https://economix.fr/

Local Whittle Analysis of Stationary Unbalanced Fractional Cointegration Systems

Document de Travail
Working Paper
2019-15

Florent Dubois
Elena Ivona Dumitrescu
Gilles de Truchis

UMR 7235

EconomiX - UMR7235
Université Paris Nanterre
Bâtiment G - Maurice Allais, 200, Avenue de la République
92001 Nanterre cedex

Email : secretariat@economix.fr

Local Whittle Analysis of Stationary Unbalanced Fractional Cointegration Systems

Gilles de Truchis[1,*], Florent Dubois[1], Elena-Ivona Dumitrescu[1]

Preliminary draft
July 2019

Abstract

In this paper we propose a local Whittle estimator of stationary bivariate unbalanced fractional cointegration systems. Unbalanced cointegration refers to the situation where the observables have different integration orders, but their filtered versions have equal integration orders and are cointegrated in the usual sense. Based on the frequency domain representation of the unbalanced version of Phillips' triangular system, we develop a semiparametric approach to jointly estimate the unbalance parameter, the long run coefficient, and the integration orders of the regressand and cointegrating errors. The paper establishes the consistency and asymptotic normality of this estimator. We find a peculiar rate of convergence for the unbalance estimator (possibly faster than $\sqrt{n}$) and a singular joint limiting distribution of the unbalance and long-run coefficients. Its good finite-sample properties are emphasized through Monte Carlo experiments. We illustrate the relevance of the developed estimator for financial data in an empirical application on the information flowing between the crude oil spot and CME-NYMEX markets.

Keywords: Unbalanced cointegration, Long memory, Stationarity, Local Whittle likelihood
JEL: C22, G10

1. Introduction

This paper addresses the estimation of a general class of models known as unbalanced cointegration systems, that encompasses the well known Phillips' triangular cointegration system. In his seminal paper of 1981, Granger establishes that two time series y_t and x_t share a common stochastic trend if (i) y_t and x_t are both integrated of order δ_2, hereafter $I(\delta_2)$ and (ii) there exists a non-null scalar β so that $e_t = y_t - \beta x_t \sim I(\delta_1)$ and $\delta_2 - \delta_1 > 0$. Engle and Granger (1987) have primarily investigated an estimation procedure in the particular case where observables are unit root processes, i.e. $I(1)$, and a linear combination of them has short memory, i.e. $I(0)$. However, the cointegration theory introduced by Granger (1981) does not constrain integration orders to be integers and is currently known as fractional cointegration. Numerous inference procedures for triangular representations of fractionally cointegrated

*Corresponding author
**We are grateful to Karim Abadir, Guillaume Chevillon, Sébastien Laurent, Anders Rahbek and the participants at the Workshop in Financial Time Series (CREST, December 2018) and the Quantitative Finance and Financial Econometrics conference (Aix-Marseille School of Economics, June 2019) for helpful comments and discussions.
Email addresses: detruchis.gilles@parisnanterre.fr (Gilles de Truchis), florent.dubois@parisnanterre.fr (Florent Dubois), elena.dumitrescu@parisnanterre.fr (Elena-Ivona Dumitrescu)
[1]EconomiX-CNRS, University of Paris Nanterre, 200 Avenue de la République, 92000 Nanterre, France.

systems have been developed to date. In a pioneer paper, Robinson (1994) discusses the estimation of the long run coefficient β. Further developments by Robinson and Marinucci (2003), Robinson and Hualde (2003), Nielsen (2005) and Christensen and Nielsen (2006), among others, account for the unknown nature of integration orders δ_2 and δ_1 in estimating β. To increase the efficiency of the estimators, a more recent strand of the literature focuses on joint estimation methods of all parameters describing the system (see e.g. Nielsen 2007, Hualde and Robinson 2007, Robinson 2008a, Hualde and Robinson 2010, Shimotsu 2012). At the same time, a parallel strand of the literature has been gauging a fundamental requirement of cointegration theory, i.e. the equality of integration orders of the observables y_t and x_t (see, e.g. Robinson and Yajima 2002, Nielsen and Shimotsu 2007, Hualde 2013, for some theoretical contributions on this topic).

Hualde (2006) discusses the consequences of relaxing this hypothesis. On the one hand, the author argues that even if pretests cannot statistically reject the equality of integration orders of the observables, their true (unobserved) values could, in certain circumstances, not be strictly equal but very close to each other. This could be qualified as a "near-fractional cointegration" situation where the differences in integration orders of the processes tend to disappear as the sample size increases. In this case, standard estimation methods remain asymptotically valid. On the other hand, if the difference between integration orders does not vanish asymptotically, the relation between the observables cannot be immediately captured by a cointegration structure although the variables are intrinsically linked. Let the two observables of such an unbalanced triangular system, y_t and x_t, be integrated of orders δ_2 and $\delta_2 + \xi$ respectively. One can say that unbalanced cointegration occurs between y_t and x_t and equivalently cointegration theory, in the usual sense, applies between y_t and $(1-L)^{\xi} x_t = x_t(\xi)$, if there exists a linear combination of the two variables which has less memory. Such a hidden long-run equilibrium relationship is empirically relevant for the same reasons as conventional fractional cointegration. Depending on the power of equality of integration orders tests, it is likely that in some cases one still estimates a spurious cointegration. In this context it is important to account for the unbalance as otherwise the least-square-type estimates of β are not consistent. They converge to 0 or diverge to infinity depending on the sign of ξ (see e.g. Robinson and Marinucci 2001).

The main question arising in this framework is how to estimate the unbalance parameter ξ. In his seminal paper, Hualde (2006) proposes a multistep time domain estimator. Nevertheless, a high enough rate of convergence for the first step estimation (integration order parameters) is a requirement for the consistency of β, but it is difficult to achieve in practice for weak cointegration ($\delta_2 - \delta_1 < 1/2$). An alternative method has been proposed by Hualde (2014) and consists in a joint nonlinear least squares estimation of the long run and unbalance parameters. The main advantage of this method is that it does not require the estimation of the integration orders, so that the estimators are more efficient. Overall, these two papers assume that the cointegration system is nonstationary, which leads to non-standard limiting distribution results. In particular, the asymptotic distribution obtained in Hualde (2014) is a functional of a modified type II fractional Brownian motion. Hualde (2014) also conjectures that his approach cannot cover the case where both observables are asymptotically stationary because his estimator should not retain consistency in this case. However, investigating stationary cointegration is fundamental from a theoretical point of view because spurious regression can occur even when y_t and x_t are stationary, as long as their integration orders sum up to a value greater than $1/2$ (see Tsay and Chung 2000). Besides, stationary triangular systems are empirically relevant and attractive mainly for financial data, as trading volume (Lobato and Velasco 2000), return volatility (Andersen and Bollerslev 1997) and electricity spot prices (Haldrup and Nielsen 2006) have integration orders in the stationary region $(0, 1/2)$. The only estimator available for stationary unbalanced triangular systems has been recently proposed by de Truchis and Dumitrescu (2019). Based on a non-linear extension of the narrow-band weighted least-squares estimator of Nielsen (2005), their approach is designed to jointly estimate the long run and unbalance parameters, β and ξ. Besides, it can also be seen as a stationary frequency-domain alternative to Hualde (2014).[2] Although the joint asymptotic distribution obtained by de Truchis and Dumitrescu (2019) is standard Gaussian and allows for simple inference relatively to that of Hualde (2014), it still depends on unknown integration orders, δ_1 and δ_2.

This paper proposes a joint semiparametric estimator of all parameters in bivariate stationary unbalanced fractional cointegration systems, i.e. long memory, long-run and unbalance parameters, which is expected to be more efficient than existing ones. We show that our local Whittle-type estimator is consistent and its limit distribution is Gaussian with block diagonal covariance matrix. Interestingly, the joint asymptotic distribution of $\hat{\beta}$ and $\hat{\xi}$ is singular, as a consequence of the linearization (in frequency-domain) of a non-linear problem that occurs in the presence of unbalance, as discussed by Hualde (2014) in time domain. Similarly to de Truchis and Dumitrescu (2019) we find that the cointegration strength affects not only the convergence rate of the long run estimator $\hat{\beta}$ but also that of the unbalance estimator. Indeed, $\hat{\xi}$ can be faster than $\sqrt{n}$-consistent although the maximum rate of semi-parametric long memory estimators is $\sqrt{m}$. But our joint estimator is more efficient than the one proposed by de Truchis and Dumitrescu (2019) and any possible extension to stationary observables of the multi-step approach developed by Hualde (2006) in non-stationary cases. Besides, as all parameters are jointly estimated in our framework, inference is straightforward. The paper can hence be seen as an extension of Nielsen (2007) to unbalanced systems and also as an improvement to the approach of de Truchis and Dumitrescu (2019) that focuses only on the estimation of β and ξ.

We investigate the finite sample properties of our estimator by means of Monte Carlo experiments for a wide range of specifications. The bias, variance and root mean squared error criteria indicate that our

[2] An alternative parametric approach in time domain for unbalanced fractional cointegrated VAR models is currently investigated by Johansen and Nielsen (2019) as an extension of Johansen and Nielsen (2012).

estimator performs very well in small samples. At the same time, we show that if one wrongly applies Nielsen (2007)'s (balanced) cointegration estimator when integration orders are actually different, he/she could draw a misleading conclusion like the absence of cointegration.

In an empirical illustration, we analyze the information flowing between the CME-NYMEX futures and the crude oil spot markets. In the spirit of Rossi and Santucci de Magistris (2013), we rely on the no-arbitrage condition to formulate the usual relationship between prices in terms of their underlying volatilities. The latter seem to exhibit stationary long memory with different integration orders, hence fitting perfectly our theoretical framework. The empirical results confirm the presence of unbalanced stationary fractional cointegration between spot and futures volatilities. Most importantly, when the futures maturity increases, the cointegration strength clearly decreases, suggesting that the information flowing mechanism between the spot market and the futures markets associated to long maturities is less efficient.

The rest of the paper is organized as follows. In Section 2 we introduce our bivariate stationary model for unbalanced cointegration, while in Section 3 we develop the joint local Whittle estimator. The consistency and the asymptotic normality of the proposed estimator are discussed in Section 4. Section 5 presents the results of the Monte Carlo studies. An empirical application is proposed in Section 6 and then, finally, we conclude. All proofs are gathered in Appendix A and B.

2. A stationary model of unbalanced cointegration

Let y_t and x_t be two unbalanced observable variables with unknown real integration orders, δ_2 and $\delta_2 + \xi_n$ respectively. They are weakly unbalanced when δ_2 and $\delta_2 + \xi_n$ do not diverge at infinity (i.e. $\xi_n \to 0$ as $n \to \infty$) and strongly unbalanced when $\xi_n = |\xi| > 0$ as $n \to \infty$ (see Hualde 2006). As our estimation procedure applies to both cases, without loss of generality, we simplify notation by using ξ to denote ξ_n. Each of these stochastic processes has stationary long memory, i.e. $\delta = \{\delta_2, \delta_2 + \xi\}$ with $\delta \in (0, 1/2)$, if its spectral density $f(\lambda)$ satisfies $f(\lambda) \sim g\lambda^{-2\delta}$ as $\lambda \to 0^+$, where $0 < g < \infty$ and "$\sim$" means that the ratio of the left and right sides converges to 1 in the limit. When $\delta = 0$ the process has short memory, while it is said to have intermediate memory when $\delta \in (-1/2, 0)$. Besides, by applying the appropriate fractional difference filter $(1-L)^\delta = \sum_{k=0}^{\infty} a_k(\delta) L^k$ with $a_k(\delta) := \Gamma(k-\delta)(\Gamma(-\delta)k!)^{-1}$ to the initial series one retrieves $I(0)$ series, e.g. $y_t(\delta_2) \sim I(0)$ and $x_t(\delta_2 + \xi) \sim I(0)$.

Now we define an unbalanced bivariate form of the triangular system introduced in Phillips (1991)

$$
\begin{aligned}
y_t &= \beta x_t(\xi) + u_{1t}(-\delta_1), \\
x_t &= u_{2t}(-\delta_2 - \xi), \qquad t = 1, 2, ..., n. \qquad (1)
\end{aligned}
$$

Although standard cointegration theory does not apply to y_t and x_t, it does to y_t and $x_t(\xi)$ when $\delta_1 < \delta_2$ and $\beta \neq 0$. Thereby, the system in (1) is a cointegration system in the sense that both series have a dominant common component with memory δ_2 that can be suitably recovered by filtering $x_t \sim I(\delta_2 + \xi)$ to obtain $x_t(\xi) \sim I(\delta_2)$. This system is very general as it encompasses various representations already analyzed in the literature, as discussed in the introduction. In particular, when $\beta \neq 0$ and $\xi = 0$ cointegration can arise in the usual sense (see e.g. Robinson and Marinucci 2003), while for $\xi \neq 0$ and $\delta_2 > 1/2$ unbalanced cointegration of non-stationary variables arises (see Hualde 2014).

In this paper we focus on the case where $\xi \neq 0$ and $\delta_2 + |\xi| < 1/2$, which corresponds to unbalanced cointegration of stationary variables, and which remains unexplored in the literature as far as we are aware of, although this case is expected to arise very often in particular with financial series (e.g. volatility, liquidity, trading volume, etc.). More formally, we work under the following mild assumptions.

Assumption 1. *y_t, x_t and $y_t - \beta x_t(\xi)$ are covariance stationary processes integrated of orders δ_2, $\delta_2 + \xi$ and δ_1 respectively with $\beta \neq 0$, and satisfying*

$$0 \leq \delta_1 < \delta_2 < \delta_2 + |\xi| < 1/2,$$

where $|\xi| < k$, with k an arbitrary real number that is small compared to δ_2.

Under Assumption 1, anti-persistent processes are left out because they clearly have limited economic relevance. Notice that $\delta_2 > \delta_1$ as otherwise β cannot be identified and $\beta \neq 0$ to ensure the identification of ξ. It follows that $z_t = (y_t - \beta x_t(\xi), x_t)'$ possesses a spectral density, $f_z(\lambda_j)$, where λ_j denotes the Fourier frequencies, $\lambda_j = 2\pi j/n$, with $j = 1, \ldots, m$ and $m = o(n)$ is the bandwidth parameter (see also de Truchis and Dumitrescu 2019).

Assumption 2. *$u_t = (u_{1t}, u_{2t})'$ has spectral density $f_u(\lambda_j)$ satisfying $f_u(\lambda_j) \sim G(1 + O(\lambda^2))$ in the neighborhood of the origin, where G is a real, symmetric, finite and positive definite matrix.*

Under Assumption 2, u_t can be a vector ARMA process or any other bivariate short memory process with a Wold representation, $u_t = C(L)\varepsilon_t$, where ε_t are further defined as martingale difference innovations and $C(L)$ is an absolutely-summable causal matrix filter satisfying $G = C(1)C(1)'(2\pi)^{-1}$. Besides, denoting $\tilde{u} = (\beta u_{2t} + u_{1t}(\delta_2 - \delta_1) \quad u_{2t})'$ and $f_{\tilde{u}}(\lambda)$ its spectral density, one observes that

$$f_{\tilde{u}}(\lambda) = \Big((\beta \quad 1)' G_{22} (\beta \quad 1)\Big)(1 + O(\lambda^{\delta_2 - \delta_1})) \sim \tilde{G}, \ \ \lambda \to 0^+$$

and that $\tilde{G}$ has reduced rank as long as $\delta_2 - \delta_1 > 0$ whether or not $\xi \neq 0$ (see Hualde 2006).

Under these Assumptions, we can avoid a parametric treatment of $f_z(\lambda)$ by relying on a local power law representation. Indeed, as interest lies in the long-run dynamics of the system, we specify the spectral density only locally around the zero frequency

$$f_z(\lambda) \sim \Big(\Lambda(\lambda;\vartheta)\Big)^{-1} G \Big(\Lambda(\lambda;\vartheta)^*\Big)^{-1}, \quad \Lambda(\lambda;\vartheta) = \text{diag}\left(\lambda^{\delta_1}, \lambda^{\delta_2+\xi}\right), \quad \text{as } \lambda \to 0^+ \tag{2}$$

where $\vartheta = (\delta_1, \delta_2 + \xi)'$ and the superscript "$*$" denotes the conjugate transpose. As Nielsen (2007) and de Truchis and Dumitrescu (2019), we assume that G is diagonal so that u_{1t} and u_{2t} are incoherent in the vicinity of the origin.[3] In contrast to Hualde (2014), no assumption is made with respect to the correlation of u_{1t} and u_{2t} away from the origin. Equation (2) further allows us to derive (in the next section) a semiparametric estimator robust to misspecification of the short-run dynamics.

3. Local Whittle estimation

In this section we introduce a joint local Whittle estimator of $\theta = (\delta_1, \delta_2, \beta, \xi)'$. Let I_z be the periodogram matrix of z_t defined as $I_z(\lambda_j;\beta,\xi) = w_z(\lambda_j;\beta,\xi)w_z(\lambda_j;\beta,\xi)^*$ with $j = 1,\ldots,n$ and $w_z(\lambda_j;\beta,\xi) = (2\pi n)^{-1/2}\sum_{t=1}^n z_t e^{it\lambda_j}$ the Fourier transform of z_t. Using only Fourier frequencies in the neighborhood of the origin,

$$I_z(\lambda_j;\beta,\xi) = \begin{pmatrix} w_y(\lambda_j) - \beta\lambda_j^{\xi} w_x(\lambda_j) \\ w_x(\lambda_j) \end{pmatrix} \begin{pmatrix} w_y(\lambda_j) - \beta\lambda_j^{\xi} w_x(\lambda_j) \\ w_x(\lambda_j) \end{pmatrix}^*, \tag{3}$$

with $j = 1,\ldots,m$, for a fixed bandwidth $m = o(n)$. Note that the presence of λ_j^{ξ} corrects for the fact that the long memory parameters of y_t and x_t are unbalanced.

Then, the discrete local Whittle approximation to the likelihood is given by

$$\begin{aligned} Q_m(\theta, G) = m^{-1}\sum_{j=1}^m \Big[&\log\det\left(\left(\Lambda(\lambda_j;\vartheta)\right)^{-1} G \left(\Lambda(\lambda_j;\vartheta)^*\right)^{-1}\right) \\ &+ \text{tr}\left(G^{-1}\Lambda(\lambda_j;\vartheta) I_z(\lambda_j;\beta,\xi)\Lambda(\lambda_j;\vartheta)^*\right)\Big], \end{aligned} \tag{4}$$

where $G \in \Theta_G$, the set of real positive definite 2×2 matrices. The objective function Q_m is minimized over Θ_G by

$$\hat{G}(\theta) = \text{Re}\left(m^{-1}\sum_{j=1}^m \Lambda(\lambda_j;\vartheta) I_z(\lambda_j;\beta,\xi)\Lambda(\lambda_j;\vartheta)^*\right). \tag{5}$$

[3] The phase parameter modeled as $\varphi = (\delta_2 - \delta_1)\pi/2$ in Robinson (2008a) and Shimotsu (2012) is null in our framework (see also Shimotsu 2007). The presence of non-null off-diagonal elements in G should imply non-negligible imaginary part of the cross-spectrum element $f_z^{ab}(\lambda)$ such that $f_z^{ab}(\lambda) \sim G_{ab}\lambda^{-\delta_a-\delta_b}e^{i(\pi-\lambda)(\delta_a-\delta_b)/2}$ as $\lambda \to 0^+$, for $a, b = 1, 2$ and where G_{ab} denotes the (a,b)th element of G.

Substituting (5) in (4) leads to the following concentrated likelihood function

$$R_m(\theta) = \log\det\hat{G}(\theta) - \frac{2(\delta_1+\delta_2+\xi)}{m}\sum_{j=1}^{m}\log\lambda_j. \tag{6}$$

The local Whittle estimator of θ satisfies $\hat{\theta} = \arg\min_{\theta\in\Theta} R_m(\theta)$, for $m \in [1, n/2]$ and where the parameter space Θ is a compact subset of $\mathbb{R}^4$ with $\Theta = \Theta_\delta \times \Theta_\beta \times \Theta_\xi$ and $\delta = (\delta_1, \delta_2)'$.

4. Limit theory

To prove the consistency of this local Whittle estimator, we introduce several assumptions fairly similar to those of Shimotsu (2007) and Nielsen (2007). In the following, θ_0 and G^0 will denote the true parameter values of θ and G. Furthermore, let $f_{ab}(\lambda)$ and G^0_{ab} denote the (a,b)th element of $f_z(\lambda)$ and G^0 respectively. Define also $\vartheta_0 = (\delta_{01}, \delta_{02}+\xi_0)'$ and ϑ_{0a} the ath element of ϑ_0.

Assumption 3. *As $\lambda \to 0^+$ the elements of the spectral density $f_z(\lambda)$ satisfy*

$$f_{ab}(\lambda) = G^0_{ab}\lambda^{-\vartheta_{0a}-\vartheta_{0b}} + o(\lambda^{-\vartheta_{0a}-\vartheta_{0b}}), \qquad a, b = \{1, 2\},$$

where matrix G^0 is finite, real, symmetric and $G^0_{ab} = G^0_{ba} = 0$.

The first part of Assumption 3 restates (2) and makes precise the conditions on the matrix G^0 while its second part implies a zero-coherence condition that applies only in the vicinity of the origin. As argued in Nielsen (2007), $G^0_{ab} = G^0_{ba} = 0$ is a less restrictive assumption than the traditional orthogonality condition encountered in the least squares theory. In particular, it allows for the presence of correlation in the errors as we move away from the origin, i.e. they can share a common short- and/or medium-term dynamics. The present estimator might be modified to account for this endogeneity issue in the spirit of Robinson (2008a) and Shimotsu (2012), but this definitely implies a nontrivial extension of our limit theory.

Assumption 4. *The sequence $z_t = (y_t - \beta x_t(\xi), x_t)$ is a linear process defined as*

$$z_t - E(z_t) = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j\varepsilon_{t-j}, \qquad \sum_{j=0}^{\infty}||A_j||^2 < \infty,$$

with $||.||$ the Euclidean norm, so that A_j is a causal square summable matrix filter. Moreover, ε_t satisfies, almost surely, $\mathbb{E}(\varepsilon_t|F_{t-1}) = 0$ and $\mathbb{E}(\varepsilon_t\varepsilon_t'|F_{t-1}) = I_2$, with F_t a σ-field generated by $\{\varepsilon_s, s \le t\}$ and there exists a random variable ε such that $\mathbb{E}(\varepsilon^2) < \infty$ and for all $\eta > 0$ and some constant $K > 0$, $Pr(||\varepsilon_t||^2 > \eta) \le KPr(\varepsilon^2 > \eta)$.

Assumption 5. *In a neighborhood of the origin, $A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$ is differentiable and*

$$\frac{\partial}{\partial\lambda}A_{a\cdot}(\lambda) = O(\lambda^{-1}||A_{a\cdot}(\lambda)||) \text{ as } \lambda \to 0^+$$

where $A_{a.}(\lambda)$ is the a-th row of $A(\lambda)$.

Assumptions 4 and 5 follow Nielsen (2007). The former imposes uniformly square integrable martingale-difference innovations with constant conditional variance, while the latter implies $\partial A_{a.}(\lambda)/\partial\lambda = O(\lambda^{-\vartheta_a - 1})$ by the Cauchy inequality

$$||A_{a.}(\lambda)|| \leq (A_{a.}(\lambda)A^*_{a.}(\lambda))^{1/2} = (2\pi f_{aa}(\lambda))^{1/2}.$$

Thereby, under Assumptions 4 and 5 we have $f_z(\lambda) = (2\pi)^{-1}A(\lambda)A(\lambda)^*$.

Assumption 6. *As $n \to \infty$, the bandwidth parameter satisfies*

$$\frac{1}{m} + \frac{m}{n} \to 0.$$

The bandwidth requirement defined in Assumption 6 ensures that m tends to ∞ as $n \to \infty$ but at a slow rate so as to remain in a neighborhood of the origin. Under these assumptions we state the consistency theorem.

Theorem 1. *Let Assumptions 1-6 hold. Define $\nu_0 = \delta_{02} - \delta_{01}$. Then, for $\theta_0 \in \Theta$, as $n \to \infty$,*

$$\begin{pmatrix} \hat{\delta}_1 \\ \hat{\delta}_2 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \delta_{01} \\ \delta_{02} \end{pmatrix}$$
$$\lambda_m^{-\nu_0}(\hat{\beta} - \beta_0) \xrightarrow{p} 0$$
$$\lambda_m^{-\nu_0}\log(\lambda_m)(\hat{\xi} - \xi_0) \xrightarrow{p} 0.$$

For the proof see Appendix A. Theorem 1 shows that the local Whittle estimator, $\hat{\theta} = \arg\min_{\theta\in\Theta} R_m(\theta)$, is consistent. In particular, the convergence rates of $\hat{\beta}$ and $\hat{\xi}$ are driven by the cointegration strength. We hence recover the usual semiparametric rate of convergence of the long-run estimator in a cointegration framework (see Nielsen 2005; 2007, Robinson and Marinucci 2003, Robinson 2008a, Shimotsu 2012). At the same time, the convergence rate of the unbalance estimator is higher.

Now, we introduce some further assumptions in view of proving the asymptotic normality of the estimator. Again, they are similar to those of Shimotsu (2007), Nielsen (2007) and Robinson (2008a).

Assumption 7. *Assumption 3 holds and also satisfies*

$$|f_z^{ab}(\lambda) - G^0_{ab}\lambda^{-\vartheta_{0a}-\vartheta_{0b}}| = O(\lambda^{\alpha-\vartheta_{0a}-\vartheta_{0b}}), \qquad a,b = \{1,2\},$$

as $\lambda \to 0^+$ and for some $\alpha \in (0,2]$.

Assumption 8. *Assumption 4 holds and we further impose that the matrices $\mu_3 = \mathbb{E}(\varepsilon_t \otimes \varepsilon_t\varepsilon_t'|F_{t-1})$ and $\mu_4 = \mathbb{E}(\varepsilon_t\varepsilon_t \otimes \varepsilon_t\varepsilon_t'|F_{t-1})$ are non-stochastic, finite and do not depend on t, with F_t a σ-field generated by $\{\varepsilon_s, s \leq t\}$.*

Assumption 9. *Assumption 5 holds.*

Assumption 10. *As $n \to \infty$, the bandwidth parameter $m = o(n)$ and $\alpha \in (0, 2]$ jointly satisfy*

$$\frac{1}{m} + \frac{m^{1+2\alpha}(\log m)^2}{n^{2\alpha}} \to 0.$$

Under Assumption 10 the bandwidth parameter, m, is theoretically bounded by $n^{4/5}$ but in practice a too small bandwidth increases the variance of the estimator while a too large m generally increases the bias.

Theorem 2. *Under Assumptions 1, 2 and 7-10, as $n \to \infty$,*

$$\sqrt{m}\, \mathrm{diag}\,(I_2) \begin{pmatrix} \hat{\delta}_1 - \delta_{01} \\ \hat{\delta}_{02} - \delta_2 \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(0, E^{-1}),$$

$$\sqrt{m}\lambda_m^{-\nu_0} \begin{pmatrix} 1 \\ \log \lambda_m \end{pmatrix}' \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\xi} - \xi_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 \\ \beta_0^{-1} \end{pmatrix} \mathcal{N}_1(0, F^{-1}).$$

The proof of Theorem 2 is given in Appendix B. Notice that the asymptotic distribution is block-diagonal. The limiting covariance block of $(\hat{\delta}_1, \hat{\delta}_2)'$ is the same as the one found by Lobato (1999). However, the joint limit distribution of $\hat{\beta}$ and $\hat{\xi}$ is singular and in particular that of $\hat{\xi}$ depends on β_0. Indeed, our frequency-domain semi-parametric estimator implies the linearization of a non-linear optimization problem with respect to β and ξ which is behind the observed singularity. This result has already been emphasized in a time domain non-linear least squares framework by Hualde (2014) and in frequency domain by de Truchis and Dumitrescu (2019). The independence between the estimates of δ and $(\beta, \xi)'$ is a direct consequence of the local orthogonality in Assumption 7. This assumption is necessary for deriving the limit distribution in Theorem (2), but one might envisage a non-trivial extension by explicitly accounting for the phase parameter and adjusting the expansion rate of the bandwidth parameter in Assumption 10 in the spirit of Robinson (2008a) and Shimotsu (2012). We conjecture that the limit distribution will still be Gaussian and singular with a non-block diagonal covariance matrix.

Unsurprisingly, since the limit theory of $\hat{\beta}$ does not depend on ξ_0, we recover the standard convergence rate $\sqrt{m}\lambda_m^{-\nu_0}$ (see e.g. Nielsen 2007, Robinson 2008a, Shimotsu 2012). In fact, its convergence rate is always higher than the standard semi-parametric $\sqrt{m}$ rate and is very close to $\sqrt{n}$ when the cointegration strength approaches 1/2. In contrast, the convergence rate of the unbalance estimator is always larger than that of the long run estimator and $\hat{\xi}$ can be superconsistent when the cointegrating gap ν_0 is close to 1/2. This is non-standard result for memory parameters in stationary semi-parametric frameworks where the maximum achievable rate is $\sqrt{m}$ and even for parametric frameworks in time and frequency domains (where the maximum rate is $\sqrt{n}$). A similar result is obtained by Hualde (2014) in a parametric non-stationary time domain framework and de Truchis and Dumitrescu (2019) in frequency domain. Our asymptotic variance is however smaller than the one obtained by de Truchis and Dumitrescu (2019) as all parameters are jointly estimated. Indeed, the ratio of the two, $1 - (1 - 2\nu_0)/(1 - \nu_0)^2$, is always inferior to unity for $\nu_0 \in (0, 1/2)$.

5. Monte Carlo experiments

This section discusses the finite sample performance of the proposed estimator by means of Monte Carlo simulations. We generate the stationary fractionally cointegrated system in (1) by using the circulant embedding method extended to multivariate fractional Gaussian noise by Helgason et al. (2011).[4] The vector u_t is generated from a bivariate normal distribution $\mathcal{N}_2(\mu, \Sigma)$ where the diagonal elements of Σ are set to 1 and its off-diagonal element is $\rho = \{0, 0.4\}$. We fix the long-run coefficient $\beta = 0.8$ and present results for $\delta_2 = 0.35$ and an unbalance coefficient $\xi = 0.1$. We investigated three stationary cointegration cases by setting $\delta_1 = \{0, 0.2, 0.3\}$.[5]

We generate $I = 10000$ replications of this system with sample sizes $n = \{256, 512, 1042, 16384\}$, where the latter should be seen as an approximation of the asymptotic behaviour of the estimator and bandwidth parameter $m = \{\lfloor n^{0.5} \rfloor, \lfloor n^{0.75} \rfloor\}$. For each simulation, we report the bias, the variance and the Root Mean Squared Error (RMSE). To deal with outliers, we follow Shimotsu (2012) and add a penalty term to the objective function, $\Pi(\beta, \tilde{\beta}) = \min(0, \beta - \tilde{\beta} + C)^4 + \max(0, \beta - \tilde{\beta} - C)^4$. Imposing $\beta \in [\tilde{\beta} \pm C]$, this penalty is equivalent to a constrained optimization on β and preserves the asymptotic results obtained in Theorem 2 if $\tilde{\beta}$ is a consistent estimator of β. We set $C = 3$ in all experiments.[6] The initial values $\tilde{\delta}_x$ and $\tilde{\delta}_y$ are obtained from the local Whittle estimator of Robinson (1995) applied to x_t and y_t respectively. Therefore, the initial value $\tilde{\xi}$ is based on the difference between $\tilde{\delta}_y$ and $\tilde{\delta}_x$ (see Hualde 2006, p. 777). Finally, the initial value $\tilde{\beta}_{LSE}$ results from the regression of y_t on $x_t(\tilde{\xi})$ and is a consistent estimator of β. The Narrow-Band Least Squares (NBLS) estimate has also been considered for the initialization of β, but it does not significantly modify the results.

Table 1 displays the bias, variance and RMSE results for $\rho = 0$. The estimates of δ_2, δ_1 and ξ are always quite precise, with bias lower than 0.09. In contrast, the local Whittle estimate for β is sensitive to the cointegration strength $\delta_2 - \delta_1$. As indicated by the econometric theory, the larger the cointegrating gap the better. The finite sample bias and variance of β decrease significantly with the larger sample size. The RMSE is decreasing in the bandwidth m for all parameters, indicating that in absence of short-run dynamics one should use frequencies further away from the origin to reduce both bias and variance.

[4] See Davidson and Hashimzade (2009) for a discussion on the benefits and limitations of existing techniques devoted to simulating type I fractional processes.

[5] We also investigated the case where $\delta_2 = 0.45$, $\xi = -0.1$ and $\delta_1 = \{0, 0.2, 0.3\}$ and we found similar results. They are available upon request.

[6] All simulation results are robust to the choice of C.

Table 1: Simulation results with $\rho = 0$ for $\delta_2 = 0.35$ and $\xi = 0.1$

$m = \lfloor n^{0.5} \rfloor$		256			512			1024			16384		
δ_1	θ	Bias	Variance	RMSE	Bias	Variance	RMSE	Bias	Variance	RMSE	Bias	Variance	RMSE
0	δ_2	-0.052	0.166	0.411	-0.033	0.093	0.307	-0.015	0.049	0.222	0.007	0.003	0.051
	δ_1	-0.078	0.051	0.239	-0.049	0.030	0.180	-0.021	0.017	0.130	0.014	0.003	0.058
	ξ	0.040	0.125	0.356	0.028	0.062	0.251	0.014	0.029	0.171	-0.005	<0.001	0.014
	β	0.043	0.406	0.638	-0.011	0.217	0.465	-0.043	0.108	0.331	-0.039	0.006	0.088
0.2	δ_2	-0.086	0.335	0.585	-0.086	0.281	0.537	-0.078	0.176	0.427	-0.009	0.013	0.116
	δ_1	-0.074	0.051	0.238	-0.051	0.028	0.175	-0.032	0.015	0.128	-0.005	0.003	0.051
	β	0.251	1.117	1.086	0.223	0.932	0.991	0.176	0.678	0.842	-0.006	0.083	0.288
	ξ	0.078	0.312	0.564	0.082	0.261	0.518	0.078	0.158	0.405	0.011	0.009	0.096
0.3	δ_2	-0.081	0.473	0.693	-0.085	0.326	0.578	-0.085	0.235	0.492	-0.076	0.065	0.266
	δ_1	-0.061	0.050	0.233	-0.044	0.027	0.171	-0.029	0.015	0.127	-0.006	0.002	0.050
	β	0.317	1.711	1.346	0.406	1.529	1.302	0.400	1.326	1.219	0.220	0.565	0.783
	ξ	0.074	0.454	0.678	0.080	0.312	0.565	0.086	0.224	0.481	0.080	0.060	0.258
$m = \lfloor n^{0.75} \rfloor$													
0	δ_2	0.004	0.026	0.162	0.008	0.006	0.080	0.008	0.002	0.051	0.002	<0.001	0.015
	δ_1	-0.016	0.006	0.079	-0.007	0.003	0.058	<0.001	0.002	0.042	0.005	<0.001	0.016
	β	-0.064	0.051	0.235	-0.056	0.026	0.171	-0.050	0.014	0.130	-0.016	0.001	0.040
	ξ	-0.015	0.020	0.143	-0.014	0.003	0.060	-0.011	0.001	0.032	-0.002	<0.001	0.007
0.2	δ_2	-0.010	0.212	0.460	-0.015	0.086	0.294	-0.010	0.030	0.173	0.001	<0.001	0.017
	δ_1	-0.024	0.006	0.081	-0.013	0.003	0.057	-0.006	0.002	0.041	-0.001	<0.001	0.013
	β	-0.054	0.307	0.557	-0.061	0.137	0.375	-0.055	0.062	0.254	-0.010	0.005	0.068
	ξ	-0.001	0.205	0.452	0.009	0.081	0.285	0.006	0.027	0.164	<0.001	<0.001	0.009
0.3	δ_2	-0.038	0.339	0.584	-0.064	0.193	0.444	-0.069	0.122	0.356	-0.004	0.002	0.041
	δ_1	-0.024	0.006	0.080	-0.014	0.003	0.057	-0.007	0.002	0.041	-0.001	<0.001	0.013
	β	0.090	0.980	0.994	0.020	0.574	0.758	-0.039	0.259	0.511	-0.015	0.017	0.130
	ξ	0.028	0.333	0.578	0.060	0.187	0.437	0.067	0.118	0.350	0.004	0.001	0.038

Note: The results are based on $I = 10000$ replications.

Table 2 presents the results for $\rho = 0.4$. The violation of the orthogonality condition introduces a bias in the estimation of β and increases its variance when $m = \lfloor n^{0.5} \rfloor$. The bias is larger when $\delta_2 - \delta_1$ is small and it worsens with a larger bandwidth because our data generating process introduces endogeneity at all frequencies. The long-run coherence between the observables and the innovations appears hence to hurt the proposed local Whittle estimator. These results suggest that our estimator is inconsistent in this setup. We do not investigate this issue further and leave it for future research.

Since a typical case of stationary fractional cointegration involves financial variables, in Table 3 we discuss the behavior of our estimator when the normality hypothesis does not hold. For this, we generate the innovations u_t as increments of a Rosenblatt process. The estimates exhibit similar patterns to those in Table 1, with a reduction in bias and variance as the cointegration strength increases. However all parameters exhibit larger bias and variance than under Gaussian innovations. This appears to be a consequence of the lower convergence rate of the Whittle-type estimator in presence of increments of a Rosenblatt process as shown by Bardet and Tudor (2014).

Table 2: Simulation results with $\rho = 0.4$ and $n = 16384$ for $\delta_2 = 0.35$ and $\xi = 0.1$

		$m = \lfloor n^{0.5} \rfloor$			$m = \lfloor n^{0.75} \rfloor$		
δ_1	θ	Bias	Variance	RMSE	Bias	Variance	RMSE
0	δ_2	0.028	0.003	0.064	0.056	<0.001	0.058
	δ_1	0.031	0.007	0.087	0.022	<0.001	0.026
	β	0.012	0.744	0.863	-0.303	0.001	0.305
	ξ	-0.038	0.001	0.044	-0.065	<0.001	0.066
0.2	δ_2	0.038	0.013	0.120	0.076	0.001	0.080
	δ_1	0.006	0.005	0.069	0.002	<0.001	0.013
	β	0.106	1.421	1.197	-0.409	0.005	0.415
	ξ	-0.044	0.010	0.107	-0.078	<0.001	0.080
0.3	δ_2	-0.005	0.092	0.304	0.018	0.021	0.147
	δ_1	-0.002	0.003	0.056	-0.001	<0.001	0.013
	β	0.186	1.638	1.293	-0.421	0.018	0.441
	ξ	0.005	0.090	0.300	-0.017	0.020	0.144

Note: The results are based on $I = 10000$ replications.

Finally, we investigate a situation where unbalanced stationary cointegration is present but a simple balanced cointegration estimator is used, i.e. Nielsen (2007)'s estimator (labeled BFC). [7] Table 4 presents the small and large-sample bias of our estimator (labeled UFC) as well as that associated with BFC. Note that one expects BFC to be inconsistent with respect to β as it assumes equality of integration orders. In particular, in view of Robinson and Marinucci (2001) and de Truchis and Dumitrescu (2019), the bias

[7] We have also envisaged a comparison of our estimator with a stationary version of Hualde (2006, p. 784)'s 3-step estimation approach for unbalanced cointegration. However, his estimator is inconsistent in such a case and to correct for that one would need to first estimate $\sqrt{n}$-consistently the integration order of x_t, $\delta_2 + \xi$, and then jointly estimate δ_1 and ξ (see Hualde 2006, eq. 40). This multi-step estimator is expected to be less efficient than ours and we do not investigate it further.

Table 3: Simulation results with $\rho = 0$ for $\delta_2 = 0.35$ and $\xi = 0.1$ under Rosenblatt distribution

$m = \lfloor n^{0.75} \rfloor$		1024			4096			16384		
δ_1	θ	Bias	Variance	RMSE	Bias	Variance	RMSE	Bias	Variance	RMSE
0	δ_2	-0.102	0.222	0.482	-0.100	0.069	0.281	-0.073	0.006	0.106
	δ_1	0.020	0.003	0.060	0.021	0.001	0.039	0.016	<0.001	0.026
	β	0.071	0.626	0.794	-0.046	0.127	0.360	-0.029	0.036	0.191
	ξ	0.029	0.214	0.464	0.029	0.064	0.254	0.003	0.004	0.065
0.2	δ_2	-0.131	0.248	0.515	-0.135	0.109	0.356	-0.101	0.038	0.22
	δ_1	-0.012	0.005	0.068	-0.008	0.002	0.042	-0.01	0.001	0.028
	β	0.238	1.136	1.092	0.029	0.423	0.651	-0.017	0.099	0.316
	ξ	0.056	0.238	0.491	0.062	0.104	0.328	0.031	0.036	0.192
0.3	δ_2	-0.149	0.369	0.626	-0.173	0.128	0.397	-0.161	0.071	0.311
	δ_1	-0.048	0.005	0.087	-0.041	0.002	0.062	-0.039	0.001	0.048
	β	0.300	1.517	1.268	0.168	0.839	0.931	0.018	0.238	0.488
	ξ	0.076	0.359	0.604	0.103	0.122	0.364	0.091	0.069	0.278

Note: The results are based on $I = 10000$ replications.

on $\hat{\beta}_{BFC}$ should be negative if $\xi > 0$ and positive otherwise. Our simulation results show that $\hat{\beta}_{BFC}$ is biased and this bias does not vanish for large sample sizes while its sign is compatible with the positive unbalance parameter of 0.1. At the same time, $\hat{\delta}_{2,BFC}$ also exhibits an asymptotic bias that corresponds to the (unestimated) unbalance parameter regardless of the cointegration strength, $\delta_2 - \delta_1$. Nielsen's balanced fractional cointegration method is hence estimating $\delta_2 + \xi$, the integration order of x_t, instead of that of y_t, and this pollutes the estimation of δ_1 too when $\delta_2 - \delta_1$ is large. All in all, since the BFC estimation method is not designed to detect hidden long-run relations, it would wrongly conclude to the absence of cointegration. This result reinforces the usefulness of our unbalanced stationary cointegration estimator in empirical applications.

6. Empirical illustration

In this section, we exploit the fact that a linear relationship exists between the spot and futures volatilities of crude oil markets in absence of arbitrage, which has implications in terms of information flow between spot and futures markets. The rational for this spot-futures volatility relationship is provided by Rossi and Santucci de Magistris (2013) who show that the no-arbitrage condition implies an analogous condition on the underlying volatility series of spot and futures asset prices. This original approach is convenient because it can be extended to commodity markets under simple assumptions that are discussed in the following.

Assumption 11. *Weak arbitrage-free hypothesis: arbitrage opportunities can exist but they are infrequent and short-lived.*

Table 4: Bias comparison with $\rho = 0$ for $\delta_2 = 0.35$ and $\xi = 0.1$

$m = \lfloor n^{0.75} \rfloor$		1024		4096		16384	
δ_1	θ	UFC	BFC	UFC	BFC	UFC	BFC
0	δ_2	0.008	0.095	0.005	0.095	0.002	0.092
	δ_1	-0.001	0.023	0.005	0.042	0.005	0.061
	β	-0.049	-0.338	-0.031	-0.373	-0.016	-0.405
	ξ	-0.011		-0.005		-0.002	
0.2	δ_2	-0.012	0.097	$<$0.001	0.099	$<$0.001	0.099
	δ_1	-0.006	$<$0.001	-0.002	0.004	-0.001	0.007
	β	-0.051	-0.444	-0.024	-0.453	-0.011	-0.469
	ξ	0.009		$<$0.001		$<$0.001	
0.3	δ_2	-0.073	0.097	-0.032	0.099	-0.004	0.100
	δ_1	-0.008	-0.005	-0.003	-0.001	-0.001	0.001
	β	-0.039	-0.643	-0.033	-0.624	-0.013	-0.620
	ξ	0.070		0.031		0.004	

Note: The results are based on $I = 10000$ replications.

Accordingly, the spot and futures log-prices should be cointegrated and do not drift too far apart. Under this hypothesis, Rossi and Santucci de Magistris (2013) show that the no-arbitrage condition is directly related to the volatility of the price of a futures contract that expires at time $t+k$ ($\sigma_{t,F}$) and the spot price ($\sigma_{t,S}$) by

$$\sigma_{t,F} = \sigma_{t,S} + b_t + u_t, \qquad b_t = (\log 2)^{-1/2}(r_{\tau_{\max}} - r_{\tau_{\min}}), \qquad t-1 < \tau \leq t \tag{7}$$

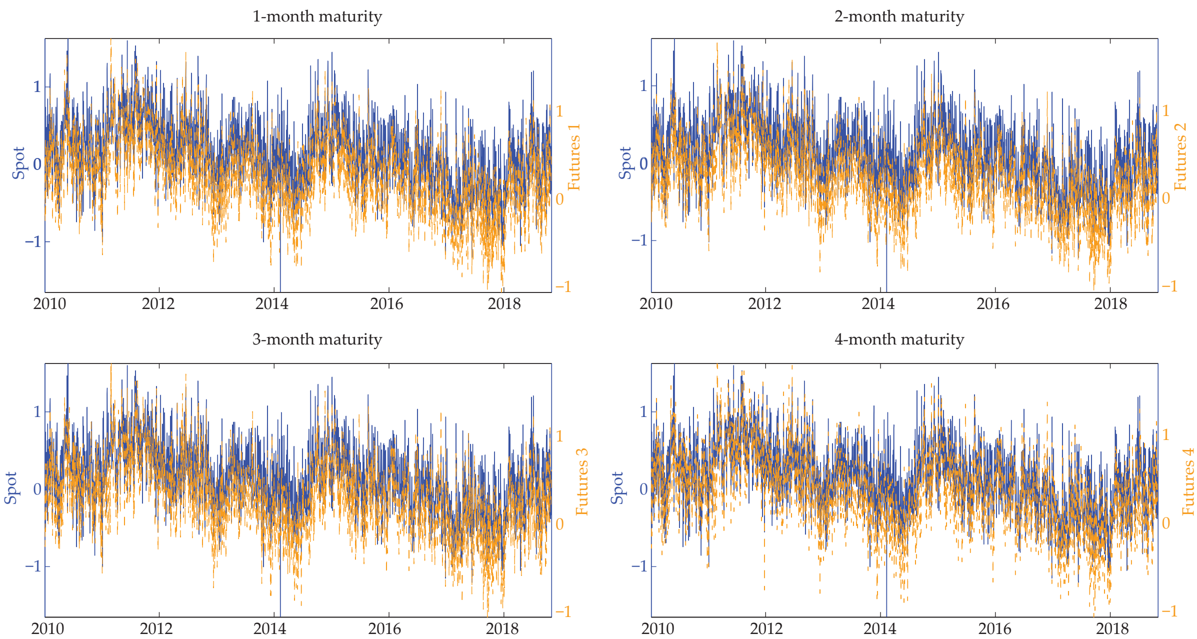
with $r_{\tau_{(.)}}$ the risk-free interest rate of the highest and lowest price in a given day and where the volatility of the risk-free asset over a day is assumed to be small such that $b_t \to 0$. The additional term u_t stands for market frictions and is expected to have zero mean and finite variance. As the persistence of volatility is very well documented in the literature (see e.g. Andersen and Bollerslev 1997, Hurvich et al. 2005, Frederiksen et al. 2012), (7) comes down to a fractional cointegration equation. The presence of long term relationship between spot and futures volatilities appears then as a way to validate (or not) Assumption 11.

The particularity of physical commodity markets consists in the presence of the net convenience yield. As the convenience yield is stochastic and unobserved, there is a vast literature trying to model it and relate it to the no-arbitrage condition. Liu and Tang (2010) show that the non-arbitrage condition holds only if the convenience yield is non-negative. Besides, it is widely assumed in the literature that the convenience yield is homoscedastic. And even when this hypothesis is relaxed, in the particular case of the crude oil market, Liu and Tang (2011) show that the degree of heteroscedasticity of the net convenience yield, i.e. the fraction of the stochastic part in the asymptotic variance, is small (around 9%). These results justify the following hypothesis from an empirical point of view.

Assumption 12. *The convenience yield is non-negative and its volatility over a day is small.*

Under Assumptions 11 and 12, Equation (7) holds for the crude oil markets up to a constant term. Furthermore, it is consistent with the informational theory of Cox (1976) and Ross (1989). Indeed, Ross (1989) shows that in absence of arbitrage opportunity, volatility reflects the information flow. Cox (1976) argues that the transaction costs on the futures market are lower than those of the spot market. Hence, the futures volatility should be the leading factor and the long term no arbitrage equation we estimate becomes $\sigma_{t,S} = \beta\sigma_{t,F}(\xi) + u_t$.[8] In particular, we expect our fractional cointegration model to bring light on the fact that the futures crude oil markets convey information about the future spot market to the current spot market.

Figure 1: Daily range volatility proxy of the crude oil market prices from January 4, 2010 to November 1, 2018.



We use Light-Sweet crude oil spot and futures prices traded in NYMEX obtained through Thomson Reuters Eikon. Our data set runs from January 4, 2010 to November 1, 2018 for a total of $n = 2177$ observations. Besides, to investigate whether the maturity of the futures contracts impacts the spot-futures relationship we consider four different maturities. The contract F_1 specifies the earliest delivery date. It expires on the third business day prior to the 25th calendar day of the month preceding the delivery month. If the 25th calendar day of the month is a non-business day, trading ceases on the third

[8]On stock markets, this hypothesis is supported by the results of Rossi and Santucci de Magistris (2013).

business day prior to the business day preceding the 25th calendar day. The contracts $F_2 - F_4$ represent the successive delivery months following the contract F_1. Figure 1 displays the daily range volatility proxy for the spot and the four futures markets. It appears that the series exhibit long term swings and similarities in their dynamics. In the following we hence thoroughly investigate the presence of stationary (unbalanced) cointegration in this framework.

Since our estimator is semi-parametric, we follow the informal approach by Robinson (2008b) to select the "optimal" bandwidth over a grid of m values. A visual inspection of the sensitivity of parameter estimates to the choice of bandwidth reveals that the optimal bandwidth value lays between $[n^{0.65}, n^{0.75}]$.

Table 5: Unbalanced stationary fractional cointegration analysis

	$m = \lfloor n^{0.65} \rfloor$				$m = \lfloor n^{0.75} \rfloor$			
	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
				UFC				
$\hat{\delta}_2$	0.493	0.494	0.483	0.483	0.417	0.416	0.413	0.408
	(0.041)	(0.041)	(0.041)	(0.041)	(0.028)	(0.028)	(0.028)	(0.028)
$\hat{\delta}_1$	0.089	0.154	0.200	0.238	0.096	0.144	0.175	0.196
	(0.041)	(0.041)	(0.041)	(0.041)	(0.028)	(0.028)	(0.028)	(0.028)
$\hat{\xi}$	-0.029	-0.027	-0.024	-0.025	-0.031	-0.031	-0.029	-0.029
	(0.006)	(0.011)	(0.018)	(0.018)	(0.014)	(0.022)	(0.029)	(0.029)
$\hat{\beta}$	0.854	0.851	0.869	0.845	0.837	0.821	0.819	0.808
	(0.013)	(0.026)	(0.042)	(0.056)	(0.022)	(0.034)	(0.045)	(0.056)
				BFC				
$\hat{\delta}_{2,BFC}$	0.463	0.465	0.459	0.457	0.385	0.384	0.383	0.378
	(0.041)	(0.041)	(0.041)	(0.041)	(0.028)	(0.028)	(0.028)	(0.028)
$\hat{\delta}_{1,BFC}$	0.109	0.160	0.201	0.236	0.109	0.148	0.176	0.196
	(0.041)	(0.041)	(0.041)	(0.041)	(0.028)	(0.028)	(0.028)	(0.028)
$\hat{\beta}_{BFC}$	1.003	1.000	1.014	1.010	1.000	1.000	0.999	1.002
	(0.023)	(0.037)	(0.057)	(0.057)	(0.034)	(0.050)	(0.064)	(0.064)
RY02	3.860	7.910	77.75	-2.880	15.59	22.80	-1461	-12.46

Note : UFC (BFC) stands for unbalanced (balanced) fractional cointegration estimators defined in Section 3 (proposed by Nielsen 2007). Asymptotic standard deviations are displayed in parentheses. RY02 stands for the homogeneity test of integration orders by Robinson and Yajima (2002). The initial values of $\hat{\beta}$ are obtained by OLS. Those of $\hat{\delta}_2$ and $\hat{\delta}_1$ are obtained by the local Whittle estimator of Robinson (1995), while $\hat{\xi}$ is initialized to the difference between the estimated integration orders of $\sigma_{t,F}$ and $\sigma_{t,S}$.

The estimation results of the no arbitrage equation for the different maturities are reported in Table 5 for the bounds of the optimal bandwidth interval.[9] We first discuss the results of our unbalanced cointegration framework (panel UFC). Notice that all volatility series are strongly persistent but remain in the stationary region of the parameter space, i.e. $\hat{\delta}_2 < 1/2$ and $\hat{\delta}_2 + \hat{\xi} < 1/2$. At the same time, the orders of integration of the residuals indicate less persistence albeit they are statistically different from 0. For

[9]Similar results have been found for intermediate values of m and are available upon request.

all these reasons and given that $\hat{\beta}$ is always statistically significant, one can conclude that the fractional cointegration hypothesis holds for all maturities and both bandwidths, which goes along Assumption 11. However, as the values of $\hat{\beta}$ are statistically different and less than 1, the presence of arbitrage opportunities in the long run cannot be neglected. More interestingly, the significance of the unbalance estimator $\hat{\xi}$ points out that unbalanced cointegration occurs in most cases for the first two maturities. The negative sign of $\hat{\xi}$ indicates that the volatility of the spot market exhibits more persistence than that of the futures while the cointegration strength diminishes with the maturity suggesting that the information flowing mechanism becomes less and less efficient. One reason behind this could be the lower level of liquidity of the futures markets at long horizons (see e.g. Darolles et al. 2017). Recall that in Theorem 2 the convergence rates of $\hat{\beta}$ and $\hat{\xi}$ depend on the cointegration strength. It follows that the reported asymptotic standard deviations for these parameters inflate with the maturity, which affects mainly the significance of $\hat{\xi}$ despite roughly constant estimates around -0.025 and -0.030. The last row of the table reports the test-statistics for the homogeneity of integration orders hypothesis proposed by Robinson and Yajima (2002). A statistics larger than a standard Normal critical value is seen as evidence against the null hypothesis irrespective of whether or not there is cointegration. The results confirm the presence of unbalanced integration orders in all cases. Besides, the (unreported) unit rank estimates by Robinson and Yajima (2002) indicate that the spot-futures system is characterized by unbalanced stationary fractional cointegration which goes along the lines of our estimations.

We also investigate the implications of a balanced fractional cointegration (BFC) approach à la Nielsen (2007) on the estimation of this long run relationship (see panel BFC in Table 5). The $\delta_{1,BFC}$ estimates are similar to ours while $\delta_{2,BFC}$ estimates are systematically smaller than the UFC ones, which is consistent with our negative estimates of ξ. What is particularly striking in this framework is that $\hat{\beta}_{BFC}$ is always very close to one although the persistence of the deviations to the long-run equilibrium is increasing with the maturity. Neglecting even small differences in integration orders ($\hat{\xi}$ are small in this illustration) appears hence to have a large effect on the estimate of the long run parameter and induce spurious market efficiency. To prevent against such (wrong) conclusions, we recommend the use of an unbalanced cointegration framework to analyse long run comovements in the spirit of a general to specific approach.

Conclusion

Cointegration estimators for unbalanced triangular systems are particularly useful from an empirical point of view as one can easily find itself in a spurious cointegration framework if integration orders equality tests are not powerful enough. Hualde (2006; 2014) covers the non-stationary region, and the only counterpart for stationary observables (de Truchis and Dumitrescu 2019) focuses on the estimation of β and ξ while neglecting the estimation of the long memory parameters.

In this paper we develop the first joint estimator of all the parameters in bivariate unbalanced stationary triangular fractional cointegration systems. It relies on the local behavior of the spectral density of the system in the vicinity of the origin, thereby allowing for a semi-parametric treatment of the high frequencies. It estimates jointly all the parameters of interest and notably the unbalance parameter, hence achieving greater efficiency than existing competitors. Our local Whittle estimator is consistent and asymptotically normally distributed with block-diagonal covariance matrix under a local orthogonality condition between the regressors and the errors. In particular, the joint limit distribution of the long run coefficient and the unbalance parameter is singular and these parameters exhibit faster rates of convergence than the regular semi-parametric $\sqrt{m}$ rate. By means of Monte Carlo simulations we show the good finite sample properties of the proposed estimator.

In a short application, we use the no-arbitrage hypothesis on the spot and CME-NYMEX futures markets to derive an analogous relation between the spot and futures volatilities. Our results reveal that the apparent unbalance of the integration orders between the spot and futures volatility series is misleading. The estimation of the unbalance parameter allows one to recover a balanced stationary cointegration relationship. The results conclude in favor of an information flowing mechanism between the two markets albeit it is not a fully efficient one. Interestingly, the cointegration strength is higher for short maturities than for long maturities, probably reflecting a reduction in the level of liquidity with the horizon.

Appendix A: Proof of Theorem 1

Proof. Let θ be the vector of admissible parameter values, θ_0 the vector of true parameter values and $S(\theta) = R_m(\theta) - R_m(\theta_0)$. Then, define the neighborhoods $\Theta_\delta^n(d) = \{\delta : \| \delta - \delta_0 \| < d\}$, $\Theta_\xi^n(e) = \{\xi : |\xi - \xi_0| < e\}$, $\Theta_\beta^n(b) = \{\beta : |\beta - \beta_0| < b\}$ and their complements $\Theta_\delta^c = \Theta_\delta \backslash \Theta_\delta^n$, $\Theta_\xi^c = \Theta_\xi \backslash \Theta_\xi^n$ and $\Theta_\beta^c = \Theta_\beta \backslash \Theta_\beta^n$ such that $\Theta^n(\varepsilon) = \Theta_\delta^n(\varepsilon) \times \Theta_\xi^n(\varepsilon^{-1}\lambda_m^{\nu_0}\log(\lambda_m)^{-1}) \times \Theta_\beta^n(\varepsilon^{-1}\lambda_m^{\nu_0})$, $\Theta^c(\varepsilon) = \Theta \backslash \Theta^n(\varepsilon)$ and where $\| \cdot \|$ denotes the Euclidean norm. Without loss of generality with respect to Assumption 1 we set

$$\max\left(\min_i \| \delta_i - \delta_{0i} \|, |\xi - \xi_0|\right) \geq d, \qquad \delta \in \Theta_\delta^c, \qquad \xi \in \Theta_\xi^c, \tag{8}$$

so that $1/2 > d \geq e > 0$. Since $\theta_0 \in \Theta^n(\varepsilon)$, it follows that

$$\Pr\left(\hat{\theta} \in \Theta^c(\varepsilon)\right) = \Pr\left(\inf_{\hat{\theta}\in\Theta^c(\varepsilon)} R_m(\theta) \leq \inf_{\hat{\theta}\in\Theta^n(\varepsilon)} R_m(\theta)\right) \leq \Pr\left(\inf_{\hat{\theta}\in\Theta^c(\varepsilon)} S(\theta) \leq 0\right).$$

Accordingly, to prove Theorem 1 it suffices to show that, as $n \to 0$, $S(\theta)$ is positive and bounded away from 0 uniformly on $\Theta^c(\varepsilon)$ so that

$$\Pr\left(\inf_{\hat{\theta}\in\Theta^c(\varepsilon)} S(\theta) \leq 0\right) \to 0. \tag{9}$$

For this, introduce $\psi_1 = \delta_1 - \delta_{01}$ and $\psi_2 = (\delta_2 + \xi) - (\delta_{02} + \xi_0)$ and develop $S(\theta)$ as

$$\begin{aligned} S(\theta) &= \log\det\hat{G}(\theta) - 2(\delta_1 + \delta_2 + \xi)m^{-1}\sum_{j=1}^m \log\lambda_j - \log\det\hat{G}(\theta_0) + 2(\delta_{01} + \delta_{02} + \xi_0)m^{-1}\sum_{j=1}^m \log\lambda_j \\ &= \log\det\hat{G}(\theta) - \log\det\hat{G}(\theta_0) - m^{-1}\sum_{j=1}^m 2\log\lambda_j \sum_{i=1}^2 \psi_i. \end{aligned}$$

Using that $m^{-1}\sum_{j=1}^m \log\lambda_j = \log\lambda_m + m^{-1}\sum_{j=1}^m \log j - \log m$ and by rearranging $S(\theta)$, we obtain $S(\theta) = S_1(\theta) + S_2(\theta) + S_3(\theta)$, where

$$\begin{aligned} S_1(\theta) &= \log\det\hat{G}(\theta) - \log\det G_0 - 2\log\lambda_m \sum_{i=1}^2 \psi_i + \sum_{i=1}^2 \log(2\psi_i + 1) \\ S_2(\theta) &= \log\det G_0 - \log\det\hat{G}(\theta_0) \\ S_3(\theta) &= 2\sum_{i=1}^2 \psi_i\left(\log m - m^{-1}\sum_{j=1}^m \log j\right) - \sum_{i=1}^2 \log(2\psi_i + 1). \end{aligned}$$

The way we split $S(\theta)$ has the advantage that $S_2(\theta)$ and $S_3(\theta)$ do not depend on β or ξ so that we can treat them by drawing on the works of Robinson (1995). To prove the boundedness of $S(\theta)$ we also make use of the decomposition

$$\Pr\left(\inf_{\hat{\theta}\in\Theta^c(\varepsilon)} S(\theta)\leq 0\right)\leq \Pr\left(\inf_{\{\hat{\theta}\in\Theta^c(\varepsilon)\}\cap\{\hat{\beta}\in\Theta_\beta\}}\Big(S_1(\theta)+S_2(\theta)+S_3(\theta)\Big)\leq 0\right) \tag{10}$$

$$+\Pr\left(\inf_{\{\hat{\theta}\in\Theta^c(\varepsilon)\}\cap\{\hat{\xi}\in\Theta_\xi\}}\Big(S_1(\theta)+S_2(\theta)+S_3(\theta)\Big)\leq 0\right) \tag{11}$$

$$+\Pr\left(\inf_{\{\hat{\theta}\in\Theta^c(\varepsilon)\}\cap\{\hat{\delta}\in\Theta_\delta\}}\Big(S_1(\theta)+S_2(\theta)+S_3(\theta)\Big)\leq 0\right) \tag{12}$$

$$+\Pr\left(\inf_{\{\hat{\theta}\in\Theta^c(\varepsilon)\}\cap\{\hat{\xi}\in\Theta_\xi\,\cup\,\hat{\beta}\in\Theta_\beta\}}\Big(S_1(\theta)+S_2(\theta)+S_3(\theta)\Big)\leq 0\right) \tag{13}$$

$$+\Pr\left(\inf_{\{\hat{\theta}\in\Theta^c(\varepsilon)\}\cap\{\hat{\delta}\in\Theta_\delta\,\cup\,\hat{\beta}\in\Theta_\beta\}}\Big(S_1(\theta)+S_2(\theta)+S_3(\theta)\Big)\leq 0\right) \tag{14}$$

$$+\Pr\left(\inf_{\{\hat{\theta}\in\Theta^c(\varepsilon)\}\cap\{\hat{\delta}\in\Theta_\delta\,\cup\,\hat{\xi}\in\Theta_\xi\}}\Big(S_1(\theta)+S_2(\theta)+S_3(\theta)\Big)\leq 0\right). \tag{15}$$

We first treat $S_2(\theta)$ as it does not depend on $\hat{\theta}$ and no uniform bound is needed. Since $|\log(1+x)|\leq 2|x|$ (see Robinson 1995, p. 1635), it follows that for $\varepsilon\leq 1$

$$\begin{aligned}\Pr\left(|S_2(\theta)|\leq\varepsilon\right)&=\Pr\left(|\log\det\hat{G}(\theta_0)-\log\det G_0|\leq\varepsilon\right)\\&\leq\Pr\left(\left|\frac{\det\hat{G}(\theta_0)-\det G_0}{\det G_0}\right|\leq\varepsilon/2\right).\end{aligned} \tag{16}$$

Accordingly, proving that $\det\hat{G}(\theta_0)-\det G_0\xrightarrow{p}0$ suffices to show that $S_2(\theta)$ is $o_p(1)$. To simplify notation, let $I^0_{jab}=I^0_{ab}(\lambda_j;\beta_0,\xi_0)$ and G^0_{ab} be the (a,b)-th element of G_0 and recall that $\vartheta_0=(\delta_{01},\delta_{02}+\xi_0)'$. Evaluating (3) and (5) at the true value, we obtain

$$I^0_j=\begin{pmatrix} I_{jyy}-2\beta_0\lambda_j^{\xi_0}I_{jxy}+\beta_0^2\lambda_j^{2\xi_0}I_{jxx} & I_{jxy}-\beta_0\lambda_j^{\xi_0}I_{jxx}\\ I_{jxy}-\beta_0\lambda_j^{\xi_0}I_{jxx} & I_{jxx}\end{pmatrix},$$

from which we implicitly take the real part. Then, under Assumptions 3-6

$$\hat{G}_{ab}(\theta_0)-G^0_{ab}=\frac{G^0_{ab}}{m}\sum_{j=1}^{m}\left(\frac{I^0_{jab}}{G^0_{ab}\lambda_j^{-\vartheta_{0a}-\vartheta_{0b}}}-1\right)$$

can be shown to be $o_p(1)$ by the analysis of Robinson (1995, p. 1635). It follows that $S_2(\theta)$ is also $o_p(1)$.

Now we turn to the analysis of $S_3(\theta)$ which can be rearranged as

$$S_3(\theta) = -2\sum_{i=1}^{2}\psi_i\left(m^{-1}\sum_{j=1}^{m}\log j - (\log m - 1)\right) + 2\sum_{i=1}^{2}\psi_i - \sum_{i=1}^{2}\log(2\psi_i + 1). \tag{17}$$

From Lemma 2 of Robinson (1995) we have that the first term of (17) is $m^{-1}\sum_{j=1}^{m}\log j - (\log m - 1) = O(m^{-1}\log m)$ so that the analysis of $S_3(\theta)$ comes down to studying a non-null lower bound of the last two terms in (17) which are of the form $f(x) = x - \log(1+x)$ for ψ_1 and $(x+y) - \log(x+y+1)$ for ψ_2. Because $\inf_x f(x) \geq x^2/6$ and $\inf_{x,y} f(x,y) \geq (x^2+y^2)/6$ for $0 < |x| < 1$, $0 < |y| < 1$, and from the condition stated in Equation (8), we can apply the analysis of Nielsen (2007, p. 437) uniformly over $\{\hat{\delta} \in \Theta_\delta^c\} \cup \{\hat{\xi} \in \Theta_\xi^c\}$. From Lütkepohl (1996, sec. 8.5.2, p. 111) and by the triangular inequality,

$$\sqrt{2}\max(|\delta_1 - \delta_{01}|, |\delta_2 - \delta_{02}|) + \sqrt{2}\max(|0|, |\xi - \xi_0|) \geq \left\|\begin{matrix}\delta_1 - \delta_{01}\\ \delta_2 - \delta_{02}\end{matrix}\right\| + \left\|\begin{matrix}0\\ \xi - \xi_0\end{matrix}\right\| \geq \| \Psi \| \geq d + e,$$

with $\Psi = (\psi_1, \psi_2)'$. Thereby, the infimum of $2\sum_{i=1}^{2}\psi_i - \sum_{i=1}^{2}\log(2\psi_i + 1)$ over $\{\hat{\delta} \in \Theta_\delta^c\} \cup \{\hat{\xi} \in \Theta_\xi^c\} \cup \{\hat{\beta} \in \Theta_\beta\}$ is no less than

$$\frac{2(d+e)}{\sqrt{2}} - \log\left(1 + \frac{2(d+e)}{\sqrt{2}}\right) \geq \frac{2(d^2+e^2)}{6}.$$

Then, given that $f(x,y)$ has a unique minimum on $\{(x,y) : y > -x-1\}$ at $(x,y) = (0,0)$,

$$\inf_{\{\hat{\delta}\in\Theta_\delta^c\}\cup\{\hat{\xi}\in\Theta_\xi^c\}\cup\{\hat{\beta}\in\Theta_\beta\}} S_3(\theta) \geq \frac{2(d^2+e^2)}{6} + O(m^{-1}\log m),$$

$$\inf_{\{\hat{\delta}\in\Theta_\delta^n\}\cup\{\hat{\xi}\in\Theta_\xi^n\}\cup\{\hat{\beta}\in\Theta_\beta\}} S_3(\theta) = o(1).$$

The two remaining cases are $\{\hat{\delta} \in \Theta_\delta^c\} \cup \{\hat{\xi} \in \Theta_\xi^n\} \cup \{\hat{\beta} \in \Theta_\beta\}$ and $\{\hat{\delta} \in \Theta_\delta^n\} \cup \{\hat{\xi} \in \Theta_\xi^c\} \cup \{\hat{\beta} \in \Theta_\beta\}$. In the former case, $S_3(\theta)$ is no less than $2d^2/6 + O(m^{-1}\log m)$ while in the latter, $S_3(\theta)$ is no less than $2e^2/6 + O(m^{-1}\log m)$.

Finally we turn to the analysis of $S_1(\theta)$. It reduces to

$$\begin{aligned}&\log\det\hat{G}(\theta) - 2\psi_1\log\lambda_m - 2\psi_2\log\lambda_m - \log\left(\det G_0(2\psi_1+1)^{-1}(2\psi_2+1)^{-1}\right)\\ &= \log\det\left(V_m\hat{G}(\theta)V_m\right) - \log\left(\det G_0(2\psi_1+1)^{-1}(2\psi_2+1)^{-1}\right),\end{aligned}$$

by (16) and where $V_m = \operatorname{diag}(\lambda_m^{-\psi_1}, \lambda_m^{-\psi_2})$ and $\det\left(V_m\hat{G}(\theta)V_m\right) = \lambda_m^{-2\psi_1-2\psi_2}(\hat{G}_{11}(\theta)\hat{G}_{22}(\theta) - \hat{G}_{12}^2(\theta))$.

Then, the analysis of $S_1(\theta)$ reduces to the study of

$$\lambda_m^{-2\psi_1-2\psi_2}\left(\hat{G}_{11}(\theta)\hat{G}_{22}(\theta)-\hat{G}_{12}^2(\theta)\right)-\left(G_{11}G_{22}-G_{12}^2\right)(2\psi_1+1)^{-1}(2\psi_2+1)^{-1}.$$

Using $\hat{G}_{ab}(\theta)=m^{-1}\sum_{j=1}^m\lambda_j^{\vartheta_a+\vartheta_b}I_{jab}$, one can rewrite $S_1(\theta)$ as $S_{11}(\theta)+S_{12}(\theta)+S_{13}(\theta)$ with

$$\begin{aligned}
S_{11}(\theta)&=\lambda_m^{-2\psi_1-2\psi_2}\left(m^{-1}\sum_{j=1}^m\lambda_j^{2\delta_1}I_{j11}\times m^{-1}\sum_{j=1}^m\lambda_j^{2(\delta_2+\xi)}I_{j22}\right),\\
S_{12}(\theta)&=-\lambda_m^{-2\psi_1-2\psi_2}\left(m^{-1}\sum_{j=1}^m\lambda_j^{\delta_1+\delta_2+\xi}\operatorname{Re}(I_{j12})\right)^2,\\
S_{13}(\theta)&=-\left(G_{11}G_{22}-G_{12}^2\right)(1+2\psi_1)^{-1}(1+2\psi_2)^{-1}.
\end{aligned}$$

Distinguishing the two summations by indexes j and k and then rearranging $S_{11}(\theta)$ and $S_{12}(\theta)$ we find

$$S_{11}(\theta)=\frac{1}{m^2}\sum_{j=1}^m\sum_{k=1}^m\left(\frac{j}{m}\right)^{2\psi_1}\left(\frac{k}{m}\right)^{2\psi_2}\frac{I_{j11}I_{k22}}{\lambda_j^{-2\delta_{01}}\lambda_k^{-2(\delta_{02}+\xi_0)}}$$

and

$$S_{12}(\theta)=-\frac{1}{m^2}\sum_{j=1}^m\sum_{k=1}^m\left(\frac{j}{m}\right)^{\psi_1+\psi_2}\left(\frac{k}{m}\right)^{\psi_1+\psi_2}\frac{\operatorname{Re}(I_{j12})\operatorname{Re}(I_{k12})}{\lambda_j^{-\delta_{01}-\delta_{02}-\xi_0}\lambda_k^{-\delta_{01}-\delta_{02}-\xi_0}}.$$

Then, we correct for the fact that $I_{11}(\lambda)$ and $I_{12}(\lambda)$ are based on estimated cointegration errors as in Nielsen (2007). Denoting by $I_{11}(\lambda)-I_{11}^0(\lambda)$ and $I_{12}(\lambda)-I_{12}^0(\lambda)=(\beta_0\lambda^{\xi_0}-\beta\lambda^{\xi})I_{22}(\lambda)$ the measurement errors from estimating β and ξ, we obtain

$$\begin{aligned}
I_{11}(\lambda)&=I_{11}^0(\lambda)-2\lambda_m^{\nu_0}\lambda_j^{\xi_0}\tilde{\beta}\operatorname{Re}(I_{12}^0(\lambda))+\lambda_m^{2\nu_0}\lambda_j^{2\xi_0}\tilde{\beta}^2I_{22}(\lambda),\\
I_{12}(\lambda)^2&=I_{12}^0(\lambda)^2-2\lambda_m^{\nu_0}\lambda_j^{\xi_0}\tilde{\beta}\operatorname{Re}(I_{12}^0(\lambda))I_{22}(\lambda)+\lambda_m^{2\nu_0}\lambda_j^{2\xi_0}\tilde{\beta}^2I_{22}(\lambda)^2
\end{aligned}$$

with $\lambda_m^{\nu_0}\lambda_j^{\xi_0}\tilde{\beta}=(\beta_0\lambda_j^{\xi_0}-\beta\lambda_j^{\xi})$ and $\nu_0=\delta_{02}-\delta_{01}$. Substituting this in $S_{11}(\theta)$ and $S_{12}(\theta)$ and after rearrangements we have

$$S_{11}(\theta) = \frac{G_{11}G_{22}}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(\frac{j}{m}\right)^{2\psi_1}\left(\frac{k}{m}\right)^{2\psi_2}\frac{I^0_{j11}I_{k22}}{G_{11}\lambda_j^{-2\delta_{01}}G_{22}\lambda_k^{-2(\delta_{02}+\xi_0)}}$$
$$-2\tilde{\beta}\frac{G_{12}G_{22}}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(\frac{j}{m}\right)^{2\psi_1-\nu_0}\left(\frac{k}{m}\right)^{2\psi_2}\frac{\mathrm{Re}(I^0_{j12})I_{k22}}{G_{12}\lambda_j^{-\delta_{01}-\delta_{02}-\xi_0}G_{22}\lambda_k^{-2(\delta_{02}+\xi_0)}}$$
$$+\tilde{\beta}^2\frac{G_{22}^2}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(\frac{j}{m}\right)^{2\psi_1-2\nu_0}\left(\frac{k}{m}\right)^{2\psi_2}\frac{I_{j22}I_{k22}}{G_{22}^2\lambda_j^{-2(\delta_{02}+\xi_0)}\lambda_k^{-2(\delta_{02}+\xi_0)}},$$
$$S_{12}(\theta) = -\frac{G_{12}^2}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(\frac{j}{m}\right)^{\psi_1+\psi_2}\left(\frac{k}{m}\right)^{\psi_1+\psi_2}\frac{\mathrm{Re}(I^0_{j12})\,\mathrm{Re}(I^0_{k12})}{G_{12}^2\lambda_j^{-\delta_{01}-\delta_{02}-\xi_0}\lambda_k^{-\delta_{01}-\delta_{02}-\xi_0}}$$
$$+2\tilde{\beta}\frac{G_{12}G_{22}}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(\frac{j}{m}\right)^{\psi_1+\psi_2}\left(\frac{k}{m}\right)^{\psi_1+\psi_2-\nu_0}\frac{\mathrm{Re}(I^0_{j12})I_{k22}}{G_{12}\lambda_j^{-\delta_{01}-\delta_{02}-\xi_0}G_{22}\lambda_k^{-2(\delta_{02}+\xi_0)}}$$
$$-\tilde{\beta}^2\frac{G_{22}^2}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}\left(\frac{j}{m}\right)^{\psi_1+\psi_2-\nu_0}\left(\frac{k}{m}\right)^{\psi_1+\psi_2-\nu_0}\frac{I_{j22}I_{k22}}{G_{22}^2\lambda_j^{-2(\delta_{02}-\xi_0)}\lambda_k^{-2(\delta_{02}-\xi_0)}}.$$

In the following, we will use the fact that $m^{-1}\sum_{j=1}^{m}(j/m)^{\alpha} = (1+\alpha)^{-1}$. Moreover, by the analysis of Robinson (1995, p. 1636-1638), we have

$$m^{-1}\sum_{j=1}^{m}\left(\frac{\mathrm{Re}(I^0_{jab})}{G_{ab}\lambda_j^{-\vartheta_{0a}-\vartheta_{0b}}}-1\right) = m^{-1}\sum_{j=1}^{m}\mathrm{Re}(I^0_{jab})G_{ab}^{-1}\lambda_j^{\vartheta_{0a}+\vartheta_{0b}} - 1 = o_p(1).$$

Finally, we can rewrite

$$S_1(\theta) = G_{12}^2\left(\int_0^1 x^{2\psi_1}dx\int_0^1 x^{2\psi_2}dx - \left(\int_0^1 x^{\psi_1+\psi_2}\right)^2\right)$$
$$+\tilde{\beta}^2G_{22}^2\left(\int_0^1 x^{2\psi_1-2\nu_0}dx\int_0^1 x^{2\psi_2}dx - \left(\int_0^1 x^{\psi_1+\psi_2-\nu_0}\right)^2\right)$$
$$+2\tilde{\beta}G_{12}G_{22}\left(\int_0^1 x^{\psi_1+\psi_2-\nu_0}dx\int_0^1 x^{\psi_1+\psi_2}dx - \int_0^1 x^{2\psi_1-\nu_0}dx\int_0^1 x^{2\psi_2}dx\right)$$
$$+G_{11}G_{22}\left((1+2\psi_1)^{-1}(1+2\psi_2)^{-1} - (1+2\psi_1)^{-1}(1+2\psi_2)^{-1}\right) + o_p(1).$$

Recall that $G_{12} = 0$ and $\nu_0 > 0$ under cointegration. Then, by the Cauchy-Schwarz inequality, $S_1(\theta)$ is bounded away from zero in all cases but $\{\hat{\theta} \in \Theta^c\} \cap \{\hat{\xi} \in \Theta_{\xi} \cup \hat{\beta} \in \Theta_{\beta}\}$. In the latter, $S_1(\theta) \geq o_p(1)$ whereas $S_3(\theta)$ is bounded away from zero, hence proving Equation (9) in view of (10)-(15) and implicitly Theorem 1.

□

Appendix B: Proof of Theorem 2

As an implication of the consistency Theorem 1, $\hat{\theta}$ satisfies

$$\frac{\partial R_m(\theta)}{\partial \theta}\Big|_{\hat{\theta}} = \frac{\partial R_m(\theta)}{\partial \theta}\Big|_{\theta_0} + \frac{\partial^2 R_m(\theta)}{\partial \theta \partial \theta'}\Big|_{\bar{\theta}}(\hat{\theta}-\theta_0) = 0 \tag{18}$$

where $||\bar{\theta}-\theta_0|| \le ||\hat{\theta}-\theta_0||$. Then, by application of the Cramer-Wold theorem we can show that $\hat{\theta}$ has the stated distribution if

$$\eta' \sqrt{m}\,\mathrm{diag}(I_2, \lambda_m^{\nu_0}, \lambda_m^{\nu_0}\log(\lambda_m)^{-1})\frac{\partial R_m(\theta)}{\partial \theta}\Big|_{\theta_0} \xrightarrow{d} \mathcal{N}(0, \eta'\Omega\eta) \tag{19}$$

and

$$\mathrm{diag}(I_2, \lambda_m^{2\nu_0}, \lambda_m^{2\nu_0}\log(\lambda_m)^{-2})\frac{\partial^2 R_m(\theta)}{\partial \theta \partial \theta'}\Big|_{\bar{\theta}} \xrightarrow{p} \Omega. \tag{20}$$

B.1. Limit of the score

In this section we investigate the limit of the score to prove (19), while (20) will be analyzed in the next section. Note also that the subscript 0, indicating the true parameter values, will be omitted unless its absence causes confusion.

Proof. By the chain rule for matrix derivatives with respect to δ_1 and δ_2 we have

$$\frac{\partial R_m(\theta)}{\partial \delta_a} = \mathrm{tr}\left(\hat{G}(\theta)^{-1}\frac{\partial \hat{G}(\theta)}{\partial \delta_a}\right) - \frac{2}{m}\sum_{j=1}^{m}\log\lambda_j,$$

with $a = \{1, 2\}$ and

$$\frac{\partial \hat{G}(\theta)}{\partial \delta_1} = \frac{\partial \frac{1}{m}\sum_{j=1}^{m}\Lambda_j \mathrm{Re}(I_j)\Lambda_j}{\partial \delta_1} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{\delta_1}\log\lambda_j \mathrm{Re}\begin{pmatrix} 2\lambda_j^{\delta_1} I_{j11} & \lambda_j^{\delta_2+\xi} I_{j12} \\ \lambda_j^{\delta_2+\xi} I_{j21} & 0 \end{pmatrix} \tag{21}$$

$$\frac{\partial \hat{G}(\theta)}{\partial \delta_2} = \frac{\partial \frac{1}{m}\sum_{j=1}^{m}\Lambda_j \mathrm{Re}(I_j)\Lambda_j}{\partial \delta_2} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{\delta_2+\xi}\log\lambda_j \mathrm{Re}\begin{pmatrix} 0 & \lambda_j^{\delta_1} I_{j12} \\ \lambda_j^{\delta_1} I_{j21} & 2\lambda_j^{\delta_2+\xi} I_{j22} \end{pmatrix} \tag{22}$$

finally yielding

$$\frac{\partial R_m(\theta)}{\partial \delta_1} = \frac{2}{m}\sum_{j=1}^{m}\nu_j \mathrm{Re}\left(\hat{G}_{1.}(\theta)^{-1}(\Lambda_j I_j \Lambda_j)_{.1} - 1\right) \tag{23}$$

$$\frac{\partial R_m(\theta)}{\partial \delta_2} = \frac{2}{m}\sum_{j=1}^{m}\nu_j \mathrm{Re}\left(\hat{G}_{2.}(\theta)^{-1}(\Lambda_j I_j \Lambda_j)_{.2} - 1\right) \tag{24}$$

where $\nu_j = \log \lambda_j - m^{-1} \sum_{j=1}^{m} \log \lambda_j$ and $\hat{G}_{a.}(\theta)^{-1}$ denotes the a-th row of the matrix $\hat{G}(\theta)^{-1}$ where Assumption 7 holds. We also have that

$$\begin{aligned}\frac{\partial R_m(\theta)}{\partial \beta} &= \frac{1}{m}\sum_{j=1}^{m} \text{tr Re}\left(\hat{G}(\theta)^{-1}\begin{pmatrix} 2\lambda_j^{2\delta_1}(\beta\lambda_j^{2\xi} I_{jxx} - \lambda_j^{\xi} I_{jxy}) & -\lambda_j^{\delta_1+\delta_2+2\xi} I_{jxx} \\ -\lambda_j^{\delta_1+\delta_2+2\xi} I_{jxx} & 0 \end{pmatrix}\right) \\ &= -\frac{2}{m}\sum_{j=1}^{m} \lambda_j^{-\nu_0} \text{Re}\left(\hat{G}_{1.}(\theta)^{-1}(\Lambda_j I_j \Lambda_j)_{.2}\right).\end{aligned}$$

Finally, the last derivative is

$$\frac{\partial R_m(\theta)}{\partial \xi} = \text{tr Re}\left(\hat{G}(\theta)^{-1}\left(\frac{\partial \tilde{G}(\theta)}{\partial \beta} + \frac{\partial \hat{G}(\theta)}{\partial \delta_2}\right)\right) - \frac{2}{m}\sum_{j=1}^{m}\log \lambda_j$$

with

$$\frac{\partial \tilde{G}(\theta)}{\partial \beta} = \frac{1}{m}\sum_{j=1}^{m} \beta\lambda_j^{\delta_1}\log\lambda_j \,\text{Re}\begin{pmatrix} 2\lambda_j^{\xi}(\beta\lambda_j^{\xi} I_{jxx} - I_{jxy}) & -\lambda_j^{\delta_2+2\xi} I_{xx} \\ -\lambda_j^{\delta_2+2\xi} I_{jxx} & 0 \end{pmatrix}$$

and $\hat{G}(\theta)/\partial\delta_2$ defined in (22), which finally gives

$$\begin{aligned}\frac{\partial R_m(\theta)}{\partial \xi} &= \frac{2}{m}\sum_{j=1}^{m}\log\lambda_j \,\text{Re}\left(\beta\lambda_j^{-\nu_0}\hat{G}_{1.}(\theta)^{-1}(\Lambda_j I_j\Lambda_j)_{.2} + \hat{G}_{2.}(\theta)^{-1}(\Lambda_j I_j \Lambda_j)_{.2} - 1\right) \\ &= \frac{\partial \tilde{R}_m(\theta)}{\partial \beta} + \frac{\partial R_m(\theta)}{\partial \delta_2} \end{aligned} \tag{25}$$

where

$$\partial \tilde{R}_m(\theta)/\partial\beta = -\frac{2}{m}\beta\sum_{j=1}^{m}\lambda_j^{-\nu_0}\log\lambda_j \,\text{Re}\left(\hat{G}_{1.}(\theta)^{-1}(\Lambda_j I_j\Lambda_j)_{.2}\right).$$

We analyze (19) for each parameter and begin with β. We proceed as in Robinson (2008a) and after rearrangements implying only negligible errors (see Lobato 1999, Appendix C) arising from the replacement of $(\Lambda_j I_j \Lambda_j)$ by $P_j I_{j\varepsilon} P_j^* := \Lambda_j A(\lambda_j) I_{j\varepsilon} A(\lambda_j)^* \Lambda_j$ and $\hat{G}(\theta_0)$ by G as $||\hat{G}(\theta_0) - G|| = O_p(m^{-1/2})$, we obtain

$$\eta_3\sqrt{m}\lambda_m^{\nu_0}\frac{\partial R_m(\theta)}{\partial\beta}\Big|_{\theta_0} = -\eta_3\lambda_m^{\nu_0}\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\left(\lambda_j^{-\nu_0} - m^{-1}\sum_{k=1}^{m}\lambda_k^{-\nu_0}\right)\text{Re}\left(G_{1.}^{-1}P_j I_{j\varepsilon}P_{j2.}^*\right) + o_p(1). \tag{26}$$

In (26) we decompose $I_{j\varepsilon} = (2\pi n)^{-1}|\sum_t^n \varepsilon_t|^2$ such that

$$\eta_3 \sqrt{m}\lambda_m^{\nu_0} \frac{\partial R_m(\theta)}{\partial \beta}\Big|_{\theta_0} = -\eta_3 \lambda_m^{\nu_0} \frac{1}{\pi\sqrt{m}} \sum_{j=1}^{m} \gamma_j \operatorname{Re}\left(G_{1.}^{-1} P_j P_{j2.}^*\right) \tag{27}$$

$$-\eta_3 \lambda_m^{\nu_0} \frac{1}{\pi\sqrt{m}} \sum_{j=1}^{m} \gamma_j \operatorname{Re}\left(G_{1.}^{-1} P_j \left(\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t \varepsilon_t' - I_2\right) P_{j2.}^*\right) \tag{28}$$

$$-\eta_3 \lambda_m^{\nu_0} \frac{2}{\sqrt{m}} \sum_{j=1}^{m} \gamma_j \operatorname{Re}\left(G_{1.}^{-1} P_j \frac{1}{2\pi n} \sum_{t=1}^{n}\sum_{s\neq t} \varepsilon_t \varepsilon_s' e^{i(t-s)\lambda_j} P_{j2.}^*\right), \tag{29}$$

with $\gamma_j = \left(\lambda_j^{-\nu_0} - m^{-1}\sum_{k=1}^{m} \lambda_k^{-\nu_0}\right)$. As $f_z(\lambda) = (2\pi)^{-1} A(\lambda) A(\lambda)^*$, by Assumpions 7 and 10, (27) is

$$O\left(\frac{1}{\sqrt{m}} \lambda_m^{\nu_0} \sum_{j=1}^{m} f_{12}(\lambda_j) \lambda_j^{\delta_1+\delta_2+\xi-\nu_0}\right) = O\left(\frac{1}{\sqrt{m}} \lambda_m^{\nu_0} \sum_{j=1}^{m} \lambda_j^{\alpha-\nu_0}\right) = O(n^{-\alpha} m^{1/2+\alpha} \log m) \to 0,$$

(28) is

$$O_p\left(\frac{1}{\sqrt{m}} \lambda_m^{\nu_0} \sum_{j=1}^{m} \frac{1}{\sqrt{n}} f_{12}(\lambda_j) \lambda_j^{\delta_1+\delta_2+\xi-\nu_0}\right) = O_p\left(\lambda_m^{1/2+\alpha} \log m\right) \xrightarrow{p} 0$$

by the law of large numbers and the remaining term (29) can be rearranged as

$$-\sum_{t=1}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \frac{\eta_3}{\pi n\sqrt{m}} \lambda_m^{\nu_0} \sum_{j=1}^{m} \gamma_j \operatorname{Re}\left(P_j' G_{.1}^{-1} e^{i(t-s)\lambda_j} \bar{P}_{j2.}\right) \varepsilon_s \tag{30}$$

where $\bar{P}_j$ denotes the conjugate of P_j. We pursue our analysis of (19) and discuss δ_1 and δ_2. By similar arguments to Lobato (1999), equations (23) and (24) have the following asymptotic equivalences

$$\sum_{a=1}^{2} \eta_a \sqrt{m} \frac{\partial R_m(\theta)}{\partial \delta_a}\Big|_{\theta_0} = \frac{2}{\sqrt{m}} \sum_{a=1}^{2} \frac{\eta_a}{2\pi n} \sum_{t=1}^{n}\sum_{s=1}^{t-1} \varepsilon_t' \varepsilon_s \sum_{j=1}^{m} \nu_j \operatorname{Re}\left(G_{a.}^{-1} P_j e^{i(t-s)\lambda_j} P_{ja.}^*\right).$$

A simple rearrangement of this equation gives

$$\sum_{t=1}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \sum_{a=1}^{2} \frac{\eta_a}{\pi n\sqrt{m}} \sum_{j=1}^{m} \nu_j \operatorname{Re}\left(P_j' G_{.a}^{-1} e^{i(t-s)\lambda_j} \bar{P}_{ja.}\right) \varepsilon_s. \tag{31}$$

Regarding ξ, as the second term in (25) is identical to the score with respect to δ_2 it will be treated in the same way, resulting in

$$\sum_{t=1}^{n} \varepsilon_t' \sum_{s=1}^{t-1} \frac{\eta_4}{\pi n\sqrt{m}} \frac{\lambda_m^{\nu_0}}{\log \lambda_m} \sum_{j=1}^{m} \nu_j \operatorname{Re}\left(P_j' G_{.2}^{-1} e^{i(t-s)\lambda_j} \bar{P}_{j2.}\right) \varepsilon_s. \tag{32}$$

We hence focus on the first term and by analogy with (26) we obtain

$$\eta_4 \sqrt{m}\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\frac{\partial\tilde{R}_m(\theta)}{\partial\beta}\Big|_{\theta_0} = -\eta_4\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\frac{\beta}{\pi\sqrt{m}}\sum_{j=1}^{m}\log\lambda_j\gamma_j \operatorname{Re}\left(G_{1.}^{-1}P_jP_{j2.}^*\right) \tag{33}$$

$$-\eta_4\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\frac{\beta}{\pi\sqrt{m}}\sum_{j=1}^{m}\log\lambda_j\gamma_j \operatorname{Re}\left(G_{1.}^{-1}P_j\left(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\varepsilon_t' - I_2\right)P_{j2.}^*\right) \tag{34}$$

$$-\eta_4\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\frac{2\beta}{\sqrt{m}}\sum_{j=1}^{m}\log\lambda_j\gamma_j \operatorname{Re}\left(G_{1.}^{-1}P_j\frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s\neq t}\varepsilon_t\varepsilon_s' e^{i(t-s)\lambda_j}P_{j2.}^*\right). \tag{35}$$

As $f_z(\lambda) = (2\pi)^{-1}A(\lambda)A(\lambda)^*$, by Assumptions (7) and (10), (33) is

$$O\left(\frac{1}{\sqrt{m}}\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\sum_{j=1}^{m}\log\lambda_j f_{12}(\lambda_j)\lambda_j^{\delta_1+\delta_2+\xi-\nu_0}\right) = O(n^{-\alpha}m^{1/2+\alpha}\log m) \to 0,$$

(34) is

$$O_p\left(\frac{1}{\sqrt{m}}\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\sum_{j=1}^{m}\log\lambda_j\frac{1}{\sqrt{n}}f_{12}(\lambda_j)\lambda_j^{\delta_1+\delta_2+\xi-\nu_0}\right) = O_p\left(\lambda_m^{1/2+\alpha}\log m\right) \xrightarrow{p} 0$$

and the remaining term (35) can be rearranged as

$$-\sum_{t=1}^{n}\varepsilon_t'\sum_{s=1}^{t-1}\frac{\eta_4}{\pi n\sqrt{m}}\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\beta\sum_{j=1}^{m}\log\lambda_j\gamma_j \operatorname{Re}\left(P_j' G_{.1}^{-1}e^{i(t-s)\lambda_j}\bar{P}_{j2.}\right)\varepsilon_s. \tag{36}$$

Using (31), (36) and (30) and Euler formula, (19) has the same asymptotic distribution as $\sum_{t=1}^{n}\varepsilon_t'\sum_{s=1}^{t-1}\Xi_{t-s,n}\varepsilon_s$, where

$$\Xi_{t-s,n} = \frac{1}{\pi n\sqrt{m}}\sum_{j=1}^{m}(\theta_{j,1}+\theta_{j,2}+\theta_{j,3})\cos((t-s)\lambda_j),$$

$$\theta_{j,1} = \nu_j\sum_{a=1}^{2}\eta_a\tilde{\theta}_a,$$

$$\theta_{j,2} = -\lambda_m^{\nu_0}\eta_3\gamma_j\tilde{\theta}_\beta,$$

$$\theta_{j,3} = \frac{\lambda_m^{\nu_0}}{\log\lambda_m}\eta_4(\nu_j\tilde{\theta}_2 - \log\lambda_j\beta\gamma_j\tilde{\theta}_\beta)$$

with $\tilde{\theta}_a = \operatorname{Re}\left(P_j' G_{.a}^{-1}\bar{P}_{ja.} + P_{j.a}G_{a.}^{-1}\bar{P}_j\right)$ and $\tilde{\theta}_\beta = \operatorname{Re}\left(P_j' G_{.1}^{-1}\bar{P}_{j2.} + P_{j.2}G_{1.}\bar{P}_j\right)$. By Assumption 3, $||\theta_{j,1}|| = O(1)$ and $||\theta_{j,2}|| = O((m/j)^{\nu_0})$. Using that the first term in $\theta_{j,3}$ is the same as $\theta_{j,1}$ when $a = 2$ it can be shown that $||\theta_{j,3}|| = O((m/j)^{\nu_0}\times(\log j)/(\log m))$. It follows that

$$\sum_{a=1}^{2}\eta_a\sqrt{m}\frac{\partial R_m(\theta)}{\partial\theta_a}\Big|_{\theta_0} + \eta_3\lambda_m^{\nu_0}\sqrt{m}\frac{\partial R_m(\theta)}{\partial\theta_3}\Big|_{\theta_0} + \eta_4\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\sqrt{m}\frac{\partial R_m(\theta)}{\partial\theta_4}\Big|_{\theta_0} = \sum_{t=1}^{n}\zeta_t + o_p(1)$$

where $\zeta_t = \varepsilon_t' \sum_{s=1}^{t-1} \Xi_{t-s,n}\varepsilon_s$ is a martingale difference array with respect to $\mathcal{F}_t = \sigma(\{\varepsilon_s, s \leq t\})$. By a standard martingale central limit theorem, (19) follows if

$$\sum_{t=1}^{n} \mathbb{E}(\zeta_t^2|\mathcal{F}_{t-1}) - \sum_{a=1}^{4}\sum_{b=1}^{4} \eta_a\eta_b\Omega_{ab} \xrightarrow{p} 0 \tag{37}$$

$$\sum_{t=1}^{n} \mathbb{E}(\zeta_t^2 \mathbb{1}(|\zeta_t| > d)) \to 0, \quad \forall d > 0. \tag{38}$$

We first show (37). For this, we use the following decomposition

$$\begin{aligned}\sum_{t=1}^{n} \mathbb{E}(\zeta_t^2|\mathcal{F}_{t-1}) &= \sum_{t=1}^{n} \mathbb{E}\Big(\sum_{s=1}^{t-1}\sum_{r=1}^{t-1} \varepsilon_s'\Xi_{t-s,n}'\varepsilon_t\varepsilon_t'\Xi_{t-r,n}\varepsilon_r|\mathcal{F}_{t-1}\Big) \\ &= \sum_{t=1}^{n}\sum_{s=1}^{t-1} \varepsilon_s'\Xi_{t-s,n}'\Xi_{t-s,n}\varepsilon_s + \sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{r=1}^{s-1} \varepsilon_s'\Xi_{t-s,n}'\Xi_{t-r,n}\varepsilon_r,\end{aligned} \tag{39}$$

where the second term has mean 0 and variance

$$O\left(n\Big(\sum_{s=1}^{n} ||\Xi_{s,n}||^2\Big)^2 + \sum_{t=3}^{n}\sum_{u=2}^{t-1}\Big(\sum_{s=1}^{u-1} ||\Xi_{u-s,n}||^2 \sum_{s=1}^{u-1} ||\Xi_{t-s,n}||^2\Big)\right) \tag{40}$$

as shown by Lobato (1999). Following Nielsen (2007), when $s < n/m$, $||\Xi_{s,n}|| = O(1/(n\sqrt{m})\sum_{j=1}^{n} ||\theta_{j,1} + \theta_{j,2} + \theta_{j,3}||) = O(n^{-1}\sqrt{m}\log m)$ and when $s > n/m$, $||\Xi_{s,n}|| = O(s^{-1}m^{-1/2}\log m)$, where for the latter we use $|\sum_j \cos(s\lambda_j)| = O(n/s)$ and therefore

$$\sum_{s=1}^{n} ||\Xi_{s,n}||^2 = O\left(\sum_{s=1}^{\lfloor n/m \rfloor} \frac{m(\log m)^2}{n^2} + \sum_{s=\lfloor n/m \rfloor+1}^{n} \frac{(\log m)^2}{s^2 m}\right) = O((\log m)^2 n^{-1})$$

such that the first term of (40) is $O((\log m)^4 n^{-1})$. Besides, the second term in (40) is $O\big(n\sum_{s=1}^{n} ||\Xi_{s,n}||^2 \sum_{s=1}^{n/2} s||\Xi_{s,n}||^2\big)$, following the analysis in Robinson (1995), where

$$\sum_{s=1}^{n/2} s||\Xi_{s,n}||^2 = O\Big(m^{-1}(\log m)^2 \log n\Big).$$

It follows immediately that (40) is $O(n^{-1}(\log m)^4 + m^{-1}(\log m)^4 \log n) \to 0$.

To complete the proof of (37) we now have to show that the mean of the first term in (39) is equal to

$\sum_{a=1}^{4}\sum_{b=1}^{4}\eta_a\eta_b\Omega_{ab}$. Since $\mathbb{E}(\varepsilon_t\varepsilon_t'|F_{t-1}) = I_2$ by Assumption (4), we can rewrite

$$\mathbb{E}(\sum_{t=1}^{n}\sum_{s=1}^{t-1}\varepsilon_s'\Xi_{t-s,n}'\Xi_{t-s,n}\varepsilon_s) = \sum_{t=1}^{n}\sum_{s=1}^{t-1}\mathbb{E}\,\mathrm{tr}\,(\Xi_{t-s,n}'\Xi_{t-s,n}\varepsilon_s\varepsilon_s') = \sum_{t=1}^{n}\sum_{s=1}^{t-1}\mathbb{E}\,\mathrm{tr}\,(\Xi_{t-s,n}'\Xi_{t-s,n})$$

and decompose it as

$$\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k=1}^{m}\frac{1}{\pi^2n^2m}\,\mathrm{tr}\,\left((\theta_{j,1}'+\theta_{j,2}'+\theta_{j,3}')(\theta_{k,1}+\theta_{k,2}+\theta_{k,3})\right)\times\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \tag{41}$$

$$=\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2n^2m}\,\mathrm{tr}\,\left(\theta_{j,1}'\theta_{j,1}\right)\cos((t-s)\lambda_j)^2 \tag{42}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2n^2m}\,\mathrm{tr}\,\left(\theta_{j,2}'\theta_{j,2}\right)\cos((t-s)\lambda_j)^2 \tag{43}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2n^2m}\,\mathrm{tr}\,\left(\theta_{j,3}'\theta_{j,3}\right)\cos((t-s)\lambda_j)^2 \tag{44}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2n^2m}2\,\mathrm{tr}\,\left(\theta_{j,1}'\theta_{j,2}\right)\cos((t-s)\lambda_j)^2 \tag{45}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2n^2m}2\,\mathrm{tr}\,\left(\theta_{j,1}'\theta_{j,3}\right)\cos((t-s)\lambda_j)^2 \tag{46}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2n^2m}2\,\mathrm{tr}\,\left(\theta_{j,2}'\theta_{j,3}\right)\cos((t-s)\lambda_j)^2 \tag{47}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2n^2m}\,\mathrm{tr}\,\left(\theta_{j,1}'\theta_{k,1}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \tag{48}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2n^2m}\,\mathrm{tr}\,\left(\theta_{j,2}'\theta_{k,2}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \tag{49}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2n^2m}\,\mathrm{tr}\,\left(\theta_{j,3}'\theta_{k,3}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \tag{50}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2n^2m}2\,\mathrm{tr}\,\left(\theta_{j,1}'\theta_{k,2}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \tag{51}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2n^2m}2\,\mathrm{tr}\,\left(\theta_{j,1}'\theta_{k,3}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \tag{52}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2n^2m}2\,\mathrm{tr}\,\left(\theta_{j,2}'\theta_{k,3}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k). \tag{53}$$

Following Lobato (1999), it can be shown that (42) is asymptotically equal to $\sum_{a=1}^{2}\sum_{b=1}^{2}\eta_a\eta_bE_{ab}$ with

$E = 2(I_2 + G \odot G^{-1})$ and (48) is asymptotically negligible. Also, (45), (49) and (51) are asymptotically equivalent to B.16, B.18 and B.19 in Nielsen (2007) and therefore asymptotically negligible. For (43), after approximating a Riemann sum by an integral and using $\sum_{t=1}^{n} \sum_{s=1}^{t-1} \cos(s\lambda_j)^2 = (n-1)^2/4$, we have that

$$\sum_{t=1}^{n} \sum_{s=1}^{t-1} \sum_{j=1}^{m} \frac{1}{\pi^2 n^2 m} \operatorname{tr}\ (\theta_{j,2}' \theta_{j,2}) \cos((t-s)\lambda_j)^2 \sim \eta_3^2 \frac{2G_{22}}{G_{11}} \frac{\nu_0^2}{(1-2\nu_0)(1-\nu_0)^2}.$$

Among the remaining terms we first analyze (44) and observe that

$$\begin{aligned}\frac{\operatorname{tr}\ (\theta_{j,3}' \theta_{j,3})}{4\pi^2} &= \operatorname{tr}\ \left(\eta_4^2 \frac{\lambda_m^{2\nu_0}}{4\pi^2 (\log \lambda_m)^2} \left(\nu_j^2 \tilde{\theta}_2' \tilde{\theta}_2 + \beta^2 \tilde{\theta}_\beta' \tilde{\theta}_\beta \gamma_j^2 - \gamma_j \nu_j \beta \tilde{\theta}_2' \tilde{\theta}_\beta - \gamma_j \nu_j \beta \tilde{\theta}_\beta' \tilde{\theta}_2 \right) \right) \\ &= \eta_4^2 \frac{\lambda_m^{2\nu_0}}{(\log \lambda_m)^2} \left(4\nu_j^2 + 2\beta^2 G_{22} G_{11}^{-1} \gamma_j^2 \right),\end{aligned}$$

as $\operatorname{tr}\ (\tilde{\theta}_2' \tilde{\theta}_\beta) = \operatorname{tr}\ (\tilde{\theta}_\beta' \tilde{\theta}_2) = 0$ by Assumption (7). Then, using $m^{-1} \sum_{j=1}^{m} \nu_j^2 = 1 + O(m^{-1}(\log m)^2)$ and the same approximation as for (43),(44) is asymptotically equal to

$$\sum_{t=1}^{n} \sum_{s=1}^{t-1} \sum_{j=1}^{m} \frac{1}{\pi^2 n^2 m} \operatorname{tr}\ (\theta_{j,3}' \theta_{j,3}) \cos((t-s)\lambda_j)^2 \sim \eta_4^2 \frac{2\beta^2 G_{22}}{G_{11}} \frac{\nu_0^2}{(1-2\nu_0)(1-\nu_0)^2}.$$

The next term is (46), for which one observes that all elements in $\operatorname{tr}\ (\theta_{j,1}' \theta_{j,3})/(4\pi^2)$ are trivially equal to 0 except

$$\frac{\operatorname{tr}\ (\theta_{j,1}^{(2)'} \theta_{j,3})}{4\pi^2} = \operatorname{tr}\ \left(\eta_2 \eta_4 \frac{\lambda_m^{\nu_0}}{4\pi^2 \log \lambda_m} \left(\nu_j^2 \tilde{\theta}_2' \tilde{\theta}_2 \right) \right) = \eta_2 \eta_4 4 \nu_j^2 \frac{\lambda_m^{\nu_0}}{\log \lambda_m},$$

where $\theta_{j,1}^{(2)}$ denotes $\theta_{j,1}$ with $a = 2$ only. Then, it is immediate that

$$\sum_{t=1}^{n} \sum_{s=1}^{t-1} \sum_{j=1}^{m} \frac{4}{n^2 m} \eta_2 \eta_3 \frac{4\nu_j^2 (n-1)^2}{4} \frac{\lambda_m^{\nu_0}}{\log \lambda_m} = o(1).$$

As the joint limiting distribution of β and ξ is singular, when analyzing (47) we find that it is asymptotically equivalent to

$$\eta_3 \eta_4 2\beta \frac{G_{22}}{G_{11}} \frac{\nu_0^2}{(1-2\nu_0)(1-\nu_0)^2}.$$

Equations (50), (52) and (53), where $j \neq k$, remain to analyze. Using that $||\theta_{j,3}|| = O((m/j)^{\nu_0} (\log j)/(\log m))$

we show that (50) is equivalent to

$$O\left(\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{n^2m}\left(\frac{m}{j}\right)^{\nu_0}\left(\frac{m}{k}\right)^{\nu_0}\frac{(\log j)^2}{(\log m)^2}\cos((t-s)\lambda_j)\cos((t-s)\lambda_k)\right)=O\left(n^{-1}m(\log m)^2\right)$$

where $\sum_{t=1}^{n}\sum_{s=1}^{t-1}\cos((t-s)\lambda_j)\cos((t-s)\lambda_k)=-n/2$ for $\lambda_j\neq\lambda_k$. For (52) we also use that $||\theta_{j,1}||=O(1)$ and obtain the following bound

$$O\left(\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{n^2m}\left(\frac{m}{k}\right)^{\nu_0}\frac{\log j}{\log m}\cos((t-s)\lambda_j)\cos((t-s)\lambda_k)\right)=O\left(n^{-1}m\log m\right).$$

For the last term we use that $||\theta_{j,2}||=O((m/j)^{\nu_0})$ and find that (53) is bounded by

$$O\left(\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{n^2m}\left(\frac{m}{j}\right)^{\nu_0}\left(\frac{m}{k}\right)^{\nu_0}\frac{\log j}{\log m}\cos((t-s)\lambda_j)\cos((t-s)\lambda_k)\right)=O\left(n^{-1}m(\log m)^2\right).$$

It remains to show (38) or equivalently the sufficient condition $\sum_{t=1}^{n}\mathbb{E}(\zeta^4)\to 0$. As our analysis of (37) is similar to Lemma 4 of Nielsen (2005), this condition can be proved under Assumption (8). We then obtain $\sum_{t=1}^{n}\mathbb{E}(\zeta^4)=O(n(\sum_{t=1}^{n}||\Xi_{tn}^2||)^2)=O(n^{-1}(\log m)^4)$ as in Nielsen (2007). This completes the proof of (19). □

B.2. Limit of the Hessian

We now derive the limit of the Hessian for any estimator $\bar{\theta}$ such that $||\bar{\theta}-\theta_0||\leq||\hat{\theta}-\theta_0||$ and prove that

$$\frac{\partial^2 R_m(\bar{\theta})}{\partial\delta_a\partial\delta_b}\xrightarrow{p}E_{ab}\tag{54}$$

$$\lambda_m^{\nu_0}\frac{\partial^2 R_m(\bar{\theta})}{\partial\delta_a\partial\beta}\xrightarrow{p}0\tag{55}$$

$$\lambda_m^{\nu_0}\log(\lambda_m)^{-1}\frac{\partial^2 R_m(\bar{\theta})}{\partial\delta_a\partial\xi}\xrightarrow{p}0\tag{56}$$

$$\lambda_m^{2\nu_0}\begin{pmatrix}1\\ \log(\lambda_m)^{-2}\end{pmatrix}'\begin{pmatrix}\partial^2 R_m(\bar{\theta})/(\partial\beta\partial\beta)\\ \partial^2 R_m(\bar{\theta})/(\partial\xi\partial\xi)\end{pmatrix}\xrightarrow{p}\begin{pmatrix}1\\ \beta^2\end{pmatrix}F.\tag{57}$$

Proof. As Nielsen (2007) we strengthen the approximation $||\hat{G}(\theta_0)-G||=O_p(m^{-1/2})$ by showing that

$$||\hat{G}(\bar{\theta})-\hat{G}(\theta_0)||=o_p(1).\tag{58}$$

We first analyze $\hat{G}_{22}(\bar{\theta})$ and proceed to the following decomposition

$$\begin{aligned}\hat{G}_{22}(\bar{\theta})-\hat{G}_{22}(\theta_0)&=\frac{1}{m}\sum_{j=1}^{m}\left(\lambda_j^{2\bar{\vartheta}_2}-\lambda_j^{2\vartheta_2}\right)I_{j22}\\&=O_p\left(\frac{1}{m}\Big(\max_{1\le j\le m}\lambda_j^{2\bar{\vartheta}_2-2\vartheta_2}-1\Big)\sum_{j=1}^{m}\lambda_j^{2\vartheta_2}I_{j22}\right)=o_p(1).\end{aligned}$$

Similarly,

$$\begin{aligned}\hat{G}_{12}(\bar{\theta})-\hat{G}_{12}(\theta_0)&=\frac{1}{m}\sum_{j=1}^{m}\left(\lambda_j^{\bar{\vartheta}_1+\bar{\vartheta}_2}\bar{I}_{j12}-\lambda_j^{\vartheta_1+\vartheta_2}I_{j12}\right)\\&=\frac{1}{m}\sum_{j=1}^{m}\left(\lambda_j^{\bar{\vartheta}_1+\bar{\vartheta}_2}(\beta\lambda_j^{\xi}-\bar{\beta}\lambda_j^{\bar{\xi}})I_{j22}\right)+\frac{1}{m}\sum_{j=1}^{m}\left(\lambda_j^{\bar{\vartheta}_1+\bar{\vartheta}_2}-\lambda_j^{\vartheta_1+\vartheta_2}\right)I_{j12}\end{aligned}\tag{59}$$

with $\bar{I}_{j12}-I_{j12}=(\beta\lambda_j^{\xi}-\bar{\beta}\lambda_j^{\bar{\xi}})I_{j22}$. The first term in (59) is bounded by

$$\frac{1}{m}\Big(\max_{1\le j\le m}\lambda_j^{\bar{\vartheta}_1+\bar{\vartheta}_2-\vartheta_1-\vartheta_2}\Big)\sum_{j=1}^{m}\lambda_j^{\vartheta_1+\vartheta_2}(\beta\lambda_j^{\xi}-\bar{\beta}\lambda_j^{\bar{\xi}})I_{j22}=o_p(1)$$

and the second term in (59) is

$$O_p\left(\frac{1}{m}\Big(\max_{1\le j\le m}\lambda_j^{\bar{\vartheta}_1+\bar{\vartheta}_2-\vartheta_1-\vartheta_2}-1\Big)\sum_{j=1}^{m}\lambda_j^{\vartheta_1+\vartheta_2}I_{j12}\right)=o_p(1)$$

Regarding the first diagonal element of $\hat{G}(\bar{\theta})$, the same bound can be obtained by Cauchy-Schwartz inequality. Then, using (58) and similar arguments to Lobato (1999) one can easily show (54). Proceeding in a similar way to (58), it can be shown that

$$\lambda_m^{\nu_0}\Big(\frac{\partial^2R_m(\bar{\theta})}{\partial\delta_a\partial\beta}-\frac{\partial^2R_m(\theta_0)}{\partial\delta_a\partial\beta}\Big)\xrightarrow{p}0\tag{60}$$

$$\lambda_m^{2\nu_0}\Big(\frac{\partial^2R_m(\bar{\theta})}{\partial\beta\partial\beta}-\frac{\partial^2R_m(\theta_0)}{\partial\beta\partial\beta}\Big)\xrightarrow{p}0\tag{61}$$

$$\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\Big(\frac{\partial^2R_m(\bar{\theta})}{\partial\delta_a\partial\xi}-\frac{\partial^2R_m(\theta_0)}{\partial\delta_a\partial\xi}\Big)\xrightarrow{p}0\tag{62}$$

and implicitly

$$\frac{\lambda_m^{2\nu_0}}{(\log\lambda_m)^2}\Big(\frac{\partial^2R_m(\bar{\theta})}{\partial\xi\partial\xi}-\frac{\partial^2R_m(\theta_0)}{\partial\xi\partial\xi}\Big)\xrightarrow{p}0.\tag{63}$$

To analyze (57) we first observe that

$$\lambda_m^{2\nu_0}\frac{\partial^2 R_m(\theta_0)}{\partial\beta\partial\beta} = \lambda_m^{2\nu_0}\,\mathrm{tr}\,\left(\hat{G}(\theta_0)^{-1}\frac{\partial^2\hat{G}(\theta_0)}{\partial\beta\partial\beta} - \hat{G}(\theta_0)^{-1}\frac{\partial\hat{G}(\theta_0)}{\partial\beta}\hat{G}(\theta_0)^{-1}\frac{\partial\hat{G}(\theta_0)}{\partial\beta}\right). \tag{64}$$

The first term can be shown to correspond to C.9 in Nielsen (2007)

$$\begin{aligned}\lambda_m^{2\nu_0}\,\mathrm{tr}\,\left(\hat{G}(\theta_0)^{-1}\frac{\partial^2\hat{G}(\theta_0)}{\partial\beta\partial\beta}\right) &= \lambda_m^{2\nu_0}\frac{2}{m}\frac{G_{22}}{G_{11}}\sum_{j=1}^{m}\lambda_j^{2(\delta_1-\delta_2)} + o_p(1)\\ &\sim \frac{2G_{22}}{G_{11}(1-2\nu_0)},\end{aligned} \tag{65}$$

where the Riemann sum was approximated by an integral. Using the same arguments and approximation as above, it can also be shown that the second term in (64) is

$$\begin{aligned}-\lambda_m^{2\nu_0}\,\mathrm{tr}\,\left(\hat{G}(\theta_0)^{-1}\frac{\partial\hat{G}(\theta_0)}{\partial\beta}\hat{G}(\theta_0)^{-1}\frac{\partial\hat{G}(\theta_0)}{\partial\beta}\right) &= -\lambda_m^{2\nu_0}\frac{2}{m^2}\frac{G_{22}}{G_{11}}\sum_{j=1}^{m}\lambda_j^{-\nu_0}\sum_{k=1}^{m}\lambda_k^{-\nu_0} + o_p(1)\\ &\sim -\frac{2G_{22}}{G_{11}(1-\nu_0)^2}\end{aligned}$$

and overall $\lambda_m^{2\nu_0}\partial^2 Rm(\theta_0)/(\partial\beta\partial\beta) \xrightarrow{p} F = 2G_{22}\nu_0^2/(G_{11}(1-2\nu_0)(1-\nu_0)^2)$, which in view of (61) proves the first term in (57). We now show that $\lambda_m^{\nu_0}\partial^2 Rm(\theta_0)/(\partial\delta_a\partial\beta) \xrightarrow{p} 0$. Observe that

$$\begin{aligned}\lambda_m^{\nu_0}\frac{\partial^2 R_m(\theta_0)}{\partial\delta_a\partial\beta} &= \mathrm{tr}\,\left(\hat{G}(\theta_0)^{-1}\frac{\partial^2\hat{G}(\theta)}{\partial\delta_a\partial\beta} - \hat{G}(\theta_0)^{-1}\frac{\partial\hat{G}(\theta)}{\partial\delta_a}\hat{G}(\theta_0)^{-1}\frac{\partial\hat{G}(\theta)}{\partial\beta}\right)\\ &= \lambda_m^{\nu_0}\frac{2}{m}\sum_{j=1}^{m}\nu_j\,\mathrm{Re}\left[G_{a.}^{-1}\left(\frac{1}{m}\sum_{k=1}^{m}\lambda_k^{\xi}\Lambda_k\begin{pmatrix}2I_{k1a} & I_{ka2}\\ I_{ka2} & 0\end{pmatrix}\Lambda_k\right)G^{-1}\Lambda_j\begin{pmatrix}I_{j12}\\ I_{j22}\end{pmatrix}\lambda_j^{\delta_a+\xi}\right]\\ &- \lambda_m^{\nu_0}\frac{4}{m}\sum_{j=1}^{m}\nu_j\,\mathrm{Re}\left[G_{a.}^{-1}\Lambda_j\begin{pmatrix}I_{ja2}\\ 0\end{pmatrix}\lambda_j^{\delta_a+\xi}\right] + o_p(1).\end{aligned}$$

Using Assumption (3), it can be easily shown that both terms are $o_p(1)$, which proves (55) in view of (60). To show that $\lambda_m^{\nu_0}(\log\lambda_m)^{-1}\partial^2 Rm(\theta_0)/(\partial\delta_a\partial\xi) \xrightarrow{p} 0$, we decompose it as

$$\frac{\lambda_m^{\nu_0}}{\log\lambda_m}\frac{\partial^2 R_m(\theta_0)}{\partial\delta_a\partial\xi} = \frac{\lambda_m^{\nu_0}}{\log\lambda_m}\left(\frac{\partial^2\tilde{R}_m}{\partial\delta_a\partial\beta} + \frac{\partial^2 R_m}{\partial\delta_a\partial\delta_2}\right).$$

The first term is $o_p(1)$ by analogy with the result on $\partial^2 Rm(\theta_0)/(\partial\delta_a\partial\beta)$ while the second term vanishes asymptotically as in Lobato (1999) when $a = 1$ and is $O(\lambda_m^{\nu_0}(\log\lambda_m)^{-1})$ when $a = 2$ since $\partial^2 R_m/\partial\delta_2^2 = 4$. By (62), (56) follows.

To complete the proof, it remains to analyze $\lambda_m^{2\nu_0}(\log\lambda_m)^{-2}\partial^2 R_m(\theta_0)/(\partial\xi\partial\xi)$. Writing it as $\lambda_m^{2\nu_0}(\log\lambda_m)^{-2}\times$

$(\partial^2\tilde{R}_m(\theta)/(\partial\beta\partial\xi)\ +\partial^2 R_m(\theta)/(\partial\delta_2\partial\xi))$, observe that the second term of the sum vanishes as shown above for $a=2$, whereas the first term is

$$\lambda_m^{2\nu_0}(\log\lambda_m)^{-2}\frac{\partial^2\tilde{R}_m(\theta_0)}{\partial\beta\partial\xi}=\lambda_m^{2\nu_0}(\log\lambda_m)^{-2}\mathrm{tr}\Big(\hat{G}(\theta_0)^{-1}\frac{\partial^2\tilde{G}(\theta_0)}{\partial\beta\partial\xi}-\hat{G}(\theta_0)^{-1}\frac{\partial\tilde{G}(\theta_0)}{\partial\beta}\hat{G}(\theta_0)^{-1}\frac{\partial\tilde{G}(\theta_0)}{\partial\xi}\Big). \quad (66)$$

The first component in (66) can be analyzed by analogy with (65) to show that in the vicinity of the origin the following approximation holds

$$\begin{aligned}\frac{\lambda_m^{2\nu_0}}{(\log\lambda_m)^2}\,\mathrm{tr}\,\left(\hat{G}(\theta_0)^{-1}\frac{\partial^2\tilde{G}(\theta_0)}{\partial\beta\partial\xi}\right)&=\lambda_m^{2\nu_0}\log(\lambda_m)^{-2}\beta^2G_{11}^{-1}G_{22}\frac{2}{m}\sum_{j=1}^{m}\log\lambda_j^2\lambda_j^{-2\nu_0}+o_p(1)\\&\sim\frac{2\beta^2G_{22}}{G_{11}(1-2\nu_0)}.\end{aligned}$$

The second component follows from the same arguments

$$\begin{aligned}-\frac{\lambda_m^{2\nu_0}}{\log(\lambda_m)^2}\,\mathrm{tr}\,\left(\hat{G}(\theta_0)^{-1}\frac{\partial\tilde{G}(\theta_0)}{\partial\beta}\hat{G}(\theta_0)^{-1}\frac{\partial\tilde{G}(\theta_0)}{\partial\xi}\right)&=-\frac{\lambda_m^{2\nu_0}}{\log(\lambda_m)^2}\frac{2\beta^2}{m^2}\frac{G_{22}}{G_{11}}\sum_{j=1}^{m}\log\lambda_j\lambda_j^{-\nu_0}\sum_{k=1}^{m}\log\lambda_k\lambda_k^{-\nu_0}+o_p(1)\\&\sim-\frac{2\beta^2G_{22}}{G_{11}(1-\nu_0)^2},\end{aligned}$$

finally yielding

$$\lambda_m^{2\nu_0}\log(\lambda_m)^{-2}\frac{\partial^2R_m(\theta_0)}{\partial\xi\partial\xi}\xrightarrow{p}\beta^2F=\frac{2\beta^2G_{22}}{G_{11}}\frac{\nu_0^2}{(1-2\nu_0)(1-\nu_0)^2},$$

which completes the proof in view of (63). Besides, as the joint limiting distribution of β and ξ is singular, we omit the superfluous covariance term. □

References

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115-158.

Bardet, J.-M., Tudor, C., 2014. Asymptotic behavior of the Whittle estimator for the increments of a Rosenblatt process. Journal of Multivariate Analysis 131, 1-16.

Cox, C.C., 1976. Futures Trading and Market Information. Journal of Political Economy 84, 1215-1237.

Cheung, Y., Lai, K., 1993. A fractional cointegration analysis of purchasing power parity. Journal of Business & Economic Statistics 11, 103-112.

Christensen, B.J., Nielsen, M.Ø., 2006. Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. Journal of Econometrics 133, 343-371.

Engle, R., Granger, C.W.J, 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55, 251-276.

Darolles, S., Le Fol, G., Mero, G., 2017. Mixture of distribution hypothesis: Analyzing daily liquidity frictions and information flows. Journal of Econometrics, 201, 367-383.

Davidson, J., Hashimzade, N., 2009. Type I and type II fractional Brownian motions: A reconsideration. Computational Statistics & Data Analysis 53, 2089-2106.

de Truchis, G., Dumitrescu, E-I., 2019. Narrow-band Weighted Nonlinear Least Squares Estimation of Unbalanced Cointegration Systems. Working paper.

Frederiksen, P., Nielsen, F.S., Nielsen, M.Ø., 2012. Local polynomial Whittle estimation of perturbed fractional processes. Journal of Econometrics 167, 426-447.

Granger, C.W.J., 1981. Some properties of time series data and their use in econometric model specification. Journal of econometrics 16, 121-130.

Haldrup, N., Nielsen, M.Ø., 2006. A regime switching long memory model for electricity prices. Journal of Econometrics 135, 349-376.

Helgason, H., Pipiras, V., Abry, P., 2011. Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding. Signal Processing 91, 1123-1133.

Hualde, J., 2006. Unbalanced Cointegration. Econometric Theory 22, 765-814.

Hualde, J., 2012. Weak convergence to a modified fractional Brownian motion. Journal of Time Series Analysis 33, 519-529.

Hualde, J., 2013. A simple test for the equality of integration orders. Economics Letters 119, 233-237.

Hualde, J., 2014. Estimation of long-run parameters in unbalanced cointegration. Journal of Econometrics 178, 761-778.

Hualde, J., Robinson, P. M., 2007. Root-n-consistent estimation of weak fractional cointegration. Journal of Econometrics, 140, 450-484.

Hualde, J., Robinson, P.M., 2010. Semiparametric inference in multivariate fractionally cointegrated systems. Journal of Econometrics 157, 492-511.

Hurvich, C.M., Moulines, E., Soulier, P., 2005. Estimating long memory in volatility. Econometrica 73, 1283-1328.

Jeganathan, P., 1999. On asymptotic inference in cointegrated time series with fractionally integrated errors. Econometric Theory 15, 583-621.

Johansen, S., Nielsen, M.Ø., 2012. Likelihood Inference for a Fractionally Cointegrated Vector Autoregressive Model. Econometrica 80, 2667-2732.

Johansen, S., Nielsen, M.Ø., 2019. A cointegrated model allowing for different fractional orders. Work in progress.

Liu, P., Tang, K., 2010. No-arbitrage conditions for storable commodities and the modeling of futures term structures. Journal of Banking & Finance 34, 1675-1687.

Liu, P., Tang, K., 2011. The stochastic behavior of commodity prices with heteroskedasticity in the convenience yield. Journal of Empirical Finance 18, 211-224.

Lobato, I.N., 1999. A semiparametric two-step estimator in a multivariate long memory model. Journal of Econometrics 90, 129-153.

Lobato, I.N., Velasco, C., 2000. Long memory in stock-market trading volume. Journal of Business & Economic

Statistics 18, 410-427.

Lütkepohl, H., 1996. Handbook of Matrices, New York: Wiley.

Marinucci, D., Robinson, P.M., 1999. Alternative forms of fractional Brownian motion. Journal of Statistical Planning and Inference 80, 111-122.

Nielsen, M.Ø., 2005. Semiparametric Estimation in Time-Series Regression with Long-Range Dependence. Journal of Time Series Analysis 26, 279-304.

Nielsen, M.Ø., 2007. Local Whittle Analysis of Stationary Fractional Cointegration and the Implied-Realized Volatility Relation. Journal of Business & Economic Statistics 25, 427-446.

Nielsen, M.Ø., Frederiksen, P., 2011. Fully modified narrow-band least squares estimation of weak fractional cointegration. The Econometrics Journal 14, 77-120.

Nielsen, M.Ø., Shimotsu, K., 2007. Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. Journal of Econometrics 141, 574-596.

Phillips, P., 1991. Optimal inference in cointegrated systems. Econometrica 59, 283-306.

Robinson, P.M., 1994. Semiparametric analysis of long-memory time series. The Annals of Statistics 22, 515-539.

Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. The Annals of statistics 23, 1630-1661.

Robinson, P.M., 2005. The distance between rival nonstationary fractional processes. Journal of Econometrics 128, 283-300.

Robinson, P.M, 2008a. Multiple local Whittle estimation in stationary systems. The Annals of Statistics 36, 2508-2530.

Robinson, P.M., 2008b. Diagnostic testing for cointegration. Journal of Econometrics 143, 206-225.

Robinson, P.M., Hualde, J., 2003. Cointegration in fractional systems with unknown integration orders. Econometrica 71, 1727-1766.

Robinson, P.M., Marinucci, D., 2001. Narrow-band analysis of nonstationary processes. The Annals of Statistics 29, 947-986.

Robinson, P.M., Marinucci, D., 2003. Semiparametric frequency domain analysis of fractional cointegration. In Time Series with Long Memory (ed. P. M. Robinson). Oxford University Press, pp. 334-373.

Robinson, P.M., Yajima, Y., 2002. Determination of cointegrating rank in fractional systems. Journal of Econometrics 106, 217-241.

Ross, S., 1989. Information and Volatility : The No-Arbitrage Martingale Approach to Timing and Resolution Irrelevancy. The Journal of Finance XLIV, 1-15.

Rossi, E., Santucci de Magistris, P., 2013. A no-arbitrage fractional cointegration model for futures and spot daily ranges. Journal of Futures Markets 33, 77-102.

Shimotsu, K., 2007. Gaussian semiparametric estimation of multivariate fractionally integrated processes. Journal of Econometrics 137, 277-310.

Shimotsu, K., 2012. Exact local Whittle estimation of fractionally cointegrated systems. Journal of Econometrics 169, 266-278.

Tsay, W. J., Chung, C. F., 2000. The spurious regression of fractionally integrated processes. Journal of Econometrics, 96, 155-182.