UvA ECONOMETRICS

Discussion Paper: 2010/11

Dynamic OLS estimation of fractionally cointegrated regressions

Paulius Stakénas

www.feb.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Dynamic OLS estimation of fractionally cointegrated regressions

Paulius Stakėnas†

University of Amsterdam

Tinbergen Institute

This version: October 31, 2010

Abstract

In this paper we study estimation and inference of cointegration vector(s) in a fractionally cointegrated system employing a regression-based approach. In "strongly cointegrated" regressions (when the difference between integration order of observables and cointegration errors exceeds 1/2) the OLS estimator of the cointegration vector does not have an optimal rate of convergence in a part of parameter space. We use the approach of Saikkonen (1991) appending the regression equation with leads and lags of filtered regressor and estimate cointegration vector with OLS in the appended regression in this way obtaining optimal convergence rate and local asymptotic mixed normal distribution of the estimator. Although the estimator depends on the values of integration order and cointegration strength, we show that use of consistent estimates does not affect asymptotic properties of the estimator. This allows to construct feasible Wald test for linear restrictions on the coefficients with nuisance-free asymptotic null distribution. Monte Carlo study illustrating finite sample properties of the estimator and Wald test is provided.

Keywords: Long memory processes, fractional cointegration, dynamic OLS estimation.

JEL classification: C13, C32.

†Address: Faculty of Economics and Business, Department of Quantitative Economics, Roetersstraat 11, 1018WB Amsterdam, The Netherlands. E-mail: P.Stakenas@uva.nl.

1 Introduction

Cointegration analysis has become one of the main tools of empirical research in economics and finance. Traditional cointegration framework assumes observed time series to be a unit root process, while cointegration errors are assumed to be a weakly stationary process. Although the implicit knowledge of integration orders of both observed time series and cointegration errors might be seen restrictive in the light of empirical studies documenting possible fractional behaviour of economic/financial time series (see Baillie (1996) for a review), most of the previous theoretical and empirical work concentrated on this standard $I(1)/I(0)$ case. Hence, it seems natural to embed cointegration analysis in the fractional framework letting time series be integrated of arbitrary order. Generally, we will say that multivariate time series X_t is fractionally integrated of order d and denote it as $X_t \sim I(d)$, if $X_t = \Delta^{-d} u_t \mathbb{1}_{t>0}, t \in \mathbb{N}$, for $d > -1/2$, where u_t has a continuous, bounded, positive semi-definite and non-zero everywhere spectral density matrix (hence we follow a so called type II definition, cf. Marinucci and Robinson (2000)). We will say that X_t is fractionally cointegrated, if $X_t \sim I(d)$, but there exists a non-zero full column rank matrix β: $\beta' X_t \sim I(d-b)$, for some $b > 0$.

One of the basic questions of interest in a fractionally cointegrated system is the value of cointegration vector. There has been a number of approaches for inference of cointegration vector suggested in the literature and they can be attributed to either likelihood-based (see Johansen (1995) for $I(0)/I(1)$ and Johansen and Nielsen (2010b) for fractional framework) or regression-based methods. To illustrate the ideas of the latter approach, suppose we observe a non-stationary bivariate fractionally cointegrated times series $X_t = (X_{1t}, X_{2t})'$. Then the normalized cointegration vector α in the regression:

$$X_{1t} = \alpha X_{2t} + u_t \tag{1}$$

can be estimated with regression methods upon observing that due to cointegration the strength of the signal X_t dominates the strength of the noise u_t (in a stochastic sense). However, it is well-known that in $I(1)/I(0)$ case, OLS estimator of (1) is second-order biased and does not bring the inference problem into locally asymptotically mixed normal (LAMN) family (cf. Phillips (1991)). The problem runs even deeper in fractional framework: OLS has slower than optimal convergence rate in the parameter space $\{2d - b \leq 1, d \geq 0, (d, b) \neq (1, 1)\}$. This can be solved with spectral regression methods: narrow-band least squares (NBLS) regression improves convergence rate and removes second-order bias in case $\psi = (d, b) = (1, 1)$ (cf. Robinson and Marinucci (2001)). Although the rate of NBLS estimator is yet not optimal, the narrow-band weighted least squares (NBWLS) regression achieves optimal rate of convergence in the parameter space $\{d \geq b > 1/2\}$ (cf. Robinson and Hualde (2003), Hualde and Robinson (2010)).

On the other hand, the problem of non-optimal rate of OLS estimator in (1) might be solved correcting for endogeneity in the regression using time-domain methods. There are two established approaches of this idea in $I(1)/I(0)$ literature: fully-modified estimation (cf. Phillips and Hansen (1990)) and dynamic OLS estimation (cf. Saikkonen (1991), Stock and Watson (1993)). The former method has been extended to fractional framework by Kim and Phillips (2001), whereas the goal of our paper is to extend the latter approach to fractional framework. Dynamic OLS estimation in $I(1)/I(0)$ is based on appending regression equation (1) with leads and lags of differenced regressor X_{2t} what removes second-order bias, whereas estimation of fractional regression appended with fractionally filtered regressor X_{2t} improves rate of convergence of the estimator (as compared to OLS estimator of (1)) yielding locally asymptotically mixed normal estimator of α. DOLS estimator was proved to be asymptotically efficient in $I(1)/I(0)$ setting in Saikkonen (1991), provided some conditions on the rate of growth of number k of appended leads and lags hold. Similar conditions for k require to hold in a fractional regression. Although in fractional setting typically neither the order of integration of the original series nor the cointegration strength are known and estimates of the fractional parameters $\psi = (d, b)$ have to be used, we show that their consistency at a rate of T^{κ}, for $\kappa > 0$, is enough for optimal feasible DOLS estimation.

The rest of the paper is organized as follows: section 2 presents the framework, discusses the model and assumptions. Section 3 outlines the idea of estimation and presents main results. The results of Monte Carlo simulations evaluating finite sample performance of the estimator are presented in the section 4, while section 5 concludes. Proofs are given in the appendices.

We use the following notation in the paper: $\xrightarrow{P}$ means convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution. The Euclidian norm of a matrix, vector or scalar Z is denoted as $||Z|| = \sqrt{tr(Z'Z)}$. We will also use operator norm: $||A||_1 = \sqrt{\lambda_{\max}(A'A)}$. $AsCov(Z)$ denotes asymptotic variance-covariance matrix of asymptotically stationary random vector Z. Finally, we use shorthand notation for filtered multivariate observables, i.e. $\Delta^d X_t$ denotes a vector X_t with its individual components filtered with Δ^d. Also, D_u will denote a first-order derivative w.r.t. u. $W_b(s)$ denotes fractional standard type II Brownian motion which is defined as follows:

$$W_b(0) = 0, \ \ a.s.$$
$$W_b(t) = \frac{1}{\Gamma(b+1)} \int_0^t (t-u)^b dW(u).$$

where $W(u)$ is a standard Brownian motion.

2 The model and preliminaries

In the paper we are concerned with generic fractionally cointegrated (FCI) time series:

Assumption 1. *Observed n-dimensional time series X_t satisfies:*

$$\Delta(L,\psi)(\beta,\gamma)' X_t = u_t, \ \ t = 0, 1, \ldots \tag{2}$$

where $\Delta(L,\psi) = blockdiag\{\Delta_+^{d-b} I_r, \Delta_+^d I_{n-r}\}$, I_r is a $r \times r$ unit matrix and $b > 0$. β is $n \times r$ matrix, (β,γ) is $n \times n$ matrix of full rank, u_t is $I(0)$ process and the expression Δ_+^{-d} is a truncated fractional differencing operator:

$$\Delta_+^{-d} u_t = (1-L)_+^{-d} u_t = \sum_{i=0}^{t} \frac{\Gamma(d+i)}{\Gamma(d)\Gamma(i+1)} u_{t-i}$$

with $\Gamma(i)$ - Gamma function.

Our characterization of fractionally integrated process coincides with type II definition, which defines fractional integration directly in terms of fractional filter, i.e. as a weakly stationary time series filtered with truncated fractional filter. Different characterization is also possible, but it leads to different asymptotic inference considerations and have different interpretation for transition mechanisms of innovation shocks (for a discussion see Shimotsu and Phillips (2006), section 7). Given this definition, the model (2) formalizes the idea of fractional cointegration in a very general way: $X_t \sim I(d)$, but $\beta' X_t \sim I(d-b)$.

In the paper we are interested in “strongly“ cointegrated fractional systems:

Assumption 2. $b > 1/2$.

Although most empirical studies are concerned with long memory type behaviour of cointegration errors $\beta' X_t$, what implies $d \geq b$, our framework in principle does not necessitate this assumption, only requiring strong cointegration. The assumption is crucial for feasible optimal estimation of cointegration vector: in a fractional regression filtered observables are $I(b)$ and hence non-stationary processes, wheras the errors are $I(0)$ and stochastic dominance of signal to noise allows feasible optimal inference. Finally, for the development of partial sum limit theory we need the following assumption on the errors:

Assumption 3. *$u_t = C(L)\varepsilon_t$, where $C(L)$ is a lag matrix polynomial such that $det(C(z)) \neq 0, \forall |z| = 1$ and coefficients of $C(L)$ and $C^{-1}(L)$ are 1/2-summable, i.e. $\sum_j \sqrt{j}||C_j|| < \infty$. ε_t is i.i.d. $(0, \Sigma)$ time series with $\Sigma > 0$, $E||\varepsilon_t||^q < \infty$ for some $q > \max\{4, (b-1/2)^{-1}\}$.*

The assumption on the error process u_t conveys somewhat restrictive cointegration framework in a sense that it does not allow multicointegration, i.e. cointegration between cointegration errors $\beta' X_t$ (note that under the assumptions long-run covariance matrix of u_t: $\Omega = C(1)\Sigma C(1)'$ is positive definite). Although it may be plausible to consider multicointegration in the empirical work, from theoretical perspective the question seems complicated and it does not seem possible to solve it with our method.[1]

Assumption 3 ensures that multivariate fractional invariance principle holds for u_t (see Marinucci and Robinson (2000), Theorem 1):

$$T^{1/2-b} \sum_{i=0}^{[Ts]} \Delta_+^{1-b} u_t \Rightarrow W_{b-1}(s). \tag{3}$$

Here $W_{b-1}(s)$ is type II fractional Brownian motion with covariance matrix $\Omega = C(1)\Sigma C(1)'$.

Our assumptions for u_t are almost identical to that of Robinson and Hualde (2003) and Hualde and Robinson (2010), where feasible optimal inference for cointegration vector in strongly cointegrated systems is also considered, ours being relatively milder requiring 1/2-summability (rather than 1-summability) for coefficients of $C(L), C^{-1}(L)$, although the setting of Hualde and Robinson (2010) is more general and allows for multicointegration.

[1] Our method implicitly relies on the following commutation of lag polynomials: $\Delta(L,d)A(L) = A(L)\Delta(L,d)$, where $\Delta(L,d)$ is a diagonal matrix with Δ_+^d on the diagonal and $A(L)$ is a conformable lag matrix polynomial.

3 Regression-based dynamic OLS estimation of cointegration vectors

The key element for the regression-based estimation is the following normalization[2] of cointegration vector: $\beta' = (I_r, -\alpha)$ and $\gamma' = (0_{n-r\times r}, I_{n-r})$. The normalization in $I(1)/I(0)$ systems was introduced by Phillips (1991) and conveys triangular representation of a cointegrated system. Similarly we obtain a fractional triangular model:

$$\Delta_+^{d-b}(X_{1t} - \alpha X_{2t}) \quad = u_{1t}, \tag{4}$$

$$\Delta_+^{d} X_{2t} \quad = u_{2t}. \tag{5}$$

where $X_t = (X_{1t}', X_{2t}')'$ and $u_t = (u_{1t}', u_{2t}')'$.

Cointegrated regression of the type $X_{1t} = \alpha X_{2t} + u_t$ has been studied in a number of papers. It is well-known that in case of endogeneity between X_{2t} and error term u_t in $I(1)/I(0)$ systems, OLS estimator has second-order bias and does not bring the inference problem into LAMN framework (cf. Phillips (1991)). However, in case of fractionally cointegrated regressions, endogeneity between regressors and errors has deeper consequences: its rate of convergence in a part of parameter space $\{2d - b \leq 1, d \geq 0, (d, b) \neq (1, 1)\}$ is slower than optimal (cf. Robinson and Marinucci (2001)). Within spectral regression framework the problem can be partly solved with narrow-band least squares or narrow-band weighted least squares estimation, the latter achieving optimal rate of convergence in non-stationary strongly cointegrated systems. Our approach uses time-domain framework and takes idea from dynamic OLS estimation introduced in Saikkonen (1991), where the regression equation is appended with lags and leads of differenced regressor thus removing endogeneity effects.

We sketch the idea of dynamic OLS estimation. Note, that absolute summability of $||C_j||$ implies absolute summability of autocovariances of u_t. On the other hand, fourth order stationary time series with absolutely summable coefficients and finite fourth moment of innovation process implies 4-th order summability of their cummulants. That and positive boundedness of spectral density of error term $f_u(\lambda) = C(e^{i\lambda})\Sigma C'(e^{-i\lambda}) = (f_{ij}(\lambda))_{i,j=1,2}$ imply (cf. Brillinger (1974), p.

[2] Validity of normalization is showed in the appendix.

296):

$$u_{1t} = \sum_{i=-\infty}^{\infty} \Pi_i u_{2t-i} + v_t, \tag{6}$$

where $\Pi(L) = \sum_{i=-\infty}^{\infty} \Pi_i L^i$ satisfies $\Pi(e^{-i\lambda}) = f_{12}(\lambda) f_{22}^{-1}(\lambda)$ and hence v_t is a stationary process such that:

- $E v_t u_{2t+k} = 0, \forall k \in \mathbb{Z}$,
- $\sum_j ||\Pi_j|| < \infty$,
- $Cov(v_t) = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$.

Where $(\Omega_{ij})_{ij=1,2}$ are the blocks of long-run covariance matrix Ω corresponding to $u_t = (u_{1t}', u_{2t}')'$. If we denote $\mathcal{F} = \sigma(\{u_{2t}, t \in \mathbb{Z}\})$ - σ-field generated by a sequence of random variables $\{u_{2t}, t \in \mathbb{Z}\}$, then v_t is the difference between u_{1t} and its linear projection onto $\mathcal{F}$ and if u_t is Gaussian, then $v_t = u_{1t} - E(u_{1t}||\mathcal{F})$. The uncorrelatedness between v_t and u_{2t} suggests that appending regression equation (4) with a finite number of leads and lags of fractionally filtered regressor X_{2t} could "almost" remove endogeneity effects yielding optimal rate estimator of α, as it does in $I(1)/I(0)$ systems removing second-order bias. Our proposed dynamic OLS estimator $\hat{\alpha}_{DOLS}(d, b)$ is defined as OLS estimator in the following regression:

$$\Delta_+^{d-b} X_{1t} = \alpha \Delta_+^{d-b} X_{2t} + \sum_{i=-k}^{k} \Pi_i \Delta_+^d X_{2t-i} + \tilde{v}_t, \ \ t = k, \dots, T-k \tag{7}$$

where $\tilde{v}_t = v_t + e_t$, $e_t = \sum_{|i|>k} \Pi_i \Delta_+^d X_{2t-i}$. Error term e_t represents the error due to autoregressive approximation of v_t with finite number of lags k. It is not difficult to show that given the true values $\psi_0 = (d_0, b_0)$, the infeasible DOLS estimator $\hat{\alpha}_{DOLS}(\psi_0)$ has optimal rate of convergence with LAMN asymptotic distribution (albeit mixing covariates being functionals of fractional Brownian motion). However, obviously in fractional setting it is rather unrealistic to know the true values of fractional parameters ψ, but we show that feasible estimator retains its asymptotic properties under the following conditions on the estimator of ψ and the rate of growth of lag length k:

Assumption 4. *Suppose* $||\hat{\psi} - \psi|| = O_p(T^{-\kappa})$, $\kappa > 0$ *and* $k \to \infty$ *are such that:*

1. $k\left(T^{-1/2} + \log T T^{-a} + T^{-\kappa}\right) = o(1)$, *for* $a = \min\{1, 2b - 1\}$,

2. $\left(\log T T^{1-b} + k\right) \sum_{|j|>k} ||\Pi_j|| = o(1)$.

Assumption 4.1 puts an upper bound for the rate k: it does not grow faster than the rate of estimator $\hat{\psi}$ and $(\log T T^{-a} + T^{-1/2})$, while assumption 4.2 puts the lower bound for the rate. Obviously, the bound is infeasible and depends on the structure of the process u_t, but for the ARMA-type processes u_t it translates into: $k \to \infty$, when $b > 1$; $k(\log\log T)^{-1} \to \infty$, when $b = 1$ and $k \log^{-1} T \to \infty$, when $b < 1$ (in case $b > 1$, the second part of the assumption is innocuous).

Given the above-stated assumptions hold, the following theorem is the main result of the paper:

Theorem 3.1. *Under assumptions 1-4 for the feasible estimator $\hat{\alpha} = \hat{\alpha}_{DOLS}(\hat{\psi})$ it holds:*

$$T^b \left(\hat{\alpha}_{DOLS} - \alpha\right) \xrightarrow{d} \left(\int_0^1 dW_{1\cdot 2} W_{b-1}'\right) \left(\int_0^1 W_{b-1} W_{b-1}'\right)^{-1}. \tag{8}$$

Since Brownian motions $W_{1\cdot 2}(s) = W_1(s) - \Omega_{12}\Omega_{22}^{-1} W_2(s)$ and $W_{b-1}(s)$ are uncorrelated (hence independent) and $W_{b-1}(s)$ is a function of $W_2(s)$ (hence independent of $W_{1\cdot 2}(s)$) we might express the limit distribution using mixture representation (cf. Phillips (1989)):

$$T^b \left(\hat{\alpha}_{DOLS} - \alpha\right) \overset{def}{=} \int_{G>0} N(0, \Omega_{11\cdot 2} \otimes G) dP(G) \tag{9}$$

where

$$G = \left(\int_0^1 W_{b-1} W_{b-1}'\right)^{-1} \text{ and } \Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}. \tag{10}$$

Asymptotic mixed normality of DOLS estimator allows standard asymptotic inference on the parameters of the model. Consider Wald statistic for the linear hypothesis $H_0 : R\text{vec}(\alpha) = r$, where R is $s \times r(n-r)$ matrix, based on the estimator $\hat{\alpha}_{DOLS} = \hat{\alpha}_{DOLS}(\psi)$:

$$W(\psi) = \left(R\text{vec}(\hat{\alpha}_{DOLS}) - r\right)' \left(R(\Omega_{11\cdot 2} \otimes (X_2 X_2')^{-1})R'\right)^{-1} \left(R\text{vec}(\hat{\alpha}_{DOLS}) - r\right) \tag{11}$$

where $X_2 = X_2(\psi) = (\Delta_+^{d-b} X_{2k}, \ldots, \Delta_+^{d-b} X_{2T-k})$. Then under the null hypothesis for the infeasible test statistic holds: $W(\psi_0) \xrightarrow{d} \chi_s^2$. Feasible test statistic requires estimation of ψ and long run covariance matrix of v_t. One way to estimate $\Omega_{11\cdot 2}$ is with HAC-type estimator (cf. Andrews (1991)) as weighted sum of the sample autocovariances of regression (7) residuals $\hat{\hat{v}}_t$:

$$\hat{\Omega}_{11\cdot 2} = T^{-1} \sum_{t_1=k}^{T-k} \sum_{t_2=k}^{T-k} \omega\left(\frac{|t_1 - t_2|}{h_T}\right) \hat{\hat{v}}_{t_1} \hat{\hat{v}}_{t_2}'. \tag{12}$$

If we impose the following conditions on the kernel function and bandwidth:

Assumption 5. *The kernel function* $\omega(\cdot) : \mathbb{R} \to [-1, 1]$ *is a continuous, even function, satisfying:*

- $\omega(0) = 1$,
- $\int_{-\infty}^{\infty} |w(x)|dx < \infty$

and the bandwidth h_T *satisfies:* $||\hat{\psi} - \psi||h_T = o_p(1)$.

then it guarantees consistency of $\hat{\Omega}_{11\cdot 2}$:

Theorem 3.2. *Suppose assumptions 1-5 hold. Then* $\hat{\Omega}_{11\cdot 2} \xrightarrow{P} \Omega_{11\cdot 2}$.

Given the latter theorem we could use estimate (12) and $\hat{\psi}$ to construct feasible Wald test statistic (11) for which: $W(\hat{\psi}) \xrightarrow{d} \chi^2_s$. Of course, (12) is not the only way for estimating $\Omega_{11\cdot 2}$: another option is to construct $\hat{\Omega}_{11\cdot 2}$ from a consistent estimate of Ω. Estimation of Ω can be achieved with HAC-type or Robinson's MAC estimator (for a comparison see Abadir et al. (2009)).

Remark 3.3. Note, that theorem 3.1 does not discuss consistency of regression (7) coefficients $\Pi = (\Pi_{-k}, \ldots, \Pi_k)$. The following proposition gives the rate of consistency of $\hat{\Pi}$:

Proposition 3.4. *Suppose assumptions 1-4 hold. Then* $||\hat{\Pi} - \Pi|| \left(T^{-1/2} + T^{-\kappa} + \sum_{|i|>k} ||\Pi_i||\right)^{-1} = O_p(\sqrt{k})$.

The proposition effectively states that in case we use $\hat{\psi}$ instead of ψ_0 for estimation of Π, the highest rate of consistency for $\hat{\Pi}$ we achieve is $T^{\kappa}/\sqrt{k}$, provided $T^{\kappa} \sum_{|i|>k} ||\Pi_i|| = o(1)$ holds. That is in contrast with the case when true ψ_0 for estimation is used: rate $\sqrt{T/k}$ consistency is achieved, given $\sqrt{T} \sum_{|i|>k} ||\Pi_i||$ holds (cf. Saikkonen (1992)). In addition, lower upper bound assumption $k^3/T = o(1)$ used in Saikkonen (1991) and Said and Dickey (1984) for consistency of $\hat{\Pi}$ seems to be not needed.

Remark 3.5. In an empirical analysis it might be reasonable to allow for different integration orders of individual components of cointegration errors, however, this is not feasible in our framework. The assumption on homogeneity of memory of error term components is crucial, since it permits the following commutation: $\Delta_+^{d-b} \alpha X_{2t} = \alpha \Delta_+^{d-b} X_{2t}$ allowing regression based estimation of α in (4).

Remark 3.6. Notice that if $b < 1$, the assumption 4.3 precludes the use of Akaike, Schwarz and other lag selection criterias to select the number of lags k in the regression (7) for ARMA type processes, since they select the lag length proportionial to $\log T$. Latter observation is important, since although assumption 4 specifies admissable growth rates, it does not solve the problem in finite samples. In case $b \geq 1$ the number of lags in finite samples can be selected using selection rules based on one of information criterias (for a comparison of lag selection rules see Kejriwal and Perron (2008)).

Remark 3.7. In specific case when the true value of ψ is known, the assumption 4 boils down to:

1. $k\left(T^{-1/2} + \log T T^{-a}\right) = o(1)$, for $a = \min\{1, 2b-1\}$,
2. $\left(\log T T^{1-b} + k\right) \sum_{|j|>k} ||\Pi_j|| = o(1)$.

and in the very special case $b = 1$ assumptions for the rate of k are: $kT^{-1/2} + k\sum_{|j|>k} ||\Pi_j|| = o(1)$ and are comparable to the assumptions in Kejriwal and Perron (2008).

Remark 3.8. In "weak" cointegration case ($b < 1/2$), optimal inference on cointegration vector has been studied in Hualde and Robinson (2007), where $\sqrt{T}$-consistent feasible estimator was derived. Although in this case inference within our framework is also possible, $\sqrt{T}$-consistency is not achievable and the best achievable rate is $\sqrt{T/k}$.

Remark 3.9. Now suppose, that instead of X_t we observe contaminated time series: $\tilde{X}_t = X_t + \xi_t$, such that assumptions 1-3 hold for X_t. The contamination term ξ_t can be interpreted as the term generated by the initial values of the series or measurement error term. Then under the following condition[3] asymptotics of $\hat{\alpha}_{DOLS}$ is not affected if we use $\tilde{X}_t$ instead of X_t in the regression (7):

$$T^{-b}\sum_{t=k}^{T-k} t^{b-1/2}(E||\Delta_+^{d-b}\xi_t||^2)^{1/2} + T^{-b}\sum_{t=k}^{T-k} E||\Delta_+^{d-b}\xi_t||^2 + kT^{-1}\sum_{t=k}^{T-k}(E||\Delta_+^{d}\xi_t||^2)^{1/2} = o(1). \quad (13)$$

Notice that assumption 1 assumes $X_0 = u_0$, i.e. the initial value of the process X_t has the same distribution as the error term u_t. Obviously, it is restrictive, but condition (13) shows that more generally we may assume that $X_0 = O_p(1)$.

4 Finite sample performance

In this section we introduce design of simulated data generating process and present simulation results. We analyze finite sample performance of proposed estimator and corresponding Wald

[3]The condition can be proved with a succession of applications of Cauchy-Schwarz inequality.

test with fractionally cointegrated bivariate time series. Error term is designed to have both contemporaneous and serial correlation effects. We present RMSE and bias of DOLS and competing estimators as well as the size of the test based on feasible test statistic.

4.1 Monte Carlo setup

We simulate bivariate fractionally cointegrated model satisfying Assumptions 1-3 without loss of generality assuming that normalized cointegration vector is $\alpha_0 = 1$:

$$X_{1t} = X_{2t} + \Delta_+^{d-b} u_{1t},$$
$$X_{2t} = \Delta_+^{d} u_{2t}$$

for $t = 1, \ldots, T$. u_t is a bivariate ARMA$(1,1)$ process:

$$u_t' = \phi(L)^{-1} \psi(L) \varepsilon_t', \;\; t = 1, \ldots, T$$

where $\phi(L) = 1 - \phi L$, $\psi(L) = 1 + \psi L$ and ε_t is Gaussian[4] i.i.d. $(0, \Omega)$ process with $\Omega = ((1,\rho)', (\rho, 1)')$. We simulate this system for the values $T = 256, 512$, $d = 0.8, 1, 1.2$, $d - b = 0, 0.2, 0.4$. The values of coefficients ρ, ψ and ϕ are presented in the Table 1. A number of 1000 Monte Carlo replications was used.

Feasible estimator is constructed in the following way: d is estimated as the memory of X_{2t} with exact local Whittle (ELW) estimator (cf. Katsumi and Phillips (2005)) maximizing over the interval $[-0.1, 2]$ with bandwidth $m = [T^{0.6}]$, while b is estimated as $\hat{d}$ minus the memory of residuals $\hat{u}_t = \hat{\beta} X_t$ which is also estimated with ELWE and the same bandwidth. Here $\hat{\beta}$ is a pre-estimate of $\beta = (1, -\alpha)$ with NBLS estimator using bandwidth $m = [T^{0.65}]$. Given remark 3.6, two lag selection rules for feasible DOLS (FDOLS) estimation were compared: Akaike information criteria (AIC) and $k = [4(T/100)^{1/4}]$, the latter being taken from Demetrescu et al. (2008), given their superior performance over information criteria-based selection rules.

Feasible DOLS estimator is compared to NBLS (using above bandwidth) and feasible NBZLS estimators (cf. Hualde and Robinson (2010))[5]. Selected bandwidth was $m = [T^{0.8}]$ where d, b are

[4]Students-t distribution with 5 d.f. was also considered, but results did not differ much.

[5]We chose to simulate simpler feasible NBZLS (”zero frequency”) estimator which does not require estimation of spectral density in the whole degenerating band in the light of Monte Carlo results in Hualde and Robinson (2006), which did not reveal significant differences in finite sample behaviour of NBZLS and NBWLS estimators.

estimated as above, whereas $\hat{f}_u(0)$ is estimated as $\hat{f}_u(0) = (2m+1)^{-1}\sum_{i=0}^{m} s_j I_{u^f}(\lambda_j)$, where I_{u^f} is a periodogram of time series $u_t^f = (\Delta_+^{\hat{d}-\hat{b}}\hat{\beta}X_t', \Delta_+^{\hat{d}}X_{2t}')'$, $s_j = 1 + \mathbb{1}_{j>0}$ and $m = [0.5 \cdot T^{0.8}]$.

We also simulated empirical size for the null $\alpha = 1$ for the above design of feasible Wald test statistic using DOLS estimate of $\hat{\beta}$ and $\hat{\Omega}$ estimated as[6] $2\pi\hat{f}_{u^f}(0)$.

Table 1: Simulation designs for weakly dependant proces u_t

Nr.	ϕ	ψ	ρ
1	0	0	0
2	0.5	0	0.5
3	0	0.5	0.5
4	0.8	0	0.8
5	0	0.8	0.8
6	0.8	0.8	0.8

4.2 Simulation results

We compare both bias and RMSE of FDOLS with AIC-based lag selection rule, NBLS and FNBZLS estimators. In terms of bias we see that, except design nr.1, FDOLS estimator generally dominates FNBZLS and NBLS estimators (the latter being always the worst out of the three). However, in case of design nr.1 NBLS is an efficient estimator, since regression errors are already "white" and results were to be expected. Overall bias tends to grow with b approaching critical value 0.5 and generally highest for design nr.6. The outcome is rather predictable, since both endogeneity and serial correlation effects are the strongest in this case and the bias induced due to biases of estimates of d, b (which are the most sensitive to changes in AR parameter) grows, while on the other hand, b approaching critical value 0.5 reduces order of stochastic dominance of signal to errors. However, even with design nr.6 FDOLS estimator tends to be more than 10 times less biased (and in some cases up to 100 times) than NBLS and more than 2 times less biased than FNBZLS. Although relative comparison of biases might not give the full picture, since they are small in absolute value, FDOLS clearly performs best of all three estimators when contemporaneous correlation in the error

[6]Unreported simulations show better properties of this estimator in compare to (12) and feasible HAC-type estimator of Ω.

Table 2: Monte Carlo simulation results

				Empirical sizes for W_{p4}			Empirical sizes for W_{AIC}			Bias$\times 10^3$			RMSE		
$Nr.$	d	$d-b$	T	0.10	0.05	0.01	0.10	0.05	0.01	FDOLS	NBLS	FNBZLS	FDOLS	NBLS	FNBZLS
	0.8	0	256	0.10	0.07	0.02	0.11	0.07	0.02	-0.67	-0.36	**0.11**	0.024	0.026	**0.021**
	0.8	0	512	0.14	0.08	0.02	0.14	0.07	0.03	-0.34	-0.22	**-0.19**	0.013	0.014	**0.012**
	0.8	0.2	256	0.20	0.14	0.07	0.21	0.15	0.07	-10.60	-1.27	**-0.07**	0.288	0.044	**0.044**
	0.8	0.2	512	0.23	0.18	0.08	0.23	0.17	0.08	-0.65	-0.45	**-0.30**	0.034	0.029	**0.028**
	1	0	256	0.11	0.07	0.02	0.13	0.07	0.02	-0.23	**0.04**	-0.07	0.010	0.013	**0.010**
	1	0	512	0.14	0.08	0.02	0.14	0.08	0.03	-0.09	-0.24	**-0.04**	0.005	0.006	**0.005**
	1	0.2	256	0.21	0.16	0.08	0.22	0.16	0.08	-0.70	**-0.18**	-0.22	0.024	0.025	**0.021**
1	1	0.2	512	0.24	0.18	0.09	0.25	0.18	0.10	-0.30	-0.51	**-0.10**	0.013	0.014	**0.012**
	1	0.4	256	0.18	0.12	0.06	0.20	0.13	0.06	-1.87	**-0.30**	0.37	0.074	0.048	**0.047**
	1	0.4	512	0.19	0.13	0.07	0.20	0.14	0.08	-0.98	-0.90	**-0.68**	0.036	0.032	**0.030**
	1.2	0	256	0.11	0.07	0.02	0.12	0.07	0.02	-0.07	**0.03**	-0.07	0.005	0.006	**0.004**
	1.2	0	512	0.13	0.08	0.03	0.13	0.08	0.03	-0.01	-0.09	**0.00**	0.002	0.003	**0.002**
	1.2	0.2	256	0.22	0.16	0.09	0.22	0.16	0.09	-0.25	**-0.02**	-0.16	0.010	0.012	**0.010**
	1.2	0.2	512	0.26	0.18	0.10	0.26	0.18	0.10	-0.07	-0.20	**-0.03**	0.005	0.006	**0.005**
	1.2	0.4	256	0.19	0.15	0.08	0.21	0.16	0.08	-0.60	**-0.01**	-0.24	0.025	0.024	**0.022**
	1.2	0.4	512	0.19	0.14	0.08	0.19	0.13	0.08	-0.41	-0.40	**-0.20**	0.013	0.014	**0.012**
	0.8	0	256	0.14	0.09	0.04	0.15	0.09	0.03	**0.29**	23.06	2.55	0.023	0.034	**0.020**
	0.8	0	512	0.16	0.10	0.03	0.17	0.11	0.04	**0.33**	14.99	0.93	0.012	0.020	**0.012**
	0.8	0.2	256	0.18	0.13	0.06	0.18	0.13	0.06	**0.55**	56.64	22.55	0.064	0.072	**0.045**
	0.8	0.2	512	0.20	0.13	0.07	0.22	0.15	0.07	**1.32**	41.20	12.87	0.035	0.050	**0.029**
	1	0	256	0.14	0.10	0.04	0.15	0.10	0.04	**-0.13**	2.23	-0.36	**0.010**	0.012	0.010
	1	0	512	0.16	0.09	0.04	0.17	0.11	0.04	**-0.01**	0.82	-0.16	**0.005**	0.006	0.005
	1	0.2	256	0.18	0.12	0.06	0.18	0.12	0.06	**-0.18**	13.68	1.42	0.023	0.027	**0.021**
2	1	0.2	512	0.20	0.13	0.06	0.20	0.14	0.06	**0.04**	7.88	0.38	0.013	0.015	**0.012**
	1	0.4	256	0.18	0.13	0.06	0.18	0.13	0.06	**-0.93**	39.96	16.65	0.062	0.061	**0.044**
	1	0.4	512	0.19	0.13	0.07	0.20	0.15	0.07	**0.45**	26.59	8.59	0.035	0.041	**0.028**
	1.2	0	256	0.14	0.09	0.03	0.14	0.08	0.03	**-0.13**	-1.73	-0.31	**0.004**	0.006	0.004
	1.2	0	512	0.15	0.09	0.03	0.16	0.09	0.04	**-0.04**	-0.97	-0.08	0.002	0.003	**0.002**
	1.2	0.2	256	0.17	0.11	0.06	0.16	0.12	0.06	**-0.34**	1.09	-0.65	**0.010**	0.011	0.010
	1.2	0.2	512	0.17	0.12	0.06	0.18	0.13	0.06	**-0.15**	0.32	-0.30	**0.005**	0.006	0.005
	1.2	0.4	256	0.16	0.11	0.06	0.16	0.11	0.06	-0.88	9.45	**0.41**	0.023	0.024	**0.021**
	1.2	0.4	512	0.17	0.11	0.06	0.18	0.12	0.06	-0.45	5.10	**-0.30**	0.013	0.014	**0.012**
	0.8	0	256	0.12	0.08	0.02	0.13	0.08	0.02	**-0.66**	26.63	2.03	0.023	0.037	**0.020**
	0.8	0	512	0.15	0.09	0.03	0.16	0.09	0.03	**-0.27**	16.64	0.43	0.012	0.022	**0.011**
	0.8	0.2	256	0.20	0.14	0.07	0.19	0.15	0.07	**-0.20**	61.63	24.63	0.062	0.076	**0.046**
	0.8	0.2	512	0.22	0.16	0.08	0.24	0.17	0.08	**1.82**	43.76	14.47	0.035	0.053	**0.029**
	1	0	256	0.14	0.08	0.03	0.14	0.08	0.03	**-0.43**	2.90	-0.76	**0.009**	0.013	0.009
	1	0	512	0.15	0.09	0.03	0.16	0.10	0.03	**-0.16**	1.04	-0.36	**0.005**	0.006	0.005
	1	0.2	256	0.20	0.15	0.08	0.20	0.15	0.08	**-0.10**	14.98	2.11	0.023	0.028	**0.020**
3	1	0.2	512	0.24	0.16	0.08	0.23	0.17	0.08	**0.26**	8.40	0.89	0.012	0.016	**0.012**
	1	0.4	256	0.20	0.15	0.08	0.21	0.16	0.08	**-1.06**	41.86	17.81	0.060	0.062	**0.044**
	1	0.4	512	0.21	0.15	0.08	0.23	0.17	0.09	**0.41**	27.42	9.39	0.034	0.041	**0.029**
	1.2	0	256	0.13	0.08	0.03	0.13	0.08	0.02	**-0.26**	-1.73	-0.41	0.004	0.006	**0.004**
	1.2	0	512	0.14	0.09	0.03	0.15	0.10	0.03	**-0.07**	-0.97	-0.12	0.002	0.003	**0.002**
	1.2	0.2	256	0.20	0.14	0.08	0.21	0.15	0.08	**-0.22**	1.27	-0.26	0.010	0.011	**0.009**
	1.2	0.2	512	0.23	0.15	0.08	0.23	0.17	0.08	**0.00**	0.37	-0.09	**0.005**	0.006	0.005
	1.2	0.4	256	0.19	0.15	0.08	0.20	0.14	0.08	-1.01	9.83	**0.89**	0.023	0.024	**0.020**
	1.2	0.4	512	0.21	0.13	0.07	0.22	0.15	0.07	-0.33	5.23	**-0.09**	0.013	0.014	**0.012**

Note: Bold font signifies the smallest number of the three in absolute value.

Table 3: Monte Carlo simulation results

				Empirical sizes for W_{p4}			Empirical sizes for W_{AIC}			Bias$\times 10^3$			RMSE		
$Nr.$	d	$d-b$	T	0.10	0.05	0.01	0.10	0.05	0.01	FDOLS	NBLS	FNBZLS	FDOLS	NBLS	FNBZLS
	0.8	0	256	0.15	0.08	0.03	0.14	0.09	0.04	**0.65**	25.90	3.66	**0.023**	0.033	0.023
	0.8	0	512	0.15	0.08	0.04	0.16	0.10	0.04	**0.52**	17.67	1.47	**0.012**	0.021	0.013
	0.8	0.2	256	0.22	0.14	0.06	0.20	0.15	0.07	**5.70**	73.83	34.85	0.060	0.084	**0.051**
	0.8	0.2	512	0.22	0.14	0.06	0.26	0.17	0.08	**5.62**	55.17	22.01	0.033	0.061	**0.033**
	1	0	256	0.10	0.06	0.03	0.09	0.05	0.03	**-0.59**	1.82	-1.08	**0.009**	0.010	0.012
	1	0	512	0.10	0.06	0.02	0.10	0.06	0.02	**-0.25**	0.77	-0.56	**0.004**	0.005	0.006
	1	0.2	256	0.14	0.08	0.03	0.13	0.08	0.03	**0.09**	17.60	2.53	**0.022**	0.026	0.023
4	1	0.2	512	0.14	0.08	0.04	0.15	0.09	0.04	**0.20**	10.85	0.86	**0.012**	0.016	0.013
	1	0.4	256	0.22	0.13	0.05	0.20	0.13	0.07	**4.53**	56.70	28.65	0.058	0.069	**0.048**
	1	0.4	512	0.22	0.14	0.06	0.24	0.17	0.08	**4.72**	39.30	17.29	0.032	0.048	**0.031**
	1.2	0	256	0.07	0.04	0.01	0.07	0.03	0.01	**-0.35**	-2.50	-0.64	**0.004**	0.005	0.005
	1.2	0	512	0.07	0.04	0.02	0.07	0.04	0.02	**-0.14**	-1.36	-0.23	**0.002**	0.003	0.002
	1.2	0.2	256	0.09	0.05	0.02	0.08	0.05	0.02	**-0.76**	1.35	-1.28	0.009	**0.009**	0.012
	1.2	0.2	512	0.09	0.05	0.02	0.09	0.05	0.02	**-0.34**	0.51	-0.70	**0.004**	0.005	0.006
	1.2	0.4	256	0.13	0.08	0.03	0.12	0.07	0.03	**-0.54**	13.73	1.61	**0.022**	0.023	0.023
	1.2	0.4	512	0.14	0.07	0.04	0.14	0.08	0.03	**-0.12**	7.87	0.35	**0.012**	0.014	0.013
	0.8	0	256	0.14	0.08	0.03	0.12	0.08	0.03	**-1.01**	42.72	2.40	0.018	0.049	**0.016**
	0.8	0	512	0.17	0.10	0.04	0.16	0.11	0.04	-0.52	26.75	**0.41**	0.010	0.030	**0.010**
	0.8	0.2	256	0.23	0.15	0.09	0.22	0.16	0.09	**2.21**	99.37	36.79	0.050	0.108	**0.049**
	0.8	0.2	512	0.27	0.17	0.09	0.29	0.21	0.10	**2.75**	70.30	21.93	**0.029**	0.076	0.031
	1	0	256	0.16	0.09	0.04	0.15	0.09	0.03	**-0.38**	4.58	-0.80	**0.007**	0.011	0.008
	1	0	512	0.17	0.11	0.04	0.17	0.11	0.05	**-0.12**	1.85	-0.38	**0.003**	0.005	0.004
	1	0.2	256	0.24	0.16	0.09	0.22	0.16	0.09	**1.15**	24.05	3.59	0.017	0.032	**0.017**
5	1	0.2	512	0.25	0.18	0.10	0.26	0.19	0.10	**0.92**	13.82	1.79	**0.010**	0.018	0.010
	1	0.4	256	0.27	0.19	0.11	0.26	0.19	0.11	**5.03**	67.14	27.70	0.046	0.079	**0.043**
	1	0.4	512	0.30	0.22	0.12	0.32	0.24	0.13	**4.11**	44.55	16.19	**0.027**	0.053	0.028
	1.2	0	256	0.15	0.08	0.03	0.13	0.08	0.03	**-0.21**	-2.79	-0.39	**0.003**	0.006	0.003
	1.2	0	512	0.16	0.10	0.03	0.17	0.11	0.04	**-0.07**	-1.49	-0.13	**0.001**	0.003	0.001
	1.2	0.2	256	0.23	0.16	0.08	0.21	0.15	0.08	0.19	2.03	**0.18**	**0.007**	0.009	0.008
	1.2	0.2	512	0.23	0.16	0.08	0.23	0.17	0.09	0.15	0.75	**0.10**	**0.004**	0.005	0.004
	1.2	0.4	256	0.24	0.16	0.10	0.22	0.16	0.09	**0.67**	15.72	2.55	**0.017**	0.025	0.017
	1.2	0.4	512	0.24	0.18	0.09	0.25	0.18	0.09	**0.42**	8.68	0.88	**0.010**	0.014	0.010
	0.8	0	256	0.14	0.08	0.03	0.13	0.08	0.04	**1.06**	25.63	3.34	**0.023**	0.033	0.024
	0.8	0	512	0.13	0.08	0.03	0.15	0.09	0.04	**0.58**	17.53	1.34	**0.012**	0.021	0.013
	0.8	0.2	256	0.21	0.14	0.05	0.20	0.14	0.08	**7.33**	73.41	33.61	0.059	0.083	**0.051**
	0.8	0.2	512	0.22	0.14	0.06	0.26	0.18	0.09	**5.93**	54.94	21.74	0.033	0.061	**0.033**
	1	0	256	0.08	0.05	0.02	0.09	0.06	0.02	**-0.56**	1.80	-1.11	**0.009**	0.010	0.012
	1	0	512	0.09	0.05	0.02	0.09	0.05	0.02	**-0.26**	0.76	-0.59	**0.004**	0.005	0.006
	1	0.2	256	0.13	0.08	0.03	0.12	0.08	0.03	**0.45**	17.54	2.36	**0.022**	0.026	0.024
6	1	0.2	512	0.14	0.08	0.04	0.15	0.09	0.03	**0.35**	10.81	0.90	**0.012**	0.016	0.013
	1	0.4	256	0.21	0.13	0.05	0.18	0.13	0.07	**5.77**	56.59	27.60	0.056	0.069	**0.048**
	1	0.4	512	0.21	0.15	0.06	0.24	0.17	0.08	**5.01**	39.23	17.04	0.032	0.048	**0.031**
	1.2	0	256	0.06	0.03	0.01	0.06	0.03	0.01	**-0.36**	-2.48	-0.63	**0.004**	0.005	0.005
	1.2	0	512	0.06	0.04	0.01	0.06	0.04	0.02	**-0.14**	-1.35	-0.24	**0.002**	0.003	0.002
	1.2	0.2	256	0.08	0.05	0.02	0.08	0.05	0.02	**-0.65**	1.37	-1.31	0.009	**0.009**	0.012
	1.2	0.2	512	0.09	0.05	0.02	0.09	0.05	0.02	**-0.33**	0.51	-0.68	**0.004**	0.005	0.006
	1.2	0.4	256	0.13	0.07	0.02	0.11	0.07	0.03	**-0.32**	13.75	1.33	**0.022**	0.023	0.024
	1.2	0.4	512	0.14	0.07	0.03	0.14	0.08	0.03	**-0.04**	7.87	0.31	**0.012**	0.014	0.013

Note: Bold font signifies the smallest number of the three in absolute value.

term u_t is present.

In terms of RMSE, FDOLS estimator performs very much on par with FNBZLS, except for the cases with $b = 0.6$ and sample size $T = 256$, while it performs clearly better than NBLS in all designs, except nr.1 when NBLS is efficient, with RMSE generally being up to 2 times smaller (and in some extreme cases more than 3). The cases with $b = 0.6$ and sample size $T = 256$ are the most problematic for FDOLS - stochastic dominance of signal to errors is close to critical value of 0.5 and improper selection of number of lags in the regression distorts the signal enough to affect the estimate considerably. However, it is reassuring to see drop of RMSE of FDOLS considerably with $T = 512$, it being comparable to that of FNBZLS. Summing up, we may conclude that in terms of RMSE it performs better than NBLS and on par with NBZLS, although it does seem to underperform with (jointly) small sample size, b close to 0.5 and no contemporaneous correlation effects.

As far as the empirical test sizes are concerned, both lag selection rules tend to have very similar sizes and none of them could be more preferred to the other. Generally, empirical sizes do not seem to vary much across designs (except maybe designs nr.5, nr.6), as a rule both of them being oversized. Overall the sizes approach nominal values with increase of b, but more worringly depend little on the sample size. Our (unreported) simulations of empirical size of infeasible DOLS estimator using ψ_0 (but $\hat{\Omega}$ estimated as described above) show very good properties and we conjecture, that uncertainty coming from the estimates of ψ is more relevant for the test statistic than the uncertainty of estimate of $\Omega_{11\cdot 2}$. Given that the above-described ELW estimator showed little to no improvement in terms of RMSE with sample size increasing from $T = 256$ to $T = 512$, that most likely explains little effect of increase of sample size to the empirical size. Hence, improvement in empirical size is expected using more advanced estimation methods of ψ (using adaptive optimal bandwidth selection procedure or an estimator approximating spectral density in degenerating band with a polynomial rather than a constant), however we do not explore those options in this study. We may conclude that given complexity of estimation procedure, feasible Wald test size shows satisfactory properties.

5 Conclusions

The paper established feasible dynamic OLS type procedure for optimal estimation of cointegration vector in "strongly" cointegrated fractional regressions with LAMN distribution of the estimator. Finite sample simulations show superior properties over NBLS estimator and rivals NBWLS estimator due to Hualde and Robinson (2010). Thus the estimator may be seen as a feasible time domain alternative to semiparametric frequency domain narrow-band estimation. Monte Carlo simulations report feasible Wald test being generally oversized, but our (unreported) simulations show that the infeasible statistic has very good size properties for the sample sizes considered, suggesting that uncertainty coming from estimation of ψ is accounted for the size problems, hence more sophisticated estimation techniques with optimal bandwidth selection could render better test properties.

Acknowledgements

Comments and help from Peter Boswijk, Søren Johansen and Morten Nielsen is grately acknowledged. Additional thanks goes to the administration of Dutch national computer cluster "LISA" and CREATES in Aarhus, where the research was partly carried out.

Appendices

A Identification of FCI system

Suppose we have a fractionally cointegrated system X_t satisfying assumptions 1-3 with $n \times r$ cointegration vector β and parameters $\psi = (d, b)$, $R = (\beta, \gamma)$ for some process u_t. It is not obvious that the cointegration vector β could be identified as $\beta' = (I_r, -\alpha)$ and we show that this is indeed true, i.e. there exists a $r \times (n-r)$ matrix α and a process $\tilde{u}_t$ satisfying assumption 3 such that:

$$\Delta(L, \psi) \begin{pmatrix} I_r & -\alpha \\ 0_{n-r \times r} & I_{n-r} \end{pmatrix} X_t = \tilde{u}_t.$$

We can write matrix R as a block matrix:

$$R = \begin{pmatrix} R_1 & R_2 \\ R_3 & R_4 \end{pmatrix}.$$

Then from different representation of X_t we have equality:

$$\begin{pmatrix} I_r\Delta_+^{d-b} & 0_{r\times n-r} \\ 0_{n-r\times r} & I_{n-r}\Delta_+^{d} \end{pmatrix} \begin{pmatrix} R_1 & R_2 \\ R_3 & R_4 \end{pmatrix} \begin{pmatrix} I_r & -\alpha \\ 0_{n-r\times r} & I_{n-r} \end{pmatrix} \begin{pmatrix} I_r\Delta_+^{-d+b} & 0_{r\times n-r} \\ 0_{n-r\times r} & I_{n-r}\Delta_+^{-d} \end{pmatrix} \tilde{u}_t = u_t.$$

Take $\alpha = R_1^{-1}R_2$. Then:

$$\begin{pmatrix} I_r\Delta_+^{d-b} & 0_{r\times n-r} \\ 0_{n-r\times r} & I_{n-r}\Delta_+^{d} \end{pmatrix} \begin{pmatrix} R_1 & R_2 \\ R_3 & R_4 \end{pmatrix} \begin{pmatrix} I_r & \alpha \\ 0_{n-r\times r} & I_{n-r} \end{pmatrix} \begin{pmatrix} I_r\Delta_+^{-d+b} & 0_{r\times n-r} \\ 0_{n-r\times r} & I_{n-r}\Delta_+^{-d} \end{pmatrix} = \begin{pmatrix} R_1 & 0_{r\times n-r} \\ \Delta_+^b R_3 & R_4 - R_3R_1^{-1}R_2 \end{pmatrix}.$$

Hence we get:

$$\tilde{u}_t = \begin{pmatrix} R_1^{-1} & 0_{r\times n-r} \\ \Delta_+^b(R_4 - R_3R_1^{-1}R_2)^{-1}R_3 & (R_4 - R_3R_1^{-1}R_2)^{-1} \end{pmatrix} u_t = A(L)C(L)\varepsilon_t.$$

Since for lag polynomials' $A(L), C(L)$ coefficients' it holds: $\sum_i \sqrt{i}(||A_i|| + ||C_i||) < \infty$, their convolution also satisfies this property and thus the assumption 3 also holds for $\tilde{u}_t$.

B Notation

We introduce notation for our subsequent analysis. We denote the true parameter ψ value as ψ_0 and decompose the error term into the error obtained because of the use of estimate of ψ_0 and the error due to autoregressive approximation:

$$\tilde{v}_t = \Delta_+^{d-b}X_{1t} - \alpha\Delta_+^{d-b}X_{2t} - \sum_{i=-k}^{k}\Pi_i\Delta_+^{d}X_{2t-i} = \Delta_+^{d-b}\left(X_{1t} - \alpha X_{2t}\right) - \Delta_+^{d-d_0}\left(\sum_{i=-k}^{k}\Pi_i\Delta_+^{d_0}X_{2t-i}\right) =$$

$$= \Delta_+^{d-b-d_0+b_0}u_{1t} - \Delta_+^{d-d_0}(u_{1t} - v_t - \sum_{|j|>k}\Pi_j\Delta_+^{d_0}X_{2t-j}) = v_t + e_{1t} + e_{2t},$$

where:

$$e_{1t}(\psi) = \sum_{|j|>k}\Pi_j\Delta_+^{d}X_{2t},$$

$$e_{2t}(\psi) = (\Delta_+^{d-b-d_0+b_0} - 1)u_{1t} - (\Delta_+^{d-d_0} - 1)\sum_{j\in\mathbb{Z}}\Pi_j u_{2t-j}.$$

Also additionally denote the following variables depending on $\psi = (d, b)$:

$$\begin{aligned}
w_{1t}(\psi) &= \Delta_+^{d-b} X_{2t}, \\
w_{2t}(\psi) &= (\Delta_+^d X'_{2t+k}, \dots, \Delta_+^d X'_{2t-k})', \\
w_t(\psi) &= (w'_{1t}, w'_{2t})', \\
W(\psi) &= (w_k, \dots, w_{T-k}), \\
\Gamma_{ij,T}(\psi) &= T^{-1} \sum_{t=k}^{T-k} \Delta_+^d X_{2t+i} \Delta_+^d X'_{2t-j}, \\
\Gamma_{ij,\infty}(\psi) &= AsCov(\Delta_+^d X_{2t+i}, \Delta_+^d X_{2t-j}), \\
\Gamma_k(\psi) &= AsCov(\omega_{2t}(\psi)) = (\Gamma_{ij,\infty})_{i,j=-k..k}, \\
D_T(\psi, n_T) &= blockdiag\{I_r T^{-b}, I_{n-r} n_T^{-1}, \dots, I_{n-r} n_T^{-1}\}, \\
R_T(\psi) &= D_T(\psi, T^{1/2}) W W' D_T(\psi, T^{1/2}), \\
R_\infty(\psi) &= blockdiag\{T^{-2b} \sum_{t=k}^{T-k} w_{1t}(\psi) w'_{1t}(\psi), \Gamma_k(\psi)\}.
\end{aligned}$$

We also denote matrices $\tilde{V}, V, E_1, E_2, W_1, W_2$ which are column-by-column stacked $\tilde{v}_t$'s, v_t's, e_{it}'s and w_{it}'s with indices running from k to $T-k$, e.g. $V = (v_k, \dots, v_{T-k})$.

C Additional lemmas

The following lemma gives two criterias for a probabilistic bound of a sum of stochastically bounded random variables depending on a parameter $\theta \in \Theta$ with growing number of terms[7].

Lemma C.1. *Suppose we have a collection of random variables $\{X_T^i(\theta), i, T \in \mathbb{N}\}$, $\theta_0 \in \Theta$, where Θ is a compact subset of Euclidian space $\mathbb{R}^s$ and a deterministic sequence $\{a_i, i \in \mathbb{N}\}$. Also suppose $\theta_T \xrightarrow{P} \theta_0 \in int(\Theta)$ and consider any open neighbourhood N_{θ_0} of θ. Then $\sum_{j=1}^{k_T} a_j X_T^j(\theta_T) = O_p(\sum_{j=1}^{k_T} |a_j|)$ for any $k_T \to \infty$, if one of the two holds:*

$$\sup_{\theta \in N_{\theta_0}} \sup_{i \in \mathbb{N}} E(X_T^i(\theta))^2 = O(1), \tag{14}$$

$$\sup_{\theta \in N_{\theta_0}} \sup_{i \in \mathbb{N}} |X_T^i(\theta)| = O_p(1). \tag{15}$$

[7] In Kejriwal and Perron (2008) this argument is missed (say, lemma A.1 (ii)): k_T terms of order $O_p(1)$ may not add up to $O_p(k_T)$ random variable for $k_T \to \infty, T \to \infty$. For example, take $\{X_T^i(u) = T \mathbb{1}_{u \in [i-1/k_T, i/k_T]}\}$, $i = 1, \dots, k_T$ for $k_T = o(T)$. Then $\sum_{i=1}^{k_T} X_T^i = T$.

Neither condition is stronger than the other. If $\theta_T = \theta_0$, the first suprema is vacuous. Also similar proposition holds for small "o".

Proof. Take any $\varepsilon_1, M > 0$. Then from Chebyshev's inequality:

$$P\left(\frac{|\sum_{i=1}^{k_T} a_i X_T^i(\theta_T)|}{\sum_{j=1}^{k_T}|a_j|} \geq M\right) \leq P(\theta_T \notin N_{\theta_0}) +$$

$$\leq P\left(\theta_T \in N_{\theta_0}, |\sum_{i=1}^{k_T} a_i(X_T^i(\theta_T) - EX_T^i(\theta_T))| \geq M\sum_{j=1}^{k_T}|a_j| - \sum_{i=1}^{k_T}|a_i|E|X_T^i(\theta_T)|\right)$$

$$\leq P(\theta_T \notin N_{\theta_0}) + \left(M\sum_{j=1}^{k_T}|a_j| - \sum_{i=1}^{k_T}|a_i|E|X_T^i(\theta_T)|\right)^{-2} Var\left(\sum_{i=1}^{k_T} a_i X_T^i(\theta_T)\right)$$

$$\leq P(\theta_T \notin N_{\theta_0}) + \left(M\sum_{j=1}^{k_T}|a_j| - \sum_{i=1}^{k_T}|a_i|\sqrt{E|X_T^i(\theta_T)|^2}\right)^{-2}\left(\sum_{i=1}^{k}|a_i|\sqrt{VarX_T^i(\theta_T)}\right)^2$$

$$\leq P(\theta_T \notin N_{\theta_0}) + \left(M - \sup_{\theta \in N_{\theta_0}}\sup_{i \in \mathbb{N}}\sqrt{E|X_T^i(\theta_T)|^2}\right)^{-2}\left(\sup_{\theta \in N_{\theta_0}}\sup_{i \in \mathbb{N}}\sqrt{VarX_T^i(\theta)}\right)^2 < \varepsilon_1$$

for every T large enough. On the other hand:

$$P\left(\frac{|\sum_{j=1}^{k_T} a_j X_T^j(\theta_T)|}{\sum_{j=1}^{k_T}|a_j|} \geq M\right) \leq P(\theta_T \notin N_{\theta_0}) + P\left(\theta_T \in N_{\theta_0}, |\sum_{j=1}^{k_T} a_j X_T^j(\theta_T)| \geq M\sum_{j=1}^{k_T}|a_j|\right)$$

$$\leq P(\theta_T \notin N_{\theta_0}) + P\left(\theta \in N_{\theta_0}, \sup_{\theta \in N_{\theta_0}}\sup_{i \in \mathbb{N}}|X_T^i(\theta)| \geq M\right) < \varepsilon_1$$

for every T large enough.

To show that neither condition is stronger consider an example of two sequences of random variables on the interval $[0,1]$ with Borel measure: $\{X_T^i(u) = \sqrt{T}\mathbb{1}_{u \in [(i-1)/T, i/T]}\}$ and $\{X_T^i(u) = T\mathbb{1}_{u \in [0,1/T]}\}$. Similarly we prove the proposition for small "o". □

The following lemma specifies the neighbourhood of the true value in which criteria (14) hold for relevant norm moments, what is at the core of derivation of probabilistic bounds of various terms. Lemma derives from the results in Johansen and Nielsen (2010a):

Lemma C.2. *Denote a η_0-neighbourhood around the true value $\psi_0 = (d_0, b_0)$: $N(\eta_0) = \{\psi : ||\psi - \psi_0|| < \eta_0\}$ for some* $\min\{1/2, b_0 - 1/2\} > \eta_0 > 0$. *Assume that the process u_t satisfies* *assumption 3 and consider linear processes:*

$$V_t^0(\psi) = \Delta_+^{d-d_0} u_t,$$
$$V_t^1(\psi) = T^{-b+0.5}\Delta_+^{-b} u_t.$$

Then for the product moments:

$$S_{ij,T}^{ab}(\psi) = T^{-1}\sum_{t=k}^{T-k} V_{t+i}^a(\psi)V_{t-j}^{b\prime}(\psi),$$

it holds,[8] *as* $k = k_T \to \infty$*:*

$$\sup_{\psi\in N(\eta_0)}\sup_{i,j\in\mathbb{Z}} E\left|\left|S_{ij,T}^{00}(\psi)\right|\right|^2 = O(1), \tag{16}$$

$$\sup_{\psi\in N(\eta_0)}\sup_{i,j\in\mathbb{Z}} E\left|\left|D_\psi S_{ij,T}^{00}(\psi)\right|\right|^2 = O(1), \tag{17}$$

$$\sup_{\psi\in N(\eta_0)}\sup_{i\in\mathbb{Z}} E\left|\left|S_{0i,T}^{10}(\psi)\right|\right|^2 = O(\log^2 T T^{-a+2\eta_0}),\ \textit{for } a = \min\{1, 2b_0 - 1\}, \tag{18}$$

$$\sup_{i\in\mathbb{Z}} E\left|\left|S_{0i,T}^{10}(\psi_0)\right|\right|^2 = O(\log^2 T T^{-a}),\ \textit{for } a = \min\{1, 2b_0 - 1\}. \tag{19}$$

Proof. To keep things somewhat clearer, in the proof we will use L_q-norm for random variables: $||X||_q = (E||X||^q)^{1/q}$. Then:

$$||S_{ij,T}^{00}(\psi)||_2 \le ||S_{00,T}^{00}(\psi)||_2 \le T^{-1}\sum_{i=k}^{T-k}||V_t^0(\psi)V_t^{\prime 0}(\psi)||_2 \le T^{-1}\sum_{i=k}^{T-k}||V_t^0(\psi)||_4||V_t^0(\psi)||_4 \le$$
$$\le K_1^2 T^{-1}\sum_{i=k}^{T-k}||V_t^0(\psi)||_2||V_t^0(\psi)||_2 \le K_1^2K_2, \forall i, i \in \mathbb{N}, \forall\psi \in N(\eta_0).$$

The last inequalities follow from multivariate extensions of Lemma B.1 and Lemma C.4 in Johansen and Nielsen (2010a) applied to the process $V_t^0(\psi)$. Similarly, we prove (17). (19) follows from multivariate extension of Lemma C.5 in Johansen and Nielsen (2010a). The same lemma could be used to prove (18) upon noting that the coefficients of $V_t^1(\psi)$ can be bounded uniformly in $N(\eta_0)$ by the coefficients of $V_t^1(\psi_0 + \eta_0) = T^{-b-\eta_0+0.5}\Delta_+^{-b-\eta_0}u_t$. □

In the rest of the section we prove various bounds and convergence results for the terms depending on the estimator $\hat\psi$ and since $\hat\psi - \psi_0 = o_p(1)$, in the following we implicitly assume that $\hat\psi$

[8] As a convention, for $i + j > T$ we will assume $S_{ij,T}^{ab}(\psi) = 0$.

is in the η_1-neighbourhood $N(\eta_1)$ of ψ_0 "small enough" for lemma C.2 to hold for various fractional processes whose order of fractionality depends on $\hat{\psi}$:

$$N(\eta_1) \subset N(\eta_0), \;\; N(\eta_1) \subset \{\psi = (d, b) : d - b - d_0 < 1/2, -1/2 < d - b - d_0 + b_0\}.$$

Lemma C.3. *Assume that assumptions 1-4 hold. Then* $\sup_{i \in \mathbb{Z}} ||S^{10}_{0i,T}(\hat{\psi})|| = O_p(\log^2 T T^{-a})$, *for* $a = \min\{1, 2b_0 - 1\}$.

Proof. Mean value theorem gives:

$$||S^{10}_{0i,T}(\hat{\psi})|| \leq ||S^{10}_{0i,T}(\psi_0)|| + ||\hat{\psi} - \psi_0|| ||D_\psi S^{10}_{0i,T}(\bar{\psi})||,$$

for some $||\bar{\psi} - \psi_0|| \leq ||\hat{\psi} - \psi_0||$. Notice that the coefficients of $D_\psi V^1_t(\psi)$ can be bounded uniformly in $N(\eta_1)$ by the coefficients of $V^1_t(b_0 + \eta_1)$ and since $||\hat{\psi} - \psi_0|| = O_p(T^{-\kappa})$ for some $\kappa > 0$, the second term is of smaller order than the first and the bound follows from Lemma C.2. □

Lemma C.4. *Assume that assumptions 1-4 hold and denote:*

$$S(\psi) = T^{-b} \sum_{t=1}^{T} v_t \Delta_+^{d-b} X'_{2t}. \tag{20}$$

Then it holds:

$$\left(T^{-1/2}\Delta_+^{-1} v_{[Ts]}, T^{-\hat{b}+1/2}\Delta_+^{\hat{d}-\hat{b}} X_{2[Ts]}, S(\hat{\psi})\right) \Rightarrow \left(W_{1\cdot 2}(s), W_{b_0-1}(s), \int_0^1 dW_{1\cdot 2} W'_{b_0-1}\right). \tag{21}$$

Proof. Weak convergence of the first component follows from the fact that v_t is a stationary (double-sided) linear time series with absolutely summable coefficients, while convergence of the second component follows from the tightness of $T^{-u+1/2}\Delta_+^{-u} u_{2[Ts]}$ in u (cf. Johansen and Nielsen (2010a)) and fractional invariance principle (3). We prove convergence of the third component. Beveridge-Nelson decomposition for v_t:

$$v_t = (I_r, 0_{r \times n-r})C(L)\varepsilon_t - \Pi(L)(0_{r \times r}, I_{n-r})C(L)\varepsilon_t = \xi(L)\varepsilon_t = \xi(1)\varepsilon_t + \Delta\bar{\xi}(L)\varepsilon_t$$

Here $\xi(L), \bar{\xi}(L)$ are double sided filters. Since $\Pi(e^{-i\lambda}) = f_{12}(\lambda) f_{22}^{-1}(\lambda)$ and coefficients of $C(L)$, $C^{-1}(L)$ are 1/2-summable, that implies 1/2-summability of coefficients of $\xi(L)$ and square summability of coeficients of $\bar{\xi}(L)$ (cf. Phillips and Solo (1992)). Then we decompose $S(\psi)$ as follows:

$$S(\hat{\psi}) = T^{-\hat{b}} \sum_{t=1}^{T} v_t \Delta_+^{\hat{d}-\hat{b}} X'_{2t} = T^{-\hat{b}} \sum_{t=1}^{T} \xi(1)\varepsilon_t (\Delta_+^{\hat{d}-\hat{b}-d_0} - 1) u'_{2t} + T^{-\hat{b}} \bar{\xi}(L)\varepsilon_T (\Delta_+^{\hat{d}-\hat{b}-d_0} - 1) u'_{2T} +$$
$$T^{-\hat{b}} \sum_{t=1}^{T} \bar{\xi}(L)\varepsilon_t (\Delta_+^{1+\hat{d}-\hat{b}-d_0} - 1) u'_{2t} + T^{-\hat{b}} \sum_{t=1}^{T} v_t u'_{2t} = A_1 + A_2 + A_3 + A_4.$$

Now due to uncorrelatedness of v_t and u_{2t}, we have $ES(\hat{\psi}) = EA_1 = EA_4 = 0$. On the other hand $\Delta_+^{\hat{d}-\hat{b}-d_0} u'_{2T} = O_p(T^{\hat{d}-\hat{b}-d_0-1/2})$ uniformly in $N(\eta_1)$, while $T^{-1/2}\bar{\xi}(L)\varepsilon_T = o_p(1)$ and hence $A_2 = o_p(1)$. That implies $EA_2 = o(1)$ and hence $EA_3 = ES(\hat{\psi}) - EA_1 - EA_2 - EA_4 = o(1)$. Now, if $b_0 \leq 3/2$, then central limit theorem (CLT) implies $A_3 = O_p(T^{-b_0+1/2}) = o_p(1)$, whereas if $b_0 > 3/2$ Chebyshev's inequality gives $A_3 = O_p(T^{-1/2})$. Finally, CLT implies $A_4 = O_p(T^{-b_0+1/2})$ and thus $S(\hat{\psi}) = A_1 + o_p(1)$. The convergence of A_1 to stochastic integral follows from application of Theorem 2.2 in Kurtz and Protter (1991). Finally, since $\xi(1) = (I_r, -\Omega_{12}\Omega_{22}^{-1})C(1)$, we have that $Var(\xi(1)\varepsilon_t) = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} = \Omega_{11\cdot 2}$ and the lemma is proved. □

Lemma C.5. *Suppose, $||\hat{\psi} - \psi_0|| = O_p(T^{-\kappa}), \kappa > 0$ and assumptions 1-4 hold. Then:*

1. $||T^{-1}W_2(\hat{\psi})W_2(\hat{\psi})' - \Gamma_k(\psi_0)||^2 = O_p(k^2(T^{-2\kappa} + T^{-1})) = o_p(1)$,
2. $||T^{-\hat{b}-0.5}W_1(\hat{\psi})W_2'(\hat{\psi})||^2 = O_p(kT^{-a}\log^2 T) = o_p(1)$,
3. $||T^{-\hat{b}}E_1(\hat{\psi})W_1'(\hat{\psi})||^2 = O_p(\log^2 T T^{1-a}) \left(\sum_{|j|>k} ||\Pi_j||\right)^2 = o_p(1)$,
4. $||T^{-1}E_1(\hat{\psi})W_2'(\hat{\psi})||^2 = O_p(k) \left(\sum_{|j|>k} ||\Pi_j||\right)^2 = o_p(1)$,
5. $||T^{-\hat{b}}E_2(\hat{\psi})W_1'(\hat{\psi})||^2 = o_p(1)$,
6. $||T^{-1}E_2(\hat{\psi})W_2'(\hat{\psi})||^2 = O_p(k^2T^{-2\kappa}) = o_p(1)$,
7. $||T^{-\hat{b}}VW_1'(\hat{\psi})||^2 = O_p(1)$,
8. $||T^{-1}VW_2'(\hat{\psi})||^2 = O_p(kT^{-1})$,

with $a = \min\{1, 2b_0 - 1\}$.

Proof. In the proof we will continuously apply Lemma C.1 for big "O" in combination with Lemma C.2: we bound the growing sum of either expectations of squared norms of product moments applying condition (14) or sum of norms applying (15). We will apply mean-value theorem for the product moments and will use short hand notation for that, e.g. $S(\psi) = S(\psi_0) + (\psi - \psi_0)D_\psi S(\bar{\psi})$, meaning that $[S(\psi)]_{ij} = [S(\psi_0)]_{ij} + (\psi - \psi_0)'D_\psi[S(\bar{\psi})]_{ij}$ for some $||\bar{\psi} - \psi|| \leq ||\psi_0 - \psi||$.

Proof of (1):

$$||T^{-1}W_2(\hat{\psi})W_2(\hat{\psi})' - \Gamma_k(\psi_0)|| \leq \sum_{i,j=-k}^{k} ||S_{ij,T}^{00}(\hat{\psi}) - \Gamma_{ij,\infty}(\psi_0)||^2 \leq 2\sum_{i,j=-k}^{k} (||A_{ij,T}||^2 + ||B_{ij,T}||^2),$$

where $A_{ij,T} = S^{00}_{ij,T}(\psi_0) - \Gamma_{ij,\infty}(\psi_0)$, $B_{ij,T} = S^{00}_{ij,T}(\hat{\psi}) - S^{00}_{ij,T}(\psi_0)$.

From Hannan (1974) we have that $\sup_{i,j\in\mathbb{Z}} ||A_{ij,T}|| = o_p(1)$ and since $||A_{ij,T}|| = O_p(T^{-1/2})$ we have $\sum_{i,j=-k}^{k} ||A_{ij,T}||^2 = O_p(k^2T^{-1})$. Next applying mean-value theorem we get: $||B_{ij,T}|| \leq ||\hat{\psi} - \psi_0||||D_\psi S^{00}_{ij,T}(\bar{\psi})||$ and from Lemma C.2 we find that $\sum_{i,j=-k}^{k} ||B_{ij,T}||^2 = O_p(k^2T^{-2\kappa})$. Thus:

$$||T^{-1}W_2(\hat{\psi})W_2(\hat{\psi})' - \Gamma_k(\psi_0)||^2 = O_p(k^2(T^{-2\kappa} + T^{-1})).$$

Proof of (2):

$$||T^{-\hat{b}-0.5}W_1(\hat{\psi})W_2(\hat{\psi})'||^2 = \sum_{i=-k}^{k} ||T^{-\hat{b}-0.5}\sum_{t=k}^{T-k} \Delta_+^{\hat{d}-\hat{b}}X_{2t}\Delta_+^{\hat{d}}X'_{2t-i}||^2 \leq \sum_{i=-k}^{k} ||S^{10}_{0i,T}(\hat{\psi})||^2 = O_p(k\log^2 TT^{-a}).$$

Similarly proof of (3):

$$||T^{-\hat{b}}E_1(\hat{\psi})W_1'(\hat{\psi})|| \leq ||T^{-\hat{b}}\sum_{t=k}^{T-k}\sum_{|j|>k}\Pi_j\Delta_+^{\hat{d}}X_{2t+j}\Delta_+^{\hat{d}-\hat{b}}X'_{2t}|| \leq \sqrt{T}\sum_{|j|>k}||\Pi_j||||T^{-\hat{b}-1/2}\sum_{t=k}^{T-k}\Delta_+^{\hat{d}}X_{2t+j}\Delta_+^{\hat{d}-\hat{b}}X'_{2t}||$$

$$\leq \sqrt{T}\sum_{|j|>k}||\Pi_j||||S^{01}_{j0,T}(\hat{\psi})|| = O_p(\log TT^{1/2-a/2})\left(\sum_{|j|>k}||\Pi_j||\right).$$

Proof of (4):

$$||T^{-1}E_1(\hat{\psi})W_2'(\hat{\psi})||^2 \leq \sum_{i=-k}^{k}||T^{-1}\sum_{t=k}^{T-k}\sum_{|j|>k}\Pi_j\Delta_+^{\hat{d}}X_{2t+j}\Delta_+^{\hat{d}}X'_{2t-i}||^2 \leq$$

$$\sum_{i=-k}^{k}\left(\sum_{|j|>k}||\Pi_j||||T^{-1}\sum_{t=k}^{T-k}\Delta_+^{\hat{d}}X_{2t+j}\Delta^{\hat{d}}X'_{2t-i}||\right)^2 = \sum_{i=-k}^{k}\left(\sum_{|j|>k}||\Pi_j||||S^{00}_{ji,T}(\hat{\psi})||\right)^2 = O_p(k)\left(\sum_{|j|>k}||\Pi_j||\right)^2.$$

Proof of (5):

$$||T^{-\hat{b}}E_2(\hat{\psi})W_1'(\hat{\psi})|| \leq ||T^{-\hat{b}}\sum_{t=k}^{T-k}(1-\Delta_+^{\hat{d}-\hat{b}-d_0+b_0})u_{1t}\Delta_+^{\hat{d}-\hat{b}}X'_{2t}|| + ||T^{-\hat{b}}\sum_{t=k}^{T-k}(\Delta_+^{\hat{d}-d_0}-1)\left(\sum_{j\in\mathbb{Z}}\Pi_ju_{2t-j}\right)\Delta_+^{\hat{d}-\hat{b}}X'_{2t}||$$

$$\leq ||S^{01}_{00,T}(\hat{\psi}_1)|| + ||S^{01}_{00,T}(\hat{\psi}_2)|| + \sum_{j\in\mathbb{Z}}||\Pi_j||(||S^{01}_{j0,T}(\hat{\psi}_3)|| + ||S^{01}_{j0,T}(\hat{\psi}_1)||) = o_p(1),$$

for $\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3 \xrightarrow{P} \psi_0$.

Similarly for (6):

$$||T^{-1}E_2(\hat{\psi})W_2(\hat{\psi})'|| \leq \sum_{i=-k}^{k}(||\hat{\psi}_1-\psi_0||||D_\psi S^{00,T}_{0i}(\bar{\psi})|| + ||\hat{\psi}_2-\psi_0||||D_\psi S^{00,T}_{0i}(\bar{\psi})||)+$$

$$+\sum_{i=-k}^{k}\sum_{j\in\mathbb{Z}}||\Pi_j||(||\hat{\psi}_3-\psi_0||||D_\psi S^{00}_{ji,T}(\hat{\psi}_3)|| + ||\hat{\psi}_1-\psi_0||||D_\psi S^{00}_{ji,T}(\hat{\psi}_1)||) = O(kT^{-\kappa}).$$

Proof for (7) follws directly from Lemma C.4. Proof for (8) follows from CLT for product moments of uncorrelated stationary linear time series and tightness of the product moment in ψ.

$\square$

Lemma C.6. *Suppose, $\hat{\psi} - \psi_0 = O_p(T^{-\kappa}), \kappa > 0$ and assumptions 1-4 hold. Then $||R_\infty^{-1}(\psi_0) - R_T^{-1}(\hat{\psi})||^2 = o_p(1)$.*

Proof. Denote $R = R_\infty(\psi_0)$ and $\hat{R} = R_T(\hat{\psi})$. Then inequality $||AB|| \leq ||A||||B||_1$ implies:

$$||R^{-1} - \hat{R}^{-1}|| \leq ||\hat{R}^{-1}||_1||R - \hat{R}||||R^{-1}||_1 \leq \left(||R^{-1} - \hat{R}^{-1}|| + ||R^{-1}||_1\right)||R - \hat{R}||||R^{-1}||_1. \quad (22)$$

Applying Lemma C.5 for $||R - \hat{R}||$ gives:

$$||R - \hat{R}|| \leq ||T^{-1}W_2(\hat{\psi})W_2'(\hat{\psi}) - \Gamma_k|| + 2||T^{-\hat{b}-1/2}W_1(\hat{\psi})W_2'(\hat{\psi})|| = o_p(1).$$

Hence (22) implies:

$$||R^{-1} - \hat{R}^{-1}|| \leq \frac{||R^{-1}||_1^2||R - \hat{R}||}{1 - ||R^{-1}||_1||R - \hat{R}||},$$

where the inequality holds since the denominator is close to 1 with arbitrary high probability due to $||R^{-1}||_1 = O_p(1)$ (Lemma A.3 in Saikkonen (1991)). Hence: $||R^{-1} - \hat{R}^{-1}|| = o_p(1)$ and the bound is proved. $\square$

Lemma C.7. *Suppose $\hat{\psi} - \psi_0 = O_p(T^{-\kappa}), \kappa > 0$ and kernel and bandwidth satisfy assumption 5. Then:*

$$\Big|\Big|\sum_{j=1}^{T-2k} w(j/h_T)\frac{1}{T}\sum_{t=k}^{T-j-k} w_{2t}(\hat{\psi})w_{2t+j}'(\hat{\psi})\Big|\Big| = O_p(k).$$

Proof.

$$\Big|\Big|\sum_{j=1}^{T-2k} w(j/h_T)\frac{1}{T}\sum_{t=k}^{T-j-k} w_{2t}(\hat\psi)w'_{2t+j}(\hat\psi)\Big|\Big| \leq$$

$$\leq \Big|\Big|\sum_{j=1}^{T-2k} w(j/h_T)\frac{1}{T}\sum_{t=k}^{T-j-k} \left(w_{2t}(\psi_0)w'_{2t+j}(\psi_0) - Ew_{2t}(\psi_0)w'_{2t+j}(\psi_0)\right)\Big|\Big|$$

$$+\Big|\Big|\sum_{j=1}^{T-2k} w(j/h_T)\frac{1}{T}\sum_{t=k}^{T-j-k} \left(w_{2t}(\hat\psi)w'_{2t+j}(\hat\psi) - w_{2t}(\psi_0)w'_{2t+j}(\psi_0)\right)\Big|\Big|$$

$$+\Big|\Big|\sum_{j=1}^{T-2k} w(j/h_T)Ew_{2t}(\psi_0)w'_{2t+j}(\psi_0)\Big|\Big| = O_p(kT^{-1/2})\sum_j |w(j/h_T)| + O_p(kT^{-\kappa})\sum_j |w(j/h_T)| +$$

$$\left(\sum_{i_1,i_2=-k}^{k}\Big|\Big|\sum_{j=1}^{T-2k} w(j/h_T)\Gamma_{i_1 i_2+j}(\psi_0)\Big|\Big|^2\right)^{1/2} = O_p(k)$$

where the first bound follows from uniform convergence of autocovariances (Hannan (1974)) and application of Lemma C.1 and the second bound can be obtained with mean-value expansion. □

D Proofs of main theorems

Proof of theorem 3.1. Actual proof of the theorem 3.1 consists of applications of Lemmas C.4,C.5 and C.6. We sketch the idea: errors of the regression (7) $\tilde v_t$ consists of error term v_t uncorrelated with regressors, approximation error term e_{1t} stemming from the autoregressive approximation and e_{2t} stemming from the use of estimates of ψ instead of true values. We show that under the assumptions the error terms are asymptotically negligible.

We estimate $A = (\alpha, \Pi_k, \ldots, \Pi_{-k})$ with feasible DOLS estimator $\hat A(\hat\psi)$ in the regression (7):

$$\hat A(\hat\psi) - A_0 = \tilde V(\hat\psi)W'(\hat\psi)\left(W(\hat\psi)W'(\hat\psi)\right)^{-1}. \tag{23}$$

For the sake of clarity in the following we will supress the dependance of matrices $W, V, \tilde V, E_1, E_2$ on $\hat\psi$. Denote $R = R_\infty(\psi_0)$, $\hat R = R_T(\hat\psi)$ then:

$$(\hat A - A_0)D_T^{-1}(1) = \tilde V W' D_T(\sqrt{T})\hat R^{-1} D_T(\sqrt{T})D_T^{-1}(1) = \tilde V W' D_T(\sqrt{T})\left(\hat R^{-1} - R^{-1}\right)D_T(\sqrt{T})D_T^{-1}(1) +$$
$$+\tilde V W' D_T(T)R^{-1}.$$

where commutativity of $D_T(n_T)$ with block-diagonal matrix R^{-1} is used. Then for the first term it holds:

$$||\tilde{V}W'D_T(\sqrt{T})\left(\hat{R}^{-1}-R^{-1}\right)D_T(\sqrt{T})D_T^{-1}(1)||^2 \le ||\tilde{V}W'D_T(T)||^2||\hat{R}^{-1}-R^{-1}||^2 \le$$
$$\le \left(||VW'D_T(T)||^2+||E_1W'D_T(T)||^2+||E_2W'D_T(T)||^2\right)||\hat{R}^{-1}-R^{-1}||^2.$$

From Lemma C.6 the order of the second term is $o_p(1)$, while from Lemma C.5: $||E_1W'D_T(T)|| = o_p(1)$, $||E_2W'D_T(T)|| = o_p(1)$ and $||VW'D_T(T)|| = O_p(1)$, hence the whole term is of order $o_p(1)$. From $||R^{-1}||_1 = O_p(1)$ and inequality $||AB|| \le ||A||||B||_1$ similarly it follows that $\tilde{V}W'D_T(T)R^{-1} = VW'D_T(T)R^{-1} + o_p(1)$. Hence:

$$(\hat{A}-A_0)D_T(1) = VW'D_T^{-1}(T)R^{-1} + o_p(1). \tag{24}$$

Since R is a block-diagonal matrix, its inverse is too and for the first block of matrix $(\hat{A}-A_0)D_T^{-1}(1)$ it holds:

$$T^{b_0}(\hat{\alpha}-\alpha_0) = T^{b_0-\hat{b}}\left(T^{-b}\sum_{t=k}^{T-k} v_t\Delta_+^{\hat{d}-\hat{b}}X_{2t}'\right)\left(T^{-2\hat{b}}\sum_{t=k}^{T-k}\Delta_+^{\hat{d}-\hat{b}}X_{2t}\Delta_+^{\hat{d}-\hat{b}}X_{2t}'\right)^{-1} + o_p(1). \tag{25}$$

Then Lemma C.4, tightness of $T^{-u+1/2}\Delta_+^{-u}u_{2t}$ in u, application of continuous mapping theorem and $T^{b_0-\hat{b}} \xrightarrow{P} 1$, $k/T = o(1)$ finish the proof of convergence (8). □

Proof of 3.2. If we show that:

$$\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\hat{\tilde{v}}_t\hat{\tilde{v}}_{t+j}' = \sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}v_tv_{t+j}' + o_p(1) \tag{26}$$

and:

$$T^{-1}\sum_{t=k}^{T-k}\hat{\tilde{v}}_t\hat{\tilde{v}}_t' = T^{-1}\sum_{t=k}^{T-k}v_tv_t' + o_p(1) \tag{27}$$

the theorem will follow.

We apply the following uniform bounds in j:

$$\Big|\Big|\sum_{t=k}^{T-j-k}\tilde{v}_t w'_{1t+j}\Big|\Big| = O_p(T^{b_0}),$$

$$\Big|\Big|\sum_{t=k}^{T-j-k}\tilde{v}_t w'_{2t+j}\Big|\Big| = O_p(\sqrt{k}T(T^{-\kappa}+\sum_{|j|>k}||\Pi_j||)) = o_p(Tk^{-1/2}),$$

$$\Big|\Big|\sum_{t=k}^{T-j-k} w_{1t} w'_{2t+j}\Big|\Big| = O_p((k\log T T^{2b_0+1-a})^{1/2}) = o_p(T^{b_0+1/2}),$$

$$\Big|\Big|\sum_{t=k}^{T-j-k} w_{1t} w'_{1t+j}\Big|\Big| = O_p(T^{2b_0}).$$

Also given Lemma C.7 and assumptions, it holds:

$$||\alpha-\hat{\alpha}|| = O_p(T^{b_0}), \tag{28}$$

$$||(W_2W_2)^{-1}||_1 = O_p(T^{-1}), \tag{29}$$

$$h_T^{-1}\textstyle\sum_{j=0}^{T}|\omega(j/h_T)| \to \int_0^\infty |\omega(x)|dx < \infty, \tag{30}$$

$$\Big|\Big|\textstyle\sum_{j=1}^{T-2k}\omega(j/h_T)T^{-1}\sum_{t=k}^{T-j-k} w_{2t}w'_{2t+j}\Big|\Big|_1 = O_p(1). \tag{31}$$

Now, note that:

$$\hat{\tilde{v}}_t = \tilde{v}_t + (A-\hat{A})w_t = \tilde{v}_t + (\alpha-\hat{\alpha})w_{1t} + \Big((\alpha-\hat{\alpha})W_1 + \tilde{V}\Big)W'_2(W_2W'_2)^{-1}w_{2t},$$

where the dependance of w_t, W, V on the estimator $\hat{\psi}$ is implicitly suppresed. From here:

$$\hat{\tilde{v}}_t\hat{\tilde{v}}'_{t+j} = \tilde{v}_t\tilde{v}'_{t+j} + \tilde{v}_t w'_{1t+j}(\alpha-\hat{\alpha})' + \tilde{v}_t w'_{2t+j}(W_2W'_2)^{-1}W_2W'_1(\alpha-\hat{\alpha})' + \tilde{v}_t w'_{2t+j}(W_2W'_2)^{-1}W_2\tilde{V}' +$$
$$(\alpha-\hat{\alpha})w_{1t}w'_{1t+j}(\alpha-\hat{\alpha})' + (\alpha-\hat{\alpha})w_{1t}w'_{2t+j}(W_2W'_2)^{-1}W_2W'_1(\alpha-\hat{\alpha})' + (\alpha-\hat{\alpha})w_{1t}w'_{2t+j}(W_2W'_2)^{-1}W_2\tilde{V}' +$$
$$(\alpha-\hat{\alpha})W_1W'_2(W_2W'_2)^{-1}w_{2t}w'_{2t+j}(W_2W'_2)^{-1}W_2W'_1(\alpha-\hat{\alpha})' +$$
$$(\alpha-\hat{\alpha})W_1W'_2(W_2W'_2)^{-1}w_{2t}w'_{2t+j}(W_2W'_2)^{-1}W_2\tilde{V}' + \tilde{V}W'_2(W_2W'_2)^{-1}w_{2t}w'_{2t+j}(W_2W'_2)^{-1}W_2\tilde{V}'$$

Applying above bounds and norm inequalities, we have:

$$\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\hat{\tilde{v}}_t\hat{\tilde{v}}'_{t+j} = \sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\tilde{v}_t\tilde{v}'_{t+j} + O_p(h_TT^{-1}) + o_p(h_TT^{-1/2}) +$$
$$\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\tilde{v}_t w'_{2t+j}(W_2W'_2)^{-1}W_2\tilde{V}' + O_p(h_TT^{-1}) + o_p(h_TT^{-1}) + O_p(h_TT^{-1/2}) + o_p(h_TT^{-1}) + o_p(1).$$

The last term can be bounded as follows:

$$\Big|\Big|\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\tilde{v}_t w'_{2t+j}(W_2W_2')^{-1}W_2\tilde{V}'\Big|\Big| = \Big|\Big|\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}(v_t+e_{1t}+e_{2t})w'_{2t+j}(W_2W_2')^{-1}W_2\tilde{V}'\Big|\Big| =$$

$$= O_p(h_T T^{-1/2}) + \Big|\Big|\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\left(\sum_{|i|>k}\Pi_i\Delta^d X_{2t+i}\right)w'_{2t+j}(W_2W_2')^{-1}W_2\tilde{V}'\Big|\Big| +$$

$$+ \Big|\Big|\sum_{j=1}^{T-k}\omega(j/h_T)\Big|\Big|\Big|\Big|T^{-1}E_2(\hat{\psi})W_2(\hat{\psi})\Big|\Big|\Big|\Big|(W_2W_2')^{-1}\Big|\Big|_1\Big|\Big|W_2\tilde{V}'\Big|\Big| \leq$$

$$\Big|\Big|\sum_{|i|>k}\Pi_i\Big|\Big|\Big|\Big|\sum_{j=1}^{T-k}\omega(j/h_T)S^{00}_{ij,T}(\hat{\psi})\Big|\Big|\Big|\Big|(W_2W_2')^{-1}\Big|\Big|_1\Big|\Big|W_2\tilde{V}'\Big|\Big| + o_p(h_T k^{-1/2}T^{-\kappa}k^{1/2}) + o_p(1) = o_p(1).$$

Hence:

$$\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\hat{\tilde{v}}_t\hat{\tilde{v}}'_{t+j} = \sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\tilde{v}_t\tilde{v}'_{t+j} + o_p(1). \tag{32}$$

Remember, $\tilde{v}_t = v_t + e_{1t} + e_{2t}$ and cross products involving e_{1t} can be treated similarly as in Kejriwal and Perron (2008), whereas cross-products involving e_{2t} can be bounded using mean-value expansion and bounds of Lemma C.2, we have:

$$\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\tilde{v}_t\tilde{v}'_{t+j} = \sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}v_t v'_{t+j} + o_p(1). \tag{33}$$

Now we prove (27). Observe:

$$\hat{\tilde{V}}\hat{\tilde{V}}' = \tilde{V}\tilde{V}' + \tilde{V}W_1'(\alpha-\hat{\alpha})' + \tilde{V}W_2'(W_2W_2')^{-1}W_2W_1'(\alpha-\hat{\alpha})' + (\alpha-\hat{\alpha})W_1W_1'(\alpha-\hat{\alpha})' +$$

$$2(\alpha-\hat{\alpha})W_1W_2'(W_2W_2')^{-1}W_2W_1'(\alpha-\hat{\alpha})' + 2(\alpha-\hat{\alpha})W_1W_2'(W_2W_2')^{-1}W_2\tilde{V}' + 2\tilde{V}W_2'(W_2W_2')^{-1}W_2\tilde{V}'.$$

Then applying Lemma C.5, we find: $T^{-1}\hat{\tilde{V}}\hat{\tilde{V}}' = T^{-1}\tilde{V}\tilde{V}' + o_p(1) = T^{-1}VV' + o_p(1)$ and the theorem is proved. □

Proof of 3.4. Take a sequence $m_T = T^{-1/2} + T^{-\kappa} + \sum_{|i|>k}||\Pi_i||$. Then similarly as in proof of 3.1, we find:

$$(\hat{A} - A_0)D_T^{-1}(m_T) = \tilde{V}W'D_T(\sqrt{T})\left(\hat{R}^{-1} - R^{-1}\right)D_T(\sqrt{T})D_T^{-1}(m_T) + \tilde{V}W'D_T(Tm_T^{-1})R^{-1}.$$

We find bound for the second term from Lemma C.5:

$$||\tilde{V}W'D_T(Tm_T^{-1})||^2 \leq ||VW'D_T(Tm_T^{-1})||^2 + ||E_1W'D_T(Tm_T^{-1})||^2 + ||E_2W'D_T(Tm_T^{-1})||^2 \leq$$
$$||T^{-\hat{b}}VW_1'||^2 + ||T^{-\hat{b}}E_1W_1'||^2 + ||T^{-\hat{b}}E_2W_1'||^2 + ||m_TT^{-1}VW_2'||^2 + ||m_TT^{-1}E_1W_2'||^2 +$$
$$+ ||m_TT^{-1}E_2W_2'||^2 = m_T^2(T^{-1} + T^{-2\kappa} + (\sum_{|i|>k} ||\Pi_i||)^2)O_p(k) = O_p(k).$$

Since the bound of the first term is of smaller stochastic order than the second, we get:

$$||(\hat{A} - A_0)D_T^{-1}(m_T)||^2 = T^{2\hat{b}}||\hat{\alpha} - \alpha_0||^2 + m_T^2||\hat{\Pi} - \Pi_0||^2 = O_p(k) + o_p(k) \quad (34)$$

and since the first term is $O_p(1)$, it follows that: $m_T||\hat{\Pi} - \Pi_0|| = O_p(\sqrt{k})$. □

References

K. M. Abadir, W. Distaso, and L. Giraitis. Two estimators of the long-run variance: Beyond short memory. *Journal of Econometrics*, 150(1):56–70, 2009.

D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–58, May 1991.

R. T. Baillie. Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1):5–59, July 1996.

D. R. Brillinger. *Time series: data analysis and theory.* Holt, Rinehart, and Winston New York, 1974.

M. Demetrescu, V. Kuzin, and U. Hassler. Long memory testing in the time domain. *Econometric Theory*, 24(01):176–215, February 2008.

E. J. Hannan. The uniform convergence of autocovariances. *Ann. Statist.*, 2(4):803–806, 1974.

J. Hualde and P. M. Robinson. Semiparametric estimation of fractional cointegration. STICERD - Econometrics Paper Series /2006/502, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE, May 2006.

J. Hualde and P.M. Robinson. Root-N-consistent estimation of weak fractional cointegration. *Journal of Econometrics*, 140(2):450 – 484, 2007.

J. Hualde and P.M. Robinson. Semiparametric inference in multivariate fractionally cointegrated systems. *Journal of Econometrics*, 157(2):492 – 511, 2010.

S. Johansen. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, 1995.

S. Johansen and M. Ø. Nielsen. Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics*, 158(1):51–56, September 2010a.

S. Johansen and M. Ø. Nielsen. Likelihood inference for a fractionally cointegrated vector autoregressive model. CREATES Research Papers 2010-24, 2010b.

S. Katsumi and P. C. B. Phillips. Exact local Whittle estimation of fractional integration. *Ann. Statist.*, 33 (4):1890–1933, 2005.

M. Kejriwal and P. Perron. Data dependent rules for selection of the number of leads and lags in the dynamic ols cointegrating regression. *Econometric Theory*, 24(5):59–73, 2008.

C. S. Kim and P. C. B. Phillips. Fully modified estimation of fractional cointegration models. Cowles Foundation Discussion Paper (preprint), 2001.

T. G. Kurtz and P. Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, 19(3):1035–1070, 1991.

D. Marinucci and P. M. Robinson. Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications*, 86(1):103–120, March 2000.

P. C. B. Phillips. Partially identified econometric models. *Econometric Theory*, 5(2):181–240, 1989.

P. C. B. Phillips. Optimal inference in cointegrated systems. *Econometrica*, 59(2):283–306, March 1991.

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

P. C. B. Phillips and V. Solo. Asymptotics for linear processes. *Annals of Statistics*, 20(2):971–1001, 1992.

P. M. Robinson and J. Hualde. Cointegration in fractional systems with unknown integration orders. *Econometrica*, 71(6):1727–1766, November 2003.

P. M. Robinson and D. Marinucci. Semiparametric fractional cointegration analysis. *Journal of Econometrics*, 105(1):225–247, 2001.

S. Said and D. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

P. Saikkonen. Asymptotically efficient estimation of cointegration regressions. *Econometric Theory*, 7(01): 1–21, March 1991.

P. Saikkonen. Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory*, 8(1):1–27, 1992.

K. Shimotsu and P. C. B. Phillips. Local Whittle estimation of fractional integration and some of its variants. *Journal of Econometrics*, 130(2):209–233, February 2006.

J. H. Stock and M. W. Watson. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61(4):783–820, 1993.