Likelihood based testing for fractional cointegration

Katarzyna Lasak*

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

August 2005

Abstract

We generalize two types of likelihood ratio tests, so-called maximum eigenvalue test and trace test, which are basic inference tools in cointegration analysis, to the fractional cointegration case. The standard cointegration analysis only considers the assumption that deviations from equilibrium can be integrated of order zero, which is very restrictive in many cases and may imply an important loss of power for fractional cointegration testing. We consider the case where equilibrium deviations can be mean reverting with order of integration possibly greater than zero. However, the order of fractional cointegration is not assumed to be known, and the asymptotic distribution of both tests is found when considering an interval of possible values. This method also provides ML estimates of the memory errors when cointegration is assumed. The power of the proposed tests under fractional alternatives is compared with Johansen’s procedure. Size accuracy provided by the asymptotic distribution in finite samples is investigated. We illustrate our method testing the cointegration between consumption and income.

Keywords: Error correction model, Gaussian VAR model, Maximum likelihood estimation, Fractional cointegration, Likelihood ratio tests, fractional Brownian motion.

JEL: C12, C15, C32.

*I am very grateful to my advisor Carlos Velasco. Financial support from Departament d’Economia i d’Historia Economica UAB is acknowledged.
1 Introduction

Cointegration is commonly thought of a stationary relation between nonstationary variables. It has become a standard tool in econometrics since the seminal paper of Granger (1981). Following the initial suggestion of Engle and Granger (1987), when the series of interest are \( I(1) \), testing for cointegration in a single-equation framework is usually conducted by means of residual based tests (cf. Phillips and Ouliaris (1990)). Residual-based tests rely on initial regressions among the levels of the relevant time series. They are designed to test the null of no cointegration by testing whether there is a unit root in the residuals against the alternative that the regression errors are \( I(0) \).

Alternatively fully parametric inference on \( I(1)/I(0) \) cointegrated systems in the framework of Error Correction Mechanism (ECM) representation has been developed by Johansen (1988, 1991, 1995). He suggests a maximum likelihood procedure based on reduced rank regressions. His methodology consists in identifying the number of cointegration vectors within the VAR by means of performing a sequence of likelihood ratio tests. If the variables are cointegrated, after selecting the rank, cointegration vectors and the speed of adjustment coefficients (short-run dynamics) are estimated. Johansen procedure can be preferred to the residual-based approach because it provides a simple way of testing for the cointegration rank and of making inference on the parameters of complex cointegrated systems.

However the assumption that deviations from equilibrium are integrated of order zero is far too restrictive. In a general set up, where errors with fractional degree of integration are allowed, it is possible to permit mean reverting errors. The case of fractionally cointegrated processes has the same economic implications, i.e. exist long-run equilibrium among variables, as when the processes are integer-valued cointegrated, except for the slower rate of convergence to the equilibrium in the former situation. Since a standard setup of \( I(1)/I(0) \) cointegrated systems ignore the fractional cointegration parameter, a fractionally integrated equilibrium error will result in a misspecified likelihood function.

There is a growing literature on fractional cointegration. A first group of contributions deal with estimation of the cointegrating coefficients in regression models, e.g. Marinucci (2000) and Marinucci and Robinson (2001), and with inference on the properties of a possibly cointegrated system based on regression residuals. Davidson (2002) models cointegration in fractionally integrated processes and considers methods for testing the existence of cointegrating rela-

Other authors have considered (Gaussian) maximum likelihood (ML) techniques. Dueker and Startz (1998) illustrate a cointegration testing methodology based on joint estimates of the fractional orders of integration of a cointegrating vector and its parent series. Andersson and Gredenhoff (1998) showed that likelihood ratio test proposed by Johansen to test for cointegrating rank has power against fractional alternatives, so using standard likelihood based approach we are likely to find the evidence of $C(1,1)$ cointegration when in reality we have fractional cointegration. At the same time standard ML approach on fractional cointegrated systems gives severe bias and large mean square errors for the "impact" matrix $\Pi$. So it is much more severe to ignore fractional cointegration than incorporating it when it is not present. Breitung and Hassler (2002) propose a variant of efficient score tests, which allows to determine the cointegration rank of possibly fractionally integrated series where the error correction terms may be fractionally integrated as well. Semiparametric methods have also been used to design tests on the cointegration rank in fractionally integrated systems, e.g. Robinson and Yajima (2002), Chen and Hurvich (2004).

Lyhagen (1998), on the basis of a fractional ECM, allowed for fractionally integrated errors and has found the asymptotic distribution of the trace test when the fractional order of cointegration is assumed to be known. He also simulated bias and mean square error of the estimators of cointegrating vector and adjustment coefficients vector when the cointegration rank is assumed to be one.

However the assumption that the order of cointegration is known is very restrictive and may have unexpected effects on the power of the test in case of misspecification, so we reconsider in this paper the method proposed by Lyhagen without that restriction. We examine the distribution of the trace test and maximum eigenvalue test when the order of cointegration is not known and is estimated by maximum likelihood under the appropriate hypothesis on the cointegration rank. The standard cointegration case is also included in our setup. We find that our generalizations of Johansen’s tests have more power
than the standard tests when the true cointegration is fractional, while in case of $C(1,1)$ both procedures have essentially the same power.

The rest of the paper is organized as follows. Section 2 describes fractional cointegration framework. Section 3 presents model considered in the paper and testing procedure for fractional cointegration. The asymptotic distribution of the trace and maximum eigenvalue tests are presented in section 4. In Section 5 we extend the model allowing for short run correlation. Section 6 presents Monte Carlo results. We give the relevant quantiles of test statistics’ asymptotic distributions. The power of the tests and the size accuracy in small samples are investigated. Section 7 illustrates our method by testing the cointegration between consumption and income. Section 8 concludes. Appendix contains the proof of Theorem 3.

2 Framework description

Let us define the fractionally integrated process $I(d)$, following Marinucci and Robinson (2001):

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

$$a_t = \Delta^{-d} \eta_t 1_{(t > 0)}, \quad t \in \mathbb{Z}, \quad d > 0, \quad (1)$$

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

$$(1 - z)^d = \frac{1}{\Gamma(-d)} \sum_{j=0}^{\infty} \frac{\Gamma(j - d) z^j}{\Gamma(j + 1)}, \quad (2)$$

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

Let us recall definition of Granger (1986):

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


In the standard cointegration setup \( d = b = 1 \) and Johansen’s procedure applies. If \( b < 1 \) we have fractional cointegration case, which calls for generalization of Johansen’s framework.

The fractional cointegration setup that we consider in this paper is based on an extension of the Johansen’s Error Correction Mechanism (ECM) framework. Johansen (1995) considers the following Vector Error Correction Model (VECM):

\[
\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t,
\]

where \( X_t \) is a vector of \( I(1) \) series of order \( k \times 1 \), \( D_t \) are deterministic terms, \( \varepsilon_t \) is a \( k \times 1 \) vector of Gaussian error with variance-covariance matrix \( \Omega \) and \( \Pi, \Gamma_1, \ldots, \Gamma_{k-1}, \Phi \) are freely varying parameters. When \( X_t \) is cointegrated we have the reduced rank condition \( \Pi = \alpha \beta' \), where the constant matrices \( \alpha \) and \( \beta \) are \( N \times r \), having rank \( r \), representing the error correction and cointegrating coefficients, respectively. So in case of cointegration we can use reduced form of the model, i.e.:

\[
\Delta X_t = \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t,
\]

what is asserted by Granger’s representation theorem.

A first generalization of VECM to the fractional case was suggested by Granger (1986). Adopted to a notation similar to Johansen’s by Davidson (2002) it has the following form:

\[
\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t,
\]

where \( X_t \) and \( \varepsilon_t \) are \( N \times 1 \), \( \varepsilon_t \sim IID \ (0, \Sigma) \), \( B(L) \) and \( D(L) \) are finite-order matrix polynomials in the lag operator with all roots outside the unit circle. Setting \( d = b = 1 \) yields the usual Johansen style VECM, but \( d \) and \( b \) can be real values with \( d > 0 \) and \( 0 < b < d \). All the elements of \( X_t \) exhibit the same order of integration and the cointegrating residuals are all integrated of order \( d - b \).

Dittmann (2004) provided the proof of existence of a fractional ECM given in slightly different form pointing out a necessary assumption, which seems to have been neglected in the literature. He finds that Granger representation theorem only holds if \( X_t \) is purely fractionally cointegrated, i.e. all the cointegrating
vectors share the same memory. If $X_t$ is multiply fractionally cointegrated, i.e. $X_t \sim FCI(d,r)$ and simultaneously $X_t \sim FCI(d_2,r_2)$ with $d_2 \neq d$, then the error correction model

$$A(L)\Delta X_t = -[1 - \Delta^{1-d}]\Delta^d \Gamma X_t + d(L)\varepsilon_t,$$

where $A$ is cointegrating matrix, $A(L)$ and $d(L)$ are linear filters and $\Gamma$ is a matrix of rank $r$, cannot be a representation of the system $X_t$.

Independently Lyhagen (1998) presents the version of the Granger Representation Theorem presented in Banerjee at al.(1993) but adjusted for the fractionally cointegrated model of the form:

$$A^1(L)\Delta X_t = \alpha D(L)\beta' X_t + \varepsilon_t,$$

where

$$D(L) = \begin{bmatrix} \Delta^{1-d_1} - \Delta & 0 \\ 0 & \Delta^{1-d_2} - \Delta \\ & & \ddots \end{bmatrix}$$

Lyhagen’s version of fractional Vector Error Correction Model (FVECM) does not require the assumption that $X_t$ is purely fractionally cointegrated. The only necessary assumptions are that $X_t$ is not stationary but $X_t$ is stationary after differencing and $0 < d \leq 1$. However Lyhagen states that it is possible to estimate the fractionally cointegrating model by using standard reduced rank regression suggested by Johansen (1988) only when all cointegrating relations share the same $d$.

3 Model and testing for cointegration

As a first natural research step we decided to reconsider the simple version (with one $d$) of fractional cointegrated model used by Lyhagen (1998):

$$\Delta X_t = \alpha \beta' X_t^d + \varepsilon_t$$

(3)

where $X_t$ is a random vector of $I(1)$ processes of order $k \times 1$, $\varepsilon_t$ is a $k \times 1$ vector of $IID$ Gaussian errors with nonsingular covariance matrix $\Omega$, $d$ is fractional differencing operator (order of the fractional cointegration) defined in (2),

$$X_t^d \overset{def}{=} (\Delta^{1-d} - \Delta) X_t 1_{(t>0)},$$
\( \alpha \) and \( \beta \) are \( k \times r \) matrices of adjustment coefficients and cointegration vectors respectively, and \( r \) is a cointegration rank. Note that the assumption of Gaussianity will be used only to derive the form of the test statistics for different hypothesis, but not to derive the asymptotic properties of the tests.

Johansen’s procedure adapted for our model is the following. Let’s define \( Z_{0t} = \Delta X_t \), \( Z_{1t}(d) = X_t^d \). The model expressed in these variables becomes:

\[
Z_{0t} = \alpha \beta' Z_{1t}(d) + \varepsilon_t, \quad t = 1, \ldots, T.
\]

The log-likelihood function apart from a constant is given by:

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

Define as well:

\[
S_{ij}(d) = T^{-1} \sum_{t=1}^{T} Z_{0t}(d) Z_{j1}(d)' \quad i, j = 0, 1
\]

and note that \( S_{ij} \) do not depend on \( d \) when \( i = j = 0 \). For fixed \( \beta \), parameters \( \alpha \) and \( \Omega \) are estimated by regressing \( Z_{0t} \) on \( \beta' Z_{1t}(d) \). Plugging the estimates into the likelihood we get:

\[
L_{\max}^{-2/T}(\alpha(\beta), \beta, \Omega(\beta), d) = L_{\max}^{-2/T}(\beta, d) = |S_{00} - S_{10}(d) \beta (\beta' S_{11}(d) \beta)^{-1} \beta' S_{10}(d)|,
\]

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

\[
|\lambda(d) S_{11}(d) - S_{10}(d) S_{00}^{-1} S_{01}(d)| = 0 \quad (4)
\]

for eigenvalues \( \lambda_i(d) \) and eigenvectors \( v_i(d) \), 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 $\beta$:

$$L_{\text{max}}^{-2/T}(d) = |S_{00}| \prod_{i=1}^{r} (1 - \lambda_i(d)).$$

Lyhagen proposed to test the cointegration rank when $d_0$ is assumed to be known using a trace statistic, which tests the null hypothesis of rank $r$ versus the alternative of a full rank $k$, i.e. he proposed to perform a sequence of tests:

$H_0 : \text{rank} = r$

$H_1 : \text{rank} = k$

$r = 0, 1, 2, \ldots, k - 1$

by means of trace statistics defined as:

$$\text{trace}(d_0) = -2 \ln [LR(r|k)] = -T \sum_{i=r+1}^{k} \log[1 - \lambda_i(d_0)],$$

He derived its asymptotic distribution under the assumption that fractional cointegration order $d_0$ is known and tabulated the critical values for his test for several values of $d_0$. However his procedure is not useful in practice because generally we do not have information on $d$ and in case of no information it is not clear which would be a reasonable choice of $d$ when we consider alternatives with $d \leq 1$.

Since the assumption that $d$ is known is very strong and impossible to be satisfied in practice we reconsider Lyhagen’s proposal including estimation of $d$ by maximum likelihood under each hypothesis, i.e.:

$$\hat{d} = \arg \max_{d \in D} L(d),$$

(5)

where $L$ is the concentrated likelihood function defined above.

Since the fact of estimating $d$ complicates the issue of deriving the distributions of tests statistics we concentrate on testing for existence of cointegration only, and we do it independently using two tests: trace and maximum eigenvalue tests. By means of a generalized version of trace we test:

$H_0 : \text{rank} = r_0 = 0$
$H_1: \text{rank} = k$

using the $LR$ statistic defined by:

\[
\text{trace}(\hat{d}_k) = -2 \ln [LR(0|k)] = -T \sum_{i=1}^{k} \log[1 - \lambda_i(\hat{d}_k)], \quad (6)
\]

where:

\[
\hat{d}_k = \arg \max_{d \in \mathcal{D}} L_k(d)
\]

and $L_k$ denotes likelihood under the hypothesis of rank $k$.

We use also generalized maximum eigenvalue ($\lambda_{\text{max}}$) statistic, which tests cointegrating rank 0 against rank 1, i.e.:

$H_0: \text{rank} = r_0 = 0$

$H_1: \text{rank} = 1$

where $\lambda_{\text{max}}$ statistic is defined by:

\[
\lambda_{\text{max}}(\hat{d}_1) = -2 \ln [LR(0|1)] = -T \ln[1 - \lambda_1(\hat{d}_1)] \quad (7)
\]

and:

\[
\hat{d}_1 = \arg \max_{d \in \mathcal{D}} L_1(d),
\]

where $L_1$ denotes likelihood under the hypothesis of rank 1.

4 Asymptotic distribution

In this section we derive asymptotic distribution of likelihood ratio tests that we proposed in (6) and (7).

First recall the fact shown by Johansen (1995):

$S_{00} \overset{P}{\rightarrow} \Omega.$

Further state assumptions about innovations necessary to derive asymptotic distribution of our likelihood ratio tests.

Assumption 1. The process $X_t, \; t = 0, \pm 1, \ldots,$ has representation $X_t = A(L)\varepsilon_t$, where $\varepsilon_t$ are independent and identically distributed vectors with mean zero, positive definite covariance matrix $\Omega$, and $E[||\varepsilon_t||^q] < \infty$, $q \geq 4$, $q > 2/(2d - 1)$, $d = \min D > \frac{1}{2}$
Using the methods of Marinucci and Robinson (2000) we obtain that under Assumption 1:

\[ T^{0.5-d} Z_{1[T,T]} \overset{\text{a}}{\to} W_d(\tau), \quad \text{for } d > 0.5, \]

where \( \overset{\text{a}}{\to} \) means convergence in the Skorohod \( J_1 \) topology of \( D[0,1] \). \( W_d \) is a fractional Brownian motion called by Marinucci and Robinson (1998) "Type II" fractional Brownian motion and defined as:

\[ W_d(\tau) = \int_0^\tau \frac{(\tau - s)^{d-1}}{\Gamma(d)} dW(s), \]

\( W(s) \) is vector Brownian motion with covariance matrix \( \Omega \).

Then by the Central Mapping Theorem we have following convergence for each \( d > 0.5 \):

\[ T^{1-2d} S_{11}(d) \xrightarrow{d} \int_0^1 W_d(\tau)W_d(\tau)' d\tau \]

and as in e.g. Robinson and Hualde (2003, Proposition 3):

\[ T^{1-d} S_{10}(d) \xrightarrow{d} \int_0^1 W_d(\tau)dW', \]

where \( \xrightarrow{d} \) denotes convergence in distribution.

\( S_{11}(d) \) is \( O_p(T^{2d-1}) \), \( S_{10}(d) \) is \( O_p(T^{d-1}) \) and \( S_{00} \) is \( O_p(1) \), so roots \( \lambda_i(d) \) of equation (4) converge to zero like \( T^{-1} \) under the null of no cointegration. This implies that:

\[ -T \sum_{i=1}^{p} \log[1 - \lambda_i(d)] = T \sum_{i=1}^{p} \lambda_i(d) + o_p(1). \]

The sum of the eigenvalues can be found as follows:

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

if and only if

\[ |\lambda(d)I - S_{11}^{-1}(d)S_{10}(d)S_{00}^{-1}S_{01}(d)| = 0, \]

which shows that:

\[ T \sum_{i=1}^{p} \lambda_i(d) = T \text{ tr}\{S_{11}^{-1}(d)S_{10}(d)S_{00}^{-1}S_{01}(d)\}. \]
From the above reasoning we find that for each $d$ the product

$$S_{11}^{-1}(d)S_{10}(d)S_{00}^{-1}S_{01}(d)$$

goes in distribution towards

$$ \left( \int_0^1 W_d(\tau) W_d(\tau)' \, d\tau \right)^{-1} \int_0^1 W_d(\tau) \, dW' \Omega^{-1} \int_0^1 (dW) W_d(\tau)' ,$$

which we can write as

$$\int_0^1 (dB) B_d(\tau)' \left[ \int_0^1 B_d(\tau) B_d(\tau)' \, d\tau \right]^{-1} \int_0^1 B_d(\tau) \, dB', \tag{8}$$

defining the standard fractional Brownian motion $B_d(\tau) = \Omega^{-1/2} W_d(\tau)$. Then we can see that asymptotic distribution of trace and maximum eigenvalue for a fixed $d$ are respectively the trace and the greatest eigenvalue of (8), i.e.:

$$\text{trace}(d) \overset{d}{\rightarrow} \text{trace} \left[ \int_0^1 (dB) B_d(\tau)' \left[ \int_0^1 B_d(\tau) B_d(\tau)' \, d\tau \right]^{-1} \int_0^1 B_d(\tau) \, dB' \right]$$

$$\lambda \text{max}(d) \overset{d}{\rightarrow} \lambda_1 \left[ \int_0^1 (dB) B_d(\tau)' \left[ \int_0^1 B_d(\tau) B_d(\tau)' \, d\tau \right]^{-1} \int_0^1 B_d(\tau) \, dB' \right]$$

In the case when $d$ is estimated the following theorem applies:

**Theorem 3** When $d$ is estimated by the maximum likelihood principle the asymptotic distributions of trace and maximum eigenvalue statistics are given respectively by:

$$\text{trace}(\hat{d}_k) \Rightarrow \sup_{d \in \mathcal{D}} \text{trace} \left[ \mathcal{L}(d) \right],$$

and

$$\lambda \text{max}(\hat{d}_1) \Rightarrow \sup_{d \in \mathcal{D}} \lambda_1 \left[ \mathcal{L}(d) \right],$$

where $\mathcal{D} \subset (0.5, 1]$ is a compact set, and

$$\mathcal{L}(d) = \int_0^1 (dB) B_d(\tau)' \left[ \int_0^1 B_d(\tau) B_d(\tau)' \, d\tau \right]^{-1} \int_0^1 B_d(\tau) \, dB'.$$
The proof is given in the appendix. Note that convergences in Theorem 3 are based on the weak convergence in $C(D)$ of the respective trace($d$) and $\lambda \max(d)$ as stochastic processes indexed by $d \in D$. Observe that we consider $d$ that belongs to a set $D \subset (0.5, 1]$, because only for these values we have non-degenerate asymptotic distribution of our test statistics.

Let us note that if the null hypothesis is not true and we have fractional cointegration, then one of the eigenvalues in (4) will be positive in the limit. Then:

$$-2 \ln [LR(0|k)] \geq -T \log \left(1 - \lambda_1(\hat{d}_k)\right) \overset{p}{\to} \infty$$

where $\hat{d}_k = \arg \max_{d \in D} L_k(d)$, and

$$-2 \ln [LR(0|1)] = -T \log \left(1 - \lambda_1(\hat{d}_1)\right) \overset{p}{\to} \infty,$$

where $\hat{d}_1 = \arg \max_{d \in D} L_1(d)$. So the asymptotic power of both tests is 1.

Note that $\hat{d}_1$ and $\hat{d}_k$ do not estimate consistently the true value of any parameter under the null hypothesis, because in this case $\beta = 0$ in (3) and the parameter $d$ is no longer identified under the null. Therefore, our tests can be interpreted as sup $LR$ tests, in the spirit of Davies (1977) and Hansen (1996).

5 Extended model

Note that we may consider extended version of the model (3), which allows for short run correlation among variables

$$\Delta X_t = \alpha \beta' X_t^d + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i-1} + \varepsilon_t. \quad (9)$$

Then to the procedure that leads to define likelihood ratio tests we have to add the preliminary step of regressing $\Delta X_t$ and $X_t^d$ on the lagged differences $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}$ and later consider the regressions residuals $R_{0t}$ and $R_{1t}$ instead of $Z_{0t}$ and $Z_{1t}$ respectively. The remaining steps of the procedure do not change.

In the case of this extended model asymptotic results for our test statistics do not change and Theorem 3 holds since for each $d > 0.5$ the main regressor $X_t^d$ in the equation (9) dominates the stationary lags $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}$ as in the standard $C(1, 1)$ case.
6 Monte Carlo

All Monte Carlo simulations were done using Ox 3.40 (see Doornik, 2001). The asymptotic distribution of the trace and maximum eigenvalue statistics were simulated using the approximation of fractional Brownian motion by fractionally integrated series based on IID Gaussian noise of length 1000. To maximize the likelihood function we used MaxSQPF procedure and optimization was done on the interval $d \in < 0.5000001; 1 >$. Quantiles of the simulated (with 100,000 repetitions) asymptotic distribution are given in Tables 1 and 2.

Table 1: Quantiles of trace statistic

<table>
<thead>
<tr>
<th></th>
<th>1%</th>
<th>2.5%</th>
<th>5%</th>
<th>10%</th>
<th>50%</th>
<th>90%</th>
<th>95%</th>
<th>97.5%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0005</td>
<td>0.0032</td>
<td>0.012</td>
<td>0.044</td>
<td>0.859</td>
<td>3.709</td>
<td>4.952</td>
<td>6.232</td>
<td>7.921</td>
</tr>
<tr>
<td>2</td>
<td>0.419</td>
<td>0.697</td>
<td>1.062</td>
<td>1.636</td>
<td>5.377</td>
<td>10.925</td>
<td>12.801</td>
<td>14.611</td>
<td>16.822</td>
</tr>
<tr>
<td>4</td>
<td>7.061</td>
<td>8.678</td>
<td>10.266</td>
<td>12.497</td>
<td>25.756</td>
<td>36.714</td>
<td>39.938</td>
<td>42.812</td>
<td>46.426</td>
</tr>
<tr>
<td>5</td>
<td>14.012</td>
<td>16.294</td>
<td>18.730</td>
<td>22.626</td>
<td>42.627</td>
<td>55.843</td>
<td>59.628</td>
<td>62.981</td>
<td>67.208</td>
</tr>
<tr>
<td>6</td>
<td>23.445</td>
<td>26.637</td>
<td>30.020</td>
<td>35.065</td>
<td>63.715</td>
<td>78.894</td>
<td>83.275</td>
<td>87.304</td>
<td>92.399</td>
</tr>
<tr>
<td>7</td>
<td>35.540</td>
<td>39.896</td>
<td>44.536</td>
<td>52.388</td>
<td>88.894</td>
<td>106.074</td>
<td>111.176</td>
<td>115.647</td>
<td>121.006</td>
</tr>
<tr>
<td>8</td>
<td>50.621</td>
<td>56.344</td>
<td>62.860</td>
<td>77.073</td>
<td>117.845</td>
<td>136.817</td>
<td>142.547</td>
<td>147.541</td>
<td>153.517</td>
</tr>
<tr>
<td>9</td>
<td>69.339</td>
<td>77.132</td>
<td>87.639</td>
<td>125.215</td>
<td>150.670</td>
<td>171.627</td>
<td>177.831</td>
<td>183.546</td>
<td>190.306</td>
</tr>
<tr>
<td>10</td>
<td>91.111</td>
<td>102.750</td>
<td>130.266</td>
<td>163.232</td>
<td>187.296</td>
<td>210.118</td>
<td>217.030</td>
<td>223.242</td>
<td>230.268</td>
</tr>
</tbody>
</table>

Table 2: Quantiles of lambda_{max}

<table>
<thead>
<tr>
<th></th>
<th>1%</th>
<th>2.5%</th>
<th>5%</th>
<th>10%</th>
<th>50%</th>
<th>90%</th>
<th>95%</th>
<th>97.5%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0005</td>
<td>0.0032</td>
<td>0.012</td>
<td>0.044</td>
<td>0.859</td>
<td>3.709</td>
<td>4.952</td>
<td>6.232</td>
<td>7.921</td>
</tr>
<tr>
<td>2</td>
<td>0.364</td>
<td>0.606</td>
<td>0.924</td>
<td>1.429</td>
<td>4.742</td>
<td>9.896</td>
<td>11.707</td>
<td>13.420</td>
<td>15.600</td>
</tr>
<tr>
<td>5</td>
<td>6.916</td>
<td>8.155</td>
<td>9.415</td>
<td>11.219</td>
<td>19.415</td>
<td>27.656</td>
<td>30.251</td>
<td>32.630</td>
<td>35.491</td>
</tr>
<tr>
<td>6</td>
<td>9.990</td>
<td>11.451</td>
<td>12.963</td>
<td>15.080</td>
<td>24.587</td>
<td>33.566</td>
<td>36.357</td>
<td>38.922</td>
<td>42.078</td>
</tr>
<tr>
<td>9</td>
<td>20.353</td>
<td>22.477</td>
<td>24.579</td>
<td>27.727</td>
<td>40.746</td>
<td>51.164</td>
<td>54.322</td>
<td>57.234</td>
<td>60.930</td>
</tr>
<tr>
<td>10</td>
<td>24.082</td>
<td>26.307</td>
<td>28.627</td>
<td>32.269</td>
<td>46.179</td>
<td>56.919</td>
<td>60.234</td>
<td>63.124</td>
<td>66.800</td>
</tr>
</tbody>
</table>
To evaluate the finite sample properties of our testing for cointegration procedure we have generated the following two equation model (see Engle, Granger (1987), Banerjee et al. (1993), p.137 or Lyhagen (1998)):

\[ x_t + \beta y_t = u_t \]

\[ x_t + \alpha y_t = e_t \]

where \( \Delta^{1-d} u_t = \varepsilon_{1t}, \Delta e_t = \varepsilon_{2t}, \varepsilon_{1t} \) and \( \varepsilon_{2t} \) are both independently and standard bivariate normally distributed with expectation zero, \( d \) is order of fractional cointegration, \( d \in (0.5000001, 1] \), \([1, \beta]'\) is a cointegrating vector, \([1 \alpha]'^\prime\) is a vector of adjustment coefficients. In the simulations reported below the cointegration vector was standardized to a form of \([1/\beta, 1]'\). In all simulations we used the same parameters \( \alpha \) and \( \beta \) equal to \(-1\) and \(2\) respectively. Small samples properties (for \( k = 2 \)) of the proposed tests were investigated by simulation with 10,000 repetitions and compared with Johansen’s tests based on the model \( \Delta X_t = \Pi X_{t-1} + \varepsilon_t \). To check the power and size of Johansen’s tests we used the critical values given by MacKinnon et al. (1998).

Size accuracy of all considered tests is presented in Table 3. Power comparison is presented in Tables 4-7 and power curve is drawn on the Figure 1.

Table 3: Small sample size accuracy (in %) for \( k=2 \). Nominal size 5%.

| \( T \) | \( \text{lambda} \) Johansen’s \( \text{lambda} \) trace JOHANSENS’ \( \text{trace} \) JOHANSENS’ \( \text{trace} \) |
|---|---|---|---|---|
| 50 | 4.74 | 5.06 | 4.60 | 4.94 |
| 100 | 4.98 | 5.33 | 4.82 | 5.18 |
| 200 | 4.71 | 4.96 | 4.91 | 5.18 |
| 500 | 5.01 | 5.17 | 5.14 | 5.09 |

The Monte Carlo simulation shows that size distortions of our tests are comparable to size distortions of Johansen’s tests.

Table 4: Power of the tests for samples of \( T=50 \) observations

| \( d \) | \( \text{lambda} \) JOHANSENS’ \( \text{lambda} \) trace JOHANSENS’ \( \text{trace} \) JOHANSENS’ \( \text{trace} \) |
|---|---|---|---|---|
| 0.1 | 6.20 | 6.42 | 6.20 | 6.42 |
| 0.2 | 13.12 | 12.66 | 13.12 | 12.54 |
| 0.3 | 27.89 | 24.29 | 27.97 | 24.01 |
| 0.4 | 52.60 | 44.65 | 51.60 | 44.61 |
| 0.5 | 77.98 | 69.80 | 76.69 | 69.46 |
| 0.6 | 93.78 | 89.08 | 92.87 | 88.52 |
| 0.7 | 99.06 | 98.14 | 98.86 | 98.01 |
| 0.8 | 99.89 | 99.89 | 99.88 | 99.83 |
| 0.9 | 100 | 100 | 100 | 100 |
Table 5: Power of the tests for a sample of $T=100$ observations

<table>
<thead>
<tr>
<th>$d$</th>
<th>lambda</th>
<th>Johansen's lambda trace</th>
<th>Johansen's trace</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>9.00</td>
<td>8.47</td>
<td>8.95</td>
</tr>
<tr>
<td>0.2</td>
<td>26.89</td>
<td>20.93</td>
<td>26.60</td>
</tr>
<tr>
<td>0.3</td>
<td>63.27</td>
<td>45.55</td>
<td>61.36</td>
</tr>
<tr>
<td>0.4</td>
<td>93.66</td>
<td>75.82</td>
<td>92.88</td>
</tr>
<tr>
<td>0.5</td>
<td>99.79</td>
<td>95.14</td>
<td>99.72</td>
</tr>
<tr>
<td>0.6</td>
<td>100</td>
<td>99.65</td>
<td>100</td>
</tr>
<tr>
<td>0.7</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 6: Power of the tests for a sample of $T=200$ observations

<table>
<thead>
<tr>
<th>$d$</th>
<th>lambda</th>
<th>Johansen's lambda trace</th>
<th>Johansen's trace</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>14.47</td>
<td>10.97</td>
<td>14.14</td>
</tr>
<tr>
<td>0.2</td>
<td>57.10</td>
<td>32.88</td>
<td>55.57</td>
</tr>
<tr>
<td>0.3</td>
<td>96.91</td>
<td>68.92</td>
<td>96.36</td>
</tr>
<tr>
<td>0.4</td>
<td>100</td>
<td>93.34</td>
<td>100</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>99.75</td>
<td>100</td>
</tr>
<tr>
<td>0.6</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 7: Power of the tests for a sample of $T=500$ observations

<table>
<thead>
<tr>
<th>$d$</th>
<th>lambda</th>
<th>Johansen's lambda trace</th>
<th>Johansen's trace</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>32.47</td>
<td>15.57</td>
<td>31.75</td>
</tr>
<tr>
<td>0.2</td>
<td>97.60</td>
<td>52.16</td>
<td>97.25</td>
</tr>
<tr>
<td>0.3</td>
<td>100</td>
<td>88.67</td>
<td>100</td>
</tr>
<tr>
<td>0.4</td>
<td>100</td>
<td>99.52</td>
<td>100</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 1: The power curve for a sample of $T=50$ observations
A Monte Carlo comparison shows that both fractional likelihood ratio tests have similar power, which is much greater than power of standard Johansen’s tests. Power of all tests increases with sample size.

7 Application

In order to illustrate the properties of our testing procedure we employ the well known example of cointegration between consumption and income. Macroeconomic theory tells us that the two variables should be cointegrated and the evidence of $CI(1,1)$ cointegration has been found in Engle and Granger (1987).

In a fractional framework this example has been studied by Marinucci and Robinson (2001). They found an integration order of both variables to be very close to one (the estimates ranging from 0.89 to 1.08 for income and from 0.89 to 1.13 for consumption) and that the $I(1)/I(0)$ framework can produce a satisfactory approximation for the behavior of the raw series, but not of the cointegrating residuals. The estimates of residual memory found by Marinucci and Robinson (2001) varied quite noticeably with the procedure adopted, ranging from 0.19 to 0.87.

These facts encouraged us to use our test to reconsider the problem. To test the cointegration between consumption and income we used the monthly data for US in the period: 1959/1 – 1996/12. The results are following. Both tests find evidence of cointegration. The $d$ estimated by our procedure is equal to the lower bound of the possible values interval $[0.5000001; 1]$, which indicates that the order of cointegration might be lower than 0.5. To estimate the residual memory we enlarged the interval of allowed values of $d$ to $[0.0000001; 1]$. Then the $d$ estimate was equal to 0.27, which is in line with findings in Marinucci and Robinson (2001). So it seems that cointegration between income and consumption really exists, but it is far from being a $CI(1,1)$ cointegration.

8 Conclusions

In this paper we considered two likelihood based tests for fractional cointegration. They are more general then other tests considered previously in the literature because of two aspects: departures from equilibrium are allowed to be fractionally cointegrated and the memory of the errors is estimated. By means of Monte Carlo simulation we demonstrated that proposed tests have better
power than the standard Maximum Likelihood procedures to detect cointegration, while size distortions are comparable and small.

There are many extensions to the setup considered in this paper to be developed in the nearest future. We would like to propose testing procedure for higher ranks, including the case of multiple fractional cointegration. The following extension would be to generalize proposed model to the case when original series considered can have integration order different than 1, either known or unknown. We are planning as well to consider the issue of $d$ estimation and of other parameters and find their asymptotic properties.

References


[7] Davies, R.B. (1977), Hypothesis Testing when a Nuisance is Present Only under the Alternative, Biometrika, 64, 247-254.


[17] Hansen, B.E. (1996), Inference when a Nuisance Parameter is not Identified under the Null Hypothesis, Econometrica, 64, 413-430.


9 Appendix

Proof. First note that for each $d$ we have:

$$\text{trace}(d) \xrightarrow{d} \text{trace}\left\{ \int_0^1 (dB_d) B'_d \left[ \int_0^1 B_d B'_d du \right]^{-1} \int_0^1 B_d (dB_d)' \right\},$$

where $\xrightarrow{d}$ denotes usual standard convergence in distribution, which follows because of the joint convergence of the matrices of sample moments to the stochastic integrals. Then by the same argument we have convergence for many $d$'s.

Second recall that $\text{trace}(d)$ is a continuous function in all elements of the matrices involved and the random processes on the right hand side are continuous in $d$. Then if we check that the process is tight in $d$, we have that:

$$\text{trace}(d) \Rightarrow \text{trace}\left\{ \int_0^1 (dB_d) B'_d \left[ \int_0^1 B_d B'_d du \right]^{-1} \int_0^1 B_d (dB_d)' \right\},$$

where $\Rightarrow$ denotes weak convergence for $d \in \mathcal{D}$ in the space $C(\mathcal{D})$ of continuous functions in $\mathcal{D}$.

Third since sup function is well defined and

$$\text{trace}(d_{\text{trace}}) = \text{trace}(\arg\max_{d \in \mathcal{D}} \text{trace}(d)) = \sup_{d \in \mathcal{D}} \text{trace}(d),$$

we get by the Continuous Mapping Theorem that the asymptotic distribution of $\sup_{d \in \mathcal{D}} \text{trace}(d)$ is the distribution of the

$$\sup_{d \in \mathcal{D}} \left( \text{trace}\left\{ \int_0^1 (dB_d) B'_d \left[ \int_0^1 B_d B'_d du \right]^{-1} \int_0^1 B_d (dB_d)' \right\} \right).$$
So to prove that Theorem 3 holds it is left to check that the process $\text{trace}(d)$ is smooth in $d$. It is enough to demonstrate that the elements of the sample moments matrices $(S_{11}(d) \text{ and } S_{10}(d))$ are smooth (tight) in $d$, since $\text{trace}(d)$ is a continuous function in all elements of the matrices involved as we stated before.

Note that $S_{00}$ does not depend on $d$ and $S_{01}(d) = S'_{10}(d)$. We demonstrate the sketch of the proof for a typical element of $S_{11}$. The tightness of $S_{10}(d)$ follows by the same arguments.

Recall that in our case:

$$S_{11}(d) = T^{-1} \sum_{t=1}^{T} Z_{1t} Z'_{1t},$$

$$Z_{1t} = Z_{1t}(d) = (\Delta^{-d} - 1) \Delta X_t = (\Delta^{-d} - 1) \varepsilon_t = \sum_{j=1}^{t} \pi_j (-d) \varepsilon_{t-j}. $$

Since

$$T^{1-2d}S_{11}(d) \overset{d}{\rightarrow} \int_{0}^{1} W_d(\tau)W_d(\tau)'d\tau$$

and we can proceed componentwise, then for tightness, by Billingsley’s (1968) Theorem 15.6, it is sufficient to check that

$$E \left| T^{1-2d} S_{11}^{r,s}(d_a) - T^{1-2d} S_{11}^{r,s}(d_b) \right|^m \leq K |d_a - d_b|^\gamma,$$

for some $m > 0$, $K < \infty$ and $\gamma > 1$, where $S_{11}^{r,s}(d)$ is the $(r,s)$ element of $S_{11}(d)$, $K$ and $\gamma$ are generic constants that do not depend on $T$ nor on $(d^a, d^b)$. We will demonstrate that it holds for $m = 2$. Then

$$S_{11}^{r,s}(d) = \sum_{t=1}^{T} Z'^{r}_{1t}(d) Z'^{s}_{1t}(d)$$

$$= T^{-1} \sum_{t=1}^{T} \left( \sum_{j=1}^{t} \pi_j (-d) \varepsilon_{t-j}^{r} \right) \left( \sum_{i=1}^{t} \pi_i (-d) \varepsilon_{t-i}^{s} \right)$$

so

$$E \left| T^{1-2d} S_{11}^{r,s}(d_a) - T^{1-2d} S_{11}^{r,s}(d_b) \right|^2$$

$$= E \left| T^{-2d} \sum_{t=1}^{T} Z'^{r}_{1t}(d_a) Z'^{s}_{1t}(d_a) - T^{-2d} \sum_{t=1}^{T} Z'^{r}_{1t}(d_b) Z'^{s}_{1t}(d_b) \right|^2$$
First cross product $AC$ is equal to

\[
T^{-4d_a} \sum_{t=1}^{T} \sum_{t'=1}^{T} E \left\{ T^{-2d_a} \left( \sum_{j=1}^{t} \pi_j \left( -d_a \right) \varepsilon_{t-j}^r \right) \left( \sum_{i=1}^{t} \pi_i \left( -d_a \right) \varepsilon_{t-i}^s \right) \times \left( \sum_{j'=1}^{t'} \pi_{j'} \left( -d_a \right) \varepsilon_{t'-j'}^r \right) \left( \sum_{i'=1}^{t'} \pi_{i'} \left( -d_a \right) \varepsilon_{t'-i'}^s \right) \right\}
\]

\[
= \sigma_{rs}^2 T^{-4d_a} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{j=1}^{t} \sum_{j'=1}^{t'} \pi_j \left( -d_a \right) \pi_{j'} \left( -d_a \right) \pi_{j+j'-(t-t')} \left( -d_a \right) \pi_{j'+(t-t')} \left( -d_a \right)
\]

\[
+ \sigma_{rs}^2 \sigma_{rs}^2 T^{-4d_a} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{j=1}^{t} \sum_{j'=1}^{t'} \pi_j \left( -d_a \right) \pi_{j+j'-(t-t')} \left( -d_a \right) \pi_{j'+(t-t')} \left( -d_a \right)
\]

\[
+ \sigma_{rs}^2 \sigma_{rs}^2 T^{-4d_a} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{j=1}^{t} \sum_{j'=1}^{t'} \pi_j \left( -d_a \right) \pi_{j+j'-(t-t')} \left( -d_a \right) \pi_{j'+(t-t')} \left( -d_a \right)
\]
The cross product $\mathbf{BC}$ is the fourth cumulant of $(\varepsilon^r_t, \varepsilon^s_t, \varepsilon^r_{t'}, \varepsilon^s_{t'})$.

The cross product $BD$ can be calculated in the same way. The cross product $BC$ is equal to:

$$ T^{-2d_a-2d_b} \sum_{t=1}^T \sum_{t'=1}^T \sum_{j=1}^T \sum_{j'=1}^T \pi_j (-d_a)^2 \pi_{j'} (-d_b)^2 \times \sum_{i=1}^t \pi_i (-d_b) \varepsilon^{r}_{t-i} \right) \left( \sum_{i=1}^t \pi_{i'} (-d_a) \varepsilon^{s}_{t-i'} \right) \times \sum_{j=1}^t \pi_j (-d_b) \varepsilon^{r}_{t-j} \right) \left( \sum_{i=1}^t \pi_{i} (-d_a) \varepsilon^{s}_{t-i} \right) \times \sum_{j=1}^t \pi_j (-d_a)^2 \pi_j (-d_b)^2 $$. 

The cross product $AD$ can be calculated in the similar way. Then:

$$ E \left[ (T^{0.5-d_a})^2 S_{11}^{r,s} (d_a) - (T^{0.5-d_b})^2 S_{11}^{r,s} (d_b) \right]^2 $$

$$ = \sigma_{rs}^2 \sigma_{rs}^2 \sum_{t=1}^T \sum_{t'=1}^T \sum_{j=1}^T \sum_{j'=1}^T \left[ \pi_j^* (-d_a)^2 - \pi_{j'}^* (-d_b)^2 \right] \pi_j (-d_a)^2 \pi_{j'} (-d_b)^2 $$

$$ + \sigma_{rr} \sigma_{ss}^2 \sum_{t=1}^T \sum_{t'=1}^T \sum_{j=1}^T \sum_{j'=1}^T \left[ \pi_j^* (-d_a)^2 - \pi_{j'}^* (-d_b)^2 \right] \pi_j (-d_a)^2 \pi_{j'} (-d_b)^2 $$

$$ + \sigma_{rr}^2 \sigma_{ss}^2 \sum_{t=1}^T \sum_{t'=1}^T \sum_{j=1}^T \sum_{j'=1}^T \left\{ \pi_j^* (-d_a) \pi_{j'}^* (-d_b) \pi_j (-d_a) \pi_{j'} (-d_b) \right\} $$

$$ + \sigma_{rr}^2 \sigma_{ss}^2 \sum_{t=1}^T \sum_{t'=1}^T \sum_{j=1}^T \sum_{j'=1}^T \left\{ \pi_j^* (-d_a) \pi_{j'}^* (-d_b) \pi_j (-d_a) \pi_{j'} (-d_b) \right\} $$

23
\[
\times \left[ \pi_{j+|\nu-t|}^* (-d_a) \pi_{j+|\nu-t|}^* (-d_b) - \pi_{j+|\nu-t|}^* (-d_b) \pi_{j+|\nu-t|}^* (-d_a) \right] \\
+ \kappa_{r,s,r,s} \sum_{t=1}^T \sum_{j=1}^t \left( \pi_j^* (-d_a)^2 - \pi_j^* (-d_b)^2 \right)^2
\]

where
\[
\pi_j^* (-d) = T^{-d} \pi_j (-d).
\]

From Lemma 4 given below we get that the value of the first term in (10) is bounded by \( K |d_a - d_b|^2 \). Using similar arguments it is possible to demonstrate that the other terms in (10) can also be bounded by \( K |d_a - d_b|^2 \). This completes the proof. \( \blacksquare \)

**Lemma 4** The value
\[
K \sum_{t=1}^T \sum_{\nu'=1}^T \sum_{j=1}^{t'} \sum_{j'=1}^{t'} \left\{ \left( \pi_{j'}^* (-d_a)^2 - \pi_{j'}^* (-d_b)^2 \right) \left( \pi_j^* (-d_a)^2 - \pi_j^* (-d_b)^2 \right) \right\}
\]

is bounded by \( K |d_a - d_b|^2 \).

**Proof.** Let’s prove Lemma 4 by applying the Mean Value Theorem to \( \pi_j^* (-d) \). First, observe that
\[
\left| \frac{\Gamma' (j + d)}{\Gamma (j + 1)} - \frac{\Gamma (j + d)}{\Gamma (j + 1)} \log j \right| \leq K j^{d-1}, \quad (10)
\]

since
\[
\frac{\Gamma' (j + d)}{\Gamma (j + 1)} - \frac{\Gamma (j + d)}{\Gamma (j + 1)} \log j = \frac{\Gamma' (j + d) \Gamma (j + 1) - \Gamma (j + d) \Gamma' (j + 1)}{\Gamma (j + d) \Gamma (j + 1)} \log j
\]
\[
= \left\{ \frac{\Gamma' (j + d)}{\Gamma (j + d)} - \log j \right\} \frac{\Gamma (j + d)}{\Gamma (j + 1)}
\]
\[
= \{ \psi (j + d) - \log j \} \frac{\Gamma (j + d)}{\Gamma (j + 1)}
\]

where \( \Gamma (j + d) \Gamma (j + 1)^{-1} \sim K j^{d-1} \) for \( j \to \infty \) and \( \psi (z) \) is the digamma function, which satisfies
\[
\psi (z) = \log z + \frac{1}{2z} + O \left( z^{-2} \right), \quad z \to \infty,
\]
so
\[
\psi (j + d) = \log (j + d) + O \left( j^{-1} \right)
\]
\[
= \log j + O \left( j^{-1} \right)
\]
as \( j \to \infty \), uniformly for \( d \in \mathcal{D} \).

Now consider
\[
T^d \left| \frac{\partial}{\partial d} \pi_j^* (-d) \right| = \left| \frac{\partial}{\partial d} \pi_j(-d) - \pi_j(-d) \log T \right|
\]
\[
= \left| -\Gamma' (d) \frac{\Gamma(j + d) + \Gamma(d) \Gamma'(j + d)}{\Gamma^2(d) \Gamma(j + 1)} - \frac{\Gamma(j + d)}{\Gamma(d) \Gamma(j + 1)} \log T \right|
\]
\[
= \left| \frac{1}{\Gamma^2(d) \Gamma(j + 1)} \{ -\Gamma'(d) \frac{\Gamma(j + d) + \Gamma(d) \Gamma'(j + d)}{\Gamma(j + 1)} - \Gamma(d) \Gamma'(j + d) \log T \} \right|
\]
\[
\leq \left| \frac{\Gamma'(d) \Gamma(j + d)}{\Gamma^2(d) \Gamma(j + 1)} \right|^{K^j d^{-1}} \leq K^j d^{-1} |\log(j/T)|,
\]
uniformly for \( j = 1, \ldots, T \) and \( d \in \mathcal{D} \), using \( \pi_j(-d) \sim K^j d^{-1} \) and (10).

Finally we get by the Mean Value Theorem that for some \( d^* \in [d_a, d_b] \)
\[
K \sum_{i=1}^{T} \sum_{j'=1}^{T} \sum_{j=1}^{t} \sum_{j'=1}^{t'} \left\{ \left( \pi_j^* (-d_a)^2 - \pi_j^* (-d_b)^2 \right) \left( \pi_j^* (-d_a)^2 - \pi_j^* (-d_b)^2 \right) \right\}^2
\]
\[
\leq KT^{-4d^*} |d_a - d_b|^2 \left\{ \sum_{i=1}^{T} \sum_{j=1}^{t} j^{2d^*-2} |\log(j/T)| \right\}^2
\]
\[
\leq KT^{-4d^*} |d_a - d_b|^2 \left\{ \sum_{i=1}^{T} |\log(t/T)| t^{2d^*-1} \sum_{j=1}^{t} (j/t)^{2d^*-2} |\log(j/t)| \right\}^{2}
\]
\[
\sim K_{d^*} T^{2d^*} \sum_{t=1}^{T} t^{2d^*-1} |\log(t/T)| (t/T)^{2d^*-1}
\]
\[
\leq KT^{-4d^*} |d_a - d_b|^2 \left\{ \sum_{i=1}^{T} \frac{|\log(t/T)| (t/T)^{2d^*-1}}{log x = c.} \right\}^{2}
\]
\[
\leq KT^{-4d^*} |d_a - d_b|^2 \left\{ \sum_{t=1}^{T} T^{2d^*} \frac{|\log(t/T)| (t/T)^{2d^*-1}}{log x = c.} \right\}^{2}
\]
\[
\leq KT^{-4d^*} |d_a - d_b|^2 T^{4d^*} \leq K |d_a - d_b|^2
\]
since \( d_b > d_a > 0.5 \) (the case \( d_a = d_b \) is trivial). \( \blacksquare \)