The London School of Economics and Political Science

Whittle Estimation of Multivariate Exponential Volatility Models

A thesis submitted to the Department of Statistics of the
London School of Economics
for the degree of
Doctor of Philosophy

by
Malvina Marchese
March 18, 2015

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Table of Contents

List of Symbols

The following notation and conventions are used through this thesis. The symbol "$\equiv$" stands for "by definition". All limits are taken for either $n \to \infty$ or $T \to \infty$, unless specified otherwise.

$R \equiv$ the field of real numbers.

$R^s \equiv$ the Euclidean p-space.

$\Pi \in [-\pi, \pi)$.

$Z \equiv$ the ring of integers.

$a.s \equiv$ almost sure.

$a.e \equiv$ almost everywhere.

$|a| \equiv$ absolute value of a if referred to a number.

$\det \mathbf{A} \equiv$ determinant of a matrix A.

$\|\mathbf{A}\| \equiv$ euclidean norm.

$\|g\|_\infty \equiv$ the supremum norm over the domain of a function g.

$diag\,(a,\ b,\ c) \equiv$ diagonal matrix with elements a, b, c.

$\to^p \equiv$ convergence in probability.

$\to^d \equiv$ convergence in distribution.

$\to^{a.s} \equiv$ convergence almost surely.

$(n, T \to \infty)_{\text{seq}} \equiv$ sequential limit of T first, followed by n.

$CLT \equiv$ central limit theorem.

$FCLT \equiv$ functional central limit theorem.

$LLN \equiv$ law of large numbers.

$C_r\ inq \equiv$ the c-r inequality.

$1\,(A) \equiv$ indicator of set A.

$I_t \equiv$ identity matrix of dimension $n \times n$.

$iid \equiv$ independent identically distributed.

$ind\ s \equiv$ independent random sequence.

$l.h.s \equiv$ left hand side.

$r.h.s \equiv$ right hand side.

$O_p(1) \equiv$ random sequence bounded in probability.

$o_p(1) \equiv$ random sequence converging to zero in probability.

$sgn \equiv$ sign function.

$[x] \equiv$ largest integer not greater than x.

$wrt \equiv$ with respect to.

$rv \equiv$ random variable.

$BM(\Omega) \equiv$ Brownian motion with covariance matrix Ω.

$BM(I) \equiv$ standardized Brownian motion.

$K_{abcd} \equiv$ fourth order cumulant of random variables a, b, c, d.

$> \equiv$ positive definiteness if applied to a matrix.

$\geq \equiv$ positive semi definiteness if applied to matrix.

$i \equiv (-1)^{1/2}$.

$K \equiv$ finite constant not always the same.

$\displaystyle\int W dW \equiv \int_0^1 W(r) dW(r).$

Abstract

The aim of this thesis is to offer some insights into two topics of some interest for time-series econometric research.

The first chapter derives the rates of convergence and the asymptotic normality of the pooled OLS estimators for linear regression panel models with mixed stationary and non-stationary regressors. This work is prompted by the consideration that many economic models of interest present a mixture of I(1) and I(0) regressors, for example models for analysis of demand system or for assessment of the relationship between growth and inequality. We present results for a model where the regressors and the regressand are cointegrated. We find that the OLS estimator is asymptotically normal with convergence rates $T\sqrt{n}$ and $\sqrt{nT}$ for respectively the non-stationary and the stationary regressors. Phillips and Moon (1990) show that in a cointegrated regression model with non-stationary regressors, the OLS estimator converges at a rate of $T\sqrt{n}$. We find that the presence of one stationary regressor in the model does not increases the rate of convergence. All the results are derived for sequential limits, with T going to infinity followed by $n,$ and under quite restrictive regularity conditions.

Chapters 3-5 focus on parametric multivariate exponential volatility models. It has long been recognized that the volatility of stock returns responds differently to good news and bad news. In particular, while negative shocks tend to increase future volatility, positive ones of the same size will increase it by less or even decrease it. This was in fact one of the chief motivations that led Nelson (1991) to introduce the univariate EGARCH model. More recently empirical studies have found that the asymmetry is a robust feature of multivariate stock returns series as well, and several multivariate volatility models have been developed to capture it. Another important property that characterizes the dynamic evolution of volatilities is that squared returns have significant au-

tocorrelations that decay to zero at a slow rate, consistent with the notion of long memory, where the autocovariances are not absolutely summable. Univariate long-memory volatility models have received a great deal of attention. However, the generalization to a multivariate long-memory volatility model has not been attempted in the literature. Chapter 3 offers a detailed literature review on multivariate volatility models. Chapter 4 and 5 introduce a new multivariate exponential volatility (MEV) model which captures long-range dependence in the volatilities, while retaining the martingale difference assumption and short-memory dependence in mean. Moreover the model captures cross-assets spillover effects, leverage and asymmetry. The strong consistency and the asymptotic normality of the Whittle estimator of the parameters in the Multivariate Exponential Volatility model is established under a variety of parameterization. The results cover both the case of exponentially and hyperbolically decaying coefficients, allowing for different degrees of persistence of shocks to the conditional variances. It is shown that the rate of convergence and the asymptotic normality of the Whittle estimates do not depend on the degree of persistence implied by the parameterization as the Whittle function automatically compensates for the possible lack of square integrability of the model spectral density.

Chapter 1 Introduction

This thesis is the outcome of my Ph.D. studies at the London School of Economics and it reflects the evolution of my research interests during the past few years. The second chapter was written during my first year as a research student and it focuses on large panel data models. The research question was prompted by the lack of an existing asymptotic theory for the pooled OLS estimator in large panel regression models with stationary and non-stationary regressors. At the time I started my research, only one paper in the literature addressed the asymptotics of the estimator in large, possibly non-stationary, panels. The paper, by Baltagi, Kao and Liu (Econometrics Journal, 2008), considers a simple one regressor error-correction model where the error and the regressor are both generated by possibly non-stationary ARMA processes. These results left room to explore models where the mixture of stationarity and non-stationarity arises from a combination of stationary and non-stationary regressors. In applied research such setting appears quite relevant. Many panel data models, for example for the assessment of the relationship between growth and inequality, have a mixture of integrated and stationary regressors. Chapter two derives the rates of convergence and the asymptotic normality of the pooled Ordinary Least Square estimator in a simple scalar model with mixed regressors. The error term is assumed stationary, indicating the existence of a cointegrating relation between the regressors and the regressand. The results are obtained under quite strong conditions, some of which are too restrictive. The estimators turn out to be asymptotically normal with convergence rates $T\sqrt{n}$ and $\sqrt{nT}$ for respectively the non-stationary and the stationary regressors. This is not unexpected. As already shown by Phillips and Moon (1999) in a linear panel regression model with non-stationary regressors under a variety of cointegrating relationship, the OLS estimator is $T\sqrt{n}$ consistent and asymptotically normal. We find that the presence of an additional stationary regressor does not alter the convergence rates of the non-stationary regressors. The results provided are for sequential limits, with T going to infinity followed by n.

During the second year of my studies I developed a strong interest in modelling multivariate volatilities. The importance of modelling comovements of financial returns is well established in the literature. The knowledge of correlation structures is vital in many financial applications, including asset pricing, optimal portfolio allocation and risk management. Moreover, as the volatilities of different assets and markets move together, modelling volatility in a multivariate framework can lead to greater statistical efficiency. Starting from the Vector Autoregression model of Bollerslev, Engle and Wooldridge (1988), several multivariate conditional volatility models have been proposed in the literature and used extensively in applied work. Over the last few years, the literature on multivariate stochastic volatility models has also developed significantly, thanks to the availability of many new numerical estimation methods. Recently empirical studies found robust evidences of asymmetric response of volatilities to positive and negative returns in multivariate asset models. A number of conditional and stochastic volatility models have been proposed to capture this inherent characteristic of volatility in a multivariate context, such as the QARCH latent factor model of Sentana (1995), the MSV-Leverage model of Asai and McAleer (2005), the asymmetric dynamic covariance (ADC) model of Kroner and Ng (1998), the Matrix Exponential GARCH model of Kawakatsu (2006), and others.

Another important property that characterizes the dynamic evolution of volatilities is that power transformations of absolute returns have significant autocorrelations that decay to zero at a slow rate. Many authors have argued that the slow decay of the autocorrelations of squared returns is consistent with the notion of long-memory, where the autocovariances are not absolutely summable. Univariate long-memory volatility models, such as the FIEGARCH model of Bollerslev and Mikklesen (1996), the nonlinear moving average model of Robinson and Zaffaroni (1996,

1997) and the long-memory stochastic volatility (LMSV) model of Ruiz and Veiga (2006), have received a great deal of attention. However, to my knowledge, no generalization to a multivariate long-memory volatility model has been attempted in the literature. The aim of this thesis is to fill in this gap. Chapter 4 introduces a new multivariate Exponential Volatility (MEV) model, which captures long-range dependence in certain nonlinear functions of the data, such as squares, while retaining the martingale difference assumption and short-memory dependence in the level. The multivariate Exponential Volatility model is an extension of the univariate exponential volatility model of Zaffaroni (2009). It nests "one-shock" conditional variance specifications and "two-shocks" stochastic volatility specifications. It captures cross-assets spillover effects, leverage and asymmetry. The choice of an exponential specification offers several advantages, the most relevant of which is that no further restriction is required to grant positive definiteness of the covariance matrix. Estimation of the MEV model by maximum likelihood methods is possible, however we advocate the use of the frequency domain Gaussian estimator in the sense of Whittle (1962). MLE estimation of nonlinear exponential models is computationally costly, and possibly unstable. Moreover its asymptotic properties depend on the invertibility of the model, which is not easy to establish in exponential models (see Straumann and Mikosh, 2006). These difficulties do not apply to the Whittle estimator, partly due to its frequency domain specification. We follow Harvey et al. (1994) and estimate a logarithmic transformation of the squared returns of the observations. The estimated model turns out to be a vector signal plus noise model, where the signal evolves according to an infinite order moving average process and the noise is an $i.i.d$ shock. The dependence structure of the MEV model implies that the signal and the noise might be correlated. Statistical literature on Whittle estimation of signal plus noise models requires at least incoherent signal and noise. In fact all the available results deal with linearly regular signal plus noise processes with parameters that can be estimated directly on the factored representation of the process spectral density (see Dunsmuir, 1979, and Hosoya and Taniguchi, 1982). Such results do not readily apply to correlated signal plus noise processes, even when the processes are linearly regular. In chapter 4, we establish the strong consistency and asymptotic normality of the Whittle estimator when the signal coefficients have an exponential decay rate, following Robinson (1978). In chapter 5, we establish the properties of the estimator when the signal coefficients have an hyperbolic decay rate that imparts long-range dependence in the squares of the MEV model, relying the on the central limit theorem for the integrated weighted periodogram of Giraitis and Taqqu (1999). As expected, the asymptotic properties of the estimator do not depend on the degree of persistence of shocks to the conditional variances, thanks to a convenient feature of the Whittle function that allows to compensate for the possible lack of square integrability of the spectral density. The results are established under the conditions of strict stationarity, ergodicity and finite fourth moments and either absolute summability of the autocovariance function or standard smoothness assumptions for the spectral density and its higher order derivatives.

The remainder of this thesis is organized as follows. Chapter 2 presents asymptotics results on the pooled OLS estimator in panel models with mixed stationary and non-stationary regressors. Chapter 3 offers a detailed literature review of multivariate volatility models, with some discussion on the most relevant multivariate exponential model, i.e. the matrix exponential GARCH model of Kawakatsu. Chapter 4 introduces the multivariate Exponential Volatility model, discusses its estimation and establishes the asymptotic properties of the estimator under fairly general conditions suitable for both “one-shock” and “two-shocks” specification of the model. Chapter 5 extends these results to the long memory MEV model.

Chapter 2 Large-panel models with stationary and non-stationary regressors

2.1 Introduction

The advantages of panel data over cross section and time series data have long been established in econometric research. Panel data sets usually give the researcher a larger number of data points than conventional cross section and time series data, thus increasing the degrees of freedom and reducing multicollinearity among explanatory variables. This results in more reliable parameter estimates and most importantly enables the researcher to specify and test more sophisticated models with less restrictive assumptions.

A panel data set contains observations on a given number of individuals (n) across time (T). As such, it is a double indexed process and any treatment of the asymptotics must take this into account. The initial focus of research has been on identifying and estimating effects from panel models with a large number of cross section and few time series observations. The asymptotics for the standard panel estimators in this setting are well established since the work of Hsiao (1986) and Chamberlain (1984). However, starting from the Nineties, empirical work has used panel data sets with a large number of time series and cross section observations. Examples of this literature range from testing growth convergence theories in macroeconomics to estimating long run relations between international financial series. These works have been enhanced and facilitated by the availability of a number of important data sets covering different individuals, regions and countries over a relatively long period of time, such as the Penn World Table. In a context of few time-series observations, the non-stationarity of the series cannot be addressed properly but with larger data sets it must be explored properly. When the time-series component of the model is assumed non-stationary, traditional limit theory is no longer valid. Phillips and Moon (1999) investigated regressions with panel data where the time series component is an integrated process of order one. Under a variety of cointegrating relations between the regressors and the regressand, they derive the consistency and asymptotic normality of the pooled OLS estimator. Their framework does not allow for the presence of stationary and non-stationary regressors in the same model. In practise however this framework is very relevant since many empirical models have a mixture of stationary and non-stationary regressors. For example demand systems models, where budget shares are regressed on relative prices and real income for different countries over time. Typically some prices are stationary and some other are trending. Money demand equations offer a similar mixture of stationary and non-stationary variables, with real income trending over time for most countries but stationary interest rates. A comprehensive limit theory for the pooled OLS estimator in this framework has not been established in the literature. Baltagi Kao and Liu (2008) derive the asymptotic properties of the most common panel estimators in a simple error-correction model, where the regressor and the remainder disturbance term are possibly non-stationary. Their results allow for a mixture of stationary and non-stationary terms in the same model, however they do not allow for the simultaneous presence of stationary and a non-stationary regressors. In this chapter we discuss the asymptotic properties of the OLS estimator in a linear panel regression model with mixed stationary and non-stationary regressors, as T and n increase to infinity sequentially. The chapter is organized as follows. Section 2.2 presents some literature review. Section 2.3 introduces the model and the assumptions, with discussion. Section 2.4 discusses the main results, namely the asymptotic normality and the convergence rates of the estimator in a cointegrated model. All proofs are contained in appendix A. Notation is fairly standard. The symbol "$\rightarrow_{a.s}$" signifies convergence almost surely, "$\Rightarrow$" denotes weak convergence. The inequality "$>$" signifies positive definiteness when applied to matrices, $||\Omega||$ is the Euclidean norm of the matrix Ω and $|K|$ is the absolute value of the scalar K, the symbol "$[.]$" denotes the largest integer part. "$(n, T \rightarrow \infty)_{\text{seq}}$" denotes sequential limits where T goes to infinity followed by n. Brownian motions on $[0, 1]$ are usually written as BM and stochastic integrals $\int_0^1 W(r)dW(r)$ are denoted as $\int WdW$.

2.2 Literature review

Since the beginning of the Nineties there has been much research on non-stationary panel data. Quah (1994), Levin and Lin (1993) considered unit root time-series regressions with non-stationary panel data and proposed a test statistic for unit roots. Pedroni (1995) studied some properties of cointegration statistics in pooled time-series panel models, Baltagi and Kramer (1997) and Kao and Emerson (2004) investigated the case of a panel time trend model. Pesaran and Smith (1995) examined the impact of non-stationary variables on cross-section regression estimates with a large number of groups and of time periods. Phillips and Moon (1999) generalized the results of Pesaran and Smith providing a fundamental framework for asymptotics of the OLS estimator in large non-stationary panels. Phillips and Moon investigate the behavior of the estimator in non-stationary panel models in the cases of no time series cointegration, heterogeneous cointegration, homogeneous cointegration and near-homogeneous cointegration. Extending Phillips (1986), they define the different degrees of cointegration on the base of the rank condition of the conditional long-run variance matrix of the regressors and the regressand. The case of no cointegrating relation is covered by the assumption of almost sure positive definiteness of the long-run conditional variance matrix, whereas a cointegrating relation of various degree exists when such matrix has deficient rank.

In absence of cointegration, Phillips and Moon find that, if panel observations with large cross-section and time-series observations are available, the noise can be characterized as independent across individuals. By pooling the cross-section and the time-series observations, the OLS estimator attenuates the strong effect of the residuals in the regression while retaining the strength of the signal and provides $\sqrt{n}$-consistent estimates of some long-run regression coefficient. This implies that in contrast to non-stationary time-series regression, large-panel regressions can identify a long-run average relation between the regressors and the regressand even in absence of cointegration. Large-panel non-stationary regressions are in fact no longer spurious. The degree of cointegration across individuals depends on the degree of randomness of the cointegrating vector in the model. Phillips and Moon show that the assumption of deficient rank of the long-run variance matrix implies the existence of a panel cointegration model,

$$Y_{i,t} = \beta_i X_{i,t} + E_{i,t}, \tag{2.1}$$

with probability one, where the cointegrating coefficient β_i is random. If no further assumption is imposed, β_i differs randomly across individuals and the cointegrating relation between the regressors and the regressand is heterogeneous. When β_i is constant across different individuals, the same long-run relation between $Y_{i,t}$ and $X_{i,t}$ applies for all i and the cointegrating relation is homogeneous. If β_i has form

$$\beta_i = \beta + \frac{\theta_i}{T\sqrt{n}}, \tag{2.2}$$

where the θ_i are a sequence of $i.i.d$ vectors with mean θ, the model allows for a near-homogeneous cointegrating relation. For all three cointegrating relations, they find that the OLS estimator is $n\sqrt{T}$-consistent and asymptotically normal. Their results are based on a panel Beveridge Nelson decomposition that generalizes Phillips and Solo (1992), and a panel functional central limit theorem for random-coefficients non-stationary models that provide a fundamental framework for any development of asymptotics in large panel models.

The development of asymptotic theory for panel data with large n and T requires assumptions on the treatment of the two indexes. Different approaches are possible. One approach is to fix one index and allow the other to pass to infinity giving an intermediate limit. By letting the other index to pass infinity subsequently a sequential limit is obtained. A second approach, known as diagonal path limit theory, lets the two indexes pass to infinity along a specific diagonal path determined by a functional relation of the type $T = T(n)$. A third approach allows both indexes to infinity simultaneously without any restriction. Joint limits require stronger conditions than sequential limits but, on the other hand, sequential limits can give misleading results. Phillips and Moon discuss a set of sufficient conditions for sequential convergence to imply joint convergence and derive the asymptotic properties of the OLS estimator for sequential and joint limits, imposing the additional rate condition of $n/T \to 0$ in the second case.

Baltagi, Kao and Liu (2008) study the asymptotic properties of the most common panel estimators in a simple error component disturbance model, with random effect assumption,

$$y_{it} = \alpha + \beta x_{it} + u_{it}, \tag{2.3}$$

$$u_{it} = \mu_i + \nu_{it}, \tag{2.4}$$

$$E(\mu_i | x_{it}) = 0, \tag{2.5}$$

with $i = 1, ..., n$ and $t = 1, ..., T$. The regressors and the remainder term are autoregressive and possibly non-stationary,

$$x_{it} = \lambda x_{it-1} + \varepsilon_{it}, \quad |\lambda| \leq 1, \tag{2.6}$$

$$\nu_{it} = \rho \nu_{it-1} + \epsilon_{it}, \quad |\rho| \leq 1, \tag{2.7}$$

and the disturbances $w_{i0} \equiv (\nu_{i0}, \varepsilon_{i0}, \epsilon_{i0})$ are independent across individuals and satisfy a multivariate panel functional central limit theorem,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \mathbf{w}_{i,t} \Longrightarrow BM_i\left(\Omega_i\right), \quad \text{as } T \to \infty \quad \text{for each } i,$$

where $BM_i\left(\Omega_i\right)$ denotes a Brownian motion with covariance matrix Ω_i. Baltagi, Kao and Liu find that the properties of the OLS estimator depend crucially on the non-stationarity of the regressor and the remainder disturbance. When the error component of the disturbance term and the regressor are both stationary ($|\rho| < 1$ and $|\lambda| < 1$) the estimator is $\sqrt{nT}$ consistent and asymptotically normal. If the disturbance is I(1) and the regressor is stationary ($\rho = 1$ and $|\lambda| < 1$) the estimator is $\sqrt{n}$ consistent and asymptotically normal. When the disturbance is stationary and the regressor is I(1) ($|\rho| < 1$ and $\lambda = 1$) the model is cointegrated and the estimator is $\sqrt{nT}$ consistent and asymptotically normal.

2.3 The model

We consider the following scalar panel regression model:

$$y_{it} = \alpha + \beta x_{it} + \gamma z_{it} + \eta_{it}, \tag{2.8}$$

for $t = 1, ..., T$ and $i = 1, ..., n$. The regressors and the regression error have common initialization at $t = 0$ and are generated recursively by

$$x_{it} = x_{it-1} + \varepsilon_{it}, \tag{2.9}$$

$$z_{it} = \rho z_{i,t-1} + u_{it} \quad \text{with } \mid \rho \mid < 1, \tag{2.10}$$

$$\eta_{it} = \lambda \eta_{it-1} + v_{it} \quad \text{with } \mid \lambda \mid < 1. \tag{2.11}$$

The regression errors in (2.11) follow a stationary process ($|\lambda| < 1$) implying that the model is cointegrated. We assume that the vector of innovations

$$\mathbf{w}'_{it} \equiv (\varepsilon_{it}, u_{it}, v_{it})'$$

is generated by the non-random coefficient linear process

$$\mathbf{w}_{it} = \sum_{j=0}^{\infty} \Psi_{i,j} \boldsymbol{\xi}_{i,t-j} \quad \text{with} \quad \Psi_{i,0} \equiv I, \quad \sum_{j=0}^{\infty} ||\Psi_{i,j}||^2 < \infty, \tag{2.12}$$

where I denotes the identity matrix, and that it satisfies the following:

Assumption 1 For each i, $\boldsymbol{\xi}_{i,t}$ is an $i.i.d$ zero mean vector with finite variance-covariance matrix Ξ and finite fourth order cumulants, $K^{\xi}_{abcd}(t_1, t_2, t_3)$, such that

$$\sum_{t_1,t_2,t_3=1}^{\infty} \left| K^{\xi}_{abcd}(t_1, t_2, t_3) \right| < \infty.$$

Assumption 2 For each i,

$$\sum_{j=0}^{\infty} j^a ||\Psi_{i,j}||^a < \infty, \quad \text{for some integer } a > 1,$$

where "$\|.\|$" denotes the Euclidian norm of the coefficient matrix $\Psi_{i,j}$.

(2.12) implies that the innovation process $\mathbf{w}_{i,t}$ admits for every fixed i the panel Beveridge

Nelson decomposition (see Phillips and Moon 1999, Lemma 2) :

$$\mathbf{w}_{i,t} = \Psi_i(1)\,\xi_{i,t} + \tilde{\mathbf{w}}_{i,t-1} - \tilde{\mathbf{w}}_{i,t}, \tag{2.13}$$

where $\Psi_i(1) \equiv \sum_{j=0}^{\infty} \Psi_{i,j}$ and

$$\tilde{\mathbf{w}}_{i,t} = \sum_{j=0}^{\infty} \tilde{\Psi}_{i,j}\boldsymbol{\xi}_{i,t-j}, \quad \tilde{\Psi}_{i,j} = \sum_{t=j+1}^{\infty} \Psi_{i,t}.$$

Assumption 1 and Assumption 2 ensure that the panel data innovations $\mathbf{w}_{it}$ display a degree of homogeneity across time strong enough for the partial sum process $T^{-1/2}\sum_{t=1}^{[Tr]}\mathbf{w}_{i,t}$ to satisfy a Multivariate Invariance Principle for each i. In what follows we directly assume that the partial sum process constructed from the innovation sequence, denoted by $\mathbf{S}_{[Tr]}$, satisfies the following multivariate large T result

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\mathbf{w}_{i,t} \Longrightarrow BM_i(\Omega_i) \quad T \to \infty \quad \text{for each } i, \tag{2.14}$$

where $r \in [0,1]$, the symbol "$[.]$" denotes the largest integer part, the symbol "$\Rightarrow$" denotes weak convergence, and $BM_i(\Omega_i)$ is a (3×1) vector Brownian motion

$$BM_i = \begin{bmatrix} BM_{\varepsilon_i} \\ BM_{u_i} \\ BM_{v_i} \end{bmatrix}, \tag{2.15}$$

with covariance matrix

$$\Omega_i = \Psi_i(1)\,\Xi\Psi_i(1)'.$$

The matrix Ω_i is known as the "long-run covariance matrix" of the innovations $\mathbf{w}_{i,t}$, and it is defined as

$$\Omega_i \equiv \lim_{T\to\infty} E(\mathbf{S}_T\mathbf{S}_T') = \begin{bmatrix} \omega_{\varepsilon_i}^2 & \omega_{\varepsilon_i u_i} & \omega_{\varepsilon_i v_i} \\ \omega_{\varepsilon_i u_i} & \omega_{u_i}^2 & \omega_{u_i v_i} \\ \omega_{\varepsilon_i v_i} & \omega_{u_i v_i} & \omega_{u_{vi}}^2 \end{bmatrix}.$$

It can be decomposed as

$$\Omega_i = \Sigma_i + \Gamma_i + \Gamma_i',$$

where $\Sigma_i = \lim_{T\to\infty} T^{-1/2}\sum_{t=1}^{T} E\left(\mathbf{w}_{i,0}\mathbf{w}_{i,0}'\right)$ and $\Gamma_i = \lim_{T\to\infty}\sum_{t=2}^{T}\sum_{j=1}^{t-1} E\left(\mathbf{w}_{i,j}\mathbf{w}_{i,t}'\right)$. When the innovations are stationary, as implied by Assumption 1, Σ_i and Γ_i reduce to

$$\begin{aligned} \Sigma_i &= E\left(\mathbf{w}_{i,0}\mathbf{w}_{i,0}'\right), \\ \Gamma_i &= \sum_{j=1}^{\infty} E\left(\mathbf{w}_{i,0}\mathbf{w}_{i,j}'\right). \end{aligned} \tag{2.16}$$

In order to rule out endogeneity, we make the following assumption,

Assumption 3 For every i, the sequences $\{\varepsilon_{i,t}\}$, $\{u_{i,t}\}$, $\{v_{i,t}\}$ are statistically independent.

Assumption 3 implies that the innovations have zero covariances in the short and in the long run, thus for every i,

$$\Sigma_i = \begin{bmatrix} \sigma_{\varepsilon_i}^2 & 0 & 0 \\ 0 & \sigma_{u_i}^2 & 0 \\ 0 & 0 & \sigma_{v_i}^2 \end{bmatrix} \text{ and } \Gamma_i = \begin{bmatrix} \gamma_{\varepsilon_i}^2 & 0 & 0 \\ 0 & \gamma_{u_i}^2 & 0 \\ 0 & 0 & \gamma_{v_i}^2 \end{bmatrix}.$$

To simplify the notation we assume constant variances across i, that is to say we set for all i,

$$\Sigma_i = \Sigma \quad \text{and} \quad \Gamma_i = \Gamma.$$

The long run variance matrix of the innovations is for every i

$$\Omega = \begin{bmatrix} \omega_{\varepsilon}^2 & 0 & 0 \\ 0 & \omega_u^2 & 0 \\ 0 & 0 & \omega_v^2 \end{bmatrix}.$$

Assumption 3 implies that the vector Brownian motion in (2.15) has independent components, $BM_{\varepsilon_i}(\omega_{\varepsilon}^2)$, $BM_{u_i}(\omega_u^2)$ and $BM_{v_i}(\omega_v^2)$ and can be rewritten as

$$BM_i(\Omega) = \Omega^{-1/2} \begin{bmatrix} W_{\varepsilon_i} \\ W_{u_i} \\ W_{v_i} \end{bmatrix},$$

where

$$\begin{bmatrix} W_{\varepsilon_i} \\ W_{u_i} \\ W_{v_i} \end{bmatrix}$$

is a standardized vector Brownian motion.

We make the further assumption of cross sectional independence in our model.

Assumption 4 For each i and j such that $i \neq j$, $\xi_{i,t}$ and $\xi_{j,t}$ are independent.

We rely on Assumption 4 to apply the strong law of large numbers for independent sequences to intermediate limits of the statistics of interest, suitably averaged over i. This condition is however very restrictive and it is an important limitation of our results.

Invariance principles such as (2.14) have been established for time series processes satisfying assumptions 1-2 by Phillips and Solow (1992) and for panel data by Phillips and Moon (1999) as an alternative to asymptotics for time series satisfying certain mixing conditions. In addition to the multivariate invariance principle (2.14), our asymptotic results rely upon the weak convergence of certain sample covariance matrix to matrix stochastic integrals of the form $\int W_i dV_i$ which has been shown by Phillips (1997).

We now introduce some notation. In what follows $\mathbf{y}$ denotes the $nT \times 1$ vector that contains all the observations on the dependent variable across individuals and time,

$$\mathbf{y}' \equiv (y_1', y_2', y_3', ..., y_n'),$$

where the vector y_i' contains all the observations for individual i from $t = 1, ..., T$

$$y_i' = (y_{i1}, y_{i2}, ..., y_{iT}).$$

We denote by $\mathbf{W}$ the $nT \times 3$ matrix that contains all the panel observations on the regressors,

$$\mathbf{W} \equiv \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ . \\ . \\ W_n \end{bmatrix},$$

where W_i is the $T \times 3$ matrix that contains observations on individual i on the regressors for $t = 1, ..., T$,

$$W_i = \begin{pmatrix} 1 & x_{i1} & z_{i1} \\ 1 & x_{i_2} & z_{i2} \\ 1 & x_{i_3} & z_{i_3} \\ . & . & . \\ . & . & . \\ 1 & x_{iT} & z_{iT} \end{pmatrix}.$$

Finally we denote by $\boldsymbol{\eta}$ the $nT \times 1$ vector that contains the disturbances for each individual across time

$$\boldsymbol{\eta} = (\eta_1', \eta_2', \eta_3', ..., \eta_n'),$$

where $\eta_i' = (\eta_{i1}, \eta_{i2}, ..., \eta_{iT})$. Model (2.8)-(2.11) is written in conventional matrix form as $\mathbf{y} = \mathbf{W}\boldsymbol{\vartheta} + \boldsymbol{\eta}$. The statistical problem is the estimation of the (3×1) vector of parameters $\boldsymbol{\vartheta}_0' = (\alpha_0, \beta_0, \gamma_0)'$ on the base of panel observations $\{x_{it}, z_{it}, y_{it}\}$ with $i = 1, ..., n$ and $t = 1, ..., T$. The pooled OLS estimator of $\boldsymbol{\vartheta}$ is defined as

$$\begin{aligned} \hat{\boldsymbol{\vartheta}} &\equiv (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} \\ &= \boldsymbol{\vartheta}_0 + (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\eta}, \end{aligned} \tag{2.17}$$

where

$$(\mathbf{W}'\mathbf{W})^{-1} = \begin{bmatrix} nT & \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it} & \sum_{i=1}^{n}\sum_{t=1}^{T} z_{it} \\ \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it} & \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^2 & \sum_{i=1}^{n}\sum_{t=1}^{T} z_{it}x_{it} \\ \sum_{i=1}^{n}\sum_{t=1}^{T} z_{it} & \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}z_{it} & \sum_{i=1}^{n}\sum_{t=1}^{T} z_{it}^2 \end{bmatrix}^{-1}, \tag{2.18}$$

and

$$\mathbf{W}'\boldsymbol{\eta} = \begin{pmatrix} \sum_{i=1}^{n}\sum_{t=1}^{T} \eta_{it} \\ \sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}\eta_{it} \\ \sum_{i=1}^{n}\sum_{t=1}^{T} z_{it}\eta_{it} \end{pmatrix}. \tag{2.19}$$

The following section discusses our main results, namely the asymptotic normality and rates of convergence of the pooled OLS estimator of ϑ in a cointegrated model. All the results are for sequential limits. We find the time series limit behavior of the double index statistics of interest, say $Y_{i,T}$, either in probability or in distribution and then, by the independence across i for all T, we establish the limit behavior of its cross sectional average. Assumptions 1-4 are not strong enough to ensure that our results readily extends to joint convergence.

2.4 Main results

The following theorem establishes the asymptotic normality of the pooled OLS estimator of ϑ in model (2.8)-(2.11).

Theorem 1. *Under assumptions 1-4, as* $(n, T \to \infty)_{\text{seq}}$

$$D^{1/2}\left(\hat{\boldsymbol{\vartheta}} - \vartheta_0\right) \to_d N\left(0, \Phi\mathbf{V}\Phi'\right),$$

where

$$\mathbf{D}^{1/2} = diag\left(\sqrt{nT}, T\sqrt{n}, \sqrt{nT}\right),$$

V *is a positive definite matrix and*

$$\Phi \equiv diag\left(1,\ \frac{\omega_{\varepsilon}^{2}}{2},\ \frac{\omega_{u}^{2}}{2}\right).$$

The proof of Theorem 1 follows from a simple application of Cramèrs's convergence Theorem. Writing (2.17) as

$$\mathbf{D}^{1/2}\left(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\right) = \left(\mathbf{D}^{-1/2}\mathbf{W}'\mathbf{W}\mathbf{D}^{-1/2}\right)^{-1}\mathbf{D}^{-1/2}\mathbf{W}'\boldsymbol{\eta},$$

the result follows once we establish the convergence in probability of $\mathbf{D}^{-1/2}\mathbf{W}'\mathbf{W}\mathbf{D}^{-1/2}$ to a positive definite matrix and the convergence in distribution of the vector $\mathbf{D}^{-1/2}\mathbf{W}'\boldsymbol{\eta}$ to a normal random variable with finite variance. Lemma 1 derives the pointwise convergence in probability of $\mathbf{D}^{-1/2}\mathbf{W}'\mathbf{W}\mathbf{D}^{-1/2}$.

Lemma 1 *Under assumptions 1-4, as* $(n, T \to \infty)_{\text{seq}}$ *the matrix*

$$\mathbf{D}^{-1/2}\mathbf{W}'\mathbf{W}\mathbf{D}^{-1/2} = \begin{bmatrix} 1 & \frac{1}{nT^{3/2}}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it} & \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} z_{it} \\ \frac{1}{nT^{3/2}}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it} & \frac{1}{T^2 n}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^2 & \frac{1}{nT^{3/2}}\sum_{i=1}^{n}\sum_{t=1}^{T} z_{it}x_{it} \\ \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} z_{it} & \frac{1}{nT^{3/2}}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}z_{it} & \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} z_{it}^2 \end{bmatrix}$$

converges in probability to Φ*.*

The proof of Lemma 1 relies on sequential limit theory in an essential way. Following Phillips and Moon the first intermediate limit is found by standard functional central limit theorems for stationary and non-stationary sequences (see Park and Phillips, 1998). Then the uniform square integrability of this intermediate limit is verified to justify the use of a strong law of large numbers as $n \to \infty$. The first intermediate limit follows by well-established asymptotic results for linear processes provided by Phillips and Durlauf (1986) and Phillips and Solo (1992). When these convergence result are not readily available we exploit the martingale difference sequence approximation of the innovation sequence implied by the Beveridge Nelson decomposition to obtain an intermediate limit as shown in Phillips and Moon (1999).

Lemma 2 derives the asymptotic normality of any linear combination of the vector $\mathbf{D}^{-1/2}\mathbf{W}'\boldsymbol{\eta}$.

Lemma 2 *Under assumptions1-4 as* $(n, T \to \infty)_{\text{seq}}$

$$\mathbf{D}^{-1/2}\mathbf{W}'\boldsymbol{\eta} \to^{d} \mathbf{N}\left(0, \mathbf{V}\right),$$

where

$$\mathbf{D}^{-1/2}\mathbf{W}'\boldsymbol{\eta} = \begin{pmatrix} \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{t} \eta_{it} \\ \frac{1}{T\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{t} x_{it}\eta_{it} \\ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{t} z_{it}\eta_{it} \end{pmatrix}. \tag{2.20}$$

The proof of Lemma 2 verifies that uniformly in T the linear combinations of the vector (2.20) satisfy a Liapunov condition. This is achieved via the martingale difference approximation of the innovation sequences which allows to make repeated use of Burkholder's inequality (see Appendix A).

2.5 Conclusion

The main result of this chapter (Theorem 1) is coherent with the results of Phillips and Moon (1999), who find that in a homogeneous cointegrated model with $\mathrm{I}(1)$ regressors, the rate of convergence of the estimators is $T\sqrt{n}$. This is not surprising since in presence of a cointegrating relation adding one more stationary regressor does not modify crucially any derivation. The coefficients of the stationary regressors converge at rate $\sqrt{nT}$, which is the standard convergence rate in large stationary panels. The presence of one non-stationary regressor does not slow down the convergence rate of the stationary regressors since the model is cointegrated. The stronger limitation of this result is the assumption of cross-section independence, which is indeed quite restrictive. In fact it rules out all instances of global shocks or interdependencies among the variables, which are a common feature of many economic models. Theorem 1 can be extended in many directions. Assuming higher order moment conditions on the $\boldsymbol{\xi}_{it}$ and convergence of the ratio $\sqrt{n}/T$ to zero, we could derive joint asymptotics. Since the ability of identifying individual heterogeneity is one of the most important advantages in the use of panel data, the introduction of an individual effect in model (2.8)-(2.11) seems highly desirable. It would be of particular interest to work out the asymptotics of the estimator under fixed effect assumptions as most economic models imply some correlation between the individual effect and one or more controls. A further step could be to include I(2) regressors in the model. The simultaneous presence of I(0), I(1), I(2) regressors is quite common in empirical literature, for example in growth models that study the effect of inequality on growth rates employing the Gini's coefficient as a measure of inequality. In such models it is frequently found that the Gini's coefficient is I(2), whereas income or purchasing power parity are first order stationary and other explanatory variables such as prices or interest rates are stationary. Park and Phillips (1988) developed a multivariate regression theory for time-series integrated processes accommodating integrated processes of different orders, however their results have not been extended to large panel models.

2.6 Appendix A: Technical lemmas

Proof of Lemma 1 The convergence in probability of $\mathbf{D}^{-1/2}\mathbf{W}'\mathbf{W}\mathbf{D}^{-1/2}$ is derived element by element. Consider the term

$$\frac{1}{nT^{3/2}}\sum_{i=1}^{n}\sum_{t=1}^{T}x_{it}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_{it}\equiv\frac{1}{n}\sum_{i=1}^{n}Y_{i,T}^{(1)}, \tag{2.21}$$

for every i. Assumptions 1-4 imply that

$$Y_{i,T}^{(1)}\Rightarrow\omega_{\varepsilon}\int_{0}^{1}W_{\varepsilon_i}(r)dr\quad T\to\infty,$$

by the Functional Central Limit Theorem (see Park and Phillips, 1988, Lemma 2.1 **(**a**)**). By the Continuous Mapping Theorem

$$\left(Y_{i,T}^{(1)}\right)^2\Rightarrow\left(\omega_{\varepsilon}\int_{0}^{1}W_{\varepsilon_i}(r)dr\right)^2\quad \text{as } T\to\infty\ .$$

Standard calculations yield

$$E\left(Y_{i,T}^{(1)}\right)^2\to E\left(\omega_{\varepsilon}\int_{0}^{1}W_{\varepsilon_i}(r)dr\right)^2,$$

and we may conclude that the sequence $Y_{i,T}^{(1)}$ is uniformly square integrable in T. Then by independence across i we can apply a strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n}Y_{i,T}^{(1)}\to_{a.s}\omega_{\varepsilon}E\left(\int_{0}^{1}W_{\varepsilon_i}(r)dr\right)=0,\quad n\to\infty.$$

Consider the term

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}z_{it}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}z_{it}\equiv\frac{1}{n}\sum_{i=1}^{n}Y_{i,T}^{(2)}. \tag{2.22}$$

for every i. Recall that by definition

$$z_{it}=\rho z_{i,t-1}+u_{it}\quad \text{with } \mid\rho\mid<1, \tag{2.23}$$

where

$$u_{it}=\sum_{s=0}^{\infty}\psi_{is}\xi_{it-s}\qquad\sum_{s=0}^{\infty}\psi_{is}^2<\infty \text{ and } \psi_{i0}=1,$$

with ξ_{it} independent and identically distributed across i and t. Moreover by Assumptions 1 and 2, u_{it} admits the following panel Beveridge-Nelson decomposition:

$$u_{it}=\psi_i\left(1\right)\xi_{it}+\tilde{\xi}_{it-1}-\tilde{\xi}_{it}, \tag{2.24}$$

where

$$\psi_i(1) \equiv \sum_{j=0}^{\infty} \psi_{ij} \quad \text{and} \quad \tilde{\xi}_{it} \equiv \sum_{s=0}^{\infty} \tilde{\psi}_{is} \xi_{it-s} \text{ with } \tilde{\psi}_{is} = \sum_{t=s+1}^{\infty} \psi_{it}$$

(see Phillips and Solo, 1992, Lemma 2). Then

$$\frac{1}{T}\sum_{t=1}^{T} u_{it} = \psi_i(1)\frac{1}{T}\sum_{t=1}^{T}\xi_{it} + \frac{1}{T}\tilde{\xi}_{i0} - \frac{1}{T}\tilde{\xi}_{iT}. \tag{2.25}$$

By recursive substitution (2.23) can be expressed as

$$z_{it} = \rho^t z_{i0} + \sum_{s=1}^{t} \rho^{|t-s|} u_{is},$$

then substituting (2.25) we express it as

$$\begin{aligned} z_{it} &= \rho^t z_{i0} + \sum_{s=1}^{t} \rho^{|t-s|}\psi_i(1)\xi_{is} + \sum_{s=1}^{t} \rho^{|t-s|}\tilde{\xi}_{is-1} - \sum_{s=1}^{t} \rho^{|t-s|}\tilde{\xi}_{is} \\ &= \rho^t z_{i0} + \psi_i(1)\sum_{s=1}^{t} \rho^{|t-s|}\xi_{is} + \rho^{|t-1|}\tilde{\xi}_{i0} - \tilde{\xi}_{it}. \end{aligned}$$

Therefore

$$\frac{1}{T}\sum_{t=1}^{T} z_{it} = \frac{1}{T}\sum_{t=1}^{T}\psi_i(1)\left(\sum_{s=1}^{t}\rho^{|t-s|}\xi_{is}\right) + \frac{1}{T}\sum_{t=1}^{T}\rho^{|t-1|}\tilde{\xi}_{i0} - \frac{1}{T}\sum_{t=1}^{T}\tilde{\xi}_{it} + \frac{1}{T}\sum_{t=1}^{T}\rho^t z_{i0}.$$

Following Phillips and Moon, Lemma 13, page 1100, let us denote

$$Q_{i,T} \equiv \frac{1}{T}\sum_{t=1}^{T}\psi_i(1)\left(\sum_{s=1}^{t}\rho^{|t-s|}\xi_{is}\right), \tag{2.26}$$

and

$$R_{i,T} \equiv \frac{1}{T}\sum_{t=1}^{T}\rho^{|t-1|}\tilde{\xi}_{i0} - \frac{1}{T}\sum_{t=1}^{T}\tilde{\xi}_{it} + \frac{1}{T}\sum_{t=1}^{T}\rho^t z_{i0}. \tag{2.27}$$

By Lemma 13 of Phillips and Moon (1999), (2.27) converges almost surely to zero as $(n, T \to \infty)_{\text{seq}}$ (see page 1101, Phillips and Moon, 1999), implying that the sequential limit of (2.22) is found by establishing the sequential limit of (2.26). Consider that for fixed i

$$Q_{i,T} \to_{a.s} Q_i = 0, \quad \text{as } T \to \infty,$$

then by the strong law of large numbers for iid sequences

$$\frac{1}{n}\sum_{i=1}^{n} Q_i \to_{a.s} 0, \quad \text{as } n \to \infty,$$

which implies that

$$\frac{1}{n}\sum_{i=1}^{n} Y_{i,T}^{(2)} \to 0, \quad \text{as } (n, T\to\infty)_{\text{seq}}.$$

Consider the term

$$\frac{1}{T^2 n}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^2}\sum_{t=1}^{T} x_{it}^2 \equiv \frac{1}{n}\sum_{i=1}^{n} Y_{i,T}^{(3)}. \tag{2.28}$$

For every i, as $T\to\infty$

$$Y_{i,T}^{(3)} = \left(\frac{1}{T^2}\sum_{t=1}^{T} x_{it}^2\right) \Rightarrow \omega_\varepsilon^2 \int_0^1 W_{\varepsilon_i}^2(r)dr,$$

by Functional Central Limit Theorem (see Lemma 2.1**(**c**)**, Park and Phillips, 1988). By the Continuos Mapping Theorem

$$\left(Y_{i,T}^{(3)}\right)^2 \Rightarrow \left(\omega_\varepsilon^2 \int_0^1 W_{\varepsilon_i}^2(r)dr\right)^2.$$

Standard calculations (see Phillips and Moon, 1999, page 1100) yield

$$E\left(Y_{i,T}^{(3)}\right)^2 \to E\left(\omega_\varepsilon^2 \int_0^1 W_{\varepsilon_i}^2(r)dr\right)^2,$$

which implies that $Y_{i,T}^{(3)}$ is a uniformly square integrable sequence in T, then the strong law of large numbers applies and as $n\to\infty$

$$\frac{1}{n}\sum_{i=1}^{n} Y_{i,T}^{(3)} \to_{a.s} E\left(\omega_\varepsilon^2 \int_0^1 W_{\varepsilon_i}^2(r)dr\right) = \frac{\omega_\varepsilon^2}{2}.$$

Consider the mixed term

$$\frac{1}{nT^{\frac{3}{2}}}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}z_{it} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T} x_{it}z_{it} \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}Y_{i,T}^{(4)}\right). \tag{2.29}$$

For every fixed i, as $T\to\infty$

$$Y_{i,T}^{(4)} = \frac{1}{T}\sum_{t=1}^{T} x_{it}z_{it} \Rightarrow \omega_\varepsilon\omega_u \int_0^1 W_i^{u_i} dW_i^{\varepsilon_i} + \Sigma_{21} + \Gamma_{21}$$

by the Functional Central Limit Theorem (see Lemma 2.1 **(**e**),** Park and Phillips, 1988)**,** where Σ_{21} and Γ_{21} are the $2nd$ row, $1st$ column elements of the matrices Σ and Γ defined in (2.16). Moreover by Assumption 3, $\Sigma_{21} = \Gamma_{21} = 0$. By the Continuous Mapping theorem

$$\left(Y_{i,T}^{(4)}\right)^2 \Rightarrow \left(\omega_\varepsilon\omega_u \int_0^1 W_i^{u_i} dW_i^{\varepsilon_i}\right)^2,$$

standard calculation shows

$$E\left(Y_{i,T}^{(4)}\right)^2 \to E\left(\omega_\varepsilon\omega_u\int_0^1 W_i^{u_i}dW_i^{\varepsilon_i}\right)^2$$

(see Baltagi, Kao, Liu, Appendix A) and we may conclude that $Y_{i,T}^{(4)}$ is uniformly square integrable in T. Then the strong law of large numbers for independent sequences implies

$$\frac{1}{n}\sum_{i=1}^{n} Y_{i,t}^{(4)} \to_{a.s} E\left(\omega_\varepsilon\omega_u\int_0^1 W_i^{u_i}dW_i^{\varepsilon_i}\right) = 0.$$

Consider the last term

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} z_{it}^2 \equiv \frac{1}{n}\sum_{i=1}^{n} Y_{i,t}^{(5)}. \tag{2.30}$$

As for (2.22), the existence of the panel Beveridge- Nelson decomposition of the sequence of disturbances ξ_{it} for every fixed i, allows us to apply Lemma 13 of Phillips and Moon and to conclude that for fixed i

$$Y_{i,t}^{(5)} \Rightarrow \frac{\omega_u^2}{2}, \quad T\to\infty,$$

which is constant across i by assumption. Then we may conclude that by the strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^{n} Y_{i,t}^{(5)} \to_{a.s} \frac{\omega_u^2}{2},$$

which concludes the proof of Lemma 1.

Proof of Lemma 2 Observe that $\mathbf{c}'\mathbf{D}^{-1/2}\mathbf{W}'\boldsymbol{\eta}$ can be written as

$$= \mathbf{c}'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \frac{1}{\sqrt{T}}\sum\limits_{t=1}^{T}\eta_{it} \\ \frac{1}{T}\sum\limits_{t=1}^{T}x_{it}\eta_{it} \\ \frac{1}{\sqrt{T}}\sum\limits_{t=1}^{T}z_{it}\eta_{it}\end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(c'Z_{i,T}\right), \tag{2.31}$$

where

$$\mathbf{c}'Z_{i,T} = \frac{c_1}{\sqrt{T}}\sum_{t=1}^{T}\eta_{it} + \frac{c_2}{T}\sum_{t=1}^{T}x_{it}\eta_{it} + \frac{c_3}{\sqrt{T}}\sum_{t=1}^{T}z_{it}\eta_{it}.$$

The independence across i of the $c'Z_{i,T}$ implies that to establish a Central Limit Theorem for

(2.31), we must show that $E\left(c'Z_{i,T}\right)=0$ and that for some $\delta>0$,

$$E\left|c'Z_{i,T}\right|^{2+\delta}<\infty.$$

By Assumptions 1-4, $E\left(c'Z_{i,T}\right)=0$. Notice that by the C_r-inequality

$$E\left|\frac{c_1}{\sqrt{T}}\sum_{t=1}^{T}\eta_{it}+\frac{c_2}{T}\sum_{t=1}^{T}x_{it}\eta_{it}+\frac{c_3}{\sqrt{T}}\sum_{t=1}^{T}z_{it}\eta_{it}\right|^{2+\delta}\leq$$

$$C_r\left(E\left|\frac{c_1}{\sqrt{T}}\sum_{t=1}^{T}\eta_{it}\right|^{2+\delta}+E\left|\frac{c_2}{T}\sum_{t=1}^{T}x_{it}\eta_{it}\right|^{2+\delta}+E\left|\frac{c_3}{\sqrt{T}}\sum_{t=1}^{T}z_{it}\eta_{it}\right|^{2+\delta}\right), \tag{2.32}$$

with $c_r=2^{r-1}$. We show that each term of (2.32) is bounded. Consider the term

$$E\left|\frac{c_1}{\sqrt{T}}\sum_{t=1}^{T}\eta_{it}\right|^{2+\delta}. \tag{2.33}$$

Since by Assumptions 1-4, η_{it} admits a Beveridge-Nelson decomposition, Burkholder's inequality yields

$$T^{-\frac{2+\delta}{2}}E\left|\sum_{t=1}^{T}\eta_{it}\right|^{2+\delta}\leq T^{-\frac{2+\delta}{2}}E\left|\sum_{t=1}^{T}\eta_{it}^{2}\right|^{\frac{2+\delta}{2}},$$

then Holder's inequality yields

$$T^{-\frac{2+\delta}{2}}E\left|\sum_{t=1}^{T}\eta_{it}^{2}\right|^{\frac{2+\delta}{2}}\leq T^{-\frac{2+\delta}{2}}E\left|\left(\sum_{t=1}^{T}\eta_{it}^{2+\delta}\right)^{\frac{2}{2+\delta}}\right|^{\frac{2+\delta}{2}},$$

and by Jensen's inequality

$$\begin{aligned}T^{-\frac{2+\delta}{2}}E\left|\left(\sum_{t=1}^{T}\eta_{it}^{2+\delta}\right)^{\frac{2}{2+\delta}}\right|^{\frac{2+\delta}{2}} &= T^{-\frac{2+\delta}{2}}E\left|\sum_{t=1}^{T}\eta_{it}^{2+\delta}\right| \\ &\leq T^{-\frac{2+\delta}{2}}\sum_{t=1}^{T}E\left|\eta_{it}\right|^{2+\delta}=O\left(T^{-\frac{\delta}{2}}\right),\end{aligned}$$

because we have assumed $E\left|\eta_{it}\right|^{2+\delta}<\infty$, which implies

$$E\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_{it}\right|^{2+\delta}<\infty.$$

Consider the last term

$$E\left|\frac{c_3}{\sqrt{T}}\sum_{t=1}^{T}z_{it}\eta_{it}\right|^{2+\delta}.$$

By Assumptions 1-4, z_{it} and η_{it} are independent, therefore the sequence $\{z_{it}\eta_{it}\}$ is a martingale difference sequence. Indeed

$$\begin{aligned} E\left[z_{it}\eta_{it}|z_{it-1}\eta_{it-1}\right] &= E\left[z_{it}|z_{it-1}\eta_{it-1}\right]E\left[\eta_{it}|z_{it-1}\eta_{it-1}\right] \\ &= E\left[z_{it}|z_{it-1}\right]E\left[\eta_{it}|\eta_{it-1}\right]=0. \end{aligned}$$

But then as above Burkholder's inequality for martingale difference sequences yields

$$T^{-\frac{2+\delta}{2}}E\left|\sum_{t=1}^{T}z_{it}\eta_{it}\right|^{2+\delta}\leq T^{-\frac{2+\delta}{2}}E\left|\sum_{t=1}^{T}z_{it}^{2}\eta_{it}^{2}\right|^{\frac{2+\delta}{2}},$$

then Holder's inequality yields

$$T^{-\frac{2+\delta}{2}}E\left|\sum_{t=1}^{T}z_{it}^{2}\eta_{it}^{2}\right|^{\frac{2+\delta}{2}}\leq T^{-\frac{2+\delta}{2}}E\left|\left(\sum_{t=1}^{T}\left(z_{it}\eta_{it}\right)^{2+\delta}\right)^{\frac{2}{2+\delta}}\right|^{\frac{2+\delta}{2}},$$

and by Jensen's inequality

$$\begin{aligned} T^{-\frac{2+\delta}{2}}E\left|\left(\sum_{t=1}^{T}\left(z_{it}\eta_{it}\right)^{2+\delta}\right)^{\frac{2}{2+\delta}}\right|^{\frac{2+\delta}{2}} &\leq T^{-\frac{2+\delta}{2}}E\left|\left(\sum_{t=1}^{T}\left(z_{it}\eta_{it}\right)^{2+\delta}\right)^{\frac{2}{2+\delta}}\right|^{\frac{2+\delta}{2}} \\ &= T^{-\frac{2+\delta}{2}}E\left|\sum_{t=1}^{T}z_{it}^{2+\delta}\eta_{it}^{2+\delta}\right| \\ &\leq T^{-\frac{2+\delta}{2}}\sum_{t=1}^{T}E\left|\eta_{it}z_{it}\right|^{2+\delta}=O\left(T^{-\frac{\delta}{2}}\right), \end{aligned}$$

because we have assumed that $E\left|\eta_{it}\right|^{2+\delta}<\infty$ and $E\left|z_{it}\right|^{2+\delta}$ and independence between z_{it} and η_{it}. Analogously it can be shown that

$$E\left|\frac{c_2}{\sqrt{T}}\sum_{t=1}^{T}x_{it}\eta_{it}\right|^{2+\delta}<\infty,$$

which concludes the proof of Lemma 2.

Chapter 3 Multivariate Volatility Models

3.1 Introduction

Recently a large body of research in financial econometrics has emerged on modeling the comovements of financial returns. Understanding the comovements of financial returns is of great importance in many applied situations. The knowledge of correlation structures is vital in asset pricing, optimal portfolio risk management and asset allocation. Moreover, as the volatilities of different assets and markets move together, modelling volatility in a multivariate framework can lead to greater statistical efficiency. Univariate modelling of random volatility has developed along two main lines of research: conditional volatility models, where the volatility is a deterministic function of the past realizations of the asset, and stochastic volatility models, where it is a latent process. Generalizations to multivariate settings have been implemented in both classes and, as a result, a wide range of multivariate GARCH and Stochastic Volatility models has been developed and applied extensively in recent years.

In both classes, multivariate generalizations have to combine different needs. Parsimony of the chosen specification is essential for relatively easy estimation of the model. However, as the dimension of the vector of returns increases, the number of parameters to estimate increases at a much faster rate, making parameter estimation computationally very intensive. Estimation of multivariate GARCH models requires numerical optimization of the likelihood function which requires to invert the conditional covariance matrix at every iteration. The conditional covariance matrix often depends on time t and has to be inverted for all t in every iteration of the numerical optimization. If the dimension of the model is large, this is computationally time consuming and numerically unstable. In contrast to multivariate GARCH models, in multivariate Stochastic Volatility models, the conditional covariance is latent and has to be integrated out from the likelihood function. As a result, the likelihood function does not have a closed form expression and its evaluation requires evaluation via numerical methods of high-dimensional integrals. As the dimension of the model increases, this brings a high computational burden. Parsimony is in general ensured by imposing directly on the model some simplifying restrictions. However if too many restrictions are imposed, parameters interpretation may become difficult and moreover the model may fail to capture important dynamics of the data. Indeed there is a trade off between flexibility of the model specification and the so called curse of dimensionality.

Another important feature that needs to be taken into account in the specification of a multivariate volatility model is the positive definiteness of the conditional covariance matrix. By definition covariance matrices must be positive semidefinite, however in order to ensure that all the possible portfolios (i.e. linear combinations of a vector of returns) have correlations between -1 and 1, positive definiteness of the covariance matrix must be ensured in the model. Imposing on the model conditions that guarantee positive definiteness of the covariance matrix across time is in practice numerically infeasible, especially in large systems. Most of the multivariate models in the literature are formulated in such a way that positive definiteness is implied directly by the model structure.

Combining these issues has been the main difficulty of the multivariate GARCH and Stochastic Volatility literature. Various approaches to address positive definiteness of the conditional covariance matrix, the curse of dimensionality, and practical implementation issues have generated different types of models in both classes. Extensive literature reviews on MGARCH models are provided in Bauwens, Laurent and Rombouts (2006) and in Silvennoinen and Terasvirta (2008). Asai, McAleer and Yu (2006) survey the main developments of multivariate Stochastic Volatility models.

3.2 Multivariate Conditional Volatility Models

Consider a $n \times 1$ random vector y_t. Denote the sigma algebra generated by the past information until time $t-1$ as F_{t-1}. The general parametric formulation of a multivariate GARCH model is given as:

$$y_t = \mu_t(\theta) + \varepsilon_t, \tag{3.1}$$

$$\varepsilon_t = H^{1/2}(\theta) z_t, \tag{3.2}$$

where z_t is an $i.i.d$ zero mean random vector with unit variance, I_n, the identity matrix of order n, and $H_t^{1/2}$ is a $n \times n$ positive definite matrix. The vector $\mu_t(\theta)$ is the conditional mean of the process, and the matrix $H_t(\theta)$ is its conditional variance. To see this consider that the assumptions on the first two moments of z_t imply

$$Var(y_t|F_{t-1}) = Var(\varepsilon_t|F_{t-1}) = H_t^{1/2} Var(z_t|F_{t-1}) \left(H_t^{1/2}\right)' = H_t. \tag{3.3}$$

$H_t^{1/2}$ is the $n \times n$ positive definite matrix such that $H_t = \left(H_t^{1/2}\right)^2$ is the conditional variance matrix and it is obtained by Cholesky decomposition of H_t. The conditional mean and the and the conditional variance of the process depend on the unknown vector of parameters θ. In most cases the parameterization is disjoint, i.e. the conditional mean and variance depend on two disjoint subvectors of the vector θ. However in GARCH-in mean models μ_t is functionally dependent on H_t. The literature on multivariate GARCH models typically reviews the models according to the different specification of H_t, taking no account of the conditional mean vector. We follow the same approach. According to the specification of the conditional variance matrix H_t, we distinguish three main classes in the literature. The first class comprises models which are a direct extension of the univariate GARCH and model directly the variance covariance matrix. The second class is that of the factor GARCH models, which are motivated both by parsimony and easiness of their economic interpretation. The models in the third class specify the conditional covariances and variances directly, including the CCC model and its dynamic extensions. They offer a straightforward interpretation of the estimated parameters.

3.2.1 VEC and BEKK models

One of the first multivariate GARCH models proposed in the literature was the VEC model of Bollerslev, Engle and Wooldridge (1988). The model is a straightforward generalization of the univariate GARCH of Bollerslev (1986). Every conditional variance and covariance is a function of lagged conditional variances and covariances as well as lagged squared returns,

$$vech\left(H_t\right) = vech\left(C\right) + \sum_{i=1}^{q} A_i vech\left(\varepsilon_{t-i}\varepsilon'_{t-i}\right) + \sum_{j=1}^{p} B_j vech\left(H_{t-j}\right), \tag{3.4}$$

where the $vech\left(H_t\right)$ operator stacks the columns of the lower triangular part of H_t into a long vector, and A_i and B_j are matrices of parameters with dimension $\left(n\left(n+1\right)/2\right) \times \left(n\left(n+1\right)/2\right)$. The generality of this model allows for a great flexibility, however it comes at the price of an excessive number of parameters to estimate. For example in a simple bivariate VEC$(1,1)$:

$$\begin{aligned} vech\left(H_t\right) &= \begin{bmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{bmatrix} \qquad\qquad (3.5) \\ &= \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon^2_{1t-1} \\ \varepsilon_{1t-1}\varepsilon_{2t-1} \\ \varepsilon^2_{2t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} h_{1t-1} \\ h_{21t-1} \\ h_{2t-1} \end{bmatrix}, \end{aligned}$$

there are 18 free parameters in the A and B matrices and 3 free parameters in the constant vector. A general n dimensional VEC(p,q) has $n\left(n+1\right)/2$ free parameters in the constant $c = vech\left(C\right)$, $\left[n\left(n+1\right)/2\right]^2$ parameters in each of the A_i and B_j matrices, and so in the overall it has $n\left(n+1\right)/2+ \left[n\left(n+1\right)/2\right]^2\left(p+q\right)$ free parameters to estimate. The model also requires relevant restrictive conditions to ensure the positive definiteness of the conditional variance matrix. Engle and Kroner (1995) provide a sufficient condition for almost sure positive definiteness of the covariance matrix using the BEKK representation of the VEC models, see below. Gourieroux (1997) provides another sufficient condition for positive definiteness of H_t, rewriting the conditional matrix as a recursive equation yielding a symmetric solution

$$H_t = C + \sum_{i=1}^{q} \left(I_n \otimes \varepsilon'_{t-i}\right) \ddddot{A}_i \left(I_n \otimes \varepsilon_{t-i}\right) + \sum_{j=1}^{p} E\left[\left(I_n \otimes \varepsilon'_{t-i}\right) \ddddot{B}_j \left(I_n \otimes \varepsilon_{t-i}\right) | F_{t-i-1}\right], \tag{3.6}$$

where the symbol "$\otimes$" denotes the Kronecker product. Positive definiteness is achieved assuming that C is positive definite and that the $\ddddot{A}_i$ and $\ddddot{B}_j$ are positive semidefinite matrices.

For empirical application Bollerslev, Engle and Wooldridge implement a simplified version of the general VEC model, setting $p = q = 1$ and imposing diagonality on the matrices A_1 and B_1. In the diagonal VEC, each element h_{ijt} depends only on its own lag and the previous value of $\varepsilon_{it}\varepsilon_{jt}$, so the number of parameters reduces to $3 \times [n(n+1)/2]$. For example a bivariate diagonal VEC model is

$$\begin{aligned} h_t &= \begin{bmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{bmatrix} \\ &= \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} \varepsilon^2_{1t-1} \\ \varepsilon_{2t-1}\varepsilon_{1t-1} \\ \varepsilon^2_{2t-1} \end{bmatrix} + \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} \begin{bmatrix} h_{1t-1} \\ h_{21t-1} \\ h_{2t-1} \end{bmatrix}. \end{aligned}$$

Bollerslev, Engle and Wooldridge argue that this restriction of the general VEC model is plausible because information about variances is usually revealed in squared residuals, if the variances are evolving slowly then past squares residuals should be able to forecast future variances. The diagonal restriction keeps under control the proliferation of parameters. However it might still not yield a positive definite covariance matrix.

Necessary and sufficient conditions for covariance stationarity of the VEC model are derived in Engle and Kroner (1995). Consider a VEC(p, q) model :

$$vech\,(H_t) = c + \sum_{i=1}^{q} A_i vech\left(\varepsilon_{t-i}\varepsilon'_{t-i}\right) + \sum_{j=1}^{p} B_j vech\,(H_{t-j})\,. \tag{3.7}$$

Defining the lag operator L such that $L\varepsilon_t = \varepsilon_{t-1}$, and defining the polynomial $A\,(L) = A_1 L + A_2 L^2 + ... + A_q L^q$ and $B\,(L) = B_1 L + B_2 L^2 + ... + B_p L^p$, (3.7) can be written as:

$$\begin{aligned} vech\,(H_t) &= c + A\,(L)\,vech\,(\varepsilon_t\varepsilon'_t) + B\,(L)\,vech\,(H_t) \\ &= c + A\,(L)\,vech\,(\varepsilon_t\varepsilon'_t) + B\,(L)\sum_{i=1}^{\infty} B\,(L)^{i-1}\left[c + A\,(L)\,vech\,(\varepsilon_t\varepsilon'_t)\right] \\ &= c + A\,(L)\,vech\,(\varepsilon_t\varepsilon'_t) + \sum_{i=2}^{\infty} B\,(L)^{i-1}\left[c + A\,(L)\,vech\,(\varepsilon_t\varepsilon'_t)\right]. \end{aligned}$$

Assuming that ε_t is a doubly infinite sequence, Engle and Kroner show that the process is covariance stationary if and only if the eigenvalues of $A(1)+B(1)$ are less than one in modulus (see Engle and Kroner, 1995, page 132).

To overcome these relevant limitations Engle and Kroner (1995) introduce the Baba-Engle-Kraft-Kroner model, know as the BEKK model. The new parameterization guarantees positive definiteness of the covariance matrix under very weak conditions, easy to impose during estimation. Moreover it is general enough to include all the positive definite diagonal VEC representations and almost all the positive definite general VEC representations. The BEKK(p,q,K) model specifies the conditional variance matrix as

$$H_t = C_0^{*'} C_0^* + \sum_{k=1}^{K}\sum_{i=1}^{q} A_{ik}^{*'} \varepsilon_{t-i}\varepsilon_{t-i}' A_{ik}^* + \sum_{k=1}^{K}\sum_{j=1}^{p} G_{ik}^{*'} H_{t-i} G_{ik}^*, \tag{3.8}$$

where C_0^*, A_{ik}^* and G_{ik}^* are $n\times n$ parameter matrices with C_0^* upper triangular. The summation limit $K, 1\leq K\leq n^2$ determines the generality of the process. In contrast with the VEC model, the parameters of the BEKK do not represent directly the impact of the different lagged terms on the element of H_t. The BEKK model can be seen as a special case of the general VEC, obtained by imposing restrictions on its parameters. Consider for example a simple BEKK $(1,1,1)$

$$H_t = C_0^{*'} C_0^* + A_{11}^{*'} \varepsilon_{t-1}\varepsilon_{t-1}' A_{11}^* + G_{11}^{*'} H_{t-1} G_{11}^*, \tag{3.9}$$

to simplify matters, consider a bivariate BEKK$(1,1,1)$

$$\begin{aligned} H_t &= C_0^{*'} C_0^* + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} \\ &\quad + \begin{bmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{bmatrix}' \begin{bmatrix} h_{1t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{2t-1} \end{bmatrix} \begin{bmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{bmatrix}. \end{aligned} \tag{3.10}$$

(3.10) specifies the individual conditional variances as

$$
\begin{aligned}
h_{11t} &= c_{11} + a_{11}^{*2}\varepsilon_{1,t-1}^2 + 2a_{11}^{*}a_{21}^{*}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{21}^{*2}\varepsilon_{2,t-1}^2 + g_{11}^{2*}h_{t-1} + 2g_{11}^{*}g_{21}^{*}h_{1,t-1}h_{2,t-1} + g_{21}^{*2}h_{2,t-1}^2, \\
h_{12t} &= c_{12} + a_{11}^{*}a_{12}^{*}\varepsilon_{1,t-1}^2 + (a_{21}^{*}a_{12}^{*} + a_{11}^{*}a_{22}^{*})\varepsilon_{1,t-1}\varepsilon_{2,t-2} + a_{21}^{*}a_{22}^{*}\varepsilon_{2,t-1}^2 + g_{11}^{*}g_{12}^{*}h_{1,t-1}^2 + g_{11}^{*}g_{12}^{*}h_{1,t-1}^2 \\
&\quad + (g_{21}^{*}g_{12}^{*} + g_{11}^{*}g_{22}^{*})h_{1,t-1}h_{2,t-2} + g_{21}^{*}g_{22}^{*}h_{2,t-1}^2, \\
h_{22t} &= c_{22} + a_{12}^{*2}\varepsilon_{1,t-1}^2 + 2a_{12}^{*}a_{22}^{*}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{22}^{*2}\varepsilon_{2,t-1}^2 + g_{12}^{2*}h_{1,t-1} + 2g_{12}^{*}g_{22}^{*}h_{1,t-1}h_{2,t-1} + g_{22}^{*2}h_{2,t-1}^2.
\end{aligned}
$$

Comparison of (3.10) with (3.5) shows that the BEKK parameterization reduces the number of parameters, imposing restrictions both across and within equations, but without necessarily restricting the dynamics to a diagonal model. A general bivariate VEC$(1,1)$ model has 18 free parameters excluding the constants, the bivariate BEKK$(1,1,1)$ model has 8 free parameters excluding the constants. A BEKK$(1,1,1)$ model has $2n^2 + n(n+1)/2$ free parameters, a general VEC$(1,1)$ has $n(n+1)/2 + 2\left[n(n+1)/2\right]^2$ free parameters. Positive definiteness and identification of the BEKK$(1,1,1)$ are achieved under simple and straightforward conditions, which can be imposed during estimation relatively easily. Engle and Kroner show that a sufficient condition for identification is that the diagonal elements of C_0^* and a_{11}^* and g_{11}^* are also restricted to be positive. This is by no means the only possible condition that guarantees identification of the model. In fact looking at the model, it is clear that the only observationally equivalent structures are obtained replacing A_{11}^* with $-A_{11}^*$ or G_{11}^* with $-G_{11}^*$, so any condition that eliminates $-\ A_{11}^*$ and $-G_{11}^*$ from the set of admissible structures will suffice to guarantee identification. For example one could impose that a_{ij}^* and g_{kl}^* are positive for some given i,j and k,l. Non negativity restrictions can be imposed in estimation by estimating the square root of the restricted parameters, making identification relatively easy for estimation. Positive definiteness of the variance matrix is ensured by the decomposition of the constant matrix, C, into $C_0^{*'}C_0^*$ where C_0^* is a triangular matrix. This is an identifiable factorization of the constant matrix C that ensures positive definiteness by construction simply by assuming that the diagonal elements of C_0^* are non null.

In empirical applications, when the number of assets can be quite large, estimation of $2n^2 + n(n+1)/2$ parameters might become computationally heavy. Most empirical applications overcome this difficulty by restricting the BEKK$(1,1,1)$ model to a diagonal BEKK$(1,1,1)$, i.e. imposing that A^*_{1k} and G^*_{1k} are diagonal matrices. This of course reduces further the generality of the model. A diagonal BEKK$(1,1,1)$ is a restricted, less general DVEC model, positive definite by construction; the parameters of the covariance equation are products of the corresponding parameters of the two variance equations and the number of free parameters reduces further to $2n + n\,(n+1)\,/2$.

Full generality of the BEKK representation can be achieved adding more positive semidefinite terms to the variance equation, assuming $K > 1$. A full BEKK(p,q,K) has $(p+q)\,Kn^2 + n\,(n+1)\,/2$ free parameters to estimate. Engle and Kroner define full generality of the BEKK representation as its ability to be equivalent to as many VEC models as possible and show that the fully general BEKK representation spans the full set of symmetric positive definite VEC representations. In order to achieve full generality, a BEKK$(1,1,K)$ model should satisfy two necessary conditions. First, K, the generality parameter, should be such that the numbers of distinct parameters in each of the A^*_{1k} and G^*_{1k} is not less than $(n\,(n+1)\,/2)^2$; this ensures that no unnecessary restrictions are being imposed. Moreover it is required that there exists a A^*_{1k} matrix that contains either the pair of nonzero elements $(a_{il,k}, a_{mj,k})$ or the pair of non zero elements $(a_{jl,k}, a_{im,k})$ for all i, j, k, m from 1 to n; this guarantees that no implicit extra restrictions are imposed in the model. For the case of $n = 2$, if none of the A_{1k} matrices contains the pair $(a_{12,}a_{21})$, the second necessary condition is violated and this violation implies the restriction that the term $\varepsilon^2_{2,t-1}$ does not appear in the covariance equation. These conditions are necessary but not sufficient to achieve full generality in a BEKK$(1,1,K)$ model. Many different set of sufficient conditions are possible, however most of them generate an identification problem. In general whenever $K > 1$, some extra restrictions need to be imposed on the model to eliminate observationally equivalent structures. Engle and Kroner provide a condition that rules out all the observationally equivalent structures, while retaining the full generality of the BEKK representation (see Engle and Kroner, 1995, Proposition 2.3).

The mathematical relationship between the parameters of a general BEKK(p,q,K) and a VEC(p,q) model can be found by vectorizing both sides of a general BEKK(p,q,K) model

$$H_t = C_0^{*'} C_0^* + \sum_{k=1}^{K}\sum_{i=1}^{q} A_{ik}^{*'} \varepsilon_{t-i}\varepsilon_{t-i}' A_{ik}^* + \sum_{k=1}^{K}\sum_{j=1}^{p} G_{ik}^{*'} H_{t-i} G_{ik}^*, \tag{3.11}$$

obtaining

$$\begin{aligned} vech(H_t) \quad &= \quad (C_0^* \otimes C_0^*)' \, vech(I_n) + \sum_{i=1}^{q} \left[\sum_{k=1}^{K} (A_{ik}^* \otimes A_{ik}^*)' \, vech \left(\varepsilon_{t-i}\varepsilon_{t-i}^{'} \right) \right] \\ &\quad + \sum_{j=1}^{p} \left[\sum_{k=1}^{K} \left(G_{jk}^* \otimes G_{jk}^* \right)' vech \left(H_{t-j} \right) \right]. \end{aligned}$$

Representations (3.11) and $(3.4)$and are equivalent if and only if there exists matrices C_0^*, A_{ik}^* and G_{ik}^* such that

$$\begin{aligned} C_0 \quad &= \quad (C_0^* \otimes C_0^*)' \, vech(I_n), \\ A_i \quad &= \quad \sum_{k=1}^{K} (A_{ik}^* \otimes A_{ik}^*)', \\ G_i \quad &= \quad \sum_{k=1}^{K} (G_{ik}^* \otimes G_{ik}^*)'. \end{aligned}$$

This is a necessary and sufficient condition for the equivalence of the BEKK and VEC representation. The VEC models for which there exists no such C_0^*, A_{ik}^* and G_{ik}^* do not have a BEKK representation. Engle and Kroner (1995) show that this class includes all the non positive definite VEC parameterizations. However all positive definite symmetric VEC representations and all positive definite diagonal VEC representations admit C_0^*, A_{ik}^* and G_{ik}^* satisfying the above. The general diagonal BEKK representation,where each of the A_{ik}^* and G_{ik}^* matrices is diagonal, includes all the possible positive definite linear VEC model and if it satisfies the sufficient condition for generality, it is always identified.

Necessary and sufficient conditions for covariance stationarity of the general BEKK are found rewriting the general model:

$$vech(H_t) = (C_0^* \otimes C_0^*)' \, vech(I_n) + \sum_{i=1}^{q} \left[\sum_{k=1}^{K} (A_{ik}^* \otimes A_{ik}^*)' \, vech \left(\varepsilon_{t-i} \varepsilon_{t-i}^{'} \right) \right]$$
$$+ \sum_{j=1}^{p} \left[\sum_{k=1}^{K} \left(G_{jk}^* \otimes G_{jk}^* \right)' vech \left(H_{t-j} \right) \right],$$

in vector ARMA form

$$vech(H_t) = C_0 + A(L) vech \left(\varepsilon_t \varepsilon_t^{'} \right) + G(L) vech(H_t),$$

where L denotes the lag operator, and the polynomial $A(L)$ and $G(L)$ are defined by

$$\begin{aligned} A(L) &= \sum_{k=1}^{K} (A_{1k}^* \otimes A_{1k}^*)' L + ... + \sum_{k=1}^{K} \left(A_{qk}^* \otimes A_{qk}^* \right)' L^q, \\ G(L) &= \sum_{k=1}^{K} (A_{1k}^* \otimes A_{1k}^*)' L + ... + \sum_{k=1}^{K} \left(A_{pk}^* \otimes A_{pk}^* \right)' L^p. \end{aligned}$$

Then ε_t is covariance stationary if and only if all the eigenvalues of $\sum_{i=1}^{q} \sum_{k=1}^{K} (A_{ik}^* \otimes A_{ik}^*) + \sum_{j=1}^{p} \sum_{k=1}^{K} \left(G_{jk}^* \otimes G_{jk}^* \right)$ are less than one in modulus. In the diagonal BEKK model the covariance stationarity of the process is determined only by the diagonal elements of the A_{ik}^* and G_{ik}^* matrices, since the model is covariance stationary if and only if $\sum_{k=1} \left(a_{ii,k}^{*2} + g_{ii,k}^{*2} \right) < 1$ for all i.

Estimation of BEKK models is typically done with maximum likelihood methods. Assuming that the errors ε_t are $i.i.d$, the problem is to maximize the sample log likelihood function $L_T(\theta)$ for the T observations, conditional on some starting value for H_0 with respect to the vector of parameter θ. The general conditional log likelihood is:

$$L_T(\theta) = \sum_{t=1}^{T} \log f \left(y_t | \theta, F_{t-1} \right), \tag{3.12}$$

where $f(y_t|\theta, F_{t-1})$ is the conditional density of y_t,

$$f(y_t|\theta, F_{t-1}) = |H_t|^{-1|2} g \left(H_t^{-1/2} y_t | \varepsilon_t \right),$$

assuming correct specification of the model. The most commonly employed distribution for the errors is the multivariate normal distribution. Empirically there is a great amount of evidence that the standardized residuals of estimated volatility models are fat tailed, so the assumption of

Gaussianity of the innovations is not innocuous and reduces efficiency. Fiorentini et al. (2004) provided a general framework for ML estimation using the t distribution. Bauwen and Laurent (2002) extend their work to a multivariate skewed student distribution, Barndorff-Nielsen and Sheppard (2001) use a generalized hyperbolic distribution. The drawback of those approaches is that if the initial assumption on the distribution is wrong, in general ML estimates are not even consistent. On the other hand, using a Gaussian likelihood retains consistency also under misspecification of the conditional density, as long as the conditional mean and the conditional variance are correctly specified. Under the assumption of normality, the conditional log likelihood is

$$L_T(\theta) = c - \frac{1}{2}\sum_{t=1}^{T} \ln |H_T| - \frac{1}{2}\sum_{t=1}^{T} y_t' H_t^{-1} y_t.$$

Maximization of this likelihood brings two types of problems. First of all, as stated by Engle and Kroner (1995), the calculation of the derivatives of the model log likelihood with respect to the vector of parameters is quite cumbersome. Engle and Kroner suggest the use of numerical derivatives to approximate the score vector. However, subsequent work by Lucchetti (2002) and Hafner and Herwartz (2003) has shown that using analytical scores in the estimation procedure improves the accuracy of the estimates and speeds up convergence. When the dimension of the vector y_t is not small, the use of numerical derivatives makes estimation of the BEKK model slow and prone to numerical errors. The second issue is that, once the score vector is obtained, the parameter vector θ can be estimated only via non linear maximization, i.e. via numerical optimization through iterative methods. Among the several possible optimization algorithms, for sake of computational simplicity, Engle and Kroner (1995) advocate the use of the Bernd, Hall, Hall and Hausman (1976) algorithm. The BHHH algorithm is an iterative optimization method that calculates the updating term by a regression of a vector of ones on the scores. So the $i+1$ iteration is obtained as:

$$\theta^{i+1} = \theta^i + \lambda_i \left[(S'S)^{-1} \right]_{\theta=\theta^i} [S']_{\theta=\theta^i} \, \iota,$$

where ι indicates a vector of ones, $S \equiv \partial L_t(\theta)/\partial\theta$, i is the iteration number and λ_i is the step length, calculated at each iteration by a line search. The algorithm has the advantage that, under the assumption of normality, the $(S'S)^{-1}$ from the final iteration can be used as a consistent estimate of the variance covariance matrix of the parameters. Other popular iteration methods include the BFGS and the OPS algorithm. All these methods require to invert the conditional covariance matrix for all t at every iteration of the optimization and, when the dimension of y_t increases, this is time consuming and numerically unstable. For example a diagonal general BEKK model has $(p+q)\,Kn^2 + n(n+1)/2$ free parameters, and its estimation involves heavy computations due to several matrix inversions. Empirical applications overcome computational difficulties setting $p = q = K = 1$ and assume that H_1 is the unconditional covariance matrix. Under ergodicity of the process this has no consequence for the asymptotic properties of the estimator. Asymptotic properties of the QMLE estimator in multivariate GARCH models are not yet firmly established, and are difficult to derive from low level assumptions. Gourieraux (1997) establishes the weak consistency of the Quasi MLE for BEKK models, relying on the martingale difference properties of the sequence of score vectors evaluated at the true parameter value. Comte and Lieberman (2003) establish its strong consistency and asymptotic normality, verifying the conditions given by Jeantheau (1998) which do not impose any restriction on the derivative of the log likelihood of the process. Avarucci et al. (2013) derive the asymptotic properties of the estimator under weaker moments conditions in a simple BEKK$(1,1)$ model.

3.2.2 Factor models

Another very popular class of multivariate GARCH models in the literature is that of multivariate factor-GARCH models. As for the BEKK-type of models this class was motivated by the need to overcome the difficulties of the VEC specification. However, in contrast to the BEKK-type of models, the factor-GARCH models are not a restricted specification of the VEC model but are based on the idea that the dynamics of the conditional variance matrix are driven by the dynamics of small number of common underlying variables, called factors. All the models in this class express the observed series of returns $\mathbf{y}_t$ as a linear and invertible transformation of a small number of unobserved factors $\mathbf{f}_t$ which follow a GARCH process.

The first multivariate factor-GARCH model in the literature was introduced by Engle, Ng and Rothschild (1990). They assume that the series of returns can be expressed as

$$\mathbf{y}_t = \sum_{k=1}^{K} \mathbf{w}_k f_{kt} + \mathbf{e}_t, \tag{3.13}$$

where $\mathbf{w}_k$, $k = 1, ..., K$, are linearly independent $n \times 1$ vectors of factor weights, known as factor loadings, the f_{kt} are K not necessarily uncorrelated factors and $\mathbf{e}_t$ is a vector of idiosyncratic shocks with constant variance matrix and uncorrelated with the factors. It is assumed that the factors have a first-order GARCH structure, so their individual conditional variances, denoted as λ_{kt}^2 evolve according to:

$$\lambda_{k,t}^2 = \omega_k + \alpha_k \left(\mathbf{w}_k' \mathbf{y}_{t-1}\right)^2 + \beta_k \lambda_{k,t-1}^2, \tag{3.14}$$

where ω_k, α_k and β_k are scalar parameters, and their unconditional variances are normalized to one. This specification implies that the dynamics of conditional covariance matrix of $\mathbf{y}_t$ are expressed as:

$$H_t = \Omega + \sum_{k=1}^{K} \mathbf{w}_k \mathbf{w}_k' \lambda_{k,t}^2, \tag{3.15}$$

where Ω is the $n \times n$ positive definite constant covariance matrix of $\mathbf{e}_t$. Parsimony of parameterization is achieved by choosing the number of factors K to be much smaller than the number of assets n. Model (3.13)-(3.15) implies that the time varying part of $\mathbf{H}_t$ has reduced rank K, but $\mathbf{H}_t$ remains of full rank because Ω is assumed positive definite. Engle et al. propose a consistent but not efficient two-steps estimation method using maximum likelihood.

The assumption of correlation between the factors in model (3.13)-(3.15) turns out to be undesirable since it allows several of the factors to capture similar characteristic of the data, possibly increasing the number of factors in the model. Motivated by this consideration, most of the models in this class assume uncorrelated factors. Uncorrelated factors have a straightforward economic interpretation since they can be directly interpreted as different common components that drive the returns; moreover their use can potentially reduce the dimensionality problem. In all the uncorrelated GARCH-factor models the original series of returns is expressed as

$$\mathbf{y}_t = \mathbf{W}\mathbf{f}_t,$$

where $\mathbf{W}$ is a $n \times n$ non-singular matrix of factor loadings and $\mathbf{f}_t$ is a vector of $n \times 1$ heteroskedastic factors which are standardized to have unit unconditional variances, i.e. $E\left(\mathbf{f}_t\mathbf{f}_t'\right) = \mathbf{I}$. The unconditional variance matrix of the returns is expressed as

$$\mathbf{H}_t = \mathbf{W}\mathbf{H}_t^{\mathbf{f}}\mathbf{W}',$$

where $\mathbf{H}_t^{\mathbf{f}}$ is the unconditional covariance matrix of the factors.

Differences between the factor models are due to the specification of the linear transformation $\mathbf{W}$ and to whether the number of factors is less than the number of assets or not. Alexander and Chibumba (1997) propose an Orthogonal (O-) GARCH-factor model where the linear transformation $\mathbf{W}$ is assumed orthogonal and invertible; Van der Weide (2002) extends the O-GARCH model to a Generalized Orthogonal (GO-) GARCH factor model specifying the linear transformation by using the singular value decomposition of $\mathbf{E}\left(\mathbf{y}_t\mathbf{y}_t'\right) = \mathbf{W}\mathbf{W}'$, that is by assuming

$$\mathbf{W} = \mathbf{U}\mathbf{Q}^{1/2}\mathbf{V},$$

where the columns of $\mathbf{U}$ hold the eigenvectors of $\mathbf{E}\left(\mathbf{y}_t\mathbf{y}_t'\right)$, the diagonal matrix $\mathbf{Q}$ holds its eigenvalues and $\mathbf{V}$ is an orthogonal matrix of parameters. The conditional covariance matrix of the factors is defined as

$$\mathbf{H}_t^{\mathbf{f}} = (\mathbf{I} - \mathbf{A} - \mathbf{B}) + \mathbf{A}\odot\left(\mathbf{f}_{t-1}\mathbf{f}_{t-1}'\right) + \mathbf{B}\mathbf{H}_{t-1}^{\mathbf{f}},$$

where $\mathbf{A}$ and $\mathbf{B}$ are diagonal $n \times n$ parameter matrices and $\odot$ denotes the Hadamard, i.e. the element by element, product. Vrontos, Dellaportas, Politis (2003) introduce a Full Factor (FF-

) GARCH model, restricting the mapping $\mathbf{W}$ to a $n \times n$ invertible triangular parameter matrix with ones on the main diagonal and propose an estimation method for $\mathbf{W}$ that only exploits the conditional information. Lanne and Sakkonen (2007) propose a generalized orthogonal factor (GOF-) GARCH model where the mapping is decomposed using the polar decomposition

$$\mathbf{W} = \mathbf{C}\mathbf{V},$$

where $\mathbf{C}$ is a symmetric $n \times n$ matrix and $\mathbf{V}$ is an orthogonal $n \times n$ matrix. Since $\mathbf{E}\left(\mathbf{y}_t\mathbf{y}_t'\right) = \mathbf{W}\mathbf{W}' = \mathbf{C}\mathbf{C}'$, the matrix $\mathbf{C}$ can be estimated making use of the spectral decomposition $\mathbf{C} = \mathbf{U}\mathbf{Q}^{1/2}\mathbf{U}'$, where the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{E}\left(\mathbf{y}_t\mathbf{y}_t'\right)$ and the diagonal matrix $\mathbf{Q}$ contains its eigenvalues.

3.2.3 CCC models

Another popular class of multivariate GARCH models is the class of conditional variances and correlation models. These models are based on the decomposition of the conditional covariance matrix into conditional standard deviations and conditional correlations matrices. The form of the conditional variance matrix is specified as:

$$\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$$

where $\mathbf{D}_t = diag\left(h_{1t}^{1/2}, ..., h_{nt}^{1/2}\right)$ is a diagonal matrix that contains on the main diagonal the individual conditional standard deviations of each element of the vector $\mathbf{y}_t$. $\mathbf{R}_t = \left[\rho_{ij}\right]_t$ is a symmetric, positive definite matrix such that $\rho_{ii} = 1$ for every i and every t. The off-diagonal elements of $\mathbf{H}_t$ are defined by

$$\left[\mathbf{H}_t\right]_{ij} = h_{it}^{1/2}h_{jt}^{1/2}\rho_{ijt} \quad i \neq j, \tag{3.16}$$

the diagonal elements by

$$\left[\mathbf{H}_t\right]_{ii} = h_{it}.$$

The choice of $\mathbf{H}_t$ in these models entails first of all the specification of the individual conditional variance models, which need not to be the same across different assets, and are in general members of the class of univariate GARCH models, and then the choice of a conditional correlation matrix, $\mathbf{R}_t$, positive definite at every t. The first model of this class is the constant conditional correlation CCC-GARCH model of Bollerslev (1990). In the CCC model the conditional correlation matrix $\mathbf{R}$ is assumed to be time invariant and (3.16) simplifies to :

$$[\mathbf{H}_t]_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij} \quad i \neq j.$$

Often the individual conditional variances are modelled as individual GARCH(p,q) processes and the vector of conditional variances is written as

$$diag\left(\mathbf{H}_t\right) = \boldsymbol{\omega} + \sum_{i=1}^{q} \mathbf{A}_i \left(\boldsymbol{\varepsilon}_{t-i} \odot \boldsymbol{\varepsilon}_{t-i}\right) + \sum_{j=1}^{p} \mathbf{B}_j \mathbf{h}_{t-j} \tag{3.17}$$

where $\boldsymbol{\omega}$ is a n-dimensional vector of constants, $\mathbf{A}_i$ and $\mathbf{B}_j$ are $n \times n$ diagonal matrices of parameters, the symbol "$\odot$" denotes the Hadamard product, and the term $(\boldsymbol{\varepsilon}_t \odot \boldsymbol{\varepsilon}_t)$ is the $n \times 1$ vector with elements ε_{it}^2. Bollerslev (1990) assumes a GARCH$(1,1)$ structure for the individual variances and ensures positive definiteness of $\mathbf{H}_t$ assuming positive definiteness of $\mathbf{R}$ and that all the elements of ω, $\mathbf{A}_i$ and $\mathbf{B}_j$ are positive. When the individual variances follow a GARCH(p,q) model with either $p > 1$ or $q > 1$, or both, the condition that the elements of $\mathbf{A}_i$ and $\mathbf{B}_j$ must be positive can be replaced by any condition ensuring the positive definiteness of the individual variances. In general $\mathbf{H}_t$ is positive definite if and only if the matrix $\mathbf{R}$ is positive definite and all the n conditional variances are positive. A CCC model with GARCH(p,q) conditional variances contains $n + n\left(p+q\right) + n\left(n-1\right)/2$ parameters and does not allow individual volatilities to depend on each other. Jeantheau (1998) generalizes the CCC-GARCH model to the Extended Constant Conditional Correlation (ECCC-GARCH) model assuming non zero off-diagonal elements in $\mathbf{A}_i$ and $\mathbf{B}_j$. This allows past squared returns and variances of all the series to enter the individual conditional variance equation. For example in the first order ECCC-GARCH model, the ith variance equation is specified as

$$h_{it} = \omega_i + a_{11}\varepsilon_{1,t-1}^2 + ... + a_{1n}\varepsilon_{n,t-1}^2 + b_{11}h_{1,t-1} + ... + b_{1n}h_{n,t-1}, \tag{3.18}$$

and generates a much richer dependence structure than in a first order CCC model. An extended constant conditional correlation model with GARCH(p,q) conditional variances contains $n\left(1+1+np+nq+(n-1)/2\right)$ parameters.

Jeantheau (1998) shows that the necessary and sufficient condition for weak stationarity of a multivariate GARCH model established by Engle and Kroner (1995) implies strict stationarity and ergodicity of the ECCC model. In the CCC model where the matrices A_i and B_j are diagonal, this condition reduces to assuming that each diagonal element of the matrices is less than one in absolute value and it is easily imposed during estimation. Francq and Zakoian (2010) generalize the result of Bourgerol and Picard (1992) on strict stationarity of a univariate GARCH (p,q) to the ECCC model. As Bourgerol and Picard, they express the condition using the Lyapunov exponent of a matrix associated with the vector of parameter θ. To obtain the Markov-chain representation of the process they write it in vector representation

$$\bar{z}_t = \bar{b}_t + A_t \bar{z}_{t-1},$$

with

$$\bar{b}_t = \begin{pmatrix} \Upsilon_t \omega \\ 0 \\ . \\ . \\ \omega \\ . \\ . \\ 0 \end{pmatrix} \in R^{n(p+q)}, \qquad \bar{z}_t = \begin{pmatrix} \boldsymbol{\varepsilon}_t^{(2)} \\ . \\ . \\ \boldsymbol{\varepsilon}_{t-q+1}^{(2)}. \\ \mathbf{h}_t \\ . \\ . \\ \mathbf{h}_{t-p+1} \end{pmatrix} \in R^{n(p+q)},$$

and

$$A_t = \begin{pmatrix} \Upsilon_t A_1 & . & . & . & \Upsilon_t A_q & \Upsilon_t B_1 & . & . & . & \Upsilon_t B_p \\ I_n & 0 & . & . & 0 & 0 & . & . & . & 0 \\ 0 & I_n & . & . & 0 & 0 & . & . & . & 0 \\ . & . & . & . & . & . & & & & . \\ 0 & . & . & I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ A_1 & . & . & . & A_q & B_1 & . & . & . & B_p \\ 0 & . & . & . & 0 & I_n & & & & \\ 0 & . & . & . & 0 & & I_n & & & \\ . & . & . & . & . & & & . & & \\ 0 & . & . & . & 0 & 0 & 0 & . & I_n & 0 \end{pmatrix},$$

where

$$\Upsilon_t \equiv \left(\begin{array}{cccc} \left(R^{1/2}z_{1t}\right)^2 & 0 & ... & 0 \\ 0 & \left(R^{1/2}z_{2t}\right)^2 & ... & 0 \\ . & . & & \\ 0 & . & . & \left(R^{1/2}z_{mt}\right)^2 \end{array} \right)$$

and $\boldsymbol{\varepsilon}_t^{(2)}$denotes the vector $(\varepsilon_{1t}^2, ..., \varepsilon_{Nt}^2)'$. The ergodicity and strict stationarity of $\mathbf{h}_t$ and $\boldsymbol{\varepsilon}_t$ are derived from the ergodicity and strict stationarity of the Markov chain $\bar{z}_t$. Using results of Meyn and Tweedie (2009), Franq and Zakoian show that a necessary and sufficient condition for the existence of a strictly stationary and ergodic solution is that the top Lyapunov exponent of the sequence $\{A_t,\ t \in Z\}$ is strictly negative.

Conditions for identification of the ECCC model are derived by Jeantheau (1998), under the assumption that the ECCC-GARCH representation is "minimal". Using the back-shift operator L, (3.17) may be written as

$$B\left(L\right) \left(\begin{array}{c} h_{1t} \\ \cdot \\ h_{Nt} \end{array} \right) = \omega + A\left(L\right) \left(\begin{array}{c} \varepsilon_{1t-i}^2 \\ \cdot \\ \varepsilon_{Nt-i}^2 \end{array} \right),$$

with invertible $B\left(L\right) = I - \sum_{j=1}^{p} \mathbf{B}_j L^j$ and $A\left(L\right) = \sum_{i=1}^{q} \mathbf{A}_i L^i$. In order to ensure identification it is necessary to ensure that there exists no other pair $\left(A_\theta\left(L\right), B_\theta\left(L\right)\right)$ of polynomial matrices with the same degree (p,q) such that $B_\theta^{-1}\left(L\right) A_\theta\left(L\right) = B_{\theta_0}^{-1}\left(L\right) A_{\theta_0}\left(L\right)$. Jeantheau (1998) provides a set of necessary and sufficient conditions on the polynomial matrices $A\left(L\right)$ and $B\left(L\right)$ that rule out any observationally equivalent structure of the model, defining the ECCC-GARCH representation that satisfies those conditions "minimal".

A great advantage of the CCC model and its extensions is the straightforward interpretation of the parameters. For example this model is extremely popular in risk premium analysis because the estimated correlations between securities are readily available in the estimation results. Moreover the use of a CCC model facilitates the comparison of correlation patterns between different periods. One can estimate independently different CCC models in sub-periods of the available sample period and then examine if the correlation patterns vary substantially from one sub-period to another.

However in some empirical applications the assumption of constant conditional correlation might seem restrictive. For this reason, the CCC model has been generalized to a Dynamic Conditional Correlation (DCC) model, that retains the decomposition of the conditional covariance matrix but allows for a time varying conditional correlation matrix. Tse and Tsui (2002) proposed a Varying Correlation (VC-) GARCH model assuming that the conditional correlations are functions of their lagged values. The time varying correlation matrix is generated by the recursion:

$$R_t = (1 - \theta_1 - \theta_2) R + \theta_1 \Psi_{t-1} + \theta_2 R_{t-1}, \tag{3.19}$$

where the $n \times n$ matrix R is symmetric, and positive definite with elements on the main diagonal equal to one, θ_1 and θ_2 are non negative scalar parameters satisfying $\theta_1 + \theta_2 \leq 1$, and the matrix Ψ_{t-1} is the sample correlation matrix of the M-lagged standardized residuals, with elements

$$\psi_{ij,t-1} = \frac{\sum_{k=1}^{M} \left(\varepsilon_{i,t-k}/\sqrt{\hat{h}_{i,t-k}}\right)\left(\varepsilon_{j,t-k}/\sqrt{\hat{h}_{j,t-k}}\right)}{\sqrt{\left(\sum_{k=1}^{M} \varepsilon_{i,t-k}/\sqrt{\hat{h}_{i,t-k}}\right)^2 \left(\sum_{k=1}^{M} \varepsilon_{j,t-k}/\sqrt{\hat{h}_{j,t-k}}\right)^2}}.$$

The conditional correlation is formulated as the weighted sum of past correlations. A different specification is that of Engle and Sheppard (2001) and Engle (2002), where the time varying conditional correlation matrix is given by

$$R_t = diag\left(q_{11,t}^{-1/2}, ..., q_{NN,t}^{-1/2}\right) Q_t diag\left(q_{11,t}^{-1/2}, ..., q_{NN,t}^{-1/2}\right), \tag{3.20}$$

where the $n \times n$ symmetric positive definite matrix Q_t is set equal to

$$(1 - \alpha - \beta) Q + \alpha \left(D_{t-1}^{-1/2} \varepsilon_{t-1}\right)\left(D_{t-1}^{-1/2} \varepsilon_{t-1}\right)' + \beta Q_{t-1}, \tag{3.21}$$

where Q is the unconditional variance matrix of the standardized residuals, and α and β are non negative scalar parameters satisfying $\alpha + \beta < 1$. The elements of Q can be estimated or set directly to their empirical counterparts to simplify estimation. The drawback of this specification of the correlation matrix is that all the conditional correlations obey to the same dynamics. Engle (2002) suggests modifying (3.21) as

$$Q_t = Q \odot (\iota\iota' - A - B) + A \odot \left(D_{t-1}^{-\frac{1}{2}} \varepsilon_{t-1}\right)\left(D_{t-1}^{-\frac{1}{2}} \varepsilon_{t-1}\right)' + B \odot Q_{t-1}, \tag{3.22}$$

where ι is a vector of ones,the symbol "$\odot$" denotes the Hadamard product, and A and B are $n \times n$ matrices of parameters, which can be defined to functionally depend on a small fixed number of parameters. Billo et al. (2003) propose a Quadratic Flexible Dynamic Conditional Correlation model, where the correlation matrix has a block diagonal structure and the dynamics of the correlations are identical only within each block. The drawback of this approach is that the block members need to be defined a priori. Kwan et al. (2009) propose a threshold extension of the VC-GARCH model of Tse and Tsai, where the transition between regimes is governed by an indicator variable which belongs to the extended information set at time $t-1$, and the number of regimes is know a priori. Silvennoionnen and Terasvirta (2005) propose a Smooth Transition Conditional Correlation model, where the conditional correlation matrix varies between two states according to a transition variable.

Estimation of the CCC and ECC models via Quasi Maximum Likelihood methods is computationally quite attractive thanks to the decomposition of the conditional variance into conditional standard deviations and constant conditional correlations matrices. The log likelihood of the T observations of the process, conditional on some starting value H_0, has the simple form

$$L_T(\theta) = c - \frac{1}{2}\sum_{t=1}^{T} \ln|D_t(\theta_1)| - \frac{1}{2}\sum_{t=1}^{T} \log|R(\theta_2)| - \frac{1}{2}\sum_{t=1}^{T} y_t' \left(D_t^{-1}(\theta_1) R^{-1}(\theta_2) D_t^{-1}(\theta_1)\right)' y_t.$$

In sharp contrast with VEC and BEKK models, the T inversions of the matrix H_t reduce to only one inversion of the matrix R. Moreover the separate parameterization allows to write the likelihood as the sum of a volatility part, depending on the vector of unknown parameters θ_1, and a correlation part, depending on the vector of unknown parameter θ_2. As a consequence, the model can be consistently estimated using a two steps approach. Engle and Sheppard (2001) show that a consistent estimator of the parameters in θ_1 can be found by replacing the matrix R with the identity matrix in the likelihood,

$$L_T(\theta_1) = c - \frac{1}{2}\sum_{t=1}^{T} \ln|D_t(\theta_1)| - \frac{1}{2}\sum_{t=1}^{T} y_t' D_t^{-1}(\theta_1) D_t^{-1}(\theta_1) y_t,$$

which can be expressed as the sum of n individual log likelihood functions,

$$L_T(\theta_1) = c - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\ln h_{it} + \frac{y_{it}^2}{h_{ii}}\right).$$

Then a consistent estimator of θ_2 is obtained maximizing:

$$L_T\left(\theta_2|\hat{\theta}_1\right) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|R| + \left(D_t^{-1}y_t\right)' R\left(D_t^{-1}y_t\right)\right).$$

The estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are not fully efficient, since they are limited information methods. Engle and Sheppard suggest the use of $\hat{\theta}_1$ and $\hat{\theta}_2$ as starting values of a Netwon-Rapson iteration algorithm to maximize the full likelihood, and obtain an asymptotically efficient estimator after one iteration.

3.3 Multivariate Stochastic Volatility Models

A different approach to modelling conditional covariances is represented by multivariate Stochastic Volatility models, which have developed significantly over the past few years. In this class of models the conditional volatility process is no longer a measurable function of the sigma algebra of the information available at time $t-1$, but it is modelled as an unobservable, latent variable. GARCH-type and Stochastic Volatility models have similar statistical properties, however they are different with respect to the observability of the conditional variance at time $t-1$. The first multivariate stochastic volatility model was proposed by Harvey et al. (1994). The model contains an unobserved vector variance component, the logarithm of which is modelled directly as a vector linear process. The $n \times 1$ random vector

$$\mathbf{y}_t = \mu_t(\theta) + \boldsymbol{\varepsilon}_t(\theta),$$

has variance described by

$$\boldsymbol{\varepsilon}_t(\theta) = \mathbf{H}_t^{1/2} z_t, \tag{3.23}$$

$$\mathbf{H}_t^{1/2} = diag\left(\exp\{h_{1t}/2\}, ..., \exp\{h_{nt}/2\}\right) = diag\left\{\exp\left(\mathbf{h}_t/2\right)\right\}, \tag{3.24}$$

$$\mathbf{h}_t = \mu + \phi \odot \mathbf{h}_{t-1} + \boldsymbol{\eta}_t, \tag{3.25}$$

$$\begin{pmatrix} z_t \\ \eta_t \end{pmatrix} | \mathbf{h}_t \sim i.i.d\, N \left[\begin{matrix} 0 \\ 0 \end{matrix}, \begin{matrix} P_z & 0 \\ 0 & \Sigma_\eta \end{matrix} \right], \tag{3.26}$$

where the "exp" operator denotes the element by element exponentiation of a vector, the vector $\mathbf{h}_t = (h_{1t}, ..., h_{nt})'$ is a $n \times 1$ vector of unobserved log volatilities, and μ and ϕ are $n \times 1$ parameter vectors. The $n \times n$ positive definite matrix Σ_η is the variance covariance matrix of the disturbances of the volatility equation, and P_z is the covariance matrix of the disturbances of the level equations, with diagonal elements $\rho_{ii} = 1$ and off diagonal elements $\left|\rho_{ij}\right| < 1$ for any $i \neq j$, with $i, j = 1, ..., n$. The model has $2n + n^2$ free parameters. Harvey et al. assume a vector AR(1) specification for h_t which has been extended by Asai et al. (2006) to a VARMA(p, q),

$$\Phi(L)\, h_t = \mu + \Theta(L)\, \eta_t,$$

with

$$\begin{aligned} \Phi(L) &= I - \sum_{i=1}^{p} \Phi_i L^i, \\ \Theta(L) &= I - \sum_{j=1}^{q} \Theta_j L^j. \end{aligned}$$

In model (3.23) to (3.26), the individual log volatilities are not independent as long as the off diagonal elements of Σ_η are non zero. The model however does not allow the covariances to evolve over time independently from the variances. As CCC-GARCH models, this MSV model constraints the conditional correlation to be constant across time, but has the advantage of representing a better discrete time approximation of the continuous time Orstain-Uhlenbeck process used in finance theory. In sharp contrast with multivariate GARCH models where the conditional covariance matrix is measurable with respect to the sigma-algebra available at time $t-1$, $\mathbf{H}_t$ in (3.25) is not. This latent feature of $\mathbf{H}_t$ makes its positive definiteness more difficult to achieve than in MGARCH models. This is mostly the reason for which the MSV literature models directly the dynamics of the logarithmic transformation of $\mathbf{H}_t$, rather than of $\mathbf{H}_t$ itself. As for the MGARCH literature, the main concerns of the MSV literature have been to introduce parsimony of the parametric specification to simplify estimation and positive definiteness of the covariance matrix. Different variants of multivariate stochastic volatility models have been proposed to address these issues: factor models, time-varying correlation models and models based on the Chowlesky decomposition of the covariance matrix. Harvey (1994), Jacquier et al. (1999), Shephard (1996), Pitt and Shephard (1999), Aguilar and West (2000) proposed different specifications of MSV-factor models which decompose the returns into two additive components. The first has a smaller number of factors than the second and captures the information relevant for pricing of all the assets, while the second component is an idiosyncratic noise, which captures the asset specific information. The additive K factors MSV model is written as

$$\mathbf{y}_t = \mathbf{D}\mathbf{f}_t + \mathbf{e}_t,$$

where $\mathbf{f}_t$ is a $K \times 1$ vector of factors, K denotes the number of factors which is constrained to be smaller than the number of assets n, $\mathbf{D}$ is a $n \times K$ dimensional matrix of factor loadings and $\mathbf{e}_t \sim \mathbf{N}\left(0, diag\left\{\sigma_1^2, ..., \sigma_n^2\right\}\right)$. The individual factors evolve according to

$$\begin{aligned} f_{it} &= \exp\left(h_{it}/2\right)\varepsilon_{it}, \quad i = 1, ..., K \\ h_{it+1} &= \mu_i + \phi_t h_{it} + \eta_{it}, \end{aligned}$$

where ε_{it} and η_{it} are mutually independent shocks such that $\varepsilon_{it} \sim N(0,1)$ and $\eta_{it} \sim N\left(0, \sigma_\eta^2\right)$. In order to guarantee identification of the model, in general it is assumed that $\mathbf{D}_{ij} = 0$ and $\mathbf{D}_{ii} = 1$ for $i = 1, ..., n$. The variance of $\mathbf{y}_t$ is by construction positive definite with form

$$\mathbf{D}\mathbf{\Sigma}_f\mathbf{D}' + diag\left\{\sigma_1^2, ..., \sigma_n^2\right\},$$

where $\mathbf{\Sigma}_f$ is the covariance matrix of the factors. It can be shown (see Yu and Meyer, 2006) that additive MSV-factor models allow for both time-varying volatilities and correlations. However since the shock $\mathbf{e}_t$ is homoskedastic and the number of assets in the portfolio is greater than the number of factors, it can be shown that there must exist portfolios whose volatilities are homoskedastic. This feature does not seem to be consistent with empirical findings on portfolio volatilities and it is the main drawback of this type of models.

MSV time-varying correlation models are based on the idea that the dynamics of covariances and variances can be modelled separately. Asai and McAleer (2004) propose a MSV time-varying correlation model extending the DCC model of Engle (2002) to a stochastic volatility setting specifying the volatility matrix of $\mathbf{y}_t$ as

$$\mathbf{D}_t \Gamma_t \mathbf{D}_t,$$

where the diagonal matrix $\mathbf{D}_t$ is defined as

$$\mathbf{D}_t = diag\left\{(\exp\left\{h_{1t}/2\right\}, ..., \exp\left\{h_{nt}/2\right\}\right) = diag\left\{\exp\left(\mathbf{h}_t/2\right)\right\},$$

and the time-varying correlation matrix Γ_t is specified as

$$\mathbf{\Gamma}_t = Q_t^{*-1} Q_t Q_t^{*-1},$$

where $Q_t^* = \left(diag\left\{vecd\left(Q_t\right)\right\}\right)^{1/2}$and $vecd$ creates a vector from the diagonal element of a matrix and

$$\begin{aligned} Q_{t+1} &= (1-\psi)\bar{Q} + \psi Q_1 + \Xi_1, \\ \Xi_1 &\sim W_n\left(v, \Lambda\right), \end{aligned}$$

where $\bar{Q}$ is the unconditional variance matrix of the standardized residuals and $W_n\left(v, \Lambda\right)$ denotes a Wishart distribution. If $\bar{Q}$ is positive definite and the scalar parameter ψ is such that $|\psi| < 1$ then time-varying correlation matrix is positive definite and stationary. In the special case in which $v = 1$, Ξ_t can be expressed as the cross product of a multivariate normal distribution with zero mean and covariance given by Λ.

Gourieroux (2006) and Philipov and Glickman (2004) developed a different type of dynamic MSV models based on the Wishart autoregressive (WAR) multivariate process. The time varying covariance matrix of $\mathbf{y}_t$ is defined as a WAR(p) process

$$\mathbf{H}_t = \sum_{k=1}^{K} \mathbf{x}_{kt}\mathbf{x}_{kt}',$$

where $K > n - 1$ and each vector $\mathbf{x}_{kt}$ follows a VAR(p) model given by

$$\mathbf{x}_{kt} = \sum_{i=1}^{p} A_i \mathbf{x}_{kt-i} + \varepsilon_{kt}, \quad \varepsilon_{kt} \sim N\left(0, \Sigma\right).$$

This type of MSV models has been widely used in empirical applications since it offers closed-form derivative prices which are employed in a number of financial problems, such as term structure of T-bonds and corporate bonds and structural models for credit risk.

Estimation of MSV models is not straightforward because of the difficulties involved in evaluating their likelihood. Feasible estimation strategies for MSV are method of moment estimation or estimation via quasi maximum likelihood of a linear state-space representation of the model via the Kalaman filter. These procedures are computationally very simple and have been extensively used in empirical applications, however they have poor finite sample properties and suboptimal efficiency with respect to direct maximum likelihood estimation of the model. Over the last few years, these approaches have been replaced by simulation-based methods (see Shephard and Pitt (1997), Durbin and Koopman (1997), Kim et al. (1998), Sandmann and Koopman (1998), and Chib et al. (2002)) that have successfully dealt with numerical evaluation of the high dimensional integrals.

3.4 Asymmetric Multivariate Volatility Models

It has long be recognized that the volatility of stock returns responds differently to good news and bad news. In particular, while bad news tends to increase the future volatility, good news of the same size will increase the future volatility by a smaller amount or might even decrease it. The different impact of past price decreases and past price increases of the same magnitude on the current volatility is know as asymmetry. The negative correlation between past returns and current volatility is know as leverage effect. Therefore leverage denotes asymmetry but not all the asymmetric effects display leverage. In the class of univariate ARCH specifications the most popular models that capture asymmetric effects are the EGARCH model, the threshold GARCH model, and its most popular variant GJR-GARCH, and the class of asymmetric power GARCH models. Many empirical studies have found robust evidences of asymmetries in multivariate stock returns series. As a consequence in the last decade there have been quite a few attempts to introduce asymmetry and leverage in multivariate random variance model. The analysis of asymmetric effects in conditional volatility models is a relatively new topic. Sentana (1995) proposed a multivariate latent factor model with QARCH-type effects on the underlying factors that capture leverage through the common latent factor. Kroner and Ng (1998) introduced a General Dynamic Covariance model to capture the covariance asymmetry in the volatility dynamics of portfolios of small and large firms. More recently Shepard (2002) and Cappiello, Engle and Sheppard (2006) proposed models which allow for asymmetric dynamics in the conditional variances as well as in the conditional correlations, based on a generalization of the multivariate Dynamic Conditional Correlation GARCH model of Engle (2002). Audrino and Barone-Adesi (2006) introduced a semiparametric multivariate GARCH model to allow for asymmetric conditional covariances and time varying conditional correlations. Dellaportas and Vrontos (2007) introduced a new class of multivariate threshold GARCH models based on a binary tree approach, where every terminal node parametrizes a local multivariate GARCH model for a specific partition of the data. Haas, Mittnik and Paolella (2008) proposed an asymmetric multivariate generalization of the class of normal mixture GARCH models. Finally Kawakatsu (2006) introduced the matrix exponential GARCH model, the only multivariate conditional volatility model with exponential specification.

The asymmetric property of stochastic volatility models is based on the direct correlation between the innovations of the mean and volatility equations. Danielsson (1998) introduced leverage effect in MSV model based on the specification considered by Harvey et al (1994) including a negative correlation between the returns and volatility innovations. Asai and McAleer (2005) proposed a multivariate dynamic asymmetric leverage (DAL) model that accommodates threshold effects. The model allows volatility to undergo discrete shifts depending on whether the return for the previous period is above or below some threshold value. Asai and McAleer (2006) used the numerical Monte Carlo Likelihood method proposed by Durbin and Koopman (1997) to estimate the basic MSV model with leverage of Danielsson (1998). Moreover they extended the model of Danielsson to incorporate the effect of the size and magnitude of the previous return into the volatility equation by using the absolute value function. Their estimation results of the MSV with leverage and size effect (SV-LSE) model show that the models fits the bivariate and multivariate returns of the S&P500, Nikkei 225 and Hang Seng indexes more accurately than any other MSV asymmetric model available.

In chapters 4 and 5 of this thesis we introduce a new multivariate volatility model, the multivariate Exponential Volatility (MEV) model, which is based on an exponential specification that allows to nest stochastic and heteroskedastic volatility specifications in the same framework. As explained in greater details in Chapter 1, the model was partly motivated by the need to capture asymmetries and leverage effect in a multivariate exponential specification that easily grants positive definiteness of the covariance matrix. It is therefore relevant to our purposes to discuss in some details the matrix exponential GARCH model of Kawawatsu (2006), which is the only MGARCH model with exponential specification, asymmetries and leverage effects. The next section briefly introduces Kawakatsu's model and discusses the main difficulties arising from its estimation via maximum likelihood methods which partly motivate estimation of the MEV model parameters in chapters 4 and 5 via Whittle methods.

3.4.1 Matrix Exponential GARCH.

The matrix exponential GARCH model of Kawakatsu (2006) specifies the dynamics in the logarithm of the conditional covariance matrix. The model is an extension of Nelson (1991) univariate EGARCH model to a multivariate model of the VEC class. The exponentiation is not the element by element exponentiation of the conditional covariance matrix, but a non linear matrix exponential transformation of the conditional covariance matrix. This specification ensures positive definiteness of the covariance matrix through the exponential transformation, and allows for a term that captures multivariate asymmetries. In the general formulation of the matrix exponential GARCH, the logarithm of the covariance matrix depends on its own past and on lagged innovations, according to

$$(\log H_t - C) = \sum_{j=1}^{p} \tilde{A}_j \odot \log \bar{H}_{t-j} + \sum_{j=1}^{q} \sum_{i=1}^{k} \tilde{B}_{ij} \varepsilon_{i,t-j} + \sum_{j=1}^{q} \sum_{i=1}^{k} \tilde{F}_{ij} \left(|\varepsilon_{i,t-j}| - E\,|\varepsilon_{i,t-j}| \right), \quad (3.27)$$

where C is a symmetric $k \times k$ matrix of constants, and A_j, B_{ij} and F_{ij} are $k \times k$ symmetric parameter matrices. This specification excludes cross products terms such as $\varepsilon_{i,t-1}\varepsilon_{j,t-1}$ and has a total of $k\,(k+1)/2 \times (1 + p + 2kq)$ free parameters. The term F_{ij} captures the asymmetries of the volatility process. A different specification of (3.27) is written using the $vech$ operator,

$$\bar{h}_t = vech\,(\log H_t - C) = \sum_{i=1}^{p} A_i \bar{h}_{t-1} + \sum_{j=1}^{q} B_j \varepsilon_{t-j} + \sum_{j=1}^{q} F_j \left(|\varepsilon_{t-j}| - E\,|\varepsilon_{t-j}| \right), \quad (3.28)$$

where A_i, B_j and F_j are parameter matrices of dimension respectively $k^* \times k^*$, $k^* \times k$, $k^* \times k$, with $k^* = k\,(k+1)/2$. The number of free parameters in this model is $k^* + {k^*}^2 p + 2kk^*q$. The $vech$ specification does not require any symmetry constraints, however it increases the number of parameters to estimate because it allows for a richer dependence structure of the volatilities. In (3.28) the volatilities don't depend on their own past values and the past values of the covariances, but also on the past values of all the other individual variances. Parameters restrictions to ensure positive definiteness are not required in (3.27) nor (3.28) since $\log H_t$ does not need to be a positive definite matrix. However neither (3.27) nor (3.28) are feasible when $k > 3$, due to the very large number of parameters to estimate. To deal with the curse of dimensionality, Kawakatsu proposes a restricted diagonal matrix exponential GARCH, where each element of $(\log H_t - C)$ evolves according to

$$\bar{h}_{rc,t} = \sum_{i=1}^{p} a_{i,rc}\bar{h}_{rc,t-i} + \sum_{j=1}^{q} \left(b_{jr,rc}\varepsilon_{r,t-j} + b_{jc,rc}\varepsilon_{c,t-j}\right) \tag{3.29}$$

$$+ \sum_{j=1}^{q} \left(f_{jr,rc}\left(|\varepsilon_{r,t-j}| - E\left[|\varepsilon_{r,t-j}|\right]\right) + f_{jc,rc}\left(|\varepsilon_{c,t-j}| - E\left[|\varepsilon_{c,t-j}|\right]\right)\right).$$

The (r, c) element of $(\log H_t - C)$ depends only on its own lagged values and lagged rth and cth innovations. In this diagonal specification the number of parameters reduces to $k^* \times (1 + p + 2k^2 q)$. Estimation of the matrix exponential GARCH model has some serious drawbacks. Kawakatsu proposes estimation of the parameters via Maximum Likelihood methods. However MLE estimation of an exponential matrix model is computationally very costly. Moreover evaluation of the derivative of the exponential of a non symmetric matrix is a computationally unstable (see Moler and Van Loan, 2003). The asymptotic properties of the MLE estimator for exponential models are not established in the literature as they require the invertibility of the model, which is extremely difficult to establish in exponential specifications (see Straumann and Mikosh, 2006).

Chapter 4 Whittle estimation of multivariate exponential volatility models

4.1 Introduction

In the previous chapter, we discussed the two main approaches to modelling multivariate volatility, i.e. the conditional volatility and the stochastic volatility approaches. In stochastic volatility models the mean equation and the volatility equation are driven by two separate shocks. In earlier specifications the two shocks are set independent, in the more recent literature the independence is relented to allow for asymmetries and leverage effects. However the two shocks are distinct and unobservable. Multivariate GARCH models, on the other hand, regardless of the specification of the conditional volatility matrix, are "one-shock" models. The aim of this chapter is to propose a class of multivariate volatility models that encompasses both "one-shock" and "two-shocks" specifications. The idea of a parameterization that nests conditional and stochastic volatility specifications can be traced back to the work of Robinson and Zaffaroni (1997, 1998). In the context of univariate volatility, Robinson and Zaffaroni introduced the nonlinear moving average class as an alternative to the ARCH(∞) class. In the univariate non-linear moving average model, the volatility evolves according to

$$h_t = \rho + \sum_{i=1}^{\infty} \alpha_i \epsilon_{t-i} \quad \sum_{i=1}^{\infty} \alpha_i^2 < \infty, \tag{4.1}$$

and the mean equation is either

$$x_t = \eta_t h_t, \tag{4.2}$$

or

$$x_t = \epsilon_t h_t, \tag{4.3}$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are sequences of zero mean $i.i.d$ random variables, independent from each other. When the level equation and the volatility equation are driven by the same shock as in (4.3) and (4.1) , the model is a "one-shock" model, when instead the shocks are distinct as in (4.2) and

(4.1), the model is a "two-shocks" model. The nonlinear moving average model was originally introduced to allow for long memory dependence in the squares of a weak dependent process. It was prompted by the consideration that financial data can have small sample autocovariances, whereas certain nonlinear functions such as squares have sample correlations that die away very slowly. Under (4.2) or (4.3) and (4.1) x_t is white noise, but if the α_i decay suitably slowly the volatility has long memory and the squares x_t^2 may also have long memory. Robinson and Zaffaroni consider a number of properties of this class of models and consider statistical inference for parameterizations allowing for long memory autocorrelated squares. They stress frequency domain Gaussian estimates in the sense of Whittle (1962) as maximum likelihood estimates are computationally very cumbersome and their asymptotic properties are extremely difficult to derive. Zaffaroni (2009) extended the nonlinear moving average model to an exponential specification, introducing a parametrization encompassing both the univariate EGARCH model of Nelson (1991) and the Stochastic Volatility model of Taylor (1986). He considers an observable satisfying

$$x_t = z_t e^{0.5h_t}, \quad t \in Z, \tag{4.4}$$

and

$$h_t = \omega_0 + \sum_{k=0}^{\infty} \psi_{0k} \epsilon_{t-k-1} \quad \text{a.s,} \quad \sum_{j=0}^{\infty} \psi_{0j}^2 < \infty \tag{4.5}$$

where the $\{z_t, \epsilon_t\}$ form a sequence of $i.i.d$ unobservable random variables which, for some $t = s$, might be correlated. Model (4.4) and (4.5) nests a "two-shocks" specification, where the shocks need not to be independent, and a "one-shock" specification, when $\epsilon_t = \epsilon(z_t)$ for some instantaneous transformation $\epsilon(.)$. It includes a large number of different exponential volatility specifications, such as the "one-shock" EGARCH and FIEGARCH models and the "two-shocks" short and long memory asymmetric volatility models of Harvey and Shephard (1996), Ruiz and Veiga (2006). Zaffaroni establishes the strong consistency and the asymptotic normality of the Whittle estimator under a set of regularity conditions general enough to allow for long memory dependence in the squares of the process and designed to apply to both classes of models. The exponential specification of volatility has several well-known advantages. For example non-negativity constraints on the parameters need not to be imposed thus permitting a wide range of cyclical behavior in the conditional variance. Moreover asymmetric effects, leading to different response of volatility to good and bad news, are easily parametrized. As noted in Asai (2006), exponential volatility models offer the best discrete approximation to continuous time asset pricing models. These advantages advocate for an extension of the exponential volatility specification to multivariate dimensions. In stochastic volatility models such generalization is quite straightforward and very well established, on the other hand, in multivariate conditional volatility models, the exponential specification is not very popular. The matrix exponential GARCH model of Kawakatsu (2006) is currently the only "one-shock" multivariate exponential volatility model in the literature. The conditional variance dynamics are specified in the matrix logarithm of the conditional covariance, and the model maintains positive definiteness of the conditional covariance matrix with no need of parameters constraints. However it is not clear if the parameterization could be generalized to include long memory dependence in the volatility. Estimation of the matrix exponential GARCH model has several drawbacks as numerically stable evaluation of the exponential of a matrix is a very delicate issue. Moreover there is a complete lack of asymptotic distribution theory for the PMLE or any other estimator of this model.

In the next chapter we introduce a class of multivariate exponential volatility models that encompasses both "one-shock" and "two-shock" specifications and allows for a wide range of degree of persistence of shocks to the conditional variance. The exponentiation is the element by element exponentiation of the diagonal variance matrix, along the line of Harvey et al. (1994). This specification permits greater flexibility than the matrix exponential one and offers the considerable advantage to encompass multivariate asymmetric Stochastic Volatility models. In the literature there are few examples of parameterizations that nest different type of multivariate models. The more comprehensive one is perhaps the Asymmetric Dynamic Covariance matrix (ADC) model of Kroner and Ng (1998), which reduces to different multivariate GARCH models under different combinations of the initial set of conditions. To our knowledge however, nesting multivariate stochastic and conditional volatility models has not been attempted yet. In the next chapter we impose exponential decay in the autocovariances of the squares of the observables and address weakly dependent parameterizations of the model. The next section introduces the model with some discussion. Section 4.3 considers statistical inference in case of finite parameterization and advocates the use of the Whittle estimator. Section 4.4 lists a first set of regularity conditions and derives the strong consistency of the estimator. The last section reinforces the assumptions and establishes the asymptotic normality of the Whittle estimates in both "one-shock" and "two-shock" specifications.

4.2 The Multivariate Exponential Volatility model

Consider an observable vector stochastic process $\mathbf{x}_t$ of dimension $n \times 1$. Let $\mathbf{z}_t$ be a sequence of unobservable real-valued independent identically distributed $(i.i.d)$ random vectors of dimension $n \times 1$. We assume that $\mathbf{z}_t$ has zero mean and positive definite variance matrix $\sum_z$ with elements on the main diagonal normalized to one. Let $\boldsymbol{\epsilon}_t$ be a sequence of $n \times 1$ independent identically distributed unobservable real-valued random vectors with zero mean and positive definite variance matrix $\sum_\epsilon$with elements on the main diagonal normalized to one. The innovations $\mathbf{u}_t' = (\mathbf{z}_t', \boldsymbol{\epsilon}_t')$ form a sequence of unobservable $i.i.d.$ random vectors of dimension $2n \times 1$. We allow $\mathbf{z}_t$ and $\boldsymbol{\epsilon}_s$ to be correlated for some $t = s$ and denote with $\sum_{\epsilon z}$ their covariance matrix at time $t = s$, we do not require this matrix to be diagonal. For every $t \neq s$, $\mathbf{z}_t$ and $\boldsymbol{\epsilon}_s$ are assumed uncorrelated. In what follows we indicate by F_t the sigma field generated by the past information until time t and by $I(A)$ the indicator of the set A. We define $\exp(x)$ as the element by element exponentiation operator and $\ln(x)$ as element by element logarithmic operator. We denote the ath element of any vector $\mathbf{x}_t$ by $\mathbf{x}_t^{(a)}$ and the (a,b) element of any matrix A by $A^{(\alpha,b)}$ and the (a,a) element of any diagonal matrix Ψ by $\Psi^{(a)}$. Our interest lies in the zero mean Multivariate Exponential Volatility (MEV) process, defined by the following equations:

$$\mathbf{x}_t = \mathbf{D}_t^{\frac{1}{2}}(\theta)\,\mathbf{z}_t, \tag{4.6}$$

$$\mathbf{D}_t^{\frac{1}{2}} = diag\left\{\exp(\mathbf{h}_t^{(a)}/2)\right\}, \quad t = 1,...,T, \quad a = 1,..n, \tag{4.7}$$

$$\mathbf{h}_t = \boldsymbol{\omega}_0 + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{0j}\boldsymbol{\epsilon}_{t-j-1} \quad \sum_{j=0}^{\infty} \|\boldsymbol{\Psi}_{0j}\|^2 < \infty, \quad \text{a.s.}, \tag{4.8}$$

$$E\left(\mathbf{z}_t'\boldsymbol{\epsilon}_t\right) = \begin{pmatrix} \sum_{0z} & \sum_{0z\epsilon} \\ \sum_{0z\epsilon} & \sum_{0\epsilon} \end{pmatrix}. \tag{4.9}$$

The vector $\boldsymbol{\omega}$ is a $n \times 1$ vector of constants and the sequence $\{\boldsymbol{\Psi}_{0j}\}_{j=0}^{\infty}$ is a sequence of diagonal and square summable $n \times n$ parameter matrices, i.e. $\boldsymbol{\Psi}_k \equiv diag\{\psi_{1k}, \psi_{2k}, ..., \psi_{nk}\}$. The individual log volatilities $\mathbf{h}_t^{(a)}$ evolve according to

$$\mathbf{h}_t^{(a)} = \boldsymbol{\omega}_0^{(a)} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{0j}^{(a)}\epsilon_{t-1-j}^{(a)} \quad \text{a.s.} \tag{4.10}$$

Equations (4.6) to (4.10) represent a class of multivariate exponential volatility processes that encompasses both "one-shock" and "two-shocks" specifications. When $\boldsymbol{\epsilon}_t \equiv \boldsymbol{\epsilon}(\mathbf{z}_t)$ for some instantaneous transformation $\boldsymbol{\epsilon}(.)$, class (4.6)-(4.10) yields a Multivariate GARCH model. In this case the level equation shocks, $\{\mathbf{z}_t\}_{t=1}^{\infty}$, drive also the evolution of the log volatilities, according to

$$\mathbf{h}_t = \boldsymbol{\omega} + \sum_{j=0}^{\infty} \Psi_j\,\boldsymbol{\epsilon}(\mathbf{z}_{t-j-1}), \quad \text{a.s.}, \quad \sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty, \tag{4.11}$$

and the conditional variance matrix of $\mathbf{x}_t$, $\mathbf{var}(\mathbf{x}_t|F_{t-1})$, is easily obtained as

$$\mathbf{H}_t = \mathbf{D}_t^{\frac{1}{2}}\boldsymbol{\Sigma}_z\mathbf{D}_t^{\frac{1}{2}}.$$

As in the Constant Conditional Correlation model of Bollerslev (1990) the conditional variance matrix is expressed as the product of a time varying conditional standard deviation matrix $\mathbf{D}_t^{\frac{1}{2}}$ and a constant conditional correlation matrix $\mathbf{\Sigma}_z$ with diagonal elements normalized to one. However, the individual conditional variances evolve according to an exponential specification rather then the standard GARCH(p,q) specification of CCC models. As a consequence, positive definiteness of $\mathbf{H}_t$ follows from positive definiteness of $\mathbf{\Sigma}_z$ with no need of constraints on the conditional standard deviation matrix. The off diagonal elements of the conditional covariance matrix are specified as

$$[\mathbf{H}_t]_{(a,b)} = \sigma_{(z)} \exp\left(\mathbf{h}_t^{(a)}/2\right) \exp\left(\mathbf{h}_t^{(b)}/2\right),$$

where $\sigma_{(z)}$ is the (a,b) element of the matrix $\mathbf{\Sigma}_z$. The specification of the function $\epsilon\left(.\right)$ is general enough to yield different evolution patterns of the individual conditional volatilities in the same model. For example, if one specifies some transformations $\boldsymbol{\epsilon}\left(.\right)$ as

$$\boldsymbol{\epsilon}_t^{(a)} = \boldsymbol{\epsilon}\left(\mathbf{z}_t^{(a)}\right) = \theta_0^{(a)}\mathbf{z}_t^{(a)} + \delta^{(a)}\left(\left|\mathbf{z}_t^{(a)}\right| - \mu_{\left|z_{(a)}\right|}\right),$$

the corresponding volatilities follow EGARCH specifications

$$\mathbf{h}_t^{(a)} = \boldsymbol{\omega} + \sum_{j=0}^{\infty} \Psi_j^{(a)}\left(\theta_0^{(a)}\mathbf{z}_{t-j-1}^{(a)} + \delta^{(a)}\left(\left|\mathbf{z}_{t-j-1}^{(a)}\right| - \mu_{\left|z_{(a)}\right|}\right)\right),$$

which, in practical applications, can be parametrized as ARMA processes of different (p,q) orders,

$$\mathbf{h}_t^{(a)} = \boldsymbol{\omega} + \frac{\left(1 + \beta_1 L + ... + \beta_p L^p\right)}{\left(1 - \Delta_1 L + ... + \Delta_q L^q\right)}\boldsymbol{\epsilon}\left(\mathbf{z}_{t-1}^{(a)}\right),$$

and exhibit different degrees of leverage. In the same model, one might specify other transformations $\boldsymbol{\epsilon}\left(.\right)$ as

$$\boldsymbol{\epsilon}_t^{(b)} = \boldsymbol{\epsilon}\left(\mathbf{z}_t^{(b)}\right) = \theta^{(b)}\mathbf{z}_t^{(b)} + \delta^{(b)}\left(\mathbf{z}_t^{(b)} I_{\left(z_t^{(b)}>0\right)} - \mu_{z_{(b)}} I_{\left(z_{t-j}^{(b)}>0\right)}\right),$$

allowing the corresponding volatilities to evolve according to the GJR-GARCH (Glosten, Jagannathan, Runkle) specification

$$\mathbf{h}_t^{(b)} = \boldsymbol{\omega} + \sum_{j=0}^{\infty} \Psi_j^{(b)}\left(\theta^{(b)}\mathbf{z}_{t-j}^{(b)} + \delta^{(b)}\left(\mathbf{z}_{t-j}^{(b)} I_{\left(z_{t-j}^{(b)}>0\right)} - \mu_{z_{(b)}} I_{\left(z_{t-j}^{(b)}>0\right)}\right)\right)$$

and displaying a different kind of asymmetric behavior. Indirect spillover effects between different assets are introduced through the simultaneous correlation of different mean shocks which implies that shocks to the return of asset b at time t are correlated with variation of volatility of asset a, $\mathbf{h}_t^{(a)} - \mathbf{h}_{t-1}^{(a)}$. On the other end, in its "two-shocks" formulation the model is a multivariate Stochastic Volatility model with leverage. The volatility shocks are no longer specified as known functions of the mean shocks, implying that the volatility is latent. Asymmetries and leverage effects in individual volatilities are introduced by assuming non zero simultaneous correlation between the level and the volatility shocks of the same asset. This induces correlation between the return $\mathbf{x}_t^{(a)}$ at time t and the volatility variation of the same asset, $\mathbf{h}_t^{(a)} - \mathbf{h}_{t-1}^{(a)}$. The model includes indirect spillover effects between different assets by allowing for non zero simultaneous correlation between the level and the volatility shocks of different assets. Asymmetric multivariate Stochastic Volatility models in the literature, such as the MSV-L of Asai and McAleer (2005), specify the volatility equation rather than an MA(∞) process as an autoregressive process of order one,

$$\mathbf{h}_t = \boldsymbol{\omega} + \boldsymbol{\phi} \circ \mathbf{h}_{t-1} + \boldsymbol{\eta}_t,$$

where $\boldsymbol{\phi}$ is a $n \times 1$ vector of parameters all satisfying $\left|\boldsymbol{\phi}^{(a)}\right| < 1$ and the operator $\circ$ denotes the Hadamard product. Furthermore they assume conditional joint Gaussianity of the innovations $\mathbf{u}_t'$ $= (\mathbf{z}_t', \boldsymbol{\epsilon}_t')$,

$$\begin{pmatrix} \mathbf{z}_t \\ \boldsymbol{\epsilon}_t \end{pmatrix} |\mathbf{h}_t \sim \mathbf{N}_{2n} \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sum_z & \sum_{z\varepsilon} \\ \sum_{z\varepsilon} & \sum_\epsilon \end{pmatrix} \right],$$

and restrict the matrix $\sum_{z\varepsilon}$ to diagonal. We do not assume the normality anywhere in our results. However it is to be noted that we do not allow shocks at time t to affect the log volatility at time t, thus while in the two-shocks formulation of the model the volatility is latent, the levels of observable $\{\mathbf{x}_t\}$ follow a martingale difference process.

Estimation of multivariate volatility models is generally based on quasi maximum likelihood methods. However asymptotic properties of the MLE estimator for exponential volatility models have not been established in the literature. A necessary condition for the observed likelihood to behave well asymptotically is invertibility of the model, which guarantees that the likelihood will not explode nor converge to zero for any value of the parameters in the parameter space. Invertibility is achieved when the mean value shocks $\mathbf{z}_t$ can be expressed as as a convergent function of the observations $\mathbf{x}_s$ for $s \leq t$; however in the exponential specification the standardized mean shocks can only be computed as $\hat{\mathbf{z}}_t(\theta) = \mathbf{x}_t / \exp\left\{\hat{h}_t(\theta)/2\right\}$ where $\hat{h}_t(\theta)$ is a function of the $\mathbf{x}_s$ as well. This recursiveness makes extremely difficult to establish the uniform convergence of the Hessian matrix in a neighborhood of the true parameter. Some specific results on the asymptotics of MLE for univariate exponential volatility models are available under highly specific assumptions that cannot readily be verified. Straumann and Mikosch (2005) provided a sufficient condition for invertibility of a low order univariate EGARCH. However they suggest that this condition is practically infeasible except when h_t is correlated with h_{t-1} but not with h_{t-s} for any $s \geq 2$. For an EGARCH (1,1), Demos and Kyriakopoulou (2014) present sufficient conditions for the supremum norm of the second order derivative of the likelihood to be finite, however these conditions restrict the admissible parameter space and are extremely difficult to verify. Kawakatsu (2006) proposes maximum likelihood estimation of the Matrix Exponential GARCH, but he does not establish its asymptotic properties. Estimation of multivariate Stochastic Volatility models has the added difficulty of the likelihood function evaluation arising from the latency of the volatility process. Because volatility is latent, in order to derive the likelihood, the vector of unobserved volatilities has to be integrated out of the joint probability distribution. This implies the evaluation of a T dimensional integral, which requires numerical methods. In recent years, the empirical literature on stochastic volatility has developed different ways of dealing with this issue, for example introducing MCMC and SML methods for numerical evaluation of the likelihood, or implementing estimation of the parameters via auxiliary models. The asymptotic properties of these methods are not firmly established in the literature.

We propose an estimation method suitable for both specifications of the MEV model and establish its asymptotic properties. We follow Harvey et al. (1994) and estimate a logarithmic transformation of the model,

$$\log \mathbf{x}_t^2 = \boldsymbol{\omega}_0 + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{0j} \boldsymbol{\epsilon}_{t-j-1} + \log \mathbf{z}_t^2. \tag{4.12}$$

In what follows we set without loss of generality $\mu_{\log x^2} = \boldsymbol{\omega}_0 + \mu_{\log z^2} = 0$ and estimate

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{0j} \boldsymbol{\epsilon}_{t-j-1} + \boldsymbol{\xi}_t. \tag{4.13}$$

where $\mathbf{y}_t \equiv \log \mathbf{x}_t^2$ and $\boldsymbol{\xi}_t \equiv \log \mathbf{z}_t^2$. The transformed model takes the form of a vector signal plus noise model, where the zero mean signal $\mathbf{h}_t$ has the one sided MA(∞) representation in (4.8) , and the noise is an $i.i.d$ process which can be correlated with the signal. We denote the variance matrix of the noise by $\Sigma_{0\boldsymbol{\xi}}$, and the covariance matrix of the signal and the noise by $\Sigma_{0\boldsymbol{\epsilon}\boldsymbol{\xi}}$. Without any loss of generality we set the vector of mean parameters ω_0 equal to zero.

Parametric estimation of model (4.12) requires to finitely parametrize the signal coefficients $\boldsymbol{\Psi}_{0j}$. We assume that we know a set of functions $\boldsymbol{\Psi}_j\,(.)$ of the $p \times 1$ vector $\boldsymbol{\zeta}$ with $p < \infty$, such that, for some unknown $\boldsymbol{\zeta}_0$,

$$\boldsymbol{\Psi}_j\,(\boldsymbol{\zeta}_0) = \boldsymbol{\Psi}_{0j} \qquad j \geq 1.$$

Analogously we parametrize the covariance matrices assuming that we know functions $\Sigma_{\boldsymbol{\epsilon}}(.)$, $\Sigma_{\boldsymbol{\xi}}(.)$, $\Sigma_{\boldsymbol{\epsilon}\boldsymbol{\xi}}(.)$ of the $q \times 1$ vector $\boldsymbol{\tau}$ with $q < \infty$, such that, for some unknown τ_0, $\Sigma_{\epsilon}(\boldsymbol{\tau}_0) = \Sigma_{0\boldsymbol{\epsilon}}$, $\Sigma_{\xi}(\boldsymbol{\tau}_0) = \Sigma_{0\boldsymbol{\xi}}$ and $\Sigma_{\epsilon\xi}(\boldsymbol{\tau}_0) = \Sigma_{0\boldsymbol{\epsilon}\boldsymbol{\xi}}$. We don't make any assumption on the joint density of the innovations $\{\boldsymbol{\xi}_t', \boldsymbol{\epsilon}_t'\}$, so $\boldsymbol{\tau}_0$ contains the $n + n\,(n-1)\,/2$ unknown parameters of $vech(\Sigma_{0\boldsymbol{\epsilon}\boldsymbol{\xi}})$, and the $n\,(n-1)\,/2$ unknown parameters of respectively $\Sigma_{0\boldsymbol{\epsilon}}$ and $\Sigma_{0\boldsymbol{\xi}}$, yielding $q = n + 3n\,(n-1)\,/2$. This specification of $\boldsymbol{\tau}_0$ however can be straightforward extended to models where the joint density of the innovations is specified up to some unknown parameters. We wish to estimate the $s\,(\equiv p+q)$ dimensional vector $\boldsymbol{\theta}_0' = (\boldsymbol{\zeta}_0', \boldsymbol{\tau}_0')'$ on the basis of a sample $\{\mathbf{y}_1, ..., \mathbf{y}_T\}$ of observations. The following section introduces the estimation method.

4.3 The Whittle likelihood

The problem is the statistical estimation of the parameter $\boldsymbol{\theta}$ in a vector linear signal plus noise model, $\mathbf{y}_t = \boldsymbol{\xi}_t + \mathbf{h}_t$, based on finite observations $\{\mathbf{y}_1, ..., \mathbf{y}_T\}$. For this purpose, the frequency domain approach rather than the time domain seems particularly effective. The Whittle estimator of $\boldsymbol{\theta}$ is the frequency domain approximation of the Gaussian log likelihood, used as a measure of distance between the periodogram of squared observations and the model spectral density $\mathbf{f}\,(\lambda, \theta)$. The spectral density matrix of $\mathbf{y}_t$ has functional form

$$\mathbf{f}(\lambda, \boldsymbol{\theta}) = \frac{\Sigma_\xi(\boldsymbol{\tau})}{2\pi} + \frac{\mathbf{k}(e^{i\lambda}, \boldsymbol{\zeta})\Sigma_\epsilon(\boldsymbol{\tau})\mathbf{k}(e^{i\lambda}, \boldsymbol{\zeta})^*}{2\pi} + \Sigma_{\epsilon\xi}(\boldsymbol{\tau})e^{-i\lambda}\mathbf{k}(e^{i\lambda}, \boldsymbol{\zeta})^* + e^{i\lambda}\mathbf{k}(e^{i\lambda}, \boldsymbol{\zeta})\Sigma'_{\epsilon\xi}(\boldsymbol{\tau}), \quad (4.14)$$

where $\mathbf{k}(e^{i\lambda}, \zeta) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j\,(\zeta)\, e^{i\lambda j}$ and $\lambda \in [-\pi, \pi]$. We denote by $\tilde{\Gamma}\,(u, \theta)$ the autocovariance matrix of the process,

$$\tilde{\boldsymbol{\Gamma}}(\boldsymbol{\theta}, u) = I_{(u=0)}\boldsymbol{\Sigma}_\xi\,(\boldsymbol{\tau}) + \boldsymbol{\Sigma}_\epsilon\,(\boldsymbol{\tau}) \sum_{j=0}^{\infty} \Psi_j\,(\boldsymbol{\zeta})\,\Psi'_{j+u}\,(\boldsymbol{\zeta}) + I_{(m\neq 0)}\Psi_{|u|-1}\,(\boldsymbol{\zeta})\,\boldsymbol{\Sigma}_{\xi\epsilon}\,(\boldsymbol{\tau})\,, \qquad (4.15)$$

for $u \geq 0$, and we denote by $\tilde{C}\,(m)$ and $\mathbf{I}_T\,(\lambda)$ respectively, the serial covariance and the periodogram matrices, that are constructed from a partial realization of $\{\mathbf{y}_1, ..., \mathbf{y}_T\}$, namely,

$$\tilde{C}\,(m) = \frac{1}{T} \sum_{t=1}^{T-|m|} \mathbf{y}_t \mathbf{y}'_{t+|m|},$$

for $0 \leq m \leq T-1$, and $\tilde{C}\,(m) = \tilde{C}\,(-m)$ for $-T+1 \leq m < 0$; and

$$\mathbf{I}_T(\lambda) = W_T(\lambda) W_t(\lambda)^*,$$

where $W_T(\lambda)$ is the discrete Fourier transform of the data

$$W_T(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} \mathbf{y}_t e^{i\lambda t},$$

and the symbol "*" denotes conjugate transposition. We avoid mean correction of the periodogram because of its translation invariance property at the Fourier frequencies $[-\pi, \pi]$. The Whittle estimator of $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}} \equiv \arg\min_{\theta \in \Theta} Q_T\,(\boldsymbol{\theta})\,,$$

where the function $Q_T(\boldsymbol{\theta})$, called the Whittle or pseudo log-likelihood function, is given as

$$Q_T(\boldsymbol{\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda,\boldsymbol{\theta})d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\}d\lambda. \tag{4.16}$$

Although we are not making Gaussianity assumptions on the $\mathbf{y}_t$, this estimator is suggested by the Gaussian pseudo maximum likelihood estimator obtained minimizing $-2L_T(\boldsymbol{\theta})/T$, where $L_T(\boldsymbol{\theta})$ is the log-likelihood function:

$$L_T(\boldsymbol{\theta}) = -\frac{T}{2}\log\left(\det\Gamma\left(\boldsymbol{\theta},u\right)\right) - \frac{T}{2}\mathbf{y}_t\Gamma^{-1}\left(\boldsymbol{\theta},u\right)\mathbf{y}_t'.$$

Minimization of this function is computationally very cumbersome. The pseudo maximum likelihood estimator is the solution of the system of equations:

$$\frac{\partial}{\partial\theta}L_T\left(\bar{\boldsymbol{\theta}}\right) = 0,$$

whose numerical evaluation requires many trials. If the covariances of the process have a slow rate of decay to zero, the covariance matrix might become almost singular. Moreover the evaluation of the inverse of the covariance matrix may be numerically unstable. The frequency domain approach to estimation seems instead particularly effective, since an approximate likelihood function in the frequency domain has a manageable expression for estimation and testing purposes. The Whittle spectral approximation to the likelihood function was originally proposed by Whittle (1952). Under the assumption of Gaussianity, the discrete Fourier transforms of the data $W_T(\lambda_t)$, at frequencies λ_t, $t = 1,...,T$, equispaced in $[-\pi,\pi]$, have a complex-valued multivariate normal distribution. For large T they are approximately independent, each with probability density function :

$$\pi^{-2}\left\{\det\left(\mathbf{f}\left(\lambda_t,\boldsymbol{\theta}\right)\right)\right\}^{-1/2}\exp\left[-\frac{1}{2}tr\left\{\mathbf{f}^{-1}\left(\lambda_t,\boldsymbol{\theta}\right)W_T\left(\lambda_t\right)W_T^*\left(\lambda_t\right)\right\}\right].$$

Because the discrete Fourier transforms constitute a sufficient statistic for θ (Hannan, 1970, pp. 224-225), an approximate log-likelihood function of $\boldsymbol{\theta}$ based on $\{\mathbf{y}_1,...,\mathbf{y}_T\}$ is given, up to constant multiplication, by:

$$-\sum_{t=1}^{T}\log\det\mathbf{f}(\lambda_t,\boldsymbol{\theta})-\sum_{t=1}^{T}tr\left\{\mathbf{f}^{-1}(\lambda_t,\boldsymbol{\theta})\mathbf{I}_T(\lambda_t)\right\}.$$

In integral form, this has the expression

$$-T\ \left[\frac{1}{\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\boldsymbol{\theta})d\lambda+\frac{1}{\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\}d\lambda\right].$$

Therefore we may either approximate $-2L_T\left(\boldsymbol{\theta}\right)$ by (4.16) or by its discretized version

$$\tilde{Q}_T(\boldsymbol{\theta})=\frac{1}{T}\sum_{t=1}^{T}\log\det\mathbf{f}(\lambda_t,\boldsymbol{\theta})+\frac{1}{T}\sum_{t=1}^{T}tr\left\{\mathbf{f}^{-1}(\lambda_t,\boldsymbol{\theta})\mathbf{I}_T(\lambda_t)\right\}, \tag{4.17}$$

where $\lambda_t=2\pi t/T,-T/2<t\leq T/2$. In practice $\tilde{Q}_T(\boldsymbol{\theta})$ will tend to be preferred to $Q_T\left(\boldsymbol{\theta}\right)$ because for $\tilde{Q}_T(\boldsymbol{\theta})$ the $W_T\left(\lambda_t\right)$ may be computed efficiently using the fast Fourier transform. In this chapter we present result for the Whittle estimator based on $Q_T\left(\boldsymbol{\theta}\right)$, however similar arguments and results hold when its discrete version is used.

A slightly different spectral approximation to the likelihood function was introduced by Dunsmuir and Hannan (1976). In the context of vector linear processes

$$\mathbf{y}_t=\sum_{l=0}^{\infty}A_l\left(\boldsymbol{\vartheta}\right)e_{t-l},\quad\sum_{l=0}^{\infty}\left\|A_l\left(\boldsymbol{\vartheta}\right)\right\|^2<\infty \tag{4.18}$$

they suggested minimization of the quantity

$$\tilde{L}_T\left(\boldsymbol{\vartheta}\right)=\log\det\mathbf{K}^e\left(\boldsymbol{\vartheta}\right)+\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\vartheta})\mathbf{I}_T(\lambda)\right\}d\lambda, \tag{4.19}$$

where $\mathbf{K}^e\left(\boldsymbol{\vartheta}\right)$ denotes the variance covariance matrix of the linear innovations of the process. However this objective function is by far less tractable than $Q_T\left(\boldsymbol{\theta}\right)$ for parameters estimation in signal plus noise processes. Indeed signal plus noise processes offer an example of possibly linearly regular processes whose spectral density function is not easily factored. Even if a signal plus noise process admits decomposition (4.18) and has spectral density with representation

$$\mathbf{f}\left(\lambda,\boldsymbol{\vartheta}\right)=\frac{1}{2\pi}\varphi\left(e^{i\lambda},\boldsymbol{\vartheta}\right)K^e\left(\boldsymbol{\vartheta}\right)\varphi\left(e^{i\lambda},\boldsymbol{\vartheta}\right)^*, \tag{4.20}$$

where $\varphi\left(e^{i\lambda},\boldsymbol{\vartheta}\right)=\sum_{l=0}^{\infty}A_l\left(\boldsymbol{\vartheta}\right)e^{il\lambda}$, it might not be possible to express parameters $\boldsymbol{\vartheta}$ in (4.20) as closed form functions of the parameters of the signal and the noise separately. In this case we say that the spectrum $\mathbf{f}\left(\lambda,\boldsymbol{\vartheta}\right)$ is not easily factored. This is not always the case for signal plus noise process, however when the signal and the noise are correlated, an estimation procedure that does not require factorization but only knowledge of the functional form of the spectral density is preferred.

Statistical literature on Whittle estimation has established the asymptotic properties of the estimator under a variety of conditions when the true underlying model is the vector linear process in (4.18), with white noise innovations e_t. For such processes there is no factorization issue, since the functional form of the spectral density and its factored representation coincide. In this context, Dunsmuir and Hannan (1976) establish the consistency and asymptotic normality of the estimator minimizing $\tilde{L}_T(\boldsymbol{\vartheta})$, assuming separate parameterization of the coefficients and the covariance matrix of the process. They partition the true parameter $\boldsymbol{\vartheta}_0$ as $(\mu_0, \phi_0) \in \Theta_\mu \times \Theta_\phi$, so that $K(\boldsymbol{\vartheta}) \equiv K(\mu)$ and $A_l(\boldsymbol{\vartheta}) \equiv A_l(\phi)$ and obtain the asymptotic normality and independence of $\sqrt{T}\left(\hat{\phi} - \phi\right)$ and $\sqrt{T}(\hat{\mu} - \mu)$. Dunsmuir (1979) extends these results to the case of non-separable parameter space noting that, while the asymptotic covariance matrix of the estimates does not depend on the fourth cumulants of the innovations in a model with separate parameterization, it does if the parameter space is no longer separable. Hosoya and Taniguchi (1982) derive the asymptotic normality of the estimator minimizing $Q_T(\boldsymbol{\vartheta})$ when the underlying model is a vector linear process with non-separate parameterization and innovations satisfying milder mixing conditions.

Extensions of these results to signal plus noise processes are limited to the case of autoregressive signal and at least incoherent signal and noise. This simplified setting allows for factorization of the spectral density function and a straightforward application of the previous results. Hosoya and Taniguchi (1982) apply their asymptotic theory to a univariate signal observed superimposed with white noise ξ_t. The signal h_t is generated by a finite autoregressive process $\sum_{j=1}^{q} b_j h_{t-j} = \eta_t$, where the η_t have zero mean, finite variance and are independent from ξ_s for every t and s. The parameter of interest is $\boldsymbol{\theta}' = \left(b_1, ..., b_q, \sigma_\eta^2, \sigma_\xi^2\right)'$. The spectral density function of the process

$$f_{\boldsymbol{\theta}}(\lambda) = \frac{1}{2\pi} \frac{\sigma_\eta^2}{\left|\sum_{j=1}^{q} b_j e^{i\lambda j}\right|} + \frac{\sigma_\xi^2}{2\pi},$$

is represented, applying Fejer-Riesz theorem (see, e.g., Archiezer, 1956, page 152), as

$$f_{\boldsymbol{\vartheta}}(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\sum_{j=1}^{q} \psi_j e^{i\lambda j}}{\sum_{j=1}^{q} b_j e^{i\lambda j}} \right|^2, \tag{4.21}$$

where the parameter $\boldsymbol{\vartheta}$ is a closed form function of $\boldsymbol{\theta}$, since σ^2 and the ψ_j can be expressed as closed form functions of $\boldsymbol{\theta}$, and the b_js are unchanged. Relying on representation (4.21), Hosoya and Taniguchi derive the asymptotic normality of $\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$. An analogous result is provided by Dunsmuir (1979), who applies his asymptotic theory to a scalar autoregressive signal plus noise process, with uncorrelated signal and noise. Dunsmuir suggests a further extension of his results to incoherent vector signal plus noise processes. He considers an uncorrelated vector signal plus noise model $\mathbf{y}_t = \mathbf{h}_t + \boldsymbol{\xi}_t$, where the signal and the noise have one sided representation

$$\begin{aligned} \mathbf{h}_t &= \sum_{j=0}^{\infty} C_j^{(h)}(\boldsymbol{\theta}_h)\, \boldsymbol{\epsilon}_{t-j}, \\ \boldsymbol{\xi}_t &= \sum_{l=0}^{\infty} C_l^{(x)}(\boldsymbol{\theta}_\xi)\, \boldsymbol{\eta}_{t-l}, \end{aligned}$$

with uncorrelated shocks $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_s$ for every t and s. Using the purely linear representation of the process,

$$\mathbf{y}_t = \sum_{k=0}^{\infty} D_k(\boldsymbol{\vartheta})\, e_{t-k}, \quad tr \sum_{k=0}^{\infty} D_k(\boldsymbol{\vartheta})\, K^e(\boldsymbol{\vartheta})\, D_k(\boldsymbol{\vartheta})^* < \infty, \tag{4.22}$$

where the $D_k(\boldsymbol{\vartheta})$ are rectangular coefficients, $D_k(\boldsymbol{\vartheta}) \equiv \left[C_k^{(h)}(\boldsymbol{\theta}_h) \vdots C_k^{(x)}(\boldsymbol{\theta}_\xi)\right]$ and the innovations are $e_t' \equiv (\boldsymbol{\epsilon}_t', \boldsymbol{\eta}_t')$, he factors the spectral density

$$\mathbf{f}_y(\lambda, \boldsymbol{\theta}) = \mathbf{f}_h(\lambda, \boldsymbol{\theta}_h) + \mathbf{f}_\xi(\lambda, \boldsymbol{\theta}_\xi)$$

as

$$\mathbf{f}(\lambda, \boldsymbol{\vartheta}) = k_y(\boldsymbol{\vartheta})\, K^e(\boldsymbol{\vartheta})\, k_y^*(\boldsymbol{\vartheta}),$$

and discusses in details a set of conditions on the disturbances of the linear representation that allow to establish the asymptotic properties of $\hat{\boldsymbol{\vartheta}}$ along the line of his previous findings. He suggests for example that the assumption of martingale difference of the innovations e_t with respect to their past might not be appropriate because, rather than the prediction of $\mathbf{y}_t$ based on its past, the prediction of the signal based on the past of the process might be of more interest. However he does not suggest how the asymptotic normality of $\hat{\boldsymbol{\theta}}' \equiv \left(\hat{\boldsymbol{\theta}}'_h, \hat{\boldsymbol{\theta}}'_\xi\right)$ could be derived from the asymptotic normality of $\hat{\boldsymbol{\vartheta}}$ using representation (4.22) or otherwise.

This chapter extends the current literature to vector signal plus noise models where the signal, $\mathbf{h}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j(\theta)\, \boldsymbol{\epsilon}_{t-j}$, and the $i.i.d$ noise can be simultaneously correlated. We establish the consistency and the asymptotic normality of the estimates under fairly general regularity conditions, easily verifiable for both "one-shock" and "two-shocks" specifications, without relying on the factorization of the spectral density but only on the knowledge of its functional form.

4.4 Consistency

This section discusses the strong consistency of the Whittle estimator of $\boldsymbol{\theta}$ in model (4.13). We first list the assumptions with some discussion, then present the main result. All the proofs of the technical lemmas are in appendix A.1 and A.2. In what follows we denote by K a generic finite constant, not always the same. The symbol " $\sim$" denotes asymptotic equivalence: $a(x) \sim b(x)$ as $x \to x_0$ when $a(x)/b(x) \to 1$. The (a, b) component of any matrix Γ is denoted as $\Gamma^{(a,b)}$, the (a, a) component of a diagonal matrix $\boldsymbol{\Psi}_{0j}$ is denoted by $\boldsymbol{\Psi}_{0j}^{(a)}$. We denote by $\boldsymbol{\epsilon}_t^{(a)}$the ath element of the vector $\boldsymbol{\epsilon}$ at time t. We assume that all the elements of $\boldsymbol{\xi}_t$, $\boldsymbol{\epsilon}_t$, $\boldsymbol{\Psi}_j(\zeta)$ are real. We define $\Pi \equiv [-\pi, \pi]$ and denote by $L_p(\Pi)$ the class of p-integrable functions defined on Π. The symbol "$\|A\|$" denotes the Euclidean norm of a matrix A, the symbol "$|a|$" denotes the absolute value of a scalar a. The symbol "> 0" denotes strict positive definiteness when applied to a matrix.

ASSUMPTION 4.1

[**A**] $\{\boldsymbol{\epsilon}_t', \boldsymbol{\xi}_t'\}$ are $i.i.d$ unobservable random vectors, and, for every $a, b = 1, ..., n$,

(i) $E\boldsymbol{\epsilon}_0^{(a)} = 0$ and $E(\boldsymbol{\epsilon}_0\boldsymbol{\epsilon}_0') = \Sigma_\varepsilon(\boldsymbol{\tau}), \Sigma_\varepsilon(\boldsymbol{\tau}) < \infty$.

(ii) $E\left|\boldsymbol{\xi}_0^{(a)}\right| < \infty$ and $E(\boldsymbol{\xi}_0\boldsymbol{\xi}_0') = \Sigma_\xi(\boldsymbol{\tau}), \Sigma_\xi(\boldsymbol{\tau}) < \infty$.

(iii) $E(\boldsymbol{\xi}_0\boldsymbol{\epsilon}_0') = \Sigma_{\xi\epsilon}(\boldsymbol{\tau}), \Sigma_{\xi\epsilon}(\boldsymbol{\tau}) < \infty$.

[**B**] $\boldsymbol{\theta}_0$ is an interior point of the compact parameter space $\Theta \in R^s$.

[**C**] For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and for all $a = 1, ..., n$,

(i) The matrices $\Sigma_\varepsilon(\boldsymbol{\tau})$, $\Sigma_\xi(\boldsymbol{\tau})$ and $\Sigma_{\xi\epsilon}(\boldsymbol{\tau})$ are continuous.

(ii) The $\boldsymbol{\Psi}_k(\zeta)$ are continuous for all k and

$$|\boldsymbol{\Psi}_k^{(a)}(\zeta)| \leq K_1|\boldsymbol{\Psi}_j^{(a)}(\zeta)|, \quad \text{for} \quad 1 \leq j \leq k, \quad k \geq 1.$$

(iii) For a boundedly differentiable function $e(\boldsymbol{\zeta}) \in (-1, 1)$,

$$\boldsymbol{\Psi}_j^{(a)}(\zeta) \sim K_2 e^j(\boldsymbol{\zeta}), \quad \text{as} \quad j \to \infty.$$

[**D**] For any $\theta \in \Theta$, and for all $a = 1, ..., n$,

(i) $\Sigma_\varepsilon(\boldsymbol{\tau})$, $\Sigma_\xi(\boldsymbol{\tau})$ and $\Sigma_{\xi\epsilon}(\boldsymbol{\tau})$ have continuous first derivatives.

(ii) For all k, the $\boldsymbol{\Psi}_k(\zeta)$ have continuous first derivatives.

(iii) For a boundedly differentiable function $e(\boldsymbol{\zeta}) \in (-1, 1)$

$$\frac{\partial}{\partial \zeta_{i_h}} \boldsymbol{\Psi}_j^{(a)}(\zeta) \sim E_1(j; \zeta)\, e^j(\boldsymbol{\zeta}) \quad as \quad j \to \infty,$$

for all $i_h = 1, ..., p$, where $|E_1(j; \zeta)| \leq K j^r$.

[**E**] For every $\theta \in \Theta$ whenever $\theta \neq \theta_0$, $\tilde{\boldsymbol{\Gamma}}(u, \boldsymbol{\theta}) \neq \tilde{\boldsymbol{\Gamma}}(u, \boldsymbol{\theta}_0)$, for all $u \geq 0$.

[**F**] For any $\boldsymbol{\tau} \in \Theta$, $\Sigma_\varepsilon(\boldsymbol{\tau})$ is a strictly positive definite matrix.

[**G**] (i) $\mathbf{f}^{-1}(\lambda, \boldsymbol{\theta})$ has elements in $L_2(\Pi)$, bounded and continuous at all $(\lambda, \theta) \in \Pi \times \Theta$.

(ii) For every $\eta > 0$, the function

$$\boldsymbol{\phi}_\eta(\lambda, \theta) \equiv \frac{\mathbf{f}(\lambda, \theta)}{\det \mathbf{f}(\lambda, \boldsymbol{\theta}) + \eta},$$

has elements in $L_2(\Pi)$, bounded and continuous at all $(\lambda, \theta) \in \Pi \times \Theta$.

Time domain assumptions are not common in the statistical literature on Whittle estimation, that typically defines regularity conditions in term of a certain degree of smoothness of the model spectral density and its higher order derivatives. Whereas in a non parametric framework those assumptions represent a natural choice, they might not be motivated in a parametric setting. In this chapter we impose most regularity conditions directly on the model. Appendix A.1. formally establishes a number of properties of its spectral density, implied by Assumption 4.1. Under assumption 4.1.[A] and square summability of the coefficients assumed in (4.8) , the $\mathbf{y}_t$ are strictly stationarity and ergodic. Strict stationarity and ergodicity of the underlying process are common assumptions in the statistical literature on Whittle estimation. Robinson (1978) replaces the assumption of strict stationarity by the weaker assumption of fourth order stationarity as is also done in Hosoya and Taniguchi (1982). Hosoya and Taniguchi moreover dispense with the explicit assumption of ergodicity and impose a Lindeberg condition. Assumption 4.1[B] is a standard assumption to ensures that $\boldsymbol{\theta}_0$ is an interior point of the compact closure of an open s-dimensional manifold (see for example Hannah (1973) and Robinson (1978)). It implies boundedness of any function of $\boldsymbol{\theta} \in \Theta$. Identification of the parameters is granted by Assumption 4.1.[E] , which rules out the possibility of two equivalent structures giving rise to the same spectral density (4.14). Assumption 4.1[F] ensures strict positivity of the spectral density at all frequencies. We follow Hannan (1973), Dunsmuir and Hannan (1976), and Hosoya and Taniguchi (1982) who, in the context of Whittle estimation of linear processes, restrict the parameter space to a subset Θ_0 where the spectrum is positive. Assumption 4.1.[C](iii) imparts the exponential decay of the parameters of the signal process. The imposition of this exact rate together with Assumption 4.1.[C](ii) implies the absolute summability of the signal process which in turn implies the absolute summability of the autocovariance function of $\mathbf{y}_t$. This latter condition is sufficient to guarantee the existence and the square integrability of the spectral density. Moreover it ensures its uniform continuity at all

$(\lambda, \theta) \in \Pi \times \Omega$. In contrast to Hannan (1973) and Dunsmuir and Hannan (1976), we follow Robinson (1978) and dispense with the assumption of Lipschitz continuity of degree $\alpha > 1/2$ in the proof the consistency of the estimator. However we directly impose smoothness conditions on the inverse of the spectral density matrix and in Assumption 4.1 [**G**] to ensure the uniform convergence of the objective function.

Consistency of the Whittle estimator is generally assumed in the literature. The only available result is due to Hannan and Dunsmuir (1976), who provide an extension of Hannan (1973) original univariate result to the vector linear process with representation (4.18) for the estimator minimizing (4.19). They establish the result assuming strict stationarity and ergodicity of the process, Lipschitz continuity of degree $\alpha > 1/2$ of the spectral density, and martingale difference linear innovations that satisfy $E\left(e_t e_s'\right) = \boldsymbol{\mu}_{ts}$ and $E\left(e_t e_s' \mid F_{t-1}\right) = K^e\left(\tau\right)$. Moreover they assume

$$\inf_{\vartheta \in \Theta} \tilde{L}_T\left(\boldsymbol{\vartheta}\right) = \tilde{L}_T\left(\boldsymbol{\vartheta}_0\right) = \log\det \mathbf{K}^e\left(\boldsymbol{\tau}_0\right) + s, \tag{4.23}$$

where s is the dimension of vector $\mathbf{y}_t$. We provide an extension to their results for an estimator minimizing (4.16) in model (4.13) with the following theorem.

Theorem 4.1 *Under Assumption 4.1, as* $T \to \infty$

$$\hat{\theta}_T \to_{a.s} \theta_0.$$

For the purpose of establishing Theorem 4.1, we define $Q\left(\theta\right)$ as

$$Q\left(\theta\right) \equiv \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \boldsymbol{\theta}) d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda, \boldsymbol{\theta})\mathbf{f}(\lambda, \boldsymbol{\theta}_0)\right\} d\lambda.$$

Additionally recall that in Assumption 4.1[**G**] we introduced for any $\eta > 0$, the function, $\boldsymbol{\phi}_\eta\left(\lambda, \theta\right) \equiv \mathbf{f}\left(\lambda, \theta\right) / \left(d\left(\lambda, \theta\right) + \eta\right)$, where $d\left(\lambda, \theta\right) \equiv \det \mathbf{f}\left(\lambda, \theta\right)$ and $\mathbf{f}\left(\lambda, \theta\right) / d\left(\lambda, \theta\right)$ is by definition the adjoint matrix of $\mathbf{f}^{-1}\left(\lambda, \theta\right)$. In what follows, whenever we replace $\mathbf{f}^{-1}\left(\lambda, \theta\right)$ by $\boldsymbol{\phi}_\eta\left(\lambda, \theta\right)$ we refer to $Q_T\left(\theta\right)$ as $Q_{T,\eta}\left(\theta\right)$ and to $Q\left(\theta\right)$ as $Q_\eta\left(\theta\right)$, denoting respectively

$$Q_{T,\eta}\left(\boldsymbol{\theta}\right) \equiv \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \boldsymbol{\theta}) d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\boldsymbol{\phi}_\eta(\lambda, \boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\} d\lambda,$$

and

$$Q_{\eta}\left(\theta\right)\equiv\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\boldsymbol{\theta})d\lambda+\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\boldsymbol{\phi}_{\eta}(\lambda,\boldsymbol{\theta})\mathbf{f}(\lambda,\theta_0)\right\}d\lambda.$$

The consistency of the estimator follows from the usual steps for consistency of M-estimators. The following lemma establishes the uniform continuity of the objective function in $(\lambda,\theta)\in\Pi\times\Omega$.

Lemma 4. 1 *If Assumption 4.1 holds, then*

(a) $\lim_{T\to\infty}Q_T\left(\boldsymbol{\theta}\right)=Q\left(\boldsymbol{\theta}\right)$ *almost surely uniformly in* $\boldsymbol{\theta}\in\Theta$*,*

(b) *for any* $\eta>0$*, uniformly in* $\boldsymbol{\theta}\in\Theta$*,* $\lim_{T\to\infty}Q_{T,\eta}\left(\boldsymbol{\theta}\right)=Q_{\eta}\left(\boldsymbol{\theta}\right)$ *almost surely.*

Dunsmuir and Hannan (1976, Lemma 1) establish a similar result for an objective function without the term $\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\boldsymbol{\theta})d\lambda$ under stronger conditions that Assumption 4.1. Moreover they assume that the minimum of the objective function is achieved at $\boldsymbol{\theta}_0$, whereas we establish it with the following lemma.

Lemma 4.2 *If Assumption 4.1 holds, then for all* $\theta\in\Theta$*,*

$$\begin{aligned}\inf_{\theta\in\Theta}Q\left(\theta\right) &= Q\left(\theta_0\right)=\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)d\lambda+\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}_0)\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right\}d\lambda\\ &= \int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta_0\right)d\lambda+T.\end{aligned}$$

We follow Giraitis et al. (2012, Chapter 8, Theorem 8.2.1) and prove Theorem 4.1 by contradiction. Suppose that $\hat{\boldsymbol{\theta}}_T$ is not consistent for $\boldsymbol{\theta}_0$. Then by the compactness of the parameter space Θ, there is a subsequence $\tilde{\boldsymbol{\theta}}_{T(M)}$ of $\hat{\boldsymbol{\theta}}_T$ converging to some $\boldsymbol{\vartheta}\in\Theta$ such that $\boldsymbol{\vartheta}\neq\boldsymbol{\theta}_0$. Then

$$\begin{aligned}&\lim\inf_{M\to\infty}Q_{T(M)}\left(\tilde{\boldsymbol{\theta}}_{T(M)}\right)\\ \geq\;&\sup_{\eta>0}\left\{\lim\inf_{M\to\infty}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\tilde{\boldsymbol{\theta}}_{T(M)}\right)d\lambda+\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left[\phi_{\eta}(\lambda,\tilde{\boldsymbol{\theta}}_{T(M)})\mathbf{I}_M(\lambda)\right]d\lambda\right\}\right\}\\ =\;&\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\vartheta}\right)d\lambda+\sup_{\eta>0}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left[\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)\phi_{\eta}\left(\lambda,\boldsymbol{\vartheta}\right)\right]d\lambda\right\},\end{aligned}$$

where the first inequality follows from the definition of $\boldsymbol{\phi}_{\eta}$, and the last equality follows from Lemma 4.1, part (b). As $\eta\to0$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\vartheta}\right)d\lambda+\sup_{\eta>0}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left[\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)\phi_{\eta}\left(\lambda,\boldsymbol{\vartheta}\right)\right]d\lambda\right\}$$

$$\to \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\vartheta}\right)d\lambda+\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left[\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\vartheta}\right)\right]d\lambda \quad \text{a.s.}$$

Assumption 4.1[**E**] and Lemma 4.2 imply that for every $\vartheta \neq \theta_0$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\vartheta}\right)d\lambda+\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left[\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\vartheta}\right)\right]d\lambda$$
$$> \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)d\lambda+T.$$

So we may conclude that

$$\lim\inf_{M\to\infty} Q_{T(M)}\left(\tilde{\boldsymbol{\theta}}_{T(M)}\right) > \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)d\lambda+T. \tag{4.24}$$

However in view of the definition of the estimator $\hat{\boldsymbol{\theta}}_T$, $Q_{T(M)}\left(\tilde{\boldsymbol{\theta}}_{T(M)}\right) \leq Q_{T(M)}\left(\boldsymbol{\theta}\right)$ for all $\boldsymbol{\theta} \in \Theta$.

Therefore

$$\begin{aligned}\lim\sup_{M\to\infty} Q_{T(M)}\left(\tilde{\boldsymbol{\theta}}_{T(M)}\right) &\leq \inf_{\theta\in\Theta}\lim\sup_{M\to\infty} Q_{T(M)}\left(\boldsymbol{\theta}\right) &(4.25)\\ &= \inf_{\theta\in\Theta} Q\left(\boldsymbol{\theta}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)d\lambda+T, &(4.26)\end{aligned}$$

where the first equality uses Lemma 4.1, part(a) and the last equality uses Lemma 4.2. Hence the contradiction:

$$\lim\sup_{M\to\infty} Q_{T(M)}\left(\tilde{\boldsymbol{\theta}}_{T(M)}\right) \leq \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)d\lambda+T < \lim\inf_{M\to\infty} Q_{T(M)}\left(\tilde{\boldsymbol{\theta}}_{T(M)}\right),$$

almost surely, which completes the proof.

The vector of mean parameters $\boldsymbol{\omega}_0$ cannot be identified by the Whittle function since its elements enter linearly in $\log x_{it}^2$ and are lost when computing the empirical autocovariances of the process. However it can be estimated using the sample mean of the vector $\mathbf{y}_t = (\log x_{1t}^2, \log x_{2t}^2, ..., \log x_{nt}^2)'$. Since $\hat{\mathbf{y}}_T = 1/T\sum_{t=1}^{T}\mathbf{y}_t$ is a $\sqrt{T}$-consistent estimate of $E\mathbf{y}_t = \boldsymbol{\omega}_0 + E\boldsymbol{\xi}_t$ under Assumption 4.1, we can obtain a $\sqrt{T}$-consistent estimate of $\boldsymbol{\omega}_0$ subtracting the Whittle estimate of $E\boldsymbol{\xi}_t$ from $\hat{\mathbf{y}}_T$.

4.5 Asymptotic Normality

This section derives the asymptotic normality of the estimator. We introduce a set of stronger conditions and reinforce the assumptions on the moments of the unobservable shocks driving the process assuming finite fourth moments. We strengthen the degree of smoothness of the spectral density and extend the regularity conditions of Assumption 4.1 to the higher order derivatives of the spectral density matrix up to the third order.

ASSUMPTION 4.2

[**A**] $\{\boldsymbol{\epsilon}_t', \boldsymbol{\xi}_t'\}$ are $i.i.d$ unobservable random vectors, and for all $a, b, c, d = 1, ..., n,$

(i) $E\boldsymbol{\epsilon}_0^{(a)} = 0$ and $E\left(\boldsymbol{\epsilon}_0^{(a)}\boldsymbol{\epsilon}_0^{(b)}\boldsymbol{\epsilon}_0^{(c)}\boldsymbol{\epsilon}_0^{(d)}\right) = K_{abcd}^{\boldsymbol{\epsilon}}(\boldsymbol{\tau}), |K_{abcd}^{\boldsymbol{\epsilon}}(\boldsymbol{\tau})| < \infty.$

(ii) $E\left|\boldsymbol{\xi}_0^{(a)}\right| < \infty$ and $E\left(\boldsymbol{\xi}_0^{(a)}\boldsymbol{\xi}_0^{(b)}\boldsymbol{\xi}_0^{(c)}\boldsymbol{\xi}_0^{(d)}\right) = K_{abcd}^{\boldsymbol{\xi}}(\boldsymbol{\tau}), |K_{abcd}^{\boldsymbol{\epsilon}}(\boldsymbol{\tau})| < \infty.$

(iii) $E\left(\boldsymbol{\xi}_0^{(a)}\boldsymbol{\epsilon}_0^{(b)}\boldsymbol{\xi}_0^{(c)}\boldsymbol{\epsilon}_0^{(d)}\right) = K_{abcd}^{\boldsymbol{\epsilon\xi}}(\boldsymbol{\tau}), \left|K_{abcd}^{\boldsymbol{\epsilon\xi}}(\boldsymbol{\tau})\right| < \infty.$

[**B**] $\boldsymbol{\theta}_0$ is an interior point of the compact parameter space $\Theta \in R^s$.

[**C**] For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and for all $a = 1, ..., n,$

(i) The matrices $\Sigma_\varepsilon(\boldsymbol{\tau})$, $\Sigma_\xi(\boldsymbol{\tau})$ and $\Sigma_{\xi\epsilon}(\boldsymbol{\tau})$ are continuous.

(ii) The $\boldsymbol{\Psi}_k(\zeta)$ are continuous for all k and

$$|\boldsymbol{\Psi}_k^{(a)}(\zeta)| \leq K_1|\boldsymbol{\Psi}_j^{(a)}(\zeta)| \quad \text{for} \quad 1 \leq j \leq k, \quad k \geq 1.$$

(iii) For a boundedly differentiable function $e(\boldsymbol{\zeta}) \in (-1, 1)$,

$$\boldsymbol{\Psi}_j^{(a)}(\zeta) \sim K_2 e^j(\boldsymbol{\zeta}) \quad \text{as} \quad j \to \infty.$$

[**D**] For any $\theta \in \Theta$, and for all $a = 1, ..., n,$

(i) $\Sigma_\varepsilon(\boldsymbol{\tau})$, $\Sigma_\xi(\boldsymbol{\tau})$ and $\Sigma_{\xi\epsilon}(\boldsymbol{\tau})$ have continuous first, second and third derivatives.

(ii) For all k, the $\boldsymbol{\Psi}_k(\zeta)$ have continuous first, second and third derivatives.

(iii) For a boundedly differentiable function $e(\boldsymbol{\zeta}) \in (-1, 1)$

$$\frac{\partial^r \boldsymbol{\Psi}_j^{(a)}(\zeta)}{\partial\zeta_{i_h}\partial\zeta_{i_h}\partial\zeta_{i_h}} \sim E_r(j;\zeta)\, e^j(\boldsymbol{\zeta}) \quad as \quad j \to \infty,$$

for all $i_h = 1, ..., p$, $h = 1, 2, 3$ and $r = 1, 2, 3$, where $|E_r(j;\zeta)| \leq Kj^r$.

[**E**] For every $\boldsymbol{\theta} \in \Theta$ whenever $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\tilde{\boldsymbol{\Gamma}}(u, \boldsymbol{\theta}) \neq \tilde{\boldsymbol{\Gamma}}(u, \boldsymbol{\theta}_0)$, for all u.

[**F**] For any $\boldsymbol{\tau} \in \Theta$, $\Sigma_\varepsilon(\boldsymbol{\tau})$ is a strictly positive definite matrix.

[**G**] (i) $\mathbf{f}(\lambda, \boldsymbol{\theta}_0)$ has elements that satisfy a Lipschitz condition of degree $\alpha > 1/2$.

(ii) $\mathbf{f}^{-1}(\lambda, \boldsymbol{\theta})$ has elements in $L_2(\Pi)$, bounded and continuous at all $(\lambda, \theta) \in \Pi \times \Theta$.

(iii) $(\partial/\partial\theta)\,\mathbf{f}^{-1}(\lambda, \boldsymbol{\theta})$ has elements in $L_2(\Pi)$, bounded and continuous at all $(\lambda, \theta) \in \Pi \times \Theta$.

Assumption 4.2[A] imparts finiteness of the fourth order cumulants of the process. Together with Assumption 4.2[C] it implies that the process has square integrable trispectrum

$$\tilde{K}^{\mathbf{y}}_{abcd}\left(\lambda_1,\lambda_2,\lambda_3\right)=\frac{1}{\left(2\pi\right)^3}\sum_{t_1,t_2,t_3=-\infty}^{\infty}\exp\left\{-i\left(\lambda_1t_1+\lambda_2t_2+\lambda_3t_3\right)\right\}K^{\mathbf{y}}_{abcd}\left(t_1,t_2,t_3\right). \quad (4.27)$$

The complex form of the spectral density matrix in (4.14) implies that $\mathbf{f}\left(\lambda,\theta\right)$ is not separately parametrized, and so we expect the fourth cumulant spectral density to appear in the asymptotic covariance matrix of the estimates, as for the cases discussed in Hosoya and Taniguchi (1982). Assumptions 4.2[G] reinforces the degree of smoothness of the spectral density assuming its Lipschitz continuity of degree $\alpha > 1/2$ at the true parameter value. This guarantees (see Hannan, 1970, page 513) that the spectral density matrix uniformly in $\lambda \in \Pi$ satisfies

$$\sup_{\lambda\in\Pi}\left\|\mathbf{f}_T\left(\lambda,\boldsymbol{\theta}_0\right)-\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right\|=O\left(T^{-\alpha}\right),$$

where $\mathbf{f}_T\left(\lambda,\boldsymbol{\theta}_0\right)$ denotes the Tth order Cesaro sum of the Fourier series of $\mathbf{f}(\lambda,\boldsymbol{\theta}_0)$. This condition allows to approximate the score vector by a simpler quadratic form whose asymptotic distribution is easily found using results on the converge of sample serial covariances. Assumptions 4.2[G] (iii) reinforces the smoothness of $\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})$ at all $\boldsymbol{\theta}$ in order to guarantee that $\log\det\mathbf{f}(\lambda,\boldsymbol{\theta})$ is twice differentiable in $\theta\in\Theta$ under the integral sign. Assumptions 4.2[D] extends the uniform continuity at all (λ,θ) to the higher order derivatives of the process and it imposes the required degree of smoothness to the Hessian. Moreover Assumption 4.2 implies (see Lemma B1.6 in Appendix B, section B.1) that

$$\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta\right)d\lambda>-\infty, \quad (4.28)$$

which ensures (see Giraitis et al., 2012, Chapter 3, Theorem 3.2.1) that $\mathbf{y}_t$ is purely non deterministic with Wold decomposition

$$\mathbf{y}_t=\sum_{l=0}^{\infty}A_l\left(\boldsymbol{\theta}\right)e_{t-l},\quad \sum_{l=0}^{\infty}\left\|A_l\left(\boldsymbol{\theta}\right)\right\|^2<\infty, \quad (4.29)$$

where the linear innovations e_t are n-dimensional white noise vectors. Statistical literature on Whittle estimation assumes regularity conditions directly on the linear innovations of the Wold decomposition (4.29). We instead represent $\mathbf{y}_t$ as a vector linear process with innovations $\boldsymbol{\xi}_t$ and $\boldsymbol{\epsilon}_t$. We express (4.13) as

$$\mathbf{y}_t = \sum_{l=0}^{\infty} \mathbf{\Phi}_l\left(\boldsymbol{\theta}\right)\tilde{e}_{t-l}, \quad \sum_{l=0}^{\infty} \left\|\Phi_l\left(\boldsymbol{\theta}\right)\right\|^2 < \infty, \tag{4.30}$$

where we define the innovations $\tilde{e}_t$ as

$$\tilde{e}_t \equiv \boldsymbol{\xi}_t, \quad l = 0, \tag{4.31}$$

$$\tilde{e}_{t-l} \equiv \boldsymbol{\epsilon}_{t-l}, \quad l \geq 1, \tag{4.32}$$

and the linear coefficients $\Phi_l\left(\theta\right)$ as

$$\mathbf{\Phi}_l\left(\theta\right) \equiv I_n, \quad l = 0, \tag{4.33}$$

$$\mathbf{\Phi}_l\left(\theta\right) \equiv \mathbf{\Psi}_{l-1}\left(\theta\right), \quad l \geq 1. \tag{4.34}$$

The linear innovations in (4.30) are a sequence of zero mean, independent, random vectors with finite fourth moment under Assumption 4.2. However they are not identically distributed at all l; more precisely, by definition the $\tilde{e}_{t-l}$ are identically distributed for all $l \geq 1$.

For the purpose of showing the asymptotic normality of the estimator we must introduce a central limit theorem for any linear combination of the quantities

$$\tilde{\tau}_{(a,b)}\left(m\right) \equiv \sqrt{T}\left(\mathbf{I}_{(a,b)}\left(\lambda\right) - \mathbf{EI}_{(a,b)}\left(\lambda\right)\right) \tag{4.35}$$

$$= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T-m} \mathbf{y}_t^{(a)}\mathbf{y}_{t+m}^{(b)} - \tilde{\Gamma}_m^{(a,b)}\right) \tag{4.36}$$

where $\tilde{\Gamma}_m^{(a,b)}$ is the (a, b) element of the autocovariance matrix of $\mathbf{y}_t$ at lag m, for $a, b = 1, .., n$ and $m = 0, \pm1, \pm2, ...$ The joint asymptotic normality of any linear combination of the $\tilde{\tau}_{(a,b)}$ has been established in the literature for a zero mean purely non deterministic linear vector process, with representation (4.29), under a variety of conditions on the process and on its linear innovations. For an ergodic and strictly stationary process, Hannan (1976) assumes that the innovations satisfy almost surely

$$
\begin{array}{l}
(i)\, E\left\{e_t|F_{t-1}\right\}=0, \\
(ii)\, E\left\{e_t^{(a)}e_t^{(b)}|F_{t-1}\right\}=\delta_{(a,b)}, \\
(iii)\, E\left\{e_t^{(a)}e_t^{(b)}e_t^{(c)}|F_{t-1}\right\}=\delta_{(a,b,c)}, \\
(iv)\, E\left(e_t^{(a)}e_t^{(b)}e_t^{(c)}e_t^{(d)}\right)=K_{abcd}^e,\, \left|K_{abcd}^e\right|<\infty,
\end{array}
\tag{4.37}
$$

and shows that the necessary and sufficient condition for the convergence of the sample serial covariances is that the diagonal elements of $\mathbf{f}\left(\lambda,\theta\right)$ be square integrable. Hosoya and Taniguchi (1982, Theorem 2.2) derive a central limit theorem for the $\tilde{\tau}_{(a,b)}\left(m\right)$ under a different set of assumptions. They replace strict stationarity with second order stationarity and ergodicity with a Lindeberg-type condition. They replace the strongly mixing conditions on the innovations with

$$
(i)\, Var\left(E\left\{e_t^{(a)}e_{t+u}^{(b)}|F_{t-\tau}\right\}-E\left\{e_t^{(a)}e_{t+u}^{(b)}\right\}\right)=O\left(\tau^{-2-\varepsilon}\right) \text{ uniformly in } t \text{ for some } \varepsilon>0,
$$

$$
(ii)\, E\left|E\left\{e_{t_1}^{(a)}e_{t_2}^{(b)}e_{t_3}^{(c)}e_{t_4}^{(d)}|F_{t-\tau}\right\}-E\left\{e_{t_1}^{(a)}e_{t_2}^{(b)}e_{t_3}^{(c)}e_{t_4}^{(d)}\right\}\right|=O\left(\tau^{-1-\eta}\right) \text{ for some positive constant } \eta,
$$

$$
(iii)\sum_{t_1,t_2,t_3=1}^{\infty}\left|K_{abcd}^e\left(t_1,t_2,t_3\right)\right|<\infty.
$$

As Hannan, they find that the necessary and sufficient condition for the result is that the diagonal elements of $\mathbf{f}\left(\lambda,\theta\right)$ be square integrable. The relationship between the two sets of conditions seems not to be straightforward. Hosoya and Taniguchi offer an extensive discussion (see Hosoya and Taniguchi, 1982, Section 2, Remark 2.1) and provide some examples of strictly stationary uniformly mixing and absolutely mixing processes that satisfy (i) and (ii).

We conjecture that Theorem 1 of Hannan (1976) can be extended to the sample serial covariances of a process with representation (4.30). The process is strictly stationary and ergodic by Lemma B1.1. The linear innovations $\tilde{e}_t$ are zero mean independent (martingale difference) vectors with finite fourth moments that satisfy conditions (4.37) at all $t-l$, with $l\geq 1$. Moreover the diagonal elements of the spectral density matrix are square integrable by Lemma B1.4. For all $a,b=1,....s$, and for $m=0,\pm 1,\pm 2,...$, we make the following conjecture

Conjecture 4.2 *If Assumption 4.2 holds, the quantities*

$$\tilde{\tau}_{(a,b)}(m) = \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T-m}\mathbf{y}_t^{(a)}\mathbf{y}_{t+m}^{(b)} - \tilde{\Gamma}_m^{(a,b)}\right), \tag{4.38}$$

have a joint asymptotic normal distribution with zero mean. The asymptotic covariance between $\tilde{\tau}_{(a,b)}(m)$ and $\tilde{\tau}_{(c,d)}(u)$ is given as

$$2\pi\int_{-\pi}^{\pi}\left\{\mathbf{f}_{(a,c)}(\lambda,\theta)\,\bar{\mathbf{f}}_{(b,d)}(\lambda,\theta)\,e^{-i(m-u)\lambda} + \mathbf{f}_{(a,d)}(\lambda,\theta)\,\bar{\mathbf{f}}_{(b,c)}(\lambda,\theta)\,e^{i(m+u)\lambda}\right\}d\lambda$$
$$+2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{(im\lambda_1+iu\lambda_2)}\tilde{K}_{abcd}^{\mathbf{y}}(-\lambda_1,\lambda_2,-\lambda_2)d\lambda_1 d\lambda_2.$$

Hannan derives the serial covariances of the limiting distribution of the $\tilde{\tau}_{(a,b)}(m)$ as a function of the trispectrum of the linear innovations

$$2\pi\int_{-\pi}^{\pi}\left\{\mathbf{f}_{(a,c)}(\lambda,\vartheta)\,\bar{\mathbf{f}}_{(b,d)}(\lambda,\vartheta)\,e^{-i(m-u)\lambda} + \mathbf{f}_{(a,d)}(\lambda,\vartheta)\,\bar{\mathbf{f}}_{(b,c)}(\lambda,\vartheta)\,e^{i(m+u)\lambda}\right\}d\lambda$$
$$+\sum_{\beta_1,\dots,\beta_4=0}^{s}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{(im\lambda_1+iu\lambda_2)}\varphi_{a\beta_1}(\lambda_1)\,\varphi_{b\beta_2}(-\lambda_1)\,\varphi_{c\beta 3}(\lambda_2)\,\varphi_{d\beta_4}(-\lambda_2)\,\tilde{K}_{\beta_1,\dots,\beta_4}^{e}(-\lambda_1,\lambda_2,-\lambda_2)\,d\lambda_1 d\lambda_2,$$

whereas Conjecture 4.2 states the result in terms of the trispectrum of the process directly (see Appendix B, Section B.3).

We can now address the asymptotic normality of the estimator. In what follows $\dot{Q}_T(\boldsymbol{\theta})$ denotes the $s\times 1$ vector of first partial derivatives of $Q_T(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, with j element $\dot{Q}_T^{(j)}(\boldsymbol{\theta})$. $\ddot{Q}_T(\boldsymbol{\theta})$ denotes the $s\times s$ matrix of second partial derivatives of $Q_T(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ with (i,j) element $\ddot{Q}_T^{(i,j)}(\boldsymbol{\theta})$. The $n\times n$ matrix of first partial derivatives of $\mathbf{f}(\lambda,\boldsymbol{\theta})$ with respect to θ_j is $\dot{\mathbf{f}}_{(j)}(\lambda,\boldsymbol{\theta})$ with (a,b) components $\dot{\mathbf{f}}_{(j)}^{(a,b)}(\lambda,\boldsymbol{\theta})$. Finally set $\ddot{\mathbf{f}}_{(i,j)}(\lambda,\boldsymbol{\theta}) = [(\partial^2/\partial\theta_i\partial\theta_j)\,\mathbf{f}(\lambda,\boldsymbol{\theta})]$. The j element of the score vector $\dot{Q}_T^{(j)}(\theta_0)$ is

$$\begin{aligned}&\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}_0)\,\dot{\mathbf{f}}_{(j)}(\lambda,\boldsymbol{\theta}_0)\right\}d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}(\lambda,\theta_0)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta_0)\,\mathbf{f}^{-1}(\lambda,\theta_0)\,I_T(\lambda)\right\}d\lambda\\ =\ &\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{g}_{(j)}(\lambda,\theta_0)\left(I_T(\lambda)-\mathbf{f}(\lambda,\theta_0)\right)\right\}d\lambda,\end{aligned} \tag{4.39}$$

where we set

$$\mathbf{g}_{(j)}(\lambda,\theta_0) \equiv \mathbf{f}^{-2}(\lambda,\theta_0)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta_0)\,.$$

The (i,j) element of the Hessian matrix $\ddot{Q}_T^{(i,j)}(\theta)$ is

$$\begin{aligned} & \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\log\det\mathbf{f}(\lambda,\boldsymbol{\theta})d\lambda+\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial^2}{\partial\theta_i\partial\theta_j'}tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\}d\lambda \qquad (4.40)\\ = & \frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\ddot{\mathbf{f}}_{(i,j)}(\lambda,\boldsymbol{\theta})\right\}d\lambda-\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\dot{\mathbf{f}}_{(i)}(\lambda,\boldsymbol{\theta})\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\dot{\mathbf{f}}_{(j)}(\lambda,\boldsymbol{\theta})\right\}d\lambda\\ & +\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\right)\mathbf{I}_T(\lambda)\right\}d\lambda. \end{aligned}$$

We can now state our main result.

Theorem 4.2 *Under Assumption 4.2,* as $T\to\infty$, *the vector* $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ *has an asymptotic normal distribution with zero mean and covariance matrix*

$$M^{-1}(\boldsymbol{\theta}_0)V(\boldsymbol{\theta}_0)M^{-1}(\boldsymbol{\theta}_0),$$

where

$$M(\boldsymbol{\theta}_0)=\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}_0)\dot{\mathbf{f}}_{(i)}(\lambda,\boldsymbol{\theta}_0)\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}_0)\dot{\mathbf{f}}_{(j)}(\lambda,\boldsymbol{\theta}_0)\right\}d\lambda,$$

and $V(\boldsymbol{\theta}_0)$ *is an* $s\times s$ *matrix with* (j,l) *element*

$$\begin{aligned} V_{(j,l)} &= \frac{1}{\pi}\int_{-\pi}^{\pi}\mathbf{tr}\left[\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\left(\frac{\partial}{\partial\theta_j}\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}_0)\right)\mathbf{f}(\lambda,\boldsymbol{\theta})\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}_0)\right)\right]d\lambda\\ &\quad+\frac{1}{2\pi}\sum_{a,b,c,d=1}^{s}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\dot{\mathbf{f}}_{(j)}^{(a,b)}(\lambda_1,\boldsymbol{\theta}_0)\dot{\mathbf{f}}_{(l)}^{(c,d)}(\lambda_2,\boldsymbol{\theta}_0)\right\}\tilde{K}_{abcd}^{\mathbf{y}}(-\lambda_1,\lambda_2,-\lambda_2,\boldsymbol{\theta}_0)d\lambda_1d\lambda_2. \end{aligned}$$

The proof of Theorem 4.2 is classical in nature. The consistency of $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, guaranteed by Theorem 4.1, implies that, as $T\to\infty$, $\hat{\boldsymbol{\theta}}$ eventually enters an arbitrary neighborhood of $\boldsymbol{\theta}_0$. By definition $\hat{\boldsymbol{\theta}}$ solves the equation $(\partial/\partial\boldsymbol{\theta})\,Q_T\left(\hat{\boldsymbol{\theta}}\right)=0$. The mean-value theorem implies that for $\mathring{\boldsymbol{\theta}}$, such that $\left\|\mathring{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\right\|\leq\left\|\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\right\|$,

$$0=T^{1/2}\dot{Q}_T(\boldsymbol{\theta})=T^{1/2}\dot{Q}_T(\boldsymbol{\theta})+\left[\ddot{Q}_T(\boldsymbol{\theta})\right]T^{1/2}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right). \qquad (4.41)$$

From (4.41) the central limit theorem for $T^{1/2}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)$ reduces to that for $\left[\ddot{Q}_T(\boldsymbol{\theta})\right]^{-1}T^{1/2}\dot{Q}_T(\boldsymbol{\theta}_0)$. Lemma 4.5 establishes the almost sure uniform convergence of $\ddot{Q}_T(\boldsymbol{\theta})$ to the $n\times n$ non singular matrix $M(\boldsymbol{\theta})$.

Lemma 4.5 *Under Assumption 4.2 as $T \to \infty$, uniformly in $\theta \in \Theta$,*

$$\ddot{Q}_T(\boldsymbol{\theta}) \to M(\boldsymbol{\theta})$$

almost surely, where $M(\theta)$ is a positive definite matrix with (i,j) element,

$$\begin{aligned} M^{(i,j)}(\boldsymbol{\theta}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\boldsymbol{\theta})\right\}d\lambda \\ &-\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\,\dot{\mathbf{f}}_{(i)}(\lambda,\boldsymbol{\theta})\,\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\,\dot{\mathbf{f}}_{(j)}(\lambda,\boldsymbol{\theta})\right\}d\lambda \\ &-\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\right)\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\,d\lambda\right\}. \end{aligned}$$

Lemma 4.5 and the consistency of $\hat{\theta}_T$ for θ_0 imply

$$\ddot{Q}_T(\boldsymbol{\theta}) \to_{a.s} M(\boldsymbol{\theta}_0). \tag{4.42}$$

In view of this result, the asymptotic normality of the estimator follows from the asymptotic normality of $\dot{Q}_T(\boldsymbol{\theta}_0)$, which is established by the following lemma.

Lemma 4.6 *Under Assumption 4.2 as $T \to \infty$,*

Part (a):

$$\sqrt{T}\left[\dot{Q}_T(\boldsymbol{\theta}_0) - E\dot{Q}_T(\boldsymbol{\theta}_0)\right] \to_d N(0, V(\boldsymbol{\theta}_0)), \tag{4.43}$$

where $V(\theta)$ is a positive definite matrix with (j,l) element,

$$V_{(j,l)}(\theta_0) = \frac{1}{\pi}\int_{-\pi}^{\pi}\mathbf{tr}\left[\mathbf{f}(\lambda,\theta_0)\left(\frac{\partial}{\partial\theta_j}\mathbf{f}^{-1}(\lambda,\theta_0)\right)\mathbf{f}(\lambda,\theta)\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}(\lambda,\theta_0)\right)\right]d\lambda$$

$$+\frac{1}{2\pi}\sum_{a,b,c,d=1}^{T}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\dot{\mathbf{f}}_{(j)}^{(a,b)}(\lambda_1,\theta_0)\dot{\mathbf{f}}_{(l)}^{(c,d)}(\lambda_2,\theta_0)\right\}\tilde{K}_{a,b,c,d}(-\lambda_1,\lambda_2,-\lambda_2,\theta_0)d\lambda_1 d\lambda_2.$$

Part (b):

$$\sqrt{T}E\dot{Q}_T(\boldsymbol{\theta}) \to 0.$$

The main idea of the proof of Lemma 4.6 is the approximation of each element of the vector $\sqrt{T}\left[\dot{Q}_T(\boldsymbol{\theta}_0) - E\dot{Q}_T(\boldsymbol{\theta}_0)\right]$ by a known function of the $\tilde{\tau}_{(a,b)}(m)$, in particular we establish, by means of an approximation, that $\sqrt{T}\left[\dot{Q}_T(\boldsymbol{\theta}_0) - E\dot{Q}_T(\boldsymbol{\theta}_0)\right]$ is, for a finite integer M, asymptotically equivalent to

$$\frac{1}{(2\pi)^2} tr\left\{\sum_{u=-M+1}^{M-1}\left(1-\frac{|u|}{M}\right)\mathbf{g}_{(j)}\left(u\right)\tilde{\tau}_{(a,b)}\left(u\right)\right\},$$

whose asymptotic normality follows from the asymptotic normality of the $\tilde{\tau}_{(a,b)}\left(u\right)$. The approximation is established in two steps. First consider that

$$\sqrt{T}\dot{Q}_T^{(j)}(\boldsymbol{\theta}_0)=\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}\left(\lambda\right)-\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right]\right\}d\lambda$$

can be expressed as

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}\left(\lambda\right)-\mathbf{EI}\left(\lambda\right)\right]\right\}d\lambda$$
$$+\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{EI}\left(\lambda\right)-\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right]\right\}d\lambda.$$

Noting that $\mathbf{EI}\left(\lambda\right)$ is the Cesaro sum of $\mathbf{f}(\lambda,\boldsymbol{\theta}_0)$ and that $\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)$ is Lipschitz continuous of degree $\alpha>1/2$ by Assumption 4.2[**G**] , we have that uniformly in $\lambda\in\Pi$

$$\sup_{\lambda\in\Pi}\left\|\mathbf{EI}\left(\lambda\right)-\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right\|=O\left(T^{-\alpha}\right),$$

(see Hannan, 1970, Theorem 3.15, page 513). Then

$$\sqrt{T}\dot{Q}_T^{(j)}(\boldsymbol{\theta}_0)=\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}\left(\lambda\right)-\mathbf{EI}(\lambda)\right]\right\}d\lambda+O\left(T^{1/2-\alpha}\right),$$

and the last term converges to zero as $T\to\infty$ and we approximate $\sqrt{T}\left[\dot{Q}_T^{(j)}(\theta_0)-E\dot{Q}_T^{(j)}(\theta_0)\right]$ by

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}\left(\lambda\right)-\mathbf{EI}(\lambda)\right]\right\}d\lambda. \tag{4.44}$$

We now introduce the quantities

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_M^{(j)}\left(\lambda,\theta_0\right)\left[\mathbf{I}\left(\lambda\right)-\mathbf{EI}(\lambda)\right]\right\}d\lambda, \tag{4.45}$$

where $\mathbf{g}_M^{(j)}\left(\lambda,\boldsymbol{\theta}_0\right)$ denotes the Cesaro sum of the Fourier series of $\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)$. The asymptotic distribution of (4.44) can be approximated by that of (4.45). To see this, put $\boldsymbol{\delta}_M^{(j)}\left(\lambda\right)=\mathbf{g}_{(j)}(\lambda,\theta_0)-\mathbf{g}_M^{(j)}(\lambda,\theta_0)$. Then

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\left[\mathbf{g}_{(j)}(\lambda,\theta_0)-\mathbf{g}_M^{(j)}(\lambda,\theta_0)\right]\left[\mathbf{I}\left(\lambda\right)-E\mathbf{I}(\lambda)\right]\right\}d\lambda$$

$$= \frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \boldsymbol{\delta}_M^{(j)} (\lambda) \left[\mathbf{I} (\lambda) - \mathbf{E}\mathbf{I}(\lambda) \right] \right\} d\lambda,$$

and we must evaluate

$$Var \left(\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \boldsymbol{\delta}_M^{(j)} (\theta) \left[\mathbf{I} (\lambda) - \mathbf{E}\mathbf{I}(\lambda) \right] \right\} d\lambda \right).$$

Lemma 4.7 *Under Assumption 4.2, as* $M \to \infty$

$$Var \left(\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \boldsymbol{\delta}_M^{(j)} (\theta) \left[\mathbf{I} (\lambda) - \mathbf{E}\mathbf{I}(\lambda) \right] \right\} d\lambda \right) = o(1).$$

By Bernstein's lemma (e.g., Hannan, 1970, page 242), Lemma 4.7 implies that the asymptotic normality of

$$\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_{(j)}(\lambda, \theta_0) \left[\mathbf{I} (\lambda) - E\mathbf{I}(\lambda) \right] \right\} d\lambda \tag{4.46}$$

is equivalent to that of

$$\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_M^{(j)} (\lambda, \theta_0) \left[\mathbf{I} (\lambda) - \mathbf{E}\mathbf{I}(\lambda) \right] \right\} d\lambda, \tag{4.47}$$

for every M. Finally,

$$\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_M^{(j)} (\lambda, \theta_0) \left[\mathbf{I} (\lambda) - \mathbf{E}\mathbf{I}(\lambda) \right] \right\} d\lambda$$

$$= \frac{1}{(2\pi)^2} tr \left\{ \sum_{u=-M+1}^{M-1} \left(1 - \frac{|u|}{M} \right) \sqrt{T} \left\{ \tilde{C}(-u) - \left(1 - \frac{|u|}{T} \right) \tilde{\Gamma}(-u) \right\} \mathbf{g}_{(j)}(u) \right\}, \tag{4.48}$$

where $\mathbf{g}_{(j)}(u) \equiv (1/2\pi) \int_{-\pi}^{\pi} e^{i\lambda u} \mathbf{g}_{(j)}(\lambda, \theta_0) d\lambda$ denotes the Fourier series of $\mathbf{g}_{(j)}(\lambda, \theta_0)$. The asymptotic normality of each element of the vector $\sqrt{T} \left[\dot{Q}_T(\boldsymbol{\theta}_0) - E\dot{Q}_T(\boldsymbol{\theta}_0) \right]$ is therefore implied by the asymptotic normality of (4.48). However (4.48) is equal to

$$\frac{1}{(2\pi)^2} tr \left\{ \sum_{u=-M+1}^{M-1} \left(1 - \frac{|u|}{M} \right) \mathbf{g}_{(j)}(u) \, \tilde{\tau}_{(a,b)}(u) \right\} + o_p(1),$$

whose asymptotic normality follows from the asymptotic normality of the $\tilde{\tau}_{(a,b)}(u)$ established by Conjecture 4.2 since M is finite.

All that remains is to evaluate the asymptotic covariance between $\sqrt{T} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_{(j)}(\lambda, \boldsymbol{\theta}_0) \mathbf{I}(\lambda) \right\} d\lambda$ and $\sqrt{T} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_{(l)}(\lambda, \boldsymbol{\theta}_0) \mathbf{I}(\lambda) \right\} d\lambda$. Indeed

$$Cov \left\{ \sqrt{T} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_{(j)}(\lambda) \mathbf{I}(\lambda) \right\} d\lambda, \sqrt{T} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_{(l)}(\lambda) \mathbf{I}(\lambda) \right\} d\lambda \right\}$$

$$= T\ Cov \left\{ \int_{-\pi}^{\pi} \sum_{a,b=1}^{n} \mathbf{g}_{(a,b)}^{(j)} (\lambda) \, \mathbf{I}_{(a,b)} (\lambda) \, d\lambda, \ \int_{-\pi}^{\pi} \sum_{c,d=1}^{n} \mathbf{g}_{(c,d)}^{(l)}(\lambda) \mathbf{I}_{(c,d)} (\lambda) \, d\lambda \right\}. \tag{4.49}$$

To proceed, we introduce the following lemma.

Lemma 4.8 *If Assumption 4.2 holds, for any scalar square integrable functions* h_1 *and* h_2 *defined on* $[-\pi, \pi]$,

$$\begin{aligned} & \lim_{T\to\infty} T\ Cov \left\{ \int_{-\pi}^{\pi} h_1(\lambda) \, I_{ab}(\lambda) \, d\lambda, \int_{-\pi}^{\pi} h_2(\lambda) \, I_{cd}(\lambda) \, d\lambda \right\} \\ = \ & 2\pi \int_{-\pi}^{\pi} h_1(\lambda) \, \bar{h}_2(\lambda) \, \mathbf{f}_{(a,c)}(\lambda) \, \bar{\mathbf{f}}_{(b,d)}(\lambda) \, d\lambda \\ & +2\pi \int_{-\pi}^{\pi} h_1(\lambda) \, \bar{h}_2(-\lambda) \, \mathbf{f}_{(a,d)}(\lambda) \, \bar{\mathbf{f}}_{(b,c)}(\lambda) \, d\lambda \\ & +2\pi \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} h_1(\lambda_1) \, \bar{h}_2(-\lambda_2) \, \tilde{K}_{abcd}(\lambda_1, \lambda_2, -\lambda_2) \, d\lambda_1 d\lambda_2. \end{aligned}$$

By Lemma 4.8, we may conclude that as $T \to \infty$, (4.49) converges to $V_{(j,l)}(\boldsymbol{\theta}_0)$.

For practical use of the asymptotic results, a consistent estimator of the asymptotic covariance matrix is required. As suggested in Zaffaroni (2009) for $M(\boldsymbol{\theta})$ such estimate can be obtained by substituting $\hat{\boldsymbol{\theta}}$ into $\ddot{Q}_{T(j)}(\boldsymbol{\theta})$. For $V(\boldsymbol{\theta})$ one can conjecture that the estimates provided by Hosoya and Taniguchi (1982, Section 5) will be consistent.

4.6 Appendix B: Technical lemmas

4.6.1 B.1 Preliminary lemmas

In this section we establish a number of properties of the model mainly in terms of its spectral density $\mathbf{f}(\lambda,\theta)$ and its derivatives. Recall that $\sim$ denotes asymptotic equivalence, tr is the trace operator, $\det$ is the determinant operator and $\|.\|$ is the Euclidean norm. Constants (not always the same) are denoted by K. Almost sure convergence and convergence in distribution are denoted respectively by $\rightarrow^{a.s.}$ and $\rightarrow^{d}$. We denote a positive integer number as r. The class of p-integrable functions on the set Π is denoted as $\mathbf{L}_p\left(\Pi\right)$.

Lemma B1.1 *Under Assumption 4.1*$[\mathbf{A}]$ *and* $(4.8\ )$, *the* y_t *are ergodic and strictly stationary.*

Proof The ergodicity and strict stationarity of the $\mathbf{x}_t$ follows from Nelson (1991, Theorem 2.1, page 251) and implies the ergodicity and strict stationarity of the $\mathbf{y}_t$. The $\mathbf{y}_t$ are covariance stationary if and only if (4.8) holds.

Lemma B1.2 *Under assumption 4.1*$[\mathbf{F}]$, $\mathbf{f}(\lambda,\theta)\neq\mathbf{f}(\lambda,\theta_0)$ *for every* $\theta\neq\theta_0$ *and for any* λ.

Proof Follows from the fact that the autocovariance function of the process uniquely identifies its power spectrum.

Lemma B1.3 *Under Assumptions 4.1,* $\mathbf{f}(\lambda,\boldsymbol{\theta})$ *has elements in* $L_2\left(\Pi\right)$, *bounded and continuous at all* $(\lambda,\boldsymbol{\theta})\in\Pi\times\Theta$.

Proof Assumption 4.1$[\mathbf{C}]$ implies that as $j\rightarrow\infty$

$$\sum_{j=0}^{\infty}\left|tr\left\{\Psi_j\left(\zeta\right)\right\}\right|<\infty.$$

However by (4.15)

$$tr\left\{\tilde{\Gamma}_u(\theta)\right\}=I_{(u=0)}tr\left\{\mathbf{\Sigma}_\xi\left(\tau\right)\right\}+tr\left\{\mathbf{\Sigma}_\epsilon\left(\tau\right)\sum_{j=0}^{\infty}\Psi_j\left(\zeta\right)\Psi_{j+u}^{'}\left(\zeta\right)\right\}+I_{(m\neq0)}tr\left\{\Psi_{|u|-1}\left(\zeta\right)\mathbf{\Sigma}_{\xi\epsilon}\left(\tau\right)\right\},$$

then Assumption 4.1$[\mathbf{A}]$ implies

$$\sum_{u=0}^{\infty}\left|tr\left\{\tilde{\Gamma}_u(\theta)\right\}\right|<\infty,$$

which is a sufficient condition for existence of the spectral density of the process. Moreover it implies its square integrability, and continuity at all $(\lambda, \theta) \in \Pi \times \Theta$ (see Giraitis et al., 2012, Chapter 2, Proposition 2.2.1, page 11). Uniform continuity and compactness of the parameter space (see Assumption 4.1[**B**]) imply that the element of $\mathbf{f}(\lambda, \boldsymbol{\theta})$ are bounded at all (λ, θ).

Lemma B1.4 *Under Assumption 4.1,* $(\partial/\partial\theta)\,\mathbf{f}\,(\lambda, \boldsymbol{\theta})$ *has elements in* $L_2\,(\Pi)$ *which are bounded and continuous at all* $(\lambda, \boldsymbol{\theta}) \in \Pi \times \Theta$.

Proof For any $j = 1, ..., s$,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}_j}\mathbf{f}\,(\lambda, \boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}_j}\left[\frac{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\,(\boldsymbol{\tau})}{2\pi}\right] + \frac{\partial}{\partial \boldsymbol{\theta}_j}\left[\frac{\mathbf{k}\left(e^{i\lambda}, \zeta\right)\boldsymbol{\Sigma}_{\epsilon}\,(\boldsymbol{\tau})\,\mathbf{k}\left(e^{i\lambda}, \zeta\right)^*}{2\pi}\right] \\ &\quad + \frac{\partial}{\partial \boldsymbol{\theta}_j}\left[\frac{\boldsymbol{\Sigma}_{\epsilon\xi}\,(\boldsymbol{\tau})\,e^{-i\lambda}\mathbf{k}\left(e^{i\lambda}, \zeta\right)^* + e^{i\lambda}\mathbf{k}\left(e^{i\lambda}, \zeta\right)\boldsymbol{\Sigma}_{\epsilon\xi}'\,(\boldsymbol{\tau})}{2\pi}\right].\end{aligned}$$

Assumption 4.1[**D**] implies that for any $j = 1, ..., s$,

$$\sum_{u=0}^{\infty}\left|tr\left\{\frac{\partial}{\partial \theta_j}\tilde{\Gamma}_u(\theta)\right\}\right| < \infty,$$

which is sufficient condition for the existence of the first derivative of $\mathbf{f}\,(\lambda, \boldsymbol{\theta})$. Moreover it implies that $(\partial/\partial\theta)\,\mathbf{f}\,(\lambda, \boldsymbol{\theta})$ has elements in $L_2\,(\Pi)$ which are continuous at all $(\lambda, \boldsymbol{\theta}) \in \Pi \times \Theta$ (see Giraitis et al, 2012, Proposition 2.2.1., page 11). By compactness of the parameter space (see Assumption 4.2[B]), the uniform continuity implies that the elements of $(\partial/\partial\theta)\,\mathbf{f}\,(\lambda, \boldsymbol{\theta})$ are bounded at all $(\lambda, \boldsymbol{\theta}) \in \Pi \times \Theta$.

Lemma B1.5 *Under Assumption 4.1*[F], $\mathbf{f}(\lambda, \theta)$ *is a strictly positive definite matrix for all* $\theta \in \Theta$, $\lambda \in \Pi$.

Proof The autocovariance function of the process defined in (4.15) is

$$\tilde{\boldsymbol{\Gamma}}(\boldsymbol{\theta}, u) = I_{(u=0)}\boldsymbol{\Sigma}_{\xi}\,(\boldsymbol{\tau}) + \boldsymbol{\Sigma}_{\epsilon}\,(\boldsymbol{\tau})\sum_{j=0}^{\infty}\Psi_j\,(\boldsymbol{\zeta})\,\Psi_{j+u}'\,(\boldsymbol{\zeta}) + I_{(m\neq 0)}\Psi_{|u|-1}\,(\boldsymbol{\zeta})\,\boldsymbol{\Sigma}_{\xi\epsilon}\,(\boldsymbol{\tau})\,.$$

Under Assumption[F], $\boldsymbol{\Sigma}_{\epsilon}\,(\boldsymbol{\tau})$ is positive definite for every value of $\boldsymbol{\tau}$ in the parameter space. Since by definition $\boldsymbol{\Sigma}_{\xi}\,(\boldsymbol{\tau})$ and $\boldsymbol{\Sigma}_{\xi\epsilon}\,(\boldsymbol{\tau})$ are positive semidefinite covariance matrices, $\tilde{\boldsymbol{\Gamma}}(\boldsymbol{\theta}, u)$ is expressed as the sum of two positive semidefinite matrices and one positive definite matrix. Therefore it is positive definite. By definition the spectrum is the unique Fourier transform of the autocovariance matrix and its positive definiteness is implied by the positive definiteness of $\tilde{\mathbf{\Gamma}}(\boldsymbol{\theta}, u)$.

Lemma B1.6 *Under Assumption 4.1, y_t is purely non deterministic with Wold decomposition*

$$\mathbf{y}_t = \sum_{l=0}^{\infty} A_l(\theta) e_{t-l}, \quad \sum_{l=0}^{\infty} \|A_l(\theta)\|^2 < \infty,$$

where the e_t are n dimensional white noise vectors.

Proof Since $\mathbf{y}_t$ is a stationary zero mean process, the result follows once we establish that

$$\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \theta) > -\infty,$$

(see Giraitis et al, 2012, Theorem 3.2.1, page 38). By the logarithm inequality

$$\left|1 - \det \mathbf{f}^{-1}(\lambda, \boldsymbol{\theta})\right| \le |\log\det \mathbf{f}(\lambda, \boldsymbol{\theta})| \le |\det \mathbf{f}(\lambda, \boldsymbol{\theta}) - 1|$$

and the result follows from the continuity of $\mathbf{f}^{-1}(\lambda, \theta)$ at all (λ, θ) by Assumption 4.1[**G**] and Lemma B1.5**.**

Lemma B1.7 *Under Assumption 4.1,* $\log\det \mathbf{f}(\lambda, \boldsymbol{\theta})$ *is differentiable in $\theta \in \Theta$ under the integral sign.*

Proof Denoting the jth unit vector in R^s by i_j, we have

$$\begin{aligned} & \frac{1}{\varepsilon}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \theta + i_j\varepsilon)\, d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \theta)\, d\lambda\right] \\ = \; & \frac{1}{\varepsilon}\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \theta + i_j\varepsilon) - \log\det \mathbf{f}(\lambda, \theta)\, d\lambda. \end{aligned}$$

By the mean value theorem the integrand is bounded by

$$\left|\frac{\partial}{\partial \theta_j^*} \log\det \mathbf{f}(\lambda, \theta^*)\right| = \left|tr\left\{\mathbf{f}^{-1}(\lambda, \theta^*) \frac{\partial}{\partial \theta_j^*} \mathbf{f}(\lambda, \theta^*)\right\}\right|,$$

where $|\theta^*(\lambda) - \theta| < |\varepsilon|$. By Assumption 4.1 [**G**], Lemma B1.4 and compactness of the parameter space

$$\left|tr\left\{\mathbf{f}^{-1}(\lambda, \theta^*) \frac{\partial}{\partial \theta_j^*} \mathbf{f}(\lambda, \theta^*)\right\}\right| < K,$$

where K is a positive constant that does not depend on θ. Then

$$\int_{-\pi}^{\pi} \left|tr\left\{\mathbf{f}^{-1}(\lambda, \theta^*) \frac{\partial}{\partial \theta_j^*} \mathbf{f}(\lambda, \theta^*)\right\}\right| d\lambda < \infty,$$

and the dominated convergence theorem implies that $\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda,\theta)$ can be differentiated under the integral sign.

Lemma B1.8 *Under assumption 4.2, for* $r=1,2,3$, $(\partial^r/\partial\theta_{j_1}...\theta_{j_r})\,\mathbf{f}(\lambda,\theta)$ *has elements in* $L_2(\Pi)$ *which are bounded and continuous at all* $(\lambda,\theta)\in\Pi\times\Theta$.

Proof For any $j_h = 1,...,s$, with $h=1,...,r$ and $r=1,2,3$,

$$\begin{aligned}\frac{\partial^r}{\partial\boldsymbol{\theta}_{j_1}...\partial\boldsymbol{\theta}_{j_r}}\mathbf{f}(\lambda,\boldsymbol{\theta}) &= \frac{\partial^r}{\partial\boldsymbol{\theta}_{j_1}...\partial\boldsymbol{\theta}_{j_r}}\left[\frac{\boldsymbol{\Sigma}_{\boldsymbol{\xi}}(\boldsymbol{\tau})}{2\pi}\right]+\frac{\partial^r}{\partial\boldsymbol{\theta}_{j_1}...\partial\boldsymbol{\theta}_{j_r}}\left[\frac{\mathbf{k}\left(e^{i\lambda},\zeta\right)\boldsymbol{\Sigma}_{\epsilon}(\boldsymbol{\tau})\,\mathbf{k}\left(e^{i\lambda},\zeta\right)^*}{2\pi}\right]\\ &+\frac{\partial^r}{\partial\boldsymbol{\theta}_{j_1}...\partial\boldsymbol{\theta}_{j_r}}\left[\frac{\boldsymbol{\Sigma}_{\epsilon\xi}(\boldsymbol{\tau})\,e^{-i\lambda}\mathbf{k}\left(e^{i\lambda},\zeta\right)^*+e^{i\lambda}\mathbf{k}\left(e^{i\lambda},\zeta\right)\boldsymbol{\Sigma}'_{\epsilon\xi}(\boldsymbol{\tau})}{2\pi}\right].\end{aligned}$$

Assumption 4.2[D] implies that for any $j_h = 1,...,s$, with $h=1,...,r$ and $r=1,2,3$

$$\sum_{u=0}^{\infty}\left|tr\left\{\frac{\partial^r}{\partial\theta_{j_1}...\theta_{j_r}}\Gamma_u(\theta)\right\}\right|<\infty,$$

which is sufficient condition for the existence of the rth derivative of $\mathbf{f}(\lambda,\theta)$. Moreover it implies that $(\partial^r/\partial\boldsymbol{\theta}_{j_1}...\boldsymbol{\theta}_{j_r})\,\mathbf{f}(\lambda,\theta)$ has elements in $L_2(\Pi)$ which are continuous at all $(\lambda,\theta)\in\Pi\times\Theta$ (see Giraitis et al, 2012, Proposition 2.2.1, page 11). By compactness of the parameter space (see Assumption 4.2[B]), the uniform continuity implies that the elements of $(\partial^r/\partial\boldsymbol{\theta}_{j_1}...\boldsymbol{\theta}_{j_r})\,\mathbf{f}(\lambda,\theta)$ are bounded at all $(\lambda,\theta)\in\Pi\times\Theta$.

Lemma B1.9 *Under Assumption 4.2,* $\mathbf{g}(\lambda,\theta)\equiv\mathbf{f}^{-1}(\lambda,\theta)\,\dot{\mathbf{f}}(\lambda,\theta)$ *is uniformly continuous in* (λ,θ).

Proof The uniform continuity of $\mathbf{g}(\lambda,\theta)$ follows once we establish

$$\sup_{\theta^*\in\Theta}\left\|\frac{\partial}{\partial\theta^*}\mathbf{g}(\lambda,\theta^*)\right\|<\infty,$$

where $|\theta^*(\lambda)-\theta|<|\varepsilon|$, (see Davidson, 1994, Theorem 21.10, page 339). However

$$\frac{\partial}{\partial\theta^*}\mathbf{g}(\lambda,\theta^*)=\mathbf{f}^{-1}(\lambda,\theta^*)\,\ddot{\mathbf{f}}(\lambda,\theta)+\left(\frac{\partial}{\partial\theta^*}\mathbf{f}^{-1}(\lambda,\theta^*)\right)\dot{\mathbf{f}}(\lambda,\theta),$$

which by Assumption 4.2[**G**] (ii) and (iii), Lemma B1.8 and compactness of the parameter space is bounded by a positive constant K for all $\theta^*\in\Theta$.

Lemma B1.10 *Under Assumption 4.2,* $\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\theta)\,d\lambda$ *can be differentiated twice under the* *integral sign.*

Proof By the same argument used in Lemma B1.7, the result follows from Assumption 4.2[**G**] (ii) and (iii) and Lemma B1.9.

Lemma B1.11 *Under Assumption 4.2, for* $1 \leq a, b, c, d \leq n$.

$$\sum_{t_1,t_2,t_3,t_4=-\infty}^{\infty} \left| K^{\mathbf{y}}_{abcd}\left(t_1,t_2,t_3,t_4\right)\right| < \infty.$$

Proof Denote as $K_{abcd}\left(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t, \mathbf{u}_t\right)$ the fourth order cumulant of elements a, b, c, d of random vectors $\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t, \mathbf{u}_t$. Set $K^{\epsilon}_{abcd} = cumulant(\epsilon^{(a)}_0, \epsilon^{(b)}_0, \epsilon^{(c)}_0, \epsilon^{(d)}_0)$ and set $K^{\xi}_{abcd} = cumulant(\xi^{(a)}_0, \xi^{(b)}_0, \xi^{(c)}_0, \xi^{(d)}_0)$.

Then $K^{\mathbf{y}}_{abcd}\left(t_1,t_2,t_3,t_4\right)$ is made by the sum of the following terms:

$$\sum_{a,b,c,d=1}^{n} \left(K^{\xi}_{abcd} \mathit{1}\left(t_1 = t_2 = t_3 = t_4\right)\right) \tag{4.51}$$
$$\sum_{r=a,b,c,d} \left(\sum_{a,b,c,d=1}^{n} K_{abcd}\left(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0, \boldsymbol{\xi}_0, \epsilon_0\right) \Psi^{(r)}_{t_4-t_1-1} \mathit{1}\left(t_1 = t_2 = t_3\right)\right)$$
$$\sum_{u,v=a,b,c,d} \left(\sum_{a,b,c,d=1}^{n} K_{abcd}\left(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0, \epsilon_0, \epsilon_0\right) \Psi^{(u,u)}_{t_3-t_2-1} \Psi^{(v)}_{t_4-t_1-1} \mathit{1}\left(t_1 = t_2\right)\right)$$
$$\sum_{u,v,z=a,b,c,d} \left(\sum_{a,b,c,d=1}^{n} K_{abcd}\left(\boldsymbol{\xi}_0, \epsilon_0, \epsilon_0, \epsilon_0\right) \Psi^{(u)}_{t_2-t_1-1} \Psi^{(v)}_{t_3-t_1-1} \Psi^{(z)}_{t_4-t_2-1}\right)$$
$$\sum_{a,b,c,d=1}^{n} \left(K^{\epsilon}_{abcd} \sum_{j=0}^{\infty} \Psi^{(a)}_{j} \Psi^{(b)}_{j+t_2-t_1} \Psi^{(c)}_{j+t_3-t_1} \Psi^{(d)}_{j+t_4-t_1}\right).$$

The absolute summability of the cumulants follows from the absolute summability of the last term in (4.51), which is implied by Assumption 4.2 [C].

Lemma B1.12 *Under Assumption 4.2, the trispectrum of* $\mathbf{y}_t$,

$$\tilde{K}^{\mathbf{y}}_{abcd}\left(\lambda_1,\lambda_2,\lambda_3\right) = \frac{1}{\left(2\pi\right)^3} \sum_{t_1,t_2,t_3=-\infty}^{\infty} \exp\left\{-i\left(\lambda_1 t_1 + \lambda_2 t_2 + \lambda_3 t_3\right)\right\} K^{\mathbf{y}}_{abcd}\left(t_1,t_2,t_3\right)$$

is square integrable.

Proof Follows from the square summability of the Fourier coefficients of $\tilde{K}^{\mathbf{y}}_{abcd}$, implied by Lemma B1.11 (see Giraitis et al., 2012, (2.1.4), page 8).

4.6.2 B.2 Consistency lemmas

This section contains the proof of Lemma 4.1, part **(b)**, and Lemma 4.2. To prove Lemma

4.1 we follow Giraitis et al.(Chapter 8, Section 8.2) and establish two preliminary results. First, in Lemma 4.3, we establish an approximation for matrix functions and then in Lemma 4.4, we establish uniform almost sure convergence of some discrete functions of the periodogram under minimal condition on the underlying model, generalizing Lemma 1 of Hannan (1973) to matrix functions. In what follows, for any matrix function $\mathbf{h}(\lambda, \theta)$, we denote by

$$\mathbf{h}_u(\theta) = \int_{-\pi}^{\pi} e^{iu\lambda} \mathbf{h}(\lambda, \theta)\, d\lambda, \quad u = 0, \pm 1, \pm 2, ...$$

its Fourier coefficients, and we denote by

$$\mathbf{q}_M(\lambda, \boldsymbol{\theta}) = \sum_{u=-M}^{M} \left(1 - \frac{|u|}{M}\right) \mathbf{h}_u(\boldsymbol{\theta})\, e^{-iu\lambda},$$

the Cesaro sum of its Fourier coefficients up to M terms.

Lemma B2.1 *Let $\mathbf{h}(\lambda, \theta)$ be a $n \times n$ matrix function, continuous in $\lambda \in \Pi$ and such that* $\mathbf{h}(-\pi, \theta) = \mathbf{h}(\pi, \theta)$ in $[-\pi, \pi]$. *Then $\mathbf{h}(\lambda, \theta)$ may be approximated uniformly in λ by $\mathbf{q}_M(\lambda, \boldsymbol{\theta})$,*

$$\sup_{\lambda \in \Pi} \|\mathbf{h}(\lambda, \theta) - \mathbf{q}_M(\lambda, \boldsymbol{\theta})\| \to 0 \quad \text{as} \quad M \to \infty.$$

If in addition $\mathbf{h}(\lambda, \theta)$ is continuous in λ uniformly in θ, the approximation may be made uniformly in θ also,

$$\sup_{\lambda, \theta} \|\mathbf{h}(\lambda, \theta) - \mathbf{q}_M(\lambda, \boldsymbol{\theta})\| \to 0 \quad \text{as} \quad M \to \infty.$$

Proof A detailed proof of this lemma for matrix functions can be found in Hannan (1970, Mathematical Appendix, Section 3).

Lemma B2.2 *Let $\mathbf{y}_t$ be a stationary, ergodic and purely non deterministic vector process, with $n \times n$ spectral density matrix $\mathbf{f}(\lambda, \theta_0)$. Let $\mathbf{h}(\lambda, \theta)$ be a $n \times n$ matrix function, continuous in $(\lambda, \theta) \in \Pi \times \Theta$ and such that $\mathbf{h}(\lambda, \theta) = \mathbf{h}(-\lambda, \theta)$. Then, uniformly in $\theta \in \Theta$ and $\lambda \in \Pi$,*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda, \boldsymbol{\theta}) \mathbf{I}_T(\lambda)\right\} d\lambda \ \to \frac{1}{2\pi} \int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda, \boldsymbol{\theta})\, \mathbf{f}(\lambda, \boldsymbol{\theta}_0)\right\} d\lambda \quad \text{a.s.}$$

Proof By Lemma B2.1, for every $\eta > 0$ we may find M large enough such that:

$$\sup_{\lambda, \theta} \|\mathbf{h}(\lambda, \theta) - \mathbf{q}_M(\lambda, \theta)\| \leq \eta.$$

Let $\eta > 0$. For sufficiently large M, uniformly in θ :

$$\frac{1}{2\pi}\left|\int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\}d\lambda - \int_{-\pi}^{\pi} tr\left\{\mathbf{q}_M(\lambda_t,\boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\}d\lambda\right|$$

$$= \frac{1}{2\pi}\left|\int_{-\pi}^{\pi} tr\left\{\left(\mathbf{h}(\lambda,\boldsymbol{\theta}) - \mathbf{q}_M(\lambda,\boldsymbol{\theta})\right)\mathbf{I}_T(\lambda)\right\}d\lambda\right|$$

$$\leq \frac{\eta}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{I}_T(\lambda)\right\}d\lambda = \frac{\eta}{2\pi}\, tr\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{u=-T+1}^{T-1}\tilde{C}\left(u\right)e^{-iu\lambda}\right\} = \frac{\eta}{2\pi} tr\left\{\tilde{C}\left(0\right)\right\}.$$

Because the process is ergodic, by the Ergodic Theorem (see Giraitis et al., Chapter 2, Section 2.5) $\tilde{C}\left(0\right)$ converges almost surely to its population analogue $\tilde{\Gamma}\left(0\right)$ as $T \to \infty$. Thus for all sufficiently large t, uniformly in $\theta \in \Theta$:

$$\left|\int_{-\pi}^{\pi} tr\left\{\mathbf{h}(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\}d\lambda - \int_{-\pi}^{\pi} tr\left\{\mathbf{q}_M(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\}d\lambda\right| \leq \eta\, tr\left\{\tilde{\Gamma}\left(0\right)\right\} \quad \text{a.s.}$$

Moreover,

$$\int_{-\pi}^{\pi} tr\left\{\mathbf{q}_M(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda)\right\}d\lambda = \int_{-\pi}^{\pi} tr\left[\sum_{u=-M}^{M}\left(1-\frac{|u|}{M}\right)\mathbf{h}_u\left(\boldsymbol{\theta}\right)e^{-iu\lambda}\mathbf{I}_T(\lambda)\right]d\lambda$$

$$= tr\left[\sum_{u=-M}^{M}\left(1-\frac{|u|}{M}\right)\mathbf{h}_u\left(\boldsymbol{\theta}\right)\tilde{C}\left(u\right)\right].$$

By the Ergodic Theorem for each $|u| \leq M$, as $T \to \infty$, $\tilde{C}\left(u\right)$ converges almost surely to $\tilde{\Gamma}\left(u\right) = \int_{-\pi}^{\pi}\mathbf{f}\left(\lambda,\theta\right)e^{-i\lambda u}d\lambda$. Therefore the above expression tends almost surely to:

$$tr\left[\sum_{u=-M}^{M}\left(1-\frac{|u|}{M}\right)\mathbf{h}_u\left(\boldsymbol{\theta}\right)\tilde{\Gamma}\left(u\right)\right]$$

$$= tr\left[\sum_{u=-M}^{M}\left(1-\frac{|u|}{M}\right)\mathbf{h}_u\left(\theta\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{f}(\lambda,\theta_0)e^{-iu\lambda}d\lambda\right)\right]$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{q}_M(\lambda,\theta)\mathbf{f}(\lambda,\theta_0)\right]d\lambda \to \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{h}(\lambda,\theta)\mathbf{f}(\lambda,\theta_0)\right]d\lambda$$

on letting $\eta \to 0$, which completes the proof.

Lemma 4.1 Part (a) *If Assumption 4.1 holds,*

$$lim_{T\to\infty}Q_T\left(\boldsymbol{\theta}\right) = Q\left(\boldsymbol{\theta}\right) \quad \textit{a.s.,}$$

and the convergence is uniform in $\theta \in \Theta$.

Proof The almost sure uniform convergence of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\mathbf{f}^{-1}(\lambda_t,\boldsymbol{\theta})\mathbf{I}_T(\lambda_t)\right)$$

to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right)d\lambda$$

follows from Lemma B2.2, taking $\mathbf{h}(\lambda,\theta) \equiv \mathbf{f}^{-1}(\lambda,\theta)$. By Assumption 4.1[**G**], $\mathbf{f}^{-1}(\lambda,\theta)$ is uniformly continuous at all $(\lambda,\theta) \in \Pi \times \Omega$, moreover it satisfies $\mathbf{f}^{-1}(-\pi,\theta) = \mathbf{f}^{-1}(\pi,\theta)$ and the conditions of Lemma B2.2 are satisfied.

Consider the first term of $Q_T(\theta)$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det\mathbf{f}(\lambda,\boldsymbol{\theta})d\lambda.$$

This term is non stochastic so its uniform convergence follows once with establish the equicontinuity property

$$\lim_{\varepsilon\to 0}\sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|\leq\epsilon}\left|\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\tilde{\theta}\right)d\lambda-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta\right)d\lambda\right|\to 0. \tag{4.52}$$

(4.52) is implied by

$$\sup_{\theta^*\in\Theta}\left|\frac{\partial}{\partial\theta}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta^*\right)d\lambda\right|<\infty,$$

where $|\theta^*(\lambda)-\theta| < |\varepsilon|$, (see Davidson, 1994, Theorem 21.10, page 339). By Lemma B1.7,

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\theta^*\right)d\lambda\\
=&\ \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta}\log\det\mathbf{f}\left(\lambda,\theta^*\right)d\lambda\\
=&\ \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\frac{\partial}{\partial\theta}\mathbf{f}\left(\lambda,\theta^*\right)\right\},
\end{aligned}$$

and the integrand is bounded by some positive constant by Assumption 4.1[**G**](i), Lemma B1.4 and compactness of the parameter space. The use of the dominated convergence theorem allows to conclude that

$$\sup_{\theta^*\in\Theta}\left|\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\frac{\partial}{\partial\theta}\mathbf{f}\left(\lambda,\theta^*\right)\right\}\right|<K,$$

which concludes the proof.

Lemma 4.1 (b) *If Assumption 4.1 holds, then for any $\eta > 0$, uniformly in $\theta \in \Theta$,*

$$\lim_{T\to\infty} Q_{T,\eta}(\theta) = Q_\eta(\theta) \quad \text{a.s.}$$

Proof The almost sure uniform convergence of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\boldsymbol{\phi}_\eta(\lambda_t, \boldsymbol{\theta})\mathbf{I}_T(\lambda_t)\right)$$

to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\boldsymbol{\phi}_\eta(\lambda, \boldsymbol{\theta})\mathbf{f}(\lambda, \boldsymbol{\theta}_0)\right) d\lambda$$

follows from Lemma 4.4, taking $\mathbf{h}(\lambda, \theta) \equiv \boldsymbol{\phi}_\eta(\lambda, \theta)$. By Assumption 4.1[**G**], $\boldsymbol{\phi}_\eta(\lambda, \theta)$ is uniformly continuous in $(\lambda, \theta) \in \Pi \times \Omega$, and it satisfies $\boldsymbol{\phi}_\eta(\lambda, \theta) = \boldsymbol{\phi}_\eta(-\lambda, \theta)$ for all $\lambda \in \Pi$, and the conditions of Lemma 4.4 are satisfied. The uniform almost sure convergence of the first term of $Q_{T,\eta}(\theta)$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \boldsymbol{\theta}) d\lambda,$$

follows from Lemma 4.2 part (a).

Lemma 4.2 *If Assumption 4.1 holds, for any $\theta \in \Theta$*

$$\inf_{\theta\in\Theta} Q(\theta) = Q(\theta_0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \boldsymbol{\theta}_0) d\lambda + T.$$

Proof

$$Q(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \boldsymbol{\theta}) d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{f}^{-1}(\lambda, \theta)\, \mathbf{f}(\lambda, \boldsymbol{\theta}_0\right] d\lambda, \tag{4.53}$$

adding and subtracting $1/2\pi \int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \theta_0)\, d\lambda$, (4.53) is equal to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda, \boldsymbol{\theta}_0) d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{f}^{-1}(\lambda, \theta)\, \mathbf{f}(\lambda, \boldsymbol{\theta}_0\right] d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{\det \mathbf{f}(\lambda, \boldsymbol{\theta}_0)}{\det \mathbf{f}(\lambda, \boldsymbol{\theta})} d\lambda,$$

because for any non-singular matrix A, $\det^{-1}(A) = \det(A^{-1})$ (Luktepohl, 1996, Section 3.4.4, Result (f)), (4.53) is equal to

$$Q(\boldsymbol{\theta}_0) - T + \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{f}^{-1}(\lambda, \theta)\, \mathbf{f}(\lambda, \boldsymbol{\theta}_0\right] d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(\det\left(\mathbf{f}^{-1}(\lambda, \boldsymbol{\theta})\right) \det\left(\mathbf{f}(\lambda, \boldsymbol{\theta}_0)\right)\right) d\lambda,$$

because for any non singular matrix A and B, $\det(A) \times \det(B) = \det(AB)$ (Luktepohl, 1996, Section 4.2.1, Result (4)), (4.53) is equal to

$$Q(\boldsymbol{\theta}_0)+\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left[\mathbf{f}^{-1}(\lambda,\theta)\,\mathbf{f}(\lambda,\boldsymbol{\theta}_0\right]d\lambda - T\right\}-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\det\left(\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right)\right)d\lambda > Q\left(\theta_0\right),$$

where the strict inequality follows, in view of Lemma B1.5, because $\log\det\left(A\right)\leq tr\left(A\right)-T$ for any positive definite matrix A with equality holding if and only if $A = I_n$ (Luktepohl, 1996, Section 4.1.2, Result (10)).

4.6.3 B.3 Asymptotic Normality lemmas

This section contains the proof of the lemmas used to establish the asymptotic normality of the estimator.

Lemma B3.1 *The asymptotic covariance between* $\tilde{\tau}_{(a,b)}\left(m\right)$ *and* $\tilde{\tau}_{(c,d)}\left(u\right)$ *is given as*

$$\begin{aligned}&2\pi\int_{-\pi}^{\pi}\left\{\mathbf{f}_{(a,c)}\left(\lambda,\theta\right)\overline{\mathbf{f}}_{(b,d)}\left(\lambda,\theta\right)e^{-i(m-u)\lambda}+\mathbf{f}_{(a,d)}\left(\lambda,\theta\right)\overline{\mathbf{f}}_{(b,c)}\left(\lambda,\theta\right)e^{i(m+u)\lambda}\right\}d\lambda\\&+2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{(im\lambda_1+iu\lambda_2)}\tilde{K}^{\mathbf{y}}_{abcd}(-\lambda_{1,}\lambda_2,-\lambda_2)d\lambda_1 d\lambda_2.\end{aligned}$$

Proof The covariance between $\tilde{\tau}_{(a,b)}\left(m\right)$ and $\tilde{\tau}_{(c,d)}\left(n\right)$ is

$$\begin{aligned}&\frac{1}{T}\sum_{u=1-T}^{T-1}\left(1-\frac{|u|}{T}\right)\left\{\tilde{\Gamma}_{(a,c)}\left(n\right)\tilde{\Gamma}_{(b,d)}\left(u+n-m\right)+\tilde{\Gamma}_{(a,d)}\left(u+n\right)\tilde{\Gamma}_{(b,c)}\left(u-m\right)\right\}\\&+\sum_{u=1-T}^{T-1}\left(1-\frac{|u|}{T}\right)K^{\mathbf{y}}_{a,b,c,d}\left(m,u,u+n\right),\end{aligned}\tag{4.54}$$

(see Hannan, 1979, page 209-211). The term

$$\frac{1}{T}\sum_{u=1-T}^{T-1}\left(1-\frac{|u|}{T}\right)\tilde{\Gamma}_{(a,c)}\left(n\right)\tilde{\Gamma}_{(b,d)}\left(u+n-m\right),\tag{4.55}$$

is the Cesaro sum, evaluated at the origin, of $(4\pi)^2$ the uth Fourier coefficient of the convolution of $\mathbf{f}_{(a,c)}\left(\lambda\right)$ with $\mathbf{f}_{(b,d)}\left(\lambda\right)e^{-i(m-n)\lambda}$. By Lemma B1.3, $\mathbf{f}\left(\lambda\right)$ has elements in L_2, so their convolution is continuous. Then (4.55) converges to

$$2\pi\int_{-\pi}^{\pi}\mathbf{f}_{(a,c)}\left(\lambda,\theta\right)\overline{\mathbf{f}}_{(b,d)}\left(\lambda,\theta\right)e^{-i(n-m)\lambda}d\lambda.$$

The same argument applies to

$$\frac{1}{T}\sum_{u=1-T}^{T-1}\left(1-\frac{|u|}{T}\right)\tilde{\Gamma}_{(a,d)}\left(u+n\right)\tilde{\Gamma}_{(b,c)}\left(u-m\right),$$

which converges to

$$2\pi \int_{-\pi}^{\pi} \mathbf{f}_{(a,d)}\left(\lambda,\theta\right) \overline{\mathbf{f}}_{(b,c)}\left(\lambda,\theta\right) e^{i(m+u)\lambda} d\lambda.$$

(4.54) is the Cesaro sum, evaluated at the zero frequency, of the Fourier coefficients of the function $\tilde{K}^{\mathbf{y}}_{abcd}\left(\lambda_1,\lambda_2,\lambda_3\right) e^{-i(n\lambda_1+m\lambda_2)}$. By Lemma B1.13, the trispectrum of the process is square integrable, implying the convergence of (4.54) to

$$2\pi \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{(in\lambda_1+im\lambda_2)} \tilde{K}^{\mathbf{y}}_{abcd}\left(-\lambda_1,\lambda_2,-\lambda_2\right) d\lambda_1 d\lambda_2.$$

Proof of Lemma 4.5 We establish the uniform convergence of $\ddot{Q}_T(\boldsymbol{\theta})$ to $M(\boldsymbol{\theta})$ pointwise. The (i,j) element of $\ddot{Q}_T(\boldsymbol{\theta})$, $\ddot{Q}_T^{(i,j)}(\boldsymbol{\theta})$ is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right) \ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\} d\lambda \tag{4.56}$$

$$-\frac{1}{2\pi} \int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right) \dot{\mathbf{f}}_{(i)}\left(\lambda,\theta\right) \mathbf{f}^{-1}\left(\lambda,\theta\right) \dot{\mathbf{f}}_{(j)}\left(\lambda,\theta\right)\right\} d\lambda \tag{4.57}$$

$$+\frac{1}{2\pi} \int_{-\pi}^{\pi} tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'} \mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\right) \mathbf{I}_T(\lambda)\right\} d\lambda. \tag{4.58}$$

The last term converges almost surely uniformly in $(\lambda,\theta) \in \Pi \times \Theta$ to

$$\frac{1}{2\pi} \int tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'} \mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\right) \mathbf{f}(\lambda,\theta_0)\right\} d\lambda$$

by Lemma 4.4, taking $\mathbf{h}\left(\lambda,\theta\right) \equiv \left(\partial^2/\partial\theta_i\partial\theta_j'\right) \mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})$, which is continuos at all $(\lambda,\theta) \in \Pi \times \Theta$ by Assumption 4.2[**G**] and symmetric around zero in $[-\pi,\pi]$.

The first two terms of (4.56) are non stochastic. Their uniform convergence in θ follows once with establish their equicontinuity property. Consider the first term. We want to show that

$$\sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|\leq\epsilon} \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[tr\left\{\mathbf{f}^{-1}\left(\lambda,\tilde{\theta}\right) \ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\tilde{\theta}\right)\right\} - tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right) \ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\}\right] d\lambda \right| \to 0 \quad \text{a.s..} \tag{4.59}$$

(4.59) is implied by

$$\sup_{\theta\in\Theta} \left| \frac{\partial}{\partial\theta} \int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right) \ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\} d\lambda \right| < \infty, \tag{4.60}$$

(see Davidson, 1994, Theorem 21.10, page 339). We must establish that $\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right) \ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\} d\lambda$

is differentiable under the integral sign. Denote the jth unit vector in R^s by i_j, and consider

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{\varepsilon}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta+i_j\varepsilon\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta+i_j\varepsilon\right)-\mathbf{f}^{-1}\left(\lambda,\theta\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\}d\lambda.$$

By the mean value theorem the integrand is dominated for each λ by

$$\left|\frac{\partial}{\partial\theta_l}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\left(\lambda\right)\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta^*\left(\lambda\right)\right)\right\}\right|, \tag{4.61}$$

where $|\theta^*\left(\lambda\right)-\theta|<|\varepsilon|$. Taking derivatives (4.61) is equal to

$$\left|tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\left(\lambda\right)\right)\left(\frac{\partial}{\partial\theta_l}\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta^*\left(\lambda\right)\right)\right)+\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}\left(\lambda,\theta^*\left(\lambda\right)\right)\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta^*\left(\lambda\right)\right)\right\}\right|. \tag{4.62}$$

By Assumption 4.2[**G**](ii) and (iii), Lemma B1.8 and compactness of the parameter space, (4.62) is at most K, where K denotes a generic positive constant. The use of the dominated convergence theorem allows to conclude that

$$\left|\frac{\partial}{\partial\theta}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\theta\right)\right\}d\lambda\right|<\infty,$$

which completes the proof of (4.59).

The equicontinuity property of the second term of (4.56) is implied by

$$\sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta_l}\frac{1}{2\pi}\int_{-\pi}^{\pi}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\theta\right)\mathbf{f}^{-1}\left(\lambda,\theta\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\theta\right)\right\}d\lambda\right|<\infty, \tag{4.63}$$

(see Davidson, 1994, Theorem 21.10, page 339). The left hand side of (4.63) is differentiable under the integral sign because for $|\theta^*\left(\lambda\right)-\theta|<|\varepsilon|$

$$\begin{aligned}
&\left|\frac{\partial}{\partial\theta_l}tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\theta^*\right)\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\theta^*\right)\right\}\right| \\
=\;& |tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\theta^*\right)\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\ddot{\mathbf{f}}_{(j,l)}\left(\lambda,\theta^*\right)\right\} \\
&+tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\theta^*\right)\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\theta^*\right)\right\} \\
&+tr\left\{\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\ddot{\mathbf{f}}_{(i,l)}\left(\lambda,\theta^*\right)\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\theta^*\right)\right\} \\
&+tr\left\{\left(\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\theta^*\right)\mathbf{f}^{-1}\left(\lambda,\theta^*\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\theta^*\right)\right\}|,
\end{aligned}$$

which by Assumption 4.2[**G**](ii) and (iii), Lemma B1.8, and compactness of the parameter space is bonded at all $\theta\in\Theta$. Then the use of the dominated convergence theorem completes the proof.

Proof of Lemma 4.8 Set

$$\tilde{h}_1(u) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_1(\lambda) e^{iu\lambda} d\lambda,$$
$$\tilde{h}_2(u) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h_2(\lambda) e^{iu\lambda} d\lambda.$$

Then

$$TCov\left\{\int_{-\pi}^{\pi} h_1(\lambda)\mathbf{I}_{ab}(\lambda)d\lambda, \int_{-\pi}^{\pi} h_2(\lambda)\mathbf{I}_{cd}(\lambda)d\lambda\right\} \tag{4.64}$$

$$= \frac{1}{T}\sum_{u_1,u_2,u_3,u_4=1}^{T} \tilde{h}_1(u_1-u_2)\tilde{h}_2(u_3-u_4)\tilde{\mathbf{\Gamma}}_{(a,c)}(u_3-u_1)\tilde{\mathbf{\Gamma}}_{(b,d)}(u_4-u_2) \tag{4.65}$$

$$+\frac{1}{T}\sum_{u_1,u_2,u_3,u_4=1}^{T} \tilde{h}_1(u_1-u_2)\tilde{h}_2(u_3-u_4)\tilde{\mathbf{\Gamma}}_{(a,d)}(u_4-u_1)\tilde{\mathbf{\Gamma}}_{(b,c)}(u_3-u_2) \tag{4.66}$$

$$+\frac{1}{T}\sum_{u_1u_2u_3u_4=1}^{T} \tilde{h}_1(u_1-u_2)\tilde{h}_2(u_3-u_4)\mathbf{K}^{\mathbf{y}}_{abcd}(u_2-u_1,u_3-u_1,u_4-u_1). \tag{4.67}$$

The convergence of (4.65) and (4.66) follows from Hannan (1976, Theorem 1, page 398). For example (4.65)

$$\left(\sum_{l=1-T}^{T+1}\left(1-\frac{|l_1|}{T}\right)\tilde{h}_1(l_1)\tilde{\Gamma}_{(b,d)}(u_4-u_2)\right)\times\left(\sum_{k=1-T}^{T+1}\left(1-\frac{|l_2|}{T}\right)\tilde{h}_2(l_2)\tilde{\Gamma}_{(a,c)}(u_3-u_1)\right),$$

which is the product of the Cesaro sums, evaluated at the origin, of the Fourier coefficients of the convolution of $h_1(\lambda)$ with $\mathbf{f}_{(b,d)}(\lambda)$ and of the convolution of $h_2(\lambda)$ with $\mathbf{f}_{(a,c)}(\lambda)$. Because $\mathbf{f}(\lambda)$ and $h(\lambda)$ are square integrable their convolution is continuous in $\lambda \in [-\pi,\pi]$. Then (4.65) converges to

$$2\pi\int_{-\pi}^{\pi} h_1(\lambda)(\lambda)\bar{\mathbf{f}}_{(b,d)}(\lambda)\bar{h}_2(\lambda)\mathbf{f}_{(a,c)}d\lambda.$$

An analogous result holds for (4.66). Set $l_1 = u_1$, $l_2 = u_2 - u_1$, $l_3 = u_3 - u_1$, $l_4 = u_4 - u_1$, (4.67) can be expressed as

$$\frac{1}{T}\sum_{l_2,l_3,l_4=1-T}^{T+1}(T - S(l_2,l_3,l_4))\tilde{h}_1(-l_2)\tilde{h}_2(l_3-l_4)K^{\mathbf{y}}_{abcd}(l_2,l_3,l_4), \tag{4.68}$$

where

$$S(l_2,l_3,l_4) = \max(|l_2|,|l_3|,|l_4|)I(sign\ l_2 = sign\ l_3 = sign\ l_4)$$
$$+\max(|l_i|,|l_j|)+|l_k|I(sign\ l_i = sign\ l_j = -sign\ l_k).$$

As $T\to\infty$, (4.68) converges to

$$\sum_{l_2,l_3,l_4=-\infty}^{+\infty}\tilde{h}_1\left(-l_2\right)\tilde{h}_2\left(l_3-l_4\right)K^{\mathbf{y}}_{abcd}\left(l_2,l_3,l_4\right) \tag{4.69}$$

$$-\lim_{T\to\infty}\frac{1}{T}\sum_{l_2,l_3,l_4=1-N}^{N+1}S\left(l_2,l_3,l_4\right)\tilde{h}_1\left(-l_2\right)\tilde{h}_2\left(l_3-l_4\right)K^{\mathbf{y}}_{abcd}\left(l_2,l_3,l_4\right). \tag{4.70}$$

However, since the functions h_1 and h_2 are square integrable in $[-\pi,\pi]$, for a certain positive constant K,

$$\begin{aligned}&\frac{1}{T}\sum_{l_2,l_3,l_4=1-T}^{T+1}S\left(l_2,l_3,l_4\right)\tilde{h}_1\left(-l_2\right)\tilde{h}_2\left(l_3-l_4\right)K^{\mathbf{y}}_{abcd}\left(l_2,l_3,l_4\right)\\ \leq\ &\frac{K^2}{N}\left|\sum_{l_2,l_3,l_4=1-T}^{T+1}\max\left(|l_2|+|l_3|+|l_4|\right)\tilde{h}_1\left(-l_2\right)K^{\mathbf{y}}_{abcd}\left(l_2,l_3,l_4\right)\right|,\end{aligned}$$

and the terms

$$\sum_{l_2,l_3,l_4=1-T}^{T+1}\frac{|l_j|}{T}\left|K^{\mathbf{y}}_{abcd}\left(l_2,l_3,l_4\right)\right|$$

for $j=1,2,3$ converge to 0 as $T\to\infty$ using Lemma B1.11. Then as $T\to\infty$, (4.67) converges to (4.69). Then, by repeated application of the Parseval equality, (4.68) converges to

$$\int_{-\pi}^{\pi}\int_{\pi}^{\pi}h_1\left(\lambda_1\right)h_2\left(-\lambda_2\right)\tilde{K}^{\mathbf{y}}_{abcd}\left(\lambda_1,\lambda_2,-\lambda_2\right).$$

Proof of Lemma 4.7 By Lemma B1.9 $\mathbf{g}^{(j)}(\lambda,\boldsymbol{\theta}_0)$ is continuous at all λ, moreover it is symmetric in $[-\pi,\pi]$. Thus, by Lemma B2.1, for any $\eta>0$, and all $a,b=1,...n$, we can always choose M large enough such that

$$\max_{a,b=1,\ldots,n}\sup_{\lambda\in\Pi}\left|\mathbf{g}^{(j)}_{M_{(a,b)}}(\lambda,\theta_0)-\mathbf{g}^{(j)}_{(a,b)}(\lambda,\boldsymbol{\theta}_0)\right|\leq\eta.$$

Consider that

$$\begin{aligned}&Var\left(\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi}tr\left\{\boldsymbol{\delta}^{(j)}_{M}(\lambda,\theta_0)\left[\mathbf{I_T}(\lambda)-\mathbf{EI}(\lambda)\right]\right\}d\lambda\right)\\ =\ &Var\left(\frac{\sqrt{T}}{2\pi}\sum_{a,b=1}^{n}\int_{-\pi}^{\pi}\left[\mathbf{I}_{(a,b)}(\lambda)-\mathbf{EI}_{(a,b)}(\lambda)\right]\boldsymbol{\delta}_{M(b,a)}(\lambda)d\lambda\right).\end{aligned}$$

As $T \to \infty$, by Lemma 4.8 this is dominated by

$$\begin{aligned}
&\frac{2}{\pi}n^2\int_{-\pi}^{\pi}\left|\boldsymbol{\delta}_{M(b,a)}(\lambda)\right|^2 \mathbf{f}_{(a,a)}\left(\lambda\right)\overline{\mathbf{f}}_{(b,b)}\left(\lambda\right)d\lambda \\
&+\frac{2}{\pi}n^2\int_{-\pi}^{\pi}\boldsymbol{\delta}_{M(b,a)}(\lambda)\overline{\boldsymbol{\delta}}_{M(a,b)}(-\lambda)\mathbf{f}_{(a,b)}\left(\lambda\right)\overline{\mathbf{f}}_{(b,a)}\left(\lambda\right)d\lambda \\
&+\frac{2}{\pi}n^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\boldsymbol{\delta}_{M(b,a)}(\lambda_1)\overline{\boldsymbol{\delta}}_{M(d,c)}(-\lambda_2)\tilde{\mathbf{K}}_{abcd}\left(\lambda_1,\lambda_2,-\lambda_2\right)d\lambda_1 d\lambda_2.
\end{aligned}$$

By compactness of the parameter space, $\boldsymbol{\delta}_{M(a,b)}(\lambda)$ is square integrable in λ,which tends to zero as $M \to \infty$, because the elements of the spectral density matrix and the trispectrum are integrable by Lemma B1.3 and Lemma B1.12.

Proof of Lemma 4.6 part (b) The jth element of $\sqrt{T}E\dot{Q}_T^{(j)}\left(\boldsymbol{\theta}_0\right)$, can be written as

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}^{(j)}(\lambda,\theta_0)\left[\mathbf{EI}(\lambda)-\mathbf{f}(\lambda,\theta_0)\right]\right\}d\lambda.$$

Note that $\mathbf{EI}\left(\lambda\right)$ is the Cesaro sum of the Fourier coefficients of $\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)$. Assumption 4.2[**G**](i) implies

$$\sup_{\lambda\in\Pi}\sum_{a,b=1}^{n}\left|\mathbf{EI}_{\mathbf{T}}^{(a,b)}\left(\lambda\right)-\mathbf{f}^{(a,b)}\left(\lambda,\boldsymbol{\theta}_0\right)\right| = O\left(T^{-\alpha}\right)$$

uniformly in $\boldsymbol{\theta}$ (see Hannan, 1970, page 513). Then

$$\begin{aligned}
\sqrt{T}E\dot{Q}_{T(j)}\left(\theta\right) &= \frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}^{(j)}(\lambda,\theta_0)\left[\mathbf{EI}\left(\lambda\right)-\mathbf{f}(\lambda,\theta_0)\right]\right\}d\lambda \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi}\max_{(a,b)}\sup_{\lambda\in\Pi}\left|\mathbf{g}_{(j)}^{(a,b)}(\lambda,\boldsymbol{\theta}_0)\right|\sum_{a,b=1}^{n}\left\{\sqrt{T}\left|\mathbf{EI}_{(a,b)}\left(\lambda\right)-\mathbf{f}_{(a,b)}\left(\lambda,\boldsymbol{\theta}_0\right)\right|\right\}d\lambda \\
&= O\left(T^{1/2-\alpha}\right),
\end{aligned}$$

which converges to zero as $T \to \infty$, since by Assumption 4.2[**G**](i), $\alpha > 1/2$.

Chapter 5 Multivariate exponential volatility models with long memory

5.1 Long-range dependence in conditional volatility models

A large body of research suggests that the conditional volatility of asset prices displays long-range persistence. Evidence of some type of persistence in financial data can be traced back to the unit root findings in estimated GARCH and EGARCH models, whose applications to high frequency data have indicated the presence of an approximate unit root in the volatility. Long-term dependencies have been found in the returns of a variety of assets classes. Helms, Kaen, and Rosenman (1984), Kao and Ma (1992), Eldridge, Bernhardt, and Mulvey (1993), Fang, Lai, and Lai (1994), Corazza, Malliaris, and Nardelli (1997) found long-term dependence in index and commodity futures returns. Greene and Fielitz (1977), Lo (1991), and Nawrocki (1995) examined long memory regularities in U.S. equity market returns. Jacobsen (1996), Cheung, Lai, and Lain (1993) found evidences of long-term dependence in European and Asian equity markets. These findings suggests that financial data display persistent features that can't be captured by standard GARCH or Stochastic Volatility models which impart short memory autocorrelation in the squares. Standard volatility models may capture such persistence only via approximation of a unit root. Out of the necessity to develop statistical methods to reproduce these presidencies in financial data, a number of authors have introduced long memory univariate volatility models. With respect to ARCH type models, various alternative have been suggested. Robinson (1991) introduced the ARCH(∞) model, a possibly long memory generalized ARCH model. Baillie et al. (1996) considered a particular case of the ARCH(∞) denominated fractionally integrated GARCH (FIGARCH) model. Arguing that the knife-edge distinction between I(0) and I(1) processes can be far too restrictive in describing persistence in conditional volatilities, Baillie et al. model the volatility as a discrete time fractionally integrated process I(d) introduced by Adenstedt (1974), Granger (1980, 1981), and Grangerand and Joyeux (1980). The propagation of shocks to the variance occurs at a slow hyperbolic decay rate when $d \in (0,1)$, as opposed to the extremes of exponential decay associated with the stationary GARCH process or the infinite persistence resulting from an IGARCH model. The FIGARCH model combines many of the features of the fractionally integrated process for the mean together with the regular GARCH process for the conditional variance. In particular, it implies a slow hyperbolic rate of decay for the lagged squared innovations in the conditional variance function, although the cumulative impulse response weights associated with the influence of a volatility shock on the optimal forecasts of the future conditional variance eventually tend to zero, a property that the model shares with weakly stationary GARCH processes. Baillie et al. present estimation results for a regular GARCH process when the true data generating process is FIGARCH. It turns out that the estimated autoregressive parameters in the misspecified GARCH models are very close to unity, indicative of IGARCH type behavior. These findings support the view that the apparent widespread IGARCH property, so frequently reported with high-frequency asset pricing data, might as well be spurious, and that the IGARCH process offers a poor diagnostic to assess the presence of long memory in conditional volatility. Recently a number of authors have pointed out some drawbacks of the FIGARCH model. For example, Davidson (2004) has shown that the in the model the persistence of shocks to volatility decreases as the long memory parameter increases. Zaffaroni (2004) has shown that the FIGARCH model cannot generate autocorrelations of squares with long memory. Consequently the Fractionally Integrated EGARCH (FIEGARCH) model of Bollerslev and Mikklsen (1996), which extends the asymmetric EGARCH model of Nelson to long memory, has been by far more popular. Robinson and Zaffaroni (1997, 1998) and Zaffaroni (2003) introduced an alternative way of modelling strong dependence in volatility by means of a family of nonlinear moving average models, that we discussed in the previous chapter. Zaffaroni (2003) established the asymptotic properties of the Gaussian estimator of the nonlinear moving average model and provided a formal framework to asses whether long memory volatility models represent a valid alternative to short memory ones.

With respect to Stochastic Volatility models, Harvey (1998) and Breidt et al. (1998) independently proposed a Long Memory Stochastic Volatility (LMSV) model in which the underlying log volatility evolves as an ARFIMA (p, d, q) process,

$$(1-B)^d \phi(B) v_t = \theta(B) \eta_t, \quad \eta_t \sim i.i.d\left(0, \sigma_\eta^2\right),$$

with $d \in (-0.5, 0.5)$. In the overall the LMSV model is very tractable and easily fit to data. Breitd et al. carried several non parametric and semiparametric tests for long memory over various market indexes of daily returns finding highly significant results. Ruiz and Vega (2006) extended the Long Memory Stochastic Volatility model to include asymmetries, allowing for correlation between the shocks of the level and volatility equation.

In the past decade the research on multivariate volatility models has produced a wide variety of specifications, however these do not yet include long range dependence in the conditional covariance matrix. Such generalization are not easily implemented in ARCH(∞) type of models. As shown for the univariate case by Giraitis at al. (2000) and Zaffaroni (2004), such models might not be adequate to capture long memory in the squares of the process. In ARCH(∞) models the necessary and sufficient condition for covariance stationarity of the levels rules out long memory in the squares of the process when the model contains a positive intercept. This implies for example that the introduction of long-memory dependence in VEC-BEKK type of models might not be straightforward. The Multivariate Exponential Volatility model defined in (4.6)-(4.10) allows for long-range dependence in the conditional covariance matrix. Since it is nested in the class of vector nonlinear moving average models, it allows for strong dependence in the squares of the process while retaining the strict stationarity and martingale difference assumption in the levels. In this chapter we investigate the asymptotic properties of the Gaussian estimator of the parameters when the sample autocorrelations of the squared returns decline very slowly. In contrast to the fast exponential decay imposed on the signal parameters in the last chapter, we allow for an hyperbolic decay rate consistent with non-summability of the autocovariances and with an unbounded spectral density at the zero frequency.

5.2 Long-range dependence in the MEV model

The vector signal plus noise representation of the Multivariate Exponential Volatility Model may display long-range persistence in the signal when its coefficients exhibit a sufficiently slow rate of decay. Indeed in the transformation

$$\mathbf{y}_t = \sum_{j=0}^{\infty} \mathbf{\Psi}_{0j} \boldsymbol{\epsilon}_{t-j-1} + \boldsymbol{\xi}_t, \quad \sum_{j=0}^{\infty} \left\| \mathbf{\Psi}_{0j} \right\|^2 < \infty, \tag{5.1}$$

the assumption of square summability allows to impart on the coefficients an hyperbolic decay rate of $D\left(j;\theta\right) j^{d(\theta)-1}$ as $j \to \infty$, where $D\left(j;\theta\right)$ is a measurable Zygmund slowly varying function at infinity and $d\left(\boldsymbol{\theta}\right) \in (-\infty, 1/2)$ is the memory parameter. When $d(\boldsymbol{\theta}) \in (0, 1/2)$ the series has long memory, when $d(\boldsymbol{\theta}) \in (-1/2, 0)$ the series has negative memory and when $d(\boldsymbol{\theta}) = 0$ the series has short memory. In this chapter we focus on signal plus noise processes with long-range dependence, setting $d(\boldsymbol{\theta}) \in (0, 1/2)$. We assume that the memory parameter is constant across the different series in $\mathbf{y}_t$, however do not consider fractional cointegration among them. The dependence structure implied by this assumption entails that the spectral density of the process might not be square integrable and might not be bounded at the zero frequency. Indeed as $\lambda \to 0^+$ the elements of the spectral density matrix behave as $KD\left(\lambda^{-1};\theta\right)\lambda^{-2d(\boldsymbol{\theta})}$. To establish the asymptotic properties of the frequency domain Gaussian estimates we can no longer rely on the uniform continuity of the spectrum and its higher order derivatives at all frequencies. However since $d\left(\boldsymbol{\theta}\right)$ is positive, the inverse of the spectral density matrix is continuous at all frequencies (see Giraitis et al., 2012, page 212), hence the strong consistency of the estimator still holds. The most relevant technical difference from Chapter 4 pertains to the asymptotic normality of the score vector. In this chapter we cannot rely on the asymptotic normality of the sample serial covariances and we introduce a different approximation that has the effect of annulling the singularity of the spectral density at the zero frequency. It must be pointed out that our results are limited to the case where the spectral density matrix has a singularity at the zero frequency and cannot be readily extended to long-memory multiple time-series possessing a variety of singularities in their spectrum. This last case is dealt with in Hosoya (1997) by means of the bracketing function approach, however Hosoya's results do not extend to the case of correlated signal and noise. The next section discusses the Whittle estimator for linear long-memory processes. Section 5.4 introduces spectral domain regularity conditions and discusses the consistency of the estimator. Section 5.5 reinforces the assumptions and derives the asymptotic normality of the estimates. All the proofs of the technical lemmas are in Appendix C.

5.3 The Whittle estimator

In the previous chapter, we discussed some advantages in the use of the Whittle estimator due to his frequency domain specification. In this chapter, as we consider a richer dependence structure in the conditional variance, a few more advantages emerge. Indeed the Gaussian frequency domain estimator has been widely used in estimation of long-range dependent models thanks to its technical properties. The Whittle function naturally takes into account the asymptotic behavior of the autocovariances as the sample size goes to infinity, so it is very sensitive to the degree of dependence of the process in second-order sense. Moreover, by construction, it automatically compensates for the possible lack of square integrability of the model spectral density that occurs when the memory parameter is between $1/2$ and $1/4$. This implies that the estimator has a rate of convergence and an asymptotic distribution that do not depend on whether long-memory holds or not. The Whittle function does not require truncation of the process for estimation, as is typically the case with MLE and PMLE where one needs to distinguish between the observable likelihood and the unobservable one. As shown in Robinson and Zaffaroni (2006), such truncation might not turn out asymptotically negligible in long-memory models since it could induce asymptotic bias.

A number of authors have developed the asymptotic properties of the Whittle estimator in long-range dependent stationary processes. Yajima (1985), Fox and Taqqu (1987) and Dahlhaus (1989) investigated univariate Gaussian processes. Giraitis and Surgailis (1990) dealt with univariate non Gaussian linear processes establishing the asymptotic normality of the Whittle estimates of the parameters of strongly dependent linear sequences. Heyde and Gay (1993) partially extended the former result to vector-valued non Gaussian processes. Giraitis and Taqqu (1999) generalized Giraitis and Surgailis (1990) to a wide class of vector linear processes satisfying minimal assumptions. Hosoya (1997) dealt with vector long-memory stationary processes with singularities not necessarily limited at the zero frequency. The key results in this literature are the central limit theorems for quadratic forms of the type

$$\sqrt{T}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}\left(\lambda,\boldsymbol{\theta}\right)\left[\mathbf{I}_{T}\left(\lambda\right)-\mathbf{E}\mathbf{I}_{T}\left(\lambda\right)\right]\right\}d\lambda \tag{5.2}$$

$$\frac{1}{\sqrt{T}}tr\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{h}_{j}\left(\left|t-s\right|\right)\left(\mathbf{y}_{t}\mathbf{y}_{s}^{'}-\Gamma\left(\left|t-s\right|\right)\right)\right\} \tag{5.3}$$

which play a crucial role in the asymptotic behavior of the Whittle estimator. The asymptotic normality of the integrated weighted periodogram in (5.2) is derived via approximation by another quadratic form which shares the same asymptotic distribution but has shorter memory. The main idea of the approximation is to impose conditions on the weight function $\mathbf{h}_{j}\left(\left|t-s\right|\right) = \int_{-\pi}^{\pi}\mathbf{g}_{(j)}\left(\lambda,\boldsymbol{\theta}\right)e^{i\lambda|t-s|}d\lambda$ that have the effect of annihilating the singularities of the spectral density in the frequency domain. To establish the validity of the approximation Fox and Taqqu (1987) rely on Gaussianity in an essential way, employing the exact expression for the cumulants of a quadratic form in Gaussian variates. Giraitis and Surgailis (1990), Giraitis and Taqqu (1999) and Hosoya (1997) relax the Guassianity assumption, and exploit the factorization of the spectrum. Hosoya establishes a limit theorem for the integrated weighted periodogram in a framework general enough to deal with a variety of multiple singularities of the spectral density, under a set of mixing conditions on the innovations which need not to satisfy a martingale difference assumption. He considers a full-rank vector linear process with spectral density

$$\mathbf{f}(\lambda) \equiv \frac{1}{2\pi}\varphi\left(e^{i\lambda},\boldsymbol{\theta}\right)K^{e}\varphi\left(e^{i\lambda},\boldsymbol{\theta}\right)^{*}$$

factored as

$$\mathbf{f}(\lambda) = \Gamma\left(e^{-\mathbf{i}\boldsymbol{\lambda}}\right)\Gamma\left(e^{-\mathbf{i}\boldsymbol{\lambda}}\right)^{*}$$

and defines the approximation to (5.2) by

$$\sqrt{T}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}'_{(j)}(\lambda,\boldsymbol{\theta})\left[\mathbf{I}'_{T}(\lambda) - \mathbf{E}\mathbf{I}'_{T}(\lambda)\right]\right\}d\lambda, \tag{5.4}$$

where $\mathbf{g}'_{(j)}(\lambda,\theta)$ and $\mathbf{I}'_{T}(\lambda,)$ are functions of a new process $\mathbf{y}'_{t}$ which by definition has spectral density

$$\mathbf{f}'(\lambda) \equiv \Gamma\left(e^{-\mathbf{i}\boldsymbol{\lambda}}\right)\mathbf{f}(\omega)\Gamma\left(e^{-\mathbf{i}\boldsymbol{\lambda}}\right)^{*}.$$

In view of the construction $\mathbf{f}'(\omega)$ is equal almost everywhere to the identity matrix and square integrable at all frequencies. The CLT for (5.4) follows using standard results. Hosoya derives the validity of the approximation by means of a general result on the convergence of covariances of quadratic forms

$$Tcov\left(\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta})\left[\mathbf{I}_{T}(\lambda) - \mathbf{E}\mathbf{I}_{T}(\lambda)\right]\right\}d\lambda, \int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(l)}(\lambda,\boldsymbol{\theta})\left[\mathbf{I}_{T}(\lambda) - \mathbf{E}\mathbf{I}_{T}(\lambda)\right]\right\}d\lambda\right). \tag{5.5}$$

To establish the convergence of (5.5) he imposes regularity conditions on the functions $\mathbf{g}_{(j)}$ and $\mathbf{g}_{(l)}$ and on the transfer function $\varphi\left(e^{i\lambda},\boldsymbol{\theta}\right)$, assuming uniform Lipschitz continuity of degree $\gamma > 0$ for the former and integrability of order $p > 2$ for the latter. Moreover he imposes that for some $\gamma_1 > 0$, the pair $\left\{\mathbf{g}_{(j)}(\lambda), \varphi\left(e^{i\lambda},\boldsymbol{\theta}\right)\right\}$ satisfies

$$\sup_{|\varepsilon|\leq\varepsilon_1}\left\|\mathbf{g}_{(j)}(\lambda)\left|\varphi\left(e^{i\lambda+\varepsilon},\boldsymbol{\theta}\right) - \varphi\left(e^{i\lambda},\boldsymbol{\theta}\right)\right|^{2}\right\|_{2} = O\left(|\varepsilon_1|^{\gamma_1}\right), \tag{5.6}$$

where "$\|.\|_2$" denotes the L^2-norm of a complex function in $[-\pi,\pi]$. Using the concept of multiple

Fejer Kernel he shows that (5.5) converges to

$$4\pi\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}\left(\lambda\right)\mathbf{f}\left(\lambda,\boldsymbol{\theta}\right)\mathbf{g}_{(l)}\left(\lambda\right)\mathbf{f}\left(\lambda,\boldsymbol{\theta}\right)\right\}d\lambda \tag{5.7}$$
$$+2\pi\sum_{a,b,c,d=1}^{s}\mathbf{g}_{(j)}\left(\lambda_1\right)\mathbf{g}_{(l)}\left(\lambda_2\right)K_{abcd}^{\mathbf{y}}\left(-\lambda_1,\lambda_2,-\lambda_2\right)d\lambda_1 d\lambda_2,$$

and relying on (5.7) he establishes that the variance of the difference between (5.4) and (5.2) converges to zero as T goes to infinity.

Giraits and Taqqu (1999) extend Giraitis and Surgailis (1990), Giraitis et al. (1996) and Giraitis and Taqqu (1997). They consider situations where there are $k\geq 1$ different scalar quadratic forms (5.2), each with its own weight sequence and possibly generated by a different process. For a linear process with spectral density $\mathbf{f}\left(\omega\right)=2\pi\sigma^2\left|\hat{a}\left(\omega\right)\right|^2$, they approximate (5.2) by

$$S_T=\sum_{k=1}^{T}Y_{1,k}Y_{2,k},$$

where the processes $Y_{i,k}$ have by definition weights $b_{i,t}\left(\lambda\right)$ such that

$$b_{i,t}\left(\lambda\right)=\int_{-\pi}^{\pi}\left|\mathbf{g}\left(\omega\right)\right|^{1/2}\left|\hat{a}\left(\omega\right)\right|e^{ik\lambda}d\lambda, \tag{5.8}$$

and square integrable spectral density $\mathbf{f}_i\left(\omega\right)=2\pi\mathbf{f}\left(\omega\right)\left|\mathbf{g}\left(\lambda\right)\right|$, where the zero frequency singularity of $\mathbf{f}\left(\omega\right)$ is compensated for by letting $\left|\mathbf{g}\left(\omega\right)\right|\to 0$ as $\omega\to\infty$. As in Hosoya (1997), the approximation depends on the form of the transfer function (5.8) which relies on the factorization of the spectral density of the original process. Zaffaroni (2003, 2009) establishes a CLT for univariate quadratic forms (5.2) without making use of the factorization of spectral density. The original process is truncated at some finite $t=N$ and the validity of the approximation is established relying on certain results on the asymptotic behavior of the trace of Toeplitz matrices (Theorem 1, Fox and Taqqu, 1987). The CLT for the quadratic forms of the new process follows from standard results on quadratic forms in N dependent variates. In Section 5 we extend Zaffaroni (2003, 2009) introducing an approximation for vector signal plus noise processes. We then rely on Giraitis and Taqqu (1999, Theorem 7.3) to establish the asymptotic normality of the approximation.

5.4 Consistency

This section establishes the consistency of the Whittle estimator. We first list the assumptions with some comparison with Assumption 4.1, then state the main result. Even if the signal component of $\mathbf{y}_t$ is a long-memory process, we still obtain a strong consistency result. The notation is unchanged from Chapter 4. In what follows, $D\left(\lambda^{-1}, \theta\right)$ denotes a Zygmund slowly varying function at infinity (ZSV), not necessarily the same.

ASSUMPTION 5.1

[**A**] $\{\boldsymbol{\epsilon}_t', \boldsymbol{\xi}_t'\}$ are $i.i.d$ unobservable random vectors, and, for every $a, b = 1, ..., n$,

(i) $E\boldsymbol{\epsilon}_0 = 0$ and $E(\boldsymbol{\epsilon}_0\boldsymbol{\epsilon}_0') = \Sigma_\varepsilon\left(\boldsymbol{\tau}\right), 0 < \Sigma_\varepsilon\left(\boldsymbol{\tau}\right) < \infty$.

(ii) $E\left|\boldsymbol{\xi}_0^{(a)}\right| < \infty$ and $E(\boldsymbol{\xi}_0\boldsymbol{\xi}_0') = \Sigma_\xi\left(\boldsymbol{\tau}\right), 0 < \Sigma_\xi\left(\boldsymbol{\tau}\right) < \infty$.

(iii) $E\left(\boldsymbol{\xi}_0\boldsymbol{\epsilon}_0'\right) = \Sigma_{\xi\epsilon}\left(\boldsymbol{\tau}\right), 0 < \Sigma_{\xi\epsilon}\left(\boldsymbol{\tau}\right) < \infty$.

[**B**] $\boldsymbol{\theta}_0$ is an interior point of the compact parameter space $\Theta \in R^s$.

[**C**] (i) $\mathbf{f}\left(\lambda, \boldsymbol{\theta}\right)$ has elements in $L_1\left(\Pi\right)$, continuous at all $\left(\lambda, \boldsymbol{\theta}\right) \in \Pi \times \Theta$ with $\lambda \neq 0$,

$$\left|\mathbf{f}^{(a,b)}\left(\lambda, \boldsymbol{\theta}\right)\right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right)|\lambda|^{-2d(\theta)} \quad \lambda \to 0^+, \quad d\left(\boldsymbol{\theta}\right) \in (0, 1/2).$$

(ii) $\mathbf{f}^{-1}\left(\lambda, \boldsymbol{\theta}\right)$ has elements in $L_1\left(\Pi\right)$, continuous at all $\left(\lambda, \boldsymbol{\theta}\right) \in \Pi \times \Theta$,

$$\left|\mathbf{f}^{-1}_{(a,b)}\left(\lambda, \boldsymbol{\theta}\right)\right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right)|\lambda|^{2d(\theta)} \quad \lambda \to 0^+, \quad d\left(\boldsymbol{\theta}\right) \in (0, 1/2).$$

(iii) The function

$$\boldsymbol{\varphi}_\eta\left(\lambda, \boldsymbol{\theta}\right) \equiv \frac{\mathbf{f}\left(\lambda, \boldsymbol{\theta}\right)}{\left(\det \mathbf{f}\left(\lambda, \boldsymbol{\theta}\right) + \eta\right)}$$

has elements in L_1, continuous at all $\left(\lambda, \boldsymbol{\theta}\right) \in \Pi \times \Theta$ for all $\eta > 0$.

[**D**] For every $\boldsymbol{\theta} \in \Theta$ whenever $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\mathbf{f}\left(\lambda, \boldsymbol{\theta}\right) \neq \mathbf{f}\left(\lambda, \boldsymbol{\theta}_0\right)$.

[**E**] For any $\boldsymbol{\theta} \in \Theta$, $\mathbf{f}\left(\lambda, \boldsymbol{\theta}\right)$ is a strictly positive definite matrix.

[**F**] $\int_{-\pi}^{\pi} \log\det \mathbf{f}\left(\lambda, \boldsymbol{\theta}\right) d\lambda$ is twice differentiable in $\boldsymbol{\theta} \in \Theta$ under the sign of integral.

[**G**] $\left(\partial/\partial\theta_j\right)\mathbf{f}\left(\lambda, \boldsymbol{\theta}\right)$ has elements in $L_1\left(\Pi\right)$, continuous at all $\left(\lambda, \theta\right) \in \Pi \times \Theta$ with $\lambda \neq 0$,

$$\left|\dot{\mathbf{f}}^{(a,b)}_{(j)}\left(\lambda, \boldsymbol{\theta}\right)\right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right)|\lambda|^{-2d(\theta)} \quad \lambda \to 0^+ \quad d\left(\boldsymbol{\theta}\right) \in (0, 1/2).$$

In contrast with Assumption 4.1, Assumption 5.1 imposes regularity conditions directly on the spectral density and its first order derivatives, defining unambiguously their behavior near the origin, as well as a form of uniform continuity away from the zero frequency. Spectral domain regularity conditions are common in long-memory parametric literature, see for example Giraitis and Surgailis (1990), Heyde and Gay (1993), and Hosoya (1997). A relevant exception is Zaffaroni (2009). In a scalar signal plus noise model (5.1) Zaffaroni imposes regularity conditions directly on the signal coefficients and their derivatives, assuming that they satisfy an exact hyperbolic decay rate, quasi monotonic convergence towards zero, and a pure-bounded variation condition. These conditions imply that the spectral density and its derivatives satisfy a form of Lipschitz continuity of degree $\alpha \geq \min\left[1, 1-2d\left(\zeta\right)\right]$ away from the zero frequency and have an exact decay rate of $L(j,\zeta)\left|\lambda\right|^{-2d(\zeta)}$ as $\lambda \to 0^{+}$, where $L(j,\zeta)$ is a slowly varying function at infinity. We impose slightly stronger regularity conditions directly on the spectral density matrix and its derivatives at the zero frequency. The function $D(\lambda)$ is a Zygmund slowly varying function at infinity, i.e. a slowly varying function at infinity in Karamata's sense such that for any $a > 0$ and for some $\lambda_0 > 0$, $\lambda^{a}D(\lambda)$ is increasing in λ and $\lambda^{-a}D(\lambda)$ is decreasing in λ , for all $\lambda \geq \lambda_0$. In this chapter we use repetitively the result that a Zygmund slowly varying function at infinity (ZSV) is $O\left(\left|\lambda\right|^{\delta}\right)$ as $\lambda \to 0$ for any $\delta > 0$ (see Feller 1971, page 277).

The main result of this section is the strong consistency of the Whittle estimator.

Theorem 5.1 *Under Assumption 5.1, as* $T \to \infty$

$$\hat{\boldsymbol{\theta}}_T \to_{a.s} \boldsymbol{\theta}_0.$$

The proof of Theorem 5.1 follows by contradiction, exactly as for Theorem 4.1, relying on the uniform convergence of the objective function to $Q\left(\boldsymbol{\theta}\right)$ and on Lemma 4.2. Lemma 4.2. holds under Assumption 5.1 as well because its proof is based on the strict positivity and the identification property of the spectral density which are directly assumed in this chapter. The uniform converge of $Q_T(\boldsymbol{\theta})$ and $Q_{T,\eta}(\boldsymbol{\theta})$ requires a slightly different proof from Chapter 4 and it is established by the following lemma.

Lemma 5.1 *If Assumption 5.1 holds, then*

(a) $\lim_{T\to\infty} Q_T(\boldsymbol{\theta}) = Q(\theta)$ almost surely uniformly in $\boldsymbol{\theta} \in \Theta$.

(b) for any $\eta > 0$, uniformly in $\theta \in \Theta$, $\lim_{T\to\infty} Q_{T,\eta}(\boldsymbol{\theta}) = Q_\eta(\boldsymbol{\theta})$ almost surely.

The almost sure uniform convergence of the first term of $Q_T(\boldsymbol{\theta})$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda_t)\right\}$$

to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right\}d\lambda$$

follows from Lemma 4.4, taking $\mathbf{h}(\lambda,\boldsymbol{\theta}) \equiv \mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})$. By Assumption 5.2[**C**](ii) $\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})$ is continuous at all $(\lambda,\boldsymbol{\theta})$, even if $\mathbf{f}(\lambda,\boldsymbol{\theta})$ is unbounded at the zero frequency. Moreover $\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}) = \mathbf{f}^{-1}(-\lambda,\boldsymbol{\theta})$. The uniform continuity of first non stochastic term of $Q_T(\theta)$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det\mathbf{f}(\lambda,\boldsymbol{\theta})d\lambda,$$

is shown by establishing its equicontinuity property

$$\lim_{\varepsilon\to 0}\sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|\le\epsilon}\left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}\left(\lambda,\tilde{\boldsymbol{\theta}}\right)d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\boldsymbol{\theta})\,d\lambda\right| \to 0. \tag{5.9}$$

(5.9) is implied by

$$\sup_{\theta^*\in\Theta}\left|\frac{\partial}{\partial\theta_j}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\boldsymbol{\theta}^*)\,d\lambda\right| < \infty, \tag{5.10}$$

where $|\theta^*(\lambda) - \theta| < |\varepsilon|$, (see Davidson, 1994, Theorem 21.10, page 339). By Assumption 5.1[**F**]

$$\begin{aligned}
&\frac{\partial}{\partial\theta_j}\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\det\mathbf{f}(\lambda,\boldsymbol{\theta}^*)\,d\lambda \\
=\;& \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta_j}\left(\log\det\mathbf{f}(\lambda,\boldsymbol{\theta}^*)\right)d\lambda \\
=\;& \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}^*)\,\dot{\mathbf{f}}_{(j)}(\lambda,\boldsymbol{\theta}^*)\right\}d\lambda.
\end{aligned}$$

However by Assumption 5.1[**C**](ii) and [**G**] the integrand is at most, ignoring constant terms,

$$\int_{-\pi}^{\pi}|\lambda|^{2(d_l-d_u)-\delta}\,d\lambda < \infty,$$

where we take $d_l = \inf_\Theta d(\boldsymbol{\theta})$ and $d_u = \sup_\Theta d(\boldsymbol{\theta})$ and so $(d_l - d_u) > -1/2$ and δ can be taken arbitrarily small. Then we choose δ such that

$$\int_{-\pi}^{\pi} |\lambda|^{2(d_l - d_u) - \delta} \, d\lambda < \infty,$$

and the use of the dominated convergence theorem concludes.

The proof of Lemma 5.1 (b) follows for $\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\boldsymbol{\varphi}_\eta(\lambda,\boldsymbol{\theta})\mathbf{I}_T(\lambda_t)\right)$ from Lemma 4.4. using assumption 5.1[G](ii) and for $\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \mathbf{f}(\lambda,\boldsymbol{\theta})d\lambda$ from Lemma 5.1 (a).

5.5 Asymptotic Normality

This section reinforces the moments condition of the process innovations and the regularity conditions on the spectral density matrix, controlling the behavior of its higher order derivatives around the pole $\lambda = 0$. The following conditions are very similar to those in Fox and Taqqu (1986) and Giraitis and Taqqu (1999). However, following Zaffaroni (2009), we also impose an exact rate of decay on certain combinations of the signal coefficients that guarantees their quasi monotonic convergence to zero.

ASSUMPTION 5.2

[**A**] $\{\boldsymbol{\epsilon}_t', \boldsymbol{\xi}_t'\}$ are $i.i.d$ unobservable random vectors, and, for all $a, b, c, d = 1, ..., n,$

(i) $E\boldsymbol{\epsilon}_0 = 0$ and $E\left(\boldsymbol{\epsilon}_0^{(a)}\boldsymbol{\epsilon}_0^{(b)}\boldsymbol{\epsilon}_0^{(c)}\boldsymbol{\epsilon}_0^{(d)}\right) = K_{abcd}^{\boldsymbol{\epsilon}}(\boldsymbol{\tau})$, $|K_{abcd}^{\boldsymbol{\epsilon}}(\boldsymbol{\tau})| < \infty.$

(ii) $E\left|\boldsymbol{\xi}_0^{(a)}\right| < \infty$ and $E\left(\boldsymbol{\xi}_0^{(a)}\boldsymbol{\xi}_0^{(b)}\boldsymbol{\xi}_0^{(c)}\boldsymbol{\xi}_0^{(d)}\right) = K_{abcd}^{\boldsymbol{\xi}}(\boldsymbol{\tau}), |K_{abcd}^{\boldsymbol{\epsilon}}(\boldsymbol{\tau})| < \infty.$

(iii) $E\left(\boldsymbol{\xi}_0^{(a)}\boldsymbol{\epsilon}_0^{(b)}\boldsymbol{\xi}_0^{(c)}\boldsymbol{\epsilon}_0^{(d)}\right) = K_{abcd}^{\boldsymbol{\epsilon\xi}}(\boldsymbol{\tau}), \left|K_{abcd}^{\boldsymbol{\epsilon\xi}}(\boldsymbol{\tau})\right| < \infty.$

(iv) For any $0 \le \eta < 1$, and all $u \ge 0$

$$\Delta(u) \equiv \left\{\sum_{j=0}^{\infty}\left|\Psi_j^{(a)}\right|^{1-\eta/2}\left|\Psi_{j+|u|}^{(b)}\right|^{1-\eta/2}\right\} \sim K\left|D(u,\boldsymbol{\theta})\right| u^{(1-\eta/2)(2d-1)} \quad \text{as} \quad u \to \infty$$

(v) For any $\eta > 0$, $\left|\Psi_N^{(a)}\right|^{\eta}$ is a positive sequence converging to zero as $N \to \infty$.

[**B**] $\boldsymbol{\theta}_0$ is an interior point of the compact parameter space $\Theta \in \mathbf{R}^s$.

[**C**] (i) $\mathbf{f}(\lambda,\boldsymbol{\theta})$ has elements in $L_1(\Pi)$, continuous at all $(\lambda,\boldsymbol{\theta}) \in \Pi \times \Theta$ with $\lambda \neq 0$,

$$\left|\mathbf{f}^{(a,b)}(\lambda,\theta)\right| \le D\left(\lambda^{-1},\boldsymbol{\theta}\right)|\lambda|^{-2d(\theta)} \quad \lambda \to 0^+, \quad d(\boldsymbol{\theta}) \in (0,1/2).$$

(ii) $\mathbf{f}^{-1}(\lambda, \boldsymbol{\theta})$ has elements in $L_1(\Pi)$, continuous at all $(\lambda, \boldsymbol{\theta}) \in \Pi \times \Theta$,

$$\left| \mathbf{f}^{-1}_{(a,b)}(\lambda, \theta) \right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right) |\lambda|^{2d(\theta)} \quad \lambda \to 0^+, \quad d(\boldsymbol{\theta}) \in (0, 1/2).$$

[**D**] For every $\boldsymbol{\theta} \in \Theta$ whenever $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $\mathbf{f}(\lambda, \boldsymbol{\theta}) \neq \mathbf{f}(\lambda, \boldsymbol{\theta}_0)$.

[**E**] For any $\boldsymbol{\theta} \in \Theta$, $\mathbf{f}(\lambda, \boldsymbol{\theta})$ is a strictly positive definite matrix.

[**F**] $\int_{-\pi}^{\pi} \log \det \mathbf{f}(\lambda, \theta)\, d\lambda$ is twice differentiable in $\theta \in \Theta$ under the integral sign.

[**G**] (i) $(\partial/\partial\theta_j)\, \mathbf{f}(\lambda, \theta)$ has elements in $L_1(\Pi)$ continuous at all (λ, θ), $\lambda \neq 0$, and

$$\left| \dot{\mathbf{f}}^{(a,b)}_{(j)}(\lambda, \theta) \right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right) |\lambda|^{-2d(\theta)} \quad \lambda \to 0^+ \quad d(\boldsymbol{\theta}) \in (0, 1/2).$$

(ii) $\left(\partial^2/\partial\theta_i\partial\theta'_j\right) \mathbf{f}(\lambda, \theta)$ has elements in $L_1(\Pi)$ continuous at all (λ, θ), $\lambda \neq 0$, and

$$\left| \ddot{\mathbf{f}}^{(a,b)}_{(i,j)}(\lambda, \theta) \right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right) |\lambda|^{-2d(\theta)} \quad \lambda \to 0^+ \quad d(\boldsymbol{\theta}) \in (0, 1/2).$$

(iii) $\left(\partial^3/\partial\theta_i\partial\theta'_j\partial\theta_l\right) \mathbf{f}(\lambda, \theta)$ has elements in $L_1(\Pi)$ continuous at all (λ, θ), $\lambda \neq 0$, and

$$\left| \dddot{\mathbf{f}}^{(a,b)}_{(i,j,l)}(\lambda, \theta) \right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right) |\lambda|^{-2d(\theta)} \quad \lambda \to 0^+ \quad d(\boldsymbol{\theta}) \in (0, 1/2).$$

(iv) $(\partial/\partial\theta_j)\, \mathbf{f}^{-1}(\lambda, \theta)$ has elements in $L_1(\Pi)$ continuous at all $(\lambda, \theta) \in \Pi \times \Theta$, and

$$\left| \frac{\partial}{\partial\theta_j} \mathbf{f}^{-1}_{(a,b)}(\lambda, \theta) \right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right) |\lambda|^{2d(\theta)} \quad \lambda \to 0^+ \quad d(\boldsymbol{\theta}) \in (0, 1/2).$$

(v) $\left(\partial^2/\partial\theta_i\partial\theta'_j\right) \mathbf{f}^{-1}(\lambda, \theta)$ has elements in $L_1(\Pi)$ continuous at all $(\lambda, \theta) \in \Pi \times \Theta$ and

$$\left| \frac{\partial^2}{\partial\theta_i\partial\theta'_j} \mathbf{f}^{-1}(\lambda, \theta) \right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right) |\lambda|^{2d(\theta)} \quad \lambda \to 0^+ \quad d(\boldsymbol{\theta}) \in (0, 1/2).$$

[**H**] (i) The function

$$\mathbf{g}_{(j)}(\lambda, \boldsymbol{\theta}_0) \equiv \mathbf{f}^{-2}(\lambda, \boldsymbol{\theta}_0)\, \dot{\mathbf{f}}_{(j)}(\lambda, \boldsymbol{\theta}_0) \quad \text{for all } j = 1, \ldots s.$$

has elements in $L_1(\Pi)$ continuous at all (λ, θ), $\lambda \neq 0$, and

$$\left| \mathbf{g}^{(b,a)}_{(j)}(\lambda, \boldsymbol{\theta}_0) \right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right) |\lambda|^{2d(\theta)} \quad \lambda \to 0^+ \quad d(\boldsymbol{\theta}) \in (0, 1/2) \quad \text{for all } j = 1, \ldots s.$$

(ii) For some $1/2 < \gamma < 1$, for any $\lambda_1$and $\lambda_2 \in \Pi$

$$\left| tr\left\{ \mathbf{g}_{(j)}(\lambda_1, \boldsymbol{\theta}_0)\, \mathbf{f}(\lambda_1, \boldsymbol{\theta}_0) - \mathbf{f}(\lambda_1 - \lambda_2, \boldsymbol{\theta}_0) \right\} \right| \leq K |\lambda_2|^{\gamma} \quad \text{for all } j = 1, \ldots s.$$

Assumption 5.2[**C**] and [**G**] are quite standard in long-memory statistical literature. We assume regularity conditions on the spectral density and its derivatives up to the third order to ensure the necessary degree of smoothness on the Hessian matrix. Rather than assuming an exact decay rate at the zero frequency, we prefer slightly stronger and more general assumptions that impose a common bound on the elements of the matrices. We assume uniform continuity away from the zero frequency rather than Lipschitz continuity as Zaffaroni (2009). Assumption 5.2[**A**](i)-(iii) imply strict stationarity and ergodicity of the process, and the existence of its fourth order spectral density. Heyde and Gay (1993) make assumptions which ensure that the fourth-order cumulant term vanishes in the asymptotic covariance matrix, however in model (5.1) this is not possible. Hosoya (1997) imposes uniform Lipschitz continuity of degree $\gamma > 0$ on the fourth order spectral density of the innovations to ensure the convergence of the covariances in (5.5). Giraitis and Surgailis (1990) directly assume the convergence of

$$\frac{1}{T}\sum_{t_1,t_2,t_3,t_4=1}^{T} h_j\left(|t_1-t_2|\right) h_l\left(|t_3-t_4|\right)\left(\gamma\left(|t_3-t_1|\right)\gamma\left(|t_4-t_2|\right)+\gamma\left(|t_4-t_1|\right)\gamma\left(|t_3-t_2|\right)\right)$$

to

$$(2\pi)^3\int_{-\pi}^{\pi}\left(f\left(\lambda,\theta\right)g\left(\lambda,\theta\right)\right)^2 d\lambda < \infty.$$

Assumption 5.2 [**A**] (iv) and (v) allow us to establish the approximation of the integrated weighted periodogram via a truncation of the original process, rather than via a factorizations of its spectral density. They imply an exact decay rate of the Fourier coefficients of certain covariances that arise as a consequence of the truncation and enable us to establish the validity of the truncation via well-known results on the asymptotic behavior of the trace of Toeplitz matrices (see Fox and Taqqu, 1987, Theorem 1). Assumption 5.2[**H**](i) implies that the weights of the integrated weighted periodogram satisfy the sufficient condition of Theorem 7.3 of Giraitis and Taqqu (1999) on which we rely to obtain the asymptotic distribution of the approximated score vector. Assumption 5.2[**H**](ii) gives a condition for the asymptotic unbiasedness of $\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\mathbf{I}_T\left(\lambda\right)\right\}d\lambda$ which enables us to approximate the score vector by another long-memory quadratic form that can be easily truncated.

The following theorem establishes the asymptotic normality of the estimator.

Theorem 5.2 *Under assumptions 5.2,* $as\ T \to \infty$

$$\sqrt{T}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \to_d N_s(0, M^{-1}(\boldsymbol{\theta}_0)V(\boldsymbol{\theta}_0)M^{-1}(\boldsymbol{\theta}_0)),$$

where

$$M\left(\theta_0\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\theta}_0\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\boldsymbol{\theta}_0\right)\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\theta}_0\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\boldsymbol{\theta}_0\right)\right\}d\lambda,$$

and the matrix $V\left(\theta_0\right)$ *has elements*

$$V_{(i,j)} = \frac{1}{\pi}\int_{-\pi}^{\pi}\mathbf{tr}\left\{\mathbf{f}(\lambda,\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}_0)\mathbf{f}(\lambda,\boldsymbol{\theta})\frac{\partial}{\partial\theta_l}\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta}_0)\right\}d\lambda$$
$$+\frac{1}{2\pi}\sum_{r,t,u,v=1}^{T}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\frac{\partial}{\partial\theta_j}f_{rq}(\lambda_1,\boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_l}f_{uv}(\lambda_2,\boldsymbol{\theta}_0)\right\}\tilde{K}^{\mathbf{y}}_{rtuv}(-\lambda_1,\lambda_2,-\lambda_2)d\lambda_1 d\lambda_2.$$

The proof of Theorem 5.2 has the same structure of that of Theorem 4.2. The consistency of $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, guaranteed by Theorem 5.1, implies that, as $T \to \infty$, $\hat{\boldsymbol{\theta}}$ eventually enters an arbitrary neighborhood of $\boldsymbol{\theta}_0$. By definition $\hat{\boldsymbol{\theta}}$ solves the equation $(\partial/\partial\boldsymbol{\theta})\, Q_T\left(\hat{\boldsymbol{\theta}}\right) = 0$. The mean-value theorem implies that for $\mathring{\boldsymbol{\theta}}$, such that $\left\|\mathring{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right\| \leq \left\|\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right\|$,

$$0 = T^{1/2}\dot{Q}_T\left(\boldsymbol{\theta}\right) = T^{1/2}\dot{Q}_T\left(\boldsymbol{\theta}_0\right) + \left[\ddot{Q}_T\left(\boldsymbol{\theta}\right)\right]T^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right),$$

thus the asymptotic distribution of $T^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$ is obtained from the asymptotic distribution of $\left[\ddot{Q}_T\left(\boldsymbol{\theta}\right)\right]T^{1/2}\dot{Q}_T\left(\boldsymbol{\theta}\right)$. To establish the latter we prove the uniform convergence of the Hessian matrix to the positive definite matrix $M\left(\boldsymbol{\theta}\right)$ in Lemma 5.3 and we conjecture the asymptotic normality of the score vector in Conjecture 5.4.

Lemma 5.3 *Under Assumption 5.2, as* $T \to \infty$*, uniformly in* $\theta \in \Theta$*,*

$$\ddot{Q}_T\left(\boldsymbol{\theta}\right) \to M\left(\boldsymbol{\theta}\right)$$

almost surely, where $M\left(\boldsymbol{\theta}\right)$ *is a positive definite matrix with elements*

$$M^{(i,j)}\left(\theta\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\theta}\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\boldsymbol{\theta}\right)\right\}d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\theta}\right)\dot{\mathbf{f}}_{(i)}\left(\lambda,\boldsymbol{\theta}\right)\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\theta}\right)\dot{\mathbf{f}}_{(j)}\left(\lambda,\boldsymbol{\theta}\right)\right\}d\lambda$$
$$-\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\theta}\right)\right)\mathbf{f}\left(\lambda,\boldsymbol{\theta}_0\right)d\lambda\right\}d\lambda.$$

Lemma 5.3 and the consistency of $\hat{\theta}_T$ for θ_0 imply

$$\ddot{Q}_T(\boldsymbol{\theta}) \rightarrow_{a.s} M(\boldsymbol{\theta}_0).$$

In view of this result, the asymptotic normality of the estimator follows from the asymptotic normality of $\dot{Q}_T(\boldsymbol{\theta}_0)$.

Conjecture 5.4 *Under Assumption 5.2 as $T \rightarrow \infty$,*

Part (a):

$$\sqrt{T}\left[\dot{Q}_T(\theta_0) - E\dot{Q}_T(\theta_0)\right] \rightarrow_d N(0, V(\theta_0)),$$

where $V(\theta)$ is a positive definite matrix with (j,l) element:

$$V_{(j,l)} = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathbf{tr}\left[\mathbf{f}(\lambda,\theta_0)\dot{\mathbf{f}}_{(j)}^{-1}(\lambda,\theta_0)\mathbf{f}(\lambda,\theta)\dot{\mathbf{f}}_{(l)}^{-1}(\lambda,\theta_0)\right]d\lambda$$
$$+\frac{1}{2\pi}\sum_{a,b,c,d=1}^{T}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\dot{\mathbf{f}}_{(j)}^{(a,b)}(\lambda_1,\theta_0)\dot{\mathbf{f}}_{(l)}^{(c,d)}(\lambda_2,\theta_0)\right\}\tilde{K}_{a,b,c,d}(-\lambda_1,\lambda_2,-\lambda_2,\theta_0)d\lambda_1 d\lambda_2.$$

Part (b):

$$\sqrt{T}E\dot{Q}_T(\theta) \rightarrow 0.$$

The proof of the asymptotic normality of the score vector is based on the idea, introduced by Giraitis and Surgailis (1990), of approximating the score vector by another quadratic form which shares the same asymptotic distribution but has a less strong dependence structure. In what follows we show that each element of the score vector

$$\sqrt{T}\dot{Q}_{T(j)}(\theta_0) = \frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}(\lambda) - \mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right]\right\}d\lambda \tag{5.11}$$

shares the same asymptotic distribution of the quantity

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}(\lambda,\tilde{\mathbf{y}}_t) - \mathbf{EI}(\lambda,\tilde{\mathbf{y}}_t)\right]\right\}d\lambda, \tag{5.12}$$

where $\mathbf{I}(\lambda,\tilde{\mathbf{y}})$ and $\mathbf{EI}(\lambda,\tilde{\mathbf{y}})$ denote the periodogram and the expected value of the periodogram of a new process $\tilde{\mathbf{y}}_t$ that we define below, and $\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0) \equiv \mathbf{f}^{-2}(\lambda,\theta_0)\dot{\mathbf{f}}_{(j)}(\lambda,\theta_0)$. As a preliminary step, the following lemma allows us to approximate the jth element of the score vector

$$\sqrt{T}\left[\dot{Q}_{T(j)}(\theta_0) - E\dot{Q}_{T(j)}(\theta_0)\right]$$

by another quadratic form of the original process

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}_T\left(\lambda\right)-\mathbf{E}\mathbf{I}_T(\lambda)\right]\right\}d\lambda.$$

Lemma 5.5 *Under Assumption 5.2, as* $T\to\infty$

$$\lim_{T\to\infty}\sqrt{T}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{E}\mathbf{I}_T(\lambda)-\mathbf{f}(\lambda,\boldsymbol{\theta}_0)\right]\right\}d\lambda=0 \quad \text{for all } j=1,...,s.$$

We now introduce the new process $\tilde{\mathbf{y}}_t$. We set

$$\tilde{\mathbf{y}}_t\equiv\sum_{j=0}^{N}\boldsymbol{\Psi}_{0j}\boldsymbol{\epsilon}_{t-j-1}+\boldsymbol{\xi}_t \quad 0<N<\infty,$$

a truncation of of the original process $\mathbf{y}_t$ at a finite integer N. Consider that $\tilde{\mathbf{y}}_t$ has autocovariance function $\tilde{\Gamma}\left(u\right)$

$$\begin{aligned} &= I_{(u=0)}\boldsymbol{\Sigma}_{\xi}\left(\tau\right)+\boldsymbol{\Sigma}_{\epsilon}\left(\tau\right)\sum_{j=0}^{N-|u|}\Psi_j\left(\zeta\right)\Psi_{j+|u|}^{'}\left(\zeta\right)+I_{(m\neq0)}\Psi_{|u|-1}\left(\zeta\right)\boldsymbol{\Sigma}_{\xi\epsilon}\left(\tau\right) \quad \text{for } 0\leq u\leq N,\\ &= 0 \quad \text{for all } u>N. \end{aligned}$$

which satisfies

$$\sum_{u=0}^{\infty}\left|tr\left\{\tilde{\Gamma}\left(u\right)\right\}\right|<\infty. \tag{5.13}$$

By Lemma B1.3, (5.13) implies that $\tilde{\mathbf{f}}\left(\lambda,\theta\right)$ has elements in $L_2\left(\Pi\right)$ which are bounded and continuous at all $(\lambda,\boldsymbol{\theta})\in\Pi\times\Theta$. Moreover by definition $\tilde{\mathbf{y}}_t$ is a signal plus noise process where the signal is a finite order MA(N) process and can be represented as

$$\tilde{\mathbf{y}}_t=\sum_{l=0}^{\infty}\Phi_l^*\left(\theta\right)e_{l-t}^*, \quad \sum_{l=0}^{\infty}\left\|\Phi_l^*\left(\theta\right)\right\|^2<\infty, \tag{5.14}$$

where we define

$$\begin{aligned} e_t^* &\equiv \boldsymbol{\xi}_t, \qquad (5.15)\\ e_{t-l}^* &= \boldsymbol{\epsilon}_{t-l}, \quad l\geq 1 \end{aligned}$$

and

$$\begin{aligned}
\Phi_l^*(\boldsymbol{\theta}) &\equiv I_n, \quad l=0, \qquad (5.16)\\
\Phi_l^*(\boldsymbol{\theta}) &\equiv \Psi_{l-1}(\boldsymbol{\theta}), \quad 1\le l\le N\\
\Phi_l^*(\boldsymbol{\theta}) &\equiv \mathbf{0}_n \quad l>N.
\end{aligned}$$

We now establish the joint asymptotic normality of the quadratic forms

$$\begin{aligned}
&\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}\left(\lambda,\tilde{\mathbf{y}}_t\right)-\mathbf{E}\mathbf{I}(\lambda,\tilde{\mathbf{y}}_t)\right]\right\}d\lambda \quad j=1,...,s \qquad (5.17)\\
&=\frac{1}{2\pi}\sum_{a,b=1}^{n}\sqrt{T}\int_{-\pi}^{\pi}\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}_{(a,b)}\left(\lambda,\tilde{\mathbf{y}}_t\right)-\mathbf{E}\mathbf{I}_{(a,b)}(\lambda,\tilde{\mathbf{y}}_t)\right]d\lambda \quad j=1,...s.\\
&=\sum_{a,b=1}^{n}T^{-1/2}\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{h}_{(b,a)}^{(j)}\left(t-s\right)\left(\tilde{\mathbf{y}}_t^{(a)}\tilde{\mathbf{y}}_s^{(b)}-\tilde{\Gamma}_{(|t-s|)}^{(a,b)}\right)\right] \quad j=1,...,s,
\end{aligned}$$

where $\mathbf{h}_{(b,a)}^{(j)}$ denotes the Fourier coefficient of $\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)$,

$$\mathbf{h}_{(b,a)}^{(j)}\left(|t-s|\right)=(1/2\pi)\int_{-\pi}^{\pi}\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)e^{i(t-s)\lambda}d\lambda.$$

Even if (5.17) for $j=1,...,s$ is a quadratic form of a short-memory, N-dependent process we cannot derive its asymptotic normality relying on Lemma 4.2 because the function $\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)$ does not satisfy the necessary regularity conditions of square integrability and uniform continuity at all $(\lambda,\theta)\in\Pi\times\Theta$. Zaffaroni (2009) derives the asymptotic normality of (5.17) for a univariate "truncated" signal plus noise process (5.14) relying on Theorem 18.5.1 of Ibragimov and Linnik (1971, page 340), who provide a central limit theorem for ϕ-mixing processes with arbitrarily fast decreasing mixing coefficients. It is not clear how such result could be extended to quadratic forms of multivariate ϕ-mixing processes with different weight sequences which have Fourier transforms possibly unbounded at the zero frequency. To establish the asymptotic normality of (5.17) we rely on Theorem 7.3 of Giraitis and Taqqu (1999, page 29). Giraitis and Taqqu derive the joint asymptotic normality of quadratic form of multivariate Appell polynomials for linear sequences with $i.i.d$ innovations and possibly different weights and linear coefficients. Taking in their notation a multivariate Appell polynomial, $P_{m,n}\left(X_t^{(i,1)},X_s^{(i,2)}\right)$ of degree equal to one, to establish the result we must verify that the scalar quantities

$$T^{-1/2}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{h}_{(b,a)}^{(j)}\left(t-s\right)\left[\tilde{\mathbf{y}}_{t}^{(a)}\tilde{\mathbf{y}}_{s}^{(b)}-\tilde{\Gamma}_{(|t-s|)}^{(a,b)}\right]\right) \tag{5.18}$$

satisfy the regularity conditions of Theorem 6.1 of Giraitis and Taqqu (1999). Consider that each component of the truncated process has representation

$$\tilde{\mathbf{y}}_{t}^{(a)}=\sum_{l=0}^{\infty}\Phi_{a,l}^{*}\left(\boldsymbol{\theta}\right)e_{l-t}^{*} \tag{5.19}$$

where $\Phi_{a,l}^{*}$ denotes the ath row of the matrix $\Phi_{l}^{*}\left(\theta\right)$ defined in (5.16), the linear innovations e_{l-t}^{*} are defined in (5.15) and satisfy Assumption 5.2[**A**]. A sufficient condition for Theorem 6.1 of Giraitis and Taqqu is provided by Theorem 6.3.2 of Giraitis et al. (2012, Section 6.3.1, Chapter 6). For a stationary scalar linear process with finite fourth moments $i.i.d$ innovations, the asymptotic normality of (5.18) is implied by

$$f\left(\lambda\right) \leq C\left|\lambda\right|^{-\alpha}, \quad \text{and} \quad \left|g\left(\lambda\right)\right|\leq C\left|\lambda\right|^{-\beta}, \quad \text{for every } \lambda\in\Pi \tag{5.20}$$

$$\text{for some } -1 < \alpha,\ \beta<1, \qquad \alpha+\beta<1/2.$$

Consider that for all $a=1,...,n$, the spectral density function of $\tilde{\mathbf{y}}_{t}^{(a)}$ trivially satisfies

$$\tilde{\mathbf{f}}^{(a)}\left(\lambda,\theta\right)\leq C\left|\lambda\right|^{-\alpha} \quad \text{for all } \lambda\in\Pi$$

choosing $\alpha=0$. Moreover the function $\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_{0})$ is by definition a real valued even function with elements in $L_{1}\left(\Pi\right)$ that satisfy

$$\left|\mathbf{g}_{(j)}^{(a,b)}(\lambda,\boldsymbol{\theta}_{0})\right|\leq D\left(\left|\lambda\right|^{-1}\right)\left|\lambda\right|^{2d(\boldsymbol{\theta})}, \quad \left|\lambda\right|\leq\pi,$$

by Assumption 5.2[**H**] . Choosing $-\beta=2d\left(\theta\right)$ and $\alpha=0$,

$$\alpha+\beta<1/2,$$

and condition (5.20) is satisfied. The linear innovations in (5.19) are a sequence of independent random variables with finite fourth moments by Assumption 5.2 [**A**] and identically distributed at all $s=t-l$, $l\geq1$. Thus we make the following conjecture.

Conjecture 5.6 *Under Assumption 5.2, for all* $j = 1, ..., s$*, as* $T \to \infty$*, the quantities*

$$\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_{(j)}(\lambda, \boldsymbol{\theta}_0) \left[\mathbf{I} \left(\lambda, \tilde{\mathbf{y}}_t\right) - \mathbf{E}\mathbf{I}(\lambda, \tilde{\mathbf{y}}_t) \right] \right\} d\lambda$$

have a joint normal asymptotic distribution.

To establish the validity of the approximation we must show that the variance of the difference between

$$\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_{(j)}(\lambda, \boldsymbol{\theta}_0) \left[\mathbf{I}_T \left(\lambda\right) - \mathbf{E}\mathbf{I}_T(\lambda) \right] \right\} d\lambda \tag{5.21}$$

and

$$\frac{\sqrt{T}}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{g}_{(j)}(\lambda, \boldsymbol{\theta}_0) \left[\mathbf{I}_T \left(\lambda, \tilde{\mathbf{y}}_t\right) - \mathbf{E}\mathbf{I}_T(\lambda, \tilde{\mathbf{y}}_t) \right] \right\} d\lambda \tag{5.22}$$

tends to 0 as $T \to \infty$ for each j, so that the former quantities have the same asymptotic distribution of the latter. To that end, we write the difference between (5.21) and (5.22) as

$$tr \left\{ T^{-1/2} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\mathbf{h}}^{(j)} \left(|t-s|\right) \left[\mathbf{y}_t \mathbf{y}_s - \tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_s \right] \right\} \tag{5.23}$$

$$= tr \left\{ T^{-1/2} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\mathbf{h}}^{(j)} \left(|t-s|\right) \left[\mathbf{y}_t \left(\mathbf{y}_s - \tilde{\mathbf{y}}_s\right) + \tilde{\mathbf{y}}_s \left(\mathbf{y}_t - \tilde{\mathbf{y}}_t\right) \right] \right\}. \tag{5.24}$$

In what follows any function of $\boldsymbol{\theta}$ is evaluated at $\boldsymbol{\theta}_0$, however to avoid an excess of notation we omit it. (5.24) is given by the sum of the following three terms

$$Var \left[tr \left\{ T^{-1/2} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\mathbf{h}}^{(j)} \left(|t-s|\right) \mathbf{y}_t \left(\mathbf{y}_s - \tilde{\mathbf{y}}_s\right) \right\} \right] \tag{5.25}$$

$$Var \left[tr \left\{ T^{-1/2} \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\mathbf{h}}^{(j)} \left(|t-s|\right) \tilde{\mathbf{y}}_s \left(\mathbf{y}_t - \tilde{\mathbf{y}}_t\right) \right\} \right] \tag{5.26}$$

$$T^{-1} Cov \left[tr \left\{ \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\mathbf{h}}^{(j)} \left(|t-s|\right) \mathbf{y}_t \left(\mathbf{y}_s - \tilde{\mathbf{y}}_s\right) \right\}, tr \left\{ \sum_{s=1}^{T} \sum_{t=1}^{T} \tilde{\mathbf{h}}^{(j)} \left(|t-s|\right) \tilde{\mathbf{y}}_s \left(\mathbf{y}_t - \tilde{\mathbf{y}}_t\right) \right\} \right]. \tag{5.27}$$

The same bound applies to (5.25) and (5.26) and, by Schwartz inequality, to (5.27) as well. Therefore we follow closely Zaffaroni (2009, Lemma 7**)** and consider (5.25). Using the definition of the trace, (5.25) is

$$Var\left[\sum_{a,b=1}^{n}\left\{T^{-1/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\mathbf{h}}_{(b,a)}^{(j)}\left(|t-s|\right)\mathbf{y}_{t}^{(a)}\left(\mathbf{y}_{s}^{(b)}-\tilde{\mathbf{y}}_{s}^{(b)}\right)\right\}\right].$$

For $a,b,c,d=1,...,n$, (5.25) is given by the sum of the following three terms

$$T^{-1}\sum_{t,s=1}^{T}\sum_{p,q=1}^{T}\tilde{\mathbf{h}}_{(b,a)}^{(j)}\left(|t-s|\right)\tilde{\mathbf{h}}_{(d,c)}^{(j)}\left(|p-q|\right)cov\left(\mathbf{y}_{t}^{(a)}\mathbf{y}_{p}^{(c)}\right)cov\left(\left(\mathbf{y}_{s}^{(b)}-\tilde{\mathbf{y}}_{s}^{(b)}\right),\left(\mathbf{y}_{q}^{(d)}-\tilde{\mathbf{y}}_{q}^{(d)}\right)\right) \tag{5.28}$$

$$+T^{-1}\sum_{t,s=1}^{T}\sum_{p,q=1}^{T}\tilde{\mathbf{h}}_{(b,a)}^{(j)}\left(|t-s|\right)\tilde{\mathbf{h}}_{(d,c)}^{(j)}\left(|p-q|\right)cov\left(\mathbf{y}_{t}^{(a)},\left(\mathbf{y}_{q}^{(d)}-\tilde{\mathbf{y}}_{q}^{(d)}\right)\right)cov\left(\mathbf{y}_{p}^{(c)},\left(\mathbf{y}_{s}^{(b)}-\tilde{\mathbf{y}}_{s}^{(b)}\right)\right) \tag{5.29}$$

$$+T^{-1}\sum_{t,s=1}^{T}\sum_{p,q=1}^{T}\tilde{\mathbf{h}}_{(b,a)}^{(j)}\left(|t-s|\right)\tilde{\mathbf{h}}_{(d,c)}^{(j)}\left(|p-q|\right)cum\left(\mathbf{y}_{t}^{(a)},\mathbf{y}_{p}^{(c)},\left(\mathbf{y}_{s}^{(b)}-\tilde{\mathbf{y}}_{s}^{(b)}\right)\left(\mathbf{y}_{q}^{(d)}-\tilde{\mathbf{y}}_{q}^{(d)}\right)\right). \tag{5.30}$$

Consider (5.28). We show that this term is $O\left(\delta_N T\right)$, for a positive sequence satisfying $\delta_N \to 0$ as $N \to \infty$. As Zaffaroni (2009, Lemma 7), we rely on Theorem 1 of Fox and Taqqu (1987). To that end, we must establish that (5.28) can be expressed as

$$1/T\left(2\pi\right)^{4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}f\left(\lambda_{1}\right)g\left(\lambda_{2}\right)f\left(\lambda_{3}\right)g\left(\lambda_{4}\right)$$
$$\times\sum_{j_{1}=0}^{T-1}...\sum_{j_{4}=0}^{T-1}e^{i(j_{1}-j_{2})\lambda_{1}}e^{i(j_{2}-j_{3})\lambda_{2}}e^{i(j_{3}-j_{4})\lambda_{3}}e^{i(j_{4}-j_{3})\lambda_{4}}d\lambda_{1}d\lambda_{2}d\lambda_{3}d\lambda_{4},$$

for functions $f\left(.\right)$ and $g\left(.\right)$ bounded on the interval $[\delta,\pi]$ for $\delta>0$, and such that

$$\left|f\left(\omega\right)\right|=O\left(\left|\omega\right|^{-a-\delta}\right)\quad as\;\omega\to 0,$$

and

$$\left|g\left(\omega\right)\right|=O\left(\left|\omega\right|^{-b-\delta}\right)\quad as\;\omega\to 0,$$

with

$$2\left(\alpha+\beta\right)<1/2.$$

Setting

$$cov\left[\mathbf{y}_t^{(a)}\mathbf{y}_p^{(c)}\right] = (1/2\pi)\int_{-\pi}^{\pi}\mathbf{f}^{(a,c)}\left(\lambda_4\right)e^{i(p-t)\lambda_4}, \tag{5.31}$$

where $\mathbf{f}\left(\lambda\right)$ denotes the spectral density of $\mathbf{y}_t$, and setting

$$cov\left[\left(\mathbf{y}_s^{(b)} - \tilde{\mathbf{y}}_s^{(b)}\right), \left(\mathbf{y}_q^{(d)} - \tilde{\mathbf{y}}_q^{(d)}\right)\right] = (1/2\pi)\int_{-\pi}^{\pi}\breve{\mathbf{f}}^{(b,d)}\left(\lambda_2\right)e^{i(q-s)\lambda_2}, \tag{5.32}$$

where $\breve{\mathbf{f}}\left(\lambda\right)$ denotes the spectral density of $\left(\mathbf{y}_t - \tilde{\mathbf{y}}_t\right)$, (5.28) may be written as

$$(1/2\pi)^4\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathbf{f}^{(b,d)}\left(\lambda_2\right)\mathbf{g}_{(j)}^{(b,a)}(\lambda_1)\breve{\mathbf{f}}^{(a,c)}\left(\lambda_4\right)\mathbf{g}_{(j)}^{(d,c)}(\lambda_3) \tag{5.33}$$

$$\times\sum_{t=0}^{T-1}\sum_{s=0}^{T-1}\sum_{q=0}^{T-1}\sum_{p=0}^{T-1}e^{i(t-s)\lambda_1}e^{i(q-s)\lambda_2}e^{i(p-t)\lambda_4}e^{i(p-q)\lambda_3}d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4. \tag{5.34}$$

Theorem 1 of Fox and Taqqu (1987) cannot be directly applied to (5.33). We follow Lemma 7 of Zaffaroni (2009) and find an upper bound to (5.33) that satisfies the regularity conditions of Fox and Taqqu.

By Lemma C.1 (see Appendix C), for any $\eta \in (0,1)$, $(5.31)$and (5.32) are bounded by

$$K\left|\mathbf{\Psi}_N^{(a)}\right|^{\eta}\left\{\sum_{j=0}^{\infty}\left|\mathbf{\Psi}_j^{(a)}\mathbf{\Psi}_{j+|p-t|}^{(c)}\right|^{1-\eta/2}\right\},$$

where $\Psi_j^{(a)}$ is the ath element of the matrix of coefficients $\mathbf{\Psi}_j$ defined in (5.1). Set $u = |p-t|$. By Assumption 5.2[**A**] as $u \to \infty$

$$\left\{\sum_{j=0}^{\infty}\left|\Psi_j^{(a)}\right|^{1-\eta/2}\left|\Psi_{j+u}^{(c)}\right|^{1-\eta/2}\right\} \sim K\left|D\left(u\right)\right|u^{(1-\eta/2)(2d-1)}, \tag{5.35}$$

where $D\left(u\right)$ is a measurable Zygmund slowly varying function at infinity. Denote the Fourier transform of the left hand side of (5.35) by $\mathbf{b}\left(\lambda\right)$. (5.33) is bounded by

$$(1/2\pi)^4\left|\psi_{0N}\right|^{\eta}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathbf{b}^{(b,d)}\left(\lambda_2\right)\mathbf{g}_{(j)}^{(b,a)}(\lambda_1)\mathbf{b}^{(a,c)}\left(\lambda_4\right)\mathbf{g}_{(j)}^{(d,c)}(\lambda_3) \tag{5.36}$$

$$\times\sum_{t=0}^{T-1}\sum_{s=0}^{T-1}\sum_{q=0}^{T-1}\sum_{p=0}^{T-1}e^{i(t-s)\lambda_1}e^{i(q-s)\lambda_2}e^{i(p-t)\lambda_4}e^{i(p-q)\lambda_3}d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4.$$

By $(2.3.8)$ of Giraitis et al. (2012, page 19), $\mathbf{b}\left(\lambda\right)$ is

$$O\left(|\lambda|^{-(1+(1-\eta/2)(2d-1))-\delta}\right), \quad \text{as} \quad \lambda \to 0^+, \quad \text{for any } \delta > 0,$$

taking in their notation $-\beta = (1-\eta/2)(2d-1)$, and using the fact that any ZSV function satisfies $\left|D\left(u\right)\right| = O\left(|u|^{\delta}\right)$ as $u \to 0$.

Moreover Assumption 5.2 [**H**] implies that the function $\mathbf{g}_{(j)}(\lambda, \boldsymbol{\theta}_0)$ has elements of order $O\left(|\lambda|^{2d(\zeta)-\delta}\right)$. Set $-a = -\left(1 + (1-\eta/2)(2d-1)\right)$ and $-b = 2d(\theta_0)$, then choose $\eta < 1/\left(1-2d(\theta_0)\right)$. Since $\eta < 1/\left(1-2d(\theta_0)\right)$ implies $2(\alpha+\beta) = 2\left(-2d(\theta_0) + 1 + (1-\eta/2)(2d-1)\right) < 1$, Theorem 1 of Fox and Taqqu holds and (5.36) is $O\left(|\psi_{0N}|^{\eta} T\right)$, where $|\psi_{0N}|^{\eta}$ is a positive sequence converging to zero as $N \to \infty$ by Assumption 5.2[**A**].

Then we conclude that (5.28) is $O\left(|\psi_{0N}|^{\eta} T\right)$. The same bound applies to (5.29).

Consider (5.30). By Lemma C.2,

$$\left|cum\left(\mathbf{y}_t^{(a)}, \mathbf{y}_p^{(c)}, \left(\mathbf{y}_q^{(d)} - \tilde{\mathbf{y}}_q^{(d)}\right), \left(\mathbf{y}_s^{(b)} - \tilde{\mathbf{y}}_s^{(b)}\right)\right)\right| \leq K\left|\psi_{0N}\right|^{\eta/2} \sum_{j=0}^{\infty}\left|\psi_j \psi_{j+|s-q|}\right|^{1-\eta/2} \sum_{j=0}^{\infty}\left|\psi_j \psi_{j+|p-t|}\right|^{1-\eta/2}.$$

Setting $u_1 = |s-q|$, and $u_2 = |p-t|$, as $u_1 \to \infty$ and $u_2 \to \infty$, by Assumption 5.2[**A**]

$$\left\{\sum_{j=0}^{\infty}\left|\psi_j \psi_{j+u_1}\right|^{1-\eta/2}\right\} \sim K\left|D(u_1)\right| u_1^{(1-\eta/2)(2d-1)}, \tag{5.37}$$

$$\left\{\sum_{j=0}^{\infty}\left|\psi_j \psi_{j+u_2}\right|^{1-\eta/2}\right\} \sim K\left|D(u_2)\right|^{2-\eta} u_2^{(1-\eta/2)(2d-1)}. \tag{5.38}$$

The Fourier transform of (5.37) and (5.38), $\mathbf{b}(\lambda)$, is

$$O\left(\lambda^{-(1+(1-\eta/2)(2d-1))-\delta}\right), \quad \lambda \to 0^{+}, \quad \text{for any } \delta > 0,$$

by $(2.3.8)$ of Giraitis et al.(2012, page 19), taking in their notation $-\beta = (1-\eta/2)(2d-1)$. Then (5.30) is bounded by

$$(1/2\pi)^4\left|\psi_{0N}\right|^{\eta/2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathbf{g}_{(j)}^{(b,a)}(\lambda_1)\mathbf{b}(\lambda_2)\,\mathbf{g}_{(j)}^{(d,c)}(\lambda_3)\mathbf{b}(\lambda_4)$$
$$\sum_{t=0}^{T-1}\sum_{s=0}^{T-1}\sum_{q=0}^{T-1}\sum_{p=0}^{T-1} e^{i(t-s)\lambda_1} e^{i(q-s)\lambda_2} e^{i(p-q)\lambda_3} d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4$$

where the functions $\mathbf{g}_{(j)}(\lambda)$ and $\mathbf{b}(\lambda)$ satisfy all the regularity conditions of Theorem 1 of Fox and Taqqu and we may conclude that(5.30) is $O\left(|\psi_{0N}|^{\eta/2} T\right)$.

Then we conclude that (5.23) is $O\left(T\left|\psi_{0N}\right|^{\eta/2}\right)$, where $|\psi_{0N}|^{\eta/2}$ is a positive sequence converging to zero as $N \to \infty$ by Assumption 5.2[**A**], and the quantities

$$\sqrt{T}\left[\dot{Q}_T^{(j)}(\boldsymbol{\theta}_0) - E\dot{Q}_T^{(j)}(\boldsymbol{\theta}_0)\right] \tag{5.39}$$

share the same asymptotic distribution of (5.22) for all $j = 1, .., s$.

To complete the discussion of Conjecture 5.4 we must evaluate

$$\lim_{T\to\infty} Cov\left(\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}_T(\lambda)-\mathbf{E}\mathbf{I}_T(\lambda)\right]\right\}d\lambda, \frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(l)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}_T(\lambda)-\mathbf{E}\mathbf{I}_T(\lambda)\right]\right\}d\lambda\right)$$

$$= \lim_{T\to\infty}\frac{T}{4\pi^2}Cov\left(\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\mathbf{I}_T(\lambda)\right\}d\lambda, \int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(l)}(\lambda,\boldsymbol{\theta}_0)\mathbf{I}_T(\lambda)\right\}d\lambda\right) \tag{5.40}$$

$$= \lim_{T\to\infty}\frac{T}{4\pi^2}Cov\left(\sum_{a,b,c,d=1}^{n}\left\{\int_{-\pi}^{\pi}\mathbf{g}_{(j)}^{(b,a)}\mathbf{I}_{ab}(\lambda)\,d\lambda, \int_{-\pi}^{\pi}\mathbf{g}_{(l)}^{(c,d)}\mathbf{I}_{cd}(\lambda)\,d\lambda\right\}\right).$$

To that end we introduce the following lemma.

Lemma 5.6 *Under Assumption 5.2, for all* $a, b, c, d = 1, ..., n$

$$\begin{aligned}
&\lim_{T\to\infty} TCov\left(\int_{-\pi}^{\pi}\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)\,\mathbf{I}_{ab}(\lambda)\,d\lambda, \int_{-\pi}^{\pi}\mathbf{g}_{(l)}^{(d,c)}(\lambda,\boldsymbol{\theta}_0)\,\mathbf{I}_{cd}(\lambda)\,d\lambda\right)\\
&= 2\pi\int_{-\pi}^{\pi}\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)\,\bar{\mathbf{g}}_{(l)}^{(d,c)}(\lambda,\boldsymbol{\theta}_0)\,\mathbf{f}^{(a,b)}(\lambda,\boldsymbol{\theta}_0)\,\bar{\mathbf{f}}^{(c,d)}(\lambda,\boldsymbol{\theta}_0)\,d\lambda\\
&\quad +2\pi\int_{-\pi}^{\pi}\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)\,\bar{\mathbf{g}}_{(l)}^{(d,c)}(-\lambda,\boldsymbol{\theta}_0)\,\mathbf{f}^{(a,d)}(\lambda,\boldsymbol{\theta}_0)\,\bar{\mathbf{f}}^{(b,c)}(\lambda,\boldsymbol{\theta}_0)\,d\lambda\\
&\quad +2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathbf{g}_{(j)}(\lambda_1,\boldsymbol{\theta}_0)\,\mathbf{g}_{(l)}(\lambda_2,\boldsymbol{\theta}_0)\,\tilde{K}_{abcd}^{\mathbf{y}}(\lambda_1,\lambda_2,-\lambda_2)\,d\lambda_1 d\lambda_2.
\end{aligned}$$

By Lemma 5.6, we may conclude that as $T\to\infty$, (5.40) converges to $V_{(j,l)}(\boldsymbol{\theta}_0)$.

5.6 Conclusion

In Chapters 4 and 5 we introduced a new multivariate exponential volatility (MEV-) model that includes as special cases both the Constant Conditional Correlation (CCC) model with EGARCH individual volatilities and the multivariate Stochastic Volatility model with Leverage (MSV-L). Under our general model the logarithm of the vector of squared returns is decomposed into the sum of a signal vector-linear process and a vector white noise. We allow for correlation between the signal and the noise since it arises in the MSV-L model as a consequence of the leverage assumption and in the CCC model by definition. We discuss parametric estimation of the model by means of the Whittle frequency domain estimator and derive its asymptotic properties under short and long-memory parameterization, extending the statistical literature on Whittle estimation to cover correlated signal plus noise vector processes. Theorems 4.1 and 4.2 establish respectively the strong consistency and asymptotic normality of the estimator under the assumption of weak dependence in the signal process providing an extension of Hosoya and Taniguchi (1982) to vector signal plus noise processes whose spectral density cannot be readily factored. Theorems 5.1 and 5.2 derive analogous results under the assumption of long memory parameterization of the signal extending Hosoya (1997) to long-memory correlated signal plus noise processes whose spectral density has singularities only at the zero frequency. As argued in Zaffaroni (2009) for the case of a univariate signal plus noise process, it turns out that the Whittle estimator has a rate of convergence and an asymptotic distribution that do not depend on weather long memory holds or not. The proof of the strong consistency of the estimator is based on the well known result on consistency of M-estimators (see for example Van der Vaart, 1998, Section 5.2) and does not significantly differ under short and long-memory parameterization. Indeed positive definiteness of the autocovariance matrix, easily imposed on the model by its decomposition into a standard deviation and a correlation matrix, guarantees strict positivity of the spectrum. This implies the uniform convergence of the inverse of the spectrum at all frequencies even when the spectrum is unbounded at the zero frequency and thus ensures the uniform convergence of the second component of the objective function,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\boldsymbol{\theta}\right)\mathbf{I}_T\left(\lambda\right)\right\}d\lambda, \tag{5.41}$$

in both the short and the long-memory case. The uniform convergence of the first non random term of the objective function,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det\mathbf{f}\left(\lambda,\boldsymbol{\theta}\right)d\lambda, \tag{5.42}$$

is straightforward in the short-memory case when the spectrum is continuous at all frequencies; in the long-memory case it is obtained by imposing a degree of smoothness on the inverse of the spectrum and on its first derivatives that implies the strong equicontinuity of (5.42) ; this is a standard condition in the statistical literature on long-memory Whittle estimation (see for example Fox and Taqqu, 1986, 1987, Giraitis and Taqqu, 1997).

The joint asymptotic normality of the integrated weighted periodograms

$$\sqrt{T}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}\left(\lambda,\boldsymbol{\theta}_0\right)\left[\mathbf{I}_T\left(\lambda,\mathbf{y}_t\right)-\mathbf{E}\mathbf{I}_T\left(\lambda,\mathbf{y}_t\right)\right]\right\}d\lambda \quad j=1,...,s \tag{5.43}$$

is the key result in order to establish the asymptotic normality of the score vector which, together with the uniform convergence of the Hessian matrix, implies the asymptotic normality of the estimator. In Chapter 4 the asymptotic normality of (5.43) follows from the asymptotic normality of the serial covariances of short-memory vector linear processes as in Dunsmuir (1979) and Hosoya and Taniguchi (1982). Since the MEV model cannot be represented as a vector linear process with identically distributed innovations, as requested both by Theorem 2.1 of Dunsmuir (1979) and by Theorem 2.2 of Hosoya and Taniguchi (1982), we follow Robinson (1979) and conjecture that the rate of convergence and the asymptotic normality of the serial covariances are unchanged under mild violations of the identity of distribution. The Lipschitz continuity of degree $\alpha > 1/2$ of the spectral density implied by the short memory parameterization of Chapter 4 ensures that (5.43) can be approximated by its Cesaro sum approximation

$$\sqrt{T}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_M^{(j)}\left(\lambda,\boldsymbol{\theta}_0\right)\left[\mathbf{I}_T\left(\lambda,\mathbf{y}_t\right)-\mathbf{E}\mathbf{I}_T\left(\lambda,\mathbf{y}_t\right)\right]\right\}d\lambda \quad j=1,...,s, \tag{5.44}$$

and the joint asymptotic normality of (5.44) readily follows from the asymptotic normality of the serial covariances of the process and the square integrability of the spectrum. In Chapter 5 the assumption of long-range dependence implies that the model spectrum might not be Lipschitz continuous of the requested degree nor square integrable. Therefore in order to establish the joint asymptotic normality of the score vector we must rely on a different approximation of (5.43). Following Giraitis and Surgailis (1990) and Zaffaroni (2003, 2009) we approximate the integrated weighted periodogram by another quadratic form that shares the same asymptotic distribution but has a less persistent degree of memory. The approximation is based on the idea of imposing conditions on the weight function in (5.43) that effectively annihilate the singularities of the spectral density. Its validity is established extending Zaffaroni (2009, Lemma 7) to long-memory vector linear processes whose spectral density cannot be easily factored. The main limitation of Chapter 5 and of this thesis is that we are conjecturing the joint asymptotic normality and the rate of convergence of the approximated weighted periodogram

$$\frac{\sqrt{T}}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)\left[\mathbf{I}\left(\lambda,\tilde{\mathbf{y}}_t\right)-\mathbf{EI}(\lambda,\tilde{\mathbf{y}}_t)\right]\right\}d\lambda, \quad j=1,...,s, \tag{5.45}$$

defined at the short-memory truncation of $\mathbf{y}_t$, $\tilde{\mathbf{y}}_t$. The truncated process is a N-dependent process by definition, and (5.45) is a quadratic form in N-depended variates with weight function $\mathbf{g}_{(j)}(\lambda,\boldsymbol{\theta}_0)$ whose elements are $O\left(|\lambda|^{2d(\theta)-1+\delta}\right)$ as $\lambda\to\infty$ for any $\delta>0$. For the univariate case, Zaffaroni states that (5.45) is ϕ-mixing with arbitrarily fast decreasing mixing coefficients, therefore its asymptotic normality follows from Theorem 18.5.1 of Ibragimov and Linnik (1971), however it is not clear how this result extend to the multivariate case. For the multivariate case, Giraitis and Taqqu (2012) provide sufficient conditions for the joint asymptotic normality of (5.43) when the underlying possibly long-memory process can be represented as a vector linear process with independent and identically distributed innovations. However identity of distribution is not allowed for in the vector linear representation of the MEV- model and thus we cannot directly rely on their results. A central limit theorem for quadratic forms of type (5.45) arising from non-identically distributed vector linear processes needs to be investigated further. Alternatively it could be of great relevance to extend Giraitis and Taqqu limit theorems for bivariate Appell polynomials to stationary, possibly long-range dependent, linear processes whose innovations are not identically distributed.

The finite sample properties of the Whittle estimator in the MEV-model must be further explored by means of Monte-Carlo exercises in both the short and long-memory case and efficiency comparsion with maximum likelihood estimate of the parameters seems desirable. We expect maximum likelihood estimates to be more efficient, however it is of interest to explore in which cases Whittle estimates perform comparably (see Perez and Zaffaroni, 2008).

A relevant direction of further investigation is to consider parametric estimation of the MEV-model allowing the memory parameter not only to be unknown but also to lie in the nonstationary region. Hualde and Robinson (2011) investigate fractionally integrated, possibly non stationary, linear processes and establish the asymptotic normality of a one-step estimator based on an initial $\sqrt{T}$ consistent estimate of the parameters. Extensions of their results to signal plus noise processes would allow to test for non stationarity in the fractionally integrated multivariate exponential volatility model, thus providing a very general framework for testing for non stationarity in multivariate stochastic and conditional volatility models.

5.7 Appendix C: Technical Lemmas

This section proves the technical lemmas used to establish asymptotic normality of the estimator.

Proof of Lemma 5.3 We establish uniform convergence of $\ddot{Q}_T(\boldsymbol{\theta})$ to $M(\boldsymbol{\theta})$ pointwise. The (i,j) element of $\ddot{Q}_T(\boldsymbol{\theta})$, $\ddot{Q}_T^{(i,j)}(\boldsymbol{\theta})$ is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\,\dot{\mathbf{f}}_{(i)}(\lambda,\theta)\,\mathbf{f}^{-1}(\lambda,\theta)\,\dot{\mathbf{f}}_{(j)}(\lambda,\theta)\right\}d\lambda \tag{5.46}$$

$$+\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\right)\mathbf{I}_T(\lambda)\right\}d\lambda. \tag{5.47}$$

(5.47) converges in almost surely, uniformly in $(\lambda,\theta)\in\Pi\times\Theta$ to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j'}\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})\right)\mathbf{f}(\lambda,\theta_0)\right\}d\lambda,$$

by Lemma 5.4, taking $\mathbf{h}(\lambda,\theta)\equiv\left(\partial^2/\partial\theta_i\partial\theta_j'\right)\mathbf{f}^{-1}(\lambda,\boldsymbol{\theta})$, which by Assumption 5.2[**G**](v) is a continuous matrix function at all $(\lambda,\theta)\in\Pi\times\Omega$ and has symmetric elements in the interval $[-\pi,\pi]$.

The two terms of (5.46) are non stochastic. Their uniform convergence in $\theta\in\Theta$ follows from establishing their equicontinuity property. For the first term, we wish to prove that

$$\lim_{\varepsilon\to 0}\sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|\le\varepsilon}\left|\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}\left(\lambda,\tilde{\theta}\right)\ddot{\mathbf{f}}_{(i,j)}\left(\lambda,\tilde{\theta}\right)-\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}d\lambda\right|=0. \tag{5.48}$$

The equicontinuity property of $A(\theta)\equiv\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}$ is implied by

$$\sup_{\theta\in\Theta}\left|\frac{\partial}{\partial\theta_l}\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}d\lambda\right|<\infty, \tag{5.49}$$

(see Davidson, 1994, Theorem 21.10, page 339). We must establish that $A(\theta)$ is differentiable under the integral sign. We follow Fox and Taqqu (1986, Lemma 6). Denote the jth unit vector in R^s by i_j, and consider

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{\varepsilon}tr\left\{\mathbf{f}^{-1}(\lambda,\theta+i_j\varepsilon)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta+i_j\varepsilon)-\mathbf{f}^{-1}(\lambda,\theta)\,\ddot{\mathbf{f}}_{(i,j)}(\lambda,\theta)\right\}$$

By the mean value theorem the integrand is dominated for each $\lambda \neq 0$ by

$$\left| \frac{\partial}{\partial \theta_l} tr \left\{ \mathbf{f}^{-1}\left(\lambda, \theta^*\left(\lambda\right)\right) \ddot{\mathbf{f}}_{(i,j)}\left(\lambda, \theta^*\left(\lambda\right)\right) \right\} \right|, \tag{5.50}$$

where $|\theta^*\left(\lambda\right) - \theta| < \varepsilon$. Taking derivatives (5.50) is equal to

$$\left| tr \left\{ \mathbf{f}^{-1}\left(\lambda, \theta^*\left(\lambda\right)\right) \left[\frac{\partial}{\partial \theta_l} \ddot{\mathbf{f}}_{(i,j)}\left(\lambda, \theta^*\left(\lambda\right)\right) \right] + \left[\frac{\partial}{\partial \theta_l} \mathbf{f}^{-1}\left(\lambda, \theta^*\left(\lambda\right)\right) \right] \ddot{\mathbf{f}}_{(i,j)}\left(\lambda, \theta^*\left(\lambda\right)\right) \right\} \right| \tag{5.51}$$

By Assumption 5.2[**C**] and [**G**] (5.51) is at most, ignoring constant terms,

$$|\lambda|^{2(d_l - d_u) - \delta},$$

where we take $d_l = \inf_\Theta d\left(\theta\right)$ and $d_u = \sup_\Theta d\left(\theta\right)$. Since $\left(d_l - d_u\right) > -1/2$ and δ can be taken arbitrarily small, we choose δ such that

$$\int_{-\pi}^{\pi} |\lambda|^{2(d_l - d_u) - \delta} \, d\lambda < \infty. \tag{5.52}$$

Hence the dominated convergence theorem implies that $A\left(\theta\right)$ is differentiable under the integral sign. Moreover (5.52) implies (5.49).

The equicontinuity property of the second term in (5.46) is implied by

$$\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_l} \frac{1}{2\pi} \int_{-\pi}^{\pi} tr \left\{ \mathbf{f}^{-1}\left(\lambda, \theta\right) \dot{\mathbf{f}}_{(i)}\left(\lambda, \theta\right) \mathbf{f}^{-1}\left(\lambda, \theta\right) \dot{\mathbf{f}}_{(j)}\left(\lambda, \theta\right) \right\} d\lambda \right| < \infty, \tag{5.53}$$

(see Davidson, 1994, Theorem 21.10, page 339). The left hand side of (5.53) is differentiable under the integral sign because for $|\theta^*\left(\lambda\right) - \theta| < \varepsilon$

$$\begin{aligned}
& \left| \frac{\partial}{\partial \theta_l} tr \left\{ \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \dot{\mathbf{f}}_{(i)}\left(\lambda, \theta^*\right) \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \dot{\mathbf{f}}_{(j)}\left(\lambda, \theta^*\right) \right\} \right| \\
= \; & | tr \left\{ \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \dot{\mathbf{f}}_{(i)}\left(\lambda, \theta^*\right) \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \ddot{\mathbf{f}}_{(j,l)}\left(\lambda, \theta^*\right) \right\} \\
& + tr \left\{ \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \dot{\mathbf{f}}_{(i)}\left(\lambda, \theta^*\right) \left(\frac{\partial}{\partial \theta_l} \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \right) \dot{\mathbf{f}}_{(j)}\left(\lambda, \theta^*\right) \right\} \\
& + tr \left\{ \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \ddot{\mathbf{f}}_{(i,l)}\left(\lambda, \theta^*\right) \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \dot{\mathbf{f}}_{(j)}\left(\lambda, \theta^*\right) \right\} \\
& + tr \left\{ \left(\frac{\partial}{\partial \theta_l} \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \right) \dot{\mathbf{f}}_{(i)}\left(\lambda, \theta^*\right) \mathbf{f}^{-1}\left(\lambda, \theta^*\right) \dot{\mathbf{f}}_{(j)}\left(\lambda, \theta^*\right) \right\} |,
\end{aligned}$$

which is at most is at most, ignoring constant terms

$$|\lambda|^{2(d_l - d_u) - \delta},$$

where we take $d_l = \inf_\Theta d\left(\theta\right)$ and $d_u = \sup_\Theta d\left(\theta\right)$. Since $\inf_\Theta d\left(\theta\right)$ and $\sup_\Theta d\left(\theta\right)$ are bounded by Assumption 5.2[B] and $(d_l - d_u) > -1/4$ and δ can be taken arbitrarily small, we choose δ such that

$$\int_{-\pi}^{\pi} |\lambda|^{2(d_l-d_u)-\delta}\, d\lambda < \infty.$$

and the use of the dominated convergence theorem concludes the proof.

Proof of Lemma 5.5 Denote by $D_T(\lambda)$ the Fejer Kernel,

$$D_T(\lambda) = \frac{\sin^2(T\lambda/2)}{\sin^2(\lambda/2)} = \left|\sum_{s=1}^{T-1} e^{i\lambda s}\right|^2 \quad \lambda \in \Pi, T \geq 1.$$

and recall that (see Giraitis et al.2012, Chapter 2, page 8)

$$|D_T(\lambda)| \leq 2\pi T (1 + T|\lambda|)^{-1}, \quad \text{for every } \lambda \in \Pi. \tag{5.54}$$

Using the definition of Fejer Kernel,

$$\mathbf{EI}(\lambda) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbf{f}(\omega) D_T^2(\omega+\lambda)\, d\omega.$$

To simplify notation we set $\mathbf{g}_{(j)}(\lambda, \boldsymbol{\theta}_0) \equiv \mathbf{g}(\lambda)$, $\mathbf{f}(\lambda, \boldsymbol{\theta}_0) \equiv \mathbf{f}(\lambda)$, then

$$\begin{aligned}
&\int_{-\pi}^{\pi} tr\left\{\mathbf{g}(\lambda)\left[\mathbf{EI}_T(\lambda) - \mathbf{f}(\lambda)\right]\right\} d\lambda \\
&= \int_{-\pi}^{\pi} tr\left\{\mathbf{g}(\lambda)\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbf{f}(\omega) D_T(\omega+\lambda)\, d\omega - \mathbf{f}(\lambda)\right]\right\} d\lambda \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} tr\left\{\mathbf{g}(\lambda)\left[\mathbf{f}(\omega) D_T^2(\omega+\lambda) - \mathbf{f}(\lambda)\right]\right\} d\lambda d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} D_T^2(\lambda)\, tr\left\{\int_{-\pi}^{\pi} \mathbf{g}(\omega)\left[\mathbf{f}(\omega) - \mathbf{f}(\omega-\lambda)\right] d\omega\right\} d\lambda.
\end{aligned}$$

By Assumption 5.2[**H**]

$$\left|tr\left\{\int_{-\pi}^{\pi} \mathbf{g}(\omega)\left[\mathbf{f}(\omega) - \mathbf{f}(\omega-\lambda)\right] d\omega\right\}\right| \leq K|\lambda|^{\gamma}, \quad \lambda \in \Pi \tag{5.55}$$

for some $1/2 < \gamma < 1$ and some finite positive constant K. This together with (5.54) implies that

$$\begin{aligned}
&T^{-1/2}\frac{1}{2\pi}\int_{-\pi}^{\pi} D_T^2(\lambda)\, tr\left\{\int_{-\pi}^{\pi} \mathbf{g}(\omega)\left[\mathbf{f}(\omega) - \mathbf{f}(\omega-\lambda)\right] d\omega\right\} d\lambda \\
&\leq CT^{-1/2}\int_{-\pi}^{\pi} \frac{T^2}{1+(T\lambda)^2}|\lambda|^{\gamma}\, d\lambda \\
&\leq CT^{1/2-\gamma}\int_{-\infty}^{\infty}\left(1+\lambda^2\right)^{-1}|\lambda|^{\gamma}\, d\lambda \\
&\leq CT^{1/2-\gamma}
\end{aligned}$$

which goes to zero as $T \to \infty$ and completes the proof.

Lemma C.1 *Under assumption 5.2, the following holds*

Part (a)

(1) $Cov\left[\mathbf{y}_t^{(a)}\mathbf{y}_p^{(c)}\right] = I_{(t=p)}\mathbf{\Sigma}_{\xi}^{(a,c)}(\tau)+\mathbf{\Sigma}_{\epsilon}^{(a,c)}(\tau)\sum_{j=0}^{\infty}\Psi_j^{(a)}(\zeta)\,\Psi_{j+|t-p|}^{\prime(c)}(\zeta)+I_{(t\neq p)}\Psi_{|t-p|-1}^{(a)}(\zeta)\,\mathbf{\Sigma}_{\xi\epsilon}^{(a,c)}(\tau)$.

(2) $Cov\left[\left(\mathbf{y}_s^{(b)}-\tilde{\mathbf{y}}_s^{(b)}\right),\left(\mathbf{y}_q^{(d)}-\tilde{\mathbf{y}}_q^{(d)}\right)\right] = \mathbf{\Sigma}_{\epsilon}^{(b,d)}(\tau)\sum_{j=N+1}^{\infty}\Psi_j^{(b,b)}(\zeta)\,\Psi_{j+|q-s|}^{\prime(d,d)}(\zeta)$.

(3) $Cov\left[\mathbf{y}_t^{(a)},\left(\mathbf{y}_q^{(d)}-\tilde{\mathbf{y}}_q^{(d)}\right)\right] = \mathbf{\Sigma}_{\epsilon}^{(a,d)}(\tau)\sum_{j=N+1}^{\infty}\Psi_j^{(a,a)}(\zeta)\,\Psi_{j+|q-t|}^{(d,d)\prime}(\zeta)+I_{(q-t>N+2)}\Psi_{q-t-1}^{(a)}(\zeta)\,\mathbf{\Sigma}_{\xi\epsilon}^{(a,d)}(\tau)$.

Part (b) Setting $u_1=|t-p|$ and $u_2=|s-q|$, and choosing some $\eta\in(0,1)$

$$\left|Cov\left[\mathbf{y}_t^{(a)}\mathbf{y}_p^{(c)}\right]\right| \leq K\left|\Psi_N^{(a)}(\zeta)\right|^{\eta}\left\{\sum_{j=0}^{\infty}\left|\Psi_j^{(a)}(\zeta)\right|^{1-\eta/2}\left|\Psi_{j+u_1}^{\prime(c)}(\zeta)\right|^{1-\eta/2}\right\},$$

and

$$\left|cov\left[\left(\mathbf{y}_s^{(b)}-\tilde{\mathbf{y}}_s^{(b)}\right),\left(\mathbf{y}_q^{(d)}-\tilde{\mathbf{y}}_q^{(d)}\right)\right]\right| \leq K\left|\Psi_N^{(b)}(\zeta)\right|^{\eta}\left\{\sum_{j=0}^{\infty}\left|\Psi_j^{(b)}(\zeta)\right|^{1-\eta/2}\left|\Psi_{j+u_2}^{\prime(d)}(\zeta)\right|^{1-\eta/2}\right\}.$$

Proof Part (a) follows simply from the definition of the processes. For part (b) it is trivial to see that

$$\left|Cov\left(\mathbf{y}_t^{(a)}\mathbf{y}_p^{(c)}\right)\right| \leq K\left\{\sum_{j=0}^{\infty}\left|\Psi_j^{(b)}(\zeta)\right|^{1-\eta/2}\left|\Psi_{j+u_2}^{\prime(d)}(\zeta)\right|^{1-\eta/2}\right\},$$

which implies the result. Moreover

$$\begin{aligned}
\left|cov\left[\left(\mathbf{y}_s^{(b)}-\tilde{\mathbf{y}}_s^{(b)}\right),\left(\mathbf{y}_q^{(d)}-\tilde{\mathbf{y}}_q^{(d)}\right)\right]\right| &\leq \mathbf{\Sigma}_{\epsilon}^{(b,d)}\left|\sum_{j=N+1}^{\infty}\Psi_j^{(b)}(\zeta)\,\Psi_{j+u_2}^{\prime(d)}(\zeta)\right| \\
&\leq \mathbf{\Sigma}_{\epsilon}^{(b,d)}\sum_{j=N+1}^{\infty}\left|\Psi_j^{(b)}(\zeta)\right|^{\eta/2}\left|\Psi_j^{(b)}(\zeta)\right|^{1-\eta/2}\left|\Psi_{j+u_2}^{\prime(d)}(\zeta)\right|^{\eta/2}\left|\Psi_{j+u_2}^{\prime(d)}(\zeta)\right|^{1-\eta/2} \\
&\leq K\left|\Psi_N^{(b)}\right|^{\eta}\left\{\sum_{j=0}^{\infty}\left|\Psi_j^{(b)}(\zeta)\right|^{1-\eta/2}\left|\Psi_{j+u_2}^{\prime(d)}(\zeta)\right|^{1-\eta/2}\right\}.
\end{aligned}$$

Lemma C.2 *Set* $K_{abcd}^{\epsilon}\equiv cumulant(\epsilon_0^{(a)},\epsilon_0^{(b)},\epsilon_0^{(c)},\epsilon_0^{(d)})$, $K_{abcd}^{\xi}\equiv cumulant(\xi_0^{(a)},\xi_0^{(b)},\xi_0^{(c)},\xi_0^{(d)})$ *and let* $K_{abcd}(\mathbf{x}_t,\mathbf{y}_t,\mathbf{z}_t,\mathbf{u}_t)$ *denotes the fourth order cumulant of the* a,b,c,d *elements of random vectors* $\mathbf{x}_t,\mathbf{y}_t,\mathbf{z}_t,\mathbf{u}_t$. *Set* $t=t_1, p=t_2, q=t_3$ and $s=t_4$. *Under Assumption 5.2*

$$cum\left[\mathbf{y}_t^{(a)},\mathbf{y}_p^{(c)},\left(\mathbf{y}_q^{(d)}-\tilde{\mathbf{y}}_q^{(d)}\right),\left(\mathbf{y}_s^{(b)}-\tilde{\mathbf{y}}_s^{(b)}\right)\right] \leq K\left|\psi_{0N}\right|^{\eta/2}\sum_{j=0}^{\infty}\left|\psi_j\psi_{j+|s-q|}\right|^{1-\eta/2}\sum_{j=0}^{\infty}\left|\psi_j\psi_{j+|p-t|}\right|^{1-\eta/2}.$$

Proof $Cum\left[\mathbf{y}_t^{(a)}, \mathbf{y}_p^{(c)}, \left(\mathbf{y}_q^{(d)} - \tilde{\mathbf{y}}_q^{(d)}\right), \left(\mathbf{y}_s^{(b)} - \tilde{\mathbf{y}}_s^{(b)}\right)\right]$ is given by the sum of the following terms

$$K_{abcd}\left(\boldsymbol{\xi}_0, \boldsymbol{\xi}_0, \epsilon_0, \epsilon_0\right) \Psi_{t_3-t_2-1}^{(d,d)} \Psi_{t_4-t_1-1}^{(b,b)} \mathit{1}\left(t_1 = t_2\right) \mathit{1}\left(t_3 - t_2 \geq N+1\right) \mathit{1}\left(t_4 - t_1 \geq N+1\right) \tag{5.56}$$

$$K_{abcd}\left(\boldsymbol{\xi}_0, \epsilon_0, \epsilon_0, \epsilon_0\right) \Psi_{t_2-t_1-1}^{(c,c)} \Psi_{t_3-t_1-1}^{(d,d)} \Psi_{t_4-t_2-1}^{(b,b)} \mathit{1}\left(t_3 - t_1 \geq N+1\right) \mathit{1}\left(t_4 - t_1 \geq N+1\right) \tag{5.57}$$

$$K_{abcd}\left(\boldsymbol{\xi}_0, \epsilon_0, \epsilon_0, \epsilon_0\right) \Psi_{t_2-t_1-1}^{(a,a)} \Psi_{t_3-t_2-1}^{(d,d)} \Psi_{t_4-t_2-1}^{(b,b)} \mathit{1}\left(t_3 - t_2 \geq N+1\right) \mathit{1}\left(t_4 - t_1 \geq N+1\right)$$

$$K_{abcd}\left(\epsilon_0, \epsilon_0, \epsilon_0, \epsilon_0\right) \sum_{j=0}^{\infty} \mathit{1}\left(j \geq \max\left\{0, N+1-t_3+t_1, N+1-t_4+t_1\right\}\right) \Psi_j^{(a)} \Psi_{j+t_2-t_1}^{(c)} \Psi_{j+t_3-t_1}^{(b)} \Psi_{j+t_4-t_1}^{(d)}.$$

The result follows by Lemma 7 of Zaffaroni (2009, page 198).

Proof of Lemma 5.6 Set

$$\tilde{\mathbf{h}}_{(j)}\left(u\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{g}_{(j)}^{(a,b)}\left(\lambda\right) e^{iu\lambda} d\lambda,$$
$$\tilde{\mathbf{h}}_{(l)}\left(u\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathbf{g}_{(l)}^{(c,d)}\left(\lambda\right) e^{iu\lambda} d\lambda.$$

Then

$$TCov\left\{\int_{-\pi}^{\pi} \mathbf{g}_{(j)}^{(a,b)}\left(\lambda\right) \mathbf{I}_{ab}\left(\lambda\right) d\lambda, \int_{-\pi}^{\pi} \mathbf{g}_{(l)}^{(c,d)}\left(\lambda\right) \mathbf{I}_{cd}\left(\lambda\right) d\lambda\right\}$$

$$= \frac{1}{T} \sum_{u_1 u_2 u_3 u_4 = 1}^{T} \tilde{\mathbf{h}}_{(j)}\left(u_1 - u_2\right) \tilde{\mathbf{h}}_{(l)}\left(u_3 - u_4\right) \boldsymbol{\Gamma}_{(a,c)}\left(u_3 - u_1\right) \boldsymbol{\Gamma}_{(b,d)}\left(u_4 - u_2\right) \tag{5.58}$$

$$+ \frac{1}{T} \sum_{u_1 u_2 u_3 u_4 = 1}^{T} \tilde{\mathbf{h}}_{(j)}\left(u_1 - u_2\right) \tilde{\mathbf{h}}_{(l)}\left(u_3 - u_4\right) \boldsymbol{\Gamma}_{(a,d)}\left(u_4 - u_1\right) \boldsymbol{\Gamma}_{(b,c)}\left(u_3 - u_2\right) \tag{5.59}$$

$$+ \frac{1}{T} \sum_{u_1 u_2 u_3 u_4 = 1}^{T} \tilde{\mathbf{h}}_{(j)}\left(u_1 - u_2\right) \tilde{\mathbf{h}}_{(l)}\left(u_3 - u_4\right) \mathbf{K}_{abcd}^{\mathbf{y}}\left(u_2 - u_1, u_3 - u_1, u_4 - u_1\right). \tag{5.60}$$

Consider (5.58). This term can be written as

$$(1/2\pi)^4 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathbf{f}^{(a,c)}\left(\lambda_2\right) \mathbf{g}_{(j)}^{(a,b)}(\lambda_1) \mathbf{f}^{(b,d)}\left(\lambda_4\right) \mathbf{g}_{(l)}^{(c,d)}(\lambda_3) \tag{5.61}$$

$$\times \sum_{u_1=0}^{T-1} \sum_{u_2=0}^{T-1} \sum_{u_3=0}^{T-1} \cdot \sum_{u_4=0}^{T-1} .e^{i(t-s)\lambda_1} e^{i(q-s)\lambda_2} e^{i(p-t)\lambda_4} e^{i(p-q)\lambda_3} d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4, \tag{5.62}$$

where by Assumption 5.2[**C**]

$$\left|\mathbf{f}^{(a,c)}\left(\lambda, \theta\right)\right| \leq D\left(\lambda^{-1}, \boldsymbol{\theta}\right) \left|\lambda\right|^{-2d(\theta)} \quad \lambda \to 0^+, \quad d\left(\boldsymbol{\theta}\right) \in \left(0, 1/2\right),$$

and

$$\left|\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)\right| \le D\left(\lambda^{-1},\boldsymbol{\theta}\right)|\lambda|^{2d(\theta)} \quad \lambda\to 0^+ \quad d(\boldsymbol{\theta})\in(0,1/2) \quad \text{for all } j=1,...,s,$$

which imply (see Fox and Taqqu, 1986, Section 4) that

$$\begin{aligned}\left|\mathbf{f}^{(a,c)}(\lambda,\theta)\right| &= O\left(|\lambda|^{-2d(\boldsymbol{\theta})-\delta}\right) \quad \lambda\to 0^+,\\ \left|\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)\right| &= O\left(|\lambda|^{2d(\boldsymbol{\theta})-\delta}\right) \quad \lambda\to 0^+.\end{aligned}$$

Set $a=2d(\boldsymbol{\theta})$ and $-b=2d(\boldsymbol{\theta})$, then since $d(\boldsymbol{\theta})\in(0,1/2)$ and $2(a+b)<1$, (5.61) satisfies the regularity conditions of Theorem 1 of Fox and Taqqu. Hence we may conclude that as $T\to\infty$, (5.61) converges to

$$2\pi\int_{-\pi}^{\pi}\mathbf{g}_{(j)}^{(b,a)}(\lambda,\boldsymbol{\theta}_0)\,\bar{\mathbf{g}}_{(l)}^{(d,c)}(\lambda,\boldsymbol{\theta}_0)\,\mathbf{f}^{(a,b)}(\lambda,\boldsymbol{\theta}_0)\,\bar{\mathbf{f}}^{(c,d)}(\lambda,\boldsymbol{\theta}_0)\,d\lambda.$$

The same result is used to show the convergence of (5.59). Consider (5.60). By Lemma 7 of Zaffaroni (2009), for all $a,b.c,d=1,...,n$,

$$\left|\mathbf{K}_{abcd}^{\mathbf{y}}(u_2-u_1,u_3-u_1,u_4-u_1)\right| \le K\sum_{j=0}^{\infty}\left|\Psi_j^{(a)}\right|\left|\Psi_{j+|u_4-u_3|}^{(b)}\right|\sum_{j=0}^{\infty}\left|\Psi_j^{(c)}\right|\left|\Psi_{j+|u_2-u_1|}^{(d)}\right|. \quad (5.63)$$

By Assumption 5.2[A], choosing $\eta=0$,

$$\Delta(u)\equiv\left\{\sum_{j=0}^{\infty}\left|\Psi_j^{(a)}\right|\left|\Psi_{j+|u|}^{(b)}\right|\right\}\sim K\left|D(u,\boldsymbol{\theta})\right|u^{(2d(\theta)-1)} \quad \text{as} \quad u\to\infty. \quad (5.64)$$

By $(2.3.8)$ of Giraitis et al. (2012, page 19), the Fourier transform of (5.64) is

$$O\left(|\lambda|^{-(1+(2d-1))-\delta}\right), \quad \text{as} \quad \lambda\to 0^+, \quad \text{for any } \delta>0,$$

where in their notation $-\beta=(2d-1)$, using the fact that any ZSV function satisfies $|D(u)|=O\left(|u|^{\delta}\right)$ as $u\to 0$. Then the Fourier transform of the right hand side of (5.63) is $O\left(|\lambda|^{-2(1+(2d-1))-\delta}\right)$. Then the regularity conditions of Theorem 1 of Fox and Taqqu are satisfied and we conclude that (5.60) as $T\to\infty$, converges to

$$2\pi\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathbf{g}_{(j)}(\lambda_1,\boldsymbol{\theta}_0)\,\mathbf{g}_{(l)}(\lambda_2,\boldsymbol{\theta}_0)\,\tilde{K}_{abcd}^{\mathbf{y}}(\lambda_1,\lambda_2,-\lambda_2)\,d\lambda_1 d\lambda_2.$$

This completes the proof of Lemma 5.6.

References for Chapter 1

Andrews, D. W. K.: Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. Econometrica 59, 817-858 (1991).

Andrews, D. W. K.: Cross-Section Regression with Common Shocks. Econometrica 73, 1551-1586 (2005).

Aroian, L. A.: The Probability Function of the Product of Two Normally Distributed Variables. Annals of Mathematical Statistics 18, 265-271 (1947).

Baltagi, B.: Econometric Analysis of Panel Data. Wiley, New York 2005.

Baltagi, B. and Kao, C.: Nonostationary Panels, Cointegration in Panels and Dynamic Panels: A Survey. Advances in Econometrics 15, 7-51 (2000).

Baltagi, B., Kao, C., Liu, L.: Asymptotic properties of Estimators for the Linear Panel Regression Model with Individual Effect. Econometrics Journal 11, 554-772 (2008).

Baltagi, B., Kramer, W.: A simple linear trend model with error component, Econometric Theory 13, 436-463 (1997).

Billingsley, P.: Probability and Measure. John Wiley New York, New York 1986.

Billingsley, P.: Convergence of Probability Measures. Wiley, New York 1999.

Breitung, J., Meyer, W.: Testing for unit roots in panel data: Are wages on different bargaining levels cointegrated? Applied Economics 26, 353-361 (1994).

Breitung, J., Pesaran, M.H.: Unit Roots and Cointegration in Panel Data. In: The Econometrics of Panel Data 3rd ed., L. Mátyás and P. Sevestre ed., Kluwer, 236-265 (2007).

Chamberlain, G.: Panel Data. In: Handbook of Econometrics Vol. 2, ed. by Z. Grilichesand M. Intriligator, North-Holland, Amsterdam 1984.

Choi, C.Y., Hu, L., Ogaki, M.: Robust Estimation for Structural Spurious Regressions and a Hausman-type Cointegration Test. Journal of Econometrics 142, 327-351 (2008).

Conley, T.: Econometric Modelling of Cross Sectional Dependence. Mimeo, 1997.

Dudley, R.: Real Analysis and Probability. Wadsworth & Brooks Cole Mathematics Series, Pacific Grove 1989.

Eicker, F.: Central Limit Theorems for Families of Sequences of Random Variables. Annals of Mathematical Statistics 34, 439-446 (1963).

Engle, R. F., Granger, C.W.J.: Co-integration and error correction: Representation, estimation,and testing. Econometrica 55, 251-276 (1987).

Gonzalo, J.: Five alternative methods of estimating long-run equilibrium relationships. Journal of Econometrics 60, 203-233 (1994).

Granger,W. J., Newbold, P.: Spurious Regression in Econometrics. Journal of Econometrics 2, 111-120 (1974).

Hansen, B.: Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. Journal of Econometrics 53, 87-121 (1992).

Hsiao, C.: Econometrics Analysis of Panel Data. Cambridge University Press, Cambridge 1986.

Haug, A.: Tests for cointegration: A Monte Carlo comparison. Journal of Econometrics 71, 89-115 (1996).

Hall, P., Heyde, C.: Martingale Limit Theory and its Applications. New York Academic Press, New York 1980.

Hannan, E.: Multiple Time Series. Wiley, New York 1970.

Im, K., Pesaran, H., Shin, Y.: Testing for unit roots in heterogeneous panels. Manuscript, Department of Applied Economics, University of Cambridge (1997).

Kao, C., Chen, B.: On the estimation and inference of a cointegrated regression in panel data when the cross-section and time-series dimensions are comparable. Manuscript, Department of Economics, Syracuse University (1995a).

Kao, C., Chen, B.: Residual-based tests for cointegration in panel data when the cross-section and time-series dimensions are comparable. Manuscript, Department of Economics, Syracuse University (1995b).

Kao, C.: Spurious Regression and Residual Based Tests for Cointegration in Panel Data. Journal of Econometrics 90, 1-44 (1999).

Kao, C., Emerson, J.: Testing for Structural Change of a Time Trend Regression in Panel Data. Advances in Econometrics 15, 179-222 (2004).

Kao, C., Trapani, L., Urga, G.: Asymptotics for Panel Models with Common Shocks. Econometric Reviews 31, 390-439 (2010).

Levin, A., Lin, C. F.: Unit root test in panel data: asymptotic and finite sample properties. Working paper, University of California at San Diego, 1992.

Levin, A., Lin, C. F.: Unit root test in panel data: new results. Working Paper, University of California at San Diego, 1993.

Magnus, J., Neudecker, H.: Matrix Differential Calculus. Wiley, New York 1988.

Matyas, L., Sevestre, H.: The Econometrics of Panel Data. Kluwer Academic Publishers, Boston, MA 1992.

McCoskey, S., Kao, C.: A residual-based test of the null of cointegration in panel data. Econometric Reviews 17, 57-84 (1998).

Muirheadr, H.: Aspects of Multivariate Statistical Theory. Wiley, New York 1982.

Ng, S.: A Simple Test for Non-Stationarity in Mixed Panels. Journal of Business and Economics Statistics 26, 113-127 (2008).

Park, J. Y., Phillips, P. C. B.: Statistical Inference in Regressions with Integrated Processes: Part I. Econometric Theory 4, 468-497 (1988).

Park, J. Y., Phillips, P. C. B.: Asymptotics for Nonlinear Transformations of Integrated Time Series. Econometric Theory 15, 269- 298 (1999).

Pedroni, P.: Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. Indiana University Working Papers in Economics, No. 95-013 1995.

Pesaran, H., Smith, R.: Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79-113 (1995).

Pesaran, M. H., Yamagata, T.: Testing Slope Homogeneity in Large Panels. Journal of Econometrics 142, 50–93 (2008).

Phillips, P. C. B.: Understanding Spurious Regression in Econometrics. Journal of Econometrics 13, 311-340 (1986).

Phillips, P. C. B.: Time series regression with a unit root. Econometrica 55, 277-301 (1987).

Phillips, P. C. B., Durlauf, S.: Multiple Time Series Regression with Integrated Processes. Review of Economic Studies 53, 473-495 (1986).

Phillips, P. C. B., Hansen, B.: Statistical Inference in Instrumental Variables Regression with I(1) Processes. Review of Economic Studies 57, 99-125 (1990).

Phillips, P. C. B, Lee, C.: Efficiency Gains from Quasi-Differerencing under Nonstationarity. In: Athens Conference on Applied Probability and Time Series, Volume II, Time Series Analysis in Memory of E. J. Hannan, P.M.Robinson and M. Rosenblatt ed, Springer-Verlag, New York 1996.

Phillips, P. C. B., Moon, H.: Linear Regression Limit Theory for Nonstationary Panel Data. Econometrica 67, 1057-1111 (1999).

Phillips, P. C. B:, Solo, V.: Asymptotics for Linear Processes. Annals of Statistics 20, 971-1001 (1992).

Phillips, P. C. B:, Ouliaris, S.: Asymptotic properties of residual based test for cointegration. Econometrica 58, 165-193 (1990).

Phillips, P. C. B.:, Perron, P.: Testing for unit root in time series regression. Biometrika 75, 335-346 (1988).

Pollard, D.: Convergence of Stochastic Processes. Springer Verlag, New York 1984.

Quah, D.: Exploiting cross section variation for unit root inference in dynamic data. Economics Letters 44, 9-19 (1994).

Robertson, D., Symons, D.: Some Strange Properties of Panel Data Estimators. Journal of Applied Econometrics 7, 175-189 (1992).

Shorack, G.,Wellner, J.: Empirical Processes with Applications to Statistics. Wiley, New York 1986.

Trapani, L.: On the Asymptotic T Test for Large Nonstationary Panel Models. Computational Statistics & Data Analysis 56, 3286-3306 (2012).

Trapani, L., Urga, G.: Optimal Forecasting with Heterogeneous Panels: A MonteCarlo Study. International Journal of Forecasting 25, 567-586 (2009).

Westerlund, J.: New Simple Tests for Panel Cointegration. Econometric Reviews 24, 297-316 (2005).

Westerlund, J., Hess, W.: A New Poolability Test for Cointegrated Panels. Journal of Applied Econometrics 26, 56-88 (2010).

References for Chapters 3, 4, 5

Alexander, C., Chibumba, M.: Multivariate orthogonal factor GARCH. University of Sussex discussion papers in Mathematics,1998.

Adenstedt, R. K.: On large-sample estimation for the mean of a stationary random sequence. Annals of Statistics 2, 1095-1107 (1974).

Aguilar, O., West, M.: Bayesian dynamic factor models and portfolio allocation. Journal of Business and Economics Statistics 18, 338-357 (2000).

Audrino, F., Barone-Adesi, G.: Average conditional correlation and tree structures for multivariate GARCH models. Journal of Forecasting 25, 579-600 (2006).

Asai, M., McAleer, M.: Dynamic asymmetric leverage in stochastic volatility models. Econometric Reviews 24, 317-332 (2005).

Asai, M., McAleer, M.: Asymmetric multivariate stochastic volatility. Econometric Reviews 25, 453-473 (2006).

Asai, M., McAleer, M., Yu, J.: Multivariate Stochastic Volatility: A Review. Econometric Reviews 25, 145-175 (2006).

Avarucci, M., Beutner, E., Zaffaroni, P.: On moments conditions for quasi-maximum likelihood estimation of multivariate ARCH models. Econometric Theory 29, 545-566 (2013).

Baillie, R. T., Bollerlsev, T., Mikkelsen, O.H.: Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 74, 3-30 (1996).

Barndorff-Nielsen, O. E., Shephard, N.: Normal Modified Stable Process. Theory of Probability and Mathematics Statististics 65, 1-19 (2001).

Bauwens, L., Laurent, S.: A new class of multivariate skew densities, with application to GARCH models. Journal of Business and Economic Statistics, 101, 55-75 (2002).

Bauwens, L., Laurent, S., Rombouts, J. V. K.: Multivariate GARCH models:a survey. Journal of Applied Econometrics 21, 79-109 (2006).

Bauwens, L.,Veredas, D.: The stochastic conditional duration model: a latent variable model for the analysis of financial durations. Journal of Econometrics 119, 381-412 (2004).

Bhansali, R.: Asymptotic properties of the Wiener-Kolmogorov predictior. Journal of the Royal Statistical Society Series B, 36, 61-73 (1974).

Billio, M., Caporin, M.: A generalized dynamic conditional correlation model for portfolio risk evaluation. Unpublished manuscript, Ca'Foscari University of Venice, Department of Economics, 2006.

Bollerlsev, T.: Modelling the coherence in short-run nominal exchange rates:a multivariate generalized ARCH model. Review of Economics and Statistics 72, 498-505 (1990).

Bollerlsev, T., Engle, R. F., Nelson, D. B.: ARCH models. In: Handbook of Econometrics, R.F. Engle and D.L. McFadden ed.,vol 4, 2959-3038, Elsevier Science, Amsterdam (1994).

Bollerlsev, T., Engle, R. F., Wooldridge, J.M.: A capital asset pricing model with time-varying covariances. The Journal of Political Economy 96, 116-131 (1998).

Bollerlsev, T., Mikkelsen, H.: Modeling and pricing long memory in stock market volatility. Journal of Econometrics 73, 151-184 (1996).

Bourgerol, P., Picard, N.: Strict stationarity of generalized autoregressive process. Journal of Econometrics, 31, 307-327 (1992).

Brandt, M., Jones, C.: Volatility forecasting with range-based EGARCH models. Journal of Business & Economic Statistics 24, 470-486 (2005).

Breidt, F., Crato, N., DeLima, P.: The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83, 325-348 (1998).

Brockwell, P., Davis, R.: Time series: theory and methods. Springer Verlag, New York 1987.

Cappiello, L., Engle, R. F., Sheppard, K.: Asymmetric dynamics in the correlations of global equity and bond returns. Journal of Financial Econometrics 4, 537-572 (2006).

Cheung, Y. W., Lai, K. S., Lain, M.: Are there long cycles in foreign stock returns? Journal of International Financial Markets 3, 33-47 (1993).

Chib, S., Nadari, F., Shephard, N.: Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics 108, 281-316 (2002).

Chib, S., Omori, Y., Asai, M.: Multivariate Stochastic Volatility. In: Handbook of Financial Time Series, T.G.Andersen and R.A.Davies ed., Springer, New York 2008.

Comte, F., Lieberman, O.: Asymptotic theory for multivariate GARCH processes. Journal of Multivariate Analysis 84, 61-84 (2003).

Corazza, M., Malliaris, A., Nardelli, C.: Searching for fractal structure in agricultural futures markets. Journal of Futures Markets 17, 433-473 (1997).

Dahlhaus, R.: Efficient parameter estimation for self-similar processes. Annals of Statistics 17, 1749-1766 (1989).

Danielsson, J.: Vakue at Risk and Extreme Returns. Tinbergen Institute Discussion Paper 98-017/2, 1998.

Darrat, A., Rahman, S., Zhong, M.: Intraday trading volume and return volatility of the DJIA stocks: a note. Journal of Banking & Finance 27, 2035-2043 (2003).

Davidson, J.: Moments and memory properties of linear conditional heteroskedasticity models, and a new model. Journal of Business and Economic Statistics 22, 16-19 (2004).

Dellaportas, P., Vrontos, I.D.: Modelling volatility asymmetries: a Bayesian analysis of a class of tree structured multivariate GARCH models. Econometrics Journal 10, 503-520 (2007).

Deo, R., Hurvich, C., Lu, Y.: Forecasting realized volatility using a long memory stochastic volatility model: estimation, prediction and seasonal adjustment. Journal of Econometrics 131, 29-58 (2006).

Diebold, F., Cheung, Y.: Maximum Likelihood Estimationof Fractionally Integrated Noise with Unknown Mean. Journal of Econometrics 62, 301-316 (1994).

Duan, J., Gauthier, G., Sasseville, C., Simonato, J.: Approximating the GJR-GARCH and the EGARCH option pricing models analytically. Journal of Computational Finance 9, (2006).

Dunsmuir, W.: A central limit theorem for parameter estimation in stationary vector time series and its applications to models for a signal observed with noise. Annals of Statistics 7, 490-506 (1979).

Dunsmuir,W., Hannan, E.: Vector Linear Time series model. Advances in applied probability 8, 339-364 (1976).

Durbin, J., Koopman, S.J.: Monte Carlo maximum likelihood estimation for non-gaussian state space models. Biometrika 84, 669-684 (1997).

Eldridge, R. M., Bernhardt, C., Mulvey, I.: Evidence of chaos in the S&P cash index. Advances in Futures and Options Research 6, 179-192 (1993).

Engle, R. F.: Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom. Econometrica 50, 987-1007 (1982).

Engle, R. F., Kroner, K. F.: Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122-150 (1995).

Engle, R. F.: Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business and Economic Statistics 20, 339-350 (2002).

Engle, R. F.: Analysis of High Frequency Financial Data. In: Handbook of Financial Econometrics, L. Hansen and Y.Ait-Sahalia ed, North Holland, Amsterdam 2004.

Engle, R. F., Ng, V.K., Rothschild, M.: Asset Pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills. Journal of Econometrics 45, 213-238 (1990).

Engle, R., Russell, J.: Autoregressive conditional duration: a new approach for irregularly spaced transaction data. Econometrica 66, 1127-1162 (1998).

Engle, R., Sheppard, K.: Theoretical and empirical properties of Dynamic Conditional Correlation Multivariate GARCH. Mimeo UCSD, 2001.

Fang, H., Lai, K. S., Lai, M.: Fractal Structure in currency futures price dynamics. Journal of Futures Markets 14, 169-181 (1994).

Fiorentini, G., Sentana, E., Calzolari, G.: On the validity of the Jarque-Bera normality test in conditionally heteroskedastic dynamic regression models. Economics Letters 83, 307-312 (2004).

Feller, W.: An introduction to Probability and its Applications, vols I, II. Wiley, New York 1971.

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

Fox, R., Taqqu, M.: Central limit theorems for quadratic forms in random variables having long-range dependence. Probability Theory and Related Fields 74, 213-240 (1987).

Francq, C., Zakoian, J. M.: GARCH models: structure, statistical inference and financial applications. Wiley, Chichester UK, 2010.

Fuller, W.: Introduction to statistical time series. Wiley, Chichester UK, 1996.

Ghysels, E., Harvey, A., Renault, E.: Stochastic Volatility. In: Handbook of Statistics vol. 14, North Holland, Amsterdam, 1995.

Giraitis, L., Kokoszka, P., Leipus, R.: Stationary ARCH models: dependence structure and central limit theorem. Econometric Theory 1, 3-22 (2000).

Giraitis, L., Koul, H. L., Surgailis, D.: Asymptotic normality of regression estimators with long memory errors. Statistics & Probability Letters 29, 317-335 (1996).

Giraitis, L., Koul, H. L., Surgailis, D.: Large Sample Inference for Long memory Processes, Imperial College Press, 2012.

Giraitis, L., Robinson, P. M.: Whittle estimation of ARCH models. Econometric Theory 17, 608-631 (2001).

Giraitis, L., Surgailis, D.: A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotically normality of Whittle's estimate. Probability Theory and Related Fields 86, 105-129 (1990).

Giraitis, L., Taqqu, M.: Limit Theorems for bivariate Appell polynomials. Part I: Central limit theorems. Probability Theory 107, 359-381 (1997).

Giraitis, L., Taqqu, M.: Whittle Estimation for Finite Variance non Gaussian Time Series with Long Memory. Annals of Statistics 27, 178-203 (1999).

Glosten, L., Jaganathan, R., Runkle, D.: On the relation between the expected value and the volatility of the nominal excess returnson stocks. Journal of Finance 48, 1779-1801(1993).

Gourieroux, C.: ARCH models and Financial Applications. Springer Series in Statistics, Springer New York, 1997.

Gourieroux, C., Jasiak, J.: Financial econometrics: problems, models, and methods. Princeton University Press, Princeton NJ 2001.

Granger, C. W. J.: Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14, 227-238 (1980).

Granger, C. W. J.: Some properties of time series data and their use in econometric model specification. Journal of Econometrics 16, 121-130 (1981).

Granger, C. W. J., Andersen, A.: On the invertibility of time series models. Stochastic Processes and their Applications 8, 87-92 (1978).

Granger, C. W. J., Joyeux, R.: An introduction to long-memory time series models and fractional differencing. Journal of time series analysis 1, 15-29 (1980).

Greene, M. T., Fielitz, B. D.: Long-term dependence in common stock returns. Journal of Financial Economics 4, 339-349 (1977).

Haas, M., Mittnik, S., Paolella, M. S.: A new approach to Markov-switching GARCH models. Journal of Financial Econometrics 2, 493-530 (2008).

Hafner, C. M., Herwartz, H.: Analytical quasi maximum likelihood inference in multivariate volatility models. Econometric Institute Report 21. Erasmus University Rotterdam 2004.

Hannan, E.: Multiple Time Series. Wiley, New York, 1970.

Hannan, E.: The asymptotic theory of linear time series models. Journal of Applied Probability 10, 130-145 (1973).

Hannan, E.: The asymptotic distribution of serial covariances. Annals of Statistics 4, 396-399 (1986).

Harvey, A.: Long memory in stochastic volatility. In: Forecasting volatility in financial markets, J.Knight and S.Satchell ed.,Butterworth-Heineman, London 1998.

Harvey, A., Ruiz, E., Shephard, N.: Multivariate Stochastic Variance Models. Review of Economic Studies 61, 247-264 (1994).

Harvey, A., Shephard, N.: Estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business & Economic Statistics 14, 429-434 (1996).

Haynes, H.,Yung, H., Zhang, Z.: An empirical investigation of the GARCH option pricing model: Hedging performance. Journal of Futures Markets 23, 1191-1207 (2003).

Helms, B. P., Kaen, F. R., Rosenman, R. E.: Memory in Commodity Futures Contracts. Journal of Futures Markets 4, 559-567 (1984).

Heyde, C., Gay, R.: Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. Stochastic processes and their Applications 45, 169-182 (2003).

Hidalgo, J., Yajima, Y.: Prediction and signal extractionof strongly dependent processes in the frequency domain. Econometric Theory 18, 584-624 (2002).

Hosoya, Y.: Estimation problems on stationary time series. Unpublished Ph.D. Dissertation, Yale

University 1974.

Hosoya, Y.: A limit theory fon stationary processes with long-range dependence and related statistical models. Annals of Statistics 25, 105-137 (1997).

Hosoya, Y., Taniguchi, M.: A central limit theorem for stationary processes and the parameter estimation of linear processes. Annals of Statistics 10, 132-153 (1982).

Hualde, H., Robinson, P. M. : Gaussian Pseudo Maximum Likelihood Estimation of Fractional Time Series Models. Annals of Statistics 39, 3152-3181 (2011).

Ibragimov, I., Linnik, J.: Independent and stationary sequences of random variables. Walters-Noordho, Groningen 1971.

Jacobsen, B.: Long term dependence in stock returns. Journal of Empirical Finance 33, 393-417 (1996).

Jacquier, E., Polson, N.G., Rossi, P.: Stochastic volatility: univariate and multivariate extensions. Unpublished manuscript. Boston College, University of Chicago, 1999.

Jasiak, J.: Persistence in intertrade durations. Journal of Finance 19, 166-195 (1998).

Jeantheau, T.: Strong consistency of estimators for multivariate ARCH models. Econometric Theory 14, 70-86 (1998).

Kao, M. A., Ma, C. K.: Memories, Heteroskedasticity, and Price Limit in Currency Futures Markets. Journal of Futures Markets 12, 679-692 (1992).

Kawakatsu, H.: Matrix exponential GARCH. Journal of Econometrics 134, 95-128 (2006).

Kim, S., Shepard, N., Chib, S.: Stochastic volatility: optimal likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361-393 (1998).

Kroner, F.K, Ng, V.K.: Modelling asymmetric comovements of asset returns. Review of Financial Studies 11, 817-844 (1998).

Leonov, V., Shiryaev, A.: On a method of calculation of semi-invariants. Theory of Probability Applications 4, 319-329 (1959).

Ling, S., McAleer, M.: Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory 19, 280-310 (2003).

Lo, A. W.: long-term memory in stock market prices. Econometrica 59, 1279-1313 (1991).

Lucchetti, R.: Analytical Score for Multivariate GARCH models. Computational Economics 19, 133-143 (2002).

Lüktepol, H. Handbook of matrices. John Wiley New York, 1996.

Meddahi, N., Renault, E.: Temporal aggregation of volatility models. Journal of Econometrics 119, 355-379 (2004).

Meyn, S., Tweedie, R.L.: Markov chains and Stochastic Stability. Cambridge University Press, Cambridge 2009.

Nawrocki, D.: R/S analysis and long term dependence in stock market indices. Managerial Finance 21, 78-91 (1995).

Nelson, D.: The time series behaviour of stock market volatility and returns. Unpublished Ph.D. Dissertation, Massachusetts Institute of Technology 1988.

Nelson, D.: Conditional heteroscedasticity in asset pricing: a new approach. Econometrica 59, 347-370 (1991).

Ng, L.: Test of the CAPM with time-varying covariances:a multivariate GARCH approach. Journal of Finance 46, 1507-1521 (1991).

Perez, A., Zaffaroni, P. : Finite-sample properties of Maximum Likelihood and Whittle Estimators in EGARCH and FIEGARCH Models. Quantitative and Qualitative Analysis in Social Science 1, 78-97 (2008).

Pitt, M., Shephard, N.: Time varying covariances: a factor stochastic volatility approach. In: Bernardo, Berger, David, Smith, eds. Bayesian Statistics 6. Oxford University Press, 1999.

Robinson, P. M.: Alternative models for stationary stochastic processes. Stochastic Processes and their Applications 8, 141-152 (1978).

Robinson, P. M.: Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. Journal of Econometrics 47, 67-84 (1991).

Robinson, P. M.: Rates of convergence and optimal bandwidth for long range dependence. Probability Theory and Related Fields 99, 443-473 (1994).

Robinson, P. M.: Semiparametric analysis of long-memory time series. Annals of Statistics 22, 515-539 (1994).

Robinson, P. M., Zaffaroni, P.: Modelling non linearity and long memory in time series. In: Cutler, C., Kaplan, D (Eds.), Nonlinear Dynamics and Time Series, Fields Institute Communications, Vol. 11. American Mathematical Society, Providence, RI, pp. 161–170 (1997).

Robinson, P. M., Zaffaroni, P.: Non linear time series with long memory: a model for stochastic volatility. Journal of Statistical Planning and Inference 68, 359-371 (1998).

Robinson, P. M., Zaffaroni, P.: Pseudo maximum likelihood estimation of ARCH(∞) models. Annals of Statistics 34,1049-1074 (2006).

Ruiz, E., Veiga, H.: Modelling long-memory volatilities with leverage effect: A-LMSV versus a FIEGARCH. Computational Statistics & Data Analysis 6, 2846-2862 (2008).

Sentana, E.: Quadratic ARCH Models. The Review of Economic Studies 4, 639-661 (1995).

Shephard, N.: Statistical aspects of ARCH and stochastic volatility. In: Cox, D. R., Hinkely, D. V., Barndorff-Nielses, O.E., eds. Time series models in Econometrics , Finance and Other Fields. Chapman & Hall, London, 1996.

Shephard, N.: Estimating quadratic variation using realized variance. Journal of Applied Econometrics 17, 457-477 (2002).

Shephard, N.: Stochastic volatility: selected readings. Oxford University Press, Oxford 2004.

Shephard, N., Pitt, M. K.,: Likelihood analysis of non-gaussian measurement time series. Biometrika 84, 653-667 (1997).

Silvennoinen, A., Terasvirta. T.: Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations. SSE/EFI Working Paper series in Economic and Finance No. 652 (2005).

Straumann, D., Mikosch, T.: Quasi-maximum likelihood estimation in heteroskedastic time series: a stochastic recurrence equations approach. Annals of Statistics 34, 2449-2495 (2006).

Taniguchi, M.: On estimation of Parameters of Gaussian Stationary Processes. Journal of Applied Probability 16, 575-591 (1979).

Taniguchi, M.: On estimation of the integrals of the fourth order cumulant spectral density. Biometrika 69, 117-122 (1982).

Taylor, S.: Modelling Financial Time Series. Wiley, Chichester UK, 1986.

Tse, K.: A test for constant correlations in a multivariate GARCH model. Journal of Econometrics 98, 107-127 (2000).

Tse, K., Tsui, K.: A note on diagnosing multivariate conditional heteroskedasticity models. Journal of Time Series Analysis 20, 679-691 (1999).

Van Der Vaart, A.: Asymptotic Statistics. Cambridge University Press, Cambridge UK, 1998.

Van Der Weider, R.: GO-GARCH: A multivariate generalized orthogonal GARCH model. Journal of Applied Econometrics 17, 549-564 (2002).

Velasco, C., Robinson, P. M.: Whittle pseudo-maximum likelihood estimation for nonstationary time series. Journal of the American Statistical Association 95, 1229-1243 (2000).

Vrontos, I. D., Dellaportas, P., Politis, D. N.: A full factor multivariate GARCH model. Econometrics Journal 6, 312-334 (2003).

Whittle, P.: Gaussian estimation in stationary time series. Bulletin of the International Statistical Institute 39, 105-129 (1962).

Whittle, P.: Prediction and regulation by linear least-squares method. Basil Blackwell, Oxford, 1983.

Wright, J.: Log-periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns. International Finance Discussion Paper 685, Board of Governors of the Federal Reserve System (2000).

Wright, J., Bollerslev, T.: Volatility forecasting, high-frequency data, and frequency domain inference. Review of Economics and Statistics 83, 596-602 (2001).

Yajima, Y.: On estimation of long-memory time series models. Australian Journal of Statistics 27, 303-320 (1985).

Yong, C. H.: Asymptotic Behaviour of Trigonometric Series.Chinese University of Hong Kong, Hong Kong, 1974.

Zaffaroni, P.: Gaussian inference on certain long-range dependent volatility models. Journal of Econometrics 115, 199-258 (2003).

Zaffaroni, P.: Stationarity and memory of ARCH(∞) models. Econometric Theory 20, 147-160 (2004).

Zaffaroni, P.: Whittle estimation of exponential volatility models. Journal of Econometrics 151, 190-200 (2009).

Zygmund, A.: Trigonometric Series, vols I, II. Cambridge University Press, Cambridge 1979.