EconomiX

https://economix.fr/

Narrow-band Weighted Nonlinear Least Squares Estimation of Unbalanced Cointegration Systems

Document de Travail
Working Paper
2019-14

Elena Ivona Dumitrescu
Gilles de Truchis

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

Email : secretariat@economix.fr

UMR 7235

Université Paris Nanterre

Narrow-band Weighted Nonlinear Least Squares Estimation of Unbalanced Cointegration Systems

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

May 2019

Abstract

We discuss cointegration relationships when covariance stationary observables exhibit unbalanced integration orders. Least squares type estimates of the long run coefficient are expected to converge either to 0 or to infinity if one does not account for the true unknown unbalance parameter. We propose a class of narrow-band weighted non-linear least squares estimators of these two parameters and analyze its asymptotic properties. The limit distribution is shown to be Gaussian, albeit singular, and it covers the entire stationary region in the particular case of the generalized non-linear least squares estimator, thereby allowing for straightforward statistical inference. A Monte Carlo study documents the good finite sample properties of our class of estimators. They are further used to provide new perspectives on the risk-return relationship on financial stock markets. In particular, we find that the variance risk premium estimated in an appropriately rebalanced cointegration system is a better return predictor than existing risk premia measures.

Keywords: Unbalanced cointegration, Long memory, Stationarity, Generalized Least Squares, Nonlinear Least Squares
JEL: C22, G10

1. Introduction

Cointegration has attracted increasing attention since the seminal paper of Granger (1981) as a large part of economic theory involves long-run equilibrium relationships between economic and/or financial variables. Initially, most of the analyses have focused on the estimation of the long-run slope coefficient in the $I(1)/I(0)$ paradigm (single unit-root in observables and short memory in cointegration errors) using OLS type regressions. There is evidence, however, that in many economic and financial applications the variables and the residuals are fractionally integrated. In the bivariate case, if y_t and x_t are $I(\delta_2)$, $\delta_2 \in \mathbb{R}^+$, and there exists $\beta \neq 0$ such that

$$y_t = \beta x_t + u_t, \tag{1}$$

where u_t is $I(\delta_1)$ with $\delta_2 > \delta_1$, then y_t and x_t are said to be fractionally cointegrated. In particular, if $\delta_2 \in (0, 1/2)$, y_t and x_t are covariance stationary with long memory. However, the risk of spurious

*Corresponding author
Email addresses: detruchis.gilles@parisnanterre.fr (Gilles de Truchis), elena.dumitrescu@parisnanterre.fr (Elena-Ivona Dumitrescu)
[1]EconomiX-CNRS, University of Paris Nanterre, 200 Avenue de la République, 92000 Nanterre, France.

regression when estimating β is present even for such processes as long as their orders of integration sum up to a value greater than $1/2$ (see Tsay and Chung 2000). This strong result has led to a vast literature on stationary fractional cointegration, which generally uses frequency domain analysis. Indeed, the autocovariance function of a stationary long memory variable has a frequency domain representation known as spectral density, $f_y(\lambda)$, so that

$$E\Big(y_t - E(y_t)\Big)\Big(y_{t+h} - E(y_t)\Big) = \int_{-\pi}^{\pi} e^{ik\lambda} f_y(\lambda) d\lambda,$$

with $\lambda = 2\pi j/n$, the Fourier frequencies for $j = 1, \dots, n$. As in a cointegration framework interest lies in the long-run dynamics of the process, i.e. the estimation of β, one can rely on the power law approximation of $f_y(\lambda)$ in the vicinity of the origin

$$f_y(\lambda) \sim g\lambda^{-2\delta_2}, \quad \text{as } \lambda \to 0^+, \tag{2}$$

where $0 < g < \infty$ and "$\sim$" means that the ratio of the left and right sides converges to 1 in the limit, to avoid specifying the short-run dynamics explicitly. Robinson (1994a) exploits this representation to propose a narrow-band version of the frequency domain least-squares estimator of β, also known as NBLS. Its asymptotic properties are investigated by Robinson and Marinucci (2003) and Christensen and Nielsen (2006). It is found that the convergence rate of β depends on the cointegration strength, $\delta_2 - \delta_1$, and statistical inference based on asymptotically Gaussian distribution is valid as long as the collective memory of x_t and u_t does not exceed $1/2$. Otherwise, the limit distribution is non-standard and presumably of Rosenblatt-type as shown by Robinson (1994b) for the OLS estimator. Motivated by such observations, Robinson and Hidalgo (1997) introduce the frequency domain (full-band) weighted least squares estimator. It is shown to encompass the frequency domain generalized least squares estimator and to converge at conventional $\sqrt{n}$ rate to a Gaussian limit distribution if y_t is covariance stationary without further assumptions regarding the long memory of u_t. As the weight function, $\phi(\lambda)$, is unknown, it can be either set to one, in which case the estimator comes down to the OLS in frequency domain, or parametrically estimated by relying on the inverse of the spectral density of u_t, $f_{u_t}(\lambda)^{-1}$, in which case one obtains the feasible generalized least squares estimator. Since the latter hinges on the correct specification of $f_{u_t}(\lambda)$, a non-parametric estimator of the weight function was introduced by Hidalgo and Robinson (2002). Finally, Nielsen (2005) adopts an intermediate semi-parametric approach based on the approximation of the weight function in the neighborhood of the origin, i.e. $\phi(\lambda) = \lambda^{2\delta}$ as $\lambda \to 0^+$ with δ the weighting parameter. For $\delta = 0$ ($\delta = \hat{\delta}_1$), the narrow-band weighted least squares (hereafter NBWLS) of Nielsen (2005) reduces to the feasible narrow-band (generalized) least squares.

All these studies assume that y_t and x_t are stationary and integrated of the same order δ_2. Nevertheless, economic theory may sometimes suggest the existence of a long run equilibrium relationship between variables, while econometric tests conclude against the equality of integration orders. Hualde (2006) analyzes this unbalanced situation and shows that one can retrieve a nontrivial balanced hidden cointegration by estimating an additional parameter, ξ, which characterizes the level of unbalance. He summarizes the asymptotic behavior of the OLS and NBLS estimators of β in a nonstationary framework by considering different balanced and unbalanced cointegrating situations. The accurate estimation of the unbalance parameter appears to be crucial for the limit properties of the long run parameter as otherwise the least-square-type estimates of β are not consistent. Indeed, they converge to 0 or diverge to infinity depending on the sign of ξ (see e.g. Robinson and Marinucci 2001). In this vein, Hualde (2014) proposes a joint nonlinear least squares (NLS) estimator of β and ξ for $\delta_2 > 1/2$ and shows that its limit distribution is very complex and depends on a modified type II fractional Brownian motion (see Hualde 2012). Besides, he conjectures that his time domain nonlinear estimator would be inconsistent in the stationary region of the long memory parameters. More generally, to the best of our knowledge, no estimator specifically designed for stationary unbalanced triangular systems exists to date.

This paper contributes to this literature by proposing a class of narrow-band weighted nonlinear least squares (NBWNLS) estimators specifically designed for stationary unbalanced fractionally cointegrated relationships. Note that one recovers Nielsen (2005)'s NBWLS linear estimator when the true unbalance parameter is null. Although nonlinear, our joint estimators of β and ξ inherit the desirable asymptotic properties of the linear weighted least squares estimator of Robinson and Hidalgo (1997). When the weight function is set to one, the NBWNLS comes down to the the narrow-band nonlinear least squares (NBNLS). Even if the NBNLS estimates are always feasible, they are generally asymptotically less efficient than the NBWNLS unless the optimal weighting parameter is $\delta = 0$. Still, one can use them to estimate the optimal weighting parameter that further leads to the feasible narrow-band generalized nonlinear least squares (NBGNLS) estimator. We show that, under regular assumptions, our class of estimators is consistent and asymptotically Gaussian (over the entire stationary region for the NBGNLS). The convergence rates of both $\hat{\beta}$ and $\hat{\xi}$ depend on the cointegration strength. Most importantly, that of the unbalance estimator also depends on the true long run parameter β, while the converse does not hold. Consequently, their joint limit distribution is singular, although both estimators are identifiable. Our class of estimators can be seen as an extension of Robinson (1995) and Nielsen (2005) to a non-linear framework, which, in this particular case, accounts for the possible unbalance of the system and as a complementary approach to Hualde (2014) in the sense that it is particularly designed for stationary regions. These theoretical findings are subsequently corroborated by Monte Carlo experiments for different levels of cointegration strength. Our class of estimators and their main (balanced) competitors are evaluated in terms of bias, variance and root-mean-square error criteria. The simulation results reveal the good small-sample properties of our estimators.

To illustrate the usefulness of our approach, we revisit the risk-return relationship by drawing on Bollerslev et al. (2013) who show that the difference between implied and realized variance measures, i.e. the variance risk premium, is a good return predictor in the long run. As our framework accounts for possible unbalance of integration orders of the variance measures, our approach provides more accurate estimates of the variance risk premium, which results in better return predictability for four main stock market indices (S&P500, Dow Jones, Russell, NASDAQ).

The rest of the paper is organized as follows. In Section 2 we introduce our narrow-band weighted nonlinear least squares estimator, while in Section 3 we discuss its consistency and asymptotic normality. The particular case of the feasible NBGNLS is investigated 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, B and C.

2. Narrow-band weighted nonlinear least squares

Let the unbalanced version of Equation (1) be

$$y_t = \beta x_t(\xi) + u_t \tag{3}$$

where $x_t(\xi) = (1-L)^{\xi} x_t = \sum_{k=0}^{\infty} a_k(\xi) L^k$, with $a_k(\xi) := \Gamma(k-\xi)(\Gamma(-\xi)k!)^{-1}$, $x_t \sim I(\delta_2+\xi)$, $y_t \sim I(\delta_2)$, and $u_t \sim I(\delta_1)$. To allow for the presence of common stationary long-run dynamics in (3), we introduce a regularity assumption on the memory parameters.

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*

$$0 \leq \delta_1 < \delta_2 < \delta_2 + |\xi| < 1/2, \tag{4}$$

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

Under Assumption 1, $\delta_2 > \delta_1$ such that β is identifiable, and $\beta \neq 0$ to ensure the identification of ξ. It follows that $f_z(\lambda_j)$, the joint spectral density of $z_t = (u_t, x_t)'$, exists, where λ_j denotes the Fourier frequencies, $\lambda_j = 2\pi j/n$, with $j = 1, \ldots, m$ and m is the bandwidth parameter. Setting $m = o(n)$, we avoid a parametric treatment of $f_z(\lambda)$ by specifying the spectral density only locally around the zero frequency

$$f_z(\lambda) \sim \Big(\Lambda(\lambda)\Big)^{-1} G \Big(\Lambda(\lambda)^*\Big)^{-1}, \qquad \Lambda(\lambda) = \text{diag}\left(\lambda^{\delta_1}, \lambda^{\delta_2+\xi}\right), \qquad \text{as } \lambda \to 0^+ \tag{5}$$

where the superscript "$*$" denotes the conjugate transpose. As Nielsen (2005) we assume that G is diagonal, thereby ruling out the possibility of any endogeneity in the vicinity of the origin but without imposing any further restrictions on frequencies away from the origin in contrast to Hualde (2014).

Now, let $I_{uu}(\lambda_j) = w_u(\lambda_j)w_u(\lambda_j)^*$ be the periodogram of u_t with $w_u(\lambda_j) = (2\pi n)^{-1/2} \sum_{t=1}^{n} u_t e^{it\lambda_j}$ the Fourier transform of u_t. In view of (3), $I_{uu}(\lambda_j)$ actually comes down to

$$I_{uu}(\lambda_j) = I_{yy}(\lambda_j) - \mathrm{Re}\left(2\beta\lambda^{\xi} I_{xy}(\lambda_j) + \beta^2\lambda^{2\xi} I_{xx}(\lambda_j)\right).$$

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. In this context, we identify two natural ways to estimate the long-run and unbalance parameters and show that they are particular cases of a more general class of NBWNLS estimators. A first approach to estimate $\theta = (\beta, \xi)'$ consists in the NBNLS method,

$$\hat{\theta}^{(0)} = \arg\min_{\theta\in\Theta} L_m(\theta), \text{ where } L_m(\theta) = m^{-1}\sum_{j=1}^{m} I_{uu}(\lambda_j) \tag{6}$$

and $\Theta = \Theta_\beta \times \Theta_\xi$ is a compact subset of $\mathbb{R}^2$. However, a more efficient alternative appears naturally from the local Whittle approximation to the likelihood associated with Equation (3),

$$Q_m(\theta, G_{uu}) = m^{-1}\sum_{j=1}^{m}\left(\log f_u(\lambda_j) + \frac{I_{uu}(\lambda_j)}{f_u(\lambda_j)}\right), \tag{7}$$

where $G_{uu} \in \Theta_{G_{uu}}$, the set of real positive numbers. The objective function Q_m is minimized over $\Theta_{G_{uu}}$ by

$$\hat{G}_{uu}(\theta;\delta_1) = \mathrm{Re}\left(m^{-1}\sum_{j=1}^{m}\lambda^{2\delta_1} I_{uu}(\lambda_j)\right), \tag{8}$$

where $\lambda^{2\delta_1}$ arises as an implicit weight function. Substituting (8) in (7) and solving for θ, we obtain the local Whittle quasi maximum likelihood estimator (QMLE) of β and ξ,

$$\hat{\theta}^{(\delta_1)} = \arg\min_{\theta\in\Theta} R_m(\theta), \tag{9}$$

where $R_m(\theta) = \log \hat{G}_{uu}(\theta;\delta_1)$ is the concentrated likelihood function. Notice that $\hat{\theta}^{(\delta_1)}$ is implicitly a NBGNLS estimator and in the sequel we will retain this label for the QMLE so as to maintain a clear connection with the NBNLS estimator in (6). Finally, we relax the definition of the weighting parameter and introduce the general class of NBWNLS estimators

$$\hat{\theta}^{(\delta)} = \arg\min_{\theta\in\Theta} W_m(\theta) \text{ with } W_m(\theta) = \log \hat{G}_{uu}(\theta;\delta). \tag{10}$$

Said otherwise, the NBNLS and NBGNLS estimators are particular cases of the NBWNLS estimator that correspond to $\delta = 0$ (i.e. no weight) and $\delta = \delta_1$, respectively.

3. Limit theory

In this section we present the limit theory of the general class of infeasible NBWNLS estimators $\hat{\theta}^{(\delta)}$ and discuss the particularities of the NBNLS and NBGNLS estimators when necessary. As Nielsen (2005) and Christensen and Nielsen (2006), we work under several assumptions commonly used in this literature.

Assumption 2. *The elements of the spectral density $f_z(\lambda)$ satisfy*

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

as $\lambda \to 0^+$ and for some $\alpha \in (0,2]$ with $\vartheta_0 = (\delta_{01}, \delta_{02} + \xi_0)'$, matrix G^0 finite, real, symmetric and where $G_{ab}^0 = G_{ba}^0 = 0$.

Assumption 2 imposes a smoothness condition on the spectral density matrix $f_z(\lambda)$ and specifies the properties of G^0, which imply a zero-coherence condition that applies in the vicinity of the origin. As argued in Nielsen (2005), $G_{ab}^0 = G_{ba}^0 = 0$ is a less restrictive assumption than the traditional orthogonality condition encountered in linear (see e.g. Robinson and Hidalgo 1997) and non-linear (see Hualde 2014) 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.

Assumption 3. *The sequence $z_t = (u_t, x_t)$ 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 $\mathbb{E}(\varepsilon_t\varepsilon_t'|F_{t-1}) = I_2$ 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 $\{\varepsilon_s, s \le t\}$.

Assumption 4. *In a neighborhood of the origin, $A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$ is differentiable and*

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

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

Assumptions 3 and 4 are standard in this literature since Robinson (1995). The former imposes uniformly square integrable martingale-difference innovations with constant conditional variance, while the latter implies $\partial A_{a.}(\lambda)/\partial\lambda = O(\lambda^{-\vartheta_{0a}-1})$ 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 3 and 4 we have $f_z(\lambda) = (2\pi)^{-1}A(\lambda)A(\lambda)^*$.

Assumption 5. *As $n \to \infty$, the bandwidth parameter $m = o(n)$ and $\alpha \in (0, 2]$ jointly satisfy*

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

The bandwidth expansion rate is restricted by Assumption 5 as one needs m to tend to ∞ as $n \to \infty$ but at a slower rate so as to remain in a neighborhood of the origin where the behaviour of the spectral density is of interest. Theoretically, the upper bound on the bandwidth parameter m is $n^{4/5}$ (i.e. $\alpha = 2$) but in practice a too small bandwidth increases the variance of the estimator while a too large m generally increases the bias.

Assumption 6. *The weight function $\phi(\lambda) = \lambda^{2\delta}$ as $\lambda \to 0^+$ and the weighting parameter δ satisfies*

$$\frac{1}{2}(\delta_2 + \delta_1 - 1/2) < \delta \leq \delta_1. \tag{11}$$

Assumption 6 restricts the weighting parameter as a function of the long memory parameters as discussed in Nielsen (2005). When $\delta_1 + \delta_2 < 1/2$, $\delta = 0$ is covered by our theory and consequently the NBWNLS estimator is equivalent to the NBNLS one and should be uniformly Gaussian over the weighting parameter space. When $\delta = \delta_1$, this assumption restricts the cointegration strength to $\delta_2 - \delta_1 < 1/2$, which is consistent with Assumption 1. Under slightly weaker versions of the assumptions above, which are generally stated in the literature, we establish the consistency of our NBWNLS class of estimators.

Theorem 1. *Let Assumptions (1)-(5) be satisfied and assume $0 \leq \delta \leq \delta_1$. Then, as $n \to \infty$,*

$$\lambda_m^{\delta_{01}-\delta_{02}} \begin{pmatrix} 1 \\ \log \lambda_m \end{pmatrix}' \begin{pmatrix} \hat{\beta}^{(\delta)} - \beta_0 \\ \hat{\xi}^{(\delta)} - \xi_0 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where $\hat{\theta}^{(\delta)} = (\hat{\beta}^{(\delta)}, \hat{\xi}^{(\delta)})'$.

In addition, the central limit theorem for $\hat{\beta}^{(\delta)}$ and $\hat{\xi}^{(\delta)}$ is stated below.

Theorem 2. *Under Assumptions* (1)-(6)*, as* $n \to \infty$*,*

$$\sqrt{m}\lambda_m^{\delta_{01}-\delta_{02}} \begin{pmatrix} 1 \\ \log \lambda_m \end{pmatrix}' \begin{pmatrix} \hat{\beta}^{(\delta)} - \beta_0 \\ \hat{\xi}^{(\delta)} - \xi_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 \\ \beta_0^{-1} \end{pmatrix} \mathcal{N}(0, E^{-1}FE^{-1}).$$

where $\hat{\theta}^{(\delta)} = (\hat{\beta}^{(\delta)}, \hat{\xi}^{(\delta)})'$ *and*

$$E = \frac{2G_{xx}}{G_{uu}(1 - 2\delta_{02} + 2\delta)},$$
$$F = \frac{2G_{xx}}{G_{uu}(1 + 4\delta - 2\delta_{01} - 2\delta_{02})}.$$

Sketch of proof. The proof of Theorem 2 relies on Theorem 1 and the usual Taylor series expansion for extremum estimators

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

where $||\bar{\theta} - \theta_0|| \leq ||\hat{\theta} - \theta_0||$. Since the joint limit distribution of $\hat{\beta}$ and $\hat{\xi}$ is singular, it is the same up to a constant term β_0^{-1}. Focusing on β, one can show that

$$\sqrt{m}\lambda_m^{\delta_{01}-\delta_{02}}(\hat{\beta} - \beta_0) = \left(\lambda_m^{2(\delta_{02}-\delta)} \frac{\partial^2 W_m(\theta)}{\partial \beta \partial \beta}\Big|_{\bar{\theta}}\right)^{-1} \left(\sqrt{m}\lambda_m^{\delta_{01}+\delta_{02}-2\delta} \frac{\partial W_m(\theta)}{\partial \beta}\Big|_{\theta_0}\right)$$

has the stated distribution, $\sqrt{m}\lambda^{\delta_{01}-\delta_{02}}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, E^{-1}FE^{-1})$, if by application of Cramer-Wold theorem

$$\sqrt{m}\lambda_m^{\delta_{01}+\delta_{02}-2\delta} \frac{\partial W_m(\theta)}{\partial \beta}\Big|_{\theta_0} \xrightarrow{d} \mathcal{N}(0, F)$$

and

$$\lambda_m^{2\delta_{02}-2\delta} \frac{\partial^2 W_m(\theta)}{\partial \beta \partial \beta}\Big|_{\bar{\theta}} \xrightarrow{p} E.$$

A similar reasoning allows one to show that $\sqrt{m}\lambda_m^{\delta_{01}-\delta_{02}} \log(\lambda_m)^{-1}(\hat{\xi} - \xi_0) \xrightarrow{d} \beta_0^{-1}\mathcal{N}(0, E^{-1}FE^{-1})$. The complete proofs of Theorems 1 and 2 are given in Appendix A and B respectively.

Unsurprisingly, we recover the standard convergence rate $\sqrt{m}\lambda_m^{\delta_{01}-\delta_{02}}$ for $\hat{\beta}^{(\delta)}$ since its limit theory does not depend on ξ_0 (see e.g. Robinson and Marinucci 2003, Nielsen 2005, Christensen and Nielsen 2006). Besides, its convergence rate approaches $\sqrt{n}$ when the cointegration strength is close to $1/2$ and it remains higher than the standard semi-parametric $\sqrt{m}$ rate in all other cases. The limit theory of $\hat{\xi}^{(\delta)}$ is however non-standard. It always converges faster than $\hat{\beta}^{(\delta)}$ and can be superconsistent when $\delta_{02} - \delta_{01}$ is close to $1/2$. This is a peculiar 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 $\sqrt{n}$). Note also that the limit distribution of $\hat{\xi}^{(\delta)}$ depends on β_0. Consequently, although Gaussian and hence simple to use for inference, the joint limit distribution of $\hat{\beta}^{(\delta)}$ and $\hat{\xi}^{(\delta)}$ is singular. In contrast, the limit distribution of Hualde (2014)'s estimator, which is also singular, involves a modified version of the type II fractional Brownian motion, which makes it hardly suitable for inference.

When Assumption 6 holds, one can compare the relative asymptotic efficiency of $\hat{\theta}^{(0)}$ and $\hat{\theta}^{(\delta)}$. Our NBWNLS estimators are always asymptotically more efficient than the NBNLS ones except for the case $\delta_1 = 0$ as their variances become identical. The maximum gain in efficiency is actually reached when $\delta = \delta_1$, i.e. for the NBGNLS estimator. In this case, the asymptotic relative efficiency of the weighted and unweighted estimators comes down to

$$\frac{\mathbb{V}(\hat{\theta}^{(0)})}{\mathbb{V}(\hat{\theta}^{(\delta_1)})} = \frac{(1-2\delta_2)^2}{(1-2\delta_2)^2 - 4\delta_1^2}.$$

A second advantage of the NBGNLS estimator over its two competitors is that simple inference is possible for the whole stationary parameter space (see Assumption 6).

4. Feasible narrow-band generalized nonlinear least squares

The main issue with the NBWNLS approach is that $\hat{\theta}^{(\delta)}$ is infeasible in practice since δ is unknown. Feasible estimates $\hat{\delta}$ are available without knowledge of the memory parameters, e.g. based on residuals from (3), but they have to fulfill Assumption 6 as well as the following requirement

Assumption 7. *As $n \to \infty$, the estimate of the weighting parameter δ satisfies*

$$(\log n)(\hat{\delta} - \delta) \xrightarrow{p} 0.$$

In particular, in view of Theorem 2, the most asymptotically efficient feasible NBWNLS estimator is $\hat{\theta}^{(\hat{\delta}_1)}$ provided that one can consistently estimate the weighting parameter δ_1 such that Assumption 7 is fulfilled. In the balanced cointegration framework, Nielsen (2005) relies on results from Velasco (2003) to state that $\hat{\delta}_1 - \delta_1 = O(m^{-1/2})$ for carefully chosen m and hence Assumption 7 holds. But unlike Nielsen (2005), in order to estimate δ_1 in our unbalanced cointegration framework one requires consistent estimators not only for β but also for ξ. For this reason, we propose to use in a first step the NBNLS estimator $\hat{\theta}^{(0)}$, whose consistency is proven in Appendix A as a Corollary of Theorem 1.

Corollary 1. *Let Assumptions (1)-(5) be satisfied and assume $\delta = 0$. Then, as $n \to \infty$,*

$$\lambda_m^{\delta_1-\delta_2} \begin{pmatrix} 1 \\ \log \lambda_m \end{pmatrix}' \begin{pmatrix} \hat{\beta}^{(0)} - \beta_0 \\ \hat{\xi}^{(0)} - \xi_0 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where $\hat{\theta}^{(0)} = (\hat{\beta}^{(0)}, \hat{\xi}^{(0)})'$.

Since $\hat{\xi}^{(0)}$ converges faster than $\hat{\beta}^{(0)}$, it is innocuous for $\hat{\delta}_1$'s rate of convergence. Consequently, for carefully chosen m, Assumption 7 will always be satisfied if one uses the NBNLS residuals to estimate semi-parametrically, in a second step, δ_1 in an unbalanced framework. The asymptotic distribution of the feasible NBWNLS estimator is given in Theorem 3.

Theorem 3. *Under Assumptions 1-6 and 7, as $n \to \infty$, Theorem 2 holds when the weighting parameter δ is replaced by $\hat{\delta}_1$.*

For a proof of Theorem 3 see Appendix C. It indicates that the error arising from the estimation of the weighting parameter does not impact the limit distribution of the parameters of interest stated in Theorem 2.

5. Monte Carlo experiments

This section discusses the finite sample performance of the NBWNLS class of estimators by means of Monte Carlo simulations. We generate a fractionally cointegrated system according to (3) with $\beta = 1$, $\xi = 0.1$, $u_t \sim \mathcal{N}(0, 1)$, and analyze three levels of cointegration strength $\nu_0 = \{0.35, 0.15, 0.05\}$ where δ_2 is fixed to 0.35. We generate 10000 replications for sample sizes $n = \{256, 512, 1024\}$ and fix the bandwidth parameter to $m = \{[n^{0.6}], [n^{0.75}]\}$. Besides, in the case of the long run parameter, β, we compare our three specifications (NBNLS, NBGNLS and NBWNLS) with two misspecified competitors, i.e. the NBLS estimator of Robinson (1994a) and the NBGLS estimator of Nielsen (2005). The NBGNLS estimator involves a weighting parameter $\hat{\delta}_1$ obtained in two steps. First, one gathers the residuals from the NBNLS estimation of β and then estimates δ_1 by a simple local Whittle procedure à la Robinson (1995).[2] The same bandwidth is used for the entire estimation procedure. At the same time, for the NBWNLS estimator the weighting parameter is arbitrarily fixed to $\delta_1/2$.

Tables 1 and 2 report the bias, variance and root-mean-square error (RMSE) for all the estimators under analysis. First, we discuss the estimates of β. Notice that in this unbalanced setup our NBWNLS class of estimators performs better than the existing balanced estimators. The NBLS and NBGLS estimators

[2]Similarly, the NBGLS estimator uses NBLS residuals in its first step.

Table 1: Simulation results for $m = \lfloor n^{0.6} \rfloor$, $\beta = 1$ and $\xi = 0.1$

$m = [n^{0.6}]$	$\hat{\beta}_{LS}$	β_{GLS}	$\hat{\beta}^{(0)}$	$\hat{\xi}^{(0)}$	$\hat{\beta}^{(\hat{\delta}_1)}$	$\hat{\xi}^{(\hat{\delta}_1)}$	$\hat{\beta}^{(\hat{\delta})}$	$\hat{\xi}^{(\hat{\delta})}$	$\hat{\beta}^{(\check{\delta})}$	$\hat{\xi}^{(\check{\delta})}$
$n = 256$										
					$\nu_0 = 0.35$					
Bias	-0.207	-0.199	0.002	0.008	0.003	0.009	0.002	0.008	0.003	0.009
Variance	0.005	0.006	0.031	0.007	0.032	0.008	0.031	0.007	0.031	0.008
RMSE	0.219	0.213	0.175	0.086	0.179	0.089	0.175	0.086	0.177	0.088
					$\nu_0 = 0.15$					
Bias	-0.207	-0.175	0.014	0.013	0.022	0.017	0.017	0.014	0.024	0.018
Variance	0.012	0.01	0.059	0.021	0.121	0.023	0.085	0.022	0.128	0.024
RMSE	0.233	0.202	0.244	0.146	0.349	0.152	0.293	0.149	0.358	0.156
					$\nu_0 = 0.05$					
Bias	-0.205	-0.161	0.019	0.011	0.027	0.02	0.019	0.014	0.029	0.021
Variance	0.02	0.014	0.094	0.043	0.079	0.034	0.078	0.035	0.090	0.038
RMSE	0.25	0.201	0.308	0.207	0.283	0.187	0.28	0.187	0.301	0.197
$n = 512$										
					$\nu_0 = 0.35$					
Bias	-0.241	-0.229	-0.004	0.005	-0.003	0.006	-0.004	0.005	-0.003	0.006
Variance	0.003	0.003	0.018	0.003	0.019	0.003	0.018	0.003	0.018	0.003
RMSE	0.247	0.235	0.134	0.056	0.136	0.057	0.134	0.056	0.136	0.057
					$\nu_0 = 0.15$					
Bias	-0.242	-0.202	0.008	0.009	0.009	0.01	0.008	0.009	0.010	0.010
Variance	0.007	0.006	0.042	0.01	0.037	0.009	0.037	0.009	0.038	0.009
RMSE	0.256	0.215	0.205	0.102	0.192	0.095	0.194	0.096	0.196	0.097
					$\nu_0 = 0.05$					
Bias	-0.241	-0.186	0.016	0.01	0.017	0.013	0.014	0.011	0.019	0.014
Variance	0.013	0.008	0.073	0.021	0.053	0.016	0.055	0.016	0.055	0.016
RMSE	0.267	0.207	0.27	0.147	0.231	0.125	0.235	0.129	0.235	0.128
$n = 1024$										
					$\nu_0 = 0.35$					
Bias	-0.274	-0.254	0	0.006	0	0.006	0	0.006	0.000	0.006
Variance	0.002	0.002	0.01	0.001	0.011	0.001	0.01	0.001	0.011	0.001
RMSE	0.277	0.257	0.102	0.037	0.103	0.037	0.102	0.037	0.103	0.037
					$\nu_0 = 0.15$					
Bias	-0.274	-0.227	0.006	0.006	0.006	0.007	0.005	0.007	0.008	0.008
Variance	0.004	0.003	0.031	0.006	0.026	0.005	0.027	0.005	0.027	0.005
RMSE	0.282	0.234	0.176	0.075	0.162	0.069	0.164	0.07	0.164	0.070
					$\nu_0 = 0.05$					
Bias	-0.274	-0.21	0.018	0.008	0.014	0.009	0.013	0.008	0.017	0.010
Variance	0.009	0.005	0.063	0.013	0.041	0.009	0.044	0.01	0.043	0.009
RMSE	0.29	0.221	0.252	0.114	0.203	0.094	0.211	0.098	0.208	0.096

Note: The results are based on $I = 10000$ replications. $\hat{\beta}_{LS}$ and $\hat{\beta}_{GLS}$ stand for the NBLS estimator of Robinson (1994a) and the NBGLS estimator of Nielsen (2005), respectively. The NBWNLS estimators are obtained by fixing the weighting parameter to $\delta_1/2$. Besides, $\nu_0 = 0.35$ corresponds to $\delta_1 = 0$ and in this case the NBWNLS estimators come down to the NBNLS ones (indicated by '-').

Table 2: Simulation results for $m = \lfloor n^{0.75} \rfloor$, $\beta = 1$ and $\xi = 0.1$

$m = [n^{0.6}]$	$\hat{\beta}_{LS}$	β_{GLS}	$\hat{\beta}^{(0)}$	$\hat{\xi}^{(0)}$	$\hat{\beta}^{(\hat{\delta}_1)}$	$\hat{\xi}^{(\hat{\delta}_1)}$	$\hat{\beta}^{(\delta)}$	$\hat{\xi}^{(\delta)}$	$\hat{\beta}^{(\check{\delta})}$	$\hat{\xi}^{(\check{\delta})}$
$n = 256$										
					$\nu_0 = 0.35$					
Bias	-0.164	-0.152	-0.012	0.006	-0.012	0.006	-0.012	0.006	-0.012	0.006
Variance	0.004	0.004	0.009	0.003	0.009	0.003	0.009	0.003	0.009	0.003
RMSE	0.176	0.166	0.094	0.056	0.095	0.057	0.094	0.056	0.095	0.057
					$\nu_0 = 0.15$					
Bias	-0.163	-0.111	-0.014	0.007	-0.013	0.007	-0.013	0.007	-0.012	0.008
Variance	0.008	0.006	0.011	0.008	0.01	0.007	0.01	0.007	0.010	0.007
RMSE	0.186	0.136	0.106	0.091	0.099	0.083	0.1	0.084	0.099	0.085
					$\nu_0 = 0.05$					
Bias	-0.162	-0.09	-0.018	0.007	-0.013	0.009	-0.014	0.008	-0.012	0.011
Variance	0.013	0.007	0.014	0.015	0.01	0.01	0.011	0.011	0.010	0.010
RMSE	0.198	0.124	0.12	0.122	0.101	0.099	0.104	0.104	0.101	0.102
$n = 512$										
					$\nu_0 = 0.35$					
Bias	-0.195	-0.177	-0.009	0.006	-0.009	0.006	-0.009	0.006	-0.009	0.006
Variance	0.002	0.002	0.005	0.001	0.005	0.001	0.005	0.001	0.005	0.001
RMSE	0.201	0.184	0.072	0.037	0.073	0.037	0.072	0.037	0.073	0.037
					$\nu_0 = 0.15$					
Bias	-0.194	-0.13	-0.01	0.006	-0.01	0.006	-0.01	0.006	-0.009	0.006
Variance	0.005	0.004	0.008	0.004	0.007	0.003	0.007	0.003	0.007	0.003
RMSE	0.207	0.144	0.089	0.064	0.082	0.057	0.083	0.059	0.083	0.059
					$\nu_0 = 0.05$					
Bias	-0.196	-0.108	-0.01	0.009	-0.009	0.008	-0.009	0.008	-0.009	0.008
Variance	0.009	0.004	0.011	0.008	0.007	0.005	0.008	0.006	0.007	0.005
RMSE	0.217	0.125	0.106	0.091	0.085	0.071	0.088	0.076	0.086	0.074
$n = 1024$										
					$\nu_0 = 0.35$					
Bias	-0.224	-0.197	-0.007	0.005	-0.007	0.005	-0.007	0.005	-0.007	0.005
Variance	0.002	0.001	0.003	0.001	0.003	0.001	0.003	0.001	0.003	0.001
RMSE	0.228	0.201	0.055	0.025	0.056	0.025	0.055	0.025	0.056	0.025
					$\nu_0 = 0.15$					
Bias	-0.223	-0.148	-0.007	0.006	-0.007	0.006	-0.007	0.006	-0.007	0.006
Variance	0.003	0.002	0.006	0.002	0.004	0.002	0.005	0.002	0.004	0.002
RMSE	0.231	0.154	0.075	0.048	0.066	0.041	0.067	0.043	0.067	0.043
					$\nu_0 = 0.05$					
Bias	-0.224	-0.124	-0.008	0.007	-0.007	0.006	-0.008	0.006	-0.007	0.006
Variance	0.006	0.002	0.009	0.005	0.005	0.003	0.006	0.003	0.005	0.003
RMSE	0.237	0.133	0.096	0.072	0.071	0.053	0.075	0.058	0.072	0.055

Note: The results are based on $I = 10000$ replications. $\hat{\beta}_{LS}$ and $\hat{\beta}_{GLS}$ stand for the NBLS estimator of Robinson (1994a) and the NBGLS estimator of Nielsen (2005), respectively. The NBWNLS estimators are obtained by fixing the weighting parameter to $\delta_1/2$. Besides, $\nu_0 = 0.35$ corresponds to $\delta_1 = 0$ and in this case the NBWNLS estimators come down to the NBNLS ones (indicated by '-').

assume that $\xi = 0$ and exhibit a strong negative bias. This follows from their limit behaviour in an unbalanced framework, where the bias of $\hat{\beta}_{GLS}$ is of order $O_p(\lambda_m^{\xi_0})$ as $m \to \infty$ since

$$\hat{\beta}_{GLS} = \beta \frac{\sum_j^m \lambda_j^{2\delta_{01}+\xi_0} I_{xx}(\lambda_j)}{\sum_j^m \lambda_j^{2\delta_{01}} I_{xx}(\lambda_j)} + O_p(\lambda_m^{\nu_0+\xi_0}) \sim O_p(\lambda_m^{\xi_0}) + O_p(\lambda_m^{\nu_0+\xi_0}). \tag{12}$$

By a similar reasoning and setting δ_1 to zero in the previous equation one can show that the asymptotic bias of $\hat{\beta}_{LS}$ is of the same order. The finite-sample bias of these estimators is increasing in the sample size and consequently so is the RMSE. By contrast, our estimators exhibit very small finite sample bias. It diminishes with n and so does the RMSE. At the same time, as expected given our limit theory, the variance of all estimators increases when the cointegration strength reduces and diminishes with the sample size.

Focusing on our three estimators, one notices that for $\nu_0 = 0.35$, i.e. $\delta_1 = 0$, there is almost no difference between them regardless of the sample size. As ν_0 decreases, the NBGNLS and NBWNLS estimators become more efficient than the NBNLS one. The performance of these two estimators also improves relatively to that of the NBNLS when n increases. In particular, the NBGNLS estimator appears to be slightly more efficient than the NBWNLS estimator for large sample sizes and low cointegration strength consistently with the asymptotic theory.

We now focus on the estimates of ξ. All three estimators exhibit similar minor bias levels but the variances of $\xi^{(\hat{\delta}_1)}$ and $\xi^{(\delta)}$ are smaller relatively to that of $\xi^{(0)}$. The variance of the unbalance parameter estimates is always much lower than that of the corresponding long-run parameter estimates. As expected, this finite-sample finding goes along the lines of the rates of convergence identified in the asymptotic theory (see Theorem 2). All results are robust to the choice of the bandwidth parameter.

6. Empirical illustration

There is a huge literature in finance that focuses on return predictability. On one hand, multiyear predictability patterns have been studied in association with the default spread, the P/E ratio and the consumption-wealth ratio. But there is extensive evidence that the predictive power of such traditional indicators has been decreasing since the 1990s (see e.g. Fama and French 1988, Campbell and Shiller 1988, Lettau and Ludvigson 2001). On the other hand, the role of stock market volatility in return forecasting, also known as volatility feedback effect, has been deeply investigated in the existing literature using both conditional and realized measures to proxy risk (see Ghysels et al. 2005, and the references therein). Merton's Intertemporal CAPM suggests that market's conditional variance should have a positive impact on the conditional expected excess return. Indeed, if volatility is priced, an anticipated increase in volatility would raise the required rate of return, in turn necessitating an immediate stock-price decline to allow for higher future returns. Existing empirical results are however often inconclusive and sometimes conflicting (see Nelson 1991, Campbell and Hentschel 1992, Bali and Peng 2006, among others).

In this context, recent studies have theoretically shown that the variance risk premium, i.e. the difference between the implied and realized variance series, should be a useful return predictor (see e.g. Bollerslev et al. 2009, Drechsler and Yaron 2011, Bollerslev et al. 2012). Interestingly, the long-run risk model developed by Bollerslev et al. (2012) accounts for the persistent nature of the unobserved integrated variance, IV_t. At the same time, the variance swap rate, mirroring the definition of the squared VIX volatility index, is defined as the risk-neutralized expectation of IV_t and inherits the long memory property of volatility. A direct implication of this fractionally integrated framework is the presence of cointegration between VIX_t^2 and IV_t, with the variance risk premium, $VRP_t = VIX_t^2 - IV_t$, exhibiting less memory than the volatility series.[3] VRP_t is expected to increase if uncertainty about volatility increases. Hence, the model theoretically predicts that the cross-correlations between the variance risk premium and future returns should be positive, reflecting the premium for bearing volatility risk. These insights are corroborated by empirical evidence on the SP500 index mainly. Indeed, Bollerslev et al. (2013) exploit this long-run relationship to extract the variance risk premium and show that it results in nontrivial return predictability over within-year horizons. Further empirical evidence that variance risk premium results in strong return predictability is documented in Bollerslev et al. (2014; 2015).

The theoretical and empirical results aforementioned hinge on model-free measures of integrated and implied (unobserved) variances. But depending on the accuracy of these measurements, the long-run dynamics of the two variance proxies might spuriously diverge. Indeed, empirical differences in the integration orders of the volatility series inherent to the finite sample context may occur and pollute the estimate of the long-run relationship. We therefore conjecture that our general unbalanced cointegration framework is more suitable than existing ones as it implicitly avoids such biases that are likely to appear when estimating the variance risk premium.

To estimate the risk-return relation we essentially follow Bollerslev et al. (2013) but rely on data sampled at monthly frequency. Our data set runs from January 2014 to March 2019 and consists in monthly realized measures of volatility for four stock market indexes, S&P500, Dow Jones 30, Russell 2000 and NASDAQ 100. For each index, the realized variance (RV), the realized kernel (RK), the median realized variance ($medRV$) and the bipower variation (BV) are used as variance proxies to control for the presence of microstructure noise or jumps that might affect the results.[4] As we construct a forward-looking measure of VRP_t, we align beginning of month VIX CBOE indexes, intended to represent a 30-days target framework for expected variance, with the monthly realized measures.[5]

[3]The fast exponential decay in the autocorrelation structure of VRP_t and the slow hyperbolic decay in the autocorrelations of volatility series are consistent with the empirical findings of Bandi and Perron (2006) and Nielsen (2007) among others.

[4]All realized measures are based on 5-minutes intraday prices and collected from Oxford-Man Institute's realized library website.

[5]All VIX data are collected from the CBOE website http://www.cboe.com/.

More appropriate for Whittle estimation and linear regressions, the log transforms $vix_t^2 = \log(VIX_t^2)$ and $rm_t = \log(\widehat{IV}_t)$ are used in the following, with $\widehat{IV}$ standing for any of the four realized measures of variance. Figure 1 reports the vix_t^2 versus two rm_t measures, RV and BV, in the $S\&P500$ case. A visual inspection clearly reveals that all series exhibit long term swings and similarities in their dynamics. A local Whittle estimation of the long memory parameters confirms that all series are fractionally integrated with estimates confined in the stationary region between 0.3 and 0.5. The bandwidth retained for the local Whittle estimator (and all narrow-band estimations hereafter) by using the informal " optimal" bandwidth selection procedure in Robinson (2008b) is $m = \lfloor n^{0.6} \rfloor$ where $n = 253$ for all stock market indexes. This intermediate bandwidth should ensure that our estimates are not contaminated by high frequencies and that they exhibit a reasonable finite sample variance. These results suggest that a long-run relationship may exist between the implied and the realized volatility in the form of $rm_t = \beta vix_t^2(\xi) + u_t$. In contrast to Bollerslev et al. (2013), our general unbalanced fractional cointegration framework allows for (possibly small) differences between vix_t^2 and rm_t long memory parameters. We also implement the (possibly misspecified) NBLS and GNBLS estimators of β. The results are reported in Table 3.

All the estimates of the long-run parameter β are statistically significant apart from the cases where Assumption 6 is violated and hence the standard errors are not available. Our NBNLS and NBGNLS estimates are quite similar, with values close to one, as expected from economic theory. However, the NBLS and NBGLS estimates are systematically lower. This finding appears to be a consequence of the integration order unbalance between the two types of variance proxy. Indeed, the estimates of the unbalance parameter ξ are always positive and significant, indicating that an unbalanced cointegration framework is more appropriate. In line with equation (12), we can hence conjecture that even a quite small unbalance (estimated to be around 0.1 here) may have a large impact on the long-run parameter β (a drop of about 0.15). Besides, the integration order of the long-run regression errors does not seem to be statistically different from zero, supporting the fact that $vrp_t = \beta vix_t^2(\xi) - rm_t$ exhibits less persistence than the (rebalanced) variance series.

We can now express the risk-return relationship in regression form as

$$r_{t+1} = \alpha_0 + \alpha_1 vrp_t + u_{t+1},$$

where r_{t+1} is the one-step-ahead monthly stock market log-return. As one might expect the association between risk and return to be more of a long-run than a short-run phenomenon, we draw on Bollerslev et al. (2013) and focus on the long-run component of each variable. Using time-domain band-pass filters, we extract from the observed series the specific low-frequency region up to the Fourier frequency $\lambda = 0.65$, which corresponds to the chosen bandwidth and equivalently to a periodicity of about 9 months. For this, we set the truncation parameter to $k = 12$, resulting in a loss of one year of observations at both ends of the sample. Each filtered series is given by $y_t^{low} = \sum\limits_{i=-k}^{k} a_i L^i y_t$, where

$$a_i = \begin{cases} \frac{\sin(i\lambda)}{i\pi} - \left(\frac{\lambda}{\pi} + 2\sum_{j=1}^{k} \frac{\sin(j\lambda)}{j\pi} - 1\right)(2k+1)^{-1}, i = \pm 1, \ldots, \pm k, \\ 1 - \sum_{h=-k}^{-1} a_h - \sum_{h=1}^{k} a_h, i = 0. \end{cases}$$

Figure 2 displays the low-pass filter returns and variance risk premium based on the realized kernel variance proxy for S&P500. This illustration emphasizes the comovement between the series and motivates the following low-frequency regression analysis.

The slope estimates of the regressions fitted with low-pass filter variables are reported in Table 4. The upper panel shows that the long-run component of the variance risk premia estimated in the unbalanced fractional cointegration (UFC) framework explains a nontrivial fraction of the low-pass returns at an intermediate 9.5 months horizon for all realized measures. More importantly, the slope parameters are positive and consistent with the trade-off between returns and variance risk premium as documented in Bollerslev et al. (2012; 2013; 2014). $\hat{\alpha}_1$ estimates are reported in Figure 3 for the Fourier frequency λ ranging in $[0.01, 0.9]$ in the spirit of a robustness analysis. The predictive power of the UFC variance risk premia is stable and reaches its maximum when the low-pass filter is defined in terms of frequencies between 0.6 and 0.8. This covers a horizon of 8 to 10.5 months and supports the usefulness of our approach for return prediction at long horizons where existing literature fails to find significant predictive power.

The middle panel reports the slope parameters in the case where the variance risk premia are computed following the naive approach in Bollerslev et al. (2013), i.e. imposing $\beta = 1$ and $\xi = 0$. Neglecting the possible unbalance of the system, their estimates of the long run parameter are close to unity and hence $vrp_t = vix_t^2 - rm_t$. The low-pass filtered naive vrp_t seems, however, less informative for the returns of the four indices in our monthly sample ($\hat{\alpha}_1$ estimates are systematically lower than those estimated with the UFC approach). In contrast, the results reported in the lower part of the table correspond to more conventional regressors, i.e. the low-pass variance risk measures. The estimates are now negative and rarely significant, in line with existing literature on the risk-return puzzle. All in all, the long-run component of variance risk premia estimated in an unbalanced framework seems to carry more useful information about future returns than competing risk proxies.

Conclusion

This paper introduces a class of narrow band weighted non-linear least squares estimators that allow for the joint estimation of the long run and unbalance parameters in a bivariate unbalanced fractionally cointegrated equation. Our estimates are shown to be asymptotically Gaussian with singular distributional properties. Moreover, they exhibit faster than regular $\sqrt{m}$ rate of convergence. A particular estimator of our class is the feasible narrow-band generalized nonlinear least squares. It offers the same appealing limit properties over the whole stationary region, thereby allowing for straightforward statistical inference. In contrast, existing balanced estimators are inconsistent in presence of unbalance in the integration orders of the observables. This supports the need for a general to specific approach to cointegration as the balanced framework is a specific case of the unbalanced one. Our contribution meets this challenge. By means of Monte Carlo simulations we further show the good finite sample properties of the proposed estimators. Finally, they are used to provide new perspectives on the risk-return relationship on financial stock markets. In particular, we find that the variance risk premium estimated in an appropriately balanced cointegration system is a better return predictor than existing risk premia measures.

Figure 1: S&P500 VIX (solid lines) and realized volatility measures (dashed lines, secondary axis)

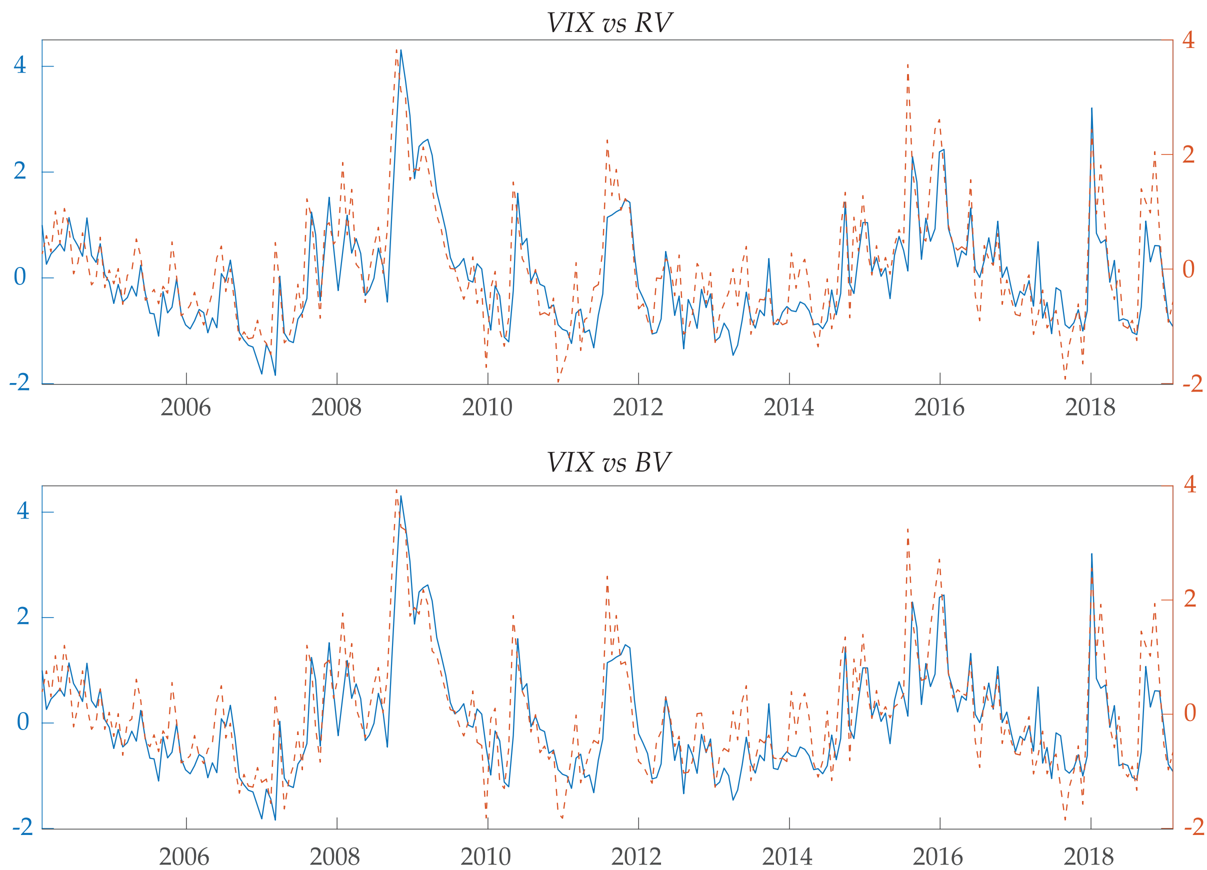

Figure 2: S&P500 r_t^{low} (solid line) and $vrp_{t,RK}^{low}$ (dotted line, secondary axis)

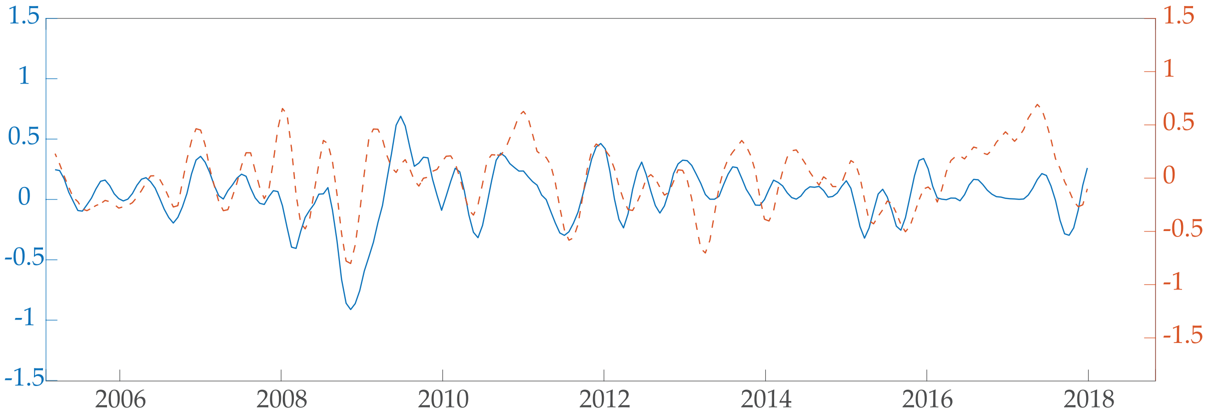

Figure 3: S&P500 slope parameters for ω ranging in $[0.01, 0.9]$

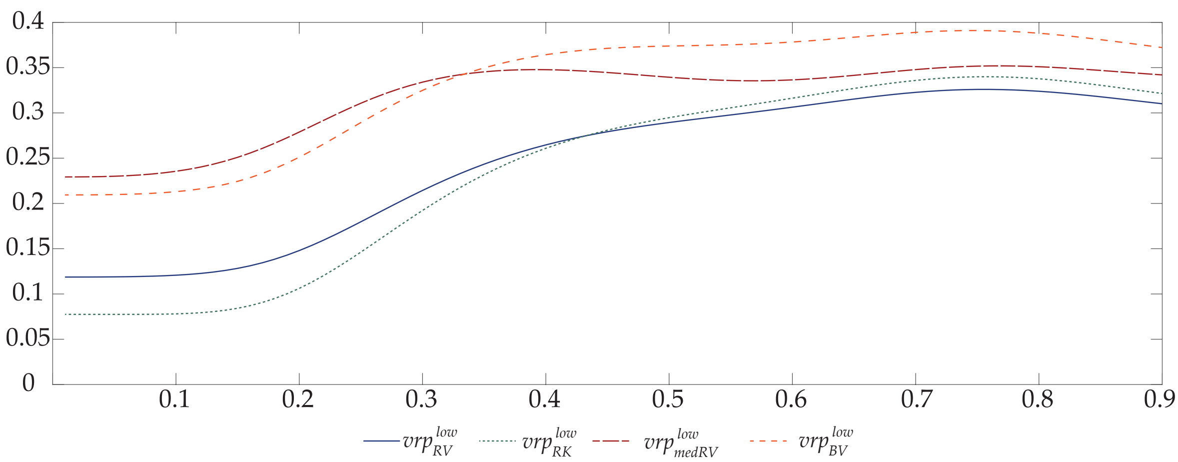

Table 3: Fractional cointegration estimates with $m = \lfloor n^{0.6} \rfloor$

	$\hat{\beta}_{LS}$	$\hat{\beta}_{GLS}$	$\hat{\beta}^{(0)}$	$\hat{\xi}^{(0)}$	$\hat{\beta}^{(\hat{\delta}_1)}$	$\hat{\xi}^{(\hat{\delta}_1)}$	$\hat{\delta}_1$	$\hat{\delta}_2$
				S&P 500				
RV	0.875	0.890	1.061	0.099	1.053	0.095	0.051	0.362
	-	(0.022)	(0.016)	(0.037)	(0.012)	(0.028)	(0.096)	(0.096)
RK	0.888	0.899	1.065	0.093	1.061	0.091	0.031	0.363
	-	(0.019)	(0.012)	(0.028)	(0.010)	(0.025)	(0.096)	(0.096)
medRV	0.873	0.879	1.050	0.095	1.060	0.100	-0.087	0.344
	-	(0.014)	(0.006)	(0.014)	(0.003)	(0.007)	(0.096)	(0.096)
BV	0.876	0.880	1.030	0.083	1.036	0.086	-0.079	0.346
	-	(0.010)	(0.005)	(0.012)	(0.002)	(0.004)	(0.096)	(0.096)
				Dow Jones				
RV	0.853	0.863	0.967	0.064	0.951	0.055	0.076	0.396
	-	(0.027)	(0.031)	(0.067)	(0.012)	(0.026)	(0.096)	(0.096)
RK	0.865	0.874	0.968	0.057	0.948	0.046	0.091	0.407
	-	(0.027)	-	-	(0.012)	(0.026)	(0.096)	(0.096)
medRV	0.847	0.852	0.985	0.077	0.993	0.082	-0.043	0.358
	-	(0.019)	(0.008)	(0.017)	(0.007)	(0.017)	(0.096)	(0.096)
BV	0.850	0.854	0.960	0.062	0.965	0.065	-0.026	0.371
	-	(0.017)	(0.007)	(0.015)	(0.007)	(0.015)	(0.096)	(0.096)
				Russell				
RV	0.817	0.831	0.987	0.105	0.967	0.092	0.080	0.298
	-	(0.035)	(0.026)	(0.057)	(0.020)	(0.043)	(0.096)	(0.096)
RK	0.851	0.855	0.957	0.064	0.962	0.067	-0.027	0.319
	(0.031)	(0.024)	(0.011)	(0.023)	(0.011)	(0.023)	(0.096)	(0.096)
medRV	0.793	0.819	1.031	0.147	1.015	0.137	0.073	0.260
	-	(0.044)	(0.031)	(0.071)	(0.026)	(0.058)	(0.096)	(0.096)
BV	0.823	0.833	0.984	0.098	0.987	0.100	-0.013	0.285
	-	(0.031)	(0.015)	(0.033)	(0.015)	(0.034)	(0.096)	(0.096)
				NASDAQ				
RV	0.856	0.856	0.991	0.076	1.009	0.086	-0.123	0.299
	(0.015)	(0.014)	(0.007)	(0.015)	(0.005)	(0.011)	(0.096)	(0.096)
RK	0.871	0.872	1.002	0.073	1.015	0.080	-0.095	0.310
	(0.016)	(0.016)	(0.007)	(0.016)	(0.006)	(0.014)	(0.096)	(0.096)
medRV	0.838	0.848	1.008	0.097	1.018	0.102	-0.043	0.305
	-	(0.027)	(0.012)	(0.027)	(0.012)	(0.028)	(0.096)	(0.096)
BV	0.848	0.847	0.994	0.083	1.016	0.095	-0.137	0.290
	(0.015)	(0.015)	(0.007)	(0.016)	(0.004)	(0.010)	(0.096)	(0.096)

Note: This table reports the balanced and unbalanced fractional cointegration estimates for the long run equation $rm_t = \beta vix_t^2(\xi) + u_t$, where rm_t stands for realized measures of variance. $\hat{\beta}_{LS}$ and $\hat{\beta}_{GLS}$ are the NBLS estimator of Robinson (1994a) and the NBGLS estimator of Nielsen (2005), respectively. $\hat{\beta}^{(0)}$ and $\hat{\xi}^{(0)}$ correspond to the NBNLS estimator, while $\hat{\beta}^{(\hat{\delta}_1)}$ and $\hat{\xi}^{(\hat{\delta}_1)}$ correspond to the NBGNLS estimator. Standard errors are reported in parentheses.

Table 4: Risk-return regressions

	S&P 500	Russell	Dow Jones	NASDAQ
	UFC approach: $vrp_t = \beta vix_t^2(\xi) - rm_t$			
$vrp_{t,RV}^{low}$	0.315	0.302	0.324	0.347
	(0.101)	(0.077)	(0.063)	(0.122)
$vrp_{t,RK}^{low}$	0.327	0.291	0.415	0.356
	(0.108)	(0.075)	(0.057)	(0.130)
$vrp_{t,medRV}^{low}$	0.342	0.300	0.278	0.290
	(0.101)	(0.077)	(0.052)	(0.102)
$vrp_{t,BV}^{low}$	0.384	0.322	0.354	0.345
	(0.115)	(0.081)	(0.058)	(0.119)
	Naive approach: $vrp_t = vix_t^2 - rm_t$			
$vrp_{t,RV}^{low}$	0.253	0.206	0.266	0.280
	(0.079)	(0.067)	(0.046)	(0.083)
$vrp_{t,RK}^{low}$	0.274	0.209	0.340	0.295
	(0.084)	(0.068)	(0.042)	(0.090)
$vrp_{t,medRV}^{low}$	0.265	0.204	0.225	0.237
	(0.075)	(0.060)	(0.040)	(0.070)
$vrp_{t,BV}^{low}$	0.289	0.220	0.278	0.274
	(0.083)	(0.065)	(0.043)	(0.079)
	Realized measures			
VIX_t^{low}	-0.057	-0.060	0.005	-0.013
	(0.047)	(0.039)	(0.034)	(0.043)
RV_t^{low}	-0.107	-0.111	-0.069	-0.070
	(0.045)	(0.036)	(0.030)	(0.048)
RK_t^{low}	-0.106	-0.107	-0.067	-0.067
	(0.045)	(0.037)	(0.029)	(0.049)
$medRV_t^{low}$	-0.112	-0.115	-0.074	-0.073
	(0.044)	(0.035)	(0.030)	(0.047)
BV_t^{low}	-0.113	-0.115	-0.072	-0.074
	(0.044)	(0.035)	(0.030)	(0.048)

Note: This table reports the slope estimates of univariate regressions of the long run components of the returns r_t on the long run components of variance risk premia and realized measures, respectively. The low-pass filtered series are obtained as in Bollerslev et al. (2013) with the long run frequency band set to $\lambda = 0.65$ and truncation parameter $k = 12$. The standard errors (reported in paranthesis) are based on the HAC covariance matrix estimator. Standard errors are reported in parentheses.

Appendix A: Proof of Theorem 1

Proof. Let θ be the vector of admissible parameter values and θ_0 the vector of true parameter values. For any $b > 0$ and $e > 0$, define the neighborhoods $\Theta^n_\xi(e) = \{\xi : |\xi - \xi_0| < e\}$, $\Theta^n_\beta(b) = \{\beta : |\beta - \beta_0| < b\}$, and their complements $\Theta^c_\xi = \Theta_\xi \backslash \Theta^n_\xi$ and $\Theta^c_\beta = \Theta_\beta \backslash \Theta^n_\beta$ such that $\Theta^n(\varepsilon) = \Theta^n_\xi(\varepsilon^{-1}\lambda_m^{\nu_0}\log(\lambda_m)^{-1}) \times \Theta^n_\beta(\varepsilon^{-1}\lambda_m^{\nu_0})$, and $\Theta^c(\varepsilon) = \Theta\backslash\Theta^n(\varepsilon)$, with $\nu_0 = \delta_{02} - \delta_{01}$ the true cointegration strength. Since $\theta_0 \in \Theta^n(\varepsilon)$, it follows that

$$\Pr\left(\hat{\theta} \in \Theta^c(\varepsilon)\right) = \Pr\left(\inf_{\hat{\theta}\in\Theta^c(\varepsilon)} W_m(\theta) \leq \inf_{\hat{\theta}\in\Theta^n(\varepsilon)} W_m(\theta)\right) \leq \Pr\left(\inf_{\hat{\theta}\in\Theta^c(\varepsilon)} W_m(\theta) - W_m(\theta_0) \leq 0\right),$$

where $\hat{\theta}$ stands for $\hat{\theta}^{(\delta)}$. Accordingly, to prove Proposition 1 it suffices to show that as $n \to 0$,

$$\begin{aligned} S(\theta) &= \left(\log \hat{G}_{uu}(\theta) - \log G_{uu}\right) + \left(\log G_{uu} - \log \hat{G}_{uu}(\theta_0)\right) \\ &= S_1(\theta) + S_2(\theta), \end{aligned}$$

is positive and bounded away from 0 uniformly on $\Theta^c(\varepsilon)$ so that

$$\Pr\left(\inf_{\hat{\theta}\in\Theta^c(\varepsilon)} S(\theta) \leq 0\right) \to 0 \tag{13}$$

where $S(\theta) = W_m(\theta) - W_m(\theta_0)$ and $\hat{G}_{uu}(\theta)$ replaces $\hat{G}_{uu}(\theta;\delta)$ to simplify notations. From the analysis of (Robinson 1995, p. 1635) and the fact that

$$\begin{aligned} \Pr\left(|S_2(\theta)| \leq \varepsilon\right) &= \Pr\left(|\log \hat{G}_{uu}(\theta_0) - \log G_{uu}| \leq \varepsilon\right) \\ &\leq \Pr\left(\left|\frac{\hat{G}_{uu}(\theta_0) - G_{uu}}{G_{uu}}\right| \leq \varepsilon/2\right) \end{aligned} \tag{14}$$

one can immediately show that $S_2(\theta)$ is $o_p(1)$. It remains to study $S_1(\theta)$. A similar reasoning to that in (14), but in terms of $S_1(\theta)$, shows that its treatment is equivalent to the analysis of

$$\hat{G}_{uu}(\theta) - G_{uu} = m^{-1}\sum_{j=1}^{m} \lambda_j^{2\delta} I_{uu}(\lambda_j) - G_{uu}. \tag{15}$$

Accounting for the measurement errors arising from estimating β and ξ,

$$I_{uu}(\lambda) - I^0_{uu}(\lambda) = \lambda_m^{2\nu_0}\lambda_j^{2\xi_0}\tilde{\beta}^2 I_{jxx}(\lambda) - 2\lambda_m^{\nu_0}\lambda_j^{\xi_0}\tilde{\beta}\,\mathrm{Re}(I^0_{jux}(\lambda)),$$

where $\lambda_m^{\nu_0}\lambda_j^{\xi_0}\tilde{\beta} = (\beta_0\lambda_j^{\xi_0} - \beta\lambda_j^{\xi})$, it follows that

$$\begin{aligned}
\hat{G}_{uu}(\theta) - G_{uu} &= \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2\delta} I_{uu}^0(\lambda_j) - 2\tilde{\beta}\frac{G_{ux}}{m}\sum_{j=1}^{m}\lambda_j^{2\delta}\frac{\mathrm{Re}(I_{ux}^0(\lambda_j))}{G_{ux}\lambda_m^{-\nu_0}\lambda_j^{-\xi_0}} + \tilde{\beta}^2\frac{G_{xx}}{m}\sum_{j=1}^{m}\lambda_j^{2\delta}\frac{I_{xx}(\lambda_j)}{G_{xx}\lambda_m^{-2\nu_0}\lambda_j^{-2\xi_0}} - G_{uu} \\
&= \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2\delta} I_{uu}^0(\lambda_j) \\
&- \lambda_m^{\nu_0}\frac{2\tilde{\beta}G_{ux}}{m}\sum_{j=1}^{m}\lambda_j^{-\nu_0+2\delta-2\delta_{01}}\frac{\mathrm{Re}(I_{ux}^0(\lambda_j))}{G_{ux}\lambda_j^{-\delta_{01}-\delta_{02}-\xi_0}} \\
&+ \lambda_m^{2\nu_0}\frac{\tilde{\beta}^2 G_{xx}}{m}\sum_{j=1}^{m}\lambda_j^{-2\nu_0+2\delta-2\delta_{01}}\frac{I_{xx}^0(\lambda_j)}{G_{xx}\lambda_j^{-2(\delta_{02}+\xi_0)}}.
\end{aligned}$$

By Proposition 1 of Nielsen (2005), as $n \to \infty$, the first term satisfies

$$\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2\delta} I_{uu}^0(\lambda_j) - \frac{G_{uu}}{1+2\delta-2\delta_1} \xrightarrow{p} 0$$

since $m^{-1}\sum_{j=1}^{m}\lambda_j^{2\delta} f_{uu}(\lambda_j) - G_{uu}/(1+2\delta-2\delta_1) = o\left(\lambda_m^{1+2\delta-2\delta_1}\right)$ is valid under our Assumptions. By similar manipulations of the other two terms using Lemma 1 of Nielsen (2005) and the analysis of Robinson (1995, p. 1638), one can show that under Assumption 5

$$\hat{G}_{uu}(\theta) - G_{uu} = \frac{G_{uu}}{1+2\delta-2\delta_{01}} - \frac{2\tilde{\beta}G_{ux}}{1+2\delta-\delta_{01}-\delta_{02}} + \frac{\tilde{\beta}^2 G_{xx}}{1+2\delta-2\delta_{02}} - G_{uu} + o_p(1).$$

By Assumptions 1, 2 and 6, the first term is positive and greater than (or equal to) G_{uu}, the second term is null, whereas the last term is bounded away from 0 when $\{\hat{\theta} \in \Theta^c\} \cap \{\hat{\xi} \in \Theta_\xi\}$ or $\{\hat{\theta} \in \Theta^c\} \cap \{\hat{\beta} \in \Theta_\beta\}$ $\forall\ 0 \le \delta \le \delta_{01}$. This completes the proof of (13) and Theorem 1. Note also that the consistency of the NBNLS emerges in the limit case when $\delta = 0$ and that of the NBGNLS when setting $\delta = \delta_1$.

□

Appendix B: Proof of Theorem 2

As an implication of the consistency result proved in Theorem 1, $\hat{\theta}_\delta$ satisfies

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

where $||\bar{\theta} - \theta_0|| \le ||\hat{\theta} - \theta_0||$. Then, we can show that the NBWNLS estimator converges to the stated distribution

$$\sqrt{m}\lambda^{\delta_1-\delta_2}\,\mathrm{diag}(1,\log(\lambda_m)^{-1})(\hat{\theta}-\theta_0)\xrightarrow{d}\begin{pmatrix}1\\ \beta_0^{-1}\end{pmatrix}\mathcal{N}(0,E^{-1}FE^{-1}).$$

if by application of the Cramer-Wold theorem

$$\eta'\sqrt{m}\lambda_m^{\delta_1+\delta_2-2\delta}\,\mathrm{diag}(1,\log(\lambda_m)^{-1})\frac{dW_m(\theta)}{d\theta}\Big|_{\theta_0}\xrightarrow{d}\begin{pmatrix}1\\ \beta_0^{-1}\end{pmatrix}\mathcal{N}(0,\eta'F\eta) \tag{16}$$

and

$$\lambda_m^{2\delta_2-2\delta}\,\mathrm{diag}(1,\log(\lambda_m)^{-2})\frac{\partial^2 W_m(\theta)}{\partial\theta\partial\theta'}\Big|_{\bar{\theta}}\xrightarrow{p}\begin{pmatrix}1\\ \beta_0^{-1}\end{pmatrix}E. \tag{17}$$

B.1. Limit of the score

In this section we investigate the limit of the score in (16), while (17) will be analysed in the next section. Note also that the subscript 0, indicating the true parameter values, will be omitted unless its absence causes confusion.

Proof. The derivative with respect to β is

$$\frac{\partial W_m(\theta_0)}{\partial\beta}=-\frac{2}{m}\sum_{j=1}^{m}\lambda_j^{2\delta-\delta_1-\delta_2}\,\mathrm{Re}\left(\hat{G}_{uu}(\theta_0)^{-1}\lambda_j^{\delta_2+\xi+\delta_1}(I_{jxy}-\beta\lambda_j^{\xi}I_{jxx})\right).$$

where $\hat{G}_{uu}(\theta)$ is used instead of $\hat{G}_{uu}(\theta;\delta)$ to simplify notation. The derivative with respect to ξ is

$$\frac{\partial W_m(\theta_0)}{\partial\xi}=-\frac{2}{m}\beta\sum_{j=1}^{m}\lambda_j^{2\delta-\delta_1-\delta_2}\log\lambda_j\,\mathrm{Re}\left(\hat{G}_{uu}(\theta_0)^{-1}\lambda_j^{\delta_2+\xi+\delta_1}(I_{jxy}-\beta\lambda_j^{\xi}I_{jxx})\right).$$

We proceed as in Robinson (2008a), albeit in a unidimensional framework, and after rearrangements implying only negligible errors (see Lobato 1999, Appendix C) arising from the replacement of $(\Lambda_j I_j\Lambda_j)$ by $P_jI_{ju}P_j^*:=\Lambda_jA(\lambda_j)I_{ju}A(\lambda_j)^*\Lambda_j$ and $\hat{G}_{uu}(\theta_0)$ by G as $|\hat{G}_{uu}(\theta_0)-G|=o_p(1)$, we obtain

$$\eta_1\sqrt{m}\lambda_m^{\delta_1+\delta_2-2\delta}\frac{\partial W_m(\theta)}{\partial\beta}\Big|_{\theta_0}=-\eta_1\lambda_m^{\delta_1+\delta_2-2\delta}\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\gamma_j\,\mathrm{Re}\left(G_{uu}^{-1}P_{j1.}I_{ju}P_{j2.}^*\right)+o_p(1). \tag{18}$$

with $\gamma_j = \lambda_j^{2\delta-\delta_1-\delta_2}$. In (18) we decompose $I_{ju} = (2\pi n)^{-1}|\sum_t^n \varepsilon_t|^2$ such that

$$\eta_1 \sqrt{m}\lambda_m^{\delta_1+\delta_2-2\delta}\frac{\partial W_m(\theta)}{\partial \beta}\Big|_{\theta_0} = -\eta_1\lambda_m^{\delta_1+\delta_2-2\delta}\frac{1}{\pi\sqrt{m}}\sum_{j=1}^m \gamma_j \operatorname{Re}\left(G_{uu}^{-1}P_{j1.}P_{j2.}^*\right) \tag{19}$$

$$-\eta_1\lambda_m^{\delta_1+\delta_2-2\delta}\frac{1}{\pi\sqrt{m}}\sum_{j=1}^m \gamma_j \operatorname{Re}\left(G_{uu}^{-1}P_{j1.}\left(\frac{1}{n}\sum_{t=1}^n \varepsilon_t\varepsilon_t' - I_2\right)P_{j2.}^*\right) \tag{20}$$

$$-\eta_1\lambda_m^{\delta_1+\delta_2-2\delta}\frac{2}{\sqrt{m}}\sum_{j=1}^m \gamma_j \operatorname{Re}\left(G_{uu}^{-1}P_{j1.}\frac{1}{2\pi n}\sum_{t=1}^n\sum_{s\neq t}\varepsilon_t\varepsilon_s' e^{i(t-s)\lambda_j}P_{j2.}^*\right), \tag{21}$$

As $f_z(\lambda) = (2\pi)^{-1}A(\lambda)A(\lambda)^*$, by Assumptions 2 and 5, (19) is

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

(20) is

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

by the law of large numbers and the remaining term (21) can be rearranged as

$$-\sum_{t=1}^n \varepsilon_t' \sum_{s=1}^{t-1}\frac{\eta_1}{\pi n\sqrt{m}}\lambda_m^{\delta_1+\delta_2-2\delta}\sum_{j=1}^m \gamma_j \operatorname{Re}\left(G_{uu}^{-1}P_{j.1}e^{i(t-s)\lambda_j}\bar{P}_{j2.}\right)\varepsilon_s \tag{22}$$

where $\bar{P}_j$ denotes the conjugate of P_j.

Regarding ξ, by analogy with (18) we obtain

$$\eta_2\sqrt{m}\frac{\lambda_m^{\delta_1+\delta_2-2\delta}}{\log\lambda_m}\frac{\partial W_m(\theta)}{\partial \xi}\Big|_{\theta_0} = -\eta_2\frac{\lambda_m^{\delta_1+\delta_2-2\delta}}{\log\lambda_m}\frac{\beta}{\pi\sqrt{m}}\sum_{j=1}^m \log\lambda_j\gamma_j \operatorname{Re}\left(G_{uu}^{-1}P_{j1.}P_{j2.}^*\right) \tag{23}$$

$$-\eta_2\frac{\lambda_m^{\delta_1+\delta_2-2\delta}}{\log\lambda_m}\frac{\beta}{\pi\sqrt{m}}\sum_{j=1}^m \log\lambda_j\gamma_j \operatorname{Re}\left(G_{uu}^{-1}P_{j1.}\left(\frac{1}{n}\sum_{t=1}^n \varepsilon_t\varepsilon_t' - I_2\right)P_{j2.}^*\right) \tag{24}$$

$$-\eta_2\frac{\lambda_m^{\delta_1+\delta_2-2\delta}}{\log\lambda_m}\frac{2\beta}{\sqrt{m}}\sum_{j=1}^m \log\lambda_j\gamma_j \operatorname{Re}\left(G_{uu}^{-1}P_{j1.}\frac{1}{2\pi n}\sum_{t=1}^n\sum_{s\neq t}\varepsilon_t\varepsilon_s' e^{i(t-s)\lambda_j}P_{j2.}^*\right). \tag{25}$$

As $f_z(\lambda) = (2\pi)^{-1}A(\lambda)A(\lambda)^*$, by Assumptions (2) and (5), (23) is

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

(24) is

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

and the remaining term (25) can be rearranged as

$$-\sum_{t=1}^{n}\varepsilon_t'\sum_{s=1}^{t-1}\frac{\eta_2}{\pi n\sqrt{m}}\frac{\lambda_m^{\delta_1+\delta_2-2\delta}}{\log\lambda_m}\beta\sum_{j=1}^{m}\log\lambda_j\gamma_j\,\mathrm{Re}\left(G_{uu}^{-1}P_{j.1}e^{i(t-s)\lambda_j}\bar{P}_{j2.}\right)\varepsilon_s. \tag{26}$$

Using (26), (22) and Euler formula, (16) has the same asymptotic distribution as $\sum_{t=1}^{n}\varepsilon_t'\sum_{s=1}^{t-1}\Xi_{t-s,n}\varepsilon_s$ where

$$\begin{aligned}\Xi_{t-s,n}&=\frac{1}{\pi n\sqrt{m}}\sum_{j=1}^{m}(\theta_{j,1}+\theta_{j,2})\cos((t-s)\lambda_j),\\ \theta_{j,1}&=-\lambda_m^{\delta_1+\delta_2-2\delta}\eta_1\gamma_j\tilde{\theta}_\beta,\\ \theta_{j,2}&=-\frac{\lambda_m^{\delta_1+\delta_2-2\delta}}{\log\lambda_m}\eta_2\log\lambda_j\beta\gamma_j\tilde{\theta}_\beta\end{aligned}$$

with $\tilde{\theta}_\beta=\mathrm{Re}\left(G_{uu}^{-1}P_{j.1}\bar{P}_{j2.}+G_{uu}^{-1}P_{j.2}\bar{P}_{j1.}\right)$. By Assumption 2, $||\theta_{j,1}||=O((m/j)^{\delta_1+\delta_2-2\delta})$ and $||\theta_{j,2}||=O((m/j)^{\delta_1+\delta_2-2\delta}\times(\log j)/(\log m))$. It follows that

$$\eta_1\lambda_m^{\delta_1+\delta_2-2\delta}\sqrt{m}\frac{\partial W_m(\theta)}{\partial\theta_1}\Big|_{\theta_0}+\eta_2\frac{\lambda_m^{\delta_1+\delta_2-2\delta}}{\log\lambda_m}\sqrt{m}\frac{\partial W_m(\theta)}{\partial\theta_2}\Big|_{\theta_0}=\sum_{t=1}^{n}\zeta_t+o_p(1)$$

where $\zeta_t=\varepsilon_t'\sum_{s=1}^{t-1}\Xi_{t-s,n}\varepsilon_s$ is a martingale difference array with respect to $\mathcal{F}_t=\sigma(\{\varepsilon_s,s\le t\})$. By a standard martingale central limit theorem, (16) follows if

$$\sum_{t=1}^{n}\mathbb{E}(\zeta_t^2|\mathcal{F}_{t-1})-\sum_{a=1}^{2}\sum_{b=1}^{2}\eta_a\eta_b\Omega_{ab}\xrightarrow{p}0, \tag{27}$$

$$\sum_{t=1}^{n}\mathbb{E}(\zeta_t^2\mathbb{1}(|\zeta_t|>d))\to 0,\quad\forall d>0, \tag{28}$$

where Ω is shown to be singular and defined as

$$\Omega=\begin{pmatrix}1&\beta_0\\ \beta_0&\beta_0^2\end{pmatrix}F,$$

with $F = 2G_{xx}/(G_{uu}(1+4\delta-2\delta_1-2\delta_2))$. We first show (27). For this, we use the following decomposition

$$\sum_{t=1}^{n}\mathbb{E}(\zeta_t^2|\mathcal{F}_{t-1}) = \sum_{t=1}^{n}\mathbb{E}\Big(\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\varepsilon_s'\Xi_{t-s,n}'\varepsilon_t\varepsilon_t'\Xi_{t-r,n}\varepsilon_r|\mathcal{F}_{t-1}\Big)$$
$$= \sum_{t=1}^{n}\sum_{s=1}^{t-1}\varepsilon_s'\Xi_{t-s,n}'\Xi_{t-s,n}\varepsilon_s + \sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{r=1}^{s-1}\varepsilon_s'\Xi_{t-s,n}'\Xi_{t-r,n}\varepsilon_r, \tag{29}$$

where the second term has mean 0 and variance

$$O\bigg(n\Big(\sum_{s=1}^{n}||\Xi_{s,n}||^2\Big)^2 + \sum_{t=3}^{n}\sum_{u=2}^{t-1}\Big(\sum_{s=1}^{u-1}||\Xi_{u-s,n}||^2\sum_{s=1}^{u-1}||\Xi_{t-s,n}||^2\Big)\bigg) \tag{30}$$

as shown by Lobato (1999). Following Nielsen (2005), when $s < n/m$, $||\Xi_{s,n}|| = O(1/(n\sqrt{m})\sum_{j=1}^{n}||\theta_{j,1}+\theta_{j,2}||) = O(n^{-1}\sqrt{m}\log m)$ 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 (30) is $O((\log m)^4 n^{-1})$. Besides, the second term in (30) 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 (30) is $O(n^{-1}(\log m)^4 + m^{-1}(\log m)^4\log n) \to 0$.

To complete the proof of (27) we now have to show that the mean of the first term in (29) is equal to $\sum_{a=1}^{2}\sum_{b=1}^{2}\eta_a\eta_b\Omega_{ab}$. Since $\mathbb{E}(\varepsilon_t\varepsilon_t'|F_{t-1}) = I_2$ by Assumption (3), 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

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

$$=\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2 n^2 m}\operatorname{tr}\left(\theta'_{j,1}\theta_{j,1}\right)\cos((t-s)\lambda_j)^2 \tag{31}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2 n^2 m}\operatorname{tr}\left(\theta'_{j,2}\theta_{j,2}\right)\cos((t-s)\lambda_j)^2 \tag{32}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2 n^2 m}2\operatorname{tr}\left(\theta'_{j,1}\theta_{j,2}\right)\cos((t-s)\lambda_j)^2 \tag{33}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2 n^2 m}\operatorname{tr}\left(\theta'_{j,1}\theta_{k,1}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \tag{34}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2 n^2 m}\operatorname{tr}\left(\theta'_{j,2}\theta_{k,2}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \tag{35}$$

$$+\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\sum_{k\neq j}^{m}\frac{1}{\pi^2 n^2 m}2\operatorname{tr}\left(\theta'_{j,1}\theta_{k,2}\right)\cos((t-s)\lambda_j)\cos((t-s)\lambda_k). \tag{36}$$

For (31), 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

$$\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2 n^2 m}\operatorname{tr}\left(\theta'_{j,1}\theta_{j,1}\right)\cos((t-s)\lambda_j)^2\sim\eta_1^2\frac{2G_{xx}}{G_{uu}(1+4\delta-2\delta_1-2\delta_2)}.$$

Among the remaining terms we first analyze (32) and observe that

$$\begin{aligned}\frac{\operatorname{tr}\left(\theta'_{j,2}\theta_{j,2}\right)}{4\pi^2}&=\operatorname{tr}\left(\eta_2^2\frac{\lambda_m^{2\delta_2-2\delta}}{4\pi^2(\log\lambda_m)^2}\beta^2\tilde{\theta}'_\beta\tilde{\theta}_\beta\gamma_j^2\right)\\&=\eta_2^2\frac{\lambda_m^{2\delta_2-2\delta}}{(\log\lambda_m)^2}2\beta^2G_{xx}G_{uu}^{-1}\gamma_j^2,\end{aligned}$$

Then, using the same approximation as for (31),(32) is asymptotically equal to

$$\sum_{t=1}^{n}\sum_{s=1}^{t-1}\sum_{j=1}^{m}\frac{1}{\pi^2 n^2 m}\operatorname{tr}\left(\theta'_{j,2}\theta_{j,2}\right)\cos((t-s)\lambda_j)^2\sim\eta_2^2\frac{2\beta^2G_{xx}}{G_{uu}(1+4\delta-2\delta_1-2\delta_2)}.$$

As the joint limiting distribution of β and ξ is singular, when analyzing (33) we find that it is asymptotically equivalent to

$$\eta_1\eta_2 2\beta\frac{G_{xx}}{G_{uu}}\frac{1}{(1+4\delta-2\delta_1-2\delta_2)}.$$

Equations (34) - (36), where $j \neq k$, remain to analyze. For (34) we use that $||\theta_{j,1}|| = O((m/j)^{\delta_1+\delta_2-2\delta})$ to show its equivalence to

$$O\left(\sum_{t=1}^{n} \sum_{s=1}^{t-1} \sum_{j=1}^{m} \sum_{k\neq j}^{m} \frac{1}{n^2 m} \left(\frac{m}{j}\right)^{\delta_1+\delta_2-2\delta} \left(\frac{m}{k}\right)^{\delta_1+\delta_2-2\delta} \cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \right) = O\big(n^{-1} m (\log m)^2\big)$$

where $\sum_{t=1}^{n} \sum_{s=1}^{t-1} \cos((t-s)\lambda_j)\cos((t-s)\lambda_k) = -n/2$ for $\lambda_j \neq \lambda_k$. Similarly, using that $||\theta_{j,2}|| = O((m/j)^{\delta_1+\delta_2-2\delta}(\log j)/(\log m))$ we show that (35) is equivalent to

$$O\left(\sum_{t=1}^{n} \sum_{s=1}^{t-1} \sum_{j=1}^{m} \sum_{k\neq j}^{m} \frac{1}{n^2 m} \left(\frac{m}{j}\right)^{\delta_1+\delta_2-2\delta} \left(\frac{m}{k}\right)^{\delta_1+\delta_2-2\delta} \frac{(\log j)^2}{(\log m)^2} \cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \right) = O\big(n^{-1} m (\log m)^2\big).$$

For the last term, we use both $||\theta_{j,1}||$ and $||\theta_{j,1}||$ and find that (36) is bounded by

$$O\left(\sum_{t=1}^{n} \sum_{s=1}^{t-1} \sum_{j=1}^{m} \sum_{k\neq j}^{m} \frac{1}{n^2 m} \left(\frac{m}{j}\right)^{\delta_1+\delta_2-2\delta} \left(\frac{m}{k}\right)^{\delta_1+\delta_2-2\delta} \frac{\log j}{\log m} \cos((t-s)\lambda_j)\cos((t-s)\lambda_k) \right) = O\big(n^{-1} m (\log m)^2\big).$$

It remains to show (28) or equivalently the sufficient condition $\sum_{t=1}^{n} \mathbb{E}(\zeta^4) \to 0$. As our analysis of (27) is similar to Lemma 4 of Nielsen (2005), this condition can be proved under Assumption (3). 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 (16). □

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

$$\lambda_m^{2\delta_2-2\delta} \begin{pmatrix} 1 \\ \log(\lambda_m)^{-2} \end{pmatrix}' \begin{pmatrix} \partial^2 W_m(\bar{\theta})/(\partial\beta\partial\beta) \\ \partial^2 W_m(\bar{\theta})/(\partial\xi\partial\xi) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 1 \\ \beta^2 \end{pmatrix} E. \tag{37}$$

Proof. As Nielsen (2007), we strengthen the approximation $|\hat{G}_{uu}(\theta_0) - G_{uu}| = o_p(1)$ by showing that $|\hat{G}_{uu}(\bar{\theta}) - \hat{G}_{uu}(\theta_0)| = o_p(1)$. Using the consistency Theorem 1 and the fact that $||\bar{\theta} - \theta_0|| \leq ||\hat{\theta} - \theta_0||$, we have $\hat{G}_{uu}(\bar{\theta}) - \hat{G}_{uu}(\theta_0) = \lambda_m^{2\nu_0} m^{-1} \sum_{j=1}^{m} \lambda_j^{2\delta+2\xi} \tilde{\beta} I_{jxx} - 2\lambda_m^{\nu_0} m^{-1} \sum_{j=1}^{m} \lambda_j^{2\delta+\xi} \tilde{\beta}\, \mathrm{Re}(I_{jxu}^0)$ with $\lambda_m^{\nu_0}\lambda_j^{\xi}\tilde{\beta} = (\bar{\beta}\lambda^{\bar{\xi}} - \beta\lambda^{\xi})$. Then, by Assumption 2,

$$\hat{G}_{uu}(\bar{\theta}) - \hat{G}_{uu}(\theta_0) = O_p\left(\lambda_m^{2\nu_0} \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2\delta-2\delta_2} - \lambda_m^{\nu_0} \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2\delta-\delta_1-\delta_2} \right) = o_p(1). \tag{38}$$

Using (38) and by similar arguments to Lobato (1999) and Nielsen (2007) one can easily show that

$$\lambda_m^{2\delta_2-2\delta}\Big(\frac{\partial^2 W_m(\bar{\theta})}{\partial\beta\partial\beta}-\frac{\partial^2 W_m(\theta_0)}{\partial\beta\partial\beta}\Big)\xrightarrow{p}0 \tag{39}$$

and implicitly

$$\frac{\lambda_m^{2\delta_2-2\delta}}{(\log\lambda_m)^2}\Big(\frac{\partial^2 W_m(\bar{\theta})}{\partial\xi\partial\xi}-\frac{\partial^2 W_m(\theta_0)}{\partial\xi\partial\xi}\Big)\xrightarrow{p}0. \tag{40}$$

To analyze (37), we first study

$$\lambda_m^{2\delta_2-2\delta}\frac{\partial^2 W_m(\theta_0)}{\partial\beta\partial\beta}=\lambda_m^{2\delta_2-2\delta}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^{m}\lambda_j^{2\delta+2\xi}I_{jxx}+o_p(1).$$

We decompose it as

$$\lambda_m^{2\delta_2-2\delta}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^{m}\lambda_j^{2\delta+2\xi}(I_{jxx}-A_{2.}(\lambda_j)I_{ju}A_{2.}(\lambda_j)^*) \tag{41}$$

$$+\lambda_m^{2\delta_2-2\delta}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^{m}\lambda_j^{2\delta+2\xi}\Big(A_{2.}(\lambda_j)\Big(\frac{1}{2\pi n}\sum_{t=1}^{n}\varepsilon_t\varepsilon_t'-I_2\Big)A_{2.}(\lambda_j)^*\Big) \tag{42}$$

$$+\lambda_m^{2\delta_2-2\delta}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^{m}\lambda_j^{2\delta+2\xi}\,\mathrm{Re}\,\Big(A_{2.}(\lambda_j)\frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s\neq t}\varepsilon_t\varepsilon_s' e^{i(t-s)\lambda_j}A_{2.}(\lambda_j)^*\Big) \tag{43}$$

$$+\lambda_m^{2\delta_2-2\delta}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^{m}\lambda_j^{2\delta+2\xi}\Big(A_{2.}(\lambda_j)A_{2.}(\lambda_j)^*\Big)+o_p(1). \tag{44}$$

By similar arguments to that of Proposition 1 in Nielsen (2005), Equations (41)-(43) are $o_p(1)$ and by Assumptions 3 and 4, the last term (44) yields to

$$\lambda_m^{2\delta_2-2\delta}\frac{\partial^2 W_m(\theta_0)}{\partial\beta\partial\beta}=\lambda_m^{2\delta_2-2\delta}\frac{2}{m}\frac{G_{xx}}{G_{uu}}\sum_{j=1}^{m}\lambda_j^{2(\delta-\delta_2)}+o_p(1). \tag{45}$$

After approximation of the Riemann sums by integrals, we obtain

$$\lambda_m^{2\delta_2-2\delta}\frac{\partial^2 W_m(\theta_0)}{\partial\beta\partial\beta}\xrightarrow{p}E=\frac{2G_{xx}}{G_{uu}(1-2\delta_2+2\delta)}.$$

implying the first part of (37) in view of (39). It remains to analyze the second part of (37) and hence

$$\lambda_m^{2\delta_2-2\delta}(\log\lambda_m)^{-2}\frac{\partial^2 W_m(\theta_0)}{\partial\xi\partial\xi}=\lambda_m^{2\delta_2-2\delta}\log(\lambda_m)^{-2}\beta^2 G_{uu}^{-1}G_{xx}\frac{2}{m}\sum_{j=1}^{m}\log\lambda_j^2\lambda_j^{2\delta-2\delta_2}+o_p(1). \tag{46}$$

By analogy with (45), (46) is asymptotically equivalent to

$$\lambda_m^{2\delta_2-2\delta}\log(\lambda_m)^{-2}\frac{\partial^2 W_m(\theta_0)}{\partial\xi\partial\xi}\xrightarrow{p}\beta^2 E=\frac{2\beta^2 G_{xx}}{G_{uu}(1-2\delta_2+2\delta)},$$

which completes the proof in view of (40). Besides, as the joint limiting distribution of β and ξ is singular, we omit the superfluous covariance term. □

Appendix C: Proof of Theorem 3

Proof. Under Assumptions 1-6, we need to show that (16) and (17) remain true when the unknown weighting parameter $\delta=\delta_1$ is replaced by $\hat{\delta}_1$. We first discuss the case of β. Regarding the limit of the Hessian, we need to prove that $\lambda_m^{2\delta_2-2\delta_1}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^m(\lambda_j^{2\hat{\delta}_1}-\lambda_j^{2\delta_1})\lambda_j^{2\xi}I_{jxx}\xrightarrow{p}0$. As (38) holds with $\hat{\delta}_1$, we replace $G_{uu}^{-1}(\bar{\theta})$ by G_{uu}^{-1}. Then, since $|\max_{1\le j\le m}\lambda_j^{2\hat{\delta}_1}-\lambda_j^{2\delta_1}-1|=O_p(|\hat{\delta}_1-\delta_1|\log n)$, as in Nielsen (2005), we have

$$\begin{aligned}\lambda_m^{2\delta_2-2\delta_1}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^m(\lambda_j^{2\hat{\delta}_1}-\lambda_j^{2\delta_1})\lambda_j^{2\xi}I_{jxx}&=O_p\left(\lambda_m^{2\delta_2-2\delta_1}\Big|\max_{1\le j\le m}\lambda_j^{2\hat{\delta}_1-2\delta_1}-1\Big|\frac{1}{m}\sum_{j=1}^m\lambda_j^{2\delta_1+2\xi}I_{jxx}\right)\\&=O_p\left(\lambda_m^{2\delta_2-2\delta_1}|\hat{\delta}_1-\delta_1|(\log n)\frac{1}{m}\sum_{j=1}^m\lambda_j^{2\delta_1-2\delta_2}\right)\\&=O_p(|\hat{\delta}_1-\delta_1|\log n).\end{aligned}\tag{47}$$

Regarding the limit of the score, we need to show that

$$\sqrt{m}\lambda_m^{\delta_2-\delta_1}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^m(\lambda_j^{2\hat{\delta}_1}-\lambda_j^{2\delta_1})\lambda_j^{\xi}\operatorname{Re}(I_{jux})\xrightarrow{p}0.$$

By similar arguments to those of (47) and Assumption 2, we obtain

$$\begin{aligned}\sqrt{m}\lambda_m^{\delta_2-\delta_1}G_{uu}^{-1}\frac{2}{m}\sum_{j=1}^m(\lambda_j^{2\hat{\delta}_1}-\lambda_j^{2\delta_1})\lambda_j^{\xi}\operatorname{Re}(I_{jux})&=O_p\left(\sqrt{m}\lambda_m^{\delta_2-\delta_1}\Big|\max_{1\le j\le m}\lambda_j^{2\hat{\delta}_1-2\delta_1}-1\Big|\frac{1}{m}\sum_{j=1}^m\lambda_j^{2\delta_1+\xi}I_{jux}\right)\\&=O_p\left(\sqrt{m}\lambda_m^{\delta_2-\delta_1}|\hat{\delta}_1-\delta_1|(\log n)\frac{1}{m}\sum_{j=1}^m\lambda_j^{\alpha+\delta_1-\delta_2}\right)\\&=O_p(\sqrt{m}\lambda_m^{\alpha}|\hat{\delta}_1-\delta_1|\log n).\end{aligned}$$

By similar reasoning, the same bounds can be obtained for the parameter ξ, which completes the proof of Theorem 3. □

References

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

Bandi, F.M., Perron, B., 2006. Long memory and the relation between implied and realized volatility. Journal of Financial Econometrics 4, 636-670.

Bali, T. G., Peng, L., 2006. Is there a risk-return trade-off? Evidence from high-frequency data. Journal of Applied Econometrics, 21, 1169-1198.

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.

Bollerslev, T., Marrone, J., Xu, L., Zhou, H., 2014. Stock return predictability and variance risk premia: statistical inference and international evidence. Journal of Financial and Quantitative Analysis, 49, 633-661.

Bollerslev, T., Xu, L., Zhou, H., 2015. Stock return and cash flow predictability: The role of volatility risk. Journal of Econometrics, 187, 458-471.

Bollerslev, T., Osterrieder, D., Sizova, N., Tauchen, G., 2013. Risk and return: Long-run relations, fractional cointegration, and return predictability. Journal of Financial Economics, 108, 409-424.

Bollerslev, T., Sizova, N., Tauchen, G., 2012. Volatility in equilibrium: Asymmetries and dynamic dependencies. Review of Finance, 16, 31-80.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. The Review of Financial Studies, 22, 4463-4492.

Campbell, J.Y., Hentschel, L., 1992. No news is goodnews: an asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281-318.

Campbell, J. Y., Shiller, R. J. Stock Prices, Earnings, and Expected Dividends. Journal of Finance, 43, 661-676.

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.

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

Drechsler, I., Yaron, A., 2011. What's vol got to do with it. The Review of Financial Studies, 24, 1-45.

Fama, E. F., French, K. R., 1988. Dividend Yields and Expected Stock Returns. Journal of Financial Economics, 22, 3-25.

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

Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. Journal of Financial Economics, 76, 509-548.

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.

Hidalgo, J., Robinson, P.M., 2002. Adapting to unknown disturbance autocorrelation in regression with long memory. Econometrica 70, 1545-1581.

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.

Lettau, M., Ludvigson., S., 2001. Consumption, AggregateWealth, and Expected Stock Returns. Journal of Finance, 56, 815-849.

Lobato, I.N., 1997. Consistency of the averaged cross-periodogram in long memory series. Journal of Time Series Analysis 18, 137-155.

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.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347-370.

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., 1994. Time series with strong dependence. In Advances in econometrics, sixth world congress (C. A. Sims, ed.) 1, 47-95. Cambridge Univ. Press.

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., Hidalgo, F., 1997. Time series regression with long-range dependence. The annals of statistics 25, 77-104.

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.

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.

Velasco, C., 2003. Gaussian semi-parametric estimation of fractional cointegration. Journal of Time Series Analysis, 24, 345-378.