

# Estimation of long-run parameters in unbalanced cointegration

Javier Hualde\*

Universidad Pública de Navarra

October 28, 2013

## Abstract

This paper analyzes the asymptotic properties of nonlinear least squares estimators of the long run parameters in a bivariate unbalanced cointegration framework. Unbalanced cointegration refers to the situation where the integration orders of the observables are different, but their corresponding balanced versions (with equal integration orders after filtering) are cointegrated in the usual sense. Within this setting, the long run linkage between the observables is driven by both the cointegrating parameter and the difference between the integration orders of the observables, which we consider to be unknown. Our results reveal three noticeable features. First, superconsistent (faster than  $\sqrt{n}$ -consistent) estimators of the difference between memory parameters are achievable. Next, the joint limiting distribution of the estimators of both parameters is singular, and, finally, a modified version of the “Type II” fractional Brownian motion arises in the limiting theory. A Monte Carlo experiment and the discussion of an economic example are included.

*JEL Classification:* C32.

*Keywords.* Unbalanced cointegration, long run parameters, nonlinear least squares, Type II fractional Brownian motion.

---

\*Corresponding author: Javier Hualde. Tel.: +34948169674; fax: +34948169721. *E-mail address:* javier.hualde@unavarra.es. I thank James Davidson, Henryk Gzyl, Søren Johansen, Paolo Paruolo and seminar participants at the Long Memory Symposium at CREATES, University of Aarhus, for helpful comments. This research is supported by the Spanish Ministerio de Economía y Competitividad ref. ECO2011-24304.

## 1. Introduction

Since the seminal paper of Engle and Granger (1987), cointegration, which has traditionally focused on the case of unit root observables with weak dependent cointegrating errors, has been a fertile field of research. This original idea has been generalized in various directions. Among these, one of the main developments is that of fractional cointegration, which given the concept of fractional integration (introduced by Granger and Joyeux, 1980), extends and encompasses naturally the standard notion of cointegration. In the simple bivariate case two processes sharing the same integration order (say  $\delta$ ) are cointegrated if there is a linear combination of them with integration order smaller than  $\delta$ . In a multivariate situation several definitions are available (see, e.g., Robinson and Yajima, 2002), although all of them share the idea of reducing-order linear combination. Inference procedures for fractional cointegration have been developed by, e.g., Jeganathan (1999), Robinson and Marinucci (2001), Robinson and Yajima (2002), Robinson and Hualde (2003), Marmol and Velasco (2004), Christensen and Nielsen (2006), Hualde and Robinson (2007, 2010), Nielsen and Frederiksen (2011), Johansen and Nielsen (2012). However, most of the previous literature has not captured the situation termed by Hualde (2006) as unbalanced cointegration (UC hereinafter), with the important exceptions of Johansen (2008) and Franchi (2010), which give conditions under which such situation might arise in a fractional vector autoregressive model, but do not develop inferential procedures. In the simple bivariate case, UC denotes a situation where the integration orders of the observables are different, but their corresponding balanced versions (where one of the series is filtered adequately so it has identical integration order to the other one) are cointegrated in the usual sense. This can be seen as a particular case of the so-called polynomial cointegration, which in the integer orders case has been studied by, e.g., Johansen (1995).

Denoting by  $\theta$  the imbalance between the integration orders of the two observables, Hualde (2006) discusses two situations, one where  $\theta = \theta_n \rightarrow 0$  as  $n \rightarrow \infty$ , ( $n$  denoting sample size), named weak UC, and the other where  $\theta$  is an unknown fixed real number different from zero, named strong UC. While the former situation is treated with a good deal of theoretical rigor, the latter (denoted simply as UC hereinafter) is just briefly discussed. UC poses interesting challenges, mainly because while in a “balanced” bivariate situation (where  $\theta = 0$ ), if there exists

cointegration, the cointegrating parameter drives the long run linkage between the observables, if there is UC (so  $\theta \neq 0$ ), it is both  $\theta$  and the cointegrating parameter which are relevant in order to explain the long run comovements of the observables. Thus, from a theoretical viewpoint, allowing for the possibility of an unknown (and possibly nonzero)  $\theta$  is relevant, especially noting that misspecification of  $\theta$  could have very distorting effects (see Hualde, 2006). In addition, empirical researchers usually admit the possibility that  $\theta = 0$  as the outcome of testing procedures (e.g. Dickey and Fuller, 1979, or Robinson and Yajima's, 2002, test for equality of orders), so, even if  $\theta = 0$ , a safer option is to take the agnostic approach of considering  $\theta$  to be an unknown parameter, and not imposing knowledge of it in the estimation procedure.

While the main focus of the paper is to present formal theoretical discussion of the limiting properties of particular estimators in UC, we would also like to motivate UC from an empirical perspective. Interestingly, we find that UC relates directly to the idea of multicointegration proposed by Granger and Lee (1989, 1990) (which can be also seen as a particular case of polynomial cointegration). The idea of multicointegration appears to be the most empirically relevant situation involving cointegration between processes with different but known integer orders of integration and, as stated by Engsted and Haldrup (1999), this phenomenon is likely to occur in stock-flow models. Here, two flow variables (usually characterized as unit roots) cointegrate in the standard way, and the cumulated cointegrating error (stock variable) cointegrates with at least one of the flow observables. In a seminal contribution, Granger and Lee (1989) applied this idea to the relationship between production and sales (flow variables) in a given industry, exploring also the possibility of cointegration between the stock of inventories (accumulated change of inventory) and sales, which would support the idea of targeting (the target level of inventory being just a fixed proportion of sales). An alternative analysis of the relationship between inventories and sales was performed by Banerjee and Mizen (2006). Other works explore the existence of multicointegration between housing starts, completions (flow variables) and housing units under construction (stock) (see Lee, 1992), government spending, revenues and debt (Leachman, 1996, Leachman and Francis, 2002, Leachman *et al.*, 2005), imports, exports and external debt (Leachman and Francis, 2000, 2002), or real per capita private consumption expenditure, real per capita dis-

posable income and stock of consumer's wealth (Siliverstovs, 2006).

The role of UC within the framework of multicointegration can be explained as follows. One of the key assumptions behind the idea of multicointegration is that the stock variable (accumulated cointegrating error) must have the same integration order as that of the flows. This necessarily implies that the cointegrating gap (that is the reduction in order due to the cointegrating relation) in the relationship between the flows be equal to one. However, empirical works in fractional cointegration show substantial evidence in favour of smaller cointegrating gaps (see, e.g., Gil-Alana and Hualde, 2009), which in terms of multicointegration means that the stock variable would have a larger integration order than that of the flows. Particularizing this possibility, e.g., to the relationship between inventories ( $h_t$ ) and sales ( $s_t$ ), even admitting the possibility that  $s_t$  is a unit root, if the cointegrating gap arising from the relationship between production and sales is  $d$ , the only interesting cointegrating possibility between inventories and sales would be that between  $h_t$  and  $\Delta^\theta s_t$ , where  $\Delta = 1 - L$ ,  $L$  being the lag operator, and  $\theta = d - 1$ , noting that  $h_t$  and  $\Delta^\theta s_t$  would share the same integration order  $2 - d$  (a proper definition of the fractional operator  $\Delta^\theta$  will be given below). If  $d \neq 1$ , this would exemplify the situation of UC. Given that  $\Delta^\theta s_t$  is a linear combination of present and past values of  $s_t$ , UC would lead to the idea of dynamic targeting, where the target level of inventories is a proportion of present and past sales.

Multicointegration is not the only setting where the idea of UC might be useful. Another motivating example is that of predictive regressions, where rates of return are regressed against the lagged values of an explanatory variable (see, e.g., Torous, Valkanov and Yan, 2004). Here, it is standard to consider the rates of return as weak dependent, whereas the assumption of weak dependence for the regressor is usually unsatisfactory. We can exemplify this situation by the forward premium anomaly, which consists on surprising negative estimates from the regression of the change in the logarithms of the spot exchange rate (considered to be weak dependent) on the forward premium (stationary long memory or nonstationary but mean reverting), where the theory predicts a value of one for that slope (see e.g. Bekaert, 1996, Bekaert, Hodrick and Marshall, 1997). Ballie and Bollerslev (2000) refer to the forward premium anomaly as an statistical problem caused by the different integration orders of dependent and explanatory

variables, and Maynard and Phillips (2001) gave theoretical justification to this phenomenon. Recently, Maynard, Smallwood and Wohar (2013) provided an interesting empirical analysis which, in particular, takes into account the possible imbalance between the memories of the dependent variable and regressor. As will be seen below (Remark 8), the results we obtain in the present paper are not directly applicable to their problem, but an alternative approach focused on modelling the relation between the spot exchange rate (possibly unit root) and the integrated forward premium (which could have memory larger than one) might fall within the UC setting. In any case, even if the situation considered is not characterized by UC, the techniques developed in the present paper can be very useful when dealing with cases, like that of Maynard, Smallwood and Wohar (2013), where there is imbalance between the integration orders of dependent variable and regressor.

The rest of the paper is organized as follows. In Section 2 we present a model of UC and estimators of the relevant parameters, justifying also their limiting properties. A Monte Carlo experiment of finite sample performance is presented in Section 3. An empirical example is discussed in Section 4 and, finally, we conclude in Section 5.

## 2. Model and estimation of long-run parameters

Before introducing our proposed model we present some definitions. We say that a vector process  $\zeta_t$  is integrated of order zero ( $I(0)$ ) if  $\zeta_t - E(\zeta_t)$  is covariance stationary with spectral density finite and nonsingular at all frequencies. Then, denoting by  $r_{it}$  the  $i$ th element of an arbitrary vector  $r_t$ , we say, as in Robinson and Gerolimetto (2006), that a scalar process  $\xi_t$  is integrated of order  $d$  ( $I(d)$ ) if for any  $l \times 1$  zero mean  $I(0)$  vector  $\zeta_t$ ,  $\xi_t - E(\xi_t) = \sum_{k=1}^l \zeta_{kt}(-d_k)$ , with  $d = \max_{1 \leq k \leq l} d_k$ , where for a scalar or vector process  $\xi_t$  and real number  $\alpha$ ,

$$\xi_t(\alpha) = \Delta^\alpha \{\xi_t 1(t > 0)\} = \sum_{j=0}^{t-1} a_j(-\alpha) \xi_{t-j}, \quad a_j(\alpha) = \frac{\Gamma(j + \alpha)}{\Gamma(\alpha) \Gamma(j + 1)}, \quad \alpha \neq 0, -1, \dots, \quad (1)$$

where  $1(\cdot)$  denotes the indicator function (so  $\xi_t 1(t > 0) = \xi_t$  if  $t > 0$ ;  $= 0$  if  $t \leq 0$ ),  $\Gamma(\cdot)$  represents the gamma function, taking  $\Gamma(\alpha) = \infty$  for  $\alpha = 0, -1, -2, \dots$ , and  $\Gamma(0)/\Gamma(0) = 1$ . Note that introducing the indicator function in (1) leads to a truncation: in particular,  $\zeta_{kt}(-d_k) = \sum_{j=0}^{t-1} a_j(d_k) \zeta_{k,t-j}$ , which can be com-

pared to the untruncated sum  $\Delta^{-d_k} \zeta_{kt} = \sum_{j=0}^{\infty} a_j(d_k) \zeta_{k,t-j}$ . The reason why the indicator is introduced here is that it ensures that processes are well defined in mean square sense. In fact,  $\Delta^{-d_k} \zeta_{kt}$  is well defined in mean square sense just if  $d_k < 1/2$ , whereas  $\zeta_{kt}(-d_k)$  is well defined for any value of  $d_k$ . Thus, the truncation allows a uniform treatment of all integration orders, although, related to the previous expression, it is certainly unnecessary when  $d_k < 1/2$ , in which case  $\Delta^{-d_k} \zeta_{kt}$  is stationary. This truncation is very standard in the fractional integration and cointegration literature and originates the so-called Type II fractional processes. Additionally, we say that two scalar processes sharing the same integration order are cointegrated if a linear combination of them has a smaller integration order.

We introduce a bivariate model of UC. Let  $y_t, x_t, t \in Z, Z = \{t : t = 0, \pm 1, \dots\}$ , be two scalar observable series generated by model

$$y_t = \mu + \nu x_t(\theta) + u_{1t}(-\gamma), \quad (2)$$

$$x_t = u_{2t}(-(\delta + \theta)). \quad (3)$$

Assuming  $u_t = (u_{1t}, u_{2t})^T$  is a zero-mean  $I(0)$  process (where superscript  $T$  denotes transposition), it is evident that if in (2), (3),  $\nu \neq 0$  (which we impose for identification, see Remark 7) and  $\delta > \gamma$ ,  $x_t$  is  $I(\delta + \theta)$ ,  $y_t$  is  $I(\delta)$ , and the combination of both processes,  $y_t - \nu x_t(\theta)$  is  $I(\gamma)$ . Thus,  $y_t$  and  $x_t(\theta)$  are (fractionally) cointegrated. Note that even if (2), (3) allows for (and emphasizes) the possibility of non-integer orders, it covers familiar situations. For example, if  $\delta = 1, \gamma = \theta = 0$ , (2), (3) is the bivariate version of Phillips' (1991) triangular form. If instead  $\delta = \theta = 1, \gamma = 0$ , (2), (3) represents the case where  $y_t, x_t$  are  $I(1), I(2)$ , respectively, but  $y_t$  and the first differences of  $x_t$  are cointegrated in the standard way, with an  $I(0)$  cointegrating error. But in general (2), (3) permits a great variety of possibilities, including nonstationary but mean reverting observables (with integration order greater or equal than  $1/2$  but smaller than 1), or asymptotically stationary observables (with integration order smaller than  $1/2$ ), see e.g. Robinson and Hualde (2003), Hualde and Robinson (2007), which motivate different situations when  $\theta = 0$ . Although other situations could be considered, we will concentrate in this paper on the case where  $\delta > 1/2, \gamma \geq 0$ , so at least one of the observables is purely nonstationary.

The behaviour of  $y_t, x_t$  in (2), (3), is driven by various different parameters:  $\mu, \theta, \nu, \delta, \gamma$  and those describing the joint structure of  $u_{1t}, u_{2t}$ . Here, we focus on the estimation of  $\nu, \theta$ , which explain the long-run linkages between observables  $y_t, x_t$ . Given a sample  $y_t, x_t, t = 1, \dots, n$ , define, for any  $e \in \mathbb{R}$ ,  $\hat{\nu}(e) = C_{yx(e)}/C_{x(e)}$ , where for any sequences  $a_t, b_t, t = 1, \dots, n$ ,  $C_{ab} = \sum_{t=1}^n (a_t - \bar{a})(b_t - \bar{b})$ ,  $C_a = C_{aa}$  and  $\bar{a} = n^{-1}\sum_{t=1}^n a_t$ . If  $\theta$  were known (e.g.,  $\theta = 0$ , as in the traditional balanced setting), the most straightforward way of estimating  $\nu$  would be by ordinary least squares (OLS)  $\hat{\nu}(\theta)$ . If  $\theta \neq 0$ ,  $\hat{\nu}(\theta)$  is the balanced version of the standard OLS estimator, but, as in general  $\theta$  is unknown, this estimator is unfeasible. Thus, defining

$$Q_n(e) = \sum_{t=1}^n (y_t - \bar{y} - \hat{\nu}(e)(x_t(e) - \bar{x}(e)))^2, \quad \bar{x}(e) = \frac{1}{n} \sum_{t=1}^n x_t(e), \quad (4)$$

the corresponding feasible estimators of  $\theta, \nu$ , are the nonlinear least squares

$$\hat{\theta} = \arg \min_{e \in \Xi} Q_n(e), \quad \hat{\nu} = \hat{\nu}(\hat{\theta}), \quad (5)$$

where, for arbitrary real numbers  $\nabla_1 < \nabla_2$ ,  $\Xi = [\nabla_1, \nabla_2]$ . There is an interesting, but not obvious (in view of (5)), interpretation of  $\hat{\theta}$ . After some algebra it is possible to show that  $\hat{\theta} = \arg \max_{e \in \Xi} C_{yx(e)}^2 / C_y C_{x(e)}$ , so  $\hat{\theta}$  represents the particular argument  $e$  which maximizes the sample linear dependence between processes  $y_t$  and  $x_t(e)$  (or, equivalently,  $\hat{\theta}$  is the argument which maximizes the  $R$ -squared from the regression of  $y_t$  on a constant and  $x_t(e)$ ). Therefore, modelling the relationship between  $y_t$  and  $x_t$  by choosing a particular a priori differentiation of  $x_t$  (e.g., leaving  $x_t$  undifferentiated) would lead to fit losses.

Before presenting the assumptions under which our results hold, we introduce some additional notation. Throughout we denote by  $I_p$  the  $p \times p$  identity matrix,  $\|\cdot\|$  denotes the Euclidean norm and  $[\cdot]$  integer part. Also, denote by  $W(r)$  the  $2 \times 1$  vector Brownian motion with covariance matrix  $\Sigma$  (see Assumption 1 (iii) below), and for any  $d > 1/2$  define the Type II fractional Brownian motion (see Marinucci and Robinson, 2000)

$$W(r; d) = \frac{1}{\Gamma(d)} \int_0^r (r-s)^{d-1} dW(s), \quad (6)$$

and the modified fractional Brownian motion (see Hualde, 2012)

$$A(r; d) = \frac{1}{\Gamma(d)} \int_0^r \log(r-s) (r-s)^{d-1} dW(s). \quad (7)$$

As will be seen below, (6) and (7) have a crucial role in characterizing the limiting properties of our estimators.

**Assumption 1.** *The process  $u_t = (u_{1t}, u_{2t})^T$ ,  $t \in Z$ , has representation  $u_t = B(L)\varepsilon_t$ , where  $B(s) = I_2 + \sum_{j=1}^{\infty} B_j s^j$ , and the  $B_j$  are  $2 \times 2$  matrices such that*

(i)  $\det\{B(s)\} \neq 0$ ,  $|s| \leq 1$ ;

(ii)  $B(e^{i\lambda})$  is differentiable in  $\lambda$  with derivative in  $Lip(\varrho)$ ,  $\varrho > 1/2$ ;

(iii) the  $\varepsilon_t$  are independent and identically distributed (iid) vectors with mean zero, positive definite covariance matrix  $\Sigma$ , and  $E\|\varepsilon_t\|^q < \infty$ ,  $q \geq 4$ ,  $q > 2/(2\delta - 1)$ .

Assumption 1 is sufficient to derive the different results given in Robinson and Marinucci (2001). In particular, it is sufficient for the conditions related to the cumulant spectral density, and imply square integrability of the univariate spectra of  $u_{1t}$ ,  $u_{2t}$  and fourth-order stationarity of  $u_t$ . Also, by (ii), the derivative of  $B(e^{i\lambda})$  has Fourier coefficients  $jB_j = O(j^{-\varrho})$  as  $j \rightarrow \infty$ . Note also that the iid assumption in (iii) can be undoubtedly relaxed, but this would require extending the results in Hualde (2012) to cover more general processes.

**Assumption 2.** *In (2), (3),  $\delta > 1/2$ ,  $\gamma \geq 0$ ,  $\nu \neq 0$  and  $\theta \in (\nabla_1, \nabla_2)$ .*

First, we establish consistency of  $\hat{\theta}$ .

**Theorem 1.** *Let Assumptions 1, 2 hold. Then, as  $n \rightarrow \infty$ ,  $\hat{\theta} \rightarrow_p \theta$ .*

The proof of Theorem 1 is given in Appendix A and it is based on a series of lemmas introduced in Appendix B. As in related settings, the proof of consistency is nonstandard (see, e.g., Robinson 1995, and Hualde and Robinson, 2011), due to the non-uniform convergence of  $Q_n(e)$  over a large admissible parameter space.

Next we establish the joint limiting distribution of  $\hat{\theta}$ ,  $\hat{\nu}$ .

**Theorem 2.** *Let Assumptions 1, 2 hold. Then, as  $n \rightarrow \infty$ ,*

$$q_n(\gamma, \delta) \begin{pmatrix} \hat{\theta} - \theta \\ \log^{-1} n (\hat{\nu} - \nu) \end{pmatrix} \rightarrow_d \begin{pmatrix} \frac{1}{\nu} \\ 1 \end{pmatrix} \Xi(\gamma, \delta), \quad (8)$$



where

$$q_n(\gamma, \delta) = \frac{n^{2\delta-1}}{\log n} \mathbf{1}(\gamma + \delta < 1) + \frac{n^{2\delta-1}}{\log^2 n} \mathbf{1}(\gamma + \delta = 1, \gamma > 0) + \frac{n}{\log n} \mathbf{1}(\delta = 1, \gamma = 0) + n^{\delta-\gamma} \mathbf{1}(\gamma + \delta > 1, \gamma > 1/2) + n^\delta \mathbf{1}(\gamma + \delta > 1, \gamma = 0), \quad (9)$$

and  $\Xi(\gamma, \delta)$  is defined in (59).

The proof of Theorem 2 is given in Appendix A. As in Robinson and Marinucci (2001), we are able to derive results for  $\gamma + \delta > 1$  only when  $\gamma = 0$  or  $\gamma > 1/2$ .

**Remark 1.** As can be inferred from (8), the joint asymptotic distribution of  $\widehat{\theta}$  and  $\widehat{\nu}$  is singular. The reason for this phenomenon is that by the mean value theorem,

$$\widehat{\nu}(\widehat{\theta}) - \nu = \widehat{\nu}(\theta) - \nu + \widehat{\nu}'(\bar{\theta})(\widehat{\theta} - \theta), \quad (10)$$

where for any real number  $p$  and functional  $f(\cdot)$ ,  $f'(p)$  denotes the first derivative of  $f(\cdot)$  evaluated at  $p$ , and also  $|\bar{\theta} - \theta| \leq |\widehat{\theta} - \theta|$ . The behaviour of  $\widehat{\nu}(\theta) - \nu$  is well known from Robinson and Marinucci (2001), and, as can be inferred from the proof of Theorem 2,

$$\frac{1}{\log n} \widehat{\nu}'(\bar{\theta}) \rightarrow_p \nu. \quad (11)$$

Then given that in all cases  $\widehat{\nu}(\theta) - \nu = o_p(q_n^{-1}(\gamma, \delta) \log n)$ , it can be shown that

$$\begin{pmatrix} q_n(\gamma, \delta) (\widehat{\theta} - \theta) \\ q_n(\gamma, \delta) \log^{-1} n (\widehat{\nu} - \nu) \end{pmatrix} = \begin{pmatrix} 1 \\ \nu \end{pmatrix} q_n(\gamma, \delta) (\widehat{\theta} - \theta) + o_p(1), \quad (12)$$

which leads to (8).

There is an intuitive reason behind the singularity of the joint limiting distribution. Linearizing the nonlinear least squares problem (5), it can be easily shown that, approximately, our nonlinear problem corresponds to a linear one where  $\widehat{\nu}$  and  $\widehat{\theta}$  are the estimated slopes corresponding to the pseudoregressors  $x_t(\theta)$  and  $\nu x'_t(\theta)$ , respectively. Then noting that  $x_t(\theta) = \sum_{j=0}^{t-1} a_j(\delta) u_{2,t-j}$ ,  $x'_t(\theta) = -\sum_{j=1}^{t-1} a'_j(\delta) u_{2,t-j}$ , where proceeding as in the proof of Lemma D.1 of Robinson and Hualde (2003),  $a'_j(\delta) = (\psi(j+\delta) - \psi(\delta)) a_j(\delta)$ ,  $\psi(\cdot)$  being the digamma function, by the properties of the digamma function and results in Marinucci and Robinson (2000), it can be shown that  $t^{1/2-\delta} x_t(\theta)$  and  $-t^{1/2-\delta} \log^{-1} t x'_t(\theta)$  converge in distribution to the same random variable. In this particular sense,

the regressors in the linearized model are asymptotically collinear and this is the crucial issue behind the singularity of the joint limiting distribution. A similar problem occurs in the regression with cointegrated regressors setting, see, e.g., Park and Phillips (1989).

**Remark 2.** Even if  $\theta$  is a memory parameter,  $\hat{\theta}$  could be superconsistent (understood as faster-than- $\sqrt{n}$ -consistent). This is a nonstandard result and, as far as we know, it has not been hinted before in the literature on memory estimation, which suggests that  $\sqrt{n}$  is the maximum rate achievable by memory estimators even by parametric procedures (see e.g. Fox and Taqqu, 1986, Dalhaus, 1989, Hualde and Robinson, 2011). The reason is that in our setting  $\theta$  represents the difference between the memories of two series, and the existence of cointegration between them is the crucial distinctive aspect leading to the superconsistency result (unlike in the case of no cointegration where the maximum rate is  $\sqrt{n}$ ). Note that, as can be inferred from Remark 1,  $\theta$  is similar to a cointegrating parameter (it is the slope of the pseudoregressor  $\nu x_t'(\theta)$ ). In fact,  $\theta$ , like  $\nu$ , characterizes a singularity which arises in the spectral density of  $\Delta^\delta(y_t, x_t)'$ . Then, the higher is the cointegrating gap  $\delta - \gamma$ , the “stronger” is this singularity, and in this sense, higher cointegrating gaps lead in general to faster convergence rates.

**Remark 3.** The limiting distribution of  $\hat{\theta}$  (but not that of  $\hat{\nu}$ ) depends on  $\nu$  (the presence of  $\nu$  in the second equation of (12) cancels the dependence of the limiting distribution of  $\hat{\nu}$  on  $\nu$ ).

**Remark 4.** Our results are valid in the  $\theta = 0$  case (where (2), (3) become a standard fractional cointegration model), noting that it is not very relevant whether  $\theta = 0$  or  $\theta \neq 0$ , but whether the information about the true value of  $\theta$  is incorporated into the estimation. Also, the rate of convergence  $q_n(\gamma, \delta)$  in (12) is, apart from a  $\log^{-1} n$  factor when  $\gamma + \delta \leq 1$ , identical to that of the estimator of  $\nu$  in the situation where  $\theta$  is correctly assumed to be known (see Robinson and Marinucci, 2001). However, even if  $\theta = 0$ , if the information about  $\theta$  is not incorporated into the estimation, the estimator of  $\nu$  suffers from a slight loss in rate of convergence. This is the main price to pay for allowing the greater flexibility that letting  $\theta$  be unknown permits. In addition, apart from the rate, the asymptotic distribution of  $\hat{\nu}$  differs substantially from that where  $\theta$  is considered to be known (see Hualde, 2006).

**Remark 5.** Different values of  $\gamma, \delta$ , lead to different characterizations of ob-

servables and cointegrating error which might arise in practice. Related to this, Theorem 2 reflects the fact that, as usual in fractional cointegration (see, e.g., Robinson and Marinucci, 2001), the limiting behaviour of the estimators depends on the integration orders of observables and cointegrating error. This dependence affects convergence rates and also the form of  $\Xi(\gamma, \delta)$  (see (59)), which also depends on the (possibly infinite-dimensional) set of short run parameters characterizing  $u_t$ . For particular cases of  $u_t$  some simplifications of the corresponding limiting distributions and convergence rates are possible (as it is clearly the case when  $u_{1t}$  and  $u_{2t}$  are mutually independent sequences, because in this case  $\Xi(\gamma, \delta) = 0$  when  $\gamma + \delta \leq 1$ ).

**Remark 6.** As it is evident from the characterization of  $\Xi(\gamma, \delta)$  given in (59), the modified version of the Type II fractional Brownian motion given in (7) appears in our limiting results. This process has been introduced by Hualde (2012) and it is the limit of a normalized modified Type II fractional process.

**Remark 7.** Condition  $\nu \neq 0$  (see Assumption 2) is necessary for identification, because if  $\nu = 0$ ,  $\theta$  is unidentified. Thus, when  $\nu$  is close to zero it is expected that our estimators of  $\theta$ ,  $\nu$  perform poorly in finite samples, and this is confirmed by an unreported Monte Carlo experiment.

**Remark 8.** An alternative way of estimating  $\nu$  is to use a two-stage approach. Given that  $\theta$  represents the difference between the memories of  $x_t$  and  $y_t$ , this can be estimated first by standard memory estimation methods (say we obtain estimator  $\bar{\theta}$ ), and then estimate  $\nu$  by  $\hat{\nu}(\bar{\theta})$ . Actually, this estimator has been already considered by Hualde (2006) (with the only difference that he did not include a constant in the OLS regression) who denoted it by  $\bar{\nu}_5$ . The properties of this estimator were described in Hualde's (2006) Theorem UCD. In view of the results given in pp. 808-809 of Hualde (2006), if  $\gamma + \delta > 1$  with  $\delta - \gamma \geq 1/2$  or  $\gamma = 0$ ,  $\delta = 1$ ,  $\bar{\nu}_5$  cannot have a faster convergence rate than  $\hat{\nu}$ , but, in most cases,  $\bar{\nu}_5$  will be much slower than  $\hat{\nu}$  (the difference in convergence rates in favour of  $\hat{\nu}$  being higher the higher the cointegrating gap  $\delta - \gamma$  and the slower the rate of convergence of  $\bar{\theta}$ ). However, if  $\gamma + \delta > 1$  with  $\delta - \gamma < 1/2$ , or  $\gamma + \delta = 1$ ,  $\gamma > 0$ , or  $\gamma + \delta < 1$ ,  $\bar{\nu}_5$  might have a faster rate of convergence than  $\hat{\nu}$  (which is at most due to a  $\log^2 n$  factor), but again, depending on the orders  $\delta$ ,  $\gamma$  and also on the rate of convergence of  $\bar{\theta}$ ,  $\hat{\nu}$  might be substantially faster than  $\bar{\nu}_5$ . It is interesting to note that the equivalent to  $\bar{\nu}_5$  in a model with constant has been considered

by Maynard, Smallwood and Wohar (2013) in setting very similar to ours, but imposing  $\delta = \gamma = 0$  (note that we consider the case where  $\delta > 1/2$ ,  $\delta > \gamma$ ) and also strong restrictions on  $u_t$ . Based on our results, we conjecture that in the particular setting they consider, our  $\widehat{\nu}$  estimator will outperform their two-stage estimator (although a formal and detailed analysis along the lines offered by our present paper is warranted).

### 3. Monte Carlo evidence

With the aim of providing evidence of the finite sample behaviour of our estimators, we run a small Monte Carlo experiment. There are two parts to our experiment. The first one presents results of Monte Carlo bias and standard deviation (SD) of  $\widehat{\theta}$ ,  $\widehat{\nu}$ . These estimates were computed as in (5) with  $\Xi = [\theta - 2, \theta + 2]$  in different  $(\gamma, \delta)$  situations. The second part focusses on the traditional  $(\gamma, \delta) = (0, 1)$  cointegrating case, comparing  $\widehat{\nu}$  with the OLS computed for three different values of  $\theta$ .

In both parts we generated 5,000 replications of  $u_t$  in (2), (3), as Gaussian white noise processes of dimensions  $n = 64, 128, 256$ , for different values of the signal-to-noise ratio  $\tau = Var(u_{2t})/Var(u_{1t})$  and correlation coefficient  $\rho$ , where in all cases we set  $Var(u_{1t}) = 1$ . Then, for different choices of  $(\gamma, \delta, \theta)$  we generated  $y_t$  and  $x_t$  from (2), (3) setting  $\mu = 0$ .

In the first part of the experiment we fixed  $\tau = 1$ , and present results for cases  $\rho = .5, 0, -.5$ ,  $(\gamma, \delta) = (0, .6), (0, 1.2), (0, 2), (.4, .6), (.4, 1.2), (.8, 2)$ ,  $\theta = 0$ ,  $\nu = 1, 10$ . As shown by Theorem 2, our estimators of  $\theta$ ,  $\nu$  are asymptotically invariant to  $\theta$ , and this was also reflected in finite samples by our experiment, so we fixed  $\theta = 0$  without loss of generality. The estimator  $\widehat{\nu}$  is also asymptotically invariant to  $\nu$ , so, without loss of generality, we just give results of  $\widehat{\nu}$  for  $\nu = 1$ . On the contrary,  $\widehat{\theta}$  is affected by  $\nu$ , hence we reported results for both  $\nu = 1$  and  $\nu = 10$  cases.

Results for  $\widehat{\theta}$  are presented in Tables 1-4. In terms of Monte Carlo bias (Tables 1, 3), results are better the larger the cointegrating gap  $(\delta - \gamma)$ , except for the  $(\gamma, \delta) = (0, .6)$  case (which is the only situation when  $\gamma + \delta < 1$ ), which induces the very slow rate of convergence  $n^{-2}/\log n$ . The endogeneity bias is controlled by  $\rho$ , which has a large influence in our results. Positive (negative) correlation generates positive (negative) bias in an approximately symmetric way, except for the  $(\gamma, \delta) = (.4, .6)$ ,  $\nu = 1$ , case, where bias is much larger in absolute value

for negative  $\rho$ . Results are best for  $\rho = 0$ , where the endogeneity bias is not present, noting that in this case  $\hat{\theta}$  could enjoy faster convergence rates than those presented in Theorem 2. As expected in view of Theorem 2, bias is much smaller for  $\nu = 10$  (approximately 10 times smaller). In all cases bias reacts in the appropriate direction when  $n$  increases.

Results for the SD of  $\hat{\theta}$  are presented in Tables 2, 4. In most cases smaller SD is associated to larger cointegrating gaps, and SD is relatively unaffected by  $\rho$  (except in the  $(\gamma, \delta) = (.4, .6)$ ,  $\nu = 1$ , case, where the estimator performs clearly best and worst for  $\rho = .5$  and  $\rho = -.5$ , respectively). As for the bias, SD is approximately 10 times smaller when  $\nu = 10$  instead of  $\nu = 1$ , and it decreases as  $n$  increases.

Results for  $\hat{\nu}$  are reported in Tables 5, 6. Bias and SD are substantially larger than those corresponding to  $\hat{\theta}$ , but, as it could have inferred from our asymptotic theory, the behaviour of both estimates is qualitatively very similar.

In the second part of the experiment we concentrate on the traditional cointegrating case  $(\gamma, \delta) = (0, 1)$ . Here, we present results for  $\tau = .5, 1, 2$ ,  $\rho = .25, 0, -.75$ , and compare  $\hat{\nu}$  with the OLS  $\tilde{\nu}_\theta = \sum_{t=1}^n x_t y_t / \sum_{t=1}^n x_t^2$ , in three different situations, corresponding to the cases where in (2), (3),  $\theta = 0, .1, -.1$ . As mentioned before, our estimators of  $\theta, \nu$  are asymptotically invariant to  $\theta$ , so results corresponding to  $\hat{\nu}$  are just given for the  $\theta = 0$  case (almost identical results were obtained if  $\theta = \pm 0.1$ ). Note that  $\tilde{\nu}_0$  is expected to perform substantially better than  $\hat{\nu}$  because it incorporates the correct information about the lack of imbalance between the observables. However,  $\tilde{\nu}_{.1}, \tilde{\nu}_{-.1}$ , are misspecified estimators, and, as our experiment shows, behave substantially worse than  $\hat{\nu}$  in finite samples. This has important implications for empirical work, because a difference (in absolute value) between the integration orders of the observables as small as 0.1, could be very difficult to detect in practice, especially if the sample size is relatively small. Results for bias are reported in Table 7. Bias of  $\hat{\nu}, \tilde{\nu}_0$ , is heavily affected by  $\tau$  and  $\rho$ : the larger and smaller are  $\tau$  and  $|\rho|$ , the better, with positive (negative)  $\rho$  associated to positive (negative) bias. On the contrary, the misspecified estimators  $\tilde{\nu}_{.1}, \tilde{\nu}_{-.1}$ , are hardly affected by these parameters. As anticipated,  $\tilde{\nu}_0$  performs better than  $\hat{\nu}$  (although they behave in a very similar way when  $\rho = 0$ ), but this latter estimator outperforms clearly the misspecified ones, especially as  $n$  increases, where the inconsistency of  $\tilde{\nu}_{.1}, \tilde{\nu}_{-.1}$  is

more noticeable. A positive (negative)  $\theta$  induces negative (positive) bias in the misspecified estimators, negative  $\theta$  having generally more perverse effects.

Results for SD are given in Table 8. Here,  $\tilde{\nu}_0$  performs best and  $\tilde{\nu}_{-1}$  is worst. Noticeably, as  $n$  increases,  $\hat{\nu}$  improves relative to the misspecified estimators, so when  $n = 128, 256$  and  $\tau = 1, 2$ ,  $\hat{\nu}$  performs clearly better than  $\tilde{\nu}_{-1}$ . Again, the misspecification caused by negative  $\theta$  appears to have stronger negative effects.

#### 4. Empirical example

Next we present an empirical application to US quarterly inventory ( $h_t$ ) and sales ( $s_t$ ) data for the period 1982Q1-2001Q4 ( $n = 80$ ). Both series were recorded in billions of chained (2000) dollars, seasonally adjusted, end-of-period quarterly totals drawn from the BEA database. Banerjee and Mizen (2006), who previously analyzed this data set, estimated a polynomially cointegrated model arising from the linear quadratic inventory model proposed by Holt *et al.* (1960), which was also analyzed by Ramey and West (1999) and Hamilton (2002). Before proceeding to analyze the data, we justify why UC can be derived from the linear-quadratic inventory model. Suppose we deal with model (1), (2), (3) of Hamilton (2002) (or (1) of Banerjee and Mizen, 2006), with the only modification that the last term of the cost function (2) is replaced by  $b_2 (h_{t-1} - b_4 - b_3 s_t(\theta))^2$  (to avoid confusion with our fractional parameters note the change in notation, so the  $a$ 's in Hamilton's, 2002, notation become  $b$ 's in ours). This term reflects the costs associated to deviations of inventories from optimal levels (assumed to be  $b_3 s_t(\theta)$ ), see Ramey and West (1999, p.909). Our flexible characterization of the optimal inventory (making it depend, in general, on present and past sales) covers the usual parameterization ( $\theta = 0$ ), but letting  $\theta$  be free controls for possible differences between the integration orders of  $h_t$  and  $s_t$ , leading to a closer approximation between realized and optimal inventories (therefore avoiding important losses). As anticipated in the Introduction, in the present situation it is likely that  $\theta < 0$ , which has important implications concerning the dependence of optimal inventories on past sales. In particular, note that  $s_t(\theta) = \sum_{j=0}^{t-1} a_j(-\theta) s_{t-j}$ , where  $a_{j+1}(-\theta) = (j - \theta) a_j(-\theta) / (j + 1)$ , so it is straightforward to show that for  $\theta < 1$  all coefficients  $a_j(-\theta)$ ,  $j \geq 1$  have opposite sign to that of  $\theta$ . Thus if  $\theta < 0$ , the optimal inventory  $b_3 s_t(\theta)$  is modelled as a linear combination of present and past sales with positive coefficients. This is possibly more appropriate than just restricting to present sales, which nevertheless is also a possibility

covered by our general framework. We further assume that  $s_t(\delta + \theta) = b_5 + v_{st}$ , where  $v_{st}$  is a white noise, so  $s_t \sim I(\delta + \theta)$ , covering both Hamilton's (2002) and Banerjee and Mizen's (2006) assumptions, where  $\delta + \theta = 1$  and  $2$ , respectively. Robinson (1978) and Granger (1980) demonstrated that fractional integration could originate from aggregation of data exhibiting heterogeneous dynamic behaviour at the individual level, so given that Ramey and West (1999) stressed that inventory and sales data are obtained by aggregating across heterogeneous firms, our fractional setting appears to be sensible.

Proceeding as Hamilton (2002), Banerjee and Mizen (2006) (and assuming for simplicity  $b_0 = 0$ , as in Hamilton, 2002), we derive the first order condition for cost minimization

$$E_t((q_t - u_{ct}) - \beta(q_{t+1} - u_{c,t+1}) + \beta b_2(h_t - b_4 - b_3 s_{t+1}(\theta))) = 0,$$

where  $q_t$  is quantity produced,  $u_{ct}$  is the shock to marginal cost of production and  $\beta$  is the discount rate. As in Hamilton (2002), we assume  $u_{ct} = s_t + v_{ct}$ , where  $v_{ct}$  is white noise, which implies that  $u_{ct}$  and  $s_t$  are cointegrated. This might occur if for example technological advance (an upward trend in  $u_{ct}$ ) generates an upward trend in sales. Noting that  $a_{j+1}(-\theta) - a_j(-\theta) = a_{j+1}(-\theta - 1)$ , it can be easily shown that  $s_{t+1}(\theta) = s_t(\theta) + s_{t+1}(\theta + 1)$ , so, given that  $q_t = \Delta h_t + s_t$ , it can be derived that

$$\beta b_2(h_t - b_4 - b_3 s_t(\theta)) + \Delta h_t - \beta E_t(\Delta h_{t+1}) = \beta b_2 b_3 E_t(s_{t+1}(\theta + 1)) + v_{ct}. \quad (13)$$

We interpret (13). Let  $\xi_t$  be an arbitrary  $I(0)$  process and  $\zeta_t = \sum_{j=0}^{t-1} a_j(d) \xi_{t-j}$ , so  $\zeta_t \sim I(d)$ . Clearly  $E_{t-1}(\zeta_t) = \zeta_t - \xi_t$ , so  $E_{t-1}(\zeta_t)$  inherits the integration order of  $\zeta_t$ . Noting that  $s_{t+1}(\theta + 1) \sim I(\delta - 1)$ , the immediate implication of (13) is that necessarily  $h_t \sim I(\delta)$  and the linear combination  $h_t - b_3 s_t(\theta)$  is cointegrating, so our particular parameterization of the optimal inventory leads to UC.

Next, we proceed to the statistical analysis. We just focus on estimating the UC relationship  $h_t - b_4 - b_3 s_t(\theta)$ , although, given that (13) suggests much richer dynamics than just those implied by UC, practitioners could consider our empirical analysis simplistic. However, a complete treatment of (13) would require more sophisticated techniques than those provided in this paper, and, at this

stage, our aim is just to propose a sensible methodology incorporating the techniques developed in the paper. In order to support the use of the UC techniques, we first test for equality of the integration orders  $\delta$  and  $\delta + \theta$  of the two observable series  $h_t$  and  $s_t$ , respectively. In particular, we estimated semiparametrically their respective integration orders by local Whittle, using second differences of the series and then adding back 2. We present these estimates in Table 9 for a wide set of bandwidths  $m_i$ ,  $i = 1, \dots, 10$  (setting  $m_1 = 16$ ,  $m_{i+1} = m_i + 1$ ). In all cases  $\widehat{\delta} > \widehat{\delta + \theta}$ , which supports the likability of a negative  $\theta$ : Based on these estimates, we tested formally for equality of the orders (that is,  $\theta = 0$ ) by means of Robinson and Yajima's (2002) test statistic  $\widehat{\tau}_{hs}$  with trimming sequence  $h(n) = 0$ . Theoretically  $h(n)$  should be a positive sequence tending to zero, but in practice, for a given sample, any positive number, no matter how small it is, satisfies the required condition for  $h(n)$ , noting that for positive choices of  $h(n)$ , the presented statistics should be slightly smaller (in absolute value). Results for the same set of bandwidths are presented in Table 9, and, although they are not entirely conclusive (note that the presented values should be compared with critical values from  $N(0, 1)$ ), they cast serious doubt on the  $\theta = 0$  hypothesis (this is clearly the case for large  $m_i$ 's).

Next, we test for UC by means of the  $X^{**}$  statistic of Robinson (2008). This statistic, unlike other proposals, is robust to distinct integration orders and, particularized to the bivariate situation, captures the possibility of UC. The values of  $X^{**}$  for bandwidths  $m_i$ ,  $i = 1, \dots, 10$  (setting  $m_1 = 26$ ,  $m_{i+1} = m_i + 1$ ) are presented in Table 9. These values need to be compared with nominal ones from a  $\chi_1^2$  distribution, so along with the results concerning the test for  $\theta = 0$ , the likability of UC is supported (at least for large  $m_i$ ). As shown by Robinson (2008),  $X^{**}$  displays very poor power for small bandwidths, which justifies our choice of relatively large  $m_i$ 's.

Finally, we estimate  $\nu$ ,  $\theta$  in (2) setting  $y_t = \Delta h_t$ ,  $x_t = \Delta s_t$ . Using model (2), (3) with raw variables is not realistic because (3) imposes a zero mean to  $x_t$ . Thus we avoid this problem by modelling first differenced observables, noting that the UC structure is preserved under differencing. In view of the local Whittle estimates provided in Table 9, the nonstationarity of  $y_t$  appears to be supported by the data. We also took this opportunity to examine the issue of truncation, which always arises when dealing with fractional models. In (2), (3),



the truncation is inherent in the model, so there is no “error” associated with it. However, the model reflects the time when the data begins, and if we were to drop the first observation, say, and start the model off at the next one, the degree of filtering applied to all subsequent observations would change, and it is possible that this could have a marked effect, especially with nonstationary data. Thus, in Table 9 we report computations of our estimates based on the last  $n' = n - j$  observations, for  $j = 0, 1, \dots, 9$ , in order to explore sensitivity to starting value. Finally, given that we are modelling first differenced data,  $\mu \neq 0$  implies the presence of a linear trend in levels. We analyze heuristically the effect of omitting this linear trend by estimating a more parsimonious model where knowledge of  $\mu = 0$  is imposed. These alternative estimators, which we denote by  $\hat{\nu}_{\mu 0}, \hat{\theta}_{\mu 0}$ , minimize a very similar loss function to  $Q_n(e)$ , with the only difference that sample means are not subtracted in (4). Also, the limiting theory for  $\hat{\nu}_{\mu 0}, \hat{\theta}_{\mu 0}$ , is almost identical to that for  $\hat{\nu}, \hat{\theta}$ , just accounting for the effect of the sample means (which just affects  $\Xi(\gamma, \delta)$ ). Results are reported in Table 9 for different values of  $n'$ . They are similar for both estimated models (possibly suggesting that  $\mu = 0$ ) and quite robust to different  $n'$ . The estimates of  $\theta$  correspond approximately to the evidence based on local Whittle estimates and  $\hat{\nu}, \hat{\nu}_{\mu 0}$ , differ heavily from the OLS of  $h_t$  on  $s_t$  (including also a constant in the regression), which takes value 2.08. Note that given that it is likely that  $\theta < 0$ , in view of our Monte Carlo experiment, we could have anticipated that OLS would suffer a severe positive bias, which might explain the huge difference between  $\hat{\nu}, \hat{\nu}_{\mu 0}$  and OLS.

Finally, our results allow us to calculate the estimated optimal inventory  $\hat{\nu}_{s_t}(\hat{\theta})$ . For example, taking  $\hat{\nu} = .657, \hat{\theta} = -.296$  (which correspond to  $n' = 76$  and model imposing  $\mu = 0$ ),  $\hat{\nu}_{s_t}(\hat{\theta}) = .657s_t + .194s_{t-1} + .126s_{t-2} + .096s_{t-3} + \dots$

## 5. Final comments

In this paper we have proposed nonlinear least squares estimators of long run parameters in a bivariate UC framework, where the integration orders of the observables might be different, but their corresponding balanced versions (with equal integration orders after filtering) are cointegrated in the usual sense. The UC setting appears to be relevant in at least two empirically relevant research areas: multicointegration and predictive regressions. We have analyzed the limiting properties of these estimators, and discovered three noticeable features:

the possible superconsistency of the estimator of the difference between memory parameters, the singularity of the joint limiting distribution of the estimators, and the particular form of the joint limiting distributions, which can be partially characterized by a modified version of the “Type II” fractional Brownian motion.

Although our bivariate system (2), (3), extends in a nontrivial direction some results given in the literature (e.g., Robinson and Marinucci, 2001), there are many different aspects which are not covered by the present paper, but which will be object of future research.

1. It is straightforward to derive a frequency domain representation for our estimators, and noting that both  $\theta$  and  $\nu$  describe long run properties of the observables, this representation leads to consideration of narrow band (NB) versions of  $\hat{\nu}$ ,  $\hat{\theta}$ . As in Robinson and Marinucci (2001), NB estimators, whose focus is the long-run components of the series, could display better asymptotic properties than  $\hat{\nu}$ ,  $\hat{\theta}$ . In particular, in view of Robinson and Marinucci (2001), it is expected that NB enjoy a faster convergence rate and different limiting distribution when  $\gamma + \delta < 1$  or  $\gamma + \delta = 1$  (with  $\gamma > 0$ ), the same rate (but different distribution) when  $\gamma = 0$ ,  $\delta = 1$ , and identical properties to those in Theorem 2 when  $\gamma + \delta > 1$ . In view of the proof of Theorem 2 and results in Robinson and Marinucci (2001), the properties of the NB should be easily derivable.
2. Both nonlinear least squares and NB estimators lead, in general, to non-standard limiting distributions, which make them unsuitable for inference. Thus, improved methods which correct for the endogeneity caused by the relation between  $u_{1t}$  and  $u_{2t}$ , in the lines developed by Robinson and Hualde (2003), Hualde and Robinson (2007) (as those hinted by Hualde (2006) in the parametric setting), could be developed, but this is definitely a nontrivial extension of the results in the present paper. It is expected that these improved estimators would enjoy in some cases faster convergence rates than those displayed by  $\hat{\nu}$ ,  $\hat{\theta}$ , or, alternatively, by their NB versions. Generalizing this improved theory to a semiparametric situation (as in Hualde and Robinson, 2010, for the balanced case) could be even more demanding, but these extensions are very relevant, because they could lead to estimators with standard asymptotic properties and, therefore, straightforward inference would apply.

3. As in Robinson and Marinucci (2001), our results only cover cases where at least one of the observables ( $y_t$ ) is purely nonstationary (because we impose  $\delta > 1/2$ ). Note that  $x_t$  is allowed to be asymptotically stationary (when  $\theta$  is negative), but we do not cover interesting cases (for example for financial data) where both observables are asymptotically stationary. In this case, our guess is that  $\hat{\nu}$ ,  $\hat{\theta}$  would be in general inconsistent, but their NB versions could retain consistency. Improved methods for this particular case could be developed in the spirit of the “weak cointegration” literature (see e.g. Hualde and Robinson, 2007, 2010).
4. Undoubtedly, our bivariate system could be extended to multivariate settings. However, non-trivial extensions allowing for different integration orders among the observables, cointegrating errors and imbalance parameters, raise important difficulties regarding the identification and estimation of a general multivariate system. Some of these difficulties were considered by Hualde and Robinson (2010), but allowing for the greater flexibility implied by the possibly unbalanced series, complicates matters substantially.

## Appendix A. Proof of Theorems

**Proof of Theorem 1.** The behaviour of  $Q_n(e)$  is not uniform across the parameter space, so it is not possible to rely on uniform convergence arguments to justify consistency. Thus our proof will be based on an alternative strategy similar to that of Hualde and Robinson (2011) (HR hereinafter). For some arbitrarily small  $\varepsilon > 0$ , such that  $\varepsilon < \min\{\theta - \nabla_1, \nabla_2 - \theta\}$ , define  $N_\varepsilon = \{e : |e - \theta| < \varepsilon\}$ ,  $\overline{N}_\varepsilon = \{e : |e - \theta| \geq \varepsilon\}$ . Then, as in Robinson (1995),

$$\Pr\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) = \Pr\left(\inf_{\overline{N}_\varepsilon \cap \Xi} Q_n(e) \leq \inf_{N_\varepsilon \cap \Xi} Q_n(e)\right) \leq \Pr\left(\inf_{\overline{N}_\varepsilon \cap \Xi} S_n(e) \leq 0\right),$$

where  $S_n(e) = n^{-2\delta}(Q_n(e) - Q_n(\theta))$ . In order to simplify subsequent notation let  $c = \delta + \theta - e$ , noting that  $x_t(e) \sim I(c)$ . Then  $\Pr(\inf_{\overline{N}_\varepsilon \cap \Xi} S_n(e) \leq 0) = \Pr(\inf_{c \in \mathcal{I}} S_n(\delta + \theta - c) \leq 0)$ , where  $\mathcal{I} = [\delta + \theta - \nabla_2, \delta - \varepsilon] \cup [\delta + \varepsilon, \delta + \theta - \nabla_1]$ .

We give the proof for the most general case where  $\delta + \theta - \nabla_2 < 1/2$ , but our proof trivially covers the  $\delta + \theta - \nabla_2 \geq 1/2$  situation, for which some of the steps described below are superfluous. We split the admissible  $c$ -interval onto four disjoint intervals:  $\mathcal{I}_1 = [\delta + \theta - \nabla_2, 1/2 - \eta]$ ,  $\mathcal{I}_2 = (1/2 - \eta, 1/2)$ ,  $\mathcal{I}_3 = [1/2, 1/2 + \eta)$ ,

$\mathcal{I}_4 = [1/2 + \eta, \delta - \varepsilon] \cup [\delta + \varepsilon, \delta + \theta - \nabla_1]$ , where  $\eta > 0$  is arbitrarily small and  $\eta < \min\{1/2, 1/2 - (\delta + \theta - \nabla_2), \delta - 1/2 - \varepsilon\}$ . The idea of splitting the  $\mathcal{I}$  interval in four subsets reflects the distinct behaviour of  $x_t(e)$  when  $c > 1/2$  or  $c < 1/2$ , whereas the behaviour of  $x_t(e)$  when  $c$  is in a neighborhood of  $1/2$  requires a special treatment. Then Theorem 1 holds on showing that

$$\Pr\left(\inf_{\mathcal{I}_i} S_n(\delta + \theta - c) \leq 0\right) \rightarrow 0 \text{ as } n \rightarrow \infty, i = 1, 2, 3, 4. \quad (14)$$

Clearly

$$\begin{aligned} S_n(\delta + \theta - c) &= \frac{\nu^2}{n^{2\delta}} \left( C_{x(\theta)} - \frac{C_{x(e)x(\theta)}^2}{C_{x(e)}} \right) - \frac{2\nu C_{x(e)x(\theta)} C_{x(e)u_1(-\gamma)}}{n^{2\delta} C_{x(e)}} - \frac{C_{x(e)u_1(-\gamma)}^2}{n^{2\delta} C_{x(e)}} \\ &\quad + \frac{2\nu C_{x(\theta)u_1(-\gamma)}}{n^{2\delta}} + \frac{C_{x(\theta)u_1(-\gamma)}^2}{n^{2\delta} C_{x(\theta)}} = \sum_{i=1}^5 p_i. \end{aligned}$$

First, we show (14) for  $i = 1, 2, 3$ . Clearly

$$\inf_{\mathcal{I}_i} S_n(\delta + \theta - c) \geq \frac{\nu^2 C_{x(\theta)}}{n^{2\delta}} - \sup_{\mathcal{I}_i} b_n(c),$$

where  $b_n(c) = \nu^2 n^{-2\delta} C_{x(e)x(\theta)}^2 C_{x(e)}^{-1} + \sum_{i=2}^5 |p_i|$ . Then

$$\Pr\left(\inf_{\mathcal{I}_i} S_n(\delta + \theta - c) \leq 0\right) \leq \Pr\left(\frac{\nu^2 C_{x(\theta)}}{n^{2\delta}} \leq \sup_{\mathcal{I}_i} b_n(c)\right).$$

Let  $D(r; \cdot)$ ,  $E(r; \cdot)$  be  $W(r; \cdot)$  or  $A(r; \cdot)$  (defined in (6), (7)),  $\tilde{D}(r; \cdot) = D(r; \cdot) - \int_0^1 D(s; \cdot) ds$  and define for any  $2 \times 1$  vectors  $v, w$

$$B_{D(c)E(d)}^{vw} = v^T B(1) \int_0^1 \tilde{D}(r; c) \tilde{E}^T(r; d) dr B^T(1) w, \quad B_{D(c)}^v = B_{D(c)D(c)}^{vv}.$$

By Assumption 1, Marinucci and Robinson (2000) (MR hereinafter) and the continuous mapping theorem  $n^{-2\delta} C_{x(\theta)} \rightarrow_d B_{W(\delta)}^\xi > 0$ , a.s., where  $\xi = (0, 1)^T$ . Then, (14) for  $i = 1, 2, 3$ , holds on showing that  $\sup_{\mathcal{I}_i} b_n(c) = o_p(1)$ . First, note that  $p_4$  and  $p_5$  are not  $c$ -dependent, and by results in Robinson and Marinucci (2001) (RM hereinafter),  $n^{-2\delta} C_{x(\theta)u_1(-\gamma)} = o_p(1)$ , so  $|p_4| + |p_5| = o_p(1)$ . Also, as

$\delta > \gamma$ ,  $p_2$  and  $p_3$  cannot be larger in order of magnitude than  $n^{-2\delta} C_{x(e)}^{-1} C_{x(e)x(\theta)}^2$ , so (14) holds on showing that  $\sup_{\mathcal{I}_i} n^{-2\delta} C_{x(e)}^{-1} C_{x(e)x(\theta)}^2 = o_p(1)$ . For  $i = 1, 2$ ,

$$\sup_{\mathcal{I}_i} \frac{C_{x(e)x(\theta)}^2}{n^{2\delta} C_{x(e)}} \leq \frac{\sup_{\mathcal{I}_i} \left( \frac{1}{n^{\delta+\frac{1}{2}}} C_{x(e)x(\theta)} \right)^2}{\inf_{\mathcal{I}_i} \frac{1}{n} C_{x(e)}} = o_p(1), \quad i = 1, 2,$$

by (60), (61) of Lemma 1 and (68), (69) of Lemma 2, for  $i = 1, 2$ , respectively. Similarly, for  $i = 3$ ,

$$\sup_{\mathcal{I}_3} \frac{C_{x(e)x(\theta)}^2}{n^{2\delta} C_{x(e)}} \leq \frac{\sup_{\mathcal{I}_3} \left( \frac{1}{n^{\delta+c}} C_{x(e)x(\theta)} \right)^2}{\inf_{\mathcal{I}_3} \frac{1}{n^{2c}} C_{x(e)}} = o_p(1),$$

by (62) of Lemma 1 and (70) of Lemma 2.

The proof for  $i = 4$  is slightly different. Given that  $C_{x(e)} \geq 0$ ,

$$\Pr \left( \inf_{\mathcal{I}_4} S_n (\delta + \theta - c) \leq 0 \right) \leq \Pr \left( \inf_{\mathcal{I}_4} \frac{C_{x(e)}}{n^{2c}} S_n (\delta + \theta - c) \leq 0 \right). \quad (15)$$

Now

$$\inf_{\mathcal{I}_4} \frac{C_{x(e)}}{n^{2c}} S_n (\delta + \theta - c) \geq \nu^2 \inf_{\mathcal{I}_4} \frac{1}{n^{2(\delta+c)}} [C_{x(\theta)} C_{x(e)} - C_{x(e)x(\theta)}^2] - \sup_{\mathcal{I}_4} \frac{1}{n^{2c}} C_{x(e)} d_n(c),$$

where  $d_n(c) = \sum_{i=2}^5 |p_i|$ . Then

$$\begin{aligned} & \Pr \left( \inf_{\mathcal{I}_4} S_n (\delta + \theta - c) \leq 0 \right) \\ & \leq \Pr \left( \nu^2 \inf_{\mathcal{I}_4} \frac{1}{n^{2(\delta+c)}} [C_{x(\theta)} C_{x(e)} - C_{x(e)x(\theta)}^2] \leq \sup_{\mathcal{I}_4} \frac{1}{n^{2c}} C_{x(e)} d_n(c) \right). \end{aligned} \quad (16)$$

Defining for any  $a_t$ ,  $\tilde{a}_t = a_t - \bar{a}$ , it is straightforward to show that

$$\begin{aligned} C_{x(\theta)} C_{x(e)} - C_{x(e)x(\theta)}^2 &= \sum_{t=1}^{n-1} \sum_{k=t+1}^n (\tilde{x}_t(\theta) \tilde{x}_k(e) - \tilde{x}_k(\theta) \tilde{x}_t(e))^2 \\ &\geq \frac{1}{n^2} \left( \sum_{t=1}^{n-1} \sum_{k=t+1}^n (\tilde{x}_t(\theta) \tilde{x}_k(e) - \tilde{x}_k(\theta) \tilde{x}_t(e)) \right)^2. \end{aligned} \quad (17)$$

Noting that  $\sum_{l=0}^j a_l(c) = a_j(c+1)$ , so  $\sum_{l=1}^t x_l(e) = x_t(e-1)$  and  $x_t(e-1) -$

$x_{t-1}(e-1) = x_t(e)$ , using repeatedly summation by parts, the right side of (17) equals  $n^{-2}(g_n(c) + h_n(c))^2$ , where

$$\begin{aligned} g_n(c) &= x_n(e-1)x_n(\theta-1) - \frac{2}{n}x_n(e-1)x_{n-1}(\theta-2) \\ &\quad - 2\sum_{t=1}^n x_t(\theta)x_t(e-1) + \frac{2}{n}x_n(\theta-1)x_{n-1}(e-2), \end{aligned}$$

$h_n(c) = \sum_{t=1}^n x_t(\theta)x_t(e)$ . Then the right side of (16) is bounded by

$$\Pr\left(\inf_{\mathcal{I}_4}\left(\frac{1}{n^{\delta+c+1}}g_n(c)\right)^2 \leq \nu^{-2}\sup_{\mathcal{I}_4}\frac{1}{n^{2c}}C_{x(e)}d_n(c) + 2\sup_{\mathcal{I}_4}\frac{1}{n^{2(c+\delta)+2}}|g_n(c)h_n(c)|\right),$$

so the required result follows on showing that for an arbitrarily small  $\epsilon > 0$

$$\Pr\left(\inf_{\mathcal{I}_4}\left(\frac{1}{n^{\delta+c+1}}g_n(c)\right)^2 > \epsilon\right) \rightarrow 1 \text{ as } n \rightarrow \infty, \quad (18)$$

$$\sup_{\mathcal{I}_4}\frac{1}{n^{2c}}C_{x(e)}d_n(c) = o_p(1), \quad (19)$$

$$\sup_{\mathcal{I}_4}\frac{1}{n^{c+\delta+1}}|g_n(c)| = O_p(1), \quad (20)$$

$$\sup_{\mathcal{I}_4}\frac{1}{n^{c+\delta+1}}|h_n(c)| = o_p(1). \quad (21)$$

First, as  $|p_4| + |p_5| = o_p(1)$ , noting (62), (19) follows by showing

$$\sup_{\mathcal{I}_4}\frac{1}{n^{2c}}C_{x(e)} = O_p(1), \quad (22)$$

$$\sup_{\mathcal{I}_4}\frac{1}{n^{c+\delta}}C_{x(e)u_1(-\gamma)} = o_p(1). \quad (23)$$

First, (22) easily follows by very similar arguments to those in the proof of (62), whereas by the Cauchy inequality

$$\sup_{\mathcal{I}_4}\frac{1}{n^{\delta+c}}C_{x(e)u_1(-\gamma)} \leq \left\{\sup_{\mathcal{I}_4}\frac{1}{n^{2c}}C_{x(e)}\frac{1}{n^{2\delta}}C_{u_1(-\gamma)}\right\}^{\frac{1}{2}} = o_p(1),$$

by (22) and results in RM to establish (23). Next, (20), (21), follow by very similar arguments to those in the proof of (62). Finally, considering  $n^{-(\delta+c+1)}g_n(c)$

as a process indexed by  $c$ , we show first that

$$\frac{1}{n^{\delta+c+1}}g_n(c) \Rightarrow g(c), \quad (24)$$

where

$$g(c) = \xi^T B(1) \left[ W(1; c+1) W^T(1; \delta+1) - 2W(1; c+1) W^T(1; \delta+2) \right. \\ \left. + 2W(1; c+2) W^T(1; \delta+1) - 2 \int_0^1 W(r; \delta) W^T(r; c+1) dr \right] B^T(1) \xi,$$

where  $\Rightarrow$  means weak convergence in the space of continuous functions on  $\mathcal{I}_4$ . We give this space the uniform topology. Note that by integration by parts  $g(\delta) = 0$ , a.s., but for  $c \in \mathcal{I}_4$ ,  $g(c) \neq 0$ , a.s. Convergence of the finite dimensional distributions of the processes involved follows by MR, whereas tightness follows by almost identical arguments to those in HR. Thus (24) follows by the continuous mapping theorem. Then  $\inf_{\mathcal{I}_4} \left( \frac{1}{n^{\delta+c+1}}g_n(c) \right)^2 \Rightarrow \inf_{\mathcal{I}_4} g^2(c)$ , again by the continuous mapping theorem, where  $\inf_{\mathcal{I}_4} g^2(c) > 0$ , a.s, because the infimum is taken over a compact set. Thus, as  $n \rightarrow \infty$ ,

$$\Pr \left( \inf_{\mathcal{I}_4} \left( \frac{1}{n^{\delta+c+1}}g_n(c) \right)^2 > \epsilon \right) \rightarrow \Pr \left( \inf_{\mathcal{I}_4} g^2(c) > \epsilon \right),$$

and (18) follows because  $\epsilon$  is arbitrarily small. Thus we conclude the proof for  $i = 4$  and thence complete the proof of the theorem.

**Proof of Theorem 2.** First, we show that  $\hat{\theta}$  is  $n^\tau$ -consistent for a fixed  $\tau$  such that  $0 < \tau < \delta - \max\{\gamma, 1/2\}$ . Noting the results in the proof of Theorem 1, the result holds on establishing that

$$\Pr \left( \inf_{\bar{\mathcal{I}}_{n^\tau}} S_n(\delta + \theta - c) \leq 0 \right) \rightarrow 0 \text{ as } n \rightarrow \infty, \quad (25)$$

where  $\bar{\mathcal{I}}_{n^\tau} = [\delta - \epsilon, \delta - \epsilon/n^\tau] \cup [\delta + \epsilon/n^\tau, \delta + \epsilon]$ . Denote  $\mathcal{I}_\epsilon = [\delta - \epsilon, \delta + \epsilon]$ . As in (15)

$$\Pr \left( \inf_{\bar{\mathcal{I}}_{n^\tau}} S_n(\delta + \theta - c) \leq 0 \right) \leq \Pr \left( \inf_{\bar{\mathcal{I}}_{n^\tau}} \frac{C_{x(\epsilon)}}{n^{2c}} S_n(\delta + \theta - c) \leq 0 \right).$$

Now

$$\inf_{\bar{\mathcal{I}}_{n^\tau}} \frac{C_{x(e)}}{n^{2c}} S_n(\delta + \theta - c) \geq \nu^2 \inf_{\bar{\mathcal{I}}_{n^\tau}} \frac{1}{n^{2(\delta+c)}} [C_{x(\theta)} C_{x(e)} - C_{x(e)x(\theta)}^2] - \sup_{\mathcal{I}_\varepsilon} \frac{1}{n^{2c}} C_{x(e)} d_n(c),$$

so that

$$\Pr \left( \inf_{\bar{\mathcal{I}}_{n^\tau}} S_n(\delta + \theta - c) \leq 0 \right) \leq \Pr \left( \nu^2 \inf_{\bar{\mathcal{I}}_{n^\tau}} \frac{C_{x(\theta)} C_{x(e)} - C_{x(e)x(\theta)}^2}{n^{2(\delta+c)}} \leq \sup_{\mathcal{I}_\varepsilon} \frac{C_{x(e)} d_n(c)}{n^{2c}} \right).$$

We introduce some notation. First, noting that  $x_t(e)/n^c = n^{-c} \sum_{j=0}^{t-1} a_j(c) u_{2,t-j}$ , denote  $f_t(c) = x_t(e)/n^c$  (there is a slight abuse of notation here because  $f_t(c)$  is  $n$ -dependent). Also, as in Remark 1,

$$a'_j(c) = (\psi(j+c) - \psi(c)) a_j(c). \quad (26)$$

Then it is simple to derive that  $f'_t(c) = df_t(c)/dc = g_t(c) - r(c) f_t(c)$ , where  $g_t(c) = n^{-c} \sum_{j=0}^{t-1} \psi(j+c) a_j(c) u_{2,t-j}$ ,  $r(c) = \psi(c) + \log n$ . For  $c \in \bar{\mathcal{I}}_{n^\tau}$ , by the mean value theorem,  $f_t(c) = f_t(\bar{c}) + (c - \bar{c}) f'_t(\bar{c})$ , where  $|\bar{c} - \delta| \leq |c - \delta|$  and  $\bar{c} \in \mathcal{I}_\varepsilon$ . Then it is straightforward to show that

$$\frac{1}{n^{2(\delta+c)}} (C_{x(\theta)} C_{x(e)} - C_{x(e)x(\theta)}^2) = C_{f(\delta)} C_{f(c)} - C_{f(c)f(\delta)}^2 = (c - \delta)^2 P_n(\bar{c}),$$

where  $P_n(\bar{c}) = C_{f(\delta)} C_{f'(\bar{c})} - C_{f(\delta)f'(\bar{c})}^2$ . Thus

$$\Pr \left( \inf_{\bar{\mathcal{I}}_{n^\tau}} S_n(\delta + \theta - c) \leq 0 \right) \leq \Pr \left( \nu^2 \inf_{\bar{\mathcal{I}}_{n^\tau}} (c - \delta)^2 \inf_{\mathcal{I}_\varepsilon} P_n(c) \leq \sup_{\mathcal{I}_\varepsilon} C_{f(c)} d_n(c) \right),$$

so given that  $\inf_{\bar{\mathcal{I}}_{n^\tau}} (c - \delta)^2 = \varepsilon^2/n^{2\tau}$ , (25) holds on showing that, as  $n \rightarrow \infty$ ,

$$n^{2\tau} \sup_{\mathcal{I}_\varepsilon} C_{f(c)} d_n(c) = o_p(1), \quad (27)$$

$$\Pr \left( \inf_{\mathcal{I}_\varepsilon} P_n(c) \leq \varepsilon \right) \rightarrow 0, \quad (28)$$

for  $\varepsilon > 0$  arbitrarily small. First, (27) holds on showing  $n^{2\tau} \sup_{\mathcal{I}_\varepsilon} C_{f(c)} \sum_{i=2}^5 |p_i| = o_p(1)$ . First, noting that as in (22),  $\sup_{\mathcal{I}_\varepsilon} C_{f(c)} = O_p(1)$ , then  $n^{2\tau} \sup_{\mathcal{I}_\varepsilon} C_{f(c)} \sum_{i=4}^5 |p_i| =$



$o_p(1)$  by results in RM. Thus, given that as in (62),  $\sup_{\mathcal{I}_\varepsilon} C_{f(c)f(\delta)} = O_p(1)$ , (27) holds on showing that  $n^{2\tau} \sup_{\mathcal{I}_\varepsilon} n^{-(\delta+c)} C_{x(e)u_1(-\gamma)} = o_p(1)$ , which holds straightforwardly by the Cauchy inequality, (22) and RM, as  $\tau < \delta - \max\{\gamma, 1/2\}$ . Next, by Lemma 3, for an arbitrarily small  $\rho > 0$ ,

$$\Pr\left(\inf_{\mathcal{I}_\varepsilon} P_n(c) \leq \epsilon\right) \leq \Pr\left(\sup_{\mathcal{I}_\varepsilon} |r_n(c)| > \rho\right) + \Pr\left(\inf_{\mathcal{I}_\varepsilon} X_n(c) \leq \epsilon + \rho\right). \quad (29)$$

By (83) the first term on the right side of (29) tends to zero as  $n \rightarrow \infty$ . In view of (82), noting that  $X_{1n}(c) \geq 0$ , the second term is bounded by

$$\Pr\left(\inf_{\mathcal{I}_\varepsilon} X_{2n}(c) + \inf_{\mathcal{I}_\varepsilon} X_{3n}(c) \leq \epsilon + \rho\right),$$

which for an arbitrarily small  $\tau > 0$  is bounded by

$$\Pr\left(\inf_{\mathcal{I}_\varepsilon} X_{3n}(c) \leq \epsilon + \rho + \tau\right) + \Pr\left(\inf_{\mathcal{I}_\varepsilon} X_{2n}(c) \leq -\tau\right). \quad (30)$$

Then, as in (24), we have

$$X_{3n}(c) \Rightarrow B_{W(\delta)}^\xi B_{A(c)}^\xi - \left(B_{W(\delta)A(c)}^{\xi\xi}\right)^2 \equiv X_3(c), \quad (31)$$

$$X_{2n}(c) \Rightarrow 2\psi(c) \left(B_{W(\delta)}^\xi B_{W(c)A(c)}^{\xi\xi} - B_{W(\delta)W(c)}^{\xi\xi} B_{W(\delta)A(c)}^{\xi\xi}\right) \equiv X_2(c). \quad (32)$$

Under Assumption 1, (31), (32) follow by the continuous mapping theorem, noting that convergence of the corresponding finite dimensional distributions of the processes involved follow by MR and Hualde (2012), whereas tightness can be justified as in HR. Then, by the continuous mapping theorem,  $\inf_{\mathcal{I}_\varepsilon} X_{3n}(c) \Rightarrow \inf_{\mathcal{I}_\varepsilon} X_3(c) > 0$ , a.s., so the first term in (30) can be made arbitrarily small on setting  $\epsilon, \rho, \tau$ , arbitrarily close to zero. Similarly

$$\inf_{\mathcal{I}_\varepsilon} X_{2n}(c) \Rightarrow \inf_{\mathcal{I}_\varepsilon} X_2(c). \quad (33)$$

Noting that  $X_2(c)$  evaluated at  $c = \delta$  is identically zero, by (33) and continuity of  $X_2(c)$ , the second probability in (30) can be made arbitrarily small on setting  $\varepsilon$  close enough to zero, to conclude the proof of (25), and hence  $n^\tau$ -consistency of  $\hat{\theta}$  is established.

Next, by the mean value theorem,  $\widehat{\theta} - \theta = -Q'_n(\theta) / Q''_n(\bar{\theta})$ , where  $|\bar{\theta} - \theta| \leq |\widehat{\theta} - \theta|$  and  $Q'_n(\theta)$ ,  $Q''_n(\bar{\theta})$ , are the first and second derivatives of  $Q_n(e)$ , evaluated at  $e = \theta$ ,  $\bar{\theta}$ , respectively. Define  $g_n(\gamma, \delta) = n^{2\delta} g_n^{-1}(\gamma, \delta)$  (see (9)). As will be shown below,  $g_n(\gamma, \delta)$  and  $n^{2\delta}$  are the exact normalizations for  $Q'_n(\theta)$ ,  $Q''_n(\theta)$ , respectively. Thus

$$n^{2\delta} g_n^{-1}(\gamma, \delta) (\widehat{\theta} - \theta) = -n^{2\delta} g_n^{-1}(\gamma, \delta) \frac{Q'_n(\theta)}{Q''_n(\bar{\theta})} + o_p(1), \quad (34)$$

holds on showing that  $n^{-2\delta} (Q''_n(\bar{\theta}) - Q''_n(\theta)) = o_p(1)$ , which follows if

$$\sup_{c \in \mathcal{I}_{n^\tau}} |n^{-2\delta} (Q''_n(e) - Q''_n(\theta))| = o_p(1), \quad (35)$$

where  $\mathcal{I}_{n^\tau}$  is defined in Lemma 4 below. Taking derivatives in (4),  $Q'_n(e) = -2\widehat{\mathcal{V}}(e) C_{yx'(e)} + 2\widehat{\mathcal{V}}^2(e) C_{x(e)x'(e)}$ , and

$$\begin{aligned} Q''_n(e) &= -2\widehat{\mathcal{V}}'(e) C_{yx'(e)} + 4\widehat{\mathcal{V}}'(e) \widehat{\mathcal{V}}(e) C_{x(e)x'(e)} - 2\widehat{\mathcal{V}}(e) C_{yx''(e)} \\ &\quad + 2\widehat{\mathcal{V}}^2(e) C_{x(e)x''(e)} + 2\widehat{\mathcal{V}}^2(e) C_{x'(e)}, \end{aligned} \quad (36)$$

where  $x'_t(e) = dx_t(e) / de$ ,  $x''_t(e) = d^2x_t(e) / de^2$  and

$$\widehat{\mathcal{V}}'(e) = \frac{d\widehat{\mathcal{V}}(e)}{de} = C_{x(e)}^{-1} C_{yx'(e)} - 2\widehat{\mathcal{V}}(e) C_{x(e)}^{-1} C_{x(e)x'(e)}. \quad (37)$$

The proof follows after tedious but mainly straightforward calculations, so we just present the main steps. By Lemma 4 and results in MR, RM, it is straightforward to obtain that

$$\begin{aligned} \sup_{\mathcal{I}_{n^\tau}} |\widehat{\mathcal{V}}(e) - \widehat{\mathcal{V}}(\theta)| &= O_p(n^{-\tau} \log n), \\ \sup_{\mathcal{I}_{n^\tau}} |\widehat{\mathcal{V}}'(e) - \widehat{\mathcal{V}}'(\theta)| &= O_p(n^{-\tau} \log^2 n). \end{aligned} \quad (38)$$

Similarly, analyzing every term in  $Q''_n(e) - Q''_n(\theta)$ , we conclude that

$$\sup_{\mathcal{I}_{n^\tau}} |Q''_n(e) - Q''_n(\theta)| = O_p(n^{2\delta-\tau} \log^3 n),$$

to conclude for (35). Then

$$n^{2\delta} g_n^{-1}(\gamma, \delta) \left( \widehat{\theta} - \theta \right) = -n^{2\delta} g_n^{-1}(\gamma, \delta) \frac{F_n(\theta)}{H_n(\theta)} + o_p(1),$$

where  $F_n(\theta) = C_{x(\theta)} Q_n'(\theta)$ ,  $H_n(\theta) = C_{x(\theta)} Q_n''(\theta)$ .  $F_n$  and  $H_n$  are introduced because in our particular setting it is simpler to consider the joint limiting behaviour of  $(n^{-4\delta} H_n(\theta), n^{-2\delta} g_n^{-1}(\gamma, \delta) F_n(\theta))$  (where the extra  $n^{-2\delta}$  in the normalization is due to the presence of  $C_{x(\theta)}$ , see RM), in order to derive the asymptotic distribution of  $\widehat{\theta}$ , than that of  $(n^{-2\delta} Q_n''(\theta), g_n^{-1}(\gamma, \delta) Q_n'(\theta))$ .

We first analyze the behaviour of  $H_n(\theta)$ . Let

$$x_{t,n}(\theta) = \sum_{j=0}^{t-1} u_{2,t-j} a_j(\delta) \sum_{k=j}^n (k + \delta)^{-1}.$$

Then, noting the approximation in Lemma 6, by Theorem 1 of MR, Theorem 1 of Hualde (2012) and the continuous mapping theorem, as  $n \rightarrow \infty$ ,

$$\frac{1}{n^{4\delta}} H_n(\theta) \rightarrow_d 2\nu^2 \left( B_{W(\delta)}^\xi B_{A(\delta)}^\xi - \left( B_{A(\delta)W(\delta)}^{\xi\xi} \right)^2 \right) > 0, \text{ a.s.} \quad (39)$$

Next, we analyze the behaviour of  $F_n(\theta)$  noting the approximation given in Lemma 5. Here, the limiting properties of  $C_{u_1(-\gamma)x(\theta)}$ ,  $C_{x(\theta)}$ , follow directly from RM, those of  $C_{x(\theta)x_n(\theta)}$  from MR, Theorem 1 of Hualde (2012) and the continuous mapping theorem, so we concentrate on the analysis of  $C_{u_1(-\gamma)x_n(\theta)}$ . This element equals

$$\begin{aligned} & \log(n+1+\delta) \sum_{t=1}^n u_{1t}(-\gamma) x_t(\theta) - \sum_{t=1}^n u_{1t}(-\gamma) \sum_{j=0}^{t-1} a_j(\delta) \log(j+\delta) u_{2,t-j} \\ & + \sum_{t=1}^n u_{1t}(-\gamma) \sum_{j=0}^{t-1} a_j(\delta) c_{j,n} u_{2,t-j} - \frac{1}{n} \sum_{t=1}^n u_{1t}(-\gamma) \sum_{t=1}^n x_{t,n}(\theta) \end{aligned} \quad (40)$$

where  $c_{j,n} = \sum_{k=j}^n \frac{1}{k+\delta} - \int_j^{n+1} \frac{1}{z+\delta} dz$ . First, we discuss the  $\gamma + \delta < 1$  case. Here, by Theorems 4.1 and 5.1 of RM, as  $n \rightarrow \infty$ ,

$$\frac{1}{n \log n} \log(n+1+\delta) \sum_{t=1}^n u_{1t}(-\gamma) x_t(\theta) \rightarrow_p \int_{-\pi}^{\pi} a(\gamma; \lambda) a(\delta; -\lambda) f_{12}(\lambda) d\lambda, \quad (41)$$

where  $a(d; \lambda) = \sum_{j=0}^{\infty} a_j(d) e^{ij\lambda}$  and  $f_{ij}(\lambda)$  the  $ij$ th element of  $f(\lambda)$  (which is the spectral density matrix of  $(u_{1t}, u_{2t})^T$ ). Next, the expectation of the second term in (40) is

$$\sum_{t=1}^n \sum_{j=0}^{t-1} a_j(\gamma) \sum_{k=0}^{t-1} a_k(\delta) \log(k+\delta) \gamma_{j-k}, \quad (42)$$

where  $\gamma_j = E(u_{10}u_{2j})$ . Under Assumption 1,  $\sum_{j=0}^{\infty} |\gamma_j| < \infty$ , so (42) can be easily shown to be  $O(n)$ . Also, by a simple extension of Theorem 5.1 of RM

$$\text{Var} \left( \sum_{t=1}^n u_{1t}(-\gamma) \sum_{j=0}^{t-1} a_j \log(j+\delta) u_{2,t-j} \right) = O(n^{2(\gamma+\delta)} \log^2 n),$$

so the second term in (40) is  $o_p(n \log n)$ . Next, noting that

$$|c_{j,n}| \leq \sum_{k=j}^n \sup_{k \leq v \leq k+1} \left( \frac{1}{k+\delta} - \frac{1}{v+\delta} \right) = \sum_{k=j}^n \frac{1}{(k+\delta)(k+1+\delta)} \leq K j^{-1}, \quad (43)$$

uniformly in  $n$ , where  $K$  denotes an arbitrarily large finite constant, it is easily shown that the third term in (40) is also  $o_p(n \log n)$ . Finally, by MR and Hualde (2012), the fourth term in (40) is also  $o_p(n \log n)$ , so that

$$\frac{1}{n \log n} C_{u_1(-\gamma)x_n(\theta)} \rightarrow_p \int_{-\pi}^{\pi} a(\gamma; \lambda) a(\delta; -\lambda) f_{12}(\lambda) d\lambda.$$

Next, we discuss the  $\gamma + \delta = 1$ ,  $\gamma > 0$  case. By Theorems 4.2 and 5.1 of RM, as  $n \rightarrow \infty$ ,

$$\frac{1}{n \log^2 n} \log(n+1+\delta) \sum_{t=1}^n u_{1t}(-\gamma) x_t(\theta) \rightarrow_p 2f_{12}(0) \sin(\delta\pi). \quad (44)$$

Next, the expectation of the second term in (40) equals

$$\begin{aligned} & 2\pi f_{12}(0) \sum_{t=1}^n \sum_{j=0}^{t-1} a_j(\gamma) a_j(\delta) \log(j+\delta) + \sum_{t=1}^n \left( \int_{-n^{-1}}^{n^{-1}} + \int_{n^{-1} \leq |\lambda| \leq \varepsilon} + \int_{\varepsilon \leq |\lambda| \leq \pi} \right) \\ & \times \sum_{j=0}^{t-1} a_j(\gamma) e^{ij\lambda} \sum_{k=0}^{t-1} a_k(\delta) \log(k+\delta) e^{-ik\lambda} \bar{f}_{12}(\lambda) d\lambda, \end{aligned} \quad (45)$$

where  $\bar{f}_{12}(\lambda) = f_{12}(\lambda) - f_{12}(0)$ , and for a fixed  $\eta > 0$ ,  $\varepsilon \in (0, \pi)$  is chosen such that  $\sup_{|\lambda| < \varepsilon} |\bar{f}_{12}(\lambda)| < \eta$ . Note that by Assumption 1, given any  $\eta$ , such  $\varepsilon$  always exists. As in the proof of Theorem 4.2 of RM the first term in (45) is asymptotically equivalent (meaning that the ration between this term and (46) below tends to one as  $n \rightarrow \infty$ ) to

$$\frac{2\pi f_{12}(0)}{\Gamma(\gamma)\Gamma(\delta)} \sum_{t=1}^n (n-t) t^{-1} \log t = f_{12}(0) \sin(\delta\pi) n \log^2 n + o(n \log^2 n), \quad (46)$$

by the reflection formula for the Gamma function, approximating summation by integrals, and noting that  $\int_1^{n+1} x^{-1} \log x dx = \log^2(n+1)/2$ . Next, we show that the contribution to the expectation of the remaining terms in (45) is  $o(n \log^2 n)$ . First, by a simple extension of Lemma 3.2 of RM, for  $0 < |\lambda| \leq \pi$ ,  $0 < d < 1$ ,

$$\left| \sum_{j=0}^{t-1} a_j(d) \log(j+\delta) e^{ij\lambda} \right| \leq K \min \left\{ t^d \log t, \frac{1}{|\lambda|}, \frac{\log |\lambda|}{|\lambda|^d} \right\}. \quad (47)$$

Then noting (47), for  $n$  large enough (such that  $n^{-1} < \varepsilon$ ), the second term in (45) is bounded by

$$K \sum_{t=1}^n \int_{-n^{-1}}^{n^{-1}} t \log t d\lambda \leq K n \log n. \quad (48)$$

Next, again by (47) and Lemma 3.2 of RM, the third term in (45) is bounded by

$$K\eta n \int_{n^{-1} \leq |\lambda| \leq \varepsilon} |\lambda|^{-1} \log |\lambda| d\lambda \leq K\eta n \log^2 n. \quad (49)$$

Finally, the fourth term in (45) is bounded by

$$K n \int_{\varepsilon}^{\pi} |\lambda|^{-2} d\lambda \leq K n \varepsilon^{-1}, \quad (50)$$

noting (47) and Lemma 3.2 of RM. Thus, given that  $\eta$  is arbitrary, the first term

in (45) is the leading one. Next, by a simple extension of Theorem 5.1 of RM

$$\text{Var} \left( \sum_{t=1}^n u_{1t} (-\gamma) \sum_{j=0}^{t-1} a_j (\delta) \log(j + \delta) u_{2,t-j} \right) = O(n^2 \log^2 n), \quad (51)$$

so

$$\frac{1}{n \log^2 n} \sum_{t=1}^n u_{1t} (-\gamma) \sum_{j=0}^{t-1} a_j (\delta) \log(j + \delta) u_{2,t-j} \rightarrow_p f_{12}(0) \sin(\delta\pi). \quad (52)$$

Next, by (43) the third term in (40) can be easily shown to be  $o_p(n \log^2 n)$ , where again by MR, the fourth term in (40) is  $o_p(n \log^2 n)$ , so noting (44), (52), we conclude that

$$\frac{1}{n \log^2 n} C_{u_1(-\gamma)x_n(\theta)} \rightarrow_p f_{12}(0) \sin(\delta\pi).$$

Next we discuss the  $\gamma = 0, \delta = 1$  case, for which

$$C_{u_1(-\gamma)x_n(\theta)} = \sum_{t=1}^n u_{1t} \sum_{j=0}^{t-1} u_{2,t-j} \sum_{k=j}^n \frac{1}{\delta + k} - \frac{1}{n} \sum_{t=1}^n u_{1t} (-\gamma) \sum_{t=1}^n x_{t,n}(\theta) \quad (53)$$

First, the expectation of the first term on the right side of (53) is

$$\begin{aligned} & n \log(n+1+\delta) \sum_{t=0}^{n-1} \gamma_{-t} - \log(n+1+\delta) \sum_{t=0}^{n-1} t \gamma_{-t} - n \sum_{t=0}^{n-1} \gamma_{-t} \log(t+\delta) \\ & + \sum_{t=0}^{n-1} t \gamma_{-t} \log(t+\delta) + \sum_{t=0}^{n-1} (n-t) \gamma_{-t} c_{t,n}. \end{aligned} \quad (54)$$

First Assumption 1 implies that  $|\gamma_t| = O(t^{-1-\epsilon})$ , so  $\sum_{t=1}^n t |\gamma_t| = O(n^{1-\epsilon})$ . Noting also (43), it is clear that the first term in (54) dominates. Then

$$\frac{1}{n \log n} E \left( \sum_{t=1}^n u_{1t} \sum_{j=0}^{t-1} u_{2,t-j} \sum_{k=j}^n \frac{1}{\delta + k} \right) \rightarrow \sum_{j=0}^{\infty} E(u_{10} u_{2,-j}),$$

as  $n \rightarrow \infty$ . Next, we show that

$$\text{Var} \left( \sum_{t=1}^n u_{1t} \sum_{j=0}^{t-1} u_{2,t-j} \sum_{k=j}^n \frac{1}{\delta + k} \right) = o(n^2 \log^2 n). \quad (55)$$

The left side of (55) is

$$\begin{aligned} & \sum_{t=1}^n \sum_{j=0}^{t-1} \sum_{k=j}^n \sum_{s=1}^n \sum_{l=0}^{s-1} \sum_{m=l}^n \frac{1}{(\delta+k)(\delta+m)} (E(u_{1t}u_{1s}) E(u_{2,t-j}u_{2,s-l}) \\ & + E(u_{1t}u_{2,s-l}) E(u_{1s}u_{2,t-j}) + \kappa), \end{aligned} \quad (56)$$

where  $\kappa$  is the fourth cumulant of  $u_{1t}, u_{1s}, u_{2,t-j}, u_{2,s-l}$ . We just analyze the contribution of the first term in brackets in (56), which is bounded by

$$\begin{aligned} & K \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left| \sum_{t=1}^n \sum_{j=0}^{t-1} \sum_{k=j}^n \frac{1}{\delta+k} e^{-it(\mu+\lambda)} e^{ij\lambda} \right|^2 d\mu d\lambda \leq K \sum_{t=1}^n \sum_{j=0}^{t-1} \left( \sum_{k=j}^n \frac{1}{\delta+k} \right)^2 \\ & \leq K \sum_{t=1}^n (n-t) \log^2 \frac{t}{n} \leq Kn^2 \int_0^1 (1-x) \log^2 x dx \leq Kn^2. \end{aligned} \quad (57)$$

By the Cauchy inequality the contribution of the other terms in brackets in (56) can be easily shown to be also bounded by  $Kn^2$ , to conclude the proof of (55). Thus, noting that by previous arguments the second term on the right of (53) is  $O_p(n)$ ,

$$\frac{1}{n \log n} C_{u_1(-\gamma)x_n(\theta)} \rightarrow_p \sum_{j=0}^{\infty} E(u_{10}, u_{2,-j}).$$

Next, if  $\gamma + \delta > 1$ ,  $\gamma > 1/2$ , by MR, Hualde (2012) and the continuous mapping theorem  $n^{-(\gamma+\delta)} C_{u_1(-\gamma)x_n(\theta)} \rightarrow_d B_{A(\delta)W(\gamma)}^{\xi\zeta}$ , where  $\zeta = (1, 0)^T$ , whereas if  $\delta > 1$ ,  $\gamma = 0$ ,  $n^{-\delta} C_{u_1(-\gamma)x_n(\theta)} \rightarrow_d D_{A(\delta)}$ , where for  $E(r; \cdot)$  being  $W(r; \cdot)$  or  $A(r; \cdot)$ ,

$$D_{E(d)} = \xi^T B(1) \int_0^1 \tilde{E}(r; d) dW^T(r) B^T(1) \zeta.$$

Collecting all these results, it is straightforward to derive the limiting distribution of  $\hat{\theta}$ , by routine application of Theorem 1 of MR, Theorem 1 of Hualde (2012), the continuous mapping theorem and Theorem 2.2 of Kurtz and Protter (1991) (see e.g. Proposition 3 of Robinson and Hualde, 2003). In order to derive the joint limiting distribution of  $\hat{\theta}, \hat{\nu}$ , in view of (10), we need to show (11) and (12).

Clearly

$$\frac{1}{\log n} \widehat{\nu}'(\bar{\theta}) = \frac{1}{\log n} \widehat{\nu}'(\theta) + \frac{1}{\log n} (\widehat{\nu}'(\bar{\theta}) - \widehat{\nu}'(\theta)). \quad (58)$$

First, by a similar analysis to that of (38), the second term in (58) is  $O_p(q_n^{-1}(\gamma, \delta) \log n) = o_p(1)$ . Next, by (37),  $\widehat{\nu}'(\theta) = (\nu - 2\widehat{\nu}(\theta)) C_{x(\theta)x'(\theta)}/C_{x(\theta)}$ . By a very similar derivation to that of Theorem ABCU of Hualde (2006),  $\log^{-1} n C_{x(\theta)x'(\theta)}/C_{x(\theta)} \rightarrow_p -1$ , so (11) holds as  $\widehat{\nu}(\theta) \rightarrow_p \nu$ . Then, noting that in (10) by the rates derived in RM,  $\widehat{\nu}(\theta) - \nu$  is in every circumstance of smaller order than  $\widehat{\nu}'(\bar{\theta})(\widehat{\theta} - \theta)$ , (12) is justified.

Finally, collecting all previous results, we specify  $\Xi(\gamma, \delta)$ . Define  $P = B_{W(\delta)}^\xi B_{A(\delta)}^\xi - \left( B_{A(\delta)W(\delta)}^{\xi\xi} \right)^2$ . Noting that by Theorems 4.1, 4.2, 4.3, 5.1 of RM, Theorem 1 of Hualde (2012) and the continuous mapping theorem, the first term in  $F_n(\theta)$  is of smaller order when  $\gamma + \delta \leq 1$ ,

$$\begin{aligned} \Xi(\gamma, \delta) &= P^{-1} B_{W(\delta)}^\xi \int_{-\pi}^{\pi} a(\gamma; \lambda) a(\delta; -\lambda) f_{12}(\lambda) d\lambda, \text{ if } \gamma + \delta < 1; \\ &= P^{-1} B_{W(\delta)}^\xi f_{12}(0) \sin(\delta\pi), \text{ if } \gamma > 0, \gamma + \delta = 1; \\ &= P^{-1} B_{W(\delta)}^\xi \sum_{j=0}^{\infty} E(u_{10}, u_{2,-j}), \text{ if } \gamma = 0, \delta = 1; \\ &= P^{-1} \left( B_{W(\delta)}^\xi B_{A(\delta)W(\gamma)}^{\xi\xi} - B_{A(\delta)W(\delta)}^{\xi\xi} B_{W(\delta)W(\gamma)}^{\xi\xi} \right), \text{ if } \gamma + \delta > 1, \gamma > 1/2; \\ &= P^{-1} \left( B_{W(\delta)}^\xi D_{A(\delta)} - B_{A(\delta)W(\delta)}^{\xi\xi} D_{W(\delta)} \right), \text{ if } \gamma + \delta > 1, \gamma = 0, \end{aligned} \quad (59)$$

to conclude the proof of the theorem.

## Appendix B. Technical lemmas

**Lemma 1 (Behaviour of  $C_{x(e)x(\theta)}$ ).** *Let Assumptions 1, 2 hold. Then, as  $n \rightarrow \infty$ ,*

$$\sup_{\mathcal{I}_1} \frac{1}{n^{\delta+\frac{1}{2}}} C_{x(e)x(\theta)} = o_p(1), \quad (60)$$

$$\sup_{\mathcal{I}_2} \frac{1}{n^{\delta+\frac{1}{2}}} C_{x(e)x(\theta)} = O_p(1), \quad (61)$$

$$\sup_{\mathcal{I}_3 \cup \mathcal{I}_4} \frac{1}{n^{\delta+c}} C_{x(e)x(\theta)} = O_p(1). \quad (62)$$

**Proof.** First we show (60). Noting that by results in RM,  $n^{-(\delta+1/2)} \sum_{t=1}^n x_t(\theta) =$



$O_p(1)$ , the result follows on showing

$$\sup_{\mathcal{I}_1} \frac{1}{n^{\delta+\frac{1}{2}}} \sum_{t=1}^n x_t(e) x_t(\theta) = o_p(1), \quad (63)$$

$$\sup_{\mathcal{I}_1} \frac{1}{n} \sum_{t=1}^n x_t(e) = o_p(1). \quad (64)$$

First, (63) holds on showing

$$\sup_{\mathcal{I}_1} \frac{1}{n^{\delta+\frac{1}{2}}} \left| \sum_{t=1}^n E(x_t(e) x_t(\theta)) \right| \leq K n^{-\eta}, \quad (65)$$

$$\sup_{\mathcal{I}_1} \frac{1}{n^{\delta+\frac{1}{2}}} \left| \sum_{t=1}^n (x_t(e) x_t(\theta) - E(x_t(e) x_t(\theta))) \right| \leq K n^{-\eta}, \quad (66)$$

where  $\eta$  was introduced when defining  $\mathcal{I}_1$ . Noting that

$$\begin{aligned} \frac{1}{n^{\delta+\frac{1}{2}}} \sum_{t=1}^n x_t(e) x_t(\theta) &= \frac{1}{n^{\delta+\frac{1}{2}}} \left( \sum_{j=0}^{n-1} a_j(c) a_j(\delta) \sum_{l=1}^{n-j} u_{2,l}^2 \right. \\ &\quad + \sum_{j=0}^{n-2} \sum_{k=j+1}^{n-1} a_j(c) a_k(\delta) \sum_{l=k-j+1}^{n-j} u_{2,l} u_{2,l-(k-j)} \\ &\quad \left. + \sum_{j=0}^{n-2} \sum_{k=j+1}^{n-1} a_j(\delta) a_k(c) \sum_{l=k-j+1}^{n-j} u_{2,l} u_{2,l-(k-j)} \right), \quad (67) \end{aligned}$$

the proof of (65) is almost immediate by simple application of Stirling's approximation, noting that  $\sup_{\mathcal{I}_1} j^{c-1} = j^{-1/2-\eta}$  and  $\sum_{k=1}^{\infty} |E(u_{2,t} u_{2,t-k})| < \infty$ . Next, (66) holds by almost identical steps to those in the proof of (2.12) of HR, just accounting for few minor differences.

Next we show (64). Clearly

$$\frac{1}{n} \sum_{t=1}^n x_t(e) = \frac{1}{n} \sum_{j=0}^{n-1} a_j(c) \sum_{l=1}^{n-j} u_{2,l}.$$

It can be easily shown that uniformly in  $j$ ,  $\sum_{l=1}^{n-j} u_{2,l} = O_p(n^{1/2})$ , so

$$E \sup_{\mathcal{I}_1} \frac{1}{n} \sum_{t=1}^n x_t(e) \leq K \frac{1}{n^{1/2}} \sup_{\mathcal{I}_1} \sum_{j=1}^n j^{c-1} \leq Kn^{-\eta},$$

to conclude the proof of (60).

The proof of (61) is omitted because it is almost identical to that of (60), just replacing  $\eta$  by 0, so the bounds of the corresponding terms above become  $K$  (instead of  $Kn^{-\eta}$ ).

Finally the proof of (62) is again almost identical to that of (60), (61), so it is again omitted. The only difference is the different normalization  $n^{\delta+c}$  (instead of  $n^{\delta+1/2}$ ) but the required result follows straightforwardly noting that  $\sup_{\mathcal{I}_3 \cup \mathcal{I}_4} \left(\frac{j}{n}\right)^c = \left(\frac{j}{n}\right)^{1/2}$ .

**Lemma 2. (Behaviour of  $C_{x(e)}$ ).** *Let Assumptions 1, 2 hold. Then, for arbitrarily large and small positive  $K, \epsilon$ , respectively, as  $n \rightarrow \infty$ ,*

$$\Pr \left( \inf_{\mathcal{I}_1} \frac{1}{n} C_{x(e)} > \epsilon \right) \rightarrow 1, \quad (68)$$

$$\Pr \left( \inf_{\mathcal{I}_2} \frac{1}{n} C_{x(e)} > K \right) \rightarrow 1, \quad (69)$$

$$\Pr \left( \inf_{\mathcal{I}_3} \frac{1}{n^{2c}} C_{x(e)} > K \right) \rightarrow 1. \quad (70)$$

**Proof.** First we show (68). Clearly  $n^{-1}C_{x(e)} = U(c) + T_n(c)$ , where  $U(c) = E \left( \sum_{j=0}^{\infty} a_j(c) \xi^T u_{t-j} \right)^2$  and

$$T_n(c) = \frac{1}{n} \sum_{t=1}^n x_t^2(e) - E \left( \sum_{j=0}^{\infty} a_j(c) \xi^T u_{t-j} \right)^2 - \left( \frac{1}{n} \sum_{t=1}^n x_t(e) \right)^2,$$

noting that for  $c \in \mathcal{I}_1$ ,  $\sum_{j=0}^{\infty} a_j(c) \xi^T u_{t-j}$  is a covariance stationary process. The proof of (68) follows on showing that for some  $\tau > 0$

$$\inf_{\mathcal{I}_1} U(c) > \tau, \quad (71)$$

and  $\sup_{\mathcal{I}_1} |T_n(c)| = o_p(1)$ . Clearly

$$U(c) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |1 - e^{i\lambda}|^{-2c} \xi^T B(e^{i\lambda}) \Sigma B^T(e^{-i\lambda}) \xi d\lambda.$$

By Assumption 1, there exists  $\phi > 0$  such that  $\inf_{\lambda} \xi^T B(e^{i\lambda}) \Sigma B^T(e^{-i\lambda}) \xi \geq \phi$ , so

$$U(c) \geq \frac{\phi}{2\pi} \int_{-\pi}^{\pi} |1 - e^{i\lambda}|^{-2c} d\lambda = \frac{\phi \Gamma(1-2c)}{2\pi \Gamma^2(1-c)}, \quad (72)$$

(see, e.g., Brockwell and Davis, 1991, p. 522), where the right side of (72) is positive for any  $c \in \mathcal{I}_1$ , so (71) holds because  $\mathcal{I}_1$  is compact.

Next, noting (64), (68) would hold on showing

$$\sup_{\mathcal{I}_1} \left| \frac{1}{n} \sum_{t=1}^n \left[ \left( \sum_{j=0}^{t-1} a_j(c) u_{2,t-j} \right)^2 - E \left( \sum_{j=0}^{t-1} a_j(c) u_{2,t-j} \right)^2 \right] \right| = o_p(1), \quad (73)$$

$$\sup_{\mathcal{I}_1} \left| \frac{1}{n} \sum_{t=1}^n \sum_{j=0}^{t-1} \sum_{k=t}^{\infty} a_j(c) a_k(c) \gamma_{j-k} \right| = o_p(1), \quad (74)$$

$$\sup_{\mathcal{I}_1} \left| \frac{1}{n} \sum_{t=1}^n \sum_{j=t}^{\infty} \sum_{k=t}^{\infty} a_j(c) a_k(c) \gamma_{j-k} \right| = o_p(1), \quad (75)$$

which follows by almost identical (but simpler) arguments to those in the proofs of (2.12), (2.13), (2.14) of HR.

Next we show (69). Clearly  $u_t = B(1) \varepsilon_t + v_{t-1} - v_t$ , where  $v_t = \sum_{j=0}^{\infty} B_j^* \varepsilon_{t-j}$ ,  $B_j^* = \sum_{k=j+1}^{\infty} B_k$ , and by Assumption 1,  $v_t$  is well defined in the mean square sense. Noting that  $a_{j+1}(c) - a_j(c) = a_{j+1}(c-1)$ ,

$$x_t(c) = \xi^T B(1) \sum_{j=0}^{t-1} a_j(c) \varepsilon_{t-j} - \sum_{j=0}^{t-1} a_j(c-1) \xi^T v_{t-j} + a_{t-1}(c) \xi^T v_0.$$

Then

$$\begin{aligned}
C_{x(e)} &\geq \sum_{t=1}^n \left( \sum_{j=0}^{t-1} a_j(c) \xi^T B(1) \varepsilon_{t-j} \right)^2 \\
&\quad - 2 \left| \xi^T B(1) \sum_{t=1}^n \sum_{j=0}^{t-1} a_j(c) \varepsilon_{t-j} \sum_{k=0}^{t-1} a_k(c-1) v_{t-k}^T \xi \right| \\
&\quad - 2 \left| \xi^T B(1) \sum_{t=1}^n \sum_{j=0}^{t-1} a_j(c) \varepsilon_{t-j} a_{t-1}(c) v_0^T \xi \right| \\
&\quad - 2 \left| \xi^T \sum_{t=1}^n \sum_{j=0}^{t-1} a_j(c-1) v_{t-j} a_{t-1}(c) v_0^T \xi \right| - n \left( \frac{1}{n} \sum_{t=1}^n x_t(e) \right)^2 \\
&= (b_1) - (b_2) - (b_3) - (b_4) - (b_5).
\end{aligned}$$

The proof of (69) follows on showing that  $\inf_{\mathcal{I}_2} n^{-1} (b_1)$  can be made arbitrarily large on setting  $\eta$  arbitrarily close to zero, and  $\sup_{\mathcal{I}_2} n^{-1} (b_i) = O_p(1)$ ,  $i = 2, 3, 4$ , given that, as in the proof of Lemma 1,  $\sup_{\mathcal{I}_2} n^{-1} (b_5) = O_p(1)$ . First, noting that  $da_j^2(c)/dc = 2(\psi(j+c) - \psi(c))a_j^2(c)$ , so  $a_j^2(c)$  is strictly increasing in  $\mathcal{I}_2$  (because  $\psi(\cdot)$  is strictly increasing), implying that

$$\begin{aligned}
\inf_{\mathcal{I}_2} \frac{1}{n} (b_1) &\geq \frac{\xi^T B(1)}{n} \sum_{t=1}^n \sum_{j=0}^{t-1} a_j^2(1/2 - \eta) \varepsilon_{t-j} \varepsilon_{t-j}^T B^T(1) \xi \\
&\quad - \frac{2}{n} \sup_{\mathcal{I}_2} \left| \xi^T B(1) \sum_{j=0}^{n-2} \sum_{k=j+1}^{n-1} a_j(c) a_k(c) \sum_{l=k-j+1}^{n-j} \varepsilon_l \varepsilon_{l-(k-j)}^T B^T(1) \xi \right|.
\end{aligned} \tag{76}$$

Then, by an almost identical treatment to that of (2.24) of HR, it can be shown that

$$\inf_{\mathcal{I}_2} \frac{1}{n} (b_1) \geq \xi^T B(1) \Sigma B^T(1) \xi \frac{\Gamma(2\eta)}{\Gamma^2(\frac{1}{2} + \eta)} - \sup_{\mathcal{I}_2} |d_n(c)|, \tag{77}$$

where  $\sup_{\mathcal{I}_2} |d_n(c)| = O_p(1)$ , so the result holds because the first term on the

right of (77) can be made arbitrarily large by setting  $\eta$  close enough to zero. Next

$$\begin{aligned}
(b_2) &\leq 2 \left| \xi^T B(1) \sum_{j=0}^{n-1} a_j(c) a_j(c-1) \sum_{l=1}^{n-j} \varepsilon_l v_l^T \xi \right| \\
&\quad + 2 \left| \sum_{j=1}^{n-1} a_j(c-1) \sum_{k=0}^{j-1} a_k(c) \sum_{l=1}^{n-j} \xi^T B(1) \varepsilon_{l+j-k} v_l^T \xi \right| \\
&\quad + 2 \left| \sum_{j=1}^{n-1} a_j(c) \sum_{k=0}^{j-1} a_k(c-1) \sum_{l=1}^{n-j} \xi^T B(1) \varepsilon_l v_{l+j-k}^T \xi \right| \\
&= (b_{21}) + (b_{22}) + (b_{23}).
\end{aligned}$$

It can be easily shown that, uniformly in  $j, n$ ,  $E \left| \sum_{l=1}^{n-j} \xi^T B(1) \varepsilon_l v_l^T \xi \right| = O(n)$ , so that

$$E \sup_{\mathcal{I}_2} \frac{1}{n} (b_{21}) \leq K \sum_{j=1}^n j^{-2} \leq K.$$

Next, it can be easily shown that, uniformly in  $j, k, n$ ,  $E \left| \sum_{l=1}^{n-j} \xi^T B(1) \varepsilon_{l+j-k} v_l^T \xi \right| = O(n^{1/2})$ , so

$$E \sup_{\mathcal{I}_2} \frac{1}{n} (b_{22}) \leq \frac{K}{n^{1/2}} \sum_{j=1}^n j^{-3/2} \sum_{k=1}^n k^{-1/2} \leq K.$$

Regarding  $(b_{23})$ , we show first that for any  $j, k$ ,

$$E \left| \sum_{l=1}^{n-j} \xi^T B(1) \varepsilon_l v_{l+j-k}^T \xi \right| \leq K \left( n \|B_{j-k}^*\| + n^{1/2} \right). \quad (78)$$

Now

$$\begin{aligned}
E \left( \sum_{l=1}^{n-j} \xi^T B(1) \varepsilon_l v_{l+j-k}^T \xi \right) &= E \left( \sum_{l=1}^{n-j} \sum_{r=0}^{\infty} \xi^T B(1) \varepsilon_l \varepsilon_{l+j-k-r}^T B_r^{*T} \xi \right) \\
&= (n-j) \xi^T B(1) \Sigma B_{j-k}^{*T} \xi \leq Kn \|B_{j-k}^*\|.
\end{aligned}$$

Next

$$\begin{aligned} \text{Var} \left( \sum_{l=1}^{n-j} \xi^T B(1) \varepsilon_l v_{l+j-k}^T \xi \right) &= \sum_{l=1}^{n-j} \sum_{m=1}^{n-j} \{ E(\xi^T B(1) \varepsilon_l \varepsilon_m^T B^T(1) \xi) E(\xi^T v_{m+j-k} v_{l+j-k}^T \xi) \\ &\quad + E(\xi^T B(1) \varepsilon_l v_{m+j-k}^T \xi) E(\xi^T B(1) \varepsilon_m v_{l+j-k}^T \xi) + \kappa \}, \end{aligned} \quad (79)$$

where  $\kappa$  is the fourth cumulant of  $\xi^T B(1) \varepsilon_l$ ,  $\xi^T B(1) \varepsilon_m$ ,  $\xi^T v_{l+j-k}$ ,  $\xi^T v_{m+j-k}$ . Clearly, the contribution of the first and third terms on the right of (79) is bounded by  $Kn$ . Regarding the contribution of the second term, this is equal to

$$\sum_{l=1}^{n-j} (E(\xi^T B(1) \varepsilon_l v_{l+j-k}^T \xi))^2 + 2 \sum_{l=2}^{n-j} \sum_{m=1}^{l-1} E(\xi^T B(1) \varepsilon_l v_{m+j-k}^T \xi) E(\xi^T B(1) \varepsilon_m v_{l+j-k}^T \xi). \quad (80)$$

The first term of (80) equals  $(n-j) (\xi^T B(1) \Sigma B_{j-k}^{*T} \xi)^2$ , whereas the second is

$$\begin{aligned} &2 \sum_{l=2}^{n-j} \sum_{m=1}^{l-1} \xi^T B(1) \Sigma B_{m-l+j-k}^{*T} \xi \xi^T B(1) \Sigma B_{l-m+j-k}^{*T} \xi 1(m-l+j-k \geq 0) \\ &= 2 \sum_{s=1}^{n-j-1} (n-j-s) \xi^T B(1) \Sigma B_{-s+j-k}^{*T} \xi \xi^T B(1) \Sigma B_{s+j-k}^{*T} \xi 1(-s+j-k \geq 0). \end{aligned}$$

Thus, noting that  $\|B_s^*\|$  is uniformly bounded, (80) is bounded by

$$Kn \left( 1 + \sum_{s=0}^{\infty} \|B_{s+j-k}^*\| \|B_{-s+j-k}^* 1(-s+j-k \geq 0)\| \right) \leq Kn \left( 1 + \sum_{s=0}^{\infty} \|B_s^*\|^2 \right) \leq Kn,$$

to conclude the proof of (78). Then

$$E \sup_{\mathcal{I}_2} \frac{1}{n} (b_{23}) \leq K \sum_{j=1}^n j^{-\frac{1}{2}} \sum_{k=1}^j k^{-\frac{3}{2}} \|B_{j-k}^*\| + \frac{K}{n^2} \sum_{j=1}^n j^{-\frac{1}{2}} \sum_{k=1}^j k^{-\frac{3}{2}}. \quad (81)$$

The second term on the right of (81) is clearly  $O(1)$ , whereas noting that by

Assumption 1 (ii),  $B_j^* = O(j^{-\varrho})$ , where  $\varrho > 1/2$ , the first is bounded by

$$\begin{aligned}
& K \sum_{j=1}^n j^{-\frac{1}{2}} \sum_{k=1}^{\lfloor j/2 \rfloor} k^{-\frac{3}{2}} (j-k)^{-\varrho} + K \sum_{j=1}^n j^{-\frac{1}{2}} \sum_{k=\lfloor j/2 \rfloor + 1}^{j-1} k^{-\frac{3}{2}} (j-k)^{-\varrho} \\
& \leq K \sum_{j=1}^n j^{-\frac{1}{2}-\varrho} \sum_{k=1}^{\infty} k^{-\frac{3}{2}} + K \sum_{j=1}^n j^{-2} \sum_{k=\lfloor j/2 \rfloor + 1}^{j-1} (j-k)^{-\varrho} \\
& \leq K + K \sum_{j=1}^n j^{-1-\varrho} \leq K,
\end{aligned}$$

to conclude that  $E \sup_{\mathcal{I}_2} n^{-1} (b_2) = O(1)$ . Next, it is easy to show that

$$E \sup_{\mathcal{I}_2} \frac{1}{n} (b_3) \leq \frac{K}{n} \sum_{t=1}^n \sum_{j=1}^t j^{-\frac{1}{2}} t^{-\frac{1}{2}} = O(1).$$

Finally,

$$E \sup_{\mathcal{I}_2} \frac{1}{n} (b_4) \leq \frac{K}{n} \sum_{t=1}^n \sum_{j=1}^t j^{-\frac{3}{2}} t^{-\frac{1}{2}} = O\left(n^{-\frac{1}{2}}\right),$$

to conclude the proof of (69).

Next the proof of (70) follows on showing that  $\inf_{\mathcal{I}_3} n^{-2c} (b_1)$  can be made arbitrarily large on setting  $\eta$  arbitrarily close to zero, and  $\sup_{\mathcal{I}_3} n^{-2c} (b_i) = O_p(1)$ ,  $i = 2, 3, 4$ . First, by almost identical arguments to those in the proof of (2.30) of HR, it can be shown that for an arbitrarily large  $K$ ,  $\Pr\left(\inf_{\mathcal{I}_3} \frac{1}{n^{2c}} (b_1) > K\right) \rightarrow 1$  as  $n \rightarrow \infty$ , by setting  $\eta$  arbitrarily close to zero. Next, the proof of  $\sup_{\mathcal{I}_3} n^{-2c} (b_i) = O_p(1)$ , is almost identical to the corresponding one for  $\mathcal{I}_2$ , to complete the proof of the lemma.

**Lemma 3. (Approximation of  $P_n(c)$ ).** *Let Assumptions 1, 2 hold. Then, as  $n \rightarrow \infty$ ,*

$$P_n(c) = X_n(c) + r_n(c) = X_{1n}(c) + X_{2n}(c) + X_{3n}(c) + r_n(c), \quad (82)$$

where

$$\begin{aligned} X_{1n}(c) &= \psi^2(c) (C_{f(\delta)} C_{f(c)} - C_{f(\delta)f(c)}^2), \\ X_{2n}(c) &= 2\psi(c) (C_{f(\delta)} C_{f(c)f_n(c)} - C_{f(\delta)f(c)} C_{f(\delta)f_n(c)}), \\ X_{3n}(c) &= C_{f(\delta)} C_{f_n(c)} - C_{f(\delta)f_n(c)}^2, \end{aligned}$$

where

$$f_{t,n}(c) = \frac{1}{n^c} \sum_{j=0}^{t-1} a_j(c) u_{2,t-j} \sum_{k=j}^n \frac{1}{k+c},$$

and

$$\sup_{\mathcal{I}_\varepsilon} |r_n(c)| = o_p(1). \quad (83)$$

**Proof.** Noting  $f'_t(c) = g_t(c) - r(c) f_t(c)$  and the recurrence formula for the digamma function, by summation by parts

$$f'_t(c) = b_{t-1}(c) f_t(c) - h_t(c), \quad (84)$$

where

$$b_{t-1}(c) = \psi(t-1+c) - r(c), \quad h_t(c) = \frac{1}{n^c} \sum_{j=0}^{t-2} \frac{1}{j+c} \sum_{i=0}^j a_i(c) u_{2,t-i}.$$

First, we show that

$$\sum_{t=1}^n (f'_t(c))^2 = b_{n-1}^2(c) \sum_{t=1}^n f_t^2(c) - 2b_{n-1}(c) \sum_{t=1}^n f_t(c) f_{t,n}(c) + \sum_{t=1}^n f_{t,n}^2(c) + o_p(1), \quad (85)$$

uniformly in  $c \in \mathcal{I}_\varepsilon$ . By (84), using the summation by parts formula repeatedly,

$$\begin{aligned} \sum_{t=1}^n (f'_t(c))^2 &= b_{n-1}^2(c) \sum_{t=1}^n f_t^2(c) - (b_{n-1}(c) + b_{n-2}(c)) \sum_{t=1}^{n-1} \frac{1}{t-1+c} \sum_{j=1}^t f_j^2(c) \\ &\quad + \sum_{t=1}^{n-2} \left( \frac{1}{t+c} + \frac{1}{t-1+c} \right) \sum_{s=1}^t \frac{1}{s-1+c} \sum_{j=1}^s f_j^2(c) + \sum_{t=1}^n h_t^2(c) \\ &\quad - 2b_{n-1}(c) \sum_{t=1}^n f_t(c) h_t(c) + 2 \sum_{t=1}^{n-1} \frac{1}{t-1+c} \sum_{j=1}^t f_j(c) h_j(c). \quad (86) \end{aligned}$$



Noting that

$$\begin{aligned}
\sum_{t=1}^{n-1} \frac{1}{t-1+c} \sum_{j=1}^t f_j(c) h_j(c) &= \sum_{t=1}^{n-1} f_t(c) h_t(c) \sum_{k=t-1}^{n-2} \frac{1}{k+c} \\
&= \sum_{t=1}^n f_t(c) h_t(c) \sum_{k=t-1}^n \frac{1}{k+c} + o_p(1), \\
\sum_{t=1}^{n-2} \left( \frac{1}{t+c} + \frac{1}{t-1+c} \right) \sum_{s=1}^t \frac{1}{s-1+c} \sum_{j=1}^s f_j^2(c) &= \sum_{t=1}^{n-2} f_t^2(c) \sum_{k=t-1}^{n-3} \frac{1}{k+c} \sum_{j=t-1}^{n-2} \frac{1}{j+c} \\
&= \sum_{t=1}^n f_t^2(c) \left( \sum_{k=t-1}^n \frac{1}{k+c} \right)^2 + o_p(1),
\end{aligned}$$

uniformly in  $c \in \mathcal{I}_\varepsilon$ , and  $\sup_{c \in \mathcal{I}_\varepsilon} |b_{n-1}(c) - b_{n-2}(c)| \leq Kn^{-1}$ , the right hand side of (86) is

$$\begin{aligned}
&b_{n-1}^2(c) \sum_{t=1}^n f_t^2(c) - 2b_{n-1}(c) \left( \sum_{t=1}^{n-1} \frac{1}{t-1+c} \sum_{j=1}^t f_j^2(c) + \sum_{t=1}^n f_t(c) h_t(c) \right) \\
&+ \sum_{t=1}^n \left( h_t(c) + f_t(c) \sum_{k=t-1}^n \frac{1}{k+c} \right)^2 + o_p(1),
\end{aligned}$$

uniformly in  $c \in \mathcal{I}_\varepsilon$ . Now, noting that for any sequence  $z_t$

$$\sum_{t=1}^{n-1} \frac{1}{t-1+c} \sum_{j=1}^t z_j f_j(c) + \sum_{t=1}^n z_t h_t(c) = \sum_{t=1}^n z_t f_{t,n}(c) - \left( \frac{1}{n+c} + \frac{1}{n-1+c} \right) \sum_{t=1}^n z_t f_t(c),$$

(see (37) in Hualde, 2012), and also that

$$h_t(c) + f_t(c) \sum_{k=t-1}^n \frac{1}{k+c} = f_{t,n}(c),$$

(85) holds. Next, by very similar arguments it is simple to show that

$$\left( \sum_{t=1}^n f_t'(c) \right)^2 = \left( b_{n-1}(c) \sum_{t=1}^n f_t(c) - \sum_{t=1}^n f_{t,n}(c) \right)^2 + o_p(1),$$

uniformly in  $c \in \mathcal{I}_\varepsilon$ , so that

$$C_{f'(c)} = b_{n-1}^2(c) C_{f(c)} - 2b_{n-1}(c) C_{f(c)f_n(c)} + C_{f_n(c)} + o_p(1),$$

uniformly in  $c \in \mathcal{I}_\varepsilon$ . By almost identical arguments, it can be shown that, uniformly in  $c \in \mathcal{I}_\varepsilon$ ,

$$C_{f(\delta)f'(c)} = b_{n-1}(c) C_{f(\delta)f(c)} - C_{f(\delta)f_n(c)} + o_p(1),$$

so that

$$\begin{aligned} P_n(c) &= b_{n-1}^2(c) (C_{f(\delta)} C_{f(c)} - C_{f(\delta)f(c)}^2) - 2b_{n-1}(c) (C_{f(\delta)} C_{f(c)f_n(c)} - C_{f(\delta)f(c)} C_{f(\delta)f_n(c)}) \\ &\quad + C_{f(\delta)} C_{f_n(c)} - C_{f(\delta)f_n(c)}^2 + o_p(1), \end{aligned}$$

uniformly in  $c \in \mathcal{I}_\varepsilon$ . Then, noting (31), (32), the lemma is established on showing that

$$\sup_{\mathcal{I}_\varepsilon} |\psi(n-1+c) - \log n| = o(1). \quad (87)$$

Clearly

$$\sup_{\mathcal{I}_\varepsilon} |\psi(n-1+c) - \log n| \leq \sup_{\mathcal{I}_\varepsilon} |\psi(n-1+c) - \psi(n)| + |\psi(n) - \log n|.$$

Now,  $\psi(n) - \log n = o(1)$ , whereas (see, e.g., Gradshteyn and Ryzhik, 2000, p. 893)

$$\begin{aligned} |\psi(n-1+c) - \psi(n)| &= \left| \sum_{k=0}^{\infty} \left( \frac{1}{n+k} - \frac{1}{n-1+c+k} \right) \right| \\ &\leq |c-1| \sum_{k=0}^{\infty} \frac{1}{(n-1+c+k)(n+k)}, \end{aligned}$$

so

$$\sup_{\mathcal{I}_\varepsilon} |\psi(n-1+c) - \psi(n)| \leq K \sum_{k=1}^{\infty} \frac{1}{k(n+k)} = K \frac{1}{n} (\psi(n+1) - \psi(1)) = o(1), \quad (88)$$

the equality in (88) following again by Gradshteyn and Ryzhik (2000, pp. 893-94), to conclude the proof of (87).

**Lemma 4.** *Let Assumptions 1, 2 hold. Then, as  $n \rightarrow \infty$  uniformly in  $t = 1, \dots, n$ ,*

$$\sup_{c \in \mathcal{I}_{n^\tau}} \left| x_t^{(i)}(e) - x_t^{(i)}(\theta) \right| = O_p(n^{\delta-1/2-\tau} \log^{i+1} n), \quad i = 0, 1, 2, \quad (89)$$

where  $\mathcal{I}_{n^\tau} = [\delta - \frac{\varepsilon}{n^\tau}, \delta + \frac{\varepsilon}{n^\tau}]$ ,  $x_t^{(0)}(e) = x_t(e)$ ,  $x_t^{(1)}(e) = x_t'(e)$ ,  $x_t^{(2)}(e) = x_t''(e)$ .

**Proof.** First, we consider the result for  $i = 0$ . By the mean value theorem, for  $t > 1$  (note that for  $t = 1$  the difference below is zero)

$$x_t(e) - x_t(\theta) = (c - \delta) \sum_{j=1}^{t-1} a'_j(\bar{c}) u_{2,t-j}, \quad (90)$$

where  $|\bar{c} - \delta| \leq |c - \delta|$ . Then, by summation by parts, the right side of (90) is

$$(c - \delta) a'_{t-1}(\bar{c}) \sum_{j=1}^{t-1} u_{2,t-j} - (c - \delta) \sum_{j=1}^{t-2} (a'_{j+1}(\bar{c}) - a'_j(\bar{c})) \sum_{i=1}^j u_{2,t-i}, \quad (91)$$

where for  $t = 2$  the second term in (91) is identically zero. Noting (26), the first term in (91) is bounded in absolute value by

$$|\psi(t-1+\bar{c}) - \psi(\bar{c})| |c - \delta| |a_{t-1}(\bar{c})| \left| \sum_{j=1}^{t-1} u_{2,t-j} \right| \leq K |c - \delta| t^{\bar{c}-1} \log t \left| \sum_{j=1}^{t-1} u_{2,t-j} \right|,$$

by Lemma D.1 of Robinson and Hualde (2003). Then, the supremum over  $\mathcal{I}_{n^\tau}$  of the absolute value of the first term in (91) is bounded by

$$K \log t \sup_{\mathcal{I}_{n^\tau}} |c - \delta| t^{\delta-1+\frac{\varepsilon}{n^\tau}} \left| \sum_{j=1}^{t-1} u_{2,t-j} \right|.$$

Noting that uniformly in  $t$ ,  $t^{\frac{\varepsilon}{n^\tau}} \leq n^{\frac{\varepsilon}{n^\tau}} \rightarrow 1$  as  $n \rightarrow \infty$ , and also that under Assumption 1,  $\sum_{j=1}^{t-1} u_{2,t-j} = O_p(t^{1/2})$ , the first term of (91) is, uniformly in  $t$ ,  $O_p(n^{\delta-1/2-\tau} \log n)$ .

For  $t > 2$ , the second term of (91) is bounded in absolute value by

$$\begin{aligned} & |c - \delta| \sum_{j=1}^{t-2} |\psi(j+1+\bar{c}) - \psi(j+\bar{c})| |a_{j+1}(\bar{c})| \left| \sum_{i=1}^j u_{2,t-i} \right| \\ & + |c - \delta| \sum_{j=1}^{t-2} |\psi(j+\bar{c}) - \psi(\bar{c})| |a_{j+1}(\bar{c}) - a_j(\bar{c})| \left| \sum_{i=1}^j u_{2,t-i} \right|. \end{aligned} \quad (92)$$

Noting (D.2), (D.7), (D.9) in Robinson and Hualde (2003),  $\sum_{i=1}^j u_{2,t-i} = O_p(j^{1/2})$ , uniformly in  $t$ , by the mean value theorem the supremum over  $\mathcal{I}_{n^\tau}$  of (92) is bounded by

$$K n^{-\tau} \sum_{j=1}^{t-2} j^{\delta + \frac{\varepsilon}{n^\tau} - 2} \log j \left| \sum_{i=1}^j u_{2,t-i} \right| = O_p(n^{\delta-1/2-\tau} \log n),$$

uniformly in  $t$ , by a similar analysis to that of the first term of (91), (89) being justified for  $i = 0$ .

The proof of (89) for  $i = 1$  is very similar to that for  $i = 0$ . For  $t > 2$  (otherwise the difference below is zero)

$$x'_t(e) - x'_t(\theta) = (c - \delta) \sum_{j=2}^{t-1} a''_j(\bar{c}) u_{2,t-j},$$

with the slight abuse of notation that the intermediate point  $\bar{c}$  could differ from that in the proof for  $i = 0$ . As in (D.5) of Robinson and Hualde (2003), it is immediate to show that

$$a''_j(\bar{c}) = (\psi'(j+\bar{c}) - \psi'(\bar{c}) + (\psi(j+\bar{c}) - \psi(\bar{c}))^2) a_j(\bar{c}), \quad (93)$$

so the proof follows as in the  $i = 0$  case noting that the term  $\psi^2(j+\bar{c}) a_j(\bar{c})$  dominates in (93), and  $\psi^2(j+\bar{c})$  contributes a further  $\log n$  factor.

Similarly, for  $t > 3$  (otherwise the difference below is zero)

$$x''_t(e) - x''_t(\theta) = (c - \delta) \sum_{j=3}^{t-1} a'''_j(\bar{c}) u_{2,t-j},$$

where the dominant term in  $a'''_j(\bar{c})$  is  $\psi^3(j+\bar{c}) a_j(\bar{c})$ , so (89) follows immediately

for  $i = 2$ .

The proofs of the following two lemmas are omitted as they are straightforward extensions of Propositions 1 and 2 of Hualde (2012), respectively.

**Lemma 5. (Approximation of  $F_n(\theta)$ ).** *Let Assumptions 1, 2 hold. Then, as  $n \rightarrow \infty$ ,*

$$F_n(\theta) = 2\nu \left( C_{u_1(-\gamma)x(\theta)} C_{x(\theta)x_n(\theta)} - C_{u_1(-\gamma)x_n(\theta)} C_{x(\theta)} \right) + o_p \left( n^{2\delta} g_n \right). \quad (94)$$

**Lemma 6. (Approximation of  $H_n(\theta)$ ).** *Let Assumptions 1, 2 hold. Then, as  $n \rightarrow \infty$ ,*

$$H_n(\theta) = 2\nu^2 \left( C_{x(\theta)} C_{x_n(\theta)} - C_{x(\theta)x_n(\theta)}^2 \right) + o_p \left( n^{4\delta} \right).$$

## References

- Baillie, R.T., Bollerslev, T., 2000. The forward premium anomaly is not as bad as you think. *Journal of International Money and Finance* 19, 471–488.
- Banerjee, A., Mizen, P., 2006. A re-interpretation of the linear quadratic model when inventories and sales are polynomially cointegrated. *Journal of Applied Econometrics* 21, 1249-1264.
- Bekaert, G., 1996. The time variation of risk and return in foreign exchange markets: a general equilibrium approach. *Review of Financial Studies* 9, 427-470.
- Bekaert, G., Hodrick, R.J., Marshall, D.A., 1997. The implications of first order risk aversion for asset market risk premiums. *Journal of Monetary Economics* 40, 3-39.
- Brockwell, P.J., Davis, R.A., 1991. *Time series: theory and methods*. Springer-Verlag.
- 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.

- Dalhaus, R., 1989. Efficient parameter estimation for self-similar processes. *Annals of Statistics* 14, 517-532.
- Dickey, D.A., Fuller, W.A., 1979. Distribution of estimators of autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427-431.
- Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction model. Representation, estimation and testing. *Econometrica* 55, 251-276.
- Engsted, T., Haldrup, N. 1999. Multicointegration in stock-flow models. *Oxford Bulletin of Economics and Statistics* 61, 237–254.
- Fox, R., Taqqu, M.S., 1986. Large-sample properties of parameter estimates for strongly dependent processes. *Annals of Statistics* 14, 517-532.
- Franchi, M., 2010. A representation theory for polynomial cofractionality in vector autoregressive models. *Econometric Theory* 26, 1201-1217.
- Gil-Alana, L.A., Hualde, J., 2009. Fractional integration and cointegration: an overview and an empirical application. *Palgrave Handbook of Econometrics*, Vol. 2, Applied Econometrics (K. Patterson and T.C. Mills, eds.). Palgrave, MacMillan, pp. 434-469.
- Gradshteyn, I.S., Ryzhik, I.M., 2000. *Table of integrals, series, and products*. Academic Press, Boston.
- Granger, C.W.J., 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14, 227-238.
- Granger, C.W.J., Joyeux, R., 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15-29.
- Granger, C.W.J., Lee, T.-H., 1989. Investigation of production sales and inventory relationships, using multicointegration and non-symmetric error correction models. *Journal of Applied Econometrics* 4, Suppl. 145-159.

- Granger, C.W.J., Lee, T.-H., 1990. Multicointegration. In G. F. Rhodes, Jr. and T. B. Fomby (eds.): *Advances in Econometrics: Cointegration, Spurious Regressions and unit Roots*, JAI Press, Greenwich, CT, 8, 71-84.
- Hamilton, J.D., 2002. On the interpretation of cointegration in the linear-quadratic inventory model. *Journal of Economic Dynamics & Control* 26, 2037-2049.
- Holt, C.C., Modigliani, F., Muth, J.F., Simon, H.A., 1960. *Planning production, inventories and work force*. Prentice-Hall: Englewood Cliffs, NJ.
- 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., 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 estimation of fractional cointegration. *Journal of Econometrics* 157, 492-511.
- Hualde, J., Robinson, P.M., 2011. Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics* 39, 3152-3181.
- Jeganathan, P., 1999. On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* 15, 583-621.
- Johansen, S., 1995. A statistical analysis of cointegration for  $I(2)$  variables. *Econometric Theory* 11, 25-59.
- Johansen, S., 2008. A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* 24, 651-676.
- Johansen, S., Nielsen, M.Ø., 2012. Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* 80, 2667-2732.
- Kurtz, T.G., Protter, P., 1991. Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035-1070.

- Leachman, L.L., 1996. New evidence on the Ricardian equivalence theorem: a multicointegration approach. *Applied Economics* 28, 695-704.
- Leachman, L.L., Bester, A., Rosas, G., Lange, P., 2005. Multicointegration and sustainability of fiscal practices. *Economic Inquiry* 43, 454-466.
- Leachman, L.L., Francis, B.B., 2000. Multicointegration analysis of the sustainability of foreign debt. *Journal of Macroeconomics* 22, 207-227.
- Leachman, L.L., Francis, B.B., 2002. Twin deficits: apparition or reality? *Applied Economics* 34, 1121-1132.
- Lee, T.-H., 1992. Stock-flow relationships in US housing construction. *Oxford Bulletin of Economics and Statistics* 54, 419-430.
- Marinucci, D., Robinson, P.M., 2000. Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* 86, 103-120.
- Marmol, F., Velasco, C., 2004. Consistent testing of cointegrating relationships. *Econometrica* 72, 1809-1844.
- Maynard, A., Phillips, P.C.B., 2001. Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics* 16 671-708.
- Maynard, A., Smallwood, A., Wohar, M.E., 2013. Long memory regressors and predictive testing: a two-stage rebalancing approach. *Econometric Reviews* 32, 318-360.
- Nielsen, M.Ø., Frederiksen, P., 2011. Fully modified narrow-band least squares estimation of weak fractional cointegration. *Econometrics Journal* 14, 77-120.
- Park, J.Y., Phillips, P.C.B., 1989. Statistical inference in regressions with integrated processes: part 2. *Econometric Theory* 5, 95-131.
- Phillips, P.C.B., 1991. Optimal inference in cointegrated systems. *Econometrica* 59, 283-306.



- Ramey, V.A., West, K.D., 1999. Inventories. Handbook of Macroeconomics, in: J. B. Taylor & M. Woodford (ed.), Handbook of Macroeconomics, edition 1, volume 1, chapter 13, pp.863-923 Elsevier.
- Robinson, P.M., 1978. Statistical inference for a random coefficient autoregressive model. Scandinavian Journal of Statistics 5, 163–168.
- Robinson, P.M., 1995. Gaussian semiparametric estimation of long-range dependence. Annals of Statistics 23, 1630-1661.
- Robinson, P.M., 2008. Diagnostic testing for cointegration. Journal of Econometrics 143, 206-225.
- Robinson, P.M., Gerolimetto, M., 2006. Instrumental variables estimation of stationary and non-stationary cointegrating regressions. Econometrics Journal 9, 291–306.
- 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. Annals of Statistics 29, 947-986.
- Robinson, P. M., Yajima, Y., 2002. Determination of cointegrating rank in fractional systems. Journal of Econometrics 106, 217-241.
- Siliverstovs, B., 2006. Multicointegration in US consumption data. Applied Economics 38, 819-833.
- Torous, W., Valkanov, R., Yan, S., 2004. On predicting stock returns with nearly integrated explanatory variables. Journal of Business 77, 937-966.

Table 1. Monte Carlo bias of  $\hat{\theta}$  for  $\nu = 1$ 

$\gamma$	$\delta \backslash \rho$	$n = 64$			$n = 128$			$n = 256$		
		.5	0	-.5	.5	0	-.5	.5	0	-.5
0	.6	.127	.003	-.149	.109	.001	-.120	.094	.000	-.098
0	1.2	.041	.000	-.036	.018	.000	-.017	.008	.000	-.008
0	2	-.001	.000	.001	.000	.000	.000	.000	.000	.000
.4	.6	.058	.003	-.132	.054	.001	-.106	.051	.001	-.093
.4	1.2	.062	-.002	-.064	.037	-.001	-.037	.022	-.001	-.022
.8	2	.014	-.002	-.016	.006	-.001	-.007	.003	.000	-.003

Notes: Monte Carlo bias over 5,000 replications of  $\hat{\theta}$  where in (2), (3)  $\mu = 0$ ,  $\nu = 1$ ,  $\theta = 0$ ,  $u_t$  is a Gaussian white noise with  $Var(u_{1t}) = Var(u_{2t}) = 1$  and correlation  $\rho$

Table 2. Monte Carlo SD of  $\hat{\theta}$  for  $\nu = 1$ 

$\gamma$	$\delta \backslash \rho$	$n = 64$			$n = 128$			$n = 256$		
		.5	0	-.5	.5	0	-.5	.5	0	-.5
0	.6	.050	.059	.063	.033	.033	.038	.023	.019	.025
0	1.2	.031	.023	.026	.014	.010	.012	.006	.004	.005
0	2	.005	.005	.005	.001	.001	.001	.000	.000	.000
.4	.6	.070	.122	.222	.052	.082	.115	.039	.060	.076
.4	1.2	.044	.048	.046	.025	.026	.025	.015	.014	.014
.8	2	.026	.030	.026	.011	.013	.011	.005	.005	.005

Notes: Monte Carlo SD over 5,000 replications of  $\hat{\theta}$  where in (2), (3)  $\mu = 0$ ,  $\nu = 1$ ,  $\theta = 0$ ,  $u_t$  is a Gaussian white noise with  $Var(u_{1t}) = Var(u_{2t}) = 1$  and correlation  $\rho$

Table 3. Monte Carlo bias of  $\hat{\theta}$  for  $\nu = 10$ 

$\gamma$	$\delta \backslash \rho$	$n = 64$			$n = 128$			$n = 256$		
		.5	0	-.5	.5	0	-.5	.5	0	-.5
0	.6	.014	.000	-.014	.012	.000	-.012	.010	.000	-.010
0	1.2	.004	.000	-.004	.002	.000	-.002	.001	.000	-.001
0	2	.000	.000	.000	.000	.000	.000	.000	.000	.000
.4	.6	.008	.000	-.008	.007	.000	-.007	.006	.000	-.007
.4	1.2	.006	.000	-.006	.004	.000	-.004	.002	.000	-.002
.8	2	.001	.000	-.002	.001	.000	-.001	.000	.000	.000

Notes: Monte Carlo bias over 5,000 replications of  $\hat{\theta}$  where in (2), (3)  $\mu = 0$ ,  $\nu = 10$ ,  $\theta = 0$ ,  $u_t$  is a Gaussian white noise with  $Var(u_{1t}) = Var(u_{2t}) = 1$  and correlation  $\rho$

Table 4. Monte Carlo SD of  $\hat{\theta}$  for  $\nu = 10$ 

$\gamma$	$\delta \backslash \rho$	$n = 64$			$n = 128$			$n = 256$		
		.5	0	-.5	.5	0	-.5	.5	0	-.5
0	.6	.005	.006	.006	.004	.003	.004	.002	.002	.002
0	1.2	.003	.002	.003	.001	.001	.001	.001	.000	.001
0	2	.001	.001	.001	.000	.000	.000	.000	.000	.000
.4	.6	.009	.011	.010	.007	.008	.007	.005	.006	.005
.4	1.2	.005	.005	.004	.003	.003	.003	.001	.001	.001
.8	2	.003	.003	.003	.001	.001	.001	.001	.001	.000

Notes: Monte Carlo SD over 5,000 replications of  $\hat{\theta}$  where in (2), (3)  $\mu = 0$ ,  $\nu = 10$ ,  $\theta = 0$ ,  $u_t$  is a Gaussian white noise with  $Var(u_{1t}) = Var(u_{2t}) = 1$  and correlation  $\rho$

Table 5. Monte Carlo bias of  $\hat{\nu}$  for  $\nu = 1$ 

$\gamma$	$\delta \backslash \rho$	$n = 64$			$n = 128$			$n = 256$		
		.5	0	-.5	.5	0	-.5	.5	0	-.5
0	.6	.455	.005	-.371	.436	.001	-.348	.418	.000	-.328
0	1.2	.120	.000	-.094	.064	.000	-.054	.033	.000	-.030
0	2	-.006	.000	.005	-.002	.000	.002	-.001	.000	.001
.4	.6	.496	.003	-.475	.487	-.003	-.465	.480	-.005	-.457
.4	1.2	.234	-.001	-.187	.161	-.001	-.134	.107	-.002	-.095
.8	2	.049	-.001	-.047	.024	-.001	-.024	.012	-.001	-.012

Notes: Monte Carlo bias over 5,000 replications of  $\hat{\nu}$  where in (2), (3)  $\mu = 0$ ,  $\nu = 1$ ,  $\theta = 0$ ,  $u_t$  is a Gaussian white noise with  $Var(u_{1t}) = Var(u_{2t}) = 1$  and correlation  $\rho$

Table 6. Monte Carlo SD of  $\hat{\nu}$  for  $\nu = 1$ 

$\gamma$	$\delta \backslash \rho$	$n = 64$			$n = 128$			$n = 256$		
		.5	0	-.5	.5	0	-.5	.5	0	-.5
0	.6	.111	.118	.095	.080	.081	.066	.059	.054	.047
0	1.2	.090	.069	.064	.047	.036	.036	.024	.018	.020
0	2	.016	.016	.016	.005	.005	.005	.002	.001	.002
.4	.6	.142	.170	.154	.115	.135	.118	.092	.110	.094
.4	1.2	.139	.130	.107	.096	.088	.074	.064	.059	.052
.8	2	.084	.090	.074	.043	.047	.039	.022	.023	.021

Notes: Monte Carlo SD over 5,000 replications of  $\hat{\nu}$  where in (2), (3)  $\mu = 0$ ,  $\nu = 1$ ,  $\theta = 0$ ,  $u_t$  is a Gaussian white noise with  $Var(u_{1t}) = Var(u_{2t}) = 1$  and correlation  $\rho$

Table 7. Monte Carlo bias of  $\hat{\nu}$ ,  $\tilde{\nu}_\theta$  for  $\nu = 1$ ,  $\gamma = 0$ ,  $\delta = 1$ ,  $\theta = 0, .1, -.1$ 

$\tau$	$\rho$	$n = 64$			$n = 128$			$n = 256$		
		.25	0	-.75	.25	0	-.75	.25	0	-.75
.5	$\hat{\nu}$	.163	.003	-.316	.112	.000	-.237	.074	-.001	-.172
	$\tilde{\nu}_0$	.029	-.001	-.083	.014	.000	-.043	.007	.000	-.022
	$\tilde{\nu}_{.1}$	-.226	-.241	-.280	-.286	-.291	-.306	-.337	-.339	-.345
	$\tilde{\nu}_{-.1}$	.305	.253	.101	.369	.337	.244	.448	.429	.376
1	$\hat{\nu}$	.111	.001	-.244	.077	.000	-.179	.051	-.001	-.128
	$\tilde{\nu}_0$	.021	.000	-.059	.010	.000	-.030	.005	.000	-.015
	$\tilde{\nu}_{.1}$	-.230	-.241	-.269	-.287	-.291	-.302	-.337	-.339	-.343
	$\tilde{\nu}_{-.1}$	.290	.253	.147	.359	.337	.272	.442	.429	.393
2	$\hat{\nu}$	.077	.001	-.183	.053	.000	-.133	.036	-.001	-.094
	$\tilde{\nu}_0$	.015	.000	-.042	.007	.000	-.021	.004	.000	-.011
	$\tilde{\nu}_{.1}$	-.233	-.240	-.262	-.289	-.290	-.299	-.338	-.339	-.342
	$\tilde{\nu}_{-.1}$	.278	.253	.179	.353	.337	.291	.438	.430	.404

Notes: Monte Carlo bias over 5,000 replications of  $\hat{\nu}$ ,  $\tilde{\nu}_\theta$ , where in (2), (3)  $\mu = 0$ ,  $u_t$  is a Gaussian white noise with  $Var(u_{1t}) = 1$ ,  $Var(u_{2t}) = \tau$  and correlation  $\rho$

Table 8. Monte Carlo SD of  $\hat{\nu}$ ,  $\tilde{\nu}_\theta$  for  $\nu = 1$ ,  $\gamma = 0$ ,  $\delta = 1$ ,  $\theta = 0, .1, -.1$ 

$\tau$	$\rho$	$n = 64$			$n = 128$			$n = 256$		
		.25	0	-.75	.25	0	-.75	.25	0	-.75
.5	$\hat{\nu}$	.141	.125	.102	.086	.072	.073	.051	.041	.053
	$\tilde{\nu}_0$	.075	.072	.083	.037	.036	.042	.019	.019	.021
	$\tilde{\nu}_{.1}$	.098	.086	.055	.070	.066	.050	.060	.057	.050
	$\tilde{\nu}_{-.1}$	.119	.137	.196	.111	.125	.164	.120	.127	.151
1	$\hat{\nu}$	.097	.088	.078	.059	.051	.057	.035	.029	.041
	$\tilde{\nu}_0$	.053	.051	.058	.026	.026	.030	.013	.013	.015
	$\tilde{\nu}_{.1}$	.085	.076	.051	.067	.063	.052	.058	.057	.052
	$\tilde{\nu}_{-.1}$	.108	.122	.167	.110	.120	.149	.121	.126	.143
2	$\hat{\nu}$	.067	.062	.060	.041	.036	.044	.024	.021	.031
	$\tilde{\nu}_0$	.038	.036	.041	.019	.018	.021	.009	.009	.011
	$\tilde{\nu}_{.1}$	.077	.070	.051	.064	.062	.054	.058	.056	.053
	$\tilde{\nu}_{-.1}$	.103	.114	.147	.110	.118	.138	.122	.125	.137

Notes: Monte Carlo SD over 5,000 replications of  $\hat{\nu}$ ,  $\tilde{\nu}_\theta$ , where in (2), (3)  $\mu = 0$ ,  $u_t$  is a Gaussian white noise with  $Var(u_{1t}) = 1$ ,  $Var(u_{2t}) = \tau$  and correlation  $\rho$

Table 9. Empirical example

	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$	$m_6$	$m_7$	$m_8$	$m_9$	$m_{10}$
$\widehat{\delta}$	1.88	1.92	1.98	2.00	2.03	2.04	2.07	2.03	1.94	1.80
$\widehat{\delta + \theta}$	1.77	1.84	1.81	1.68	1.73	1.66	1.65	1.67	1.64	1.53
$\widehat{\tau}_{hs}$	-.605	-.501	-1.02	-1.96	-1.89	-2.44	-2.73	-2.45	-2.10	-2.02
$X^{**}$	.715	1.12	1.33	2.31	3.33	5.66	7.04	6.81	6.81	7.70
$n'$	80	79	78	77	76	75	74	73	72	71
$\widehat{\nu}$	.651	.652	.701	.679	.653	.652	.665	.667	.688	.660
$\widehat{\theta}$	-.329	-.329	-.324	-.310	-.296	-.287	-.272	-.271	-.286	-.309
$\widehat{\nu}_{\mu 0}$	.579	.579	.588	.625	.657	.676	.709	.711	.682	.626
$\widehat{\theta}_{\mu 0}$	-.327	-.327	-.323	-.308	-.296	-.289	-.278	-.278	-.285	-.304

Notes:  $\widehat{\delta}$ ,  $\widehat{\delta + \theta}$ : local Whittle estimates of the memories of the observables  $h_t$ ,  $s_t$ , respectively (bandwidths:  $m_1 = 16$ ,  $m_{i+1} = m_i + 1$ );  $\widehat{\tau}_{hs}$ : Robinson and Yajima's (2002) test for equality of orders of integration (bandwidths:  $m_1 = 16$ ,  $m_{i+1} = m_i + 1$ );  $X^{**}$ : Robinson's (2008) test statistic for the null of no cointegration (bandwidths:  $m_1 = 26$ ,  $m_{i+1} = m_i + 1$ );  $n'$ : sample size;  $\widehat{\nu}$ ,  $\widehat{\theta}$ : estimates of the long run parameters (model with constant);  $\widehat{\nu}_{\mu 0}$ ,  $\widehat{\theta}_{\mu 0}$ : estimates of the long run parameters (model without constant)