Fractional integration and cointegration in financial time series

Stakėnas, P.

Publication date
2012

Chapter 4

Dynamic OLS estimation of fractionally cointegrated regressions

4.1 Introduction

Cointegration analysis has become one of the main tools of empirical research in economics and finance. The traditional cointegration framework assumes the observed time series to be a vector unit root process, while cointegration errors are assumed to be a short-memory 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 an arbitrary order. Generally, we will say that a 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 main questions of interest in a fractionally cointegrated system is the value of the cointegration vector. A number of approaches to inference on cointegration vectors has been suggested in the literature, and they can be attributed to either likelihood-based (see Johansen (1995) for $I(0)/I(1)$ cointegration and Johansen and Nielsen (2010) for the 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:

$$X_{1t} = \alpha X_{2t} + u_t \tag{4.1}$$

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 the $I(1)/I(0)$ case, the OLS estimator of (4.1) is second-order biased and does not bring the inference problem into the locally asymptotically mixed normal (LAMN) family (cf. Phillips (1991)). The problem runs even deeper in the fractional framework: OLS has a 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 the convergence rate and removes the second-order bias in case $\psi = (d, b) = (1, 1)$ (cf. Robinson and Marinucci (2001a)). Although the rate of the NBLS estimator is not yet optimal, the narrow-band weighted least squares (NBWLS) regression achieves the 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 the non-optimal rate of the OLS estimator in (4.1) might be solved by correcting for endogeneity in the regression using time-domain methods. There are two established approaches of this idea in the $I(1)/I(0)$ literature: fully-modified estimation (cf. Phillips and Hansen (1990)) and dynamic OLS (DOLS) estimation (cf. Saikkonen (1991), Stock and Watson (1993)). The former method has been extended to the fractional framework by Kim and Phillips (2001), whereas the goal of this chapter is to extend the latter approach to the fractional framework. Dynamic OLS estimation in $I(1)/I(0)$ systems is based on appending regression equation (4.1) with leads and lags of the differenced regressor X_{2t} which removes the second-order bias, whereas estimation of a fractional regression appended with a fractionally filtered regressor X_{2t} improves the rate of convergence of the estimator (as compared to the OLS estimator of (4.1)) yielding an asymptotically mixed normal estimator of α. The DOLS estimator was proved to be asymptotically efficient in the $I(1)/I(0)$ setting in Saikkonen (1991), provided some conditions on the rate of growth of the number k of appended leads and lags hold. Similar conditions for k are required to hold in a fractional regression. Although in a 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 chapter is organized as follows: Section 4.2 presents the framework and discusses the model and assumptions. Section 4.3 outlines the idea of estimation and presents the main results. The results of Monte Carlo simulations evaluating the finite sample performance of the estimator are presented in Section 4.4, Section 4.5 presents an empirical application, while Section 5 concludes. Proofs are given in the appendices.

We use the following notation in this chapter: $\xrightarrow{P}$ means convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution. The Euclidian norm of a matrix, vector or scalar A is denoted as $||A|| = \sqrt{tr(A'A)}$. We will also use the spectral norm: $||A||_1 = \sqrt{\lambda_{\max}(A'A)}$. K is a generic positive constant bounded from below and bounded away from zero. $AsCov(Z_t)$ denotes asymptotic variance-covariance matrix of an asymptotically stationary random vector series Z_t. 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, $D_u f$ denotes a first-order derivative of a function f w.r.t. u. $W_b(s)$ denotes fractional standard type II Brownian motion which is defined as follows:

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

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

4.2 The model and preliminaries

In this chapter we are concerned with generic fractionally cointegrated (FCI) time series:

Assumption 4.1. *The observed n-dimensional time series X_t satisfies:*

$$\Delta(\psi)(\beta, \gamma)' X_t = u_t, \ \ t = 0, 1, \ldots \tag{4.2}$$

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

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

where $\Gamma(i)$ is the Gamma function.

Our characterization of fractionally integrated processes coincides with the type II definition, which defines fractional integration directly in terms of a fractional filter, i.e.

as a weakly stationary time series filtered with a truncated fractional filter. A different characterization is also possible, but it leads to different asymptotic inference considerations and has a different interpretation for the transition mechanisms of innovation shocks (for a discussion see Shimotsu and Phillips (2006), Section 7). Given this definition, the model (4.2) 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 this chapter we are interested in "strongly" cointegrated fractional systems:

Assumption 4.2. $b > 1/2$.

Although most empirical studies are concerned with long memory type behaviour of cointegration errors $\beta' X_t$, which 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 the cointegration vector: in a fractional regression, filtered observables are $I(b)$ and hence non-stationary processes, whereas the errors are $I(0)$, and stochastic dominance of the signal to the noise allows feasible optimal inference. Finally, for the development of partial sum limit theory we need the following assumption on the errors:

Assumption 4.3. *$u_t = C(L)\varepsilon_t$, where $C(z)$ is a matrix power series such that $det(C(z)) \neq 0$, $\forall |z| = 1$ and the coefficients of $C(z)$ and $C^{-1}(z)$ are $1/2$-summable, i.e. $\sum_j \sqrt{j}||C_j|| < \infty$. ε_t is an i.i.d. $(0, \Sigma)$ time series with $\Sigma > 0$ and $E||\varepsilon_t||^q < \infty$ for some $q > \max\{4, (b-1/2)^{-1}\}$.*

The assumption on the error process u_t conveys a 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, the 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 empirical work, from a theoretical perspective the question seems complicated and it does not seem possible to solve it with our method.[1]

Assumption 4.3 ensures that a multivariate fractional invariance principle holds for u_t (see Marinucci and Robinson (2000), Theorem 1):

$$T^{1/2-b} \sum_{i=0}^{[Ts]} \Delta_+^{1-b} u_t \Rightarrow W_{b-1}(s). \tag{4.3}$$

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

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

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

4.3 Regression-based dynamic OLS estimation of cointegration vectors

The key element for the regression-based estimation is the following normalization[2] of the 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 a triangular representation of a cointegrated system. Similarly we obtain a fractional triangular model:

$$\Delta_+^{d-b}(X_{1t} - \alpha X_{2t}) = u_{1t}, \tag{4.4}$$

$$\Delta_+^{d} X_{2t} = u_{2t}, \tag{4.5}$$

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 of X_{2t} in $I(1)/I(0)$ systems, the OLS estimator has a second-order bias. Within the spectral regression framework the problem can be partly solved with narrow-band least squares or narrow-band weighted least squares estimation, the latter achieving the optimal rate of convergence in non-stationary strongly cointegrated systems. Our approach uses a time-domain framework and takes the idea of dynamic OLS estimation introduced in Saikkonen (1991), where the regression equation is appended with lags and leads of the 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 the autocovariances of u_t. On the other hand, fourth order stationary time series with absolutely summable coefficients and finite fourth moments of the innovation process implies 4-th order summability of their cumulants. That and

[2]Validity of this normalization is shown in the appendix. In practice it may be achieved by testing for no cointegration among X_{2t}.

positive boundedness of the spectral density of the 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. 296):

$$u_{1t} = \sum_{i=-\infty}^{\infty} \Pi_i u_{2t-i} + v_t, \tag{4.6}$$

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

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

Here $(\Omega_{ij})_{ij=1,2}$ are the blocks of the long-run covariance matrix Ω corresponding to $u_t = (u'_{1t}, u'_{2t})'$. If we denote $\mathcal{F} = \sigma(\{u_{2t}, t \in \mathbb{Z}\})$, the 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 the regression equation (4.4) with a finite number of leads and lags of the fractionally filtered regressor X_{2t} could "almost" remove endogeneity effects, yielding an 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 the OLS estimator in the following regression:

$$\Delta_+^{d-b} X_{1t} = \alpha \Delta_+^{d-b} X_{2t} + \sum_{i=-k}^{k} \Pi_i \Delta_+^{d} X_{2t-i} + \tilde{v}_t, \;\; t = k, \ldots, T-k \tag{4.7}$$

where $\tilde{v}_t = v_t + e_t$, $e_t = \sum_{|i|>k} \Pi_i \Delta_+^d X_{2t-i}$. The error term e_t represents the error due to truncation of a doubly infinite sum at the lag 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 an optimal rate of convergence with an asymptotic mixed normal distribution (albeit with mixing covariates being functionals of a fractional Brownian motion). However, obviously in a fractional setting it is rather unrealistic to know the true values of the fractional parameters ψ, but we show that a feasible estimator retains its asymptotic properties under the following conditions on the estimator of ψ and the rate of growth of the lag length k:

Assumption 4.4. *Suppose that* $||\hat{\psi} - \psi|| = O_p(T^{-\kappa})$, $\kappa > 0$ *and* $k \to \infty$ *are such that:*

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

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

Assumption 4.4.1 puts an upper bound on the rate k: it does not grow faster than the rate of the estimator $\hat{\psi}$ and $(\log TT^{-a} + T^{-1/2})$, while Assumption 4.4.2 puts the lower bound on 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 $k(\log T)^{-1} \to \infty$, when $b < 1$ (in case $b > 1$, the second part of the assumption is innocuous).

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

Theorem 4.1. *Under Assumptions 4.1-4.4, for the feasible estimator $\hat{\alpha} = \hat{\alpha}_{DOLS}(\hat{\psi})$ it holds:*

$$T^b\left(\hat{\alpha}_{DOLS} - \alpha\right) \xrightarrow{d} \left(\int_0^1 dW_{1\cdot 2}W'_{2,b-1}\right)\left(\int_0^1 W_{2,b-1}W'_{2,b-1}\right)^{-1}, \tag{4.8}$$

where $W_b(s) = (W'_{1b}(s), W'_{2b}(s))'$ is partitioned according to $u_t = (u_{1t}, u_{2t})$, $W_{1\cdot 2}(s) = W_1(s) - \Omega_{12}\Omega_{22}^{-1}W_2(s)$, where $W(s) = (W'_1(s), W'_2(s))'$ is Brownian motion with covariance matrix Ω.

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

$$\left(\int_0^1 dW_{1\cdot 2}W'_{2,b-1}\right)\left(\int_0^1 W_{2,b-1}W'_{2,b-1}\right)^{-1} \sim \int_{G>0} N(0, \Omega_{11\cdot 2}\otimes G)dP(G), \tag{4.9}$$

where:

$$G = \left(\int_0^1 W_{2,b-1}W'_{2,b-1}\right)^{-1} \text{ and } \Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}. \tag{4.10}$$

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

$$W_\alpha(\psi) = (R\text{vec}(\hat{\alpha}_{DOLS}) - r)'\left(R(\Omega_{11\cdot 2}\otimes (X_2X'_2)^{-1})R'\right)^{-1}(R\text{vec}(\hat{\alpha}_{DOLS}) - r), \tag{4.11}$$

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: $W_\alpha(\psi_0) \xrightarrow{d} \chi^2_s$. The feasible test statistic requires estimation of ψ and the long run covariance matrix of v_t. One way to estimate $\Omega_{11\cdot 2}$ is with a HAC-type estimator (cf. Andrews (1991)), as weighted sum of the sample autocovariances of the regression (4.7) residuals $\hat{\hat{v}}_t$:

$$\hat{\Omega}_{11\cdot 2} = T^{-1} \sum_{t_1=k}^{T-k} \sum_{t_2=k}^{T-k} \omega \left(\frac{|t_1 - t_2|}{h_T} \right) \hat{\hat{v}}_{t_1} \hat{\hat{v}}'_{t_2}. \tag{4.12}$$

The following conditions on the kernel function and bandwidth guarantee consistency of $\hat{\Omega}_{11\cdot 2}$:

Assumption 4.5. *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: $||\hat{\psi} - \psi|| h_T = o_p(1)$.

Theorem 4.2. *Suppose that Assumptions 4.1-4.5 hold. Then $\hat{\Omega}_{11\cdot 2} \xrightarrow{P} \Omega_{11\cdot 2}$.*

Given the latter theorem we could use estimate (4.12) and $\hat{\psi}$ to construct a feasible Wald test statistic (4.11) for which: $W_\alpha(\hat{\psi}) \xrightarrow{d} \chi^2_s$ under the null H_0. Of course, (4.12) is not the only way of 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 a HAC-type or Robinson's MAC estimator (for a comparison see Abadir et al. (2009)).

Remark 4.1. Note, that Theorem 4.1 does not discuss consistency of the regression (4.7) coefficients $\Pi = (\Pi_{-k}, \ldots, \Pi_k)$. The following theorem gives the rate of consistency of $\hat{\Pi}$:

Theorem 4.3. *Suppose that Assumptions 4.1-4.4 hold. Then:*

$$||\hat{\Pi} - \Pi|| \left(T^{-1/2} + T^{-\kappa} + \sum_{|i|>k} ||\Pi_i|| \right)^{-1} = O_p(\sqrt{k}).$$

The theorem effectively states that when 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 the true ψ_0 for estimation is used: the rate $\sqrt{T/k}$ of consistency is achieved, given $\sqrt{T} \sum_{|i|>k} ||\Pi_i|| = o(1)$ holds (cf. Saikkonen (1992)). In addition, the upper bound assumption $k^3/T = o(1)$ used in Saikkonen (1991) and Saïd and Dickey (1984) for consistency of $\hat{\Pi}$ seems to be not needed.

Remark 4.2. In an empirical analysis it might be reasonable to allow for different integration orders of individual components of the cointegration errors, however, this is not feasible in our framework. The assumption on homogeneity of memory of the error term components is crucial, since the commutation $\Delta(L,\delta)\alpha X_{2t} = \alpha\Delta(L,\delta)X_{2t}$ with $\Delta(\delta) = diag(\Delta_+^{\delta_1},\ldots,\Delta_+^{\delta_r})$, allowing regression based estimation of α in (4.4), is permitted only with homogenous integration orders: $\delta_1 = \ldots = \delta_r$.

Remark 4.3. Notice that if $b < 1$, Assumption 4.3 precludes the use of Akaike, Schwarz and other lag selection criteria to select the number of lags k in the regression (4.7) for ARMA type processes, since they select the lag length proportional to $\log T$. The latter observation is important, since although Assumption 4.4 specifies admissible 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 the information criteria (for a comparison of lag selection rules see Kejriwal and Perron (2008)).

Remark 4.4. In the specific case when the true value of ψ is known, Assumption 4.4 boils down to:

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

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

Remark 4.5. The model (4.4) does not consider any deterministic terms and is restrictive in this respect. Let us consider the following system:

$$\Delta_+^{d-b}(X_{1t} - \alpha X_{2t} + \mu) = u_{1t}, \tag{4.13}$$
$$\Delta_+^{d}(X_{2t} + \delta) = u_{2t}, \tag{4.14}$$

where $(\mu', \delta')' \in \mathbb{R}^n$. In this case, the DOLS regression is the following:

$$\Delta_+^{d-b}X_{1t} = \alpha\Delta_+^{d-b}X_{2t} - \Delta_+^{d-b}\mu + \sum_{i=-k}^{k}\Pi_i\Delta_+^{d}X_{2t-i} + \Delta_+^{d}\delta\left(\sum_{i=-\infty}^{\infty}\Pi_i\right) + \tilde{v}_t,\ \ t = k,\ldots,T-k. \tag{4.15}$$

One could show that the term $\Delta_+^{d}\delta(\sum_{i=-\infty}^{\infty}\Pi_i)$ is asymptotically negligible in the regression. If $d - b > 1/2$, the term $\Delta_+^{d-b}\mu$ is also asymptotically negligible and for the estimator $\hat{\alpha}_{DOLS}$ convergence (4.8) still holds. However, if $d-b < 1/2$, then the following convergence holds:

$$\left(\hat{\alpha}_{DOLS}-\alpha,\hat{\mu}_{DOLS}-\mu\right)\begin{pmatrix}T^{b}I_{n-r} & 0_{n-r\times 1}\\ 0_{1\times n-r} & T^{1/2-d+b}\end{pmatrix}\xrightarrow{d}\left(\int_0^1 dW_{1\cdot 2}Q'\right)\left(\int_0^1 QQ'\right)^{-1}, \quad (4.16)$$

where $Q(s)=(W_{2,b-1}'(s),F_{d-b}(s))'$, $F_{d-b}(s)=s^{-d+b}/\Gamma(1-d+b)$. In this case for $\hat{\alpha}_{DOLS}$ it holds:

$$T^{b}\left(\hat{\alpha}_{DOLS}-\alpha\right)\xrightarrow{d}\left(\int_0^1 d\bar{W}_{1\cdot 2}\bar{W}_{2,b-1}'\right)\left(\int_0^1 \bar{W}_{2,b-1}\bar{W}_{2,b-1}'\right)^{-1}, \quad (4.17)$$

where:

$$\bar{W}_{1\cdot 2}=W_{1\cdot 2}-\int_0^1 W_{1\cdot 2}F_{d-b}\left(\int_0^1 F_{d-b}F_{d-b}\right)^{-1}F_{d-b}, \quad (4.18)$$

$$\bar{W}_{2,b-1}=W_{2,b-1}-\int_0^1 W_{2,b-1}F_{d-b}\left(\int_0^1 F_{d-b}F_{d-b}\right)^{-1}F_{d-b}. \quad (4.19)$$

In any case, one can test restrictions on α with test statistic (4.11), with $X_2=X_2(\psi)=(\Delta_+^{d-b}\bar{X}_{2k},\ldots,\Delta_+^{d-b}\bar{X}_{2T-k})$, where $\Delta_+^{d-b}\bar{X}_{2t}$ is $\Delta_+^{d-b}X_{2t}$ regressed on $\Delta_+^{d-b}1=F_{d-b}(t)(1+o(1))$. We may also study inference with deterministic terms of arbitrary order in the model (4.4) and the results would correspond to those derived Robinson and Iacone (2005).

Remark 4.6. Now suppose, that instead of X_t we observe a contaminated time series: $\tilde{X}_t=X_t+\xi_t$, such that Assumptions 4.1-4.3 hold for X_t. The contamination term ξ_t can be interpreted as the term generated by the initial values of the series or a measurement error term. Then under the following condition[3] the asymptotics of $\hat{\alpha}_{DOLS}$ is not affected if we use $\tilde{X}_t$ instead of X_t in the regression (4.7):

$$T^{-b}\sum_{t=k}^{T-k}t^{b-1/2}(E||\Delta_+^{d-b}\xi_t||^2)^{1/2}+T^{-b}\sum_{t=k}^{T-k}E||\Delta_+^{d-b}\xi_t||^2+kT^{-1}\sum_{t=k}^{T-k}(E||\Delta_+^{d}\xi_t||^2)^{1/2}=o(1). \quad (4.20)$$

Notice that (4.4) implies $X_0=u_0$, 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 (4.20) shows that more generally we may assume that $X_0=O_p(1)$.

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

4.4 Finite sample performance

In this section we introduce the design of the simulated data generating process and present simulation results. We analyze the finite sample performance of the proposed estimator and the corresponding Wald test with a fractionally cointegrated bivariate time series. The error term is designed to have both contemporaneous and serial correlation effects. We present RMSE of DOLS and competing estimators as well as the size of the test based on the feasible test statistic.

4.4.1 Monte Carlo setup

We simulate a bivariate fractionally cointegrated model satisfying Assumptions 4.1-4.3, without loss of generality assuming that the normalized cointegration vector is $\alpha_0 = 1$:

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

for $t = 1, \ldots, T$. $u_t = (u_{t1}, u_{t2})'$ 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 a Gaussian[4] i.i.d. $(0, \Omega)$ process with $\Omega = ((1,\rho)', (\rho, 1)')$. We simulate this system for the values $T = 256, 512$, $d = 0.8, 1, 1.2$, $d - b = 0, 0.2, 0.4$. We have considered three simulation designs: $\rho = \psi = \phi = 0$, $\rho = \psi = \phi = 0.5$, $\rho = \psi = \phi = 0.8$ (simulation designs Nr. 1, 2, 3, respectively). A number of 1000 Monte Carlo replications was used.

The feasible estimator is constructed in the following way: d is estimated as an average of the integration orders of X_{1t} and X_{2t} with the exact local Whittle estimator (ELWE), given in (2.18), maximizing over the interval $[d-2, d+2]$ with bandwidth $m = [T^{0.6}]$, while b is estimated as $\hat{d}$ minus the memory of the 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 the NBLS estimator using bandwidth $m = [T^{0.65}]$. The deterministic lag selection rule $p4 = [4(T/100)^{1/4}]$, taken from Demetrescu et al. (2008), was used in the feasible DOLS (FDOLS) estimation[5].

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

[5] Given Remark 4.3, AIC selection rule for FDOLS estimation was compared to the deterministic rule $p4$, but little difference was found.

The feasible DOLS estimator is compared to NBLS (using the above bandwidth) and feasible NBZLS estimators (cf. Hualde and Robinson (2010))[6]. The selected bandwidth is $m = [T^{0.8}]$, where d, b are estimated as above, whereas $f_u(0)$ is estimated as $\hat{f}_u(0) = \sum_{i=-m}^{m} I_{u^f}(\lambda_j)$, where I_{u^f} is the periodogram of the time series $u_t^f = (\Delta_+^{\hat{d}-\hat{b}}\hat{\beta}X_t', \Delta_+^{\hat{d}}X_{2t}')'$ and $m = [0.5 \cdot T^{0.8}]$.

The simulation experiment requires a selection of four different bandwidths, what obviously influences the outcome of the whole experiment. Since, to the best knowledge of the author, there are no adaptive bandwidth selection procedures of the estimators we use, we have chosen bandwidth selection rules typically used in the literature.

We also simulated the empirical size of the feasible Wald test statistic W_α^1 using the DOLS estimate of $\hat{\beta}$ and $\hat{\Omega}$ estimated as[7] $2\pi\hat{f}_{u^f}(0)$ for the null $\alpha = 1$ with the above design. We have also simulated the size of the test statistic W_α^2 which is constructed identically, but using the estimate of b, obtained by imposing the null $\alpha = 1$ in the estimation procedure, i.e. b estimated with ELWE as an integration order of $\beta_0 X_t$.

4.4.2 Simulation results

We compare RMSE of FDOLS with NBLS and FNBZLS. The 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 marginally better than NBLS in all designs, except 1 when NBLS is efficient. The cases with $b = 0.6$ and sample size $T = 256$ are the most problematic for FDOLS - stochastic dominance of the signal to the errors is close to the critical value of 0.5 and improper selection of number of lags in the regression distorts the signal enough to affect the estimator considerably. However, it is reassuring to see the drop of the RMSE of FDOLS considerably with $T = 512$, it being comparable to that of FNBZLS. Summing up, we may conclude that in terms of RMSE, FDOLS performs marginally better than NBLS and on par with NBZLS, although it does seem to underperform with (jointly) small sample size, b close to 0.5.

As far as the empirical test sizes are concerned, as expected, empirical sizes of the statistic W_α^2 are closer to the nominal size than that of W_α^1. We also see that the empirical sizes decrease with increase in parameters and our unreported simulations show that out

[6] We chose to simulate a simpler feasible NBZLS ("zero frequency") estimator which does not require estimation of the 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 performance of the NBZLS and NBWLS estimators.

[7] Unreported simulations show better properties of this estimator in compare to (4.12) and feasible HAC-type estimator of Ω.

Table 4.1: Monte Carlo simulation results

				Empirical sizes for W_α^1			Empirical sizes for W_α^2			RMSE		
$Nr.$	d	$d-b$	T	10%	5%	1%	10%	5%	1%	FDOLS	NBLS	FNBZLS
1	0.8	0	256	0.12	0.07	0.03	0.11	0.06	0.02	0.025	0.026	**0.022**
			512	0.14	0.08	0.03	0.13	0.07	0.02	0.013	0.014	**0.012**
		0.2	256	0.21	0.14	0.07	0.17	0.11	0.05	0.073	**0.044**	0.045
			512	0.23	0.17	0.08	0.20	0.14	0.06	0.037	0.029	**0.028**
	1	0	256	0.12	0.08	0.03	0.11	0.07	0.02	0.010	0.013	**0.010**
			512	0.15	0.09	0.03	0.13	0.07	0.03	0.005	0.006	**0.005**
		0.2	256	0.22	0.16	0.09	0.18	0.12	0.05	0.024	0.025	**0.022**
			512	0.25	0.17	0.09	0.21	0.14	0.06	0.013	0.014	**0.012**
		0.4	256	0.18	0.12	0.07	0.13	0.08	0.03	0.077	0.048	**0.048**
			512	0.19	0.13	0.07	0.14	0.09	0.04	0.038	0.032	**0.030**
	1.2	0	256	0.12	0.08	0.03	0.11	0.06	0.02	0.004	0.006	**0.004**
			512	0.14	0.09	0.03	0.13	0.07	0.03	0.002	0.003	**0.002**
		0.2	256	0.22	0.16	0.09	0.18	0.12	0.05	0.010	0.012	**0.010**
			512	0.26	0.18	0.10	0.21	0.14	0.06	0.005	0.006	**0.005**
		0.4	256	0.20	0.15	0.08	0.14	0.09	0.04	0.025	0.024	**0.022**
			512	0.19	0.14	0.07	0.14	0.09	0.04	0.013	0.014	**0.012**
2	0.8	0	256	0.13	0.09	0.04	0.12	0.07	0.02	0.026	0.033	**0.021**
			512	0.16	0.10	0.04	0.15	0.08	0.03	0.013	0.020	**0.012**
		0.2	256	0.17	0.10	0.04	0.13	0.07	0.02	0.081	0.071	**0.043**
			512	0.20	0.13	0.06	0.16	0.09	0.03	0.040	0.050	**0.027**
	1	0	256	0.14	0.09	0.04	0.12	0.06	0.02	**0.010**	0.012	0.010
			512	0.15	0.10	0.04	0.14	0.08	0.03	**0.005**	0.006	0.005
		0.2	256	0.16	0.11	0.05	0.12	0.07	0.02	0.026	0.027	**0.022**
			512	0.18	0.12	0.06	0.15	0.09	0.03	0.013	0.015	**0.012**
		0.4	256	0.17	0.11	0.05	0.12	0.07	0.02	0.076	0.061	**0.043**
			512	0.20	0.13	0.06	0.15	0.09	0.03	0.039	0.040	**0.028**
	1.2	0	256	0.14	0.09	0.04	0.11	0.06	0.02	**0.004**	0.006	0.005
			512	0.15	0.10	0.04	0.13	0.08	0.03	**0.002**	0.003	0.002
		0.2	256	0.16	0.10	0.05	0.12	0.07	0.02	**0.010**	0.011	0.011
			512	0.17	0.12	0.05	0.14	0.08	0.03	**0.005**	0.006	0.005
		0.4	256	0.16	0.11	0.05	0.12	0.07	0.02	0.026	0.024	**0.022**
			512	0.18	0.11	0.06	0.14	0.08	0.03	0.013	0.014	**0.013**
3	0.8	0	256	0.12	0.07	0.03	0.10	0.06	0.01	0.028	0.033	**0.024**
			512	0.16	0.09	0.03	0.14	0.07	0.01	**0.014**	0.021	0.014
		0.2	256	0.18	0.10	0.04	0.13	0.07	0.02	0.076	0.083	**0.042**
			512	0.21	0.14	0.04	0.18	0.10	0.02	0.038	0.061	**0.026**
	1	0	256	0.09	0.06	0.02	0.08	0.03	0.01	0.011	**0.010**	0.014
			512	0.10	0.06	0.02	0.09	0.05	0.01	**0.005**	0.005	0.006
		0.2	256	0.13	0.07	0.03	0.10	0.06	0.01	0.027	0.026	**0.025**
			512	0.16	0.09	0.03	0.14	0.07	0.01	**0.013**	0.016	0.014
		0.4	256	0.17	0.10	0.04	0.13	0.07	0.02	0.074	0.069	**0.040**
			512	0.21	0.13	0.05	0.18	0.10	0.02	0.038	0.048	**0.026**
	1.2	0	256	0.06	0.04	0.01	0.06	0.03	0.00	**0.004**	0.005	0.006
			512	0.06	0.04	0.02	0.06	0.03	0.01	**0.002**	0.003	0.002
		0.2	256	0.09	0.05	0.02	0.08	0.04	0.01	0.011	**0.009**	0.014
			512	0.10	0.06	0.02	0.09	0.05	0.01	0.005	**0.005**	0.006
		0.4	256	0.13	0.07	0.02	0.10	0.06	0.01	0.027	**0.023**	0.025
			512	0.16	0.09	0.03	0.14	0.07	0.01	**0.014**	0.014	0.015

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

of the parameters ϕ, ψ, ρ, the parameter ϕ influences the decrease of empirical size the most. Hence generally, the empirical sizes tend to decrease with increase of persistence in the short-run noise u_t. In addition, the empirical sizes approach their nominal values with an increase of b, but more worryingly they depend little on the sample size and, in most cases, even tend to move away from the nominal levels increasing the sample size T. Our (unreported) simulations of empirical size for the infeasible DOLS estimator using ψ_0 and Ω_0 show very good properties, hence there are two possible reasons for the poor asymptotic properties of the test statistics: it is either the uncertainty coming from the estimates of ψ or that of Ω. Unreported simulations show that the estimate of Ω has bad asymptotic properties, confirming findings in Abadir et al. (2009), hence we hypothesize that poor asymptotic properties of $\hat{\Omega}$ are responsible for the poor asymptotic properties of test statistics and thus an improvement in the empirical size is expected using better estimation methods of Ω.

4.5 Empirical application

Section 2.5 analyzed the U.S. Treasury yields dataset described in Section 1.3 and did not reject the null hypothesis that the yields are processes integrated of order 1, while in Section 3.4 we rejected the null that the bivariate time series $(M1, M3)$, $(M1, M6)$ are not cointegrated. In this section we estimate the cointegration vector in bivariate series by DOLS regression and test structural hypothesis on the cointegration vector implied by the expectations hypothesis (EH).

EH implies that the bivariate systems cointegrate with cointegration vector $\beta = (1, -1)'$ or, in terms of normalized triangular system (4.4), $\alpha = 1$. That constitutes our null hypothesis. Before analysis we visually inspect graphs (see Figure 4.1) of term spreads, i.e. the time series $\beta' X_t$. We see that even though the term spreads do not display trending behaviour, their mean is clearly not zero. Hence, in the following we analyze a system with intercept terms (4.13). We also assume that the integration order of interest rates is $d = 1$, based on findings in Section 3.4.

The DOLS estimate $\hat{\alpha}_{DOLS}$ is constructed in the following way: b is estimated as d minus the estimated memory of $\hat{\beta}' X_t$ using the local Whittle estimator (see Robinson (1995)) with bandwidth m_1, where $\hat{\beta}$ is the NBLS estimate with bandwidth $m_2 = [T^{0.65}]$. We select the bandwidth m_1 by applying the iterative procedure in Henry (2001), where an optimal bandwidth is selected for the $ARFIMA(1, d, 0)$ model. We iterate the procedure once starting with an initial value $m_1^0 = [T^{0.6}]$. In this way, we have selected the bandwidths $m_1 = 59, 41$ for the spreads of $(M1, M3)$, $(M1, M6)$, respectively. The number of lags

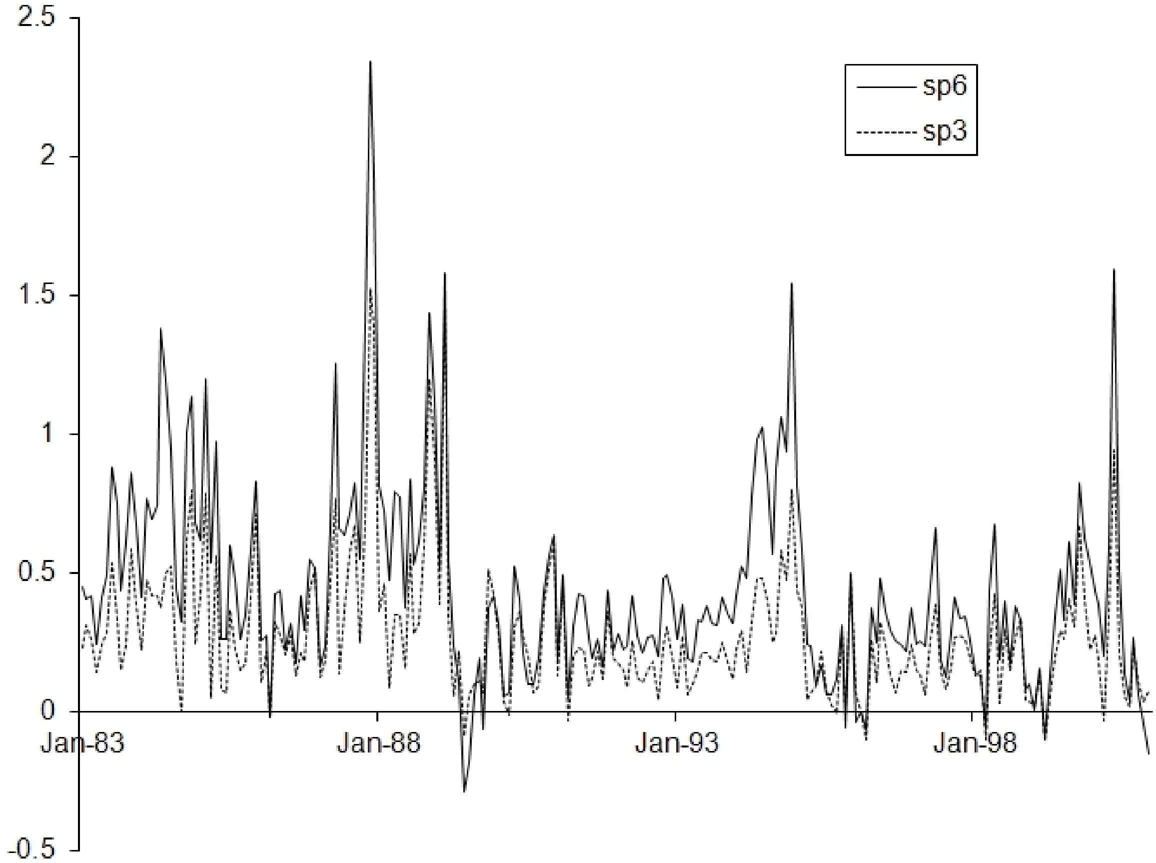

Figure 4.1: Graph of spread terms

selected in the DOLS regression is $p4 = [4(T/100)^{0.25}] = 4$. A Wald test statistic W_α is constructed using a periodogram estimate for the long-run covariance matrix of the error term: $\hat{\Omega} = 2\pi \hat{f}_{u^f}(0)$, where I_{u^f} is the periodogram of the time series $u_t^f = (\Delta_+^{d-\hat{b}} \hat{\beta} X_t', \Delta_+^d X_{2t}')'$ and a number of ordinates used is $m_3 = [0.5 \cdot T^{0.8}]$. We have selected parameters k, m_2, m_3 arbitrarily, since these parameters, unlike m_1, do not greatly affect the results of the DOLS estimator, as our finite sample simulations have shown.

Results of the tests are reported in Table 4.4. Firstly, we note that $\hat{b} > 0.5$ for both bivariate systems, indicating relations of strong cointegration and thus validity of the DOLS estimator. We also see that in both cases the null hypothesis $\alpha = 1$ is not rejected

Table 4.2: Test results for the null $\alpha = 1$

System	$d - \hat{b}$	$\hat{\alpha}_{DOLS}$	W_α	p-value
(M3,M1)	0.209	1.027	1.556	0.212
(M6,M1)	0.407	1.014	0.073	0.787

with p-values of the test 0.212 and 0.787 for $(M1, M3)$, $(M1, M6)$, respectively. This shows that the implication of expectations hypothesis holds for bivariate systems with terms spread displaying long memory behaviour, yet being stationary. A natural research question is to model a trivariate system jointly, which will be done in Chapter 5.

4.6 Conclusions

This chapter established a feasible dynamic OLS type procedure for optimal estimation of the cointegration vectors in "strongly" cointegrated fractional regressions with a mixed normal asymptotic distribution of the estimator. Finite sample simulations show superior properties over the NBLS estimator and rivals the NBZLS 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 the feasible Wald test being generally oversized, but our (unreported) simulations show that the infeasible statistic has very good size properties even for small sample sizes, suggesting that uncertainty coming from the estimation of Ω accounts for the size problems. An empirical application to U.S. interest rates testing implications of EH is also provided, which does not reject the cointegration structure implied by EH.

4.7 Appendix

4.7.1 Preliminaries and notation

Suppose we have a fractionally cointegrated system X_t satisfying Assumptions 4.1-4.3 with $n \times r$ cointegration matrix β and parameters $\psi = (d, b)$, $R = (\beta, \gamma)$ for some process u_t. It is not obvious that the cointegration matrix β could be identified as $\beta' = (I_r, -\alpha)$ and we show that this is indeed true, i.e. there exists an $r \times (n - r)$ matrix α and a process $\tilde{u}_t$ satisfying Assumption 4.3 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 representations of X_t we have the equality:

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

Take[8] $\alpha = -R_1^{-1}R_2$. 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 the 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 Assumption 4.3 also holds for $\tilde{u}_t$.

Next 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 an estimate of ψ_0 and the error due to truncation:

$$\tilde{v}_t = \Delta_+^{d-b}X_{1t} - \alpha\Delta_+^{d-b}X_{2t} - \sum_{i=-k}^{k}\Pi_i\Delta_+^{d}X_{2t-i} = \Delta_+^{d-b}\left(X_{1t} - \alpha X_{2t}\right) - \Delta_+^{d-d_0}\left(\sum_{i=-k}^{k}\Pi_i\Delta_+^{d_0}X_{2t-i}\right)$$
$$= \Delta_+^{d-b-d_0+b_0}u_{1t} - \Delta_+^{d-d_0}(u_{1t} - v_t - \sum_{|j|>k}\Pi_j\Delta_+^{d_0}X_{2t-j}) = v_t + e_{1t} + e_{2t},$$

where:

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

[8] It is always possible to arrange the coordinates of X_t in such a way that R_1 has a full rank.

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

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

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. $V = (v_k, \ldots, v_{T-k})$.

4.7.2 Additional lemmas

The following lemma gives two criteria for a probabilistic bound on a sum of stochastically bounded random variables depending on a parameter $\theta \in \Theta$ with growing number of terms[9].

Lemma 4.1. *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 neighborhood N_{θ_0} of θ_0. 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 following conditions hold:*

$$\sup_{\theta \in N_{\theta_0}} \sup_{i \in \mathbb{N}} E(X_T^i(\theta))^2 = O(1), \tag{4.21}$$

$$\sup_{\theta \in N_{\theta_0}} \sup_{i \in \mathbb{N}} |X_T^i(\theta)| = O_p(1). \tag{4.22}$$

---

[9] In Kejriwal and Perron (2008) this argument is missing (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 $\{X_T^i(u) = T\mathbb{1}_{u \in [i-1/k_T, i/k_T]}\}$, $i = 1, \ldots, k_T$ 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 supremum is vacuous. Also a similar proposition holds for small "o".

Proof. Take any $\varepsilon_1 > 0$ and $M > \sup_{\theta \in N_{\theta_0}} \sup_{i \in \mathbb{N}} \sqrt{E|X_T^i(\theta_T)|^2}$. Then from Chebyshev's inequality:

$$\begin{aligned}
&P\left(\frac{|\sum_{i=1}^{k_T} a_i X_T^i(\theta_T)|}{\sum_{j=1}^{k_T}|a_j|} \geq M\right) \leq P(\theta_T \notin N_{\theta_0}) \\
&+ P\left(\theta_T \in N_{\theta_0}, |\sum_{i=1}^{k_T} a_i (X_T^i(\theta_T) - EX_T^i(\theta_T))| \geq M\sum_{j=1}^{k_T}|a_j| - \sum_{i=1}^{k_T}|a_i| E|X_T^i(\theta_T)|\right) \\
&\leq P(\theta_T \notin N_{\theta_0}) + \left(M\sum_{j=1}^{k_T}|a_j| - \sum_{i=1}^{k_T}|a_i| E|X_T^i(\theta_T)|\right)^{-2} Var\left(\sum_{i=1}^{k_T} a_i X_T^i(\theta_T)\right) \\
&\leq P(\theta_T \notin N_{\theta_0}) + \left(M\sum_{j=1}^{k_T}|a_j| - \sum_{i=1}^{k_T}|a_i| \sqrt{E|X_T^i(\theta_T)|^2}\right)^{-2} \left(\sum_{i=1}^{k}|a_i|\sqrt{VarX_T^i(\theta_T)}\right)^2 \\
&\leq P(\theta_T \notin N_{\theta_0}) + \left(M - \sup_{\theta \in N_{\theta_0}} \sup_{i \in \mathbb{N}} \sqrt{E|X_T^i(\theta_T)|^2}\right)^{-2} \left(\sup_{\theta \in N_{\theta_0}} \sup_{i \in \mathbb{N}} \sqrt{VarX_T^i(\theta)}\right)^2 < \varepsilon_1,
\end{aligned}$$

for every T large enough. On the other hand:

$$\begin{aligned}
&P\left(\frac{|\sum_{j=1}^{k_T} a_j X_T^j(\theta_T)|}{\sum_{j=1}^{k_T}|a_j|} \geq M\right) \leq P(\theta_T \notin N_{\theta_0}) + P\left(\theta_T \in N_{\theta_0}, |\sum_{j=1}^{k_T} a_j X_T^j(\theta_T)| \geq M\sum_{j=1}^{k_T}|a_j|\right) \\
&\leq P(\theta_T \notin N_{\theta_0}) + P\left(\theta \in N_{\theta_0}, \sup_{\theta \in N_{\theta_0}} \sup_{i \in \mathbb{N}} |X_T^i(\theta)| \geq M\right) < \varepsilon_1,
\end{aligned}$$

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 neighborhood of the true value in which criteria (4.21) hold for relevant norm moments, which is at the core of the derivation of probabilistic bounds of various terms. The lemma derives from the results in Johansen and Nielsen (2010):

Lemma 4.2. *Denote a η_0-neighborhood 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 4.3 and consider the 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:

$$S_{ij,T}^{rs}(\psi) = T^{-1} \sum_{t=k}^{T-k} V_{t+i}^r(\psi) V_{t-j}^{s\prime}(\psi),$$

it holds,[10] *as $k = k_T \to \infty$:*

$$\sup_{\psi \in N(\eta_0)} \sup_{i,j \in \mathbb{Z}} E \left|\left| S_{ij,T}^{00}(\psi) \right|\right|^2 = O(1), \tag{4.23}$$

$$\sup_{\psi \in N(\eta_0)} \sup_{i,j \in \mathbb{Z}} E \left|\left| D_\psi S_{ij,T}^{00}(\psi) \right|\right|^2 = O(1), \tag{4.24}$$

$$\sup_{\psi \in N(\eta_0)} \sup_{i \in \mathbb{Z}} E \left|\left| S_{0i,T}^{10}(\psi) \right|\right|^2 = O(\log^2 T T^{-a+2\eta_0}), \text{ for } a = \min\{1, 2b_0 - 1\}, \tag{4.25}$$

$$\sup_{i \in \mathbb{Z}} E \left|\left| S_{0i,T}^{10}(\psi_0) \right|\right|^2 = O(\log^2 T T^{-a}), \text{ for } a = \min\{1, 2b_0 - 1\}. \tag{4.26}$$

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

$$||S_{ij,T}^{00}(\psi)||_2 \leq ||S_{00,T}^{00}(\psi)||_2 \leq T^{-1} \sum_{i=k}^{T-k} ||V_t^0(\psi) V_t^{\prime 0}(\psi)||_2 \leq T^{-1} \sum_{i=k}^{T-k} ||V_t^0(\psi)||_4 ||V_t^0(\psi)||_4 \leq$$
$$\leq KT^{-1} \sum_{i=k}^{T-k} ||V_t^0(\psi)||_2 ||V_t^0(\psi)||_2 \leq K, \forall i, j \in \mathbb{N}, \forall \psi \in N(\eta_0).$$

The last inequalities follow from multivariate extensions of Lemma B.1 and Lemma C.4 in Johansen and Nielsen (2010) applied to the process $V_t^0(\psi)$. Similarly, we prove (4.24). (4.26) follows from multivariate extension of Lemma C.5 in Johansen and Nielsen (2010). The same lemma could be used to prove (4.25) upon noting that the coefficients of the linear process $V_t^1(\psi)$ 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

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

assume that $\hat{\psi}$ is in the η_1-neighborhood $N(\eta_1)$ of ψ_0 "small enough" for Lemma 4.2 to hold for various fractional processes whose order of fractionality depends on $\hat{\psi}$:

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

Also, although not mentioned explicitly, all lemmas in this section are proved under Assumptions 4.1-4.4.

Lemma 4.3. $\sup_{i\in\mathbb{Z}} ||S^{10}_{0i,T}(\hat{\psi})|| = O_p(\log^2 T T^{-a})$*, for* $a = \min\{1, 2b_0 - 1\}$.

Proof. The 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 4.2. □

Lemma 4.4. *Denote:*

$$S(\psi) = T^{-b} \sum_{t=1}^{T} v_t \Delta_+^{d-b} X'_{2t}. \tag{4.27}$$

Then it holds:

$$\left(T^{-1/2}\Delta_+^{-1} v_{[Ts]}, T^{-\hat{b}+1/2}\Delta_+^{\hat{d}-\hat{b}} X_{2[Ts]}, S(\hat{\psi})\right) \Rightarrow \left(W_{1\cdot 2}(s), W_{2,b_0-1}(s), \int_0^1 dW_{1\cdot 2} W'_{2,b_0-1}\right). \tag{4.28}$$

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 (2010)) and the fractional invariance principle (4.3). We prove convergence of the third component. Consider the Beveridge-Nelson decomposition for v_t:

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

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

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

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 $A_2 = o_p(1)$. Also, the Cauchy-Schwarz inequality 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 the 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, the 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 the stochastic integral follows from an 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 4.5. *It holds:*

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 repeatedly apply Lemma 4.1 for big "O" in combination with Lemmas 4.2, 4.3: we bound the growing sum of either expectations of squared norms of product moments applying condition (4.21) or sum of norms applying (4.22). We will apply the mean value theorem for the product moments and will use a 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):

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

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 $\sum_{i,j=-k}^{k} ||A_{ij,T}||^2 = O_p(k^2 T^{-1})$. Next applying the mean value theorem we get: $||B_{ij,T}|| \leq ||\hat{\psi} - \psi_0|| \, ||D_\psi S^{00}_{ij,T}(\bar{\psi})||$ and from Lemma 4.2 we find that $\sum_{i,j=-k}^{k} ||B_{ij,T}||^2 = O_p(k^2 T^{-2\kappa})$. 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):

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

Similarly, proof of (3):

$$||T^{-\hat{b}}E_1(\hat{\psi})W_1'(\hat{\psi})|| \leq ||T^{-\hat{b}} \sum_{t=k}^{T-k} \sum_{|j|>k} \Pi_j \Delta_+^{\hat{d}} X_{2t+j} \Delta_+^{\hat{d}-\hat{b}} X'_{2t}||$$
$$\leq \sqrt{T} \sum_{|j|>k} ||\Pi_j|| \, ||T^{-\hat{b}-1/2} \sum_{t=k}^{T-k} \Delta_+^{\hat{d}} X_{2t+j} \Delta_+^{\hat{d}-\hat{b}} X'_{2t}|| = \sqrt{T} \sum_{|j|>k} ||\Pi_j|| \, ||S^{01}_{j0,T}(\hat{\psi})||$$
$$= O_p(\log T T^{1/2-a/2}) \left(\sum_{|j|>k} ||\Pi_j|| \right).$$

Proof of (4):

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

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

$$= O_p(k)\left(\sum_{|j|>k}||\Pi_j||\right)^2.$$

Proof of (5):

$$||T^{-\hat{b}}E_2(\hat{\psi})W_1'(\hat{\psi})|| \leq ||T^{-\hat{b}}\sum_{t=k}^{T-k}(1-\Delta_+^{\hat{d}-\hat{b}-d_0+b_0})u_{1t}\Delta_+^{\hat{d}-\hat{b}}X_{2t}'||$$

$$+ ||T^{-\hat{b}}\sum_{t=k}^{T-k}(\Delta_+^{\hat{d}-d_0}-1)\left(\sum_{j\in\mathbb{Z}}\Pi_j u_{2t-j}\right)\Delta_+^{\hat{d}-\hat{b}}X_{2t}'|| \leq ||S_{00,T}^{01}(\hat{\psi}_1)|| + ||S_{00,T}^{01}(\hat{\psi}_2)||$$

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

for $\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3 \xrightarrow{P} \psi_0$. Similarly for (6):

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

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

Proof of (7) follows directly from the Lemma 4.4. Proof of (8) follows from the CLT for product moments of uncorrelated stationary linear time series and tightness of the product moment in ψ.

□

Lemma 4.6. $||\bar{R}_T^{-1}(\psi_0) - R_T^{-1}(\hat{\psi})||^2 = o_p(1)$.

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

$$||R^{-1}-\hat{R}^{-1}|| \leq ||\hat{R}^{-1}||_1||R-\hat{R}||||R^{-1}||_1 \leq \left(||R^{-1}-\hat{R}^{-1}|| + ||R^{-1}||_1\right)||R-\hat{R}||||R^{-1}||_1. \tag{4.29}$$

Applying Lemma 4.5 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 (4.29) 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. □

Lemma 4.7. *Suppose that the kernel and bandwidth satisfy Assumption 4.5. 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.

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

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

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

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

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

where the first bound follows from uniform convergence of autocovariances (Hannan (1974)) and application of Lemma 4.1, and the second bound can be obtained with the mean value expansion. □

4.7.3 Proofs of main theorems

Proof of Theorem 4.1. The actual proof of Theorem 4.1 consists of applications of Lemmas 4.4, 4.5 and 4.6. We sketch the idea: the error of the regression (4.7) $\tilde{v}_t$ consist of an error term v_t uncorrelated with regressors, an 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 the feasible DOLS estimator $\hat{A}(\hat{\psi})$ in the regression (4.7):

$$\hat{A}(\hat{\psi}) - A_0 = \tilde{V}(\hat{\psi})W'(\hat{\psi})\left(W(\hat{\psi})W'(\hat{\psi})\right)^{-1}. \tag{4.30}$$

For the sake of clarity in the following we will suppress the dependence of matrices $W, V, \tilde{V}, E_1, E_2$ on $\hat{\psi}$. Denote $R = \bar{R}_T(\psi_0)$, $\hat{R} = R_T(\hat{\psi})$ then:

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

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

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

From Lemma 4.6, the order of the second term is $o_p(1)$, while from Lemma 4.5: $||E_1W'D_T(T)|| = o_p(1)$, $||E_2W'D_T(T)|| = o_p(1)$ and $||VW'D_T(T)|| = O_p(1)$, hence the whole term is of order $o_p(1)$. Similarly, from $||R^{-1}||_1 = O_p(1)$ and inequality $||AB|| \leq ||A||||B||_1$ it follows that $\tilde{V}W'D_T(T)R^{-1} = VW'D_T(T)R^{-1} + o_p(1)$. Hence:

$$(\hat{A} - A_0)D_T(1) = VW'D_T^{-1}(T)R^{-1} + o_p(1). \tag{4.31}$$

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:

$$T^{b_0}(\hat{\alpha}-\alpha_0)=T^{b_0-\hat{b}}\left(T^{-b}\sum_{t=k}^{T-k}v_t\Delta_+^{\hat{d}-\hat{b}}X_{2t}'\right)\left(T^{-2\hat{b}}\sum_{t=k}^{T-k}\Delta_+^{\hat{d}-\hat{b}}X_{2t}\Delta_+^{\hat{d}-\hat{b}}X_{2t}'\right)^{-1}+o_p(1). \tag{4.32}$$

Then Lemma 4.4, tightness of $T^{-u+1/2}\Delta_+^{-u}u_{2t}$ in u, application of the continuous mapping theorem and $T^{b_0-\hat{b}}\xrightarrow{P}1$, $k/T=o(1)$ finish the proof of convergence (4.8). □

Proof of Theorem 4.2. If we show that:

$$\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\hat{\tilde{v}}_t\hat{\tilde{v}}_{t+j}'=\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}v_tv_{t+j}'+o_p(1) \tag{4.33}$$

and:

$$T^{-1}\sum_{t=k}^{T-k}\hat{\tilde{v}}_t\hat{\tilde{v}}_t'=T^{-1}\sum_{t=k}^{T-k}v_tv_t'+o_p(1) \tag{4.34}$$

the theorem will follow.

We apply the following uniform bounds in j:

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

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

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

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

Also given Lemma 4.7 and assumptions, it holds:

$$||\alpha - \hat{\alpha}|| = O_p(T^{b_0}), \tag{4.35}$$

$$||(W_2 W_2)^{-1}||_1 = O_p(T^{-1}), \tag{4.36}$$

$$h_T^{-1} \sum_{j=0}^{T} |\omega(j/h_T)| \to \int_0^\infty |\omega(x)| dx < \infty, \tag{4.37}$$

$$\Big|\Big| \sum_{j=1}^{T-2k} \omega(j/h_T) T^{-1} \sum_{t=k}^{T-j-k} w_{2t} w'_{2t+j} \Big|\Big|_1 = O_p(1). \tag{4.38}$$

Note that:

$$\hat{\tilde{v}}_t = \tilde{v}_t + (A - \hat{A}) w_t = \tilde{v}_t + (\alpha - \hat{\alpha}) w_{1t} + \Big((\alpha - \hat{\alpha}) W_1 + \tilde{V} \Big) W'_2 (W_2 W'_2)^{-1} w_{2t},$$

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

$$\begin{aligned}
\hat{\tilde{v}}_t \hat{\tilde{v}}'_{t+j} &= \tilde{v}_t \tilde{v}'_{t+j} + \tilde{v}_t w'_{1t+j} (\alpha - \hat{\alpha})' + \tilde{v}_t w'_{2t+j} (W_2 W'_2)^{-1} W_2 W'_1 (\alpha - \hat{\alpha})' + \tilde{v}_t w'_{2t+j} (W_2 W'_2)^{-1} W_2 \tilde{V}' \\
&+ (\alpha - \hat{\alpha}) w_{1t} w'_{1t+j} (\alpha - \hat{\alpha})' + (\alpha - \hat{\alpha}) w_{1t} w'_{2t+j} (W_2 W'_2)^{-1} W_2 W'_1 (\alpha - \hat{\alpha})' \\
&+ (\alpha - \hat{\alpha}) w_{1t} w'_{2t+j} (W_2 W'_2)^{-1} W_2 \tilde{V}' \\
&+ (\alpha - \hat{\alpha}) W_1 W'_2 (W_2 W'_2)^{-1} w_{2t} w'_{2t+j} (W_2 W'_2)^{-1} W_2 W'_1 (\alpha - \hat{\alpha})' \\
&+ (\alpha - \hat{\alpha}) W_1 W'_2 (W_2 W'_2)^{-1} w_{2t} w'_{2t+j} (W_2 W'_2)^{-1} W_2 \tilde{V}' + \tilde{V} W'_2 (W_2 W'_2)^{-1} w_{2t} w'_{2t+j} (W_2 W'_2)^{-1} W_2 \tilde{V}'.
\end{aligned}$$

Applying above bounds and norm inequalities, we have:

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

The last term can be bounded as follows:

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

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

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

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

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

Hence:

$$\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\hat{\tilde{v}}_t\hat{\tilde{v}}'_{t+j} = \sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\tilde{v}_t\tilde{v}'_{t+j} + o_p(1). \tag{4.39}$$

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 a mean value expansion and bounds of Lemma 4.2, we have:

$$\sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}\tilde{v}_t\tilde{v}'_{t+j} = \sum_{j=1}^{T-k}\omega(j/h_T)T^{-1}\sum_{t=1}^{T-j}v_t v'_{t+j} + o_p(1). \tag{4.40}$$

Now we prove (4.34). Observe:

$$\begin{aligned}\hat{\tilde{V}}\hat{\tilde{V}}' &= \tilde{V}\tilde{V}' + \tilde{V}W_1'(\alpha-\hat{\alpha})' + \tilde{V}W_2'(W_2W_2')^{-1}W_2W_1'(\alpha-\hat{\alpha})' + (\alpha-\hat{\alpha})W_1W_1'(\alpha-\hat{\alpha})' \\ &+ 2(\alpha-\hat{\alpha})W_1W_2'(W_2W_2')^{-1}W_2W_1'(\alpha-\hat{\alpha})' + 2(\alpha-\hat{\alpha})W_1W_2'(W_2W_2')^{-1}W_2\tilde{V}' \\ &+ 2\tilde{V}W_2'(W_2W_2')^{-1}W_2\tilde{V}'.\end{aligned}$$

Applying Lemma 4.5, 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 Theorem 4.3. Take a sequence $m_T = T^{-1/2} + T^{-\kappa} + \sum_{|i|>k}||\Pi_i||$. Then similarly as in proof of Theorem 4.1, 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 a bound for the second term from Lemma 4.5:

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

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

$$||(\hat{A}-A_0)D_T^{-1}(m_T)||^2=T^{2\hat{b}}||\hat{\alpha}-\alpha_0||^2+m_T^2||\hat{\Pi}-\Pi_0||^2=O_p(k)+o_p(k),\tag{4.41}$$

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