Maximum likelihood estimation of fractionally cointegrated systems

Katarzyna Łasak*
Department d’Economia i d’Historia Economica
Universitat Autonoma de Barcelona
Edifici B - Campus de la Autonoma
08193 Bellaterra, Spain
E-mail: klasak@idea.uab.es

June 24, 2007

Abstract

In this paper we consider a fractionally cointegrated error correction model and investigate asymptotic properties of the maximum likelihood (ML) estimators of the matrix of the cointegration relations, the degree of fractional cointegration, the matrix of the speed of adjustment to the equilibrium parameters and the variance-covariance matrix of the error term. We show that using ML principles to estimate jointly all parameters of the fractionally cointegrated system we obtain consistent estimates and provide their asymptotic distributions. The cointegration matrix is asymptotically mixed normal distributed, while the degree of fractional cointegration and the speed of adjustment to the equilibrium matrix have joint normal distribution, which proves the intuition that the memory of the cointegrating residuals affects the speed of convergence to the long-run equilibrium, but does not have any influence on the long-run relationship. The rate of convergence of the estimators of the long-run relationships depends on the cointegration degree but is optimal for the strong cointegration case considered. We also prove that misspecification of the degree of fractional cointegration does not affect the consistency of the estimators of the cointegration relationships, although usual inference rules are not valid. We illustrate our results in finite samples by Monte Carlo analysis.

Keywords: Error correction model, Gaussian VAR model, Maximum likelihood estimation, Fractional cointegration

JEL: C13, C32.

*This research has been carried out partially under financial support from the Departament d’Economia i d’Historia Economica, UAB, Spain.
1 Introduction

Cointegration is thought of a stationary relation between nonstationary variables. It has become a standard tool in econometrics since the seminal paper of Granger (1981). One of the most commonly used procedures in econometric practice, fully parametric inference on $I(1)/I(0)$ cointegrated systems in the framework of Vector Error Correction Mechanism (VECM) representation, has been developed by Johansen (1988, 1991, 1995). He suggests a maximum likelihood (ML) procedure based on reduced rank regressions. His methodology consists in identifying the number of cointegration vectors within the Vector Autoregressive (VAR) model by means of performing a sequence of likelihood ratio (LR) tests. If the variables are cointegrated, after selecting the rank, cointegration vectors and the speed of adjustment to the equilibrium coefficients are estimated.

However the assumption that deviations from equilibrium are integrated of order zero is far too restrictive. In a general set up it is possible to permit errors with fractional degree of integration. This is an important generalisation, since fractional cointegration has the same economic implications as when the processes are integer-valued cointegrated, in the sense that there exist a long-run equilibrium among the variables. The only difference is that the rate of convergence to the equilibrium is slower in the fractional than in the standard case. Moreover since an $I(1)/I(0)$ cointegration setup ignores the fractional cointegration parameter, a fractionally integrated equilibrium error results in a misspecified likelihood function.

It has been studied what happens if we use standard VECM models to make an inference on fractionally cointegrated systems. Gonzalo and Lee (1998) have found that likelihood ratio tests based on the standard models find spurious cointegration between independent variables that are not unit root processes. Further Andersson and Gredenhoff (1998) have shown by simulation that LR test based on the standard model has power against fractional alternatives, so using ML techniques we are likely to find the evidence of $C(1,1)$ cointegration when in reality we have fractional cointegration. At the same time the ML approach based on standard models gives the estimates of the "impact" matrix $\Pi = \alpha \beta$ that are severely biased and have large mean square errors if the variables are fractionally cointegrated. So it can be much more severe to ignore fractional cointegration than to incorporate it, when it is not present, in a fractional framework like the one we consider, that nests the standard case.

In Lasak (2005) we have developed an asymptotic theory for LR tests based on the fractional VECM. The procedure that leads to construction of LR tests simultaneously produces ML estimates of all the parameters of the fractional VECM, the fractional cointegration degree, the cointegrating vectors, speed of adjustment to the equilibrium coefficients matrix, short run correlation parameters and the variance-covariance matrix of the error term. Knowledge of the properties of those estimators would allow us to propose more complex and complete analysis of fractionally cointegrated systems in line of the analysis in Johansen (1988, 1991, 1995). Therefore in this paper we examine the properties of ML estimators of the fractional VECM.

The list of other work treating inference problems of cointegrated systems in a fractional context, without pretension of completeness, includes the following papers. Robinson (1994) have estab-
lished the consistency for frequency domain narrow-band estimates of the fractional cointegrating relationship in the stationary bivariate case. In a nonstationary framework the properties of this estimator have been studied by Marinucci and Robinson (2001) and Robinson and Marinucci (2001, 2003). Robinson and Hualde (2003) have considered estimation of the cointegrating relationship using GLS estimator, which is asymptotically mixed normal and leads to a Wald test statistic with a standard $\chi^2$ distribution under the null. Their model assumes "strong cointegration", similarly to the model we consider in this paper. The asymptotic properties of the cointegrating vector in "weak cointegration" case have been established in Hualde and Robinson (2006b). In the latter case the cointegrating vector is not superconsistent, in spite of the result in Robinson and Hualde (2003) that $\sqrt{T}$-consistent and asymptotic normal estimate can be obtained.

Other works allow for deterministic components whose presence implies a competition between stochastic and deterministic trends as discussed in Marinucci and Robinson (2000). Robinson and Iacone (2005) have developed an asymptotic theory for the cointegrating vector in systems generated by polynomial trends and processes that may be fractionally integrated. Chen and Hurvich (2003) have derived an asymptotic distribution of a tapered narrow-band least squares estimator of the cointegrating parameter. Hassler, Marmol and Velasco (2006) have examined bivariate regressions of nonstationary variables dominated by linear time trends.

Cointegration among stationary long memory processes is especially of interest in financial applications. Christensen and Nielsen (2006) have found that the asymptotic distribution of narrow band least squares (NBLS) is normal if regressors and errors obey the condition that their collective memory is less than 0.5 and their coherency is zero at the origin. Nielsen and Frederiksen (2007) have shown that if the zero coherence assumption is not satisfied then a bias term appears in the mean of the asymptotic distribution. They have also proposed a fully modified NBLS estimator in the spirit of Phillips and Hansen (1990) that does not have this drawback. Nielsen (2007) have shown consistency of joint local Whittle quasi ML estimators of integration orders, regressors, errors and the cointegration vector.

In this paper we consider a fractionally cointegrated system and investigate the asymptotic properties of the ML estimators of the cointegration relations, the degree of fractional cointegration, speed of adjustment to the equilibrium parameters and the variance-covariance matrix of the error term. We show that using ML principles to estimate jointly all the parameters of the fractionally cointegrated model we obtain consistent estimates with known asymptotic distribution. The cointegration matrix estimate results to be asymptotically mixed normal distributed, while the degree of fractional cointegration and the speed of adjustment to the equilibrium matrix have joint normal distribution. This proves the intuition that the memory of the cointegrating residuals affects the speed of convergence to the long-run equilibrium, but does not have any influence on the long-run relationship. However the rate of convergence of the estimators of the long-run relationships depends on the cointegration degree. We also prove that misspecification of the degree of fractional cointegration does not affect the consistency of the cointegration relationships estimators, although usual inference rules are not valid.

The rest of the paper is organised as follows. Section 2 describes the fractional cointegration framework. Section 3 presents the model considered in the paper and the procedure that gives
us estimates of the fractionally cointegrated systems. Section 4 describes main results on the consistency and the asymptotic distribution of all the estimators of the system obtained jointly. Section 5 discusses a model with short run correlation. Section 6 presents Monte Carlo simulation. Section 7 concludes. Appendix A contains all the lemmas. In Appendix B and C proofs of main results of this paper are given under different assumptions.

2 Framework description

We use the following definition of fractionally integrated process $I(\delta)$ like in Marinucci and Robinson (2001).

**Definition 1** We say that a scalar process $a_t$, $t \in Z$, is an $I(\delta)$ process, $\delta > 0$, if there exists a zero mean scalar process $\eta_t$, $t \in Z$, with positive and bounded spectral density at zero, such that

$$a_t = \Delta^{-\delta} \eta_t 1_{t>0}, \quad t \in Z, \quad \delta > 0,$$  

(1)

where $1_{(\cdot)}$ is the indicator function, $\Delta = 1 - L$, $L$ is the lag operator and the fractional difference filter is defined formally by:

$$(1 - z)^\delta = \frac{1}{\Gamma(-\delta)} \sum_{j=0}^{\infty} \frac{\Gamma(j - \delta)}{\Gamma(j + 1)} z^j,$$

(2)

$\Gamma(\cdot)$ is gamma function: $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$.

The process $a_t$ is said to be asymptotically stationary when $\delta < \frac{1}{2}$, since it is nonstationary only due to the truncation on the right-hand side of (1). The truncation is designed to cater for cases $\delta \geq \frac{1}{2}$, because otherwise the right-hand side of (1) does not converge in mean square and hence $a_t$ is not well defined.

We follow with the definition of cointegration by Granger (1986):

**Definition 2** A set of $I(\delta)$ variables is said to be cointegrated, or $CI(\delta,d)$, if there exists a linear combination that is $I(\delta - d)$ for $d > 0$.

In the standard cointegration setup $\delta = d = 1$ and we can use ML techniques as in Johansen (1998, 1991, 1995). However if $\delta \neq 1$ or $d < 1$ we have fractional cointegration, which calls for a generalization of the standard cointegration framework, since inference based on the standard VECM may not be valid.

Johansen (2005a) has shown how fractional VECM representations could be derived. Assume that $X_t$ is a $p \times 1$ vector fractionally integrated of order $\delta$ and there are $r$ linear combinations $\beta$, that are of order $\delta - d$, and

$$\xi^\delta X_t = u_{1t},$$

$$\beta^\delta \Delta^{\delta-d} X_t = u_{2t},$$

(3)
where \( u_t = (u_{1t}', u_{2t}')' \) is i.i.d. \((0, \Sigma)\), \( \xi \) is \( p \times (p - r) \) so that \( (\xi, \beta) \) has rank \( p \) and \( X_t = 0 \) for \( t \leq 0 \).

Then using the identity
\[
\xi_\perp (\beta' \xi_\perp)^{-1} \beta' + \beta_\perp (\xi' \beta_\perp)^{-1} \xi' = I_p
\]
we can show that
\[
\Delta^\delta X_t = \beta_\perp (\xi' \beta_\perp)^{-1} u_{1t} + \xi_\perp (\beta' \xi_\perp)^{-1} \Delta^d u_{2t} = \beta_\perp (\xi' \beta_\perp)^{-1} u_{1t} + \xi_\perp (\beta' \xi_\perp)^{-1} \Delta^d u_{2t} - \xi_\perp (\beta' \xi_\perp)^{-1} (1 - \Delta^d) u_{2t} = (1 - \Delta^d) \alpha \beta' \Delta^{\delta-d} X_t + \varepsilon_t
\]
where \( \varepsilon_t = \beta_\perp (\xi' \beta_\perp)^{-1} u_{1t} + \xi_\perp (\beta' \xi_\perp)^{-1} u_{2t} \) is i.i.d. Recall that \( \alpha \) is a \( p \times r \) matrix of speed of adjustment to the equilibrium coefficients, \( \alpha = -\xi_\perp (\beta' \xi_\perp)^{-1} \) and it satisfies \( \beta' \alpha = -I_r, r \) is a cointegration rank. The formulation (3) allows for modelling and estimating both the cointegrating vectors \( \beta \) and "common trends" vectors \( \xi \) and has also been used by Breitung and Hassler (2002).

To make a model more flexible it is a natural idea to add a lag structure to the model (4). Granger (1986) have included lags of \( \Delta^\delta X_t \) and have proposed a model that can be presented as
\[
A^*(L) \Delta^\delta X_t = (1 - \Delta^d) \Delta^{\delta-d} \alpha \beta' X_{t-1} + d(L) \varepsilon_t,
\]
where \( A^*(L) \) and \( d(L) \) are usual lag polynomials.

Johansen (2005 a,b) have proposed another model that comes from adding the fractional lag operator \( L_d = 1 - (1 - L)^d \) to model (4) and has the following form
\[
A(L_d) \Delta^\delta X_t = (1 - \Delta^d) \Delta^{\delta-d} \alpha \beta' X_t + \varepsilon_t.
\]

An alternative model that allows for short run correlation in both the fractional cointegration relationship and in the levels has been proposed in Avarucci (2007)
\[
\Delta^\delta X_t = \alpha \beta' (\Delta^{-d} - 1) A(L) \Delta^\delta X_t + (I - A(L)) \Delta^\delta X_t + \varepsilon_t,
\]
with a usual lag polynomial \( A(L) \), of order \( k \), that can be also expressed as
\[
\Delta^\delta X_t = \alpha \beta' (\Delta^{-d} - 1) \Delta^\delta X_t + \sum_{j=1}^{k} L^j B_j \{(\Delta^{-d} - 1) \Delta^\delta X_t\} + \sum_{j=1}^{k} L^j A_j \Delta^\delta X_t + \varepsilon_t,
\]
with the restriction \( B_j = -A_j I \). We use this model in Section 4.

### 3 Model and ML estimation

In this paper as a first natural research step we consider the simplest version of the fractional VECM model without lagged differences, which is obviously a special case of models (5), (6) and (7). We use the VECM representation
\[
\Delta X_t = \alpha \beta' (\Delta^{1-d} - \Delta) X_t + \varepsilon_t
\]
\(^1\)Recall for a \( p \times m \) matrix \( a \) we define \( a_{\perp} \) to be a \( p \times (p-m) \) matrix of rank \( p-m \), for which \( a' a_{\perp} = 0 \).
together with the representation given by (3). Note that we assume the Gaussianity of the errors, but only to define the likelihood function.

To estimate the parameters of model (9) we follow the procedure described in Johansen (1995), but adjusted for the case of fractional VECM that has been already presented in Lasak (2005). Let’s define $Z_{0t} = \Delta X_t$, $Z_{1t}(d) = (\Delta^{1-d} - \Delta) X_t$. The model expressed in these variables becomes

$$Z_{0t} = \alpha \beta^0 Z_{11}(d) + \varepsilon_t, \quad t = 1, \ldots, T.$$  

The log-likelihood function apart from a constant for the model (9) is given by

$$L(\alpha, \beta, \Omega, d) = -\frac{1}{2} T \log |\Omega| - \frac{1}{2} \sum_{t=1}^{T} [Z_{0t} - \alpha \beta^0 Z_{11}(d)]' \Omega^{-1} [Z_{0t} - \alpha \beta^0 Z_{11}(d)].$$  

Define as well

$$S_{ij}(d_a, d_b) = T^{-1} \sum_{t=1}^{T} Z_{it}(d_a) Z_{jt}(d_b)' \quad i, j = 0, 1,$$

where $S_{11}(d) = S_{11}(d, d)$ and note that $S_{ij}$ do not depend on $d$ when $i = j = 0$. For fixed $d$ and $\beta$, parameters $\alpha$ and $\Omega$ are estimated by regressing $Z_{0t}$ on $\beta^0 Z_{11}(d)$ and

$$\hat{\alpha}(\beta) = S_{01}(d)\beta(\beta^0 S_{11}(d)\beta)^{-1}$$

while

$$\hat{\Omega}(\beta) = S_{00} - S_{01}(d)\beta(\beta^0 S_{11}(d)\beta)^{-1}\beta^0 S_{10}(d) = S_{00} - \hat{\alpha}(\beta)(\beta^0 S_{11}(d)\beta)\hat{\alpha}(\beta)'$$

(10)

Plugging the estimates into the likelihood we get:

$$L_{\text{max}}^{-2/T}(\hat{\alpha}(\beta), \beta, \hat{\Omega}(\beta), d) = L_{\text{max}}^{-2/T}(\beta, d) = |S_{00} - S_{01}(d)\beta(\beta^0 S_{11}(d)\beta)^{-1}\beta^0 S_{10}(d)|,$$

and finally the maximum of the likelihood is obtained by solving the following eigenvalue problem:

$$|\lambda(d)S_{11}(d) - S_{10}(d)S_{00}^{-1}S_{01}(d)| = 0$$

(12)

for eigenvalues $\lambda_i(d)$ and eigenvectors $v_i(d)$, for a given $d$, such that :

$$\lambda_i(d)S_{11}(d)v_i(d) = S_{10}(d)S_{00}^{-1}S_{01}(d)v_i(d),$$

and $v_j(d)S_{11}(d)v_i(d) = 1$ if $i = j$ and 0 otherwise. Note that the eigenvectors diagonalize the matrix $S_{10}(d)S_{00}^{-1}S_{01}(d)$ since

$$v_j(d)S_{10}(d)S_{00}^{-1}S_{01}(d)v_i(d) = \lambda_i(d)$$

if $i = j$ and 0 otherwise. Thus by simultaneously diagonalizing the matrices $S_{11}(d)$ and $S_{10}(d)S_{00}^{-1}S_{01}(d)$ we can estimate the $r$—dimensional cointegrating space as the space spanned by the eigenvectors corresponding to the $r$ largest eigenvalues. With this choice of $\beta$ we can estimate $d$ by maximizing the log-likelihood, i.e.

$$d = \arg \max_{d \in \mathcal{D}} L_{\text{max}}(d),$$

(13)

where

$$L_{\text{max}}(d) = |S_{00}| \prod_{i=1}^{r} \left(1 - \hat{\lambda}_i(d)\right)^{-\frac{1}{2}}$$

and $\mathcal{D} \subset (0.5, 1]$. Note that we assume that the cointegration rank is known already, or alternatively we can establish it using for example the sequence of the tests proposed in Lasak and Velasco (2007).
4 Consistency and Asymptotic Distribution

First let us define for \(d \in (0.5, 1]\) and omitting the dependence on the true value of \(d, d_0\),

\[
\lim_{t \to \infty} \text{Var} \begin{bmatrix}
Z_{0t} \\
\beta' Z_{1t}^{(1)} (d) \\
\beta' Z_{1t}^{(1)} (d)
\end{bmatrix} =
\begin{bmatrix}
\Sigma_{00} & \Sigma_{0\beta} (d) & \Sigma_{0\beta} (d) \\
\Sigma_{\beta0} (d) & \Sigma_{\beta\beta} (d) & \Sigma_{\beta\beta} (d) \\
\Sigma_{\beta0} (d) & \Sigma_{\beta\beta} (d) & \Sigma_{\beta\beta} (d)
\end{bmatrix}
\tag{14}
\]

where

\[
Z_{1t}^{(1)} (d) := \frac{\partial}{\partial d} Z_{1t} (d).
\]

The exact form of the elements of this covariance matrix is given in Lemma 7 in the Appendix A.

Further we assume as in Robinson and Hualde (2003):

**Assumption 1** For some estimate \(\hat{d}\)

\[
d - d_0 = O_p (T^{-\kappa}), \quad \kappa > 0,
\]

\[
\left| \hat{d} - d_0 \right| \leq K \omega p_1,
\]

and \(\varepsilon_t\) are independent and identically distributed vectors with mean zero, positive definite covariance matrix \(\Omega\), and \(E\|\varepsilon_t\|^q < \infty, q \geq 4, q > 2/ (2d_0 - 1), d_0 > \frac{1}{2}\).

Note that we assume that we have a \(T^\kappa\) consistent pre-estimate of \(d\), which can be obtained for example by semiparametric memory estimates based on single equation OLS residuals. Its particular choice does not affect the generality of our results. The parametric space of \(\hat{\beta}, \hat{\alpha}\) and \(\hat{\Omega}\) we consider is unrestricted, but \(\hat{d} \in D_T\), where \(D_T = \left[ \hat{d} - cT^{-\kappa}, \hat{d} + cT^{-\kappa} \right]\), so \(\hat{d}\) is also assumed to be \(T^\kappa\) consistent.

We also use the following normalization of \(\hat{\alpha}\) and \(\hat{\beta}\) as in Johansen’s (1995). We choose the coordinate system \((\beta, \gamma)\) and expand

\[
\tilde{\beta} = \beta \tilde{\beta} + \gamma \gamma' \tilde{\beta},
\]

where \(\tilde{\beta} = \beta (\beta' \beta)^{-1}\) etc. and define the estimator

\[
\tilde{\beta} = \tilde{\beta} \left( \beta' \beta \right)^{-1} = \beta + \gamma \gamma' \tilde{\beta} = \beta + \gamma U_T
\]

where \(U_T = \gamma' \tilde{\beta}\). This way of normalizing is convenient for the analysis, since it has the property that \(\tilde{\beta} - \beta\) is contained in the space spanned by \(\gamma\) and hence orthogonal to \(\beta\). Note that since \(\tilde{\beta}\) is just a linear transformation of the columns of \(\beta\) it also maximizes the likelihood function and hence \(\tilde{\beta}\) satisfies the likelihood equations. The normalization depends on \(\beta\), so for practice it is not so useful, but it is convenient in the analysis. We define \(\tilde{\alpha} = \tilde{\alpha} \hat{\beta}'\hat{\beta}\) so that \(\tilde{\alpha} \tilde{\beta}' = \hat{\alpha} \hat{\beta}'\).

Under Assumption 1 and for \(d_0 \in D \subset (0.5, 1]\), where \(D\) is a closed set, we demonstrate following Johansen (1995) that
Theorem 1 The estimators $\hat{\beta} = \hat{\beta} \left( \hat{\beta}' \hat{\beta} \right)^{-1}$, $\hat{\alpha} = \hat{\alpha} \hat{\beta}$, $\hat{\Omega}$ are consistent. Moreover $\hat{\beta} - \beta = o_P(T^{\frac{1}{2} - d_0})$.

Note that from Theorem 1 we get the consistency of all the parameters of the fractional VECM we have proposed to estimate jointly by ML.

Theorem 2 For any fixed $d$, $d \neq d_0$, $d > 0.5$ so that $q > 2/(2d - 1)$ in Assumption 1 the estimator $\hat{\beta} = \hat{\beta} \left( \hat{\beta}' \hat{\beta} \right)^{-1}$ remains consistent with a rate $\hat{\beta} - \beta = o_P(T^{\frac{1}{2} - d})$, but $\hat{\alpha}$ and $\hat{\Omega}$ are not consistent anymore.

Theorem 2 tells us that if instead of estimating $d$ we plug in any fixed $d$, $d > 0.5$ to estimate other parameters of fractional VECM we will still obtain a consistent estimate of $\beta$, but not of $\alpha$ and $\Omega$, which shows that the bias and large mean square errors of the estimator of the impact matrix $\Pi = \alpha \beta'$ found by Andersson and Gredenhoff (1998) came from the estimation of $\alpha$ rather than $\beta$.

Theorem 3 Under Assumption 1 and for $d_0 \in \text{Int}\mathbb{D} \subset (0.5, 1]$ the asymptotic distribution of $\hat{\beta}$ is mixed Gaussian and given by

$$T^{d_0} U_T = \gamma' \left( \hat{\beta} - \beta \right) \rightarrow_d \left[ \gamma' C \int_0^1 W_{d_0} (\tau) W_{d_0} (\tau)' \, d\tau C' \gamma \right]^{-1} \gamma' C \int_0^1 W_{d_0} (\tau) \, dV_{\alpha} \gamma', $$

where $W_{d_0} (\tau)$ is $p$-dimensional standard Fractional Brownian motion $W_d (\tau) = \Gamma^{-1} (d) \int_0^\tau (\tau - z)^{d - 1} \, dW (z)$ with parameter $d \in (0.5, 1]$, $W_{d_0} (\tau)$ and $dV_{\alpha} (\tau)$ are independent and $dV_{\alpha} (\tau) = \left( \alpha' \Omega^{-1} \alpha \right)^{-1} \alpha' \Omega^{-1} \, dW (\tau)$ with $W$ a Brownian Motion with covariance matrix $\Omega$. The conditional variance of the limit distribution is given by

$$C \int_0^1 W_{d_0} (\tau) W_{d_0} (\tau)' \, d\tau C' \left[ \left( \alpha' \Omega^{-1} \alpha \right)^{-1} \right]$$

and $C = \beta \left( \alpha' \beta \right)^{-1} \alpha'$.

We can observe that the distribution of $\hat{\beta}$ given by Theorem 3 is similar to the distribution found in Johansen (1995) for $d_0 = 1$ fixed. It is also equal to the distribution that Hualde and Robinson (2004) found for their GLS estimator when $r = 1$. The convergence rate of $\hat{\beta}$ is optimal, hence $\hat{\beta} - \beta \in O_P \left( T^{-d_0} \right)$.

We would like to emphasize the fact that the estimate $\hat{\beta}$ is asymptotically independent of other estimates, which means that estimation of the other parameters of the system do not affect the estimate of long run-relationship.

Note that since the asymptotic distribution of $\hat{\beta}$ remains mixed normal, we can test for the values of cointegration vector using Wald test that will be $\chi^2$ distributed. Thus following Johansen (1991) we state the Theorem 4.

Theorem 4 If only 1 cointegrating vector $\beta$ is present ($r = 1$), and if we want to test the hypothesis $K' \beta = 0$, then the test statistic $T^{d_0} \left( K' \hat{\beta} \right)^2 \left( (\hat{\lambda}_1^{-1} - 1)(K' \hat{v} v' K)^{-1} \right)$ is asymptotically $\chi^2$ with 1 degree
of freedom. Here $\lambda_1$ is the maximal eigenvalue and $\hat{\beta}$ the corresponding eigenvector of the equation (12). The remaining eigenvectors form $\tilde{v}$.

In Section 6 we perform a simulation of the Wald test and check that it has proper size and good power in finite samples to test the values of the cointegration vector.

**Theorem 5** The joint asymptotic distribution of $\tilde{\alpha}$ and $\hat{d}$ is given by

$$
\begin{bmatrix}
T^{\frac{1}{2}}(\hat{d} - d_0) \\
T^{\frac{1}{2}}\text{vec}(\tilde{\alpha} - \alpha)
\end{bmatrix} \rightarrow_d N(0, \Psi),
$$

where

$$\Psi = \begin{bmatrix}
\omega^{-1} & c_0\omega^{-1}\text{vec}(\alpha)' \\
\omega^{-1}\text{vec}(\alpha) & \frac{1}{\omega_0} \left( \Sigma_{\beta\beta}^{-1} \otimes \Omega \right) + \frac{\sigma^2_{0\alpha}}{\omega_0} \text{vec}(\alpha) \text{vec}(\alpha)'
\end{bmatrix}
$$

and

$$\omega = \frac{\pi^2}{6} (1 - \rho_0^2) \text{tr} \left( \Sigma_{\beta\beta} \alpha'\Omega^{-1}\alpha \right),
$$

$$\rho_0^2 = \frac{\sigma^2_{0\alpha}}{a_0 \pi^2 / 6}, \quad \hat{\Sigma}_{\beta\beta} = \beta' \Omega \beta, \quad a_0 = \sum_{j=1}^{\infty} \{ \pi_j(d_0) \}^2, \quad c_0 = -\sum_{j=1}^{\infty} j^{-1} \pi_j(d_0)
$$

The asymptotic distribution of $\tilde{\alpha}$ is root-$T$ consistent and we can observe that it is related with the asymptotic distribution of $\hat{d}$. Therefore estimation of the degree of the fractional cointegration $d$ affects the speed of the adjustment to the equilibrium coefficients, which agrees with common intuition about the speed of the convergence to the long run equilibrium. The asymptotic variance is the usual result when $d_0 = 1$ is known with the extra multiplicative term $a_0$ and the contribution from estimation of $d$, $(\sigma^2_{0\alpha}/\omega_0) \text{vec}(\alpha) \text{vec}(\alpha)'$.

The cointegration degree estimator $\hat{d}$ is also root-$T$ consistent and has asymptotic normal distribution. The asymptotic variance includes the factor $(1 - \rho_0^2)^{-1} > 1$ due to estimation of $\alpha$, the factor $\text{tr} \left( \Sigma_{\beta\beta} \alpha'\Omega^{-1}\alpha \right)^{-1}$ due to estimation inside the ECM and finally, the factor $(\pi^2 / 6)^{-1}$ is the usual asymptotic variance for MLEs of memory parameters in univariate ARFIMA$(0, d, 0)$. Note that $\frac{\pi^2}{6} (1 - \rho_0^2) \Sigma_{\beta\beta} = \hat{\Sigma}_{\beta\beta}(d_0) - \sigma^2_{0\alpha}(d_0) \Sigma_{\beta\beta}^{-1}(d_0) \hat{\Sigma}_{\beta\beta}(d_0)$.

We present proofs of Theorem 1, 2, 3 and 5 in Appendix B. In fact the same conclusions can be obtained without resourcing to (15) and (16) in Assumption 1 and using standard results on the existence of a consistent sequence of solutions to stochastic optimization problems, such as Lemma 1 in Andrews and Sun (2004). In Appendix C this is investigated under the assumption that $\Omega$ is known and $r = 1$ not to complicate in excess the presentation.

### 5 Short run correlation

We use the model (7)-(8) proposed in Avarucci (2007), that allows for short run correlation in both the fractional cointegration relationship and in the levels. Note that this model can be shown to
encompass triangular models used in the literature (cf. Robinson and Hualde (2003)) and has nice representations if the roots of the equation \(|A(z)| = 0\) are out of the unit circle, \(\delta > d\). Basically this model implies that there is fractional cointegration among the prewhitened series \(X_t^j = A(L)X_t\).

It can also be seen as a multivariate extension of Hualde and Robinson’s (2006b) bivariate cointegrated model.

The model (7) is nonlinear in \(\Pi\) and \(A_1, \ldots, A_k\), but we propose to estimate the unrestricted linear model (8) without imposing \(B_j = -A_j\Pi\). Then the estimation procedure runs as in Johansen (1995), but with an initial step to prewhiten the main series \(\Delta^d X_t\) and \(\Pi(\Delta^{-d} - 1)\Delta^d X_t\) on \(k\)-lags of both \(\{(\Delta^{-d} - 1)\Delta^d X_t\}\) and \(\Delta^d X_t\) as in equation (8). This estimate is inefficient compared with the MLE, but much simpler to compute and analyze.

Let’s maintain the assumption that \(\delta\) is known and fix \(\delta = 1\) to easy notation. We are interested in the asymptotic distributions of \(\beta, \tilde{d}\) and the linear parameter estimates \(\left(\hat{\alpha}, \hat{A}_1, \ldots, \hat{A}_k\right)\). If we employ unrestricted estimation, then we could investigate the properties of \(\left(\hat{\alpha}, \hat{A}_1, \ldots, \hat{A}_k, \hat{B}_1, \ldots, \hat{B}_k\right)\), though \(B_j\) are redundant parameters. We can derive all asymptotics results in a similar way to the case with no lag estimation, but obviously the distributions are affected by lag correction compared to those of Theorem 3.

To derive the asymptotic results we should make appropriate changes in the Appendix B. For instance replace \(\Sigma_{\beta\beta} (d)\) by the limit variance of the residuals of the projection of \((\Delta^{-d} - 1)\Delta \beta' X_t\) on \(k\) lags of \(\{(\Delta^{-d} - 1)\Delta X_t\}\) and \(\Delta X_t\), but with that change the nice covariance structure in terms of constants \(a_0, c_0\) need not be kept. The asymptotic properties of \(\tilde{d}\) can be deduced from the expansion (28), where now \(\Sigma_{\beta\beta} (d_0)\), \(\Sigma_{\beta\beta} (d_0)\) and \(\Sigma_{\beta\beta} (d_0)\) have to be replaced by the limit variance and covariances of \((\beta' Z_{1t} (d_0), \beta' Z_{1t}^{(1)} (d_0))\), \(\Sigma_{\beta\beta} (d_0)\), etc., when projected on \(k\) lags of \(\{(\Delta^{-d} - 1)\Delta X_t\}\) and \(\Delta X_t\). Then the following theorem holds.

**Theorem 6** Under Assumption 1 and model 7 \(\tilde{\beta}\) has the same properties as in Theorems 1 and 3, and \(\tilde{d}, \tilde{\alpha}, \tilde{A}_1, \ldots, \tilde{A}_k\) have an asymptotic normal joint distribution.

The asymptotic distribution of \(\tilde{d}\) is

\[
T^{1/2}(\tilde{d} - d_0) \to_d N(0, \tilde{\omega}^{-1})
\]

where

\[
\tilde{\omega} = tr \left( \left[ \Sigma_{\beta\beta}^+ (d_0) - \Sigma_{\beta\beta} (d_0) \Sigma_{\beta\beta}^{-1} (d_0) \Sigma_{\beta\beta}^+ (d_0) \right] \alpha' \Omega^{-1} \alpha \right).
\]

For example \(\Sigma_{\beta\beta}^+ (d_0)\), can be estimated consistently by

\[
\frac{1}{T} \sum_{t=1}^{T} \tilde{\beta}' Z_{1t}^+ (\tilde{d}) Z_{1t}^+ (\tilde{d}) \tilde{\beta}
\]

where \(Z_{1t}^+ (\tilde{d})\) are the OLS residuals of projecting \(Z_{1t} (\tilde{d})\) against \(k\) lags of \(\{(1 - \Delta^{-\tilde{d}})\Delta X_t\}\) and \(\Delta X_t, t = 1, \ldots, T\) and \(\tilde{\beta}\) and \(\tilde{d}\) are MLE estimates of \(\beta\) and \(d\).
For $\alpha$ we could obtain a similar expression to (29), in terms of the projected series, and for $\tilde{A}_j$ a parallel result as in Johansen (1995), Theorem 13.5, but corrected for the $d$ estimation increment as in Theorem 3.

6 Monte Carlo

To evaluate the small sample properties of the ML estimators of cointegrated fractional VECM model we have designed the following Monte Carlo experiment. We have generated the two equation model (see Engle, Granger (1987), Banerjee et al. (1993), p.137 or Lyhagen (1998) and also Lasak (2005))

\[
x_t + by_t = u_t
\]

\[
x_t + ay_t = e_t
\]

where $\Delta^{1-d} u_t = \varepsilon_{1t}$, $\Delta e_t = \varepsilon_{2t}$ and $\varepsilon_{1t}$, $\varepsilon_{2t}$ are both independently and standard bivariate normally distributed with expectation zero. $d$ is the cointegration degree and we have considered $d \in [0.5, 1]$. Note that if $d = 1$ then we are in the special case of Johansen’s unit root framework. $\beta = [1, b]'$ is the cointegrating vector, $\alpha = [1 - a]'$ is the vector of speed of adjustment to the equilibrium coefficients. In all simulations we used the same parameters $a$ and $b$ equal to 1 and 2 respectively.

Note that model (17) is a special case of the model (3), with $\alpha = \xi$.

All Monte Carlo simulations were done using Ox 3.40 or Ox 4.04 (see Doornik and Ooms (2001) and Doornik (2002)). To maximize the likelihood function we used the MaxSQPF procedure and optimization was done on the interval $D = [0.5000001; 1]$. For all the simulations we have made 10,000 iterations.

We have calculated bias and standard deviation of the estimators $\tilde{d}$, $\tilde{\beta}$ and $\tilde{\alpha}$ for the values of the true $d$, $d_0 = 0.5$, 0.6, 0.7, 0.8, 0.9, 1 and sample sizes of 50, 100, 200 and 500 observations. Results are reported in Tables 1-4 and compared with the same statistics for the ML estimators of $\beta_j$ and $\alpha_j$ (betaJ and alphaJ) assuming $d = 1$ like in the standard unit root model. Note that the asymptotic theory we have developed in fact does not cover the case when $d_0 = 0.5$ nor $d_0 = 1$ (in Theorem 3).

<table>
<thead>
<tr>
<th>$d_0$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias $\tilde{d}$</td>
<td>0.262</td>
<td>0.191</td>
<td>0.135</td>
<td>0.082</td>
<td>0.028</td>
<td>-0.038</td>
</tr>
<tr>
<td>std $\tilde{d}$</td>
<td>0.216</td>
<td>0.205</td>
<td>0.186</td>
<td>0.160</td>
<td>0.127</td>
<td>0.089</td>
</tr>
</tbody>
</table>

| bias $\beta$ | 0.063 | 0.036 | 0.006 | 0.010 | 0.005 | 0.002 |
| std $\beta$   | 3.847 | 0.861 | 0.828 | 0.155 | 0.072 | 0.054 |
| bias $\beta_j$ | 0.055 | 0.047 | 0.018 | 0.008 | 0.005 | 0.002 |
| std $\beta_j$ | 3.302 | 0.965 | 0.266 | 0.098 | 0.068 | 0.052 |
| bias $\alpha$ | -0.004 | -0.173 | -0.039 | -0.226 | -0.069 | -0.007 |
| std $\alpha$   | 11.735 | 17.487 | 8.604 | 10.504 | 4.116 | 5.252 |
| bias $\alpha_j$ | -0.115 | -0.098 | -0.179 | -0.219 | -0.108 | -0.020 |
| std $\alpha_j$ | 15.487 | 8.118 | 8.833 | 10.717 | 1.639 | 7.122 |
Table 2. Bias and standard deviation of the estimators $\hat{d}$, $\hat{\beta}$ and $\hat{\alpha}$ and $T=100$ observations

<table>
<thead>
<tr>
<th>$d_0$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias $\hat{d}$</td>
<td>0.182</td>
<td>0.127</td>
<td>0.095</td>
<td>0.065</td>
<td>0.028</td>
<td>-0.029</td>
</tr>
<tr>
<td>std $\hat{d}$</td>
<td>0.181</td>
<td>0.177</td>
<td>0.162</td>
<td>0.136</td>
<td>0.100</td>
<td>0.061</td>
</tr>
<tr>
<td>bias $\hat{\beta}$</td>
<td>0.014</td>
<td>0.006</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>std $\hat{\beta}$</td>
<td>0.145</td>
<td>0.087</td>
<td>0.064</td>
<td>0.046</td>
<td>0.034</td>
<td>0.026</td>
</tr>
<tr>
<td>bias $\beta_J$</td>
<td>0.016</td>
<td>0.010</td>
<td>0.004</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>std $\beta_J$</td>
<td>0.593</td>
<td>0.119</td>
<td>0.071</td>
<td>0.048</td>
<td>0.033</td>
<td>0.025</td>
</tr>
<tr>
<td>bias $\hat{\alpha}$</td>
<td>-0.130</td>
<td>-0.175</td>
<td>-0.054</td>
<td>-0.041</td>
<td>-0.035</td>
<td>-0.025</td>
</tr>
<tr>
<td>std $\hat{\alpha}$</td>
<td>4.562</td>
<td>6.274</td>
<td>2.207</td>
<td>0.278</td>
<td>0.178</td>
<td>0.141</td>
</tr>
<tr>
<td>bias $\alpha_J$</td>
<td>0.296</td>
<td>-0.208</td>
<td>-0.072</td>
<td>-0.043</td>
<td>-0.035</td>
<td>-0.025</td>
</tr>
<tr>
<td>std $\alpha_J$</td>
<td>41.337</td>
<td>6.128</td>
<td>2.754</td>
<td>0.282</td>
<td>0.176</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Table 3. Bias and standard deviation of the estimators $\hat{d}$, $\hat{\beta}$ and $\hat{\alpha}$ and $T=200$ observations

<table>
<thead>
<tr>
<th>$d_0$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias $\hat{d}$</td>
<td>0.107</td>
<td>0.075</td>
<td>0.058</td>
<td>0.046</td>
<td>0.027</td>
<td>-0.021</td>
</tr>
<tr>
<td>std $\hat{d}$</td>
<td>0.124</td>
<td>0.133</td>
<td>0.127</td>
<td>0.108</td>
<td>0.079</td>
<td>0.042</td>
</tr>
<tr>
<td>bias $\hat{\beta}$</td>
<td>0.006</td>
<td>0.003</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>std $\hat{\beta}$</td>
<td>0.072</td>
<td>0.049</td>
<td>0.036</td>
<td>0.025</td>
<td>0.017</td>
<td>0.012</td>
</tr>
<tr>
<td>bias $\beta_J$</td>
<td>0.013</td>
<td>0.005</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>std $\beta_J$</td>
<td>0.178</td>
<td>0.065</td>
<td>0.041</td>
<td>0.026</td>
<td>0.017</td>
<td>0.012</td>
</tr>
<tr>
<td>bias $\hat{\alpha}$</td>
<td>-0.197</td>
<td>-0.033</td>
<td>-0.023</td>
<td>-0.019</td>
<td>-0.012</td>
<td>-0.011</td>
</tr>
<tr>
<td>std $\hat{\alpha}$</td>
<td>14.832</td>
<td>1.188</td>
<td>0.132</td>
<td>0.108</td>
<td>0.089</td>
<td>0.079</td>
</tr>
<tr>
<td>bias $\alpha_J$</td>
<td>-0.070</td>
<td>-0.045</td>
<td>-0.028</td>
<td>-0.020</td>
<td>-0.012</td>
<td>-0.011</td>
</tr>
<tr>
<td>std $\alpha_J$</td>
<td>7.814</td>
<td>0.765</td>
<td>0.174</td>
<td>0.109</td>
<td>0.089</td>
<td>0.079</td>
</tr>
</tbody>
</table>

Table 4. Bias and standard deviation of the estimators of $\hat{d}$, $\hat{\beta}$ and $\hat{\alpha}$ and $T=500$ observations

<table>
<thead>
<tr>
<th>$d_0$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias $\hat{d}$</td>
<td>0.056</td>
<td>0.030</td>
<td>0.024</td>
<td>0.024</td>
<td>0.019</td>
<td>-0.015</td>
</tr>
<tr>
<td>std $\hat{d}$</td>
<td>0.069</td>
<td>0.082</td>
<td>0.077</td>
<td>0.071</td>
<td>0.058</td>
<td>0.027</td>
</tr>
<tr>
<td>bias $\hat{\beta}$</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>std $\hat{\beta}$</td>
<td>0.039</td>
<td>0.026</td>
<td>0.017</td>
<td>0.011</td>
<td>0.007</td>
<td>0.005</td>
</tr>
<tr>
<td>bias $\beta_J$</td>
<td>0.003</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>std $\beta_J$</td>
<td>0.063</td>
<td>0.035</td>
<td>0.020</td>
<td>0.012</td>
<td>0.008</td>
<td>0.005</td>
</tr>
<tr>
<td>bias $\hat{\alpha}$</td>
<td>-0.018</td>
<td>-0.011</td>
<td>-0.009</td>
<td>-0.006</td>
<td>-0.005</td>
<td>-0.005</td>
</tr>
<tr>
<td>std $\hat{\alpha}$</td>
<td>0.102</td>
<td>0.081</td>
<td>0.070</td>
<td>0.059</td>
<td>0.052</td>
<td>0.047</td>
</tr>
<tr>
<td>bias $\alpha_J$</td>
<td>-0.039</td>
<td>-0.015</td>
<td>-0.011</td>
<td>-0.007</td>
<td>-0.005</td>
<td>-0.005</td>
</tr>
<tr>
<td>std $\alpha_J$</td>
<td>0.190</td>
<td>0.099</td>
<td>0.075</td>
<td>0.060</td>
<td>0.052</td>
<td>0.047</td>
</tr>
</tbody>
</table>

Bias and standard error of $\hat{d}$, $\hat{\beta}$ and $\hat{\alpha}$ are all decreasing with $d_0$ and with sample size $T$. For $\beta$ we obtain very good estimates already for moderate values of $d_0$ in larger samples. We can estimate $\beta$ much better than $\alpha$ even for small values of $d_0$ where $\hat{\beta}$ has convergence rate close to $T^{\frac{1}{2}}$. Estimates
obtained on the basis of the fractional model have smaller bias and standard error than estimates from the standard VECM model.

Further we have simulated the size and the power of the Wald test given in Theorem 4. We have used again the system described by (17). To check the size we have tested the true restriction $K'\beta = 0$ with $K = [-2, 1]$, while to check the power we have tested the false linear restriction $K'\beta = 0$ with $K = [-3, 1]$. We have compared the properties of the Wald test based on the fractional VECM with Wald test based on standard VECM. Size accuracy of both considered tests is presented in Table 5. Power comparison is presented in Table 6. We can easily observe that if we base Wald test on standard VECM model in case when we have fractionally cointegrated system the test does not have a proper size. Size distortions in this case are significant even for values of $d_0$ relatively close to 1.

Table 5. Size of the Wald test

<table>
<thead>
<tr>
<th>$T / d$</th>
<th>estimated $\hat{d}$</th>
<th>fixed $d = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5 0.6 0.7 0.8 0.9 1</td>
<td>0.5 0.6 0.7 0.8 0.9 1</td>
</tr>
<tr>
<td>50</td>
<td>21.34 18.55 16.86 14.37 10.72 6.83</td>
<td>42.01 35.33 29.01 22.14 14.43 8.23</td>
</tr>
<tr>
<td>100</td>
<td>4.22 4.48 5.16 5.64 6.94 7.61</td>
<td>45.06 40.56 33.06 24.77 14.07 6.45</td>
</tr>
<tr>
<td>200</td>
<td>3.75 4.80 6.18 7.92 7.75 4.26</td>
<td>51.56 47.09 39.9 28.31 15.51 5.62</td>
</tr>
</tbody>
</table>

Table 6. Power of the Wald test

<table>
<thead>
<tr>
<th>$T / d$</th>
<th>estimated $\hat{d}$</th>
<th>fixed $d = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5 0.6 0.7 0.8 0.9 1</td>
<td>0.5 0.6 0.7 0.8 0.9 1</td>
</tr>
<tr>
<td>50</td>
<td>65.55 76.07 86.53 93.74 97.72 99.36</td>
<td>88.46 94.45 97.86 99.21 99.80 99.90</td>
</tr>
<tr>
<td>100</td>
<td>68.98 85.47 95.54 99.20 99.94 100</td>
<td>96.34 99.30 99.93 99.99 100 100</td>
</tr>
<tr>
<td>200</td>
<td>79.72 95.64 99.83 99.99 100 100</td>
<td>99.27 99.96 100 100 100 100</td>
</tr>
</tbody>
</table>

We could also think of testing the value of $d$ using t-test. However based on our simulations we have observed that this test have much distorted size, so we do not report the results. We expect that it happens due to the fact that estimates of $d$ have significant bias.

We have computed the following feasible variance $\omega_T^{-1}$ of the asymptotic distribution of the estimator of $d$, where

$$\omega_T = \frac{\pi^2}{6} (1 - \rho_0^2) \text{tr} \left( \hat{\beta}' S_{11} \left( \hat{d} \right) \hat{\beta} \hat{\Omega}^{-1} \hat{\alpha} \right),$$

$$\rho_0^2 = \frac{\sigma_0^2}{a_0 T \pi^2/6}, \quad \sigma_0 T = \sum_{j=1}^T \pi_j \left( \hat{d} \right)^2, \quad \sigma_0 = - \sum_{j=1}^T j^{-1} \pi_j \left( \hat{d} \right)$$

and have compared average value of the standard deviation obtained throughout the iterations with the corresponding true value calculated for given sample size $T$ and true value of cointegration degree $d_0$. Results are presented in Tables 7 and 8.
Table 7. Asymptotic standard deviation of $\tilde{d}$, $(\sqrt{T} \omega_T(d_0))^{-1}$

<table>
<thead>
<tr>
<th>$T / d$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.522</td>
<td>0.391</td>
<td>0.307</td>
<td>0.249</td>
<td>0.207</td>
<td>0.176</td>
</tr>
<tr>
<td>100</td>
<td>0.370</td>
<td>0.277</td>
<td>0.217</td>
<td>0.176</td>
<td>0.147</td>
<td>0.125</td>
</tr>
<tr>
<td>200</td>
<td>0.262</td>
<td>0.196</td>
<td>0.154</td>
<td>0.125</td>
<td>0.104</td>
<td>0.088</td>
</tr>
<tr>
<td>500</td>
<td>0.166</td>
<td>0.124</td>
<td>0.097</td>
<td>0.079</td>
<td>0.066</td>
<td>0.056</td>
</tr>
<tr>
<td>1000</td>
<td>0.117</td>
<td>0.088</td>
<td>0.069</td>
<td>0.056</td>
<td>0.046</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Table 8. Average standard error of $\tilde{d}$, $(\sqrt{T} \omega_T(\tilde{d}))^{-1}$

<table>
<thead>
<tr>
<th>$T / d$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.347</td>
<td>0.291</td>
<td>0.248</td>
<td>0.215</td>
<td>0.189</td>
<td>0.171</td>
</tr>
<tr>
<td>100</td>
<td>0.289</td>
<td>0.233</td>
<td>0.192</td>
<td>0.161</td>
<td>0.139</td>
<td>0.123</td>
</tr>
<tr>
<td>200</td>
<td>0.227</td>
<td>0.179</td>
<td>0.143</td>
<td>0.118</td>
<td>0.100</td>
<td>0.088</td>
</tr>
<tr>
<td>500</td>
<td>0.154</td>
<td>0.119</td>
<td>0.094</td>
<td>0.077</td>
<td>0.064</td>
<td>0.056</td>
</tr>
<tr>
<td>1000</td>
<td>0.112</td>
<td>0.086</td>
<td>0.068</td>
<td>0.055</td>
<td>0.046</td>
<td>0.040</td>
</tr>
</tbody>
</table>

We observe that for small sample sizes the estimated standard deviation of the asymptotic distribution of $\tilde{d}$ is underestimated, but for larger samples the results get closer to their theoretical values, which confirms our expectation that problems with size of t-test comes from the bias of $\tilde{d}$ rather than from its standard deviation.

Note also that the asymptotic theory we have developed does not cover the cases when $d_0 = 0.5$ nor $d_0 = 1$. In fact when $d_0 = 0.5$ the value reported in Table 7 is finite only due to the truncation at $T$, otherwise the infinite sum in $a_0$ would not converge.

7 Conclusions

In this paper we have considered a generalization of the analysis of cointegrated systems to the fractional case. We have investigated asymptotic properties of the ML estimators of fractional VECM models and have proven that all estimators can be estimated consistently. We have shown that the asymptotic distributions of the estimators of the cointegration vectors remain mixed normal, hence we can test for the values of cointegration vectors using Wald test. The asymptotic distributions of the estimators of the speed of the adjustment to the equilibrium coefficients $\alpha$ and cointegration degree $d$ are joint normal.

The most natural extension would be to consider more general models that allow for different persistence of the series, with memory different from one and possibly unknown. We could also think of introducing trends and structural breaks. Note that we did not consider the case when true $d$ is on the boundary, in particular $d_0 = 1$, $D = [d, 1]$. However note that such case is mainly of interest technically, because we could always allow for $D = [\tilde{d}, \tilde{d}]$ for $\tilde{d} > 1$, if we want to make inference under $H_0 : d = 1$. Then all arguments go through as far as $\tilde{d} - \tilde{d} < \frac{1}{2}$. 

8 Appendix A

We give in Lemma 7 the formulas for those elements of (14) which we need in this paper.

Lemma 7 Under the triangular model (3), so that \( \beta' \Delta X_t = \beta' \Delta d \varepsilon_t \), we have

\[
\Sigma_{\beta 0} (d) = \beta' \Omega \alpha' \sum_{j=1}^{\infty} \{\pi_j (d_0) - \pi_j (d_0 - d)\} \pi_j (d_0) =: \Sigma_{\beta 0} \alpha' \cdot b (d, d_0)
\]

\[
\Sigma_{\beta \beta} (d) = \beta' \Omega \beta \sum_{j=1}^{\infty} (\pi_j (d_0) - \pi_j (d_0 - d))^2 := \Sigma_{\beta \beta} \alpha \cdot a (d, d_0)
\]

where \( \Sigma_{\beta \beta} = \beta' \Omega \beta \). Then \( a (d_0, d_0) = b (d_0, d_0) = a_0 \) and

\[
\Sigma_{\beta 0} (d_0) = \beta' \Omega \beta \alpha' \sum_{j=1}^{\infty} \pi_j (d_0)^2 = \Sigma_{\beta 0} \alpha' \cdot a_0
\]

\[
\Sigma_{\beta \beta} (d_0) = \beta' \Omega \beta \sum_{j=1}^{\infty} (\pi_j (d_0))^2 := \Sigma_{\beta \beta} \cdot a_0
\]

\[
\dot{\Sigma}_{\beta \beta} (d_0) = \Sigma_{\beta \beta} \cdot c_0
\]

\[
\ddot{\Sigma}_{\beta \beta} (d_0) = \Sigma_{\beta \beta} \cdot \frac{\pi^2}{6}
\]

where \( c_0 = \sum_{j=1}^{\infty} j^{-1} \pi_j (d_0) \).

Proof. Let us demonstrate the result for \( \Sigma_{\beta \beta} (d) = \lim_{t \to \infty} \text{Var}(\beta' Z_{1t} (d)) \).

\[
\text{Var}(\beta' Z_{1t} (d)) = E \left\{ \frac{1}{T} \sum_{t=1}^{T} \beta' Z_{1t} (d) Z_{1t} (d) \beta' \right\} = \frac{1}{T} \sum_{t=1}^{T} E \{ (\Delta^{-d}_t - 1) \beta' \Delta X_t \} \{ \Delta X_t' \beta (\Delta^{-d} - 1) \}
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} E \{ (\Delta^{d_0-d}_t - \Delta^{d_0}) \beta' \varepsilon_t \} \{ \varepsilon_t' \beta (\Delta^{d_0-d} - \Delta^{d_0}) \}
\]

which converges to

\[
\beta' \Omega \beta \sum_{j=1}^{\infty} (\pi_j (d_0 - d) - \pi_j (d_0))^2.
\]

Other elements of (14) could be calculated in a similar way noting for example that

\[
\beta' Z_{1t}^{(1)} (d) = \beta' \frac{\partial}{\partial d} Z_{1t} (d) = \frac{\partial}{\partial d} \{ (\Delta^{-d} - 1) \beta' \Delta X_t \}
\]

\[
= \frac{\partial}{\partial d} \{ (\Delta^{-d}) \Delta^{d_0} \beta' \varepsilon_t = - \log \Delta (\Delta^{d_0} - 1) \beta' \varepsilon_t \}
\]

Define

\[
S_{10}^{(i)} (d) = T^{-1} \sum_{t=1}^{T} \left\{ (\partial / \partial d)^i Z_{1t} (d) \right\} Z_{1t}^{(i)}
\]
and
\[ S^{(i,j)}_{11}(d_a, d_b) = T^{-1} \sum_{t=1}^{T} \left\{ (\partial / \partial d)^j Z_{it}(d_a) \right\} \left\{ (\partial / \partial d)^i Z_{it}(d_b) \right\}^\prime. \]

**Lemma 8** Under the triangular model (3), so that \( \beta' \Delta X_t = \beta' \Delta d_0 \varepsilon_t \), we have that, uniformly in \( d \in D \subset (0.5, 1) \),
\[
\begin{align*}
(\text{a}) & \quad \beta' S_{11}(d, d) \beta \to \rho a (d, d_0) \Sigma \beta, \\
(\text{b}) & \quad \beta' S^{(i)}_{12}(d) = O_p \left( T^{-1/2} \right), \quad i = 0, 1, 2, \\
(\text{c}) & \quad T^{1/2-\delta} \beta' S^{(i)}_{12}(d) \gamma \to p0, \quad i, j = 0, 1, 2, \\
(\text{d}) & \quad T^{1/2-\delta} \beta' S^{(i)}_{12}(d) = O_p \left( T^{-1/2} \right), \quad i = 0, 1, 2. \\
\end{align*}
\]

**Proof.** We first give the proof for (a). We have that \( X_t (d) = (\Delta^{-d} - 1) \Delta X_t \), so that
\[
\beta' X_t (d) = (\Delta^{-d} - 1) \beta' \Delta X_t = (\Delta^{-d} - 1) \beta' \Delta d_0 \varepsilon_t = (\Delta d_0^{-d} - \Delta d_0) \beta' \varepsilon_t = \sum_{j=1}^{t-1} \phi_j (d) \beta' \varepsilon_{t-j},
\]
where \( \phi_j (d) = \pi_j (d_0 - d) - \pi_j (d_0) \). Then
\[
E \left[ \beta' S_{11}(d) \beta \right] = \beta' \Omega \beta \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \left\{ \pi_j (d_0 - d) - \pi_j (d_0) \right\}^2
\]
\[
= a (d, d_0) \beta' \Omega \beta + o (1).
\]
Then we can write
\[
\beta' S_{11}(d) \beta - a (d, d_0) \beta' \Omega \beta = \beta' S_{11}(d) \beta - E \left[ \beta' S_{11}(d) \beta \right]
\]
\[
+ E \left[ \beta' S_{11}(d) \beta \right] - a (d, d_0) \beta' \Omega \beta
\]
where the second line converges uniformly in \( d \) to 0, and writing \( B_T (d) = \beta' S_{11}(d) \beta - E \left[ \beta' S_{11}(d) \beta \right] \), it is easy to show that \( B_T (d) = o_p (1) \) for each fixed \( d \). Now we show tightness in \( d \) of \( B_T (d) \) implying that \( \sup_d |B_T (d)| = o_p (1) \). For a typical element of \( B_T (d) \) and \( d_a, d_b \in D \), we have that
\[
E \left[ B_T^{(r,s)} (d_a) - B_T^{(r,s)} (d_b) \right]^2
\]
\[
\leq \left( [\beta' \Omega \beta]^r s \right)^2 \left[ \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \left( \phi_j (d_a) - \phi_j (d_b) \right)^2 \right]^2
\]
\[
+ \frac{1}{T^2} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \sum_{j'=1}^{t-1} \sum_{j''=1}^{t-1} E \left[ u_{t-j} v_{t-j'j} u_{t-j''} v_{t-j''} \right]
\]
\[
\times \left\{ \phi_j (d_a) \phi_i (d_a) - \phi_j (d_b) \phi_i (d_b) \right\} \left\{ \phi_j (d_a) \phi_i (d_a) - \phi_j (d_b) \phi_i (d_b) \right\} \quad (18)
\]
where \( u_t = [\beta' \varepsilon_1]^r \), \( v_t = [\beta' \varepsilon_1]^s \).

Now we have that
\[
\frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \left( \phi_j (d_a) - \phi_j (d_b) \right)^2
\]
\[
= \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \left( \pi_j^2 (d_0) \pi_j^2 (d_0) - 2 \pi_j (d_0) (\pi_j (d_0) - \pi_j (d_0) \right)
\]
\[
= 16
\]
and for an intermediate point $d^*$ between $d_a$ and $d_b$ and $\hat{\pi}_j = (\partial/\partial x) \pi_j (x)$ this in absolute value is not larger than

$$\frac{K}{T |d_a - d_b|} \sum_{t=1}^{T} \sum_{j=1}^{t-1} \left( \hat{\pi}_j (d_0 - d^*) \pi_j (d_0 - d^*) - 2\pi_j (d_0) \hat{\pi}_j (d_0 - d^*) \right) \leq K |d_a - d_b|$$

uniformly in $T$ because $\hat{\pi}_j (d_0 - d^*) \pi_j (d_0 - d^*)$ and $\pi_j (d_0) \pi_j (d_0 - d^*)$ are square summable and can be bounded by $K_{j^{-\eta-1/2}}$, for some $\eta > 0$, because $d_0, d_a, d_b \in (0.5, 1]$ and therefore $|d_0 - d^*| < 0.5$.

On the other hand (18) has terms with four typical forms, cf. proof of Theorem 3 in Lasak (2005). The difference with respect to this case is that the weight functions $\phi_j (d)$ are now square summable for any combination of parameters and these can be bounded by $K_{j^{-\eta-1/2}}$, some $\eta > 0$, while the differences $|\phi_j (d_a) - \phi_j (d_b)|$ can be bounded by $|d_a - d_b|K_{j^{-\eta-1/2}}$, some $\eta > 0$, and uniformly in $j$. Then the contribution of (18) is of order of magnitude

$$\left( \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{t-1} |d_a - d_b| \left( K_{j^{-\eta-1/2}} \right)^2 \right) \leq K |d_a - d_b|^2,$$

which shows the tightness of $B_T$ and the uniformity of (a).

For the proof of (b) we note that $E \left[ \beta' S_{1\varepsilon}^{(i)} (d) \right] = 0$, while the variance of a typical element of $S_{1\varepsilon}^{(0)} (d)$ is

$$\text{Var} \left[ \{ \beta' S_{1\varepsilon} (d) \} \right] \leq \frac{K}{T^2} \sum_t \sum_j \phi_j (d)^2 = O \left( T^{-1} \right)$$

and the uniformity in $d$ for any $i$ can be shown using similar techniques. For terms involving derivatives, note that the asymptotic approximations for the derivatives of $\pi_j (\cdot)$ for large $j$ are like those for $\pi_j (\cdot)$ up to logarithmic terms.

For the proof of (c) we note that

$$\xi' X_i (d) = (\Delta^{-d} - 1) \xi' \Delta X_i = (\Delta^{-d} - 1) \xi' \varepsilon_t = (\Delta^{-d} - 1) \xi' \varepsilon_t = \sum_{j=1}^{t-1} \pi_j (-d) \xi' \varepsilon_{t-j},$$

so that for $i, j = 0$,

$$T^{1/2-d} \beta' S_{11} (d, d) \gamma = T^{-1/2-d} \beta' \sum_{t=1}^{T} \sum_{j=1}^{t} \sum_{i=1}^{t} \phi_j (d) \pi_i (-d) \varepsilon_{t-j} \varepsilon_{t-i} \gamma.$$

We first note that

$$E \left[ T^{1/2-d} \beta' S_{11} (d, d) \gamma \right] = T^{-1/2-d} \beta' \sum_{t=1}^{T} \sum_{j=1}^{t} \phi_j (d) \pi_j (-d) \Omega \gamma = O \left( T^{-1/2-d} \sum_{t=1}^{T} \sum_{j=1}^{t} \gamma_{-3/2+d-\epsilon} \right) = O (T^{-\epsilon}) = o (1)$$

for some $\epsilon > 0$, and similarly we can show that for each $d$, $\text{Var} \left[ T^{1/2-d} \beta' S_{11} (d, d) \gamma \right] = o (1)$ as $T \to \infty$. Then tightness follows as in the proof of Theorem 1 in Lasak (2005) and thus

$$\sup_d |T^{1/2-d} \beta' S_{11} (d, d) \gamma| = o_p (1).$$

The argument for other values of $i$ and $j$ is similar.
The proof of (d) follows combining ideas of the proofs of (b) and (c).

**Lemma 9** Under the triangular model (3), so that $\beta'\Delta X_t = \beta'\Delta^d\varepsilon_t$, we have that, uniformly in $d$ such that $|d - d_0| \leq T^{-\kappa}$, some $\kappa > 0$, and for all $\eta > 0$,

\[
\begin{align*}
(a) & \quad \beta' S_{11} (d, d) \beta \to_p \Sigma_{\beta \beta} (d_0) = a_0 \Sigma_{\beta \beta} \\
& \quad \beta' S_{11}^{(1,0)} (d, d) \beta \to_p \tilde{\Sigma}_{\beta \beta} (d_0) = c_0 \tilde{\Sigma}_{\beta \beta} \\
& \quad \beta' S_{11}^{(1,1)} (d, d) \beta \to_p \bar{\Sigma}_{\beta \beta} (d_0) = \frac{\pi^2}{6} \bar{\Sigma}_{\beta \beta} \\
(b) & \quad \beta' S_{1c}^{(i)} (d) = O_p \left( T^{-1/2} \right), \quad i = 0, 1, 2. \\
(c) & \quad \beta' S_{11}^{(i,j)} (d, d) \tilde{\gamma} = O_p \left( T^{d_0 - 1 + \eta} \right), \quad i, j = 0, 1, 2. \\
(d) & \quad \tilde{\gamma}' S_{1c} (d) = O_p \left( T^{d_0 - 1 + \eta} \right), \quad i = 0, 1, 2. \\
(e) & \quad \beta' \left\{ S_{11}^{(i)} (d_0) - S_{1c}^{(i)} (d) \right\} = o_p \left( T^{-1/2} \right), \quad i = 0, 1. \\
& \quad \tilde{\gamma}' \left\{ S_{1c} (d_0) - S_{1c} (d) \right\} = o_p \left( T^{d_0 - 1} \right). \\
\end{align*}
\]

When $d_0 = 1$ we can set $\eta = 0$.

**Proof.** Omitted, the proofs of (a) – (b) being similar to Lemma A. For the proof of (e) follow the methods of the proof in Appendix B in Lasak (2005).

**Lemma 10** Let the process $X_t$ be given by (3), choose $\gamma$ orthogonal to $\beta$ such that $(\beta, \gamma)$ has full rank $p$. Then for $d_0 \in (0.5, 1]$ as $T \to \infty$ and $u \in [0, 1]$

\[
T^{1/2} \tilde{d}_c \tilde{\gamma}' X_{[Tu]} \xrightarrow{d} \tilde{\gamma}' CW_{d_0} (u),
\]

\[
T^{1 - d_0} \tilde{\gamma}' S_{1c}(d_0) \xrightarrow{d} \tilde{\gamma}' C \int_0^1 W_{d_0} (\tau) dW (\tau)'
\]

\[
T^{1 - 2d_0} \tilde{\gamma}' S_{11}(d_0, d_0) \tilde{\gamma} \xrightarrow{d} \tilde{\gamma}' C \int_0^1 W_{d_0} (\tau) W_{d_0} (\tau)' d\tau C' \tilde{\gamma}
\]

where $C = \beta_{\perp} (\alpha'_{\perp} \beta_{\perp})^{-1} \alpha'_{\perp}$.

**Proof.** The result follows by similar arguments as in Theorem B.13 of Johansen (1995) and weak convergence follows from Lemma (10) and Marinucci and Robinson (2000).

**Lemma 11** Under Assumption 1

\[
T^{1/2} tr \left\{ \alpha \left[ \Sigma_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) \beta' S_{1c} (d_0) - \beta' S_{1c}^{(i)} (d_0) \right] \Omega^{-1} \right\} \to_d N (0, \omega^2).
\]
Proof. Use the martingale Central Limit Theorem and that
\[
\lim_{T \to \infty} \text{Var} \left( T^{1/2} \text{tr} \left\{ \alpha \left[ \hat{\Sigma}_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) \beta' S_{11} (d_0) - \beta' S_{11}^{(1)} (d_0) \right] \Omega^{-1} \right\} \right)
\]
\[
= \lim_{T \to \infty} \text{Var} \left( \frac{1}{T} \sum_{t=1}^{T} \left[ \hat{\Sigma}_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) \beta' Z_t (d_0) - \beta' Z_t^{(1)} (d_0) \right] \right)
\]
\[
= \frac{1}{T} \sum_{t=1}^{T} \text{tr} \left\{ \left[ \frac{\hat{\Sigma}_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) \beta' Z_t (d_0) - \beta' Z_t^{(1)} (d_0) \right] \right\}
\]
\[
= \text{tr} \left\{ \frac{\pi^2}{6} \hat{\Sigma}_{\beta \beta} (d_0) - \frac{\pi^2}{a_0} \Sigma_{\beta \beta} \right\} \alpha' \Omega^{-1} \alpha = \omega.
\]

\[\boxdot\]

9 Appendix B

Proof. (of Theorem 1) Define the matrix \( A_T (d) = \left( \beta, T^{\frac{1}{2}-d} \hat{\gamma} \right) \). For any value of \( d, d > 0.5 \) the ordered eigenvalues of
\[
|\lambda (d) A_T (d) S_{11} (d) A_T (d) - A_T (d) S_{10} (d) S_{00}^{-1} S_{01} (d) A_T (d) | = 0.
\]
converge to those of
\[
|\lambda (d) \Sigma_{\beta \beta} (d) - \Sigma_{\beta \beta} (d) \Sigma_{\beta \beta}^{-1} (d) | \ | \lambda (d) \int_{0}^{1} W_d (r) W_d (r') \, dr | = 0
\]
and the space spanned by the \( r \) first eigenvectors of (19) converges to the space spanned by the first unit vectors or equivalently to the space spanned by vectors with zeros in the last \( p - r \) coordinates. The space spanned by the first \( r \) eigenvectors of (19) is \( \text{sp}(A_T^{-1} (d) \tilde{\beta}) = \text{sp}(A_T^{-1} (d) \tilde{\beta}) \), where \( A_T^{-1} \tilde{\beta} = \left( \beta, T^{\frac{1}{2}+d} \hat{\gamma} \right)' \beta = (I, T^{\frac{1}{2}+d} U_T)' \). Thus we find that \( T^{-\frac{1}{2}+d} U_T \xrightarrow{P} 0 \). This shows consistency of \( \beta \) and moreover that \( \tilde{\beta} - \beta = o_P (T^{\frac{1}{2}-d}) \). Note that (20) has \( p - r \) zero roots and \( r \) positive roots given by the solutions of
\[
|\lambda (d) \Sigma_{\beta \beta} (d) - \Sigma_{\beta \beta} (d) \Sigma_{\beta \beta}^{-1} (d) | = 0,
\]
which can be expressed as
\[
|\lambda (d) \Sigma_{\beta \beta} (d_0) \frac{a (d, d_0)}{a_0} - \Sigma_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) \frac{b (d, d_0)}{a_0} |^2 = 0.
\]
Moreover if \( d = \tilde{d} \), \( \tilde{d} \) is a consistent estimate of \( d \) then (21) converges to
\[
|\lambda (d_0) \Sigma_{\beta \beta} (d_0) - \Sigma_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) | = 0.
\]
Next recall \( \tilde{\beta} = \hat{\gamma} U_T \), so
\[
\beta' S_{11} (d) \tilde{\beta} = (\beta + \hat{\gamma} U_T)' S_{11} (d) (\beta + \hat{\gamma} U_T)
\]
\[
= \beta' S_{11} (d) \beta + \beta' S_{11} (d) \hat{\gamma} U_T + (\hat{\gamma} U_T)' S_{11} (d) \beta + (\hat{\gamma} U_T)' S_{11} (d) (\hat{\gamma} U_T).
\]
Since \( U_T = o_P(T^{\frac{1}{d}-d}) \), by consistency of \( \hat{d} \) and, by Lemma 9, we have that for all \( \eta > 0 \)
\[
\beta' S_{11} \left( \tilde{d} \right) \tilde{\beta} = \beta' S_{11} \left( \tilde{d} \right) \beta + O_P(T^{d_0-1+\eta})o_P(T^{\frac{1}{d}-d_0}) + o_P(T^{1-2d_0})O_P(T^{2d_0-2+\eta})
\]
\[
= \beta' S_{11} \left( \tilde{d} \right) \beta + o_P(T^{\eta/2}) + o_P(T^{\eta-1})
\]
\[
= \beta' S_{11} \left( d_0 \right) \beta + o_P(1) \xrightarrow{P} \Sigma_{\beta\beta} \left( d_0 \right).
\]
and also
\[
\hat{\beta} S_{10} \left( \tilde{d} \right) = (\beta + \gamma U_T)' S_{10} \left( \tilde{d} \right) = \beta' S_{10} \left( \tilde{d} \right) + o_P(T^{\frac{1}{d}-d_0}) \xrightarrow{P} \Sigma_{\beta0} \left( d_0 \right).
\]
Further consider \( \tilde{\alpha} = S_{01} \left( \tilde{d} \right) \tilde{\beta} \left( \beta' S_{11} \left( \tilde{d} \right) \tilde{\beta} \right)^{-1} \), which converges towards
\[
\Sigma_{0\beta} \left( d_0 \right) \Sigma_{\beta\beta}^{-1} \left( d_0 \right) = \alpha \Sigma_{\beta\beta} a_0 \left[ \Sigma_{\beta\beta} a_0 \right]^{-1} = \alpha.
\]
and
\[
\tilde{\Omega} = S_{00} - S_{01} \left( \tilde{d} \right) \tilde{\beta} \left( \beta' S_{11} \left( \tilde{d} \right) \tilde{\beta} \right)^{-1} \beta' S_{10} \left( \tilde{d} \right),
\]
which converges towards
\[
\Sigma_{00} - \Sigma_{0\beta} \left( d_0 \right) \Sigma_{\beta\beta}^{-1} \left( d_0 \right) \Sigma_{\beta0} \left( d_0 \right) = \Sigma_{00} - \alpha \Sigma_{\beta\beta} a_0 \left[ \Sigma_{\beta\beta} a_0 \right]^{-1} \Sigma_{\beta\beta} a' a_0 = \Sigma_{00} - \alpha \Sigma_{\beta\beta} a_0 a' = \Sigma_{00} - \alpha \Sigma_{\beta\beta} \left( d_0 \right) a' = \Omega.
\]

\[\square\]

**Proof.** (of Theorem 2) Using Lemma 8 we have consistency of \( \tilde{\beta} \) and \( \tilde{\beta} - \beta = o_P(T^{\frac{1}{d}-d}) \) for any \( d, d > 0.5 \), however we do not have consistency of estimators of \( \alpha \) and \( \Omega \) for a fixed \( d \neq \tilde{d} \), because \( \tilde{\alpha} \) converges towards
\[
\Sigma_{0\beta} \left( d_0 \right) \Sigma_{\beta\beta}^{-1} \left( d_0 \right) = \alpha \Sigma_{\beta\beta} b \left( d, d_0 \right) \left[ \Sigma_{\beta\beta} a \left( d, d_0 \right) \right]^{-1} = \frac{b \left( d, d_0 \right)}{a \left( d, d_0 \right)},
\]
while \( \tilde{\Omega} \) converges towards
\[
\Sigma_{00} - \Sigma_{0\beta} \left( d_0 \right) \Sigma_{\beta\beta}^{-1} \left( d_0 \right) \Sigma_{\beta0} \left( d_0 \right) = \Sigma_{00} - \alpha \Sigma_{\beta\beta} a_0 \left[ \Sigma_{\beta\beta} a_0 \right]^{-1} \Sigma_{\beta\beta} a' a_0 = \Sigma_{00} - \alpha a_0 \left[ a_0 \right]^{-1} \Sigma_{\beta\beta} \left( d_0 \right) a' = \Omega.
\]

\[\square\]

**Proof.** (of Theorems 3 and 5) The estimators \( \tilde{\alpha}, \tilde{\beta}, \tilde{d} \) and \( \tilde{\Omega} \) satisfy the likelihood equations, so we derive expressions for the derivatives of \( L(\alpha, \beta, d, \Omega) \), the concentrated log-likelihood function, with respect to \( \beta, \alpha \) and \( d \).

The expressions for the derivatives of \( L(\alpha, \beta, d, \Omega) \) with respect to \( \beta \) and \( \alpha \) in the directions \( b \) and \( a \) are respectively:
\[
D_\beta L(\alpha, \beta, d, \Omega) \left( b \right) = \text{tr} \left\{ \Omega^{-1} \left( \sum_{t=1}^{T} \hat{\epsilon}_t Z_{1t}' \left( d \right) ba' \right) \right\} = T \text{tr} \left\{ a' \Omega^{-1} \left( S_{01} \left( d \right) - \alpha \beta' S_{11} \left( d \right) \right) b \right\},
\]
\[
D_\alpha L(\alpha, \beta, d, \Omega) \left( a \right) = \text{tr} \left\{ \Omega^{-1} \left( \sum_{t=1}^{T} \hat{\epsilon}_t Z_{1t}' \left( d \right) ba' \right) \right\} = T \text{tr} \left\{ \Omega^{-1} \left( S_{01} \left( d \right) - \alpha \beta' S_{11} \left( d \right) \right) \beta a' \right\},
\]
where \( \hat{e}_t = Z_{0t} - \alpha \beta' Z_{1t}(d) \)
and the expression for the derivative with respect to \( d \) is
\[
D_d L(\alpha, \beta, d, \Omega)(d) = \sum_{t=1}^{T} [Z_{0t} - \alpha \beta' Z_{1t}(d)]' \Omega^{-1} \Omega \Omega^{-1} \Omega^{-1} [\alpha \beta' Z_{1t}^{(1)}(d)].
\]

From these results we can derive the first order conditions that are satisfied at a maximum point.
At the point \((\alpha, \beta, \tilde{d})\) the derivatives are zero in all directions hence the likelihood equations are:
\[
\begin{align*}
\tilde{\alpha}' & \Omega^{-1} \left( S_{01} \left( \tilde{d} \right) - \tilde{\alpha} \beta' S_{11} \left( \tilde{d} \right) \right) = 0 \quad (23) \\
\left( S_{01} \left( \tilde{d} \right) - \tilde{\alpha} \beta' S_{11} \left( \tilde{d} \right) \right) \tilde{\beta} & = 0 \\
tr \left\{ \tilde{\alpha} \beta' S_{11}^{(1)} \left( \tilde{d} \right) - \tilde{\alpha} \beta' S_{11}^{(1,0)} \left( \tilde{d} \right) \tilde{\alpha}' \tilde{\Omega}^{-1} \right\} & = 0,
\end{align*}
\]

Now substitute \( S_{11}^{(1)} \left( \tilde{d} \right) = S_{11}^{(1)}(\tilde{d}) + S_{11}^{(1,0)}(\tilde{d}, d_0) \beta \alpha' \) in the third equation, with the obvious definition for \( S_{11}^{(1)}(d) \),
\[
tr \left\{ \tilde{\alpha} \beta' S_{11}^{(1)} \left( \tilde{d} \right) - \tilde{\alpha} \beta' S_{11}^{(1,0)} \left( \tilde{d}, d_0 \right) \beta \alpha' \right\} = 0,
\]
and using Taylor expansion, consistency of \( \tilde{\beta} \) and Lemma 9,
\[
\tilde{\beta}' \left[ S_{11}^{(1,0)} \left( \tilde{d} \right) - S_{11}^{(1,0)}(\tilde{d}, d_0) \right] \tilde{\beta} = \tilde{\beta}' S_{11}^{(1,0)}(\tilde{d}, d_0) \beta (\tilde{d} - d_0) + O_{\theta} \left( (\tilde{d} - d_0)^2 \right)
\]
\[
= \frac{\pi^2}{6} \Sigma_{\beta \beta} \left( \tilde{d} - d_0 \right) (1 + o_p(1)),
\]
we get that
\[
tr \left\{ \tilde{\alpha} \beta' S_{11}^{(1)} \left( \tilde{d} \right) - \tilde{\alpha} \beta' S_{11}^{(1,0)} \left( \tilde{d}, d_0 \right) \beta (\tilde{d} - d_0) \right\} = 0,
\]
so by consistency of \( \tilde{\alpha} \) and \( \tilde{\Omega} \),
\[
\tilde{d} - d_0 = tr \left[ \frac{\pi^2}{6} \Sigma_{\beta \beta} \alpha' \Omega^{-1} \alpha \right]^{-1} (1 + o_p(1)) \times \tilde{\alpha}' \tilde{\beta}' S_{11}^{(1,0)}(\tilde{d}, d_0) \beta (\tilde{d} - d_0) \left( \tilde{d} - d_0 \right) (1 + o_p(1)),
\]
and therefore
\[
T^\frac{1}{2}(\tilde{d} - d_0) = \left[ \frac{\pi^2}{6} \Sigma_{\beta \beta} \alpha' \Omega^{-1} \alpha \right]^{-1} \times (1 + o_p(1)) \times \tilde{\alpha}' \tilde{\beta}' S_{11}^{(1,0)}(\tilde{d}, d_0) \left( \tilde{d} - d_0 \right) (1 + o_p(1)) \tilde{\Omega}^{-1} \left( \tilde{d} - d_0 \right) \left( \tilde{d} - d_0 \right) (1 + o_p(1)),
\]
\[
T^\frac{1}{2}(\tilde{d} - d_0) = O_p(1) + O_p \left( T^\frac{1}{2} \right) \left( \tilde{d} - d_0 \right) \left( \tilde{d} - d_0 \right) (1 + o_p(1)) \tilde{\alpha}' \tilde{\beta}' \Omega^{-1} \left( \tilde{d} - d_0 \right) \left( \tilde{d} - d_0 \right) (1 + o_p(1)).
\]
Then, using Lemma 9 and consistency of \( \tilde{\beta} \), we get
\[
T^{1/2}(\tilde{d} - d_0) = O_p(1) + O_p \left( T^{1/2} \right) \left( \tilde{d} - d_0 \right) \left( \tilde{d} - d_0 \right) (1 + o_p(1)) \tilde{\alpha}' \tilde{\beta}' \Omega^{-1} \left( \tilde{d} - d_0 \right) \left( \tilde{d} - d_0 \right) (1 + o_p(1)).
\]

Then, using Lemma 9 and consistency of \( \tilde{\beta} \), we get
\[
T^{1/2}(\tilde{d} - d_0) = O_p(1) + O_p \left( T^{1/2} \right) \left( \tilde{d} - d_0 \right) \left( \tilde{d} - d_0 \right) (1 + o_p(1)) \tilde{\alpha}' \tilde{\beta}' \Omega^{-1} \left( \tilde{d} - d_0 \right) \left( \tilde{d} - d_0 \right) (1 + o_p(1)).
\]

21
Next consider the second equation in (23) and insert 

\[ S_{01}(\bar{d}) = \alpha \beta' S_{11}(d_0, \bar{d}) + S_{z1}(\bar{d}), \]

0 = \( (S_{z1}(\bar{d}) + \alpha \beta' S_{11}(d_0, \bar{d}) - \bar{\alpha} \bar{\beta}' S_{11}(\bar{d})) \bar{\beta} \)

= \( S_{z1}(\bar{d}) \bar{\beta} - (\bar{\alpha} - \alpha) \beta' S_{11}(d_0, \bar{d}) \bar{\beta} + \alpha \beta' S_{11}(d_0, \bar{d}) \bar{\beta} - \alpha \beta S_{11}(\bar{d}) \bar{\beta}. \)

Then, standardizing \( \bar{\alpha} - \alpha \) we obtain that

\[
T^{\frac{1}{2}}(\bar{\alpha} - \alpha) = \left\{ T^{\frac{1}{2}} S_{z1}(\bar{d}) \bar{\beta} + T^{\frac{1}{2}} S_{z1}(\bar{d}) (\bar{\beta} - \beta) + T^{\frac{1}{2}} \alpha \beta' S_{11}(d_0, \bar{d}) \bar{\beta} - T^{\frac{1}{2}} \alpha \beta S_{11}(\bar{d}) \bar{\beta} \right\} \\
\times \left[ \bar{\beta} S_{11}(\bar{d}) \bar{\beta} \right]^{-1} \\
= \left\{ T^{\frac{1}{2}} S_{z1}(\bar{d}) \bar{\beta} + T^{\frac{1}{2}} S_{z1}(\bar{d}) (\bar{\beta} - \beta) - T^{\frac{1}{2}} \alpha \left( \bar{\beta} - \beta \right) S_{11}(d_0, \bar{d}) \bar{\beta} \right\} \\
+ T^{\frac{1}{2}} \alpha \bar{\beta}' \left\{ S_{11}(d_0, \bar{d}) - S_{11}(\bar{d}) \right\} \bar{\beta} \\
\times \left[ a_0 \Sigma_{\beta \beta} \right]^{-1} (1 + o_p(1)).
\]

Then using Taylor expansion and Lemma 9,

\[
\bar{\beta}' \left\{ S_{11}(d_0, \bar{d}) - S_{11}(\bar{d}) \right\} \bar{\beta} = -\bar{\beta}' \left\{ S^{(1,0)}(d_0, \bar{d}) \right\} \bar{\beta} (\bar{d} - d_0) + O_p \left( (\bar{d} - d_0)^2 \right) \\
= c_0 \Sigma_{\beta \beta} (\bar{d} - d_0) (1 + o_p(1))
\]

so that using again Lemma 9, it holds for all \( \eta > 0, \)

\[
T^{\frac{1}{2}}(\bar{\alpha} - \alpha) = O_p(1) + T^{1/2} O_p(T^{d_0 - 1 + \eta}) o_p \left( T^{1/2 - d_0} \right) + O_p \left( T^{\frac{1}{2} T^{-\eta}} \right) \\
= O_p(1) + o_p(T^\eta) + O_p \left( T^{\frac{1}{2} T^{-\eta}} \right)
\]

so \( \bar{\alpha} - \alpha = O_p \left( T^{-\eta} + T^{\eta - 1/2} \right). \) In fact

\[
T^{\frac{1}{2}}(\bar{\alpha} - \alpha) = \left\{ T^{\frac{1}{2}} S_{z1}(\bar{d}) \bar{\beta} + T^{\frac{1}{2}} \alpha \Sigma_{\beta \beta} (d_0) (\bar{d} - d_0) + O_p \left( T^{d_0 - 1/2 + \eta} \right) \left( \bar{\beta} - \beta \right) \right\} \Sigma_{\beta \beta}^{-1}(d_0) (1 + o_p(1))
\]

Consider now the first equation (23) and insert \( S_{01}(\bar{d}) = \alpha \beta' S_{11}(d_0, \bar{d}) + S_{z1}(\bar{d}) \) to get

\[
0 = \bar{\alpha}' \bar{\Omega}^{-1} \left( S_{z1}(\bar{d}) + \alpha \beta' S_{11}(d_0, \bar{d}) - \bar{\alpha} \bar{\beta}' S_{11}(\bar{d}) \right) \\
= \bar{\alpha}' \bar{\Omega}^{-1} \left( S_{z1}(\bar{d}) + \alpha \beta' S_{11}(d_0, \bar{d}) - S_{11}(\bar{d}) \right) - \bar{\alpha} (\bar{\beta} - \beta)' S_{11}(\bar{d}) - (\bar{\alpha} - \alpha) \beta' S_{11}(\bar{d}).
\]

We next multiply by \( \gamma \) from the right and insert \( \bar{\beta} = \beta = \gamma U_T, \)

\[
0 = \bar{\alpha}' \gamma \left( S_{z1}(\bar{d}) + \alpha \beta' S_{11}(d_0, \bar{d}) - S_{11}(\bar{d}) \right) \gamma - \bar{\alpha} U_T \gamma S_{11}(\bar{d}) \gamma - (\bar{\alpha} - \alpha) \beta' S_{11}(\bar{d}) \gamma
\]

so that

\[
T^{d_0} U_T' = \left( \bar{\alpha}' \bar{\Omega}^{-1} \bar{\alpha} \right)^{-1} \left( \bar{\alpha}' \bar{\Omega}^{-1} T^{1 - d_0} S_{z1}(\bar{d}) \gamma \right) \\
+ \bar{\alpha}' \bar{\Omega}^{-1} T^{1 - d_0} \alpha \beta' S_{11}(d_0, \bar{d}) - S_{11}(\bar{d}) \gamma \\
- \bar{\alpha}' \bar{\Omega}^{-1} T^{1 - d_0} (\bar{\alpha} - \alpha) \beta' S_{11}(\bar{d}) \gamma \left( T^{1 - 2d_0} S_{11}(\bar{d}) \gamma \right)^{-1}
\]
Then, following Lemma 9, for any \( \eta > 0 \),
\[
\beta'\{S_{11}\left(d_0, \tilde{d}\right) - S_{11}\left(\tilde{d}\right)\} \tilde{\gamma} = -\beta'\{S_{11}^{(1,0)}\left(d_0, \tilde{d}\right)\} \tilde{\gamma} (\tilde{d} - d_0) + O_p\left((\tilde{d} - d_0)^2\right)
\]
\[
= O_p\left(T^{d_0 - 1 + \eta}\right) (\tilde{d} - d_0)
\]
and by Lemmas 10 and 9, consistency of \( \tilde{\alpha} \), \( \tilde{\Omega} \) and the rate of convergence for \( \tilde{d} \),
\[
T^{d_0} U'_1 = O_p\left(1\right) \left\{ O_p\left(1\right) + O_p\left(T^{\eta}\right) \left\| \tilde{d} - d_0 + \| \tilde{\alpha} - \alpha \| \right\| \right\}
\]
\[
= O_p\left(1\right) \left\{ O_p\left(1\right) + O_p\left(T^{\eta - \kappa}\right) + O_p\left(T^{\eta - 1/2}\right) \right\}
\]
\[
= O_p\left(1\right)
\]
and therefore \( \tilde{\beta} - \beta = O_p\left(T^{-d_0}\right) \), 0.5 < \( d_0 \) ≤ 1.

Now substituting (25) into (24) and ignoring the negligible terms in \( \tilde{\beta} - \beta \), we find that
\[
(\tilde{d} - d_0) \frac{1}{\alpha} \left\{ \left[ \tilde{\Sigma}_{\beta \beta} (d_0) - \tilde{\Sigma}_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) \Sigma_{\beta \beta} (d_0) \right] \alpha' \Omega^{-1} \alpha \right\}
\]
\[
= -\frac{1}{\alpha} \left\{ \alpha \left[ \tilde{\Sigma}_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) \beta' S_{1z} (\tilde{d}) - \beta' S_{1z}^{(1)} (\tilde{d}) \right] \right\} (1 + o_p\left(1\right))
\]
and therefore, using Lemma 9.(e),
\[
\tilde{d} - d_0 = -\omega^{-1} \left\{ \Omega^{-1} \alpha \left[ \tilde{\Sigma}_{\beta \beta} (d_0) \Sigma_{\beta \beta}^{-1} (d_0) \beta' S_{1z} (d_0) - \beta' S_{1z}^{(1)} (d_0) \right] \right\} (1 + o_p\left(1\right))
\]
and the distribution of \( \tilde{d} \) follows using Lemma 11.

For the distribution of \( \tilde{\alpha} \) we can first write
\[
\tilde{d} - d_0 = -\omega^{-1} \left\{ \Omega^{-1} \alpha \left[ \frac{c_0}{a_0} \beta' S_{1z} (d_0) - \beta' S_{1z}^{(1)} (d_0) \right] \right\} (1 + o_p\left(1\right))
\]
\[
= -\omega^{-1} \left(1 + o_p\left(1\right)\right) \frac{1}{\alpha} \sum_{t=1}^{T} \frac{\varepsilon_t}{\sqrt{t}} \Omega^{-1} \alpha \left[ \frac{c_0}{a_0} \beta' Z_{1t} (d_0) - \beta' Z_{1t}^{(1)} (d_0) \right]
\]
so that
\[
T^{1/2} (\tilde{\alpha} - \alpha) = \left(1 + o_p\left(1\right)\right) \frac{1}{\omega^{1/2}} \sum_{t=1}^{T} \left\{ \omega \left[ \frac{1}{a_0} \varepsilon_t Z_{1t}^{(1)} (d_0) \beta \right] \Omega^{-1} \alpha \left[ \frac{c_0}{a_0} \beta' Z_{1t} (d_0) - \beta' Z_{1t}^{(1)} (d_0) \right] \right\} \Sigma_{\beta \beta}^{-1} (d_0)
\]
\[
= \left(1 + o_p\left(1\right)\right) \frac{1}{\omega^{1/2}} \sum_{t=1}^{T} \left\{ -\frac{1}{\omega a_0} \varepsilon_t \Omega^{-1} \alpha \left[ \frac{c_0}{a_0} \beta' Z_{1t} (d_0) - \beta' Z_{1t}^{(1)} (d_0) \right] \right\} .
\]

Taking vec's and using that vec(AXB) = (B' ⊗ A) vec(X) , tr (A'BCD') = vec (A) (D ⊗ B) vec (C) , and ignoring \( o_p\left(1\right) \) terms,
\[
T^{1/2} \text{vec}(\tilde{\alpha} - \alpha) = T^{-1/2} \sum_{t=1}^{T} \left\{ -\frac{1}{\omega a_0} \varepsilon_t \Omega^{-1} \alpha \left[ \frac{c_0}{a_0} \beta' Z_{1t} (d_0) - \beta' Z_{1t}^{(1)} (d_0) \right] \varepsilon_t \right\}
\]
\[
= T^{-1/2} \sum_{t=1}^{T} \left\{ -\frac{1}{\omega a_0} \varepsilon_t \Omega^{-1} \alpha \left[ \frac{c_0}{a_0} \beta' Z_{1t} (d_0) - \beta' Z_{1t}^{(1)} (d_0) \right] \right\} \text{vec}(\alpha)
\]

\[ T^{-\frac{1}{2}} \sum_{t=1}^{T} \frac{1}{a_0} \left( \Sigma_{\beta \beta}^{-1} \beta' Z_{1t} (d_0) \otimes I \right) \varepsilon_t \]  
\[ = T^{-\frac{1}{2}} \frac{c_0}{\omega_0} \text{vec}(\alpha) \text{vec}(\alpha)' \sum_{t=1}^{T} \left( \left[ \frac{c_0}{a_0} \beta' Z_{1t} (d_0) - \beta' Z_{1t}^{(1)} (d_0) \right] \otimes I \right) \Omega^{-1} \varepsilon_t. \quad (31) \]

Then the distribution for \( T^{\frac{1}{2}} \text{vec}(\tilde{\alpha} - \alpha) \) follows by a standard martingale difference CLT, noting that the contributions to its asymptotic variance are

\[ \text{Cov} ((30), (31)) = -\frac{c_0}{\omega_0} \text{vec}(\alpha) \text{vec}(\alpha)' \times \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E \left( \left[ \frac{c_0}{a_0} \beta' Z_{1t} (d_0) - \beta' Z_{1t}^{(1)} (d_0) \right] \otimes I \right) \left( Z_{1t}^{(1)} (d_0) \beta \Sigma_{\beta \beta}^{-1} \otimes I_n \right) \]

Finally recall (26), by Lemma 10 and the \( T^{\frac{1}{2}} \) consistency of \( \tilde{\alpha} \) for \( d_0 > \frac{1}{2} \), the last term of (26) converges in probability to zero and the consistency of \( \tilde{\Omega} \) then implies that

\[ T^{d_0} U_T = \left[ \varepsilon' T^{1-2d_0} S_{11} (d_0) \varepsilon \right]^{-1} \varepsilon' T^{1-d_0} S_{11}' (d_0) \Omega^{-1} \alpha (\alpha' \Omega^{-1} \alpha)^{-1} + o_P (1), \]

which converges in \( d \) towards the limit given in the theorem by Lemma 10.

\section{Appendix C}

\textbf{Proof.} Assuming \( r = 1 \), the first derivatives of the log-likelihood \( L(\alpha, \beta, \Omega, d) \) are

\[ L_\beta (\theta) = \frac{\partial}{\partial \beta} L (\alpha, \beta, d) = T (S_{10} (d) - S_{11} (d) \beta \alpha') \Omega^{-1} \alpha \]

\[ L_\alpha (\theta) = \frac{\partial}{\partial \alpha} L (\alpha, \beta, d) = T \Omega^{-1} (S_{01} (d) - \alpha \beta' S_{11} (d)) \beta \]

\[ L_d (\theta) = \frac{\partial}{\partial d} L (\alpha, \beta, d) = T \text{tr} \left\{ \alpha \beta' \left( S_{10}^{(1)} (d) - S_{11}^{(1)} (d) \beta \alpha' \right) \Omega^{-1} \right\} \]
where $\theta = (\beta', \alpha', d)'$, while the second derivatives are

\[
L_{\beta\beta} (\theta) = -T \alpha' \Omega^{-1} \alpha S_{11} (d) \\
L_{\beta\alpha} (\theta) = -T \{ 2 S_{11} (d) \beta \alpha' - S_{10} (d) \} \Omega^{-1} \\
L_{\beta d} (\theta) = -T \left\{ 2 \alpha' S_{11}^{(1)} (d) \beta - S_{01} (d) \right\} \Omega^{-1} \\
L_{\alpha\alpha} (\theta) = -T \beta' S_{11} (d) \beta \Omega^{-1} \\
L_{dd} (\theta) = -T \text{tr} \left\{ \alpha' \beta' \left( S_{11}^{(1)} (d) \beta \alpha' + S_{11}^{(2)} (d) \beta \alpha' - S_{10} (d) \right) \Omega^{-1} \right\}.
\]

We check the conditions of Lemma 1 in Andrews and Sun (2004) for $B_T = \text{diag} \left( (\gamma T^{d_0} \beta T^{1/2}) , I_{k+1} T^{1/2} \right)$. Now we have that

\[
\begin{align*}
\gamma' T^{-d_0} L_{\beta} (\theta_0) & \rightarrow \gamma' T^{1-d_0} S_{1z} (d_0) \Omega_0^{-1} \alpha_0 \rightarrow_d \gamma' \Omega_0^{-1/2} \int_0^1 W_{d_0} W_0' \Omega_0^{-1/2} \alpha_0 \\
T^{-1/2} L_{\alpha} (\theta_0) & \rightarrow T^{1/2} \Omega^{-1} S_{11} (d_0) \beta \rightarrow_d N \left( 0, \alpha (d_0) \Sigma_{\beta\beta} \Omega^{-1} \right) \\
T^{-1/2} L_d (\theta_0) & \rightarrow T^{1/2} \text{tr} \left\{ \alpha' \beta' S_{11}^{(1)} (d_0) \Omega^{-1} \right\} \rightarrow_d N \left( 0, \frac{\pi^2}{6} \Sigma_{\beta\beta} \text{tr} \left\{ \alpha' \Omega^{-1} \right\} \right)
\end{align*}
\]

and

\[
\beta' T^{-1/2} L_{\beta} (\theta_0) \rightarrow \beta' T^{1/2} S_{1z} (d_0) \Omega_0^{-1} \alpha_0 \rightarrow_d N \left( 0, \alpha (d_0) \Sigma_{\beta\alpha} \alpha_0 \Omega_0^{-1/2} \alpha_0 \right)
\]

while

\[
\begin{align*}
T^{-2d_0} \gamma' L_{\beta} (\theta_0) & \rightarrow \gamma' \gamma \int_0^1 W_{d_0} W_0' \Omega_0^{1/2} \\
T^{-d_0} \gamma' L_{\beta\alpha} (\theta_0) & \rightarrow -T^{d_0+1/2} \gamma' S_{11} (d) \beta \alpha' \Omega^{-1} \rightarrow 0 \\
T^{-d_0} \gamma' L_{\beta d} (\theta_0) & \rightarrow -T^{d_0+1/2} \gamma' S_{11}^{(1)} (d) \beta \alpha' \Omega^{-1} \rightarrow 0 \\
T^{-1} L_{\alpha\alpha} (\theta_0) & \rightarrow -T^{1/2} \text{tr} \left\{ \alpha' \beta' S_{11}^{(1)} (d_0) \beta \alpha' + S_{1z}^{(2)} (d_0) \right\} \Omega^{-1} \rightarrow_d N \left( 0, \frac{\pi^2}{6} \text{tr} \left\{ \Sigma_{\beta\beta} \alpha' \Omega^{-1} \right\} \right)
\end{align*}
\]

and

\[
\begin{align*}
T^{-d_0} (\beta' L_{\beta} (\theta_0) & \rightarrow \beta' \Omega^{-1} \alpha \Omega^{-1} \alpha \Sigma_{\beta\beta} \\
T^{-1} (\beta' L_{\beta\alpha} (\theta_0) & \rightarrow -\beta' S_{11} (d) \beta \alpha' \Omega^{-1} \rightarrow_d -\alpha (d_0) \Sigma_{\beta\beta} \alpha' \Omega^{-1} \\
T^{-1} (\beta' L_{\beta d} (\theta_0) & \rightarrow -\beta' S_{11}^{(1)} (d) \beta \alpha' \Omega^{-1} \rightarrow_d -c (d_0) \Sigma_{\beta\beta} \Sigma_{\beta\alpha} \alpha' \Omega^{-1} \alpha.
\end{align*}
\]

Therefore $(B_T^{-1})' L_{\theta} (\theta_0) B_T^{-1}$ converges to a matrix that is positive definite with probability one.

The fourth point in Andrews and Sun’s Lemma can be checked if

\[
T^{1-2d_0} \| \gamma' \left\{ S_{11} (d) - S_{11} (d_0) \right\} \| \rightarrow \rho_0 \\
\| \beta' S_{11} (d) - \beta' S_{11} (d_0) \beta' \| \rightarrow \rho_0
\]

25
and if the second statement also holds with \( S_{11} \) replaced by \( S_{11}^{(1)}, S_{11}^{(1,1)} \) or \( S_{11}^{(2,0)} \), uniformly for \( \theta \) so that \( \| (B_T^{-1}(\theta - \theta_0) \| \leq k_T \) for some \( k_T \to \infty \) (e.g. \( k_T = \log T \)). The first statement is equivalent to show that

\[
\sup_{|d-d_0| \leq K T^{-1/2} \log T} T^{1-2d_0} \| \gamma' \{ S_{11} (d) - S_{11} (d_0) \} \gamma \| \to_p 0,
\]

which follows by pointwise convergence and tightness of \( T^{1-2d_0} \gamma' S_{11} (d) \gamma \), cf. Theorem 1 of Lasak (2005).

For the second statement, we have that by the triangle inequality

\[
\| b' S_{11} (d) b - \beta' S_{11} (d_0) \beta \| \leq \| b' S_{11} (d) b - b' S_{11} (d_0) b \|
\]

\[+ \| b' S_{11} (d_0) b - \beta' S_{11} (d_0) \beta \|.
\]

Now, for \( \| b - \beta \| \leq T^{-d_0} \log T \),

\[
T^{1-2d_0} b' \{ S_{11} (d) - S_{11} (d_0) \} b = \sum_{t=1}^{T} \{ (b \pm \beta)' \{ Z_{1t} (d) Z_{1t}^t (d) \pm Z_{1t} (d_0) Z_{1t}^t (d_0) - Z_{1t} (d_0) Z_{1t}^t (d_0) \} (b \pm \beta). \]

A typical term is then

\[
T^{1-2d_0} \sum_{t=1}^{T} (b - \beta)' Z_{1t} (d) \{ Z_{1t}^t (d) - Z_{1t}^t (d_0) \} \beta
\]

\[= T^{1-2d_0} \sum_{t=1}^{T} (b - \beta)' Z_{1t} (d) \{ \Delta^{-d} - \Delta^{-d_0} \} \Delta X_t' \beta
\]

\[= T^{1-2d_0} \sum_{t=1}^{T} (b - \beta)' \{ \Delta^{-d} - 1 \} \Delta X_t \{ \Delta^{-d} - \Delta^{-d_0} \} \Delta^{d_0} X_t' \beta
\]

\[= T^{1-2d_0} \sum_{t=1}^{T} (b - \beta)' \sum_{j=1}^{ \log j \to \infty} \phi_j \varepsilon_t \{ \Delta^{d_0-d} - 1 \} \varepsilon_t' \beta
\]

(32)

where \( \phi_j \sim j^{d-1} \) as \( j \to \infty \), whereas the weights of the filter \( \Delta^{d_0-d} - 1 \) can be bounded by \( |d_0 - d| j^{-1} \log j \) for \( |d_0 - d| \leq T^{-1/2} \log T \). Then

\[
\sum_{t=1}^{T} \sum_{j=1}^{ \log j \to \infty} \phi_j \varepsilon_t \{ \Delta^{d_0-d} - 1 \} \varepsilon_t'
\]

is \( O_p (T) \) using the same techniques as in the proof of Lemma 8, and therefore (32) is \( o_p (1) \) uniformly in \( d \) and \( b \). ■

References


[26] Lasak, K. (2005), Likelihood based testing for fractional cointegration, mimeo.


