in the short memory component.

Under the break in δ, all the test statistics diverged, albeit the performances were rather different.

All the three differences $\widehat{\delta}_1(\tau_1) - \widehat{\delta}_2(\tau_2)$ detected the presence of the break satisfactorily or well. One can view the case $\tau_1 = \tau_2 \neq \tau_0$ as the case in which the researcher has some information about the breakpoint, but that is not exact: as we saw when discussing Theorem 3.6, the test statistic may then diverge either because $\widehat{\delta}_1(\tau_1)$ estimates the higher δ_1 while $\widehat{\delta}_2(\tau_2)$ estimates the lower δ_2 (Theorem 3.6, (ii)), or because although both the estimates converge to δ_1, one of them is subject to the lower order bias (Theorem 3.6, (iii)), and in our result this difference was relevant, as can be seen by comparing the cases $\tau = 1/4$ and $\tau = 3/4$. Also notice that when the guess about the location of the break was more precise, as in Model 4 compared to Model 3, the power was higher.

It is better to try to choose τ in such a way that the two estimates converge to different limits, as in Theorem 3.6, (ii): one way of doing it is by setting $\tau_1 \neq \tau_2$ and by keeping them very distant (we put $\tau_1 = 1/4$, $\tau_2 = 3/4$, but one could consider 1/8 and 7/8 instead). Yet of course the researcher has to trade off this with the larger variance associated with the fact that many fewer observations are used: notice that even with the rather moderate 1/4, 3/4 split, the power is less than if $\widehat{t}^2$ is used.

The test based on $\widetilde{t}$ performed badly, the power being at most 35% even in the largest sample. This is far from surprising: even for Models 3 and 4, where we did actually observe a certain difference between $\widehat{\delta}_{(2)}$ and $\widehat{\delta}$, the difference was still relatively small, and always well within one standard deviation (this can be found in Table 3.6 in the case Model 1). That test fared even worse when δ_1 was so small that $\widehat{\delta}_{(2)}$ too was subject to the lower order bias: in that case the test had nearly no power at all. As a result we then think that a test based on $\widetilde{t}$ should only be preferred in very large samples, when the power can be reliable and the burden to compute the test based on $\widehat{t}$ may be excessive. In samples of dimension comparable to the ones we used, a rejection of the null of no break in δ based on $\widetilde{t}$ could depend on a particularly poor estimate $\widehat{\delta}_{(2)}$ (recall how much the sample variance exceeded the theoretical one) as easily as on an effective break in δ.

Finally, the test based on $\widehat{t}^2$, of which the elements $\widehat{\delta}_1(\tau_1) - \widehat{\delta}_2(\tau_2)$ are the building blocks. The power of the $\widehat{t}^2$ test was higher the larger the gap ϑ, and it increased with the sample size. Not surprisingly, given that the test statistic is computed for each τ, the test was not sensitive to the location of τ_0. As we already conjectured in the discussion of Theorems 3.6 and 3.7, the $\widehat{t}^2$ test proved to be the most powerful when the location of the break is unknown, especially when the sample was small or the gap ϑ little.

When evidence of a break has been established, its location may be estimated by $\widehat{\tau}$. We report, as usual, the bias and the sample standard deviation of $\widehat{\tau}$. Anyway, since we did not derive the limit distribution of $\widehat{\tau}$, nor we established existence of the second moment, we supplement this information with three nonparametric measures: the first and last 5% quantile and the median.

The median was always centred on the correct location of the breakpoint, and the distribution collapsed on it rather quickly. It can be suspected that the estimation of the location of the break is more difficult when ϑ is relatively small, and indeed the range between the top and last 5% quantile was bigger in our experiment when $\delta_1 = 0.2$ than when $\delta_1 = 0.4$; the position of τ_0 on the other hand did not affect the estimates.

This information is mirrored in the bias and standard deviation. Indeed, if we were to compute the first and last 5% quantile using the sample averages and standard deviations as if under normality we would get approximately the same intervals.

Overall, $\widehat{\tau}$ seemed satisfactory unless the sample was very small or the gap very little (although in this last case it does not seem to be a big loss, because the mistake made using the wrong δ is relatively small).

3.4 Inflation persistence in the euro-area

In this section we study the persistency in inflation over the years 1972-2004 for the countries that constitute the European Monetary Union (EMU).

A number of events that could have potentially caused a structural change in the long term dynamics of inflation took place during these years.

Some of these are indeed changes of policy regimes. The first European Exchange Rate Mechanism (ERM) was established in March 1979, and the

central parities had often been revised over time; the whole ERM actually changed, increasing the number of countries participating in it: some of the current members of the euro-area were not even members of the European Union in 1979, let alone of the ERM. Moreover, monetary policy was managed locally by the national central banks, so other potential breakpoints could be considered, each one different for each country.

Other potential structural changes did not originate from monetary policy decisions: the two oil shocks, or the major exchange rate crisis in 1992, might have affected the persistence of the long term dynamics of inflation; again, local, country specific shocks, intervened as well.

Undoubtedly also other countries experienced potential shocks in the past, so we motivate our preference for this particular example with the peculiarity of the Eurosystem itself: its whole existence depends on the assumption that it can do a better job at controlling inflation than the banks that conferred their powers to it. Several arguments could be proposed to explain it: first, by fostering the integration of the markets, the monetary union should favour the transmission of policy impulses; second, the institution of a single authority implicitly removes the possibility of conflicting policies between different countries (and the incentive to do so); third, it is often argued that some of the monetary authorities that the Eurosystem replaced lacked the credibility (or the appetite) for strong control of inflation.

The Eurosystem seems to be well aware of the importance of studying inflation persistence in the euro-area, and it stimulated applied research on the topic. The results, however, are apparently inconclusive: Angeloni, Aucremanne, Ehrmann, Gali, Levin and Smets (2004) associated inflation persistence with unstable monetary regimes (in which case less persistence should be found in the last part of the sample, especially after 1999), but O'Reilly and Whelan (2004) argued that persistence remained fairly stable and the

institution of the Eurosystem did not *per se* reduce it.

We then analysed the inflation figures for these years considering the potential breakpoints as unknown, but allowing for a major change in policy in 1999, when the authority to decide monetary policy was transferred from the local central bank to the Eurosystem. We used January 1999 for Greece as well, although it only joined the EMU the following year: it can be argued that Greece already benefited from increased stability at that point, because in 1999 it was clear it was about to join the EMU soon anyway; besides, as we saw when discussing Theorem 3.6, the test detects the presence of a break consistently even computed at $\tau \neq \tau_0$.

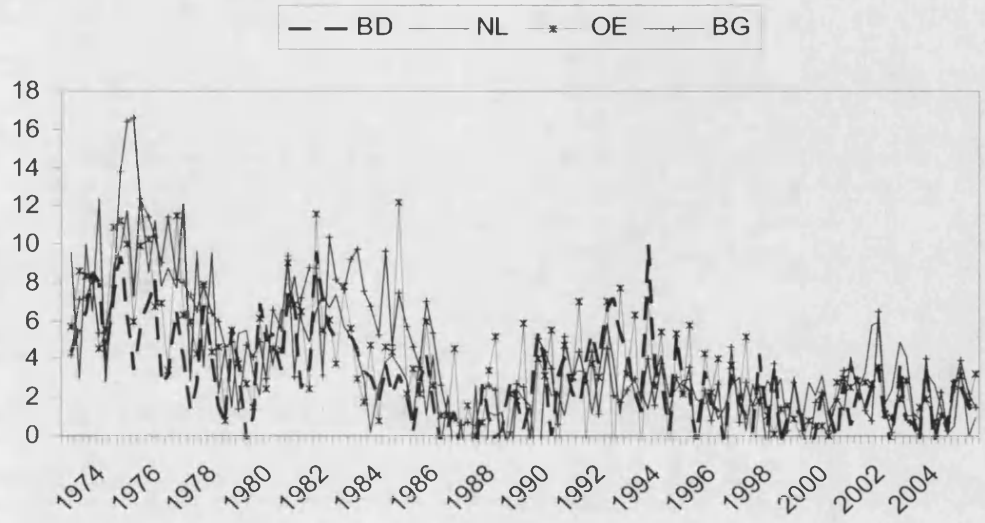
We analysed monthly inflation in eleven of the countries constituting the EMU: these are Portugal (PT), Spain (ES), France (FR), Ireland (IR), Italy (IT), Belgium (BE), Netherlands (NL), Germany (BD), Finland (FN), Greece (GR) and Austria (OE) (within brackets we indicated the way we shortened the names in the figures and in the tables). The data were computed from price indices collected from Datastream: these have codes PTCONPRCF, ESCONPRCF, FRCP....F, IRCONPRCF, ITCONPRCF, BGCONPRCF, NLCONPRCF, BDCONPRCF, FNCONP95F, GRCONPRCF, OECP..96F. We did not include Luxembourg because not enough data were available.

The dataset covers the years 1972-2004 (inclusive), so $n = 395$ when levels of inflation were considered, and $n = 394$ when first differences were computed. The starting date depended on the availability of data for France on Datastream: we preferred to have all the samples covering the same period to make comparison easier.

The plots of inflation are in Figures 3.1 to 3.3.

These figures have been obtained by transforming the frequency to quarterly, where for each quarter the price index was obtained by averaging the

Figure 3.1: Inflation: BD, NL, OE, BG

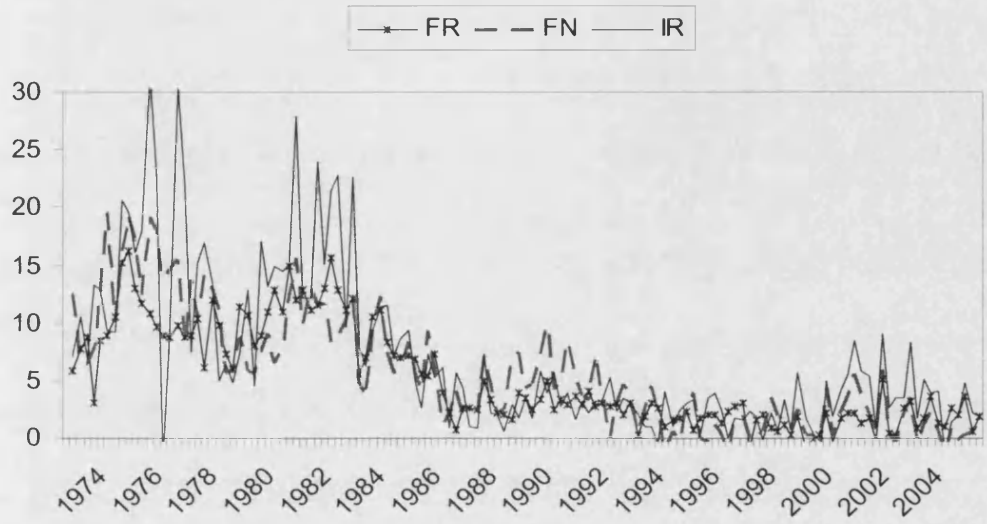


monthly index for each country: we then computed the growth rate by taking first differences of the logarithms, and then multiplied the result by 400 in order to get a measure of the inflation at annual rate. This was only done in order to remove from the figures some short term volatility, thus making it simpler to observe the long run dynamics: we used monthly data in the empirical analysis, inflation being computed as the first difference of the logarithm of the price index.

We analysed the data by estimating $\widehat{\delta}_{(2)}$ and $\widehat{\delta}$, by computing the test statistics $\widetilde{t}$ and $\widehat{t}^2$ to detect a break and eventually by estimating $\widehat{\tau}$. We kept $\tau \in [1/8, 7/8]$, so we searched for a break approximately between 1976 and 2000.

We also estimated $\widehat{\delta}_1(\tau_{99})$, $\widehat{\delta}_2(\tau_{99})$ and then computed $\widehat{t}(\tau_{99})^2$, where τ_{99} is the τ to test for a break in January 1999. This point was already in the set in which we considered a potential break, but by using $\widehat{t}(\tau_{99})^2$ we treated it differently because we then assumed that the potential breakpoint was known: the limit distribution of $\widehat{t}(\tau_{99})^2$ is a simple χ^2_1, with a much smaller critical value.

Figure 3.2: Inflation: FR, FN, IR



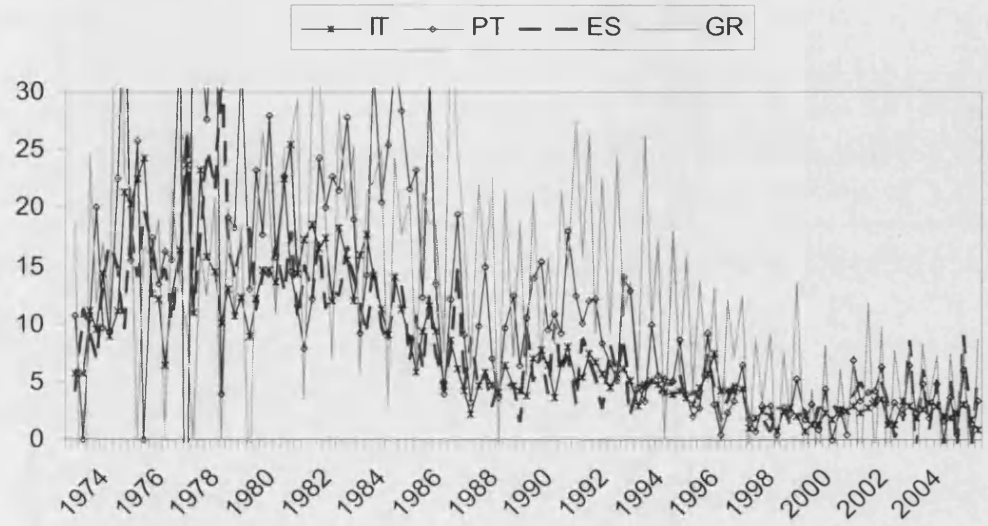
Contrary to the Monte Carlo exercise, we set a rather conservative bandwidth, $m = 0.28n^{0.79} = 31$ and $m_2 = 0.28n^{0.49} = 5$, in order avoid the influence of short term dynamics, especially considering the strong seasonal component between $m = 32$ and $m = 33$. Since the estimated order of integration was in some cases potentially high, we also analysed the first differences of the data.

All the results are presented in Table 3.2.

Notice that if inflation has a non-zero mean, the estimation of $\delta(\tau)$ from data in levels may be subject to a lower order bias because, by truncating the sample and padding it with zeros, we induced a change in the mean. This could have been corrected by taking mean-differences before truncating the sample, but we think it was not necessary in our example: as we saw in Chapter 2, the potential lower order bias is going to be a problem only when δ is relatively small, a case that, judging on the full sample estimates $\widehat{\delta}$, we can safely exclude.

The estimates $\widehat{\delta}$ of the orders of integration ranged between 0.48 for Germany and 0.83 for Italy and Finland. Germany was also the only country with

Figure 3.3: Inflation: IT, PT, ES, GR



all the estimates ($\widehat{\delta}$, $\widehat{\delta}_{(2)}$, $\widehat{\delta}_1(\tau_{99})$ and $\widehat{\delta}_2(\tau_{99})$) below 0.5, thus not requiring first differences to be considered. In general, we used level or first differences according to whether the estimated value was in the interval (-0.5,0.75) or not, because this is the range for which Robinson (1995b) and Velasco (1999b) established root-m consistency and limit normal distribution; in the cases in which we had to confront estimates, as in the tests $\widehat{t}^2$, $\widehat{t}(\tau_{99})^2$ and $\widetilde{t}$, the decision to refer to either levels or first differences also depended on the other estimates, and we then considered levels only for Greece and Austria, besides the aforementioned Germany.

Overall, it seems that inflation persistence remained stable over time, rather than showing the decay that the Eurosystem might have expected. Indeed, the only country for which the hypothesis of stability was rejected with the Andrews-type test $\widehat{t}^2$ at the 5% size was Italy, although for France too the test statistic was very high, and actually significant at 10% size (we used 7.33 as the critical value, interpolating it from the tables in Andrews (1993)). Notice that neither for Italy nor France did we consider the test statistic computed

Table 3.2: Estimates of persistence and stability tests for EMU countries

		$\widehat{\delta}_{(2)}$	$\widehat{\delta}$	$\widehat{\delta}_1(\tau_{99})$	$\widehat{\delta}_2(\tau_{99})$	$\widehat{t}^2(\tau_{99})$	$\widehat{t}^2$	$\widetilde{t}$	$\widehat{\tau}n$
BD	levels	0.26	0.48	0.47	0.46	0.00	2.09	-0.95	342
	first diff.	-0.37	-0.34	-0.70	-0.20	4.27	9.79	-0.11	345
IT	levels	0.93	0.83	0.87	0.71	0.39	2.47	0.44	124
	first diff.	-0.04	-0.12	-0.13	-0.59	3.56	13.60	0.39	122
FR	levels	0.96	0.80	0.84	0.66	0.53	5.51	0.73	154
	first diff.	-0.10	-0.26	-0.27	-0.76	4.09	7.87	0.74	153
PT	levels	0.99	0.53	0.55	0.86	1.64	3.91	2.07	258
	first diff.	-0.76	-0.48	-0.50	-0.10	2.66	6.58	-1.25	49
ES	levels	1.47	0.62	0.66	0.72	0.05	1.35	3.80	170
	first diff.	0.45	-0.39	-0.32	-0.29	0.01	5.03	3.75	60
NL	levels	0.53	0.79	0.84	0.51	1.92	6.74	-1.17	118
	first diff.	0.05	-0.29	-0.32	-0.17	0.35	3.17	1.52	341
FN	levels	0.59	0.83	0.80	0.53	1.27	7.26	-1.04	221
	first diff.	-0.64	-0.30	-0.30	-0.53	0.84	7.07	-1.55	72
IR	levels	0.61	0.57	0.61	0.70	0.16	4.25	0.16	117
	first diff.	-0.37	-0.42	-0.36	-0.40	0.02	8.43	0.23	142
GR	levels	0.84	0.47	0.49	0.43	0.05	0.93	1.62	340
	first diff.	-0.15	-0.48	-0.52	-0.64	0.28	6.86	1.47	67
OE	levels	0.52	0.54	0.60	0.71	0.20	2.92	-0.09	139
	first diff.	-0.59	-0.54	-0.66	-0.23	3.13	7.12	-0.25	237
BG	levels	0.44	0.67	0.75	0.43	1.76	8.81	-1.03	153
	first diff.	-1.04	-0.52	-0.41	-0.27	0.32	1.28	-2.33	83

for the levels of inflation, because the point estimates $\widehat{\delta}$ were quite above 0.75. The interpretation of the estimates was less straightforward for Ireland and Belgium, because on the basis of the estimated values $\widehat{\delta}$ one could conjecture $\delta \in (1/2, 3/4)$ if there is no break, so the results should have been similar whether levels or first differences are applied, but in the case of the $\widehat{t}^2$ test they were conflicting. We decided to consider the possibility of a break anyway, and we estimated the location of the potential breakpoint for these two countries too.

The point estimates of the breaks, $\widehat{\tau}n$, were all concentrated in the first part of the sample: mid 1982 for Italy, late 1983 for Ireland and late 1984 for France and Belgium. Notice these were not actually discussed in Section 3.2,

where we only derived consistency of $\widehat{\tau}$ but not the rate of convergence to τ. Yet we think that the estimation and comparison of the estimates before and after the potential break may still be of interest on occasions, especially when, as in this case, additional information is available. It is generally agreed that if a change in δ took place, this was for a reduction of persistence: even if the location of the break was not correctly estimated, then, one should have expected $\widehat{\delta}_1(\tau) > \widehat{\delta}_2(\tau)$. We present the results in Table 3.3: in all the cases the break was indeed associated with a reduction in the order of integration and, especially for France and Italy, with a rather large one.

The test for a break in 1999 was only significant for Italy (at 10%) and France (at 5%), but whether it really reflected a change in δ at that point or responded to the same break detected by the $\widehat{t}^2$ statistic cannot be concluded on the basis of Table 3.2 alone. Table 3.3 however provided an additional piece of relevant information: the estimated persistence after the 1982/1984 break was approximately the same as that in the sample for 1999 onwards only, so one should not expect another break in 1999.

Regardless of this we proceeded to repeat the analysis for France and Italy, solely on the sample 1985-2004 and still testing for a break in January 1999: again the reliability of the test depends on the precision of the estimation of the month of the break, but notice that given the assumption of a non increasing δ, if the actual break took place before 1985 then $\widehat{\delta}_1(\tau_{99})$ still estimated δ correctly, while if the break took place after that date then $\widehat{\delta}_1(\tau_{99})$ overestimated δ_2 and then yielded a larger $\widehat{t}(\tau_{99})^2$, thus making the rejection of the null of no change in persistence more likely. Setting $m = 19$ (slightly stronger than $m = 0.28n^{0.79}$, but necessary to exclude the peak of seasonality), the realised $\widehat{t}(\tau_{99})^2$ statistics were 0.01 and 0.12 for levels and first differences for Italy, and 0.31 and 0.42 for levels and first differences for France, so the hypothesis

of a break in 1999 was clearly rejected.

Table 3.3: Estimates of persistence on selected periods

		$\widehat{\delta}_1(\tau_b)$	$\widehat{\delta}_2(\tau_b)$
IT	first diff.	0.91	0.65
	first diff	-0.14	-0.48
FR	level	0.91	0.49
	first diff	-0.22	-0.69
IR	level	0.63	0.57
	first diff.	-0.39	-0.53
BG	level	0.91	0.36
	first diff.	-0.23	-0.44

The breakpoint $\tau_b n$ is 122 for Italy,
153 for France and Belgium, 142 for Ireland.

Turning to $\widetilde{t}$ (see again Table 3.2), given the very small bandwidth ($m_2 = 5$ only) and the small power in the Monte Carlo exercise, we did not expect the test statistic to be significant for any country. It was on the contrary significant at 5% for Portugal and Spain, and it was just below the critical value for Greece (although well above the 10% critical value). This is even more surprising considering that the more powerful test $\widehat{t}^2$ failed to detect any change in persistence in these cases. We then interpreted the results in a different way. Greece Portugal and Spain were the countries characterised by the highest volatility and average inflation, within the sample: the plot of their inflation shows a remarkably strong drop in the levels, from about 20% or more to 4% or less. Taking into account the analysis we run in Chapter 2 about the effect on low frequencies of a time-varying deterministic component, then, we think that $\widetilde{t}$ detected that drop in the mean rather than a change in δ.

Summarising, according to our results there has been a certain reduction in the persistence of the shocks to inflation after 1972, but this seems to be

more related to the initial inflation stabilisation than to the centralisation of the monetary policy under the Eurosystem.

3.5 Discussion

We have studied the local Whittle estimate of the memory parameter in the presence of a structural break in the stochastic component.

We have found that the semiparametric design has the advantage of being robust to any form of instability in the short memory component. When the order of integration itself changes, the highest one is estimated; the lower order of integration may at most induce a lower order bias, which can be avoided by removing the highest frequencies. We have proposed two tests to detect a change in the long term dynamics, but the Monte Carlo exercise and the empirical application both showed that only the Chow-type test with unknown breakpoint is really reliable, the one based on the comparison of estimates with different bandwidths having low power and being too sensitive to time-varying deterministic components. Upon having knowledge of a break in δ, we have also proposed a consistent estimate of the breakpoint.

1. Some of the comments, conjectures and potential extensions that we discussed in Chapter 2 are valid here as well. We only considered a Type I integrated process, but we expect that all the results carry through for a Type II as well. Also, we focused on the range of δ more often considered in the literature, but we think that a wider range for δ could be treated, following Velasco (1999b): indeed we assumed this result when discussing the empirical example. Finally, we discussed the local Whittle estimate, but we also think that the same results apply for the log-periodogram regression estimate. Contrary to the case of a time-varying

mean, though, we do not expect that the results discussed in this chapter apply to fully parametric estimates: we think that this difference may really motivate a preference for the semiparametric approach in some cases.

2. The theoretical literature seems to treat the two problems of detecting a break and of estimating its location as distinct: Andrews (1993) for example discussed the former but not the latter. We have followed the same approach: we have only proposed a consistent estimate of the breakpoint τ. It would be interesting to provide a rate of convergence for $\widehat{\tau}$, and possibly to compare it to the maximum likelihood estimate of the breakpoint discussed by Bai and Perron (1998).

3.6 Appendix to Chapter 3

Proof of Theorem 3.1. Consider $E\left(I_{\xi 1}\left(\lambda_j\right)\right)$ first. The expectation is

$$\begin{aligned} E\left(I_{\xi 1}\left(\lambda_j\right)\right) &= E\left[\left(\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{[\tau_0 n]}\xi_{1t}e^{i\lambda_j t}\right)\left(\frac{1}{\sqrt{2\pi n}}\sum_{s=1}^{[\tau_0 n]}\xi_{1s}e^{-i\lambda_j s}\right)\right] \\ &= \frac{1}{2\pi n}\sum_{t,s=1}^{[\tau_0 n]}\gamma_{\xi 1}\left(t-s\right)e^{i\lambda_j(t-s)} \end{aligned} \tag{3.66}$$

where $\gamma_{\xi 1}\left(k\right)$ is the autocovariance. Since $\gamma_{\xi 1}\left(k\right)=\int_{-\pi}^{\pi}f_{\xi 1}\left(\lambda\right)e^{-i\lambda k}d\lambda$,

$$E\left(I_{\xi 1}\left(\lambda_j\right)\right)=\frac{1}{2\pi n}\sum_{t,s=1}^{[\tau_0 n]}\left(\int_{-\pi}^{\pi}f_{\xi 1}\left(\lambda\right)e^{-i\lambda(t-s)}d\lambda\right)e^{i\lambda_j(t-s)} \tag{3.67}$$

which we rewrite as

$$\left(\int_{-\pi}^{\pi}\left(f_{\xi 1}(\lambda)-f_{\xi 1}(\lambda_j)\right)\frac{1}{2\pi n}\sum_{t,s=1}^{[\tau_0 n]}e^{-i(\lambda-\lambda_j)(t-s)}d\lambda\right) \tag{3.68}$$

$$+\left(\int_{-\pi}^{\pi}f_{\xi 1}(\lambda_j)\frac{1}{2\pi n}\sum_{t,s=1}^{[\tau_0 n]}e^{-i(\lambda-\lambda_j)(t-s)}d\lambda\right). \tag{3.69}$$

The term in (3.69) is

$$f_{\xi 1}(\lambda_j)\frac{1}{2\pi n}\int_{-\pi}^{\pi}\sum_{|r|\leq[\tau_0 n]}\left([\tau_0 n]-|r|\right)e^{-i(\lambda-\lambda_j)r}d\lambda$$

$$= f_{\xi 1}(\lambda_j)\frac{1}{2\pi n}\left([\tau_0 n]\sum_{|r|\leq[\tau_0 n]}\int_{-\pi}^{\pi}e^{-i(\lambda-\lambda_j)r}d\lambda-\sum_{|r|\leq[\tau_0 n]}\int_{-\pi}^{\pi}|r|\,e^{-i(\lambda-\lambda_j)r}d\lambda\right) \tag{3.70}$$

$$= f_{\xi 1}(\lambda_j)\frac{1}{2\pi n}[\tau_0 n]\,2\pi=\tau_0^{\triangle}f_{\xi 1}(\lambda_j) \tag{3.71}$$

where we used $\int_{-\pi}^{\pi}e^{-i(\lambda-\lambda_j)r}d\lambda=0$ for any $r\neq 0$. To show that (3.68) is $O\left(j^{-1}\ln j\lambda_j^{-2d}\right)$ we notice that, since the absolute value of the Dirichlet kernel is bounded by $O\left(|\lambda|^{-1}\right)$ at any non-zero frequency, we can still follow the proof of Robinson (1995a).

The same argument can be applied for $E\left(I_{\xi 2}(\lambda_j)\right)$.

Finally, for $E\left(I_{\xi 12}(\lambda_j)\right)$, this is

$$E\left(I_{\xi 12}(\lambda_j)\right) = E\left[\left(\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{[\tau_0 n]}\xi_{1t}e^{i\lambda_j t}\right)\left(\frac{1}{\sqrt{2\pi n}}\sum_{s=[\tau_0 n]+1}^{n}\xi_{2s}e^{-i\lambda_j s}\right)\right]$$

$$= \frac{1}{2\pi n}\sum_{t=1}^{[\tau_0 n]}\sum_{s=[\tau_0 n]+1}^{n}\gamma_{\xi 12}(t-s)\,e^{i\lambda_j(t-s)} \tag{3.72}$$

$$= \frac{1}{2\pi n}\int_{-\pi}^{\pi}\sum_{t=1}^{[\tau_0 n]}\sum_{s=[\tau_0 n]+1}^{n}\left(f_{\xi 12}(\lambda)-f_{\xi 12}(\lambda_j)\right)e^{i(\lambda_j-\lambda)(t-s)}d\lambda \quad (3.73)$$

$$+\frac{1}{2\pi n}\int_{-\pi}^{\pi}\sum_{t=1}^{\tau_0 n}\sum_{s=\tau_0 n+1}^{n} f_{\xi 12}(\lambda_j)\, e^{i(\lambda_j-\lambda)(t-s)}d\lambda \quad (3.74)$$

and notice that (3.74) is 0 because $f_{\xi 12}(\lambda_j)\sum_{t=1}^{\tau_0 n}\sum_{s=\tau_0 n+1}^{n}\int_{-\pi}^{\pi} e^{i(\lambda_j-\lambda)(t-s)}d\lambda = 0$ since $t > s$. The result then follows using $\left|\sum_{s=u}^{v} e^{is\lambda}\right| = O\left(|\lambda|^{-1}\right)$ as for example in Robinson and Marinucci (2001), Lemma 3.2, and following again the proof in Robinson (1995a).

Proof of Theorem 3.2. We consider $Plim\ \widehat{\delta}_2(\tau)$ when $\tau < \tau_0$, the other cases can be treated in the same way. We follow the proof in Robinson (1995b), Theorem 1, replacing $H - 1/2$, $H_0 - 1/2$ and G_0 by d, δ_1 and $\left(\tau_0^{\Delta} - \tau^{\Delta}\right) G_{\xi 1}$ respectively; we refer to the original article for a definition of Θ_1 and Θ_2 and of $S(d)$. Introduce

$$\zeta_{2t} = \begin{cases} \xi_{1t} \text{ if } [\tau n]+1 \leq t < [\tau_0 n] \\ \\ 0 \text{ otherwise} \end{cases} \quad (3.75)$$

and

$$g_{\zeta 2}(\lambda_j) = \left(\tau_0^{\Delta} - \tau^{\Delta}\right) G_{\xi 1}\lambda^{-2\delta_1}, \quad (3.76)$$

and consider the set Θ_1 first. We can follow the proof in the original paper up to Robinson's equation (3.13): when $\Theta = \Theta_1$ then $\widehat{\delta}_2(\tau) \to_p \delta_1$ as $n \to \infty$ if

$$6\sum_{r=1}^{m-1}\left(\frac{r}{m}\right)^{2(\Delta-\delta_1)+1}\frac{1}{r^2}\left|\sum_{j=1}^{r}\left(\frac{I_{z2}(\lambda_j)}{g_{\zeta 2}(\lambda_j)}-1\right)\right| \quad (3.77)$$

is $o_p(1)$. We then Rewrite (3.14) of Robinson as

$$\begin{aligned}
&\frac{I_{z2}(\lambda_j)}{g_{\zeta 2}(\lambda_j)} - 1 \\
= &\left(1 - \frac{g_{\zeta 2}(\lambda_j)}{f_{\zeta 2}(\lambda_j)}\right) \frac{I_{z2}(\lambda_j)}{g_{\zeta 2}(\lambda_j)} \\
&+ \frac{1}{f_{\zeta 2}(\lambda_j)} \left(I_{z2}(\lambda_j) - |\alpha_{1j}|^2 I_{\varepsilon}(\lambda_j)\right) \\
&+ \left(\frac{2\pi}{\left(\tau_0^{\Delta} - \tau^{\Delta}\right)} I_{\varepsilon}(\lambda_j) - 1\right) \qquad (3.78)
\end{aligned}$$

where

$$\alpha_{1j} = \sum_{l=0}^{\infty} \alpha_{1l} e^{i\lambda_j l},\ f_{\zeta 2}(\lambda) = \left(\tau_0^{\Delta} - \tau^{\Delta}\right) f_{\xi 1}(\lambda) \text{ and } I_{\varepsilon}(\lambda) = |F_{\varepsilon}(\lambda)|^2, \qquad (3.79)$$

$$\varepsilon_t = \begin{cases} \varepsilon_{1t} \text{ if } [\tau n] + 1 \leq t < [\tau_0 n] \\ \\ 0 \text{ otherwise.} \end{cases} \qquad (3.80)$$

Notice that $f_{\zeta 2}(\lambda)$ is not actually a spectral density because $\zeta 2_t$ is not stationary. The result

$$E\left|\frac{I_{z2}(\lambda_j)}{g_{\zeta 2}(\lambda_j)}\right| \leq C \quad j = 1, ..., m \qquad (3.81)$$

for a generic, positive finite constant C still follows using (3.19), so

$$6\sum_{r=1}^{m-1} \left(\frac{r}{m}\right)^{2(\Delta - \delta_1)+1} \frac{1}{r^2} \left|\sum_{j=1}^{r} \left(1 - \frac{g_{\zeta 2}(\lambda_j)}{f_{\zeta 2}(\lambda_j)}\right) \frac{I_{z2}(\lambda_j)}{g_{\zeta 2}(\lambda_j)}\right| \leq \frac{C\eta}{2(\Delta - \delta_1) + 1} \qquad (3.82)$$

for any $\eta > 0$. Next rewrite

$$\begin{aligned}
E\left|I_{z2}(\lambda_j) - |\alpha_{1j}|^2 I_{\varepsilon}(\lambda_j)\right| \leq\ & E\left|I_{\zeta 2}(\lambda_j) - |\alpha_{1j}|^2 I_{\varepsilon}(\lambda_j)\right| \qquad (3.83) \\
&+ 2E\left|\mathrm{Re}\, I_{\zeta 2x2}(\lambda_j)\right| + E\left|I_{x2}(\lambda_j)\right| \qquad (3.84)
\end{aligned}$$

where $I_{\zeta 2}(\lambda)$ is the periodogram of ζ_{2t} and $I_{\zeta 2x2}(\lambda) = F_{\zeta 2}(\lambda) F_{x2}(-\lambda)$.

The contribution of $I_{\zeta 2}(\lambda_j) - |\alpha_{1j}|^2 I_\varepsilon(\lambda_j)$ can be discussed as in equation (3.17) from Robinson, using the same argument as in Theorem 3.1 to show that $\left|I_{\zeta 2}(\lambda_j) - |\alpha_{1j}|^2 I_\varepsilon(\lambda_j)\right| = O_p\left(f_{\zeta 2}(\lambda_j)(\ln(j+1)/j)^{1/2}\right)$ and

$$E\left\{6\sum_{r=1}^{m-1}\left(\frac{r}{m}\right)^{2(\Delta-\delta_1)+1}\frac{1}{r^2}\left|\sum_{j=1}^{r} f_{\zeta 2}^{-1}(\lambda_j)\left(I_{\zeta 2}(\lambda_j) - |\alpha_{1j}|^2 I_\varepsilon(\lambda_j)\right)\right|\right\} = o(1) \tag{3.85}$$

following the same steps of Robinson. We are then left with

$$6\sum_{r=1}^{m-1}\left(\frac{r}{m}\right)^{2(\Delta-\delta_1)+1}\frac{1}{r^2}\left|\sum_{j=1}^{r} f_{\zeta 2}^{-1}(\lambda_j)\left(2\,\mathrm{Re}\, I_{\zeta 2x2}(\lambda_j) + I_{x2}(\lambda_j)\right)\right| \tag{3.86}$$

which we bound using

$$6\sum_{r=1}^{m-1}\left(\frac{r}{m}\right)^{2(\Delta-\delta_1)+1}\frac{1}{r^2}\left|\sum_{j=1}^{r} f_{\zeta 2}^{-1}(\lambda_j)\, 2\,\mathrm{Re}\, I_{\zeta 2x2}(\lambda_j)\right| \tag{3.87}$$

$$+6\sum_{r=1}^{m-1}\left(\frac{r}{m}\right)^{2(\Delta-\delta_1)+1}\frac{1}{r^2}\left|\sum_{j=1}^{r} f_{\zeta 2}^{-1}(\lambda_j) I_{x2}(\lambda_j)\right|. \tag{3.88}$$

The order of magnitude of (3.87) is

$$= O\left(\sum_{r=1}^{m-1}\left(\frac{r}{m}\right)^{2(\Delta-\delta_1)+1}\frac{1}{r^2}\sum_{j=1}^{r}\left(\frac{j}{n}\right)^{2\delta_1}\left(\frac{j}{n}\right)^{-\delta_1-\delta_2} j^{-1}\ln j\right) \tag{3.89}$$

$$= \begin{cases} O\left(\left(\frac{1}{n}\right)^{\delta_1-\delta_2}\left(\frac{1}{m}\right)^{2(\Delta-\delta_1)+1}\right) \text{ if } 2\Delta-\delta_1-\delta_2<0 \\ O\left(\left(\frac{m}{n}\right)^{\delta_1-\delta_2}\frac{\ln^2 m}{m}\right) \text{ if } 2\Delta-\delta_1-\delta_2=0 \\ O\left(\left(\frac{m}{n}\right)^{\delta_1-\delta_2}\frac{\ln m}{m}\right) \text{ if } 2\Delta-\delta_1-\delta_2>0 \end{cases} \tag{3.90}$$

and all the elements in (3.90) are $o_p(1)$ using $\delta_2 < \delta_1$ and $m/n \to 0$ (also notice that $\delta_1 < \Delta + 1/2$). In a similar way

$$6\sum_{r=1}^{m-1}\left(\frac{r}{m}\right)^{2(\Delta-\delta_1)+1}\frac{1}{r^2}\sum_{j=1}^{r} f_{\zeta 2}^{-1}(\lambda_j)\, E\left|I_{\xi 2}(\lambda_j)\right| = O\left((m/n)^{2(\delta_1-\delta_2)}\right). \tag{3.91}$$

To deal with the final contribution, we notice that

$$\frac{2\pi}{\left(\tau_0^{\triangle}-\tau^{\triangle}\right)}I_{\varepsilon}\left(\lambda_j\right)-1 \tag{3.92}$$

$$= \frac{1}{\left(\tau_0^{\triangle}-\tau^{\triangle}\right)n}\sum_{t=[\tau n]}^{[\tau_0 n]}\left(\varepsilon_{1t}^2-1\right)+\frac{1}{\left(\tau_0^{\triangle}-\tau^{\triangle}\right)n}\sum_{s\neq t\ [\tau n]}^{[\tau_0 n]}\left[\cos\left(s-t\right)\lambda_j\right]\varepsilon_{1s}\varepsilon_{1t}, \tag{3.93}$$

where with $\sum_{s\neq t\ [\tau n]}^{[\tau_0 n]}$ we indicated that summation is done both for s and for t, provided that $t\neq s$. A law of large number argument delivers

$$\frac{1}{\left(\tau_0^{\triangle}-\tau^{\triangle}\right)n}\sum_{t=[\tau n]}^{[\tau_0 n]}\left(\varepsilon_{1t}^2-1\right)\rightarrow_p 0 \tag{3.94}$$

while for the second term

$$E\left(\sum_{j=1}^{r}\sum_{s\neq t\ [\tau n]}^{[\tau_0 n]}\left[\cos\left(s-t\right)\lambda_j\right]\varepsilon_{1s}\varepsilon_{1t}\right)^2=2\sum_{s\neq t\ [\tau n]}^{[\tau_0 n]}\left(\sum_{j}^{r}\left[\cos\left(s-t\right)\lambda_j\right]\right)^2 \tag{3.95}$$

$$\leq\ 2\sum_{s\neq t\ 1}^{n}\left(\sum_{j=1}^{r}\left[\cos\left(s-t\right)\lambda_j\right]\right)^2 \tag{3.96}$$

which is exactly the term in the proof of Robinson so the rest of his argument applies without modifications.

If Θ_2 is empty, this implies that $\widehat{\delta}_2\left(\tau\right)\rightarrow_p\delta_1$. If Θ_2 is not empty, we have also to show that

$$P\left(\inf_{\Theta_2}S\left(d\right)\leq 0\right)\rightarrow 0. \tag{3.97}$$

Following the proof in Robinson,

$$P\left(\inf_{\Theta_2}S\left(d\right)\leq 0\right)\leq P\left(\frac{1}{m}\sum_{j=1}^{m}(a_j-1)\lambda_j^{2\delta_1}I_{z2}(\lambda_j)\leq 0\right) \tag{3.98}$$

with

$$a_j = \begin{cases} \left(\frac{j}{p}\right)^{2(\Delta-\delta_1)} & 1 \le j \le p \\ \left(\frac{j}{p}\right)^{2(\Delta_1-\delta_1)} & p < j \le m \end{cases} \tag{3.99}$$

and,

$$p = \exp(\frac{1}{m}\sum_{j=1}^{m} \ln j) \text{ so that } p \sim m/e \text{ when } m \to \infty. \tag{3.100}$$

Since

$$\frac{1}{m}\sum_{j=1}^{m}(a_j - 1) \ge \frac{1}{e\left(2\left(\Delta - \delta_1\right) + 1\right)} - 1 > 1, \tag{3.101}$$

choosing $\Delta < \delta_1 - 1/2 + 1/(4e)$ there is $\iota > 0$ such that

$$\frac{1}{m}\sum_{j=1}^{m}(a_j - 1) \ge 1 + \iota. \tag{3.102}$$

We then rewrite the bound in (3.98) as

$$P\left(\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\lambda_j^{2\delta_1} I_{z2}(\lambda_j) \le 0, \right. \tag{3.103}$$
$$\left. \frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\lambda_j^{2\delta_1}\left(I_{\xi 2}(\lambda_j) + 2\,\mathrm{Re}\left(I_{\xi 2\zeta 2}\left(\lambda_j\right)\right)\right) < -\iota\right)$$

$$+P\left(\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\lambda_j^{2\delta_1} I_{z2}(\lambda_j) \le 0, \right. \tag{3.104}$$
$$\left. \frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\lambda_j^{2\delta_1}\left(I_{\xi 2}(\lambda_j) + 2\,\mathrm{Re}\left(I_{\xi 2\zeta 2}\left(\lambda_j\right)\right)\right) \ge -\iota\right).$$

Clearly, the probability in (3.103) can be bounded by

$$P\left(\frac{1}{m}\sum_{j=1}^{m}(a_j - 1)\lambda_j^{2\delta_1}\left(I_{\xi 2}(\lambda_j) + 2\,\mathrm{Re}\left(I_{\xi 2\zeta 2}\left(\lambda_j\right)\right)\right) < -\iota\right) \tag{3.105}$$

$$\leq P\left(\frac{1}{m}\sum_{j=1}^{m}(a_j+1)\lambda_j^{2\delta_1}\left|I_{\xi 2}(\lambda_j)+2\,\mathrm{Re}\left(I_{\xi 2\zeta 2}\left(\lambda_j\right)\right)\right|>\iota\right) \tag{3.106}$$

which goes to zero because the argument of (3.106) converges in probability to 0. We show this discussing each term separately: first,

$$\left|\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2\delta_1}I_{\xi 2}(\lambda_j)\right| = O_p\left(\left(\frac{m}{n}\right)^{2(\delta_1-\delta_2)}\right)=o_p\left(1\right), \tag{3.107}$$

$$\left|\frac{2}{m}\sum_{j=1}^{m}\lambda_j^{2\delta_1}\,\mathrm{Re}\left(I_{\xi 2\zeta 2}\left(\lambda_j\right)\right)\right| = O_p\left(\frac{\ln m}{m}\left(\frac{m}{n}\right)^{\delta_1-\delta_2}\right)=o_p\left(1\right), \tag{3.108}$$

while for the two remaining terms,

$$\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m}a_j\lambda_j^{2\delta_1}\left(2\,\mathrm{Re}\,I_{\zeta 2\xi 2}\left(\lambda_j\right)\right)= \\
&= O_p\left(\frac{1}{m}\sum_{j=1}^{p}\left(\frac{j}{p}\right)^{2(\Delta-\delta_1)}\left(\frac{j}{n}\right)^{2\delta_1}\left(\frac{j}{n}\right)^{-\delta_1-\delta_2}\frac{\ln j}{j}\right. \\
&\quad\left.+\frac{1}{m}\sum_{j=p}^{m}\left(\frac{j}{p}\right)^{2(\Delta_1-\delta_1)}\left(\frac{j}{n}\right)^{2\delta_1}\left(\frac{j}{n}\right)^{-\delta_1-\delta_2}\frac{\ln j}{j}\right) \\
&= \begin{cases} O_p\left(\left(\frac{1}{n}\right)^{\delta_1-\delta_2}\left(\frac{1}{m}\right)^{2(\Delta-\delta_1)+1}+\left(\frac{m}{n}\right)^{\delta_1-\delta_2}\frac{\ln m}{m}\right) \text{ if } 2\Delta-\delta_1-\delta_2<0 \\ O_p\left(\left(\frac{m}{n}\right)^{\delta_1-\delta_2}\frac{\ln^2 m}{m}\right) \text{ if } 2\Delta-\delta_1-\delta_2=0 \\ O_p\left(\left(\frac{m}{n}\right)^{\delta_1-\delta_2}\frac{\ln m}{m}\right) \text{ if } 2\Delta-\delta_1-\delta_2>0 \end{cases}
\end{aligned} \tag{3.109}$$

and

$$\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m}a_j\lambda_j^{2\delta_1}I_{\xi 2}\left(\lambda_j\right)= \\
&= O_p\left(\frac{1}{m}\sum_{j=1}^{p}\left(\frac{j}{p}\right)^{2(\Delta-\delta_1)}\left(\frac{j}{n}\right)^{2\delta_1}\left(\frac{j}{n}\right)^{-2\delta_2}\right. \\
&\quad\left.+\frac{1}{m}\sum_{j=p}^{m}\left(\frac{j}{p}\right)^{2(\Delta_1-\delta_1)}\left(\frac{j}{n}\right)^{2\delta_1}\left(\frac{j}{n}\right)^{-2\delta_2}\right) \\
&= O_p\left((m/n)^{2(\delta_1-\delta_2)}\right)=o_p\left(1\right).
\end{aligned} \tag{3.110}$$

The probability in (3.106) then goes to 0. The probability in (3.104) can be bounded by

$$P\left(\frac{1}{m}\sum_{j=1}^{m}(a_j-1)\lambda_j^{2\delta_1}I_{z2}(\lambda_j)\leq 0\right), \tag{3.111}$$

which Robinson showed to tend to 0.

Proof of Theorem 3.3. As before, we discuss $\widehat{\delta}_2(\tau)$ when $\tau<\tau_0$. Robinson (1995b), Theorem 2, considered the expansion based on the mean value theorem

$$\left(\widehat{\delta}_2(\tau)-\delta_1\right)=-\left(\frac{\partial^2 R_2(\tau)}{\partial d^2}\right)^{-1}\Bigg|_{d=\delta_m}\frac{\partial R_2(\tau)}{\partial d}\Bigg|_{d=\delta_1} \tag{3.112}$$

with $|\delta_m-\delta_1|\leq\left|\widehat{\delta}_2(\tau)-\delta_1\right|$: following the same argument of the original proof,

$$\left(\frac{\partial^2 R_2(\tau)}{\partial d^2}\right)^{-1}\Bigg|_{d=\delta_m}\rightarrow_p 4 \tag{3.113}$$

and

$$m^{1/2}\frac{\partial R_2(\tau)}{\partial d}\Bigg|_{d=\delta_1}=2m^{-1/2}\sum_{j=1}^{m}\nu_j\left(\frac{I_{z2}(\lambda_j)}{g(\lambda_j)}-1\right)(1+o_p(1)) \tag{3.114}$$

where

$$\nu_j=\ln j-\frac{1}{m}\sum_{j=1}^{m}\ln j. \tag{3.115}$$

Decomposing $I_{z2}(\lambda_j)=I_{\zeta 2}(\lambda_j)+2\left|\mathrm{Re}\,I_{\zeta 2\xi 2}(\lambda_j)\right|+I_{\xi 2}(\lambda_j)$ as before, we can

rewrite (3.98) as

$$2m^{-1/2}\sum_{j=1}^{m}\nu_j\left(\frac{I_{\zeta 2}(\lambda_j)}{g_{\zeta 2}(\lambda_j)}-1\right)(1+o_p(1)) \tag{3.116}$$

$$2m^{-1/2}\sum_{j=1}^{m}\nu_j\left(\frac{2\left|\mathrm{Re}\, I_{\zeta 2\xi 2}(\lambda_j)\right|}{g_{\zeta 2}(\lambda_j)}\right)(1+o_p(1)) \tag{3.117}$$

$$2m^{-1/2}\sum_{j=1}^{m}\nu_j\left(\frac{I_{\xi 2}(\lambda_j)}{g_{\zeta 2}(\lambda_j)}\right)(1+o_p(1)) \tag{3.118}$$

Making use of (3.78) and of the decomposition following (4.11) in Robinson, (3.116) is

$$\left(2\frac{m^{-1/2}}{\left(\tau_0^{\Delta}-\tau^{\Delta}\right)}\sum_{t=[\tau n]}^{[\tau_0 n]}q_t+o_p(1)\right)(1+o_p(1)) \tag{3.119}$$

where $q_t=\varepsilon_{1t}\sum_{s=[\tau n]}^{t-1}\varepsilon_{1s}c_{t-s}$ replaces z_t in the original proof of Robinson but the rest follows in the same way, so $m^{-1/2}\sum_{t=[\tau n]}^{[\tau_0 n]}q_t$ converges in distribution to a normal $N(0,(\tau_0-\tau))$.

The result then holds if the two remaining terms are negligible: using Theorem 3.1 the term in (3.117) is $O_p\left((m/n)^{\vartheta}\ln m/m^{1/2}\right)=o_p(1)$, while the other one is $O_p\left((m/n)^{2\vartheta}m^{1/2}\right)$, which is only negligible under Assumption 3.5'.

Proof of Theorem 3.6 *(i).* Clearly $\widehat{\delta}_1(\tau)-\widehat{\delta}_2(\tau)$ is asymptotically normal, being it the sum of two normally distributed random variables; the variance is

$$\begin{aligned} & Var\left(\widehat{\delta}_1(\tau)-\widehat{\delta}_2(\tau)\right). \\ = \; & Var\left(\widehat{\delta}_1(\tau_1)\right)+Var\left(\widehat{\delta}_2(\tau_2)\right)-2Cov\left(\widehat{\delta}_1(\tau_1),\widehat{\delta}_2(\tau_2)\right) \end{aligned} \tag{3.120}$$

To compute the last element, we use the same decomposition as in Theorem

2.3,

$$m^{1/2}\left(\widehat{\delta}_1(\tau_1)-\delta\right) = \left(\frac{1}{4}+o_p(1)\right)\left(2\frac{m^{-1/2}}{\tau_1^{\Delta}}\sum_{t=1}^{[\tau_1 n]}q_t+o_p(1)\right)(1+o_p(1)) \tag{3.121}$$

$$m^{1/2}\left(\widehat{\delta}_2(\tau_2)-\delta\right) = \left(\frac{1}{4}+o_p(1)\right)\left(2\frac{m^{-1/2}}{1-\tau_2^{\Delta}}\sum_{t=[\tau_2 n]+1}^{n}q_t+o_p(1)\right)(1+o_p(1)), \tag{3.122}$$

where here $\delta_1 = \delta_2 = \delta$ because we assumed no break. The q_t in (3.121) and (3.122) are martingale differences so $Cov\left(\widehat{\delta}_1(\tau_1),\widehat{\delta}_2(\tau_2)\right) = 0$ because $\tau_2 \geq \tau_1$. Therefore, as $n \to \infty$,

$$mVar\left(\widehat{\delta}(\tau_1,\tau_2)\right) \to \frac{1}{4\tau_1}+\frac{1}{4(1-\tau_2)} = \frac{\tau_1+1-\tau_2}{4\tau_1(1-\tau_2)}. \tag{3.123}$$

Proof of Theorem 3.8. As before we only discuss the case $\delta_1 > \delta_2$.

The proof follows from the standard argument for implicitly defined extremum estimates as for example in Newey and McFadden (1994), Theorem 2.1. Let

$$Q_0(\tau) = \tau\delta_1 + (1-\tau)\left(\delta_1 1(\tau<\tau_0)+\delta_2 1(\tau\geq\tau_0)\right), \tag{3.124}$$

clearly $\tau_0 = \arg\min Q_0(\tau)$ and $[\tau_l,\tau_h]$ is a compact set. Sufficient conditions for consistency are (i) upper semicontinuity of $(-Q_0(\tau))$, (iia) pointwise convergence of the loss function in τ_0, i.e.

$$\widehat{Q}_n(\tau_0) \to_p Q_0(\tau_0), \tag{3.125}$$

and, (iib), that as $n \to \infty$, for $\varepsilon > 0$,

$$P\left(\left(-\widehat{Q}_n(\tau)\right) < \left(-Q_0(\tau)\right) + \varepsilon\right) \to 1 \text{ for all } \tau \in [\tau_l, \tau_h]. \tag{3.126}$$

Upper semicontinuity can be easily checked, and pointwise convergence (3.125) is also immediate. To prove (3.126), we show the equivalent statement that

$$P\left(\left(-\widehat{Q}_n(\tau)\right) > \left(-Q_0(\tau)\right) + \varepsilon\right) \to 0. \tag{3.127}$$

We bound the expression (3.127) as

$$\begin{aligned} & P\left(\left(-\widehat{Q}_n(\tau)\right) > \left(-Q_0(\tau)\right) + \varepsilon,\ \tau \in [\tau_l, \tau_h]\right) \\ \leq\ & P\left(Q_0(\tau) - \widehat{Q}_n(\tau) > \varepsilon,\ \tau \in [\tau_l, \tau_0)\right) && (3.128) \\ & + P\left(Q_0(\tau) - \widehat{Q}_n(\tau) > \varepsilon,\ \tau \in [\tau_0, \tau_h]\right). && (3.129) \end{aligned}$$

Notice that, for $\tau \in [\tau_l, \tau_0)$,

$$Q_0(\tau) - \widehat{Q}_n(\tau) = \delta_1 - \tau\widehat{\delta}_1(\tau) + (1-\tau)\widehat{\delta}_2(\tau). \tag{3.130}$$

The probability in (3.128) can be bounded by

$$\begin{aligned} & P\Big(\delta_1 - \tau\widehat{\delta}_1(\tau) + (1-\tau)\widehat{\delta}_2(\tau) > \varepsilon, \\ & \left|\widehat{\delta}_1(\tau) - \delta_1\right| < \varepsilon/2,\ \left|\widehat{\delta}_2(\tau) - \delta_1\right| < \varepsilon/2,\ \tau \in [\tau_l, \tau_0)\Big) && (3.131) \\ & + P\left(\left|\widehat{\delta}_1(\tau) - \delta_1\right| > \varepsilon/2,\ \tau \in [\tau_l, \tau_0)\right) && (3.132) \\ & + P\left(\left|\widehat{\delta}_2(\tau) - \delta_1\right| > \varepsilon/2,\ \tau \in [\tau_l, \tau_0)\right) && (3.133) \end{aligned}$$

and (3.132) and (3.133) go to 0 from Theorem 3.2, while (3.131) is bounded

by

$$P\left(\tau\left|\widehat{\delta}_1(\tau)-\delta_1\right|+(1-\tau)\left|\widehat{\delta}_2(\tau)-\delta_1\right|>\varepsilon,\right.$$
$$\left.\left|\widehat{\delta}_1(\tau)-\delta_1\right|<\varepsilon/2,\ \left|\widehat{\delta}_2(\tau)-\delta_1\right|<\varepsilon/2,\ \tau\in[\tau_l,\tau_0)\right) \quad (3.134)$$
$$= 0.$$

With a similar argument (3.129) goes to 0 too.

Notes on Tables 3.4 - 3.9 below

Notes on Tables 3.4 - 3.7:

‡ : Consistent for $\delta_1 = 0.4$ or 0.2 but subject to a lower order bias;

† : True value $\delta_2 = 0$.

Notes on Tables 3.7 and 3.8:

the columns represent t tests, so each estimate actually refers to the number of occurrences in which the standardised t test exceeded the critical value:

$\widehat{\delta}_{(2)}$: $\left|2\sqrt{m_2}\left(\widehat{\delta}_{(2)} - \delta_1\right)\right| \geq 1.96,$

$\widehat{\delta}$: $\left|2\sqrt{m\tau_0}\left(\widehat{\delta} - \delta_1\right)\right| \geq 1.96,$

$\widehat{\delta}_1(1/4)$: $\left|2\sqrt{m(1/4)}\left(\widehat{\delta}_1(1/4) - \delta_1\right)\right| \geq 1.96,$

$\widehat{\delta}_2(1/4)$: $\left|2\sqrt{m(\tau_0 - 1/4)}\left(\widehat{\delta}_2(1/4) - \delta_1\right)\right| \geq 1.96,$

$\widehat{\delta}_1(3/4)$: $\left|2\sqrt{m\tau_0}\left(\widehat{\delta}_1(3/4) - \delta_1\right)\right| \geq 1.96,$

$\widehat{\delta}_2(3/4)$: $\left|2\sqrt{m(1/4)}\left(\widehat{\delta}_2(3/4) - \delta_2\right)\right| \geq 1.96,$

$\widehat{\delta}_1(1/4) - \widehat{\delta}_2(1/4)$: $\left|2\sqrt{\frac{1/4(1-1/4)}{1+1/4-1/4}m}\left(\widehat{\delta}_1(1/4) - \widehat{\delta}_2(1/4)\right)\right| \geq 1.96,$

$\widehat{\delta}_1(1/4) - \widehat{\delta}_2(3/4)$: $\left|2\sqrt{\frac{1/4(1-3/4)}{1+1/4-3/4}m}\left(\widehat{\delta}_1(1/4) - \widehat{\delta}_2(3/4)\right)\right| \geq 1.96,$

$\widehat{\delta}_1(3/4) - \widehat{\delta}_2(3/4)$: $\left|2\sqrt{\frac{3/4(1-3/4)}{1+3/4-3/4}m}\left(\widehat{\delta}_1(3/4) - \widehat{\delta}_2(3/4)\right)\right| \geq 1.96.$

Also, notice the difference between Table 3.7 and 3.8: in Table 3.7 we assumed knowledge of the probability limit of the estimate, while in Table 3.8 we assumed that $\widehat{\delta}_1(\tau_1)$ and $\widehat{\delta}_2(\tau_1)$ had the same probability limit, which was incorrect under the break in δ.

Note on Table 3.9:

the columns represent t tests, so each estimate actually refers to the number of occurrences in which the standardised t test exceeded the critical value:

$\widehat{t}^2$: $\widehat{t}^2 \geq 9.01$

$\widetilde{t}$: $2\sqrt{m_2}\left(\widehat{\delta}_{(2)} - \widehat{\delta}\right) \geq 1.65$

Table 3.4: Monte Carlo bias: estimates on the whole sample and on selected sections only ($\tau = 1/4$ and $\tau = 3/4$)

Model	n	$\widehat{\delta}_{(2)}$	$\widehat{\delta}$	$\widehat{\delta}_1(1/4)$	$\widehat{\delta}_2(1/4)$	$\widehat{\delta}_1(3/4)$	$\widehat{\delta}_2(3/4)$
	64	-0.0517	-0.0283	0.0243	0.0177	0.0143	0.0250
	128	-0.0239	-0.0166	0.0351	0.0148	0.0171	0.0218
1	256	-0.0210	-0.0116	0.0351	0.0120	0.0139	0.0225
	512	-0.0071	-0.0053	0.0305	0.0099	0.0121	0.0213
	1024	0.0021	-0.0026	0.0231	0.0104	0.0091	0.0249
	64	-0.0340	-0.0208	0.0243	0.0183	0.0171	0.0250
	128	-0.0217	-0.0145	0.0351	0.0136	0.0183	0.0218
2	256	-0.0208	-0.0100	0.0351	0.0115	0.0153	0.0225
	512	-0.0032	-0.0040	0.0305	0.0096	0.0136	0.0213
	1024	0.0082	-0.0008	0.0231	0.0107	0.0107	0.0249
	64	-0.0058	-0.1177‡	0.0270	-0.1767‡	-0.0621‡	-0.0307†
	128	0.0036	-0.1014‡	0.0358	-0.1595‡	-0.0501‡	-0.0286†
3	256	0.0077	-0.0901‡	0.0372	-0.1490‡	-0.0451‡	-0.0196†
	512	0.0337	-0.0795‡	0.0321	-0.1288‡	-0.0388‡	-0.0119†
	1024	0.0161	-0.0782‡	0.0205	-0.1240‡	-0.0400‡	-0.0106†
	64	-0.0786	-0.1774‡	0.0270	-0.3234‡	-0.1333‡	-0.0307†
	128	-0.0476	-0.1578‡	0.0358	-0.3035‡	-0.1142‡	-0.0286†
4	256	-0.0355	-0.1423‡	0.0372	-0.2792‡	-0.1024‡	-0.0196†
	512	-0.0008	-0.1308‡	0.0321	-0.2605‡	-0.0937‡	-0.0119†
	1024	-0.0063	-0.1255‡	0.0205	-0.2459‡	-0.0897‡	-0.0107†
	64	-0.1146‡	-0.1059‡	-0.0419	-0.1410‡	-0.0773‡	-0.0307†
	128	-0.0937‡	-0.0958‡	-0.0277	-0.1316‡	-0.0654‡	-0.0286†
5	256	-0.0730‡	-0.0801‡	-0.0090	-0.1196‡	-0.0526‡	-0.0196†
	512	-0.0546‡	-0.0737‡	-0.0055	-0.1092‡	-0.0466‡	-0.0119†
	1024	-0.0547‡	-0.0714‡	-0.0055	-0.1054‡	-0.0444‡	-0.0107†

Table 3.5: Monte Carlo standard deviation: estimates on the whole sample and on selected sections only ($\tau = 1/4$ and $\tau = 3/4$)

Model	n	$\widehat{\delta}_{(2)}$	$\widehat{\delta}$	$\widehat{\delta}_1(1/4)$	$\widehat{\delta}_2(1/4)$	$\widehat{\delta}_1(3/4)$	$\widehat{\delta}_2(3/4)$
	64	0.4643	0.1533	0.2817	0.1604	0.1673	0.2885
	128	0.3000	0.1115	0.2156	0.1199	0.1178	0.2222
1	256	0.2346	0.0762	0.1616	0.0863	0.0855	0.1625
	512	0.1826	0.0552	0.1096	0.0638	0.0627	0.1172
	1024	0.1415	0.0409	0.0894	0.0465	0.0468	0.0858
	64	0.4803	0.1535	0.2817	0.1643	0.1767	0.2885
	128	0.3080	0.1168	0.2156	0.1253	0.1254	0.2222
2	256	0.2442	0.0801	0.1616	0.0890	0.0901	0.1625
	512	0.1880	0.0571	0.1096	0.0649	0.0662	0.1172
	1024	0.1448	0.0424	0.0894	0.0477	0.0496	0.0858
	64	0.4702	0.1726‡	0.2922	0.1991‡	0.1839‡	0.2667†
	128	0.3363	0.1350‡	0.2215	0.1684‡	0.1402‡	0.2009†
3	256	0.2873	0.1015‡	0.1589	0.1289‡	0.1053‡	0.1504†
	512	0.2277	0.0708‡	0.1099	0.0962‡	0.0738‡	0.1102†
	1024	0.1800	0.0539‡	0.0838	0.0746‡	0.0559‡	0.0799†
	64	0.4907	0.1849‡	0.2922	0.1768‡	0.2004‡	0.2667†
	128	0.3680	0.1470‡	0.2215	0.1586‡	0.1561‡	0.2009†
4	256	0.3138	0.1120‡	0.1589	0.1312‡	0.1171‡	0.1504†
	512	0.2584	0.0815‡	0.1099	0.1097‡	0.0852‡	0.1102†
	1024	0.2100	0.0618‡	0.0838	0.0922‡	0.0637‡	0.0799†
	64	0.4479‡	0.1531‡	0.2816	0.1697‡	0.1687‡	0.2667†
	128	0.3151‡	0.1141‡	0.2168	0.1317‡	0.1226‡	0.2009†
5	256	0.2634‡	0.0857‡	0.1518	0.0971‡	0.0941‡	0.1504†
	512	0.2080‡	0.0574‡	0.1003	0.0699‡	0.0640‡	0.1102†
	1024	0.1567‡	0.0442‡	0.0794	0.0529‡	0.0496‡	0.0799†

Table 3.6: St. dev. from asymptotic theory: estimates on the whole sample and on selected sections only ($\tau = 1/4$ and $\tau = 3/4$)

Model	n	$\widehat{\delta}_{(2)}$	$\widehat{\delta}$	$\widehat{\delta}_1(1/4)$	$\widehat{\delta}_2(1/4)$	$\widehat{\delta}_1(3/4)$	$\widehat{\delta}_2(3/4)$
	64	0.2236	0.1118	0.2236	0.1291	0.1291	0.2236
	128	0.1768	0.0857	0.1715	0.0990	0.0990	0.1715
1	256	0.1508	0.0651	0.1302	0.0752	0.0752	0.1302
	512	0.1291	0.0493	0.0985	0.0569	0.0569	0.0985
	1024	0.1066	0.0374	0.0747	0.0432	0.0432	0.0747
	64	0.2236	0.1118	0.2236	0.1291	0.1291	0.2236
	128	0.1768	0.0857	0.1715	0.0990	0.0990	0.1715
2	256	0.1508	0.0651	0.1302	0.0752	0.0752	0.1302
	512	0.1291	0.0493	0.0985	0.0569	0.0569	0.0985
	1024	0.1066	0.0374	0.0747	0.0432	0.0432	0.0747
	64	0.3162	0.1581‡	0.2236	0.2236‡	0.1581‡	0.2236†
	128	0.2500	0.1213‡	0.1715	0.1715‡	0.1213‡	0.1715†
3	256	0.2132	0.0921‡	0.1302	0.1302‡	0.0921‡	0.1302†
	512	0.1826	0.0697‡	0.0985	0.0985‡	0.0697‡	0.0985†
	1024	0.1508	0.0529‡	0.0747	0.0747‡	0.0529‡	0.0747†
	64	0.3873	0.1936‡	0.2236	0.3873‡	0.1936‡	0.2236†
	128	0.3062	0.1485‡	0.1715	0.2970‡	0.1485‡	0.1715†
4	256	0.2611	0.1127‡	0.1302	0.2255‡	0.1127‡	0.1302†
	512	0.2236	0.0853‡	0.0985	0.1707‡	0.0853‡	0.0985†
	1024	0.1846	0.0647‡	0.0747	0.1295‡	0.0647‡	0.0747†
	64	0.3162‡	0.1581‡	0.2236	0.2236‡	0.1581‡	0.2236†
	128	0.2500‡	0.1213‡	0.1715	0.1715‡	0.1213‡	0.1715†
5	256	0.2132‡	0.0921‡	0.1302	0.1302‡	0.0921‡	0.1302†
	512	0.1826‡	0.0697‡	0.0985	0.0985‡	0.0697‡	0.0985†
	1024	0.1508‡	0.0529‡	0.0747	0.0747‡	0.0529‡	0.0747†

Table 3.7: Empirical sizes of 5% t tests: estimates on the whole sample and on selected sections only ($\tau = 1/4$ and $\tau = 3/4$)

Model	n	$\widehat{\delta}_{(2)}$	$\widehat{\delta}$	$\widehat{\delta}_1(1/4)$	$\widehat{\delta}_2(1/4)$	$\widehat{\delta}_1(3/4)$	$\widehat{\delta}_2(3/4)$
	64	31.4	14.6	12.1	12.0	13.1	12.2
	128	22.8	13.4	12.0	11.0	8.7	13.2
1	256	18.9	9.6	11.3	8.9	8.6	12.0
	512	16.4	8.0	8.6	8.3	8.2	10.3
	1024	13.8	7.4	11.1	7.1	7.4	9.7
	64	32.0	16.6	12.1	11.7	15.1	12.2
	128	24.3	14.1	12.0	12.5	11.7	13.2
2	256	21.2	11.0	11.3	9.4	10.4	12.0
	512	17.5	10.0	8.6	9.0	10.3	10.3
	1024	15.3	9.0	11.1	9.2	11.0	9.7
	64	15.8	13.9‡	13.6	10.2‡	10.8‡	11.2†
	128	13.9	15.5‡	14.3	15.1‡	11.4‡	9.5†
3	256	14.3	19.2‡	11.4	20.9‡	11.9‡	9.2†
	512	11.6	21.4‡	8.8	23.7‡	10.9‡	8.5†
	1024	9.4	32.3‡	9.3	37.5‡	13.5‡	7.6†
	64	12.2	13.7‡	13.6	0.9‡	11.6‡	11.2†
	128	11.1	17.6‡	14.3	3.6‡	13.4‡	9.5†
4	256	10.2	23.9‡	11.4	9.4‡	15.0‡	9.2†
	512	8.6	33.4‡	8.8	26.7‡	20.5‡	8.5†
	1024	8.9	48.2‡	9.3	47.3‡	28.0‡	7.6†
	64	15.5‡	10.4‡	11.4	3.4‡	8.9‡	11.2†
	128	12.3‡	11.7‡	12.1	7.3‡	9.6‡	9.5†
5	256	12.1‡	12.3‡	9.2	7.8‡	9.7‡	9.2†
	512	10.1‡	13.4‡	6.5	11.4‡	8.6‡	8.5†
	1024	6.8‡	23.1‡	6.9	21.6‡	11.1‡	7.6†

Table 3.8: Empirical sizes of 5% t tests: differences of estimates on selected sections ($\tau = 1/4$ and $\tau = 3/4$)

Model	n	$\widehat{\delta}_1(1/4) - \widehat{\delta}_2(1/4)$	$\widehat{\delta}_1(1/4) - \widehat{\delta}_2(3/4)$	$\widehat{\delta}_1(3/4) - \widehat{\delta}_2(3/4)$
	64	8.9	10.4	11.0
	128	8.7	9.2	10.3
1	256	8.9	9.3	8.7
	512	6.3	8.1	8.0
	1024	7.1	8.0	6.3
	64	9.4	10.4	11.0
	128	8.5	9.2	10.1
2	256	9.2	9.3	8.4
	512	6.9	8.1	7.4
	1024	6.8	8.0	6.0
	64	19.1	35.0	32.4
	128	22.1	49.8	46.1
3	256	27.6	63.2	66.1
	512	31.4	85.3	86.6
	1024	40.7	96.7	98.3
	64	32.7	35.0	27.7
	128	42.2	49.8	38.9
4	256	55.1	63.2	52.3
	512	69.6	85.3	74.4
	1024	78.6	96.7	93.7
	64	12.7	15.9	16.1
	128	14.0	15.8	15.6
5	256	14.2	25.0	23.0
	512	16.7	33.4	32.3
	1024	24.3	49.6	48.7

Table 3.9: Empirical sizes of 5% tests to detect a break and sample statistics for $\widehat{\tau}$

Model	n	$\widetilde{t^2}$	$\widetilde{t}$	$\widehat{\tau}$, bias	$\widehat{\tau}$, st. dev.	$\widehat{\tau}$, low 5%	$\widehat{\tau}$, median	$\widehat{\tau}$, up 5%
	64	14.1	16.9					
	128	13.9	13.0					
1	256	10.4	10.1					
	512	7.6	10.9					
	1024	7.7	10.4					
	64	13.8	18.4					
	128	14.7	13.7					
2	256	11.9	11.1					
	512	8.8	11.2					
	1024	8.7	10.5					
	64	45.3	25.2	-0.0259	0.1908	0.1406	0.4688	0.8438
	128	58.0	24.9	-0.0294	0.1548	0.1875	0.4844	0.7813
3	256	78.7	26.4	-0.0283	0.1173	0.2461	0.4844	0.6875
	512	94.7	30.6	-0.0186	0.0721	0.3438	0.4941	0.5703
	1024	99.7	31.5	-0.0145	0.0445	0.4150	0.4951	0.5293
	64	42.4	23.5	0.0733	0.2107	0.1406	0.3281	0.8438
	128	58.0	25.2	0.0354	0.1725	0.1484	0.3203	0.7969
4	256	74.6	30.4	-0.0008	0.1238	0.1641	0.3203	0.6016
	512	93.0	33.9	-0.0155	0.0629	0.2207	0.3242	0.4180
	1024	99.8	41.2	-0.0103	0.0420	0.2607	0.3291	0.3682
	64	23.1	17.5	-0.0209	0.2318	0.1250	0.4688	0.8594
	128	25.2	14.0	-0.0320	0.2138	0.1406	0.4531	0.8438
5	256	32.0	15.0	-0.0403	0.2080	0.1484	0.4531	0.8398
	512	41.1	15.7	-0.0356	0.1722	0.1641	0.4688	0.8145
	1024	61.4	12.4	-0.0319	0.1316	0.2246	0.4746	0.7314

Chapter 4

Cointegration in fractional systems with deterministic trends

4.1 Introduction

We noticed earlier that economic time series are very often characterised by a time-varying mean. In this chapter we discuss a model of fractional cointegration for data that are contaminated by one or more fractional deterministic trends.

We consider OLS and GLS estimation, in the first case also taking into account the situation in which the deterministic component is neglected by the researcher. We choose OLS because it is very fast to compute, not requiring detailed modelling of the structure of the cointegrated variables and of the error term, nor the preliminary estimation of parameters or any other transformation of the data, so it is often used in applied analysis to provide an initial indication about the true value of the cointegrating parameter, albeit possibly a rather inefficient one. It is then important to assess if the OLS estimate is

consistent when the trends are neglected, and to see when the specification of the deterministic structure as an additional vector of regressors improves the rate of convergence. And of course we discuss a GLS type estimate because it is more efficient, and indeed optimal under Gaussianity.

In this chapter we consider fractional integration according to the Type II definition. This is because, as we saw in Chapter 1, Type I fractional Brownian motion is only defined for a certain range of orders of integration, while we want to allow for a potentially larger range. Moreover, this is also the definition considered by Robinson and Marinucci (2001) (which we refer to as RM in the rest of the chapter) and by Robinson and Hualde (2003) (which we refer to as RH in the rest of the chapter) in two very important works in which NBLS and GLS estimations of a fractionally cointegrated model are described: by keeping the same structure we also make comparison with those works easier.

For the fractional trend, we say w_t is a $J(d)$ (deterministic) sequence if

$$w_t \sim c_1 \frac{t^{d-1/2}}{\Gamma(d+1/2)}, \text{ as } t \to \infty, \ |w_t - w_{t+1}| \ \leq C w_t/t \ , t \geq 1, \qquad (4.1)$$

for some finite, non-zero c_1. The definitions of $I(d)$ and $J(d)$ match in the sense that:

(i) If w_t is $I(d)$ then $\Delta^c w_t^{\#}$ is $I(d-c)$; if w_t is $J(d)$ then $\Delta^c w_t^{\#}$ is $J(d-c)$.

(ii) If $d > 0$, then: if w_t is $I(d)$, $Ew_t^2 \sim Ct^{2d-1}$ as $t \to \infty$; if w_t is $J(d)$, $w_t^2 \sim Ct^{2d-1}$ as $t \to \infty$.

We expect then, in a system containing $I(d)$ and $J(e)$ sequences, that (fractional) differencing has a similar effect on both and that there will be a tendency for an $I(d)$ component to dominate a $J(e)$ one for $d > e$, and vice versa when $d < e$. This is most simply seen in an additive model. We consider

a bivariate observable sequence (x_t, y_t) given by the components model

$$x_t = g_{1t}(\phi_1) + h_{1t}(\delta_1),\ t \geq 1, \tag{4.2}$$

$$y_t = g_{2t}(\phi_2) + h_{2t}(\delta_2),\ t \geq 1, \tag{4.3}$$

$$g_{1t}(\phi_1) = h_{1t}(\delta_1) = g_{2t}(\phi_2) = h_{2t}(\delta_2) \equiv 0,\ t \leq 0, \tag{4.4}$$

where $g_{it}(d)$ is $J(d)$ and $h_{it}(d)$ is $I(d)$, $i = 1, 2$ and the $g_{it}(\phi_i)$, $h_{it}(\delta_i)$ are unobservable sequences. Robinson and Marinucci (2000) discussed a fractional model of form (4.2)-(4.4). In particular they considered the asymptotic behaviour of the sample covariance matrix of (x_t, y_t), based on observations at $t = 1, 2, ..., n$, and of the averaged periodogram of (x_t, y_t), being the normalized sum of the periodogram matrix of (x_t, y_t) over m Fourier frequencies close to the origin, where $1/m + m/n \to 0$ as $n \to \infty$; their stress is on cases where $\phi_i > 0$, $\delta_i > 1/2$, $i = 1, 2$, when the sample covariance matrix and the averaged periodogram have the same asymptotic behaviour (since they are equal when $m = n$, and low frequency components dominate the sample covariance matrix). Robinson and Marinucci (2000) described how stochastic or deterministic components dominate, depending on the values of the ϕ_i, δ_i. They also considered estimating the cointegration parameter by a frequency domain regression of y_t on x_t, over a possibly narrow band of frequencies, in case the deterministic trend of x_t dominates the stochastic one, the reverse takes place in the error, and this term is $I(d)$ with $d > 1/2$.

Regression relations also arise in case of cointegration. Consistency in cointegrated systems involving no deterministic components is due to the domination of the stochastic component of the cointegrating error by that of the regressor, while in the case of Robinson and Marinucci (2000) the estimate of the regression coefficient is consistent and asymptotically normal due to the dominating effect of the deterministic component of x_t.

Here we look at implications of more general combinations of deterministic and stochastic trends as well. On the basis of (4.2)-(4.4), it is supposed that $\delta_1 = \delta_2 = \delta$ and there exists $\nu \neq 0$ such that $h_{2t}(\delta) - \nu h_{1t}(\delta)$ is $I(\gamma)$, $\gamma < \delta$.

Irrespective of whether or not we commence from the component model (4.2)-(4.4), we assume the following cointegrated system as data generating mechanism,

$$y_t = \nu x_t + \sum_{j=1}^{p_1} \mu_{1j} t^{\phi_{1j}-1/2} + u_{1t}(-\gamma), \tag{4.5}$$

$$x_t = \sum_{j=1}^{p_2} \mu_{2j} t^{\phi_{2j}-1/2} + u_{2t}(-\delta), \tag{4.6}$$

where in (4.5), (4.6), $u_t = (u_{1t}, u_{2t})'$ is a jointly covariance stationary process with zero-mean and spectral density matrix $f_u(\lambda)$ that is nonsingular and continuous at all frequencies;

$$\nu \neq 0, \tag{4.7}$$

$$\delta > \max(\gamma, \frac{1}{2}). \tag{4.8}$$

This is then the same cointegration set-up of RH, augmented for the two groups of deterministic trends.

The truncations in (4.5), (4.6) imply that $x_t = y_t = 0$, $t \leq 0$, and that x_t, y_t $t \geq 1$, have finite variance. Without truncation they would not be well-defined in mean square, since $\delta > 1/2$, while $\Delta^{-\gamma} u_{1t}$ is not well-defined in mean square when $\gamma \geq 1/2$. In particular, x_t, and thus y_t, have variance that increases with t (like $t^{2\delta-1}$). Note that the elements of $u_t^{\#}$ are $I(0)$ processes, while $u_{1t}(-\gamma)$ is an $I(\gamma)$ process and $u_{2t}(-\delta)$ is an $I(\delta)$ process.

For the deterministic components, we assume that the ϕ_{ij} are real numbers

satisfying

$$\phi_{11} > ... > \phi_{1p_1} > 0, \tag{4.9}$$

$$\phi_{21} > ... > \phi_{2p_2} > 0, \tag{4.10}$$

and also, for $0 \leq p_{11} \leq p_1$, $0 \leq p_{21} \leq p_2$,

$$\phi_{1p_{11}} > \gamma, \ \phi_{1,p_{11}+1} < \gamma, \tag{4.11}$$

$$\phi_{2p_{21}} > \delta, \ \phi_{2,p_{21}+1} < \delta; \tag{4.12}$$

for $p_{11} = p_1$ $(p_{21} = p_2)$, the quantities $p_{11} + 1$ $(p_{21} + 1)$ are not defined, so the second inequality of (4.11) ((4.12)) is irrelevant, while for $p_{11} = 0$ $(p_{21} = 0)$ all ϕ_{1j} (ϕ_{2j}) are less than γ (δ), so the first inequality of (4.11) ((4.12)) is irrelevant. In (4.5), (4.6) an intercept term appears when $\phi_{1j} = 1/2$, $\phi_{2j} = 1/2$, respectively, while integer powers are also possible, but we allow for the ϕ_{ij} to be any real values satisfying (4.9)-(4.12).

The convention that powers of t be denoted $\phi_{ij} - 1/2$ rather than ϕ_{ij} is to enable convenient comparison with integration orders, as indicated by our definitions of $I(d)$ and $J(d)$ sequences. It is possible that one or more of the μ_{ij} are actually zero, though we do not know this, and we define

$$j^\dagger = \min\left\{j : \mu_{1j} \neq 0\right\}, \ \ j^\ddagger = \min\left\{j : \mu_{2j} \neq 0\right\}. \tag{4.13}$$

We allow for this possibility because we wish (4.5), (4.6) to nest the working model used in estimation, and it is possible that one or more regressors $t^{\phi_{ij}-1/2}$ will be included in the latter whose coefficient is zero. For brevity write $\phi_{1\dagger} = \phi_{1j^\dagger}$, $\phi_{2\ddagger} = \phi_{2j^\ddagger}$, $\mu_{1\dagger} = \mu_{1j^\dagger}$, $\mu_{2\ddagger} = \mu_{2j^\ddagger}$.

We could replace the deterministic terms in (4.5), (4.6) by more general

ones, and in particular could specify s_{1t} and s_{2t} in terms of bounds, rather than precise quantities, or model other particular deterministic components like trends with breaks. Yet the current specification allows a very simple way to present the results, and it is very often a useful mean of practical guidance in case of the existence of other deterministic components (including breaks) as well, whereas the explicit formal treatment of a more general model may have to come at the cost of a much more obscure statement of the results.

While we allow for non-integer powers $\phi_{ij} - 1/2$, this is not on the basis of arguing that these are likely to be of great practical value (though they may turn out to provide improved approximations to some data), but rather because it affords a precise treatment of the competition between stochastic and deterministic trends. Indeed, though we mostly regard γ and δ as unknown, the ϕ_{ij} are assumed known; there are difficulties with asymptotic theory for, say, OLS estimation of the ϕ_{ij} due to lack of uniform convergence of the objective function.

Define, for $i = 1, 2$, the $p_{i1} \times 1$ vector sequences

$$g_i(t) = (t^{\phi_{i1}-1/2}, ..., t^{\phi_{ip_{i1}}-1/2})', \ t \geq 1, \tag{4.14}$$

$$= (0, ..., 0)', \ t \leq 0, \tag{4.15}$$

and denote by μ_i, $i = 1, 2$, the $p_{i1} \times 1$ vectors whose jth elements are respectively μ_{ij}, $i = 1, 2$. In view of (4.13), (4.14) note that $\mu_1' g_1(t)$ is a $J(\phi_{1\dagger})$ sequence and $\mu_2' g_2(t)$ is a $J(\phi_{2\dagger})$ sequence.

The working models that are estimated are as follows.

Working Model I This is

$$y_t = \nu x_t + v_t, \tag{4.16}$$

where ν is estimated by non-intercept OLS, as if the errors v_t were serially uncorrelated random variables orthogonal to x_t. The main issues here are the effect of the misspecification error caused by neglecting the deterministic component in (4.5) and simultaneous equation bias due to (4.6); also, notice that the error may have a stochastic component subject to autocorrelation due to (4.5).

Working Model II This is

$$y_t = \nu x_t + \mu_1' g_1(t) + v_t, \tag{4.17}$$

where ν and μ_1 are simultaneously estimated by multiple OLS, again as if the errors v_t were serially uncorrelated random variables orthogonal to x_t. Here, one issue is the robustness to possible omission of the component $s_{1t} = \sum_{j=p_{11}+1}^{p_1} \mu_{1j} t^{\phi_{1j}-1/2}$. Indeed the μ_{1j}, $p_{11}+1 \leq j \leq p_1$, are not consistently estimable. Another is the simultaneous equations bias due to (4.6) and error autocorrelation due to (4.5).

Working Model III This is

$$y_t = \nu x_t + \mu_1' g_1(t) + v_{1t}, \tag{4.18}$$

$$x_t = \mu_2' g_2(t) + v_{2t}, \tag{4.19}$$

where ν, μ_1 and μ_2 are simultaneously estimated by a form of generalized least squares (GLS) as if $v_{1t} = u_{1t}(-\gamma)$, $v_{2t} = u_{2t}(-\delta)$, properly accounting for both autocorrelation and simultaneity, and with either γ and/or δ assumed known or suitably estimated in a side calculation. For our discussion of GLS we assume also that

$$\beta > \frac{1}{2}, \tag{4.20}$$

where $\beta = \delta - \gamma$. The requirement (4.20) includes the usual case of $I(0)$ cointegrating errors and $I(1)$ x_t; the case $\beta \leq \frac{1}{2}$ leads to quite different asymptotics, see e.g. Jeganathan (1999), Hualde and Robinson (2002).

The Models I and II extend the set-up of Robinson and Marinucci (2000), who considered only a constant in the estimated model; it is also an extension because they focused on $\gamma > \max\left\{\phi_{1\dagger}, 1/2\right\}$, $\phi_{2\ddagger} > \delta$ only.

The GLS method used to treat Model III is an extension of that of RH, in which $\mu_1 = 0$, $\mu_2 = 0$ was correctly assumed *a priori*. Comparison with (4.5), (4.6) indicates that in fact $v_{1t} = s_{1t} + u_{1t}(-\gamma)$, $v_{2t} = s_{2t} + u_{2t}(-\delta)$, where $s_{2t} = \sum_{j=p_{21}+1}^{p_2} \mu_{2j} t^{\phi_{2j}-1/2}$, so that we have to show that ignoring s_{1t}, s_{2t} has no asymptotic effect; in fact none of the μ_{ij} in s_{1t}, s_{2t} is consistently estimable. Thus, robustness of RH's estimates to omission of such s_{it} will be implied.

In the next section we discuss rates of convergence (if any) and asymptotic distributional properties of the estimates of working models I, II and III respectively. Proof details are left to the appendices. Section 4.3 contains a small Monte Carlo study of finite sample performance, and Section 4.4 an empirical application to testing the PPP hypothesis on the basis of data for three US cities. In Section 4.5 we discuss our results and some related topics.

4.2 Estimation of the cointegrating vector

4.2.1 Least squares estimates of Model I

The OLS estimate of ν in Model I is

$$\overline{\nu} = M_{xy}/M_{xx}, \tag{4.21}$$

where for any column vector or scalar sequences a_t, b_t, $M_{ab} = \sum_{t=1}^{n} a_t b_t'$. The estimate $\overline{\nu}$ has the advantage of not requiring knowledge of δ, β and of $g_1(t)$. Under (4.5), (4.6) with $\mu_{ij} = 0$ for all i, j, RM showed under mild additional conditions that $\overline{\nu}$ is $n^{\min(2\delta-1,\beta)}$-consistent (except when both $\delta > \beta$ and $2\delta - \beta = 1$ hold, in which case it is $(n^{\beta}/\log n)-$consistent).

We introduce

Assumption 4.1. *The process* u_t, $t = 0, \pm 1, ...,$ *has representation*

$$u_t = A(L)\,\varepsilon_t, \tag{4.22}$$

where

$$A(s) = I_2 + \sum_{j=1}^{\infty} A_j s^j, \tag{4.23}$$

and the A_j *are* 2×2 *matrices such that :*

(i)

$$\det\{A(s)\} \neq 0, \quad |s| = 1; \tag{4.24}$$

(ii) $A(e^{i\lambda})$ *is differentiable in* λ *with derivative in* $Lip(\eta)$, $\eta > \frac{1}{2}$;

(iii) the ε_t *are independent and identically distributed vectors with mean zero, positive definite covariance matrix* Ω, *and* $E\|\varepsilon_t\|^q < \infty$, $q \geq 4$, $q > 2/(2\delta - 1)$.

This is a fairly general short memory specification for u_t, and it includes, among others, the very popular autoregressive moving average (ARMA) models. It was introduced by RH, who explained that *(ii)* implies $\sum_{j=1}^{\infty} j\,\|A_j\| < \infty$ and $\sum_{j=1}^{\infty} j\,\|A_j\|^2 < \infty$: with *(iii)* it is then sufficient for the weak convergence of fractional transforms, or central limit theorems, as discussed by Marinucci and Robinson (2000). In connection with this, denote by $W(r)$ the 2×1

vector Brownian motion with covariance matrix Ω, and define, for $d > 0$, the Type II fractional Brownian motion

$$W(r;d) = \int_0^r \frac{(r-s)^{d-1}}{\Gamma(d)} dW(s), \tag{4.25}$$

and then define $\overline{W}(r;d)$ and $\widetilde{W}(r;d)$ to respectively be the first and second elements of the vector $A(1)W(r;d)$.

Define

$$\Phi_1 = \int_0^1 \widetilde{W}(r;\delta)^2 dr, \tag{4.26}$$

$$\Phi_2 = \mu_{2\ddagger} \int_0^1 r^{\phi_{2\ddagger}-1/2} \widetilde{W}(r;\delta) dr, \tag{4.27}$$

$$\Phi_3 = \mu_{2\ddagger}^2 (2\phi_{2\ddagger})^{-1}, \tag{4.28}$$

and

$$\Psi_1 = \int_{-\pi}^{\pi} (1-e^{i\lambda})^{-\gamma} (1-e^{-i\lambda})^{-\delta} f_{12}(\lambda) d\lambda, \tag{4.29}$$

$$\Psi_2 = f_{12}(0) \sin \delta\pi, \tag{4.30}$$

where $f_{12}(\lambda)$ is the $(1,2)th$ element of $f_u(\lambda)$, Ψ_3 is such that for $\gamma > 1/2$

$$\Psi_3 = \int_0^1 \overline{W}(r;\gamma) \widetilde{W}(r;\delta) dr, \tag{4.31}$$

and for $\gamma \leq 1/2$ Ψ_3 is an $O_p(1)$ random variable and

$$\Psi_4 = \begin{cases} \mu_{2\ddagger} \int_0^1 r^{\phi_{2\ddagger}-1/2} \overline{W}(r;\gamma)\, dr & \text{if } \gamma > 0, \\ \mu_{2\ddagger} \int_0^1 r^{\phi_{2\ddagger}-1/2} d\overline{W}(r) & \text{if } \gamma = 0, \end{cases} \tag{4.32}$$

$$\Psi_5 = \mu_{1\dagger} \int_0^1 r^{\phi_{1\dagger}-1/2} \widetilde{W}(r;\delta) dr, \tag{4.33}$$

$$\Psi_6 = \frac{\mu_{1\dagger}\mu_{2\ddagger}}{\phi_{1\dagger}+\phi_{2\ddagger}}. \tag{4.34}$$

Also define

$$\Phi_1^{**} = \Phi_1 1(\delta \geq \phi_{2\ddagger}), \tag{4.35}$$

$$\Phi_2^{**} = \Phi_2 1(\delta = \phi_{2\ddagger}), \tag{4.36}$$

$$\Phi_3^{**} = \Phi_3 1(\delta \leq \phi_{2\ddagger}), \tag{4.37}$$

and

$$\Psi_1^{**} = \Psi_1 1(\gamma+\delta<1,\ \gamma+\phi_{2\ddagger}\leq 1,\ \phi_{1\dagger}+\delta\leq 1,\ \phi_{1\dagger}+\phi_{2\ddagger}\leq 1), \tag{4.38}$$

$$\Psi_2^{**} = \Psi_2 1(\gamma+\delta=1,\ \gamma\geq\phi_{1\dagger},\ \phi_{2\ddagger}\leq\delta<1), \tag{4.39}$$

$$\Psi_3^{**} = \Psi_3 1(\gamma+\delta>1,\ \gamma\geq\phi_{1\dagger},\ \delta\geq\phi_{2\ddagger}), \tag{4.40}$$

$$\begin{aligned} \Psi_4^{**} = {} & \Psi_4 1(\gamma+\delta>1,\ \phi_{1\dagger}\leq\gamma,\ \delta\leq\phi_{2\ddagger}) \\ & +\Psi_4 1(\gamma+\delta=1 \text{ and } \gamma>0,\ \phi_{1\dagger}\leq\gamma,\ \delta<\phi_{2\ddagger}) \\ & +\Psi_4 1(\gamma+\delta<1,\ \phi_{1\dagger}\leq\gamma,\ 1\leq\phi_{2\ddagger}+\gamma), \end{aligned} \tag{4.41}$$

$$\begin{aligned}
\Psi_5^{**} &= \Psi_5 1(\gamma+\delta>1 \text{ or } \gamma=0 \text{ and } \delta=1,\ \phi_{1\dagger}\geq\gamma,\ \phi_{2\ddagger}\leq\delta) \\
&\quad +\Psi_5 1(\gamma+\delta=1 \text{ and } \gamma>0,\ \gamma<\phi_{1\dagger},\ \phi_{2\ddagger}\leq\delta) \\
&\quad +\Psi_5 1(\gamma+\delta<1,\ \phi_{2\ddagger}\leq\delta,\ 1\leq\phi_{1\dagger}+\delta), \qquad (4.42) \\
\Psi_6^{**} &= \Psi_6 1(\gamma+\delta>1 \text{ or } \gamma=0 \text{ and } \delta=1,\ \phi_{1\dagger}\geq\gamma,\ \phi_{2\ddagger}\geq\delta) \\
&\quad +\Psi_6 1(\gamma+\delta=1 \text{ and } \gamma>0,\ \phi_{1\dagger}\geq\gamma,\ \phi_{2\ddagger}\geq\delta,\ \phi_{1\dagger}+\phi_{2\ddagger}>1) \\
&\quad +\Psi_6 1(\gamma+\delta<1,\ \phi_{1\dagger}\geq\gamma,\ \phi_{2\ddagger}\geq\delta,\ \phi_{1\dagger}+\phi_{2\ddagger}\geq 1). \qquad (4.43)
\end{aligned}$$

Introduce the sequences

$$\begin{aligned}
k_n &= n^{\max(\delta,\phi_{2\ddagger})}, \qquad (4.44) \\
\ell_n &= n^{\max(1,\gamma+\delta,\gamma+\phi_{2\ddagger},\delta+\phi_{1\dagger},\phi_{1\dagger}+\phi_{2\ddagger})}+n\log n 1(\gamma+\delta=1,\gamma>0). \qquad (4.45)
\end{aligned}$$

Theorem 4.1. *Let (4.5) - (4.8) and Assumption 4.1 hold. Then as* $n\to\infty$,

$$(k_n^2/\ell_n)\,(\overline{\nu}-\nu)\Rightarrow\left\{\Psi_1^{**}+\Psi_2^{**}+\Psi_3^{**}+\Psi_4^{**}+\Psi_5^{**}+\Psi_6^{**}\right\}/\left\{\Phi_1^{**}+2\Phi_2^{**}+\Phi_3^{**}\right\}, \qquad (4.46)$$

where by "$\Rightarrow$*" we mean convergence in the Skorohod* J_1 *topology.*

We can deduce from Theorem 4.1 exact rates of convergence, if any, of $\overline{\nu}$ to ν. We have:

$$\begin{aligned}
\overline{\nu} &= \nu+O_e\left(n^{1-2\delta}\right), \text{ if } \delta\geq\phi_{2\ddagger},\ \gamma+\delta<1 \text{ and } \delta+\phi_{1\dagger}\leq 1 \qquad (4.47) \\
\overline{\nu} &= \nu+O_e\left(n^{1-2\delta}\log n\right), \text{ if } \delta\geq\phi_{2\ddagger},\ \gamma+\delta=1 \text{ and } \gamma\geq\phi_{1\dagger} \qquad (4.48) \\
\overline{\nu} &= \nu+O_e\left(n^{\gamma-\delta}\right), \text{ if } \delta\geq\phi_{2\ddagger},\ \gamma\geq\phi_{1\dagger} \text{ and } \gamma+\delta>1 \qquad (4.49)
\end{aligned}$$

$$\overline{\nu} = \nu + O_e\left(n^{\phi_{1\dagger}-\delta}\right), \text{ if } \delta \geq \phi_{2\ddagger},\ \phi_{1\dagger} \geq \gamma \text{ and } \gamma + \delta > 1 \tag{4.50}$$

$$\text{or } \delta \geq \phi_{2\ddagger},\ \phi_{1\dagger} + \delta \geq 1 \text{ and } \gamma + \delta < 1$$

$$\overline{\nu} = \nu + O_e\left(n^{1-2\phi_{2\ddagger}}\right), \text{ if } \phi_{2\ddagger} \geq \delta,\ \gamma + \phi_{2\ddagger} < 1 \text{ and } \phi_{1\dagger} + \phi_{2\ddagger} \leq 1 \tag{4.51}$$

$$\overline{\nu} = \nu + O_e\left(n^{\gamma-\phi_{2\ddagger}}\right), \text{ if } \phi_{2\ddagger} \geq \delta,\ \gamma \geq \phi_{1\dagger} \text{ and } \gamma + \phi_{2\ddagger} > 1 \tag{4.52}$$

$$\overline{\nu} = \nu + O_e\left(n^{\phi_{1\dagger}-\phi_{2\ddagger}}\right), \text{ if } \phi_{2\ddagger} \geq \delta,\ \phi_{1\dagger} \geq \gamma \text{ and } \phi_{1\dagger} + \phi_{2\ddagger} > 1. \tag{4.53}$$

It follows that $\overline{\nu}$ is not even consistent when

$$\phi_{1\dagger} \geq \max\left(\delta, \phi_{2\ddagger}\right), \tag{4.54}$$

(see cases (4.50) and (4.53)) so that the deterministic trend in the cointegrating equation (4.5) dominates both the stochastic and deterministic trends in x_t. Otherwise, $\overline{\nu}$ is consistent.

Notice that $\phi_{2\ddagger} > \phi_{1\dagger}$ means that although both y_t and x_t have a $J\left(\phi_{2\ddagger}\right)$ component, the cointegration residuals $y_t - \nu x_t$ have only a $J\left(\phi_{1\dagger}\right)$ term, a situation that Ogaki and Park (1997) referred to as deterministic cointegration.

The case $\max\left(\delta, \phi_{2\ddagger}\right) > \phi_{1\dagger}$ is very common in applied works, where usually $\phi_{1\dagger} = 1/2$ and often $\phi_{2\ddagger} = 3/2$ or at least $\delta > 1/2$. If, for instance, we consider the four popular examples discussed by Engle and Granger (1987), we always have $\phi_{1\dagger} = 1/2$ and $\delta > 1/2$, and in three cases (consumption and income, wages and prices, money and income; the fourth case is long and short term interest rates) we can also conjecture $\phi_{2\ddagger} = 3/2$. OLS could then provide a valid albeit inefficient first step indication of the value of the cointegrating vector in many applications.

It is also of interest to see in which other situations the deterministic component worsens the rates of convergence, and in which ones it improves them,

with respect to those found by RM. In (4.50), with $\gamma < \phi_{1\dagger} < \delta$, the deterministic trend in (4.5) worsens matters relative to the rates in RM, whereas in (4.51) and (4.52), with $\phi_{2\ddagger} > \delta$, the dominating deterministic trend in $\dot{x}_t$ improves rates; in (4.53) both deterministic trends dominate the corresponding stochastic ones, and the improvement on the situation of no trends depends on whether $\phi_{1\dagger} - \gamma > \phi_{2\ddagger} - \delta$ or the reverse. Finally, in cases (4.47)-(4.49) the same rates are achieved, stochastic trends dominating.

We can deduce from Theorem 4.1 more precise results. These are complicated in case of equalities $\phi_{2\ddagger} = \delta$ and/or $\phi_{1\dagger} = \gamma$, so we look only at strict inequalities. With $\gamma > \phi_{1\dagger}$, $\delta > \phi_{2\ddagger}$ the limit distributions corresponding to (4.47), (4.48) and (4.49) are identical to those of RM Propositions 6.1, 6.2 and 6.5. In (4.52), with $\gamma > \phi_{1\dagger}$ and $\phi_{2\ddagger} > \delta$, stochastic trends dominate in (4.5) and deterministic ones in (4.6), and $n^{\phi_{2\ddagger}-\gamma}(\bar{\nu}-\nu)$ converges to the normal variates $\Phi_3^{-1}\Psi_4$, while in (4.53) with $\phi_{1\dagger} > \gamma$, $\phi_{2\ddagger} > \delta$, $n^{\phi_{2\ddagger}-\phi_{1\dagger}}(\bar{\nu}-\nu)$ converges to the constant $\Phi_3^{-1}\Psi_6$.

The familiar case in which x_t contains a unit root plus linear trend, and cointegrating errors that are $I(0)$ ($\gamma = 0$, $\delta = 1$, $\phi_{1\dagger} = 0.5$, $\phi_{2\ddagger} = 1.5$) comes under (4.53).

The frequency domain regression by Robinson and Marinucci (2000) falls under (4.52) with $\gamma > \phi_{1\dagger}$, $\phi_{2\ddagger} > \delta$: they assumed $\gamma > 1/2$, $\delta + \gamma > 1$ and in that case the simultaneous equation bias does not affect the rate of convergence of the estimate, irrespective of the fact that only a narrow band or all the frequencies are used.

RM though also showed that with no deterministic components and $\gamma+\delta < 1$ (or $\gamma > 0$, $\gamma + \delta = 1$) the simultaneous equation bias does indeed result in a slower convergence of OLS when compared to a frequency domain regression on a shrinking subset of frequencies: that would apply to the situations (4.47), (4.48) and (4.51).

4.2.2 Least squares estimates of Model II

The OLS estimate of $\nu_+ = (\nu, \mu_1')'$ in Model II (4.17) is

$$\widetilde{\nu}_+ = M_{x_+x_+}^{-1} M_{x_+y} \ , \tag{4.55}$$

where $x_{+t} = (x_t, g_1'(t))'$. Define the sequence

$$m_n = \left\{ n^{\max(1,\gamma+\delta,\gamma+\phi_{2\ddagger})} + n \log n 1(\gamma + \delta = 1, \ \gamma \neq 0) \right\}, \tag{4.56}$$

and the matrix sequences

$$B_n = \begin{bmatrix} k_n & 0 \\ 0 & D_{1n}(0) \end{bmatrix}, \quad C_n = \begin{bmatrix} m_n/k_n & 0 \\ 0 & n^{\gamma} I_{p_{11}} \end{bmatrix}, \tag{4.57}$$

$$D_{in}(d) = diag\left\{ n^{\phi_{i1}+d}, ..., n^{\phi_{ip_{i1}}+d} \right\}, i = 1, 2. \tag{4.58}$$

For $i, j = 1, 2$, define the $p_{i1} \times p_{j1}$ matrix $\Xi_{ij}(c, d)$, having $(k, l)th$ element $(\phi_{ik} + \phi_{jl} - c - d)^{-1}$, and the $p_{11} \times 1$ vector Υ, having kth element $(\phi_{1k} + \phi_{2\ddagger})^{-1}$. Also define

$$\Phi_4 = \int_0^1 g_1(r) \widetilde{W}(r; \delta) \, dr, \tag{4.59}$$

$$\Phi_5 = \mu_{2\ddagger} \Upsilon, \tag{4.60}$$

$$\Phi_6 = \Xi_{11}(0, 0); \tag{4.61}$$

$$\Psi_7 = \int_0^1 \widetilde{W}(r; 1) d\overline{W}(r) + \sum_{j=-\infty}^{0} Cov(u_{10}, u_{2j}), \quad \delta = 1, \tag{4.62}$$

$$= \int_0^1 \widetilde{W}(r; \delta) d\overline{W}(r), \qquad \delta > 1; \tag{4.63}$$

$$\Psi_8 = \begin{cases} \int_0^1 g_1(r)\overline{W}(r;\gamma)\,dr & \text{if } \gamma > 0 \\ \int_0^1 g_1(r)d\overline{W}(r) & \text{if } \gamma = 0. \end{cases} \tag{4.64}$$

Let

$$\Phi_4^{**} = \Phi_4 1(\delta \geq \phi_{2\ddagger}), \tag{4.65}$$

$$\Phi_5^{**} = \Phi_5 1(\delta \leq \phi_{2\ddagger}), \tag{4.66}$$

and

$$\Phi = \begin{pmatrix} \Phi_1^{**} + \Phi_2^{**} + \Phi_3^{**} & \Phi_4^{**\prime} + \Phi_5^{**\prime} \\ \Phi_4^{**} + \Phi_5^{**} & \Phi_6 \end{pmatrix}, \tag{4.67}$$

and also

$$\Psi_1^{***} = \Psi_1 1(\gamma + \delta < 1,\ \gamma + \phi_{2\ddagger} \leq 1), \tag{4.68}$$

$$\Psi_2^{***} = \Psi_2 1(\gamma + \delta = 1,\ \phi_{2\ddagger} \leq \delta < 1), \tag{4.69}$$

$$\Psi_3^{***} = \Psi_3 1(\gamma + \delta > 1,\ \delta \geq \phi_{2\ddagger}, \gamma > 0), \tag{4.70}$$

$$\begin{aligned} \Psi_4^{***} = {} & \Psi_4 1(\gamma + \delta > 1,\ \delta \leq \phi_{2\ddagger}) \\ & + \Psi_4 1(\gamma + \delta = 1 \text{ and } \gamma > 0,\ \delta < \phi_{2\ddagger}) \\ & + \Psi_4 1(\gamma + \delta < 1 \text{ or } \gamma = 0 \text{ and } \delta = 1,\ \phi_{2\ddagger} + \gamma \geq 1), \end{aligned} \tag{4.71}$$

$$\Psi_7^{***} = \Psi_7 1(\gamma = 0,\ \delta \geq \max(\phi_{2\ddagger}, 1)), \tag{4.72}$$

and

$$\Psi = \begin{pmatrix} \Psi_1^{***} + \Psi_2^{***} + \Psi_3^{***} + \Psi_4^{***} + \Psi_7^{***} \\ \Psi_8 \end{pmatrix}. \tag{4.73}$$

To avoid multicollinearity in case x_t is dominated by an element of $g_2(t)$ we introduce

Assumption 4.2. *If* $\delta < \phi_{2\ddagger}$ *then* $\phi_{2\ddagger} \neq \phi_{1k}$ *for* $k \in \{1, ..., p_{11}\}$.

This assumption is slightly different from (4.54) because it refers not only to $\phi_{1\dagger}$ but to the other ϕ_{1k}, $k = 1, ..., p_{11}$, as well, and because it replaces the inequality with a milder not equal. Assumption 4.2 should be met in many empirical applications: this is the case for all the examples discussed by Engle and Granger (1987), and indeed of the very general framework $\delta > 1/2$, $\phi_{2\ddagger} = 3/2$ or $1/2$, $\phi_{1\dagger} = 1/2$. It is also not uncommon in the theoretical literature, albeit it is not often explicitly stated: Hansen (1992) and West (1988) implicitly had it, because they considered $(\mu_{11}, ..., \mu_{p1})' = 0$ in (4.5), and so did Johansen (1991), because he discussed $\phi_{1\dagger} = 1/2$, $\delta = 1$. Park and Phillips (1988) on the other hand considered several combinations of stochastic and deterministic trends, including the case $\delta = 1$, $\phi_{2\ddagger} = 3/2$ $\phi_{1\dagger} = 3/2$, in which this assumption is not met. Notice that in that case the rate of convergence of the regression estimate of the cointegrating parameter is lower than stated in Theorems 4.2 and 4.3 below: we refer to the comments of those theorems and to Subsection 4.2.4 for a more detailed explanation of why this should be the case.

Theorem 4.2. *Let (4.5) - (4.8) and Assumptions 4.1 and 4.2 hold. Then as* $n \to \infty$,

$$C_n^{-1} \Phi B_n (\widetilde{\nu}_+ - \nu_+) \Rightarrow \Psi. \tag{4.74}$$

The cumbersome norming (where indeed Φ can be stochastic) is to enable a neat, general statement, C_n and B_n not commuting with the non-diagonal matrix Φ, but we can readily deduce more comprehensible conclusions. Due to the sufficient accounting for deterministic trends in (4.5), $\widetilde{\nu}$ is always consistent. We have the following cases, that we classify according to whether the

stochastic or deterministic component of x_t dominates:

$$\begin{aligned} S &: \delta \geq \phi_{2\ddagger}; \\ D &: \delta \leq \phi_{2\ddagger}. \end{aligned} \tag{4.75}$$

S :

$$\widetilde{\nu} - \nu = O_e(n^{1-2\delta}), \quad \text{if } \delta \geq \phi_{2\ddagger},\ \gamma + \delta < 1, \tag{4.76}$$

$$\widetilde{\nu} - \nu = O_e\left(n^{1-2\delta} \log n\right), \quad \text{if } \delta \geq \phi_{2\ddagger},\ \gamma + \delta = 1,\ \gamma > 0, \tag{4.77}$$

$$\widetilde{\nu} - \nu = O_e\left(n^{-1}\right), \quad \text{if } \gamma = 1 - \delta = 0,\ \phi_{2\ddagger} \leq 1, \tag{4.78}$$

$$\widetilde{\nu} - \nu = O_e\left(n^{\gamma-\delta}\right), \quad \text{if } \delta \geq \phi_{2\ddagger},\ \gamma + \delta > 1, \tag{4.79}$$

D :

$$\widetilde{\nu} - \nu = O_e\left(n^{1-2\phi_{2\ddagger}}\right), \quad \text{if } \delta < \phi_{2\ddagger},\ \gamma + \phi_{2\ddagger} \leq 1, \tag{4.80}$$

$$\widetilde{\nu} - \nu = O_e\left(n^{-1}\right), \quad \text{if } \delta < \phi_{2\ddagger} = 1,\ \gamma = 0, \tag{4.81}$$

$$\widetilde{\nu} - \nu = O_e\left(n^{\gamma-\phi_{2\ddagger}}\right), \quad \text{if } \delta \leq \phi_{2\ddagger},\ \gamma + \phi_{2\ddagger} > 1. \tag{4.82}$$

When $\delta > \phi_{2\ddagger}$ the limit distributions corresponding to (4.76)-(4.79) are those of RM (see Propositions 6.1, 6.2, 6.3, 6.4 and 6.5). When $\phi_{2\ddagger} > \max(\delta, 1-\gamma)$ we deduce that

$$\begin{bmatrix} n^{\phi_{2\ddagger}-\gamma} & 0 \\ 0 & D_{1n}(-\gamma) \end{bmatrix} (\widetilde{\nu}_+ - \nu_+) \rightarrow_d \begin{bmatrix} \Phi_3 & \Phi_5' \\ \Phi_5 & \Phi_6 \end{bmatrix}^{-1} \begin{bmatrix} \Psi_4 \\ \Psi_8 \end{bmatrix}, \tag{4.83}$$

where the right side is a multivariate normal vector.

The case $\gamma = 0$, $\delta = 1$, $\phi_{1\dagger} = 1/2$, $\phi_{2\ddagger} = 3/2$ comes under (4.82), and the vector $\widetilde{\nu}_+$ has the multivariate normal distribution (4.83).

Notice that once that $g_1(t)$ is taken into account, the deterministic component never worsens the rate of convergence with respect to RM, and it actually

improves it when $\phi_{2\ddagger} > .\delta$. This result is hardly surprising: OLS (and GLS) can be intuitively described as an attempt to explain the dynamics of the dependent variable using the dynamics of the other ones, and this is easier the stronger is the signal in the explanatory variables as opposed to the noise in the residuals. The deterministic trend, when present in x_t but not in the residual $y_t - \nu x_t$, contributes to make the signal stronger. This also provides an intuitive explanation of why Assumption 4.2 is needed: otherwise the trend in x_t would not necessarily provide an indication for the long term dynamics of y_t.

The presence of a deterministic trend is also important because, as RM showed, the rate of convergence of the OLS estimate may be effectively worsened by the simultaneous equation bias when $\gamma + \delta \leq 1$ (except if $\gamma = 0$, $\delta = 1$) and there is no trend: this though does not happen if there is a trend and $\phi_{2\ddagger} + \gamma \geq 1$.

4.2.3 Generalised least squares estimates of Model III

The two OLS estimates are computationally convenient, especially as they both avoid the necessity of knowledge of memory parameters and do not require their estimation. However, even when they converge, their rates are not in general optimal, and their limit distributions are for the most part inconvenient for practical use.

To remedy these drawbacks we consider GLS estimation. For c, $d \geq 0$, define

$$z_t(c,d) = (y_t(c), x_t(d))', \tag{4.84}$$

$$w_t(c,d) = \begin{pmatrix} \Delta^c x_t & \Delta^c g_1'(t) & 0 \\ 0 & 0 & \Delta^d g_2'(t) \end{pmatrix}', \tag{4.85}$$

$$\nu_{++} = (\nu, \mu_1', \mu_2')'. \tag{4.86}$$

Thus (4.18), (4.19) can be written

$$z_t(\gamma, \delta) = w_t'(\gamma, \delta)\nu_{++} + v_t, \tag{4.87}$$

where now $v_t = (v_{1t}(\gamma), v_{2t}(\delta))'$. In RH the right hand side was simplified by the correct assumption that $\mu_1 = 0$, $\mu_2 = 0$ and $v_t = (u_{1t}, u_{2t})'$. Here we simultaneously estimate ν with μ_1, μ_2 and show that s_{1t}, s_{2t}, the trends which have orders $\phi_{1k} < \gamma$, $\phi_{2j} < \delta$ and are not included in $g_1(t)$, $g_2(t)$, exert negligible effect.

As in RH we construct two kinds of estimate of GLS type that allow for flexible parametric modelling of $f_u(\lambda)$, that is, the autocorrelation structure of u_t, and that either depend on known γ, δ or allow substitution of estimates of these without affecting limiting distributional properties. One kind of estimate is 'time-domain', the other is 'frequency-domain', and the practitioner's choice between them is based on computational considerations and taste.

The time-domain estimate involves autoregressive (AR) transformation. From Assumption 4.1, u_t has an AR representation

$$B(L)u_t = \varepsilon_t, \tag{4.88}$$

with $B(s) = I_2 - \sum_{j=1}^{\infty} B_j s^j$, such that the B_j are unknown 2×2 matrices. We know functions $\Omega(h)$, $B_j(h)$, where $h \in \mathbb{R}^p$, $p \geq 1$, such that $\Omega = \Omega(\theta)$, $B_j \equiv B_j(\theta)$ for some $\theta \in \mathbb{R}$. Define $B(s;h) = I_2 - \sum_{j=1}^{\infty} B_j(h)s^j$ and then

$$\widetilde{a}(c,d,h) = \sum_{t=1}^{n} \{B(L;h)w_t'(c,d)\}' \Omega(h)^{-1} \{B(L;h)z_t(c,d)\}, \tag{4.89}$$

$$\widetilde{b}(c,d,h) = \sum_{t=1}^{n} \{B(L;h)w_t'(c,d)\}' \Omega(h)^{-1} \{B(L;h)w_t'(c,d)\}. \tag{4.90}$$

Each of the AR transformations is truncated since $w_t(c,d) = 0$, $z_t(c,d) = 0$, $t \leq 0$. Now write

$$\widetilde{\nu}_{++}(c,d,h) = \widetilde{b}(c,d,h)^{-1}\widetilde{a}(c,d,h) \tag{4.91}$$

and consider

$$\widetilde{\nu}_{++}(\gamma,\delta,\theta),\ \widetilde{\nu}_{++}(\gamma,\delta,\widehat{\theta}),\ \widetilde{\nu}_{++}(\widehat{\gamma},\delta,\widehat{\theta}),\ \widetilde{\nu}_{++}(\gamma,\widehat{\delta},\widehat{\theta}),\ \widetilde{\nu}_{++}(\widehat{\gamma},\widehat{\delta},\widehat{\theta}), \tag{4.92}$$

for given estimates $\widehat{\gamma}$, $\widehat{\delta}$, $\widehat{\theta}$. These estimates of ν_{++} respectively cover the cases in which γ, δ and θ are all known, γ and δ are known but θ is not, only δ is known, only γ is known, and γ, δ and θ are all unknown. Thus $\widetilde{\nu}_{++}(\gamma,\delta,\widehat{\theta})$ covers the familiar case where $\gamma = 0$, $\delta = 1$ is known, and u_t is, say, white noise or AR(1); $\widetilde{\nu}_{++}(\widehat{\gamma},\delta,\widehat{\theta})$ with $\delta = 1$ accepts the evidence of unit root behaviour suggested by a number of empirical studies of macroeconomic variables but allows for the possibility of long memory or mean-reversion in the cointegrating error; $\widetilde{\nu}_{++}(0,\widehat{\delta},\widehat{\theta})$ insists only that the cointegrating error has short memory, possibly white noise; and $\widetilde{\nu}_{++}(\widehat{\gamma},\widehat{\delta},\widehat{\theta})$ is completely agnostic.

When u_t is a not a finite-degree AR process $\widetilde{\nu}_{++}$ can still be computed, but the following frequency-domain estimates may be preferable, making use of the neat form of the spectral density matrix $f_u(\lambda)$ when u_t is a finite-degree moving average (MA) sequence or a more general ARMA process, or in some other models. Let $f_u(\lambda;h)$ be a known function of $\lambda \in (-\pi,\pi]$ and $h \in \mathbb{R}^p$, such that $f_u(\lambda;\theta) = f_u(\lambda)$, so

$$f_u(\lambda;h) = (2\pi)^{-1}B(e^{i\lambda};h)^{-1}\Omega(h)B(e^{-i\lambda};h)^{-1'}. \tag{4.93}$$

Let $F_{w(c,d)}(\lambda)$ and $F_{z(c,d)}(\lambda)$ be the Fourier transforms of the vectors $w_t(c,d)$

and $z_t(c,d)$, and put

$$a(c,d,h) = \sum_{j=1}^{n} F_{w(c,d)}(\lambda_j) f_u(\lambda_j;h)^{-1} F_{z(c,d)}(-\lambda_j), \tag{4.94}$$

$$b(c,d,h) = \sum_{j=1}^{n} F_{w(c,d)}(\lambda_j) f_u(\lambda_j;h)^{-1} F'_{w(c,d)}(-\lambda_j). \tag{4.95}$$

Define

$$\widehat{\nu}_{++}(c,d,h) = b(c,d,h)^{-1} a(c,d,h). \tag{4.96}$$

Consider the frequency-domain variants of the five estimates (4.92),

$$\widehat{\nu}_{++}(\gamma,\delta,\theta),\ \widehat{\nu}_{++}(\gamma,\delta,\widehat{\theta}),\ \widehat{\nu}_{++}(\widehat{\gamma},\delta,\widehat{\theta}),\ \widehat{\nu}_{++}(\gamma,\widehat{\delta},\widehat{\theta}),\ \widehat{\nu}_{++}(\widehat{\gamma},\widehat{\delta},\widehat{\theta}). \tag{4.97}$$

When $v_t = (v_{1t}, v_{2t})'$ is *a priori* white noise, $\widetilde{\nu}(c,d,h) \equiv \widehat{\nu}(c,d,h)$.

To handle the last four estimates in (4.92) and (4.97) we introduce the following further assumptions. Denote by Θ the compact set of all admissible values of $\widehat{\theta}$.

Assumption 4.3.

(i) $f_u(\lambda;\theta) = f_u(\lambda)$.

(ii) $f_u(\lambda;h)$ *has determinant bounded away from zero on* $([-\pi,\pi]\times\Theta)$.

(iii) $f_u(\lambda;h)$ *is boundedly differentiable in* h *on* $([-\pi,\pi]\times\Theta)$, *with derivative that is continuous in* h *at* $h=\theta$ *for all* λ.

(iv) $f_u(\lambda;\theta)$ *is differentiable in* λ, *with derivative satisfying a Lipschitz condition of order greater than* $1/2$ *in* λ.

(v) $(\partial/\partial h) f_u(\lambda;h)$ *is differentiable in* λ *at* $h=\theta$, *with derivative satisfying a Lipschitz condition of order greater than* $1/2$ *in* λ.

Assumption 4.4.

(i) There exists $C < \infty$ *such that*

$$|\widehat{\gamma}| + \left|\widehat{\delta}\right| \leq C, \tag{4.98}$$

and $\rho > \max(0, 1 - \beta)$ *such that*

$$\widehat{\gamma} = \gamma + O_p\left(n^{-\rho}\right), \ \widehat{\delta} = \delta + O_p\left(n^{-\rho}\right). \tag{4.99}$$

(ii)

$$\widehat{\theta} = \theta + O_p(n^{-\frac{1}{2}}), \text{ where } \theta \in \Theta. \tag{4.100}$$

As for Assumption 4.1 above, these too are identical to assumptions in RH.

Assumption 4.3 seems very mild, and it is satisfied by any stationary and invertible ARMA model.

As for Assumption 4.4, condition (4.98) is standard and it is met for example when γ and δ are assumed to lie in a compact set, as is the case when they are estimated by implicitly defined extremum estimates, but (4.99) and (4.100) need a comment. Considering the second part of (4.99) first, the estimation of δ in RH can be based on the Whittle estimate discussed by Velasco and Robinson (2000) (indeed, given the assumption $\beta > 1/2$, even some semi-parametric procedures may satisfy the requirement for $\widehat{\delta}$). If the data are contaminated by a known polynomial trend, it can be removed by a preliminary regression of x_t on $g_2(t)$, so that δ is then estimated from the residuals. For this purpose we should first prove that the Whittle estimate discussed by Velasco and Robinson is root-n consistent even when computed using those residuals rather than a zero-mean, $I(\delta)$ stochastic process. This seems indeed to be the case but, as is often the case with justifying insertion of residuals

in implicitly-defined estimation procedures, the proof is rather tedious so the details are not pursued.

Alternatively, if all $\phi_{2j} - \frac{1}{2}$ in $g_2(t)$ are integers, the use of Velasco and Robinson's (2000) procedure based on the raw x_t, but using a Kolmogorov taper of sufficiently high order, will exactly remove such a polynomial trend. Strictly, this does not require knowledge of ϕ_{21}, but rather of the largest value we might anticipate for ϕ_{21}. On the other hand, we might carry out the initial OLS estimation by including such arbitrarily large powers of t, so there may not be a great deal to choose between the two approaches. The higher the taper order, the greater the imprecision in estimation of δ, while inclusion of unnecessary regressors in the OLS approach is liable to have similar effect. The OLS approach has the advantage over tapering that non-integer powers of t may be employed. Tapering seems to be needed anyway to estimate large enough values of δ in the Velasco and Robinson (2000) approach, unless some preliminary information on the range in which δ lies is available, and then large δ can be estimated after differencing the data an appropriate number of times.

Velasco and Robinson's (2000) approach can again be considered to estimate γ, with the use of residuals from the regression of y_t on x_t and $g_2(t)$ apparently necessary, while tapering is then unnecessary if $\gamma < \frac{1}{2}$ is anticipated. Notice though that discussing the same problem in a situation in which no trends were present, Robinson (2005b) showed root-n consistent estimation of γ (and of θ) is possible when the residuals are computed using an estimate of the cointegrating parameter that converges to ν fast enough, the required rate being $n^{\delta-\gamma-\varepsilon}$ for $\varepsilon > 0$ and provided that $\delta > \gamma + 1/2$. Without deterministic trends in the model, an OLS estimate only meets this criterion when $\gamma+\delta \geq 1$, although the NBLS also meets it when $\gamma + \delta < 1$; when deterministic trends are present and $\phi_{2\ddagger} > \delta$, the milder $\phi_{2\ddagger} + \gamma \geq 1$ may be required.

An alternative approach to estimate γ can be based on Chen and Hurvich (2003). They employed tapers in estimating the cointegrating coefficient in a fractional system with deterministic trends. The idea of tapering the data in a regression framework is not new: earlier, Robinson (1986) had employed tapering in reducing errors-in-variables bias in band-spectrum regression in which (contrary to Robinson (1994b)) bands do not degenerate asymptotically and the processes have short memory. Chen and Hurvich (2003), however, are concerned with estimating the relation between underlying (possibly non-stationary) stochastic components that are corrupted by deterministic trends. These (assumed to be polynomial in t with integer exponents) are handled by differencing, with tapering then employed in a modification of the narrow-band least squares regression estimate of Robinson (1994b). The estimate of ν is not optimal but has the rate of convergence required by Robinson (2005b) to obtain root-n consistent estimates of γ.

The estimation of θ can be based on the same procedures: indeed the Whittle estimation that we proposed for δ and γ should also give root-n consistent estimates for those θ parameters that are in the pseudo spectral densities of $u_{1t}(-\gamma)$ and of $u_{2t}(-\delta)$. This procedure does not include the elements of the pseudo cross-spectrum, because Velasco and Robinson (2000) only considered univariate processes, but it could be extended by considering the two residuals processes jointly, in the same way as Lobato (1999) did for the univariate Gaussian semiparametric estimate of the memory parameter of Robinson (1995b).

Define

$$\Lambda_i(d) = diag\left\{\frac{\Gamma(\phi_{i1}+1/2)}{\Gamma(\phi_{i1}-d+1/2)},...,\frac{\Gamma(\phi_{ip_{i1}}+1/2)}{\Gamma(\phi_{ip_{i1}}-d+1/2)}\right\},\quad i=1,2, \quad (4.101)$$

$$Q(r)$$
$$= \begin{pmatrix} \widetilde{W}(r;\beta)\,1(\delta \geq \phi_{2\ddagger}) + \mu_{2\ddagger}\frac{\Gamma(\phi_{2\ddagger}+1/2)}{\Gamma(\phi_{2\ddagger}-\gamma+1/2)} r^{\phi_{2\ddagger}-\gamma-1/2} 1(\delta \leq \phi_{2\ddagger}) & 0 \\ \Lambda_1(\gamma) g_1(r) r^{-\gamma} & 0 \\ 0 & \Lambda_2(\delta) g_2(r) r^{-\delta} \end{pmatrix} \tag{4.102}$$

and introduce the matrix sequence

$$D_n = \begin{pmatrix} n^{\max(\delta,\phi_{2\ddagger})-\gamma} & 0 & 0 \\ 0 & D_{1n}(-\gamma) & 0 \\ 0 & 0 & D_{2n}(-\delta) \end{pmatrix}. \tag{4.103}$$

Theorem 4.3. *Let (4.5) - (4.8) and Assumptions 4.1 - 4.4 hold, with $q > 1/(2\beta-1)$ in Assumption 4.1. Then, denoting by $\widehat{\nu}^{**}_{++}$ any of the estimates in (4.92) or (4.97), we have as $n \to \infty$,*

$$D_n\left(\widehat{\nu}^{**}_{++} - \nu_{++}\right) \Rightarrow \left\{\int_0^1 Q(r) f_u(0)^{-1} Q'(r) dr\right\}^{-1} 2\pi \int_0^1 Q(r) B(1)' \Omega^{-1} dW(r). \tag{4.104}$$

As in RH we find that we can estimate ν, along with μ_1, μ_2, as well without knowing γ and/or δ and/or θ as knowing them, so that efficiency of estimation of γ, δ and θ does not matter.

When $\delta > \phi_{2\ddagger}$, we have precisely the same limit distribution and rate of convergence for $\hat{\nu}^{**}$ as the estimate of RH, which ignores the possibility of trends, and the same limit distribution. The distribution is changed when $\delta = \phi_{2\ddagger}$, but the rate of convergence is not. When $\delta < \phi_{2\ddagger}$ the rate is faster than in RH. This confirms the finding we made discussing Model II that when the deterministic component is relevant (i.e. when $\phi_{2\ddagger} > \delta$), it does contribute

to increase the precision and it should then not be removed (whether by differencing, tapering, or by filtering the data with a preliminary regression).

When $\delta > \phi_{2\ddagger}$ there is mixed normal asymptotics as in RH, as there is also for $\delta = \phi_{2\ddagger}$, while when $\delta < \phi_{2\ddagger}$ the limit distribution is normal, as in the classical case $\gamma = 0$, $\delta = 1$, $\phi_{1\dagger} = 1/2$, $\phi_{2\ddagger} = 3/2$. Thus in all cases we can expect Wald tests on ν_{++}, for example tests on ν (such as $\nu = 1$ as in PPP testing) or on μ_1 (e.g. $\mu_1 = 0$, to test whether deterministic trends affect y_t only through x_t), to have standard, χ^2, asymptotics. For this purpose, we introduce the null hypothesis:

$$H_0 : \quad E\nu_{++} = e, \tag{4.105}$$

where E is a given $q \times (1 + p_{11} + p_{21})$ matrix of rank $q < 1 + p_{11} + p_{21}$ and e is a given $q \times 1$ vector.

Theorem 4.4. *Denoting by b^{**} any of the quantities $\widetilde{b}(c,d,h)$, $b(c,d,h)$, with $c = \gamma$ or $\widehat{\gamma}$, $d = \delta$ or $\widehat{\delta}$ and $h = \theta$ or $\widehat{\theta}$, under (4.105) the Wald statistics*

$$\left(E\widehat{\nu}^{**}_{++} - e\right)' \left(Eb^{**-1}E'\right)^{-1} \left(E\,\widehat{\nu}^{**}_{++} - e\right) \to_d \chi^2_q, \quad \text{as} \quad n \to \infty. \tag{4.106}$$

4.2.4 Cointegration with common deterministic trend

When Assumption 4.2 is not met, the explanatory variables x_t and $t^{\phi_{1k}-1/2}$ are collinear in large samples for some $k \in \{1...p_{11}\}$, so $PlimB_n^{-1}M_{x_+x_+}B_n^{-1}$ cannot be inverted.

We solve the problem by removing from x_t those trends that are also present in $g_1(t)$. Let $\kappa_1^{\Diamond}$ be a p_{11} dimensional vector having kth element

$$\kappa^{\Diamond}_{1k} = \begin{cases} \mu_{2j} \text{ if } \mu_{1k} \neq 0 \text{ and } \phi_{1k} = \phi_{2j} \text{ for a } j \in \{1...p_{21}\} \\ 0 \text{ otherwise,} \end{cases} \tag{4.107}$$

and introduce

$$g_2^{\diamond}(t) = g_2(t) - \kappa_1^{\diamond\prime} g_1(t) = \sum_{j=1}^{p_{21}} \mu_{2j}^{\diamond} t^{\phi_{2j}-1/2} \tag{4.108}$$

where the weights are

$$\mu_{2j}^{\diamond} = \begin{cases} 0 \text{ if } \phi_{2j} = \phi_{1k} \text{ for a } k \in \{1...p_{11}\} \text{ and such that } \mu_{1k} \neq 0 \\ \mu_{2j} \text{ otherwise,} \end{cases} \tag{4.109}$$

so they are the original weights unless the trend is common both to $g_1(t)$ and to $g_2(t)$, in which case they are replaced by 0. We can then define

$$j^{\diamond} = \min\left\{j : j \in \{1...p_{21}\}, \mu_{2j}^{\diamond} \neq 0\right\} \tag{4.110}$$

and introduce the abbreviation $\phi_{2\diamond} = \phi_{2j^{\diamond}}$: this indicates the trend with higher exponent among those of x_t that are not present in $g_1(t)$. Introducing the (invertible) matrices

$$G_+ = \begin{bmatrix} 1 & -\kappa_1^{\diamond\prime} \\ 0 & I_{p_{11}} \end{bmatrix} \tag{4.111}$$

$$G_{++} = \begin{bmatrix} 1 & -\kappa_1^{\diamond\prime} & 0 \\ 0 & I_{p_{11}} & 0 \\ 0 & 0 & I_{p_{21}} \end{bmatrix} \tag{4.112}$$

we can then transform (4.17) and (4.87) in

$$y_t = x_{+t}'\nu_+ + v_t = x_{+t}' G_+' G_+^{-1\prime}\nu_+ + v_t = x_{+t}^{\diamond\prime}\nu_+^{\diamond} + v_t \tag{4.113}$$

$$z_t(\gamma,\delta) = w_t'(\gamma,\delta)\nu_{++} + v_t = w_t^{\diamond\prime}(\gamma,\delta)\nu_{++}^{\diamond} + v_t \tag{4.114}$$

where $x_{+t}^{\diamond} = G_+ x_{+t}$, $\nu_+^{\diamond} = G_+^{-1\prime}\nu_+$, $w_t^{\diamond}(\gamma,\delta) = G_{++} w_t^{\diamond}(\gamma,\delta)$, $\nu_{++}^{\diamond} = G_{++}^{-1\prime}\nu_{++}$.

Theorems 4.2 and 4.3 can then be formulated for $\widetilde{\nu}_{+}^{\Diamond}$ and for $\widehat{\nu}_{++}^{**\Diamond}$ and the properties of these estimates can be derived. This essentially entails reformulating the elements in C_n, Φ, B_n, Ψ, D_n, $Q(r)$ in terms of $\phi_{2\Diamond}$ rather than of $\phi_{2\ddagger}$.

The statements are then omitted because we prefer to keep the focus on the discussion rather than on the notation.

In order to comment on the effects of the transformations (4.113) and (4.114), let $\widetilde{\nu}^{\Diamond}$ and $\widehat{\nu}^{**\Diamond}$ be the OLS and GLS estimates of $\nu^{\Diamond}$, the coefficient of $x_t - \kappa_1^{\Diamond\prime} g_1(t)$ in the transformed models. Since, in both the cases, $\nu = \nu^{\Diamond}$, we can derive the properties of the estimates of the cointegrating parameter in the original models simply by looking at the estimates in the transformed models instead.

The main conclusion is that, since $\phi_{2\Diamond} > \phi_{2\ddagger}$, the rate of convergence of $\widetilde{\nu}^{\Diamond}$ to ν is lower than what is stated in (4.80) - (4.82); in the same way, the rate of convergence of $\widehat{\nu}^{**\Diamond}$ to ν is lower than $n^{\phi_{2\ddagger}-\gamma}$. Also notice that the new rates may depend on δ as well, because there may be cases in which $\phi_{2\ddagger} > \delta$ but $\delta > \phi_{2\Diamond}$.

We illustrate these comments with an example. Consider

$$\begin{cases} y_t = \nu x_t + \mu_1 t + \varepsilon_{1t} \\ x_t = \mu_2 t + \sum_{s=1}^{t} \varepsilon_{2s} \end{cases} \tag{4.115}$$

where ε_{1t}, ε_{2t} are $I(0)$ processes and $\mu_1 \neq 0$, $\mu_2 \neq 0$, so that $\gamma = 0$, $\delta = 1$,

$\phi_{11} = 3/2$, $\phi_{21} = 3/2$. Then

$$
\begin{aligned}
& Plim B_n^{-1} M_{x+x+} B_n^{-1} \\
= \; & Plim \begin{bmatrix} n^{-3/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \begin{bmatrix} \sum_{t=1}^n x_t^2 & \sum_{t=1}^n x_t t \\ \sum_{t=1}^n x_t t & \sum_{t=1}^n t^2 \end{bmatrix} \begin{bmatrix} n^{-3/2} & 0 \\ 0 & n^{-3/2} \end{bmatrix} \\
= \; & \frac{1}{3} \begin{bmatrix} \mu_2^2 & \mu_2 \\ \mu_2 & 1 \end{bmatrix} \qquad (4.116)
\end{aligned}
$$

which, as we anticipated, cannot be inverted. By applying

$$
G_+ = \begin{bmatrix} 1 & -\mu_2 \\ 0 & 1 \end{bmatrix}, \qquad (4.117)
$$

the new regression model is

$$
y_t = \nu^{\Diamond} x_t^{\Diamond} + \mu_1^{\Diamond} t + \varepsilon_{1t}, \qquad (4.118)
$$

where actually $x_t^{\Diamond} = \sum_{s=1}^t \varepsilon_{2s}$, $\nu^{\Diamond} = \nu$, $\mu_1^{\Diamond} = \mu_1 + \nu\mu_2$ and $x_{t+}^{\Diamond} = \left(\sum_{s=1}^t \varepsilon_{2s}, t\right)'$. Since Assumption 4.2 holds in the transformed model, Theorem 4.2 can then be applied and $\widetilde{\nu}^{\Diamond} - \nu^{\Diamond} = O_e(n^{-1})$. Moreover, since $\nu^{\Diamond} = \nu$, it also holds that $\widetilde{\nu} - \nu = O_e(n^{-1})$. The rate of convergence n is then less than the $n^{3/2}$ we would have obtained if $\mu_1 = 0$.

A similar treatment of course allows the calculation of the rate of convergence and of the limit distribution of the GLS estimates $\widehat{\nu}^{**\Diamond}$ and then of $\widehat{\nu}^{**}$: in this example, $\widehat{\nu}^{**} - \nu = O_e(n^{-1})$.

These results have a fairly intuitive explanation, that complements the comment on the faster rate of convergence induced by some deterministic trends. We already noticed that, when $\phi_{2\ddagger} > \delta$, the information about the long term dynamics in x_t is mainly conveyed by the deterministic trend $t^{\phi_{2\ddagger}-1/2}$, but when

Assumption 4.2 is not met, the contribution of that component of x_t cannot be distinguished by the term in $g_1(t)$, so other components of x_t, of a lower order, must be used to derive a pattern for y_t that is common to x_t only: in the example (4.115), the linear trend characterises both x_t and $g_1(t)$, so $\sum_{s=1}^{t} \varepsilon_{2s}$ has to be used instead.

4.3 Monte Carlo simulations

We have encountered convergence rates and limit distributions that can vary substantially across both types of estimate and memory characteristics of the data generating process. In order to investigate how reliable a guide the asymptotic theory is in moderate-sized samples, a small Monte Carlo study was carried out. We generated data from (4.5) and (4.6) for several combinations of stochastic and deterministic trends. Two different specifications for each equation were employed. For the cointegrating equation (4.5) these were:

$$DT1a: \quad p_1 = 1;\ \mu_{11} = 1;\ \phi_{11} = 0.5.$$
$$DT1b: \quad p_1 = 1;\ \mu_{11} = \mu_{12} = 1;\ \phi_{12} = 0.5,\ \phi_{11} = 1.5.$$

For the x equation (4.6) they were

$$DT2a: \quad p_2 = 1;\ \mu_{21} = 1;\ \phi_{21} = 1.5.$$
$$DT2b: \quad p_2 = 2;\ \mu_{21} = \mu_{22} = 1;\ \phi_{22} = 1.5,\ \phi_{21} = 2.5.$$

Thus $DT1a$ consists only of an intercept, while $DT1b$ is a time trend; $DT2a$ and $DT2b$ include no constants, the former consisting only of a linear term, the latter a linear and a quadratic. We employed all four combinations, $DTa = DT1a \times DT2a$, $DTb = DT1a \times DT2b$, $DTc = DT1b \times DT2a$ and $DTd = DT1b \times DT2b$. The stochastic component of the model was specified as follows. We took $A_j \equiv 0$, $j \geq 1$, in Assumption 1 where the covariance matrix Ω of

Gaussian $u_t = \varepsilon_t$ was given by

$$\Omega = \begin{bmatrix} \tau & \rho\tau^{\frac{1}{2}} \\ \rho\tau^{\frac{1}{2}} & 1 \end{bmatrix} \tag{4.119}$$

for varying ρ, τ. Our stochastic trends for the cointegrating equation and the x equation were then determined by the following six choices of (γ, δ):

STa: $(\gamma, \delta) = (0, 0.6)$

STb: $(\gamma, \delta) = (0, 1.2)$

STc: $(\gamma, \delta) = (0, 2)$

STd: $(\gamma, \delta) = (0.4, 0.6)$

STe: $(\gamma, \delta) = (0.4, 1.2)$

STf: $(\gamma, \delta) = (0.4, 2)$

Notice that STd is not covered by the theory for our GLS estimate. We considered each of the 24 combinations of the 4 DT's and 6 ST's. Finally we took $\nu = 1$.

Our design includes cases where the deterministic trends improve, leave unchanged, or reduce the rate for $\overline{\nu}$, or even make it inconsistent, and cases where rates for $\widetilde{\nu}$ and $\widehat{\nu}$ are either unchanged or improved. Nevertheless, it would be possible to choose combinations that might seem more "interesting" in view of the various outcomes reported in Subsections 4.2.1-4.2.4. Our choice is motivated by two factors. One is to enable comparison with the design of RH, who used precisely the same ST's, and values of ρ and τ, with no DT's, and computed $\overline{\nu}$ and the GLS estimate of Subsection 4.2.3 simplified by (correctly) assuming no deterministic trends. The other is that non-fractional powers of t in the DT's seem rather typical of current macroeconometric practice. However, our simulations fall very far short indeed of providing a comprehensive

study, especially as behaviour will vary with the μ_{ij} as well as the ϕ_{ij}, not to mention being affected by the presence of short memory autocorrelation.

Tables 4.1-4.3 indicate rates of convergence of the various estimates of ν: $\overline{\nu}$, the OLS estimate in Subsection 4.2.1; $\widetilde{\nu}$, the first element of the OLS estimate $\widetilde{\nu}_+$ in Subsection 4.2.2; $\widehat{\nu}$, the first element of the generic GLS estimate $\widehat{\nu}^{**}_{++}$ in Subsection 4.2.3 (notice indeed that in our case of white noise AR, (4.92) and (4.97) are identical). For comparison we include also the rates when there are no deterministic trends in either (4.5) or (4.6).

Table 4.1: Convergence rates (powers of n) of $\overline{\nu}$

	STa	STb	STc	STd	STe	STf
0^+	0.2	1.2	2	$0.2^\dagger$	0.8	1.6
DTa	1	1	1.5	1	1	1.5
DTb	2	2	2	2	2	2
DTc	X	X	0.5	X	X	0.5
DTd	1	1	1	1	1	1

$^+$: Refers to the case of no trend, as in RM.
†: The rate is actually $n^{0.2}/\ln n$.
X: Inconsistent.

Table 4.2: Convergence rates (powers of n) of $\widetilde{\nu}$

	STa	STb	STc	STd	STe	STf
DTa	1.5	1.5	2	1.1	1.1	1.6
DTb	2.5	2.5	2.5	2.1	2.1	2.1
DTc	$0.2^\ddagger$	$1.2^\ddagger$	2	$0.2^{\ddagger\dagger}$	$0.8^\ddagger$	1.6
DTd	2.5	2.5	2.5	2.1	2.1	2.1

‡: Assumption 4.2 not met.
†: The rate is actually $n^{0.2}/\ln n$.

We computed the OLS estimates $\overline{\nu}$ and $\widetilde{\nu}$ described above, as well as infeasible and feasible GLS estimates of ν, namely $\widehat{\nu}_I = \widehat{\nu}(\gamma, \delta, \widehat{\theta})$ and $\widehat{\nu}_F = \widehat{\nu}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta})$ where $\widehat{\nu}(c, d, h)$ is the first element of $\widehat{\nu}_{++}(c, d, h)$ with $\widehat{\gamma}$, $\widehat{\delta}$ and $\widehat{\theta}$ as follows.

Table 4.3: Convergence rates (powers of n) of $\widehat{\nu}$

	STa	STb	STc	STd	STe	STf
DTa	1.5	1.5	2		1.1	1.6
DTb	2.5	2.5	2.5		2.1	2.1
DTc		$1.2^{\ddagger}$	2		$0.8^{\ddagger}$	1.6
DTd	2.5	2.5	2.5		2.1	2.1

‡: Assumption 4.2 not met.

Convergence rates not reported when Assumption 4.4 or $\beta > 0.5$ is not met.

Having obtained $\widetilde{\nu}_+ = (\widetilde{\nu}, \widetilde{\mu}_2')'$, we computed the $\widetilde{v}_{1t} = y_t - \widetilde{\nu}_+' x_{+t}$ and $\widetilde{v}_{2t} = x_t - \widetilde{\mu}_2' g_2(t)$. For given (c, d) define $\widetilde{u}_t(c, d) = (\widetilde{v}_{1t}(c), \widetilde{v}_{2t}(d))'$. Since u_t is white noise, θ parameterises only Ω, for which we employed the estimates $\widetilde{\Omega} = \Omega_{(\gamma,\delta)}$ and $\widehat{\Omega} = \Omega_{(\widehat{\gamma},\widehat{\delta})}$, where $\Omega_{(c,d)} = n^{-1}\sum_{t=1}^{n} \widetilde{u}_t(c, d)\widetilde{u}_t'(c, d)$, the former referring to the case (γ, δ) known (i.e. $\widehat{\nu}_I$), the latter to the case (γ, δ) is estimated by $(\widehat{\gamma}, \widehat{\delta})$ (i.e. $\widehat{\nu}_F$). Here $\widehat{\gamma}$ was the Whittle pseudo-maximum likelihood estimate of Velasco and Robinson (2000) applied to the series $\widetilde{v}_{1t}$ (without tapering the data). Likewise $\widehat{\delta}$ was obtained by the same method, but applied to the first difference of $\widetilde{v}_{2t}$, then adding back 1 when $\delta = 1.2$ or $\delta = 2$ (as an alternative to tapering, again as in the simulations of RH).

For the null hypothesis (4.105), we took $\nu = 1$, and in the Wald statistic (4.106) we computed W_I based on $\widehat{\nu}_I$ and $b(\gamma, \delta, \widetilde{\theta})$ and W_F based on $\widehat{\nu}_F$ and $b(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta})$, $\widetilde{\theta}$ and $\widehat{\theta}$ respectively denoting the vectors consisting of the three distinct elements of $\widetilde{\Omega}$ and $\widehat{\Omega}$. We employed sample sizes $n = 64$, 125 and 256, with 1000 replications.

We present two groups of simulations: in the first we kept $\tau = 1$ and $\rho = 0.5$ fixed, and considered the 24 combinations of deterministic and stochastic trends, while in the second group we focused on $\delta = 0.6$, $\gamma = 0$ and studied the effect of alternative combinations of τ and ρ.

In Tables 4.7 and 4.8 we present the Monte Carlo bias (the difference between the average of the estimates and ν), and in Tables 4.9 and 4.10 the

Monte Carlo sample standard deviation for the first group of simulations.

Consistency of $\overline{\nu}$ fails altogether if $\phi_{1\dagger} \geq \max\left(\phi_{2\ddagger}, \delta\right)$, so under DTc for STa, STb, STd, STe: in our example the estimate should converge to 2, thus having a bias of 1, and this is indeed what we observed. Notice also the much larger standard deviation when $\delta = 1.2$ than when $\delta = 0.6$: intuitively the properties of the estimates in the examples we considered depend on the fact that $\phi_{2\ddagger}$ dominated δ, so the effect is stronger (the dispersion around the limit value 2 is smaller) the larger $\phi_{2\ddagger} - \delta$.

In all the other situations the estimate is consistent: the correlation between the deterministic component omitted from the specification and the one in x_t should generate a lower order bias, and given the values we considered for μ_{11} and μ_{21} this should always be positive. This was indeed the case, but notice that it was always very little when not negligible altogether: in the worst situation, that is under DTc for STc and STf, it was still below 0.07 in the smallest sample and it was below 0.02 when 256 observations were used.

When $g_1(t)$ is correctly specified the OLS estimate $\widetilde{\nu}$ is always consistent. The stochastic component dominates in x_t in 4 of the 24 combinations considered, namely when $\delta = 2$ and $\phi_{21} = 1.5$ (for any γ, $g_1(t)$), and in this situation $\widetilde{\nu}$ should have the same rate of convergence as in RM. Comparing our results to the ones in RH, who performed a similar exercise for the case with no trends, we found for these four cases that the biases were indeed very little and of approximately similar size both with and without deterministic trends in the model. The standard deviations on the other hand were of comparable dimension or even smaller in our simulations but under DTc, when they were rather larger: this may have been the effect of a certain correlation that remains between x_t and $g_1(t)$ at least in finite samples, due to the presence of t in both the terms.

In the remaining 20 situations the estimates are dominated by the deter-

ministic component in x_t.

Under DTc, Assumption 4.2 does not hold in these cases: the multicollinearity between the deterministic trend in x_t and in $g_1(t)$ meant that it was not possible to use that piece of information to make inference about ν, so the rate of convergence is the same as that of RM. The lower order bias, due to the correlation between x_t and v_t, was still present, and indeed both the bias and the dispersion were even bigger than in the corresponding simulations of RH with the same combination of δ and γ.

Under DTa, DTb and DTd, on the other hand, the improvement on RH in terms of bias and dispersions was really remarkable: considering for example $\delta = 0.6$, $\gamma = 0$ (STa) with $n = 64$, the bias then dropped from 0.194 to 0.0009 and the standard deviation from 0.100 to 0.0069.

It is also interesting to compare the effect of a different value of δ in all the cases in which Assumption 4.2 is met and $\delta < \phi_{2\ddagger}$. RH found that the rate of convergence increased with δ, and in their simulation they found that it resulted in smaller bias and dispersion; we saw that in our case the rate of convergence does not change, the stochastic part of x_t acting as a noise (albeit a very persistent one, of course), and this was indeed the case in our Monte Carlo exercise.

Increasing γ from 0 to 0.4 on the other hand had a visible, if rather small, effect. This can be checked comparing STa, STb and STc with STd, STe and STf. The largest effect on the bias, indeed the only one that really could not be neglected, was for the increase of γ when $\delta = 0.6$ under DTc. This may seem surprising because it is actually the only situation in which the rate of convergence should be nearly unaffected, the only change being the additional factor $(\ln n)^{-1}$, but it is similar to the results in RH. In most of the cases however $\gamma + \phi_{2\ddagger} > 1$ and Assumption 4.2 is met, so the change in γ should primarily affect the rate of convergence: this was reflected in a slightly higher

dispersion of $\widetilde{\nu}$ when $\gamma = 0.4$. Again, these effects were not affected by the size of δ, so long as the deterministic component dominated in x_t.

Our theory for the GLS estimate does not cover the cases STd, in which $\beta < 1/2$, nor the situation $STa - DTc$, in which the OLS estimate $\widetilde{\nu}$ is less than root-n consistent, so that the estimates of γ, θ based on $y_t - \widetilde{\nu}_+ x_{+t}$ may fail to meet Assumption 4.4.

In all the remaining cases $\widehat{\nu}_F$ was more precise (it had a smaller dispersion) than $\widetilde{\nu}$, and $\widehat{\nu}_I$ was more precise than $\widehat{\nu}_F$. Notice anyway that in all the cases the order of magnitude did not change whether the OLS $\widetilde{\nu}$ or the GLS was used, so the changes in the dispersion, although visible, were never dramatic. A similar ranking of the estimates $\widehat{\nu}_I$, $\widehat{\nu}_F$, $\widetilde{\nu}$, emerged for the biases, but this was weaker because in many situations the bias of $\widetilde{\nu}$ was so little that introducing the GLS could not have any effect anyway. Yet in both the cases in which OLS still presented a visible lower order bias, $STb - DTc$ and $STe - DTc$, the GLS correction took care of it: in the most averse situation, with $\gamma = 0.4$ and $n = 64$, the bias dropped from approximately 0.08 to 0.02; in all the other cases it was completely removed altogether.

Comparing the results of our simulations for $\widehat{\nu}_F$ with $\widehat{\nu}_I$ with RH, notice that the pattern is basically the same as for $\widetilde{\nu}$: much more precise estimates when $\phi_{2\ddagger} > \delta$ and Assumption 4.2 is met, slightly less precise estimates otherwise.

Although our treatment of the GLS estimate required both Assumption 4.4 and $\beta > 1/2$, it is interesting to see what happens when these conditions are not met. Assumption 4.4 is not met under $STa - DTc$ and $STd - DTc$: in both the cases the GLS correction did not succeed in removing the bias completely, although it still provided a sensible improvement on the original OLS especially under STa. Under STd, on the other hand, $\beta > 1/2$ is not met, but notice that, for DTa, DTb and DTd, $\phi_{2\ddagger} - \gamma > 1/2$ is met, and indeed we still did

not observe any relevant bias nor an unusually large standard deviation.

In Tables 4.11 and 4.12 we present empirical sizes, for nominal sizes α =0.05 and α =0.10 respectively, of $\widehat{\nu}_I$ and $\widehat{\nu}_F$. As in RH, those for $\widehat{\nu}_I$ were fairly precise, but those for $\widehat{\nu}_F$ tended to be too big, though the difference got smaller the larger the sample was. The difference in rates of convergence did not seem to affect the pattern, which was rather stable among all combinations.

When Assumption 4.4 is not met, the approximation of Theorem 4.4 does not hold for the GLS estimate, thus duplicating the same outcome of RH; on the other hand if only $\beta > 1/2$ is not met but $\phi_{2\ddagger} - \gamma > 1/2$ is, then in our simulations the limit χ^2 approximation still held.

In the second part of the simulation exercise, we considered the effects of alternative specifications of the matrix Ω.

Since we had $E(u_{2t}) = 1$, altering τ affected the dispersion of u_{1t}, thus inverting the design of RH. In fact, due to the dominating deterministic component in x_t, changing the variance of u_{2t} would not affect the results in a relevant way: we already remarked that even increasing δ from 0.6 to 1.2 (and to 2 under DTd) for given γ had no major effect on variability, and this would be much more the case if δ was left unchanged.

We set $\delta = 0.6$, $\gamma = 0$, and ran the simulations for $\rho = 0$, 0.5, 0.75, -0.5, $\tau = 0.5$, 1 and 2 and all combinations of the deterministic components: the values of δ and γ are deliberately little, because we intended to investigate the effect of the lower order distortion induced by the simultaneous equation bias.

We present bias in Tables 4.13 and 4.14, standard deviation in Tables 4.15 and 4.16 and empirical size in Tables 4.17 and 4.18.

To facilitate readability we excluded DTc because Assumption 4.4 was not met and $\rho = 0.5$ because the results were not much different than for $\rho = 0.75$. In Tables 4.13, 4.14, 4.17 and 4.18 we also excluded results for $\tau = 2$ and $\tau = 0.5$ because, as RH also found in their simulation, they did not vary

much, while in Tables 4.15 and 4.16 we excluded DTb and DTd because the deterministic trend was so strong that the standard deviation was too small to indicate any effect.

The correlation ρ is a potential source of bias in OLS when $\rho \neq 0$, but the effect, presented in Tables 4.13 and 4.14, was minimal, mainly visible in small samples, and almost only for DTa, which had the lowest rate of convergence. In the simulations of RH a small fraction of the bias of $\widetilde{\nu}$ passed to $\widehat{\nu}_F$, but here the preliminary estimate of ν was so precise, due to the faster rate of convergence, that basically no bias was incurred already at the first stage. Altering τ only affected dispersion: in Tables 4.15 and 4.16 the sample variance increased with τ for given ρ and DT. The low impact of changes of ρ on precision was also important because it left the empirical size nearly unaffected: it is not surprising that on average the best approximation to the nominal size is for $\rho = 0$, but even with the rather extreme $\rho = 0.75$ the effect on empirical size was often much less than 0.01. We again notice, for comparison, that changes in ρ in the same situation in RH had a stronger effect, and we explain it by the larger lower order bias in that case.

4.4 Empirical analysis of the PPP hypothesis

We applied our methodology to analyse the Purchasing Power Parity (PPP) hypothesis in three US cities: New York, Boston and Philadelphia. The PPP hypothesis indicates that arbitrage should induce prices of the same items to be the same in different places, and to react together to shocks affecting one of the two: if y_t is the local price and x_t is the price of the same good elsewhere converted to the local currency (both prices being in logarithms), then $\mu_{11} = 0$, $\nu = 1$ and $\gamma < 1$ in (4.5) and (4.6). Strictly speaking, cointegration is not prescribed, but in practice this is necessarily the case because it is gen-

erally acknowledged that $\delta \geq 1$ (a lower level of δ would imply antipersistence of inflation); empirical experience of positive inflation rates also suggests a deterministic component of order about $\phi_{2\ddagger} = 3/2$, i.e. a linear trend.

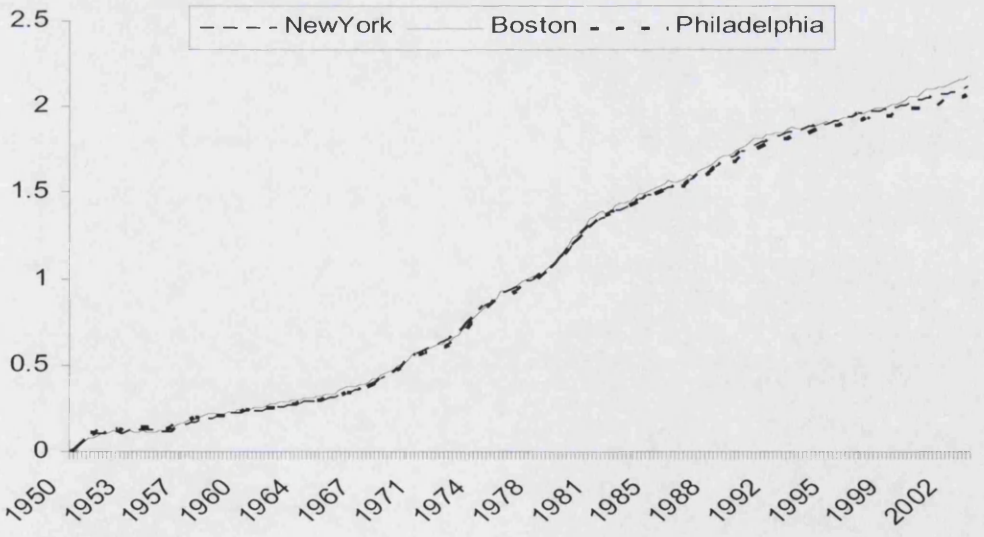
The intuition behind PPP is easy to grasp and in line with common sense. The implications for both economic theory and policy advice are important, because implicitly it can also be interpreted as indicating market integration, so it is not surprising that it has been widely analysed in the applied economic literature. The empirical failure of the PPP hypothesis, at least as a short run phenomenon, is well documented: it can be due to the use of price indices, rather than effective prices, their differences reflecting the difference in preferences of economic agents. In addition, the arbitrage effect may be reduced and delayed by the cost of actively searching on the neighbour market and of eventually shipping the good to the home market. The latter argument also suggests that some deviation should be allowed, at least in the short run, and the PPP model then quickly became a classical case-study for cointegration. Corbae and Ouliaris (1988) and Johansen and Juselius (1992) assumed $(\gamma, \delta) = (0, 1)$. They followed two different approaches: Corbae and Ouliaris tested for a unit root with a Dickey-Fuller statistic on the difference $y_t - x_t$, while Johansen and Juselius first estimated the potential cointegrating vector and then tested the restriction. In both the cases the joint restrictions $\nu = 1$, $\gamma = 0$ were rejected. The PPP hypothesis largely remains a puzzle in cointegration analysis, evidence being still dubious.

The restriction on γ imposed by the $(\gamma, \delta) = (0, 1)$ paradigm is stronger than economic theory implies, disregarding a wide class of mean- (or conditional-mean-) reverting processes that are indexed by a different γ. Allowing for $0 < \gamma < 1$, semiparametric fractional cointegration analysis of PPP was carried out by Cheung and Lai (1993), who estimated ν by OLS and γ by log-periodogram regression. They discussed the existence of a cointegrating rela-

tion but did not test $\nu = 1$. The GLS methodology presented in the previous sections on the other hand allows for a more efficient estimation and testing of ν, but the trend must then be taken explicitly into account.

We employed the same data that we used in Chapter 2: logged quarterly price indexes of Boston, New York and Philadelphia for 1950 (1) through 2003 (3). This time $n = 215$ in the (OLS or GLS) estimation of the cointegrating parameter because we did not have to take first differences. Data were then normalised, dividing each series by the first observation, and logarithms were finally taken. The normalisation was introduced to make the series visually comparable: as seen in Figure 4.1, they are dominated by a long-term deterministic component.

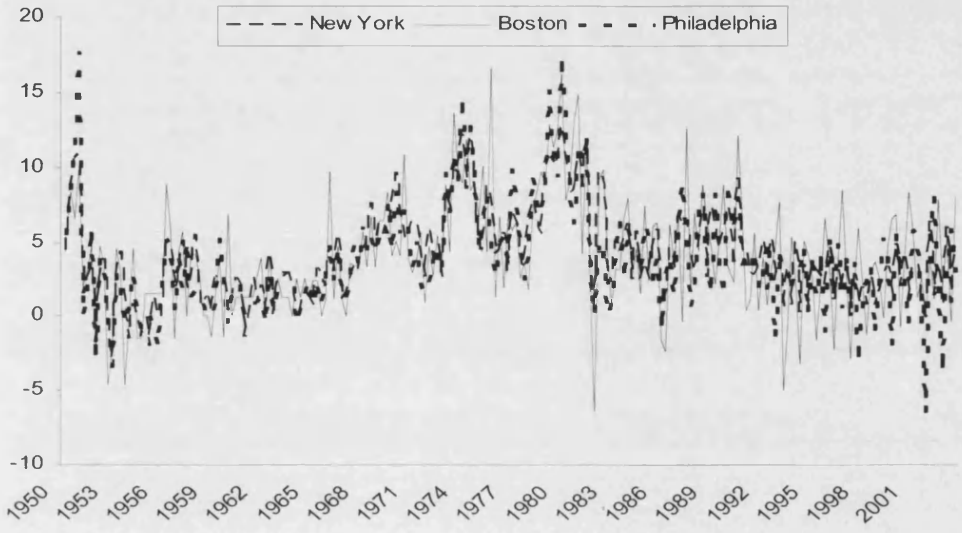
Figure 4.1: Level of prices (logs): BY, Bo, Ph



Since the three paths cannot be clearly distinguished, we also present, in Figure 4.2, annualised first differences, corresponding to the inflation rate (this is the same as in Figure 2.4). First differencing transforms the linear time trend in a constant, so any long run co-movement in the data is due to a common stochastic trend: visual inspection of the three plots seems to suggest that such

a common stochastic trend is indeed present and it is the main force driving the long term dynamics.

Figure 4.2: Annualised quarterly inflation: BY, Bo, Ph



Our assumption is that long term inflation remained stable over the whole period: phases of high inflation, such as those observed after the two oil shocks, are still possible, but on adopting a very long perspective they seem to be temporary (albeit still persistent) phenomena rather than substantial, non-reverting breaks. As Diebold and Inoue (2001) showed, allowing for random, occasional breaks, and a long enough time perspective, long memory can be a convenient way to describe the data (notice, again, the difference with respect to the analysis we did in Chapter 2, in which the number of breaks is fixed).

Finally, notice that we analysed cointegration both parametrically and semiparametrically, in the latter case estimating $\overline{\nu}$ using non-intercept OLS and $\widetilde{\delta}$, $\widetilde{\gamma}$ using local Whittle estimation as in Robinson (1995b). Although in Section 4.2 we only used deterministic trends with constant coefficients, we mainly did it to keep the limit distributions simple. It is fair anyway to conjecture that the orders of magnitude would not change if a break in μ_{21}

was introduced (nor indeed in μ_{11} but the case is not relevant here). Since we already showed in Chapter 2 that for these data local Whittle estimation of δ is robust to a change in the mean, we think that semiparametric analysis is also important because it provides results that are quite robust to changes in the long term growth rate of prices.

We analysed cointegration pairwise, denoting the three cases Bo-NY, Ph-NY and Ph-Bo for New York and Boston, New York and Philadelphia, and Boston and Philadelphia, respectively. Of course if two pairs are both cointegrated the third one will be too and it is then redundant, but considering all three is sensible, especially in the preliminary phase of semiparametric analysis, where the estimates are robust to model misspecification but inefficient. The distinction between explanatory and dependent variables has no econometric implications in our framework, but we took New York as x_t throughout, and Boston as x_t in the Ph-Bo model. The nominal size for the tests is set at 5%.

We first tested the usual $\gamma = 0$, $\delta = 1$ framework. Though inflation is sometimes modelled as an $I(1)$ process, it is usually taken to be $I(0)$, implying $\delta = 1$, as this is consistent with a monetary policy model in which the central bank aims to stabilise the growth rate of prices in the medium-run, and as it is often supported by empirical tests. We tested $\delta = 1$ by the augmented Dickey-Fuller test, allowing for a constant in the AR model in levels (thus a linear trend in (4.6)), with an AR(4) selected by sequential testing. We applied the same procedure (without an intercept) to the first differences with $\delta = 2$ in mind but overall our results supported $\delta = 1$. The vector AR for the Johansen procedure was determined by inspecting the sequential likelihood ratio test (LR), the Schwarz (SC) and the Hannan and Quinn (HQ) statistics, these pointing in general to 5 lags. Using the procedure of Johansen (1991)

cointegration was rejected in all the three models.

Evidence against PPP is often interpreted as indication of a lack of integration between markets, and cannot be explained by trade or cultural barriers, regulations or exchange rate instability in the present case. Since naive inspection of the data is suggestive of cointegration we thus investigate whether this can emerge in a fractional framework. In particular, we first examine the existence of a cointegrating relationship, in such a way as to avoid the consequences of misspecification of high frequency behaviour. The robust, but inefficient, estimates of γ, δ, ν that are involved will also provide a reference with the more efficient ones subsequently obtained. The results are presented in Table 4.4.

We estimated δ_x and δ_y, the orders of integration of x_t and of y_t, by means of the local Whittle estimates, $\widetilde{\delta}_x$ and $\widetilde{\delta}_y$, as described in Robinson (1995b) and in (1.53) but, in view of the anticipated nonstationarity we applied the method to first differences, then adding back 1. Any deterministic linear trend was thereby removed too. The bandwidth was $m = 0.24n^{4/5} = 16$, which is approximately MSE - optimal when the process is a ARFIMA(1,δ,0) with AR coefficient 0.5 as we already discussed in Chapter 2. The estimates of δ were 1.53 (NY), 1.45 (Bo) and 1.40 (Ph). We then tested for the equality of the orders of integration using the statistics of Robinson and Yajima (2001) $\widetilde{T}_{yx}$, as in (1.88). The hypothesis $\delta_y = \delta_x$ was rejected at 5% for Ph-NY though the statistic depends on a trimming number, and if this is not large enough the rejection could be due to the presence of cointegration. In the other two cases equality of the orders was not rejected. We then computed the Hausman-type statistics $\widetilde{H}_y$ and $\widetilde{H}_x$ for no-cointegration of Marinucci and Robinson (2001), described here in (1.75): we rejected the hypothesis of no cointegration in case of Ph-NY and of Ph-Bo but not for Bo-NY. This seeming inconsistency could reflect a type two error: in Marinucci and Robinson's Monte Carlo experiment

the power of a comparable case was only about 50%. Note also that for Bo-NY the memory parameter estimate $\widetilde{\delta}_*$ used in the test did not lie between the individual estimates.

For each of the two pairs, $\overline{\nu}$ was computed and in each case found to be very close to 1. No-intercept OLS accords with economic theory, and the preliminary Johansen analysis suggested that the level of persistence γ, even if less than δ, may be between between 1/2 and 1. The possibility of such a high γ would suggest estimating it from first difference of the OLS residuals $\overline{v}_{1t} = y_t - \overline{\nu} x_t$. On the basis of $\widetilde{\gamma}$, the local Whittle estimate of γ, (4.20) is satisfied in each case but only barely for the combination Bo-NY; point estimates were also below 1 for Ph-NY and Ph-Bo, again in accordance with economic theory, while it was just above 1 for Bo-NY.

Table 4.4: PPP: semiparametric analysis of fractional cointegration

	$\widetilde{\delta}_y$	$\widetilde{\delta}_x$	$\widetilde{\delta}_*$	$\left(\widetilde{T}_{yx}\right)^2$	$\left(\widetilde{H}_y\right)^2$	$\left(\widetilde{H}_x\right)^2$	$\overline{\nu}$	$\widetilde{\gamma}$
Ph-NY	1.40	1.53	1.35	4.83	0.26	4.00	0.99	0.82
Bo-NY	1.45	1.53	1.38	1.28	2.83	0.68	1.02	1.05
Ph-Bo	1.40	1.45	1.21	0.67	4.53	7.66	0.97	0.83

Note: critical values for χ^2_1: 3.84 (5%) and 6.63 (1%).

Our semiparametric analysis suggested that the PPP model could be reasonable for the three cities in fractional context, the rejection of the $(\gamma, \delta) = (0, 1)$ version being due to the high persistence of deviations from the long-run relation. We then proceeded to a parametric analysis along the lines described in Section 4.3 with the aim of ultimately testing $\nu = 1$.

We consider the Ph-NY and Bo-NY relations only, the results appearing in Table 4.5. In (4.18) and (4.19) we took $g_1(t) = 0$, $g_2(t) = t$. Using the previously obtained estimates of γ, δ (pooled), ν, and denoting by $\widetilde{\mu}_{21}$ the

estimate of μ_{21} obtained by OLS regression of x_t on t, we formed the residuals $\widetilde{v}_{1t} = y_t - x_t\overline{\nu}$, $\widetilde{v}_{2t} = x_t - \widetilde{\mu}_{21}t$, and then took $\widetilde{\gamma}$ and $\widetilde{\delta}$ fractional differences respectively, labelling them as $\overline{u}_{1t} = \widetilde{v}_{1t}(\widetilde{\gamma})$, $\overline{u}_{2t} = \widetilde{v}_{2t}(\widetilde{\delta})$. We then determined the structure $B(L;h)$ using the LR, SC and HQ procedures on the series $\overline{u}_{1t}, \overline{u}_{2t}$, concluding in favour of a vector AR(4). Since the coefficients of the second and third lags were small, we took them to be zero, representing the short term dynamics by, in effect, an AR(1) combined with a quarterly effect.

Now define, for given c, d, $\widetilde{u}_t(c,d) = (\widetilde{v}_{1t}(c), \widetilde{v}_{2t}(d))'$ and $\widetilde{\varepsilon}_t(c,d) = B(L;\widetilde{\theta}(c,d))\,\widetilde{u}_t(c,d)$, where $\widetilde{\theta}(c,d)$ indicates the OLS estimates of the AR coefficients for given c, d, dropping $\widetilde{u}_1(c,d)$ because this term is not differenced at all. We then took $(\widehat{\gamma},\widehat{\delta}) = \arg\min n^{-1}\left|\sum_{t=1}^{n}\widetilde{\varepsilon}_t(c,d)\widetilde{\varepsilon}_t'(c,d)\right|$, and $\widehat{\theta} = \widetilde{\theta}(\widehat{\gamma},\widehat{\delta})$, $\widehat{\Omega} = n^{-1}\sum_{t=1}^{n}\widetilde{\varepsilon}_t(\widehat{\gamma},\widehat{\delta})\widetilde{\varepsilon}_t'(\widehat{\gamma},\widehat{\delta})$.

Table 4.5: PPP: ML estimates of the long term parameters

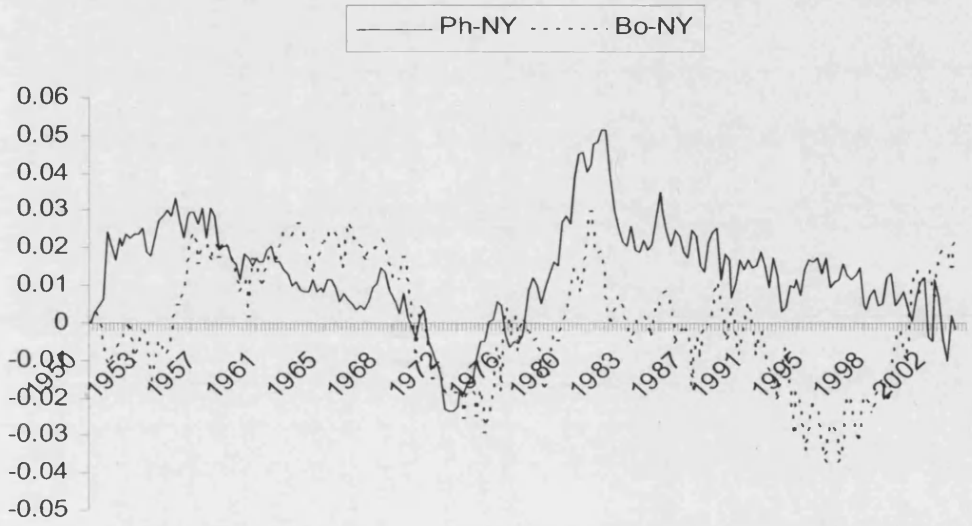
	$\widehat{\delta}$	$\widehat{\gamma}$	$\widehat{\nu}$	$\widehat{\mu}_{21}$
Ph-NY	1.42	0.66	0.98	0.0045
Bo-NY	1.43	0.69	1.02	0.0041

Note: estimates from a model with AR(4) structure for the lags but in which the second and third lag are excluded.

Hypothesis on γ, δ, θ can be tested with a likelihood ratio statistic. These, and the hypothesis on ν, are discussed below and summarised in Table 4.6. The parametric analysis confirmed the restrictions on the AR(4) model for u_t, and $\widehat{\gamma}$ and $\widehat{\delta}$ were similar to the semiparametric estimates. We strongly rejected the joint hypothesis that $\delta = 1, \gamma = 0$. On the other hand the hypothesis $\gamma = 1$ was rejected against the alternative $\gamma \neq 1$. We then applied the time domain GLS procedure to estimate ν: the estimates $\widehat{\nu}$ were in both cases close to 1, and the null hypothesis was not rejected.

In Figures 4.3 and 4.4 we plotted the GLS residuals $y_t - \widehat{\nu} x_t$ (where $\widehat{\nu}$ is the GLS estimate of ν) and the restricted residuals $y_t - x_t$ (assuming $\nu = 1$), respectively. In both cases the residual series only occasionally crossed 0 (the value that we assumed to be the mean the disturbances).

Figure 4.3: unconstrained GLS residuals



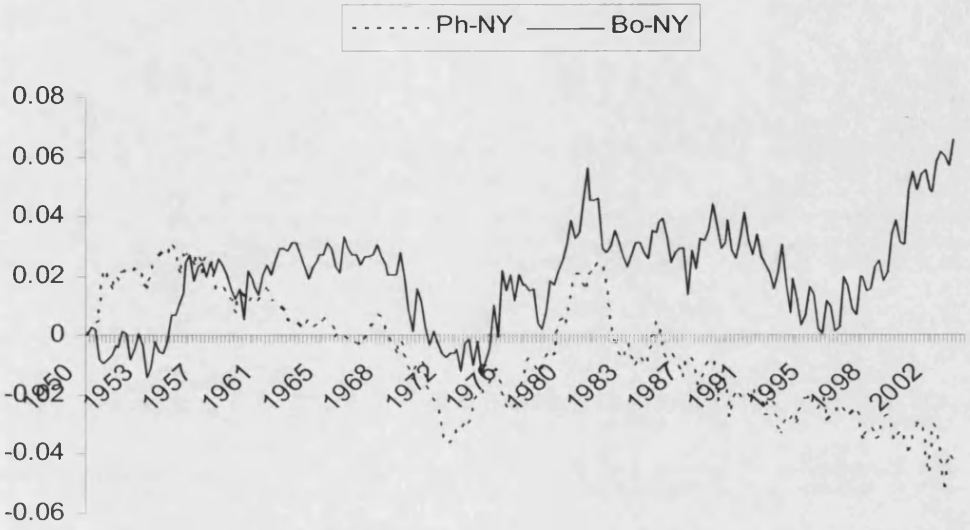
The patterns of GLS and restricted residuals are very similar, but autocorrelation appeared to be a little stronger for the latter: iterating the parametric procedure, the estimates $\widehat{\gamma}$ using GLS residuals $y_t - \widehat{\nu} x_t$ were 0.65 for the pair Ph-NY and 0.69 for the pair Bo-NY, while using the restricted residuals they were both 0.73.

Table 4.6: PPP: LR tests

	LR_B	$LR_{\{\delta=1,\gamma=0\}}$	$LR_{\{\gamma=1\}}$	$LR_{\{\nu=1\}}$
Ph-NY	9.69 [0.287]	21.50 [0.000]	9.83 [0.002]	3.74 [0.053]
Bo-NY	4.51 [0.808]	35.87 [0.000]	9.85 [0.002]	2.21 [0.137]

Note: LR_B is the likelihood ratio statistic for the hypothesis that the second and third lag are excluded; $LR_{\{\delta=1,\gamma=0\}}$, $LR_{\{\gamma=1\}}$, $LR_{\{\nu=1\}}$ are the likelihood ratio statistics for the hypotheses $H_0 : \{\delta = 1, \gamma = 0\}$, $H_0 : \{\gamma = 1\}$, $H_0 : \{\nu = 1\}$; P-values are in brackets.

Figure 4.4: residuals of the restricted model



As a final small exercise we investigated the role of the deterministic trend in increasing precision in the estimation of ν, bearing in mind the findings of the Monte Carlo exercise. We thus also computed the GLS estimate of RH, using detrended data.

The factors $((1,0)(b^{**})^{-1}(1,0)')^{1/2}$ (see Theorem 4.4) were 0.0107 in case of Ph-NY, and 0.0151 in case of Bo-NY, whereas the $(b^{**})^{-1/2}$ defined from RH for the detrended data were more than twice as big, being 0.0304 and 0.0415 respectively.

4.5 Discussion

We have studied the estimation of a cointegrating parameter in a bivariate process when the data may have been contaminated by a deterministic trend. We have discussed no-intercept OLS, finding that in some cases it may be subject to omitted-variable bias, while in others it may even be more efficient than if the data did not have deterministic components at all: although the final outcome then depends on the precise situation, we have noticed that in

the most popular models the rate of convergence is improved by the presence of the deterministic trends. Afterwards, we discussed OLS and infeasible and feasible GLS when the deterministic trend is correctly taken into account, finding that the estimates are always consistent and in some cases the rate of convergence is better than in the situation without deterministic trends, while in the remaining ones it is just as good.

We conclude by discussing a few arguments that can be related to the problem and some potential extensions.

1. We focused our semiparametric analysis on OLS estimates motivating the approach on the ground of simplicity. Yet we also saw that in some cases the rate of convergence of the estimate may be too slow to allow for a root-n consistent estimation of the remaining parameters of the model (θ) in the GLS procedure. RM showed for the model without trends that when $\gamma + \delta < 1$ a faster rate of convergence may be achieved estimating the cointegrating parameter by a NBLS procedure: it may then be worth exploring this opportunity when the semiparametric analysis is preliminary to a GLS estimation rather than a quick way to get a first glance at the relation between the two time series.

2. We only focused on $\beta > 1/2$. Yet when deterministic components are present that may be too strong a condition, and the milder $\max\{\delta, \phi_{2\diamond}\} - \gamma > 1/2$ should be discussed. Our Monte Carlo exercise seems to confirm that this extension is feasible.

3. We could generalise the results by allowing for a wider class of deterministic terms: the trends that we consider do anyway provide an interesting benchmark, and in many situations more complicated structures (like trends subject to breaks) would grossly generate the same type of

predictions while making the proofs and the statement of the theorems much less clear.

4. We only considered a scalar x_t. Like RH, we did not consider a larger model because an interesting treatment of this case would entail, in our fractional setting, allowance for differing integration orders in the elements of x_t, and also two or more cointegrating relationships, possibly with different integration orders. Yet an additional comment should be still given when deterministic terms are present in a multivariate x_t. Consider the model

$$y_t = \nu_2 x_{2t} + \nu_3 x_{3t} + \mu_1 t^{\phi_1 - 1/2} + u_{1t}(-\gamma), \tag{4.120}$$

$$x_{2t} = \mu_2 t^{\phi_2 - 1/2} + u_{2t}(-\delta_2), \tag{4.121}$$

$$x_{3t} = \mu_3 t^{\phi_3 - 1/2} + u_{3t}(-\delta_3), \tag{4.122}$$

with $\mu_1 \neq 0$, $\mu_2 \neq 0$, $\mu_3 \neq 0$, $\phi_1 > \gamma$, $\phi_2 > \delta_2$, $\phi_3 > \delta_3$, $\delta_2 \geq \delta_3$. Here Assumption 4.2 is much more restrictive, because it requires not only

$$\phi_2 \neq \phi_1 \text{ and } \phi_3 \neq \phi_1, \tag{4.123}$$

but also

$$\phi_2 \neq \phi_3, \tag{4.124}$$

otherwise the problem of multicollinearity among regressors remains.

If, in fact, (4.123) is met but (4.124) is not, as it is when $\phi_2 = \phi_3 = \phi$,

then the regression model becomes

$$\begin{aligned} y_t &= \nu_2 \left(x_{2t} \pm \mu_2/\mu_3 x_{3t} \right) + \nu_3 x_{3t} + \mu_1 t^{\phi_1 - 1/2} + v_t, \\ y_t &= \nu_2 \left(u_{2t} \left(-\delta_2 \right) - \mu_2/\mu_3 u_{3t} \left(-\delta_3 \right) \right) + \left(\nu_3 + \nu_2 \mu_2/\mu_3 \right) x_{3t} + \mu_1 t^{\phi_1 - 1/2} + v_t, \end{aligned} \tag{4.125}$$

so, letting $\nu_4 = \nu_3 + \nu_2 \mu_2/\mu_3$ and indicating by $\widehat{\nu}_2$, $\widehat{\nu}_3$ $\widehat{\nu}_4$ the GLS estimates of ν_2, ν_3 and ν_4, (and assuming proper generalisation of the regularity conditions, so that a multivariate version of Theorem 4.3 is possible) then $\widehat{\nu}_2 = \nu_2 + O_e\left(n^{\gamma - \delta_2}\right)$ but for the linear combination ν_4 we have $\widehat{\nu}_4 = \nu_4 + O_e\left(n^{\gamma - \phi}\right)$, which is then faster. Also notice that with a similar argument we may find that $\widehat{\nu}_3 = \nu_3 + O_e\left(n^{\gamma - \delta_2}\right)$, therefore faster than the rate of convergence obtained when there is no trend at all. This still has an intuitive explanation, because $t^{\phi - 1/2}$ conveys information about x_{2t} and x_{3t} simultaneously and it cannot, then, be used to estimate ν_2 or ν_3 separately, but only for the linear combination $(\nu_2 \mu_2 + \nu_3 \mu_3)$.

Assumption 4.2 is then a condition that becomes stronger the larger the dimension of x_t, and if we only consider linear trends it is not met when x_t is not scalar.

4.6 Appendix to Chapter 4

Proof of Theorem 4.1 Define $g(t) = \left(t^{\phi_1 - \frac{1}{2}}, ..., t^{\phi_r - \frac{1}{2}}\right)'$, $D_{gn} = diag\left\{n^{\phi_1}, ..., n^{\phi_r}\right\}$, for $0 < \phi_1 < ... < \phi_r$. For $d \geq 0$ we have

$$n^{-d}\left(D_{gn}^{-1} \otimes I_2\right) \sum_{t=1}^{n} g(t) \otimes u_t(-d) \rightarrow_d \int_0^1 g(r) \otimes d\left(\overline{W}(r; d+1), \widetilde{W}(r; d+1)\right)', \tag{4.126}$$

where $\otimes$ denotes Kronecker product. For $d > \frac{1}{2}$ (4.126) follows from Theorem 1 of Marinucci and Robinson (2000) (hereafter MR), and the continuous mapping theorem. For $0 \leq d \leq \frac{1}{2}$ (4.126) follows from a central limit theorem; note that the right side of (4.126) is in any case a $2r-$dimensional normal variate. We have

$$x_t = u_{2t}(-\delta) + s_{xt}, \tag{4.127}$$

where

$$s_{xt} = \sum_{j=1}^{p_2} \mu_{2j} t^{\phi_{2j} - \frac{1}{2}}. \tag{4.128}$$

From (4.126),

$$\begin{aligned} n^{-\phi_{2\ddagger} - \delta} \sum_{t=1}^{n} s_{xt} u_{2t}(-\delta) &= n^{-\phi_{2\ddagger} - \delta} \mu_{2\ddagger} \sum_{t=1}^{n} t^{\phi_{2\ddagger} - \frac{1}{2}} u_{2t}(-\delta) + o_p(1) \\ &\rightarrow_d \Phi_2. \end{aligned} \tag{4.129}$$

By Theorem 1 of MR and the continuous mapping theorem

$$n^{-2\delta} \sum_{t=1}^{n} u_{2t}^2(-\delta) \rightarrow_d \Phi_1, \tag{4.130}$$

and by integral approximation

$$\begin{aligned} n^{-2\phi_{2\dagger}} \sum_{t=1}^{n} s_{xt}^2 &= n^{-2\phi_{2\dagger}} \mu_{2\ddagger}^2 \sum_{t=1}^{n} t^{2\phi_{2\dagger}-1} + o(1) \\ &\to \Phi_3. \end{aligned} \tag{4.131}$$

Thus

$$k_n^{-2} M_{xx} \to_d \Phi_1^{**} + 2\Phi_2^{**} + \Phi_3^{**}. \tag{4.132}$$

Next, the numerator of $\overline{\nu} - \nu$ is

$$a = \sum_{t=1}^{n} \left(s_{xt} + u_{2t}(-\delta)\right) \left(s_{yt} + u_{1t}(-\gamma)\right) \tag{4.133}$$

where

$$s_{yt} = \sum_{j=1}^{p_1} \mu_{1j} t^{\phi_{1j} - \frac{1}{2}}. \tag{4.134}$$

Integral approximation gives

$$\begin{aligned} n^{-\phi_{1\dagger}-\phi_{2\ddagger}} \sum_{t=1}^{n} s_{xt} s_{yt} &= n^{-\phi_{1\dagger}-\phi_{2\ddagger}} \mu_{1\dagger} \mu_{2\ddagger} \sum_{t=1}^{n} t^{\phi_{1\dagger}+\phi_{2\ddagger}-1} + o(1) \\ &\to \Psi_6 \end{aligned} \tag{4.135}$$

and (4.126) gives

$$n^{-\gamma-\phi_{2\ddagger}} \sum_{t=1}^{n} s_{xt} u_{1t}(-\gamma) \to_d \Psi_4. \tag{4.136}$$

$$n^{-\delta-\phi_{1\dagger}} \sum_{t=1}^{n} s_{yt} u_{2t}(-\delta) \to_d \Psi_5, \tag{4.137}$$

To deal with $b = \sum_{t=1}^{n} u_{1t}(-\gamma) u_{2t}(-\delta)$, we make use of results of RM. From Lemmas 4.1, 4.2 and 5.1 of RM (see also their Propositions 6.1 and 6.2) we

have

$$n^{-1}b \to_p \Psi_1, \quad \text{if } \gamma+\delta < 1, \tag{4.138}$$

$$\frac{n^{-1}}{\log n}b \to_p \Psi_2, \quad \text{if } \gamma+\delta = 1,\ \gamma > 0. \tag{4.139}$$

From Lemmas 4.3, 4.5 and 5.1 of RM, $b = O_p(n^{\delta})$ for $\gamma = 0$, $\delta \geq 1$, but since $\phi_{1\dagger} > 0$ it follows that in this case $a = o_p\left(n^{\delta+\phi_{1\ddagger}}\right)$ so Ψ_5 will dominate. Finally

$$n^{-\gamma-\delta}b \to_d \Psi_3, \quad \gamma+\delta > 1,\ \gamma > 0, \tag{4.140}$$

where Theorem 1 of MR and the continuous mapping theorem covers the case $\gamma > \frac{1}{2}$, and Lemmas 4.5 and 5.1 of MR the case $\gamma \leq \frac{1}{2}$ (RM discuss the problem of representation of Ψ_3 in this case). It follows that

$$\ell_n^{-1}a \to_d \Psi_1^{**} + \Psi_2^{**} + \Psi_3^{**} + \Psi_4^{**} + \Psi_5^{**} + \Psi_6^{**}, \tag{4.141}$$

noting that in case $\gamma + \delta = 1$, $\gamma > 0$, $n\log n$ dominates $n = n^{\gamma+\delta}$, and dominates $n^{\gamma+\phi_{2\ddagger}}$, $n^{\delta+\phi_{1\dagger}}$, $n^{\phi_{1\dagger}+\phi_{2\ddagger}}$ if and only if $\delta \geq \phi_{2\ddagger}$, $\gamma \geq \phi_{1\dagger}$, $\phi_{1\dagger}+\phi_{2\ddagger} \leq 1$ respectively.

Proof of Theorem 4.2 The proof that

$$B_n^{-1}M_{x_+x_+}B_n^{-1} \to_d \Phi \tag{4.142}$$

straightforwardly uses results employed in showing (4.126), and is omitted. We are left to consider

$$M_{x_+v} = \sum_{t=1}^{n} x_{+t}\left\{u_{1t}(-\gamma) + s_{1t}\right\} = \sum_{t=1}^{n}\left(u_{2t}(-\delta) + s_{xt}, g_1'(t)\right)'\left\{u_{1t}(-\gamma) + s_{1t}\right\}. \tag{4.143}$$

By again employing results from the proof of Theorem 4.1,

$$D_{1n}^{-1}(\gamma)\sum_{t=1}^{n} g_1(t)u_{1t}(-\gamma) \rightarrow_d \Psi_8 \tag{4.144}$$

while

$$D_{1n}^{-1}(\gamma)\sum_{t=1}^{n} g_1(t)s_{1t} \rightarrow 0 \tag{4.145}$$

because, for $j > p_{11}$ and $i+1...p_{11}$

$$n^{-\gamma-\phi_{1i}}\sum_{t=1}^{n} t^{\phi_{1i}+\phi_{1j}} = O\left(n^{\phi_{1j}-\gamma}\right) = o(1) \tag{4.146}$$

because $\gamma > \phi_{1j}$ for $j > p_{11}$. Next,

$$\sum_{t=1}^{n} s_{xt}s_{1t} = O\left(n^{\phi_{2\ddagger}+\phi_{1,p_{11}+1}}\right) = o\left(n^{\phi_{2\ddagger}+\gamma}\right). \tag{4.147}$$

Then from the proof of Theorem 4.1,

$$m_n^{-1}\sum_{t=1}^{n}\left\{u_{2t}(-\delta)+s_{xt}\right\}u_{1t}(-\gamma) \rightarrow_d \Psi_1^{***}+\Psi_2^{***}+\Psi_3^{***}+\Psi_4^{***}+\Psi_7^{***} \tag{4.148}$$

with Ψ_7^{***} coming from MR Theorem 1 and Kurtz and Protter (1991), see also MR Propositions 6.3 and 6.4 (unlike in Theorem 4.1, this contribution is not always dominated).

Further, we may obtain the joint result

$$C_n^{-1}B_n^{-1}M_{x+v} \rightarrow_d \Psi. \tag{4.149}$$

From the commutativity properties of diagonal matrices

$$\widetilde{\nu}_+ - \nu = B_n^{-1}\left(B_n^{-1}M_{x_+x_+}B_n^{-1}\right)^{-1}C_n\left(C_n^{-1}B_n^{-1}M_{x+v}\right) \tag{4.150}$$

so the proof is completed by application of (4.142) and (4.150).

Proof of Theorem 4.3 Consider the case of white noise u_t. From (4.5), (4.6), (4.87), (4.94) and (4.95) we can write

$$\widehat{\nu}_{++}(\gamma,\delta,\theta) - \nu = b\,(\gamma,\delta,\theta)^{-1}\,\{e(\gamma,\delta,\theta) + f(\gamma,\delta,\theta)\}, \tag{4.151}$$

where

$$e(\gamma,\delta,\theta) = \sum_{t=1}^{n} w_t(\gamma,\delta,\theta)\Omega^{-1}u_t, \quad f(\gamma,\delta,\theta) = \sum_{t=1}^{n} w_t(\gamma,\delta,\theta)\Omega^{-1}\widetilde{s}_t, \tag{4.152}$$

where $\widetilde{s}_t = (\widetilde{s}_{1t},\, \widetilde{s}_{2t})' = (\Delta^{\gamma}s_{1t},\, \Delta^{\delta}s_{2t})'$. From Lemma 1 of Robinson (2005a), with $v_t = t^c$,

$$v_t(d) = \frac{\Gamma(c+1)}{\Gamma(c-d+1)}t^{c-d} + O\left(t^{c-m-1}\right), \tag{4.153}$$

where m is the integer such that $d-1 < m \leq d$. Then

$$\begin{aligned}
\sum_{t=1}^{n}\Delta^{\gamma}g_2'(t)\mu_2\widetilde{s}_{1t} &= O\left(\sum_{t=1}^{n} t^{\phi_{2\dagger}-2\gamma+\phi_{1,p_{11}+1}-1}\right) \\
&= O\left(n^{\phi_{2\dagger}+\phi_{1,p_{11}+1}-2\gamma}\right), && (4.154)\\
\sum_{t=1}^{n}\Delta^{\gamma}g_{1j}(t)\widetilde{s}_{1t} &= O\left(\sum_{t=1}^{n} t^{\phi_{1j}-2\gamma+\phi_{1,p_{11}+1}-1}\right) \\
&= O\left(n^{\phi_{1j}+\phi_{1,p_{11}+1}-2\gamma}\right), \quad j=1,...,p_{11}, && (4.155)\\
\sum_{t=1}^{n}\Delta^{\delta}g_{2j}(t)\widetilde{s}_{1t} &= O\left(\sum_{t=1}^{n} t^{\phi_{2j}-\gamma-\delta+\phi_{1,p_{11}+1}-1}\right) \\
&= O\left(n^{\phi_{2j}+\phi_{1,p_{11}+1}-\gamma-\delta}\right), \quad j=1,...,p_{21}, && (4.156)
\end{aligned}$$

and similarly $\sum_{t=1}^{n}\Delta^{\gamma}g_2'(t)\mu_2\widetilde{s}_{2t} = O\left(n^{\phi_{2\dagger}+\phi_{2,p_{21}+1}-\gamma-\delta}\right)$, $\sum_{t=1}^{n}\Delta^{\gamma}g_{1j}(t)\widetilde{s}_{2t} = O\left(n^{\phi_{1j}+\phi_{2,p_{21}+1}-\gamma-\delta}\right)$ for $j = 1...p_{11}$, $\sum_{t=1}^{n}\Delta^{\delta}g_{2j}(t)\widetilde{s}_{2t} = O\left(n^{\phi_{2j}+\phi_{2,p_{21}+1}-2\delta}\right)$ for $j = 1...p_{21}$.

On the other hand, much as in the proof of Theorem 4.1,

$$\sum_{t=1}^{n} u_{2t}(\gamma-\delta)\widetilde{s}_{1t} = O_p\left(n^{\delta-2\gamma+\phi_{1,p_{11}+1}+\frac{1}{2}}\right), \quad \sum_{t=1}^{n} u_{2t}(\gamma-\delta)\widetilde{s}_{2t} = O_p\left(n^{-\gamma+\phi_{2,p_{21}+1}+\frac{1}{2}}\right). \tag{4.157}$$

It straightforwardly follows that

$$D_n^{-1} f(\gamma, \delta, \theta) \to 0. \tag{4.158}$$

From (4.153) and routine arguments

$$n^{\frac{1}{2}} D_{in}^{-1}(-d)\Delta^d g_i([rn]) \to \Lambda_i(d)\frac{g_i(r)}{r^d}, \quad r \in (0,1], \quad d \geq 0, \ i = 1,2. \tag{4.159}$$

From this and MR Theorem 1

$$n^{\frac{1}{2}} D_n^{-1} w_{[rn]}(\gamma, \delta) \Rightarrow Q(r), \quad r \in (0,1]. \tag{4.160}$$

Thus from the continuous mapping theorem and Kurtz and Protter (1991)

$$D_n^{-1} b(\gamma, \delta, \theta) D_n^{-1} \Rightarrow \int_0^1 Q(r)Q'(r)dr, \quad D_n^{-1} e(b, \delta, \theta) \Rightarrow \int_0^1 Q(r)dW(r) \tag{4.161}$$

to complete the proof for $\widehat{\nu}_{++}(\gamma, \delta, \theta)$. The application of Assumption 4.3 to prove the Theorem for the remaining quantities in (4.97), in the white noise u_t case, is straightforward, and thus omitted.

A good deal of the proof detail in RH is concerned with justifying the general short memory autocorrelation in u_t described in Subsection 4.2.3. It is clearly unnecessary to repeat this for our extended estimate, and it suffices merely to consider the implications for the deterministic components we have introduced. These are somewhat different from the treatment of stochastic

trends. Consider the quantity

$$\widetilde{e}(\gamma,\delta,\theta)=\sum_{j=1}^{n}F_{\widetilde{w}(\gamma,\delta)}(\lambda_j)f_u(\lambda_j)^{-1}F_u(-\lambda_j), \tag{4.162}$$

where

$$F_{\widetilde{w}(\gamma,\delta)}(\lambda)=\frac{1}{(2\pi n)^{\frac{1}{2}}}\sum_{t=1}^{n}\widetilde{w}_t(\gamma,\delta)e^{it\lambda},\ F_u(\lambda)=\frac{1}{(2\pi n)^{\frac{1}{2}}}\sum_{t=1}^{n}u_te^{it\lambda}, \tag{4.163}$$

with $\widetilde{w}_t(c,d)$ defined like $w_t(c,d)$ in (4.85) but with $\Delta^c x_t$ replaced by $\Delta^c s_{xt}$. Denote by $\psi_L(\lambda)=\sum_{\ell=-L}^{L}\psi_\ell(1-|\ell|/L)e^{-i\ell\lambda}$ the Cesaro sum, to L terms, of the Fourier series of $f_u(\lambda)^{-1}$. Define $\widetilde{D}_n$ like D_n but with $n^{\max(\delta,\phi_{2\dagger})}$ replaced by $n^{\phi_{2\dagger}}$ and

$$\widehat{e}(\gamma,\delta,\theta)=\sum_{j=1}^{n}F_{\widetilde{w}(\gamma,\delta)}(\lambda_j)\psi_L(\lambda_j)F_u(-\lambda_j). \tag{4.164}$$

Then $\widetilde{D}_n^{-1}\{\widetilde{e}(\gamma,\delta,\theta)-\widehat{e}(\gamma,\delta,\theta)\}$ has mean zero and covariance matrix

$$\frac{1}{2\pi n}\int_{-\pi}^{\pi}\left\{\widetilde{D}_n^{-1}\sum_{j=1}^{n}F_{\widetilde{w}(\gamma,\delta)}(\lambda_j)\left\{f_u(\lambda_j)^{-1}-\psi_L(\lambda_j)\right\}\sum_s e^{i(\lambda_j-\lambda)s}\right\}f_u(\lambda)$$
$$\times\left\{\widetilde{D}_n^{-1}\sum_{j=1}^{n}F_{\widetilde{w}(\gamma,\delta)}(-\lambda_j)\left\{f_u(-\lambda_j)^{-1}-\psi_L(-\lambda_j)\right\}\sum_s e^{i(\lambda-\lambda_j)s}\right\}'d\lambda. \tag{4.165}$$

Using the properties of the complex exponential function, this has norm bounded by a constant times

$$\sum_{j=1}^{n}\left\|\widetilde{D}_n^{-1}F_{\widetilde{w}(\gamma,\delta)}(\lambda_j)\left\{f(\lambda_j)^{-1}-\psi_L(\lambda_j)\right\}\right\|^2\le\varepsilon^2\sum_{j=1}^{n}\left\|\widetilde{D}_n^{-1}F_{\widetilde{w}(\gamma,\delta)}(\lambda_j)\right\|^2 \tag{4.166}$$

for arbitrary $\varepsilon>0$, on choosing L large enough and noting the continuity of

$f_u(\lambda)^{-1}$. For any sequence c_t

$$n^{-1}\sum_{j=1}^{n}\left|\sum_{t=1}^{n}c_t e^{it\lambda_j}\right|^2=\sum_{t=1}^{n}c_t^2, \tag{4.167}$$

so applying again (4.161), it straightforwardly follows that (4.166) $= O(\varepsilon^2)$. Thus

$$\begin{aligned}\widehat{e}(\gamma,\delta,\theta) &= \sum_{\ell=-L}^{L}\sum_{1\le t,t-\ell\le n}\widetilde{w}_t(\gamma,\delta)\psi_\ell\left(1-\frac{|\ell|}{L}\right)u_{t-\ell} &(4.168)\\ &+\sum_{\ell=1}^{L}\sum_{t=1}^{\ell}\widetilde{w}_t(\gamma,\delta)\psi_\ell\left(1-\frac{\ell}{L}\right)u_{t-\ell}, &(4.169)\\ &+\sum_{\ell=-L}^{-1}\sum_{t=n-\ell+1}^{n}\widetilde{w}_t(\gamma,\delta)\psi_\ell\left(1+\frac{\ell}{L}\right)u_{t-\ell}. &(4.170)\end{aligned}$$

The sums over t in (4.169) and (4.170) include only ℓ terms, and, with L fixed and $n\to\infty$, will turn out to be dominated by (4.168). To deal with this, note from boundedness of f that for $c>0$, $d>0$ and any j,

$$\sum_{t=1}^{n}\left\{\Delta^d t^c-\frac{\Gamma(c+1)t^{c-d}}{\Gamma(c-d+1)}\right\}u_{t-j} \tag{4.171}$$

has mean zero and variance bounded by a constant times

$$\sum_{t=1}^{n}\left\{\Delta^d t^c-\frac{\Gamma(c+1)t^{c-d}}{\Gamma(c-d+1)}\right\}^2=O\left(\sum_{t=1}^{n}t^{2(c-m-1)}\right) \tag{4.172}$$

where m is as described after (4.153). Then (4.172) is $O(1)$ for $2c-2m<1$, $O(\log n)$ for $2c-2m=1$ and $O\left(n^{2c-2m+1}\right)$ for $2c-2m>1$. It follows that we may replace $\widetilde{D}_n^{-1}\widehat{e}(\gamma,\delta,\theta)$ by

$$\widetilde{D}_n^{-1}\sum_{\ell=-L}^{L}\sum_{1\le t,t-\ell\le n}\widehat{w}_t(\gamma,\delta)\psi_\ell\left(1-\frac{|\ell|}{L}\right)u_{t-\ell}, \tag{4.173}$$

where

$$\widehat{w}_t(\gamma,\delta) = \begin{bmatrix} \mu_{2\ddagger}\frac{\Gamma(\phi_{2\ddagger}+\frac{1}{2})}{\Gamma(\phi_{2\ddagger}-\gamma+\frac{1}{2})}t^{\phi_{2\ddagger}-\gamma-\frac{1}{2}} & \{\Lambda_1(\gamma)g_1(t)\}'t^{-\gamma} & 0 \\ 0 & 0 & \{\Lambda_2(\delta)g_2(t)\}'t^{-\delta} \end{bmatrix}', \tag{4.174}$$

the above arguments indicating that other contributions from the top left hand element of $\widetilde{w}_t(\gamma,\delta)$ can be neglected. The asymptotic normality, for fixed L, of (4.173) follows standardly (see e.g. Hannan, 1970, Chapter 7), whence Bernstein's lemma, with $L \to \infty$, completes the central limit theorem for $\widetilde{D}_n^{-1}\widetilde{e}(\gamma,\delta,\theta)$. The proof that we can neglect contributions from s_{1t} and s_{2t} follows much as above, as does, using also RH, the limiting behaviour of $D_n^{-1}b(\gamma,\delta,\theta)D_n^{-1}$. The proof that we can replace γ,δ,θ by $\widehat{\gamma},\widehat{\delta},\widehat{\theta}$ is lengthy but relies basically on RH and standard arguments to cope with the deterministic components. As in RH, the proof for the 'time - domain' estimates is similar but slightly simpler, and is omitted.

Table 4.7: Monte Carlo bias for OLS, $\tau = 1$, $\rho = 0.5$

		$n = 64$		$n = 128$		$n = 256$	
ST	DT	$\overline{\nu}$	$\widetilde{\nu}$	$\overline{\nu}$	$\widetilde{\nu}$	$\overline{\nu}$	$\widetilde{\nu}$
	a	.0235	.0009	.0117	.0003	.0059	.0001
a	b	.0004	.0000	.0001	.0000	.0000	.0000
	c	1.0232	.2980	1.0112	.2515	1.0057	.2076
	d	.0194	.0000	.0097	.0000	.0049	.0000
	a	.0247	.0003	.0121	.0000	.0060	.0000
b	b	.0004	.0000	.0001	.0000	.0000	.0000
	c	1.0912	.0176	1.0505	.0048	1.0347	.0006
	d	.0194	.0000	.0097	.0000	.0049	.0000
	a	.0016	-.0017	.0005	-.0005	.0001	-.0001
c	b	.0004	.0000	.0001	.0000	.0000	.0000
	c	.0668	-.0057	.0481	-.0016	.0193	-.0004
	d	.0195	.0000	.0097	.0000	.0049	.0000
	a	.0237	.0010	.0116	.0001	.0058	.0001
d	b	.0004	.0000	.0001	.0000	.0000	.0000
	c	1.0233	.4212	1.0112	.3907	1.0056	.3598
	d	.0194	.0000	.0097	.0000	.0049	.0000
	a	.0238	.0005	.0116	.0000	.0057	.0001
e	b	.0004	.0000	.0001	.0000	.0000	.0000
	c	1.0903	.0761	1.0501	.0416	1.0344	.0240
	d	.0194	.0000	.0097	.0000	.0049	.0000
	a	.0027	-.0021	.0006	-.0008	.0002	-.0002
f	b	.0004	.0000	.0001	.0000	.0000	.0000
	c	.0678	-.0078	.0482	-.0030	.0193	-.0009
	d	.0195	.0000	.0097	.0000	.0049	.0000

Table 4.8: Monte Carlo bias for GLS, $\tau = 1$, $\rho = 0.5$

		$n = 64$		$n = 128$		$n = 256$	
ST	DT	$\widehat{\nu}_I$	$\widehat{\nu}_F$	$\widehat{\nu}_I$	$\widehat{\nu}_F$	$\widehat{\nu}_I$	$\widehat{\nu}_F$
	a	.0000	.0000	.0000	.0000	.0000	.0000
a	b	.0000	.0000	.0000	.0000	.0000	.0000
	c	.1549	.1634	.1145	.1201	.0804	.0841
	d	.0000	.0000	.0000	.0000	.0000	.0000
	a	.0001	.0000	.0000	.0000	.0000	.0000
b	b	.0000	.0000	.0000	.0000	.0000	.0000
	c	-.0002	-.0040	.0006	-.0007	.0004	.0002
	d	.0000	.0000	.0000	.0000	.0000	.0000
	a	-.0003	-.0006	.0000	-.0001	.0000	.0000
c	b	.0000	.0000	.0000	.0000	.0000	.0000
	c	-.0006	-.0020	.0000	-.0003	.0000	.0000
	d	.0000	.0000	.0000	.0000	.0000	.0000
	a	.0001	.0001	-.0002	-.0001	.0001	.0001
d	b	.0000	.0000	.0000	.0000	.0000	.0000
	c	.3969	.3911	.3657	.3622	.3344	.3319
	d	.0000	.0000	.0000	.0000	.0000	.0000
	a	-.0003	-.0004	-.0004	-.0003	.0000	.0000
e	b	.0000	.0000	.0000	.0000	.0000	.0000
	c	.0194	.0224	.0090	.0096	.0043	.0051
	d	.0000	.0000	.0000	.0000	.0000	.0000
	a	-.0001	-.0007	.0000	-.0001	.0001	.0000
f	b	.0000	.0000	.0000	.0000	.0000	.0000
	c	-.0003	-.0024	-.0001	-.0006	.0001	.0000
	d	.0000	.0000	.0000	.0000	.0000	.0000

Table 4.9: Monte Carlo s.d. for OLS, $\tau = 1$, $\rho = 0.5$

		$n = 64$		$n = 128$		$n = 256$	
ST	DT	$\overline{\nu}$	$\widetilde{\nu}$	$\overline{\nu}$	$\widetilde{\nu}$	$\overline{\nu}$	$\widetilde{\nu}$
	a	.0031	.0069	.0011	.0024	.0004	.0008
a	b	.0001	.0001	.0000	.0000	.0000	.0000
	c	.0298	.1120	.0160	.0778	.0091	.0547
	d	.0001	.0004	.0000	.0001	.0000	.0000
	a	.0080	.0080	.0029	.0025	.0012	.0009
b	b	.0001	.0001	.0000	.0000	.0000	.0000
	c	.3691	.0512	.2614	.0223	.2163	.0088
	d	.0001	.0004	.0000	.0001	.0000	.0000
	a	.0132	.0041	.0047	.0012	.0016	.0002
c	b	.0001	.0001	.0000	.0000	.0000	.0000
	c	.5489	.0100	.4008	.0025	.2748	.0007
	d	.0014	.0004	.0005	.0001	.0002	.0000
	a	.0137	.0189	.0067	.0087	.0031	.0040
d	b	.0003	.0003	.0001	.0001	.0000	.0000
	c	.0269	.1520	.0145	.1183	.0082	.0922
	d	.0003	.0009	.0001	.0002	.0000	.0000
	a	.0143	.0214	.0065	.0092	.0029	.0042
e	b	.0003	.0003	.0001	.0001	.0000	.0000
	c	.3629	.0951	.2589	.0550	.2148	.0307
	d	.0002	.0009	.0001	.0002	.0000	.0000
	a	.0153	.0087	.0052	.0034	.0017	.0009
f	b	.0003	.0003	.0001	.0001	.0000	.0000
	c	.5489	.0189	.4005	.0062	.2747	.0019
	d	.0014	.0009	.0005	.0002	.0002	.0001

Table 4.10: Monte Carlo s.d. for GLS, $\tau = 1$, $\rho = 0.5$

		$n = 64$		$n = 128$		$n = 256$	
ST	DT	$\widehat{\nu}_I$	$\widehat{\nu}_F$	$\widehat{\nu}_I$	$\widehat{\nu}_F$	$\widehat{\nu}_I$	$\widehat{\nu}_F$
	a	.0060	.0062	.0021	.0022	.0007	.0007
a	b	.0001	.0001	.0000	.0000	.0000	.0000
	c	.1171	.1197	.0780	.0793	.0492	.0502
	d	.0004	.0004	.0001	.0001	.0000	.0000
	a	.0071	.0071	.0022	.0023	.0008	.0008
b	b	.0001	.0001	.0000	.0000	.0000	.0000
	c	.0462	.0492	.0204	.0213	.0083	.0085
	d	.0004	.0004	.0001	.0001	.0000	.0000
	a	.0032	.0033	.0009	.0010	.0002	.0002
c	b	.0001	.0001	.0000	.0000	.0000	.0000
	c	.0084	.0088	.0020	.0020	.0005	.0005
	d	.0004	.0004	.0001	.0001	.0000	.0000
	a	.0168	.0171	.0076	.0077	.0035	.0035
d	b	.0003	.0003	.0001	.0001	.0000	.0000
	c	.1586	.1658	.1241	.1277	.0951	.0971
	d	.0008	.0008	.0002	.0002	.0000	.0000
	a	.0176	.0183	.0073	.0075	.0034	.0034
e	b	.0003	.0003	.0001	.0001	.0000	.0000
	c	.0855	.0910	.0476	.0500	.0250	.0269
	d	.0008	.0008	.0002	.0002	.0000	.0000
	a	.0075	.0076	.0026	.0027	.0008	.0008
f	b	.0002	.0002	.0001	.0001	.0000	.0000
	c	.0166	.0168	.0054	.0055	.0016	.0017
	d	.0008	.0008	.0002	.0002	.0000	.0000

Table 4.11: Empirical sizes of Wald test, $\tau = 1$, $\rho = 0.5$, $\alpha = 0.05$

		$n = 64$		$n = 128$		$n = 256$	
ST	DT	W_I	W_F	W_I	W_F	W_I	W_F
	a	.065	.186	.057	.134	.053	.095
a	b	.070	.182	.055	.129	.053	.102
	c	.412	.474	.472	.513	.520	.535
	d	.068	.227	.052	.130	.051	.109
	a	.062	.194	.058	.130	.057	.094
b	b	.068	.191	.052	.127	.054	.096
	c	.060	.169	.058	.139	.047	.082
	d	.071	.228	.050	.136	.058	.111
	a	.054	.199	.056	.151	.052	.101
c	b	.068	.198	.059	.137	.048	.087
	c	.077	.227	.044	.147	.052	.120
	d	.072	.229	.045	.139	.051	.111
	a	.071	.191	.061	.146	.053	.105
d	b	.066	.182	.064	.139	.051	.099
	c	.889	.887	.960	.958	.990	.991
	d	.060	.219	.049	.139	.047	.106
	a	.070	.192	.054	.135	.045	.101
e	b	.069	.196	.059	.132	.045	.102
	c	.094	.228	.082	.176	.076	.137
	d	.060	.221	.054	.136	.054	.106
	a	.071	.211	.058	.147	.043	.096
f	b	.072	.204	.068	.145	.055	.112
	c	.068	.235	.056	.144	.041	.103
	d	.057	.236	.054	.144	.051	.105

Table 4.12: Empirical sizes of Wald tests, $\tau = 1$, $\rho = 0.5$, $\alpha = 0.10$

		$n = 64$		$n = 128$		$n = 256$	
ST	DT	W_I	W_F	W_I	W_F	W_I	W_F
	a	.113	.255	.113	.191	.103	.148
a	b	.122	.252	.108	.189	.108	.169
	c	.497	.554	.572	.608	.633	.646
	d	.113	.304	.095	.201	.106	.171
	a	.112	.267	.114	.188	.102	.148
b	b	.119	.248	.113	.178	.100	.164
	c	.121	.263	.126	.199	.090	.153
	d	.116	.297	.100	.202	.106	.168
	a	.112	.272	.113	.215	.096	.168
c	b	.114	.260	.112	.201	.095	.158
	c	.120	.302	.101	.230	.102	.189
	d	.119	.308	.096	.205	.109	.168
	a	.117	.254	.116	.209	.108	.168
d	b	.124	.268	.105	.204	.107	.167
	c	.925	.921	.973	.972	.995	.993
	d	.102	.291	.105	.197	.099	.169
	a	.120	.259	.114	.199	.111	.161
e	b	.122	.265	.102	.194	.114	.175
	c	.175	.305	.160	.242	.122	.215
	d	.109	.295	.095	.195	.103	.161
	a	.127	.269	.105	.199	.096	.162
f	b	.126	.268	.120	.213	.105	.172
	c	.124	.306	.114	.216	.097	.165
	d	.113	.309	.101	.199	.104	.172

Table 4.13: Monte Carlo bias for OLS, STa, $\tau = 1$

		$n = 64$		$n = 128$		$n = 256$	
ρ	DT	$\overline{\nu}$	$\widetilde{\nu}$	$\overline{\nu}$	$\widetilde{\nu}$	$\overline{\nu}$	$\widetilde{\nu}$
	a	.0233	-.0001	.0116	.0000	.0058	.0000
0	b	.0004	.0000	.0001	.0000	.0000	.0000
	d	.0194	.0000	.0097	.0000	.0049	.0000
	a	.0237	.0015	.0118	.0004	.0059	.0001
0.75	b	.0004	.0000	.0001	.0000	.0000	.0000
	d	.0194	.0000	.0097	.0000	.0049	.0000
	a	.0230	-.0011	.0116	-.0003	.0058	-.0001
-0.5	b	.0004	.0000	.0001	.0000	.0000	.0000
	d	.0194	.0000	.0097	.0000	.0049	.0000

Table 4.14: Monte Carlo bias for GLS, STa, $\tau = 1$

		$n = 64$		$n = 128$		$n = 256$	
ρ	DT	$\widehat{\nu}_I$	$\widehat{\nu}_F$	$\widehat{\nu}_I$	$\widehat{\nu}_F$	$\widehat{\nu}_I$	$\widehat{\nu}_F$
	a	.0000	.0000	.0000	.0000	.0000	.0000
0	b	.0000	.0000	.0000	.0000	.0000	.0000
	d	.0000	.0000	.0000	.0000	.0000	.0000
	a	.0000	.0000	.0000	.0000	.0000	.0000
0.75	b	.0000	.0000	.0000	.0000	.0000	.0000
	d	.0000	.0000	.0000	.0000	.0000	.0000
	a	-.0001	.0000	.0000	.0000	.0000	.0000
-0.5	b	.0000	.0000	.0000	.0000	.0000	.0000
	d	.0000	.0000	.0000	.0000	.0000	.0000

Table 4.15: Monte Carlo s. d. for OLS, STa, DTa

		$n = 64$		$n = 128$		$n = 256$	
τ	ρ	$\overline{\nu}$	$\widetilde{\nu}$	$\overline{\nu}$	$\widetilde{\nu}$	$\overline{\nu}$	$\widetilde{\nu}$
	0	.0034	.0068	.0012	.0024	.0004	.0008
1	0.75	.0030	.0069	.0011	.0024	.0004	.0009
	-0.5	.0036	.0067	.0013	.0023	.0005	.0008
	0	.0047	.0096	.0017	.0034	.0006	.0012
2	0.75	.0043	.0097	.0016	.0033	.0006	.0012
	-0.5	.0050	.0094	.0018	.0033	.0006	.0012
	0	.0024	.0048	.0009	.0017	.0003	.0006
0.5	0.75	.0020	.0049	.0007	.0017	.0003	.0006
	-0.5	.0027	.0047	.0009	.0016	.0003	.0006

Table 4.16: Monte Carlo s. d. for GLS, STa, DTa

		$n = 64$		$n = 128$		$n = 256$	
τ	ρ	$\widehat{\nu}_I$	$\widehat{\nu}_F$	$\widehat{\nu}_I$	$\widehat{\nu}_F$	$\widehat{\nu}_I$	$\widehat{\nu}_F$
	0	.0068	.0068	.0024	.0024	.0008	.0008
1	0.75	.0048	.0052	.0017	.0018	.0006	.0006
	-0.5	.0060	.0061	.0021	.0021	.0007	.0007
	0	.0097	.0097	.0034	.0034	.0012	.0012
2	0.75	.0068	.0073	.0024	.0025	.0008	.0008
	-0.5	.0085	.0086	.0030	.0030	.0010	.0011
	0	.0048	.0048	.0017	.0017	.0006	.0006
0.5	0.75	.0034	.0037	.0012	.0012	.0004	.0004
	-0.5	.0042	.0043	.0015	.0015	.0005	.0005

Table 4.17: Empirical sizes of Wald tests for STa, $\tau = 1$, $\alpha = 0.05$

		$n = 64$		$n = 128$		$n = 256$	
ρ	DT	W_I	W_F	W_I	W_F	W_I	W_F
0	a	.062	.183	.061	.121	.058	.099
	b	.062	.190	.056	.121	.048	.091
	d	.067	.218	.052	.126	.057	.107
0.75	a	.062	.192	.059	.144	.053	.111
	b	.064	.179	.056	.124	.055	.104
	d	.064	.225	.044	.136	.047	.111
-0.5	a	.060	.196	.058	.122	.055	.103
	b	.055	.171	.046	.111	.046	.089
	d	.068	.245	.057	.143	.058	.114

Table 4.18: Empirical sizes of Wald tests for STa, $\tau = 1$, $\alpha = 0.10$

		$n = 64$		$n = 128$		$n = 256$	
ρ	DT	W_I	W_F	W_I	W_F	W_I	W_F
0	a	.115	.258	.110	.180	.105	.159
	b	.114	.248	.111	.172	.091	.148
	d	.119	.305	.100	.190	.107	.166
0.75	a	.115	.259	.110	.207	.111	.163
	b	.121	.245	.106	.197	.117	.163
	d	.115	.310	.094	.204	.101	.181
-0.5	a	.100	.258	.109	.193	.097	.158
	b	.101	.241	.089	.178	.088	.149
	d	.128	.312	.101	.196	.104	.167

Bibliography

[1] Adenstedt, R.K., 1974. On large-sample estimation of the mean of a stationary random sequence, *Annals of Statistics* 2, 1095-1107.

[2] Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821-856.

[3] Angeloni, I., L. Aucremanne, M. Ehrmann, J. Gali, A. Levin and F. Smets, 2004. Inflation persistence in the euro area: preliminary summary of findings, *ECB, mimeo.*

[4] Bai, J., and P. Perron, 1998. Estimating and testing linear models with multiple structural changes, *Econometrica* 66, 47-78.

[5] Baillie , R.T., 1996. Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5-59.

[6] Bhattacharya, R.N., V.K. Gupta and E. Waymire, 1983. The Hurst effect under trends, *Journal of Applied Probability* 20, 649-662.

[7] Beran, J., 1995. Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models, *Journal of the Royal Statistical Society, Series B,* 57, 659-672.

[8] Breitung, J., and U. Hassler, 2002. Inference on the cointegrated rank in fractionally integrated processes, *Journal of Econometrics* 110, 167–185.

[9] Brockwell, P.J., and R.A. Davies, 1987. *Time series: theory and methods*, Springer-Verlag.

[10] Busetti, F., and A.M.R. Taylor, 2004. Tests of stationarity against a change in persistence, *Journal of Econometrics* 123, 33-66 .

[11] Campbell, J.Y., and R.J. Shiller, 1987. Cointegration and tests of present value models, *Journal of Political Economy* 95, 1062–1088.

[12] Campos, J., N.R. Ericsson and D.F. Hendry, 1996. Cointegration in the presence of structural breaks. *Journal of Econometrics* 70, 187-220.

[13] Chen, W.W., and C.M. Hurvich, 2003. Estimating fractional cointegration in the presence of polynomial trends. *Journal of Econometrics* 117, 95-121.

[14] Cheung, Y.W., and K.S. Lai, 1993. A fractional cointegration analysis of purchasing power parity, *Journal of Business and Economic Statistics* 11, 103-112.

[15] Clarida, R., J. Gali and M. Gertler, 2000. Monetary policy rules and macroeconomic stability: evidence and some theory, *The Quarterly Journal of Economics* 115, 147–180.

[16] Corbae, D., and S. Ouliaris, 1988. Cointegration and tests of purchasing power parity. *Review of Economics and Statistics* 70, 508-511.

[17] Davies, R.B., and D.S. Harte, 1987. Tests for Hurst effect, *Biometrika* 74, 95-101.

[18] Diebold, F.X., and A. Inoue, 2001. Long memory and structural change, *Journal of Econometrics* 105, 131-159.

[19] Dominguez, E., and A. Novales, 2000. Testing the expectations hypothesis in Eurodeposits, *Journal of International Money and Finance* 19, 713–736.

[20] Engle, R.F., and C.W.J. Granger, 1987. Co-integration and error correction: representation, estimation and testing, *Econometrica* 55, 251-276.

[21] Engsted, T., and C. Tangaard, 1994. Cointegration and the US term structure, *Journal of Banking and Finance* 18, 167–181.

[22] Fox, R., and M.S. Taqqu, 1986. Large sample properties of parameter estimates for strongly dependent stationary gaussian time series, *Annals of Statistics* 14, 517-532.

[23] 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.

[24] Giraitis, L., P. Kokoszka and R. Leipus, 2001. Testing for long memory in the presence of a general trend, *Journal of Applied Probability* 38, 1033-1054.

[25] Giraitis, L., and D. Surgalis, 1990. A central limit theorem for quadratic forms in strongly dependent random variables and its application to asymptotical normality of Whittle's estimate, *Probability Theory and Related Fields* 86, 87-104.

[26] Giraitis, L., and P.M. Robinson, 2003. Edgeworth expansions for semiparametric Whittle estimation of long memory, *Annals of Statistics* 31, 1325-1375.

[27] Granger, C.W.J., 1966. The typical spectral shape of an economic variable, *Econometrica* 34, 150-161.

[28] Granger, C.W.J., and N. Hyung, 2004. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance* 11, 399–421.

[29] Granger, C.W.J., and R. Joyeux, 1980. An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-39.

[30] Hall, A.D., H.M. Anderson, and C.W.J. Granger, 1992. A cointegration analysis of Treasury bills yields, *The Review of Economics and Statistics* 74, 116–126.

[31] Hamori, S., and A. Tokihisa, 1997. Testing for a unit root in the presence of a variance shift, *Economics Letters* 57, 245-253.

[32] Hannan, E.J., 1963. Regression for time series with errors of measurement, *Biometrika* 50, 293–302.

[33] Hannan, E.J., 1970. *Multiple Time Series.* Wiley, New York.

[34] Hannan, E.J., 1973. The asymptotic theory of linear time series models, *Journal of Applied Probability* 10, 130-145.

[35] Hansen, B.E., 1992. Efficient estimation and testing of cointegration vectors in the presence of deterministic trends, *Journal of Econometrics* 53, 87-121.

[36] Hansen, P.R., 2003. Structural changes in the cointegrated vector autoregressive model, *Journal of Econometrics* 114, 261 – 295.

[37] Hassler, U., F. Marmol and C. Velasco, 2006. Residual log-periodogram inference for long-run relationships, *Journal of Econometrics* 130, 165-207.

[38] Harvey, D., S.J. Leybourne and A.M.R. Taylor, 2004. Modified tests for a change in persistence, *preprint.*

[39] Henry, M., and P.M. Robinson, 1996. Bandwidth choice in Gaussian semiparametric estimation of long-range dependence, in *Athens Conference on Applied Probability and Time Series Analysis, Vol. II:*

Time Series Analysis (P. M. Robinson and M. Rosenblatt, eds), New York: Springer-Verlag, pp. 220–32.

[40] Heyde, C.C., and W. Dai, 1996. On the robustness to small trends of estimation based on the smoothed periodogram, *Journal of Time Series Analysis* 17, 141-150.

[41] Hidalgo, J., and P.M. Robinson, 1996. Testing for structural change in a long memory environment, *Journal of Econometrics* 70, 159-174.

[42] Hidalgo, J., and P.M. Robinson, 2002. Adapting to unknown disturbance autocorrelation in regression with long memory, *Econometrica* 70, 1545-1581.

[43] Hoffman, D.L., and R.H. Rasche, 1991. Long run income and interest elasticities of of money demand in the United States. *The Review of Economics and Statistics* 73, 665-674.

[44] Hosking, J.R.M., 1981. Fractional differencing, *Biometrika* 68, 165-176.

[45] Hualde, J., and P.M. Robinson, 2002. Root-n-consistent estimation of weak fractional cointegration, *preprint.*

[46] Hurst, H.E., 1951. Long-Term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers* 116, 770-799.

[47] Hurvich, C.M., R. Deo and J. Brodsky, 1998. The mean squared error of Geweke and Porter Hudak's estimate of the memory parameter of a long memory time series, *Journal of Time Series Analysis* 19, 19-46.

[48] Hurvich, C.M., and B.K. Ray, 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes, *Journal of Time Series Analysis* 16, 17–42.

[49] Jeganathan, P., 1999. On asymptotic inference in cointegrated time series with fractionally integrated errors, *Econometric Theory* 15, 583-621.

[50] Johansen, S., 1991. Estimation and hypothesis testing of co-integrating vectors in Gaussian vector autoregressive models, *Econometrica* 59, 1551-1580.

[51] Johansen, S., and K. Juselius, 1992. Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for the UK, *Journal of Econometrics* 53, 211-244.

[52] Kim, C.J., and C.R. Nelson, 1999. Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle, *Review of Economics and Statistics* 81, 608-616.

[53] Kim, J.Y., 2000. Detection of a change in persistence in a linear time series, *Journal of Econometrics* 95, 97-116.

[54] Kim, J.Y., J. Belaire-Franch, and R. Badilli-Amador, 2002. Corrigendum to 'Detection of a change in persistence in a linear time series', *Journal of Econometrics* 109, 389-392.

[55] Künsch, H.R., 1986. Discrimination between monotonic trends and long range dependence, *Journal of Applied Probability* 23, 1025-1030.

[56] Künsch, H.R., 1987. Statistical aspects of self-similar processes, *Proceedings of the First World Congress of the Bernoulli Society*, VNU Science Press, 1, 67–74.

[57] Kurtz, T.G., and P. Protter, 1991. Weak limit theorems for stochastic integrals and stochastic differential equations, *Annals of Probability* 19, 1035-1070.

[58] Lanne, M., 2000. Near unit root, cointegration and the term structure of interest rates, *Journal of Applied Econometrics* 15, 513–529.

[59] Lo, A.W., 1991. Long Term Memory In Stock Market Prices, *Econometrica* 59, 1279-1313.

[60] Lobato, I.N., 1999. A semiparametric two-step estimator in a multivariate long memory model, *Journal of Econometrics* 90, 129-153.

[61] Lobato, I.N., and N.E. Savin, 1998. Real and spurious long memory properties of the stock market data, *Journal of Business and Economic Statistics* 16, 261-283.

[62] Lucas, R., 1976. Econometric policy evaluation: a critique, *Carnegie-Rochester Series on Public Policy* 1, 19-46.

[63] Mandelbrot, B.B., and J.W. Van Ness, 1968. Fractional Brownian motions, fractional noises and applications, *SIAM Review* 10, 422-437.

[64] Marinucci, D., and P.M. Robinson, 1999. Alternative forms of fractional Brownian motion, *Journal of Statistical Planning and Inference* 80, 111–122.

[65] Marinucci, D., and P.M. Robinson, 2000. Weak convergence of multivariate fractional processes, *Stochastic Processes and their Applications* 86, 103-120.

[66] Marinucci, D., and P.M. Robinson, 2001. Semiparametric fractional cointegration analysis, *Journal of Econometrics* 105, 225-247.

[67] Mikosch, T., and C. Stărică, 1999. Change of structure in financial time series, long range dependence and the GARCH Model, *Working paper*, Department of Statistics, University of Pennsylvania.

[68] Nelson, C.M., and C.I. Plosser, 1982. Trends and random walk in macroeconomic time series, *Journal of Monetary Economics* 10, 139-162.

[69] Newey, W.K., and D. McFadden, 1994. Large sample estimation and hypothesis testing, in *Handbook of Econometrics* (R.F. Engle and D.L. McFadden eds.), Elsevier Science B.V., pp. 2113-2145.

[70] Ogaki, M., and J.Y. Park, 1997. A cointegration approach to estimating preference parameters, *Journal of Econometrics* 82, 107-134.

[71] O'Reilly, G., and K. Whelan, 2004. Has euro-area inflation persistence changed over time?, *ECB,* W. P. 335.

[72] Park, J.Y., and P.C.B. Phillips, 1988. Statistical inference in regressions with integrated processes: part 1, *Econometric Theory* 4, 468-497.

[73] Perror, P., 1991. The great crash, the oil price shock and the unit root hypothesis, *Econometrica*, 57, 1361-1401.

[74] Perron, P., and J.Y. Campbell, 1993. A note on Johansen's procedure when trends are present, *Empirical Economics* 18, 777-789.

[75] Phillips, P.C.B., 1991. Optimal inference in co-integrated systems, *Econometrica* 59, 283-306.

[76] Phillips, P.C.B., and S. Ouliaris, 1988. Testing for cointegration using principal components methods, *Journal of Economic Dynamics and Control* 12, 205–230.

[77] Phillips, P.C.B., and M. Loretan, 1991. Estimating long-run economic equilibria, *Review of Economic Studies* 58, 407-436.

[78] Robinson, P.M., 1986. On the errors-in-variables problem for time series. *Journal of Multivariate Analysis* 19, 240-250.

[79] Robinson, P.M., 1994a. Time series with strong dependence, in *Advances in Econometrics: Sixth World Congress* vol. I, (C.A. Sims ed.), New York: Cambridge University Press, pp. 47-95.

[80] Robinson, P.M., 1994b. Semiparametric analysis of long-memory time series, *Annals of Statistics* 22, 515–39.

[81] Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence, *Annals of Statistics* 23, 1048 - 1072.

[82] Robinson, P.M., 1995b. Gaussian semiparametric estimation of long range dependence, *Annals of Statistics* 23, 1630 - 1661.

[83] Robinson, P.M., 2003. Long memory time series, in *Time series with long memory* (P.M. Robinson ed.), Oxford University press, 4-32.

[84] Robinson, P.M., 2005a. Efficiency improvements in inference on stationary and nonstationary fractional time series, *Annals of Statistics* 33, 1800-1842.

[85] Robinson, P.M., 2005b. The distance between rival nonstationary fractional processes, *Journal of Econometrics* 128, 283-300.

[86] Robinson, P.M., and J. Hualde, 2003. Cointegration in fractional systems with unknown integration orders, *Econometrica* 71, 1727-1766.

[87] Robinson, P.M., and D. Marinucci, 2000. The averaged periodogram for nonstationary vector stochastic processes, *Statistical Inference on Stochastic Processes* 3, 149-160.

[88] Robinson, P.M., and D. Marinucci, 2001. Narrow band analysis of nonstationary processes, *Annals of Statistics* 29, 947-86.

[89] Robinson, P.M., and D. Marinucci, 2003. Semiparametric frequency domain analysis of fractional cointegration, in *Time series with long memory* (P.M. Robinson ed.), Oxford University press, pp. 334-373.

[90] Robinson, P.M., and Y. Yajima, 2002. Determination of cointegrating ranking fractional systems, *Journal of Econometrics* 106, 217–241.

[91] Said, S.E., and D.A. Dickey, 1984. Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* 71, 599-607.

[92] Söderlind, P., and A. Vredin, 1996. Applied cointegration analysis in the mirror of macroeconomic theory, *Journal of Applied Econometrics* 11, 363-381.

[93] Sowell, F.B., 1992. Maximum likelihood estimation of stationary univatiate fractionally integrated time series models, *Journal of Econometrics* 53, 165–88.

[94] Stock, J., and M.W. Watson, 1988. Testing for common trends, *Journal of the American Statistical Society* 83, 1097-1107.

[95] Stock, J., and M.W. Watson, 1993. A simple estimate of cointegrating vectors in higher order integrated systems, *Econometrica* 61, 783-820.

[96] Svensson, L.E.O., 1997. Inflation forecast targeting: implementing and monitoring inflation targets, *European Economic Review* 41, 1111–1146.

[97] Taylor, J.B., 1993. Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy* 39, 195–214.

[98] Teverovsky, V., and M. Taqqu, 1997. Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimate, *Journal of Time Series Analysis* 18, 279-304.

[99] Velasco, C., 1999a. Non-stationary log-periodogram regression, *Journal of Econometrics* 91, 325-371.

[100] Velasco, C., 1999b. Gaussian semiparametric estimation of nonstationary time series, *Journal of Time Series Analysis* 20, 87-127.

[101] Velasco, C., 2000. Non gaussian log-periodogram regression, *Econometric Theory* 16, 44-79.

[102] Velasco, C., and P.M. Robinson, 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series, *Journal of the American Statistical Association* 95, 1229-1243.

[103] West, K.D., 1988. Asymptotic normality when regressors have a unit root, *Econometrica* 56, 1397-1417.

[104] Yong, C.H., 1974. Asymptotic behaviour of trigonometric series, Hong Kong: Chinese University.

[105] Zivot, E., and D.W.K. Andrews, 1992. Further evidence on the great crash, the oil price shock and the unit root hypothesis, *Journal of Business and Economic Statistics* 10, 251–270.

[106] Zygmund, A., 1988. *Trigonometric series*, Cambridge University press.