Fractional Cointegration Analysis of Nonlinear Time Series with Long Memory

Afonso Gonçalves da Silva

London School of Economics and Political Science

London, April 2008

A thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of Doctor of Philosophy

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

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is my own work. Parts of Chapters 2, 3, and 4 were undertaken as joint work with Professor Peter M. Robinson.

Afonso Gonçalves da Silva

Abstract

This thesis develops theoretical tools for fractional cointegration analysis of nonlinear time series. These tools are employed to establish consistency of narrow band versions of Least Squares and Principal Components, in situations when the observables do not follow traditional linear process assumptions. Chapter 1 introduces the problem, and Chapter 2 reviews the tools and techniques used in the literature for analysing stationary fractional cointegration, emphasizing methods that will be the focus of subsequent chapters. Chapter 3 considers a bivariate factor model, where the unobservable common factor and idiosyncratic errors are stationary and serially uncorrelated, but have strong dependence in higher moments. Assuming the latent variables are driven by long memory stochastic volatility models, and that the underlying persistence is higher in the factor than in the errors, a fractional cointegrating relationship can be recovered by suitable transformation of the data. We consider a narrow band semiparametric estimate of the factor loadings, which is shown to be consistent with a rate of convergence. Chapter 4 contains two Monte Carlo experiments: the first illustrates the performance of the Narrow Band Least Squares estimate in the setting of the previous chapter, while the second attempts to fill the gap in theoretical distributional results for nonlinear processes, by analysing distributional properties of the more general Weighted Narrow Band Least Squares estimate, under linear and nonlinear settings. Chapter 5 extends the techniques of Chapter 3 to a general multivariate setting, with more than two observables and multiple common factors. A narrow band version of the Principal Components estimate is introduced and shown to converge to the space spanned by the factor loadings, allowing their consistent estimation under suitable linear restrictions. A Monte Carlo study of finite sample performance and an empirical application to European equity indices are also presented.

Acknowledgements

I thank my supervisor, Peter M. Robinson, for his support, patience, and careful supervision during these years. I will always remember him as an inspiring and helpful mentor. I thank the remaining faculty who taught, challenged, and inspired me during my time in LSE. In particular, I thank Oliver Linton, Javier Hidalgo, Greg Connor, and Andrew Patton for their support and helpful advice during the stressful job market period, and Sue Kirkbride for her efficient and invaluable support. I thank my friends and fellow students in the department for endless intellectual coffee breaks, and enduring me through the highs and lows of the PhD experience; especially Violetta Dalla, Fabrizio Iacone, Dennis Kristensen, and David Jacho-Chavez, who shared with me the love for (and occasional frustration with) Econometric Theory. I gratefully acknowledge financial support from Fundação para a Ciência e a Tecnologia, through grant SFRH/BD/4783/2001, and the ESRC, through grant R000239936. I thank Concordia Advisors for granting me access to the Bloomberg data used in Section 5.6. Finally, I thank my parents, Amélia and António, for their encouragement and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The analysis, estimation, and testing of cointegration relationships, introduced by Granger (1981) and developed, among others, by Granger and Weiss (1983), Engle and Granger (1987), has grown to be one of the most important and active areas of research in time series econometrics. It was noticed by these authors that in a variety of different situations, notably in macroeconomic models, certain linear combinations of nonstationary time series may be stationary, and that such situations may be interpreted as a long run relationship between variables driven by common factors.

In much of the early literature on the subject, the concept of nonstationarity has been used interchangeably with the presence of unit roots; likewise, the only type of stationary processes considered were short memory ones, such as autoregressive moving average (ARMA) processes, or more generally those satisfying certain strong mixing conditions. In recent times, some of the focus has shifted to a more general framework, where the orders of integration of both the observable time series and the cointegration residuals are allowed to be real valued, rather than integers. Obviously, such a framework nests the unit root case; perhaps more importantly, it accurately represents additional types of persistence that are

excluded from unit root cointegration analysis. One may observe nonstationary time series that nonetheless revert to a stable, long run mean value; or stationary series where shocks appear to decay very slowly over time. Within this framework, one can even entertain the concept of cointegration between stationary variables, wherein estimates such as ordinary least squares (OLS), frequently employed in the unit root setting, become invalid.

As the analysis of fractional cointegration is a fairly recent field, most of the asymptotic statistical theory has been developed under assumptions that may not be satisfied in practice. A range of parametric methods is available; however, these methods require correct specification of the short run dynamics of the data, which is not directly relevant to the intrinsically long run phenomenon of cointegration. We focus instead on semiparametric estimates, which are valid under substantially weaker conditions, at the cost of a slower rate of convergence. Still, most analysis of these estimates has relied on restrictions on the heterogeneity and dependence of the innovations in the Wold representation of the processes. These restrictions may range from Gaussianity to the conditional homogeneity of some moments, and are in general invalid for nonlinear processes. In this thesis, we aim to develop theoretical tools applicable to a wide variety of nonlinear models, where observables are non-trivial transformations of underlying stationary long memory processes. For concreteness, and also because of its empirical relevance, we focus throughout on a particular framework where observable time series follow a statistical factor model, and the underlying components of the model follow stochastic volatility (SV) models. This framework, while relatively general, is particularly well suited to describe properties of multivariate asset returns, and presents a clear violation of the type of assumptions mentioned earlier.

Financial time series, such as asset returns, are commonly found to be approximately uncorrelated but not independent across time. Much of this dependence

can be traced to the fact that volatilities are time dependent, with highly volatile observations grouped in some periods, and relatively low volatilities elsewhere. A great deal of attention has focused on modelling the consequent conditional heteroscedasticity. Influential early contributions were the autoregressive conditional heteroscedasticity (ARCH) model of Engle (1982) (applied there to inflation data), the generalised ARCH (GARCH) extension of Bollerslev (1986), and along a different line, the SV model of Taylor (1986). Empirical evidence has suggested a higher degree of persistence than these models entail, leading to Engle and Bollerslev's (1986) introduction of the integrated GARCH (IGARCH) model. However, the persistence implied by this model (and other unit root based ones, such as integrated exponential GARCH, IEGARCH) seems too extreme. On the one hand, the absence of mean reversion in the second moments implies permanent shifts to long term volatility forecasts, which is theoretically implausible. On the other, empirical investigation of volatility measures, such as absolute values and squares of observations, suggests they are better explained as stationary processes with long memory, indicating the need for a more flexible model of volatility persistence; see, for example, Whistler (1990), Ding, Granger, and Engle (1993), Ding and Granger (1996), Andersen and Bollerslev (1997).

Several parametric models for this phenomenon have been proposed. Robinson (1991) extended the GARCH framework to an ARCH(∞) model that can explain greater persistence. Other models within this framework include Ding and Granger (1996), Baillie, Bollerslev, and Mikkelsen (1996), Bollerslev and Mikkelsen (1996). Other authors have extended Taylor's (1986) SV model to explain long memory in squares, e.g. Andersen and Bollerslev (1997), Harvey (1998), Breidt, Crato, and de Lima (1998).

In a parallel line of research, asset pricing models assume the existence of one or more common factors explaining asset returns. The classical capital asset pricing

model (CAPM) of Sharpe (1964) decomposes returns into a single factor, interpreted as the market return, and an idiosyncratic component. The intertemporal CAPM (ICAPM) of Merton (1973) and the arbitrage pricing theory (APT) of Ross (1976) show that, under more realistic assumptions, multiple factors need to be considered as determinants of returns. Estimation of the ICAPM requires correct specification of the factors, there assumed to be observable state variables; the APT uses asymptotics on the cross-sectional dimension (i.e. the number of assets) for its theoretical implications, and also to allow estimation of the unobservable factors and respective loadings.

The methods introduced in the following chapters allow for identification and estimation of factor loadings when cross sectional asymptotics are not available, by exploring the persistence in higher moments implied by the SV models employed. We analyse narrow band versions of least squares and of principal components, under quite general semiparametric assumptions. We relax the linear process assumptions used in the past, which are invalid in this setting, and consider instead general nonlinear transformations of underlying Gaussian processes. Our techniques allow for nonparametric nonlinearity, in the form of generalised SV models, and semiparametric dependence structures of underlying components, where the short run dynamics are essentially unrestricted.

The following chapter reviews the tools and techniques used for analysing stationary long memory and fractional cointegration, emphasizing semiparametric methods that will be the focus of subsequent chapters. After introducing the main concepts, we describe how they can be applied in the context of nonlinear processes. A simplified factor model (in the spirit of those mentioned above) is presented and used to exemplify how the long memory properties of underlying factor model components can give rise to fractional cointegration in higher moments of observables. Some of the challenges associated with this setting are

discussed, namely the lack of orthogonality and the nonlinearity implied by the power transformation. The leading semiparametric estimates of fractional cointegration are then introduced, together with a sketch of the arguments typically used to establish consistency. A brief description of semiparametric memory estimation techniques is also provided.

Chapter 3 presents our main theoretical results. A more general version of the bivariate factor model introduced in Chapter 2 is described, where the unobservable common factor and idiosyncratic errors are stationary and serially uncorrelated, but are generated by SV models with strong dependence in higher moments. We introduce an approximation result for cross-moments of nonlinear functions of Gaussian variables, which is then used to establish consistency of the Narrow Band Least Squares estimate. Under fairly general assumptions, the rate of convergence for nonlinear processes is shown to be comparable to that of linear processes.

Two Monte Carlo studies of finite sample properties are shown in Chapter 4. The first of these studies illustrates the performance of the Narrow Band Least Squares estimate in the setting of Chapter 3, showing encouraging performances for small to moderate sample sizes across a variety of specifications. The second study attempts to fill the gap in theoretical distributional results for nonlinear processes, by analysing distributional properties of the more general Weighted Narrow Band Least Squares estimate, under linear and nonlinear settings. Results indicate that asymptotic distributional results may be misleading in finite samples, and that under a number of different specifications Weighted Narrow Band Least Squares may be dominated by its unweighted counterpart.

Chapter 5 applies the theoretical tools developed in Chapter 3 to a multivariate setting, extending the bivariate model presented therein to allow for more than two observables and multiple common factors. A narrow band version of principal components is introduced and shown to be consistent for the space spanned by

the factor loadings, under assumptions comparable to those of Chapter 3. Under a suitable set of linear restrictions on the parameters, it is possible to consistently estimate the original factor loadings. The practical application and finite sample performance of these techniques are illustrated in a Monte Carlo study and an empirical study of risk exposures for a set of European large-cap equity indices.

Chapter 2

Stationary long memory and fractional cointegration in nonlinear models

2.1 Introduction

This chapter presents a general discussion of the tools commonly employed in the analysis of long memory and fractional cointegration, with special emphasis on problems raised by nonlinearity. In Sections 2 and 3 we introduce the concepts of long memory and fractional cointegration. Section 4 describes some basic properties of nonlinear transformations of long memory processes, while Section 5 discusses the possibility of fractional cointegration between such transformations. Section 6 reviews methods of estimating cointegrating coefficients (the stress being on relatively simple "single equation" methods). Section 7 briefly describes some leading semiparametric estimates of the memory parameter. Section 8 concludes.

2.2 Long memory processes

As noted in Chapter 1, the modelling of dependence in stationary economic variables has until recently relied heavily on models satisfying certain weak dependence conditions, notably the (stationary and invertible) ARMA specification. These models are necessarily associated with absolutely summable autocovariances, and a spectral density bounded away from zero and infinity. Nevertheless, various time series across the natural and social sciences have been found to display a degree of persistence that violates these assumptions, leading to a diverging spectral density and hyperbolically decaying autocovariances. An early example was found in Hydrology by Hurst (1951), in his study of the Nile river yearly minima over a period spanning several centuries. In Economics, Granger (1966) noted that "the typical spectral shape of an economic variable" is dominated by a sharp peak at frequency zero, and other authors (see e.g. Mandelbrot, 1969) have argued for the need to explicitly account for these effects in econometric models.

The early literature on long memory processes is closely related to self-similarity, in particular to fractional Brownian motion, studied in detail by Mandelbrot and van Ness (1968). We first introduce the fundamental properties used in the definition of fractional Brownian motion, wherein "equality in distribution" signifies equality of all finite-dimensional distributions of the processes.

Definition 1 *A continuous time process $\{X_t\}_{t\in\mathbb{R}}$ is self-similar with index $H > 0$ if, for any $a > 0$, $\{X_{at}\}_{t\in\mathbb{R}}$ and $\{a^H X_t\}_{t\in\mathbb{R}}$ are equal in distribution.*

Definition 2 *A continuous time process $\{X_t\}_{t\in\mathbb{R}}$ has stationary increments if, for any $h \in \mathbb{R}$, $\{X_{t+h} - X_h\}_{t\in\mathbb{R}}$ and $\{X_t - X_0\}_{t\in\mathbb{R}}$ are equal in distribution.*

Both self-similarity and stationary increments are invariance properties of the process between different time intervals. Intuitively, a process with stationary increments moves in the same manner in any interval of fixed length t, irrespectively

of the starting point h. A self-similar process behaves similarly in intervals of any length: keeping the starting point at zero, if the length of the time interval, t, is expanded or shrunk by a factor a, then scaling the units of the process by a factor a^H leaves its distribution unchanged. Fractional Brownian motion is simply defined as a Gaussian process with these two properties.

Definition 3 *A continuous time process $\{B_t\}_{t\in\mathbb{R}}$ is a fractional Brownian motion if*

(i) It is a mean zero Gaussian process with $B_0 = 0$ a.s.;

(ii) It has stationary increments;

(iii) It is self-similar with index $0 < H < 1$.

It can be shown that, for a given H, all fractional Brownian motion processes are equal in distribution up to a multiplicative scaling factor; furthermore, the variance of B_t is given by

$$E(B_t^2) = \sigma^2|t|^{2H}, \tag{2.1}$$

where $\sigma^2 = E(B_1^2)$, while the autocovariance function is

$$E(B_tB_s) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} + |t-s|^{2H}\right). \tag{2.2}$$

Since a mean zero Gaussian process can be uniquely characterised by (2.2), or equivalently by (2.1) and the stationary increments property, alternative (and equivalent) definitions of fractional Brownian motion may be employed, wherein (2.1) or (2.2) replace the requirement of self-similarity. Taking the usual Brownian motion (which in this setting corresponds to the above definition with $H = \frac{1}{2}$) as a primitive, fractional Brownian motion may also be defined as a fractional integral thereof. We will not pursue such alternative definitions here.

Fractional Brownian motion is a nonstationary process, but its increments form a stationary, Gaussian, and mean zero process by definition. We therefore focus on the (discrete time) increment process.

Definition 4 *If* $\{B_t\}_{t\in\mathbb{R}}$ *is a fractional Brownian motion, the process*

$$X_t = B_{t+1} - B_t, \tag{2.3}$$

where $t \in \mathbb{Z} = \{0, \pm 1, \ldots\}$, *is a fractional Brownian noise.*

From (2.2), it is easily found that the autocovariance function of X_t is

$$\begin{aligned} \rho_j &= E(X_0 X_j) \\ &= \frac{\sigma^2}{2}\left(|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\right), \qquad j \in \mathbb{Z}. \end{aligned} \tag{2.4}$$

For $H = \frac{1}{2}$, this reduces to $\rho_j = \sigma^2 1(j = 0)$, where $1(\cdot)$ denotes the identity function. This is a well known result; the usual Brownian motion has independent increments. However, for $H \neq \frac{1}{2}$ the autocorrelation function ρ_j is non-zero for all j, and furthermore

$$\rho_j \sim \sigma^2 H(2H-1)|j|^{2H-2}, \qquad \text{as } j \to \infty, \tag{2.5}$$

where "$\sim$" indicates that the ratio of left- and right-hand sides tends to one. We find that the autocovariance function of fractional Brownian motion decays hyperbolically. If $H > \frac{1}{2}$, this decay is so slow that the ρ_j are not absolutely summable.

Alternative parametric models displaying long memory, inspired by the popular integrated ARMA (ARIMA) models, were developed by Adenstedt (1974), and later popularised in Economics by Granger and Joyeux (1980), Hosking (1981),

and Granger (1981). We first introduce the Wold Representation Theorem (here restricted to scalar processes for simplicity).

Theorem 2.1 *Any covariance stationary scalar process z_t may be uniquely represented as*

$$z_t = \mu + \eta_t + \Phi(L)\varepsilon_t, \qquad t \in \mathbb{Z}, \tag{2.6}$$

where:

(i) $\mu = E(z_t)$;

(ii) η_t *is a deterministic component;*

(iii) ε_t *is a white noise process with variance* σ_ε^2, *i.e.*

$$E(\varepsilon_t) = 0,$$
$$E(\varepsilon_t \varepsilon_s) = \sigma_\varepsilon^2 1(t = s);$$

(iv) $\Phi(L)$ *is an infinite order polynomial on the lag operator* L, *i.e.*

$$\Phi(L) = \sum_{j=0}^{\infty} \varphi_j L^j,$$

where $\varphi_0 = 1$, *and furthermore the Wold coefficients* φ_j *are square summable,*

$$\sum_{j=1}^{\infty} \varphi_j^2 < \infty.$$

We will assume throughout this chapter that the process z_t in (2.6) is mean zero and purely non-deterministic, so that $\mu = 0$, $\eta_t \equiv 0$ a.s., and we can write z_t as a linear filter on ε_t,

$$z_t = \Phi(L)\varepsilon_t. \tag{2.7}$$

In the special case of stationary ARMA processes, we have

$$\Phi(L) = A^{-1}(L)B(L), \tag{2.8}$$

where $A(L)$ and $B(L)$ are finite dimensional polynomials, and all roots of $A(\cdot)$ lie outside the unit circle; in this case, the Wold coefficients φ_j decay exponentially to zero with j, or are indeed identically zero for large enough j if $A(L) \equiv 1$. It follows that the autocovariance function, $\rho_j = \mathrm{Cov}(z_0, z_j)$, is absolutely summable. To see this, let $\sigma_\varepsilon^2 = 1$ without loss of generality, and note that

$$\rho_j = \sum_{k=0}^{\infty} \varphi_k \varphi_{k+j}, \qquad j \geq 0, \tag{2.9}$$

so we may write

$$\begin{aligned} \sum_{j=0}^{\infty} |\rho_j| = \sum_{j=0}^{\infty} \left| \sum_{k=0}^{\infty} \varphi_k \varphi_{k+j} \right| &\leq \sum_{k=0}^{\infty} \left(|\varphi_k| \sum_{j=0}^{\infty} |\varphi_{k+j}| \right) \\ &\leq \sum_{k=0}^{\infty} |\varphi_k| \sum_{j=0}^{\infty} |\varphi_j| = \left(\sum_{j=0}^{\infty} |\varphi_j| \right)^2, \end{aligned} \tag{2.10}$$

and therefore absolute summability of ρ_j follows from absolutely summability of φ_j.

Adenstedt (1974), Granger and Joyeux (1980), and Hosking (1981) discussed processes where instead $\Phi(L)$ is the fractional differencing operator,

$$\Phi(L) = (1 - L)^{-d}, \tag{2.11}$$

where $-\frac{1}{2} < d < \frac{1}{2}$, which implies that the weights φ_j can be written, by a binomial

expansion, as functionals of the Gamma function $\Gamma(\cdot)$,

$$\varphi_j = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}. \tag{2.12}$$

By analogy with ARIMA models, this family of processes is referred to as fractionally integrated ARMA (ARFIMA) $(0, d, 0)$, or simply fractionally integrated noise.

Definition 5 *A scalar process z_t is said to be an ARFIMA$(0, d, 0)$, process, for $-\frac{1}{2} < d < \frac{1}{2}$, if it is covariance stationary and*

$$z_t = (1-L)^{-d}\varepsilon_t, \qquad t \in \mathbb{Z},$$

where ε_t is a white noise process.

The asymptotic behaviour of the φ_j sequence may be derived from Stirling's approximation,

$$\Gamma(x) \sim \left(\frac{2\pi}{x}\right)^{\frac{1}{2}} \left(\frac{x}{e}\right)^x, \qquad \text{as } x \to \infty. \tag{2.13}$$

While processes generated by (2.7), (2.11) are still stationary and invertible, when $d \neq 0$ the φ_j decay hyperbolically,

$$\varphi_j \sim \frac{j^{d-1}}{\Gamma(d)}, \qquad \text{as } j \to \infty. \tag{2.14}$$

The autocovariance function, $\rho_j = \mathrm{Cov}(z_0, z_j)$, is

$$\rho_j = \sigma_\varepsilon^2 \frac{(-1)^j \Gamma(1-2d)}{\Gamma(1-d+j)\Gamma(1-d-j)}, \tag{2.15}$$

so using Stirling's approximation again we may derive a power law for ρ_j, namely

$$\rho_j \sim C_\rho j^{2d-1}, \qquad \text{as } j \to \infty, \tag{2.16}$$

for some finite, non-zero constant C_ρ that depends on σ_ε^2 and d. Furthermore, the spectral density $f(\lambda)$ of z_t, which satisfies $\rho_j = \int_{-\pi}^{\pi} f(\lambda)\cos(j\lambda)\mathrm{d}\lambda$, is

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}\left(2\sin\frac{\lambda}{2}\right)^{-2d}, \tag{2.17}$$

for $0 < \lambda \leq \pi$, and therefore (since $\sin x \sim x$ as $x \to 0$)

$$f(\lambda) \sim C_f \lambda^{-2d}, \qquad \text{as } \lambda \to 0^+, \tag{2.18}$$

for a finite constant $C_f > 0$.

For $d < 0$, the autocovariances ρ_j are all negative for $j \neq 0$ (thus $C_\rho < 0$), and $f(0) = 0$, implying that $\sum_{j=0}^{\infty} \rho_j = 0$. For $d = 0$, z_t trivially reduces to white noise, as $\Phi(L) \equiv 1$. For $d > 0$, the autocovariances ρ_j are all positive (thus $C_\rho > 0$), the Wold coefficients and the autocovariances are not summable, and therefore the spectral density diverges around $\lambda = 0$. Extension to the general ARFIMA(p, d, q) family is straightforward, replacing the white noise ε_t in (2.7) by a stationary ARMA(p, q) process, whence (2.16), (2.18) still hold for finite $C_\rho \neq 0$ and $C_f > 0$ respectively.

Definition 6 *A scalar process z_t is said to be an ARFIMA(p, d, q) process, for $-\frac{1}{2} < d < \frac{1}{2}$, if it is covariance stationary and*

$$(1 - L)^d A(L) z_t = B(L)\varepsilon_t, \qquad t \in \mathbb{Z},$$

where ε_t is a white noise process, and $A(L)$, $B(L)$ are polynomials of degree p, q respectively, with no common roots.

The crucial property (2.18) is invariant to short range dynamics. Assume that

$$z_t = (1 - L)^{-d} v_t, \qquad t \in \mathbb{Z}, \tag{2.19}$$

where now v_t is any covariance stationary process with spectral density, $g(\lambda)$, bounded away from zero and infinity at the origin. The process v_t thus defined is quite general; in particular it may display slower-than-exponential decay of autocorrelations and Wold coefficients, which is excluded by the ARMA(p,q) specification. The spectral density of z_t is

$$f(\lambda) = \left(2\sin\frac{\lambda}{2}\right)^{-2d} g(\lambda), \tag{2.20}$$

for $0 < \lambda \leq \pi$, yielding

$$f(\lambda) \sim g(0)\lambda^{-2d}, \qquad \text{as } \lambda \to 0^+, \tag{2.21}$$

satisfying (2.18). As such, it is possible to disentangle the long run persistence of the process from the short run dynamic properties, captured here by $g(\lambda)$. Analysis of parametric ARFIMA(p,d,q) processes is sensitive to specification of the orders p and q, assumed to be finite, and estimates derived under such assumptions are likely to be inconsistent otherwise. By contrast, methods relying only on the asymptotic behaviour of $f(\lambda)$ around the origin (or ρ_j as $j \to \infty$) are valid under significantly weaker assumptions, and do not require specification or estimation of high frequency dynamics at all. As is typical of semiparametric analysis, the additional robustness is attained at the cost of slower rates of convergence. We present two alternative semiparametric definitions of long memory.

Definition 7 *(frequency domain) A scalar process z_t is said to be integrated of order d, denoted $I(d)$, for $d \in (-\frac{1}{2}, \frac{1}{2})$ if (2.18) holds for some $C_f \in (0, \infty)$.*

Definition 8 *(time domain) A scalar process z_t is said to be integrated of order d, denoted $I(d)$, for $d \in (-\frac{1}{2}, \frac{1}{2})$ if either:*

(i) $d = 0$ and $0 < \sum_{j=0}^{\infty} \rho_j < \infty$, or

(ii) $d \neq 0$ *and (2.16) holds for some finite* $C_\rho \neq 0$.

Note that, under mild regularity conditions (see Theorem III-12 of Yong, 1974), satisfied in particular by ARFIMA models, these two definitions are equivalent, and indeed we will use them interchangeably throughout this chapter. Some authors employ slightly more general versions of (2.16), (2.18) replacing the finite constants C_ρ, C_f by possibly diverging functions, assumed to be slowly varying at infinity or zero, respectively. We call d the "memory parameter" of z_t; if $d < 0$ the process is said to to be anti-persistent or negative memory; if $d = 0$ the process is said to be weakly correlated or short memory; if $d > 0$ the process is said to be strongly correlated or long memory. While the negative memory case may be relevant in some special situations, namely in the analysis of overdifferenced time series, its properties are less relevant in the context of cointegration, and we will mainly focus throughout the thesis on the case $d \geq 0$. Comparison of (2.16) with (2.5) indicates that the decay of the autocovariance function (and therefore the persistence of the processes) for fractionally integrated noise and fractional Brownian noise will be comparable when $d = H - \frac{1}{2}$, and indeed this reparameterisation allows either definition to nest fractional Brownian noise, as well as ARFIMA models.

Unfortunately, the semiparametric definitions presented are too broad for useful asymptotic theory to be developed, and a number of additional technical assumptions have been employed in the literature. The ones most widely used assume a linear process structure for z_t; while we know (by Theorem 2.1) that such a representation can always be obtained for a covariance stationary process, Theorem 2.1 only ensures that the innovations ε_t are serially uncorrelated. Stronger versions are obtained by imposing additional structure on the innovation process.

The most restrictive assumption under which semiparametric analysis of long memory processes has been pursued is that of Gaussianity. If z_t is a Gaussian

process, the following assumption holds.

Assumption 2.1 *The process z_t is generated by (2.7), where the innovations ε_t are iid Gaussian.*

An alternative, weaker specification assumes that z_t is a linear filter of an iid process, without requiring Gaussianity.

Assumption 2.2 *The process z_t is generated by (2.7), where the innovations ε_t are iid.*

It is possible to allow added generality by stating assumptions only on specific moments of the innovations, in particular imposing neither serial independence nor strict stationarity of ε_t. These assumptions (used by e.g. Robinson, 1994a) may take various forms, depending on the particular application, but are usually restrictions on the dependence and heterogeneity of such moments (conditional or otherwise) stated using the well known concept of martingale difference processes.

Definition 9 *A process ε_t is said to be a martingale difference sequence if*

$$E(\varepsilon_{t+1}|\mathcal{F}_t) = 0, \qquad t \in \mathbb{Z},$$

where $\mathcal{F}_t$ is the σ-algebra generated by $\{\varepsilon_s, s \leq t\}$.

Assumption 2.3 *The process z_t is generated by (2.7), where ε_t, $\varepsilon_t^2 - \sigma_\varepsilon^2$ are martingale difference sequences.*

This assumption is frequently extended to guarantee constant conditional skewness and kurtosis, i.e. that $\varepsilon_t^k - E(\varepsilon_t^k)$ is also a martingale difference for $k = 3, 4$.

Finally, some authors abandon the linear process assumption altogether, establishing asymptotic theory for nonlinear transformations of processes satisfying either (2.1) or (2.2). We discuss these models in Section 2.4.

2.3 Fractional cointegration

The concept of cointegration, first introduced by Granger (1981) and developed by Granger and Weiss (1983), Engle and Granger (1987), plays a crucial role in the modern analysis of economic time series. Many economic models suggest that certain variables, which may be highly persistent or even nonstationary, obey an equilibrium relationship. This relationship is not expected to hold exactly at any particular point in time, but deviations from it trigger mechanisms that drive the economy back towards the equilibrium state. For a $q \times 1$ column vector $Z_t = (z_{1t}, \ldots, z_{qt})'$, $t \in \mathbb{Z}$, suppose that the the linear constraint

$$\alpha' Z_t = 0 \tag{2.22}$$

represents the equilibrium relationship. As stated before, (2.22) is not expected to hold exactly, but one would expect the deviations from this equilibrium state, say

$$u_t = \alpha' Z_t, \tag{2.23}$$

to be small in some sense, and to revert to zero relatively quickly. In particular, for the concept of equilibrium to hold any value in the description of the underlying processes, one would expect u_t not to share the persistence and/or nonstationarity of the original processes Z_t.

While Engle and Granger (1987) (and subsequent authors, see e.g. Phillips and Durlauf, 1986; Stock, 1987; Johansen, 1988) focused much of their attentions on the case of unit root Z_t and short memory u_t, they did note that their results can be generalised to (stationary or nonstationary) fractionally integrated processes. We will frame their definition in the case of interest for this thesis, where both the observables Z_t and the errors u_t are stationary fractionally integrated $I(d)$

processes, as introduced in the previous section.

Definition 10 *(Engle and Granger, 1987) Z_t is said to be cointegrated of orders d and b, denoted $CI(d,b)$, if:*

(i) all the components of Z_t are $I(d)$,

(ii) there exists a $q \times 1$ vector $\alpha \neq 0$ such that (2.23) holds, where u_t is an $I(d-b)$ process with $0 < b \leq d$.

The (usually unknown) α is called a cointegrating vector, and the equilibrium error, u_t, is called a cointegrating residual. To highlight the distinction between the traditional literature on cointegration and the analysis of long memory processes, when d and/or b are not integers Z_t is said to be fractionally cointegrated. In the context of fractional cointegration, it is possible to generate the traditional behaviour of nonstationary observables and stationary errors (when $d \geq \frac{1}{2}$ and $d - b < \frac{1}{2}$), but other combinations are possible, namely where observables and errors are all nonstationary ($d - b \geq \frac{1}{2}$) or all stationary ($d < \frac{1}{2}$). We will focus on the latter case throughout the thesis.

This definition may be extended to allow the elements of Z_t to have differing memory parameters. We will adopt instead the alternative definition used by e.g. Robinson and Marinucci (2003) in the study of fractional cointegration of nonstationary Z_t; alternative definitions were reviewed and discussed by Robinson and Yajima (2002).

Definition 11 *(Robinson and Marinucci, 2003) Z_t is said to be (fractionally) cointegrated if:*

(i) z_{it} is $I(d_i)$, $0 < d_i < \frac{1}{2}$, $i = 1, \ldots, q > 1$,

(ii) there exists a $q \times 1$ vector $\alpha \neq 0$ such that (2.23) holds, where u_t is an $I(d_u)$ process with $0 \leq d_u < \min_i d_i$.

The existence of cross-correlation between the z_{it} is a necessary condition for cointegration, but we will presently avoid specifying the nature of this, except to note that by the Schwarz inequality the cross-spectral density at frequency λ between z_{it} and z_{jt} has modulus of order no greater than $|\lambda|^{-d_i-d_j}$ as $\lambda \to 0$. If $q = 2$, a necessary condition for fractional cointegration is that $d_1 = d_2$, so the two definitions are identical; however, this is not necessarily the case for $q > 2$, as we will illustrate below, and in the context of real valued memory parameters it may be unreasonable to make such restrictions.

Engle and Granger (1987) also noted that in the $q > 2$ case there may be multiple cointegrating vectors α satisfying the definition. The set of vectors α for which the (2.23) constitute cointegrating residuals is a linear subspace of $\mathbb{R}^q$, and can therefore be characterised by a base of linearly independent vectors. The dimension of this subspace, k, is called the cointegrating rank, and necessarily $k < q$. If $k = q$, then any vector α would be a cointegrating vector, including $\alpha = (1, 0, \ldots, 0)$, which would yield $u_t = z_{1t}$ and contradict the requirement that $d_u < d_1$.

The study of multivariate fractional cointegration involves delicate issues that are largely absent in the unit root case. Following work in the unit root framework (e.g. Stock and Watson, 1988), Robinson and Yajima (2002) suggested a representation of Z_t in terms of unobservable components, which is also relevant in the modelling of fractional cointegration. Whenever Z_t is cointegrated, it can be expressed as

$$Z_t = AF_t + U_t, \tag{2.24}$$

where F_t is a $J \times 1$ column vector of (unobservable) common components, for $J = q - k$, U_t is a q column vector of unobservable residuals, and A is a full rank $q \times J$ matrix of coefficients. Note that no row of A is identically zero, as that

would contradict our definition of cointegration, and denote by A_i the i-th column of A. Assume, for the time being, that in this representation F_t is a vector of (not cointegrated) $I(d)$ processes, while all the elements of U_t are $I(d_u)$, with $d > d_u$. It follows that all elements of Z_t are $I(d)$, the cointegrating residuals are $I(d_u)$, and the cointegrating vectors are given by the null space of A, i.e. the solutions to the equation $\alpha' A = 0$. This setting has been studied in great detail in the unit root framework, but here we need not assume that d, d_u are integers.

Now assume that instead F_{it} is d_i, $i = 1, \ldots, J$, and that $d_1 > d_2 > \ldots > d_J > d_u$. It is still true that the $(q - J)$-dimensional null space of A contains vectors for which (2.23) is a linear combination of U_t, and therefore $I(d_u)$. Still, other vectors may satisfy our definition of fractional cointegration. Using, for instance the partition $A = [A^* | A_J]$, where $A^* = [A_1 \ldots A_{J-1}]$ is a full rank $q \times (J-1)$ matrix, then any α in the $(q - J + 1)$-dimensional null space of A^* will generate residuals in (2.23) which are either $I(d_u)$ or $I(d_J)$. These will therefore be cointegrating residuals, provided that no row of A^* is identically zero, and thus all the observables are at least $I(d_{J-1})$. Following this argument to the extreme case, if all the elements of A_1 are non-zero, so that all the Z_t are $I(d_1)$, any α orthogonal to A_1 generates residuals that are (at most) $I(d_2)$. In this case, the cointegrating rank of Z_t is $q - 1$, and indeed one could let F_{it}, $i = 2, \ldots, J$, be "absorbed" into U_t to yield a single factor version of (2.24). If, on the other hand, some of the coefficients of A_1 are zero, then the corresponding elements of Z_t will have memory parameters lower than d_1, while still being part of meaningful cointegrating relationships. In this case (and also in the general case where some of the d_i may be equal and the elements of U_t may have different memory parameters), there may be cointegration between processes of different memory, and the cointegrating residuals themselves may have differing memory for alternative cointegrating vectors.

2.4 Nonlinear models of long memory

It is desirable to reconcile these properties of long memory and fractional cointegration with a more fundamental modelling of Z_t, which is plausible in financial series. Consider first a strictly stationary scalar process η_t, such that

$$z_t = g(\eta_t), \tag{2.25}$$

where $g(\cdot)$ is a scalar nonlinear function. Since our definition of long memory requires the existence of a spectral density, in what follows we will assume that both z_t and η_t have finite second moments. Furthermore, and without loss of generality, we will assume that both z_t and η_t are mean zero (otherwise subtract the means and redefine g accordingly). We are interested in inferring properties of z_t, in particular its memory parameter, in this setting. To do this, we must consider not only the memory parameter of the underlying η_t, but also the nature of the transformation g and the joint distributional properties of the $\{\eta_t\}$ process. There is a large literature on the asymptotic behaviour of partial sums of long memory processes such as (2.25) (see e.g. Taqqu, 1975; Taqqu, 1979; Dobrushin and Major, 1979; Breuer and Major, 1983; Ho and Sun, 1987) when $\{\eta_t\}$ is assumed to be Gaussian, enabling the use of a Hermite expansion.

Definition 12 *For integer $k \geq 0$, the Hermite polynomials $H_k(\cdot)$ are given by*

$$H_k(x)\phi(x) = (-1)^k \frac{\partial^k}{\partial x^k}\phi(x), \tag{2.26}$$

where $\phi(\cdot)$ denotes the standard normal density.

Assume $\{\eta_t\}$ is Gaussian, and furthermore (without loss of generality) that η_t is $N(0,1)$. The process $\{\eta_t\}$ is thus fully characterised by its autocorrelation

function $\gamma_t = E(\eta_0\eta_t)$, $t > 0$. For functions $g(\cdot)$ such that $E(z_t^2) < \infty$, we may express z_t as

$$z_t = \sum_{k=r}^{\infty} \frac{G_k}{k!} H_k(x), \tag{2.27}$$

where the Hermite coefficients G_k are given by

$$G_k = E[g(\eta_t)H_k(\eta_t)], \tag{2.28}$$

and the Hermite rank r is defined as

$$r = \min\{k \geq 0 : G_k \neq 0\}. \tag{2.29}$$

Note that $G_0 = E(z_t)$ is zero by assumption, and therefore in this setting $r \geq 1$.

Hermite polynomials constitute an orthogonal base (under the Gaussian probability measure) for the space of square integrable functions $g(\cdot)$, and furthermore it is well known that

$$E[H_j(\eta_0)H_k(\eta_t)] = k!\gamma_t^k 1(j = k), \qquad j, k, t \in \mathbb{Z}. \tag{2.30}$$

It follows that, if η_t is $I(d)$ for $d > 0$, and therefore its autocovariance satisfies (2.16), then

$$E[H_k(\eta_0)H_k(\eta_t)] \sim k!C_\gamma^k t^{k(2d-1)}, \qquad \text{as } t \to \infty. \tag{2.31}$$

Now, from (2.27) the autocovariance of z_t is

$$E(z_0 z_t) = \sum_{k=r}^{\infty} \left(\frac{G_k}{k!}\right)^2 E[H_k(\eta_0)H_k(\eta_t)] = \sum_{k=r}^{\infty} \frac{G_k^2}{k!}\gamma_t^k. \tag{2.32}$$

The contribution of the j-th term in the expansion is proportional to the j-th power of the autocorrelation of η_t. As $t \to \infty$, $\gamma_t \to 0$, and since $\sum_{k=r}^{\infty} G_k^2/k! =$

$E(z_t^2) < \infty$, the leading term of (2.32) can be shown to dominate the remaining ones. Therefore,

$$\text{Cov}(z_0, z_t) \sim Ct^{2d^*-1}, \qquad \text{as } t \to \infty, \tag{2.33}$$

for some finite constant C, where

$$d^* = \frac{1}{2} - r\left(\frac{1}{2} - d\right), \tag{2.34}$$

which follows trivially from setting $2d^* - 1 = r(2d-1)$. If $d^* > 0$, z_t is a long memory process with memory parameter d^*, while if $d^* < 0$ the autocovariance function is summable and thus z_t is short memory. The process z_t in (2.27) will share the memory of η_t if $G_1 = E[\eta_t g(\eta_t)] \neq 0$, say for $g(x) = e^x$. In the special case where $G_1 = 0$ (in particular, if g is symmetric around the origin), but $G_2 = E[(\eta_t^2 - 1)g(\eta_t)] \neq 0$ and $d > \frac{1}{4}$, z_t will be $I(2d - \frac{1}{2})$, and thus of lower memory than η_t; this would arise for instance if $g(x) = |x|^a$, $a \neq 0$. While theoretically possible, functions orthogonal to both H_1 and H_2 are less likely to be of practical interest.

Robinson (2001) provided a multivariate extension of (2.32), in the general setting where the quantity of interest is the covariance between two nonlinear functions of a multivariate Gaussian process. This expansion is then used to approximate the autocovariance function of models of the form $z_t = g_1(\eta_{1t})g_2(\eta_{2t})$. Models such as Taylor's (1986) SV model are nested by this form; it can therefore be seen as a generalised SV model if, say, $E[g_1(\eta_{1t})] = 0$, η_{1t} is iid, and η_{1t} is independent of past values of η_{2t}, so that z_t satisfies the martingale property. Moreover, Robinson (2001) noted that $|z_t|^a$ may also be represented within this family, but in this case $E[|g_1(\eta_{1t})|^a] > 0$, for any (non-trivial) g_1 and a, and thus $|z_t|^a$ may inherit some of the autocorrelation assumed for η_{2t}. In this general setting, the existence of long memory in z_t, and the actual value of its memory

parameter, will depend as above on the nature of g_1 and g_2, namely a multivariate extension of (2.29), as well as the memory parameters (and dependence) of η_{1t} and η_{2t}.

Most of the literature on nonlinear transformations of long memory time series assumes Gaussianity of the underlying processes which, as previously illustrated, greatly facilitates theoretical analysis. Nonetheless, some asymptotic theory has been developed where the Gaussian η_t is replaced by a linear process with non-Gaussian iid innovations, see e.g. Giraitis and Taqqu (1997) and Giraitis, Taqqu, and Terrin (1998). The non-Gaussian case is considerably more complicated to handle theoretically, and the added generality is offset by strong restrictions on g, typically assumed to be a finite order polynomial. For these reasons, we will assume Gaussianity of η_t wherever the use of Hermite expansions is called for, noting that the nonparametric nature of g adds enough flexibility to this specification to lead to an essentially unrestricted distribution for z_t.

2.5 Fractional cointegration in nonlinear models

It may be possible, further, to infer a cointegrating relation for Z_t from an underlying structural relation in terms of η_t. In the spirit of the previous section, consider a jointly strictly stationary $s \times 1$ vector process η_t, for $s \geq q$, such that

$$z_{it} = g_i(\eta_t), \qquad i = 1, \ldots, q, \tag{2.35}$$

where the g_i are nonlinear functions. We consider perhaps the simplest interesting case leading to a nonlinear cointegrating relationship. We take $q = 2$, $s = 4$, write $\eta_t = (\eta_{1t}, \eta_{2t}, \eta_{3t}, \eta_{4t})'$, and assume it is Gaussian. Suppose that the $\{\eta_{it}\}$ are mutually independent processes, that for $i = 1, 2, 3$ the η_{it} are iid with zero mean and variance σ_i^2, and that η_{4t} is an $I(d_4)$ process, for $d_4 > 0$. Suppose that

we observe sequences x_t, y_t, generated by

$$x_t = \beta_1 \zeta_t + \eta_{1t}, \tag{2.36}$$

$$y_t = \beta_2 \zeta_t + \eta_{2t}, \tag{2.37}$$

where $\beta_1, \beta_2 \neq 0$ and

$$\zeta_t = \eta_{3t} h(\eta_{4t}), \tag{2.38}$$

where h is a possibly nonlinear function, with $E\{h(\eta_{4t})^2\} < \infty$. The common factor ζ_t thus follows a SV model of the type discussed previously.

This setup can be interpreted as a factor model for asset returns, x_t and y_t, where ζ_t is the (unobservable) market return, and β_1, β_2 are the market risk exposures of x_t and y_t, respectively. Since memory properties (of volatilities, in this case) are invariant to temporal aggregation (see Chambers, 1998), (2.36)-(2.38) should be a reasonable model across all sampling frequencies. Now ζ_t is not an iid sequence but it is a square-integrable martingale difference, and thus uncorrelated, sequence, as therefore are x_t and y_t. Thus x_t and y_t exhibit an ideal property of asset returns, say. Because x_t and y_t are therefore $I(0)$ sequences, and all linear combinations of them are also $I(0)$, they are not cointegrated. However, we can deduce a cointegrating relation between the squares $z_{1t} = x_t^2$, $z_{2t} = y_t^2$. We have

$$\begin{aligned} z_{2t} &= (\beta_2 \zeta_t + \eta_{2t})^2 \\ &= \theta z_{1t} + u_t, \end{aligned} \tag{2.39}$$

where $\theta = \beta_2^2 / \beta_1^2$ and

$$u_t = \eta_{2t}^2 + 2\beta_2 \eta_{2t} \zeta_t - 2\theta \beta_1 \eta_{1t} \zeta_t - \theta \eta_{1t}^2. \tag{2.40}$$

Clearly u_t has no autocorrelation, and is thus an $I(0)$ process. We have

$$z_{1t} = \beta_1^2\zeta_t^2 + 2\beta_1\eta_{1t}\zeta_t + \eta_{1t}^2. \tag{2.41}$$

The last two terms on the right are also $I(0)$. However, for suitable h the leading term $\beta_1^2\zeta_t^2$ has long memory, and thence so has z_{1t}. Noting that

$$\zeta_t^2 = [\eta_{3t}^2 - E(\eta_{3t}^2)]h^2(\eta_{4t}) + E(\eta_{3t}^2)h^2(\eta_{4t}), \tag{2.42}$$

where the first term is a martingale difference sequence, the memory of $\beta_1^2\zeta_t^2$ is given by using the Hermite rank of h^2 in (2.34). For example if $\zeta_t = \eta_{3t}\eta_{4t}^2$ and $d_4 > \frac{1}{4}$, z_{1t} is $I(2d_4 - \frac{1}{2})$, or if $\zeta_t = \eta_{3t}e^{\eta_{4t}}$, z_{1t} is $I(d_4)$. In either case, z_{2t} has the same memory parameter as z_{1t}, and $Z_t = (z_{1t}, z_{2t})'$ is fractionally cointegrated, with cointegrating vector $\alpha = (-\theta, 1)'$. A similar conclusion is drawn if, even more simply, η_{1t} is missing from (2.36), in which case $z_{1t} = \theta_1^2\zeta_t^2$ and $u_t = \eta_{2t}^2 + 2\beta_2\eta_{2t}\zeta_t$.

Recall that ζ_t is generated by a SV model and plays the role of a common factor. Fractional cointegration can also arise if the idiosyncratic components η_{1t} and/or η_{2t} are themselves replaced by processes with SV, allowing u_t to have long memory if it is still dominated by ζ_t^2. This situation will be fully explored in the following chapter.

Though (2.39) is expressed in the form of a regression model, it does not possess the classical properties. The unobservable sequence u_t actually has nonzero mean (as does z_{1t}), but this situation is rectified by introducing an intercept. More importantly, however, u_t is not orthogonal to the right-hand side observable z_{1t}:

$$\mathrm{Cov}(z_{1t}, u_t) = -2\theta\sigma_1^2\left\{\sigma_1^2 + 2E(\zeta_t^2)\right\} < 0, \tag{2.43}$$

taking $\beta_1 = 1$ with no loss of generality, since the scale of the unobservable ζ_t is

indetermined in (2.36), (2.37).

The absence of η_{1t} in this simple setting would ensure orthogonality (though not independence) between z_{1t} and u_t, from which classical linear regression assumptions would trivially follow. However, such a simplifying assumption is not realistic in a situation where the designation of left-hand variable is arbitrary. For general q, after rewriting $\alpha' Z_t = u_t$ in regression form, then even in the absence of an underlying structure like (2.36), (2.37) there is no reason to suppose that orthogonality between cointegrating errors and right-hand side regressors obtains.

Alternatively, it would be possible to specify SV models where issues associated with nonlinearity can be avoided. Deo and Hurvich (2001), Hurvich and Soulier (2002), Hurvich and Ray (2003), Arteche (2004), and Hurvich, Moulines, and Soulier (2005) discussed semiparametric estimation and testing of the memory parameter under the long memory SV model of Harvey (1998) and Breidt, Crato, and de Lima (1998). Following these authors, if instead of (2.36)-(2.38) we assume, say,

$$x_t = h_1(\beta_1 \eta_{1t} + \eta_{2t}), \tag{2.44}$$

$$y_t = h_2(\beta_2 \eta_{1t} + \eta_{3t}), \tag{2.45}$$

for known and invertible functions h_1, h_2, and furthermore that the memory of η_{2t}, η_{3t} is dominated by that of η_{1t}, then a fractional cointegrating relationship between $z_{1t} = h_1^{-1}(x_t)$ and $z_{2t} = h_2^{-1}(y_t)$ could be explored by imposing suitable linearity assumptions on η_t. In particular, such a relationship would follow (for $h_1(x) = h_2(x) = \exp(x)$) if the long memory SV model held for both sequences, with linearly related long memory components (yielding $I(0)$ η_{2t} and η_{3t}). However, in a general setting where h_1, h_2 are unknown or non-monotonic, or if additive errors are present as in (2.36), (2.37), no such "linearising" transformation can be

assumed to exist, and nonlinearity needs to be accounted for explicitly.

2.6 Estimation of cointegrating vectors

Assuming the existence of a cointegrating relationship such as (2.23), and furthermore that the q-th element of α is non-zero (adopting an arbitrary normalization), so that we may designate z_{qt} as the left-hand side variable, rewrite the cointegrating relation as

$$Y_t = \theta' X_t + u_t, \tag{2.46}$$

where $Y_t = z_{qt}$, $X_t = (z_{1t}, \ldots, z_{q-1,t})'$ and θ is a $(q-1) \times 1$ vector. It is desired to estimate the unknown $\theta = (\theta_1, \ldots, \theta_{q-1})'$, on the basis of observables Z_t, $t = 1, \ldots, n$.

The most obvious estimate of θ is OLS with intercept correction (bearing in mind that u_t may have non-zero mean, as the discussion of the previous section suggests). This is

$$\hat{\theta}_O = \left\{ \sum_{t=1}^{n} (X_t - \bar{X}) X_t' \right\}^{-1} \sum_{t=1}^{n} (X_t - \bar{X}) Y_t, \tag{2.47}$$

where $\bar{X} = n^{-1} \sum_{t=1}^{n} X_t$. However, the correlation envisaged between u_t and X_t makes $\hat{\theta}_O$ inconsistent for θ. We can write

$$\hat{\theta}_O - \theta = \left\{ \sum_{t=1}^{n} (X_t - \bar{X}) X_t' \right\}^{-1} \sum_{t=1}^{n} (X_t - \bar{X}) u_t, \tag{2.48}$$

where stationarity of X_t and u_t implies (under mild additional assumptions) that $n^{-1} \sum_{t=1}^{n} (X_t - \bar{X}) X_t'$ will converge in probability to a constant positive definite covariance matrix, while $n^{-1} \sum_{t=1}^{n} (X_t - \bar{X}) u_t$ does not converge to zero unless u_t and X_t are assumed orthogonal. This outcome differs from the familiar one in

which Z_t has a unit root and u_t is $I(0)$. In this case, suitably scaled versions of the sample moments, namely $n^{-2}\sum_{t=1}^{n}(X_t-\bar{X})X_t'$ and $n^{-1}\sum_{t=1}^{n}(X_t-\bar{X})u_t$, converge in distribution to random matrices that can be expressed as certain functionals of Brownian motions (see e.g. Phillips and Durlauf, 1986; Stock, 1987). The asymptotic dominance of sums of squares of u_t by those of X_t, embedded in the different scaling factors above, overwhelms the simultaneous equation bias, leading to n-consistency of $\hat{\theta}_O$.

Robinson (1994a) proposed an estimate of θ that, under linearity assumptions, achieves consistency. Before introducing this estimate, some additional notation is required. For a vector sequence a_t, define the discrete Fourier transform

$$w_a(\lambda) = (2\pi n)^{-\frac{1}{2}}\sum_{t=1}^{n} a_t e^{it\lambda}, \tag{2.49}$$

and for a vector sequence b_t, possibly the same as a_t, define the (cross-) periodogram matrix

$$I_{ab}(\lambda) = w_a(\lambda)w_b'(-\lambda), \tag{2.50}$$

and the averaged (cross-) periodogram

$$\hat{F}_{ab}(\lambda_m) = \frac{2\pi}{n}\sum_{j=1}^{m} I_{ab}(\lambda_j), \tag{2.51}$$

where $\lambda_j = 2\pi j/n$ are the Fourier frequencies. We can interpret the averaged cross-periodogram as the portion of the sample covariance of a_t and b_t pertaining to frequencies up to λ_m, in much the same way as the integrated spectrum relates to the population variance. In particular, note that by orthogonality of the complex exponential

$$\hat{F}_{ab}(\lambda_{n-1}) = n^{-1}\sum_{t=1}^{n}(a_t-\bar{a})b_t, \tag{2.52}$$

where the mean correction arises from the exclusion of frequency $\lambda_0 = 0$ in (2.51),

as it can be easily seen that $2\pi n^{-1} I_{ab}(0) = \bar{a}\bar{b}$.

Hannan (1963) first introduced "band spectrum regression," later developed by Engle (1974), wherein the traditional time domain OLS specification in (2.47) is replaced by frequency domain analogues of the sample moments. While those authors focused their attentions on a fixed, nondegenerate band of frequencies, Robinson (1994a) followed a similar approach but considered instead a degenerating band of frequencies around zero. For a sequence of bandwidths, $m = m(n)$, such that

$$m \leq \frac{n}{2}, \tag{2.53}$$

$$\frac{1}{m} + \frac{m}{n} \to 0, \qquad \text{as } n \to \infty, \tag{2.54}$$

define the narrow-band least squares (NBLS) estimate of β,

$$\hat{\theta}_{NB} = \operatorname{Re}\left\{\hat{F}_{XX}(\lambda_m)\right\}^{-1} \operatorname{Re}\left\{\hat{F}_{XY}(\lambda_m)\right\}, \tag{2.55}$$

where $\operatorname{Re}(\cdot)$ is the real part operator. Note that if in contrast to (2.53), (2.54), we have instead $m = n-1$, then (2.52) yields $\hat{\theta}_{NB} = \hat{\theta}_O$. However, condition (2.54) is crucial to the consistency of $\hat{\theta}_{NB}$. The basic intuition for consistency is as follows. By linearity of the Fourier transform, $\hat{F}_{XY}(\lambda_m) = \hat{F}_{XX}(\lambda_m)\theta + \hat{F}_{Xu}(\lambda_m)$, which allows the estimation error of $\hat{\theta}_{NB}$ to be expressed as

$$\hat{\theta}_{NB} - \theta = \operatorname{Re}\left\{\hat{F}_{XX}(\lambda_m)\right\}^{-1} \operatorname{Re}\left\{\hat{F}_{Xu}(\lambda_m)\right\}. \tag{2.56}$$

For $i = 1, \ldots, q-1$, each component of the latter term can be bounded by the Cauchy inequality,

$$\operatorname{Re}\left\{\hat{F}_{z_i u}(\lambda_m)\right\} \leq \left\{\hat{F}_{z_i z_i}(\lambda_m)\hat{F}_{uu}(\lambda_m)\right\}^{\frac{1}{2}}, \tag{2.57}$$

and under suitable conditions (ensuring convergence of the averaged periodogram to the corresponding integrated spectrum) the right-hand side is

$$O_e\left(\left(\int_0^{\lambda_m}\lambda^{-2d_i}d\lambda\int_0^{\lambda_m}\lambda^{-2d_u}d\lambda\right)^{\frac{1}{2}}\right)=O_e\left(\left(\frac{m}{n}\right)^{1-d_i-d_u}\right), \tag{2.58}$$

where $O_e(\cdot)$ indicates an exact order of magnitude. On the other hand, under suitable conditions, for $\Lambda_m=\mathrm{diag}\{\lambda_m^{d_1},\ldots,\lambda_m^{d_{q-1}}\}$,

$$\lambda_m^{-1}\Lambda_m\,\mathrm{Re}\left\{\hat{F}_{XX}(\lambda_m)\right\}\Lambda_m\to_p\Omega, \tag{2.59}$$

where Ω is a constant positive definite matrix. It follows that

$$\hat{\theta}_{NB,i}-\theta_i=O_p\left(\left(\frac{m}{n}\right)^{d_i-d_u}\right),\qquad i=1,\ldots,q-1, \tag{2.60}$$

where θ_i and $\hat{\theta}_{NB,i}$ are the i-th elements of θ and $\hat{\theta}_{NB}$ respectively. Since cointegration entails $d_u<d_i$, $i=1,\ldots,q-1$, (2.54) ensures that $\hat{\theta}_{NB}$ is consistent for θ. This result is intuitively comparable to the unit root case discussed above: while stationarity implies that (full-band) sums of squares of u_t and $z_{1t},\ldots,z_{q-1,t}$ diverge at the same rate, the spectral densities of $z_{1t},\ldots,z_{q-1,t}$ dominate that of u_t in the neighbourhood of zero, and such dominance overwhelms any potential simultaneous equation bias (whose contribution is bounded by (2.57)).

Consistency of $\hat{\theta}_{NB}$ was first shown by Robinson (1994a) in case $q=2$ and by Lobato (1997) for general q, while Robinson and Marinucci (2003) established the rate in (2.60), also for general q. The conditions they imposed to deduce the crucial properties (2.58) and (2.59) were that Z_t is generated by a linear moving average in conditionally homoscedastic martingale differences, as in Assumption 2.3. As previously noted, this is inconsistent with the SV setup (2.36)-(2.38) illustrated in the previous section, albeit compatible with the more restrictive model (2.44),

(2.45). The estimate $\hat{\theta}_{NB}$ is desirably computationally simple, it does not involve estimation of ancillary quantities such as memory parameters, and the exclusion of high-frequency contributions makes this estimate robust to contamination by short run dynamics, such as those introduced by microstructure noise. It has been applied in fractional cointegration analysis of implied and realised volatility by Christensen and Nielsen (2006), Bandi and Perron (2006).

In general the rate in (2.60) is sharp, and indeed under additional conditions it seems that, for each i, $\lambda_m^{d_u-d_i}(\hat{\theta}_{NBi} - \theta_i)$ converges in distribution not to a non-degenerate random variable, but to a constant. This is due to the presumed coherence between X_t and u_t around zero frequency, under which the bound in (2.57) cannot be improved. Without such coherence, asymptotic normality and a faster rate of convergence are possible. Christensen and Nielsen (2006) supposed that the cross-spectral density between z_{it} and u_t is $o(|\lambda|^{-d_i-d_u})$, as $\lambda \to 0$, rather than having real part behaving precisely like $|\lambda|^{-d_i-d_u}$. Assuming also that $d_i + d_u < \frac{1}{2}$, $i = 1, \ldots, q-1$, they deduced that $m^{\frac{1}{2}}\lambda_m^{d_u}\Lambda_m^{-1}(\hat{\theta}_{NB} - \theta)$ is asymptotically multivariate normal; they assumed Z_t is linear in homoscedastic martingale differences, as in Robinson (1994a), Robinson and Marinucci (2003).

Though the model constructed in the previous section, (2.39) based on (2.36)-(2.38) and $z_{1t} = X_t = x_t^2$, $z_{2t} = Y_t = y_t^2$, does not satisfy the linearity assumption of Christensen and Nielsen (2006), it does satisfy a lack-of-coherence assumption that corresponds to theirs. It is easily seen that $\mathrm{Cov}(z_0, u_t) = 0$ if $t \neq 0$, so in view of (2.41), the cross-spectral density of z_{1t} and u_t is finite and constant, and $o(|\lambda|^{-\delta})$, where $\delta > 0$ represents the memory parameter of z_{1t}. (In the cases discussed after (2.41), the possibilities that $\delta = d_4$ and $\delta = 2d_4 - \frac{1}{2}$ emerged.)

Violation of orthogonality represents an important way in which (2.46) disobeys classical regression conditions, but it is not the only one. Though the simple set-up with $q = 2$ analysed in the previous section ensured that u_t has no

autocorrelation (see (2.40)), more generally u_t can be not only autocorrelated but even have long memory, as implied by Robinson (2001). In the absence of simultaneous equations bias, a suitable weighted frequency domain estimate will be more efficient. In (2.46) with short memory u_t orthogonal to X_t, Hannan (1963) showed that weighting inversely with respect to a nonparametric estimate of f_u can achieve the same asymptotic efficiency as generalised least squares based on a correctly specified parametric model for f_u. Robinson and Hidalgo (1997) derived a central limit theorem for long memory u_t, under the assumption of a known parametric form for f_u (allowing for estimation of d_u and other parameters), while Hidalgo and Robinson (2002) extended these findings by considering both long memory u_t and nonparametric estimation of f_u. These results are obtained under orthogonality of errors and regressors, which we have previously argued to be an undesirable assumption in a cointegration context, and otherwise Hidalgo and Robinson (2002) noted that these "full-band" estimates will incur similar simultaneous equations bias and inconsistency to $\hat{\theta}_O$.

Nevertheless, it is worth considering whether some such weighting can improve on $\hat{\theta}_{NB}$, since f_u changes even over the interval $[\lambda_1, \lambda_m]$. Smith and Chen (1996) proposed the weighted narrow-band least squares (WNBLS) estimate

$$\hat{\theta}_{WNB} = \tilde{\theta}(\hat{d}_u), \tag{2.61}$$

where

$$\tilde{\theta}(d) = \left(\sum_{j=1}^{m} \lambda_j^{2d} \operatorname{Re}\{I_{XX}(\lambda_j)\} \right)^{-1} \sum_{j=1}^{m} \lambda_j^{2d} \operatorname{Re}\{I_{XY}(\lambda_j)\}, \tag{2.62}$$

and $\hat{d}_u$ is a consistent estimate of d_u (some popular choices of $\hat{d}_u$ are discussed in the following section). Note that $\tilde{\theta}(0) = \hat{\theta}_{NB}$, while otherwise the averaged periodogram in (2.55) is replaced by a weighted version thereof. Smith and Chen (1996) in fact proposed $\hat{\theta}_{WNB}$ in a more traditional regression setting, with u_t or-

thogonal to X_t, and did not establish any asymptotic properties. Recently, Nielsen (2005), under the same kind of incoherence-near-zero assumption as Christensen and Nielsen (2006), established asymptotic normality of $m^{\frac{1}{2}}\lambda_m^{d_u}\Lambda_m^{-1}(\tilde{\theta}(d)-\theta)$, for a fixed d satisfying

$$\max_{1\le i<q}\frac{d_i+d_u}{2}-\frac{1}{4}<d\le d_u, \tag{2.63}$$

and furthermore that the same result holds for the feasible (2.61), provided that (as in Robinson, 1994a)

$$(\log n)(\hat{d}_u-d_u)\to_p 0, \qquad \text{as } n\to\infty, \tag{2.64}$$

which can readily be justified in view of asymptotic theory for various memory parameter estimates. Nielsen (2005) also discussed the relative efficiency of $\tilde{\theta}(d)$ and $\hat{\theta}_{NB}$, noting some circumstances in which $\tilde{\theta}(d)$ can be the more efficient even when $d\neq d_u$.

Still, d_u is clearly an optimal choice of d, and given that d_u is unknown it is natural to focus on $\hat{\theta}_{WNB}$ which, like $\hat{\theta}_{NB}$, should still be consistent in the presence of coherence between u_t and X_t, violating Nielsen's (2005) condition. We have, say,

$$\left|\sum_{j=1}^{m}\lambda_j^{2\hat{d}_u}\,\mathrm{Re}\,\{I_{z_iu}(\lambda_j)\}\right|\le\left\{\sum_{j=1}^{m}\lambda_j^{2\hat{d}_u}I_{z_iz_i}(\lambda_j)\sum_{j=1}^{m}\lambda_j^{2\hat{d}_u}I_{uu}(\lambda_j)\right\}^{\frac{1}{2}}. \tag{2.65}$$

Under (2.64),

$$\lambda_j^{2\hat{d}_u}=\lambda_j^{2d_u}\lambda_j^{2(\hat{d}_u-d_u)}\le\lambda_j^{2d_u}n^{o((\log n)^{-1})}\le\lambda_j^{2d_u}e^{o(1)}\le 2\lambda_j^{2d_u}, \tag{2.66}$$

for n sufficiently large. It is then readily seen, under suitable conditions, that the

right-hand side of (2.65) is

$$O_e\left(n\left\{\int_0^{\lambda_m}\lambda^{2(d_u-d_i)}d\lambda\lambda_m\right\}^{\frac{1}{2}}\right)=O_e\left(n\left(\frac{m}{n}\right)^{1+d_u-d_i}\right). \tag{2.67}$$

Also under (2.64), and similar conditions to those giving (2.59), we can justify the step

$$\frac{1}{m}\lambda_m^{-2d_u}\Lambda_m\sum_{j=1}^{m}\left(\lambda_j^{2\hat{d}_u}-\lambda_j^{2d_u}\right)\mathrm{Re}\left\{I_{XX}(\lambda_j)\right\}\Lambda_m\to_p 0, \tag{2.68}$$

and then that

$$\frac{1}{m}\lambda_m^{-2d_u}\Lambda_m\sum_{j=1}^{m}\lambda_j^{2d_u}\,\mathrm{Re}\left\{I_{XX}(\lambda_j)\right\}\Lambda_m \tag{2.69}$$

converges in probability to a constant positive definite matrix.

Notice that in the model (2.39) derived from (2.36)-(2.38), $d_u=0$ so we expect no improvement of $\hat{\theta}_{WNB}$ over $\hat{\theta}_{NB}$. However, as mentioned before this model can be extended to allow at the same time $d_u>0$ and incoherence at frequency zero between regressors and errors, namely by replacing η_{1t}, η_{2t} in (2.36), (2.37) by SV processes comparable to (2.38).

At least for linear processes, bias and autocorrelation can be corrected simultaneously by more elaborate methods. These are based on a full system of q equations that expresses also the long memory properties of the z_{it}, $i=1,\ldots,q-1$, and lead to estimates of θ which depend not only on $\hat{d}_u$, but also on estimates of the d_i, $i=1,\ldots,q-1$. Such estimates of θ were developed by Hualde and Robinson (2007); they are asymptotically normal (centered at θ) with the same rate as described for $\hat{\theta}_{NB}$ and $\hat{\theta}_{WNB}$ under the incoherence-near-zero assumption, but without imposing that. This thesis focuses mainly on the "single-equation" estimates (based on (2.46)) we have discussed above, partly due to their computational simplicity, but also because incoherence-near-zero can often be justified in a factor model context, as discussed above, whence $\hat{\theta}_{NB}$ and $\hat{\theta}_{WNB}$ enjoy a

reasonably fast rate of convergence.

2.7 Estimation of memory parameters

Even if simple estimates of θ are used, there may be interest in estimation of the d_i, as well as in estimation of d_u, as is required for $\hat{\theta}_{WNB}$. In particular, such estimates are useful in determining the existence and extent of cointegration, as described by Robinson and Yajima (2002). In this multivariate setting, efficiency gains are possible by estimating memory parameters jointly, especially if prior equality constraints are placed on the d_i. However, joint estimates have principally been developed under the assumption of no cointegration (e.g. Robinson, 1995a; Lobato, 1999), and if there is cointegration they are liable to be inconsistent; see, however, Velasco (2003) for an application of Lobato's (1999) procedure to joint estimation of the memory parameters of observables and error in a bivariate cointegrating relation. Thus we briefly describe some leading univariate semiparametric estimates. We introduce a generic univariate stationary process v_t, which can represent any of the z_{it} or, where estimation of d_u is concerned, residuals $y_t - \tilde{\theta}' X_t$, such that $\tilde{\theta}$ represents one of our consistent estimates of θ. Denote by d the unknown memory parameter of v_t (or d_u when v_t represents regression residuals).

Geweke and Porter-Hudak (1983) proposed an estimate of d with an intuitive linear regression interpretation. Taking logs of (2.18), we obtain

$$\ln f(\lambda) \sim c - 2d \ln \lambda, \qquad \text{as } \lambda \to 0^+, \tag{2.70}$$

for $c = \ln C_f$. Geweke and Porter-Hudak (1983) originally considered a different approximation, where λ is replaced by $2\sin(\lambda/2)$. The simplification above, due to Robinson (1995a), is easily shown to share the same properties. Intuitively,

there is an approximate linear relationship between $\ln f(\lambda)$ and $-2\ln\lambda$ in the neighbourhood of zero, and the slope coefficient in this relationship is d. Replacing $f(\lambda)$ by $I_{vv}(\lambda)$, and evaluating (2.70) at Fourier frequencies only, then

$$\ln I_{vv}(\lambda_j) = c - 2d\ln\lambda_j + U_j, \tag{2.71}$$

where the errors U_j contain both approximation errors in (2.70) and estimation errors, which can be represented as logs of the normalised periodograms, $\ln\{I_{vv}(\lambda_j)/f(\lambda_j)\}$. Geweke and Porter-Hudak (1983) suggested interpreting (2.70) as a traditional linear regression, where the observations are the logged periodograms within a vanishing neighbourhood of zero, and using the OLS slope as an estimate for d. Defining sequences of trimming numbers $l = l(n)$ and bandwidths $m = m(n)$, where the latter satisfy at least (2.53), (2.54), and denoting $a_j = -2\ln\lambda_j$, $\bar{a} = (m-l+1)^{-1}\sum_{j=l}^{m} a_j$, we may write this "log-periodogram regression" estimate as

$$\hat{d}_{LP} = \frac{\sum_{j=l}^{m}(a_j - \bar{a})\ln I_{vv}(\lambda_j)}{\sum_{j=l}^{m}(a_j - \bar{a})^2}. \tag{2.72}$$

The suggested interpretation of (2.71) as a linear regression model may be deceptively simple, as establishing properties for the U_j sequence is not trivial. In addition to having non-zero mean, which can be easily corrected by suitably redefining the intercept c, the well known asymptotic unbiasedness and orthogonality of the periodogram at Fourier frequencies are invalidated by the presence of long memory (see Künsch, 1986), and thus, even asymptotically, the normalised periodograms cannot be treated as an iid sequence. Still, Robinson (1995a) established bounds on the magnitude of these effects that tighten as j increases, allowing him to prove consistency and asymptotic normality of OLS for slowly diverging l and m. These are assumed to satisfy a suitable condition on their growth rates, which implies

in particular that $m = o(n^{4/5})$, while the conditions on l are satisfied for $l = m^b$, $\frac{1}{2} < b < 1$. Assuming further that v_t is Gaussian, Robinson (1995a) showed that

$$m^{\frac{1}{2}}(\hat{d}_{LP} - d) \to_d N\left(0, \frac{\pi^2}{6}\right), \qquad \text{as } n \to \infty. \tag{2.73}$$

Hurvich, Deo, and Brodsky (1998) established consistency and the same limiting distribution under different assumptions and without the need for trimming, i.e. setting $l = 1$ in (2.72); Velasco (2000) reached the same results, for a tapered version of $\hat{d}_{LP}$, replacing the Gaussian v_t by a stationary linear process in non-Gaussian iid innovations, as in Assumption 2.2.

An efficiency improvement is possible, for the same m sequence, via the "local Whittle" estimate of Künsch (1987). This estimate maximises the Whittle approximation to the Gaussian likelihood, using the semiparametric model (2.18) over a vanishing neighbourhood of frequency zero:

$$(\hat{d}_{LW}, \hat{C}_f) = \arg \min_{d \in \mathcal{D}, C_f \in \mathcal{C}} \sum_{j=1}^{m} \left\{ \frac{I_{vv}(\lambda_j)}{C_f \lambda_j^{-2d}} + \ln(C_f \lambda_j^{-2d}) \right\}, \tag{2.74}$$

where $\mathcal{D}$ and $\mathcal{C}$ are compact subsets of $(-\frac{1}{2}, \frac{1}{2})$ and $(0, \infty)$ respectively. Concentrating out $\hat{C}_f = m^{-1} \sum_{j=1}^{m} \lambda_j^{2d} I_{vv}(\lambda_j)$, the estimate for d is

$$\hat{d}_{LW} = \arg \min_{d \in \mathcal{D}} \left\{ \ln \left(\sum_{j=1}^{m} \lambda_j^{2d} I_{vv}(\lambda_j) \right) - 2dm^{-1} \sum_{j=1}^{m} \ln \lambda_j \right\}. \tag{2.75}$$

This was shown by Robinson (1995b) to satisfy

$$m^{\frac{1}{2}}(\hat{d}_{LW} - d) \to_d N\left(0, \frac{1}{4}\right), \qquad \text{as } n \to \infty, \tag{2.76}$$

under a condition on the bandwidth implying $m = o(n^{4/5})$. Even though (2.74) is an approximation to the Gaussian likelihood, asymptotic theory does not re-

quire the assumption of Gaussianity, and is indeed established under variations of Assumption 2.3. Consistency is proved for linear processes in conditionally homoskedastic martingale differences, while the central limit theorem requires in addition finite and constant conditional skewness and unconditional kurtosis.

Another approach to estimation of the memory parameter relies directly on properties of the averaged periodogram (2.51). Under the assumption that v_t is a linear process in conditionally homoskedastic martingale differences, Robinson (1994a) showed that

$$\frac{\hat{F}_{vv}(\lambda_m)}{F_{vv}(\lambda_m)} \to_p 1, \qquad \text{as } n \to \infty, \tag{2.77}$$

where $F_{vv}(\lambda) = \int_0^\lambda f(x)dx$. From (2.18),

$$F_{vv}(\lambda) \sim C_f(1-2d)^{-1}\lambda^{1-2d}, \qquad \text{as } \lambda \to 0^+, \tag{2.78}$$

and hence, for any $q > 0$,

$$\frac{F_{vv}(q\lambda)}{F_{vv}(\lambda)} \to q^{1-2d}, \qquad \text{as } \lambda \to 0^+, \tag{2.79}$$

which does not depend on any unknown parameters other than d. Solving for d and replacing the integrated spectrum with the averaged periodogram, which is justified by (2.77), Robinson (1994a) proposed the estimate

$$\hat{d}_{AP} = \frac{1}{2} - \frac{\ln \hat{F}_{vv}(q\lambda_m) - \ln \hat{F}_{vv}(\lambda_m)}{2 \ln q}, \tag{2.80}$$

where $0 < q < 1$ is imposed without loss of generality, noting that $\hat{d}_{AP}$ remains identical when (q, m) is replaced by (q^{-1}, qm), and is undefined for $q = 1$. As Robinson (1994a) noted, consistency of $\hat{d}_{AP}$ is a direct consequence of (2.77) and (2.78). Under Gaussianity and for $m = o(n^{4/5})$, Lobato and Robinson (1996) established limiting distributions for $\hat{d}_{AP}$, which unlike the previous two estimates

crucially depend on d. For $d < \frac{1}{4}$,

$$m^{\frac{1}{2}}(\hat{d}_{AP} - d) \to_d N\left(0, \frac{(1 - 2q^{-2d} + q^{-1})\left(\frac{1}{2} - d\right)^2}{(\ln q)^2(1 - 4d)}\right), \qquad \text{as } n \to \infty, \quad (2.81)$$

and furthermore they tabulated the minimum mean squared error (MSE) choice of q for different values of d. However, for $\frac{1}{4} < d < \frac{1}{2}$ (and under stronger regularity conditions) the limiting distribution is found to be neither normal nor centered around zero, rendering $\hat{d}_{AP}$ less useful for inference than the alternative estimates described above.

Various modifications, in particular bias corrections, have been proposed. The condition $m = o(n^{4/5})$ imposed in deriving (2.73), (2.76), (2.81) can be understood by noting that, under mild smoothness assumptions on (2.18), the leading bias term is $O_e(\lambda_m^2)$, while the asymptotic variance is $O_e(m^{-1})$. The minimum-MSE rate for the bandwidth (for which bias and variance decay at the same rate) is therefore $m = O_e(n^{4/5})$, in which case an asymptotic bias term will appear in the above distributions, while for faster rates asymptotic bias will dominate, and so scaled estimation errors will, in general, converge to a constant. The modifications discussed below ensure that, for some $r > 1$, the first $r - 1$ terms in the bias expansion vanish, so that the bias is now $O_e(\lambda_m^{2r})$, and therefore that the optimal rate becomes $m = O_e(n^{4r/(4r+1)})$. If sufficient smoothness is present in the spectral approximation (2.18), the convergence rate of $m^{\frac{1}{2}}$ may be improved until it is arbitrarily close to $n^{\frac{1}{2}}$. It should nonetheless be noted that, as a general feature of bias reduction methods, both asymptotic and finite sample variances tend to be inflated. This effect may be potentially so large that the advantages of the method are completely nullified for finite samples.

Robinson and Henry (2003) introduced a very flexible class of higher-order kernel M-estimates, which allows for: pooling of periodograms; replacing (2.18) with

an alternative specification, $C_f g(\lambda)^{-2d}$, such that $g(\lambda) \sim \lambda$; the choice of a suitable functional form for the first order conditions (nesting in particular both $\hat{d}_{LP}$ and $\hat{d}_{LW}$); and most importantly the filtering of the periodogram by higher-order kernels. Robinson and Henry (2003) noted that conditions used in asymptotic theory bear similarities to those used in spectral estimation of short memory processes, here imposed on the approximation errors $h(\lambda) = f(\lambda)C_f^{-1}g^{2d}(\lambda)$, and that higher-order kernels may explore assumed smoothness therein. Assuming Gaussianity, they heuristically derived formulae for asymptotic bias, variance, MSE, and corresponding optimal bandwidth of such estimates, and if h is assumed sufficiently smooth ($2r$-times continuously differentiable, for $r \geq 1$), then the use of rth-order kernel leads to the bias and optimal MSE reductions discussed above.

Another bias reduction method expresses (2.18) as

$$f(\lambda) = g(\lambda)|\lambda|^{-2d}, \tag{2.82}$$

where $g(\lambda)$ is an even function such that $0 < g(0) < \infty$, and refines the approximation by using the first $r-1 \geq 1$ terms of the Taylor expansion of $\ln g(\lambda)$ around zero, i.e.

$$\ln f(\lambda) \sim \ln g(0) - 2d \ln \lambda + \sum_{k=1}^{r-1} b_{2k}\lambda^{2k}, \qquad \text{as } \lambda \to 0^+, \tag{2.83}$$

for some coefficients $b_2, \ldots, b_{2r-2}$, where the odd powers are absent because $\ln g(\lambda)$ is even. This enhanced approximation may be used to derive corresponding versions of $\hat{d}_{LP}$, replacing (2.71) with a multiple regression containing $r-1$ additional regressors, and $\hat{d}_{LW}$, replacing $C_f \lambda_j^{-2d}$ in (2.74) by the approximation above. Assuming that $g(\lambda)$ is $2r$-times continuously differentiable, but otherwise under assumptions comparable to Robinson (1995a) and Robinson (1995b), respectively, consistency and asymptotic normality were established for the modified estimates by Andrews and Guggenberger (2003) (log-periodogram regression) and Andrews

and Sun (2004) (local Whittle). These authors focused on the optimal bandwidth introduced previously, $m = O_e(n^{4r/(4r+1)})$, and showed that the reduction in optimal MSE is obtained at the cost of a known multiplicative factor in the asymptotic variance, which increases with r.

Guggenberger and Sun (2006) proposed instead to reduce bias by computing either the local Whittle or the log-periodogram estimate at a finite grid of different bandwidths, and averaging these estimates using a particular set of weights, which are essentially discrete versions of the higher-order kernels used by Robinson and Henry (2003). They obtained rates of convergence (using optimal bandwidths) comparable to Andrews and Guggenberger (2003) and Andrews and Sun (2004), under the same assumptions, but with the potential for reducing asymptotic variance inflation. In theory, weights may be selected so that the asymptotic variance is actually smaller than those in (2.73), (2.76); however, the authors noted that in finite samples this may imply the use of large bandwidths that undermine the bias reduction properties of the modified estimate.

Note that the conditions imposed to deduce (2.73), (2.76), (2.81) do not cover the SV setup described in the previous section. However, as mentioned in the end of Section 2.4, some authors have examined the properties of memory estimates under a particular nonlinear model, inspired by the long memory SV model of Harvey (1998) and Breidt, Crato, and de Lima (1998). Assume

$$z_t = \eta_{1t} \exp \eta_{2t}/2, \tag{2.84}$$

where $\{\eta_{2t}\}$ is a long memory process, while $\{\eta_{1t}\}$ is iid and independent of $\{\eta_{2t}\}$. Defining $v_t = \ln z_t^2$, we have

$$v_t = \ln \eta_{1t}^2 + \eta_{2t}, \tag{2.85}$$

and thus estimation of the memory parameter of v_t can be interpreted more gen-

erally as a signal-plus-noise problem, where observables are the sum of a long memory process (on which appropriate linearity or Gaussianity assumptions may be imposed) and an iid sequence. For $\hat{d}_{LP}$ (with Gaussian η_{2t}), Deo and Hurvich (2001) and Hurvich and Soulier (2002) showed (2.73), while for $\hat{d}_{LW}$ (with η_{2t} linear in conditionally homoskedastic martingale differences) Arteche (2004) established (2.76). Although the limit distribution remains unchanged, it should be noted that the bounds on bandwidth growth, and thus on rate of convergence, are tighter in this case, and become stricter as $d \to 0$. The spectral density of v_t is simply the sum of the spectral densities of $\ln \eta_{1t}^2$ and η_{2t}, where the former will be constant, so we may write

$$f(\lambda) \sim C_f(a + \lambda^{-2d}), \qquad \text{as } \lambda \to 0^+. \tag{2.86}$$

For this reason, Hurvich and Ray (2003) and Hurvich, Moulines, and Soulier (2005) refined (2.74) to take the signal-plus-noise structure explicitly into account, leading to the following "modified local Whittle" estimate,

$$(\hat{d}_{MLW}, \hat{a}) = \arg \min_{(d,a)\in\mathcal{D}\times\mathcal{A}} \left\{ \ln \left(\sum_{j=1}^{m} \frac{I_{vv}(\lambda_j)}{a + \lambda_j^{-2d}} \right) + \frac{1}{m} \sum_{j=1}^{m} \ln(a + \lambda_j^{-2d}) \right\}, \tag{2.87}$$

where $\mathcal{D}$ and $\mathcal{A}$ are compact subsets of $(0, \frac{1}{2})$ and $[0, \infty)$ respectively. Under similar assumptions on η_{2t} as Robinson (1995b), and allowing $\ln \eta_{1t}^2$ to be a white noise process, Hurvich, Moulines, and Soulier (2005) established

$$m^{\frac{1}{2}}(\hat{d}_{MLW} - d) \to_d N\left(0, \frac{(1+d)^2}{16d^2}\right), \qquad \text{as } n \to \infty. \tag{2.88}$$

A comparable extension of the log-periodogram regression estimate, (2.71), to a signal-plus-noise setting had been proposed earlier by Sun and Phillips (2003). Following a similar approach to Andrews and Guggenberger (2003), they included in

the regression an additional term, $b\lambda^{2d}$, derived from a first order Taylor expansion of the approximation errors, $\ln\{\lambda^{2d}f(\lambda)\}$, under the assumption of a signal-plus-noise structure. They suggested performing a nonlinear regression for both d and b, where the latter can be interpreted as the noise-to-signal ratio at frequency zero. Assuming that signal and noise are independent and Gaussian, and restricting attention to the optimal bandwidth case, they showed joint asymptotic normality of the estimates, where the covariance matrix also depends on d.

Very little is known about the asymptotic properties of these semiparametric estimates in the absence of linearity. Recently, Dalla, Giraitis, and Hidalgo (2006) established general sufficient conditions for consistency of $\hat{d}_{LW}$ that do not require a linear process specification for v_t. Under (2.18), they show that $\hat{d}_{LW}$ will be consistent if an ergodicity condition on the "renormalised periodograms,"

$$m^{-1}\sum_{j=1}^{m}\frac{I_{vv}(\lambda_j)}{C_f\lambda^{-2d}} \to_p 1, \qquad \text{as } n \to \infty, \tag{2.89}$$

holds for every bandwidth sequence satisfying (2.54). Under mild additional conditions, they also establish a rate of convergence and provide an asymptotic expansion that may be used in deriving limiting distributions for specific models. These results are applied to a general signal-plus-noise setting, where they establish (2.76) for linear (in iid innovations) signal, under minimal conditions for the noise process (assumed only to be covariance stationary with less memory than the signal), and under no restriction on the dependence between the two. They also consider nonlinear transformations of Gaussian variables, of the type discussed in Section 2.4, establishing consistency of $\hat{d}_{LW}$ under a mild spectral smoothness assumption, and (2.76) under additional conditions including the Hermite rank being one. Finally, the two applications are combined to analyse powers of SV models such as (2.84), for Gaussian η_{2t}, which can be expressed as a signal-plus-

noise model with a nonlinear signal.

2.8 Final comments

Fractional cointegration analysis is increasingly found to be a promising tool for dimensionality reduction in financial time series. On the one hand, series of asset returns may have little autocorrelation, whereas instantaneous nonlinear functions, such as squares, can exhibit evidence of long memory. Considering series on several assets, it is possible that there exists a linear combination of the nonlinear functions that has shorter memory. Then there is said to be fractional cointegration. Note that here, as implied by many SV models, series are supposed to be stationary. By contrast, in traditional cointegration analysis of macroeconomic time series, levels are typically believed to be nonstationary with a unit root, and cointegration exists when there is a linear combination that is stationary (with short memory).

A variety of tools for analysing fractional cointegration in stationary series is becoming available. The main stress has been on semiparametric methods, such as those reviewed in this chapter. These avoid full parameterisation of autocorrelation, in favour of a local power law for the spectral density around zero frequency. Estimates of memory parameters can be rendered inconsistent by misspecification of short memory properties. Moreover, when the cointegrating relation is expressed in regression form, with one of the observables on the left-hand side, the other observables cannot plausibly be assumed orthogonal to the cointegrating errors. In a stationary environment, (full-band) time domain procedures that do not explicitly account for this dependence (such as least squares) will inconsistently estimate the cointegrating vector, while more elaborate methods that aim to correct it depend on preliminary estimates of various ancillary parameters. This

leads to a focus on methods based on a vanishing neighbourhood of zero in the frequency domain, such that the number, m, of Fourier frequencies used increases with sample size n, but more slowly. An undesirable consequence of this semi-parametric strategy is rates of convergence (in case of both memory parameters and cointegrating vector estimates) that are slower than would be possible in a fully parametric setting. However, parametric estimates of memory parameters and (due to stationarity) cointegrating vectors can only converge at rate $n^{\frac{1}{2}}$ (there is no super-consistency), and the slower rates of the semiparametric methods (depending on m) may be acceptable when n is very large indeed, as is the case with many financial time series.

Asymptotic theory for the semiparametric estimates has been developed mainly under the assumption that the innovations in the Wold representation of the process of interest are essentially conditionally homoscedastic martingale differences (see Robinson, 1994a; Lobato, 1997; Robinson and Marinucci, 2003; Nielsen, 2005; Christensen and Nielsen, 2006). These assumptions are justified if, for example, series are Gaussian, but are unfortunately implausible in this setting. Recall that in financial series the long memory property, and the possibility of fractional cointegration, has tended to emerge only for certain nonlinear functions, namely for measures of volatility such as squared, absolute, or log-squared asset returns.

It is possible (see e.g. Hurvich, Moulines, and Soulier, 2005) to specify SV models for which the log-squares transformation yields a linear representation, on which linear filter assumptions might be plausible. Note, however, that common factor structures in the levels often follow from behavioural foundations, as in the CAPM literature, while in volatilities they are typically used just as a convenient assumption for dimensionality-reduction. The presence of additive errors, which seems realistic, would render this type of "linearisation" impossible. Furthermore, these results crucially hinge on particular parametric specifications for the SV

model, which are proposed, at least in part, for reasons of technical convenience. As a result, linear-in-martingale-difference representations should not be assumed to necessarily hold for volatility measures. Models for them can be articulated, in terms of underlying independent and identically distributed (iid) sequences, say, but the nonlinearity makes derivation of asymptotic properties (already a delicate matter in the linear setting) extremely complicated and lengthy. Moreover, due to second order bias that affects some estimates, useful limit distribution theory is unavailable.

Chapter 3

Consistency of Narrow Band Least Squares in long memory stochastic volatility models

3.1 Introduction

This chapter presents a proof of consistency of NBLS for a bivariate factor model with a single unobservable common factor. It is shown that, under some conditions, persistence in higher moments can allow consistent estimation of the ratio of factor loadings. The bivariate setup is chosen for simplicity; the theoretical tools developed here will be applied to a multivariate model in Chapter 5.

We extend the model (2.36), (2.37) discussed in the previous chapter to allow for long memory in the idiosyncratic errors, as well as the factor. Suppose two observable scalar time series, y_t and x_t, $t \in \mathbb{Z}$, are generated by

$$y_t = \beta_1 \zeta_t + \varepsilon_t, \tag{3.1}$$

$$x_t = \beta_2 \zeta_t + \delta_t, \tag{3.2}$$

where β_1, β_2 are unknown, $\beta_2 \neq 0$, and ζ_t, ε_t, δ_t are unobservable stationary processes, generated by SV models such as (2.38). As we mentioned in Section 2.5, this may be interpreted as an asset pricing model, where y_t and x_t would be asset returns, ζ_t the (unobservable) market return, and β_1, β_2 the market risk exposures of y_t and x_t, respectively. Since the scale of ζ_t cannot be identified, we only aim to estimate $\beta = \beta_1/\beta_2$; equivalently, β_2 could be normalised to unity by suitably rescaling ζ_t. In the suggested interpretation, knowledge of the relative risk exposures of the assets would allow the researcher to compare (and reduce, if necessary) the total exposure to market risk of portfolios containing the two assets. In particular, a portfolio could be derived which completely hedges against the common source of risk.

The OLS estimate of y_t on x_t suffers from errors-in-variables inconsistency for β, as exemplified in a simpler setting by (2.43), due to the δ_t component in x_t. Indeed, our assumptions will imply that ζ_t, ε_t, δ_t are white noise sequences, so in no meaningful sense can (3.1), (3.2) be described as a cointegrating relation. However, ζ_t, ε_t, δ_t are not serially independent, but exhibit persistence in higher moments. In particular, for some integer $p > 1$, our assumptions imply that x_t^p and y_t^p are cointegrated long memory $I(d_1)$ processes, $0 < d_1 < 1/2$, with cointegrating coefficient $\theta = \beta^p$, and cointegrating errors are $I(d_u)$ for $0 \leq d_u < d_1$. Squares of asset returns are typically found to display the underlying persistence, so a cointegrating relationship of the type described in Section 2.5 could be present, with $p = 2$. Still, x_t^p and y_t^p are stationary, so the OLS estimate is inconsistent for θ, unlike under the traditional assumption of $I(1)$ observables and $I(0)$ cointegrating errors. The usual instrumental variables estimates employed in time series models (e.g. with x_{t-1}^p as instrument) will also be inconsistent here, as the assumed persistence renders all available instruments invalid.

We argued in Section 2.6 that the existing asymptotic theory for NBLS re-

quires linearity assumptions (in conditionally homoskedastic martingale differences) that are not compatible with this intrinsically nonlinear framework. Nevertheless, NBLS has been applied to financial data (see e.g. Christensen and Nielsen, 2006; Bandi and Perron, 2006), as have other models for fractional cointegration in volatility (such as the parametric FIGARCH model of Brunetti and Gilbert, 2000). The present chapter aims to fill this gap in the theoretical properties of the NBLS estimate of θ, by establishing consistency under more relevant assumptions. While a model such as (2.44), (2.45) would allow a "linearising" transformation, and therefore sidestep the problems associated with nonlinearity, our specification may be more realistic, as discussed in the previous chapter. Our approach allows the presence of additive errors, does not require a specific (or known) shape for the volatility function, and indeed allows that shape to vary between the common and idiosyncratic components.

A key component of the proof of consistency is an approximation for expectations of products of nonlinear functions of Gaussian processes (Theorem 3.1), which may be of independent interest and is presented in the following section. Section 3 describes the SV setting. Section 4 details the NBLS estimation procedure and our consistency result, which is proved in a series of propositions stated and proved in Appendix A, using lemmas in Appendix B, as well as Theorem 3.1. Section 5 contains concluding remarks.

3.2 Approximating cross-moments of nonlinear functions of Gaussian variables

With the objective of examining the memory of SV models similar to those introduced in Sections 2.4 and 2.5, Robinson (2001) established an asymptotic expansion for the covariance between nonlinear functions of multivariate normal

random vectors. Here we need a (non-trivial) extension to cross-moments of more than two real functions.

Recall the methodology introduced in Section 2.4, in particular Definition 12. For a function $f(\cdot)$ satisfying $\int_{\mathbb{R}} f^2(x)\phi(x)\mathrm{d}x < \infty$, define the j-th Hermite coefficient $G_j = \int_{\mathbb{R}} f(x)H_j(x)\phi(x)\mathrm{d}x$ and the Hermite rank $r = \min\{j \geq 0 : G_j \neq 0\}$. Define $P_q = \{i \in \mathbb{N} : i \leq q\}$, where $\mathbb{N} = \{1, 2, \ldots\}$, $Q_q = \{(i,j) \in P_q^2 : i < j\}$, and $R_{q,k} = \{(i,j) \in Q_q : i = k \text{ or } j = k\}$ for $k \in P_q$.

Theorem 3.1 *For integer $J > 1$, let μ_j, $j \in P_J$ be jointly normally distributed with zero mean, unit variance, and covariances $\rho_{jk} = \mathrm{Cov}(\mu_j, \mu_k)$, $j \neq k$; let $f_j = f_j(\mu_j)$ be a function such that $E(f_j^2) < \infty$, with k-th Hermite coefficient $G_{j,k}$ and Hermite rank r_j. Then*

$$E\left(\prod_{j\in P_J} f_j\right) = \sum_{q=0}^{\infty} a_q, \tag{3.3}$$

where

$$a_q = \sum_{\substack{v_\alpha \geq 0:\\ \sum v_\alpha = q,\\ \alpha \in Q_J}} \prod_{j\in P_J} G_{j,w_j} \prod_{\alpha\in Q_J} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!}, \qquad w_j = \sum_{\alpha \in R_{J,j}} v_\alpha. \tag{3.4}$$

If in addition $\tau = 2\sum_{\alpha\in Q_J} |\rho_\alpha| < 1$, then

$$a_q = 0, \qquad 2q < r, \tag{3.5}$$

$$|a_q| \leq \sigma \prod_{j\in P_J} \left(\sum_{\alpha\in R_{J,j}} |\rho_\alpha|\right)^{\frac{r_j}{2}} \tau^{q-\frac{r}{2}}, \qquad 2q \geq r, \tag{3.6}$$

$$\sum_{i=q}^{\infty} |a_i| \leq \sigma \prod_{j\in P_J} \left(\sum_{\alpha\in R_{J,j}} |\rho_\alpha|\right)^{\frac{r_j}{2}} \frac{\tau^{q-\frac{r}{2}}}{1-\tau}, \qquad 2q \geq r, \tag{3.7}$$

where $r = \sum_{j\in P_J} r_j$ and $\sigma = \{\prod_{j\in P_J} E(f_j^2)\}^{1/2}$.

Proof. Throughout the proof, we denote $P = P_J$, $Q = Q_J$, and $R_j = R_{J,j}$,

$j \in P$. Furthermore, all sums and products run over P unless otherwise stated. We have

$$E\left(\prod_j f_j\right) = \int_{\mathbb{R}^J} \prod_j f_j \phi_J(\mu;\Omega)\mathrm{d}\mu, \tag{3.8}$$

where $\phi_J(\mu;\Omega)$ denotes the density function of $\mu = (\mu_1,\ldots,\mu_J)'$ and $\Omega = E(\mu\mu')$. From (22) of Slepian (1972) and Definition 12,

$$\begin{aligned}
\phi_J(\mu;\Omega) &= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j \left\{ \left(\frac{\partial}{\partial\mu_j}\right)^{w_j} \phi(\mu_j) \right\} \\
&= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j \left\{ (-1)^{w_j} H_{w_j}(\mu_j)\phi(\mu_j) \right\} \\
&= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j \left\{ H_{w_j}(\mu_j)\phi(\mu_j) \right\}, \qquad (3.9)
\end{aligned}$$

since $\sum_j w_j = 2\sum_{\alpha\in Q} v_\alpha$ is even. Using (3.9) in (3.8),

$$\begin{aligned}
E\left(\prod_j f_j\right) &= \int_{\mathbb{R}^J} \prod_j f_j(\mu_j) \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j \left\{ H_{w_j}(\mu_j)\phi(\mu_j) \right\} \mathrm{d}\mu \\
&= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \int_{\mathbb{R}^J} \prod_j \left\{ f_j(\mu_j) H_{w_j}(\mu_j)\phi(\mu_j) \right\} \mathrm{d}\mu \\
&= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \prod_j E\left\{ f_j(\mu_j) H_{w_j}(\mu_j) \right\} \\
&= \sum_{v_\alpha=0:\alpha\in Q}^{\infty} \prod_j G_{j,w_j} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!}.
\end{aligned}$$

This proves (3.4). For the remainder of the proof, we use the Cauchy-Schwarz inequality in

$$\begin{aligned}
|a_q| &\le \sum_{\substack{v_\alpha\ge 0: \\ \Sigma v_\alpha = q, \alpha\in Q}} \left| \prod_j G_{j,w_j} \prod_{\alpha\in Q} \frac{\rho_\alpha^{v_\alpha}}{v_\alpha!} \right| \\
&\le \sum_{\substack{v_\alpha\ge 0: \\ \Sigma v_\alpha = q, \alpha\in Q}} \prod_j \frac{|G_{j,w_j}|}{\sqrt{w_j!}} \prod_j \sqrt{w_j!} \prod_{\alpha\in Q} \frac{|\rho_\alpha|^{v_\alpha}}{v_\alpha!}
\end{aligned}$$

$$\leq (A_q B_q)^{\frac{1}{2}}, \tag{3.10}$$

where

$$A_q = \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = q, \alpha \in Q \\ w_j \geq r_j, j \in P}} \prod_j \frac{G_{j,w_j}^2}{w_j!}, \qquad B_q = \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = q, \alpha \in Q \\ w_j \geq r_j, j \in P}} \prod_j \left(w_j! \prod_{\alpha \in R_j} \frac{|\rho_\alpha|^{v_\alpha}}{v_\alpha!} \right).$$

The A_q term is bounded since

$$A_q \leq \prod_j \sum_{w_j = r_j}^{\infty} \frac{G_{j,w_j}^2}{w_j!} \leq \prod_j E(f_j^2) \leq \sigma^2. \tag{3.11}$$

If $2q < r$, there always exists a j in (3.4) such that $w_j < r_j$, implying (3.5).

For $2q \geq r$, the multinomial theorem yields

$$\begin{aligned}
B_q &\leq \sum_{\substack{w_j \geq r_j: \\ \Sigma w_j = 2q, j \in P}} \sum_{\substack{v_\alpha \geq 0: \Sigma v_\alpha = w_j \\ \alpha \in R_j, j \in P}} \prod_j \left(w_j! \prod_{\alpha \in R_j} \frac{|\rho_\alpha|^{v_\alpha}}{v_\alpha!} \right) \\
&\leq \sum_{\substack{w_j \geq r_j: \\ \Sigma w_j = 2q, j \in P}} \prod_j \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = w_j, \alpha \in R_j}} w_j! \prod_{\alpha \in R_j} \frac{|\rho_\alpha|^{v_\alpha}}{v_\alpha!} \\
&\leq \sum_{\substack{w_j \geq r_j: \\ \Sigma w_j = 2q, j \in P}} \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{w_j} \\
&\leq \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{r_j} \cdot \sum_{\substack{w_j \geq 0: \\ \Sigma w_j = 2q - r, j \in P}} \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{w_j} \\
&\leq \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{r_j} \sum_{\substack{w_j \geq 0: \\ \Sigma w_j = 2q - r, j \in P}} (2q - r)! \prod_j \frac{\left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{w_j}}{w_j!} \\
&\leq \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{r_j} \left(\sum_{j,k: j \neq k} |\rho_{jk}| \right)^{2q - r}
\end{aligned}$$

$$\leq \prod_j \left(\sum_{\alpha \in R_j} |\rho_\alpha| \right)^{r_j} \tau^{2q-r}. \tag{3.12}$$

Using (3.11), (3.12) in (3.10) gives (3.6). Then (3.7) follows from $\tau < 1$. □

The bounds (3.5), (3.6) reflect the individual, possibly differing, Hermite ranks r_j of the f_j. The weakest version of Theorem 3.1 arises when $r_j \equiv 0$ (i.e. when $E(f_j) \neq 0$ for all j), and because this would be relevant also when the r_j are unknown, we present it in the following Corollary, whose proof follows from the inequality $\sum_{\alpha \in R_{j,J}} |\rho_\alpha| \leq \tau$.

Corollary 3.1

$$|a_q| \leq \sigma \tau^q, \qquad \sum_{i=q}^{\infty} |a_i| \leq \sigma \frac{\tau^q}{1-\tau}.$$

As in Robinson (2001) in case $J = 2$, Theorem 3.1 provides a valid asymptotic expansion when $\tau \to 0$. Robinson (1994a) established consistency of the NBLS estimate using L^1 arguments enabled by linear process (in conditionally homoscedastic martingale difference innovations) assumptions. Since those are unavailable to us, we use L^2 arguments. These were also employed by Robinson (1994b) in studying the mean squared error of the averaged periodogram, but in case of Gaussian and linear (in iid innovations) assumptions. In the SV setting introduced in the following section, matters are considerably more complicated, and we are led to consider various cross-moments of nonlinear functions of Gaussian processes. Theorem 3.1 is crucial in obtaining sufficiently sharp bounds on these cross-moments to establish consistency.

3.3 Long memory stochastic volatility setup

To describe the structure of the latent processes ζ_t, ε_t, δ_t in (3.1), (3.2), we first refine the definition of $I(d)$ processes originally presented as Definition 8.

Definition 13 *We say a process z_t is $I(d)$, with memory parameter $d \in [0, 1/2)$, if it is stationary with finite variance, and has autocovariance function $\rho_j = \mathrm{Cov}(z_0, z_j)$ satisfying*

$$\sum_{j=0}^{\infty} |\rho_j| < \infty, \tag{3.13}$$

if $d = 0$, and

$$\rho_j \sim C_\rho j^{2d-1} \text{ as } j \to \infty, \text{ for } C_\rho > 0, \tag{3.14}$$

$$|\rho_j - \rho_{j+1}| \leq K\frac{|\rho_{j+1}|}{j}, \ j \geq 0, \tag{3.15}$$

if $0 < d < 1/2$, where K throughout denotes a generic, arbitrarily large finite constant.

The alternative, frequency domain Definition 7 holds as a consequence of (3.13) or (3.14), (3.15). For $d = 0$, the spectral density $f(\lambda)$ of z_t is continuous for all λ, whereas for $0 < d < 1/2$, Theorem III-12 of Yong (1974) indicates that

$$f(\lambda) \sim C_f \lambda^{-2d} \text{ as } \lambda \to 0^+, \tag{3.16}$$

where

$$C_f = \pi^{-1}\Gamma(2d)\sin\left\{(1-2d)\frac{\pi}{2}\right\}C_\rho,$$

so that $f(\lambda)$ diverges at $\lambda = 0$. We can therefore use Definitions 7 and 8 interchangeably under the additional quasi-monotonicity condition (3.15). Stationary ARMA processes satisfy (3.13), and stationary ARFIMA processes satisfy (3.14), (3.15).

Assumption 3.1 *For $t \in \mathbb{Z}$,*

$$\zeta_t = \eta_{1t} g_t, \qquad \delta_t = \nu_{1t} h_t, \qquad \varepsilon_t = \xi_{1t} l_t, \tag{3.17}$$

where for real-valued functions g, h, l,

$$g_t = g(\eta_{2t}), \qquad h_t = h(\nu_{2t}), \qquad l_t = l(\xi_{2t}), \tag{3.18}$$

and:

(i) $\{\eta_{1t}\}$, $\{\nu_{1t}\}$, $\{\xi_{1t}\}$ *are jointly iid processes with zero mean;*

(ii) $\{\eta_{2t}\}$ *is* $I(d_1)$, $\{\nu_{2t}\}$ *is* $I(d_2)$, *and* $\{\xi_{2t}\}$ *is* $I(d_3)$, *for* $d_2 \geq 0$, $d_3 \geq 0$, *where* $\max\{d_2, d_3\} < d_1 < 1/2$;

(iii) $\{\eta_{2t}\}$, $\{\nu_{2t}\}$, $\{\xi_{2t}\}$ *are standard Gaussian processes, independent of each other and of* $\{\eta_{1t}\}$, $\{\nu_{1t}\}$, $\{\xi_{1t}\}$;

(iv) For some integer $p > 1$,

$$E(\eta_{1t}^p) E\left\{g^p(\eta_{2t})\eta_{2t}\right\} \neq 0, \tag{3.19}$$

and for $j = 1, \ldots, p-1$,

$$E(\eta_{1t}^j \nu_{1t}^{p-j}) E\left\{g^j(\eta_{2t})\eta_{2t}\right\} = E(\eta_{1t}^j \xi_{1t}^{p-j}) E\left\{g^j(\eta_{2t})\eta_{2t}\right\} = 0; \tag{3.20}$$

(v) $\{\eta_{1t}\}$, $\{\nu_{1t}\}$, $\{\xi_{1t}\}$, $\{g_t\}$, $\{h_t\}$, $\{l_t\}$ *have finite* $4p$*-th moments.*

It follows that ζ_t, ε_t, δ_t, described by SV models in (3.17), are serially uncorrelated but not serially independent. In particular, x_t^p is $I(d_1)$, due to (3.19), which entails $E(\eta_{1t}^p) \neq 0$ and $g_t^p - E(g_t^p)$ having Hermite rank one. Condition (3.20) ensures a valid cointegrating relationship between x_t^p and y_t^p, since it implies that the cointegrating error has memory smaller than d_1. If η_{1t} is independent of ν_{1t}, ξ_{1t}, the smallest integer satisfying (3.19) will also satisfy (3.20). It is assumed that p is known, which imposes some restrictions on g; in practice it may be reasonable

to suppose that $p = 2$. As discussed in Section 2.4, the most notable exception would occur if g is a symmetric function, e.g. $g_t = |\eta_{2t}|^a$, $\alpha > 0$, but then no finite p satisfies (3.19). This does not rule out a cointegrating relationship of the type that we study below, but the associated conditions would be extremely complex, involving the magnitudes of the d_i, $i = 1, 2, 3$, and the Hermite ranks of each centered power of g_t, h_t, l_t. Note that for $\gamma \neq 0$, $g_t = |\gamma + \eta_{2t}|^a$ gives $p = 2$. Further discussion concerning the Hermite rank for functional forms in SV models with long memory can be found in Section 2.4 and Robinson (2001).

An advantage of a low p is that the moment conditions in part (v) of Assumption 3.1 increase in strength with p. Even for $p = 2$, the 8-th moment condition that is required seems stringent for most financial data: Jansen and de Vries (1991) and Loretan and Phillips (1994), among others, suggested that several financial time series may have infinite fourth moments. Other parts of Assumption 3.1 might be relaxed at cost of substantial lengthening of the proof, in particular the mutual independence assumptions of (iii). A consistency result under weaker versions of (iii) could surely be provided with the same theoretical tools, but enumeration of all relevant cross-moments would be a tedious exercise with little added value. The Gaussianity assumption on η_{2t}, ν_{2t}, ξ_{2t} is mitigated by allowing g, h, l to be quite general functions, and without Gaussianity the details would be considerably more complex; of course Gaussianity frequently plays a role in short memory SV models also. We do not assume Gaussianity of η_{1t}, ν_{1t}, ξ_{1t}.

3.4 Consistency of Narrow Band Least Squares

We generalise the techniques of Section 2.5, namely the use of (2.39), to transform (3.1), (3.2) to

$$Y_t = \theta X_t + U_t, \tag{3.21}$$

where

$$
\begin{aligned}
Y_t = y_t^p &= \sum_{j=0}^{p} \binom{p}{j} \beta_1^j \zeta_t^j \varepsilon_t^{p-j}, \\
X_t = x_t^p &= \sum_{j=0}^{p} \binom{p}{j} \beta_2^j \zeta_t^j \delta_t^{p-j}, \\
\theta = \beta^p &= \frac{\beta_1^p}{\beta_2^p}, \\
U_t = y_t^p - \theta x_t^p &= \sum_{j=0}^{p} \binom{p}{j} (\beta_1^j \zeta_t^j \varepsilon_t^{p-j} - \beta^p \beta_2^j \zeta_t^j \delta_t^{p-j}) \\
&= \sum_{j=0}^{p-1} \binom{p}{j} \beta_2^j \zeta_t^j (\beta^j \varepsilon_t^{p-j} - \beta^p \delta_t^{p-j}).
\end{aligned}
$$

It will follow from (3.21) and Assumption 3.1 that Y_t and X_t are cointegrated $I(d_1)$ processes. As an example, if $p = 2$, then U_t is given by a variation of (2.40),

$$U_t = \varepsilon_t^2 - \beta^2 \delta_t^2 + 2\beta_2 \zeta_t (\beta \varepsilon_t - \beta^2 \delta_t).$$

The memory parameters of δ_t^2, ε_t^2 are bounded by d_2, d_3, respectively, and therefore smaller than d_1 by part (ii) of Assumption 3.1. Condition (3.20) guarantees that either g_t has Hermite rank greater than one, reducing the memory of the last term by virtue of Theorem 3.1, or that both $\zeta_t \varepsilon_t$ and $\zeta_t \delta_t$ contain a zero mean and serially uncorrelated multiplicative error, and are therefore white noise. By contrast, (3.19) ensures that g_t^2 in X_t has Hermite rank one, and thus retains the memory, d_1, of its underlying volatility process.

Since X_t, Y_t, $t = 1, \ldots, n$ are now scalar sequences, the NBLS estimate (2.55) of Robinson (1994a) for θ may be written as

$$\hat{\theta}_m = \frac{\mathrm{Re}\left\{\hat{F}_{XY}(\lambda_m)\right\}}{\hat{F}_{XX}(\lambda_m)}, \qquad 1 \leq m \leq \frac{n}{2}, \tag{3.22}$$

We can estimate β by $\hat{\beta}_m = \hat{\theta}_m^{1/p}$, though only up to an unknown sign when p is even. As discussed in Section 2.6, for consistency we require that the band of frequencies considered slowly degenerates to zero.

Assumption 3.2 *The bandwidth sequence $m = m(n)$ satisfies*

$$\frac{1}{m} + \left(\frac{m}{n}\right)^{\epsilon} \log n \to 0 \text{ as } n \to \infty, \tag{3.23}$$

for all $\epsilon > 0$.

This assumption is slightly stronger than (2.54), reproduced here for convenience,

$$\frac{1}{m} + \frac{m}{n} \to 0 \text{ as } n \to \infty. \tag{3.24}$$

We need (3.23) over (3.24) only in order to handle powers of g_t, h_t, l_t with particular combinations of memory parameters and Hermite ranks, notably for $d_i = 1/4$. This case presents no special problems with the method of proof in Robinson (1994a), and is excluded in Robinson (1994b).

For integers $j \in [1, p-1]$ and $k \in [0, p-1]$, denote the Hermite rank of centered g^j, h^{p-k}, l^{p-k} by r_{gj}, r_{hk}, r_{lk} respectively, and introduce the sets

$$\begin{aligned}
S_g &= \left\{j : \beta^j E(\eta_{1t}^j \varepsilon_t^{p-j}) \neq \beta^p E(\eta_{1t}^j \delta_t^{p-j}), \quad 0 < j < p\right\}, \\
S_h &= \left\{k : E(\nu_{1t}^{p-k} \zeta_t^k) \neq 0, \quad 0 \leq k < p\right\}, \\
S_l &= \left\{k : E(\xi_{1t}^{p-k} \zeta_t^k) \neq 0, \quad 0 \leq k < p\right\}, \\
S_{gh} &= \left\{j : E(\eta_{1t}^j \nu_{1t}^{p-j}) \neq 0, \quad 0 < j < p\right\}, \\
S_{gl} &= \left\{j : E(\eta_{1t}^j \xi_{1t}^{p-j}) \neq 0, \quad 0 < j < p\right\}.
\end{aligned}$$

Intuitively, U_t will be expanded as a sum of terms involving the basic processes described in Assumption 3.1. This allows us to express the autocovariance function

of U_t as a linear combination of covariances of powers of g_t, h_t, l_t and products of these covariances. The memory of U_t will depend only on those terms associated with a nonzero coefficient, in particular for which the white noise component has nonzero mean. These five sets group the particular exponents for which this occurs: S_g, S_h, S_l for terms including only the covariances of powers of g_t, h_t, l_t respectively, and S_{gh}, S_{gl} for cross-products of said covariances. Note that U_t does not contain products of ε_t and δ_t, and therefore interactions between h_t and l_t do not occur. Using the convention that the maximum over an empty set is $-\infty$, the slowest rate of decay corresponding to each source is defined by

$$d_g^* = \max_{j \in S_g} \left\{ \frac{1}{2} - r_{gj} \left(\frac{1}{2} - d_1 \right) \right\}, \tag{3.25}$$

$$d_h^* = \max_{k \in S_h} \left\{ \frac{1}{2} - r_{hk} \left(\frac{1}{2} - d_2 \right) \right\} - 1(d_2 = 0), \tag{3.26}$$

$$d_l^* = \max_{k \in S_l} \left\{ \frac{1}{2} - r_{lk} \left(\frac{1}{2} - d_3 \right) \right\} - 1(d_3 = 0), \tag{3.27}$$

$$d_{gh}^* = \max_{j \in S_{gh}} \left\{ \frac{1}{2} - r_{gj} \left(\frac{1}{2} - d_1 \right) - r_{hj} \left(\frac{1}{2} - d_2 \right) \right\}, \tag{3.28}$$

$$d_{gl}^* = \max_{j \in S_{gl}} \left\{ \frac{1}{2} - r_{gj} \left(\frac{1}{2} - d_1 \right) - r_{lj} \left(\frac{1}{2} - d_3 \right) \right\}, \tag{3.29}$$

where $1(\cdot)$ throughout denotes the identity function, and

$$d^* = \max\{d_g^*, d_h^*, d_l^*, d_{gh}^*, d_{gl}^*\}. \tag{3.30}$$

Theorem 3.2 *Under Assumptions 3.1 and 3.2, as* $n \to \infty$

$$\hat{\theta}_m - \theta = O_p\left(\left(\frac{m}{n} \right)^{d_1 - d_u} \right), \tag{3.31}$$

where $d_u = d^* 1(d^* > 0) + \epsilon 1(d^* = 0)$, *for any* $\epsilon > 0$.

Proof. As in Section 5.3 of Robinson (1994a),

$$|\hat{\theta}_m - \theta| \leq \left\{ \frac{\hat{F}_{UU}(\lambda_m)}{\hat{F}_{XX}(\lambda_m)} \right\}^{\frac{1}{2}}.$$

By Proposition 3.2,

$$\left(\frac{m}{n}\right)^{2d_u-1} \hat{F}_{UU}(\lambda_m) = O_p(1),$$

while by Propositions 3.1, 3.3, and Slutsky's Theorem,

$$\frac{\left(\frac{m}{n}\right)^{1-2d_1}}{\hat{F}_{XX}(\lambda_m)} \xrightarrow{p} \frac{1}{C^*} < \infty. \square$$

Proofs of Propositions 3.1, 3.2, and 3.3 can be found in Appendix A, using lemmas in Appendix B. Since ϵ is arbitrarily small and $d^* < d_1$, it follows that $\hat{\theta}_m$ is consistent for θ. Moreover, when $d^* > 0$, we can write $d_1 - d_u = d_1 - d^*$, which is the difference between the integration orders of X_t and U_t. In this case, the rate in (3.31) corresponds to that of Robinson and Marinucci (2003). For some particular combinations of memory parameters and Hermite ranks, yielding zeros in (3.25)-(3.29), the autocorrelation function is $O(j^{-1})$, and an additional $\log n$ factor arises. When such a process dominates in the expansion of U_t, (3.23) is required to derive (3.31), justifying the appearance of ϵ in the above rate of convergence.

3.5 Final comments

To our knowledge this represents the first formal treatment of fractional cointegration in the context of nonlinear processes. The stationary environment, the SV models employed, and the NBLS estimate seem well motivated by applications

in finance. Our model is semiparametric both in the sense that only assumptions about low frequency behavior are required, and the volatility functions are nonparametric. While the nonlinear setting necessitates a considerably more complex proof of consistency of NBLS than earlier ones, a comparable result is obtained, with rate of convergence depending essentially on the strength of the cointegrating relation, namely the gap between integration orders of observables and cointegrating error.

As always, consistency results are reassuring only in very large data sets. Though these do exist in finance, one would like a limit distributional result that could be used in statistical inference. Christensen and Nielsen (2006) have achieved this in a simpler setting, indeed with regressor and disturbance assumed incoherent at frequency zero, and linear process (in conditionally homoscedastic martingale difference innovations) assumptions. In general, not only is the proof likely to be much more complicated than even our proof of Theorem 3.2, but the limit distribution is likely to be non-standard for various combinations of memory parameters (as is the case for the averaged periodogram estimate of the memory parameter, discussed in Section 2.7), though a bootstrap procedure might be investigated. By analogy with experience in $I(1)/I(0)$ cointegrated models (e.g. Johansen, 1991; Phillips, 1991), it may be possible to obtain estimates with nicer asymptotic distributional properties, in particular leading to Wald statistics with null limiting χ^2 distributions. However, in our nonlinear setting it is not immediately obvious that the sort of transformations used in those references to achieve the necessary "whitening" will be successful, the estimates would require preliminary estimation of memory parameters, and proofs would be significantly more complicated. In the following chapter, we present some simulation results which seem to indicate that the finite sample distribution of NBLS, under this nonlinear setting, is far from Gaussian. Still, those wishing to embark on limit distribu-

tional proofs for NBLS or other estimates in our SV setting should find techniques described in the present chapter relevant.

The bulk of the fractional and non-fractional cointegration literature assumes nonstationary observables. The motivation usually comes from macroeconomics, but nonstationarity can often appear in financial time series also. The modelling of nonstationary series via analogues of (3.21) is itself a somewhat open topic, but given that X_t has a kind of $I(d)$ property, for $d \geq 1/2$, some of the arguments of Robinson and Marinucci (2001) should be relevant in establishing rates of convergence of NBLS. Indeed, these authors, following Stock (1987) in the $I(1)/I(0)$ case, found OLS also to be consistent here, though in some circumstances NBLS has bias of smaller order. The nonstationary X_t case is in some respects technically easier than the stationary one, because consistency of OLS follows from the domination of sums of squares of U_t by those of X_t.

3.A Propositions for Theorem 3.2

We denote the Dirichlet kernel by $D_m(\lambda) = \sum_{j=1}^{m} e^{ij\lambda}$, for $m \geq 1$, and will use the fact that

$$D_n(\lambda_j) = n1(j = 0, \bmod n). \tag{3.32}$$

We also use the abbreviating notation

$$S_m(a,b) = E\left\{\hat{F}_{ab}(\lambda_m)\right\} = \frac{1}{n^2}\sum_{s,t=1}^{n} \mathrm{Cov}(a_s, b_t) D_m(\lambda_{t-s}),$$

from (3.32), and

$$S'_m(a,b;a',b') = \frac{1}{n^2}\sum_{s,t=1}^{n} \mathrm{Cov}(a_s, b_t)\,\mathrm{Cov}(a'_s, b'_t) D_m(\lambda_{t-s}),$$

where a_t, b_t, a'_t, b'_t, $t = 1, \ldots, n$ are scalar sequences with finite second moments. The integer part of x is denoted throughout as $[x]$.

Proposition 3.1 *Under (3.2) and Assumptions 3.1 and 3.2,*

$$\left(\frac{m}{n}\right)^{2d_1-1} E\left\{\hat{F}_{XX}(\lambda_m)\right\} \to C^* \text{ as } n \to \infty,$$

where

$$\begin{aligned} C^* &= 2\frac{(2\pi)^{-2d_1}\Gamma(2d_1)}{1-2d_1}\sin\left\{(1-2d_1)\frac{\pi}{2}\right\}\beta_2^{2p}E(\eta_1^p)^2E\left\{g^p(\eta_{2t})\eta_{2t}\right\}^2 C_\rho > 0, \\ C_\rho &= \lim_{j\to\infty} E(\eta_{20}\eta_{2j})j^{1-2d_1}. \end{aligned}$$

Proof. Write

$$X_t = \sum_{j=0}^{p} A_{jt}B_{jt}, \qquad A_{jt} = \binom{p}{j}\beta_2^j\eta_{1t}^j\nu_{1t}^{p-j}, \qquad B_{jt} = g_t^j h_t^{p-j}.$$

Using Lemma 3.2, since $\{A_{jt}\}$ is independent of $\{B_{kt}\}$, for any j and k,

$$\begin{aligned} \operatorname{Cov}(X_s, X_t) &= \sum \operatorname{Cov}(A_{js}B_{js}, A_{kt}B_{kt}) \\ &= \sum \left\{E(A_{js})E(A_{kt})\operatorname{Cov}(B_{js}, B_{kt}) + \operatorname{Cov}(A_{js}, A_{kt})E(B_{js}B_{kt})\right\}, \end{aligned}$$

where $\sum$ denotes $\sum_{j,k=0}^{p}$ throughout the proof.

Now define $a_j = E(A_{jt})$, $b_{g,j} = a_j E(h_t^{p-j})$, and $b_{h,j} = a_j E(g_t^j)$. Since $\{A_{jt}\}$ is iid, using Lemma 3.2 again, for $s \neq t$, $\operatorname{Cov}(X_s, X_t)$ is

$$\begin{aligned} \sum a_j a_k \operatorname{Cov}(B_{js}, B_{kt}) = \sum \Big\{ & b_{g,j}b_{g,k}\operatorname{Cov}(g_s^j, g_t^k) \\ & + b_{h,j}b_{h,k}\operatorname{Cov}(h_s^{p-j}, h_t^{p-k}) + a_j a_k \operatorname{Cov}(g_s^j, g_t^k)\operatorname{Cov}(h_s^{p-j}, h_t^{p-k})\Big\}. \end{aligned} \tag{3.33}$$

For $s = t$, denote by Λ the difference between $\mathrm{Var}(X_t)$ and (3.33). It follows that $E\{\hat{F}_{XX}(\lambda_m)\}$ is

$$\sum \{b_{g,j}b_{g,k}S_m(g^j, g^k) + b_{h,j}b_{h,k}S_m(h^{p-j}, h^{p-k}) + a_j a_k S'_m(g^j, g^k; h^{p-j}, h^{p-k})\} + \frac{m}{n}\Lambda.$$

From (3.19), (3.20), and Lemma 3.4,

$$b_{g,j}b_{g,k}S_m(g^j, g^k) = o\left(\left(\frac{m}{n}\right)^{1-2d_1}\right),$$

if either $j < p$ or $k < p$, while

$$b_{g,p}^2 S_m(g^p, g^p) = C^* \left(\frac{m}{n}\right)^{1-2d_1} + o\left(\left(\frac{m}{n}\right)^{1-2d_1}\right).$$

Lemma 3.4 and $d_2 < d_1$ imply that

$$b_{h,j}b_{h,k}S_m(h^{p-j}, h^{p-k}) = o\left(\left(\frac{m}{n}\right)^{1-2d_1}\right),$$

and by Lemma 3.5,

$$a_j a_k S'_m(g^j, g^k; h^{p-j}, h^{p-k}) = o\left(\left(\frac{m}{n}\right)^{1-2d_1}\right),$$

which concludes the proof. □

Proposition 3.2 *Under (3.1), (3.2), and Assumptions 3.1 and 3.2,*

$$E\left\{\hat{F}_{UU}(\lambda_m)\right\} = O\left(\left(\frac{m}{n}\right)^{1-2d_u}\right).$$

Proof. Write

$$U_t = \sum_{j=0}^{p-1} A_{\varepsilon,jt}B_{\varepsilon,jt} - A_{\delta,jt}B_{\delta,jt},$$

where

$$A_{\varepsilon,jt} = \binom{p}{j}\beta_2^j\beta^j\eta_{1t}^j\xi_{1t}^{p-j}, \qquad B_{\varepsilon,jt} = g_t^j l_t^{p-j}$$
$$A_{\delta,jt} = \binom{p}{j}\beta_2^j\beta^p\eta_{1t}^j\nu_{1t}^{p-j}, \qquad B_{\delta,jt} = g_t^j h_t^{p-j}.$$

Using Lemma 3.2 repeatedly, since $\{A_{\varepsilon,jt}\}$, $\{A_{\delta,jt}\}$ are independent of $\{B_{\varepsilon,kt}\}$, $\{B_{\delta,kt}\}$, for any j and k, $\mathrm{Cov}(U_s, U_t)$ is

$$\begin{aligned}
&\sum \{\mathrm{Cov}(A_{\varepsilon,js}B_{\varepsilon,js}, A_{\varepsilon,kt}B_{\varepsilon,kt}) + \mathrm{Cov}(A_{\delta,js}B_{\delta,js}, A_{\delta,kt}B_{\delta,kt}) \\
&\quad - \mathrm{Cov}(A_{\varepsilon,js}B_{\varepsilon,js}, A_{\delta,kt}B_{\delta,kt}) - \mathrm{Cov}(A_{\delta,js}B_{\delta,js}, A_{\varepsilon,kt}B_{\varepsilon,kt})\} \\
&= \sum \{E(A_{\varepsilon,js})E(A_{\varepsilon,kt})\,\mathrm{Cov}(B_{\varepsilon,js}, B_{\varepsilon,kt}) + \mathrm{Cov}(A_{\varepsilon,js}, A_{\varepsilon,kt})E(B_{\varepsilon,js}B_{\varepsilon,kt}) \\
&\quad + E(A_{\delta,js})E(A_{\delta,kt})\,\mathrm{Cov}(B_{\delta,js}, B_{\delta,kt}) + \mathrm{Cov}(A_{\delta,js}, A_{\delta,kt})E(B_{\delta,js}B_{\delta,kt}) \\
&\quad - E(A_{\varepsilon,js})E(A_{\delta,kt})\,\mathrm{Cov}(B_{\varepsilon,js}, B_{\delta,kt}) - \mathrm{Cov}(A_{\varepsilon,js}, A_{\delta,kt})E(B_{\varepsilon,js}B_{\delta,kt}) \\
&\quad - E(A_{\delta,js})E(A_{\varepsilon,kt})\,\mathrm{Cov}(B_{\delta,js}, B_{\varepsilon,kt}) - \mathrm{Cov}(A_{\delta,js}, A_{\varepsilon,kt})E(B_{\delta,js}B_{\varepsilon,kt})\}\,,
\end{aligned}$$

where $\sum$ denotes $\sum_{j,k=0}^{p-1}$ throughout the proof.

Now define $a_{\varepsilon j} = E(A_{\varepsilon,jt})$, $a_{\delta j} = E(A_{\delta,jt})$, $b_{gj} = a_{\varepsilon j}E(l_t^{p-j}) - a_{\delta j}E(h_t^{p-j})$, $b_{hj} = a_{\delta j}E(g_t^j)$, and $b_{lj} = a_{\varepsilon j}E(g_t^j)$. Since $\{A_{\varepsilon,jt}\}$, $\{A_{\delta,jt}\}$ are jointly iid, using Lemma 3.2 again, for $s \neq t$, $\mathrm{Cov}(U_s, U_t)$ is

$$\begin{aligned}
&\sum \{a_{\varepsilon j}a_{\varepsilon k}\,\mathrm{Cov}(B_{\varepsilon,js}, B_{\varepsilon,kt}) + a_{\delta j}a_{\delta k}\,\mathrm{Cov}(B_{\delta,js}, B_{\delta,kt}) \\
&\quad - a_{\varepsilon j}a_{\delta k}\,\mathrm{Cov}(B_{\varepsilon,js}, B_{\delta,kt}) - a_{\delta j}a_{\varepsilon k}\,\mathrm{Cov}(B_{\delta,js}, B_{\varepsilon,kt})\} \\
&= \sum \Big\{b_{gj}b_{gk}\,\mathrm{Cov}(g_s^j, g_t^k) + b_{hj}b_{hk}\,\mathrm{Cov}(h_s^{p-j}, h_t^{p-k}) \\
&\quad + b_{lj}b_{lk}\,\mathrm{Cov}(l_s^{p-j}, l_t^{p-k}) + a_{\delta j}a_{\delta k}\,\mathrm{Cov}(g_s^j, g_t^k)\,\mathrm{Cov}(h_s^{p-j}, h_t^{p-k}) \\
&\quad + a_{\varepsilon j}a_{\varepsilon k}\,\mathrm{Cov}(g_s^j, g_t^k)\,\mathrm{Cov}(l_s^{p-j}, l_t^{p-k})\Big\}. \qquad (3.34)
\end{aligned}$$

For $s = t$, denote by Λ the difference between $\mathrm{Var}(U_t)$ and (3.34). It follows that $E\{\hat{F}_{UU}(\lambda_m)\}$ is

$$\sum \{b_{gj}b_{gk}S_m(g^j, g^k) + b_{hj}b_{hk}S_m(h^{p-j}, h^{p-k}) + b_{lj}b_{lk}S_m(l^{p-j}, l^{p-k}) + a_{\delta j}a_{\delta k}S'_m(g^j, g^k; h^{p-j}, h^{p-k}) + a_{\epsilon j}a_{\epsilon k}S'_m(g^j, g^k; l^{p-j}, l^{p-k})\} + \frac{m}{n}\Lambda.$$

By (3.25), applying Lemma 3.4 to each (j, k) pair with non-zero coefficient,

$$b_{gj}b_{gk}S_m(g^j, g^k) = O\left(\left(\frac{m}{n}\right)^{1-2\max\{d_g^*,0\}} (\log n)^{1(d_g^*=0)}\right).$$

Similarly, by (3.26) and (3.27), Lemma 3.4 yields

$$b_{hj}b_{hk}S_m(h^{p-j}, h^{p-k}) = O\left(\left(\frac{m}{n}\right)^{1-2\max\{d_h^*,0\}} (\log n)^{1(d_h^*=0)}\right),$$
$$b_{lj}b_{lk}S_m(l^{p-j}, l^{p-k}) = O\left(\left(\frac{m}{n}\right)^{1-2\max\{d_l^*,0\}} (\log n)^{1(d_l^*=0)}\right).$$

Finally, Lemma 3.5, (3.28), and (3.29) give

$$a_{\delta j}a_{\delta k}S'_m(g^j, g^k; h^{p-j}, h^{p-k}) = O\left(\left(\frac{m}{n}\right)^{1-2\max\{d_{gh}^*,0\}} (\log n)^{1(d_{gh}^*=0)}\right),$$
$$a_{\epsilon j}a_{\epsilon k}S'_m(g^j, g^k; l^{p-j}, l^{p-k}) = O\left(\left(\frac{m}{n}\right)^{1-2\max\{d_{gl}^*,0\}} (\log n)^{1(d_{gl}^*=0)}\right).$$

By (3.20), $d_g^* < d_1$. Since d_h^* and d_{gh}^* are bounded by $d_2 < d_1$ while d_l^* and d_{gl}^* are bounded by $d_3 < d_1$, we have $d^* < d_1$. The bound for $d^* = 0$ follows from Assumption 3.2. □

Proposition 3.3 *Under (3.2) and Assumptions 3.1 and 3.2,*

$$\mathrm{Var}\left\{\hat{F}_{XX}(\lambda_m)\right\} = o\left(\left(\frac{m}{n}\right)^{2-4d_1}\right).$$

Proof. Define $\rho_t = E(\eta_{20}\eta_{2t})$; wherever time indices t_i, $i = 1, \ldots, 4$ are used, it will be convenient to write also $\gamma_{ij} = \rho_{t_j - t_i}$. Denoting $Z_t = X_t - E(X_t)$, there exists a Gaussian $I(d_1)$ process V_t such that the bounds in Lemma 3.6 hold. Lemmas 7 and 10 in Robinson (1994b) and Lemma 3.7 imply that

$$\mathrm{Var}\left\{\hat{F}_{VV}(\lambda_m)\right\} = o\left(\left(\frac{m}{n}\right)^{2-4d_1}\right),$$

so we need to show that the approximation error satisfies

$$A = \mathrm{Var}\left\{\hat{F}_{XX}(\lambda_m)\right\} - \mathrm{Var}\left\{\hat{F}_{VV}(\lambda_m)\right\} = o\left(\left(\frac{m}{n}\right)^{2-4d_1}\right). \tag{3.35}$$

Since

$$\begin{aligned} n^2[\hat{F}_{XX}(\lambda_m) - E\{\hat{F}_{XX}(\lambda_m)\}] &= \sum_{t_1,t_2=1}^{n} \{X_{t_1}X_{t_2} - E(X_{t_1}X_{t_2})\}\, D_m(\lambda_{t_2-t_1}) \\ &= \sum_{t_1,t_2=1}^{n} \{Z_{t_1}Z_{t_2} - E(Z_{t_1}Z_{t_2})\}\, D_m(\lambda_{t_2-t_1}) \end{aligned}$$

by (3.32), we have

$$\mathrm{Var}\left\{\hat{F}_{XX}(\lambda_m)\right\} = \frac{1}{n^4} \sum_{t_1,t_2,t_3,t_4=1}^{n} \mathrm{Cov}(Z_{t_1}Z_{t_2}, Z_{t_3}Z_{t_4}) D_m(\lambda_{t_2-t_1})\overline{D_m(\lambda_{t_4-t_3})},$$

and therefore A is

$$\frac{1}{n^4} \sum_{t_1,t_2,t_3,t_4=1}^{n} \{\mathrm{Cov}(Z_{t_1}Z_{t_2}, Z_{t_3}Z_{t_4}) - \mathrm{Cov}(V_{t_1}V_{t_2}, V_{t_3}V_{t_4})\}\, D_m(\lambda_{t_2-t_1})\overline{D_m(\lambda_{t_4-t_3})}.$$

We now decompose A into sums where the time indices conform to cases (a) to (g) in Lemma 3.6. Using Lemmas 3.3 and 3.6 repeatedly, the approximation error for each case is bounded by:

(a)

$$\sum_{\alpha_1,\alpha_2\in Q_4:\alpha_1\neq\alpha_2} \frac{K}{n^4} \sum_{t_1,t_2,t_3,t_4=1}^{n} \gamma_{\alpha_1}^2|\gamma_{\alpha_2}||D_m(\lambda_{t_2-t_1})||D_m(\lambda_{t_4-t_3})|.$$

If α_1 is either $(1,2)$ or $(3,4)$, each element in the first summation is bounded by

$$\begin{aligned}
&\frac{K}{n^2}\sum_{j=1}^{n}\rho_j^2|D_m(\lambda_j)|\sum_{j=1}^{n}|\rho_j||D_m(\lambda_j)| + \frac{K}{n^3}\sum_{j=1}^{n}\rho_j^2|D_m(\lambda_j)|\sum_{j=1}^{n}|\rho_j|\sum_{j=1}^{n}|D_m(\lambda_j)| \\
&\leq K\left(\frac{m}{n}\right)^{2-2d_1}\left(1+\frac{\log m}{m^{1-2d_1}}\right)\left\{1+(\log n)1\left(d_1=\frac{1}{4}\right)+\left(\frac{n}{m}\right)^{4d_1-1}1\left(d_1>\frac{1}{4}\right)\right\},
\end{aligned}$$

while if α_1 is not equal to $(1,2)$ or to $(3,4)$, we have a bound

$$\begin{aligned}
&\frac{K}{n^3}\sum_{j=1}^{n}\rho_j^2\sum_{j=1}^{n}|\rho_j||D_m(\lambda_j)|\sum_{j=1}^{n}|D_m(\lambda_j)| + \frac{K}{n^4}\sum_{j=1}^{n}\rho_j^2\sum_{j=1}^{n}|\rho_j|\left\{\sum_{j=1}^{n}|D_m(\lambda_j)|\right\}^2 \\
&\leq K\left(\frac{m}{n}\right)^{2-2d_1}\frac{\log m}{m}\left(1+\frac{\log m}{m^{1-2d_1}}\right)\left\{1+(\log n)1\left(d_1=\frac{1}{4}\right)+n^{4d_1-1}1\left(d_1>\frac{1}{4}\right)\right\}.
\end{aligned}$$

(b)

$$\begin{aligned}
&\frac{K}{n^4}\sum_{t_1,t_2,t_3=1}^{n}(\gamma_{12}^2+\gamma_{13}^2+\gamma_{23}^2)|D_m(\lambda_{t_2-t_1})||D_m(\lambda_{t_3-t_1})| \\
\leq&\frac{K}{n^3}\sum_{j=1}^{n}\rho_j^2|D_m(\lambda_j)|\sum_{j=1}^{n}|D_m(\lambda_j)| + \frac{K}{n^4}\sum_{j=1}^{n}\rho_j^2\left\{\sum_{j=1}^{n}|D_m(\lambda_j)|\right\}^2 \\
\leq&K\left(\frac{m}{n}\right)^2\frac{\log m}{m}\left[\left(1+\frac{\log m}{m}\right)\left\{1+(\log n)1\left(d_1=\frac{1}{4}\right)\right\}\right. \\
&\left.+\left\{\left(\frac{n}{m}\right)^{4d_1-1}+n^{4d_1-1}\frac{\log m}{m}\right\}1\left(d_1>\frac{1}{4}\right)\right].
\end{aligned}$$

(c)

$$\begin{aligned}
&\frac{K}{n^4}\sum_{t_1,t_2,t_3=1}^{n}(\gamma_{12}^2+\gamma_{13}^2+\gamma_{23}^2)|D_m(0)||D_m(\lambda_{t_3-t_2})| \\
\leq&K\frac{m}{n^2}\sum_{j=1}^{n}\rho_j^2|D_m(\lambda_j)| + K\frac{m}{n^3}\sum_{j=1}^{n}\rho_j^2\sum_{j=1}^{n}|D_m(\lambda_j)|
\end{aligned}$$

$$\leq K\left(\frac{m}{n}\right)^2\left[\left(1+\frac{\log m}{n}\right)\left\{1+(\log n)1\left(d_1=\frac{1}{4}\right)\right\}\right.$$
$$\left.+\left\{\left(\frac{n}{m}\right)^{4d_1-1}+n^{4d_1-1}\frac{\log m}{n}\right\}1\left(d_1>\frac{1}{4}\right)\right].$$

(d), (e), (f) For any $a=a(t_1,t_2)$ and $b=b(t_1,t_2)$,

$$\frac{K}{n^4}\sum_{t_1,t_2=1}^{n}|\gamma_{12}||D_m(a)||D_m(b)|\leq K\frac{m^2}{n^3}\sum_{j=1}^{n}|\rho_j|\leq K\left(\frac{m}{n}\right)^2 n^{2d_1-1}.$$

(g)

$$\frac{K}{n^4}\sum_{t_1=1}^{n}|D_m(0)|^2\leq K\left(\frac{m}{n}\right)^2 n^{-1}.$$

Since cases (a) to (g) satisfy (3.35), the proof is complete. □

3.B Technical lemmas for Appendix 3.A

Lemma 3.1 *Let* $|\rho_j-\rho_{j+1}|\leq K|\rho_{j+1}|/j$ *and* $|\gamma_j-\gamma_{j+1}|\leq K|\gamma_{j+1}|/j$, *for all* $j\geq 1$. *Then, for any positive integers* r, s, *and* j,

$$|\rho_j^r-\rho_{j+1}^r|\leq K\frac{|\rho_{j+1}^r|}{j}, \tag{3.36}$$

$$|\rho_j^r\gamma_j^s-\rho_{j+1}^r\gamma_{j+1}^s|\leq K\frac{|\rho_{j+1}^r\gamma_{j+1}^s|}{j}. \tag{3.37}$$

Proof. First note that

$$(a-b)^k=\sum_{i=0}^{k}\binom{k}{i}a^i(-b)^{k-i}=\sum_{i=0}^{k}\binom{k}{i}(a^i-b^i)(-b)^{k-i},$$

since

$$\sum_{i=0}^{k}\binom{k}{i}b^i(-b)^{k-i}=(b-b)^k=0.$$

Hence,

$$|a^k - b^k| = \left|(a-b)^k - \sum_{i=1}^{k-1}\binom{k}{i}(a^i - b^i)(-b)^{k-i}\right|$$
$$\leq |a-b|^k + \sum_{i=1}^{k-1}\binom{k}{i}|b|^{k-i}|a^i - b^i|.$$

Proceeding by induction, suppose (3.36) holds for $r = 1, 2, \ldots, k-1$. Then

$$|\rho_j^k - \rho_{j+1}^k| \leq |\rho_j - \rho_{j+1}|^k + \sum_{i=1}^{k-1}\binom{k}{i}|\rho_{j+1}^{k-i}||\rho_j^i - \rho_{j+1}^i|$$
$$\leq K\frac{|\rho_{j+1}^k|}{j^k} + K\sum_{i=1}^{k-1}|\rho_{j+1}^{k-i}|\frac{|\rho_{j+1}^i|}{j} \leq K\frac{|\rho_{j+1}^k|}{j},$$

proving (3.36). To prove (3.37) we use (3.36):

$$|\rho_j^r\gamma_j^s - \rho_{j+1}^r\gamma_{j+1}^s| = |(\rho_j^r - \rho_{j+1}^r)(\gamma_j^s - \gamma_{j+1}^s) + \gamma_{j+1}^s(\rho_j^r - \rho_{j+1}^r) + \rho_{j+1}^r(\gamma_j^s - \gamma_{j+1}^s)|$$
$$\leq K\frac{|\rho_{j+1}^r|}{j}\frac{|\gamma_{j+1}^s|}{j} + K|\gamma_{j+1}^s|\frac{|\rho_{j+1}^r|}{j} + K|\rho_{j+1}^r|\frac{|\gamma_{j+1}^s|}{j} \leq K\frac{|\rho_{j+1}^r\gamma_{j+1}^s|}{j}. \quad \square$$

Lemma 3.2 *If* (a_1, b_1) *is independent of* (a_2, b_2) *and* $E(a_i^2 + b_i^2) < \infty$,

$$\operatorname{Cov}(a_1a_2, b_1b_2) = \operatorname{Cov}(a_1, b_1)E(a_2)E(b_2) + E(a_1b_1)\operatorname{Cov}(a_2, b_2)$$
$$= \operatorname{Cov}(a_1, b_1)E(a_2)E(b_2) + E(a_1)E(b_1)\operatorname{Cov}(a_2, b_2) + \operatorname{Cov}(a_1, b_1)\operatorname{Cov}(a_2, b_2).$$

Proof. By independence,

$$\operatorname{Cov}(a_1a_2, b_1b_2) = E(a_1a_2b_1b_2) - E(a_1a_2)E(b_1b_2)$$
$$= E(a_1b_1)E(a_2b_2) - E(a_1)E(a_2)E(b_1)E(b_2).$$

The proof follows by writing $E(a_i b_i) = \mathrm{Cov}(a_i, b_i) + E(a_i)E(b_i)$, for $i = 1, 2$, and rearranging terms. □

Lemma 3.3 *Let* $\rho_j = O(j^{2d-1})$, $d < 1/2$, $a > 0$, $b \geq 1$, $m \leq n/2$, *and* $d^+ = (a-1)/2a$. *Then,*

$$\sum_{j=1}^{n} |\rho_j|^a = O\left(1 + (\log n)1(d = d^+) + n^{a(2d-1)+1}1(d > d^+)\right),$$

$$\sum_{j=1}^{n} |D_m(\lambda_j)|^b = O\left(n\left\{\log m + m^{b-1}1(b > 1)\right\}\right),$$

$$\sum_{j=1}^{n} |\rho_j|^a |D_m(\lambda_j)|^b = O\left(m^b\left\{1 + (\log n)1(d = d^+) + \left(\frac{n}{m}\right)^{a(2d-1)+1} 1(d > d^+)\right\}\right).$$

Proof. The first equation follows from

$$\sum_{j=1}^{n} |\rho_j|^a \leq K \sum_{j=1}^{n} j^{a(2d-1)} \leq \begin{cases} K, & \text{if } d < d^+, \\ K \log n, & \text{if } d = d^+, \\ K n^{a(2d-1)+1}, & \text{if } d > d^+, \end{cases}$$

noting that $d = d^+$ is equivalent to $a(2d-1) = -1$.

From e.g. Zygmund (1977, p. 11) and an elementary inequality,

$$|D_m(\lambda_j)| \leq K \min\left\{m, \frac{n}{|j|}\right\}, \qquad |j| \leq \frac{n}{2}. \tag{3.38}$$

Using this bound repeatedly,

$$\begin{aligned} \sum_{j=1}^{n} |D_m(\lambda_j)|^b &\leq K \sum_{j=1}^{[n/m]} m^b + K \sum_{j=[n/m]+1}^{n} \left(\frac{n}{j}\right)^b \\ &\leq K n m^{b-1} + K n^b \sum_{j=[n/m]+1}^{n} j^{-b} \end{aligned}$$

$$\leq \begin{cases} Knm^{b-1} + Kn^b \log m, & \text{if } b = 1, \\ Knm^{b-1} + Km^b, & \text{if } b > 1, \end{cases}$$

and

$$\begin{aligned} \sum_{j=1}^{n} |\rho_j|^a |D_m(\lambda_j)|^b &\leq K \sum_{j=1}^{[n/m]} m^b j^{a(2d-1)} + K \sum_{j=[n/m]+1}^{n} \left(\frac{n}{j}\right)^b j^{a(2d-1)} \\ &\leq Km^b \sum_{j=1}^{[n/m]} j^{a(2d-1)} + Kn^b \sum_{j=[n/m]+1}^{n} j^{a(2d-1)-b} \\ &\leq \begin{cases} Km^b + Km^b(n/m)^{a(2d-1)+1}, & \text{if } d < d^+ \\ Km^b \log n + Km^b(n/m)^{a(2d-1)+1}, & \text{if } d = d^+ \\ Km^b(n/m)^{a(2d-1)+1} + Km^b(n/m)^{a(2d-1)+1}, & \text{if } d > d^+ \end{cases} \\ &\leq \begin{cases} Km^b + Km^b, & \text{if } d < d^+, \\ Km^b \log n + Km^b, & \text{if } d = d^+, \\ Km^b(n/m)^{a(2d-1)+1}, & \text{if } d > d^+. \end{cases} \end{aligned}$$

Note that the last sum retains the same form in all cases, because our assumptions imply that $a(2d-1) - b < -1$. □

Lemma 3.4 *For $j = 1, 2$, define $g_{j,t} = g_j(\mu_t)$, where μ_t is a standard Gaussian $I(d)$ process, and $\rho_t = E(\mu_0 \mu_t)$. Assume $E(g_{j,t}^2) < \infty$. Denote by $G_{j,k}$ the k-th Hermite coefficient of $g_j(\cdot)$, and let*

$$r = \min\{k \in \mathbb{N} : G_{1,k} G_{2,k} \neq 0\}. \tag{3.39}$$

If $d > 0$, define

$$d^* = \frac{1}{2} - r\left(\frac{1}{2} - d\right), \qquad C_\rho = \lim_{j \to \infty} \rho_j j^{1-2d}.$$

Let $A = S_m(g_1, g_2)$, where m satisfies Assumption 3.2 if $d^* = 1/(2r+2)$ or just

(3.24) otherwise. Then,

$$A = O\left(\frac{m}{n}\{1(d=0)+1(d^*<0)+(\log n)1(d^*=0)\}\right) + \left\{C^*\left(\frac{m}{n}\right)^{1-2d^*} + o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right)\right\}1(d^*>0), \tag{3.40}$$

where

$$C^* = 2\frac{(2\pi)^{-2d^*}\Gamma(2d^*)}{1-2d^*}\frac{G_{1,r}G_{2,r}}{r!}\sin\left\{(1-2d^*)\frac{\pi}{2}\right\}C_\rho^r \neq 0.$$

Proof. Let $\gamma_t = \mathrm{Cov}(g_{1,0}; g_{2,t})$. Then

$$A = \frac{1}{n^2}\sum_{s,t=1}^{n}\gamma_{t-s}D_m(\lambda_{t-s}) \tag{3.41}$$

$$= \frac{1}{n}\sum_{u=1-n}^{n-1}\left(1-\frac{|u|}{n}\right)\gamma_u D_m(\lambda_u). \tag{3.42}$$

We will make repeated use of (3.38) and of $\rho_u^r = Ku^{r(2d-1)} = O(u^{2d^*-1})$. By Theorem 3.1 and (3.39),

$$\gamma_u = \sum_{k=1}^{\infty}\frac{G_{1,k}G_{2,k}}{k!}\rho_u^k = C\rho_u^r + O(|\rho_u^{r+1}|),$$

where $C = G_{1,r}G_{2,r}/r!$.

(a) If $d = 0$, then $\gamma_u = O(|\rho_u^r|)$ are summable. Similarly, if $d^* < 0$, then $\gamma_u = O(|\rho_u^r|) = O(u^{2d^*-1})$ are summable. In either case,

$$A \leq \frac{K}{n}\sum_{u=1-n}^{n-1}\left(1-\frac{|u|}{n}\right)|\gamma_u||D_m(\lambda_u)|$$

$$\leq K\frac{m}{n}\sum_{u=1-n}^{n-1}|\gamma_u| = O\left(\frac{m}{n}\right). \tag{3.43}$$

(b) If $d^* = 0$, $\gamma_u = O(|\rho_u^r|) = O(u^{-1})$, hence

$$A \leq K\frac{m}{n}\sum_{u=1-n}^{n-1}|\gamma_u| = O\left(\frac{m}{n}\log n\right). \tag{3.44}$$

(c) If $d^* > 0$,

$$|\gamma_u - C\rho_u^r| \leq K|\rho_u^{r+1}| \leq K|\rho_u^r|^{1+\omega},$$

where $\omega = r^{-1}$. Defining

$$B_1 = \frac{1}{n}\sum_{u=1-n}^{n-1}\left(1 - \frac{|u|}{n}\right)\rho_u^r D_m(\lambda_u),$$

we get

$$\begin{aligned}
|A - CB_1| &\leq \frac{1}{n}\sum_{u=1-n}^{n-1}\left(1 - \frac{|u|}{n}\right)|\gamma_u - C\rho_u^r||D_m(\lambda_u)| \\
&\leq \frac{K}{n}\sum_{u=1-n}^{n-1}|\rho_u^r|^{1+\omega}|D_m(\lambda_u)| \\
&\leq K\frac{m}{n} + \frac{K}{n}\sum_{u=1}^{n}|\rho_u^r|^{1+\omega}|D_m(\lambda_u)|.
\end{aligned}$$

Therefore, setting $d^+ = \omega/(2+2\omega)$ in Lemma 3.3,

$$\begin{aligned}
|A - CB_1| &= O\left(\frac{m}{n}\left\{1 + (\log n)1(d^* = d^+) + \left(\frac{m}{n}\right)^{(1+\omega)(1-2d^*)-1}1(d^* > d^+)\right\}\right) \\
&= o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right),
\end{aligned}$$

choosing $0 < \epsilon < 2d^*$ in Assumption 3.2 if $d^* = d^+$. Now, write

$$\begin{aligned}
B_1 &= \frac{1}{n}\sum_{|u|<n}\left(1 - \frac{|u|}{n}\right)\rho_u^r D_m(\lambda_u) \\
&= \frac{1}{n}\sum_{|u|<n}\rho_u^r D_m(\lambda_u) - \frac{1}{n^2}\sum_{|u|<n}|u|\rho_u^r D_m(\lambda_u)
\end{aligned}$$

$$= \frac{1}{n}\sum_{u=-\infty}^{\infty} \rho_u^r D_m(\lambda_u) - \frac{1}{n}\sum_{|u|\geq n} \rho_u^r D_m(\lambda_u) - \frac{1}{n^2}\sum_{|u|<n} |u|\rho_u^r D_m(\lambda_u)$$

$$= B_2 + B_3 + B_4,$$

where

$$B_2 = \frac{2\pi}{n}\sum_{j=1}^{m} f(\lambda_j), \qquad f(\lambda_j) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} \rho_u^r e^{iu\lambda_j}, \tag{3.45}$$

$$B_3 = -\frac{1}{n}\sum_{u=n}^{\infty} \rho_u^r \left\{ D_m(\lambda_u) + \overline{D_m(\lambda_u)} \right\}, \tag{3.46}$$

$$B_4 = -\frac{1}{n^2}\sum_{u=1}^{n-1} u\rho_u^r \left\{ D_m(\lambda_u) + \overline{D_m(\lambda_u)} \right\}. \tag{3.47}$$

Then

$$|A - CB_2| = |A - C(B_1 - B_3 - B_4)| \leq |A - CB_1| + C|B_3| + C|B_4|. \tag{3.48}$$

Note that, for any u,

$$\left| \sum_{k=1}^{u} D_m(\lambda_k) \right| = \left| \sum_{k=1}^{u}\sum_{j=1}^{m} e^{ij\lambda_k} \right| \leq \sum_{j=1}^{m} |D_u(\lambda_j)| \leq K\sum_{j=1}^{m} \frac{n}{j} \leq Kn\log m. \tag{3.49}$$

Using summation by parts, (3.49), and Lemma 3.1 in (3.47),

$$\begin{aligned}
|B_4| &\leq \frac{K}{n^2}\sum_{u=1}^{n-1} \left\{ |u\rho_u^r - (u+1)\rho_{u+1}^r| \left| \sum_{k=1}^{u} D_m(\lambda_k) \right| \right\} + \frac{K}{n^2} \left| \sum_{j=1}^{n-1} D_m(\lambda_j) \right| n|\rho_n^r| \\
&\leq \frac{K}{n^2}\sum_{u=1}^{n-1} (u|\rho_u^r - \rho_{u+1}^r| + |\rho_{u+1}^r|) n\log m + Kn^{2d^*-1}\log m \\
&\leq K\frac{\log m}{n}\sum_{u=1}^{n-1} u^{2d^*-1} + Kn^{2d^*-1}\log m \\
&\leq Kn^{2d^*-1}\log m = o\left(\left(\frac{m}{n}\right)^{1-2d^*} \right).
\end{aligned} \tag{3.50}$$

Using the partial summation formula for infinite sums, (3.49), and Lemma 3.1 in (3.46),

$$
\begin{aligned}
|B_3| &\leq \frac{K}{n}\sum_{u=n}^{\infty}|\rho_u^r-\rho_{u+1}^r|\left|\sum_{k=1}^{u}D_m(\lambda_k)\right|+\frac{K}{n}\left|\sum_{j=1}^{n-1}D_m(\lambda_j)\right||\rho_n^r| \\
&\leq K\log m\sum_{u=n}^{\infty}u^{2d^*-2}+Kn^{2d^*-1}\log m \\
&\leq Kn^{2d^*-1}\log m = o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right). \qquad (3.51)
\end{aligned}
$$

Lemma 3.1 implies that $f(\lambda)\sim C_f\lambda^{-2d^*}$ as $\lambda\to 0^+$, where

$$
C_f=\pi^{-1}\Gamma(2d^*)\sin\left\{(1-2d^*)\frac{\pi}{2}\right\}C_\rho^r.
$$

Thus, by Proposition 1 in Robinson (1994a),

$$
\begin{aligned}
B_2 &= \frac{2\pi}{n}\sum_{j=1}^{m}f(\lambda_j)\sim\int_0^{\lambda_m}f(t)\mathrm{d}t \\
&\sim C_f\frac{\lambda_m^{1-2d^*}}{1-2d^*}\sim C_f\frac{(2\pi)^{1-2d^*}}{1-2d^*}\left(\frac{m}{n}\right)^{1-2d^*},
\end{aligned}
$$

which together with (3.48), (3.50), (3.51) gives (3.40). □

Lemma 3.5 *For $i,j=1,2$, define $g_{ij,t}=g_{ij}(\mu_{it})$, where μ_{it} is a standard Gaussian $I(d_i)$ process and $\rho_{i,t}=E(\mu_{i0}\mu_{it})$. Assume $E(g_{ij,t}^2)<\infty$. Denote by $G_{ij,k}$ the k-th Hermite coefficient of $g_{ij}(\cdot)$, with*

$$
r_i=\min\{k>0: G_{i1,k}G_{i2,k}\neq 0\}. \qquad (3.52)
$$

Let $d_1\geq d_2$ without loss of generality, and define

$$
d^*=\frac{1}{2}-r_1\left(\frac{1}{2}-d_1\right)-r_2\left(\frac{1}{2}-d_2\right), \qquad C_{i\rho}=\lim_{j\to\infty}\rho_{ij}j^{1-2d_i}.
$$

Let $A = S'_m(g_{11}, g_{12}; g_{21}, g_{22})$, *where* m *satisfies Assumption 3.2 if* $d^* + d_1 = 1/2$ *or just (3.24) otherwise. Then,*

$$A = O\left(\frac{m}{n}\{1 + (\log n)1(d^* = 0)\}\right) + \left\{C^*\left(\frac{m}{n}\right)^{1-2d^*} + o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right)\right\} 1(d^* > 0), \tag{3.53}$$

where

$$C^* = 2\frac{(2\pi)^{-2d^*}\Gamma(2d^*)}{1-2d^*}\frac{G_{11,r_1}G_{12,r_1}}{r_1!}\frac{G_{21,r_2}G_{22,r_2}}{r_2!}\sin\left\{(1-2d^*)\frac{\pi}{2}\right\} C_{1\rho}^{r_1}C_{2\rho}^{r_2} \neq 0.$$

Proof. Let $\gamma_{i,t} = \mathrm{Cov}(g_{i1,0}; g_{i2,t})$. Then, similarly to (3.41),

$$A = \frac{1}{n}\sum_{u=1-n}^{n-1}\left(1 - \frac{|u|}{n}\right)\gamma_{1u}\gamma_{2u}D_m(\lambda_u).$$

By Theorem 3.1 and (3.52),

$$\gamma_{iu} = \sum_{k=1}^{\infty}\frac{G_{i1,k}G_{i2,k}}{k!}\rho_{iu}^k = C_i\rho_{iu}^{r_i} + O(|\rho_{iu}^{r_i+1}|),$$

where $C_i = G_{i1,r_i}G_{i2,r_i}/r_i!$.

(a) If $d_1 d_2 = 0$, then $\gamma_{1u}\gamma_{2u} = O(|\rho_{1u}^{r_1}\rho_{2u}^{r_2}|)$ are summable. Similarly, if $d_1 d_2 > 0$ but $d^* < 0$, then $\gamma_{1u}\gamma_{2u} = O(|\rho_{1u}^{r_1}\rho_{2u}^{r_2}|) = O(u^{r_1(2d_1-1)+r_2(2d_2-1)}) = O(j^{2d^*-1})$ are summable. In either case, writing $\gamma_{1u}\gamma_{2u}$ instead of γ_u in (3.43) yields $A = O(m/n)$.

(b) If $d^* = 0$, $\gamma_{1u}\gamma_{2u} = O(j^{-1})$, hence (3.44) holds for $\gamma_{1u}\gamma_{2u}$ instead of γ_u.

(c) If $d^* > 0$,

$$\begin{aligned}|\gamma_{1u}\gamma_{2u} - C_1C_2\rho_{2u}^{r_1}\rho_{2u}^{r_2}| &\leq |\gamma_{1u}||\gamma_{2u} - C_2\rho_{2u}^{r_2}| + C_2|\rho_{2u}^{r_2}||\gamma_{1u} - C_1\rho_{1u}^{r_1}| \\ &\leq K|\rho_{1u}^{r_1}\rho_{2u}^{r_2+1}| + K|\rho_{2u}^{r_2}\rho_{1u}^{r_1+1}|\end{aligned}$$

$$\leq K|\rho_{1u}^{r_1+1}\rho_{1u}^{r_2}| \leq K|\rho_{1u}^{r_1}\rho_{2u}^{r_2}|^{1+\omega},$$

where $\omega = (1-2d_1)/(1-2d^*)$, since $d_1 \geq d_2$. Then (3.53) follows from the proof of case (c) of Lemma 3.4, writing $\rho_{1u}^{r_1}\rho_{2u}^{r_2}$ instead of ρ_u^r. □

Lemma 3.6 *Under (3.2) and Assumption 3.1, let* $Z_t = X_t - E(X_t)$. *For* t_1, t_2, t_3, t_4 *distinct, define* $\rho_{ij} = E(\eta_{2t_i}\eta_{2t_j})$ *and* $\rho'_{ij} = E(\nu_{2t_i}\nu_{2t_j})$. *Then there exists a mean-zero Gaussian* $I(d_1)$ *process* V_t *such that:*

(a) $\mathrm{Cov}(Z_{t_1}Z_{t_2}, Z_{t_3}Z_{t_4}) - \mathrm{Cov}(V_{t_1}V_{t_2}, V_{t_3}V_{t_4}) = O(\sum_{\alpha_1,\alpha_2\in Q_4:\alpha_1\neq\alpha_2}\rho_{\alpha_1}^2|\rho_{\alpha_2}|)$;

(b) $\mathrm{Cov}(Z_{t_1}Z_{t_2}, Z_{t_1}Z_{t_3}) - \mathrm{Cov}(V_{t_1}V_{t_2}, V_{t_1}V_{t_3}) = O(\rho_{12}^2+\rho_{13}^2+\rho_{23}^2)$;

(c) $\mathrm{Cov}(Z_{t_1}^2, Z_{t_2}Z_{t_3}) - \mathrm{Cov}(V_{t_1}^2, V_{t_2}V_{t_3}) = O(\rho_{12}^2+\rho_{13}^2+\rho_{23}^2)$;

(d) $\mathrm{Cov}(Z_{t_1}Z_{t_2}, Z_{t_1}Z_{t_2}) - \mathrm{Cov}(V_{t_1}V_{t_2}, V_{t_1}V_{t_2}) = O(|\rho_{12}|)$;

(e) $\mathrm{Cov}(Z_{t_1}^2, Z_{t_2}^2) - \mathrm{Cov}(V_{t_1}^2, V_{t_2}^2) = O(|\rho_{12}|)$;

(f) $\mathrm{Cov}(Z_{t_1}^2, Z_{t_1}Z_{t_2}) - \mathrm{Cov}(V_{t_1}^2, V_{t_1}V_{t_2}) = O(|\rho_{12}|)$;

(g) $\mathrm{Cov}(Z_{t_1}^2, Z_{t_1}^2) - \mathrm{Cov}(V_{t_1}^2, V_{t_1}^2) = O(1)$.

Proof. By the law of iterated expectations, all Z_t covariances in (a) to (g) can be written as linear combinations of

$$E\left(\prod_{i=1}^4 Z_{t_i}^{k_i}\right) = E\left\{E\left(\prod_{i=1}^4 Z_{t_i}^{k_i}\middle| g_{t_j}, h_{t_j},\ j=1,\ldots,4\right)\right\}$$
$$= E\left\{\prod_{i=1}^4 E(Z_{t_i}^{k_i}|g_{t_j}, h_{t_j},\ j=1,\ldots,4)\right\} = E\left\{\prod_{i=1}^4 E(Z_{t_i}^{k_i}|g_{t_i}, h_{t_i})\right\}, \qquad (3.54)$$

where, conditionally on g_s and h_s, Z_s is independent of Z_t, g_t, and h_t, for any $t \neq s$.

In what follows, let s_i, $i = 1,\ldots,4$ denote (not necessarily distinct) elements of $\{t_1, t_2, t_3, t_4\}$. Wherever u_i and s_i are both defined, let

$$A_i = \binom{p}{u_i}\beta_2^{u_i}\eta_{1s_i}^{u_i}\nu_{1s_i}^{p-u_i}, \qquad B_i = g_{s_i}^{u_i}h_{s_i}^{p-u_i},$$

and define $c_i = E(A_i)$, $c_{ij} = E(\tilde{A}_i\tilde{A}_j)$, $c_{ijk} = E(\tilde{A}_i\tilde{A}_j\tilde{A}_k)$, where throughout the proof $\tilde{z}_t^u = z_t^u - E(z_t^u)$. We first compute $E(Z_t^k|g_t, h_t)$ for $k = 1, \ldots, 3$. Setting $s_1 = s_2 = s_3 = t$, but omitting time subscripts for convenience,

$$\begin{aligned} Z &= \sum_{u_1=0}^{p} \{A_1B_1 - E(A_1B_1)\} \\ &= \sum_{u_1=0}^{p} \left\{\tilde{A}_1B_1 + c_1B_1 - c_1E(B_1)\right\} \\ &= \sum_{u_1=0}^{p} (\tilde{A}_1B_1 + c_1\tilde{B}_1). \end{aligned}$$

Therefore, independence of A_1 and B_1 yields

$$E(Z|g, h) = \sum_{u_1=0}^{p} c_1\tilde{B}_1. \tag{3.55}$$

Similarly,

$$Z^k = \prod_{i=1}^{k}\sum_{u_i=0}^{p}(\tilde{A}_iB_i + c_i\tilde{B}_i),$$

so by independence of A_i and B_j, for all i, j

$$E(Z^2|g, h) = \sum_{u_1,u_2=0}^{p} (c_1c_2\tilde{B}_1\tilde{B}_2 + c_{12}B_1B_2), \tag{3.56}$$

$$\begin{aligned} E(Z^3|g, h) = \sum_{u_1,u_2,u_3=0}^{p} &(c_1c_2c_3\tilde{B}_1\tilde{B}_2\tilde{B}_3 + c_{123}B_1B_2B_3 \\ &+ c_{12}c_3B_1B_2\tilde{B}_3 + c_{13}c_2B_1B_3\tilde{B}_2 + c_{23}c_1B_2B_3\tilde{B}_1). \end{aligned} \tag{3.57}$$

Unless otherwise noted, we will use $\sum$ to mean $\sum_{u_1,u_2,u_3,u_4=0}^{p}$ for the remainder of the proof. Using (3.55), (3.56), (3.57) in (3.54), we can write $E(\prod_{i=1}^{4} Z_{s_i})$ as follows:

(i) if $s_i = t_i$, $i = 1, \ldots, 4$,

$$E(Z_{t_1} Z_{t_2} Z_{t_3} Z_{t_4}) = \sum c_1 c_2 c_3 c_4 E(\tilde{B}_1 \tilde{B}_2 \tilde{B}_3 \tilde{B}_4); \tag{3.58}$$

(ii) if $s_1 = s_2 = t_1$, $s_3 = t_2$, $s_4 = t_3$,

$$E(Z_{t_1}^2 Z_{t_2} Z_{t_3}) = \sum \left\{ c_1 c_2 c_3 c_4 E(\tilde{B}_1 \tilde{B}_2 \tilde{B}_3 \tilde{B}_4) + c_{12} c_3 c_4 E(B_1 B_2 \tilde{B}_3 \tilde{B}_4) \right\}; \tag{3.59}$$

(iii) if $s_1 = s_2 = t_1$, $s_3 = s_4 = t_2$,

$$\begin{aligned} E(Z_{t_1}^2 Z_{t_2}^2) = \sum \Big\{ & c_1 c_2 c_3 c_4 E(\tilde{B}_1 \tilde{B}_2 \tilde{B}_3 \tilde{B}_4) + c_{12} c_{34} E(B_1 B_2 B_3 B_4) \\ & + c_{12} c_3 c_4 E(B_1 B_2 \tilde{B}_3 \tilde{B}_4) + c_1 c_2 c_{34} E(\tilde{B}_1 \tilde{B}_2 B_3 B_4) \Big\}; \end{aligned} \tag{3.60}$$

(iv) if $s_1 = s_2 = s_3 = t_1$, $s_4 = t_2$,

$$\begin{aligned} E(Z_{t_1}^3 Z_{t_2}) = \sum \Big\{ & c_1 c_2 c_3 c_4 E(\tilde{B}_1 \tilde{B}_2 \tilde{B}_3 \tilde{B}_4) + c_{123} c_4 E(B_1 B_2 B_3 \tilde{B}_4) \\ & + c_{12} c_3 c_4 E(B_1 B_2 \tilde{B}_3 \tilde{B}_4) + c_{13} c_2 c_4 E(B_1 B_3 \tilde{B}_2 \tilde{B}_4) \\ & + c_{23} c_1 c_4 E(B_2 B_3 \tilde{B}_1 \tilde{B}_4) \Big\}. \end{aligned} \tag{3.61}$$

We can also write $E(Z_{s_1} Z_{s_2})$ as follows:

(v) if $s_1 = t_i$, $s_2 = t_j$, $i \neq j$,

$$E(Z_{t_i} Z_{t_j}) = \sum_{u_1, u_2 = 0}^{p} c_1 c_2 E(\tilde{B}_1 \tilde{B}_2); \tag{3.62}$$

(vi) if $s_1 = s_2 = t_i$,

$$E(Z_{t_i}^2) = \sum_{u_1, u_2 = 0}^{p} \left\{ c_1 c_2 E(\tilde{B}_1 \tilde{B}_2) + c_{12} E(B_1 B_2) \right\}. \tag{3.63}$$

We now proceed to expand B_i and $\tilde{B}_i$ in terms of g and h. Wherever u_i and s_i are both defined, we use the following notation: $\chi_i = \tilde{g}_{s_i}^{u_i}$, $\psi_i = \tilde{h}_{s_i}^{p-u_i}$; $\chi_{ij} = g_{s_i}^{u_i+u_j}$, $\psi_{ij} = h_{s_i}^{2p-u_i-u_j}$; $\chi_{123} = g_{s_1}^{u_1+u_2+u_3}$, $\psi_{123} = h_{s_1}^{3p-u_1-u_2-u_3}$; $\chi_{12,34} = g_{s_1}^{u_1+u_2} g_{s_3}^{u_3+u_4}$, $\psi_{12,34} = h_{s_1}^{2p-u_1-u_2} h_{s_3}^{2p-u_3-u_4}$. Note that

$$\tilde{B}_i = c_{h,i}\chi_i + \chi_i\psi_i + c_{g,i}\psi_i, \tag{3.64}$$

where $c_{h,i} = E(h_{s_i}^{p-u_i})$, $c_{g,i} = E(g_{s_i}^{u_i})$. Four forms of expectations need to be accounted for in (3.58) to (3.61).

1. $E(\tilde{B}_1\tilde{B}_2\tilde{B}_3\tilde{B}_4)$ will be a linear combination of 81 terms, all of them expectations of products of χ_i and ψ_i, $i = 1,\ldots,4$. Denoting by $\langle 1\rangle$, $\langle 2\rangle$, $\langle 3\rangle$, $\langle 4\rangle$ any permutation of P_4, those terms can be separated into the following categories: terms that vanish due to $E(\chi_i) = E(\psi_i) = 0$, namely $E(\chi_{\langle 1\rangle}\psi_{\langle 2\rangle}\psi_{\langle 3\rangle}\psi_{\langle 4\rangle})$, $E(\psi_{\langle 1\rangle}\chi_{\langle 2\rangle}\chi_{\langle 3\rangle}\chi_{\langle 4\rangle})$, $E(\chi_{\langle 1\rangle}\psi_1\psi_2\psi_3\psi_4)$, and $E(\psi_{\langle 1\rangle}\chi_1\chi_2\chi_3\chi_4)$; non-vanishing terms with four factors, namely $E(\chi_1\chi_2\chi_3\chi_4)$, $E(\psi_1\psi_2\psi_3\psi_4)$, and $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\psi_{\langle 3\rangle}\psi_{\langle 4\rangle})$; non-vanishing terms with five factors, namely $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\psi_{\langle 1\rangle}\psi_{\langle 3\rangle}\psi_{\langle 4\rangle})$ and $E(\psi_{\langle 1\rangle}\psi_{\langle 2\rangle}\chi_{\langle 1\rangle}\chi_{\langle 3\rangle}\chi_{\langle 4\rangle})$; terms with six factors, namely $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\psi_1\psi_2\psi_3\psi_4)$, $E(\psi_{\langle 1\rangle}\psi_{\langle 2\rangle}\chi_1\chi_2\chi_3\chi_4)$, and $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\chi_{\langle 3\rangle}\psi_{\langle 1\rangle}\psi_{\langle 2\rangle}\psi_{\langle 4\rangle})$; terms with seven factors, namely $E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle}\chi_{\langle 3\rangle}\psi_1\psi_2\psi_3\psi_4)$ and $E(\psi_{\langle 1\rangle}\psi_{\langle 2\rangle}\psi_{\langle 3\rangle}\chi_1\chi_2\chi_3\chi_4)$; a term with eight factors, namely $E(\chi_1\chi_2\chi_3\chi_4\psi_1\psi_2\psi_3\psi_4)$. It can be seen from (3.64) that, for each $i = 1,\ldots,4$, the corresponding coefficient will include a factor $c_{h,i}$ if only χ_i is present or $c_{g,i}$ if only ψ_i is present.

2. $E(B_1B_2\tilde{B}_3\tilde{B}_4)$ will be a linear combination of 9 terms. Denoting by $\langle 3\rangle$, $\langle 4\rangle$ any permutation of $\{3,4\}$, these can be grouped in the following categories: $E(\chi_{12}\psi_{12}\psi_3\psi_4)$, $E(\psi_{12}\chi_{12}\chi_3\chi_4)$, $E(\chi_{12}\chi_{\langle 3\rangle}\psi_{12}\psi_{\langle 4\rangle})$, $E(\chi_{12}\chi_{\langle 3\rangle}\psi_{12}\psi_3\psi_4)$, $E(\psi_{12}\psi_{\langle 3\rangle}\chi_{12}\chi_3\chi_4)$, and $E(\chi_{12}\chi_3\chi_4\psi_{12}\psi_3\psi_4)$. As in the previous case, (3.64) implies that, for each $i = 3,4$, the corresponding coefficient will include a factor $c_{h,i}$

if only χ_i is present or $c_{g,i}$ if only ψ_i is present.

3. $E(B_1B_2B_3\tilde{B}_4) = c_{h,4}E(\chi_{123}\chi_4\psi_{123}) + c_{g,4}E(\chi_{123}\psi_{123}\psi_4) + E(\chi_{123}\chi_4\psi_{123}\psi_4)$.

4. $E(B_1B_2B_3B_4) = E(\chi_{12,34}\psi_{12,34})$.

In (3.62) and (3.63), the relevant expectations are $E(\tilde{B}_1\tilde{B}_2) = c_{h,1}c_{h,2}E(\chi_1\chi_2)+$ $c_{g,1}c_{g,2}E(\psi_1\psi_2) + E(\chi_1\chi_2\psi_1\psi_2)$ and $E(B_1B_2) = E(\chi_{12}\psi_{12})$.

We can now use Theorem 3.1 to expand each these expectations as $\sum_{q=0}^{\infty} a_q$. Let μ_t represent either η_{2t} or ν_{2t}, with $\gamma_{ij} = E(\mu_{t_i}\mu_{t_j})$, and define $f_{i,t} = f_i(\mu_t)$, $f^*_{ij,t} = f^*_{ij}(\mu_t)$ such that $E(f_{i,t}) = 0$. Denote by $G_{i;q}$, $G_{ij;q}$, $G_{ijk;q}$, $G^*_{ij;q}$ the q-th Hermite coefficient of $f_{i,t}$, $f_{i,t}f_{j,t}$, $f_{i,t}f_{j,t}f_{k,t}$, $f^*_{ij,t}$ respectively.

For $E(f_{1,t_1}f_{2,t_2}f_{3,t_3}f_{4,t_4})$, we have

$$a_q = \sum_{\substack{v_\alpha \geq 0: \\ \Sigma v_\alpha = q, \alpha \in Q_4}} \prod_{i=1}^{4} G_{i;w_i} \prod_{\alpha \in Q_4} \frac{\gamma_\alpha^{v_\alpha}}{v_\alpha!}, \qquad w_k = \sum_{\alpha \in R_{4,k}} v_\alpha,$$

$$a_0 = a_1 = 0, \qquad a_2 = G_{1;1}G_{2;1}G_{3;1}G_{4;1}(\gamma_{12}\gamma_{34} + \gamma_{13}\gamma_{24} + \gamma_{14}\gamma_{23}).$$

Since $G_{i;0} = 0$, Theorem 3.1 yields

$$\sum_{q=3}^{\infty} |a_q| \leq K \prod_{i=1}^{4} \left(\sum_{\alpha \in R_{4,i}} |\gamma_\alpha| \right)^{\frac{1}{2}} \sum_{\alpha \in Q_4} |\gamma_\alpha|.$$

Label the elements of P_4 as $\langle 1\rangle$, $\langle 2\rangle$, $\langle 3\rangle$, $\langle 4\rangle$, such that $|\gamma_{\langle 1\rangle\langle 2\rangle}|$ is the largest absolute correlation. Then $\sum_{\alpha \in Q_4} |\gamma_\alpha| \leq K|\gamma_{\langle 1\rangle\langle 2\rangle}|$ and

$$\prod_{i=1}^{4} \sum_{\alpha \in R_{4,i}} |\gamma_\alpha| \leq K\gamma^2_{\langle 1\rangle\langle 2\rangle}(|\gamma_{\langle 1\rangle\langle 3\rangle}| + |\gamma_{\langle 2\rangle\langle 3\rangle}| + |\gamma_{\langle 3\rangle\langle 4\rangle}|)(|\gamma_{\langle 1\rangle\langle 4\rangle}| + |\gamma_{\langle 2\rangle\langle 4\rangle}| + |\gamma_{\langle 3\rangle\langle 4\rangle}|).$$

Choosing the second largest absolute correlation, we have a bound of the form

$$\sum_{q=4}^{\infty} |a_q| \leq K\gamma^2_{\langle 1\rangle\langle 2\rangle}|\gamma_{\langle 3\rangle\langle 4\rangle}| \qquad \text{or} \qquad \sum_{q=4}^{\infty} |a_q| \leq K\gamma^2_{\langle 1\rangle\langle 2\rangle}|\gamma_{\langle 1\rangle\langle 3\rangle}|.$$

Therefore, taking all possible permutations for $\langle 1\rangle$, $\langle 2\rangle$, $\langle 3\rangle$, $\langle 4\rangle$, $\sum_{q=3}^{\infty}|a_q| = O(e_3)$, where $e_3 = \sum_{\alpha_1,\alpha_2\in Q_4:\alpha_1\neq\alpha_2}\gamma_{\alpha_1}^2|\gamma_{\alpha_2}|$, yielding

$$E(f_{1,t_1}f_{2,t_2}f_{3,t_3}f_{4,t_4}) = \prod_{i=1}^{4} G_{i;1}(\gamma_{12}\gamma_{34}+\gamma_{13}\gamma_{24}+\gamma_{14}\gamma_{23}) + O(e_3). \tag{3.65}$$

Again from Theorem 3.1 but using Corollary 3.1, defining $e_2 = \gamma_{12}^2+\gamma_{13}^2+\gamma_{23}^2$,

$$E(f_{1,t_1}f_{2,t_1}f_{3,t_2}f_{4,t_3}) = G_{12;0}G_{3;1}G_{4;1}\gamma_{23} + O(e_2), \tag{3.66}$$

$$E(f_{1,t_1}f_{2,t_1}f_{3,t_2}f_{4,t_2}) = G_{12;0}G_{34;0} + O(|\gamma_{12}|), \tag{3.67}$$

$$E(f_{1,t_1}f_{2,t_1}f_{3,t_1}f_{4,t_2}) = O(|\gamma_{12}|), \tag{3.68}$$

$$E(f_{1,t_1}f_{2,t_2}f_{3,t_3}) = O(e_2), \tag{3.69}$$

$$E(f_{1,t_1}f_{2,t_1}f_{3,t_2}) = G_{12;1}G_{3;1}\gamma_{12} + O(\gamma_{12}^2), \tag{3.70}$$

$$E(f_{1,t_1}f_{2,t_1}f_{3,t_1}) = G_{123;0}, \tag{3.71}$$

$$E(f_{1,t_1}f_{2,t_2}) = G_{1;1}G_{2;1}\gamma_{12} + O(\gamma_{12}^2), \tag{3.72}$$

$$E(f_{1,t_1}f_{2,t_1}) = G_{12;0}, \tag{3.73}$$

$$E(f^*_{12,t_1}f_{3,t_2}f_{4,t_3}) = G^*_{12;0}G_{3;1}G_{4;1}\gamma_{23} + O(e_2), \tag{3.74}$$

$$E(f^*_{12,t_1}f_{3,t_2}f_{4,t_2}) = G^*_{12;0}G_{34;0} + O(|\gamma_{12}|), \tag{3.75}$$

$$E(f^*_{12,t_1}f_{3,t_2}) = G^*_{12;1}G_{3;1}\gamma_{12} + O(\gamma_{12}^2), \tag{3.76}$$

$$E(f^*_{12,t_1}) = G^*_{12;0}, \tag{3.77}$$

$$E(f^*_{12,t_1}f^*_{34,t_2}) = G^*_{12;0}G^*_{34;0} + O(|\gamma_{12}|). \tag{3.78}$$

Now let the G and G^* coefficients in (3.65) to (3.78) apply to the case

$$f_{i,t} = \tilde{g}_t^{u_i}, \qquad f^*_{ij,t} = g_t^{u_i+u_j}, \tag{3.79}$$

while corresponding G' and $G^{*\prime}$ coefficients apply to

$$f_{i,t} = \tilde{h}_t^{p-u_i}, \qquad f^*_{ij,t} = h_t^{2p-u_i-u_j}. \tag{3.80}$$

We can approximate each term in the expansion of (3.58) to (3.63) using (3.65) to (3.78):

(i) $E(Z_{t_1}Z_{t_2}Z_{t_3}Z_{t_4})$. Denote by $\langle 1\rangle$, $\langle 2\rangle$, $\langle 3\rangle$, $\langle 4\rangle$ any permutation of P_4. Using (3.65), (3.69), (3.72), the only terms that are not $O(e_3)$ are:

$$\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &= G_{1;1}G_{2;1}G_{3;1}G_{4;1}(\rho_{12}\rho_{34}+\rho_{13}\rho_{24}+\rho_{14}\rho_{23})+O(e_3),\\
E(\psi_1\psi_2\psi_3\psi_4) &= G'_{1;1}G'_{2;1}G'_{3;1}G'_{4;1}(\rho'_{12}\rho'_{34}+\rho'_{13}\rho'_{24}+\rho'_{14}\rho'_{23})+O(e_3),\\
E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle})E(\psi_{\langle 3\rangle}\psi_{\langle 4\rangle}) &= G_{\langle 1\rangle;1}G_{\langle 2\rangle;1}G'_{\langle 3\rangle;1}G'_{\langle 4\rangle;1}\rho_{\langle 1\rangle\langle 2\rangle}\rho'_{\langle 3\rangle\langle 4\rangle}+O(e_3).
\end{aligned}$$

(ii) $E(Z_{t_1}^2Z_{t_2}Z_{t_3})$. Using (3.66), (3.69), (3.70), (3.72), (3.73), the only terms in $E(\tilde{B}_1\tilde{B}_2\tilde{B}_3\tilde{B}_4)$ that are not $O(e_2)$ are:

$$\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &= G_{12;0}G_{3;1}G_{4;1}\rho_{23}+O(e_2),\\
E(\psi_1\psi_2\psi_3\psi_4) &= G'_{12;0}G'_{3;1}G'_{4;1}\rho'_{23}+O(e_2),\\
E(\chi_1\chi_2)E(\psi_3\psi_4) &= G_{12;0}G'_{3;1}G'_{4;1}\rho'_{23}+O(e_2),\\
E(\psi_1\psi_2)E(\chi_3\chi_4) &= G'_{12;0}G_{3;1}G_{4;1}\rho_{23}+O(e_2),\\
E(\chi_1\chi_2)E(\psi_1\psi_2\psi_3\psi_4) &= G_{12;0}G'_{12;0}G'_{3;1}G'_{4;1}\rho'_{23}+O(e_2),\\
E(\psi_1\psi_2)E(\chi_1\chi_2\chi_3\chi_4) &= G'_{12;0}G_{12;0}G_{3;1}G_{4;1}\rho_{23}+O(e_2).
\end{aligned}$$

Using (3.74), (3.76), (3.77) the only terms in $E(B_1B_2\tilde{B}_3\tilde{B}_4)$ that are not $O(e_2)$ are:

$$E(\chi_{12})E(\psi_{12}\psi_3\psi_4) = G^*_{12,0}G^{*\prime}_{12;0}G'_{3;1}G'_{4;1}\rho'_{23}+O(e_2),$$

$$E(\psi_{12})E(\chi_{12}\chi_3\chi_4) = G^{*\prime}_{12;0}G^{*}_{12;0}G_{3;1}G_{4;1}\rho_{23} + O(e_2).$$

(iii) $E(Z^2_{t_1}Z^2_{t_2})$. From (3.67), (3.70), (3.72), (3.73) it follows that all terms in $E(\tilde{B}_1\tilde{B}_2\tilde{B}_3\tilde{B}_4)$ will be $O(|\rho_{12}|)$ except the ones only involving (3.67) and (3.73):

$$\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &= G_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\psi_1\psi_2\psi_3\psi_4) &= G'_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_1\chi_2)E(\psi_3\psi_4) &= G_{12;0}G'_{34;0},\\
E(\psi_1\psi_2)E(\chi_3\chi_4) &= G'_{12;0}G_{34;0},\\
E(\chi_1\chi_2)E(\psi_1\psi_2\psi_3\psi_4) &= G_{12;0}G'_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_3\chi_4)E(\psi_1\psi_2\psi_3\psi_4) &= G_{34;0}G'_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\psi_1\psi_2)E(\chi_1\chi_2\chi_3\chi_4) &= G'_{12;0}G_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\psi_3\psi_4)E(\chi_1\chi_2\chi_3\chi_4) &= G'_{34;0}G_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\chi_1\chi_2\chi_3\chi_4)E(\psi_1\psi_2\psi_3\psi_4) &= G_{12;0}G_{34;0}G'_{12;0}G'_{34;0} + O(|\rho_{12}|).
\end{aligned}$$

Similarly, from (3.75), (3.76), (3.77), (3.78), all terms in $E(B_1B_2\tilde{B}_3\tilde{B}_4)$, $E(\tilde{B}_1\tilde{B}_2B_3B_4)$, and $E(B_1B_2B_3B_4)$ will be $O(|\rho_{12}|)$ except the following:

$$\begin{aligned}
E(\chi_{12})E(\psi_{12}\psi_3\psi_4) &= G^{*}_{12;0}G^{*\prime}_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_{34})E(\psi_{34}\psi_1\psi_2) &= G^{*}_{12;0}G^{*\prime}_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\psi_{12})E(\chi_{12}\chi_3\chi_4) &= G^{*\prime}_{12;0}G^{*}_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\psi_{34})E(\chi_{34}\chi_1\chi_2) &= G^{*\prime}_{12;0}G^{*}_{12;0}G_{34;0} + O(|\rho_{12}|),\\
E(\chi_{12}\chi_3\chi_4)E(\psi_{12}\psi_3\psi_4) &= G^{*}_{12;0}G_{34;0}G^{*\prime}_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_{34}\chi_1\chi_2)E(\psi_{34}\psi_1\psi_2) &= G^{*}_{12;0}G_{34;0}G^{*\prime}_{12;0}G'_{34;0} + O(|\rho_{12}|),\\
E(\chi_{12,34})E(\psi_{12,34}) &= G^{*}_{12;0}G^{*}_{34;0}G^{*\prime}_{12;0}G^{*\prime}_{34;0} + O(|\rho_{12}|).
\end{aligned}$$

(iv) $E(Z_{t_1}^3 Z_{t_2})$. Using (3.68), (3.70), (3.71), (3.72), (3.73) in $E(\tilde{B}_1\tilde{B}_2\tilde{B}_3\tilde{B}_4)$ note that at least one factor in each term necessarily involves t_2. Therefore, one of (3.68), (3.70), (3.72) will apply, making all terms $O(|\rho_{12}|)$.

Similarly, in $E(B_1B_2\tilde{B}_3\tilde{B}_4)$, $E(B_1\tilde{B}_2B_3\tilde{B}_4)$, $E(\tilde{B}_1B_2B_3\tilde{B}_4)$, $E(B_1B_2B_3\tilde{B}_4)$, at least one factor in each term necessarily involves t_2. Thus, (3.76) will apply for some function $f^*_{ij,t}$, not necessarily one given in (3.79) or (3.80), making all terms $O(|\rho_{12}|)$.

(v) $E(Z_{t_i}Z_{t_j})$. Using (3.72), the following are not $O(\rho_{ij}^2)$:

$$E(\chi_1\chi_2) = G_{1;1}G_{2;1}\rho_{ij} + O(\rho_{ij}^2), \qquad E(\psi_1\psi_2) = G'_{1;1}G'_{2;1}\rho'_{ij} + O(\rho_{ij}^2).$$

(vi) $E(Z_{t_i}^2)$. Using (3.73), $E(\tilde{B}_1\tilde{B}_2)$ and $E(B_1B_2)$ include the following terms:

$$E(\chi_1\chi_2) = G_{12;0}, \qquad E(\chi_1\chi_2)E(\psi_1\psi_2) = G_{12;0}G'_{12;0},$$
$$E(\psi_1\psi_2) = G'_{12;0}, \qquad E(\chi_{12})E(\psi_{12}) = G^*_{12;0}G^{*\prime}_{12;0}.$$

We now compute the coefficients of the leading terms listed above. Define

$$L_i = c_i c_{h,i} G_{i;1}, \qquad \bar{L}_1 = \sum_{u_i=0}^{p} L_i, \qquad L'_i = c_i c_{g,i} G'_{i;1}, \qquad \bar{L}'_1 = \sum_{u_i=0}^{p} L'_i;$$
$$L_{ij} = c_i c_{h,i} c_j c_{h,j} G_{ij;0}, \quad \bar{L}_2 = \sum_{u_i,u_j=0}^{p} L_{ij}, \quad L'_{ij} = c_i c_{g,i} c_j c_{g,j} G'_{ij;0}, \quad \bar{L}'_2 = \sum_{u_i,u_j=0}^{p} L'_{ij};$$
$$L^*_{ij} = c_i c_j G_{ij;0} G'_{ij;0}, \quad \bar{L}^*_2 = \sum_{u_i,u_j=0}^{p} L^*_{ij}, \quad L^{**}_{ij} = c_{ij} G^*_{ij;0} G^{*\prime}_{ij;0}, \quad \bar{L}^{**}_2 = \sum_{u_i,u_j=0}^{p} L^{**}_{ij}.$$

Note that $L_p = c_p c_{h,p} G_{p;1} = \beta_2^p E(\eta_{1t}^p) E\{g^p(\eta_{2t}) H_1(\eta_{2t})\} \neq 0$ by assumption, but $L_i = 0$ for any $i < p$. Hence $\bar{L}_1 = L_p \neq 0$. The contributions of the non-negligible terms will be:

(i) $E(Z_{t_1}Z_{t_2}Z_{t_3}Z_{t_4})$

$$\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &: L_1L_2L_3L_4(\rho_{12}\rho_{34}+\rho_{13}\rho_{24}+\rho_{14}\rho_{23})+O(e_3);\\
E(\psi_1\psi_2\psi_3\psi_4) &: L'_1L'_2L'_3L'_4(\rho'_{12}\rho'_{34}+\rho'_{13}\rho'_{24}+\rho'_{14}\rho'_{23})+O(e_3);\\
E(\chi_{\langle 1\rangle}\chi_{\langle 2\rangle})E(\psi_{\langle 3\rangle}\psi_{\langle 4\rangle}) &: L_{\langle 1\rangle}L_{\langle 2\rangle}L'_{\langle 3\rangle}L'_{\langle 4\rangle}\rho_{\langle 1\rangle\langle 2\rangle}\rho'_{\langle 3\rangle\langle 4\rangle}+O(e_3).
\end{aligned}$$

Thus, $E(Z_{t_1}Z_{t_2}Z_{t_3}Z_{t_4})$ is

$$\begin{aligned}
&\sum\{L_1L_2L_3L_4(\rho_{12}\rho_{34}+\rho_{13}\rho_{24}+\rho_{14}\rho_{23})+L'_1L'_2L'_3L'_4(\rho'_{12}\rho'_{34}+\rho'_{13}\rho'_{24}+\rho'_{14}\rho'_{23})\\
&+L_1L_2L'_3L'_4\rho_{12}\rho'_{34}+L'_1L'_2L_3L_4\rho'_{12}\rho_{34}+L_1L'_2L_3L'_4\rho_{13}\rho'_{24}+L'_1L_2L'_3L_4\rho'_{13}\rho_{24}\\
&+L_1L'_2L'_3L_4\rho_{14}\rho'_{23}+L'_1L_2L_3L'_4\rho'_{14}\rho_{23}\}+O(e_3)\\
=&\sum\{(L_1L_2\rho_{12}+L'_1L'_2\rho'_{12})(L_3L_4\rho_{34}+L'_3L'_4\rho'_{34})\\
&+(L_1L_3\rho_{13}+L'_1L'_3\rho'_{13})(L_2L_4\rho_{24}+L'_2L'_4\rho'_{24})\\
&+(L_1L_4\rho_{14}+L'_1L'_4\rho'_{14})(L_2L_3\rho_{23}+L'_2L'_3\rho'_{23})\}+O(e_3)\\
=&\left(\bar{L}_1^2\rho_{12}+\bar{L}_1'^2\rho'_{12}\right)\left(\bar{L}_1^2\rho_{34}+\bar{L}_1'^2\rho'_{34}\right)+\left(\bar{L}_1^2\rho_{13}+\bar{L}_1'^2\rho'_{13}\right)\left(\bar{L}_1^2\rho_{24}+\bar{L}_1'^2\rho'_{24}\right)\\
&+\left(\bar{L}_1^2\rho_{14}+\bar{L}_1'^2\rho'_{14}\right)\left(\bar{L}_1^2\rho_{23}+\bar{L}_1'^2\rho'_{23}\right)+O(e_3).
\end{aligned}\tag{3.81}$$

(ii) $E(Z_{t_1}^2Z_{t_2}Z_{t_3})$

$$\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &: L_{12}L_3L_4\rho_{23}+O(e_2);\\
E(\chi_1\chi_2)E(\psi_1\psi_2\psi_3\psi_4) &: L^*_{12}L'_3L'_4\rho'_{23}+O(e_2);\\
E(\psi_1\psi_2\psi_3\psi_4) &: L'_{12}L'_3L'_4\rho'_{23}+O(e_2);\\
E(\psi_1\psi_2)E(\chi_1\chi_2\chi_3\chi_4) &: L^*_{12}L_3L_4\rho_{23}+O(e_2);\\
E(\chi_1\chi_2)E(\psi_3\psi_4) &: L_{12}L'_3L'_4\rho'_{23}+O(e_2);\\
E(\chi_{12})E(\psi_{12}\psi_3\psi_4) &: L^{**}_{12}L'_3L'_4\rho'_{23}+O(e_2);\\
E(\psi_1\psi_2)E(\chi_3\chi_4) &: L'_{12}L_3L_4\rho_{23}+O(e_2);
\end{aligned}$$

$$E(\psi_{12})E(\chi_{12}\chi_3\chi_4) : L_{12}^{**}L_3L_4\rho_{23} + O(e_2).$$

Thus,

$$\begin{aligned}
E(Z_{t_1}^2 Z_{t_2} Z_{t_3}) &= \sum (L_{12}L_3L_4\rho_{23} + L'_{12}L'_3L'_4\rho'_{23} + L_{12}L'_3L'_4\rho'_{23} + L'_{12}L_3L_4\rho_{23} \\
&\quad + L^*_{12}L'_3L'_4\rho'_{23} + L^*_{12}L_3L_4\rho_{23} + L^{**}_{12}L'_3L'_4\rho'_{23} + L^{**}_{12}L_3L_4\rho_{23}) + O(e_2) \\
&= \sum (L_{12} + L'_{12} + L^*_{12} + L^{**}_{12})(L_3L_4\rho_{23} + L'_3L'_4\rho'_{23}) + O(e_2) \\
&= \left(\bar{L}_2 + \bar{L}'_2 + \bar{L}^*_2 + \bar{L}^{**}_2\right)\left(\bar{L}_1^2\rho_{23} + \bar{L}_1'^2\rho'_{23}\right) + O(e_2). \qquad (3.82)
\end{aligned}$$

(iii) $E(Z_{t_1}^2 Z_{t_2}^2)$

$$\begin{aligned}
E(\chi_1\chi_2\chi_3\chi_4) &: L_{12}L_{34} + O(|\rho_{12}|); \\
E(\chi_1\chi_2)E(\psi_1\psi_2\psi_3\psi_4) &: L^*_{12}L'_{34} + O(|\rho_{12}|); \\
E(\psi_1\psi_2\psi_3\psi_4) &: L'_{12}L'_{34} + O(|\rho_{12}|); \\
E(\chi_3\chi_4)E(\psi_1\psi_2\psi_3\psi_4) &: L'_{12}L^*_{34} + O(|\rho_{12}|); \\
E(\chi_1\chi_2)E(\psi_3\psi_4) &: L_{12}L'_{34}; \\
E(\psi_1\psi_2)E(\chi_1\chi_2\chi_3\chi_4) &: L^*_{12}L_{34} + O(|\rho_{12}|); \\
E(\psi_1\psi_2)E(\chi_3\chi_4) &: L'_{12}L_{34}; \\
E(\psi_3\psi_4)E(\chi_1\chi_2\chi_3\chi_4) &: L_{12}L^*_{34} + O(|\rho_{12}|); \\
E(\chi_{12})E(\psi_{12}\psi_3\psi_4) &: L^{**}_{12}L'_{34} + O(|\rho_{12}|); \\
E(\chi_{12}\chi_3\chi_4)E(\psi_{12}\psi_3\psi_4) &: L^{**}_{12}L^*_{34} + O(|\rho_{12}|); \\
E(\chi_{34})E(\psi_{34}\psi_1\psi_2) &: L'_{12}L^{**}_{34} + O(|\rho_{12}|); \\
E(\chi_{34}\chi_1\chi_2)E(\psi_{34}\psi_1\psi_2) &: L^*_{12}L^{**}_{34} + O(|\rho_{12}|); \\
E(\psi_{12})E(\chi_{12}\chi_3\chi_4) &: L^{**}_{12}L_{34} + O(|\rho_{12}|); \\
E(\chi_1\chi_2\chi_3\chi_4)E(\psi_1\psi_2\psi_3\psi_4) &: L^*_{12}L^*_{34} + O(|\rho_{12}|);
\end{aligned}$$

$$E(\psi_{34})E(\chi_{34}\chi_1\chi_2) : L_{12}L_{34}^{**} + O(|\rho_{12}|);$$
$$E(\chi_{12,34})E(\psi_{12,34}) : L_{12}^{**}L_{34}^{**} + O(|\rho_{12}|).$$

Thus,

$$\begin{aligned}
E(Z_{t_1}^2 Z_{t_2}^2) &= \sum (L_{12}L_{34} + L'_{12}L'_{34} + L_{12}L'_{34} + L'_{12}L_{34} + L^*_{12}L'_{34} \\
&\quad + L'_{12}L^*_{34} + L^*_{12}L_{34} + L_{12}L^*_{34} + L^*_{12}L^*_{34} + L^{**}_{12}L'_{34} + L^{**}_{12}L_{34} \\
&\quad + L^{**}_{12}L^*_{34} + L'_{12}L^{**}_{34} + L_{12}L^{**}_{34} + L^*_{12}L^{**}_{34} + L^{**}_{12}L^{**}_{34}) + O(|\rho_{12}|) \\
&= \sum (L_{12} + L'_{12} + L^*_{12} + L^{**}_{12})(L_{34} + L'_{34} + L^*_{34} + L^{**}_{34}) + O(|\rho_{12}|) \\
&= \left(\bar{L}_2 + \bar{L}'_2 + \bar{L}^*_2 + \bar{L}^{**}_2\right)^2 + O(|\rho_{12}|). \qquad (3.83)
\end{aligned}$$

(iv) $E(Z_{t_1}^3 Z_{t_2})$

$$E(Z_{t_1}^3 Z_{t_2}) = O(|\rho_{12}|). \qquad (3.84)$$

(v) $E(Z_{t_i} Z_{t_j})$, $i \neq j$

$$E(\chi_1\chi_2) : L_1L_2\rho_{ij} + O(\rho_{ij}^2); \qquad E(\psi_1\psi_2) : L'_1L'_2\rho'_{ij} + O(\rho_{ij}^2).$$

Thus,

$$E(Z_{t_i} Z_{t_j}) = \sum_{u_1,u_2=0}^{p} (L_1L_2\rho_{ij} + L'_1L'_2\rho'_{ij}) + O(\rho_{ij}^2) = \bar{L}_1^2\rho_{ij} + \bar{L}_1'^2\rho'_{ij} + O(\rho_{ij}^2). \quad (3.85)$$

(vi) $E(Z_{t_i}^2)$

$$E(\chi_1\chi_2) : L_{12}; \qquad E(\chi_1\chi_2)E(\psi_1\psi_2) : L^*_{12};$$
$$E(\psi_1\psi_2) : L'_{12}; \qquad E(\chi_{12})E(\psi_{12}) : L^{**}_{12}.$$

Thus,

$$E(Z_{t_i}^2) = \sum_{u_1,u_2=0}^{p} (L_{12} + L'_{12} + L^*_{12} + L^{**}_{12}) = \bar{L}_2 + \bar{L}'_2 + \bar{L}^*_2 + \bar{L}^{**}_2. \tag{3.86}$$

Define V_t as a mean-zero Gaussian $I(d_1)$ process with $E(V_t^2) = \bar{L}_2+\bar{L}'_2+\bar{L}^*_2+\bar{L}^{**}_2$ and $E(V_{t_i}V_{t_j}) = \bar{L}_1^2\rho_{ij} + \bar{L}_1'^2\rho'_{ij}$, for $i \neq j$. Using equations (3.81) to (3.86) to compute the covariances of interest in each case, they are easily shown to be identical to

$$\mathrm{Cov}(V_{s_1}V_{s_2}, V_{s_3}V_{s_4}) = E(V_{s_1}V_{s_3})E(V_{s_2}V_{s_4}) + E(V_{s_1}V_{s_4})E(V_{s_2}V_{s_3}),$$

up to the desired approximation errors. □

Lemma 3.7 *If V_t is Gaussian $I(1/4)$, under (3.24),*

$$\mathrm{Var}\left\{\hat{F}_{VV}(\lambda_m)\right\} = O\left(\frac{\log m}{n}\right).$$

Proof. Let $\rho_j = \mathrm{Cov}(V_0, V_j)$ and assume $\rho_0 = 1$, without loss of generality. By assumption, $|\rho_j| \leq Kj^{-1/2}$. We will use similar methods to the proof of Lemma 10 in Robinson (1994b), including the decomposition

$$\begin{aligned}
\mathrm{Var}\left\{\hat{F}_{VV}(\lambda_m)\right\} &= \frac{1}{n^4}\sum_{s,t,u,v=1}^{n} \mathrm{Cov}(V_sV_t, V_uV_v)D_m(\lambda_{t-s})\overline{D_m(\lambda_{v-u})} \\
&= \frac{1}{n^4}\sum_{s,t,u,v=1}^{n} (\rho_{u-s}\rho_{v-t} + \rho_{v-s}\rho_{u-t})D_m(\lambda_{t-s})\overline{D_m(\lambda_{v-u})} \\
&= \frac{1}{n^4}\sum_{j,k=1}^{m} (W_{j,k-j}W_{k,j-k} + W_{j,-j-k}W_{-k,j+k}),
\end{aligned} \tag{3.87}$$

where

$$W_{j,k} = \sum_{u=1-n}^{n-1} \rho_u e^{i\lambda_j u}T_k(u), \qquad T_k(u) = \sum_{t=1+u^+}^{n-u^-} e^{it\lambda_k},$$

denoting the positive and negative parts of u by $u^+ = (|u|+u)/2$ and $u^- = (|u|-u)/2$ respectively. Note that Robinson (1994b) has a typo in this decomposition, using k instead of $-k$ in the first index of the last W. However, the correct expression is used in the remainder of his proof.

To bound $W_{j,0}$, for $j = 1, \ldots, m$, note that $T_0(u) = n - |u|$. Summation by parts gives

$$\begin{aligned} W_{j,0} &= \sum_{u=1-n}^{n-1} (n-|u|)\rho_u e^{i\lambda_j u} \\ &= n + \sum_{u=1}^{n-1} (n-u)\rho_u (e^{i\lambda_j u} + e^{-i\lambda_j u}) \\ &= n + \sum_{u=1}^{n-1} \left\{(n-u)\rho_u - (n-u-1)\rho_{u+1}\right\} \left\{D_u(\lambda_j) + \overline{D_u(\lambda_j)}\right\}, \end{aligned}$$

so using (3.38) we get

$$\begin{aligned} |W_{j,0}| &\le n + K \sum_{u=1}^{n-1} \left\{(n-u)|\rho_u - \rho_{u+1}| + |\rho_{u+1}|\right\} |D_u(\lambda_j)| \\ &\le n + K \sum_{u=1}^{n-1} \frac{n}{u} |\rho_{u+1}||D_u(\lambda_j)| \\ &\le n + Kn \sum_{u=1}^{[n/j]} u^{-\frac{1}{2}} + K\frac{n^2}{j} \sum_{u=[n/j]}^{n} u^{-\frac{3}{2}} \\ &\le n + Kn \left(\frac{n}{j}\right)^{\frac{1}{2}} + K\frac{n^2}{j}\left(\frac{n}{j}\right)^{-\frac{1}{2}} \le K\frac{n^{\frac{3}{2}}}{j^{\frac{1}{2}}}. \end{aligned} \tag{3.88}$$

For $k \neq 0$ and $u > 0$, (3.32) implies that $T_k(0) = 0$,

$$\begin{aligned} T_k(u) &= \sum_{t=1+u}^{n} e^{it\lambda_k} = \sum_{t=1}^{n} e^{it\lambda_k} - \sum_{t=1}^{u} e^{it\lambda_k} = -D_u(\lambda_k), \\ T_k(-u) &= \sum_{t=1}^{n-u} e^{it\lambda_k} = \sum_{t=1}^{n} e^{it\lambda_k} - e^{i\lambda_k} \sum_{t=n-u}^{n-1} e^{it\lambda_k} = -e^{i\lambda_k}\overline{D_u(\lambda_k)}. \end{aligned}$$

Therefore, using summation by parts,

$$\begin{aligned}W_{j,k} &= \sum_{u=1}^{n-1} \rho_u \left\{ e^{i\lambda_j u} T_k(u) + e^{-i\lambda_j u} T_k(-u) \right\} \\ &= \sum_{u=1}^{n-1} (\rho_u - \rho_{u+1}) \sum_{q=1}^{u} \left\{ -e^{i\lambda_j q} D_q(\lambda_k) - e^{i(\lambda_k - \lambda_j q)} \overline{D_q(\lambda_k)} \right\} \\ &\quad + \rho_n \sum_{q=1}^{n-1} \left\{ -e^{i\lambda_j q} D_q(\lambda_k) - e^{i(\lambda_k - \lambda_j q)} \overline{D_q(\lambda_k)} \right\},\end{aligned}$$

implying

$$|W_{j,k}| \le K \sum_{u=1}^{n-1} |\rho_u - \rho_{u+1}| \left| \sum_{q=1}^{u} e^{i\lambda_j q} D_q(\lambda_k) \right| + K|\rho_n| \left| \sum_{q=1}^{n-1} e^{i\lambda_j q} D_q(\lambda_k) \right|.$$

Since

$$\begin{aligned}\sum_{q=1}^{u} e^{i\lambda_j q} D_q(\lambda_k) &= \sum_{q=1}^{u} e^{i\lambda_j q} \sum_{t=1}^{q} e^{it\lambda_k} = \sum_{q=1}^{u} e^{i\lambda_j q} \frac{1 - e^{iq\lambda_k}}{e^{-i\lambda_k} - 1} \\ &= \frac{1}{e^{-i\lambda_k} - 1} \sum_{q=1}^{u} (e^{i\lambda_j q} - e^{i\lambda_{j+k} q}) = \frac{D_u(\lambda_j) - D_u(\lambda_{j+k})}{e^{-i\lambda_k} - 1},\end{aligned}$$

and $|e^{-i\lambda} - 1| \sim |\lambda|$ as $\lambda \to 0$, we have

$$\left| \sum_{q=1}^{u} e^{i\lambda_j q} D_q(\lambda_k) \right| \le K \frac{n}{|k|} \left\{ |D_u(\lambda_j)| + |D_u(\lambda_{j+k})| \right\}.$$

So, using (3.38) and for $a = \min\{|j|, |j+k|\}$,

$$\begin{aligned}|W_{j,k}| &\le K \frac{n}{|k|} \sum_{u=1}^{[n/a]} |\rho_{u+1}| + K \frac{n}{|k|} \sum_{u=[n/a]+1}^{n-1} \frac{|\rho_{u+1}|}{u} \frac{n}{a} + K|\rho_n| \frac{n^2}{|k|a} \\ &\le K \frac{n}{|k|} \sum_{u=1}^{[n/a]} u^{-\frac{1}{2}} + K \frac{n^2}{|k|a} \sum_{u=[n/a]+1}^{n-1} u^{-\frac{3}{2}} + K \frac{n^{\frac{3}{2}}}{|k|a} \\ &\le K \frac{n}{|k|} \left(\frac{n}{a} \right)^{\frac{1}{2}} + K \frac{n^2}{|k|a} \left(\frac{n}{a} \right)^{-\frac{1}{2}} + K \frac{n^{\frac{3}{2}}}{|k|a} \le K \frac{n^{\frac{3}{2}}}{|k| a^{\frac{1}{2}}},\end{aligned}$$

yielding

$$|W_{j,k-j}W_{k,j-k}| \leq K\frac{n^3}{(j-k)^2\min\{j,k\}}, \qquad 1 \leq j,k \leq m,\ j \neq k, \tag{3.89}$$

$$|W_{j,-j-k}W_{-k,j+k}| \leq K\frac{n^3}{(j+k)^2\min\{j,k\}}, \qquad 1 \leq j,k \leq m. \tag{3.90}$$

Thus, using (3.88), (3.89), (3.90) in (3.87),

$$\begin{aligned}
\mathrm{Var}\left\{\hat{F}_{VV}(\lambda_m)\right\} \leq & \frac{K}{n^4}\sum_{j=1}^{m}\frac{n^3}{j} + \frac{K}{n^4}\sum_{\substack{j,k=1\\ j\neq k}}^{m}\frac{n^3}{(j-k)^2\min\{j,k\}} \\
& + \frac{K}{n^4}\sum_{j,k=1}^{m}\frac{n^3}{(j+k)^2\min\{j,k\}} \\
\leq & K\frac{\log m}{n} + \frac{K}{n^4}\sum_{\substack{j,k=1\\ j<k}}^{m}\frac{n^3}{(j-k)^2 j} + \frac{K}{n^4}\sum_{\substack{j,k=1\\ j\leq k}}^{m}\frac{n^3}{(j+k)^2 j} \\
\leq & K\frac{\log m}{n} + \frac{K}{n}\sum_{j=1}^{m-1} j^{-1}\sum_{a=1}^{m-j} a^{-2} + \frac{K}{n}\sum_{j=1}^{m} j^{-1}\sum_{a=2j}^{m+j} a^{-2} \\
\leq & K\frac{\log m}{n},
\end{aligned}$$

which completes the proof. □

Chapter 4

Finite sample performance of Narrow Band Least Squares

4.1 Introduction

We now present two Monte Carlo studies of finite sample performance of NBLS. For linear processes, Robinson and Marinucci (2003) reported simulation experiments of NBLS with $I(1)$ observables and $I(0)$ cointegrating errors, while Marinucci and Robinson (2001) explored different cases of $I(d_x)$ nonstationary observables and $I(d_e)$ stationary errors. Bandi and Perron (2006) examined NBLS for the regression between realized and implied volatility, generating the data from a discretised continuous time SV model.

The following section aims to evaluate in detail the validity in finite samples of the asymptotic results derived in the previous chapter. We make use of the modelling assumptions proposed in that chapter, for moderate sample sizes ranging from 256 to 2048 observations.

Section 3 attempts to fill the gap in theoretical distributional results, by providing a comparative study of distributional properties of weighted and simple

NBLS, as described in Section 2.6, under a variety of different settings. We compare nonlinear SV models with the linear setting, and explore the impact of the weighting scheme in WNBLS, for both fixed and data dependent weights. Given the focus on distributional properties, we make use of sample sizes ranging from 512 to 8192 observations, and present kernel density estimates of the finite sample distribution under various settings.

4.2 Finite sample performance of Narrow Band Least Squares under a stochastic volatility setting

In this section, we employ 50,000 replications of series of various lengths n generated by (see (2.37), (2.36) and (3.1), (3.2))

$$y_t = \beta\zeta_t + \varepsilon_t, \qquad x_t = \zeta_t + \delta_t, \tag{4.1}$$

where we set $\beta = 1$ and (see (3.17), (3.18))

$$\zeta_t = \eta_{1t} g(\eta_{2t}), \qquad \delta_t = \nu_{1t} h(\nu_{2t}), \qquad \varepsilon_t = \xi_{1t} l(\xi_{2t}). \tag{4.2}$$

All basic processes in (4.2) are independent of each other, and standard Gaussian. Processes η_{1t}, ν_{1t}, ξ_{1t} in were generated as iid, while the Davies and Harte (1987) algorithm was used to generate η_{2t} as ARFIMA$(0, d_1, 0)$ and ν_{2t}, ξ_{2t} as ARFIMA$(0, d_2, 0)$. In most cases h and l are constant functions, as in (2.37), (2.36), and ν_{2t}, ξ_{2t} are not required. For all functions g considered, $p = 2$ satisfies Assumption 3.1.

Our goal is to estimate β, or at least $\theta = \beta^2$. Denoting $Y_t = y_t^2$ and $X_t = x_t^2$,

we compare the performance of NBLS estimates (2.55),

$$\hat{\theta}_m = \frac{\mathrm{Re}\left\{\hat{F}_{XY}(\lambda_m)\right\}}{\hat{F}_{XX}(\lambda_m)}, \qquad 1 \leq m \leq \frac{n}{2}, \tag{4.3}$$

with OLS estimates (2.47) obtained either from levels,

$$\tilde{\beta} = \frac{\sum(x_t - \bar{x})y_t}{\sum(x_t - \bar{x})^2}, \tag{4.4}$$

or squares,

$$\tilde{\theta} = \frac{\sum(X_t - \bar{X})Y_t}{\sum(X_t - \bar{X})^2}. \tag{4.5}$$

To make the three estimates comparable, we take $\tilde{\beta}^2$ as an estimate of θ. Of course, neither (4.4) nor (4.5) are consistent for β or θ respectively, but they are simple estimates that a practitioner might optimistically compute. One can furthermore interpret (4.5) as the full band version of the proposed NBLS estimate, i.e. (4.3) for $m = n/2$. We report the bias, standard deviation (SD), and root mean squared error (RMSE) for each estimate. On occasion, relative quantities are reported, meaning the ratio between the corresponding quantity for NBLS and (4.5), which dominates (4.4) in every experiment.

Bandwidth choice

Theorem 3.2 highlights the relationship between bandwidth m and rate of convergence. In the first experiment, we present the evolution of relative bias, SD, and RMSE for different m and d_1. We set $n = 256$, $d_1 = 0.1, 0.2, 0.3, 0.4$, $g(x) = \exp(kx)$, with k chosen to satisfy $\mathrm{Var}(\zeta_t) = 2$, and $h(x) = l(x) = 1$. We chose this value for $\mathrm{Var}(\zeta_t)$ in several experiments in order to balance the contributions of bias and SD to RMSE; the impact of the signal to noise ratio is

explored later.

Figure 4.1 shows the bias reduction achieved by NBLS relative to OLS. Not surprisingly, it is greater for small m and large d_1. It is only around frequency zero that the spectral density of X_t dominates that of U_t; frequencies further from the origin are more contaminated by the correlation between X_t and U_t, and contribute more to bias. Also, a higher d_1 indicates a stronger cointegrating relationship, increasing the spectral density of X_t around the origin and thus the averaged periodogram.

The increase in SD of NBLS relative to OLS, displayed in Figure 4.2, is a consequence of discarding high frequency information, and is decreasing in m. The influence of d_1 on relative SD appears to be small, specially if compared to Figure 4.1.

The different profiles of bias and SD give rise to the traditional trade-off in bandwidth choice. Figure 4.3 presents the relative RMSE of NBLS. For most m, NBLS dominates OLS. For this particular n, a low d_1 does not provide enough information for NBLS to work, due to the modest bias reductions displayed in Figure 4.1, making the improvement over OLS negligible. The RMSE is essentially a flat function of m, implying that any m above a certain threshold, thereby taking in OLS, attains similar RMSE. However, note that an increase in n should have a similar effect to an increase in d_1 on RMSE, although it will be minimized at a different m. This effect is explored in the next subsection. Higher d_1 lead to very low values for the optimal m, and more significant improvements in RMSE. For $d_1 = 0.4$, a noticeable reduction is already achieved, of over 10% for a number of different m. It should also be noted that if the bandwidth selection is larger than optimal, it is still possible to considerably reduce RMSE, while choosing too small an m can lead to an undesirably large SD.

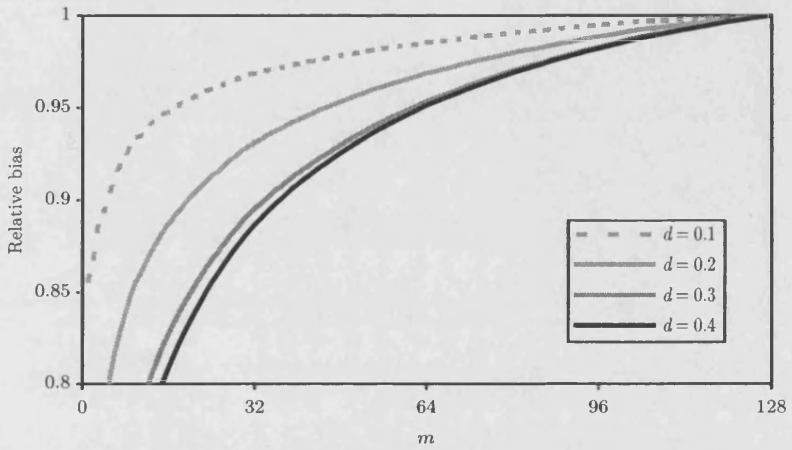


Figure 4.1: Relative bias of NBLS versus OLS, for varying m and d_1.

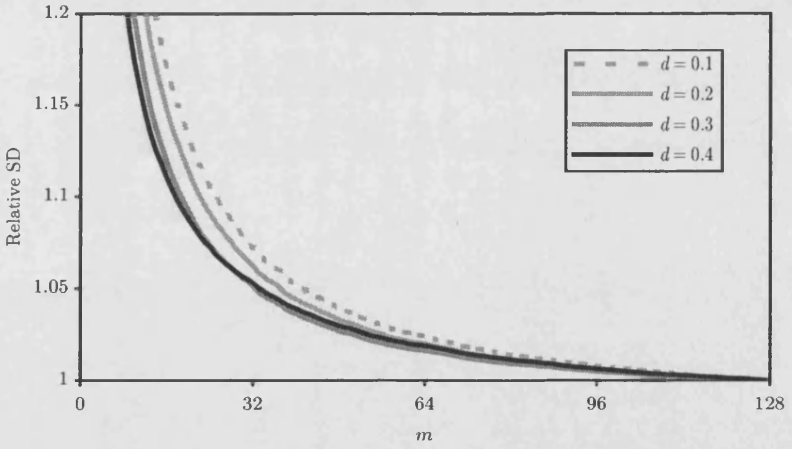


Figure 4.2: Relative SD of NBLS versus OLS, for varying m and d_1.

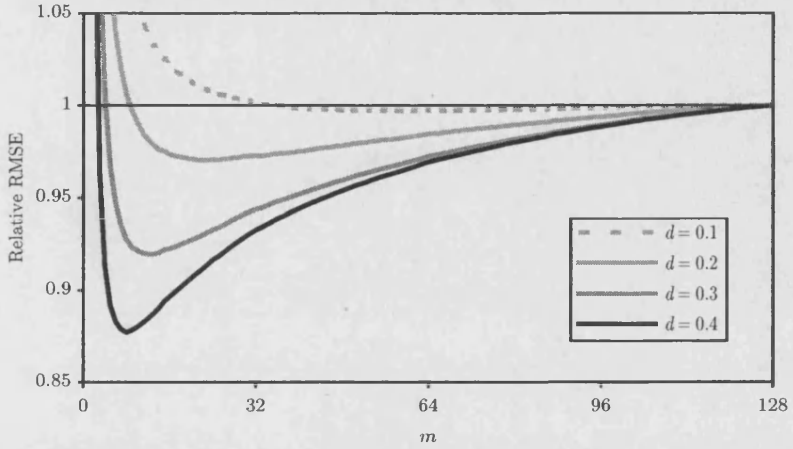


Figure 4.3: Relative RMSE of NBLS versus OLS, for varying m and d_1.

Memory in signal

We now investigate the influence of n and d_1 on the performance of the three estimates. We consider $n = 256, 512, 1024, 2048$ and $d_1 = 0.1, 0.2, 0.3, 0.4$. As before, $g(x) = \exp(kx)$, with k chosen to satisfy $\mathrm{Var}(\zeta_t) = 2$, and $h(x) = l(x) = 1$. In this experiment and in the following ones, we evaluate NBLS at the bandwidth m^* that minimizes RMSE. Although this is not a feasible choice in the usual sense, it gives an indication of potential gains; the effects of deviating from this optimal bandwidth should be qualitatively similar to those presented in Figure 4.3. Table 4.1 summarizes the results.

As expected, the RMSE of all estimates improves with n. For even moderate n, NBLS has the lowest RMSE, being clearly less biased than OLS; while OLS in levels attains the lowest SD, especially for small n, its larger bias makes it the worst.

Both bias and SD of OLS increase with d_1. Both also decrease with n, but while SD seems to be rapidly converging to zero, bias decreases rather slowly and

d_1	n	256				512				1024				2048			
		m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE
0.1	$\tilde{\beta}^2$	—	-0.560	0.085	0.567	—	-0.558	0.062	0.561	—	-0.557	0.045	0.559	—	-0.556	0.033	0.557
	$\tilde{\theta}$	—	-0.269	0.179	0.323	—	-0.244	0.141	0.282	—	-0.226	0.111	0.252	—	-0.213	0.087	0.230
	$\hat{\theta}_{m^*}$	58	-0.264	0.184	0.322	53	-0.234	0.149	0.278	50	-0.212	0.123	0.245	64	-0.197	0.097	0.220
0.2	$\tilde{\beta}^2$	—	-0.564	0.098	0.573	—	-0.561	0.075	0.566	—	-0.559	0.057	0.562	—	-0.558	0.044	0.559
	$\tilde{\theta}$	—	-0.279	0.184	0.334	—	-0.251	0.145	0.290	—	-0.231	0.115	0.258	—	-0.216	0.090	0.234
	$\hat{\theta}_{m^*}$	22	-0.253	0.203	0.324	23	-0.215	0.164	0.270	22	-0.184	0.136	0.228	24	-0.160	0.111	0.194
0.3	$\tilde{\beta}^2$	—	-0.577	0.127	0.591	—	-0.571	0.107	0.581	—	-0.568	0.090	0.575	—	-0.564	0.076	0.570
	$\tilde{\theta}$	—	-0.309	0.199	0.367	—	-0.276	0.160	0.319	—	-0.252	0.129	0.283	—	-0.233	0.102	0.254
	$\hat{\theta}_{m^*}$	12	-0.246	0.232	0.338	12	-0.191	0.188	0.268	14	-0.152	0.147	0.212	13	-0.113	0.121	0.165
0.4	$\tilde{\beta}^2$	—	-0.611	0.181	0.637	—	-0.604	0.168	0.627	—	-0.599	0.156	0.619	—	-0.594	0.146	0.611
	$\tilde{\theta}$	—	-0.404	0.239	0.469	—	-0.368	0.208	0.423	—	-0.339	0.183	0.385	—	-0.314	0.159	0.352
	$\hat{\theta}_{m^*}$	8	-0.291	0.290	0.411	7	-0.205	0.247	0.321	8	-0.148	0.192	0.242	8	-0.096	0.147	0.175

Table 4.1: Monte Carlo bias, SD, RMSE of NBLS for varying n and d_1.

appears to stabilize at some substantial non-zero value. For small n, changes in d_1 similarly affect NBLS, but for larger n, the small m^* makes bias decrease with d_1.

The bias reduction of NBLS becomes quite large with n, while the variance penalty is always of small magnitude. In fact, when $d_1 = 0.4$ and $n = 2048$, NBLS actually dominates OLS in both SD and bias. The improvement in performance for high d_1 and the rate of decay of RMSE seem compatible with the asymptotic result of Theorem 3.2.

While Figure 4.3 and Table 4.1 both illustrate the high sensitivity of m^* to d_1, caused by the different scope for bias reduction in each case, m^* does not appear to grow with n. This is surely a small sample effect, as NBLS is only consistent if $m \to \infty$. As a consequence, m^* will diverge when bias becomes negligible compared to SD, a situation which does not occur in the sample sizes considered. Since Theorem 3.2 suggests that convergence of $\hat{\theta}_m$ is faster the slower m grows, this phenomenon is not entirely surprising.

Memory in signal and noise

In Table 4.2, d_1 is kept constant, while we introduce long memory in the errors. We set $g(x) = \exp(k_1 x)$ and $h(x) = l(x) = \exp(k_2 x)$, with k_1, k_2 chosen to satisfy $\mathrm{Var}(\zeta_t) = 10$ and $\mathrm{Var}(\delta_t) = \mathrm{Var}(\varepsilon_t) = 2$. These values were again chosen to balance contributions of bias and SD to RMSE. We consider $n = 256$, 512, 1024, 2048, $d_1 = 0.4$ and $d_2 = 0$, 0.1, 0.2, 0.3.

The results are very similar to the previous experiment, but here $d_1 - d_2$ takes the role of d_1. As before, RMSE improves with n for all estimates. OLS displays similar patterns of bias and SD across $d_1 - d_2$ and n, with the exception that SD decays much more slowly with n. The bias of NBLS decreases with $d_1 - d_2$ for all n; for $n > 256$ even SD decreases with $d_1 - d_2$. A surprising fact in this case is

d_2	n	256				512				1024				2048			
		m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE
0.0	$\tilde{\beta}^2$	—	-0.520	0.276	0.588	—	-0.498	0.261	0.562	—	-0.476	0.246	0.536	—	-0.455	0.232	0.511
	$\tilde{\theta}$	—	-0.281	0.318	0.424	—	-0.225	0.277	0.357	—	-0.176	0.232	0.291	—	-0.133	0.189	0.231
	$\hat{\theta}_{m^*}$	12	-0.228	0.327	0.398	11	-0.156	0.266	0.308	9	-0.097	0.205	0.227	8	-0.054	0.145	0.155
0.1	$\tilde{\beta}^2$	—	-0.519	0.276	0.588	—	-0.497	0.261	0.561	—	-0.475	0.247	0.536	—	-0.455	0.232	0.511
	$\tilde{\theta}$	—	-0.280	0.318	0.423	—	-0.224	0.276	0.356	—	-0.175	0.232	0.291	—	-0.133	0.189	0.231
	$\hat{\theta}_{m^*}$	14	-0.236	0.327	0.403	11	-0.162	0.271	0.316	9	-0.104	0.211	0.235	9	-0.062	0.152	0.164
0.2	$\tilde{\beta}^2$	—	-0.516	0.278	0.586	—	-0.495	0.262	0.560	—	-0.474	0.247	0.535	—	-0.454	0.233	0.510
	$\tilde{\theta}$	—	-0.276	0.317	0.421	—	-0.222	0.276	0.354	—	-0.174	0.232	0.290	—	-0.132	0.189	0.231
	$\hat{\theta}_{m^*}$	19	-0.245	0.326	0.408	13	-0.173	0.277	0.327	13	-0.120	0.219	0.250	13	-0.078	0.164	0.181
0.3	$\tilde{\beta}^2$	—	-0.506	0.282	0.579	—	-0.487	0.266	0.555	—	-0.468	0.251	0.530	—	-0.449	0.236	0.508
	$\tilde{\theta}$	—	-0.266	0.316	0.413	—	-0.215	0.274	0.348	—	-0.169	0.231	0.286	—	-0.129	0.189	0.229
	$\hat{\theta}_{m^*}$	35	-0.251	0.323	0.409	26	-0.189	0.278	0.336	27	-0.141	0.227	0.267	35	-0.102	0.180	0.206

Table 4.2: Monte Carlo bias, SD, RMSE of NBLS for varying n and d_2, with $d_1 = 0.4$.

related to the variance/bias trade-off of NBLS. While this can be found in small samples, as n increases it starts dominating OLS in both bias and variance. The evolution of m^* is also similar to the previous section.

We do not directly address the impact of short run dynamics in finite samples, as we expect its consequences to be qualitatively analogous to the linear case. As reported in Robinson and Marinucci (2003), the presence of short memory positive autocorrelation in the common factor should boost the spectral peak in small samples, reducing bias in a similar manner to a higher d_1 in Table 1; conversely, negative autocorrelation in the common factor should have a similar effect to a lower d_1. The impact of short run dynamics in the idiosyncratic components would predictably be the opposite: positive correlation should worsen performance similarly to an increase in d_2 in Table 2, while negative correlation would be associated with a dampened d_2. Both effects should become negligible as n grows.

Signal to noise ratio

This experiment investigates the influence of the signal to noise (S2N) ratio on the performance of NBLS. We use $g(x) = \exp(kx)$, such that $\mathrm{Var}(\zeta_t) = 2$, and $h(x) = l(x) = \sigma$, so that $\mathrm{Var}(\delta_t) = \mathrm{Var}(\varepsilon_t) = \sigma^2$, for $\sigma^2 = 0.25$, 0.5, 1, 2, 4. The results obtained for different d_1 were qualitatively similar, so we report results only for $d_1 = 0.3$ and $n = 256$, 512, 1024, 2048. Since it is unreasonable to compare absolute performance for different S2N ratios, in Table 4.3 we focus on relative performance only. We also report the ratio between bias and SD. Although we refer to $\mathrm{Var}(\zeta_t)/\mathrm{Var}(\delta_t)$ as the S2N ratio, for simplicity, it is only an accurate description for the regression in levels. For $x_t^2 = \zeta_t^2 + 2\zeta_t\delta_t + \delta_t^2$, the dominant term is ζ_t^2, not ζ_t, and even there η_{1t}^2 could be considered a multiplicative noise. Hence, the definition of the "true" S2N ratio would be ambiguous, but it would be arguably smaller than the one in levels.

n	256			512			1024			2048		
S2N	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
0.5	0.994	1.280	0.998	0.981	1.960	0.992	0.961	2.924	0.980	0.927	4.141	0.956
1	0.910	1.406	0.959	0.818	1.682	0.902	0.737	1.773	0.829	0.621	1.955	0.731
2	0.797	1.163	0.920	0.691	1.175	0.839	0.604	1.144	0.749	0.485	1.181	0.650
4	0.834	1.050	0.969	0.738	1.057	0.922	0.652	1.048	0.857	0.557	1.059	0.781
8	0.996	1.000	1.000	0.879	1.013	0.989	0.773	1.019	0.965	0.683	1.016	0.922
		Bias / SD			Bias / SD			Bias / SD			Bias / SD	
S2N	m^*	OLS	NBLS	m^*	OLS	NBLS	m^*	OLS	NBLS	m^*	OLS	NBLS
0.5	52	-9.29	-7.22	27	-11.42	-5.72	17	-14.19	-4.66	12	-17.26	-3.86
1	12	-3.38	-2.19	8	-3.72	-1.81	8	-4.14	-1.72	7	-4.69	-1.49
2	12	-1.55	-1.06	12	-1.73	-1.02	14	-1.95	-1.03	13	-2.28	-0.94
4	26	-0.82	-0.65	28	-0.93	-0.65	33	-1.09	-0.68	35	-1.31	-0.69
8	121	-0.43	-0.43	87	-0.49	-0.42	86	-0.57	-0.43	90	-0.69	-0.47

Table 4.3: Monte Carlo relative bias, SD, RMSE of NBLS versus OLS, for varying n and S2N, with $d_1 = 0.3$.

NBLS performs best when bias and SD are balanced. The regressor X_t consists of two parts: a long memory component containing a dominating pole at frequency zero, and a component with less memory not orthogonal to U_t. In this case, it is actually short memory, since δ_t is iid. If the S2N ratio is very large, the first component will dominate the second even at frequencies distant from zero. As a result, any large enough m will perform well; even with OLS, bias will contribute very little to RMSE and gains from NBLS will be small. On the other hand, for very small S2N, the second component will be relatively large, dominating the signal even at frequencies close to zero. In small samples, an attempt to reduce bias by only choosing informative frequencies would imply the use of very small m, which would force SD to be too high (see Figure 4.2). In this case, NBLS would also provide little gains, as the cost (in terms of SD) of reducing bias is too high for RMSE.

With OLS the ratio between bias and SD increases with n. This is expected, since OLS still converges in probability to a constant. In NBLS, the ratio is very close to that of OLS in small samples. From that point, it increases with n if it was originally small, but decreases if it was originally large. It appears that this ratio will stabilize at some value close to unity for large enough n, and from that point on NBLS will have a noticeable RMSE improvement over OLS.

Nonlinearity

To investigate the influence of nonlinearity on NBLS, Table 4.4 reports its performance in three different settings, for $n = 256$, 512, 1024, 2048 and $d_1 = 0.1$, 0.2, 0.3, 0.4. The nonlinear setting (NL), already used in the first two subsections, has $g(x) = \exp(kx)$, with k chosen to satisfy $\mathrm{Var}(\zeta_t) = 2$, and $h(x) = l(x) = 1$. In

the other two we deviate from (4.1), (4.2), using instead

$$Y_t = X_t + u_t, \qquad X_t = f_t + v_t, \tag{4.6}$$

where u_t, v_t are generated as iid mean zero Gaussian with $\mathrm{Var}(u_t) = 20$, $\mathrm{Var}(v_t) = 6$, and $\mathrm{Cov}(u_t, v_t) = -10$. In a fully linear setting (L), we generate f_t as a Gaussian mean zero ARFIMA$(0, d_1, 0)$, with $\mathrm{Var}(f_t) = 44$. In a linear setting with a multiplicative noise (MN), we set $f_t = \eta_{1t}^2 z_t$, where η_{1t} is iid standard Gaussian while z_t is independently generated as a Gaussian ARFIMA$(0, d_1, 0)$, with $E(z_t) = 2$ and $\mathrm{Var}(z_t) = 12$. The chosen moments replicate those of corresponding processes in the nonlinear setting.

Both OLS and NBLS perform much better under L than NL, while performance under MN falls in the middle. A similar ordering is found in relative performance (not shown), since a relatively stable, large bias of OLS estimates throughout makes variations in RMSE smaller than for NBLS. Although some of the gap in performance should be a consequence of nonlinearity, significant excess kurtosis in NL and MN is arguably the dominant factor, since it directly affects the variance of the periodogram. In MN, the kurtosis of f_t is around 77, while in NL it is around 3523 for f_t, 36 for v_t, and 30 for u_t. A more detailed comparison of linear and nonlinear settings can be found in the next section.

Volatility function

Finally, we explore the impact of the functional form of the volatility function g. considering $g(x) = \exp(kx)$, $(1 + kx)^2$, $|1 + kx|$, with k chosen in each case so that $\mathrm{Var}(\zeta_t) = 2$. We set $h(x) = l(x) = 1$ and $d_1 = 0.1, 0.2, 0.3, 0.4$. Table 4.5 presents the results for $n = 512$, where the properties of each estimate seem robust to the choice of volatility function. Normalizing $\mathrm{Var}(\zeta_t)$ appears to be

	n	256				512				1024				2048			
d_1		m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE	m^*	Bias	SD	RMSE
	L	20	-0.151	0.086	0.174	19	-0.132	0.083	0.156	23	-0.120	0.072	0.140	24	-0.106	0.066	0.125
0.1	MN	41	-0.207	0.088	0.225	37	-0.190	0.079	0.206	41	-0.179	0.069	0.192	34	-0.167	0.070	0.181
	NL	58	-0.264	0.184	0.322	53	-0.234	0.149	0.278	50	-0.212	0.123	0.245	64	-0.197	0.097	0.220
	L	13	-0.101	0.090	0.136	16	-0.083	0.073	0.111	17	-0.065	0.063	0.091	22	-0.054	0.050	0.074
0.2	MN	17	-0.178	0.111	0.210	17	-0.152	0.098	0.181	19	-0.133	0.085	0.158	20	-0.114	0.077	0.138
	NL	22	-0.253	0.203	0.324	23	-0.215	0.164	0.270	22	-0.184	0.136	0.228	24	-0.160	0.111	0.194
	L	11	-0.072	0.084	0.111	14	-0.054	0.064	0.084	17	-0.040	0.050	0.063	22	-0.030	0.038	0.048
0.3	MN	12	-0.151	0.121	0.193	13	-0.118	0.099	0.154	15	-0.094	0.081	0.124	15	-0.069	0.070	0.098
	NL	12	-0.246	0.232	0.338	12	-0.191	0.188	0.268	14	-0.152	0.147	0.212	13	-0.113	0.121	0.165
	L	9	-0.061	0.087	0.106	12	-0.042	0.062	0.075	15	-0.028	0.045	0.053	20	-0.019	0.032	0.037
0.4	MN	8	-0.133	0.140	0.193	10	-0.098	0.105	0.143	11	-0.067	0.082	0.106	13	-0.046	0.062	0.077
	NL	8	-0.291	0.290	0.411	7	-0.205	0.247	0.321	8	-0.148	0.192	0.242	8	-0.096	0.147	0.175

Table 4.4: Monte Carlo bias, SD, RMSE of NBLS in settings L, MN, NL, for varying n and d_1.

d_1	g	$\exp(x)$ m^*	Bias	SD	RMSE	$(1+x)^2$ m^*	Bias	SD	RMSE	$\lvert 1+x \rvert$ m^*	Bias	SD	RMSE
0.1	$\tilde{\beta}^2$	—	-0.558	0.062	0.561	—	-0.557	0.059	0.560	—	-0.556	0.056	0.559
	$\tilde{\theta}$	—	-0.244	0.141	0.282	—	-0.262	0.122	0.289	—	-0.291	0.112	0.312
	$\hat{\theta}_{m^*}$	53	-0.234	0.149	0.278	47	-0.250	0.136	0.284	52	-0.280	0.128	0.308
0.2	$\tilde{\beta}^2$	—	-0.561	0.075	0.566	—	-0.560	0.072	0.564	—	-0.559	0.066	0.563
	$\tilde{\theta}$	—	-0.251	0.145	0.290	—	-0.267	0.125	0.295	—	-0.295	0.114	0.316
	$\hat{\theta}_{m^*}$	23	-0.215	0.164	0.270	17	-0.219	0.159	0.271	23	-0.255	0.148	0.295
0.3	$\tilde{\beta}^2$	—	-0.571	0.107	0.581	—	-0.570	0.105	0.580	—	-0.566	0.094	0.574
	$\tilde{\theta}$	—	-0.276	0.160	0.319	—	-0.285	0.136	0.316	—	-0.308	0.122	0.332
	$\hat{\theta}_{m^*}$	12	-0.191	0.188	0.268	12	-0.190	0.174	0.258	13	-0.220	0.174	0.280
0.4	$\tilde{\beta}^2$	—	-0.604	0.168	0.627	—	-0.603	0.172	0.627	—	-0.593	0.152	0.612
	$\tilde{\theta}$	—	-0.368	0.208	0.423	—	-0.358	0.181	0.401	—	-0.359	0.150	0.390
	$\hat{\theta}_{m^*}$	7	-0.205	0.247	0.321	8	-0.190	0.219	0.290	8	-0.205	0.219	0.299

Table 4.5: Monte Carlo bias, SD, RMSE of NBLS for varying $g(x)$ and d_1, with $n = 512$.

sufficient to capture most of the differences across functions. Results for other n are qualitatively similar and not shown.

Final comments

Though this section briefly addressed the choice of bandwidth m, it would evidently be desirable to develop a feasible rule for bandwidth selection. In a Gaussian or linear setting, Robinson (1994b) developed formulae for minimum-MSE bandwidth with respect to the basic averaged periodogram statistic, and these were further analyzed by Delgado and Robinson (1996). In principle these could be extended to the NBLS estimate, though the formulae will be highly complex, and feasible versions would require estimating memory parameters and other quantities. As in other circumstances, sensitivity to choice of m can be assessed by a "window-closing" approach, computing NBLS over a sensibly chosen grid of m values; since discrete Fourier transforms at all Fourier frequencies can be obtained simultaneously by the Fast Fourier Transform, and NBLS is algebraically

simple, this can cheaply be achieved, indeed a simple recursion deals with unit or other increases in m.

4.3 Finite sample properties of Weighted Narrow Band Least Squares under linear and stochastic volatility frameworks

We present Monte Carlo results for two settings, linear and nonlinear, where the nonlinear framework is similar to the one in the previous section. Under the linear model, we generate

$$Y_t = \theta\zeta_t + \varepsilon_t, \tag{4.7}$$

$$X_t = \zeta_t + \delta_t, \tag{4.8}$$

where we use the abbreviated notation $\zeta_t = \eta_{1t}$, $\delta_t = \eta_{2t}$, $\varepsilon_t = \eta_{3t}$, and for $i = 1, 2, 3$, $\{\eta_{it}\}$ is a zero mean Gaussian ARFIMA$(0, d_i, 0)$ process with variance σ_i^2. In the nonlinear case (see (4.1), (4.2)), we use

$$Y_t = (\beta\zeta_t + \varepsilon_t)^2, \tag{4.9}$$

$$X_t = (\zeta_t + \delta_t)^2, \tag{4.10}$$

where $\zeta_t = \xi_{1t}h(\eta_{1t})$, $\delta_t = \xi_{2t}h(\eta_{2t})$, $\varepsilon_t = \xi_{3t}h(\eta_{3t})$, and for $i = 1, 2, 3$, $\{\xi_{it}\}$ is an independent standard Gaussian sequence, and $\{\eta_{it}\}$ a zero mean Gaussian ARFIMA$(0, d_i, 0)$ with variance σ_i^2. In both models, the basic processes $\{\xi_{it}\}$ and $\{\eta_{it}\}$, $i = 1, 2, 3$, are all generated independently of each other, and we will denote the variances of ζ_t, δ_t, ε_t by σ_ζ^2, σ_δ^2, σ_ε^2 respectively.

Under each model, we employ 1,000 replications of series of length $n = 2048$

(when evaluating the impact of sample size, $n = 512, 8192$ are also used) and estimate θ by narrow-band regressions of Y_t on X_t, where $\theta \equiv \beta^2$ in the nonlinear setting. Note that both models can be written (see (2.39) and (3.21)) as

$$Y_t = \theta X_t + U_t, \tag{4.11}$$

with

$$U_t = \varepsilon_t - \theta\delta_t \tag{4.12}$$

in the linear setting and

$$U_t = \varepsilon_t^2 - \theta\delta_t^2 + 2\beta\zeta_t(\varepsilon_t - \beta\delta_t) \tag{4.13}$$

in the nonlinear setting. We present bias, standard deviation (SD) and root mean squared error (RMSE) of WNBLS estimates (2.62),

$$\tilde{\theta}(d) = \frac{\sum_{j=1}^{m} \lambda_j^{2d} \operatorname{Re}\{I_{XY}(\lambda_j)\}}{\sum_{j=1}^{m} \lambda_j^{2d} I_{XX}(\lambda_j)}, \tag{4.14}$$

for various values of d, both fixed and estimated (see (2.61)). All are evaluated at the bandwidth, m^*, that minimises RMSE.

Asymptotic theory

We first examine the performance of Nielsen's (2005) asymptotic theory for WNBLS, under the linear model, when δ_t is absent in (4.8). We set $\theta = 1$, $d_1 = 0.4$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$ and $\sigma_\varepsilon^2 = 2$. This simulation is comparable to his model A, although we focus on full-band estimates, i.e. $m = n/2$. (Given the independence between U_t and X_t, this choice dominates any other value of m.) Table 4.6 reports asymptotic (Asy.) and Monte Carlo (MC) SD for different values

d	Asy. SD	MC SD	Ratio
0.10	0.0176	0.0213	1.211
0.15	0.0155	0.0203	1.307
0.20	0.0152	0.0201	1.323
0.25	0.0154	0.0204	1.328
0.30	0.0157	0.0209	1.332
0.35	0.0161	0.0216	1.337
0.40	0.0166	0.0223	1.341
0.45	0.0171	0.0230	1.344

Table 4.6: Asymptotic and Monte Carlo SD of WNBLS, for varying d; linear setting with δ_t absent.

of d. Monte Carlo bias is negligible in this setting and therefore omitted. Note that Nielsen's (2005) theory requires (see (2.63))

$$(2d_1 + 2d_3 - 1)/4 < d \leq d_3, \tag{4.15}$$

which in this case is equivalent to $0.05 < d \leq 0.2$, but we compute his asymptotic SD also for $d > 0.2$. Here we find that Monte Carlo SD is almost always over 30% larger than the asymptotic one, so the asymptotic theory is not a good approximation even when $n = 2048$.

More realistically, a complete factor model such as (4.7), (4.8) allows the explanatory variable X_t to include an idiosyncratic component, δ_t. The discrepancy between X_t and the ideal explanatory variable, ζ_t, can be interpreted as a case of measurement error (ME), causing X_t to be correlated with U_t. While still compatible with Nielsen's (2005) assumptions, this would increase the Monte Carlo SD even further without changing the asymptotic one (as long as $d_2 < d_3$), thereby widening the gap between them. Figure 4.4 plots the theoretical and Monte Carlo SD of $\tilde{\theta}(d)$ relative to that of $\tilde{\theta}(d_3)$, for different values of d. Although the asymptotic and Monte Carlo levels in Table 4.6 substantially differ, their ratios across d are comparable, and $d = d_3$ is the optimal choice in both.

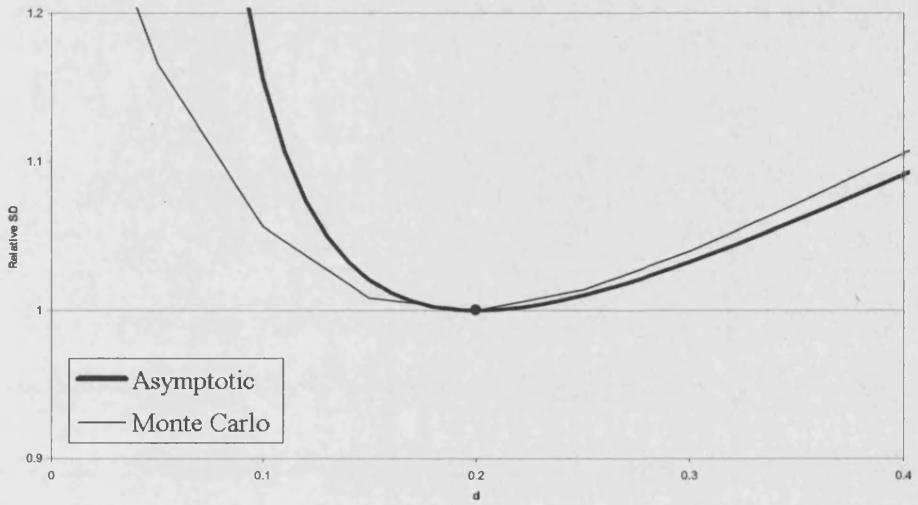


Figure 4.4: Asymptotic and Monte Carlo relative SD of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d; linear setting with δ_t absent.

Variation in measurement error

We present results for different types of ME, namely: no ME, i.e. δ_t absent in (4.8) or (4.10); antipersistent ME ($d_2 = -0.2$); iid ME ($d_2 = 0$); and long memory ME ($d_2 = 0.2$). In the nonlinear model, the antipersistent case would still generate $I(0)$ ME in (4.10) and is therefore omitted. In both settings we use $\theta = 1$, $d_1 = 0.4$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$, $\sigma_\varepsilon^2 = \sigma_\delta^2 = 2$, and $h(x) = \exp(x)$ as the volatility function for the nonlinear setting.

Table 4.7 reports Monte Carlo optimal bandwidth, bias and RMSE, under the linear setting, for various regression estimates of θ: unweighted NBLS, $\tilde{\theta}(0)$; the theoretically optimal but infeasible weighted estimate, $\tilde{\theta}(d_3)$; and feasible versions of it, $\tilde{\theta}(\hat{d}_3)$, where $\hat{d}_3$ is a consistent estimate of d_3. In the feasible cases, d_3 is estimated using the log-periodogram (LP) estimate (2.72), the Local Whittle (LW) estimate (2.75), or the Modified Local Whittle (MLW) estimate (2.87), based on the regression residuals from a first step unweighted NBLS regression, denoted

$\hat{U}_t$. These memory estimates are described in more detail in Section 2.7; the same bandwidth m is used in the first and second steps. Due to the modified spectral approximation in (2.87), when using MLW we compute WNBLS as

$$\tilde{\theta}(\hat{d},\hat{a}) = \left(\sum_{j=1}^{m} \frac{I_{XX}(\lambda_j)}{\hat{a}+\lambda_j^{-2\hat{d}}}\right)^{-1} \sum_{j=1}^{m} \frac{\mathrm{Re}\{I_{XY}(\lambda_j)\}}{\hat{a}+\lambda_j^{-2\hat{d}}}, \tag{4.16}$$

instead of (4.14). Table 4.8 reports bias and RMSE for these preliminary estimates of d_3.

In the model without ME all regression estimates have, as expected, virtually no bias and perform best in the full-band case. Here, $\tilde{\theta}(d_3)$ clearly exhibits an efficiency gain over $\tilde{\theta}(0)$, which is equivalent to OLS (4.5). However, as progressively more persistent ME is introduced, both estimates have increasing bias, and the RMSE of $\tilde{\theta}(d_3)$ grows much faster than that of $\tilde{\theta}(0)$. Indeed, in the presence of ME, simple NBLS always outperforms the weighted estimate. Here and throughout all experiments, estimates are biased towards zero, due to the negative correlation between X_t and U_t caused by ME. The feasible versions of WNBLS seem to closely match the infeasible one in both RMSE and bias, in many cases even appearing slightly better. This behaviour arises because whenever ME is present, the optimal weighting is actually obtained for $d < d_3$, so the negative bias of LP and LW, seen in Table 4.8, can actually work to their advantage. Although MLW actually displays positive bias, the weights in (4.16) do not depend on $\hat{d}_2$ alone but also on $\hat{a}$ in (2.87), allowing it to still outperform the infeasible estimate for $d_2 = -0.2$. The optimal bandwidths for each estimate are lower the more persistent the ME is, since frequencies closer to zero become more contaminated with the correlation between X_t and U_t.

Table 4.8 shows that both LP and LW perform relatively well throughout. The small biases are insufficient for the bias reduction properties of MLW to make a

	δ_t absent			$d_2 = -0.2$			$d_2 = 0$			$d_2 = 0.2$		
$\tilde{\theta}$	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
NBLS	1024	-0.0005	0.0273	81	-0.0279	0.0642	25	-0.0470	0.0897	12	-0.1301	0.1752
True d_3	1024	-0.0001	0.0201	53	-0.0283	0.0652	23	-0.0555	0.0933	10	-0.1326	0.1789
LP	1024	-0.0001	0.0209	53	-0.0297	0.0651	23	-0.0524	0.0928	10	-0.1322	0.1799
LW	1024	-0.0001	0.0204	53	-0.0296	0.0650	23	-0.0524	0.0930	10	-0.1321	0.1800
MLW	1024	0.0001	0.0205	53	-0.0301	0.0650	23	-0.0539	0.0937	10	-0.1339	0.1807

Table 4.7: Monte Carlo bias and RMSE of regression estimates, for different types of measurement error; linear setting.

		δ_t absent		$d_2 = -0.2$		$d_2 = 0$		$d_2 = 0.2$	
$\widehat{d}_3$	m	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
LP	80	-0.0020	0.0806	-0.0485	0.0934	-0.0574	0.0993	-0.0070	0.0821
LW	80	-0.0072	0.0628	-0.0519	0.0821	-0.0613	0.0892	-0.0108	0.0675
MLW	200	0.0491	0.1002	0.0569	0.1464	0.0184	0.1177	0.0418	0.0908

Table 4.8: Monte Carlo bias and RMSE of residual memory estimates, for different types of measurement error; linear setting.

	δ_t absent			$d_2 = 0$			$d_2 = 0.2$		
$\tilde{\theta}$	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
NBLS	973	-0.0042	0.0840	8	-0.1495	0.2717	8	-0.1944	0.3210
True d_3	973	-0.0042	0.0855	8	-0.1589	0.2829	8	-0.2020	0.3290
LP	973	-0.0042	0.0840	8	-0.1516	0.2753	8	-0.1969	0.3243
LW	973	-0.0043	0.0842	8	-0.1514	0.2747	8	-0.1966	0.3237
MLW	973	-0.0043	0.0847	8	-0.1532	0.2776	8	-0.1987	0.3266

Table 4.9: Monte Carlo bias and RMSE of regression estimates, for different types of measurement error; nonlinear setting.

difference; in fact, this estimate displays larger bias than LP and LW in three of the four cases. As expected, the much lower SD of LW makes it the best in RMSE. Although some of the bias can be attributed to estimation error, most of it surely comes from the "signal-plus-noise" nature of the residuals, as seen in (4.12). When δ_t is absent or when δ_t has the same memory as ε_t, LP and LW are essentially unbiased, while for $d_2 = -0.2, 0$ some bias is present.

Tables 4.9 and 4.10 present results for the nonlinear setting. Here it can be seen that the weighted estimate is always outperformed by NBLS, with ME causing much more significant bias. Even in the absence of ME, the optimal bandwidth is slightly below the full-band case, possibly as a consequence of U_t being orthogonal to but not independent of X_t, as can be seen by setting $\delta_t = 0$ in (4.13). All feasible weighted estimates outperform the infeasible one, which can again be explained by the negative biases found in Table 4.10. Biases are stronger here than in the linear setting, partly because of the estimation error and the nonlinear setting, but also because of the signal-plus-noise structure. Note that in this setting the $I(0)$ noise in (4.13) does not vanish even if δ_t is absent. For both LP and LW, bias is the main component of RMSE. Therefore, the bias reduction provided by MLW allows it to dominate the other estimates in the presence of ME. Again, the inferior performance of the weighted estimate relative to simple NBLS demonstrates that $d = d_3$ is not the optimal choice in this setting.

		δ_t absent		$d_2 = 0$		$d_2 = 0.2$	
$\widehat{d}_3$	m	Bias	RMSE	Bias	RMSE	Bias	RMSE
LP	80	-0.1512	0.1704	-0.1827	0.2016	-0.1640	0.1861
LW	80	-0.1562	0.1690	-0.1847	0.1969	-0.1666	0.1810
MLW	200	-0.0411	0.1827	-0.0986	0.1840	-0.0697	0.1788

Table 4.10: Monte Carlo bias and RMSE of residual memory estimates, for different types of measurement error; nonlinear setting.

Naturally, if d_3 is no longer the optimal choice for d, the usefulness of estimating it from the data can be questioned. This is verified in Figures 4.5 and 4.6, which show the RMSE of $\tilde{\theta}(d)$ relative to that of $\tilde{\theta}(d_3)$, for different values of d, in the linear and nonlinear settings. Only in the linear case without ME is $d = d_3$ optimal; in all other cases, the optimal value is smaller, and it is reduced the more persistent the ME is. In the nonlinear case the optimal values for d are always negative, and in a region excluded by (4.15). It should also be noted that, in the absence of information on the optimal d, NBLS should be chosen over $\tilde{\theta}(d_3)$ (or its feasible versions). Tables 4.11 and 4.12 report optimal bandwidth, bias and RMSE for $\tilde{\theta}(d)$, with $d = 0$, 0.2 and the values of d that minimise RMSE in each case (indicated in bold-face), in the linear and nonlinear settings. The degradation in performance with more persistent ME can still be seen here, and bias is often slightly smaller for the optimal d. However, the variation in bias across d is relatively small, and most of the variation in RMSE can be explained by variations in SD.

The minimization of RMSE at values different from $d = d_3$ is surprising since it does not conform to the asymptotic theory. A frequency domain generalised least squares approach will weigh the contribution of each frequency by the inverse of their approximate SD, thereby "whitening" the observations. A possible explanation for the discrepancy lies in the approximation error in $f_U(\lambda) \sim C|\lambda|^{-2d}$, which in the limit theory is made irrelevant by assuming enough smoothness in

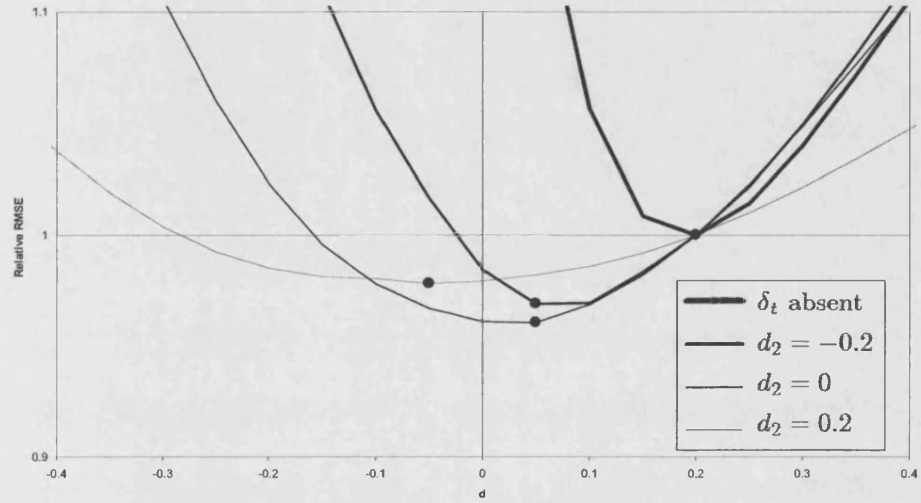


Figure 4.5: Relative RMSE of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d and d_2; linear setting.

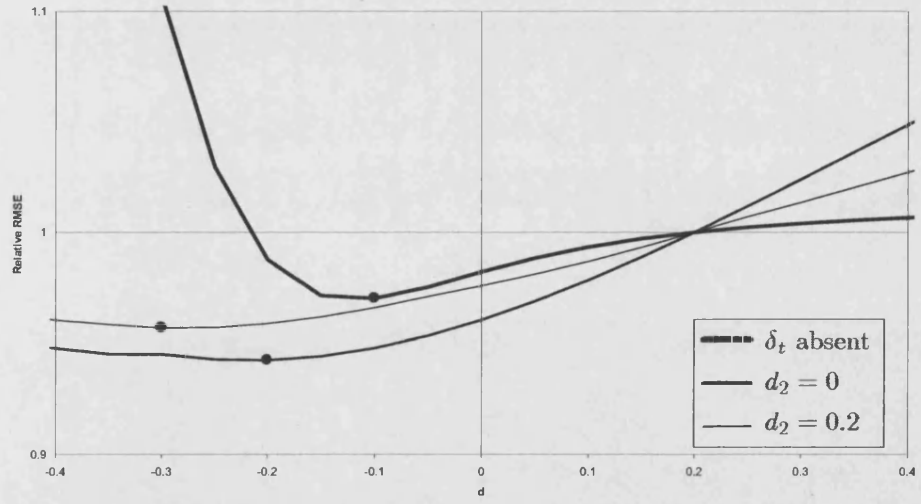


Figure 4.6: Relative RMSE of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d and d_2; nonlinear setting.

the spectral density, but can play a major role in finite samples. The whitening approach will give low weight to the frequencies closer to zero, where variance is higher but the previous approximation is more accurate, and will boost the impact of more distant frequencies where the approximation is not so accurate. Another relevant factor is the coherence between X_t and U_t, here generated by δ_t, which is the leading source of bias. Being of smaller order than the spectral pole, it will be irrelevant asymptotically, but this also means higher frequencies are more contaminated than lower ones. Again, decreasing the weight of the lowest frequencies is likely to worsen the estimation. Both these factors lead to an optimal d that will tend to be lower than d_2; in some circumstances they can outweigh the heteroscedasticity in the periodogram, and the optimal d will be negative, as can be seen in Figures 4.5 and 4.6 and Tables 4.11 and 4.12.

Variation in sample size

Failure of asymptotic theory to provide a good approximation in finite samples is further explored by changing the sample size. Figures 4.7 and 4.8 and Tables 4.13 and 4.14 present similar results to Figures 4.5 and 4.6 and Tables 4.11 and 4.12, for $n = 512, 2048, 8192$. We set $\theta = 1$, $d_1 = 0.4$, $d_2 = 0$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$, $\sigma_\varepsilon^2 = \sigma_\delta^2 = 2$, and use $h(x) = \exp(x)$ as the volatility function for the nonlinear setting. In both the linear and nonlinear settings, the optimal value for d increases with n, but not dramatically. Even for $n = 8192$, the optimal d is not only below d_3, but also outside the parameter range in (4.15). For all values of d, there is a strong improvement in both bias and RMSE as n increases. While in the linear case the optimal bandwidth for each d increases with n, in the nonlinear setting it is often higher for $n = 512$ than for $n = 8192$. Bandwidths for $n = 2048$ are the lowest of the three sample sizes, suggesting a "U-shaped" bandwidth profile that will continue diverging to infinity as the theory requires.

	δ_t absent			$d_2 = -0.2$			$d_2 = 0$			$d_2 = 0.2$		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.05	1024	-0.0005	0.0331	67	-0.0303	0.0663	33	-0.0519	0.0902	**12**	**-0.1269**	**0.1750**
0.00	1024	-0.0005	0.0273	55	-0.0279	0.0642	25	-0.0470	0.0897	12	-0.1301	0.1752
0.05	1024	-0.0004	0.0235	**52**	**-0.0294**	**0.0632**	**25**	**-0.0502**	**0.0896**	11	-0.1298	0.1757
0.20	**1024**	**-0.0001**	**0.0201**	39	-0.0283	0.0652	23	-0.0555	0.0933	10	-0.1326	0.1789

Table 4.11: Monte Carlo bias and RMSE of $\tilde{\theta}(d)$, for varying d and different types of measurement error; linear setting. The minimum RMSE choice of d is indicated in bold-face.

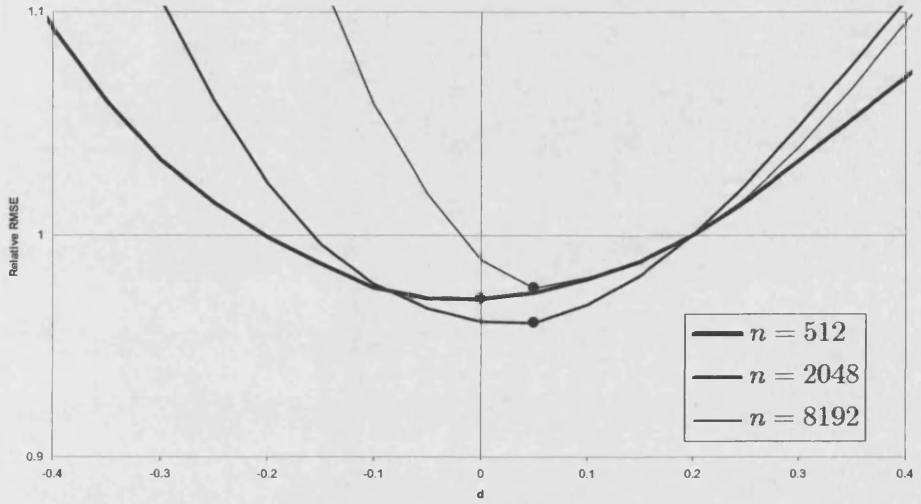


Figure 4.7: Relative RMSE of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d and n; linear setting.

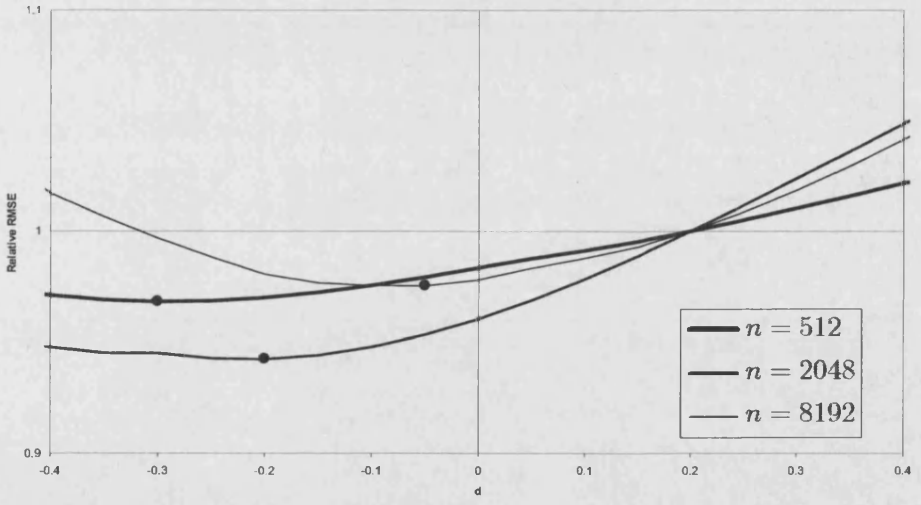


Figure 4.8: Relative RMSE of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d and n; nonlinear setting.

	δ_t absent			$d_2 = 0$			$d_2 = 0.2$		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.30	1022	-0.0025	0.0946	14	-0.1476	0.2674	**14**	**-0.1901**	**0.3148**
-0.20	976	-0.0033	0.0845	**8**	**-0.1386**	**0.2666**	14	-0.1963	0.3154
-0.10	**973**	**-0.0039**	**0.0830**	8	-0.1441	0.2681	14	-0.2022	0.3178
0.00	973	-0.0042	0.0840	8	-0.1495	0.2717	8	-0.1944	0.3210
0.20	973	-0.0042	0.0855	8	-0.1589	0.2829	8	-0.2020	0.3290

Table 4.12: Monte Carlo bias and RMSE of $\tilde{\theta}(d)$, for varying d and different types of measurement error; nonlinear setting. The minimum RMSE choice of d is indicated in bold-face.

n	512			2048			8192		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
0.00	**15**	**-0.0950**	**0.1555**	25	-0.0470	0.0897	61	-0.0271	0.0532
0.05	13	-0.0905	0.1560	**25**	**-0.0502**	**0.0896**	**61**	**-0.0296**	**0.0525**
0.20	12	-0.0977	0.1601	23	-0.0555	0.0933	41	-0.0265	0.0538

Table 4.13: Monte Carlo bias and RMSE of $\tilde{\theta}(d)$, for varying d and n; linear setting. The minimum RMSE choice of d is indicated in bold-face.

Figures 4.9 and 4.10 illustrate the distributional properties of NBLS by plotting kernel density estimates for varying n, under the linear and nonlinear setting. Density estimates are computed for a sequence of $s = 50,000$ NBLS estimates b_i, $i = 1, \ldots, s$, using

$$\widehat{f}(b) = \frac{1}{sh} \sum_{i=1}^{s} \phi\left(\frac{b_i - b}{h}\right), \tag{4.17}$$

where $\phi(\cdot)$ is the standard Gaussian density function and the bandwidth h is chosen using (3.31) of Silverman (1986),

$$h = 0.9 s^{-1/5} \min(SD, IQR/1.34), \tag{4.18}$$

where SD and IQR are the sample standard deviation and interquartile range of the b_i. Estimates for other values of d yield very similar shapes and are thus omitted. However, unlike for other values of d, NBLS is not covered by (4.15). Still, in the linear case all curves in Figure 4.9 seem to be fairly close in shape to

n		512			2048			8192	
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.30	**15**	**-0.3060**	**0.4520**	14	-0.1476	0.2674	16	-0.0583	0.1419
-0.20	15	-0.3145	0.4528	**8**	**-0.1386**	**0.2666**	13	-0.0586	0.1396
-0.05	15	-0.3261	0.4572	8	-0.1468	0.2697	**11**	**-0.0594**	**0.1389**
0.00	15	-0.3296	0.4591	8	-0.1495	0.2717	11	-0.0608	0.1392
0.20	8	-0.3096	0.4669	8	-0.1589	0.2829	8	-0.0555	0.1424

Table 4.14: Monte Carlo bias and RMSE of $\tilde{\theta}(d)$, for varying d and n; nonlinear setting. The minimum RMSE choice of d is indicated in bold-face.

that of a normal density. On the contrary, densities in Figure 4.10 are all highly skewed to the left, even for $n = 8192$, suggesting that the asymptotic distribution under the nonlinear setting might not be normal. In both settings, bias and SD seem to be decaying at the same rate, which is natural given our minimum RMSE bandwidth choice.

Variation in the signal to noise ratio

Figures 4.11 through 4.14 and Tables 4.15 through 4.18 can be interpreted in the same way as Figures 4.5 and 4.6 and Tables 4.11 and 4.12, for the linear and nonlinear settings, where we first change the variance of the ME, then the variance of the signal. In both experiments we start with $\theta = 1$, $d_1 = 0.4$, $d_2 = 0$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$, $\sigma_\varepsilon^2 = \sigma_\delta^2 = 2$, and $h(x) = \exp(x)$ as the volatility function for the nonlinear setting. The variance of the ME in the first experiment is then set to $\sigma_\delta^2 = 1/2$, 2, 8, by varying σ_2^2 in the linear setting, and by using $h_k(x) = k\exp(x)$, with $k = 1/2$, 1, 2, as the volatility function for δ_t, while keeping σ_2^2 constant, in the nonlinear setting. The resulting sequences δ_t are consequently the same, up to a multiplicative factor, for each value of σ_δ^2. In the second experiment, the variance of the signal is changed by choosing σ_1^2 so that $\sigma_\zeta^2 = 2$, 4, 8.

These parameters affect the accuracy of the estimates by influencing the relative variance of X_t and U_t in (4.7), which can be interpreted as a signal to noise

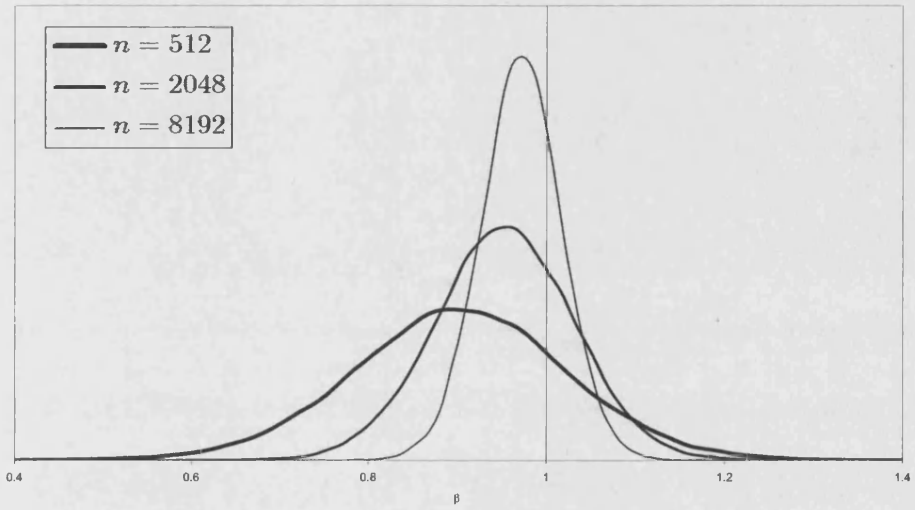


Figure 4.9: Kernel density estimates of NBLS for varying n; linear setting.

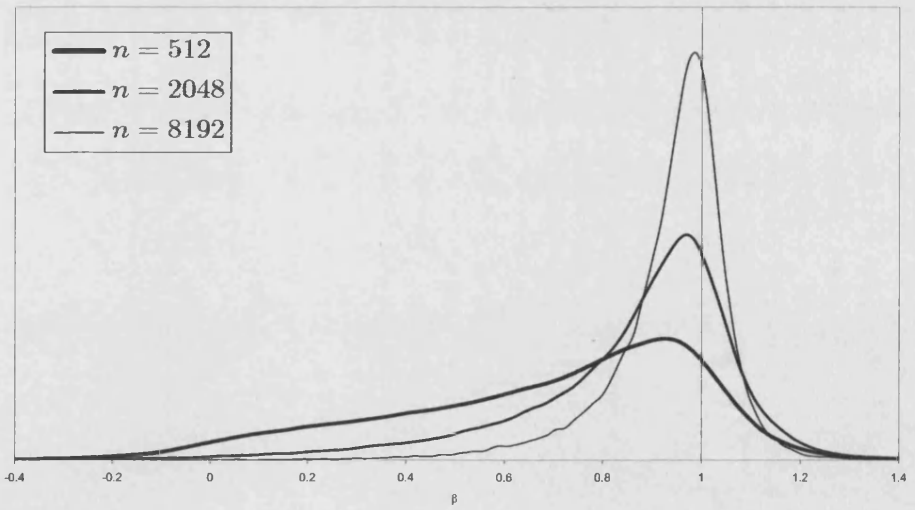


Figure 4.10: Kernel density estimates of NBLS for varying n; nonlinear setting.

σ_δ^2		1/2			2			8	
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.10	142	-0.0297	0.0629	39	-0.0524	0.0913	**10**	**-0.0897**	**0.1516**
0.00	112	-0.0329	0.0584	25	-0.0470	0.0897	10	-0.0969	0.1527
0.05	82	-0.0301	0.0579	**25**	**-0.0502**	**0.0896**	10	-0.1005	0.1541
0.10	**81**	**-0.0326**	**0.0578**	25	-0.0533	0.0904	10	-0.1041	0.1560
0.20	55	-0.0280	0.0596	23	-0.0555	0.0933	7	-0.0893	0.1595

Table 4.15: Monte Carlo bias and RMSE of $\tilde{\theta}(d)$, for varying d and σ_δ^2; linear setting. The minimum RMSE choice of d is indicated in bold-face.

σ_δ^2		1/2			2			8	
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.40	1022	-0.0477	0.1333	18	-0.1454	0.2682	**4**	**-0.4607**	**0.5858**
-0.20	98	-0.0444	0.1233	**8**	**-0.1386**	**0.2666**	4	-0.4703	0.5880
-0.10	**68**	**-0.0447**	**0.1216**	8	-0.1441	0.2681	4	-0.4751	0.5903
0.00	66	-0.0491	0.1228	8	-0.1495	0.2717	4	-0.4798	0.5933
0.20	63	-0.0555	0.1308	8	-0.1589	0.2829	4	-0.4889	0.6007

Table 4.16: Monte Carlo bias and RMSE of $\tilde{\theta}(d)$, for varying d and σ_δ^2; nonlinear setting. The minimum RMSE choice of d is indicated in bold-face.

ratio, and the covariance between X_t and U_t, which can be seen in (4.12) and (4.13) to depend crucially on δ_t.

Figures 4.11 and 4.12 and Tables 4.15 and 4.16 show that both m^* and the optimal d decrease rather heavily as σ_δ^2 increases, especially in the nonlinear setting. For large values of σ_δ^2, the common component in X_t and U_t becomes very important, influencing even frequencies relatively close to zero. As a result, both the bandwidth and the weights should adjust so that only the lowest frequencies (where the spectral pole still dominates) have significant influence. Tables 4.15 and 4.16 display a strong degradation in both bias and RMSE, caused by the increased coherence between regressor and residuals.

While increasing σ_δ^2 influences both U_t and X_t, scaling up the common component in both, increasing the cointegrating parameter θ boosts the weight of the common component in U_t alone, keeping X_t constant. Still, this provokes a comparable increase in correlation, causing very similar effects to those reported for

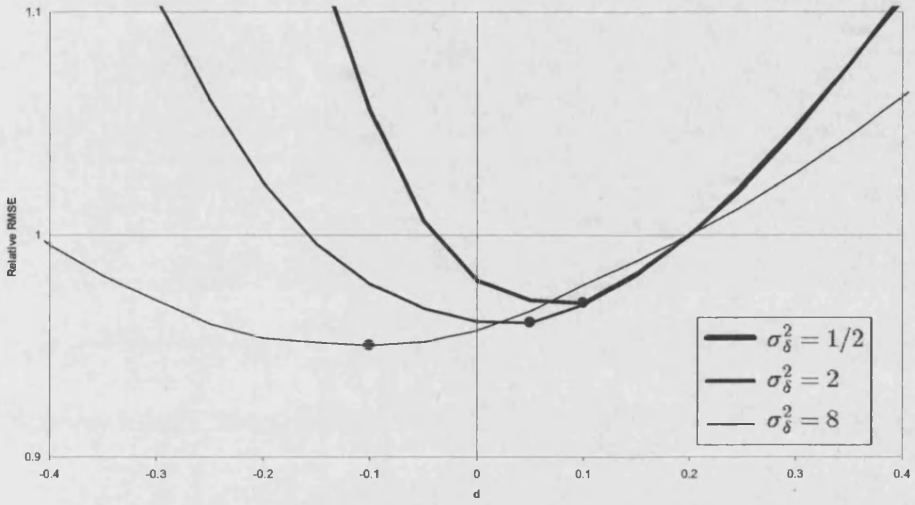


Figure 4.11: Relative RMSE of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d and σ_δ^2; linear setting.

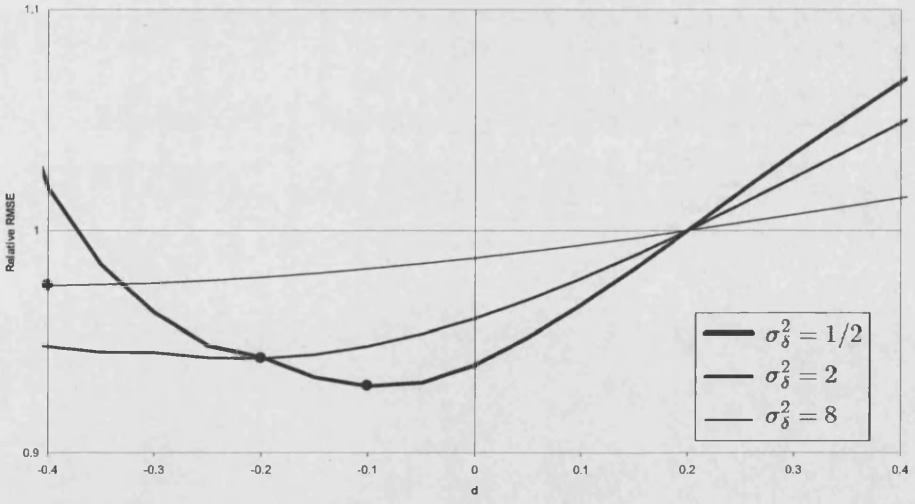


Figure 4.12: Relative RMSE of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d and σ_δ^2; nonlinear setting.

σ^2_ζ	2			4			8		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.05	**25**	**-0.0843**	**0.1380**	33	-0.0519	0.0902	42	-0.0310	0.0585
0.00	23	-0.0852	0.1382	25	-0.0470	0.0897	**40**	**-0.0327**	**0.0579**
0.05	22	-0.0883	0.1392	**25**	**-0.0502**	**0.0896**	39	-0.0346	0.0580
0.20	19	-0.0946	0.1455	23	-0.0555	0.0933	25	-0.0305	0.0603

Table 4.17: Monte Carlo bias and RMSE of $\tilde{\theta}(d)$, for varying d and σ^2_ζ; linear setting. The minimum RMSE choice of d is indicated in bold-face.

σ^2_δ. Monte Carlo results for this case are therefore omitted.

Figures 4.13 and 4.14 and Tables 4.17 and 4.18 display the effect of the strength of the signal ζ_t. In the linear case, this scales up the signal in X_t without affecting U_t. In the nonlinear case, both are affected, but since the SV model used generates heavily leptokurtic processes (implying that the variance of ζ_t^2 is the major contribution to the variance of X_t) and ζ_t only affects U_t through a white noise component (thus having a bounded contribution to the spectrum around the zero frequency), the impact on U_t will be minimal compared to that on X_t. In both models, increasing σ^2_ζ will have the double effect of increasing the variance of X_t, thereby making the observables more correlated at all frequencies, and scaling up the spectral pole caused by the memory in η_{1t}, improving the local signal to noise ratio. While both effects will have a clearly positive influence on the accuracy of the estimates, as seen in Tables 4.17 and 4.18, the effect on m^* and on the optimal d is not clear, as even frequencies distant from zero become less contaminated by the dependence between X_t and U_t. As a result, Figures 4.13 and 4.14 show very little variation on relative RMSE with σ^2_ζ.

Distributional properties of residual memory estimates

While the previous experiments show that estimates of residual memory are not necessarily useful for choosing d in (4.14), they might still be relevant for other purposes, namely to verify if a cointegrating relationship exists at all. The

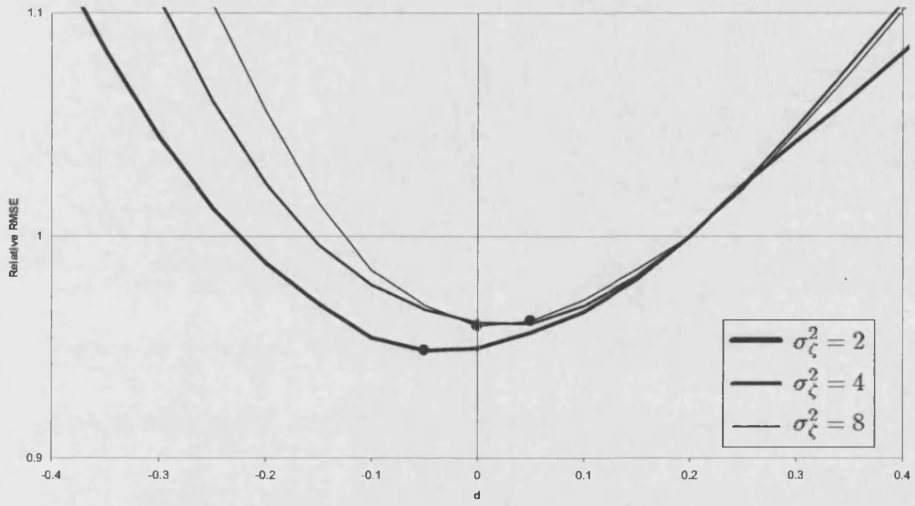


Figure 4.13: Relative RMSE of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d and σ_ζ^2; linear setting.

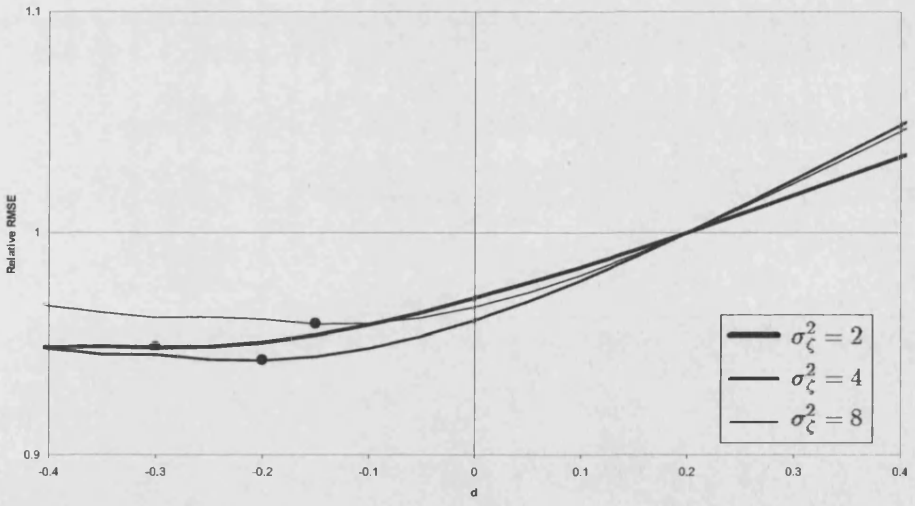


Figure 4.14: Relative RMSE of $\tilde{\theta}(d)$ versus $\tilde{\theta}(d_3)$, for varying d and σ_ζ^2; nonlinear setting.

σ^2_ζ	2			4			8		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.30	**8**	**-0.3350**	**0.4722**	14	-0.1476	0.2674	18	-0.0702	0.1696
-0.20	8	-0.3453	0.4733	**8**	**-0.1386**	**0.2666**	8	-0.0628	0.1695
-0.15	8	-0.3504	0.4750	8	-0.1414	0.2671	**8**	**-0.0642**	**0.1691**
0.00	8	-0.3649	0.4833	8	-0.1495	0.2717	8	-0.0680	0.1705
0.20	7	-0.3714	0.4980	8	-0.1589	0.2829	8	-0.0726	0.1764

Table 4.18: Monte Carlo bias and RMSE of $\tilde{\theta}(d)$, for varying d and σ^2_ζ; nonlinear setting. The minimum RMSE choice of d is indicated in bold-face.

use of the LP and LW estimates is well established by now, and their finite-sample properties have been examined in various settings (see e.g. Robinson and Henry, 1999; Nielsen and Frederiksen, 2005). In finite samples, LW is generally found to have bias of similar magnitude but lower variance than LP, to conform with the asymptotic distributions (2.73) and (2.76). However, the recent MLW estimate has not yet been directly compared to LW. The findings of Hurvich, Moulines, and Soulier (2005), Hurvich and Ray (2003), and Table 4.10, indicate that, even for moderate sample sizes, MLW can successfully reduce bias in the presence of a "signal-plus-noise" structure, but at the cost of a substantially higher SE than LW. We now present a short comparison of finite-sample distributional properties of LW and MLW in the context of residual memory estimation, for $n = 512, 2048, 8192$. Residuals are obtained from $s = 1,000$ replications of NBLS regression in the linear and nonlinear settings, with $\theta = 1$, $d_1 = 0.4$, $d_2 = 0$, $d_3 = 0.2$, $\sigma^2_\zeta = 4$, $\sigma^2_\varepsilon = \sigma^2_\delta = 2$, and $h(x) = \exp(x)$ as the volatility function for the nonlinear setting. The minimum RMSE bandwidths reported in Tables 4.13 and 4.14 are used in this step. Then, LW and MLW estimates are constructed from the residuals for a grid of bandwidths (from 10 to $n/2$, with increments of 10), allowing us to approximately locate the minimum RMSE bandwidth for each memory estimate. Figures 4.15 through 4.18 show kernel density estimates (see (4.17), (4.18)) of LW and MLW, under the linear and nonlinear settings, using the

	Linear setting			Nonlinear setting		
n	512	2048	8192	512	2048	8192
LW	140	270	360	240	320	300
MLW	240	940	3720	240	680	4010

Table 4.19: Approximate minimum RMSE bandwidths of $\hat{d}_{LW}$ and $\hat{d}_{MLW}$, for varying n; linear and nonlinear settings.

approximately optimal bandwidths given in Table 4.19.

Table 4.19 shows that while LW works best with a narrow-band approach, MLW has optimal bandwidth rather close to $n/2$. This is possible because, unlike LW, MLW corrects for the presence of iid noise, and thus its spectral approximation is relatively accurate throughout all frequencies considered. However, for higher frequencies to be informative, the absence of short memory dynamics is crucial; the inclusion of, say, ARMA dynamics in any of the $\{\eta_{it}\}$ would undoubtedly require MLW bandwidths to be much lower.

All curves in Figures 4.15 and 4.16 suggest that the finite-sample density of LW is fairly close in shape to that of a normal density, but heavily biased downwards. While in the linear setting both bias and SD are substantially reduced when n increases, estimation in the nonlinear one seems surprisingly insensitive to sample size; even for $n = 8192$ the mean is much closer to 0 than to 0.2. Figures 4.17 and 4.18 highlight a potential problem of MLW in finite samples. In several cases, the distribution of MLW is bimodal, with peaks close to 0 and 1/2, the boundaries of the parameter space. In the nonlinear setting, this behaviour is apparent even for $n = 8192$, with a small mode close to the true parameter value being barely distinguishable. Performance in the linear setting is more encouraging: for $n = 8192$, the "boundary" modes disappear and are replaced by an essentially unbiased unimodal density. Still, it is worth noting that the SD in this case is roughly twice that of LW, and that the tails of the density are still moderately asymmetric. The findings of bimodality and higher SD in MLW are maintained in alternative

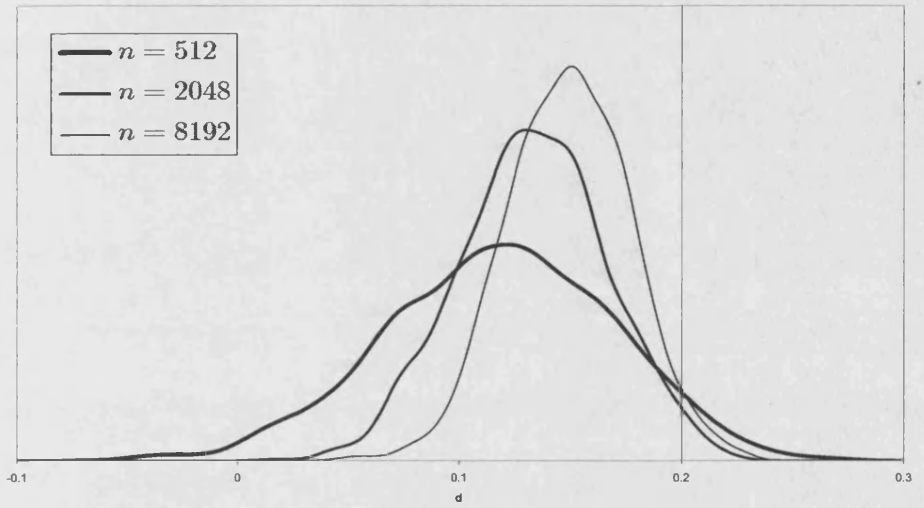


Figure 4.15: Kernel density estimates of LW for varying n; linear setting.

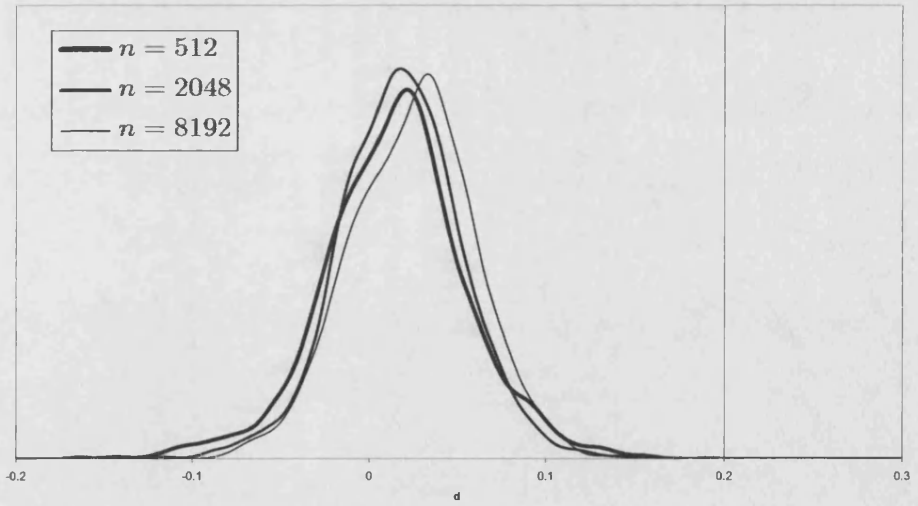


Figure 4.16: Kernel density estimates of LW for varying n; nonlinear setting.

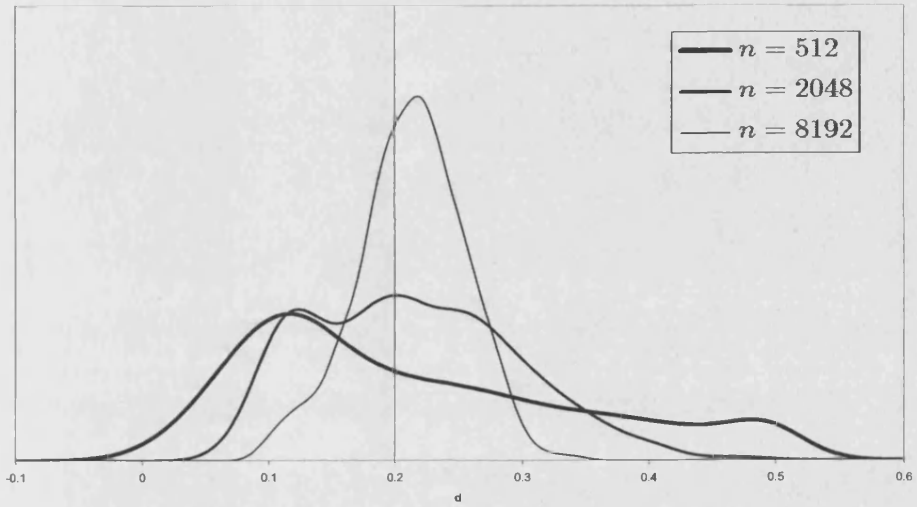


Figure 4.17: Kernel density estimates of MLW for varying n; linear setting.

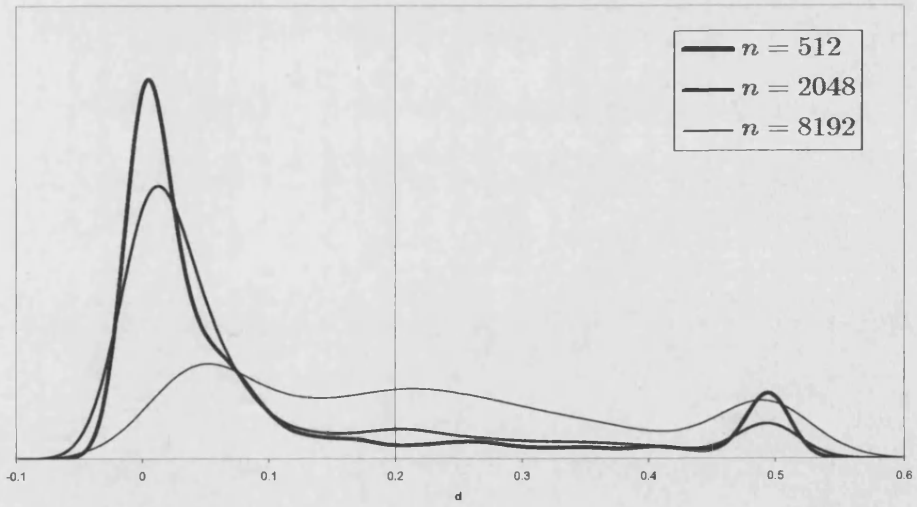


Figure 4.18: Kernel density estimates of MLW for varying n; nonlinear setting.

(unreported) experiment designs, suggesting that they are linked to the additive noise structure itself, not to first step estimation error or nonlinearity. Estimation error in the first step regression actually contaminates the true errors (4.12), (4.13) with a higher memory component (in this case, of memory $d_1 = 0.4$), which should induce a positive contribution to both bias (thereby reducing the LW bias) and SD.

Final comments

The results presented indicate that asymptotic theory should not necessarily be expected to provide a good approximation to finite-sample performance.

We first showed that, even in a standard setting, where error and regressor are independent Gaussian processes, Monte Carlo SD deviates substantially from its asymptotic counterpart. While in this setting $d = d_3$ is the optimal choice for WNBLS, further results demonstrate that the introduction of nonlinearity or ME makes this choice sub-optimal, and indeed dominated by simple NBLS. Furthermore, the nonlinear setting always yields a negative optimal d, even in the absence of ME. Although optimal bandwidths somewhat vary, they appear to be lower than those implied by commonly used feasible rules. For instance, Nielsen (2005) uses $m = [n^{0.4}]$ and $m = [n^{0.5}]$, yielding $m = 21, 45$ for $n = 2048$, which would be clearly too high for most of the nonlinear settings considered. While in the linear setting the RMSE profiles seem to be relatively sensitive to the choice of d, in the nonlinear one a wide range of values for d perform comparably; this is possibly a consequence of the lower bandwidths used. The optimal choice of d seems to be sensitive to most parameters in the model, so a feasible rule would undoubtedly require preliminary estimation of these.

All the finite sample results were generated under assumptions which might not be realistic in practice, such as Gaussianity and independence of the under-

lying, unobservable processes, and the absence of short memory dynamics. These assumptions constitute a best-case scenario, and relaxing them might well widen the gap between theoretical predictions and finite sample performance. More elaborate methods, such as those of Hualde and Robinson (2007), exhibit more desirable asymptotic properties under conditions that are in some sense weak, though it is not clear to what extent they can be justified when the linearity assumptions underlying them are relaxed. Heavy dependence on preliminary estimates may also hamper their finite sample performance.

A brief comparison of residual memory estimates was also presented. It seems that, while MLW is found to dominate LW in RMSE for large enough n, due to the large negative bias of the latter, it displays high dispersion and bimodality, which can be especially misleading in cointegration analysis, where the focus is often on the difference between memory estimates obtained from observables and residuals. On the contrary, LW, being biased downwards in both cases, might yield more accurate inference on the existence and degree of fractional cointegration. Evaluation of these issues is beyond the scope of this thesis.

Chapter 5

Narrow Band Principal Components estimation of multivariate factor models with long memory

5.1 Introduction

The present chapter is an application of the techniques developed in Chapter 3 to estimation of a multivariate factor model. We extend the bivariate model presented there to allow for more than two observables and multiple common factors, i.e. more than one cointegrating relation, as illustrated in Section 2.3. On the one hand, cointegrating relations between a potentially large number of asset returns can be of interest, while on the other, Ross (1976) and others suggest the need for additional unobservable factors in asset pricing models. We prove that, under analogous conditions to the bivariate setting, long memory in higher moments can allow consistent estimation of the space spanned by the factor loadings. We also

show how the factor loadings may be identified by suitable linear restrictions, as in the classical factor analysis literature (see e.g. Anderson and Rubin, 1956).

We use the Narrow Band Principal Components (NBPC) estimate of the cointegrating vectors. This estimate, introduced by Phillips and Ouliaris (1988) in the $I(1)/I(0)$ setting, and used by Morana (2004) in the case of stationary long memory, relies on the fact that the spectral density matrix of a cointegrated multivariate time series is singular at frequency zero, and its null space corresponds to the cointegrating vectors. Chen and Hurvich (2003), Chen and Hurvich (2006) employ a fixed band version of the averaged periodogram matrix, computed using a tapered Fourier transform, but they otherwise also rely on singularity of the respective limit matrix. We extend the results of these authors to a setting where linear process assumptions are unavailable, and furthermore we allow common components with possibly differing memory parameters, as in Chen and Hurvich (2006). In addition to nonlinearity, our setup also generalises the model of Chen and Hurvich (2006) by allowing the memory of factors to be equal in sets, for cointegration between observables of differing memory (what Robinson and Yajima (2002) termed "polynomial cointegration"), and for the presence of individual error terms, possibly correlated and with potentially differing memory parameters.

The use of regression based estimates of cointegrating vectors, namely NBLS (2.55) and multivariate generalisations thereof, is more established in the fractional cointegration literature than principal components. Nonetheless, the factor model context introduced in Sections 2.3 and 2.5 seems to warrant treatment by methods adapted from classical factor analysis. On the one hand, we have argued throughout Chapter 2 that in our setting the designation of the left-hand side variable is arbitrary. Unlike regression based estimators, the NBPC approach we pursue in this chapter is invariant to the labelling of observables, and in particular does not require assumptions of nonzero factor exposures of a subset of observ-

ables (though such assumptions may still be useful in this setting, to resolve part of the indeterminacy in the estimates). On the other hand, the transformation of the structural factor model to a regression specification will necessarily introduce errors-in-variables bias, as in (2.43). While we show in Chapter 3 that this bias can be made asymptotically negligible by means of local dominance of averaged periodograms around zero frequency, a number of experiments in Chapter 4 indicate that substantial small sample bias may still be present in NBLS. By contrast, if the underlying idiosyncratic errors have an approximately scalar covariance matrix (which may not be unreasonable when the observables are squared asset returns and a sufficient number of factors is used), a NBPC estimate may be expected to produce lower bias in small samples. Finally, principal components estimation is a commonly used approach in the APT literature (see e.g. Chamberlain and Rothschild, 1983; Connor and Korajczyk, 1986), and furthermore this extension illustrates the flexibility of the theoretical tools of Chapter 3, which can be adapted with relative ease to establish asymptotic properties of estimates other than NBLS.

The following section describes the model and the main assumptions on its underlying components. Section 3 describes the model transformation and establishes properties of the averaged periodogram matrix, using propositions and lemmas in Appendix A, also relying on results from Chapter 3. Section 4 introduces the NBPC estimate, proves its consistency for the space spanned by the factor loadings, and describes how these may be identified via suitable linear restrictions. Section 5 contains a small Monte Carlo study of finite sample properties. Section 6 presents an empirical application of both NBPC and NBLS to estimation of exposures to common risk factors for a set of European large-cap equity indices. Section 7 contains some concluding remarks.

5.2 A multivariate factor model with long memory in higher moments

Suppose a vector time series $z_t = (z_{1t}, \ldots, z_{qt})'$, $t \in \mathbb{Z}$, is generated by a multivariate factor model with J factors, for integer J satisfying $1 \leq J < q$. Write

$$z_{it} = \sum_{j=1}^{J} \beta_{ij} \zeta_{jt} + \delta_{it}, \tag{5.1}$$

or in vector form

$$z_t = \beta \zeta_t + \delta_t, \tag{5.2}$$

where β is a $q \times J$ matrix of factor loadings, ζ_t is a $J \times 1$ vector of unobservable factors, and δ_t is a $q \times 1$ vector of idiosyncratic innovations. We will assume that J is known throughout the chapter, though we briefly discuss estimation of J in Section 4. As in (2.36), (2.37) and (3.1), (3.2), this may be interpreted as a factor model for asset returns, now in the spirit of APT: the unobservable factors ζ_t would be different sources of risk, the loadings β represent the exposures of each asset to each common risk source, and the δ_t encompass asset-specific sources of risk. The following assumption generalises Assumption 3.1 to this multivariate setting.

Assumption 5.1 *For $j = 1, \ldots, J$, $i = 1, \ldots, q$, $t \in \mathbb{Z}$,*

$$\zeta_{jt} = \eta_{jt} g_{jt}, \qquad \delta_{it} = \xi_{it} h_{it}, \tag{5.3}$$

where for real-valued functions g_j, h_i,

$$g_{jt} = g_j(\nu_{jt}), \qquad h_{it} = h_i(\chi_{it}), \tag{5.4}$$

and:

(i) $\{\eta_{jt}\}$, $\{\xi_{it}\}$ *are jointly iid processes with zero mean, and* $\{\eta_{jt}\}$ *is independent of* $\{\eta_{kt}\}$, $\{\xi_{lt}\}$, *for any* $k = 1, \ldots, J$ *such that* $k \neq j$, *and* $l = 1, \ldots, q$;

(ii) $\{\nu_{jt}\}$ *is* $I(d_j)$, $\{\chi_{it}\}$ *is* $I(b_i)$, *and*

$$\tfrac{1}{2} > d_1 \geq \ldots \geq d_J > b_1 \geq \ldots \geq b_q \geq 0; \tag{5.5}$$

(iii) $\{\nu_{jt}\}$, $\{\chi_{it}\}$ *are standard Gaussian processes, independent of each other and of* $\{\eta_{kt}\}$, $\{\xi_{lt}\}$ *for any* $k = 1, \ldots, J$, $l = 1, \ldots, q$;

(iv) There is a finite integer p *such that, for all* $j = 1, \ldots, J$,

$$p = \min\left\{k \in \mathbb{N} : E(\eta_{jt}^k) E\left\{g_j^k(\nu_{jt})\nu_{jt}\right\} \neq 0\right\}; \tag{5.6}$$

(v) $\{\eta_{jt}\}$, $\{\xi_{it}\}$, $\{g_{jt}\}$, $\{h_{it}\}$ *have finite* $4p$*-th moments.*

This assumption shares the main features of its bivariate counterpart. The unobservable processes ζ_{jt}, δ_{it} follow SV models, and are therefore serially uncorrelated but not necessarily independent. Each individual factor and idiosyncratic component may have a different volatility function g_j or h_i. The underlying persistence of the common factors ζ_{jt} dominates that of the idiosyncratic components δ_{it} by (5.5), which together with assumption (iv) will ensure the existence of (at least) $q - J$ cointegrating relationships in the p-th powers of (5.1). The exponent p is assumed known, as in Chapter 3. In most practical applications it may be reasonable to assume that $p = 2$ satisfies (5.6), in which case the cointegrating relationships are present in the volatilities.

Due to the more complex nature of the proof, (i) restricts each η_{jt} to be independent of all other underlying processes; in part (i) of Assumption 3.1 we only

used the jointly iid property for corresponding processes, not restricting contemporaneous dependence between them in any way but that implied by (iv). This is a simplifying assumption, which could arguably be relaxed with a suitable strengthening of (iv). However, unlike the bivariate case (where only one factor and two idiosyncratic errors are present, and the latter do not interact) these dependences would give rise to the appearance of cross-product terms involving more than two underlying processes. The need for special treatment of these terms seems to indicate that such an extension is not trivial, and might require additional theoretical tools.

5.3 The transformed model and the averaged periodogram matrix

Given observations z_{it}, $i = 1, \ldots, q$, $t = 1, \ldots, n$, transform the data by

$$\begin{aligned} Z_{it} = z_{it}^p &= p! \sum_{\substack{p_j : 0 \le p_j \le p \\ \Sigma p_j \le p, j \in P_J}} \frac{\delta_{it}^{p - \Sigma p_j}}{(p - \Sigma p_j)!} \prod_{j=1}^{J} \frac{\beta_{ij}^{p_j} \zeta_{jt}^{p_j}}{p_j!} \\ &= \sum_{j=1}^{J} \theta_{ij} f_{jt} + U_{it}, \end{aligned} \tag{5.7}$$

where

$$f_{jt} = \zeta_{jt}^p, \tag{5.8}$$

$$\theta_{ij} = \beta_{ij}^p, \tag{5.9}$$

$$U_{it} = p! \sum_{\substack{p_j : 0 \le p_j < p \\ \Sigma p_j \le p, j \in P_J}} \frac{\delta_{it}^{p - \Sigma p_j}}{(p - \Sigma p_j)!} \prod_{j=1}^{J} \frac{\beta_{ij}^{p_j} \zeta_{jt}^{p_j}}{p_j!}, \tag{5.10}$$

and write (5.7) in vector notation,

$$Z = \theta f + U. \tag{5.11}$$

In this transformed form, Z is the new $q \times n$ matrix of observables, f is the $J \times n$ matrix of the latent factors, $\theta = (\theta_1, \ldots, \theta_J)$ is the $q \times J$ matrix of (transformed) factor loadings, and U is a $q \times n$ matrix of errors. The transformed Z, θ, f are simply the (element-by-element) p-th powers of z_t, β, ζ_t in (5.2), while elements of U are expressed as polynomials in elements of δ_t and ζ_t. This transformation converts specification (5.1), which contains only white noise components by Assumption 5.1, to one where the observables display long memory and cointegrating relationships are present. Indeed, (5.6) and Theorem 3.1 ensure that all components in the f matrix display the full memory of the underlying Gaussian $I(d_j)$ process, ν_{jt}.

Assumption 5.2 *θ is full rank.*

Assumption 5.2 is a minimal requirement for identification of θ in the above system. The presence of linear dependence in a group of factors would imply that one of them (the one with the lowest d_j) could be "absorbed" by the remaining ones without affecting the memory of the errors U_{it} or of the remaining factors.

Recall the averaged periodogram statistic introduced in (2.51),

$$\hat{F}_{ab}(\lambda_m) = \frac{2\pi}{n} \sum_{j=1}^{m} I_{ab}(\lambda_j), \tag{5.12}$$

We will omit the dependency on the bandwidth m throughout the chapter for brevity, and write $\hat{F}_{ab} = \hat{F}_{ab}(\lambda_m)$. We impose the same conditions on m as in Chapter 3, and indeed Assumption 3.2 is reproduced here as Assumption 5.3 for convenience.

Assumption 5.3 *The bandwidth sequence $m = m(n)$ satisfies*

$$\frac{1}{m} + \left(\frac{m}{n}\right)^{\epsilon} \log n \to 0 \text{ as } n \to \infty, \tag{5.13}$$

for all $\epsilon > 0$.

Using (5.7), we may write

$$Z_{i_1 t_1} Z_{i_2 t_2} = \sum_{j_1, j_2 = 1}^{J} \theta_{i_1 j_1} \theta_{i_2 j_2} f_{j_1 t_1} f_{j_2 t_2} + \sum_{j=1}^{J} (\theta_{i_1 j} f_{j t_1} U_{i_2 t_2} + \theta_{i_2 j} U_{i_1 t_1} f_{j t_2}) + U_{i_1 t_1} U_{i_2 t_2}.$$

By linearity of the Fourier transform,

$$\hat{F}_{Z_{i_1}, Z_{i_2}} = \sum_{j_1, j_2 = 1}^{J} \theta_{i_1 j_1} \theta_{i_2 j_2} \hat{F}_{f_{j_1}, f_{j_2}} + \sum_{j=1}^{J} (\theta_{i_1 j} \hat{F}_{f_j, U_{i_2}} + \theta_{i_2 j} \hat{F}_{U_{i_1}, f_j}) + \hat{F}_{U_{i_1}, U_{i_2}},$$

or in matrix notation,

$$\hat{F}_{ZZ} = \theta \hat{F}_{ff} \theta' + \hat{F}_{Uf} \theta' + \theta \hat{F}_{fU} + \hat{F}_{UU}, \tag{5.14}$$

where $\hat{F}_{ZZ}$ is $q \times q$, θ is $q \times J$, $\hat{F}_{ff}$ is $J \times J$, $\hat{F}_{Uf}$ is $q \times J$, $\hat{F}_{fU}$ is $J \times q$, and $\hat{F}_{UU}$ is $q \times q$. As long as the first term in the right-hand side retains rank J and dominates the others, the first J eigenvectors of $\hat{F}_{ZZ}$ will converge to a linear transformation of θ. We state the asymptotic properties of (5.14) as Theorem 5.1. Dominance of $\hat{F}_{UU}$ by $\hat{F}_{ff}$ is guaranteed by (5.5), but dominance of the cross-terms $\hat{F}_{Uf}, \hat{F}_{fU}$ by $\hat{F}_{ff}$ requires the following additional assumption on memory parameters.

Assumption 5.4

$$d_1 - d_J \leq d_J - \max_{1 \leq i \leq q} b_i',$$

where b_i' is the memory parameter of U_i (as defined in Proposition 5.2, Appendix 5.A).

Theorem 5.1 *Under Assumptions 5.1 through 5.4,*

$$\hat{F}_{ZZ} = \theta \hat{F}_{ff} \theta' (1 + o_p(1)), \tag{5.15}$$

as $n \to \infty$, *and*

$$\Lambda\left(\tfrac{m}{n}\right) \hat{F}_{ff} \Lambda\left(\tfrac{m}{n}\right) \to_p \Omega, \tag{5.16}$$

where

$$\Lambda(\lambda) = \operatorname{diag}\{\lambda^{d_1 - \frac{1}{2}}, \ldots, \lambda^{d_J - \frac{1}{2}}\}, \tag{5.17}$$

$$\Omega = \operatorname{diag}\{\omega_1, \ldots, \omega_J\}, \tag{5.18}$$

for positive constants ω_j, $j = 1, \ldots, J$.

Proof. Proposition 5.1 establishes the asymptotic behaviour of $\hat{F}_{ff}$ summarised in (5.16). Analogously, Proposition 5.2 provides an asymptotic bound for diagonal elements of $\hat{F}_{UU}$. The remaining elements of $\hat{F}_{UU}$, and the cross terms in $\hat{F}_{Uf}$ and $\hat{F}_{fU}$, are negligible by Assumption 5.4 and the Cauchy inequality. □

Theorem 5.1 establishes asymptotic properties of the $\hat{F}_{ZZ}$ matrix that allow us to establish limits for the relevant eigenvalues and eigenvectors. Furthermore, since $\hat{F}_{ff}$ converges to a diagonal matrix, its imaginary part vanishes asymptotically, so $\operatorname{Re}(\hat{F}_{ZZ})$ and $\operatorname{Re}(\hat{F}_{ff})$ also satisfy (5.15) and (5.16), respectively.

5.4 Narrow Band Principal Components estimation of factor loadings

Denote by $\hat{\gamma}_1 \geq \ldots \geq \hat{\gamma}_q \geq 0$ the ordered eigenvalues of $\mathrm{Re}(\hat{F}_{ZZ})$, i.e. the roots of the characteristic equation

$$|\,\mathrm{Re}(\hat{F}_{ZZ}) - \gamma I_q| = 0, \tag{5.19}$$

where I_q is the $q \times q$ identity matrix. For each $\hat{\gamma}_j$, define the associated eigenvector $\hat{c}_j$ as a non-zero solution to

$$\{\mathrm{Re}(\hat{F}_{ZZ}) - \hat{\gamma}_j I_q\}c = 0, \tag{5.20}$$

normalised to satisfy $\hat{c}_j'\hat{c}_j = 1$. For any $\hat{\gamma}_j \neq \hat{\gamma}_k$, it is known that $\hat{c}_j'\hat{c}_k = 0$. If some of the eigenvalues have multiplicity greater than one, choose corresponding eigenvectors that constitute an orthonormal base of the space of solutions, so that $\hat{c}_j'\hat{c}_k = 1(j = k)$ holds for all $1 \leq j, k \leq q$. The NBPC estimate of the space spanned by θ is

$$\hat{c} = (\hat{c}_1, \ldots, \hat{c}_J), \tag{5.21}$$

where again we omit the dependence on m for brevity.

In order to allow for full generality in the memory parameters of the factors, we need to introduce some notation grouping factors with the same memory into $l \leq J$ partitions. Define $k_0 < \ldots < k_l$ and $d_{(1)} > \ldots > d_{(l)}$, such that $k_0 = 0$, $k_l = J$ and for $i = 1, \ldots, l$

$$d_{(i)} = d_{k_{i-1}+1} = d_{k_i}. \tag{5.22}$$

Partition the limiting matrix Ω in Theorem 5.1 into l diagonal blocks, $\Omega =$

$\text{diag}\{\Omega_{(1)}, \ldots, \Omega_{(l)}\}$ where

$$\Omega_{(i)} = \text{diag}\{\omega_{k_{i-1}+1}, \ldots, \omega_{k_i}\}. \tag{5.23}$$

Similarly, let $\theta = (\theta_{(1)}, \ldots, \theta_{(l)})$ where

$$\theta_{(i)} = (\theta_{k_{i-1}+1}, \ldots, \theta_{k_i}), \tag{5.24}$$

and $\hat{c} = (\hat{c}_{(1)}, \ldots, \hat{c}_{(l)})$ where

$$\hat{c}_{(i)} = (\hat{c}_{k_{i-1}+1}, \ldots, \hat{c}_{k_i}). \tag{5.25}$$

For $i = 2, \ldots, l$, let

$$\theta_{*(i)} = (\theta_1, \ldots, \theta_{k_{i-1}}), \tag{5.26}$$

and define the orthogonal projections

$$P_{(i)} = \theta_{*(i)}(\theta'_{*(i)}\theta_{*(i)})^{-1}\theta'_{*(i)}\theta_{(i)}, \tag{5.27}$$

which are by construction linear combinations of the columns of $\theta_{*(i)}$. For notational simplicity, we will use the convention that $P_{(1)} = 0$, and we will also write $\Delta k_i = k_i - k_{i-1}$.

To link the properties established in 5.1 to the asymptotic behaviour of the eigenvalues and eigenvectors of $\hat{F}_{ZZ}$ (and thus of $\text{Re}(\hat{F}_{ZZ})$), we introduce an additional assumption, which excludes the possibility of limiting eigenvalues with multiplicities greater than one within each group of factors with the same memory. Although asymptotic theory for principal components may still be developed when the limiting eigenvalues are equal in sets (see Anderson, 1963), these circumstances are unlikely to occur in this context, and as such the additional generality

would add little value.

Assumption 5.5 *For all $1 \leq i \leq l$, the Δk_i positive eigenvalues of*

$$(\theta_{(i)} - P_{(i)})\Omega_{(i)}(\theta_{(i)} - P_{(i)})'$$

are all distinct.

Theorem 5.2 *Let Assumptions 5.1 through 5.5 hold. Then, for $j > J$,*

$$\left(\frac{m}{n}\right)^{2d_J - 1} \hat{\gamma}_j \to_p 0, \qquad as\ n \to \infty, \tag{5.28}$$

while for $1 \leq i \leq l$ and $k_{i-1} < j \leq k_i$, there exist full rank $\Delta k_i \times \Delta k_i$ matrices $\Phi_{(i)}$ and sign sequences $s_{jn} \in \{-1, 1\}$, such that:

(i)

$$\hat{\gamma}_j \sim r_j \left(\frac{m}{n}\right)^{1-2d_j}, \qquad as\ n \to \infty, \tag{5.29}$$

where r_j is the $(j - k_i)$-th diagonal element of $\Phi_{(i)}\Omega_{(i)}\Phi'_{(i)}$;

(ii)

$$s_{jn}\hat{c}_j \to_p c_j, \qquad as\ n \to \infty, \tag{5.30}$$

where

$$c_{(i)} = (c_{k_{i-1}+1}, \ldots, c_{k_i}) = (\theta_{(i)} - P_{(i)})\Phi_{(i)}^{-1}, \tag{5.31}$$

and therefore each c_j is a linear combination of $\theta_{*(i+1)}$ orthogonal to $\theta_{*(i)}$.

Proof. Write

$$A_{in} = \left(\frac{m}{n}\right)^{2d_{(i)}-1} \mathrm{Re}(\hat{F}_{ZZ}).$$

From Theorem 5.1,

$$A_{1n} = \theta_{(1)}\Omega_{(1)}\theta'_{(1)}(1 + o_p(1)).$$

It follows that A_{1n} converges in probability to $\theta_{(1)}\Omega_{(1)}\theta'_{(1)}$, which is a positive semidefinite matrix of rank k_1. By Slutsky's Theorem, the first k_1 eigenvalues of A_{1n} converge in probability to the positive eigenvalues of $\theta_{(1)}\Omega_{(1)}\theta'_{(1)}$, while the remaining $J - k_1$ are $o_p(1)$. Therefore, the first k_1 eigenvalues of $\mathrm{Re}(\hat{F}_{ZZ})$ are $O_e((\frac{m}{n})^{1-2d_1})$, while the remaining ones are $o_p((\frac{m}{n})^{1-2d_1})$. The k_1 eigenvectors of $\theta_{(1)}\Omega_{(1)}\theta'_{(1)}$ associated with non-zero eigenvalues span the same subspace as $\theta_{(1)}$; they can thus be written as $\theta_{(1)}\Phi_{(1)}^{-1}$, for some full rank $k_1 \times k_1$ matrix $\Phi_{(1)}$, while the corresponding eigenvalues are the elements of the diagonal matrix $\Phi_{(1)}\Omega_{(1)}\Phi'_{(1)}$. Since these eigenvalues are assumed all distinct, the transformation $\Phi_{(1)}$ is essentially unique (up to sign), and the first k_1 eigenvectors of A_{1n} converge, in the sense of (5.30), to the respective columns of $\theta_{(1)}\Phi_{(1)}^{-1}$, yielding (5.29) and (5.30), (5.31) for $i = 1$.

Now, proceeding by induction, fix $1 < i \leq l$, assume that (5.29)-(5.31) hold for $1, \ldots, i-1$, and define the $k_{i-1} \times k_{i-1}$ matrices

$$\begin{aligned} \Omega_{*(i)} &= \mathrm{diag}\{\omega_1, \ldots, \omega_{k_{i-1}}\}, \\ B_{in} &= \mathrm{diag}\left\{\left(\frac{m}{n}\right)^{d_{(i)}-d_1}, \ldots, \left(\frac{m}{n}\right)^{d_{(i)}-d_{k_i-1}}\right\}, \end{aligned}$$

where $d_{(i)} - d_{k_i-1} < 0$ by construction. Theorem 5.1 implies that

$$\begin{aligned} A_{in} &= (\theta_{*(i)}B_{in}\Omega_{*(i)}B_{in}\theta'_{*(i)} + \theta_{(i)}\Omega_{(i)}\theta'_{(i)})(1 + o_p(1)) \\ &= \{\theta_{*(i)}B_{in}\Omega_{*(i)}B_{in}\theta'_{*(i)} + P_{(i)}\Omega_{(i)}P'_{(i)} \\ &\quad + (\theta_{(i)} - P_{(i)})\Omega_{(i)}(\theta_{(i)} - P_{(i)})'\}(1 + o_p(1)), \end{aligned}$$

since $P_{(i)}$ and $\theta_{(i)} - P_{(i)}$ are orthogonal. By assumption, the first k_{i-1} eigenvectors of A_{in} converge in probability to a linear combination of $\theta_{*(i)}$, implying that the remaining limiting eigenvectors must be orthogonal to $\theta_{*(i)}$ (and thus to $P_{(i)}$).

For $k_{i-1} < j \leq k_i$ the j-th eigenvalue of A_{in} converges in probability to the $(j - k_{i-1})$-th eigenvalue of $(\theta_{(i)} - P_{(i)})\Omega_{(i)}(\theta_{(i)} - P_{(i)})'$, while the remaining $J - k_i$ are $o_p(1)$. This shows that the j-th eigenvalue of $\mathrm{Re}(\hat{F}_{ZZ})$ is $O_e((\frac{m}{n})^{1-2d_{(i)}})$ for $k_{i-1} < j \leq k_i$, and $o_p((\frac{m}{n})^{1-2d_{(i)}})$ for $j > k_i$. As before, the first Δk_i eigenvectors of $(\theta_{(i)} - P_{(i)})\Omega_{(i)}(\theta_{(i)} - P_{(i)})'$ span the same subspace as $(\theta_{(i)} - P_{(i)})$, and can thus be written as $(\theta_{(i)} - P_{(i)})\Phi_{(i)}^{-1}$, for a full rank $\Delta k_i \times \Delta k_i$ matrix $\Phi_{(i)}$, while the associated eigenvalues are the elements of the diagonal matrix $\Phi_{(i)}\Omega_{(i)}\Phi'_{(i)}$. Since all eigenvectors of $(\theta_{(i)} - P_{(i)})\Omega_{(i)}(\theta_{(i)} - P_{(i)})$ are orthogonal to $\theta_{*(i)}$ by construction, $\hat{c}_{(i)}$ converges (up to sign) to $(\theta_{(i)} - P_{(i)})\Phi_{(i)}^{-1}$, yielding (5.29) and (5.30), (5.31) for $i = 2, \ldots, l$; (5.28) follows when $i = l$. □

Some of the indeterminacy present in $\hat{c}_1$ may be resolved if $d_1 > d_2$. In this case, $\Phi_{(1)} = \|\theta_1\|$ (where $\|\cdot\|$ denotes the Euclidean norm) and $\hat{c}_1$ will be consistent for θ_1, up to arbitrary scale and sign. This mimics the indeterminacy found in Theorem 3.2, avoided therein by focusing on estimation of a ratio. Note, however, that subsequent principal components are still affected by more serious indeterminacy, irrespectively of the relationships between memory parameters, since θ cannot be reasonably assumed to be an orthogonal matrix. For even p, all elements of θ are non-negative, which rules out orthogonality if any of the observables is exposed to more than one factor. As a result, even when $d_{(i)}$ is unique, and therefore $\Phi_{(i)}$ is a scalar, $P_{(i)}$ cannot be assumed to vanish from (5.31). Still, when the memory parameters of all factors, $d_1, \ldots, d_J$, are distinct, Theorem 5.2 can be substantially simplified. We state this result as a Corollary, whose proof follows from $k_i = i$, and the fact that our chosen eigenvector normalisation requires $|\Phi_{(j)}| = \|\theta_j - P_{(j)}\|$.

Corollary 5.1 *Let Assumptions 5.1 through 5.4 hold, and furthermore let* $d_1 > \ldots > d_J$. *Then, for* $1 \leq j \leq J$, *there exist sign sequences* $s_{jn} \in \{-1, 1\}$, *such that*

$$\hat{\gamma}_j \sim \|\theta_j - P_{(j)}\|^2 \omega_j \left(\frac{m}{n}\right)^{2d_j - 1}, \qquad \text{as } n \to \infty,$$

and

$$s_{jn}\hat{c}_j \to_p \frac{\theta_j - P_{(j)}}{\|\theta_j - P_{(j)}\|}, \qquad as\ n \to \infty.$$

Theorem 5.2 indicates that, in general, $\hat{c}$ does not converge to θ, but the probability limit of $\hat{c}$ spans the space generated by θ (in Corollary 5.1, $\hat{c}$ converges to the Gram–Schmidt orthonormalisation of θ). This would suffice if, say, one was interested in the cointegrating vectors as exemplified in Section 2.3. In this case, for any $1 \leq i \leq l$, all vectors of coefficients, α, in the null space of the matrix $(c_1, \ldots, c_{k_i})$ will eliminate the k_i leading factors, and lead to linear combinations αZ_t with memory at most $d_{(i+1)}$. Still, in many applications the interest is on the factor loadings themselves, as is the case in asset pricing models, and it may be desirable to identify θ via suitable linear constraints, as in Anderson and Rubin (1956).

Assume that $d_1 > \ldots > d_J$, so that Corollary 5.1 holds, and denote the probability limit of $\hat{c}$ by $c = (c_1, \ldots, c_j)$ (where we omit the sign factor in each $\hat{c}_j$ for notational simplicity). We know from Corollary 5.1 that each $\hat{c}_j$ converges to a linear combination of $\theta_1, \ldots, \theta_j$, where the weight on θ_j is nonzero; therefore, there exists a full rank, upper triangular $J \times J$ matrix Φ such that $\theta = c\Phi$, or

$$\theta_j = \sum_{k=1}^{j} \phi_{kj} c_k,$$

for $j = 1, \ldots, J$. The $J(J+1)/2$ unknown elements of the matrix Φ are free, but identification for θ may be achieved by imposing enough restrictions on θ to determine Φ. The simplest (and perhaps most useful) structure that can be conceived to identify θ relies on $J(J-1)/2$ exclusion restrictions and J normalisation restrictions. Let Q denote a $q \times J$ submatrix of a $q \times q$ permutation matrix, i.e. for distinct indices $i_1, \ldots, i_J$ satisfying $1 \leq i_j \leq q$, define $Q = (\iota_{i_1}, \ldots, \iota_{i_J})$, where ι_i is the i-th column of I_q. Assume that $Q'\theta$ (a certain subset of the rows of θ)

has the lower triangular form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ * & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ * & * & \cdots & 1 & 0 \\ * & * & \cdots & * & 1 \end{bmatrix},$$

where the elements denoted by $*$ are unrestricted. Formally, θ satisfies the restrictions

$$\theta_{i_j j} = 1, \qquad \text{for } j = 1, \ldots, J, \tag{5.32}$$

$$\theta_{i_j k} = 0, \qquad \text{for } 1 \leq j < k \leq J. \tag{5.33}$$

In this structure, $z_{i_1 t}$ is assumed to have nonzero exposure to ζ_{1t} (where this exposure is normalised to unity) and zero exposures to ζ_{jt}, $j = 2, \ldots, J$; $z_{i_2 t}$ is assumed to have nonzero exposure to ζ_{2t} (again, normalised to unity) and zero exposures to ζ_{jt}, $j = 3, \ldots, J$; and so forth up to $z_{i_J t}$, whose exposure to ζ_{Jt} is assumed nonzero and normalised to unity.

Suppose we observe c, and wish to recover the coefficients Φ such that $\theta = c\Phi$ satisfies these restrictions. Since $\theta_1 = \phi_{11} c_1$, the restriction $\theta_{i_1 1} = 1$ will be uniquely satisfied by $\phi_{11} = c_{i_1 1}^{-1}$, where $c_{i_1 1}$ is nonzero by assumption. For $1 < j \leq J$, define the $q \times j$ submatrices $\theta_* = (\theta_1, \ldots, \theta_j)$, $Q_* = (\iota_{i_1}, \ldots, \iota_{i_j})$, $c_* = (c_1, \ldots, c_j)$, and the $j \times 1$ column vectors $\Phi_* = (\phi_{1j}, \ldots, \phi_{jj})'$, $\iota_* = (0, \ldots, 0, 1)'$. We may write

$$Q_*' \theta_* \iota_* = Q_*' c_* \Phi_* = \iota_*,$$

where the $j \times j$ matrix $Q_*' c_*$ can be expressed as a product of the square submatrices of $Q'\theta$ and Φ^{-1} corresponding to the first j rows and columns. Since both $Q'\theta$ and

Φ^{-1} are full rank triangular matrices, $Q_*'c_*$ is invertible, and these j restrictions yield a unique solution for Φ_*, namely the j-th column of $(Q_*'c_*)^{-1}$. Constructing $\hat{\Phi}$ in this manner from the estimates $\hat{c}$, it follows from Corollary 5.1 and Slutsky's Theorem that $\hat{\theta} = \hat{c}\hat{\Phi} \to_p \theta$. Note, however, that in finite samples nothing prevents this method from yielding negative values for some of the $\hat{\theta}_{ij}$, which would be nonsensical for even p.

We conclude this section with a brief discussion of the choice of J. The following Theorem adapts the model selection procedure of Robinson and Yajima (2002) and Chen and Hurvich (2003) to our setting, enabling consistent estimation of J if the memory parameters of factors and errors are known. Define

$$L(u) = v_n u - \sum_{i=1}^{u} \hat{\delta}_i, \tag{5.34}$$

where $v_n > 0$ is a deterministic sequence, and estimate J by

$$\hat{J} = \arg \min_{1 \le u < p} L(u). \tag{5.35}$$

Assumption 5.6 *For some $K > 0$ and $0 < d_v < \frac{1}{2}$,*

$$v_n \sim K \left(\frac{m}{n}\right)^{2d_v - 1}, \qquad \text{as } n \to \infty.$$

Theorem 5.3 *Under assumptions 5.1 through 5.6, if $d_i > d_v > d_{i+1}$,*

$$P(\hat{J} = i) \to 1, \qquad \text{as } n \to \infty,$$

while if $d_J > d_v > (d_1 + \max b_i')/2$,

$$P(\hat{J} = J) \to 1, \qquad \text{as } n \to \infty.$$

Proof. We follow the proof of Theorem 4 of Robinson and Yajima (2002), and Theorem 3 of Chen and Hurvich (2003). We have

$$P(\hat{J} > i) \leq \sum_{u=i+1}^{p-1} P\{L(u) < L(i)\}$$
$$\leq pP(\hat{\delta}_{i+1} > v_n),$$

while on the other hand

$$P(\hat{J} < i) \leq \sum_{u=1}^{i-1} P\{L(u) < L(i)\}$$
$$\leq iP(\hat{\delta}_i < v_n).$$

Noting (5.29) and Assumptions 5.3, 5.6, if $d_i > d_v > d_{i+1}$ then $P(\hat{J} \neq i) \to 0$ as $n \to \infty$. If $d_J > d_v$, $P(\hat{J} < i) \to 0$ by (5.29). Theorem 5.2 does not provide an exact order of magnitude for $\hat{\gamma}_{J+1}$; note, however, that by Propositions 5.1, 5.2 and the Cauchy inequality, each element of $\theta'\hat{F}_{Uf} + \theta\hat{F}_{fU} + \hat{F}_{UU}$ is bounded by $K\left(\frac{m}{n}\right)^{1-\max d_j - \max b'_i}$. Theorem 5.2 ensures that the first J eigenvectors of $\hat{F}_{ZZ}$ converge to a base of the space spanned by θ, and therefore the remaining eigenvectors are asymptotically orthogonal to θ. From (5.14), the eigenvalues associated with the last $(q - J)$ eigenvectors of $\hat{F}_{ZZ}$ are bounded by the first eigenvalue of $\theta'\hat{F}_{Uf} + \theta\hat{F}_{fU} + \hat{F}_{UU}$. Thus, if $d_v > (d_1 + \max b'_i)/2$, $P(\hat{\delta}_{J+1} > v_n) \to 0$, completing the proof. □

Since the d_j and b'_i are typically unknown, the usefulness of Theorem 5.3 to estimate J in practice is limited by the sensitivity to d_v and K. Still, $\hat{J}$ may be better interpreted as an estimate for the number of factors with memory above d_v. Under this interpretation, a researcher may not be concerned with whether the cointegrating residuals admit a further dimensionality reduction if their persistence is already below the user-specified threshold d_v. In particular, d_1 may be readily

estimated from the data using the methods described in Section 2.7 (e.g. using the sequential testing procedure of Robinson and Yajima, 2002), and d_v may be set to a fraction of this estimate. Estimation of $d_2, \ldots, d_J$ is a more complex problem and beyond the scope of this thesis.

5.5 Finite sample properties

We now present a Monte Carlo study of finite-sample performance. Data are generated from (5.1)-(5.4), where for $i = 1, \ldots, q$ and $j = 1, \ldots, J$, we set $g_j(x) = h_i(x) = \exp(x)$, $\{\eta_{jt}\}$ and $\{\chi_{it}\}$ as independent standard normal processes, $\{\nu_{jt}\}$ as Gaussian ARFIMA$(0, d_j, 0)$ with $\mathrm{Var}(\nu_{jt}) = 4$, and the vector process $\{\xi_t\}$ as jointly iid $N(0, \Sigma)$. All innovations are mutually independent, except for time dependence in the factor volatilities $\{\nu_{jt}\}$ and contemporaneous dependence between elements of ξ_t. In each setup we set q, J, $d = (d_1, \ldots, d_J)$, and report results for $n = 512$, 1024, 2048 and the following four cases of Σ: homoscedastic and independent, heteroscedastic and independent, heteroscedastic and positively correlated, heteroscedastic and negatively correlated. We compute two estimates: the traditional principal components method, using the J eigenvectors of $z'z$ associated with the largest eigenvalues as an estimate for the space generated by β; and the NBPC method for squared data, using the J eigenvectors of $\mathrm{Re}(\hat{F}_{ZZ})$ associated with the largest eigenvalues as an estimate for the space generated by θ.

For both estimates, we report the average proportion of remaining variance (or its narrow band equivalent) explained by each eigenvector. We denote by R_j the ratio of $\hat{\gamma}_j$ and the sum of the smallest j eigenvalues,

$$R_j = \frac{\hat{\gamma}_j}{\sum_{k=j}^{q} \hat{\gamma}_j}, \tag{5.36}$$

which indicates the relative importance of each explanatory factor.

There are two problems in comparing these estimates. Since they are only estimating a base for the corresponding linear space, the accuracy of a given estimate cannot be evaluated simply with bias and variance. We address this issue by reporting D_j, defined as the Euclidean distance between $\hat{c}_j$ and the true parameter space, i.e.

$$D_j = \left\| (I_q - \theta(\theta'\theta)^{-1}\theta')\hat{c}_j \right\|^2 . \tag{5.37}$$

Although Corollary 5.1 would allow us to define a more demanding convergence measure, namely the distance between $\hat{c}_j$ and the space spanned by $(\theta_1, \ldots, \theta_j)$, such a criterion could not be used for the full-band estimate. The second problem arises from the different parameter matrices being estimated. In general, and without using additional restrictions as described in the previous section, the two estimates only aim to provide a base for the linear subspace spanned by β or θ. However, due to the nonlinear nature of the transformation, a vector within the space spanned by β will not be transformed, in general, to a vector within the space spanned by θ. Still, the chosen normalisation bounds D_j between zero and one, and we choose values for β such that the structures of β and θ are similar, so comparing the estimates based on these distances should still be meaningful.

Monte Carlo standard deviation of both R_j and D_j are also presented in parentheses. NBPC is evaluated at the optimal bandwidth m^*, chosen to minimise the average distance between the last relevant eigenvector, $\hat{c}_J$, and the true parameter space.

Tables 5.1 and 5.2 report R_1, D_1, and m^* for $q = 2$, $J = 1$, $d_1 = 0.4$, $\beta = \theta = (1, 1)'$, and four covariance structures for ξ_t:

1. Independent and homoscedastic components, i.e. $\Sigma = I_2$;

2. Independent components ξ_{it} with variances $\Sigma = \text{diag}\{1, 4\}$;

1. Homoscedastic errors

	n	512	1024	2048
R_1	Lev	0.928	0.946	0.959
		(0.098)	(0.084)	(0.069)
	NB	0.971	0.982	0.990
		(0.070)	(0.054)	(0.039)
D_1	Lev	0.033	0.019	0.009
		(0.087)	(0.060)	(0.024)
	NB	0.057	0.031	0.015
		(0.145)	(0.100)	(0.055)
m^*	NB	33	26	30

2. Heteroscedastic errors

	n	512	1024	2048
R_1	Lev	0.909	0.925	0.938
		(0.089)	(0.082)	(0.072)
	NB	0.963	0.973	0.984
		(0.066)	(0.057)	(0.044)
D_1	Lev	0.125	0.095	0.073
		(0.180)	(0.151)	(0.127)
	NB	0.112	0.073	0.046
		(0.208)	(0.165)	(0.121)
m^*	NB	9	8	6

Table 5.1: Monte Carlo average eigenvalue ratio (R_j), average distance to parameter subspace (D_j), and optimal NBPC bandwidth (m^*) of PC in levels and NBPC in squares, for varying n and Σ; $q = 2$, $J = 1$. Monte Carlo standard deviation in parentheses.

3. Same variances as above, with $\text{Corr}(\xi_{1t}, \xi_{2t}) = 0.5$;

4. Same variances as above, with $\text{Corr}(\xi_{1t}, \xi_{2t}) = -0.5$.

In all cases R_1 is higher for the narrow band approach than for the levels. It is fairly close to one for both estimates, ranging from 0.896 to 0.959 in the levels, and from 0.961 to 0.990 in the squares. For both estimates, this ratio gives a clear indication of the relative importance of the two components, and approaches one as the sample size increases.

For covariance structure 1, the estimate in levels is on average closer to the true parameter space than NBPC. This is a consequence of the particular covariance structure chosen, for which both estimates are consistent. To see this, write the

model in levels in matrix notation, $z = \beta\zeta + \delta$, yielding

$$\begin{aligned}\frac{z'z}{n} &= \beta\frac{\zeta\zeta'}{n}\beta' + \beta\frac{\zeta\delta'}{n} + \frac{\delta\zeta'}{n}\beta' + \frac{\delta\delta'}{n} \\ &\rightarrow_p \beta\,\mathrm{Var}(\zeta_t)\beta' + \mathrm{Var}(\delta_t),\end{aligned}$$

since ζ_t, δ_t are stationary, with all moments finite, and mutually independent. Assuming all eigenvalues are distinct, the eigenvectors of $z'z$ will converge (up to sign) to the eigenvectors of $\beta\,\mathrm{Var}(\zeta_t)\beta' + \mathrm{Var}(\delta_t)$, while β and the eigenvectors of $\beta\,\mathrm{Var}(\zeta_t)\beta'$ will span the same subspace. Principal components is still consistent if $\mathrm{Var}(\delta_t)$ is a scalar matrix, as both sets of eigenvectors will coincide. The empirical findings illustrate this fact; both estimates are approaching the true parameter space as n increases.

With the introduction of heteroscedasticity, $\mathrm{Var}(\delta_t)$ is a non-scalar matrix, and principal components in the levels is no longer consistent. The reported distances are higher for the levels than for the narrow band estimate, and NBPC distances seem to decrease at a much faster rate.

The introduction of correlation between the disturbance terms has almost no impact on NBPC. The average distances for all sample sizes are extremely similar for structures 2, 3, and 4, as are their Monte Carlo standard deviations. For the estimate in levels, matters are considerably different: positive correlation improves estimation, while negative correlation worsens it. This effect should depend on how "similar" the matrices $\beta\,\mathrm{Var}(\zeta_t)\beta'$ and $\mathrm{Var}(\delta_t)$ are, which may influence how close the limiting eigenvectors of $z'z$ are to the true parameters.

The optimal bandwidths, m^*, seem to be insensitive to sample size. Although all values are quite small, ranging from 6 to 33, they are mostly affected by the presence of heteroscedasticity. The presence and direction of correlation have little impact on bandwidth.

3. Positive correlation

	n	512	1024	2048
R_1	Lev	0.923	0.936	0.947
		(0.075)	(0.071)	(0.062)
	NB	0.963	0.974	0.983
		(0.066)	(0.058)	(0.046)
D_1	Lev	0.112	0.084	0.066
		(0.156)	(0.128)	(0.110)
	NB	0.110	0.070	0.043
		(0.207)	(0.162)	(0.121)
m^*	NB	10	9	11

4. Negative correlation

	n	512	1024	2048
R_1	Lev	0.896	0.914	0.929
		(0.100)	(0.092)	(0.082)
	NB	0.961	0.972	0.983
		(0.069)	(0.061)	(0.042)
D_1	Lev	0.145	0.108	0.080
		(0.213)	(0.177)	(0.145)
	NB	0.114	0.073	0.045
		(0.207)	(0.164)	(0.126)
m^*	NB	11	13	15

Table 5.2: Monte Carlo average eigenvalue ratio (R_j), average distance to parameter subspace (D_j), and optimal NBPC bandwidth (m^*) of PC in levels and NBPC in squares, for varying n and Σ; $q = 2$, $J = 1$. Monte Carlo standard deviation in parentheses.

Tables 5.3 and 5.4 report R_j, D_j, and m^* for $j = 1, 2, 3$, corresponding to the three largest eigenvalues (of which only the first two should be relevant), $q = 5$, $J = 2$, $d = (0.4, 0.3)$,

$$\beta = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ \frac{1}{3} & \frac{2}{3} & 1 & \frac{4}{3} & \frac{5}{3} \end{bmatrix}', \qquad \theta = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ \frac{1}{9} & \frac{4}{9} & 1 & \frac{16}{9} & \frac{25}{9} \end{bmatrix}',$$

and four covariance structures for ξ_t:

1. Independent and homoscedastic components, i.e. $\Sigma = I_5$;

2. Independent components ξ_{it} with variances $\Sigma = \text{diag}\{\frac{1}{4}, \frac{1}{2}, 1, 2, 4\}$;

3. Same variances as above, with $\text{Corr}(\xi_{it}, \xi_{jt}) = 0.5$ for $1 \leq i < j \leq 5$;

1. Homoscedastic errors

	n	512			1024			2048		
	j	3	2	1	3	2	1	3	2	1
R_j	Lev	0.428	0.735	0.966	0.403	0.791	0.970	0.388	0.829	0.972
		(0.057)	(0.200)	(0.028)	(0.041)	(0.181)	(0.021)	(0.034)	(0.163)	(0.018)
	NB	0.651	0.905	0.974	0.633	0.931	0.974	0.621	0.955	0.976
		(0.134)	(0.145)	(0.039)	(0.129)	(0.132)	(0.036)	(0.134)	(0.105)	(0.034)
D_j	Lev	0.982	0.103	0.003	0.991	0.054	0.002	0.998	0.027	0.001
		(0.075)	(0.193)	(0.004)	(0.057)	(0.135)	(0.002)	(0.020)	(0.070)	(0.001)
	NB	0.967	0.163	0.007	0.979	0.098	0.004	0.990	0.055	0.002
		(0.106)	(0.278)	(0.008)	(0.091)	(0.223)	(0.007)	(0.059)	(0.159)	(0.003)
m^*	NB		15			50			87	

2. Heteroscedastic errors

	n	512			1024			2048		
	j	3	2	1	3	2	1	3	2	1
R_j	Lev	0.612	0.731	0.963	0.617	0.773	0.968	0.620	0.805	0.971
		(0.066)	(0.162)	(0.032)	(0.053)	(0.159)	(0.023)	(0.042)	(0.152)	(0.019)
	NB	0.767	0.904	0.973	0.772	0.931	0.973	0.766	0.949	0.975
		(0.119)	(0.126)	(0.040)	(0.117)	(0.115)	(0.038)	(0.113)	(0.101)	(0.034)
D_j	Lev	0.947	0.262	0.005	0.967	0.183	0.003	0.974	0.138	0.002
		(0.099)	(0.308)	(0.008)	(0.071)	(0.256)	(0.006)	(0.069)	(0.221)	(0.002)
	NB	0.951	0.207	0.009	0.967	0.135	0.005	0.979	0.084	0.003
		(0.126)	(0.316)	(0.016)	(0.113)	(0.259)	(0.010)	(0.089)	(0.209)	(0.004)
m^*	NB		15			16			45	

Table 5.3: Monte Carlo average eigenvalue ratio (R_j), average distance to parameter subspace (D_j), and optimal NBPC bandwidth (m^*) of PC in levels and NBPC in squares, for varying n and Σ; $q = 5$, $J = 2$. Monte Carlo standard deviation in parentheses.

4. Same variances as above, with $\text{Corr}(\xi_{it}, \xi_{jt}) = -0.2$ for $1 \leq i < j \leq 5$.

All values of R_j are higher for NB than for the levels. In both cases, the first principal component is clearly dominant, with values of R_1 ranging from 0.962 to 0.972 in levels, and from 0.973 to 0.976 in squares. However, R_2 and R_3 take substantially different values between levels and squares. In each setup, R_2 for the squares goes from around 0.9 for $n = 512$ to around 0.95 for $n = 2048$, while corresponding values for levels are around 0.73 for $n = 512$ and around 0.81 for $n = 2048$. These values indicate that the second principal component is less relevant in levels than in squares. On the other hand, R_3 (which should be low because the model has two factors only) is also higher for squares than for levels. In the homoscedastic case, it takes values in 0.62-0.65 for squares and in 0.39-0.43 for levels. Noting that R_3 is bounded from below by $\frac{1}{3}$, the model in levels clearly indicates the irrelevance of the third component, while it still exhibits some explanatory power in squares. In the three remaining cases, where errors are heteroscedastic, it takes values in 0.76-0.77 for squares and in 0.61-0.62 for levels. Here both estimates display some explanatory power in the third component, possibly making it harder to distinguish the number of factors in situations where it is unknown. Still, it appears that the contributions of the first two factors grow at a reasonably fast rate with sample size, while the irrelevant third factor has a more stable contribution.

For both estimates, the first principal component is extremely close to the subspace generated by the true parameters. Values of D_1 range from 0.005 to 0.001 for the levels, and from 0.009 to 0.002 for NB, decreasing as the sample size increases. Estimation of the second component is understandably less precise. For the case of homoscedastic errors, where the estimate in levels is still consistent, it outperforms NB by a clear margin: D_2 goes from 0.103 to 0.027 as n increases, while for NB D_2 goes from 0.163 to 0.055. However, the presence of heteroscedas-

ticity and autocorrelation inverts the situation. In these three cases, NB always yields lower values for D_2 than the levels, and in addition they appear to decrease at a faster rate with sample size. The presence of positive autocorrelation seems to improve the performance of both estimates, while negative autocorrelation worsens it. The first non-significant component is approximately orthogonal to the parameter space in all cases, as indicated by the high values for D_3 (from 0.947 to 0.998), which become higher with sample size.

Optimal bandwidths for homoscedastic errors tend to increase clearly with n. However, there does not appear to be a clear ordering of bandwidth sizes from positive to negative correlation. In these cases, the increase in bandwidth from $n = 512$ to $n = 1024$ is small or non-existent, while the increase from $n = 1024$ to $n = 2048$ is more substantial. Still, additional results (not shown) indicate that bandwidths close to those presented yield very similar values for D_j.

Tables 5.5 and 5.6 report R_j, D_j, and m^* for $j = 1, 2, 3, 4$, corresponding to the four largest eigenvalues (of which only the first three should be relevant), for $q = 9$, $J = 3$, $d = (0.4, 0.4, 0.3)$,

$$\beta = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & 1 & \frac{6}{5} & \frac{7}{5} & \frac{8}{5} & \frac{9}{5} \\ \frac{1}{5} & \frac{3}{5} & 1 & \frac{7}{5} & \frac{9}{5} & \frac{7}{5} & 1 & \frac{3}{5} & \frac{1}{5} \end{bmatrix}',$$

$$\theta = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \frac{1}{25} & \frac{4}{25} & \frac{9}{25} & \frac{16}{25} & 1 & \frac{36}{25} & \frac{49}{25} & \frac{64}{25} & \frac{81}{25} \\ \frac{1}{25} & \frac{9}{25} & 1 & \frac{49}{25} & \frac{81}{25} & \frac{49}{25} & 1 & \frac{9}{25} & \frac{1}{25} \end{bmatrix}',$$

and four covariance structures for ξ_t:

1. Independent and homoscedastic components, i.e. $\Sigma = I_9$;

3. Positive correlation

	n	512			1024			2048		
	j	3	2	1	3	2	1	3	2	1
R_j	Lev	0.615	0.751	0.966	0.618	0.793	0.969	0.619	0.824	0.972
		(0.070)	(0.163)	(0.028)	(0.058)	(0.157)	(0.021)	(0.045)	(0.149)	(0.018)
	NB	0.764	0.908	0.974	0.768	0.933	0.973	0.762	0.952	0.975
		(0.118)	(0.125)	(0.039)	(0.118)	(0.111)	(0.037)	(0.115)	(0.097)	(0.034)
D_j	Lev	0.950	0.235	0.005	0.972	0.161	0.003	0.979	0.118	0.002
		(0.102)	(0.299)	(0.008)	(0.069)	(0.243)	(0.006)	(0.062)	(0.202)	(0.002)
	NB	0.952	0.200	0.008	0.965	0.132	0.005	0.981	0.077	0.002
		(0.122)	(0.315)	(0.018)	(0.122)	(0.264)	(0.011)	(0.079)	(0.198)	(0.003)
m^*	NB		23			35			37	

4. Negative correlation

	n	512			1024			2048		
	j	3	2	1	3	2	1	3	2	1
R_j	Lev	0.609	0.724	0.962	0.616	0.767	0.967	0.620	0.798	0.970
		(0.064)	(0.161)	(0.033)	(0.053)	(0.158)	(0.024)	(0.042)	(0.153)	(0.019)
	NB	0.766	0.903	0.974	0.771	0.928	0.973	0.769	0.947	0.975
		(0.119)	(0.125)	(0.039)	(0.119)	(0.118)	(0.039)	(0.110)	(0.102)	(0.034)
D_j	Lev	0.946	0.270	0.005	0.965	0.192	0.003	0.972	0.145	0.002
		(0.100)	(0.310)	(0.007)	(0.076)	(0.262)	(0.004)	(0.070)	(0.227)	(0.002)
	NB	0.951	0.211	0.009	0.968	0.137	0.006	0.978	0.085	0.003
		(0.129)	(0.318)	(0.012)	(0.106)	(0.258)	(0.011)	(0.093)	(0.208)	(0.004)
m^*	NB		16			16			51	

Table 5.4: Monte Carlo average eigenvalue ratio (R_j), average distance to parameter subspace (D_j), and optimal NBPC bandwidth (m^*) of PC in levels and NBPC in squares, for varying n and Σ; $q = 5$, $J = 2$. Monte Carlo standard deviation in parentheses.

1. Homoscedastic errors

	n	512				1024				2048			
	j	4	3	2	1	4	3	2	1	4	3	2	1
R_j	Lev	0.251	0.591	0.784	0.937	0.233	0.641	0.803	0.937	0.216	0.706	0.827	0.933
		(0.049)	(0.236)	(0.142)	(0.039)	(0.044)	(0.232)	(0.126)	(0.036)	(0.032)	(0.214)	(0.107)	(0.034)
	NB	0.530	0.769	0.914	0.942	0.543	0.811	0.923	0.938	0.559	0.873	0.935	0.931
		(0.153)	(0.206)	(0.107)	(0.073)	(0.155)	(0.207)	(0.096)	(0.075)	(0.163)	(0.172)	(0.083)	(0.077)
D_j	Lev	0.975	0.148	0.017	0.003	0.986	0.085	0.010	0.002	0.997	0.035	0.005	0.001
		(0.099)	(0.238)	(0.031)	(0.002)	(0.074)	(0.169)	(0.022)	(0.001)	(0.023)	(0.077)	(0.011)	(0.001)
	NB	0.925	0.349	0.040	0.006	0.939	0.259	0.025	0.004	0.955	0.164	0.013	0.003
		(0.174)	(0.389)	(0.090)	(0.007)	(0.168)	(0.357)	(0.059)	(0.005)	(0.152)	(0.304)	(0.039)	(0.004)
m^*	NB	26				31				39			

2. Heteroscedastic errors

	n	512				1024				2048			
	j	4	3	2	1	4	3	2	1	4	3	2	1
R_j	Lev	0.531	0.552	0.729	0.920	0.539	0.570	0.749	0.926	0.550	0.600	0.779	0.925
		(0.081)	(0.138)	(0.139)	(0.057)	(0.073)	(0.147)	(0.131)	(0.043)	(0.065)	(0.159)	(0.113)	(0.038)
	NB	0.667	0.753	0.883	0.937	0.657	0.781	0.902	0.936	0.666	0.817	0.925	0.931
		(0.135)	(0.168)	(0.125)	(0.079)	(0.131)	(0.175)	(0.110)	(0.077)	(0.132)	(0.176)	(0.090)	(0.078)
D_j	Lev	0.840	0.557	0.107	0.005	0.857	0.471	0.075	0.003	0.870	0.376	0.046	0.002
		(0.254)	(0.410)	(0.135)	(0.014)	(0.248)	(0.420)	(0.095)	(0.003)	(0.247)	(0.408)	(0.058)	(0.002)
	NB	0.882	0.508	0.097	0.014	0.904	0.397	0.054	0.006	0.936	0.270	0.025	0.004
		(0.210)	(0.396)	(0.169)	(0.056)	(0.206)	(0.402)	(0.114)	(0.007)	(0.170)	(0.360)	(0.062)	(0.005)
m^*	NB	15				21				31			

Table 5.5: Monte Carlo average eigenvalue ratio (R_j), average distance to parameter subspace (D_j), and optimal NBPC bandwidth (m^*) of PC in levels and NBPC in squares, for varying n and Σ; $q = 9$, $J = 3$. Monte Carlo standard deviation in parentheses.

2. Independent components ξ_{it} with variances

$$\Sigma = \text{diag}\{\tfrac{1}{16}, \tfrac{1}{8}, \tfrac{1}{4}, \tfrac{1}{2}, 1, 2, 4, 8, 16\};$$

3. Same variances as above, with $\text{Corr}(\xi_{it}, \xi_{jt}) = 0.5$ for $1 \leq i < j \leq 9$;

4. Same variances as above, with $\text{Corr}(\xi_{it}, \xi_{jt}) = -0.1$ for $1 \leq i < j \leq 9$.

The simulation results are qualitatively similar to the previous case, though significantly less accurate. Values of R_j are higher for NB than the levels, both in the case of relevant (R_1, R_2, R_3) and irrelevant (R_4) components. The first principal component yields high values of R_1 in both estimates with a relatively small gap between levels and NB. For the two remaining components, the values of R_2 and R_3 display a much larger gap between NB and Levels. In the case homoscedastic errors, the irrelevant component displays little influence on the levels, with R_4 decreasing from 0.251 to 0.216 with n, while still having some impact in NB, with R_4 increasing from 0.530 to 0.559 with n. For heteroscedastic errors, both NB and levels have R_4 increasing with sample size and of moderate size: it ranges from 0.530 to 0.557 for levels, and from 0.656 to 0.684 for NB. The small gap found between R_4 and R_3 in these experiments indicates that estimation of the number of factors could be a difficult task, since the corresponding eigenvalues will be of a similar order of magnitude.

Both parameter estimates are significantly more accurate in the case of homoscedastic errors; in the other cases, positive autocorrelation induces marginally better results than negative one. Naturally, all values of D_j decrease with sample size for relevant components (D_1, D_2, D_3), and increase for irrelevant ones (D_4) The coefficients of the first principal component are extremely close to the parameter space, with the estimate in levels outperforming NB in every case. For the levels, D_1 takes values between 0.001 and 0.005, while for NB it ranges between

0.003 and 0.014. The other two components display substantially different behaviour in the homoscedastic and heteroscedastic cases. In the homoscedastic case, the estimate in levels attains very fast convergence in all factors, clearly outperforming NB. As n increases from 512 to 2048, D_2 for the levels decreases from 0.017 to 0.005, while D_3 decreases from 0.148 to 0.035. Results for NB are much less encouraging, with D_2 decreasing from 0.040 to 0.013 and D_3 decreasing from 0.349 to 0.164. However, the presence of heteroscedasticity, and thus inconsistency of the estimate in levels, makes D_2 and D_3 lower for NB than for the levels, and decreasing at an apparently faster rate. For $n = 2048$, while the estimate in levels can only attain values of D_2 ranging from 0.038 to 0.048 and of D_3 ranging from 0.313 to 0.387, the NB estimate yields D_2 ranging from 0.023 to 0.025 and of D_3 ranging from 0.259 to 0.275. It should be noted that the performance for the third component would not be satisfactory for most purposes; accurate estimation of three components under these conditions would require substantially larger sample sizes. This fact is also reflected in the values for D_4, which especially for $n = 512$ are reasonably far from unity.

In all cases, optimal bandwidths display a steady, albeit slow, increase with n. Unlike the previous settings, here homoscedastic errors lead to a slower increase of bandwidth with n than the other cases. For heteroscedastic errors, negative autocorrelation seems to warrant slightly higher bandwidths, though this ordering is reversed for $n = 512$. Unreported experiments again show that estimation results are essentially insensitive to moderate variations in bandwidth.

5.6 Empirical application

We apply the methods described in this chapter to estimate common risk exposures for a set of European large-cap equity indices. The series are obtained

3. Positive correlation

	n	512				1024				2048			
	j	4	3	2	1	4	3	2	1	4	3	2	1
R_j	Lev	0.541	0.569	0.748	0.923	0.547	0.591	0.766	0.927	0.557	0.624	0.793	0.926
		(0.085)	(0.146)	(0.134)	(0.053)	(0.076)	(0.156)	(0.125)	(0.042)	(0.065)	(0.167)	(0.108)	(0.037)
	NB	0.684	0.755	0.886	0.936	0.673	0.787	0.904	0.936	0.683	0.831	0.925	0.930
		(0.142)	(0.167)	(0.124)	(0.083)	(0.134)	(0.173)	(0.107)	(0.078)	(0.130)	(0.168)	(0.092)	(0.079)
D_j	Lev	0.852	0.499	0.090	0.005	0.871	0.414	0.063	0.003	0.894	0.313	0.038	0.002
		(0.249)	(0.409)	(0.113)	(0.012)	(0.245)	(0.410)	(0.078)	(0.003)	(0.235)	(0.384)	(0.046)	(0.001)
	NB	0.887	0.498	0.096	0.013	0.903	0.394	0.052	0.006	0.936	0.259	0.023	0.004
		(0.196)	(0.389)	(0.166)	(0.048)	(0.206)	(0.398)	(0.114)	(0.007)	(0.172)	(0.356)	(0.057)	(0.005)
m^*	NB	12				19				20			

4. Negative correlation

	n	512				1024				2048			
	j	4	3	2	1	4	3	2	1	4	3	2	1
R_j	Lev	0.530	0.549	0.725	0.920	0.538	0.566	0.746	0.925	0.549	0.596	0.776	0.925
		(0.080)	(0.135)	(0.141)	(0.056)	(0.072)	(0.145)	(0.132)	(0.044)	(0.065)	(0.158)	(0.114)	(0.038)
	NB	0.672	0.757	0.883	0.937	0.656	0.779	0.901	0.936	0.662	0.808	0.924	0.931
		(0.133)	(0.166)	(0.125)	(0.081)	(0.129)	(0.176)	(0.111)	(0.077)	(0.134)	(0.181)	(0.091)	(0.077)
D_j	Lev	0.834	0.571	0.111	0.005	0.858	0.478	0.077	0.003	0.866	0.387	0.048	0.002
		(0.258)	(0.410)	(0.138)	(0.014)	(0.246)	(0.419)	(0.099)	(0.003)	(0.250)	(0.411)	(0.064)	(0.002)
	NB	0.888	0.509	0.099	0.013	0.898	0.402	0.054	0.006	0.934	0.275	0.024	0.004
		(0.204)	(0.391)	(0.172)	(0.042)	(0.216)	(0.404)	(0.114)	(0.007)	(0.178)	(0.368)	(0.055)	(0.006)
m^*	NB	12				23				45			

Table 5.6: Monte Carlo average eigenvalue ratio (R_j), average distance to parameter subspace (D_j), and optimal NBPC bandwidth (m^*) of PC in levels and NBPC in squares, for varying n and Σ; $q = 9$, $J = 3$. Monte Carlo standard deviation in parentheses.

i	Index	Bloomberg ticker	Country/Region
1	AEX 25	AEX	Netherlands
2	BEL 20	BEL20	Belgium
3	CAC 40	CAC	France
4	DAX 30	DAX	Germany
5	FTSE 100	UKX	UK
6	IBEX 35	IBEX	Spain
7	OMXH 25	HEX25	Finland
8	OMXS 30	OMX	Sweden
9	S&P/MIB 40	SPMIB	Italy
10	SMI 20	SMI	Switzerland
	STOXX 600	SXXP	Europe

Table 5.7: Data description.

from Bloomberg and described in Table 5.7. All indices include a relatively small number of blue chip stocks, which nevertheless account for the majority of traded volume and market valuation within each exchange. With the exception of the FTSE 100, all other national indices include between 20 and 40 assets, while the Pan-european index includes 600. Prices for indices in currencies other than EUR (GBP for the FTSE 100, SEK for the OMXS 30, and CHF for the SMI 20) are corrected for exchange rate movements. In order to account for the different holiday schedules of local exchanges, we kept in the sample those dates for which only one or two national exchanges were closed, filling the price with the last available value. If three or more exchanges were closed we deleted the date from the sample, implying that returns were aggregated over two trading days for exchanges that remained open. After 36 days were excluded by this process, the sample contains 2321 daily price observations, ranging from 4 January 1999 to 10 March 2008.

Figure 5.1 shows the evolution of the indices over the sample period. Not surprisingly, there is clear comovement across all the indices, and the broad direction of the market is reflected in all of them. Still, there are strong differences in particular periods, most notably in the peak of the dot-com bubble, from 1999

Figure 5.1: Total returns for European indices from 1999 to 2008.

to 2001. In this period, both the OMXS 30 and the OMXH 25 displayed strong sensitivity to the sharp increase and fall in technology stocks, while most other indices had more moderate movements, and in particular the BEL 20 appeared to have no exposure to the bubble. These patterns illustrate the difficulty of indices such as the STOXX 600, or indeed of univariate factor models, in capturing the pervasive sources of risk in European markets. Despite the similarity of asset inclusion criteria across all indices considered, the presence of different exposures to particular sectors, international linkages, and other risk sources may cause national exchanges to diverge substantially from each other, more so than can be explained by exposures to a single risk factor. By contrast, and at first glance, co-movement in the period from 2003 to 2008 appears much stronger, and potentially explainable simply by different exposures to aggregate market risk.

Figure 5.2 further illustrates these differences by plotting monthly realised

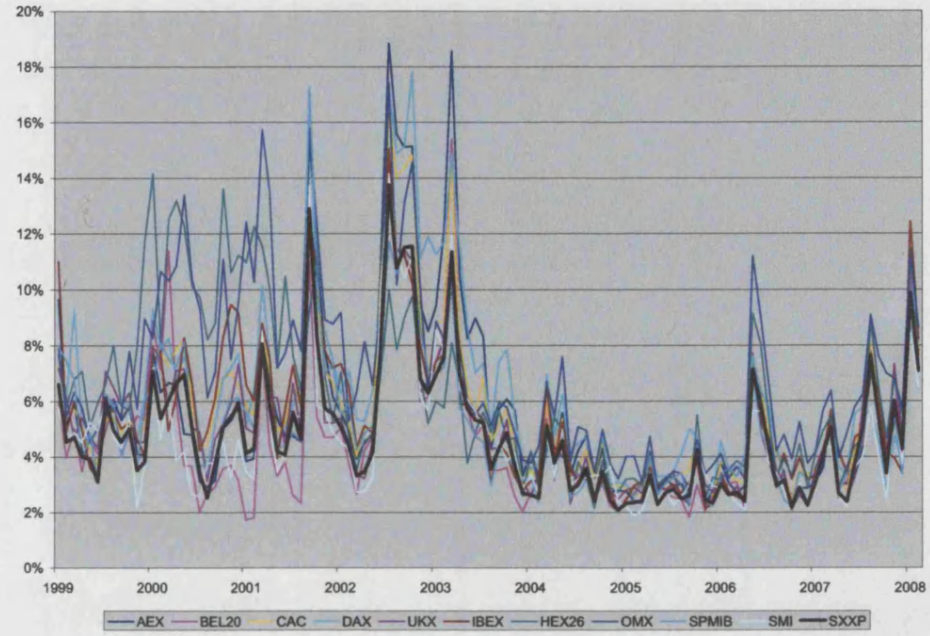


Figure 5.2: Monthly realised SD for European indices from 1999 to 2008.

standard deviations, $\hat{\sigma}_{ik}$, computed as

$$\hat{\sigma}_{ik} = \left(\sum_{t \in T_k} z_{it}^2 \right)^{\frac{1}{2}}, \tag{5.38}$$

where z_{it} denotes the returns of index i in date t and T_k is the set of dates belonging to month k. The volatilities again display the comovement present in the levels, but this comovement was much stronger in the low volatility, post-2003 period than it was during the dot-com bubble. As before, the wide variations in volatility in the earlier period can be justified by the different sector exposures of individual indices, suggesting the need for multivariate analysis of risk factors.

The data we collected allowed us to examine the performance of NBPC in two different settings: one where returns appear to be well explained by a single risk factor, and another where other factors are potentially present. In the following analysis, we obtained estimation results for the full sample of 2320 daily logarith-

mic returns, and also for two subsamples of 1160 observations each, corresponding to periods before and after 11 August 2003. When using regression methods, we used the STOXX 600 as a proxy for European market returns. We estimated factor loadings using OLS (4.4) regression of daily national index returns on STOXX 600 returns, NBLS (4.3) regression of corresponding squared returns, and NBPC (5.21) on the multivariate time series of squared index returns. For both NBLS and NBPC we used the bandwidth $m = [n^{0.5}]$, corresponding to $m = 48$ for the full sample and $m = 34$ for each subsample. We estimated memory parameters of squared returns and squared regression residuals using Local Whittle (2.75), with the bandwidth $m = [n^{0.8}]$, corresponding to $m = 492$ for the full sample and $m = 282$ for each subsample. The results are not sensitive to these choices; results for alternative bandwidths are relegated to Figures 5.3 to 5.23 in Appendix B.

Number of factors

While we have not developed a feasible method for estimation of J in the context of NBPC, this subsection heuristically evaluates the eigenvalue ratios R_j, $j = 1, \ldots 9$, defined in (5.36). Table 5.8 indicates the existence of a clearly dominant factor, but also that this factor was more important in the second half of the sample than on the first. These facts are in line with our initial observation that the indices appeared to display stronger comovement in the post-2003 period than during the dot-com bubble. The relevance of a second risk factor cannot be analysed convincingly in the absence of a formal statistical test. However, and as predicted, results for the full sample and the first half of the sample seem to support the inclusion of a second factor, while the second half of the sample seems more consistent with a single factor model. Factors beyond the first two do not display any potential for inclusion. We therefore report NBPC results assuming $J = 2$, with the proviso that the second factor may be weak or non-existent during

j	Full sample	First half	Second half
1	0.811	0.779	0.901
2	0.590	0.530	0.461
3	0.364	0.372	0.411
4	0.290	0.302	0.411
5	0.356	0.368	0.366
6	0.439	0.445	0.321
7	0.355	0.362	0.369
8	0.495	0.455	0.474
9	0.562	0.538	0.613

Table 5.8: Proportion of remaining averaged periodogram (R_j) explained by the j-th NBPC eigenvector.

the latter period.

Factor exposures

Estimates of factor exposures using OLS in returns and NBLS in squared returns are presented in Table 5.9. We convert NBLS estimates to market exposures by simply taking the positive square root, since all exposures can be reasonably assumed positive. Exposures are concentrated around one (which is natural in this setting), and this concentration increases substantially from the first subsample to the second. Most NBLS estimates are slightly higher that those obtained by OLS, arguably reflecting the stronger OLS downward bias suggested by the Monte Carlo results of Section 4.2. The notable exception is HEX25, to which NBLS attributes an exposure of 0.49 in the first half of the sample. While this is a surprisingly low value, Figure 5.2 illustrates the abnormal behaviour of this index in this period: it displayed significant excess volatility over the STOXX 600 during the peak of the bubble, from 1999 to the end of 2001, but it was one of the least volatile assets throughout 2002 and 2003. The volatility profiles during that subsample cast doubts over the crucial assumption of cointegration in volatility between the two indices.

	OLS			NBLS		
	Full sample	First half	Second half	Full sample	First half	Second half
AEX	1.139	1.175	1.043	1.360	1.391	1.093
BEL20	0.778	0.736	0.886	1.028	1.030	1.054
CAC	1.128	1.140	1.095	1.193	1.184	1.109
DAX	1.159	1.183	1.095	1.324	1.333	1.034
UKX	0.989	0.987	0.993	1.018	1.010	0.989
IBEX	0.970	0.968	0.975	1.019	0.960	1.145
HEX25	0.951	0.927	1.015	0.800	0.489	1.174
OMX	1.177	1.175	1.183	1.177	1.088	1.190
SPMIB	0.935	0.953	0.887	1.013	1.009	0.886
SMI	0.799	0.802	0.789	1.002	1.028	0.816

Table 5.9: Univariate OLS and NBLS estimates of market exposures.

Table 5.10 reports the first two eigenvectors obtained from the NBPC procedure, denoted PC1 and PC2. The first component clearly corresponds to a measure of market exposure, and follows very closely the relative magnitudes of NBLS estimates. In particular, HEX25 displays an extremely low exposure in the first half of the sample, and above average exposure in the second half. The second principal component is harder to interpret due to the indeterminacy discussed in Section 5.4. Still, for the full sample and first half of the sample the bulk of the (normalised) exposure is assigned to HEX25. This is consistent with previous results: there appears to be a second factor explaining the cross-section of returns, to which HEX25 and OMX have particularly strong exposures, at least during the first half of the sample.

We now illustrate the methods used in Section 5.4 to remove the indeterminacy in the coefficients and allow estimation of the squared factor loadings. Assuming different memory parameters of the two unobservable factors, identification of exposures to the first factor requires only one normalisation restriction. Taking into account the results of Table 5.9, and also the relative importance of each individual index, we normalised the exposure of UKX to one. The estimates obtained by univariate regression for this loading range from 0.987 to 1.018; in addition, the FTSE 100 has the largest market value and traded volume of all the national indices considered, and is substantially more diversified than any of the others.

Identification of exposures to the second factor is not as straightforward, as it requires two restrictions on the coefficients. While the normalisation restriction poses no problems, the additional exclusion restriction is somewhat arbitrary. In the full sample and in the first subsample, HEX25 has the lowest weight on the first component and the largest (absolute) weight on the second component; we therefore normalised its loading on the second component to one. AEX is chosen

	PC1			PC2		
	Full sample	First half	Second Half	Full sample	First half	Second Half
AEX	0.456	0.495	0.333	0.335	0.265	0.280
BEL20	0.262	0.277	0.308	0.127	0.121	-0.004
CAC	0.351	0.356	0.340	0.089	0.043	0.172
DAX	0.438	0.455	0.305	0.186	0.109	0.405
UKX	0.251	0.253	0.256	0.021	-0.011	-0.150
IBEX	0.256	0.231	0.370	-0.087	-0.087	0.389
HEX25	0.177	0.063	0.381	-0.755	-0.754	-0.244
OMX	0.353	0.300	0.397	-0.483	-0.560	-0.700
SPMIB	0.252	0.252	0.215	-0.035	-0.061	-0.058
SMI	0.244	0.264	0.184	0.127	0.091	0.020

Table 5.10: First two NBPC eigenvectors.

for the exclusion restriction, since it is the only asset for which such a restriction yielded non-negative values for all squared loadings; AEX it is also the index with the highest exposure to the first component. In the second subsample, we kept the normalisation restriction on HEX25 for comparison purposes, but the exclusion restriction was imposed on DAX to obtain non-negative exposures throughout. As noted before, the second factor does not appear to be strong in this subsample, and therefore results relating to these exposures should be interpreted with caution.

Table 5.11 reports estimated factor loadings using the described restrictions and a square root transformation. As for NBLS, the loadings for the first factor may be assumed positive, and we may use the positive root for direct comparison with Table 5.9. However, there is no reason to expect weights on the second factor to be all of the same sign, and we may only interpret the resulting loadings as estimates of absolute exposures. The weights for the first factor are essentially identical to the NBLS estimates in Table 5.9, with discrepancies all below 0.06. The weights of the second factor for the full sample and the first half seem to indicate the strength of deviations from the market during the dot-com bubble. In particular, the Nordic indices HEX25 and OMX both display strong exposures consistently over both subsamples. The relative ranking of other individual exposures is hard to evaluate, since it is likely to be sensitive to the restriction chosen for identification.

Memory parameters

Both NBLS and NBPC depend crucially on the assumption that the persistence in volatility of the leading factor (or factors) dominates that of the errors. To assess the validity of this assumption, Table 5.12 reports Local Whittle estimates of the memory of squared returns and squared regression residuals, where the latter are obtained from both OLS and NBLS. To make the residual memory

	First factor			Second factor		
	Full sample	First half	Second Half	Full sample	First half	Second Half
AEX	1.348	1.398	1.140	*0*	*0*	0.466
BEL20	1.021	1.047	1.096	0.271	0.186	0.742
CAC	1.183	1.187	1.152	0.437	0.432	0.610
DAX	1.322	1.341	1.090	0.392	0.413	*0*
UKX	*1*	*1*	*1*	0.430	0.431	0.808
IBEX	1.010	0.955	1.202	0.557	0.517	0.370
HEX25	0.840	0.499	1.219	*1*	*1*	*1*
OMX	1.186	1.089	1.244	0.916	0.957	1.279
SPMIB	1.002	0.998	0.915	0.499	0.499	0.676
SMI	0.986	1.022	0.846	0.242	0.252	0.547

Table 5.11: NBPC estimates of absolute exposures to first two factors, under normalisation and exclusion restrictions.

estimation fully comparable, we computed NBLS residuals by first applying the square root transformation to the estimates, obtaining excess market returns using those estimates, and squaring the resulting sequence. We also report asymptotic SD for the estimates (see (2.76)), and summary statistics for the ten national index estimates.

All Local Whittle estimates in Table 5.12 indicate the presence of stationary long memory, being significantly different from 0 and 0.5. Memory estimates of squared returns appear consistent with identity of memory parameters. Differences between memory estimates for OLS and NBLS squared residuals are negligible, with only three of the thirty residual estimate pairs displaying discrepancies larger than 0.02. Most residual estimates are substantially lower than the corresponding squared returns estimate, indicating the validity of the cointegration assumption. A notable exception is HEX25, for which the memory stays essentially the same after regression in the full sample, and is even slightly higher for squared residuals in the first subsample. This fact sheds some light on the unusually low exposure found by both NBLS and NBPC for these sample periods. The second subsample is substantially different, with a sizeable reduction in HEX25 memory, from 0.3 in squared returns to 0.1 in regression residuals. The estimated memory reductions for BEL20 and OMX are also relatively low for the full sample and the first subsample, rising to substantial values for the second subsample. All other assets appear to yield strong memory reductions in all samples, ranging from 0.09 (CAC, full sample) to 0.27 (CAC, second subsample). The indices for which the first subsample appears to invalidate the cointegration hypothesis can be immediately identified in Figure 5.1: both HEX25 and OMX were initially found to have excess sensitivity to the dot-com phenomenon, while BEL20 appeared to display no sensitivity at all.

	Squared returns			OLS residuals			NBLS residuals		
	Full sample	First half	Second half	Full sample	First half	Second half	Full sample	First half	Second half
AEX	0.307	0.275	0.360	0.198	0.124	0.124	0.168	0.080	0.114
BEL20	0.271	0.232	0.389	0.230	0.199	0.142	0.228	0.196	0.119
CAC	0.282	0.242	0.379	0.194	0.105	0.119	0.196	0.108	0.113
DAX	0.319	0.279	0.354	0.190	0.144	0.174	0.170	0.128	0.181
UKX	0.331	0.289	0.331	0.171	0.134	0.150	0.171	0.135	0.150
IBEX	0.294	0.236	0.382	0.193	0.120	0.153	0.188	0.122	0.166
HEX25	0.226	0.171	0.304	0.230	0.185	0.093	0.225	0.190	0.100
OMX	0.254	0.197	0.286	0.183	0.137	0.099	0.183	0.142	0.099
SPMIB	0.306	0.248	0.309	0.209	0.146	0.130	0.197	0.133	0.130
SMI	0.288	0.263	0.254	0.187	0.147	0.119	0.186	0.159	0.118
SXXP	0.320	0.277	0.376	—	—	—	—	—	—
Asy. SD	*0.023*	*0.030*	*0.030*	*0.023*	*0.030*	*0.030*	*0.023*	*0.030*	*0.030*
Maximum	0.331	0.289	0.389	0.230	0.199	0.174	0.228	0.196	0.181
Average	0.288	0.243	0.335	0.198	0.144	0.130	0.191	0.139	0.129
Minimum	0.226	0.171	0.254	0.171	0.105	0.093	0.168	0.080	0.099
SD	0.031	0.037	0.046	0.019	0.029	0.025	0.021	0.035	0.028

Table 5.12: Local Whittle estimates for squared returns, squared OLS residuals, and squared NBLS residuals.

5.7 Final comments

We have presented an application of theoretical techniques developed in Chapter 3 to a multivariate setting. In the context of a multivariate factor model, where both the factors and idiosyncratic errors are driven by general SV models, a narrow band version of principal components is shown to converge to vectors spanning the same space as the transformed factor loadings. Monte Carlo results are encouraging in a number of alternative specifications; however, they also highlight the superiority of the "traditional" principal components approach in the levels when the idiosyncratic errors are uncorrelated and homoscedastic across the observable series. This contrasts with the regression setting of Chapter 3, where errors-in-variables inconsistency is always present for the OLS estimate. The empirical application of the previous section illustrates the use of NBPC, there applied to equity indices. We find strong support for the need of more than one factor in modelling equity returns, at least during the peak of the dot-com bubble. NBLS estimation by regression on a proxy for market returns yields reassuringly close results to the first factor of NBPC, which does not rely on pre-specified risk factors.

The indeterminacy in parameter estimates highlighted in Theorem 5.2 and Corollary 5.1, coupled with the transformation (5.8), may limit the usefulness of the estimates. If $d_1 > d_2$, the different rates of growth of $\hat{F}_{ff}$ components in (5.16) allow consistent estimation of the vector θ_1 up to scale. However, even when $d_2 > d_3$, the second principal component can only be shown to converge to a linear combination of θ_1 and θ_2, and so on for subsequent vectors. The imposition of linear restrictions on elements of θ may resolve the indeterminacy, but such prior knowledge may not be available to the researcher in practical applications. Furthermore, since the elements of θ are related to the original β by a power transformation, any linear restrictions involving more than one element of β will

be translated into nonlinear restrictions on θ. While a sufficient number of such restrictions may still fully identify β, specifying general conditions for identification is not trivial in this setting.

Our choice of a principal components approach instead of a regression based estimate, such as NBLS, seems to be in line with widespread practice in APT applications. This approach does not require the separation of variables into dependent and explanatory, which would be arbitrary under (5.1). Furthermore, tests or estimates for the number of factors frequently rely on functions of eigenvalues, and principal components has the potential of being integrated on a full data-dependent modeling technique, where the number of factors and corresponding cointegrating vectors are estimated jointly. Nonetheless, establishing consistency of a multivariate extension for NBLS would be of considerable independent interest, and indeed such a result is likely to follow closely the theoretical steps presented here.

In practical applications of multivariate factor models, the number of common factors is rarely known to the researcher, and must be determined by the data. We presented a method of estimating J if the memory parameters of factors and residuals are known, which is not a realistic assumption. In Section 5, we also suggested certain ratios of eigenvalues, denoted R_j, as a possible means of determining cointegrating rank. Obviously, our use of these quantities is purely heuristic and would require a formal theoretical justification before being applicable in practical situations, such as the empirical analysis undertaken in the previous section. It would be obviously desirable to develop methods of formally testing hypotheses on the number of factors and other parameters, for which asymptotic distributional theory would be required. As was the case for NBLS, finding limiting distributions in our nonlinear setting is far from trivial, and would require considerable further work; we leave these topics for future research.

5.A Propositions and lemmas

Proposition 5.1 *For* $1 \leq j_1, j_2 \leq J$,

$$\hat{F}_{f_{j_1}, f_{j_2}} = C^* \left(\frac{m}{n}\right)^{1-d_{j_1}-d_{j_2}} + o_p\left(\left(\frac{m}{n}\right)^{1-d_{j_1}-d_{j_2}}\right). \tag{5.39}$$

where $C^* \neq 0$ *iff* $j_1 = j_2$.

Proof. For $k = 1, 2$, write $f_{j_k,t} = A_{kt}B_{kt}$, where $A_{kt} = \eta_{j_k,t}$ and $B_{kt} = g_{j_k}(\nu_{j_k,t})$, and apply Lemma 5.1 to $\mathrm{Cov}(f_{j_1,t}, f_{j_2,s})$. Then (5.39) follows immediately from from Lemma 5.3, noting that $C^* \neq 0$ iff $j_1 = j_2$ due to (5.6). □

For the following Proposition some notation is needed. Fix $1 \leq i \leq q$, write $\mathbf{p} = (p_1, \ldots, p_J)'$, and for $1 \leq j \leq J$ denote by $r^j_{\mathbf{p}}$ and $r^0_{\mathbf{p}}$ the Hermite ranks of $g^{p_j}_{jt} - E(g^{p_j}_{jt})$ and $h^{p-\Sigma p_j}_{it} - E(h^{p-\Sigma p_j}_{it})$ respectively. Using the convention $0^0 = 1$, define

$$\begin{aligned}
\mathbf{P} &= \{\mathbf{p} : 0 \leq p_j < p, \sum_{j=1}^{J} p_j \leq p\}, \\
S_1 &= \left\{\mathbf{p} \in \mathbf{P} : E\left\{\xi^{p-\Sigma p_j}_{it} \textstyle\prod_{j=1}^{J} (\beta_{ij}\eta_{jt})^{p_j}\right\} \neq 0\right\}, \\
S_{2\mathbf{p}} &= \left\{1 < j < J : E(g^{p_j}_{jt}) \neq 0\right\}, \\
d^*_{\mathbf{p}} &= \max_{j \in S_{2\mathbf{p}}} \left\{\frac{1}{2} - r^j_{\mathbf{p}}\left(\frac{1}{2} - d_j\right)\right\}, \\
b^*_{\mathbf{p}} &= \frac{1}{2} - r^0_{\mathbf{p}}\left(\frac{1}{2} - b_i\right) - 1\left\{E(h^{p-\Sigma p_j}_{it})b_i = 0\right\}, \\
d^* &= \max_{\mathbf{p} \in S_1}\{d^*_{\mathbf{p}}, b^*_{\mathbf{p}}\},
\end{aligned}$$

where we omit the dependence on i for convenience.

Proposition 5.2 *For* $1 \leq i \leq q$, *defining* $b'_i = \max\{d^*, 0\}$,

$$\hat{F}_{U_i,U_i} = O_p\left(\left(\frac{m}{n}\right)^{1-2b'_i} (\log n)^{1(d^*=0)}\right). \tag{5.40}$$

Proof. Write

$$U_{it} = \sum_{\mathbf{p}\in\mathbf{P}} c_{\mathbf{p}} A_{\mathbf{p}t} B_{\mathbf{p}t},$$

where the dependence on i is omitted throughout for convenience, and

$$A_{\mathbf{p}t} = \xi_{it}^{p-\Sigma p_j} \prod_{j=1}^{J} \eta_{jt}^{p_j}, \qquad B_{\mathbf{p}t} = h_{it}^{p-\Sigma p_j} \prod_{j=1}^{J} g_{jt}^{p_j},$$

$$c_{\mathbf{p}} = \frac{p!}{(p-\sum_{j=1}^{J} p_j)!} \prod_{j=1}^{J} \frac{\beta_{ij}^{p_j}}{p_j!}.$$

Noting that for any $\mathbf{p}, \mathbf{q} \in \mathbf{P}$, $\{A_{\mathbf{p}t}\}$ is independent of $\{B_{\mathbf{q}t}\}$, and $A_{\mathbf{p}t}$ is independent of $A_{\mathbf{q}s}$ unless $t = s$, Lemma 5.1 yields

$$\begin{aligned}\operatorname{Cov}(U_{is}, U_{it}) &= \sum_{\mathbf{p},\mathbf{q}\in\mathbf{P}} c_{\mathbf{p}} c_{\mathbf{q}} \operatorname{Cov}(A_{\mathbf{p}s} B_{\mathbf{p}s}, A_{\mathbf{q}t} B_{\mathbf{q}t}) \\ &= \sum_{\mathbf{p},\mathbf{q}\in\mathbf{P}} c_{\mathbf{p}} c_{\mathbf{q}} E(A_{\mathbf{p}t}) E(A_{\mathbf{q}t}) \operatorname{Cov}(B_{\mathbf{p}s}, B_{\mathbf{q}t}) + \Lambda 1(s = t),\end{aligned}$$

where $\Lambda = \sum_{\mathbf{p},\mathbf{q}\in\mathbf{P}} c_{\mathbf{p}} c_{\mathbf{q}} \operatorname{Cov}(A_{\mathbf{p}t}, A_{\mathbf{q}t}) E(B_{\mathbf{p}t} B_{\mathbf{q}t}) < \infty$. Thus,

$$E(\hat{F}_{U_i,U_i}) = \sum_{\mathbf{p},\mathbf{q}} c_{\mathbf{p}} c_{\mathbf{q}} E(A_{\mathbf{p}t}) E(A_{\mathbf{q}t}) E(\hat{F}_{B_{\mathbf{p}},B_{\mathbf{q}}}) + \Lambda \frac{m}{n}. \tag{5.41}$$

For $j = 1, \ldots, J$ denote

$$\begin{aligned} cov_{\mathbf{pq}}^{j}(t-s) &= \operatorname{Cov}(g_{js}^{p_j}, g_{jt}^{q_j}), & e_{\mathbf{p}}^{j} &= E(g_{jt}^{p_j}), \\ cov_{\mathbf{pq}}^{0}(t-s) &= \operatorname{Cov}(h_{is}^{p-\Sigma p_j}, h_{it}^{p-\Sigma q_j}), & e_{\mathbf{p}}^{0} &= E(h_{it}^{p-\Sigma p_j}). \end{aligned}$$

Using Lemma 5.1 again, $\operatorname{Cov}(B_{\mathbf{p}s}, B_{\mathbf{q}t})$ can be written

$$\prod_{j=0}^{J} \{e_{\mathbf{p}}^{j} e_{\mathbf{q}}^{j} + cov_{\mathbf{pq}}^{j}(t-s)\} - \prod_{j=0}^{J} e_{\mathbf{p}}^{j} e_{\mathbf{q}}^{j}$$

$$= \sum_Q e^0_{\mathbf{p}} e^0_{\mathbf{q}} \prod_{j \notin Q} (e^j_{\mathbf{p}} e^j_{\mathbf{q}}) \prod_{j \in Q} cov^j_{\mathbf{pq}}(t-s)$$
$$+ \prod_{j=1}^{J} (e^j_{\mathbf{p}} e^j_{\mathbf{q}}) cov^0_{\mathbf{pq}}(t-s)$$
$$+ \sum_Q \prod_{j \notin Q} (e^j_{\mathbf{p}} e^j_{\mathbf{q}}) \prod_{j \in Q} cov^j_{\mathbf{pq}}(t-s) cov^0_{\mathbf{pq}}(t-s),$$

where $\sum_Q$ denotes summation over all non-empty subsets Q of $\{1, \ldots, J\}$.

It follows that $E(\hat{F}_{B_{\mathbf{p}}, B_{\mathbf{q}}})$ is

$$\sum_Q e^0_{\mathbf{p}} e^0_{\mathbf{q}} \prod_{j \notin Q} (e^j_{\mathbf{p}} e^j_{\mathbf{q}}) \left\{ \frac{1}{n^2} \sum_{s,t=1}^{n} D_m(\lambda_{t-s}) \prod_{j \in Q} cov^j_{\mathbf{pq}}(t-s) \right\} \tag{5.42}$$
$$+ \prod_{j=1}^{J} (e^j_{\mathbf{p}} e^j_{\mathbf{q}}) \left\{ \frac{1}{n^2} \sum_{s,t=1}^{n} D_m(\lambda_{t-s}) cov^0_{\mathbf{pq}}(t-s) \right\} \tag{5.43}$$
$$+ \sum_Q \prod_{j \notin Q} (e^j_{\mathbf{p}} e^j_{\mathbf{q}}) \left\{ \frac{1}{n^2} \sum_{s,t=1}^{n} D_m(\lambda_{t-s}) \prod_{j \in Q} cov^j_{\mathbf{pq}}(t-s) cov^0_{\mathbf{pq}}(t-s) \right\} \tag{5.44}$$

If $\mathbf{p} \notin S_1$ or $\mathbf{q} \notin S_1$, the corresponding $\mathbf{p}, \mathbf{q}$ term will vanish from (5.41). Similarly, unless $Q \subset S_{2\mathbf{p}} \cap S_{2\mathbf{q}}$, the Q terms will vanish from (5.42) and (5.44), while if $e^0_{\mathbf{p}} e^0_{\mathbf{q}} = 0$ (5.42) disappears. For the remaining values of $\mathbf{p}, \mathbf{q}, Q$, Lemma 5.3 implies that

$$\frac{1}{n^2} \sum_{s,t=1}^{n} D_m(\lambda_{t-s}) \prod_{j \in Q} cov^j_{\mathbf{pq}}(t-s)$$
$$= O\left(\left(\frac{m}{n} \right)^{1 - 2\max\{d^*_{\mathbf{pq}Q}, 0\}} (\log n)^{1(d^*_{\mathbf{pq}Q} = 0)} \right),$$
$$\frac{1}{n^2} \sum_{s,t=1}^{n} D_m(\lambda_{t-s}) cov^0_{\mathbf{pq}}(t-s)$$
$$= O\left(\left(\frac{m}{n} \right)^{1 - 2\max\{b^*_{\mathbf{pq}}, 0\}} (\log n)^{1(b^*_{\mathbf{pq}} = 0)} \right),$$
$$\frac{1}{n^2} \sum_{s,t=1}^{n} D_m(\lambda_{t-s}) \prod_{j \in Q} cov^j_{\mathbf{pq}}(t-s) cov^0_{\mathbf{pq}}(t-s)$$

$$= O\left(\left(\frac{m}{n}\right)^{1-2\max\{d^{\times}_{\mathbf{pqQ}},0\}} (\log n)^{1(d^{\times}_{\mathbf{pqQ}}=0)}\right),$$

where

$$\begin{aligned}
d^{*}_{\mathbf{pqQ}} &= \frac{1}{2} - \sum_{j\in Q} r^{j}_{\mathbf{pq}}\left(\frac{1}{2} - d_j\right), \\
b^{*}_{\mathbf{pq}} &= \frac{1}{2} - r^{0}_{\mathbf{pq}}\left(\frac{1}{2} - b_i\right), \\
d^{\times}_{\mathbf{pqQ}} &= \frac{1}{2} - \sum_{j\in Q} r^{j}_{\mathbf{pq}}\left(\frac{1}{2} - d_j\right) - r^{0}_{\mathbf{pq}}\left(\frac{1}{2} - b_i\right), \\
r^{j}_{\mathbf{pq}} &= \min\{k > 0 : G^{j}_{\mathbf{p},k} G^{j}_{\mathbf{q},k} \neq 0\}, \\
r^{0}_{\mathbf{pq}} &= \min\{k > 0 : G^{0}_{\mathbf{p},k} G^{0}_{\mathbf{q},k} \neq 0\},
\end{aligned}$$

and $G^{j}_{\mathbf{p},k}$, $G^{0}_{\mathbf{p},k}$ are the k-th Hermite coefficients of $g^{p_j}_{js}$, $h^{p-\Sigma p_j}_{it}$ respectively. Notice that since $r^{j}_{\mathbf{pq}} \leq r^{j}_{\mathbf{p}}$, $j = 0, \ldots, J$, we have $d^{\times}_{\mathbf{pqQ}} \leq d^{*}_{\mathbf{pqQ}} \leq d^{*}_{\mathbf{ppQ}}$ and $b^{*}_{\mathbf{pq}} \leq b^{*}_{\mathbf{p}}$. Also, notice that for any $j \in Q$, we have $d^{*}_{\mathbf{ppQ}} \leq d^{*}_{\mathbf{pp}\{j\}} \leq d^{*}_{\mathbf{p}}$. It follows that all terms in (5.42) to (5.44) can be bounded by (5.40), completing the proof. □

The following Lemmas are straightforward generalisations of results proved in Chapter 3, namely Lemmas 3.1, 3.2, 3.4, and 3.5, to a multivariate setting.

Lemma 5.1 *Let (a_i, b_i), $i = 1, \ldots, k$, be mutually independent random variables such that $E(a_i^2 + b_i^2) < \infty$. Then,*

$$\begin{aligned}
E\left(\prod_{i=1}^{k} a_i b_i\right) &= \prod_{i=1}^{k} \{\mathrm{Cov}(a_i, b_i) + E(a_i)E(b_i)\} \\
&= \sum_{Q} \prod_{i\notin Q} \{E(a_i)E(b_i)\} \prod_{i\in Q} \mathrm{Cov}(a_i, b_i),
\end{aligned}$$

where $\sum_Q$ denotes summation over all subsets Q of $\{1, \ldots, k\}$.

Proof. Straightforward. □

Lemma 5.2 *For $i = 1, \ldots, k$, assume $|\rho_{i,t} - \rho_{i,t+1}| \leq Kt^{-1}|\rho_{i,t+1}|$ for all $t \geq 1$. Then, for positive integers $r_1, \ldots, r_k$ and t,*

$$\left| \prod_{i=1}^{k} \rho_{i,t}^{r_i} - \prod_{i=1}^{k} \rho_{i,t+1}^{r_i} \right| \leq Kt^{-1} \prod_{i=1}^{k} |\rho_{i,t+1}^{r_i}|. \tag{5.45}$$

Proof. By Lemma 3.1, (5.45) holds for $k = 1, 2$. Proceeding by induction, suppose it holds for $k = 1, \ldots, s-1$. Then,

$$\begin{aligned} \left| \prod_{i=1}^{s} \rho_{i,t}^{r_i} - \prod_{i=1}^{s} \rho_{i,t+1}^{r_i} \right| &\leq |\rho_{s,t}^{r_s}| \left| \prod_{i=1}^{s-1} \rho_{i,t}^{r_i} - \prod_{i=1}^{s-1} \rho_{i,t+1}^{r_i} \right| + |\rho_{s,t}^{r_s} - \rho_{s,t+1}^{r_s}| \prod_{i=1}^{s} |\rho_{i,t+1}^{r_i}| \\ &\leq Kt^{-1} \prod_{i=1}^{s} |\rho_{i,t+1}^{r_i}|. \ \square \end{aligned}$$

Lemma 5.3 *For $i = 1, \ldots, k$, $j = 1, 2$, define $g_{ij,t} = g_{ij}(\mu_{it})$, where μ_{it} is a standard Gaussian $I(d_i)$ process and $\rho_{i,t} = E(\mu_{i0}\mu_{it})$. Assume $E(g_{ij,t}^2) < \infty$. Denote by $G_{ij,k}$ the k-th Hermite coefficient of $g_{ij}(\cdot)$, with*

$$r_i = \min\{k > 0 : G_{i1,k} G_{i2,k} \neq 0\}. \tag{5.46}$$

Let $d_1 = \max_i d_i$ without loss of generality, and define

$$d^* = \frac{1}{2} - \sum_{i=1}^{k} r_i \left(\frac{1}{2} - d_i \right), \qquad C_{i\rho} = \lim_{j \to \infty} \rho_{ij} j^{1-2d_i}.$$

Let

$$A = \frac{1}{n^2} \sum_{s,t=1}^{n} D_m(\lambda_{t-s}) \prod_{i=1}^{k} \mathrm{Cov}(g_{i1,s}, g_{i2,t}),$$

where m satisfies Assumption 5.3. Then,

$$A = O\left(\frac{m}{n} \{1 + (\log n) 1(d^* = 0)\} \right)$$

$$+\left\{C^*\left(\frac{m}{n}\right)^{1-2d^*}+o\left(\left(\frac{m}{n}\right)^{1-2d^*}\right)\right\}1(d^*>0), \tag{5.47}$$

where

$$C^*=2\frac{(2\pi)^{-2d^*}\Gamma(2d^*)}{1-2d^*}\sin\left\{(1-2d^*)\frac{\pi}{2}\right\}\prod_{i=1}^{k}\frac{G_{i1,r_i}G_{i2,r_i}}{r_i!}C_{i\rho}^{r_i}\neq 0.$$

Proof. Let $\gamma_{i,t}=\mathrm{Cov}(g_{i1,0};g_{i2,t})$. Then, similarly to (3.41),

$$A=\frac{1}{n}\sum_{u=1-n}^{n-1}\left(1-\frac{|u|}{n}\right)D_m(\lambda_u)\prod_{i=1}^{k}\gamma_{iu}.$$

By Theorem 3.1 and (5.46),

$$\gamma_{iu}=\sum_{k=1}^{\infty}\frac{G_{i1,k}G_{i2,k}}{k!}\rho_{iu}^{k}=C_i\rho_{iu}^{r_i}+O(|\rho_{iu}^{r_i+1}|),$$

where $C_i=G_{i1,r_i}G_{i2,r_i}/r_i!$.

(a) If $\prod_{i=1}^{k}d_i=0$, then $\prod_{i=1}^{k}\gamma_{iu}=O(\prod_{i=1}^{k}|\rho_{iu}^{r_i}|)$ are summable. Similarly, if $\prod_{i=1}^{k}d_i>0$ but $d^*<0$, then $\prod_{i=1}^{k}\gamma_{iu}=O(\prod_{i=1}^{k}|\rho_{iu}^{r_i}|)=O(j^{2d^*-1})$ are summable. In either case, writing $\prod_{i=1}^{k}\gamma_{iu}$ instead of γ_u in (3.44) yields $A=O(m/n)$.

(b) If $d^*=0$, $\prod_{i=1}^{k}\gamma_{iu}=O(j^{-1})$, hence (3.44) holds for $\gamma_{1u}\gamma_{2u}$.

(c) If $d^*>0$,

$$\prod_{i=1}^{k}\gamma_{iu}=\prod_{i=1}^{k}\gamma_{iu}C_i\rho_{iu}^{r_i}+O\left(|\rho_{1u}|\prod_{i=1}^{k}|\rho_{iu}^{r_i}|\right),$$

so

$$\left|\prod_{i=1}^{k}\gamma_{iu}-\prod_{i=1}^{k}C_i\rho_{iu}^{r_i}\right|\leq K\left(\prod_{i=1}^{k}|\rho_{iu}^{r_i}|\right)^{1+\omega},$$

where $\omega=(1-2d_1)/(1-2d^*)$. Then (3.53) follows from the proof of case (c) of Lemma 3.4, writing $\prod_{i=1}^{k}\rho_{iu}^{r_i}$ instead of ρ_u^r, and making use of Lemma 5.2. □

5.B Sensitivity of empirical results to bandwidth choice

This appendix expands on the empirical results of Section 5.6 by analysing the impact of bandwidth choice on various estimates. Figures 5.3 to 5.23 display the evolution of NBLS estimates, NBPC eigenvalue ratios, the first two NBPC eigenvectors, and Local Whittle memory estimates. Despite the natural instability for very small bandwidths, all estimates are fairly stable in the neighbourhood of those used in Section 5.6.

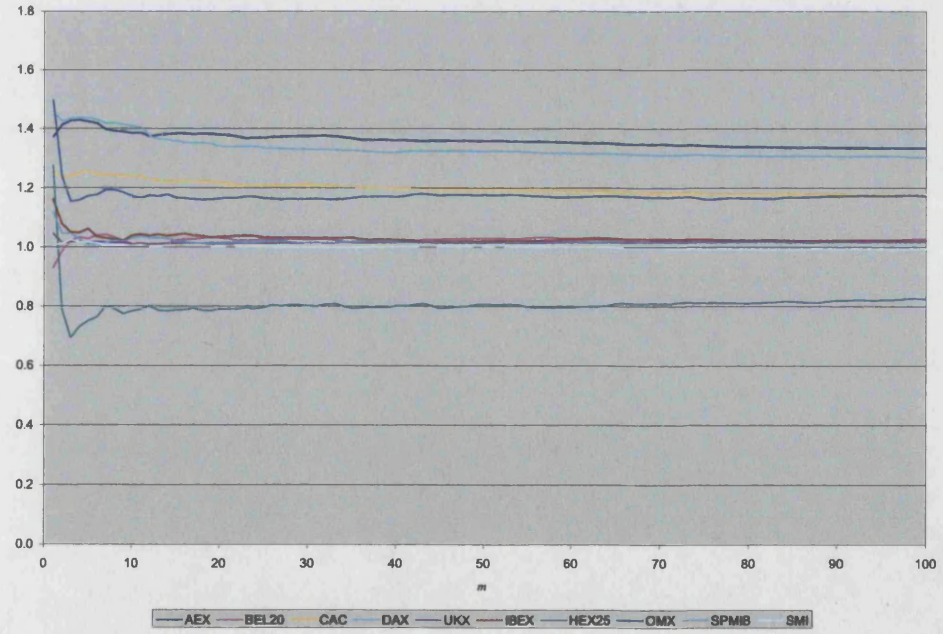


Figure 5.3: NBLS estimates of market exposures for varying m; full sample.

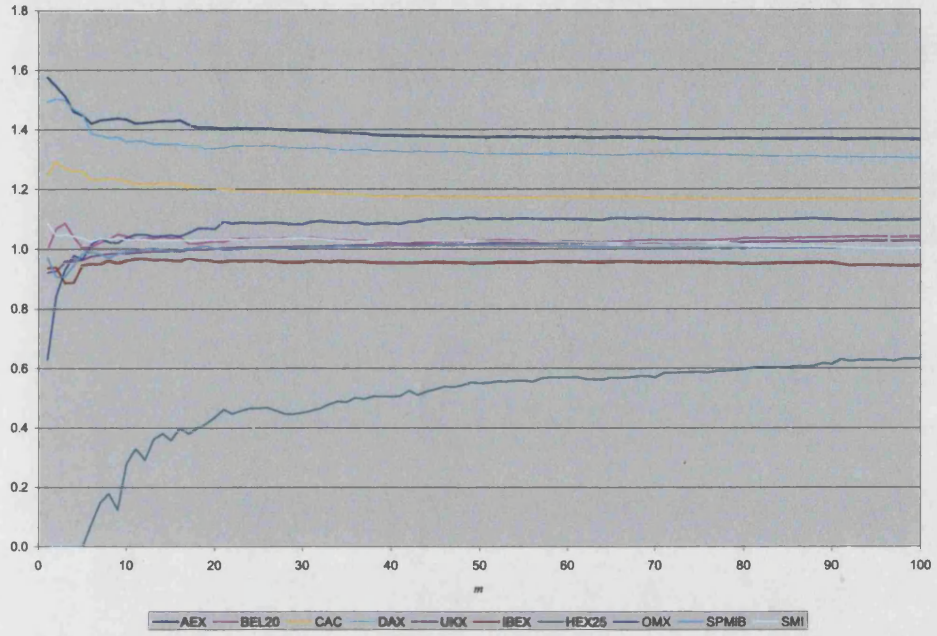


Figure 5.4: NBLS estimates of market exposures for varying m; first subsample.

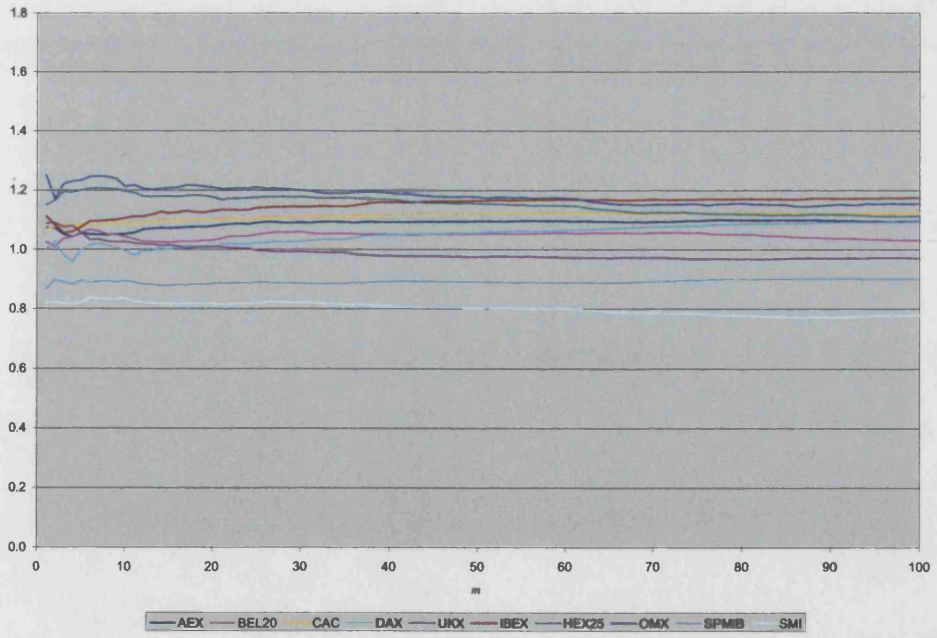


Figure 5.5: NBLS estimates of market exposures for varying m; second subsample.

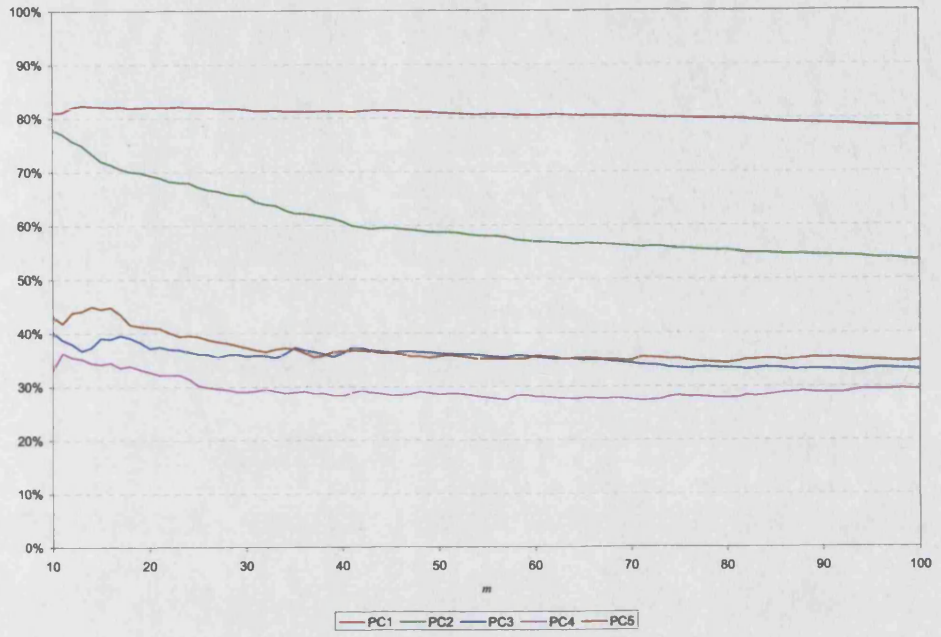


Figure 5.6: Proportion of remaining averaged periodogram (R_j) explained by the j-th NBPC eigenvector, for varying m; full sample.

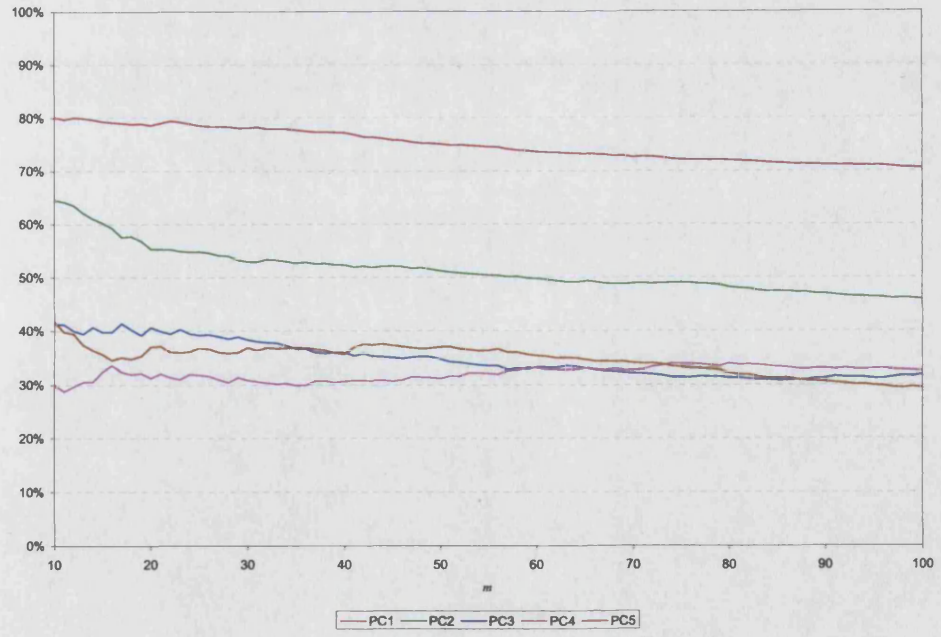


Figure 5.7: Proportion of remaining averaged periodogram (R_j) explained by the

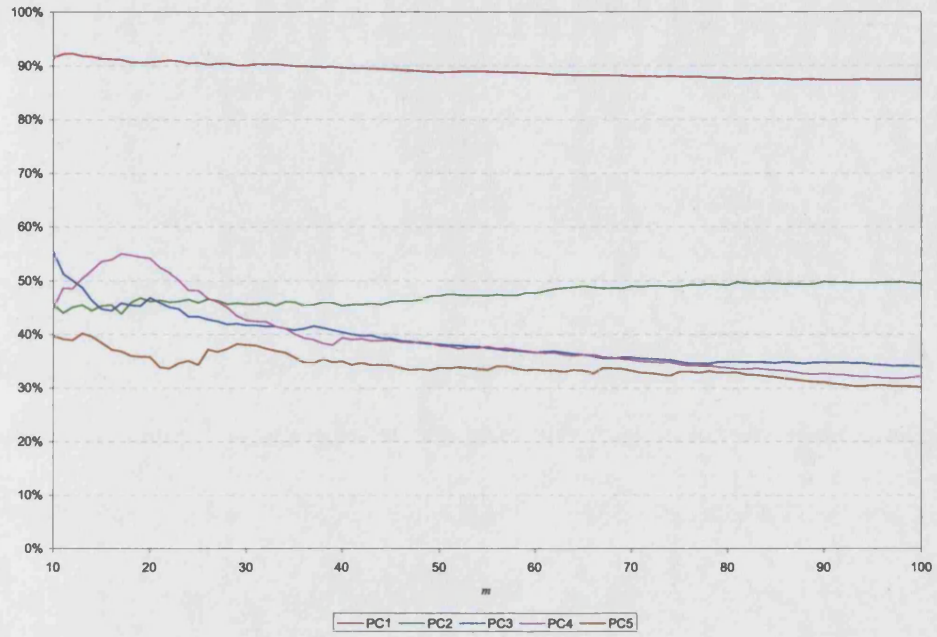


Figure 5.8: Proportion of remaining averaged periodogram (R_j) explained by the j-th NBPC eigenvector, for varying m; second subsample.

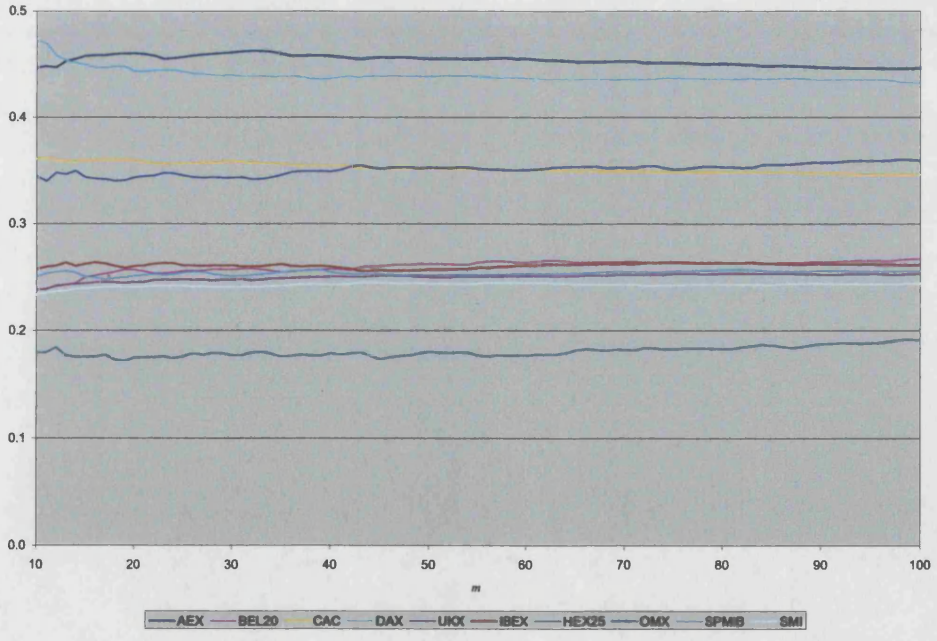


Figure 5.9: First NBPC eigenvector for varying m; full sample.

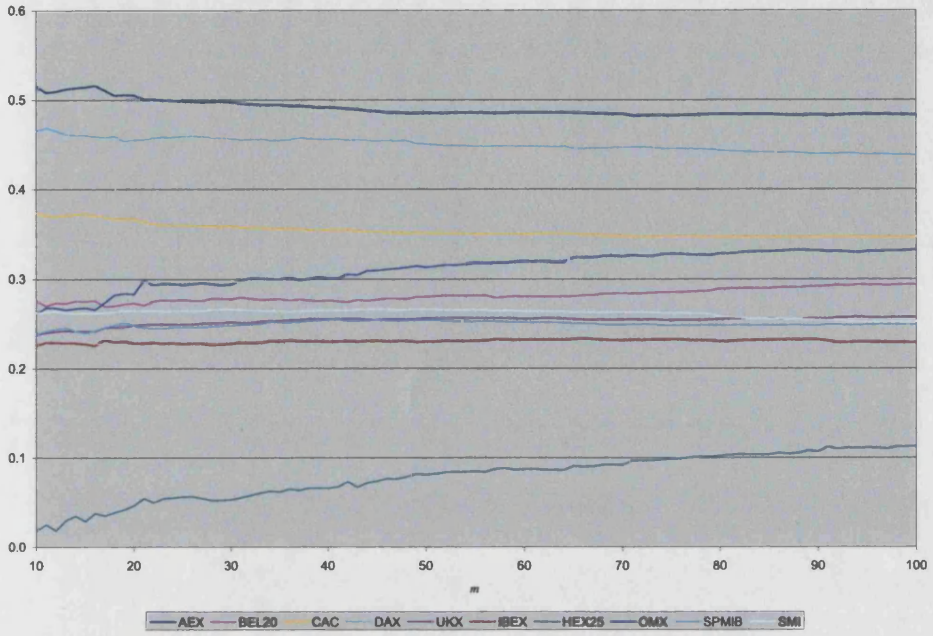


Figure 5.10: First NBPC eigenvector for varying m; first subsample.

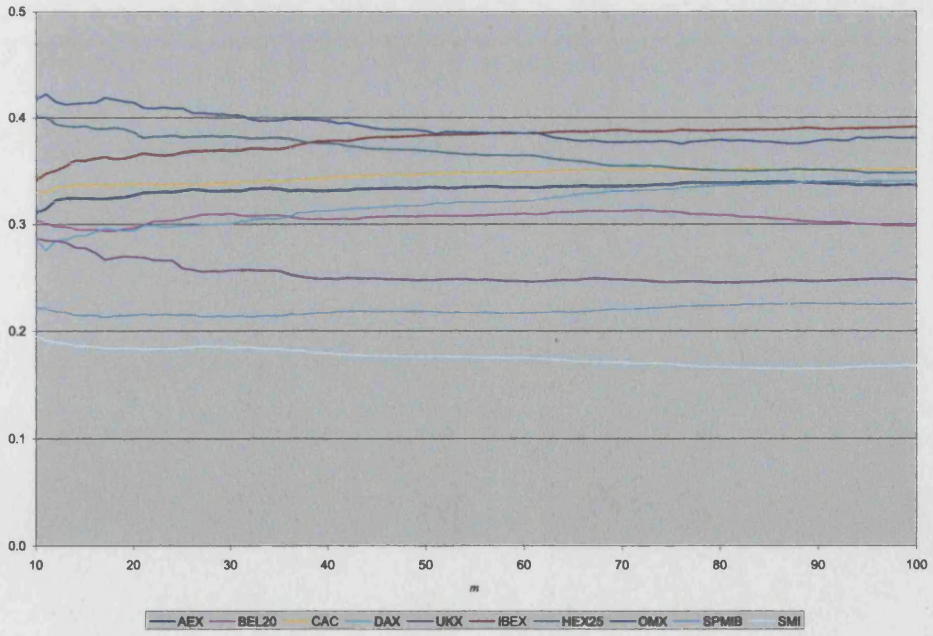


Figure 5.11: First NBPC eigenvector for varying m; second subsample.

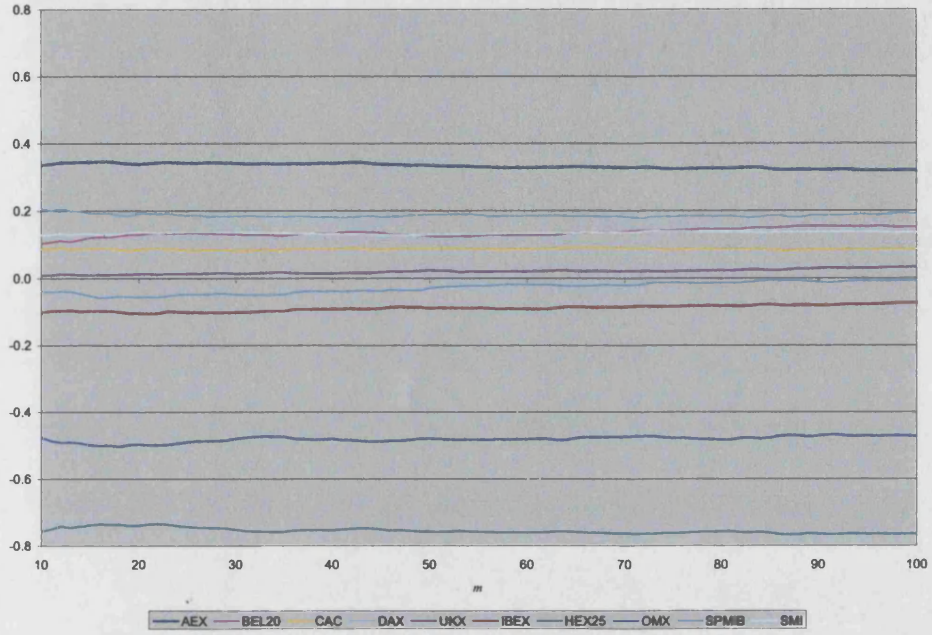


Figure 5.12: Second NBPC eigenvector for varying m; full sample.

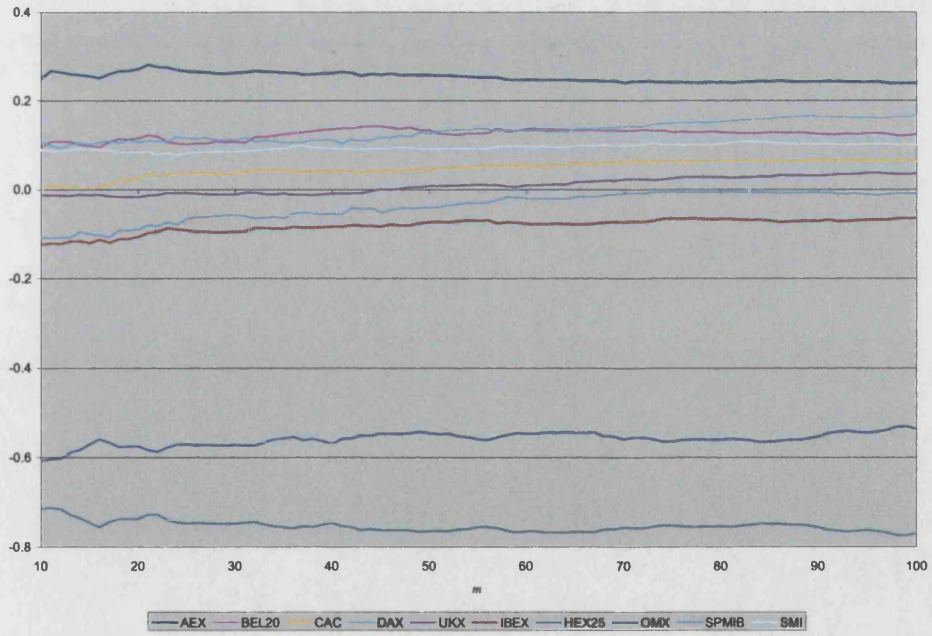


Figure 5.13: Second NBPC eigenvector for varying m; first subsample.

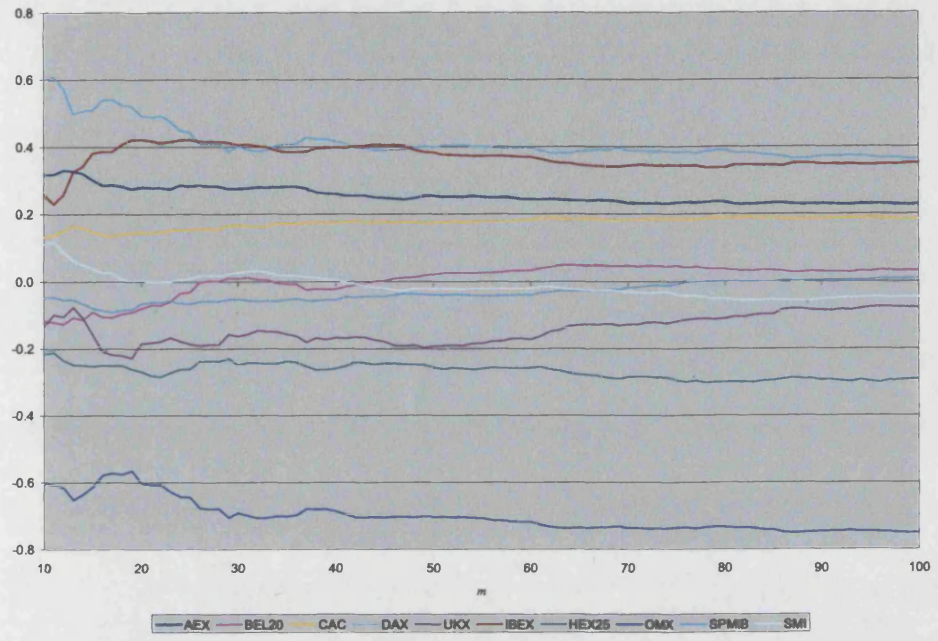


Figure 5.14: Second NBPC eigenvector for varying m; second subsample.

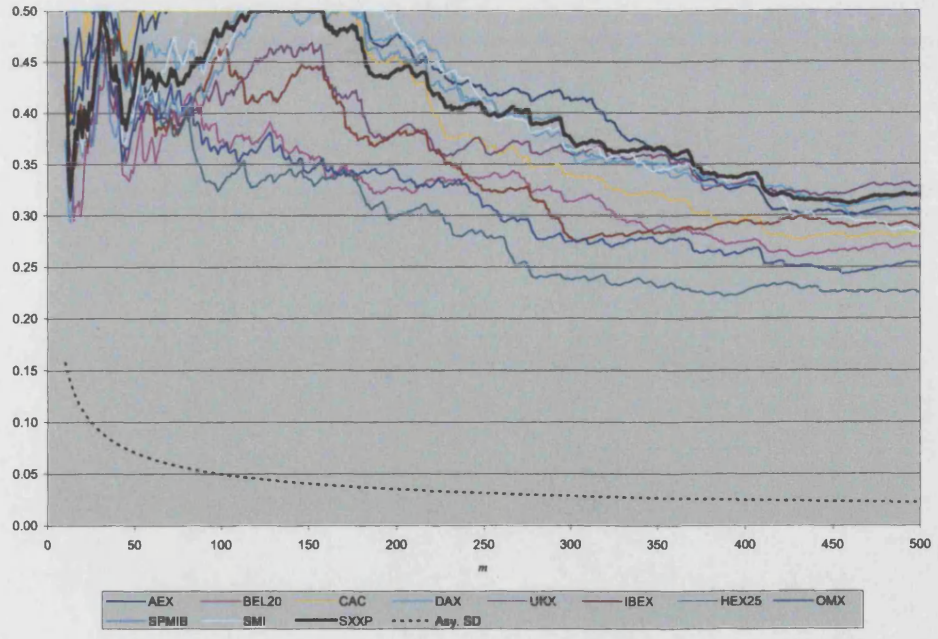


Figure 5.15: Local Whittle memory estimates of squared returns; full sample.

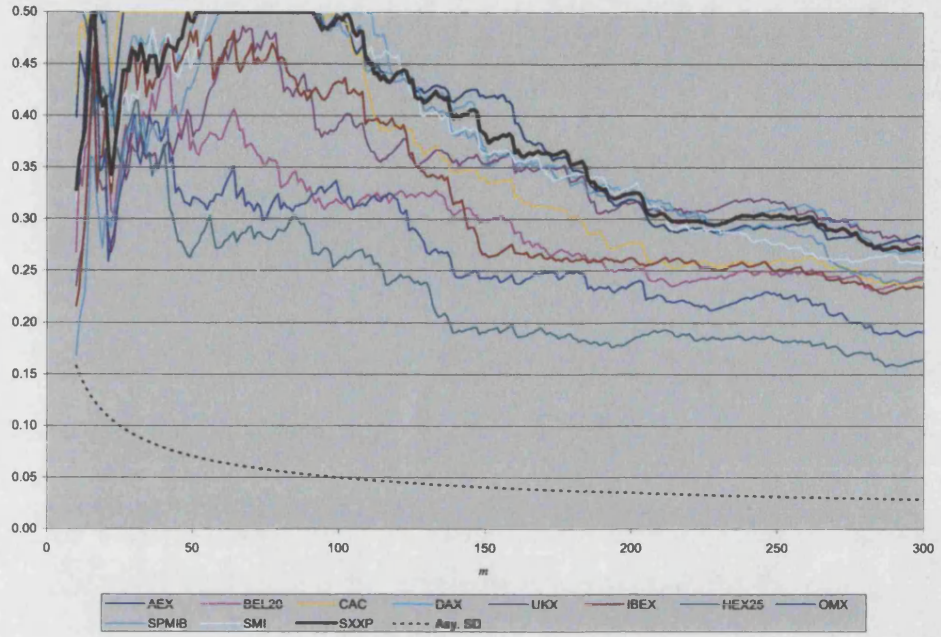


Figure 5.16: Local Whittle memory estimates of squared returns; first subsample.

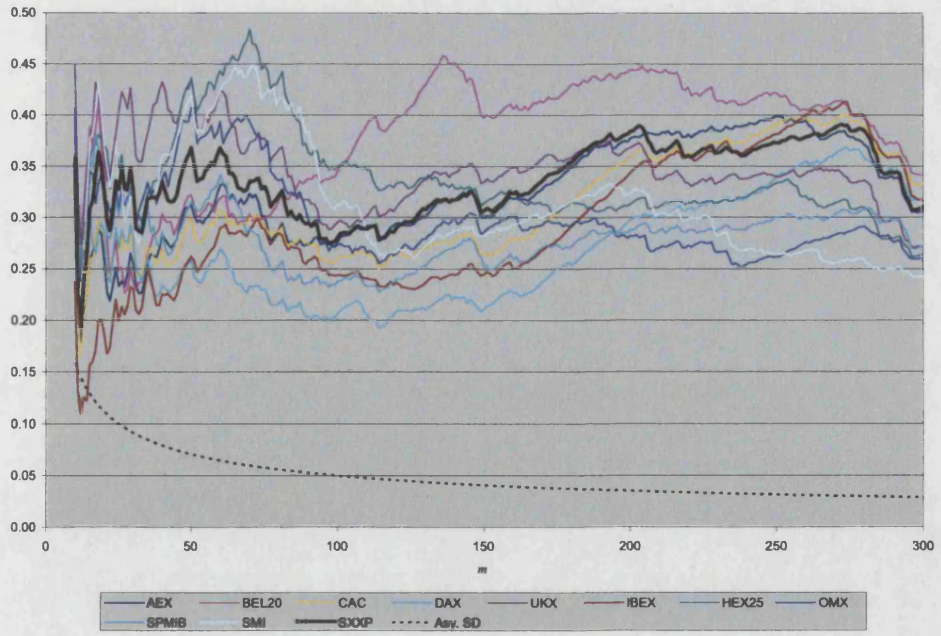


Figure 5.17: Local Whittle memory estimates of squared returns; second subsample.

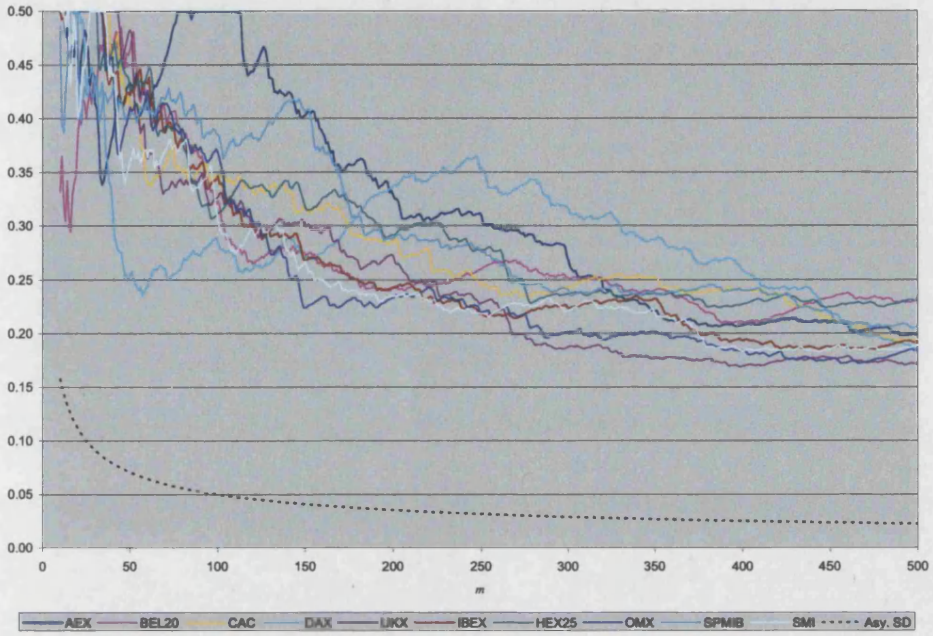
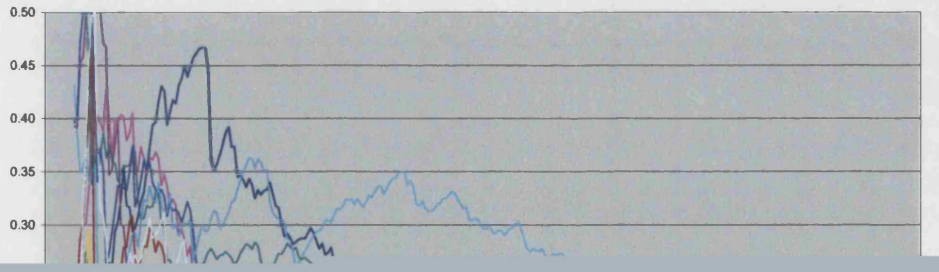


Figure 5.18: Local Whittle memory estimates of OLS residuals; full sample.



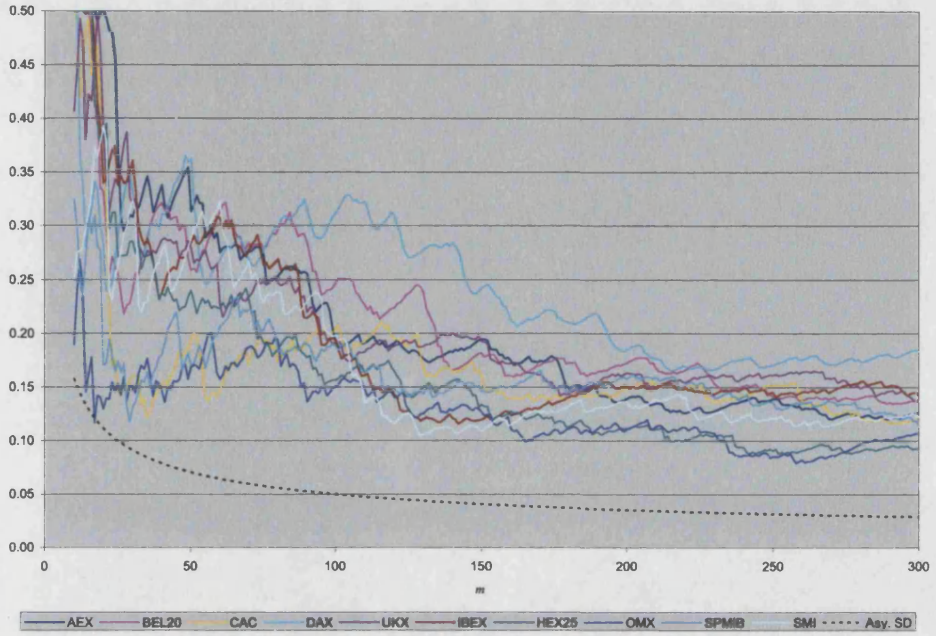


Figure 5.20: Local Whittle memory estimates of OLS residuals; second subsample.

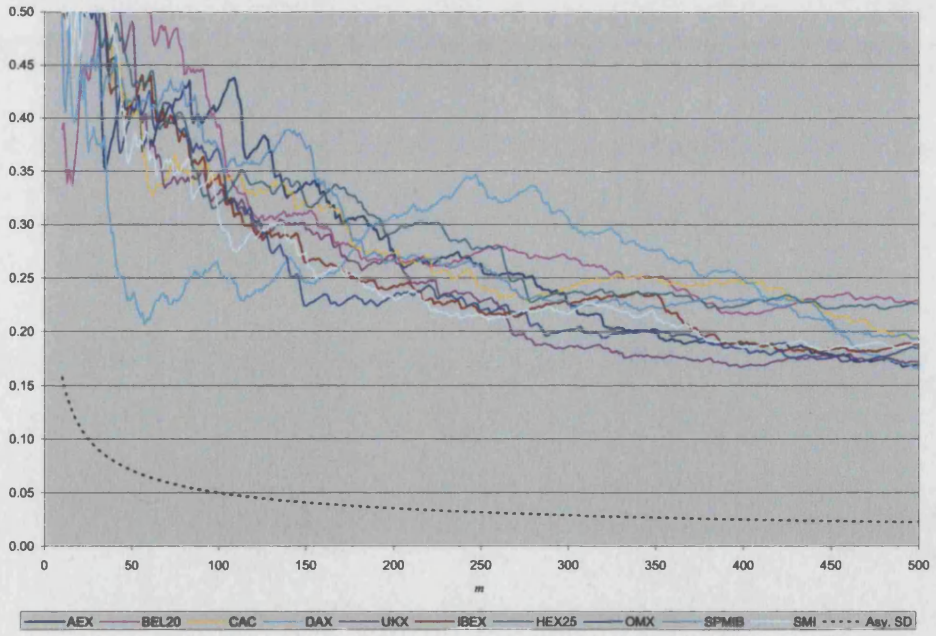


Figure 5.21: Local Whittle memory estimates of NBLS residuals; full sample.

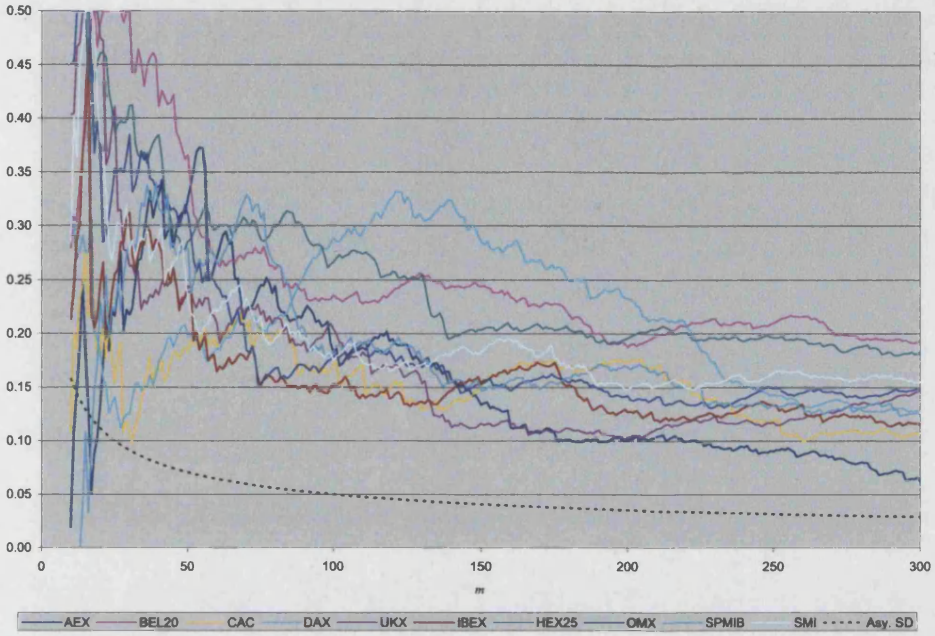
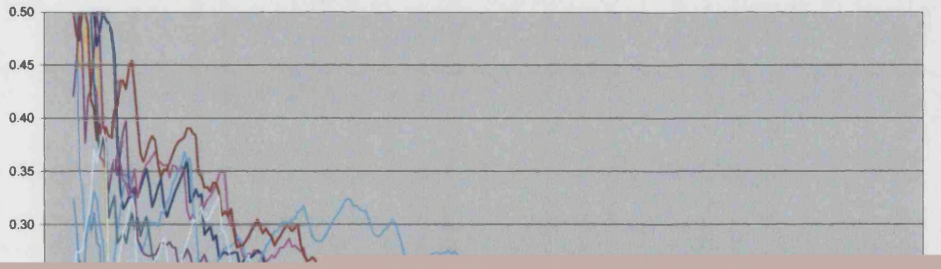


Figure 5.22: Local Whittle memory estimates of NBLS residuals; first subsample.



Bibliography

ADENSTEDT, R. K. (1974): "On large-sample estimation for the mean of a stationary random sequence," *The Annals of Statistics*, 2(6), 1095–1107.

ANDERSEN, T. G., AND T. BOLLERSLEV (1997): "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4(2), 115–158.

ANDERSON, T. W. (1963): "Asymptotic theory for principal component analysis," *The Annals of Mathematical Statistics*, 34(1), 122–148.

ANDERSON, T. W., AND H. RUBIN (1956): "Statistical inference in factor analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, vol. 5, pp. 111–150, Berkeley and Los Angeles. University of California Press.

ANDREWS, D. W. K., AND P. GUGGENBERGER (2003): "A bias-reduced log-periodogram regression estimator for the long-memory parameter," *Econometrica*, 71(2), 675–712.

ANDREWS, D. W. K., AND Y. SUN (2004): "Adaptive local polynomial Whittle estimation of long-range dependence," *Econometrica*, 72(2), 569–614.

ARTECHE, J. (2004): "Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models," *Journal of Econometrics*, 119(1), 131–154.

BAILLIE, R. T., T. BOLLERSLEV, AND H. O. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74(1), 3–30.

BANDI, F. M., AND B. PERRON (2006): "Long memory and the relation between implied and realized volatility," *Journal of Financial Econometrics*, 4(4), 636–670.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31(3), 302–327.

BOLLERSLEV, T., AND H. O. MIKKELSEN (1996): "Modeling and pricing long memory in stock market volatility," *Journal of Econometrics*, 73(1), 151–184.

BREIDT, F. J., N. CRATO, AND P. DE LIMA (1998): "The detection and estimation of long memory in stochastic volatility," *Journal of Econometrics*, 83(1-2), 325–348.

BREUER, P., AND P. MAJOR (1983): "Central limit theorems for non-linear functionals of Gaussian fields," *Journal of Multivariate Analysis*, 13(3), 425–441.

BRUNETTI, C., AND C. L. GILBERT (2000): "Bivariate FIGARCH and fractional cointegration," *Journal of Empirical Finance*, 7(5), 509–530.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets," *Econometrica*, 51(5), 1281–1304.

CHAMBERS, M. J. (1998): "Long memory and aggregation in macroeconomic time series," *International Economic Review*, 39(4), 1053–1072.

CHEN, W. W., AND C. M. HURVICH (2003): "Semiparametric estimation of multivariate fractional cointegration," *Journal of the American Statistical Association*, 98(463), 629–642.

——— (2006): "Semiparametric estimation of fractional cointegrating subspaces," *The Annals of Statistics*, 34(6), 2939–2979.

CHRISTENSEN, B. J., AND M. Ø. NIELSEN (2006): "Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting," *Journal of Econometrics*, 133(1), 343–371.

CONNOR, G. C., AND R. A. KORAJCZYK (1986): "Performance measurement with the Arbitrage Pricing Theory: A new framework for analysis," *Journal of Financial Economics*, 15(3), 373–394.

DALLA, V., L. GIRAITIS, AND J. HIDALGO (2006): "Consistent estimation of the memory parameter for nonlinear times series," *Journal of Time Series Analysis*, 27(2), 211–251.

DAVIES, R. B., AND D. S. HARTE (1987): "Tests for Hurst effect," *Biometrika*, 74(1), 95–101.

DELGADO, M. A., AND P. M. ROBINSON (1996): "Optimal spectral bandwidth for long memory," *Statistica Sinica*, 6(1), 97–112.

DEO, R. S., AND C. M. HURVICH (2001): "On the log-periodogram regression estimator of the memory parameter in long memory stochastic volatility models," *Econometric Theory*, 17(4), 686–710.

DING, Z., AND C. W. J. GRANGER (1996): "Modeling volatility persistence of speculative returns: A new approach," *Journal of Econometrics*, 73(1), 185–215.

DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1(1), 83–106.

DOBRUSHIN, R. L., AND P. MAJOR (1979): "Non-central limit theorems for non-linear functionals of Gaussian fields," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 50(1), 27–52.

ENGLE, R. F. (1974): "Band spectrum regression," *International Economic Review*, 15(1), 1–11.

——— (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50(4), 987–1008.

ENGLE, R. F., AND T. BOLLERSLEV (1986): "Modelling the persistence of conditional variances," *Econometric Reviews*, 5(1), 1–50, 81–87.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, 55(2), 251–276.

GEWEKE, J., AND S. PORTER-HUDAK (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4, 221–238.

GIRAITIS, L., AND M. S. TAQQU (1997): "Limit theorems for bivariate Appell polynomials. Part I: Central limit theorems," *Probability Theory and Related Fields*, 107(3), 359–381.

GIRAITIS, L., M. S. TAQQU, AND N. TERRIN (1998): "Limit theorems for bivariate Appell polynomials. Part II: Non-central limit theorems," *Probability Theory and Related Fields*, 110(3), 333–367.

GRANGER, C. W. J. (1966): "The typical spectral shape of an economic variable," *Econometrica*, 34(1), 150–161.

——— (1981): "Some properties of time series data and their use in econometric model specification," *Journal of Econometrics*, 16(1), 121–130.

GRANGER, C. W. J., AND R. JOYEUX (1980): "An introduction to long-memory time series and fractional differencing," *Journal of Time Series Analysis*, 1, 15–30.

GRANGER, C. W. J., AND A. A. WEISS (1983): "Time series analysis of error-correcting models," in *Studies in Econometrics, Time Series, and Multivariate Statistics,*, ed. by S. Karlin, T. Amemiya, and L. A. Goodman, pp. 255–278. Academic Press, New York.

GUGGENBERGER, P., AND Y. SUN (2006): "Bias-reduced log-periodogram and Whittle estimation of the long-memory parameter without variance inflation," *Econometric Theory*, 22(5), 863–912.

HANNAN, E. J. (1963): "Regression for time series," in *Time series analysis*, ed. by M. Rosenblatt, pp. 17–77. Wiley, New York.

HARVEY, A. C. (1998): "Long memory in stochastic volatility," in *Forecasting volatility in financial markets*, ed. by J. L. Knight, and S. E. Satchell, Quantitative Finance Series, chap. 12, pp. 307–320. Butterworth-Heinemann, London.

HIDALGO, J., AND P. M. ROBINSON (2002): "Adapting to unknown disturbance autocorrelation in regression with long memory," *Econometrica*, 20(4), 1545–1581.

HO, H.-C., AND T.-C. SUN (1987): "A central limit theorem for non-instantaneous filters of a stationary Gaussian process," *Journal of Multivariate Analysis*, 22(1), 144–155.

HOSKING, J. R. M. (1981): "Fractional differencing," *Biometrika*, 68(1), 165–176.

HUALDE, J., AND P. M. ROBINSON (2007): "Root-n-consistent estimation of weak fractional cointegration," *Journal of Econometrics*, 140(2), 450–484.

HURST, H. E. (1951): "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, 116, 770–779.

HURVICH, C. M., R. DEO, AND J. BRODSKY (1998): "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series," *Journal of Time Series Analysis*, 19(1), 19–46.

HURVICH, C. M., E. MOULINES, AND P. SOULIER (2005): "Estimating long memory in volatility," *Econometrica*, 73(4), 1283–1328.

HURVICH, C. M., AND B. K. RAY (2003): "The Local Whittle estimator of long-memory stochastic volatility," *Journal of Financial Econometrics*, 1(3), 445–470.

HURVICH, C. M., AND P. SOULIER (2002): "Testing for long memory in volatility," *Econometric Theory*, 18(6), 1291–1308.

JANSEN, D. W., AND C. G. DE VRIES (1991): "On the frequency of large stock returns: putting booms and busts into perspective," *Review of Economics and Statistics*, 73(1), 18–24.

JOHANSEN, S. (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12(2-3), 231–254.

——— (1991): "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica*, 59(6), 1551–1580.

KÜNSCH, H. R. (1986): "Discrimination between monotonic trends and long-range dependence," *Journal of Applied Probability*, 23(4), 1025–1030.

——— (1987): "Statistical aspects of self-similar processes," in *Proceedings of the first world congress of the Bernoulli Society*, ed. by Y. Prohorov, and V. Sazonov, vol. 1, pp. 67–75, Utrecht. VNU Science Press.

LOBATO, I. N. (1997): "Consistency of the averaged cross-periodogram in long memory series," *Journal of Time Series Analysis*, 18(2), 137–155.

——— (1999): "A semiparametric two-step estimator in a multivariate long memory model," *Journal of Econometrics*, 90(1), 129–153.

LOBATO, I. N., AND P. M. ROBINSON (1996): "Averaged periodogram estimation of long memory," *Journal of Econometrics*, 73(1), 303–324.

LORETAN, M., AND P. C. B. PHILLIPS (1994): "Testing the covariance stationarity of heavy-tailed time series," *Journal of Empirical Finance*, 1(2), 211–248.

MANDELBROT, B. B. (1969): "Long-run linearity, locally Gaussian process, H-spectra and infinite variances," *International Economic Review*, 10(1), 82–111.

MANDELBROT, B. B., AND J. W. VAN NESS (1968): "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, 10(4), 422–437.

MARINUCCI, D., AND P. M. ROBINSON (2001): "Semiparametric fractional cointegration analysis," *Journal of Econometrics*, 105(1), 225–247.

MERTON, R. C. (1973): "An intertemporal capital asset pricing model," *Econometrica*, 41(5), 867–887.

MORANA, C. (2004): "Frequency domain principal components estimation of fractionally cointegrated processes," *Applied Economics Letters*, 11(13), 837–842.

NIELSEN, M. Ø. (2005): "Semiparametric estimation in time-series regression with long-range dependence," *Journal of Time Series Analysis*, 26(2), 279–304.

NIELSEN, M. Ø., AND P. H. FREDERIKSEN (2005): "Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration," *Econometric Reviews*, 24(4), 405–443.

PHILLIPS, P. C. B. (1991): "Optimal inference in cointegrated systems," *Econometrica*, 59(2), 283–306.

PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple time series regression with integrated processes," *Review of Economic Studies*, 53(4), 473–495.

PHILLIPS, P. C. B., AND S. OULIARIS (1988): "Testing for cointegration using principal components methods," *Journal of Economic Dynamics and Control*, 12(2-3), 205–230.

ROBINSON, P. M. (1991): "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression," *Journal of Econometrics*, 47(1), 67–84.

——— (1994a): "Semiparametric analysis of long-memory time series," *The Annals of Statistics*, 22(1), 515–539.

——— (1994b): "Rates of convergence and optimal spectral bandwidth for long range dependence," *Probability Theory and Related Fields*, 99, 443–473.

——— (1995a): "Log-periodogram regression of time series with long range dependence," *The Annals of Statistics*, 23(3), 1048–1072.

——— (1995b): "Gaussian semiparametric estimation of long range dependence," *The Annals of Statistics*, 23(5), 1630–1661.

——— (2001): "The memory of stochastic volatility models," *Journal of Econometrics*, 101(2), 195–218.

ROBINSON, P. M., AND M. HENRY (1999): "Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels," *Econometric Theory*, 15(3), 299–336.

——— (2003): "Higher-order kernel semiparametric M-estimation of long memory," *Journal of Econometrics*, 114(1), 1–27.

ROBINSON, P. M., AND J. HIDALGO (1997): "Time series regression with long-range dependence," *The Annals of Statistics*, 25(1), 77–104.

ROBINSON, P. M., AND D. MARINUCCI (2001): "Narrow band analysis of nonstationary processes," *The Annals of Statistics*, 29(4), 947–986.

——— (2003): "Semiparametric frequency domain analysis of fractional cointegration," in *Time series with long memory*, ed. by P. M. Robinson, Advanced Texts in Econometrics, chap. 14, pp. 334–373. Oxford University Press, Oxford.

ROBINSON, P. M., AND Y. YAJIMA (2002): "Determination of cointegrating rank in fractional systems," *Journal of Econometrics*, 106(2), 217–241.

ROSS, S. A. (1976): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13(3), 341–360.

SHARPE, W. F. (1964): "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Finance*, 19(3), 425–442.

SILVERMAN, B. W. (1986): *Density estimation for statistics and data analysis*, vol. 26 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

SLEPIAN, D. (1972): "On the symmetrized Kronecker power of a matrix and extensions of Mehler's formula for Hermite polynomials," *SIAM Journal on Mathematical Analysis*, 3, 606–616.

SMITH, R. L., AND F.-L. CHEN (1996): "Regression in long-memory time series," in *Athens Conference on Applied Probability and Time Series, Volume II: Time Series Analysis in Memory of E.J. Hannan*, ed. by P. M. Robinson, and M. Rosenblatt, vol. 115 of *Lecture Notes in Statistics*, pp. 378–391, New York. Springer.

STOCK, J. H. (1987): "Asymptotic properties of least squares estimators of cointegrating vectors," *Econometrica*, 55(5), 1035–1056.

STOCK, J. H., AND M. W. WATSON (1988): "Testing for common trends," *Journal of the American Statistical Association*, 83(404), 1097–1107.

SUN, Y., AND P. C. B. PHILLIPS (2003): "Nonlinear log-periodogram regression for perturbed fractional processes," *Journal of Econometrics*, 115(2), 355–389.

TAQQU, M. S. (1975): "Weak convergence to fractional Brownian motion and to the Rosenblatt process," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 31(4), 287–302.

——— (1979): "Convergence of integrated processes of arbitrary Hermite rank," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 50(1), 53–83.

TAYLOR, S. J. (1986): *Modelling financial time series*. John Wiley and Sons, Chichester.

VELASCO, C. (2000): "Non-Gaussian log-periodogram regression," *Econometric Theory*, 16(1), 44–79.

——— (2003): "Gaussian semi-parametric estimation of fractional cointegration," *Journal of Time Series Analysis*, 24(3), 345–378.

WHISTLER, D. E. N. (1990): "Semiparametric models of daily and intradaily exchange rate volatility," PhD thesis, University of London, London School of Economics and Political Science, Department of Economics.

YONG, C. H. (1974): *Asymptotic behaviour of trigonometric series.* Chinese University of Hong Kong, Hong Kong.

ZYGMUND, A. (1977): *Trigonometric series.* Cambridge University Press, Cambridge.