Queen’s Economics Department Working Paper No. 1226

Fully Modified Narrow-Band Least Squares Estimation of Weak Fractional Cointegration

Morten Ørregaard Nielsen
Queen‘s University and CREATES

Per Frederiksen
Nordea Markets

Department of Economics
Queen’s University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

5-2010

Fully Modified Narrow-Band Least Squares Estimation of Weak Fractional Cointegration*

Morten Ørregaard Nielsen†
Queen's University and CREATES

Per Frederiksen
Nordea Markets

May 12, 2010

Abstract

We consider estimation of the cointegrating relation in the weak fractional cointegration model, where the strength of the cointegrating relation (difference in memory parameters) is less than one-half. A special case is the stationary fractional cointegration model, which has found important application recently, especially in financial economics. Previous research on this model has considered a semiparametric narrow-band least squares (NBLS) estimator in the frequency domain, but in the stationary case its asymptotic distribution has been derived only under a condition of non-coherence between regressors and errors at the zero frequency. We show that in the absence of this condition, the NBLS estimator is asymptotically biased, and also that the bias can be consistently estimated. Consequently, we introduce a fully modified NBLS estimator which eliminates the bias, and indeed enjoys a faster rate of convergence than NBLS in general. We also show that local Whittle estimation of the integration order of the errors can be conducted consistently based on NBLS residuals, but the estimator has the same asymptotic distribution as if the errors were observed only under the condition of non-coherence. Furthermore, compared to much previous research, the development of the asymptotic distribution theory is based on a different spectral density representation, which is relevant for multivariate fractionally integrated processes, and the use of this representation is shown to result in lower asymptotic bias and variance of the narrow-band estimators. We present simulation evidence and a series of empirical illustrations to demonstrate the feasibility and empirical relevance of our methodology.

Keywords: Fractional cointegration, frequency domain, fully modified estimation, long memory, semiparametric.

JEL Classifications: C22.

*We are grateful to the editor Pierre Perron, two anonymous referees, Javier Hualde, Lealand Morin, Benoit Perron, Katsumi Shimotsu, and participants at the 2006 Summer Econometrics Workshop in Aarhus and the 2006 European Meeting of the Econometric Society in Vienna for useful and constructive comments, and to the Social Sciences and Humanities Research Council of Canada (SSHRC grant no. 410-2009-0183), the Danish Social Sciences Research Council (FSE grant no. 275-05-0220) and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) for financial support. A previous version of this paper was circulated under the title "Fully modified narrow-band least squares estimation of stationary fractional cointegration."

†Corresponding author. Address: Department of Economics, Dunning Hall Room 307, 94 University Avenue, Queen's University, Kingston, Ontario K7L 3N6, Canada; phone: +1 613 533 2262; email: `mon@econ.queensu.ca`

1 Introduction

Recently, the concept of fractional cointegration has attracted increasing attention from both theoretical and empirical researchers in economics and finance. In this theory, a p-vector time series z_t is said to be cointegrated if the elements of z_t are integrated of order d, denoted $I(d)$, but there exists a linear combination that is $I(d-\delta)$ with $\delta > 0$. Originally, the concept of cointegration does not restrict d and δ to be integers, e.g. Granger (1981), but estimation methods have been developed mostly for the so-called $I(1) - I(0)$ cointegration, where it is assumed that $d = \delta = 1$.

For a precise statement, a covariance stationary time series $x_t \in I(d)$, $d < 1/2$, if $\Delta^d x_t = (1-L)^d x_t = v_t$, where $v_t \in I(0)$, i.e. has continuous spectral density that is bounded and bounded away from zero at all frequencies. In this case, $\{x_t\}$ has spectral density

$$f(\lambda) \sim g\lambda^{-2d} \text{ as } \lambda \to 0^+, \tag{1}$$

where $g \in (0, \infty)$ is a constant and the symbol "$\sim$" means that the ratio of the left- and right-hand sides tends to one in the limit. The parameter d determines the memory of the process: if $d \in (0, 1/2)$ the process is covariance stationary with long memory, but if $d = 0$ the process has only weak dependence. For surveys see, e.g., Baillie (1996) and Robinson (2003).

We consider estimation of the single-equation cointegrating regression

$$y_t = \alpha + \beta' x_t + u_t, \quad t = 1, ..., T, \tag{2}$$

where both the regressors and the errors have long memory, and can in fact be nonstationary ($d > 1/2$), but the errors have less memory than the regressors, i.e. where $x_t \in I(d_x)$ and $u_t \in I(d_u)$ with $d_x > d_u \geq 0$. In particular, we consider the model with $d_x - d_u < 1/2$ which is termed *weak fractional cointegration* in Hualde & Robinson (2010).

To accommodate potential nonstationarity, we let $\gamma \geq 0$ and consider[1]

$$\Delta^\gamma y_t = \alpha + \beta' \Delta^\gamma x_t + \Delta^\gamma u_t, \quad t = 1, ..., T. \tag{3}$$

Now $\Delta^\gamma x_t \in I(d_x - \gamma)$ and $\Delta^\gamma u_t \in I(d_u - \gamma)$, and γ is any real number which transforms a potentially nonstationary model (like (2)) into one with stationary regressors (so that $d_x - \gamma < 1/2$), where, additionally, the cointegrating error in the transformed model has nonnegative memory (so that $d_u - \gamma \geq 0$). The interpretation is that γ is a user-chosen number whose choice affects the estimation procedure, and it is connected to the model (2) which generates the data only through the two requirements $1/2 > d_x - \gamma > d_u - \gamma \geq 0$. That way x_t and possibly also u_t may be nonstationary.[2] Different choices of γ lead to different estimators, but one choice in particular leads to the best estimator in a GLS sense. This is $\gamma = d_u$ which, given that we are under weak cointegration, is always an appropriate choice in the sense that $d_x - \gamma < 1/2$ and $d_u - \gamma \geq 0$. Of course, d_u is unknown, but a feasible version of the estimator with $\gamma = d_u$ may be implemented by replacing d_u with a suitable estimator, $\hat{d}_u$. Another interesting special case is $\gamma = 0$ in which case we require $d_x < 1/2$, termed *stationary fractional cointegration* by Robinson (1994).

After appropriate differencing our model is stationary, so a comparison with the standard time series regression model with weakly dependent regressors is natural. It is well known that, in the standard case, under a wide variety of regularity conditions, the ordinary least squares (OLS) estimator of β in (2) is asymptotically normal, see e.g. Hannan (1979). The new complication is that, as pointed out by Robinson (1994) and Robinson & Hidalgo (1997), when the regressors and the errors both have long memory and are possibly non-orthogonal, the OLS estimator in (2) is in

[1] There is a slight abuse of notation in (3) since $\Delta^\gamma \alpha$ is a constant which we also call α. However, because our estimators are functions of the periodogram at non-zero frequencies only, this is irrelevant.

[2] When defining nonstationary fractionally integrated processes there is a choice of type I or type II variants. In our model, either choice will lead to identical first order asymptotic results, see Robinson (2005). Therefore, we do not consider this issue further, and below state our assumptions in terms of $\Delta^\gamma x_t$ and $\Delta^\gamma u_t$ which are stationary.

general no longer consistent. To deal with this issue, Robinson (1994) proposed a semiparametric narrow-band least squares (NBLS) estimator in the frequency domain (as opposed to a fixed band estimator as considered by e.g. Phillips (1991) in a cointegration context). The NBLS estimator assumes only a multivariate version of (1), and essentially performs OLS on a degenerating band of frequencies around the origin. The consistency of the estimator in the stationary case was proved by Robinson (1994). Christensen & Nielsen (2006) showed that its asymptotic distribution is normal when the collective memory of the regressors and the error term is less than 1/2, i.e. when $d_x + d_u < 1/2$, and under the condition that the regressors and the errors have zero coherence at the origin. In contrast, Robinson & Marinucci (2003) consider several cases where the regressors are nonstationary fractionally integrated and the limiting distributions for the NBLS estimator involve fractional Brownian motion, and Chen & Hurvich (2003) add deterministic trends.

We follow a semiparametric approach characterized by assuming only a local model such as (1) for the spectral density, and using a degenerating part of the periodogram around the origin for estimation. This approach has the advantage of being invariant to short-term dynamics (and mean terms since the zero frequency is usually left out). Specifically, a local Whittle estimator of the memory parameter d based on maximization of a local Whittle approximation to the likelihood using (1) has been developed by Künsch (1987) and Robinson (1995*a*). Of course, a fully parametric estimator would be more efficient, but is inconsistent if the parametric model is misspecified.

The methods described above are combined by Marinucci & Robinson (2001*b*) and Christensen & Nielsen (2006), who suggest conducting a (stationary) fractional cointegration analysis in several steps. First, the integration orders of the observed data are estimated by the local Whittle estimator. Secondly, the NBLS estimator of the cointegrating vector is calculated, and finally the integration order of the residuals is estimated assuming that the local Whittle approach is equally valid when based on residuals. Hypothesis testing is then conducted on d_u as if u_t were observed, and on β as if d_u (which enters in the limiting distribution of the NBLS estimator) were known. Moreover, the distribution theory for the NBLS estimator developed by Christensen & Nielsen (2006) assumes that the long-run (zero frequency) coherence between the regressors and the errors is zero.

In this paper, we extend the stationary setting of Marinucci & Robinson (2001*b*) and Christensen & Nielsen (2006) to that of weak fractional cointegration. We develop the asymptotic distribution theory based on a different spectral density representation, which is relevant for multivariate fractionally integrated processes, and the use of this representation is shown to result in lower asymptotic bias and variance of the narrow-band estimators. We show that in the non-zero coherence case a bias term appears in the mean of the asymptotic normal distribution of the NBLS estimator. The bias term is proportional to the square-root of the bandwidth, with factor of proportionality depending on the integration orders and the coherence at frequency zero. However, we show that the bias can be estimated and hence removed by a fully modified type procedure in the spirit of Phillips & Hansen (1990). The result is a fully modified NBLS (FMNBLS) estimator, which has no asymptotic bias and the same asymptotic variance as the NBLS estimator. However, the FMNBLS estimator will have a better rate of convergence in general, i.e. the same rate as the NBLS estimator under non-coherence as in Christensen & Nielsen (2006).

We also consider inference on the integration order of the error term in the cointegrating relation, and show that in the case of stationary errors it can be consistently estimated by the local Whittle estimator based on the residuals from a NBLS regression. However, the local Whittle estimator converges at a slower rate than if the errors were observed except if there is no long-run coherence between regressors and errors. In the latter case the asymptotic distribution theory for the local Whittle estimator is unaffected by the fact that the estimator is based on residuals.

Extensions of the well known fully modified least squares procedure of Phillips & Hansen (1990) to nonstationary fractional cointegration have been examined by Dolado & Marmol (1996), Kim & Phillips (2001), and Davidson (2004) in parametric frameworks. An alternative fully modified procedure for the $I(1)-I(0)$ model was suggested by Marinucci & Robinson (2001*a*), who considered the estimator of Phillips & Hansen (1990) based on NBLS residuals rather than OLS residuals.

However, the approach taken in the present paper is more direct. We derive an expression for the asymptotic bias term, which depends on the integration orders of the regressors and the errors and also on the coherence matrix at the zero frequency. We show that under appropriate conditions on the bandwidth parameters the bias term can be estimated consistently, e.g., by running an auxiliary NBLS regression, and this can be used to modify the initial NBLS estimator to eliminate the bias.

In a simulation study we document the finite sample feasibility of the proposed FMNBLS estimator. The simulations demonstrate the superiority in terms of bias of FMNBLS relative to NBLS in the presence of non-zero long-run coherence between the regressor and the error, which comes at the cost of an increased finite sample variance. In terms of RMSE, FMNBLS also clearly outperforms NBLS in most cases with non-zero long-run coherence.

To demonstrate the empirical relevance of our methodology, we include several brief empirical illustrations. We first revisit the long-run unbiasedness question in the implied-realized volatility relation. We then consider the relation between inflation rates of European Union countries, exemplified by the harmonized consumer price indexes of France and Spain. Lastly, we investigate the relationship between the volatilities of the General Electric stock and two stock indexes.

The remainder of the paper is laid out as follows. Next, we describe NBLS estimation of (2) and (3) and derive the relevant asymptotic distribution theory. We also discuss inference using the local Whittle estimator of the integration order of the errors when the errors are not observed and residuals are used instead. In Section 3 we consider the FMNBLS modification to the NBLS estimator. Sections 4 and 5 present simulation evidence and empirical illustrations, respectively, and Section 6 offers some concluding remarks. All proofs are gathered in the appendices.

2 Narrow-Band Least Squares Estimation

We begin with some remarks about the spectral representation of multivariate long memory models. Suppose the spectral density of the covariance stationary process $w_t = (\Delta^\gamma x_t', \Delta^\gamma u_t)'$ is

$$f(\lambda) \sim \Lambda(\lambda)^{-1} G \bar{\Lambda}(\lambda)^{-1} \text{ as } \lambda \to 0^+, \tag{4}$$

where the bar denotes complex conjugation, $\Lambda(\lambda) = \text{diag}(e^{-i\pi d_1/2}\lambda^{d_1}, ..., e^{-i\pi d_p/2}\lambda^{d_p})$, and G is a real, symmetric, positive definite matrix. The spectral density representation (4) is motivated by the multivariate stationary fractionally integrated model with $d_a \in (-1/2, 1/2), a = 1, ..., p$:

$$\begin{bmatrix} (1-L)^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-L)^{d_p} \end{bmatrix} \begin{bmatrix} w_{1t} - Ew_{1t} \\ \vdots \\ w_{pt} - Ew_{pt} \end{bmatrix} = \begin{bmatrix} v_{1t} \\ \vdots \\ v_{pt} \end{bmatrix}, \quad t = 1, ..., T, \tag{5}$$

where $v_t = (v_{1t}, ..., v_{pt})'$ is a covariance stationary process with spectral density matrix that is finite and bounded away from zero (in the sense of positive definite matrices) at all frequencies, i.e. v_t is $I(0)$. When v_t is an ARMA model, w_t is a multivariate fractional ARIMA model. This class of models is very popular in both theoretical and applied time series analysis. Since $(1-e^{i\lambda})^d = \lambda^d e^{-i\pi d/2}(1+O(\lambda))$ as $\lambda \to 0$ the representation (4) follows by defining $G = \lim_{\lambda \to 0} f_v(\lambda)$.

A typical element of (4) is

$$f_{ab}(\lambda) \sim G_{ab}\lambda^{-d_a-d_b}e^{i\pi(d_a-d_b)/2} \text{ as } \lambda \to 0^+, \quad a, b = 1, ..., p,$$

where d_a and d_b appear in both in the power decay and in the phase shift. Note that $f_{ab}(\lambda)$ differs from the simpler representation

$$h_{ab}(\lambda) \sim G_{ab}\lambda^{-d_a-d_b} \text{ as } \lambda \to 0^+, \quad a,b = 1,...,p, \tag{6}$$

applied by, e.g., Robinson (1995*b*) and Lobato & Robinson (1998) for inference on the integration orders and by Robinson & Marinucci (2003) and Christensen & Nielsen (2006) in the context of stationary fractional cointegration. The most important difference is that $f(\lambda)$ has non-zero complex part even at the origin unless $d_a = d$ for all $a = 1,...,p$, and neglecting the complex part is a source of misspecification. For a detailed comparison of $f(\lambda)$ and $h(\lambda)$, see Shimotsu (2007) and Robinson (2008) who derive multivariate local Whittle estimators based on (4).

We remark here that the assumptions of Christensen & Nielsen (2006) (and hence also those of, e.g., Lobato & Robinson (1998) and Lobato (1999)) and much subsequent research, unfortunately, appear incompatible. The reason is that the real-valued cross-spectral density (6) imposed in their Assumption A implies that cross-autocorrelations are symmetric with respect to time, which implies a two-sided moving average with equal lead and lag weights and not a one-sided moving average as imposed in their Assumption B. The assumptions of Christensen & Nielsen (2006) (and subsequent research on narrow-band estimation of stationary fractional cointegration) can be made compatible, however, in light of their condition that $G_{ap} = G_{pa} = 0$, by assuming that the integration orders of the regressors are all equal, i.e. that $d_a = d_x$ for $a = 1,\dots,p-1$ and $d_x > d_p$. In that special case, the representations (4) and (6) are equivalent and their results correct.

To consider frequency domain least squares inference on β in the cointegrating relation (2) or the pre-differenced regression (3), we define the cross-periodogram matrix between the observed vectors $\{\Delta^\gamma q_t, t = 1,\dots,T\}$ and $\{\Delta^\gamma r_t, t = 1,\dots,T\}$,

$$I_{qr}(\gamma,\lambda) = \frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\Delta^\gamma q_t)(\Delta^\gamma r_s)' e^{-i(t-s)\lambda}. \tag{7}$$

We then form the discretely averaged co-periodogram

$$\hat{F}_{qr}(\gamma,k,l) = \frac{2\pi}{T}\sum_{j=k}^{l}\operatorname{Re}\left(I_{qr}(\gamma,\lambda_j)\right), \quad 0 \le k \le l \le T-1, \tag{8}$$

for $\lambda_j = 2\pi j/T$. By setting $k \ge 1$ and thus excluding the zero frequency, the estimator becomes invariant to non-zero means, i.e. invariant to α in (2) and (3).

With $\hat{F}$ defined in (8) we consider the frequency domain least squares estimator

$$\hat{\beta}_m(\gamma) = \hat{F}_{xx}^{-1}(\gamma,1,m)\,\hat{F}_{xy}(\gamma,1,m) \tag{9}$$

of β in the regression (3). Notice that, by this definition, $\hat{\beta}_{T-1}(0)$ is algebraically identical to the usual OLS estimator of β in (2) with allowance for a non-zero mean. On the other hand, if

$$\frac{1}{m} + \frac{m}{T} \to 0 \text{ as } T \to \infty, \tag{10}$$

then $\hat{\beta}_m(\gamma)$ is a NBLS estimator using only a degenerating band of frequencies near the origin. We need m to tend to infinity to gather information, but we also need to remain in a neighborhood of zero where we have assumed knowledge about the spectral density, so m/T must tend to zero.

When $\gamma = 0$, the NBLS estimator $\hat{\beta}_m(0)$ in (9) is the estimator defined by Robinson (1994). On the other hand, with $\gamma = d_u$, $\hat{\beta}_m(d_u)$ is a GLS-type estimator similar to the one discussed in Nielsen (2005), who also shows that the latter in fact also corresponds to a local Whittle quasi-maximum likelihood estimator of β. Of course $\hat{\beta}_m(d_u)$ is infeasible, but a feasible version will be discussed in the next section.

To prove our main results we assume, with obvious implications for $\Delta^\gamma y_t$, the following conditions on $w_t = (\Delta^{\gamma} x_t', \Delta^{\gamma} u_t)'$ and the bandwidth parameter. Here and throughout, the memory parameters d_a, d_b, and d_p are used to refer to w_{at}, w_{bt}, and w_{pt}, respectively. I.e., the memory parameters $d_a, a = 1, \ldots, p$, belong to the transformed regression (3) and are related to the original memory parameters $d_{x,a}, d_u$, and the pre-differencing parameter γ by $d_a = d_{x,a} - \gamma, a = 1, \ldots, p-1$, and $d_p = d_u - \gamma$.

Assumption 1 *The spectral density matrix of w_t given in (4) with typical element $f_{ab}(\lambda)$, the cross-spectral density between w_{at} and w_{bt}, satisfies*

$$|f_{ab}(\lambda) - G_{ab}\lambda^{-d_a-d_b} e^{i(\pi-\lambda)(d_a-d_b)/2}| = O(\lambda^{\phi-d_a-d_b}) \text{ as } \lambda \to 0^+, \ a, b = 1, ..., p, \tag{11}$$

for some $\phi \in (0, 2]$. The matrix G is positive definite.

Assumption 2 *The memory parameters satisfy $0 \leq d_a < 1/2$ for $a = 1, ..., p$, $d_a + d_p < 1/2$ for $a = 1, ..., p-1$, and $\min_{1 \leq a \leq p-1} d_a - d_p = \min_{1 \leq a \leq p-1} d_{x,a} - d_u = \delta_{\min} > 0$.*

Assumption 3 *w_t is a linear process, $w_t = \mu + \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$, with square summable coefficient matrices, $\sum_{j=0}^{\infty} \|A_j\|^2 < \infty$. The innovations ε_t satisfy $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = I_p$, $E(\varepsilon_t \otimes \varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = \mu_3$, and $E(\varepsilon_t \varepsilon_t' \otimes \varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = \mu_4$, almost surely, where μ_3 and μ_4 are non-stochastic, finite, and do not depend on t, and $\mathcal{F}_t = \sigma(\{\varepsilon_s, s \leq t\})$.*

Assumption 4 *Let $A_a(\lambda)$ denote the a'th row of $A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$. Then, as $\lambda \to 0^+$,*

$$\frac{\partial A_a(\lambda)}{\partial \lambda} = O(\lambda^{-1} ||A_a(\lambda)||), \ a = 1, ..., p.$$

Assumption 5 *The bandwidth parameter $m_0 = m_0(T)$ satisfies*

$$\frac{1}{m_0} + \frac{m_0^{1+2\min(1,\phi)}}{T^{2\min(1,\phi)}} \to 0 \text{ as } T \to \infty.$$

Our assumptions are a multivariate generalization of those in Robinson (1994, 1995*a*), see also Lobato (1999) and Christensen & Nielsen (2006). Since our assumptions are semiparametric in nature they naturally differ from those employed by e.g. Robinson & Hidalgo (1997) in their parametric setup, and are at least in some respects weaker than standard parametric assumptions. In particular, we avoid standard assumptions (from stationary time series regression) of independence or uncorrelatedness between x_t and u_t as well as complete and correct specification of $f(\lambda)$.

The first part of Assumption 1 specializes (4) by imposing smoothness conditions on the spectral density matrix of w_t commonly employed in the literature. They are satisfied with $\phi = 2$ if, for instance, w_t is a vector fractional ARIMA process. The more precise approximation offered by Assumption 1 relative to (4) reflects the approximation $(1 - e^{i\lambda})^d = |2\sin(\lambda/2)|^d e^{-i(\pi-\lambda)d/2} = \lambda^d e^{-i(\pi-\lambda)d/2}(1 + O(\lambda^2))$ as $\lambda \to 0$, see Shimotsu (2007). The positive definiteness condition on G is a no multicollinearity or no cointegration condition within the components of x_t, which is typical in single-equation cointegration models and in regression models.

The single-equation cointegrating regression model (2) is similar to the usual cointegrating regression model in the $I(1) - I(0)$ case, and the nature of the regression setup is subject to the same advantages and disadvantages. Two important issues, given a set of more than two variables, are to justify the single-equation regression and to justify the choice of the left-hand side variable. For the latter issue, it is likely that economic theory can be used as guidance and in any case this should be done on a case-by-case basis. For the former issue, since cointegration among the regressors is ruled out by Assumption 1 (as is standard in cointegrating regression models), in practice one would have to establish that only one cointegrating relationship exists among the given set of variables. This could be done, e.g., by the approach of Robinson & Yajima (2002) as in the empirical application in Section 5.3 below.

Much of the previous literature on semiparametric frequency domain inference in the fractional cointegration model distinguish (either explicitly or implicitly) between cases of coherence and non-coherence between the regressors and the error process at the zero frequency, e.g. Robinson & Marinucci (2003), Christensen & Nielsen (2006), and Robinson (2008). In the present notation this condition is $G_{ap} = G_{pa} = 0$, for $a = 1, ..., p-1$. Indeed, in the stationary case, asymptotic distribution theory for the NBLS estimator is only available in the case with non-coherence at the zero frequency, see Christensen & Nielsen (2006). Our assumptions avoid the non-coherence condition and thus allow correlation between the errors and regressors at any frequency.

The conditions on d_a translate into conditions on γ given $d_{x,a}$ and d_u. In particular, γ satisfies

$$\frac{1}{2}(\max_{1\leq a\leq p-1} d_{x,a} + d_u - 1/2) < \gamma \leq d_u. \tag{12}$$

For instance, we find that $\gamma = 0$ is permitted only in the stationary case when also $d_{x,a} + d_u < 1/2, a = 1, \ldots, p-1$. In the important GLS case with $\gamma = d_u$, Assumption 2 reduces to

$$0 < \delta_{\min} = \min_{1\leq a\leq p-1} d_{x,a} - d_u \leq \max_{1\leq a\leq p-1} d_{x,a} - d_u < 1/2, \tag{13}$$

which is exactly the weak fractional cointegration assumption such that the condition $d_a + d_p < 1/2$ is redundant in that case because $d_p = 0$.

In view of the results from, e.g., Fox & Taqqu (1986, Prop. 1) and Lobato & Robinson (1996), showing that quadratic forms of long memory processes with square-summable autocovariances ($2d < 1/2$) are asymptotically Gaussian, we assume that $d_a + d_p < 1/2$ in Assumption 2. The last condition of Assumption 2 is the essential assumption of cointegration, with $\delta_{\min}$ denoting the strength of the cointegrating relation.

Assumptions 3 and 4 follow Robinson (1995*a*) and Lobato (1999) in imposing a linear structure on w_t with square summable coefficients and martingale difference innovations with finite fourth moments. The assumption of constant conditional variance for the innovations could presumably be relaxed by assuming boundedness of higher moments as in Robinson & Henry (1999). Under Assumption 3 we can write the spectral density matrix of w_t as

$$f(\lambda) = \frac{1}{2\pi} A(\lambda) A^*(\lambda), \tag{14}$$

where the asterisk denotes transposed complex conjugation. Assumption 4 is a smoothness condition imposing differentiability of the spectral density near the origin, analogous to those imposed on the spectral density at any frequency in parametric frameworks, see e.g. Fox & Taqqu (1986). The condition is satisfied, e.g., by fractional ARIMA models.

The statements of Assumptions 1 and 4 are made in the frequency domain whereas the statement of Assumption 3 is in the time domain, which follows the tradition in the literature on semiparametric estimation in long memory models. Clearly, the assumptions are closely related, and in particular the matrix G in Assumption 1 is a function of the lag weights $\{A_j, j \geq 0\}$ from Assumption 3. The connection between the representations (4) and (6) (or Assumption 1) and the lag weights in the linear process (Assumption 3) is explored in Theorems 1 and 2 of Robinson (2008). In particular, it is shown there that our Assumptions 1 and 3 are compatible.[3]

Finally, Assumption 5 restricts the expansion rate of the bandwidth parameter $m_0 = m_0(T)$. The bandwidth is required to tend to infinity for consistency, but at a slower rate than T to remain

[3] Note that we could alternatively write our Assumptions 1-4 in terms of the model (5) and the errors v_t, as in e.g. Shimotsu & Phillips (2005).

in a neighborhood of the origin, where we have assumed some knowledge of the form of the spectral density. When ϕ is high, (11) is a better approximation to (14) as $\lambda \to 0^+$, and hence (by the second term of Assumption 5) a higher expansion rate of the bandwidth can be chosen. The weakest constraint is implied by $\phi \geq 1$, in which case the condition is $m_0 = o(T^{2/3})$. A slightly weaker bandwidth condition was employed by Christensen & Nielsen (2006) due to their assumption of real-valued spectral density at the origin.

We next derive the distribution of the NBLS estimator in the fractional cointegration model (2)-(3). This generalizes the consistency (with rates) result of Robinson & Marinucci (2003) (when $\gamma = 0$) and the asymptotic normality results of Nielsen (2005) and Christensen & Nielsen (2006) (who assumed non-coherence at the origin and a different spectral density model).

Theorem 1 *Let Assumptions 1-5 be satisfied. Then the NBLS estimator $\hat{\beta}_{m_0}(\gamma)$ in (9) satisfies*

$$\sqrt{m_0}\left(\lambda_{m_0}^{d_u}\Lambda_{m_0}^{-1}(\hat{\beta}_{m_0}(\gamma)-\beta)-K(\gamma)^{-1}H(\gamma)\right) \xrightarrow{D} N\left(0, K(\gamma)^{-1}J(\gamma)K(\gamma)^{-1}\right) \text{ as } T\to\infty, \tag{15}$$

where $\Lambda_m = \operatorname{diag}(\lambda_m^{d_{x,1}},...,\lambda_m^{d_{x,p-1}})$ and, for $a,b=1,...,p-1$, $K(\gamma)=(K_{ab}(\gamma))$, $H(\gamma)=(H_a(\gamma))$, and $J(\gamma)=(J_{ab}(\gamma))$ are given by

$$K_{ab}(\gamma) = \frac{G_{ab}}{1-d_{x,a}-d_{x,b}+2\gamma}\cos\left(\frac{\pi}{2}\left(d_{x,a}-d_{x,b}\right)\right),$$
$$H_a(\gamma) = \frac{G_{ap}}{1-d_{x,a}-d_u+2\gamma}\cos\left(\frac{\pi}{2}\left(d_{x,a}-d_u\right)\right),$$
$$J_{ab}(\gamma) = \frac{G_{ap}G_{bp}}{2\left(1-d_{x,a}-d_{x,b}-2d_u+4\gamma\right)}\cos\left(\frac{\pi}{2}\left(d_{x,a}+d_{x,b}-2d_u\right)\right)+\frac{G_{ab}G_{pp}}{2\left(1-d_{x,a}-d_{x,b}-2d_u+4\gamma\right)}\cos\left(\frac{\pi}{2}\left(d_{x,a}-d_{x,b}\right)\right).$$

Proof. See Appendix A.1. ■

Theorem 1 refines the results of Nielsen (2005) and Christensen & Nielsen (2006) in three ways: first, our result uses the representation (4) of the multivariate spectral density, secondly we allow for non-zero coherence at the origin, and thirdly we generalize the result to the weak fractional cointegration model. The cosine terms in the asymptotic distribution are a result of using the representation (4) rather than the simpler (6), in which case these terms would not be present. In the absence of any coherence between the regressors and the errors at the origin, the distribution theory follows from the above results by setting $G_{ap} = G_{pa} = 0$ for $a = 1, ..., p-1$. Also note that the theorem presents a simple and closed form expression for the asymptotic bias term $K(\gamma)^{-1}H(\gamma)$. In the next section we show that $K(\gamma)^{-1}H(\gamma)$ can be estimated consistently with a sufficient rate such that the bias can be removed and a centered distribution can be obtained.

To illustrate the distribution theory and the developments leading to the below FMNBLS estimator, we consider briefly an illustrative example. Consider the two-variable stationary case, i.e. the regression (2) or (3) with only one regressor. Denote the integration orders d_x and d_u and the spectral density matrix at the origin $G = (G_{ab})$ with $a, b = x, u$. In this case (15) reduces to

$$\sqrt{m_0}(\lambda_{m_0}^{d_u-d_x}(\hat{\beta}_{m_0}(\gamma)-\beta)-\eta(\gamma)) \xrightarrow{D} N\left(0,\omega(\gamma)^2\right), \tag{16}$$

where the asymptotic bias and variance terms are given by

$$\eta(\gamma) = \frac{G_{xu}}{G_{xx}}\frac{(1-2d_x+2\gamma)}{(1-d_x-d_u+2\gamma)}\cos\left(\frac{\pi}{2}\left(d_x-d_u\right)\right),$$
$$\omega(\gamma)^2 = \frac{(1-2d_x+2\gamma)^2}{2\left(1-2d_x-2d_u+4\gamma\right)}\left(\frac{G_{uu}}{G_{xx}}+\frac{G_{ux}^2}{G_{xx}^2}\cos\left(\pi\left(d_x-d_u\right)\right)\right).$$

Note that, if the spectral representation (6) were used instead of (4), the cosine terms in both $\eta(\gamma)$ and $\omega(\gamma)$ would be replaced by unity, their upper bound. The increased variance obtained using (6) when the true model is (4) is a consequence of the misspecification of the spectral density at the origin since the non-zero complex part in (4) is ignored in (6). Hence, using the correct representation (4) results in a distribution theory that is more precise, both in terms of bias and variance, as shown in Theorem 1.

Both the bias $\eta(\gamma)$ and the variance $\omega(\gamma)^2$ depend on γ, and it is easily verified that $\eta(\gamma)$ is increasing in γ whereas $\omega(\gamma)^2$ is decreasing in γ for $\gamma \in [0, d_u]$. Thus, as expected, the minimum variance is attained for the GLS estimator. The fact that the bias is increasing in γ will be inconsequential since our FMNBLS estimator eliminates the bias term.

In addition to the bias, the absence of the zero coherence condition results in an additive variance inflation of

$$\frac{(1-2d_x+2\gamma)^2}{2(1-2d_x-2d_u+4\gamma)}\frac{G_{ux}^2}{G_{xx}^2}\cos\left(\pi\left(d_x-d_u\right)\right) \geq 0.$$

However, consistency of the estimator is not affected by the presence of non-zero coherence between the regressors and the errors at the zero frequency, and the rate result established by Robinson (1994) and Robinson & Marinucci (2003) (for $\gamma = 0$) is, in fact, sharp in this case as conjectured by Robinson & Marinucci (2003). This is easily seen from (16), where

$$\hat{\beta}_{m_0}(\gamma) - \beta = \frac{\lambda_{m_0}^{d_x-d_u}}{\sqrt{m_0}}\omega(\gamma)Z + \lambda_{m_0}^{d_x-d_u}\eta(\gamma) + o_P(m_0^{-1/2}\lambda_{m_0}^{d_x-d_u}),$$

for $Z \sim N(0,1)$. The consistency and rate of $\hat{\beta}_{m_0}(\gamma)$ follows immediately, and in particular, $\lambda_{m_0}^{d_u-d_x}(\hat{\beta}_{m_0}(\gamma)-\beta) \xrightarrow{P} \eta(\gamma)$. That is, when normalized as in Robinson (1994) and Robinson & Marinucci (2003), the NBLS estimator converges to a degenerate distribution (a constant) in the case of non-zero coherence between the regressors and the errors at the origin. However, in the absence of coherence between the regressors and the errors at the origin and normalized by an additional $\sqrt{m_0}$, the NBLS estimator has an asymptotic normal distribution.

To estimate the asymptotic bias term $\eta(\gamma)$ and to feasibly implement the GLS version of the FMNBLS estimator $\hat{\beta}_{m_0}(d_u)$ we would need an estimate of the memory parameter d_u of the error term. To that end, we next consider the local Whittle estimator $\hat{d}_u$ based on NBLS residuals with $\gamma = 0$. We will give the results for the stationary case where $d_u < d_{x,a} < 1/2$. Similar results have been derived by, e.g., Velasco (2003) for nonstationary fractional cointegration.

Thus, suppose d_u is estimated by

$$\begin{aligned} \hat{d}_u &= \arg\min_{d\in\Delta}\hat{R}(d), \qquad (17)\\ \hat{R}(d) &= \log\hat{G}(d) - \frac{2d}{m_1}\sum_{j=1}^{m_1}\log\lambda_j, \quad \hat{G}(d) = \frac{1}{m_1}\sum_{j=1}^{m_1}\lambda_j^{2d}I_{\hat{u}\hat{u}}(0,\lambda_j), \end{aligned}$$

where $\Delta = [0, \Delta_2], 0 < \Delta_2 < 1/2$, is the parameter space and

$$I_{\hat{u}\hat{u}}(\gamma,\lambda_j) = I_{uu}(\gamma,\lambda_j) + (\beta-\hat{\beta}_m(\gamma))'\operatorname{Re}(I_{xx}(\gamma,\lambda_j))(\beta-\hat{\beta}_m(\gamma)) + 2(\beta-\hat{\beta}_m(\gamma))'\operatorname{Re}(I_{xu}(\gamma,\lambda_j)) \qquad (18)$$

is the periodogram of the differenced residual series $\Delta^\gamma\hat{u}_t = \Delta^\gamma y_t - \hat{\beta}_m(\gamma)'\Delta^\gamma x_t = \Delta^\gamma u_t + (\beta - \hat{\beta}_m(\gamma))'\Delta^\gamma x_t$. The lower bound of the parameter space reflects prior information that $d_u \geq 0$, which seems reasonable from a practical/empirical point of view. This condition could be relaxed at the cost of a longer proof of the following theorem.

We introduce the following condition on the expansion rate of the bandwidth parameter $m_1 = m_1(T)$ used for the local Whittle estimator of d_u :

$$(\log T)^4(\log m_1)\left(\frac{m_0}{m_1}\right)^{\delta_{\min}} + \frac{m_1^{1+2\phi}(\log m_1)^2}{T^{2\phi}} \to 0 \text{ as } T\to\infty, \qquad (19)$$

where m_0 is the bandwidth parameter for $\hat{\beta}$ from Assumption 5 and ϕ is the smoothness parameter from Assumption 1. The first part of (19) is essentially satisfied if m_1 diverges to infinity at a faster rate than m_0. The second part is the standard assumption on the bandwidth parameter for local Whittle estimation, e.g. Robinson (1995*a*).

Theorem 2 *Let Assumptions 1 and 3-5 be satisfied with $\gamma = 0$ and suppose $\hat{d}_u$ is given by (17) using bandwidth m_1 satisfying (19) and based on residuals $\hat{u}_t = y_t - \hat{\beta}_{m_0}(0)'x_t$, where $\hat{\beta}_{m_0}(0)$ is the NBLS estimator (9). Suppose d_u belongs to the interior of Δ and $d_u < \min_{1\leq a\leq p-1} d_{x,a} \leq \max_{1\leq a\leq p-1} d_{x,a} < 1/2$. Then, as $T \to \infty$,*

$$\hat{d}_u - d_u = O_P\left((\log m_1)\,(m_0/m_1)^{\delta_{\min}} \right) \xrightarrow{P} 0. \tag{20}$$

If, in addition, $G_{ap} = G_{pa} = 0$ for $a = 1, \ldots, p-1$ and $(m_0/m_1)^{2\delta_{\min}}\sqrt{m_1}/m_0 \to 0$, then

$$\sqrt{m_1}(\hat{d}_u - d_u) \xrightarrow{D} N\,(0, 1/4) \text{ as } T \to \infty.$$

Proof. See Appendix A.2. ■

The second part of Theorem 2 shows that, in the absence of long-run coherence between regressors and errors and under an additional (weak) restriction on the bandwidth, the local Whittle estimator of the integration order of the errors is unaffected by the fact that it is based on NBLS residuals. In the general case the local Whittle estimator remains consistent, although it converges at a slower rate. Moreover, the result in Theorem 2 shows that in fact the three step procedure employed by Marinucci & Robinson (2001*b*) and Christensen & Nielsen (2006) is valid only when there is no long-run coherence, as assumed in Christensen & Nielsen (2006). That is, in their setup inference on d_u may be conducted based on our distributional result in Theorem 2 and is equivalent to disregarding the fact that the estimator is based on residuals, as long as the bandwidth parameter is chosen according to our assumptions.

To conclude this section we make the following assumption regarding estimation of d_u.

Assumption 6 *The memory parameter d_u of the error term is estimated semiparametrically based on NBLS residuals $\hat{u}_t = y_t - \hat{\beta}_{m_0}(0)'x_t$ and using bandwidth parameter m_1 satisfying (19). The resulting estimator $\hat{d}_u$ satisfies (20).*

By Theorem 2 it follows that Assumption 6 is satisfied when $d_u < \min_{1\leq a\leq p-1} d_{x,a} \leq \max_{1\leq a\leq p-1} d_{x,a} < 1/2$. Similar results for the nonstationary case are provided by, e.g., Velasco (2003).

3 Fully Modified NBLS Estimation

We next consider estimation of the bias in NBLS from Theorem 1, i.e. estimation of $K(\gamma)^{-1}H(\gamma)$. From the definitions of $K(\gamma)$ and $H(\gamma)$ in Theorem 1 and its proof, we can equivalently write

$$K(\gamma) = \lambda_m^{-1}\Lambda_m F_{xx}(\gamma, \lambda_m)\,\Lambda_m,\;\; H(\gamma) = \lambda_m^{d_p-1}\Lambda_m F_{xu}(\gamma, \lambda_m)\,, \tag{21}$$

where $F_{qr}(\gamma, \lambda) = \int_0^\lambda \mathrm{Re}\,(f_{qr}(\gamma, \theta))\,d\theta$ and $f_{qr}(\gamma, \theta)$ is the cross-spectral density between $\Delta^\gamma q_t$ and $\Delta^\gamma r_t$. Thus, $K(\gamma)$ is the (scaled) integrated co-spectrum of $\Delta^\gamma x_t$ and $H(\gamma)$ is the (scaled) integrated co-spectrum between $\Delta^\gamma x_t$ and $\Delta^\gamma u_t$. By rewriting $K(\gamma)$ and $H(\gamma)$ in this way, the bias term $K(\gamma)^{-1}H(\gamma)$ is recognized to be the (scaled) population equivalent to the coefficient estimator in a regression of the errors from (3) on the regressors. This mimics the corresponding well-known result from ordinary least squares when the errors and regressors are correlated. However, in our weak fractional cointegration setup the bias term can be estimated and hence eliminated.

It follows that a natural estimator of the bias can be based on

$$\Gamma_{m_2}(\gamma) = \hat{F}_{xx}^{-1}(\gamma, 1, m_2)\,\hat{F}_{xu}(\gamma, 1, m_2)\,,$$

using bandwidth parameter $m_2 = m_2(T)$. However, the estimator $\Gamma_{m_2}(\gamma)$ is infeasible since the errors u_t are unobserved. Instead, the residuals from an initial NBLS regression, $\hat{u}_t$, may be used. Defining $\hat{F}_{x\hat{u}}(\gamma, l, m) = \frac{2\pi}{T}\sum_{j=l}^{m} \mathrm{Re}(I_{x\hat{u}}(\gamma, \lambda_j))$ and noting that $\hat{F}_{x\hat{u}}(\gamma, 1, m_0) = 0$ from the first order condition for $\hat{\beta}_{m_0}(\gamma)$, yields the feasible estimator

$$\hat{\Gamma}_{m_2}(\gamma) = \hat{F}_{xx}^{-1}(\gamma, m_0 + 1, m_2)\,\hat{F}_{x\hat{u}}(\gamma, m_0 + 1, m_2). \tag{22}$$

Thus, estimation of $K(\gamma)^{-1}H(\gamma)$ can be based on simply calculating the coefficient estimator in an auxiliary NBLS regression of the (differenced) residuals from the initial NBLS regression on the same set of regressors, $\Delta^\gamma x_t$, i.e. on NBLS estimation of the auxiliary regression

$$\Delta^\gamma \hat{u}_t = \kappa + \Gamma'\Delta^\gamma x_t + \Delta^\gamma v_t, \quad t = 1, ..., T. \tag{23}$$

The discussion of the representations (4), (6), and (11) suggest that

$$\tilde{F}_{qr}(\gamma, k, l) = \frac{2\pi}{T}\sum_{j=k}^{l} \mathrm{Re}(e^{i\lambda_j(d_q - d_r)/2} I_{qr}(\gamma, \lambda_j)), \quad 0 \le k \le l \le T - 1,$$

should more precisely approximate the integrated co-spectrum $F_{qr}(\gamma, \lambda)$, c.f. Assumption 1. Thus we also consider the estimator

$$\tilde{\Gamma}_{m_2}(\gamma) = \tilde{F}_{xx}^{-1}(\gamma, m_0 + 1, m_2)\,\tilde{F}_{x\hat{u}}(\gamma, m_0 + 1, m_2). \tag{24}$$

For the estimation of the bias term we need the following condition on the bandwidth m_2.

Assumption 7 *The bandwidth parameter $m_2 = m_2(T)$ satisfies*

$$\frac{m_0}{m_2} + \frac{m_2}{T} \to 0 \text{ as } T \to \infty,$$

where m_0 is the bandwidth from Assumption 5.

The first term in Assumption 7 ensures that (22) is based on an increasing number of periodogram ordinates, $m_2 - m_0$. The second term ensures that estimation is conducted in a neighborhood of the origin, which is sufficient for consistent NBLS estimation. We can now state the following result regarding the estimation of the NBLS bias term.

Theorem 3 *Let Assumptions 1-5 and 7 be satisfied and assume that $\hat{\Gamma}_{m_2}(\gamma)$ in (22) and $\tilde{\Gamma}_{m_2}(\gamma)$ in (24) are based on residuals $\hat{u}_t = y_t - \hat{\beta}_{m_0}(\gamma)'x_t$, where $\hat{\beta}_{m_0}(\gamma)$ is the NBLS estimator (9). Then, as $T \to \infty$,*

$$\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\hat{\Gamma}_{m_2}(\gamma) - K(\gamma)^{-1}H(\gamma) = O_P\left(\left(\frac{m_0}{m_2}\right)^{\delta_{\min}} + m_0^{-1/2}(\log T)^{-1} + \left(\frac{m_2}{T}\right)^{\min(1,\phi)}\right) \xrightarrow{P} 0,$$

$$\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\tilde{\Gamma}_{m_2}(\gamma) - K(\gamma)^{-1}H(\gamma) = O_P\left(\left(\frac{m_0}{m_2}\right)^{\delta_{\min}} + m_0^{-1/2}(\log T)^{-1} + \left(\frac{m_2}{T}\right)^{\phi}\right) \xrightarrow{P} 0.$$

Proof. See Appendix A.3. ■

This result implies that $\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\hat{\Gamma}_{m_2}(\gamma)$ and $\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\tilde{\Gamma}_{m_2}(\gamma)$ based on residuals are both consistent estimators of $K(\gamma)^{-1}H(\gamma)$. The theorem also implies, in conjunction with Theorem 2, that the bias $\lambda_{m_0}^{-d_u}\Lambda_{m_0}K(\gamma)^{-1}H(\gamma)$ of the NBLS estimator in Theorem 1 can be consistently estimated. It is even possible, based on Theorems 2 and 3, to obtain a rate result for the bias estimator, which we shall apply in the derivation of the fully modified estimator.

The FMNBLS estimator is based on a new bandwidth parameter $m_3 = m_3(T)$. In particular,

$$\tilde{\beta}_{m_3}(\gamma) = \hat{\beta}_{m_3}(\gamma) - \lambda_{m_3}^{-\hat{d}_u}\hat{\Lambda}_{m_3}\lambda_{m_2}^{\hat{d}_u}\hat{\Lambda}_{m_2}^{-1}\tilde{\Gamma}_{m_2}(\gamma), \tag{25}$$

where $\hat{\Lambda}_m = \mathrm{diag}(\lambda_m^{\hat{d}_{x,1}}, \ldots, \lambda_m^{\hat{d}_{x,p-1}})$. That is, the fully modified estimator $\tilde{\beta}_{m_3}(\gamma)$ is simply the NBLS estimator $\hat{\beta}_{m_3}(\gamma)$ corrected for the asymptotic bias. All the estimators of the integration orders are based on the bandwidth m_1. The bias correction term $\tilde{\Gamma}_{m_2}(\gamma)$ is estimated using bandwidth m_2 for (24) and bandwidth m_0 is needed to obtain the residuals upon which both (24) and $\hat{d}_u$ are based. We could also have used $\hat{\Gamma}_{m_2}(\gamma)$ in (25), but Theorem 3 shows that $\tilde{\Gamma}_{m_2}(\gamma)$ converges at a faster rate than $\hat{\Gamma}_{m_2}(\gamma)$. Note that in Theorem 3 the estimator of $K(\gamma)^{-1}H(\gamma)$ is based on periodograms integrated over $\lambda_{m_0+1}, \ldots, \lambda_{m_2}$ and therefore truncates the first m_0 Fourier frequencies, which may introduce variance inflation in finite samples. For example, Hurvich, Deo & Brodsky (1998) report Monte Carlo variance inflation from trimming the lowest frequencies in log-periodogram regression, even though theoretically trimming the lowest frequencies has no detrimental effect. However, as noted above, we cannot use the lowest m_0 frequencies due to the first order condition for the initial NBLS estimator. This differs from the fully modified estimator in Phillips & Hansen (1990), which uses the frequencies closest to the origin to estimate the bias term.

For the bandwidth $m_3 = m_3(T)$ of the FMNBLS estimator, we need the following condition.

Assumption 8 *The bandwidth parameter $m_3 = m_3(T)$ satisfies*

$$\frac{1}{m_3} + \frac{m_3^{1+2\min(1,\phi)}}{T^{2\min(1,\phi)}} + m_3\left(\frac{m_0}{m_2}\right)^{2\delta_{\min}} + m_3\left(\frac{m_2}{T}\right)^{2\phi} + (\log T)^{-2}\frac{m_3}{m_0} + (\log T)^2(\log m_1)^2 m_3\left(\frac{m_0}{m_1}\right)^{2\delta_{\min}} \to 0$$

as $T \to \infty$, where m_0, m_1, and m_2 are the bandwidth parameters from Assumptions 5-7, and ϕ is the smoothness parameter from Assumption 1.

The condition on m_3 is in some ways complicated and in others quite mild and simple. The first two terms state that m_3 has to satisfy the NBLS assumption for the bandwidth, c.f. Assumption 5. At the same time, m_3 must diverge to infinity at a rate no faster than that of m_0 (fifth term on the left-hand side) and at a slower rate than m_1 and m_2 (sixth and third terms on the left-hand side). Note that if m_1 and m_2 diverge to infinity at much faster rates than m_0 and the cointegrating strength, $\delta_{\min}$, is large, Assumption 8 is less restrictive. In fact, Assumption 8 is simple and easily satisfied because it is always feasible to choose $m_3 = m_0$, in which case there is no need to obtain a new NBLS estimator upon which to base the FMNBLS estimator (25). In that case the condition simplifies significantly, and in particular the relevant assumption then becomes

$$\frac{m_0^{1+2\delta_{\min}}}{m_2^{2\delta_{\min}}} + m_0\left(\frac{m_2}{T}\right)^{2\phi} + (\log T)^2(\log m_1)^2\frac{m_0^{1+2\delta_{\min}}}{m_1^{2\delta_{\min}}} \to 0 \text{ as } T \to \infty, \tag{26}$$

in addition to Assumptions 5-7 already placed on m_0, m_1, and m_2. To illustrate the restriction placed on the bandwidths by (26), suppose $\phi \geq 1$ and that we are in the empirically relevant (see Section 5 below) situation $\delta_{\min} = 0.4$. Then choosing $m_0 = m_3 = T^{0.3}$ is feasible if at the same time $m_1 = T^{0.675+\psi_1}$ and $m_2 = T^{0.675+\psi_2}$ for any $\psi_1 > 0$ and $\psi_2 \in (0, 0.175)$. On the other hand, if $m_1 = m_2 = T^{0.8}$ then choosing $m_0 = m_3 = T^{32/90-\psi_0}$ for any $\psi_0 > 0$ is feasible which is only slightly restrictive in light of Assumption 5 on m_0. Also note that it is in fact feasible in some cases to choose m_2 to diverge faster than $T^{0.8}$, which is even faster than the rate allowed for asymptotically normal NBLS estimation, c.f. Assumption 5.

In any case, the rate of convergence of $\tilde{\beta}_{m_3}(\gamma)$ in the following theorem is mostly affected by the cointegration strength $\delta_{\min}$ and not so much by the choice of $m_0 = m_3$. For example, if $\delta_{\min} = 0.4$ and $m_0 = m_3 = T^{0.3}$, the rate of convergence of $\tilde{\beta}_{m_3}(\gamma)$ in (27) is $T^{0.43}$ which is close to the usual $\sqrt{T}$-convergence in spite of the low bandwidth rate for m_3. In general, when $m_0 = m_3 = T^{\zeta}$, the rate of convergence of $\tilde{\beta}_{m_3}(\gamma)$ is $T^{\zeta(0.5-\delta_{\min})+\delta_{\min}}$. Therefore, for any ζ, when $\delta_{\min} \to 1/2$ the rate of convergence of $\tilde{\beta}_{m_3}(\gamma)$ approaches $\sqrt{T}$, which is the best rate attainable for fully parametric estimators based on complete and correct specification of the spectral density at all frequencies.

Theorem 4 *Let Assumptions 1-7 as well as either Assumption 8 or (26) be satisfied and let $\tilde{\beta}_{m_3}(\gamma)$ be the FMNBLS estimator (25). Then*

$$\sqrt{m_3}\lambda_{m_3}^{d_u}\Lambda_{m_3}^{-1}(\tilde{\beta}_{m_3}(\gamma)-\beta) \xrightarrow{D} N\left(0, K(\gamma)^{-1}J(\gamma)K(\gamma)^{-1}\right) \text{ as } T\to\infty, \tag{27}$$

where $K(\gamma)$ and $J(\gamma)$ are defined in Theorem 1.

Proof. See Appendix A.4. ■

Finally, we prove in the next theorem that a feasible version of the GLS-version of the FMNBLS estimator can be implemented in the weak fractional cointegration model using an estimate of the memory of the error term, e.g. from Theorem 2. That is, we show that $\tilde{\beta}_{m_3}(\hat{d}_u)$ has the same asymptotic distribution as $\tilde{\beta}_{m_3}(d_u)$, as long as $\hat{d}_u$ satisfies Assumption 6.

Theorem 5 *Let Assumptions 1, 3-7, as well as either Assumption 8 or (26) be satisfied, let $\tilde{\beta}_{m_3}(\gamma)$ be the FMNBLS estimator (25), and suppose the memory parameters satisfy (13). Then*

$$\sqrt{m_3}\lambda_{m_3}^{d_u}\Lambda_{m_3}^{-1}(\tilde{\beta}_{m_3}(\hat{d}_u)-\beta) \xrightarrow{D} N\left(0, K(d_u)^{-1}J(d_u)K(d_u)^{-1}\right) \text{ as } T\to\infty, \tag{28}$$

where $K(d_u)$ and $J(d_u)$ are defined in Theorem 1.

Proof. See Appendix A.5. ■

The results in Theorems 4 and 5 demonstrate that it is possible to obtain an asymptotically unbiased estimator of the cointegration vector in the weak fractional cointegration model (2), where the memory parameters satisfy (13); even in the presence of long-run coherence. The feasible estimator $\tilde{\beta}_{m_3}(\hat{d}_u)$ in Theorem 5 has the minimum variance among the estimators $\tilde{\beta}_{m_3}(\gamma)$ in Theorem 4 for $\gamma \in [0, d_u]$, c.f. the discussion following Theorem 1 in Section 2 above.

In the stationary case, Theorems 4 and 5 prove that it is possible to consistently estimate (with a centered asymptotic distribution) the relation between stationary time series even when the regressors and the errors are correlated at any frequency. A necessary condition is that the time series in question are fractionally cointegrated. Results similar to Theorems 4 and 5 are obtained by Hualde & Robinson (2010) who derive the asymptotic distribution theory for a related inverse spectral density weighted estimator, see also Nielsen (2005). In a different setup, Robinson (2008) developed joint multiple local Whittle (MLW) estimation of the memory parameters, the cointegration coefficient, and a phase parameter in a bivariate stationary fractionally cointegrated system. The MLW estimator of β also has a centered asymptotic distribution and converges at the same rate as our FMNBLS estimator. The multivariate method enjoys the usual advantages of a systems approach, but being based on numerical optimization of a multiparameter objective function it is computationally more demanding than our regression approach and the objective function may have multiple local optima. Finite sample performance of the MLW estimator of β and our FMNBLS estimator is compared in Section 4.

Compared to the NBLS estimator of Theorem 1, the fully modified estimator incurs no asymptotic variance inflation, only bias correction. Indeed, the FMNBLS estimator enjoys a faster rate of convergence than the NBLS estimator in the general case with non-zero coherence between the regressors and the errors at the origin. In particular, in the notation of the example following Theorem 1, the asymptotic mean squared error of the two estimators are related as

$$AMSE(\hat{\beta}_{m_3}(\gamma)) = m_3\lambda_{m_3}^{2d_u-2d_x}E(\hat{\beta}_{m_3}(\gamma)-\beta)^2 = \omega(\gamma)^2 + m_3\eta(\gamma)^2 = AMSE(\tilde{\beta}_{m_3}(\gamma)) + m_3\eta(\gamma)^2.$$

Thus, FMNBLS with the asymptotic distribution theory of Theorems 4 and 5 constitutes a much more useful inferential tool for the weak fractional cointegration model than the NBLS estimator, which is commonly used in previous work and applied especially in financial economics. Furthermore, Theorem 5 shows that the FMNBLS estimator is in fact applicable in the more general weak fractional cointegration model, and not just in the stationary cointegration case.

Consistent estimation of the parameters appearing in the variance of the limiting distribution in (27) can be based on Theorem 2 in conjunction with the estimator

$$\hat{G}_{ab}\left(\beta(\gamma), d\right) = \frac{1}{m_2} \sum_{j=1}^{m_2} \mathrm{Re}\left(\lambda_j^{d_{x,a}+d_{x,b}-2\gamma} e^{i(\lambda_j-\pi)(d_{x,a}-d_{x,b})/2} I_{ab}\left(\lambda_j\right)\right),$$

where $d = (d_{x,1}, ..., d_{x,p-1}, d_u)$ and $I_{ab}(\lambda_j)$ is the (a, b)'th element of $I_{ww}(\lambda_j) = I_{ww}(0, \lambda_j)$; the periodogram matrix of $w_t = (\Delta^{\gamma} x_t', \Delta^{\gamma} u_t)'$. Note that β enters in $I_{ab}(\gamma, \lambda_j)$ if $a = p$ and/or $b = p$. Specifically, if $\tilde{d}_u$ is the local Whittle estimator of d_u based on $\tilde{u}_t$ and $\tilde{I}(\tilde{d}_u, \lambda_j)$ is the periodogram matrix of $(\Delta^{\tilde{d}_u} x_t, \Delta^{\tilde{d}_u} \tilde{u}_t)$, where $\tilde{u}_t$ denotes FMNBLS residuals $\tilde{u}_t = y_t - \tilde{\beta}_{m_3}(\hat{d}_u)' x_t$, we have

$$\hat{G}_{ab}(\tilde{\beta}_{m_3}(\hat{d}_u), \tilde{d}) = \frac{1}{m_2} \sum_{j=1}^{m_2} \mathrm{Re}\left(\lambda_j^{\hat{d}_{x,a}+\hat{d}_{x,b}-2\tilde{d}_u} e^{i(\lambda_j-\pi)(\hat{d}_{x,a}-\hat{d}_{x,b})/2} \tilde{I}_{ab}(\tilde{d}_u, \lambda_j)\right) \xrightarrow{P} G_{ab}$$

as $T \to \infty$. The proof of this statement is omitted since it follows as in Propositions 2 and 3 of Robinson & Yajima (2002) by noting that $\tilde{\beta}_{a,m_3}(\hat{d}_u) - \beta_a = O_P(m_3^{-1/2} \lambda_{m_3}^{d_{x,a}-d_u})$.[4]

4 Simulation Evidence

In this section we investigate the finite sample behavior of the GLS-versio5n of the FMNBLS estimator $\tilde{\beta}_{m_3}(\hat{d}_u)$ introduced in Theorem 5 above and compare with the performance of the NBLS estimator $\hat{\beta}_{m_3}(0)$ and the MLW estimator of Robinson (2008).[5] We consider the following three two-dimensional generating mechanisms for x_t and u_t in the cointegrating relation (2),

$$\begin{array}{lll} \text{Model A} & : & x_t = (1-L)^{-d_x} \varepsilon_{1t}, \quad u_t = (1-L)^{-d_u} \varepsilon_{2t}, \\ \text{Model B} & : & x_t = (1-L)^{-d_x} v_{1t}, \quad u_t = (1-L)^{-d_u} \varepsilon_{2t}, \quad v_{1t} = a_1 v_{1,t-1} + \varepsilon_{1t}, \\ \text{Model C} & : & x_t = (1-L)^{-d_x} \varepsilon_{1t}, \quad u_t = (1-L)^{-d_u} v_{2t}, \quad v_{2t} = a_2 v_{2,t-1} + \varepsilon_{2t}, \end{array}$$

where $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}]'$ is independently and identically $N(0, \Omega)$ distributed with

$$\Omega = \left[\begin{array}{cc} \xi & \rho\xi^{1/2} \\ \rho\xi^{1/2} & 1 \end{array} \right].$$

Thus, $\xi = var\left(\varepsilon_{1t}\right) / var\left(\varepsilon_{2t}\right)$ is the signal-to-noise ratio and $\rho = corr(\varepsilon_{1t}, \varepsilon_{2t})$ is the contemporaneous correlation between the innovations ε_{1t} and ε_{2t}.

Based on the pair (x_t, u_t) we generate y_t from (2) with $\alpha = 0$ and $\beta = 1$. For all the simulations we generate the data with $(d_x, d_u) = (0.4, 0)$ which is close to what is expected in many practical situations concerning e.g. financial volatility series. This choice is made to facilitate comparison with the MLW estimator, and is also supported by the empirical applications below where we find estimates very close to these values in almost all cases. Unreported simulations reveal that the bias in NBLS is more severe when the integration orders are closer, e.g. $(d_x, d_u) = (0.3, 0.1)$, which also reduces the effectiveness of the bias correction procedure. However, the bias reduction in FMNBLS relative to NBLS remains noteworthy in that case, and for larger sample sizes the bias reduction works as well as with $(d_x, d_u) = (0.4, 0)$.

Models A, B, and C satisfy all the assumptions of the model, and are increasing in complexity

[4] Also note that, as in Theorem 2, local Whittle estimation of the integration order of the errors based on FMNBLS residuals is consistent and, if $m_0 = m_3$, then $\tilde{d}_u - d_u = O_P(m_0^{-1/2} (\log m_1) (m_0/m_1)^{\delta_{\min}})$, which converges faster than when based on NBLS residuals.

[5] We thank the editor and an anonymous referee for suggesting the comparison with the MLW estimator.

with Model A having no short-run dynamics whereas Models B and C include short-run dynamics. Model B adds short-run dynamics to the regressor and thus disturbs the signal due to the contamination of the low frequencies of x_t from the higher frequencies which are dominated by the short-run dynamics. In Model C short-run dynamics is present in u_t instead of x_t. Note that

$$G = \begin{bmatrix} \xi(1-a_1)^{-2} & \rho\xi^{1/2}(1-a_1)^{-1}(1-a_2)^{-1} \\ \rho\xi^{1/2}(1-a_1)^{-1}(1-a_2)^{-1} & (1-a_2)^{-2} \end{bmatrix} \quad (29)$$

such that when $\rho \neq 0$ the G matrix is not diagonal and the distribution theory for NBLS from Christensen & Nielsen (2006) no longer applies, see Theorem 1. However, the NBLS estimator is still consistent when $\rho \neq 0$. On the other hand, FMNBLS should be able to handle the presence of the long-run endogeneity that is due to $\rho \neq 0$, as shown in Theorems 4 and 5 above.

We also consider the three-dimensional generating mechanism

$$\text{Model D:} \quad x_{1t} = (1-L)^{-d_{x,1}} \varepsilon_{1t}, \quad x_{2t} = (1-L)^{-d_{x,2}} \varepsilon_{2t}, \quad u_t = (1-L)^{-d_u} \varepsilon_{3t},$$

where y_t is generated by (2) with $\alpha = 1$ and $\beta = [1,0]'$, $d_u = 0$, and $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}]'$ is independently and identically $N(0,\Omega)$ distributed with

$$\Omega = \begin{bmatrix} 2 & 0 & -0.75 \\ 0 & 2 & -0.75 \\ -0.75 & -0.75 & 1 \end{bmatrix}.$$

Note that in Model D the cointegrating regression (2) is $y_t = x_{1t} + u_t$, i.e., $x_{1t} \in I(d_{x,1}), x_{2t} \in I(d_{x,2}), u_t \in I(0)$, and $y_t \in I(d_{x,1})$. Hence, this is a three-dimensional model where the integration orders of the regressors are not necessarily the same, but all assumptions are satisfied because one of the regressors is not included in the DGP for y_t; in particular there is no cointegration among the regressors. The model illustrates a situation where one of the included regressor variables is in fact not part of the cointegrating regression, and demonstrates how the estimation of the associated coefficient (with true value equal to zero) depends on the parameters of the model, in particular on the memory parameters, bandwidth parameters, and sample size.

For each model we use $10,000$ replications for sample sizes $T = 128$ and $T = 512$, which are close to what is found in practical applications, see also the following section, although many applications in finance will have much larger sample sizes. The bandwidth parameters chosen for the simulation study are $m_i = \lfloor T^{\psi_i} \rfloor, i = 0,1,2,3$, where $\psi_0 \in \{0.4, 0.5\}$, $\psi_1 \in \{0.6, 0.7\}$, $\psi_2 = 0.8$, $\psi_3 = \psi_0$, and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Tables 1-3 present the Monte Carlo bias and root mean squared error (RMSE) results for Models A-C. As expected from (15) and (29), we find that changing the sign of the contemporaneous correlation ρ only causes the bias to change sign but does not change the size of the bias or RMSE, so we only report results for $\rho \leq 0$. For comparison, we also report the corresponding results for the MLW estimator of Robinson (2008) with bandwidth m_1 and using the NBLS, d_x, and d_u estimates also applied in FMNBLS as starting values, see Robinson (2008, Remark 3). For the MLW estimator the phase parameter is set as $(d_x - d_u)\pi/2$, i.e., fractional integration, and the MLW objective function is optimized over the three-dimensional parameter (β, d_x, d_u) by the BFGS algorithm and terminated when the convergence criterion $\epsilon = 10^{-6}$ is satisfied or after 100 iterations.[6]

Table 1 presents the results for Model A. A general finding is that increasing the signal-to-noise ratio ξ from 1 to 2, reduces the bias of NBLS and also improves the bias-reducing ability of the FMNBLS procedure. This is due to the fact that the contemporaneous covariance between ε_{1t} and

[6] In the case of non-convergence after 100 iterations the replication in question was not included in the calculation of bias and RMSE for the MLW estimator. Increasing the number of iterations required before termination of the numerical optimization substantially worsens the results for the MLW estimator.

Table 1: Simulation Results for Model A

			$\rho = -0.75$						$\rho = 0$					
	Bandwidths		NBLS		FMNBLS		MLW		NBLS		FMNBLS		MLW	
ξ	m_0	m_1	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $T = 128$														
1	$\lfloor T^{0.4} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.248	0.267	-0.053	0.160	0.212	0.943	-0.000	0.136	-0.001	0.211	0.004	0.927
		$\lfloor T^{0.7} \rfloor$			-0.056	0.156	0.195	0.772			-0.000	0.205	-0.006	0.576
	$\lfloor T^{0.5} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.296	0.309	-0.030	0.146	0.185	1.118	0.001	0.114	0.003	0.198	0.007	0.925
		$\lfloor T^{0.7} \rfloor$			-0.032	0.145	0.193	0.810			0.002	0.193	0.001	0.349
2	$\lfloor T^{0.4} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.175	0.188	-0.037	0.114	0.157	0.692	0.001	0.097	0.002	0.151	-0.001	0.715
		$\lfloor T^{0.7} \rfloor$			-0.040	0.111	0.137	0.587			0.001	0.147	0.000	0.384
	$\lfloor T^{0.5} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.209	0.218	-0.021	0.103	0.141	0.727	0.001	0.081	0.001	0.139	-0.002	0.514
		$\lfloor T^{0.7} \rfloor$			-0.023	0.102	0.136	0.579			0.001	0.136	0.004	0.527
Panel B: $T = 512$														
1	$\lfloor T^{0.4} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.169	0.176	-0.022	0.072	0.059	0.258	0.000	0.064	0.000	0.092	0.000	0.208
		$\lfloor T^{0.7} \rfloor$			-0.024	0.071	0.045	0.137			0.001	0.091	0.001	0.072
	$\lfloor T^{0.5} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.203	0.208	0.003	0.066	0.058	0.231	0.000	0.057	0.001	0.090	-0.002	0.154
		$\lfloor T^{0.7} \rfloor$			0.000	0.066	0.045	0.124			0.001	0.089	0.001	0.072
2	$\lfloor T^{0.4} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.120	0.125	-0.016	0.051	0.036	0.149	-0.000	0.047	-0.000	0.067	-0.001	0.101
		$\lfloor T^{0.7} \rfloor$			-0.017	0.051	0.030	0.086			-0.000	0.066	-0.001	0.051
	$\lfloor T^{0.5} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.144	0.147	0.002	0.047	0.036	0.148	-0.000	0.040	-0.001	0.063	-0.001	0.103
		$\lfloor T^{0.7} \rfloor$			-0.000	0.047	0.030	0.086			-0.001	0.063	-0.001	0.051

Note: The simulations are based on 10,000 replications under the empirically relevant scenario $(d_x, d_u) = (0.4, 0)$, with bandwidths $m_2 = \lfloor T^{0.8} \rfloor$ and $m_3 = m_0$.

ε_{2t} is reduced when ξ increases from 1 to 2. Furthermore, estimating $K(\gamma)^{-1}H(\gamma)$ in (15) when it is in fact zero because $\rho = 0$ inflates the variance (and hence the RMSE) of FMNBLS relative to that of NBLS, but the fully modified procedure still yields unbiased estimates of β. For $\rho = -0.75$ (and $\rho = 0.75$), the FMNBLS procedure bias-corrects NBLS although this comes at the expense of an increase in the finite sample standard error of up to 50%. However, the RMSE of FMNBLS in that case is (much) lower than that of NBLS. For the larger sample size, $T = 512$, FMNBLS yields almost unbiased estimates for all bandwidths with RMSEs much smaller than those of NBLS, except when $\rho = 0$. Even though the bias of NBLS increases (and becomes fairly large) for larger m_0, the fully modified procedure is still able to correct this, and indeed the bias of FMNBLS is smaller when m_0 $(= m_3)$ is larger. Since there is no short-run dynamics, the choice of m_1 appears less important. The MLW estimator performs quite poorly compared to both NBLS and FMNBLS, especially for $T = 128$. Interestingly, the sign of the bias of MLW is opposite that of NBLS.

Table 2 presents the simulation results for Model B with autoregressive coefficients $a_1 = -1/2$ or $a_1 = 1/2$.[7] Now, (4) is a worse approximation to (14) when moving only a short distance away from the origin, due to the contamination from higher frequencies (short-run dynamics), and we therefore expect the bias of NBLS (and possibly also of FMNBLS) to be larger than for the case of no short-run dynamics. Interestingly, for Model B it appears that the biases and RMSEs of NBLS and FMNBLS are lower than for Model A when $a_1 = 1/2$ and higher than for Model A when $a_1 = -1/2$. In Model B the MLW estimator is sometimes equal to or better than FMNBLS in terms of RMSE in some cases with $m_1 = \lfloor T^{0.7} \rfloor$. In general, though, it does not perform as well as FMNBLS, and in some cases it even has convergence problems marked by asterisks in the table.

Next, we turn to Model C. Table 3 presents the simulation results, which are quite different for $a_2 = -1/2$ and $a_2 = 1/2$. Compared to the results of Model A, the NBLS estimator is actually less biased in this setup when $a_2 = -1/2$. This suggests that negative autocorrelation in u_t offsets some of the bias in the NBLS estimator introduced by contemporaneous covariance between x_t and

[7]For Models B and C we report the simulation results for $\xi = 2$ only. The results for $\xi = 1$ are qualitatively similar, see also Table 1.

Table 2: Simulation Results for Model B

			$\rho=-0.75$						$\rho=0$					
	Bandwidths		NBLS		FMNBLS		MLW		NBLS		FMNBLS		MLW	
a_1	m_0	m_1	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $T=128$														
$-1/2$	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.271	0.291	-0.134	0.211	0.297	1.345*	-0.000	0.143	-0.001	0.225	0.008	1.066
		$\lfloor T^{0.7}\rfloor$			-0.124	0.204	0.342	1.528*			-0.000	0.226	0.001	0.765
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.328	0.342	-0.188	0.246	0.243	1.203*	0.001	0.120	0.002	0.205	0.019	1.131
		$\lfloor T^{0.7}\rfloor$			-0.178	0.241	0.333	1.546*			0.002	0.206	0.008	0.855
1/2	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.078	0.086	-0.017	0.053	-0.007	0.144	0.000	0.050	0.001	0.071	0.002	0.181
		$\lfloor T^{0.7}\rfloor$			-0.027	0.054	-0.027	0.058			0.000	0.065	-0.000	0.050
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.089	0.095	0.037	0.066	-0.006	0.127	0.000	0.043	0.001	0.072	0.000	0.145
		$\lfloor T^{0.7}\rfloor$			0.024	0.054	-0.027	0.058			0.000	0.066	-0.000	0.050
Panel B: $T=512$														
$-1/2$	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.182	0.190	-0.039	0.083	0.090	0.354	0.000	0.068	0.000	0.098	-0.001	0.224
		$\lfloor T^{0.7}\rfloor$			-0.034	0.080	0.086	0.235			0.001	0.099	0.001	0.080
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.221	0.227	-0.035	0.079	0.092	0.376	0.000	0.060	0.001	0.094	-0.001	0.206
		$\lfloor T^{0.7}\rfloor$			-0.030	0.078	0.086	0.235			0.001	0.095	0.001	0.080
1/2	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.057	0.059	-0.016	0.028	-0.005	0.032	-0.000	0.023	-0.000	0.033	0.000	0.030
		$\lfloor T^{0.7}\rfloor$			-0.023	0.032	-0.019	0.028			-0.000	0.030	-0.000	0.021
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.066	0.068	0.003	0.024	-0.005	0.032	-0.000	0.020	-0.000	0.032	0.000	0.030
		$\lfloor T^{0.7}\rfloor$			-0.006	0.023	-0.019	0.028			-0.000	0.030	-0.000	0.021

Note: The simulations are based on 10,000 replications under the empirically relevant scenario $(d_x, d_u) = (0.4, 0)$, with bandwidths $m_2 = \lfloor T^{0.8}\rfloor$ and $m_3 = m_0$. An asterisk indicates that MLW did not converge for 5-10% of the replications.

u_t, see (29). Consequently, the FMNBLS procedure works very well and generally yields almost unbiased estimates and also large reductions in RMSEs when $a_2 = -1/2$. When $a_2 = 1/2$, Model C results in extremely large biases for NBLS. For the small sample size the NBLS biases when $\rho = -0.75$ range from 0.37 to 0.43 in absolute value, and for the large sample size the biases are still about two-thirds of the bias for the smaller sample size. For the small sample size, this yields an imprecise estimate of $K(\gamma)^{-1}H(\gamma)$, and as a result FMNBLS is still biased, although the fully modified procedure generally still manages to reduce the bias quite considerably and has a smaller RMSE than NBLS. For $T = 512$ FMNBLS has low bias and the RMSE is again (much) smaller than that of NBLS. The performance of MLW is similar to that in Table 2 with convergence problems for the small sample size when $a_2 = 1/2$, and performance equal to or better than that of FMNBLS only when $a_2 = -1/2$ and at the same time $m_1 = \lfloor T^{0.7}\rfloor$ and $T = 512$.

Finally, we turn to Model D with two regressors with memory parameters $(d_{x,1}, d_{x,2})$. In this case, as in Model A, the bandwidth m_1 has no significant effect since there is no short-run dynamics. Increasing the bandwidth m_0 appears to worsen the results for NBLS but improve those for FMNBLS, both in terms of bias and RMSE. The most interesting aspect of Model D is the comparison across different values of $(d_{x,1}, d_{x,2})$. In this respect we find for both NBLS and FMNBLS that bias and RMSE are higher for the coefficient on the variable with the lowest memory parameter. This finding is in line with theory and with unreported simulations of Models A-C with $(d_x, d_u) = (0.3, 0.1)$. The results appear symmetric with respect to the variable that is included in the cointegrating regression (x_{1t}) and that which is excluded (x_{2t}).

Overall, the simulations clearly demonstrate the superiority (in terms of both bias and RMSE) of the fully modified estimator relative to NBLS in the presence of non-zero long-run coherence between the regressor and the error. In all models, the bias-reduction of FMNBLS relative to NBLS is considerable, and for the larger sample size the bias practically disappears. The cost of this bias correction is an increase in the finite sample standard deviation of approximately 30-50%

Table 3: Simulation Results for Model C

	Bandwidths		$\rho = -0.75$						$\rho = 0$					
			NBLS		FMNBLS		MLW		NBLS		FMNBLS		MLW	
a_2	m_0	m_1	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
Panel A: $T = 128$														
$-1/2$	$\lfloor T^{0.4} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.113	0.122	-0.003	0.072	0.055	0.349	-0.000	0.064	-0.000	0.100	0.005	0.451
		$\lfloor T^{0.7} \rfloor$			-0.010	0.069	0.012	0.145			-0.000	0.093	0.001	0.106
	$\lfloor T^{0.5} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.133	0.139	0.036	0.070	0.047	0.337	0.000	0.055	0.001	0.097	0.001	0.325
		$\lfloor T^{0.7} \rfloor$			0.032	0.074	0.012	0.154			0.001	0.091	0.002	0.142
$1/2$	$\lfloor T^{0.4} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.369	0.394	-0.293	0.384	0.198	1.928**	0.002	0.187	0.002	0.339	-0.048	2.479*
		$\lfloor T^{0.7} \rfloor$			-0.289	0.393	0.005	1.911**			0.000	0.366	-0.001	1.951*
	$\lfloor T^{0.5} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.430	0.446	-0.421	0.477	0.076	1.865**	0.001	0.152	0.001	0.281	-0.005	1.539
		$\lfloor T^{0.7} \rfloor$			-0.443	0.503	0.093	1.717**			-0.001	0.302	-0.001	1.958*
Panel B: $T = 512$														
$-1/2$	$\lfloor T^{0.4} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.079	0.082	-0.009	0.034	0.016	0.085	0.000	0.030	0.000	0.044	-0.001	0.109
		$\lfloor T^{0.7} \rfloor$			-0.012	0.034	0.003	0.040			0.000	0.042	0.000	0.032
	$\lfloor T^{0.5} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.093	0.096	0.007	0.032	0.016	0.085	0.000	0.027	0.000	0.043	-0.000	0.064
		$\lfloor T^{0.7} \rfloor$			0.005	0.032	0.003	0.040			0.000	0.042	0.000	0.032
$1/2$	$\lfloor T^{0.4} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.249	0.258	-0.123	0.158	0.239	0.794	-0.000	0.093	-0.000	0.139	-0.005	0.376
		$\lfloor T^{0.7} \rfloor$			-0.108	0.150	0.389	1.282			0.000	0.152	0.002	0.528
	$\lfloor T^{0.5} \rfloor$	$\lfloor T^{0.6} \rfloor$	-0.301	0.308	-0.177	0.204	0.231	0.735	-0.001	0.080	-0.001	0.128	-0.004	0.279
		$\lfloor T^{0.7} \rfloor$			-0.183	0.212	0.360	1.056			-0.002	0.139	-0.005	0.502

Note: The simulations are based on 10,000 replications under the empirically relevant scenario $(d_x, d_u) = (0.4, 0)$, with bandwidths $m_2 = \lfloor T^{0.8} \rfloor$ and $m_3 = m_0$. One and two asterisks indicate that MLW did not converge for 5-10% of the replications and 10-25% of the replications, respectively.

for the models considered here. However, the results indicate that this is more than off-set by the large bias reduction when $\rho \neq 0$ thus yielding reductions in the RMSE. The simulations also suggest that the GLS-version of the FMNBLS estimator is superior to the MLW estimator in many circumstances. This could possibly be due to the extra flexibility of the FMNBLS estimator from using separate bandwidths for estimation of the cointegration vector, the integration orders, and the asymptotic bias term.

5 Empirical Illustrations

We apply NBLS and FMNBLS to three different empirically relevant examples.[8]

5.1 The Implied-Realized Volatility Relation

Recent contributions by, e.g., Comte & Renault (1998), Bandi & Perron (2006), Christensen & Nielsen (2006), and Berger, Chaboud & Hjalmarsson (2009) including empirical evidence, have pointed towards viewing the predictive regression between implied volatility (IV) and realized volatility (RV) as one of stationary fractional cointegration. However, the possible existence of a volatility risk premium that is correlated with IV can bias the NBLS estimator in a regression of RV on IV, which ultimately can lead to a wrongful rejection of the long-run unbiasedness hypothesis, see Bandi & Perron (2006). Furthermore, the existence of an unobserved risk premium can also imply a negative intercept in the regression, and thus long-run unbiasedness is typically upheld if the cointegrating coefficient is $\beta = 1$ regardless of the presence of the intercept.

We sample S&P500 index options (SPX) data from the Berkeley options data base covering the period January 1988 through December 1995 and calculate $T = 412$ weekly Black-Scholes implied volatilities and the corresponding S&P500 realized volatilities, see Christensen & Nielsen (2006) for

[8] Henry & Zaffaroni (2003) survey empirical applications of fractional integration and long memory in macroeconomics and financial economics.

Table 4: Simulation Results for Model D

(d_1,d_2)	Bandwidths m_0	m_1	NBLS Bias_1	Bias_2	RMSE_1	RMSE_2	FMNBLS Bias_1	Bias_2	RMSE_1	RMSE_2
Panel A: $T=128$										
$(0.25,0.40)$	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.207	-0.125	0.235	0.146	-0.118	-0.034	0.220	0.117
		$\lfloor T^{0.7}\rfloor$					-0.118	-0.034	0.216	0.115
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.233	-0.149	0.248	0.162	-0.126	-0.020	0.199	0.105
		$\lfloor T^{0.7}\rfloor$					-0.127	-0.021	0.197	0.104
$(0.40,0.25)$	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.124	-0.211	0.145	0.238	-0.033	-0.126	0.116	0.221
		$\lfloor T^{0.7}\rfloor$					-0.032	-0.125	0.113	0.217
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.147	-0.235	0.160	0.250	-0.016	-0.132	0.105	0.204
		$\lfloor T^{0.7}\rfloor$					-0.017	-0.132	0.103	0.201
$(0.40,0.40)$	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.127	-0.127	0.152	0.152	-0.036	-0.034	0.127	0.126
		$\lfloor T^{0.7}\rfloor$					-0.037	-0.036	0.124	0.125
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.151	-0.150	0.166	0.166	-0.021	-0.020	0.114	0.115
		$\lfloor T^{0.7}\rfloor$					-0.022	-0.021	0.112	0.113
Panel B: $T=512$										
$(0.25,0.40)$	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.165	-0.085	0.176	0.092	-0.081	-0.015	0.122	0.051
		$\lfloor T^{0.7}\rfloor$					-0.082	-0.015	0.121	0.050
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.187	-0.102	0.193	0.107	-0.087	-0.001	0.118	0.047
		$\lfloor T^{0.7}\rfloor$					-0.088	-0.002	0.118	0.047
$(0.40,0.25)$	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.086	-0.165	0.093	0.176	-0.015	-0.081	0.053	0.123
		$\lfloor T^{0.7}\rfloor$					-0.015	-0.081	0.052	0.122
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.103	-0.187	0.108	0.193	-0.002	-0.087	0.048	0.119
		$\lfloor T^{0.7}\rfloor$					-0.003	-0.087	0.048	0.118
$(0.40,0.40)$	$\lfloor T^{0.4}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.087	-0.086	0.095	0.095	-0.016	-0.015	0.057	0.057
		$\lfloor T^{0.7}\rfloor$					-0.016	-0.015	0.056	0.056
	$\lfloor T^{0.5}\rfloor$	$\lfloor T^{0.6}\rfloor$	-0.103	-0.103	0.109	0.109	-0.002	-0.002	0.052	0.052
		$\lfloor T^{0.7}\rfloor$					-0.003	-0.003	0.051	0.052

Note: The simulations are based on 10,000 replications with $d_3 = 0$ and bandwidths $m_2 = \lfloor T^{0.8}\rfloor$ and $m_3 = m_0$.

details. In particular, Christensen & Nielsen (2006) find that the log-volatilities are stationary, with insignificantly different long memory estimates, and that NBLS regression yields a cointegrating coefficient β ranging from 0.84 to 0.89 for different bandwidth choices.

Panel A of Table 5 shows the memory estimates for the two log-volatility series. As found by Christensen & Nielsen (2006), the series are stationary ($d < 1/2$) and exhibit long memory.

In Panel B of the table we show estimates (with asymptotic standard errors in parentheses) of the (stationary) fractional cointegration relation between the two log-volatility series, IV and RV, for a variety of bandwidth parameters: $m_0 = m_3 \in \{\lfloor T^{0.4}\rfloor, \lfloor T^{0.5}\rfloor\}$, $m_1 \in \{\lfloor T^{0.6}\rfloor, \lfloor T^{0.7}\rfloor, \lfloor T^{0.8}\rfloor\}$, and $m_2 = \lfloor T^{0.8}\rfloor$. The NBLS estimates are of course in line with the results of Christensen & Nielsen (2006), with the parameter of interest, β, estimated to be $0.81-0.84$. For $m_0 = \lfloor T^{0.5}\rfloor, m_1 = \lfloor T^{0.8}\rfloor$ it is significantly less than unity when applying the asymptotic distribution theory in Theorem 1. Note that in two cases $\hat{d}_x + \hat{d}_u \geq 1/2$ so that Theorem 1 does not apply to the NBLS estimator and the asymptotic standard error is denoted by $(-)$. The FMNBLS procedure corrects for the possible correlation between the regressor and the error term; those estimates are displayed in the final columns. We obtain point estimates of β that are now above unity, but insignificantly different from unity except when $m_1 = \lfloor T^{0.6}\rfloor$. Thus, our estimates generally tend to support the long-run unbiasedness hypothesis, $\beta = 1$. Finally, we notice that with $m_1 = \lfloor T^{0.8}\rfloor$ both the NBLS and FMNBLS estimates support an $I(d) - I(0)$ relation with d around $0.35 - 0.4$, although, c.f. Theorem 2, the usual asymptotic distribution may not apply for $\hat{d}_u$ and $\tilde{d}_u$ ($\hat{d}_u$ denotes the estimate based on NBLS residuals $\hat{u}_t$ and $\tilde{d}_u$ denotes the estimate based on FMNBLS residuals $\tilde{u}_t$).

Table 5: Implied-Realized Volatility Application

Panel A: Long Memory Estimates, $\hat{d}$

Bandwidth	Realized volatility $y_t = \ln \sigma_{RV,t}$	Implied volatility $x_t = \ln \sigma_{IV,t}$
$m_1 = \lfloor T^{0.6} \rfloor$	0.4476	0.4527
	(0.0822)	(0.0822)
$m_1 = \lfloor T^{0.7} \rfloor$	0.4162	0.3503
	(0.0606)	(0.0606)
$m_1 = \lfloor T^{0.8} \rfloor$	0.4180	0.2801
	(0.0449)	(0.0449)

Panel B: Cointegration Analysis

	NBLS			FMNBLS		
Bandwidths	$\hat{\alpha}_{m_3}(0)$	$\hat{\beta}_{m_3}(0)$	$\hat{d}_u$	$\tilde{\alpha}_{m_3}(\hat{d}_u)$	$\tilde{\beta}_{m_3}(\hat{d}_u)$	$\tilde{d}_u$
$m_0 = \lfloor T^{0.4} \rfloor, m_1 = \lfloor T^{0.6} \rfloor$	−0.9403	0.8364	0.1046	−0.0390	1.2792	0.1778
		(−)	(0.0822)		(0.1305)	(0.0822)
$m_0 = \lfloor T^{0.4} \rfloor, m_1 = \lfloor T^{0.7} \rfloor$	−0.9403	0.8364	0.0987	0.0289	1.3126	0.1341
		(0.1325)	(0.0606)		(0.1746)	(0.0606)
$m_0 = \lfloor T^{0.4} \rfloor, m_1 = \lfloor T^{0.8} \rfloor$	−0.9403	0.8364	0.0718	−0.2893	1.1562	0.0616
		(0.1227)	(0.0449)		(0.1554)	(0.0449)
$m_0 = \lfloor T^{0.5} \rfloor, m_1 = \lfloor T^{0.6} \rfloor$	−0.9990	0.8076	0.1145	−0.2018	1.1992	0.1525
		(−)	(0.0822)		(0.0976)	(0.0822)
$m_0 = \lfloor T^{0.5} \rfloor, m_1 = \lfloor T^{0.7} \rfloor$	−0.9990	0.8076	0.1079	−0.1359	1.2316	0.1181
		(0.1242)	(0.0606)		(0.1317)	(0.0606)
$m_0 = \lfloor T^{0.5} \rfloor, m_1 = \lfloor T^{0.8} \rfloor$	−0.9990	0.8076	0.0805	−0.3470	1.1279	0.0582
		(0.1044)	(0.0449)		(0.1264)	(0.0449)

Note: Panel A reports local Whittle estimates of the fractional integration orders as described in Robinson (1995*a*). Numbers in parentheses are asymptotic standard errors using $\sqrt{m_1}(\hat{d} - d) \xrightarrow{D} N(0, 1/4)$. Panel B reports NBLS and FMNBLS estimates with $m_2 = \lfloor T^{0.8} \rfloor$ and $m_3 = m_0$. The asymptotic standard errors for the NBLS and FMNBLS estimates are based on (15) and (27), respectively. Standard errors for $\hat{d}_u$ and $\tilde{d}_u$ are based on the same asymptotic distribution as $\hat{d}$, and should be used with caution, see Theorem 2.

5.2 Inflation Rate Harmonization in the European Union

We also examine consumer price indexes of France and Spain. Methods for calculating the consumer price index vary across different countries, which makes international comparison more difficult, and because of this we use the harmonized index for consumer prices (HICP) developed within the European Union based on a coordinated methodology.

Since the differentials between the inflation rates of individual member countries of the European Union are constrained, we expect that there exists a stable relationship between the inflation rates. Furthermore, based on evidence of long memory in inflation rates in Doornik & Ooms (2004) we expect that relationship to be one of stationary fractional cointegration. We calculate $T = 159$ monthly inflation rates based on the HICP of France and Spain. This data was obtained from Eurostat and covers the period January 1992 through April 2005.

Panel A of Table 6 shows that the memory estimates decrease as the bandwidth increases. This may be due to an added noise perturbation or, more likely, due to the distinct seasonal patterns in inflation series; possibly reflecting seasonal long memory, see Doornik & Ooms (2004). Instead of filtering this out by ad hoc procedures, we focus on the results for the lowest bandwidth, $m_1 = \lfloor T^{0.5} \rfloor$, which should be less sensitive to contamination from higher (e.g. seasonal) frequencies. For this bandwidth, the memory estimates for both inflation rates imply that the series are stationary.

Panel B of Table 6 again supports the notion of $I(d) - I(0)$ cointegration with d around 0.35. Here, the FMNBLS estimates are much higher than the NBLS estimates. In particular, the FMNBLS estimates of the cointegration coefficient are significantly higher than unity at the 1% level in both cases, implying that the long-run rate of inflation in Spain is higher than that in France (by over 80% according to the point estimates). In addition, the estimates of d for the residuals are

Table 6: Inflation Rate Harmonization Application

Panel A: Long Memory Estimates, $\hat{d}$		
	Spain	France
Bandwidth	$y_t = \pi_{S,t}$	$x_t = \pi_{F,t}$
$m_1 = \lfloor T^{0.5} \rfloor$	0.4007 (0.1443)	0.3048 (0.1443)
$m_1 = \lfloor T^{0.6} \rfloor$	0.0990 (0.1118)	−0.0690 (0.1118)
$m_1 = \lfloor T^{0.7} \rfloor$	−0.1847 (0.0857)	−0.1377 (0.0857)

Panel B: Cointegration Analysis						
	NBLS			FMNBLS		
Bandwidths	$\hat{\alpha}_{m_3}(0)$	$\hat{\beta}_{m_3}(0)$	$\hat{d}_u$	$\tilde{\alpha}_{m_3}(\hat{d}_u)$	$\tilde{\beta}_{m_3}(\hat{d}_u)$	$\tilde{d}_u$
$m_0 = \lfloor T^{0.3} \rfloor, m_1 = \lfloor T^{0.5} \rfloor$	0.0011	1.1395 (0.3139)	0.0852 (0.1443)	0.0001	1.8619 (0.2797)	0.0100 (0.1443)
$m_0 = \lfloor T^{0.4} \rfloor, m_1 = \lfloor T^{0.5} \rfloor$	0.0012	1.0577 (0.2965)	0.1048 (0.1443)	0.0011	1.8272 (0.2470)	0.0099 (0.1443)

Note: Panel A reports local Whittle estimates of the fractional integration orders as described in Robinson (1995*a*). Numbers in parentheses are asymptotic standard errors using $\sqrt{m_1}(\hat{d} - d) \xrightarrow{D} N(0, 1/4)$. Panel B reports NBLS and FMNBLS estimates with $m_2 = \lfloor T^{0.8} \rfloor$ and $m_3 = m_0$. The asymptotic standard errors for the NBLS and FMNBLS estimates are based on (15) and (27), respectively. Standard errors for $\hat{d}_u$ and $\tilde{d}_u$ are based on the same asymptotic distribution as $\hat{d}$, and should be used with caution, see Theorem 2.

lower for FMNBLS than for NBLS although all appear insignificantly different from zero (again, the usual asymptotic distribution may not apply, see Theorem 2).

5.3 Realized Volatility Relations

Finally, we analyze the relation between the realized volatility of the General Electric (GE) stock and those of the Dow Jones Industrial Average (DJIA) and NASDAQ 100 indexes. I.e., there are three variables in this application. The realized volatilities are monthly and are constructed based on daily returns calculated as the difference in log-open and log-close prices. The sample covers January 1990 to December 2008, i.e. $T = 228$.

Panel A of Table 7 shows that the memory estimates of the three realized volatilities are very similar and stable across bandwidths with point estimates around 0.4, except for the middle bandwidth where point estimates are higher and suggest nonstationarity. A test of the hypothesis that all memory parameters are equal, see Robinson & Yajima (2002, section 3), is insignificant at conventional levels for all bandwidth choices in the table. In Panel B of Table 7 we present cointegration rank statistics from Robinson & Yajima (2002) using bandwidth m_0 for rank statistics and m_1 to estimate memory parameters. In the remainder of the table we ignore bandwidth $m_1 = \lfloor T^{0.7} \rfloor$ to be able to apply their results. In particular, using the notation of Robinson & Yajima (2002, section 3), Panel B presents the eigenvalues of the correlation-type matrix P, and the value of the model determination function $L(u)$ using $v(T) = m_0^{-0.4}$. The rank can be determined by $\arg\min L(u)$, which suggests that the rank is one. Thus, we conclude that a regression approach is appropriate in this multivariate system.

In Panel C we report estimates of the stationary fractional cointegration relation between the realized volatilities of GE and the DJIA and NASDAQ indexes. Clearly the volatility of GE should be related to the broader market volatility, so it seems reasonable to assume that the volatility of GE enters in the cointegrating regression with a non-zero coefficient. Moreover, we are interested in analyzing how the volatility of GE depends on the volatilities of the two indexes, so we choose y_t to be the realized volatility of GE. From the results, it appears that the NBLS estimator underestimates the slope coefficient on DJIA (x_{1t}) in the cointegrating relation. Both the NBLS

Table 7: Realized Volatility Relations Application

Panel A: Long Memory Estimates, $\hat{d}$

Bandwidth	GE $y_t = \sigma^2_{GE,t}$	Dow Jones $x_{1t} = \sigma^2_{DJ,t}$	NASDAQ $x_{2t} = \sigma^2_{ND,t}$
$m_1 = \lfloor T^{0.6} \rfloor$	0.4350 (0.1000)	0.3526 (0.1000)	0.4383 (0.1000)
$m_1 = \lfloor T^{0.7} \rfloor$	0.5041 (0.0762)	0.4080 (0.0762)	0.5980 (0.0762)
$m_1 = \lfloor T^{0.8} \rfloor$	0.3958 (0.0585)	0.4277 (0.0585)	0.4026 (0.0585)

Panel B: Cointegration Rank Analysis

	Eigenvalues of P			$L(u)$		
Bandwidths	1	2	3	$u = 0$	$u = 1$	$u = 2$
$m_0 = \lfloor T^{0.4} \rfloor, m_1 = \lfloor T^{0.6} \rfloor$	2.3523	0.5889	0.0588	−1.6942	−2.0706	−1.9170
$m_0 = \lfloor T^{0.4} \rfloor, m_1 = \lfloor T^{0.8} \rfloor$	2.3522	0.5889	0.0588	−1.6942	−2.0706	−1.9170
$m_0 = \lfloor T^{0.5} \rfloor, m_1 = \lfloor T^{0.6} \rfloor$	2.4420	0.4764	0.0816	−1.9561	−2.2224	−2.0940
$m_0 = \lfloor T^{0.5} \rfloor, m_1 = \lfloor T^{0.8} \rfloor$	2.4419	0.4764	0.0816	−1.9561	−2.2224	−2.0940

Panel C: Cointegration Regression Analysis

	NBLS				FMNBLS			
Bandwidths	$\hat{\alpha}_{m_3}(0)$	$\hat{\beta}_{m_3}(0)$		$\hat{d}_u$	$\tilde{\alpha}_{m_3}(\hat{d}_u)$	$\tilde{\beta}_{m_3}(\hat{d}_u)$		$\tilde{d}_u$
$m_0 = \lfloor T^{0.4} \rfloor, m_1 = \lfloor T^{0.6} \rfloor$	0.0001	1.6478 (0.1321)	0.1825 (0.0211)	0.0192 (0.1000)	−0.0003	1.8591 (0.1327)	0.1828 (0.0182)	−0.0004 (0.1000)
$m_0 = \lfloor T^{0.4} \rfloor, m_1 = \lfloor T^{0.8} \rfloor$	0.0001	1.6478 (0.1318)	0.1825 (0.0332)	0.0409 (0.0585)	−0.0002	1.8349 (0.1281)	0.1837 (0.0348)	0.0561 (0.0585)
$m_0 = \lfloor T^{0.5} \rfloor, m_1 = \lfloor T^{0.6} \rfloor$	0.0003	1.4828 (0.1203)	0.2061 (0.0257)	0.0362 (0.1000)	−0.0001	1.6043 (0.1134)	0.2311 (0.0166)	−0.0023 (0.1000)
$m_0 = \lfloor T^{0.5} \rfloor, m_1 = \lfloor T^{0.8} \rfloor$	0.0003	1.4828 (0.1188)	0.2061 (0.0300)	0.0402 (0.0585)	−0.0001	1.5824 (0.1069)	0.2344 (0.0296)	0.0442 (0.0585)

Note: Panel A reports local Whittle estimates of the fractional integration orders as described in Robinson (1995*a*). Numbers in parentheses are asymptotic standard errors using $\sqrt{m_1}(\hat{d} - d) \xrightarrow{D} N(0, 1/4)$. Panel B reports rank statistics from Robinson & Yajima (2002) and Panel C reports NBLS and FMNBLS estimates with $m_2 = \lfloor T^{0.8} \rfloor$ and $m_3 = m_0$. The asymptotic standard errors for the NBLS and FMNBLS estimates are based on (15) and (27), respectively. Standard errors for $\hat{d}_u$ and $\tilde{d}_u$ are based on the same asymptotic distribution as $\hat{d}$, and should be used with caution, see Theorem 2.

and the FMNBLS results indicate that the volatility of GE most strongly follows that of the DJIA.

6 Concluding Remarks

We have considered estimation of the cointegration vector under weak fractional cointegration. A special case is the stationary fractional cointegration model which has found important application recently, especially in financial economics. Previous research has considered Robinson's (1994) semiparametric frequency domain narrow-band least squares (NBLS) estimator, for which a condition of non-coherence between regressors and errors at the zero frequency has sometimes been imposed, e.g. Christensen & Nielsen (2006). We have shown that in the absence of such condition, NBLS suffers from asymptotic bias although it remains consistent as proven by Robinson (1994). We also showed that the bias can be consistently estimated, and consequently we introduced a fully modified NBLS (FMNBLS) estimator which eliminates the bias but has the same asymptotic variance as NBLS. Indeed, FMNBLS enjoys a faster rate of convergence than NBLS in general.

We also conducted a simulation study of the proposed FMNBLS estimator, which clearly demonstrated the superiority with respect to bias of the fully modified estimator relative to NBLS in the presence of non-zero long-run coherence between regressors and errors. Although this comes at the cost of increased finite sample variance, FMNBLS is superior in terms of RMSE in simulations with long-run coherence between regressors and errors. The simulations also indicate that the bias correction method works well in the presence of short-run dynamics in regressors and errors. The empirical relevance of our methodology was demonstrated through a series of brief empirical illustrations, all of which support the notion of a stationary fractional cointegration relation.

Appendix A: Proof of Theorems

A.1 Proof of Theorem 1

First write $\sqrt{m_0}\lambda_{m_0}^{d_u}\Lambda_{m_0}^{-1}(\hat{\beta}_{m_0}(\gamma)-\beta)$ as

$$\left(\Lambda_{m_0}\lambda_{m_0}^{-1-2\gamma}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(I_{xx}\left(\gamma,\lambda_j\right)\right)\Lambda_{m_0}\right)^{-1}\Lambda_{m_0}\lambda_{m_0}^{d_u-1-2\gamma}\sqrt{m_0}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(I_{xu}\left(\gamma,\lambda_j\right)\right).$$

Let $I_{ab}(\lambda_j)$ denote the (a,b)'th element of $I_{ww}(0,\lambda_j)$; the periodogram matrix of $w_t=(\Delta^\gamma x_t',\Delta^\gamma u_t)'$. Then the (a,b)'th element of $\Lambda_{m_0}\lambda_{m_0}^{-1-2\gamma}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(I_{xx}\left(\gamma,\lambda_j\right)\right)\Lambda_{m_0}$ is $\lambda_{m_0}^{d_a+d_b-1}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(I_{ab}\left(\lambda_j\right)\right)$, which converges in probability to $K_{ab}(\gamma)$ by Lemma 6(c). Note that G, and thus the leading $(p-1)\times(p-1)$ submatrix of G and therefore $K(\gamma)$, is invertible by Assumption 1.

For the second term we show that

$$\sqrt{m_0}\left(\Lambda_{m_0}\lambda_{m_0}^{d_u-1-2\gamma}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(I_{xu}\left(\gamma,\lambda_j\right)\right)-H(\gamma)\right)\overset{D}{\to}N\left(0,J(\gamma)\right).$$

By the Cramer-Wold device, for any $(p-1)$-vector η, we need to examine

$$\begin{aligned}
&\eta'\sqrt{m_0}\left(\lambda_{m_0}^{d_u-1-2\gamma}\Lambda_m\hat{F}_{xu}\left(\gamma,1,m_0\right)-H(\gamma)\right)\\
=&\sum_{a=1}^{p-1}\eta_a\sqrt{m_0}\left(\lambda_{m_0}^{d_a+d_p-1}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(I_{ap}\left(\lambda_j\right)\right)-H_a(\gamma)\right)\\
=&\sum_{a=1}^{p-1}\eta_a\sqrt{m_0}\lambda_{m_0}^{d_a+d_p-1}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(I_{ap}\left(\lambda_j\right)-A_a\left(\lambda_j\right)I_{\varepsilon\varepsilon}\left(0,\lambda_j\right)A_p^*\left(\lambda_j\right)\right) &(30)\\
&+\sum_{a=1}^{p-1}\eta_a\sqrt{m_0}\lambda_{m_0}^{d_a+d_p-1}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(A_a\left(\lambda_j\right)I_{\varepsilon\varepsilon}\left(0,\lambda_j\right)A_p^*\left(\lambda_j\right)-f_{ap}\left(\lambda_j\right)\right) &(31)\\
&+\sum_{a=1}^{p-1}\eta_a\sqrt{m_0}\left(\lambda_{m_0}^{d_a+d_p-1}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(f_{ap}\left(\lambda_j\right)\right)-H_a(\gamma)\right), &(32)
\end{aligned}$$

where $I_{\varepsilon\varepsilon}\left(0,\lambda_j\right)$ is the periodogram matrix of ε_t from Assumption 3.

By Lemma 6(a) it follows that (30) is $O_P(m_0^{-1/6}\left(\log m_0\right)^{2/3}+m_0^{-1/2}\left(\log m_0\right)+T^{-1/4})$, and by Lemma 6(b) that (32) is $O(m_0^{\min(1,\phi)+1/2}T^{-\min(1,\phi)})$. Thus, both are $o_P\left(1\right)$ by Assumption 5.

Eq. (31) is

$$\begin{aligned}
&\sum_{a=1}^{p-1}\eta_a\sqrt{m_0}\lambda_{m_0}^{d_a+d_p-1}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(A_a\left(\lambda_j\right)\frac{1}{2\pi}\left(T^{-1}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t'-I_p\right)A_p^*\left(\lambda_j\right)\right) &(33)\\
&+\sum_{a=1}^{p-1}\eta_a\sqrt{m_0}\lambda_{m_0}^{d_a+d_p-1}\frac{2\pi}{T}\sum_{j=1}^{m_0}\mathrm{Re}\left(A_a\left(\lambda_j\right)\frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s\neq t}\varepsilon_t\varepsilon_s'e^{-i(t-s)\lambda_j}A_p^*\left(\lambda_j\right)\right). &(34)
\end{aligned}$$

Note that $D=T^{-1}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t'-I_p$ satisfies $||D||=O_P(T^{-1/2})$ since $\varepsilon_t\varepsilon_t'-I_p$ is a martingale difference sequence with finite second moments. Then, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
(33) &= O_P\left(\max_{1\le a\le p-1} m_0^{-1/2}\left\|D\right\|\lambda_{m_0}^{d_a+d_p}\left(\sum_{j=1}^{m_0}\left\|A_a\left(\lambda_j\right)\right\|^2\right)^{1/2}\left(\sum_{j=1}^{m_0}\left\|A_p\left(\lambda_j\right)\right\|^2\right)^{1/2}\right)\\
&= O_P\left(\max_{1\le a\le p-1} m_0^{-1/2}T^{-1/2}\lambda_{m_0}^{d_a+d_p}\left(\sum_{j=1}^{m_0} f_{aa}\left(\lambda_j\right)\right)^{1/2}\left(\sum_{j=1}^{m_0} f_{pp}\left(\lambda_j\right)\right)^{1/2}\right),
\end{aligned}
$$

where the second equality follows since $||A_a(\lambda)|| = O(\sqrt{f_{aa}(\lambda)})$. Thus (33) is $O_P(\lambda_{m_0}^{1/2})$.

Next, the term inside the parenthesis in eq. (34) can be rewritten as

$$
\begin{aligned}
&\frac{1}{2\pi T}A_a\left(\lambda_j\right)\left(\sum_{t=2}^{T}\sum_{s=1}^{t-1}\varepsilon_t\varepsilon_s' e^{-i(t-s)\lambda_j}+\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\varepsilon_t\varepsilon_s' e^{-i(t-s)\lambda_j}\right)A_p^*\left(\lambda_j\right)\\
&= \frac{1}{2\pi T}A_a\left(\lambda_j\right)\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(\varepsilon_t\varepsilon_s' e^{-i(t-s)\lambda_j}+\varepsilon_s\varepsilon_t' e^{i(t-s)\lambda_j}\right)A_p^*\left(\lambda_j\right),
\end{aligned}
$$

so that

$$
\begin{aligned}
(34) &= \sum_{a=1}^{p-1}\eta_a\frac{\sqrt{m_0}}{T^2}\lambda_{m_0}^{d_a+d_p-1}\sum_{j=1}^{m_0}\sum_{t=2}^{T}\varepsilon_t'\sum_{s=1}^{t-1}\mathrm{Re}\left(A_a'\left(\lambda_j\right)\bar{A}_p\left(\lambda_j\right)e^{-i(t-s)\lambda_j}+A_p^*\left(\lambda_j\right)A_a\left(\lambda_j\right)e^{i(t-s)\lambda_j}\right)\varepsilon_s\\
&= \sum_{t=2}^{T}\varepsilon_t'\sum_{s=1}^{t-1}c_{t-s}\varepsilon_s,
\end{aligned}
$$

where

$$
\begin{aligned}
c_{t-s} &= \frac{1}{2\pi T\sqrt{m_0}}\sum_{j=1}^{m_0}\theta_j,\\
\theta_j &= \mathrm{Re}\left(\sum_{a=1}^{p-1}\eta_a\lambda_{m_0}^{d_a+d_p}A_a'\left(\lambda_j\right)\bar{A}_p\left(\lambda_j\right)e^{-i(t-s)\lambda_j}+\sum_{a=1}^{p-1}\eta_a\lambda_{m_0}^{d_a+d_p}A_p^*\left(\lambda_j\right)A_a\left(\lambda_j\right)e^{i(t-s)\lambda_j}\right)\\
&= \mathrm{Re}\left(\omega_j e^{-i(t-s)\lambda_j}+\omega_j' e^{i(t-s)\lambda_j}\right)\\
&= \left(\mathrm{Re}\,\omega_j+\mathrm{Re}\,\omega_j'\right)\cos\left(\left(t-s\right)\lambda_j\right)+\left(\mathrm{Im}\,\omega_j-\mathrm{Im}\,\omega_j'\right)\sin\left(\left(t-s\right)\lambda_j\right),
\end{aligned}
$$

and we have defined $\omega_j = \sum_{a=1}^{p-1}\eta_a\lambda_{m_0}^{d_a+d_p}A_a'\left(\lambda_j\right)\bar{A}_p\left(\lambda_j\right)$. By defining the triangular array (subscript T is omitted for brevity) $z_1 = 0$ and $z_t = \varepsilon_t'\sum_{s=1}^{t-1}c_{t-s}\varepsilon_s, t = 2, ..., T$, we can apply the martingale difference central limit theorem of Brown (1971) and Hall & Heyde (1980, chp. 3.2) if

$$\sum_{t=1}^{T}E\left(z_t^2\middle|\mathcal{F}_{t-1}\right)-\sum_{a=1}^{p-1}\sum_{b=1}^{p-1}\eta_a\eta_b J_{ab}(\gamma)\xrightarrow{P}0, \tag{35}$$

$$\sum_{t=1}^{T}E\left(z_t^4\right)\to 0, \tag{36}$$

since z_t is a martingale difference array with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, $\mathcal{F}_t = \sigma\left(\{\varepsilon_s, s\le t\}\right)$.

We first show (35). The first term on the left-hand side is

$$\sum_{t=2}^{T}E\left(\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}\varepsilon_s' c_{t-s}'\varepsilon_t\varepsilon_t' c_{t-r}\varepsilon_r\middle|\mathcal{F}_{t-1}\right)=\sum_{t=2}^{T}\sum_{s=1}^{t-1}\varepsilon_s' c_{t-s}' c_{t-s}\varepsilon_s+\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{r\neq s}\varepsilon_s' c_{t-s}' c_{t-r}\varepsilon_r. \tag{37}$$

By slight modification of Lemma 4 of Nielsen (2005) the second term on the right-hand side of (37) is $o_P(1)$. Following the method of Robinson (1995a), we need to show that the mean of the first term on the right-hand side of (37) is asymptotically equal to $\sum_{a=1}^{p-1}\sum_{b=1}^{p-1}\eta_a\eta_b J_{ab}(\gamma)$. Thus,

$$\begin{aligned}\sum_{t=2}^{T}\sum_{s=1}^{t-1} E\,\mathrm{tr}\left(c_{t-s}'c_{t-s}\varepsilon_s\varepsilon_s'\right) &= \sum_{t=2}^{T}\sum_{s=1}^{t-1}\mathrm{tr}\left(c_{t-s}'c_{t-s}\right)\\ &= \sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=1}^{m_0}\frac{1}{4\pi^2T^2m_0}\,\mathrm{tr}\left(\theta_j'\theta_j\right) \qquad (38)\\ &\quad+\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=1}^{m_0}\sum_{k\neq j}\frac{1}{4\pi^2T^2m_0}\,\mathrm{tr}\left(\theta_j'\theta_k\right). \qquad (39)\end{aligned}$$

Note that, from standard trigonometric identities, see also Lemma 3 of Shimotsu (2007),

$$\begin{array}{l}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\cos\left(s\lambda_j\right)\cos\left(s\lambda_k\right)=O\left(T\right), \quad j\neq k,\\ \sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sin\left(s\lambda_j\right)\sin\left(s\lambda_k\right)=O\left(T\right), \quad j\neq k,\\ \sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\cos\left(s\lambda_j\right)\sin\left(s\lambda_k\right)=O(T^2\left(j+k\right)^{-1}+T^2\left|j-k\right|^{-1}), \quad j\neq k,\\ \sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\cos^2\left(s\lambda_j\right)=\frac{T^2}{4}+o\left(T^2\right), \quad j=1,...,m,\\ \sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sin^2\left(s\lambda_j\right)=\frac{T^2}{4}+o\left(T^2\right), \quad j=1,...,m.\end{array}$$

It is thus easily seen that (39) is of smaller order than (38), so we focus on (38) for which

$$\begin{aligned}\mathrm{tr}\left(\theta_j'\theta_j\right) &= \mathrm{tr}\left(\left(\mathrm{Re}\,\omega_j+\mathrm{Re}\,\omega_j'\right)\left(\mathrm{Re}\,\omega_j+\mathrm{Re}\,\omega_j'\right)\cos^2\left(\left(t-s\right)\lambda_j\right)\right)\\ &\quad+\mathrm{tr}\left(\left(\mathrm{Im}\,\omega_j-\mathrm{Im}\,\omega_j'\right)'\left(\mathrm{Im}\,\omega_j-\mathrm{Im}\,\omega_j'\right)\sin^2\left(\left(t-s\right)\lambda_j\right)\right)\\ &\quad+\mathrm{tr}\left(\left(\mathrm{Re}\,\omega_j+\mathrm{Re}\,\omega_j'\right)\left(\mathrm{Im}\,\omega_j-\mathrm{Im}\,\omega_j'\right)\cos\left(\left(t-s\right)\lambda_j\right)\sin\left(\left(t-s\right)\lambda_j\right)\right)\\ &\quad+\mathrm{tr}\left(\left(\mathrm{Im}\,\omega_j-\mathrm{Im}\,\omega_j'\right)'\left(\mathrm{Re}\,\omega_j+\mathrm{Re}\,\omega_j'\right)\cos\left(\left(t-s\right)\lambda_j\right)\sin\left(\left(t-s\right)\lambda_j\right)\right).\end{aligned}$$

The last two terms cancel and the sum of the first two terms can be written as

$$\begin{aligned}\mathrm{tr}\left(\theta_j'\theta_j\right) &= \mathrm{tr}\left(\left(\mathrm{Re}\,\omega_j+\mathrm{Re}\,\omega_j'\right)^2\cos^2\left(\left(t-s\right)\lambda_j\right)\right)-\mathrm{tr}\left(\left(\mathrm{Im}\,\omega_j-\mathrm{Im}\,\omega_j'\right)^2\sin^2\left(\left(t-s\right)\lambda_j\right)\right)\\ &= \mathrm{tr}\left(\left(\mathrm{Re}\,\omega_j+\mathrm{Re}\,\omega_j'\right)^2-\left(\mathrm{Im}\,\omega_j-\mathrm{Im}\,\omega_j'\right)^2\right)\cos^2\left(\left(t-s\right)\lambda_j\right)\\ &\quad-\mathrm{tr}\left(\left(\mathrm{Im}\,\omega_j-\mathrm{Im}\,\omega_j'\right)^2\right)\left(\sin^2\left(\left(t-s\right)\lambda_j\right)-\cos^2\left(\left(t-s\right)\lambda_j\right)\right),\end{aligned}$$

where the second term is of smaller order by the trigonometric relations above. Using that $2\,\mathrm{Re}\left(X\right)=X+\bar{X}$ and $2i\,\mathrm{Im}\left(X\right)=X-\bar{X}$ for any complex matrix X, the first term is

$$\begin{aligned}&\mathrm{tr}\left(2^{-2}\left(\omega_j+\bar{\omega}_j+\omega_j'+\omega_j^*\right)^2-\left(2i\right)^{-2}\left(\omega_j-\bar{\omega}_j-\omega_j'+\omega_j^*\right)^2\right)\cos^2\left(\left(t-s\right)\lambda_j\right)\\ =\;&\frac{1}{2}\,\mathrm{tr}\left(\omega_j^2+\omega_j^{*2}+\omega_j\omega_j^*+\omega_j^*\omega_j+\bar{\omega}_j^2+\omega_j'^2+\bar{\omega}_j\omega_j'+\omega_j'\bar{\omega}_j\right)\cos^2\left(\left(t-s\right)\lambda_j\right)\\ =\;&\mathrm{tr}\left(\omega_j^2+\omega_j^{*2}+\omega_j\omega_j^*+\omega_j^*\omega_j\right)\cos^2\left(\left(t-s\right)\lambda_j\right).\end{aligned}$$

Hence, we have found that (38) is asymptotically negligibly different from

$$\begin{aligned}&\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=1}^{m_0}\frac{1}{4\pi^2T^2m_0}\,\mathrm{tr}\left(\omega_j^2+\omega_j^{*2}+\omega_j\omega_j^*+\omega_j^*\omega_j\right)\cos^2\left(\left(t-s\right)\lambda_j\right)\\ =\;&\sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{1}{4\pi^2T^2m_0}\sum_{a=1}^{p-1}\sum_{b=1}^{p-1}\eta_a\eta_b\lambda_{m_0}^{d_a+d_b+2d_p}\end{aligned}$$

$$\times \sum_{j=1}^{m_0} 4\pi^2 \left(f_{pa}(\lambda_j) f_{pb}(\lambda_j) + f_{ap}(\lambda_j) f_{bp}(\lambda_j) + f_{pp}(\lambda_j) f_{ba}(\lambda_j) + f_{pp}(\lambda_j) f_{ab}(\lambda_j)\right) \cos^2((t-s)\lambda_j)$$

$$= \frac{1}{4m_0} \sum_{a=1}^{p-1}\sum_{b=1}^{p-1} \eta_a \eta_b \lambda_{m_0}^{d_a+d_b+2d_p} \sum_{j=1}^{m_0} \left(f_{pa}(\lambda_j) f_{pb}(\lambda_j) + f_{ap}(\lambda_j) f_{bp}(\lambda_j) + f_{pp}(\lambda_j) f_{ba}(\lambda_j) + f_{pp}(\lambda_j) f_{ab}(\lambda_j)\right),$$

where the equalities follow from (14) and the trigonometric identities above. Approximating the sum over j by an integral, applying Assumptions 1, 2, and that $\cos(x) = (e^{ix} + e^{-ix})/2$, this equals

$$\frac{1}{4}\sum_{a=1}^{p-1}\sum_{b=1}^{p-1} \eta_a \eta_b \frac{1}{1-d_a-d_b-2d_p}$$
$$\times \left(G_{ap}G_{bp}(e^{i\pi(d_a+d_b-2d_p)/2} + e^{-i\pi(d_a+d_b-2d_p)/2}) + G_{ab}G_{pp}(e^{i\pi(d_a-d_b)/2} + e^{-i\pi(d_a-d_b)/2})\right) + o(1)$$
$$= \frac{1}{2}\sum_{a=1}^{p-1}\sum_{b=1}^{p-1} \eta_a \eta_b \frac{1}{1-d_a-d_b-2d_p} \left(G_{ap}G_{bp}\cos(\pi(d_a+d_b-2d_p)/2) + G_{ab}G_{pp}\cos(\pi(d_a-d_b)/2)\right) + o(1)$$
$$= \sum_{a=1}^{p-1}\sum_{b=1}^{p-1} \eta_a \eta_b J_{ab}(\gamma) + o(1).$$

Finally, to show (36),

$$\sum_{t=1}^{T} E\left(z_t^4\right) = \sum_{t=2}^{T} E\left(\sum_{s=1}^{t-1} \varepsilon_s' c_{t-s} \varepsilon_t \varepsilon_t' \sum_{r=1}^{t-1} c_{t-r}\varepsilon_r \sum_{p=1}^{t-1} \varepsilon_p' c_{t-p} \varepsilon_t \varepsilon_t' \sum_{q=1}^{t-1} c_{t-q}\varepsilon_q\right)$$
$$\leq C\left(\sum_{t=2}^{T} \operatorname{tr}\left(\sum_{s=1}^{t-1} c_{t-s}' c_{t-s} c_{t-s}' c_{t-s}\right) + \sum_{t=2}^{T} \operatorname{tr}\left(\sum_{s=1}^{t-1} c_{t-s}' \sum_{r=1}^{t-1} c_{t-r} c_{t-r}' c_{t-s}\right)\right)$$

for some constant $C > 0$ by Assumption 3. Using the arguments of Lemma 4 of Nielsen (2005), this expression can be bounded by $O(T(\sum_{t=1}^{T} ||c_t^2||)^2) = O(T^{-1})$, which completes the proof.

A.2 Proof of Theorem 2

First we show that $(\log T)(\hat{d}_u - d_u) \xrightarrow{P} 0$. Since $\gamma = 0$ it holds that $w_t = (x_t', u_t)'$ such that $d_{x,a} = d_a$ and $d_u = d_p$. Rewriting equations (A.1)-(A.4), (A.24), (A.25), and (A.30) from the proof of Theorem 3 of Robinson (1997) it suffices to show that

$$m_1^{2(d_u-\Delta_1)-1} \sum_{j=1}^{m_1} j^{2(\Delta_1-d_u)} |h_j| \xrightarrow{P} 0 \quad \text{for } 0 \leq \Delta_1 < d_u, \tag{40}$$

$$(\log T)^2 m_1^{2\tau-1} \sum_{j=1}^{m_1} j^{-2\tau} |h_j| \xrightarrow{P} 0 \quad \text{for some } \tau > 0, \tag{41}$$

$$\frac{(\log T)^2}{m_1} \sum_{j=1}^{m_1} |h_j| \xrightarrow{P} 0, \tag{42}$$

$$\frac{1}{m_1} \sum_{j=1}^{m_1} \left((j/\varrho)^{2(\Delta_1-d_u)} - 1\right) h_j \xrightarrow{P} 0, \tag{43}$$

where $\varrho = \exp(m_1^{-1}\sum_{j=1}^{m_1} \log j)$ and

$$h_j = \frac{I_{\hat{u}\hat{u}}(0,\lambda_j) - I_{uu}(0,\lambda_j)}{G_{pp}\lambda_j^{-2d_u}} \tag{44}$$

measures the impact of using the periodogram of residuals (from NBLS with $\gamma = 0$) instead of that of the errors. Our assumption $d_u \geq 0$ allows a simplification of conditions (40)-(43) compared to their counterparts in Robinson (1997), and could be relaxed at the expense of a longer proof.

From our Theorem 1 and Robinson & Marinucci (2003, Theorem 3.1) it holds that $\hat{\beta}_{a,m_0}(0) - \beta_a = O_P(\lambda_{m_0}^{d_{x,a}-d_u})$ when $d_u < \min_{1\leq a\leq p-1} d_{x,a} \leq \max_{1\leq a\leq p-1} d_{x,a} < 1/2$. Using that result along with Assumption 1, (18), and the proof of Theorem 2 of Robinson (1995*b*), h_j satisfies

$$|h_j| = O_P((j/m_0)^{-\delta_{\min}} + (j/m_0)^{-2\delta_{\min}}). \tag{45}$$

Applying (45) and the fact that

$$\sup_{-1\leq\alpha\leq C} \left| m^{-\alpha-1} (\log m)^{-1} \sum_{j=1}^{m} j^{\alpha} \right| = O(1) \text{ for } C \in (1, \infty), \tag{46}$$

which is used without special reference in what follows, it is easy to show that

$$(40) = O_P\left(m_1^{2(d_u-\Delta_1)-1} \sum_{j=1}^{m_1} j^{2(\Delta_1-d_u)-\delta_{\min}} m_0^{\delta_{\min}} (1 + j^{-\delta_{\min}} m_0^{\delta_{\min}}) \right) = O_P\left((\log m_1) \left(\frac{m_0}{m_1}\right)^{\delta_{\min}} \right),$$

and similarly (41) and (42) are both $O_P((\log T)^2 (\log m_1)(m_0/m_1)^{\delta_{\min}})$. Using the fact that $\varrho \sim m_1/e$ ($e = 2.71\ldots$) as $T \to \infty$, the left-hand side of (43) is bounded, for large T, by

$$\frac{1}{m_1}\sum_{j=1}^{m_1} \left(\frac{ej}{m_1}\right)^{2(\Delta_1-d_u)} |h_j| + \frac{1}{m_1}\sum_{j=1}^{m_1} |h_j|,$$

which is negligible by (40) and (42).

Thus, we have shown $(\log T)$-consistency of $\hat{d}_u$ and proceed to prove the rate and asymptotic distribution results. With probability approaching one as $T \to \infty$, $\hat{d}_u$ satisfies

$$0 = \frac{\partial \hat{R}(\hat{d}_u)}{\partial d} = \frac{\partial \hat{R}(d_u)}{\partial d} + \frac{\partial^2 \hat{R}(\bar{d}_u)}{\partial d^2}(\hat{d}_u - d_u),$$

where $|\bar{d}_u - d_u| \leq |\hat{d}_u - d_u|$. Following Robinson (1995*a*, pp. 1641-1644) we have that

$$\frac{\partial^2 \hat{R}(d)}{\partial d^2} = \frac{4(\tilde{G}_{0,\hat{u}}(d)\,\tilde{G}_{2,\hat{u}}(d) - \tilde{G}_{1,\hat{u}}(d)^2)}{\tilde{G}_{0,\hat{u}}(d)^2} = \frac{4(\tilde{F}_{0,\hat{u}}(d)\,\tilde{F}_{2,\hat{u}}(d) - \tilde{F}_{1,\hat{u}}(d)^2)}{\tilde{F}_{0,\hat{u}}(d)^2},$$

where

$$\tilde{G}_{k,q}(d) = \frac{1}{m_1}\sum_{j=1}^{m_1} (\log\lambda_j)^k \lambda_j^{2d} I_{qq}(0,\lambda_j) \text{ and } \tilde{F}_{k,q}(d) = \frac{1}{m_1}\sum_{j=1}^{m_1} (\log j)^k \lambda_j^{2d} I_{qq}(0,\lambda_j).$$

If we show that

$$\sup_{d\in\Delta\cap N_\zeta} \left| \frac{\tilde{G}_{0,\hat{u}}(d) - \tilde{G}_{0,u}(d)}{\bar{G}(d)} \right| = o_P((\log m_1)^{-10}), \tag{47}$$

$$\left| \tilde{F}_{k,\hat{u}}(d_u) - \tilde{F}_{k,u}(d_u) \right| \xrightarrow{P} 0, \quad k = 0, 1, 2, \tag{48}$$

with $\bar{G}(d) = G_{pp}\frac{1}{m_1}\sum_{j=1}^{m_1} \lambda_j^{2(d-d_u)}$ and $N_\zeta = \{d : |d_u - d| < \zeta\}$ for $0 < \zeta < 1/2$, then

$$\frac{\partial^2 \hat{R}(\bar{d}_u)}{\partial d^2} \xrightarrow{P} 4. \tag{49}$$

Note that, following Andrews & Sun (2004, p. 600), in our eq. (47) we use $(\log m_1)^{-10}$ rather than

$(\log m_1)^{-6}$ as in Robinson's (1995a) eq. (4.6). By (4.7) in Robinson (1995a), (47) follows if

$$(\log m_1)^{10} \sum_{j=1}^{m_1} \left(\frac{j}{m_1}\right)^{1-2\tau} j^{-2} \left|\sum_{k=1}^{j} h_k\right| \xrightarrow{P} 0 \quad \text{for some } \tau > 0,$$

which holds by (41) and (42) above. The left-hand side of (48) is bounded by

$$\left|\frac{G_{pp}}{m_1} \sum_{j=1}^{m_1} (\log j)^k h_j\right| \leq \frac{G_{pp} (\log m_1)^k}{m_1} \sum_{j=1}^{m_1} |h_j| = O_P\left((\log m_1)^{k+1} \left(\frac{m_0}{m_1}\right)^{\delta_{\min}}\right)$$

by the same arguments as applied to (42) above. This proves (49).

Having established (49) it follows that

$$\sqrt{m_1}(\hat{d}_u - d_u) = (4 + o_P(1))^{-1} \sqrt{m_1} \frac{\partial \hat{R}(d_u)}{\partial d}, \tag{50}$$

and the first statement of the theorem will follow below by examining the right-hand side of (50). In order to prove the second statement of the theorem we have to show that

$$\sqrt{m_1} \left|\frac{\partial R_{\hat{u}}(d_u)}{\partial d} - \frac{\partial R_u(d_u)}{\partial d}\right| \xrightarrow{P} 0, \tag{51}$$

where

$$\begin{aligned} \frac{\partial R_q(d)}{\partial d} &= 2\frac{\tilde{G}_{1,q}(d)}{\tilde{G}_{0,q}(d)} - \frac{2}{m_1} \sum_{j=1}^{m_1} \log \lambda_j = 2\frac{\tilde{H}_q(d)}{\tilde{G}_{0,q}(d)}, \\ \tilde{H}_q(d) &= \frac{1}{m_1} \sum_{j=1}^{m_1} \nu_j \lambda_j^{2d} I_{qq}(0, \lambda_j), \end{aligned}$$

and $\nu_j = \log j - m_1^{-1} \sum_{j=1}^{m_1} \log j$. Now we write the left-hand side of (51) as

$$\begin{aligned} & 2\sqrt{m_1} \left|\frac{\tilde{H}_{\hat{u}}(d_u) - \tilde{H}_u(d_u)}{\tilde{G}_{0,\hat{u}}(d_u)} - \frac{\tilde{H}_u(d_u)}{\tilde{G}_{0,u}(d_u)} \frac{(\tilde{G}_{0,\hat{u}}(d_u) - \tilde{G}_{0,u}(d_u))}{\tilde{G}_{0,\hat{u}}(d_u)}\right| \\ \leq\ & 2\sqrt{m_1} \left|\tilde{G}_{0,\hat{u}}(d_u)\right|^{-1} \left|\tilde{H}_{\hat{u}}(d_u) - \tilde{H}_u(d_u)\right| \qquad (52) \\ & +2\sqrt{m_1} \left|\tilde{G}_{0,\hat{u}}(d_u)\right|^{-1} \left|\tilde{G}_{0,\hat{u}}(d_u) - \tilde{G}_{0,u}(d_u)\right| \left|\frac{\tilde{H}_u(d_u)}{\tilde{G}_{0,u}(d_u)}\right|. \qquad (53) \end{aligned}$$

To show that (53) is $o_P(1)$ note that $\frac{\tilde{H}_u(d_u)}{\tilde{G}_{0,u}(d_u)} = \frac{1}{2}\frac{\partial R_u(d_u)}{\partial d}$, i.e. the score for the estimation problem with observed series, such that $\left|\frac{\tilde{H}_u(d_u)}{\tilde{G}_{0,u}(d_u)}\right| = O_P(m_1^{-1/2})$ as in Robinson (1995a, p. 1644). Furthermore, based on the previous results we get

$$\left|\tilde{G}_{0,\hat{u}}(d_u) - \tilde{G}_{0,u}(d_u)\right| \leq \frac{1}{m_1} \sum_{j=1}^{m_1} \left|\lambda_j^{2d_u} (I_{\hat{u}\hat{u}}(0, \lambda_j) - I_{uu}(0, \lambda_j))\right| \leq \frac{|G_{pp}|}{m_1} \sum_{j=1}^{m_1} |h_j| = O_P\left((m_0/m_1)^{\delta_{\min}}\right),$$

which is $o_P(1)$ by Assumption 6. Since $\tilde{G}_{0,\hat{u}}(d_u) = G_{pp} + o_P(1)$ by (48) with $k = 0$ and Robinson (1995a), we have established that (53) $= O_P\left((m_0/m_1)^{\delta_{\min}}\right) = o_P(1)$. It also follows that (52) is of the same order as $\sqrt{m_1}|\tilde{H}_{\hat{u}}(d_u) - \tilde{H}_u(d_u)|$ which is equal to

$$\frac{G_{pp}}{\sqrt{m_1}} \left|\sum_{j=1}^{m_1} \nu_j h_j\right| = O_P\left(\frac{(\log m_1)}{\sqrt{m_1}} \sum_{j=1}^{m_1} |h_j|\right) = O_P\left((\log m_1)\sqrt{m_1}\,(m_0/m_1)^{\delta_{\min}}\right).$$

Hence, (51) is $O_P((\log m_1)\sqrt{m_1}\,(m_0/m_1)^{\delta_{\min}})$ in general. By (50) it then follows that $\sqrt{m_1}(\hat{d}_u - d_u) = O_P((\log m_1)\sqrt{m_1}\,(m_0/m_1)^{\delta_{\min}})$ which proves the first statement of the theorem.

To prove the second statement of the theorem, we need to show that in fact $\sqrt{m_1}|\tilde{H}_{\hat{u}}(d_u) - \tilde{H}_u(d_u)| \xrightarrow{P} 0$ if $G_{ap} = G_{pa} = 0$ for $a = 1, \ldots, p-1$. Thus, $\sqrt{m_1}|\tilde{H}_{\hat{u}}(d_u) - \tilde{H}_u(d_u)|$ is equal to

$$
\begin{aligned}
&\frac{G_{pp}}{\sqrt{m_1}}\left|\sum_{j=1}^{m_1}\nu_j h_j\right| \\
\leq\ &\frac{G_{pp}}{\sqrt{m_1}}\left|\sum_{j=1}^{m_1}\nu_j\lambda_j^{2d_u}\left((\beta-\hat{\beta}_{m_0}(0))'\operatorname{Re}(I_{xx}(0,\lambda_j))(\beta-\hat{\beta}_{m_0}(0))/2 + (\beta-\hat{\beta}_{m_0}(0))'\operatorname{Re}(I_{xu}(0,\lambda_j))\right)\right| \\
\leq\ &\frac{G_{pp}}{2\sqrt{m_1}}\left|\sum_{j=1}^{m_1}\nu_j\lambda_j^{2d_u}\sum_{a=1}^{p-1}\sum_{b=1}^{p-1}(\beta_a-\hat{\beta}_{a,m_0}(0))(\beta_b-\hat{\beta}_{b,m_0}(0))\operatorname{Re}(I_{ab}(\lambda_j))\right| \qquad (54)\\
&+\frac{G_{pp}}{\sqrt{m_1}}\left|\sum_{j=1}^{m_1}\nu_j\lambda_j^{2d_u}\sum_{a=1}^{p-1}(\beta_a-\hat{\beta}_{a,m_0}(0))\operatorname{Re}(I_{ap}(\lambda_j))\right|. \qquad (55)
\end{aligned}
$$

First, using summation by parts,

$$
\sum_{j=1}^{m_1}\nu_j\lambda_j^{2d_u}\operatorname{Re}(I_{ap}(\lambda_j)) = \nu_{m_1}\sum_{j=1}^{m_1}\lambda_j^{2d_u}\operatorname{Re}(I_{ap}(\lambda_j)) - \sum_{j=1}^{m_1-1}(\nu_{j+1}-\nu_j)\sum_{k=1}^{j}\lambda_k^{2d_u}\operatorname{Re}(I_{ap}(\lambda_k)),
$$

and for ν_j we know that $\nu_{m_1} = O(1)$ and $|\nu_{j+1}-\nu_j| = O(j^{-1})$ uniformly in j (by a mean value expansion). In the present case with $G_{ap} = G_{pa} = 0$ for $a = 1, \ldots, p-1$ we know from Theorem 1 that $\hat{\beta}_{a,m_0}(0) - \beta_a = O_P(m_0^{-1/2}\lambda_{m_0}^{d_{x,a}-d_u})$. This implies, in conjunction with Lemma 6(c) with $G_{ap} = G_{pa} = 0$ for $a = 1, \ldots, p-1$, that (55) is

$$
\begin{aligned}
&O_P\left(\frac{1}{\sqrt{m_1}}\sum_{a=1}^{p-1}\frac{\lambda_{m_0}^{d_{x,a}-d_u}}{\sqrt{m_0}}\lambda_{m_1}^{d_u-d_{x,a}}\left(m_1^{1+\min(1,\phi)}T^{-\min(1,\phi)} + m_1^{1/2}(\log m_1)\right)\right) \\
&+O_P\left(\frac{1}{\sqrt{m_1}}\sum_{a=1}^{p-1}\frac{\lambda_{m_0}^{d_{x,a}-d_u}}{\sqrt{m_0}}\sum_{j=1}^{m_1-1}j^{-1}\lambda_j^{d_u-d_{x,a}}\left(j^{1+\min(1,\phi)}T^{-\min(1,\phi)} + j^{1/2}(\log j)\right)\right) \\
=\ &O_P\left(\frac{1}{\sqrt{m_0}}\left(\frac{m_0}{m_1}\right)^{\delta_{\min}}\left(m_1^{1/2+\min(1,\phi)}T^{-\min(1,\phi)} + (\log m_1)\right)\right),
\end{aligned}
$$

which is negligible by Assumption 6. Similarly, we get that (54) is also negligible since

$$
O_P\left(\sqrt{m_1}\sum_{a=1}^{p-1}\sum_{b=1}^{p-1}\frac{\lambda_{m_0}^{d_{x,a}+d_{x,b}-2d_u}}{m_0}\lambda_{m_1}^{2d_u-d_{x,a}-d_{x,b}}\right) = O_P\left(\left(\frac{m_0}{m_1}\right)^{2\delta_{\min}}\frac{\sqrt{m_1}}{m_0}\right).
$$

A.3 Proof of Theorem 3

To derive the asymptotic order of $\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\hat{\Gamma}_{m_2}(\gamma) - K(\gamma)^{-1}H(\gamma)$, first write $\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\hat{\Gamma}_{m_2}(\gamma)$ as

$$
\left(\lambda_{m_2}^{-2\gamma}\Lambda_{m_2}\frac{1}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\operatorname{Re}(I_{xx}(\gamma,\lambda_j))\Lambda_{m_2}\right)^{-1}\lambda_{m_2}^{d_u-2\gamma}\Lambda_{m_2}\frac{1}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\operatorname{Re}(I_{x\hat{u}}(\gamma,\lambda_j)).
$$

We then show that

$$\lambda_{m_2}^{-2\gamma}\Lambda_{m_2}\frac{1}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}\left(I_{xx}\left(\gamma,\lambda_j\right)\right)\Lambda_{m_2}-K(\gamma) = O_P\left(l(m_0,m_2)\right), \quad (56)$$

$$\Lambda_{m_2}\lambda_{m_2}^{d_u-2\gamma}\frac{1}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}(I_{x\hat{u}}\left(\gamma,\lambda_j\right))-H(\gamma) = O_P(l(m_0,m_2)+(m_0/m_2)^{\delta_{\min}}), \quad (57)$$

where

$$l(m_0,m_2)=\left(\frac{m_2}{T}\right)^{\min(1,\phi)}+m_2^{-1/2}(\log m_2)+\left(\frac{m_0}{m_2}\right)^{1-2(\max_{1\le a\le p-1}d_{x,a}-\gamma)}\left(\left(\frac{m_0}{T}\right)^{\min(1,\phi)}+m_0^{-1/2}(\log m_0)\right),$$

which is sufficient to prove the desired result since

$$\begin{aligned}&K(\gamma)^{-1}\left(1+O_P\left(l(m_0,m_2)\right)\right)^{-1}H(\gamma)(1+O_P(l(m_0,m_2)+(m_0/m_2)^{\delta_{\min}}))\\ =\;&K(\gamma)^{-1}H(\gamma)\left(1+O_P\left(l(m_0,m_2)\right)\right)(1+O_P(l(m_0,m_2)+(m_0/m_2)^{\delta_{\min}}))\\ =\;&K(\gamma)^{-1}H(\gamma)(1+O_P(l(m_0,m_2)+(m_0/m_2)^{\delta_{\min}}))\end{aligned}$$

and since $\max_{1\le a\le p-1}d_{x,a}-\gamma<1/2$ implies $(m_0/m_2)^{1-2(\max_{1\le a\le p-1}d_{x,a}-\gamma)}(\log m_0)=O((\log T)^{-1})$.

The (a,b)'th element of the left-hand side of (56) is

$$\begin{aligned}&\frac{\lambda_{m_2}^{d_a+d_b}}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}\left(I_{ab}\left(\lambda_j\right)\right)-K_{ab}(\gamma)\\ =\;&\frac{1}{m_2-m_0}\left(\lambda_{m_2}^{d_a+d_b}\sum_{j=1}^{m_2}\mathrm{Re}\left(I_{ab}\left(\lambda_j\right)\right)-K_{ab}(\gamma)\right)-\frac{1}{m_2-m_0}\left(\lambda_{m_2}^{d_a+d_b}\sum_{j=1}^{m_0}\mathrm{Re}\left(I_{ab}\left(\lambda_j\right)\right)-K_{ab}(\gamma)\right)\\ =\;&O_P\left((m_2/T)^{\min(1,\phi)}+m_2^{-1/2}(\log m_2)\right)+O_P\left((m_0/m_2)^{1-d_a-d_b}((m_0/T)^{\min(1,\phi)}+m_0^{-1/2}(\log m_0))\right)\end{aligned}$$

by application of Lemma 6(c).

To prove (57) we write the a'th element of the left-hand side as

$$\frac{\lambda_{m_2}^{d_a+d_p}}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}(\hat{I}_{ap}\left(\lambda_j\right))-H_a(\gamma) = \frac{\lambda_{m_2}^{d_a+d_p}}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}(\hat{I}_{ap}\left(\lambda_j\right)-I_{ap}\left(\lambda_j\right)) \quad (58)$$

$$+\frac{\lambda_{m_2}^{d_a+d_p}}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}\left(I_{ap}\left(\lambda_j\right)\right)-H_a(\gamma), \quad (59)$$

where $\hat{I}_{ap}\left(\lambda_j\right)$ is the cross-periodogram between w_{at} and $\hat{w}_{pt}=\Delta^{\gamma}\hat{u}_t$. Since $\hat{I}_{ap}\left(\lambda_j\right)=I_{ap}\left(\lambda_j\right)+\sum_{b=1}^{p-1}I_{ab}(\lambda_j)(\beta_b-\hat{\beta}_{b,m_0}(\gamma))$, eq. (58) depends on $\beta_b-\hat{\beta}_{b,m_0}(\gamma)$ which is $O_P(\lambda_{m_0}^{d_{x,b}-d_u})=O_P(\lambda_{m_0}^{d_b-d_p})$ by Theorem 1. Thus,

$$\begin{aligned}(58)\;&=\;\sum_{b=1}^{p-1}(\beta_b-\hat{\beta}_{b,m_0}(\gamma))\frac{\lambda_{m_2}^{d_a+d_p}}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}\left(I_{ab}\left(\lambda_j\right)\right)\\ &=\;O_P\left(\sum_{b=1}^{p-1}\lambda_{m_2}^{d_a+d_p}\lambda_{m_0}^{d_b-d_p}\lambda_{m_2}^{-d_a-d_b}\right)=O_P\left(\left(\frac{m_0}{m_2}\right)^{\delta_{\min}}\right).\end{aligned}$$

Lastly, the term (59) is $O_P(l(m_0,m_2))$ by the same argument as for (56).

The same proof can be applied for $\tilde{\Gamma}_{m_2}$, although Lemma 6(c) must be modified as

$$\lambda_r^{d_a+d_b-c-1}\int_0^{\lambda_r}\mathrm{Re}\left(e^{i\lambda(d_a-d_b)/2}\lambda^c f_{ab}\left(\lambda\right)\right)d\lambda$$

$$
\begin{aligned}
&= \lambda_r^{d_a+d_b-c-1}\int_0^{\lambda_r} G_{ab}\lambda^{c-d_a-d_b}\,\mathrm{Re}(e^{i\pi(d_a-d_b)/2})\left(1+O\left(\lambda^{\phi}\right)\right)d\lambda \\
&= \lambda_r^{d_a+d_b-c-1}\int_0^{\lambda_r} G_{ab}\lambda^{c-d_a-d_b}\cos\left(\pi\left(d_a-d_b\right)/2\right)(1+O(\lambda^{\phi}))d\lambda \\
&= \frac{(1-d_a-d_b)}{(1+c-d_a-d_b)}K_{ab}(\gamma)(1+O(\lambda_r^{\phi})).
\end{aligned}
$$

A.4 Proof of Theorem 4

The result follows by application of the previous theorems. From (25) and (27),

$$
\begin{aligned}
\sqrt{m_3}\lambda_{m_3}^{d_u}\Lambda_{m_3}^{-1}(\tilde{\beta}_{m_3}(\gamma)-\beta) &= \sqrt{m_3}\lambda_{m_3}^{d_u}\Lambda_{m_3}^{-1}(\hat{\beta}_{m_3}(\gamma)-\lambda_{m_3}^{-\hat{d}_u}\hat{\Lambda}_{m_3}\lambda_{m_2}^{\hat{d}_u}\hat{\Lambda}_{m_2}^{-1}\tilde{\Gamma}_{m_2}(\gamma)-\beta) \\
&= \sqrt{m_3}\lambda_{m_3}^{d_u}\Lambda_{m_3}^{-1}(\hat{\beta}_{m_3}(\gamma)-\beta) \\
&\quad -\sqrt{m_3}\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\tilde{\Gamma}_{m_2}(\gamma)(1+O_P((\log T)\left(\log m_1\right)(m_0/m_1)^{\delta_{\min}})) \\
&= \sqrt{m_3}\lambda_{m_3}^{d_u}\Lambda_{m_3}^{-1}(\hat{\beta}_{m_3}(\gamma)-\beta)-\sqrt{m_3}\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\tilde{\Gamma}_{m_2}(\gamma)+o_P\left(1\right), \quad (60)
\end{aligned}
$$

where the third equality is by Assumption 8 (or (26) if $m_3=m_0$) and the second follows from

$$
\begin{aligned}
\lambda_{m_3}^{\hat{d}_{x,a}-d_{x,a}} &= 1+O_P((\log T)\,m_1^{-1/2}), \quad a=1,...,p-1, \quad (61) \\
\lambda_{m_2}^{\hat{d}_u-d_u} &= 1+O_P((\log T)\left(\log m_1\right)(m_0/m_1)^{\delta_{\min}}), \quad (62)
\end{aligned}
$$

which are consequences of Robinson (1995*a*) and Theorem 2 above. From Theorem 3 it follows that

$$
\begin{aligned}
\sqrt{m_3}\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}\tilde{\Gamma}_{m_2}(\gamma) &= \sqrt{m_3}K(\gamma)^{-1}H(\gamma)+\sqrt{m_3}O_P\left(\left(\frac{m_0}{m_2}\right)^{\delta_{\min}}+m_0^{-1/2}(\log T)^{-1}+\left(\frac{m_2}{T}\right)^{\phi}\right) \\
&= \sqrt{m_3}K(\gamma)^{-1}H(\gamma)+o_P\left(1\right)
\end{aligned}
$$

by Assumption 8 (or (26) if $m_3=m_0$). The desired result now follows from Theorem 1.

A.5 Proof of Theorem 5

We need to show that $\sqrt{m_3}\lambda_{m_3}^{d_u}\Lambda_{m_3}^{-1}(\tilde{\beta}_{m_3}(\hat{d}_u)-\tilde{\beta}_{m_3}(d_u))=o_P(1)$, which from (60) follows if

$$
\begin{aligned}
\sqrt{m_3}\lambda_{m_3}^{d_u}\Lambda_{m_3}^{-1}(\hat{\beta}_{m_3}(\hat{d}_u)-\hat{\beta}_{m_3}(d_u)) &= o_P(1), \quad (63) \\
\sqrt{m_3}\lambda_{m_2}^{d_u}\Lambda_{m_2}^{-1}(\tilde{\Gamma}_{m_2}(\hat{d}_u)-\tilde{\Gamma}_{m_2}(d_u)) &= o_P(1). \quad (64)
\end{aligned}
$$

First note that by the mean value theorem,

$$
I_{qr}(\hat{d}_u,\lambda)=I_{qr}(d_u,\lambda)+(\hat{d}_u-d_u)\frac{\partial}{\partial d}I_{qr}(\bar{d}_u,\lambda),
$$

where $\frac{\partial}{\partial d}I_{qr}(\bar{d}_u,\lambda)=\frac{\partial}{\partial d}I_{qr}(d,\lambda)|_{d=\bar{d}_u}$ and $\bar{d}_u$ is an intermediate value satisfying $|\bar{d}_u-d_u|\leq|\hat{d}_u-d_u|$. Setting $\theta=\bar{d}_u-d_u$ we have that

$$
\frac{\partial}{\partial d}I_{qr}(\bar{d}_u,\lambda)=\frac{1}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}(\log(1-L)\Delta^{\theta}\Delta^{d_u}q_t)(\Delta^{\theta}\Delta^{d_u}r_s)'e^{-i(t-s)\lambda},
$$

see also Shimotsu & Phillips (2005, p. 1912). Adapting the last displayed equation on p. 1914 of Shimotsu & Phillips (2005) to our notation, using their eq. (59), and the fact that their function $J_T(e^{i\lambda_j})=O(\log T)$, we find that uniformly for $\theta\in M=\{\theta:|\theta|\leq(\log T)^{-4}\}$ it holds that

$$
\frac{1}{m_3}\sum_{j=1}^{m_3}\left|\mathrm{Re}\left(\frac{\partial}{\partial d}I_{qr}(\bar{d}_u,\lambda_j)\right)\right|=O_P\left(\frac{(\log T)}{m_3}\sum_{j=1}^{m_3}\left|\mathrm{Re}\left(I_{qr}(d_u,\lambda_j)\right)\right|\right).
$$

It follows that, uniformly for $\theta\in M$, $\Lambda_{m_3}\lambda_{m_3}^{-1-2d_u}\frac{2\pi}{T}\sum_{j=1}^{m_3}\mathrm{Re}(I_{xx}(\hat{d}_u,\lambda_j)-I_{xx}(d_u,\lambda_j))\Lambda_{m_0}$ has

(a,b)'th element

$$\lambda_{m_3}^{d_{x,a}+d_{x,b}-2d_u} O_P(|\hat{d}_u - d_u|) O_P\left(\frac{(\log T)}{m_3}\sum_{j=1}^{m_3} \mathrm{Re}(\lambda_j^{-d_{x,a}-d_{x,b}+2d_u})\right) = O_P\left(|\hat{d}_u - d_u|(\log T)\right)$$

and $\Lambda_{m_3}\lambda_{m_3}^{d_u-1-2d_u}\sqrt{m_3}\frac{2\pi}{T}\sum_{j=1}^{m_3}\mathrm{Re}(I_{xu}(\hat{d}_u,\lambda_j) - I_{xu}(d_u,\lambda_j))$ has a'th element

$$\lambda_{m_3}^{d_{x,a}-d_u} O_P(|\hat{d}_u - d_u|) O_P\left(\frac{(\log T)}{\sqrt{m_3}}\sum_{j=1}^{m_3} \mathrm{Re}(\lambda_j^{-d_{x,a}+d_u})\right) = O_P\left(\sqrt{m_3}|\hat{d}_u - d_u|(\log T)\right).$$

Both terms are negligible by Assumptions 6 and 8, and thus (63) holds uniformly for $\theta \in M$.

Similarly, uniformly for $\theta \in M$, $\lambda_{m_2}^{-2d_u}\Lambda_{m_2}\frac{1}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}(I_{xx}(\hat{d}_u,\lambda_j) - I_{xx}(d_u,\lambda_j))\Lambda_{m_2}$ has (a,b)'th element

$$\lambda_{m_2}^{d_{x,a}+d_{x,b}-2d_u} O_P(|\hat{d}_u - d_u|) O_P\left(\frac{(\log T)}{m_2-m_0}\sum_{j=m_0+1}^{m_2} \mathrm{Re}(\lambda_j^{-d_{x,a}-d_{x,b}+2d_u})\right) = O_P\left(|\hat{d}_u - d_u|(\log T)\right)$$

and

$$\begin{aligned}
& \Lambda_{m_2}\lambda_{m_2}^{d_u-2d_u}\frac{1}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}(I_{x\hat{u}}(\hat{d}_u,\lambda_j) - I_{x\hat{u}}(d_u,\lambda_j)) \\
= & \Lambda_{m_2}\lambda_{m_2}^{-d_u}\frac{1}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}(I_{xu}(\hat{d}_u,\lambda_j) - I_{xu}(d_u,\lambda_j)) \\
& +\Lambda_{m_2}\lambda_{m_2}^{-d_u}\frac{1}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\mathrm{Re}(I_{xx}(\hat{d}_u,\lambda_j) - I_{xx}(d_u,\lambda_j))(\beta - \hat{\beta}_{m_0}(\hat{d}_u))
\end{aligned}$$

has a'th element

$$\begin{aligned}
& \lambda_{m_2}^{d_{x,a}-d_u} O_P(|\hat{d}_u - d_u|) O_P\left(\frac{(\log T)}{m_2-m_0}\sum_{j=m_0+1}^{m_2} \mathrm{Re}(\lambda_j^{-d_{x,a}+d_u})\right) \\
& +\lambda_{m_2}^{d_{x,a}-d_u} O_P(|\hat{d}_u - d_u|) O_P\left(\frac{(\log T)}{m_2-m_0}\sum_{j=m_0+1}^{m_2}\sum_{b=1}^{p-1} \mathrm{Re}(\lambda_j^{-d_{x,a}-d_{x,b}+2d_u}\lambda_{m_0}^{d_{x,b}-d_u})\right) \\
= & O_P\left(|\hat{d}_u - d_u|(\log T)\right) + O_P\left(|\hat{d}_u - d_u|(\log T)\,(m_0/m_2)^{\delta_{\min}}\right),
\end{aligned}$$

and again both are negligible by Assumptions 6 and 7, which proves (64) uniformly for $\theta \in M$.

From Theorem 2 it holds that $\theta \in M$ with probability tending to one. Therefore the above results also hold with probability tending to one, which proves the result.

Appendix B: Technical Lemma

Lemma 6 *Under Assumptions 1-4, as $T \to \infty$, for $1 \le r \le m$ and $0 \le c \le d_a + d_b$,*

$$\text{(a)} \quad \max_{a,b} \lambda_r^{d_a+d_b-c}\sum_{j=1}^{r}\mathrm{Re}\left(\lambda_j^c[I_{ab}(\lambda_j) - A_a(\lambda_j)\, I_{\varepsilon\varepsilon}(0,\lambda_j)\, A_b^*(\lambda_j)]\right)$$

$$= O_P(r^{1/3}(\log r)^{2/3} + (\log r) + r^{1/2}T^{-1/4}),$$

(b) $$\max_{a,b} \lambda_r^{d_a+d_b-c} \sum_{j=1}^{r} \operatorname{Re}\left(\lambda_j^c f_{ab}(\lambda_j) - \lambda_r^{c-d_a-d_b} \frac{(1-d_a-d_b)}{(1+c-d_a-d_b)} K_{ab}(\gamma) \right) = O_P\left(r^{1+\min(1,\phi)} T^{-\min(1,\phi)} \right),$$

(c) $$\max_{a,b} \lambda_r^{d_a+d_b-c} \sum_{j=1}^{r} \operatorname{Re}\left(\lambda_j^c I_{ab}(\lambda_j) - \lambda_r^{c-d_a-d_b} \frac{(1-d_a-d_b)}{(1+c-d_a-d_b)} K_{ab}(\gamma) \right) = O_P\left(r^{1+\min(1,\phi)} T^{-\min(1,\phi)} + r^{1/2}(\log r) \right),$$

where $I_{\varepsilon\varepsilon}(0,\lambda_j)$ is the periodogram matrix of ε_t from Assumption 3 and $I_{ab}(\lambda_j)$ is the (a,b)'th element of $I_{ww}(0,\lambda_j)$; the periodogram matrix of $w_t = (\Delta^\gamma x_t', \Delta^\gamma u_t)'$.

Proof. Decompose the terms inside the real operator as

$$\begin{aligned} H_{1j} &= \lambda_j^c [I_{ab}(\lambda_j) - A_a(\lambda_j) I_{\varepsilon\varepsilon}(0,\lambda_j) A_b^*(\lambda_j)], \\ H_{2j} &= \lambda_j^c [A_a(\lambda_j) I_{\varepsilon\varepsilon}(0,\lambda_j) A_b^*(\lambda_j) - f_{ab}(\lambda_j)], \\ H_{3j} &= \lambda_j^c f_{ab}(\lambda_j) - \lambda_r^{c-d_a-d_b} \frac{(1-d_a-d_b)}{(1+c-d_a-d_b)} K_{ab}(\gamma). \end{aligned}$$

The proof of Lemma 1(b) in Shimotsu (2007) applies also to our terms H_{1j} and H_{2j} which shows that (a) holds and that $\max_{a,b} |\sum_{j=1}^r H_{2j}| = O_P(r^{1/2}(\log r))$. For H_{3j} we use Assumptions 1 and 2 and the fact that $\operatorname{Re}(e^{i\lambda z}) = 1 + O(\lambda^2), \operatorname{Im}(e^{i\lambda z}) = O(\lambda)$ as $\lambda \to 0$ for any $z \in \mathbb{R}$, which imply

$$\begin{aligned} \operatorname{Re}(e^{i(\pi-\lambda)(d_a-d_b)/2}) &= \operatorname{Re}(e^{i\pi(d_a-d_b)/2}) \operatorname{Re}(e^{-i\lambda(d_a-d_b)/2}) - \operatorname{Im}(e^{i\pi(d_a-d_b)/2}) \operatorname{Im}(e^{-i\lambda(d_a-d_b)/2}) \\ &= \cos(\pi (d_a - d_b)/2)(1 + O(\lambda^2)) - \sin(\pi (d_a - d_b)/2) O(\lambda), \end{aligned}$$

such that

$$\begin{aligned} \lambda_r^{d_a+d_b-c} r^{-1} \sum_{j=1}^{r} \operatorname{Re}\left(\lambda_j^c f_{ab}(\lambda_j)\right) &= \lambda_r^{d_a+d_b-c-1} \int_0^{\lambda_r} \operatorname{Re}\left(\lambda^c f_{ab}(\lambda)\right) d\lambda + R_T \\ = \lambda_r^{d_a+d_b-c-1} \int_0^{\lambda_r} G_{ab} \lambda^{c-d_a-d_b} \operatorname{Re}(e^{i(\pi-\lambda)(d_a-d_b)/2}) \left(1 + O\left(\lambda^\phi\right)\right) d\lambda + R_T \\ = \lambda_r^{d_a+d_b-c-1} \int_0^{\lambda_r} G_{ab} \lambda^{c-d_a-d_b} \cos(\pi (d_a - d_b)/2) (1 + O(\lambda^{\min(1,\phi)})) d\lambda + R_T \\ = \frac{(1-d_a-d_b)}{(1+c-d_a-d_b)} K_{ab}(\gamma) (1 + O(\lambda_r^{\min(1,\phi)})) + R_T. \end{aligned}$$

The approximation error R_T is $O(T^{c-d_a-d_b-1}(\log r))$ uniformly in r. ■

References

Andrews, D. W. K. & Sun, Y. (2004), 'Adaptive local polynomial Whittle estimation of long-range dependence', *Econometrica* **72**, 569–614.

Baillie, R. T. (1996), 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics* **73**, 5–59.

Bandi, F. M. & Perron, B. (2006), 'Long memory and the relation between implied and realized volatility', *Journal of Financial Econometrics* **4**, 636–670.

Berger, D., Chaboud, A. & Hjalmarsson, E. (2009), 'What drives volatility persistence in the foreign exchange market', *Journal of Financial Economics* **94**, 192–213.

Brown, B. M. (1971), 'Martingale central limit theorems', *Annals of Mathematical Statistics* **42**, 59–66.

Chen, W. W. & Hurvich, C. M. (2003), 'Estimating fractional cointegration in the presence of polynomial trends', *Journal of Econometrics* **117**, 95–121.

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.

Comte, F. & Renault, E. (1998), 'Long-memory in continuous-time stochastic volatility models', *Mathematical Finance* **8**, 291–323.

Davidson, J. (2004), 'Convergence to stochastic integrals with fractionally integrated integrator processes: Theory and applications to fractional cointegration', *Preprint, University of Exeter*.

Dolado, J. J. & Marmol, F. (1996), 'Efficient estimation of cointegrating relationships among higher order and fractionally integrated processes', *Working Paper 9617, Banco de Espana*.

Doornik, J. A. & Ooms, M. (2004), 'Inference and forecasting for ARFIMA models with an application to US and UK inflation', *Studies in Nonlinear Dynamics & Econometrics* **8**(2). Article 14.

Fox, R. & Taqqu, M. S. (1986), 'Large-sample properties of parameter estimates for strongly dependent stationary gaussian series', *Annals of Statistics* **14**, 517–532.

Granger, C. W. J. (1981), 'Some properties of time series data and their use in econometric model specification', *Journal of Econometrics* **16**, 121–130.

Hall, P. & Heyde, C. C. (1980), *Martingale Limit Theory and its Application*, Academic Press, New York.

Hannan, E. J. (1979), 'The central limit theorem for time series regression', *Stochastic Processes and their Applications* **9**, 281–289.

Henry, M. & Zaffaroni, P. (2003), The long range dependence paradigm for macroeconomics and finance, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, 'Theory and Applications of Long-Range Dependence', Birkhäuser, Boston, pp. 417–438.

Hualde, J. & Robinson, P. M. (2010), 'Semiparametric inference in multivariate fractionally cointegrated systems', *Forthcoming in Journal of Econometrics*.

Hurvich, C. M., Deo, R. S. & Brodsky, J. (1998), 'The mean squared error of Geweke and Porter-Hudak's estimator of the memory paramater of a long memory time series', *Journal of Time Series Analysis* **19**, 19–46.

Kim, C. S. & Phillips, P. C. B. (2001), 'Fully modified estimation of fractional cointegration models', *Preprint, Yale University*.

Künsch, H. R. (1987), Statistical aspects of self-similar processes, *in* Y. Prokhorov & V. V. Sazanov, eds, 'Proceedings of the First World Congress of the Bernoulli Society', VNU Science Press, Utrecht, pp. 67–74.

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

Lobato, I. N. & Robinson, P. M. (1996), 'Averaged periodogram estimation of long memory', *Journal of Econometrics* **73**, 303–324.

Lobato, I. N. & Robinson, P. M. (1998), 'A nonparametric test for I(0)', *Review of Economic Studies*

65, 475–495.

Marinucci, D. & Robinson, P. M. (2001 *a*), 'Finite-sample improvements in statistical inference with I(1) processes', *Journal of Applied Econometrics* **16**, 431–444.

Marinucci, D. & Robinson, P. M. (2001 *b*), 'Semiparametric fractional cointegration analysis', *Journal of Econometrics* **105**, 225–247.

Nielsen, M. Ø. (2005), 'Semiparametric estimation in time-series regression with long-range dependence', *Journal of Time Series Analysis* **26**, 279–304.

Phillips, P. C. B. (1991), Spectral regression for cointegrated time series, *in* W. A. Barnett, J. Powell & G. E. Tauchen, eds, 'Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics', Cambridge University Press, Cambridge, pp. 413–435.

Phillips, P. C. B. & Hansen, B. E. (1990), 'Statistical inference in instrumental variables regression with I(1) variables', *Review of Economic Studies* **57**, 99–125.

Robinson, P. M. (1994), 'Semiparametric analysis of long-memory time series', *Annals of Statistics* **22**, 515–539.

Robinson, P. M. (1995*a*), 'Gaussian semiparametric estimation of long range dependence', *Annals of Statistics* **23**, 1630–1661.

Robinson, P. M. (1995*b*), 'Log-periodogram regression of time series with long range dependence', *Annals of Statistics* **23**, 1048–1072.

Robinson, P. M. (1997), 'Large-sample inference for nonparametric regression with dependent errors', *Annals of Statistics* **25**, 2054–2083.

Robinson, P. M. (2003), Long-memory time series, *in* P. M. Robinson, ed., 'Time Series With Long Memory', Oxford University Press, Oxford, pp. 4–32.

Robinson, P. M. (2005), 'The distance between rival nonstationary fractional processes', *Journal of Econometrics* **128**, 283–300.

Robinson, P. M. (2008), 'Multiple local Whittle estimation in stationary systems', *Annals of Statistics* **36**, 2508–2530.

Robinson, P. M. & Henry, M. (1999), 'Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels', *Econometric Theory* **15**, 299–336.

Robinson, P. M. & Hidalgo, F. J. (1997), 'Time series regression with long-range dependence', *Annals of Statistics* **25**, 77–104.

Robinson, P. M. & Marinucci, D. (2003), Semiparametric frequency domain analysis of fractional cointegration, *in* P. M. Robinson, ed., 'Time Series With Long Memory', Oxford University Press, Oxford, pp. 334–373.

Robinson, P. M. & Yajima, Y. (2002), 'Determination of cointegrating rank in fractional systems', *Journal of Econometrics* **106**, 217–241.

Shimotsu, K. (2007), 'Gaussian semiparametric estimation of multivariate fractionally integrated processes', *Journal of Econometrics* **137**, 277–310.

Shimotsu, K. & Phillips, P. C. B. (2005), 'Exact local Whittle estimation of fractional integration', *Annals of Statistics* **33**, 1890–1933.

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