

# DEPARTMENT OF ECONOMICS

## Working Paper

Efficient Inference in Multivariate Fractionally  
Integrated Time Series Models

Morten Ørregaard Nielsen

Working Paper No. 2002-6



ISSN 1396-2426

**UNIVERSITY OF AARHUS • DENMARK**

# **INSTITUT FOR ØKONOMI**

AFDELING FOR NATIONALØKONOMI - AARHUS UNIVERSITET - BYGNING 350  
8000 AARHUS C - ☎ 89 42 11 33 - TELEFAX 86 13 63 34

## **WORKING PAPER**

### **Efficient Inference in Multivariate Fractionally Integrated Time Series Models**

Morten Ørregaard Nielsen

Working Paper No. 2002-6

## **DEPARTMENT OF ECONOMICS**

SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350  
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

# Efficient Inference in Multivariate Fractionally Integrated Time Series Models\*

Morten Ørregaard Nielsen  
University of Aarhus

October 9, 2002

## Abstract

A computationally simple maximum likelihood procedure for multivariate fractionally integrated time series models is introduced. This allows, e.g., efficient estimation of the memory parameters of fractional models or efficient testing of the hypothesis that two or more series are integrated of the same possibly fractional order. In particular, we show the existence of a local time domain maximum likelihood estimator and its asymptotic normality, and under Gaussianity asymptotic efficiency. The likelihood-based test statistics (Wald, likelihood ratio, and Lagrange multiplier) are derived and shown to be asymptotically equivalent and chi-squared distributed under local alternatives, and under Gaussianity locally most powerful. The finite sample properties of the likelihood ratio test are evaluated by Monte Carlo experiments, which show that rejection frequencies are very close to the asymptotic local power for samples as small as  $n = 100$ .

*JEL Classification:* C32

*Keywords:* Asymptotic Local Power; Efficient Estimation; Efficient Test; Fractional Integration; Multivariate ARFIMA model; Multivariate Fractional Unit Root; Nonstationarity

---

\*This paper has benefitted from valuable comments by Joerg Breitung, Bent Jesper Christensen, Niels Haldrup, Uwe Hassler, and Katsumi Shimotsu. A first draft was completed while the author was visiting Yale University and the Cowles Foundation, their hospitality is gratefully acknowledged. Address correspondence to: Morten Ørregaard Nielsen, Department of Economics, Building 322, University of Aarhus, DK-8000 Aarhus C, Denmark; e-mail: monielsen@econ.au.dk.

# 1 Introduction

In this paper we propose a class of maximum likelihood estimators and tests in multivariate fractionally integrated time series models. The development of multivariate procedures that account for contemporaneous effects between the elements of multiple time series is important for applied econometric work, which, to a large extent, is multivariate in nature. However, previous literature has focused on univariate models, see Robinson (1991, 1994), Agiakloglou & Newbold (1994), Beran (1995), Tanaka (1999), and Nielsen (2001), among others.

With no simple procedures available to test or estimate the order of fractional integration for multiple time series, applied researchers have been forced to conduct univariate analyses of the individual elements of multiple time series. This is not only cumbersome, but ignores potentially important correlations between the series which could be used to increase power and efficiency in a multivariate setting. The present paper develops the required generalizations to multiple time series. Our generalizations of the univariate procedures in some ways resemble the work of, e.g., Phillips & Durlauf (1986) and Choi & Ahn (1999), who extend the autoregressive unit root and stationarity tests to multiple time series.

To fix ideas, suppose we observe the  $K$ -dimensional vector time series  $\{y_t, t = 1, 2, \dots, n\}$  generated by the linear model

$$y_t = \beta x_t + u_t, \quad (1)$$

where  $\{x_t\}$  is a  $q$ -vector of purely deterministic components and  $\{u_t\}$  is an unobserved  $K$ -dimensional error component. The generating mechanism of  $\{u_t\}$  is

$$(1 - L)^{d_k + \theta_k} u_{kt} = e_{kt} \mathbb{I}(t \geq 1), \quad k = 1, \dots, K, \quad (2)$$

where  $\mathbb{I}(\cdot)$  denotes the indicator function, the fractional filter  $(1 - z)^d$  is defined by the binomial expansion

$$(1 - z)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j - d)}{\Gamma(-d) \Gamma(j + 1)} z^j, \quad \Gamma(f) = \int_0^{\infty} t^{f-1} e^{-t} dt,$$

and  $\{e_t\}$  is an  $I(0)$  process. Here, a process is  $I(0)$  if it is covariance stationary and has finite and strictly positive spectral density at the origin. The process  $\{u_t\}$  defined by (2) is well defined for all  $d, \theta$  and is sometimes called a type II fractionally integrated process, see Marinucci & Robinson (1999).

Two leading cases for the deterministic terms are  $x_t = 1$  and  $x_t = (1, t)'$ , which yields the models  $y_{kt} = \beta_{k0} + u_{kt}$  and  $y_{kt} = \beta_{k0} + \beta_{k1}t + u_{kt}$ , respectively, but other terms like seasonal dummies or polynomial trends are also allowed for. As in Definition 2 of Robinson (1994) and Assumption 2 of Nielsen (2001), we require only that  $\sum_{t=1}^n \tilde{x}_t \tilde{x}_t'$  is positive definite for  $n$  sufficiently large, where  $\tilde{x}_{kt} = (1 - L)^{d_k} x_{kt}$ .

In the remainder of the paper we may assume that  $\{u_t\}$  is an observed process by the results of Robinson (1994) and Nielsen (2001) who showed that estimating  $\beta$  by maximum likelihood or an asymptotically equivalent method does not influence asymptotic inference on  $\theta$ . In particular,  $\beta$  is estimated by OLS of  $\tilde{y}_t$  on  $\tilde{x}_t$ , where  $\tilde{y}_{kt} = (1 - L)^{d_k} y_{kt}$ , i.e.

$$\hat{\beta} - \beta = \left( \sum_{t=1}^n e_t \tilde{x}_t' \right) \left( \sum_{t=1}^n \tilde{x}_t \tilde{x}_t' \right)^{-1}.$$

Then we construct the residual process  $\hat{u}_t = y_t - \hat{\beta}x_t = u_t + (\beta - \hat{\beta})x_t$ , which we may treat as if we observed  $\{u_t\}$ , see Robinson (1994) and Nielsen (2001) for the details.

The errors  $\{e_t\}$  are initially assumed to be independently and identically distributed with mean zero and positive definite covariance matrix  $\Sigma$ , *i.i.d.*(0,  $\Sigma$ ), with finite fourth moments. In section 4 we relax this assumption and let  $\{e_t\}$  follow a stationary vector autoregressive process of order  $p$ , VAR( $p$ ). Notice that positive definiteness of  $\Sigma$  rules out cointegration among  $\{u_t\}$  and thus also among  $\{y_t\}$ .

We assume that the  $d_k$ 's are prespecified and wish to test the hypothesis

$$H_0 : \theta = 0 \tag{3}$$

against the alternatives  $H_1 : \theta \neq 0$  or  $H_2 : \theta > 0$  (when  $\theta$  is one-dimensional).

In related work for the univariate model, Ling & Li (2001) merge the unit root in the autoregressive polynomial for  $\{e_t\}$  and assume that the fractional difference parameter is in the stationary region. They obtain standard asymptotics for the fractional difference parameter, but nonstandard Dickey-Fuller type asymptotics for the estimate of the unit root, see also Phillips & Durlauf (1986), the Handbook of Econometrics chapter by Watson (1994), and the references therein for an analysis of the multivariate autoregressive unit root model. On the other hand, Robinson (1994), Beran (1995), Tanaka (1999), and Nielsen (2001) all merge the unit root in the fractional difference parameter. Thus, some knowledge about the dynamics of the process is assumed, since no unit root must be estimated in the autoregressive polynomial, and hence avoiding the nonsmooth behavior of the model near these unit roots, which makes it possible to obtain standard asymptotics. For the full discussion, see Ling & Li (2001, pp. 739-741).

Our tests are a multivariate version of the univariate tests in Robinson (1994), Tanaka (1999), and Nielsen (2001), who considered testing for a unit root in a fractional integration framework, i.e. testing the parameter  $\theta$  in a univariate version of (2) in the frequency and time domains, respectively. They have shown that Lagrange Multiplier (LM) or score tests within the univariate fractionally integrated class have desirable power properties. The latter two also consider the estimation of the model under the alternative and were thus able to derive the equivalent Wald (W) and likelihood ratio (LR) tests of the fractional unit root hypothesis. They showed that their tests are asymptotically normal or chi-squared

distributed and, under invariance conditions and Gaussianity, that the tests are locally most powerful and indeed uniformly most powerful in case  $\theta$  is scalar. Simulations showed that in finite samples the time domain tests are superior to Robinson's (1994) frequency domain LM test with respect to both size and power.

As we shall see below their techniques generalize to our multivariate setup, allowing us to conduct standard inference on the integration orders in the multivariate fractionally integrated time series model (1) – (2). Some examples of interesting hypotheses that can be tested within this class of models are: (i) The unit root hypothesis. The standard unit root hypothesis nested in a fractionally integrated model is (3) with  $d_k = 1$  and can be considered an alternative to testing for a unit root nested in an autoregressive framework as in Phillips & Durlauf (1986). (ii) Short memory. Using  $d_k = 0$  we can test the hypothesis that the components of  $\{u_t\}$  have only short memory, i.e. an alternative to the multivariate autoregressive stationarity tests by Choi & Ahn (1999). (iii)  $I(2)$ . The hypothesis that the variables are jointly  $I(2)$  is (3) with  $d_k = 2$ .

More importantly, our multivariate setup allows us to test whether two or more series are integrated of the same (possibly fractional) order, with or without specifying what this common integration order should be. Such a test is an important motivation for the current study.

The inferential procedures proposed and analyzed here are intended mainly for preliminary data analysis. For instance, failing to reject (against fractional alternatives) the null of stationarity or  $I(0)$ -ness would allow for standard causality, structural VAR, or impulse response analyses. Furthermore, the estimation of the fractional integration parameter(s) would indicate the appropriate transformation of the data in order to conduct such analyses, i.e. the required order of (fractional) differencing.

We establish the desirable distributional and optimality properties of the local time domain maximum likelihood estimator (MLE) and related (LR, W, and LM) test statistics. We show that the likelihood theory in the time domain is tractable and that desirable properties from standard statistical analysis apply. In particular, there exists a local MLE which is asymptotically normal, and the test statistics are asymptotically equivalent and chi-squared distributed under local alternatives. Under the additional assumption of Gaussianity, the local MLE is asymptotically efficient in the sense that its variance attains the Cramér-Rao lower bound, and the tests are asymptotically efficient in the Pitman sense, i.e. have maximal noncentrality parameter and thus maximal asymptotic relative efficiency. In a simulation study we examine the properties of the LR test in finite samples, and find that the size of the test is close to the nominal level and that rejection frequencies are close to the asymptotic local power, even for samples as small as  $n = 100$ .

The rest of the paper is laid out as follows. Next, we consider the simple case where  $\theta$  is the same

across equations, i.e.  $\theta_k = \theta, k = 1, \dots, K$ , and in section 3 we move to the general case with a different  $\theta_k$  for each equation. Sections 2 and 3 assume *i.i.d.* errors. In section 4 we derive the analogues of the results in sections 2 and 3, but allowing the errors to be generated by a VAR( $p$ ) process. In section 5 we consider testing for a common unspecified integration order, both with *i.i.d.* and autocorrelated errors. Section 6 presents the results of the simulation study and section 7 offers some concluding remarks. Proofs are collected in appendix A while appendix B contains some technical lemmas.

## 2 Same $\theta$ Across Equations

This is the case where it is known that the integration orders are the same for all variables, and we wish to estimate this common integration order or test if it is equal to some prespecified value, e.g.  $d = 1$ . This could easily be generalized to different values for each  $d_k$ , but that seems to be of little practical interest without different  $\theta_k$ 's as well, see section 3.

The Gaussian log-likelihood function is, apart from constants,

$$L(\theta, \Sigma) = -\frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^n e_t' \Sigma^{-1} e_t. \quad (4)$$

Note that the asymptotic results derived later only impose *i.i.d.* errors and not Gaussianity. We use Gaussianity only to choose a likelihood function and later as a benchmark to derive efficiency results. Concentrating  $L$  with respect to  $\Sigma$  we obtain

$$l(\theta) = -\frac{n}{2} \ln |\Sigma(\theta)|,$$

where  $\Sigma(\theta) = n^{-1} \sum_{t=1}^n ((1-L)^{d+\theta} u_t)((1-L)^{d+\theta} u_t)'$ . We equivalently consider the maximization of

$$g(\theta) = -\frac{n}{2} \ln \left[ 1 - \frac{1}{n} \frac{n |\Sigma(0)| - n |\Sigma(\theta)|}{|\Sigma(0)|} \right]. \quad (5)$$

Thus, the MLE minimizes the conditional sum of squares function  $|\Sigma(\theta)|$  and is computationally very simple. In contrast, Hosoya (1996) considers a rather complicated exact frequency domain MLE and derive the asymptotics based on the bracketing function approach.

First, assume that there exists a  $\delta$  such that  $\delta = \sqrt{n}\theta$ , the existence of  $\delta$  will be proven shortly. Then we have the following.

**Theorem 2.1** *Let  $g(\theta)$  be given by (5). Then, under  $\theta = \delta/\sqrt{n}$ ,*

$$\begin{aligned} (i) \quad & g(\theta) \rightarrow_d W_s(\delta) = \frac{\delta}{2} \left( 2\sqrt{\mathcal{I}_s}Z - \mathcal{I}_s\delta \right), \\ (ii) \quad & \frac{\partial g(\theta)}{\partial \delta} \rightarrow_d \frac{\partial W_s(\delta)}{\partial \delta} = \sqrt{\mathcal{I}_s}Z - \mathcal{I}_s\delta, \\ (iii) \quad & \frac{\partial^2 g(\theta)}{\partial \delta^2} \rightarrow_p -\mathcal{I}_s, \end{aligned}$$

as  $n \rightarrow \infty$ , where  $\mathcal{I}_s = \pi^2 K/6$ ,  $Z$  is a standard normal random variable, and the subscript  $s$  denotes same  $\theta$  across equations.

Thus, we need to prove the existence of a local MLE  $\hat{\theta}$  such that  $\sqrt{n}\hat{\theta} = \hat{\delta} = O_p(1)$ . Following Sargan & Bhargava (1983) and Tanaka (1999) it suffices to show that

$$P\left(\frac{\partial g(\delta_1/\sqrt{n})}{\partial \delta} \geq 0\right) \leq \varepsilon \quad (6)$$

$$P\left(\frac{\partial g(\delta_2/\sqrt{n})}{\partial \delta} \leq 0\right) \leq \varepsilon \quad (7)$$

for any  $\varepsilon > 0$ ,  $n \geq n_0$  ( $n_0$  fixed) and for some  $\delta_1 > 0$ ,  $\delta_2 < 0$ . It follows from Theorem 2.1 that

$$\begin{aligned} P\left(\frac{\partial g(\delta_1/\sqrt{n})}{\partial \delta} \geq 0\right) & \rightarrow P\left(\frac{\partial W_s(\delta_1)}{\partial \delta} \geq 0\right) \\ & = P\left(\frac{\partial W_s(\delta_1)}{\partial \delta} - E\frac{\partial W_s(\delta_1)}{\partial \delta} \geq -E\frac{\partial W_s(\delta_1)}{\partial \delta}\right) \\ & \leq \frac{V\left(\frac{\partial W_s(\delta_1)}{\partial \delta}\right)}{\left(E\frac{\partial W_s(\delta_1)}{\partial \delta}\right)^2} \end{aligned}$$

by Chebyshev's inequality. In view of Theorem 2.1(ii) the last expression is  $1/\delta_1^2\mathcal{I}_s$ , and similarly  $P(\partial g(\delta_2/\sqrt{n})/\partial \delta \leq 0) \leq 1/\delta_2^2\mathcal{I}_s + o(1)$ . Thus, (6) and (7) hold by appropriate choices of  $\delta_1, \delta_2$ , and  $n_0$ , and the existence of the local MLE  $\hat{\theta}$  is ensured.

**Theorem 2.2** *There exists a local maximizer  $\hat{\theta}_n$  of (5), which satisfies, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}\hat{\theta}_n \rightarrow_d N(0, \mathcal{I}_s^{-1}), \quad (8)$$

where  $\mathcal{I}_s$  is defined in Theorem 2.1. Under the additional assumption of Gaussianity of  $\{e_t\}$ ,  $\hat{\theta}_n$  is asymptotically efficient in the sense that its asymptotic variance attains the Cramér-Rao lower bound.

The estimation of  $\theta$  as considered in this paragraph is important prior to a fractional cointegration analysis. Most procedures currently available to conduct such fractional cointegration analyses assume a triangular structure and require a prespecified value of the common integration order of the right hand



side variables, e.g. Jeganathan (1999), Breitung & Hassler (2002), and Nielsen (2002). Thus, suppose the model is

$$\begin{aligned}(1-L)^{d-b}(y_{1t}-\beta'y_{2t}) &= e_{1t}\mathbb{I}(t\geq 1), \\ (1-L)^d y_{2t} &= e_{2t}\mathbb{I}(t\geq 1),\end{aligned}$$

where  $\Sigma_2$ , the contemporaneous covariance matrix of  $e_{2t}$ , is assumed to be positive definite. I.e., the components of  $y_{2t}$  are assumed to be noncointegrated and integrated of the same order,  $d$ . Theorem 2.2 shows that an efficient estimate of this common integration order can be readily obtained using the techniques in this section.

Next, we examine the properties of the classical LR, LM, and Wald test statistics. The LR test statistic is

$$LR = 2\left(L(\hat{\theta}, \hat{\Sigma}) - L(0, \tilde{\Sigma})\right), \quad (9)$$

where  $\hat{\Sigma} = \Sigma(\hat{\theta})$  and  $\tilde{\Sigma} = \Sigma(0)$ . The Wald test statistic based on Theorem 2.2 is

$$W = n\hat{\theta}'\mathcal{I}_s\hat{\theta}, \quad (10)$$

where  $\mathcal{I}_s$  defined in Theorem 2.1 is the normalized Fisher information under Gaussianity. Finally, based on the normalized score statistic, see Tanaka (1999) and Breitung & Hassler (2002),

$$\begin{aligned}S_n &= \frac{1}{\sqrt{n}} \left. \frac{\partial L(\theta, \Sigma)}{\partial \theta} \right|_{\theta=0, \Sigma=\tilde{\Sigma}} = -\frac{1}{\sqrt{n}} \sum_{t=1}^n (\ln(1-L)e'_t) \tilde{\Sigma}^{-1} e_t \\ &= \sqrt{n} \sum_{j=1}^{n-1} j^{-1} \text{tr} \left( \tilde{\Sigma}^{-1} C(j) \right),\end{aligned} \quad (11)$$

where  $C(j) = n^{-1} \sum_{t=j+1}^n e_t e'_{t-j}$ , we can form the LM test statistic

$$LM = S'_n \mathcal{I}_s^{-1} S_n. \quad (12)$$

Corresponding to  $S_n$ , whose distribution is easily derived from Theorem 2.1, we could also define a one-sided Wald test, but they will make little sense in the next sections with multi-dimensional  $\theta$ , so we will not consider them here. When  $K = 1$ , i.e. when the observed time series is univariate, the score statistic in (11) reduces to the univariate time domain score statistic,  $s_n = \sqrt{n} \sum_j j^{-1} \hat{\rho}(j)$ , where  $\hat{\rho}(j)$  is the  $j$ 'th sample autocorrelation of  $\{e_t\}$ , see e.g. Tanaka (1999) or Nielsen (2001). In parallel to its univariate counterpart, this test bears some resemblance to the multivariate portmanteau statistic,  $P = n \sum_j \text{tr}(\hat{C}(j) \hat{\Sigma}^{-1} \hat{C}(j) \hat{\Sigma}^{-1})$ ,  $\hat{C}(j) = n^{-1} \sum_{t=j+1}^n \hat{e}_t \hat{e}'_{t-j}$ , for checking the whiteness of the residuals of a multivariate ARMA time series model, see e.g. Hosking (1980, p. 605).

The derivative found above,  $\sqrt{n} \sum_j j^{-1} \text{tr}(\hat{\Sigma}^{-1} C(j))$ , is also derived by Breitung & Hassler (2002, p. 171) for the model (2) with  $\theta_k = \theta$ ,  $k = 1, \dots, K$ . However, instead of examining the properties of this derivative or the generalizations in the next sections, they consider a multivariate variant of the fractional Dickey-Fuller test also considered by Dolado, Gonzalo & Mayoral (2002). They call it the score statistic for testing  $H_0 : \theta = 0$  and show that the corresponding quadratic form is chi-squared with  $K^2$  degrees of freedom. However, their test is not equivalent to the multivariate LM test of (3), as demonstrated for the univariate test by Breitung & Hassler (2002). Indeed, the main aim of Breitung & Hassler (2002) is to construct a fractional trace statistic similar to Johansen (1988), just as the Dickey-Fuller test generalizes to Johansen's (1988) trace statistic.

In the following theorem we present the distribution of the test statistics under alternatives local to the null,  $H_{1n} : \theta = \delta/\sqrt{n}$ , where  $\delta$  is a fixed scalar.

**Theorem 2.3** *Under  $\theta = \delta/\sqrt{n}$ , the LR, W, and LM test statistics defined by (9), (10), and (12) are asymptotically equivalent and distributed as  $\chi_1^2(\mathcal{I}_s \delta^2)$ , where  $\mathcal{I}_s$  is defined in Theorem 2.1. Under the additional assumption of Gaussianity the tests are efficient in the Pitman sense.*

Thus it is seen that standard statistical results apply in our model, unlike in the autoregressive unit root models in e.g. Phillips & Durlauf (1986) and Choi & Ahn (1999). In particular, the equivalence of the tests follows from the information matrix equality which holds in our case, contrary to AR-based unit root models.

### 3 Different $\theta$ 's Across Equations

In this section we consider the full model in (1) – (2) where the  $\theta_k$  are potentially different for each equation. Thus, we no longer force the integration orders of the variables to be equal.

Again, we assume first that we are in a neighborhood of the true value, i.e. that there exists a  $K$ -vector  $\delta$  such that  $\delta = \theta\sqrt{n}$ . The function  $g(\theta)$  is still given by (5), but  $\theta$  is now a vector and  $\Sigma(\theta)$  is redefined accordingly. Then we have the following theorem, where  $\odot$  denotes the Hadamard (see appendix B or Magnus & Neudecker (1999)).

**Theorem 3.1** *Let  $g(\theta)$  be given by (5), where  $\theta$  is now a  $K$ -vector. Then, under  $\theta = \delta/\sqrt{n}$ ,*

$$\begin{aligned} (i) \quad & g(\theta) \rightarrow_d W_d(\delta) = \frac{\delta'}{2} \left( 2\mathcal{I}_d^{1/2} Z - \mathcal{I}_d \delta \right), \\ (ii) \quad & \frac{\partial g(\theta)}{\partial \delta} \rightarrow_d \frac{\partial W_d(\delta)}{\partial \delta} = \mathcal{I}_d^{1/2} Z - \mathcal{I}_d \delta, \\ (iii) \quad & \frac{\partial^2 g(\theta)}{\partial \delta \partial \delta'} \rightarrow_p -\mathcal{I}_d, \end{aligned}$$

as  $n \rightarrow \infty$ , where  $\mathcal{I}_d = (\Sigma \odot \Sigma^{-1}) \pi^2/6$ ,  $Z$  is a  $K$ -dimensional standard normal random vector, and the subscript  $d$  denotes different  $\theta$ 's across equations.

When  $\theta$  is a vector of parameters we need to generalize the approach of Sargan & Bhargava (1983) and Tanaka (1999) to prove the existence of a local MLE  $\hat{\theta}_n$  such that  $\sqrt{n}\hat{\theta}_n = \hat{\delta} = O_p(1)$ . Let  $\eta$  be a  $p \times 1$  direction vector, i.e. satisfying  $\|\eta\| = 1$ , where  $\|\cdot\|$  is the Euclidean norm, and let  $\delta = \|\delta\| \eta$ . Generalizing the scalar approach by Sargan & Bhargava (1983) and Tanaka (1999), see also section 2 above, it suffices to show that

$$P \left( \eta' \frac{\partial g(\delta/\sqrt{n})}{\partial \delta} \geq 0 \right) \leq \varepsilon \quad (13)$$

for any direction vector  $\eta$ ,  $\varepsilon > 0$ , and  $n \geq n_0$  ( $n_0$  fixed), and for some  $\|\delta\| > 0$ . Note that  $\eta' \partial g(\delta/\sqrt{n}) / \partial \delta$  is the directional derivative of  $g$  at  $\delta/\sqrt{n}$ , i.e. the rate of change of  $g$  at  $\delta/\sqrt{n}$  in the direction  $\eta$ .

Thus, for all direction vectors  $\eta$ , moving some distance  $\|\delta\|$  in the direction  $\eta$  from the true value, the directional derivative of  $g$  in the same direction  $\eta$  should be negative for sufficiently large  $n$ . In the one-dimensional case  $\eta = \pm 1$  and (13) reduces to (6) and (7), i.e. the corresponding conditions of Sargan & Bhargava (1983) and Tanaka (1999). It follows from Theorem 3.1 that

$$\begin{aligned} P \left( \eta' \frac{\partial g(\delta/\sqrt{n})}{\partial \delta} \geq 0 \right) & \rightarrow P \left( \eta' \frac{\partial W(\delta)}{\partial \delta} \geq 0 \right) \\ & = P \left( \eta' \frac{\partial W(\delta)}{\partial \delta} - E \eta' \frac{\partial W(\delta)}{\partial \delta} \geq -E \eta' \frac{\partial W(\delta)}{\partial \delta} \right) \\ & \leq \frac{\text{Var} \left( \eta' \frac{\partial W(\delta)}{\partial \delta} \right)}{\left( E \eta' \frac{\partial W(\delta)}{\partial \delta} \right)^2} \\ & = \frac{1}{\eta' \Psi \eta \|\delta\|^2}, \end{aligned}$$

which can be made arbitrarily small by selecting  $\|\delta\|$  large. Thus, (13) holds by appropriate choices of  $\|\delta\|$  and  $n_0$ , and the existence of the local MLE  $\hat{\theta}_n$  is ensured.

**Theorem 3.2** *There exists a local maximizer  $\hat{\theta}_n$  (now a  $K$ -vector), which satisfies, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}\hat{\theta}_n \rightarrow_d N(0, \mathcal{I}_d^{-1}), \quad (14)$$

where  $\mathcal{I}_d$  is defined in Theorem 3.1. Under the additional assumption of Gaussianity of  $\{e_t\}$ ,  $\hat{\theta}_n$  is asymptotically efficient in the sense that its asymptotic variance attains the Cramér-Rao lower bound.

In the present case with different integration orders the score statistic is

$$\begin{aligned}
S_n &= \frac{1}{\sqrt{n}} \left. \frac{\partial L(\theta, \Sigma)}{\partial \theta} \right|_{\theta=0, \Sigma=\tilde{\Sigma}} = -\frac{1}{\sqrt{n}} \sum_{t=1}^n \text{diag}(\ln(1-L)e_t) \tilde{\Sigma}^{-1} e_t \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} J'_K(I_K \otimes \tilde{\Sigma}^{-1})(e_{t-j} \otimes e_t) \\
&= \sqrt{n} \sum_{j=1}^{n-1} j^{-1} J'_K(I_K \otimes \tilde{\Sigma}^{-1}) \text{vec } C(j)
\end{aligned} \tag{15}$$

by use of  $\text{vec}(ABC) = (C' \otimes A) \text{vec } B$  and property 1 of Lemma 2. We denote by  $\text{diag}(a)$  the diagonal matrix having the vector  $a$  on the diagonal and  $J_K$  is defined in Lemma 2. Exactly as in section 2, the score statistic (15) reduces to the univariate score statistic when  $K = 1$ .

Thus, we can form the LR, W, and LM test statistics as in (9) – (12), but with  $S_n$  defined in (15) and  $\mathcal{I}_s$  replaced by  $\mathcal{I}_d$ , and where the likelihood function is now in terms of different  $\theta_k$  for each equation. We proceed to consider the distribution of the test statistics under alternatives local to the null,  $H_{1n} : \theta = \delta/\sqrt{n}$ , where  $\delta$  is now a fixed  $K$ -vector.

**Theorem 3.3** *Under  $\theta = \delta/\sqrt{n}$ ,  $\delta$  a fixed  $K$ -vector, the LR, W, and LM test statistics defined by (9), (10), and (12), with  $S_n$  defined in (15) and  $\mathcal{I}_s$  replaced by  $\mathcal{I}_d$ , are asymptotically equivalent and distributed as  $\chi_K^2(\delta' \mathcal{I}_d \delta)$ , where  $\mathcal{I}_d$  is defined in Theorem 3.1. Under the additional assumption of Gaussianity the tests are efficient in the Pitman sense.*

Where the results in section 2 were useful in estimating a common integration order of the 'right hand side variables' in the triangular model prior to a fractional cointegration analysis, the results in this section have enabled us to test whether the integration orders are indeed all equal to some pre-assigned value without requiring them to be equal under the alternative. Thus, by setting  $d_k = 1$  or  $d_k = 2$  for all  $k$ , this could indeed be considered a valuable diagnostic test prior to a standard  $I(1)$  or  $I(2)$  cointegration analysis.

In section 5 we combine these procedures to derive a test of the hypothesis that the integration orders are equal without specifying the common integration order, i.e. testing if the variables have a common unspecified integration order.

## 4 Inference with Autocorrelated Errors

In this section we allow the errors  $\{e_t\}$  to follow to finite vector autoregressive process,

$$A(L)e_t = \varepsilon_t, \quad (16)$$

where  $\{\varepsilon_t\}$  satisfies the assumptions of  $\{e_t\}$  before. Here,  $A(z)$  is a matrix polynomial of order  $p$ , such that  $\{e_t\}$  is a stationary VAR( $p$ ) process and  $A(1)$  has full rank. An ARMA process could be used for  $\{e_t\}$  but it adds significantly to the notational burden and to the difficulty of our proofs, and in empirical applications it is often found that a short autoregressive polynomial is sufficient to describe data in fractional models. The parameters of  $A(z)$  are gathered in the  $K^2p$ -vector  $\psi = \text{vec}(A_1, \dots, A_p)$  with true value  $\psi_0$ .

We consider first the case with the same  $\theta$  across equations as in section 2. The log-likelihood function with autocorrelated errors is

$$L(\theta, \psi, \Sigma) = -\frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^n (A(L)e_t)' \Sigma^{-1} (A(L)e_t), \quad (17)$$

and concentrating with respect to  $\Sigma$  we obtain

$$l(\theta, \psi) = -\frac{n}{2} \ln |\Sigma(\theta, \psi)|,$$

where  $\Sigma(\theta, \psi) = n^{-1} \sum_{t=1}^n (A(L)(1-L)^{d+\theta} u_t)(A(L)(1-L)^{d+\theta} u_t)'$ . Equivalently, we consider the maximization of

$$g(\theta, \psi) = -\frac{n}{2} \ln \left[ 1 - \frac{1}{n} \frac{n |\Sigma(0, \psi_0)| - n |\Sigma(\theta, \psi)|}{|\Sigma(0, \psi_0)|} \right]. \quad (18)$$

With autocorrelated errors the LR test statistic is

$$LR = 2 \left( L(\hat{\theta}, \hat{\psi}, \hat{\Sigma}) - L(0, \tilde{\psi}, \tilde{\Sigma}) \right), \quad (19)$$

where  $\hat{\Sigma} = \Sigma(\hat{\theta}, \hat{\psi})$  and  $\tilde{\Sigma} = \Sigma(0, \tilde{\psi})$  and the Wald test statistic and LM test statistics are given by (10) and (12) as before, but with a different  $\mathcal{I}_s$  and with the score statistic

$$S_n = \frac{1}{\sqrt{n}} \frac{\partial L(\theta, \psi, \Sigma)}{\partial \theta} \Big|_{\theta=0, \psi=\tilde{\psi}, \Sigma=\tilde{\Sigma}} = \sqrt{n} \sum_{j=1}^{n-1} j^{-1} \text{tr}(\tilde{\Sigma}^{-1} \tilde{C}(j)), \quad (20)$$

where  $\tilde{C}(j) = n^{-1} \sum_{t=j+1}^n \tilde{\varepsilon}_t \tilde{\varepsilon}'_{t-j}$  is the estimated residual autocovariance function under the null.

We are now able to prove joint asymptotic normality of  $\hat{\theta}$  and  $\hat{\psi}$ , the local MLEs of  $\tau_0 = (\theta_0, \psi_0)'$ , and under Gaussianity we achieve efficiency as before. Furthermore, the results of Theorem 2.3 continue to hold in the present case with autocorrelated errors, although the noncentrality parameter is different.

**Theorem 4.1** *There exists a local maximizer  $\hat{\tau} = (\hat{\theta}, \hat{\psi}')'$  of (18), which satisfies, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{\tau} - \tau_0) \rightarrow_d N(0, \Xi^{-1}), \quad (21)$$

and in particular

$$\sqrt{n}\hat{\theta}_n \rightarrow_d N(0, \mathcal{I}_s^{-1}), \quad (22)$$

where

$$\begin{aligned} \Xi &= \begin{bmatrix} \frac{\pi^2 K}{6} & (\text{vec } \Phi')' \\ \text{vec } \Phi' & \Gamma \otimes \Sigma^{-1} \end{bmatrix} \\ \mathcal{I}_s &= \frac{\pi^2 K}{6} - \text{tr } \Phi \Gamma^{-1} \Phi' \Sigma. \end{aligned} \quad (23)$$

Here,  $\Gamma$  is the covariance matrix of  $(e'_t, \dots, e'_{t-p+1})'$ ,  $\Phi = (\Phi'_1, \dots, \Phi'_p)'$ ,  $\Phi_i = \sum_{j=i}^{\infty} j^{-1} a_{j-i}$ , and  $a_i$  is the coefficient to  $z^i$  in the Wold representation of  $\{e_t\}$ .

The LR, W, and LM test statistics in (19), (10), and (12), with  $S_n$  defined in (20) and  $\mathcal{I}_s$  defined in (23), are asymptotically equivalent and distributed as  $\chi_1^2(\mathcal{I}_s \delta^2)$  under the local alternatives  $H_{1n} : \theta = \delta/\sqrt{n}$ .

Under the additional assumption of Gaussianity,  $\hat{\tau}$  is asymptotically efficient in the sense that its asymptotic variance attains the Cramér-Rao lower bound, and the tests are efficient in the Pitman sense.

Setting  $\{\varepsilon_t\}$  equal to  $\{e_t\}$ , i.e.  $\Gamma = \Phi = 0$ , the results in section 2 with *i.i.d.* errors appear as a special case of this more general theorem. Furthermore,  $\Gamma \otimes \Sigma^{-1}$  is the normalized Fisher information for  $\psi$  in the VAR( $p$ ) model (16) in case  $\{e_t\}$  was an observed process.

As a simple example consider the VAR(1),  $e_t = A e_{t-1} + \varepsilon_t = \sum_{j=0}^{\infty} A^j \varepsilon_{t-j}$ . In this case, (23) reduces to  $\pi^2 K/6 - \text{tr } \Phi_1 \Gamma^{-1} \Phi_1' \Sigma$ , where  $\Phi_1 = I_K + \sum_{j=2}^{\infty} j^{-1} A^{j-1}$  and  $\Gamma = E(e_t e_t')$  can be estimated by  $n^{-1} \sum_{t=1}^n \hat{e}_t \hat{e}_t'$  or recovered from the relation  $\text{vec } \Gamma = (I_{K^2} - A \otimes A)^{-1} \text{vec } \Sigma$ . When  $A$  is diagonalizable, with eigenvalues denoted  $\lambda_1, \dots, \lambda_K$ ,  $\Phi_1$  is easily calculated by diagonalization of  $A$ , viz.

$$\begin{aligned} \Phi_1 &= I_K + \sum_{j=2}^{\infty} j^{-1} P \text{diag}(\lambda_1^{j-1}, \dots, \lambda_K^{j-1}) P^{-1} \\ &= P \text{diag}(-\lambda_1^{-1} \ln(1 - \lambda_1), \dots, -\lambda_K^{-1} \ln(1 - \lambda_K)) P^{-1}, \end{aligned}$$

where  $P$  is the matrix having the eigenvectors of  $A$  as its columns.

Proceeding to the case with a different  $\theta_k$  for each equation, an important caveat applies in our multivariate setup as pointed out by Lobato (1997) in a different context. Namely that the ordering of the autoregressive polynomial and the differencing operator matters. In our multivariate ARFIMA( $p, d, 0$ )

time series model in (1) – (2), (16), it is apparent that  $u_{kt}$  (and thus also  $y_{kt}$ ) is integrated of order  $d_k$  for all  $k = 1, \dots, K$ . However, suppose instead that the model (bivariate for simplicity) was given by

$$\begin{pmatrix} (1-L)^{d_1} & 0 \\ 0 & (1-L)^{d_2} \end{pmatrix} \begin{pmatrix} a_{11}(L) & a_{12}(L) \\ a_{21}(L) & a_{22}(L) \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \quad (24)$$

where the autoregressive polynomial and the differencing operator have been interchanged compared to (1) – (2), (16). Then we can write  $u_{1t}$  as

$$(a_{11}(L) a_{22}(L) - a_{12}(L) a_{21}(L)) (1-L)^{d_1+d_2} u_{1t} = a_{22}(L) (1-L)^{d_2} \varepsilon_{1t} - a_{12}(L) (1-L)^{d_1} \varepsilon_{2t} \quad (25)$$

and thus  $u_{1t} \in I(d_2)$  if  $d_1 < d_2$  and  $a_{12}(1) \neq 0$ , and  $u_{1t} \in I(d_1)$  otherwise. Similarly,  $u_{2t} \in I(d_1)$  if  $d_2 < d_1$  and  $a_{21}(1) \neq 0$ , and  $u_{2t} \in I(d_2)$  otherwise. The model (24) is equivalent to our model only in the univariate setup or when  $d_k = d$  for all  $k = 1, \dots, K$ , i.e. when the situation is as in section 2.

In the present case with autocorrelated errors and a different  $\theta_k$  for each equation, the function  $g(\theta, \psi)$  is still given by (18), but  $\theta$  is now a vector and  $\Sigma(\theta, \psi)$  is redefined accordingly. We then have the following theorem.

**Theorem 4.2** *There exists a local maximizer  $\hat{\tau} = (\hat{\theta}, \hat{\psi})'$ , where  $\hat{\theta}$  is now a  $K$ -vector, which satisfies, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{\tau} - \tau_0) \rightarrow_d N(0, \Xi^{-1}), \quad (26)$$

and in particular

$$\sqrt{n}\hat{\theta}_n \rightarrow_d N(0, \mathcal{I}_d^{-1}), \quad (27)$$

where

$$\begin{aligned} \Xi &= \begin{bmatrix} \frac{\pi^2}{6} \Sigma \odot \Sigma^{-1} & J'_K(\Phi \otimes I_K) \\ (\Phi' \otimes I_K) J_K & \Gamma \otimes \Sigma^{-1} \end{bmatrix}, \\ \mathcal{I}_d &= \frac{\pi^2}{6} \Sigma \odot \Sigma^{-1} - (\Sigma \Phi \Gamma^{-1} \Phi' \Sigma) \odot \Sigma^{-1}, \end{aligned} \quad (28)$$

$\Gamma$  and  $\Phi$  are defined in Theorem 4.1, and  $J_K$  is defined in Lemma 2. The LR, W, and LM test statistics in (19), (10), and (12), with  $C(j)$  replaced by  $\tilde{C}(j)$  and  $\mathcal{I}_s$  replaced by  $\mathcal{I}_d$  defined in (28), are asymptotically equivalent and distributed as  $\chi_{K'}^2(\delta' \mathcal{I}_d \delta)$  under the local alternatives. Under the additional assumption of Gaussianity,  $\hat{\tau}$  and the tests are asymptotically efficient.

Consistent estimates of the parameters  $\Phi_i$  and  $\Gamma$  appearing in the asymptotic distributions in Theorems 4.1 and 4.2 can be obtained from the formulae given above using estimates  $\hat{\tau}$  or  $\tilde{\tau} = (0, \tilde{\psi})$ .

We end this section by remarking that the LR test statistic has the computational advantage that there is no need to calculate the covariance matrices in  $\mathcal{I}_s$  and  $\mathcal{I}_d$ , i.e. the  $\Phi$  and  $\Gamma$  matrices, which can be quite cumbersome when  $p \geq 2$ .

## 5 Testing Common Unknown Integration Order

Now we consider the case where  $d$  is a nuisance parameter instead of a prespecified parameter, and therefore  $d$  must be estimated under the null and alternative hypotheses. Alternatively, we can formulate the hypothesis in the framework of section 3 as

$$H_0 : R\theta = 0, \quad R = [I_{K-1}; -\iota], \quad (29)$$

where  $\iota = [1, \dots, 1]'$ . Thus, again only  $\theta$  is of interest and  $d$  returns to being prespecified. Notice that  $\text{rank}(R) = K - 1$  such that there are  $K - 1$  restrictions. Under  $H_0$  we estimate a common  $\theta = \tilde{\theta}$  (scalar) as in Theorems 2.2 and 4.1, and under the alternative we estimate  $\theta = \hat{\theta}$  ( $K$ -vector) as in Theorems 3.2 and 4.2. The test statistics are

$$LR = 2 \left( L(\hat{\theta}, \hat{\psi}, \hat{\Sigma}) - L(\tilde{\theta}, \tilde{\psi}, \tilde{\Sigma}) \right), \quad (30)$$

$$W = n\hat{\theta}R' (R\mathcal{I}_d^{-1}R')^{-1} R\hat{\theta}, \quad (31)$$

$$LM = \frac{\partial L(\theta, \psi, \Sigma)}{\partial \theta'} \Big|_{H_0} \left[ -E \frac{\partial^2 L(\theta, \psi, \Sigma)}{\partial \theta \partial \theta'} \Big|_{H_0} \right]^{-1} \frac{\partial L(\theta, \psi, \Sigma)}{\partial \theta} \Big|_{H_0}, \quad (32)$$

where  $L$  is the likelihood function with different  $\theta$ 's and  $\mathcal{I}_d$  is defined in Theorems 3.1 and 4.2. The limiting distributions under local alternatives is given in the following theorem.

**Theorem 5.1** *Under  $R\theta = \delta/\sqrt{n}$ ,  $\delta$  a fixed  $(K-1)$ -vector, the test statistics in (30), (31), and (32) are asymptotically equivalent and  $\chi_{K-1}^2(\delta'R\mathcal{I}_dR'\delta)$  distributed, where  $\mathcal{I}_d$  is defined in Theorems 3.1 and 4.2. Under the additional assumption of Gaussianity the tests are efficient in the Pitman sense.*

## 6 Finite Sample Performance

In this section we compare the finite sample properties of the LR test with the approximation offered by asymptotic theory in sections 3 and 4. The asymptotic local power of the test is easily derived from Theorems 3.3 and 4.2 as

$$P(LR > \chi_{K,1-\alpha}^2) = 1 - F_\lambda(\chi_{K,1-\alpha}^2), \quad (33)$$



where  $\chi_{K,1-\alpha}^2$  is the 100(1 -  $\alpha$ )% point of the central  $\chi^2$  distribution with  $K$  degrees of freedom and  $F_\lambda$  is the distribution function of the noncentral  $\chi^2$  distribution with  $K$  degrees of freedom and noncentrality parameter  $\lambda = \delta' \mathcal{I}_d \delta$  defined in Theorems 3.1 and 4.2. Setting  $\delta = \theta \sqrt{n}$  in (33) we can compare the asymptotic local power with the finite sample rejection frequencies for any fixed value of  $\theta$ .

We report only the results for the LR test since Nielsen (2001) demonstrated the superiority of the LR test in the univariate setting and since it has the computational advantage that there is no need to compute the covariance matrices as argued in the previous sections.

The models we consider for the simulation study are the bivariate fractional unit root models,

$$\text{Model A : } \begin{bmatrix} (1-L)^{1+\theta_1} & 0 \\ 0 & (1-L)^{1+\theta_2} \end{bmatrix} u_t = \varepsilon_t \mathbb{I}(t \geq 1), \quad (34)$$

$$\text{Model B : } (I_2 - AL) \begin{bmatrix} (1-L)^{1+\theta_1} & 0 \\ 0 & (1-L)^{1+\theta_2} \end{bmatrix} u_t = \varepsilon_t \mathbb{I}(t \geq 1), \quad A = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}, \quad (35)$$

where the  $\{\varepsilon_t\}$  are *i.i.d.*  $N(0, \Sigma)$ . The contemporaneous covariance matrix  $\Sigma$  is normalized such that the diagonal elements equal unity and the correlation  $\rho$  is 0 or 0.6. The sensitivity of the test to the parameters in the coefficient matrix  $A$  is examined in Table 3 below. Throughout, we fix the nominal size (type I error) of the test at 5% and the number of replications at 10,000. We consider the sample sizes  $n = 100$  and  $n = 250$ . All calculations were made in Ox version 3.20 including the Arfima package version 1.01 (see Doornik (2001) and Doornik & Ooms (2001)).

### Tables 1-2 about here

In Table 1 the finite sample rejection frequencies of the LR test with *i.i.d.* errors is presented, i.e. Model A with  $\rho = 0$  and  $\rho = 0.6$ . The asymptotic local power for the same value of  $(\theta_1, \theta_2)$  is given in parenthesis. In each table the entry in bold corresponding to  $\theta_1 = \theta_2 = 0$  gives the finite sample size of the test. For both  $\rho = 0$  and  $\rho = 0.6$ , the finite sample size of the test is very close to the nominal 5% level, and the rejection frequencies are very close to the corresponding asymptotic local power. Notice that with positive contemporaneous correlation the power of the test (both finite sample and asymptotic) is especially high when  $\theta_1$  and  $\theta_2$  are of opposite sign.

Table 2 shows the finite sample rejection frequencies and asymptotic local power of the LR test for Model B with  $a = 0.5$ . In this case the test is slightly size distorted, especially when  $\rho = 0.6$  where the finite sample sizes are 0.0741 and 0.0838 for  $n = 100$  and  $n = 250$ , respectively. For both sample sizes the rejection frequencies are close to the asymptotic local power even in this model with autocorrelated errors. As before, power is particularly high in the correlated case when  $\theta_1$  and  $\theta_2$  are of opposite sign, and it also seems that the finite sample power is highest against negative values of  $\theta_1$  and  $\theta_2$ .

### Table 3 about here

To evaluate the sensitivity to the particular value of the coefficient matrix in the autoregressive specification in Model B, Table 3 presents the finite sample sizes of the LR test for different specifications of the coefficient matrix  $A$  in (35). In particular the values  $a = -0.8, -0.6, \dots, 0.8$  are considered. Notice that the column  $a = 0$  corresponds to the case where a VAR(1) is estimated for  $\{e_t\}$  even though it is really an *i.i.d.* sequence. Samples of  $n = 100$ ,  $n = 250$ , and  $n = 500$  are considered, and overall the size distortions are small. E.g. for samples of  $n = 100$  the finite sample size ranges from 0.0483 to 0.0899 for a nominal 5% test.

Rejection frequencies for the size corrected version of the LR test have been computed (analogously to Tables 1 and 2), and can be obtained from the author upon request. The results are qualitatively the same as without size correction since the test has only very little size distortion.

Overall, the LR test has good finite sample size and power properties as documented by the simulation results in Tables 1-3. Even for samples of  $n = 100$  the size of the test is good and rejection frequencies are very close to the asymptotic local power calculated using the asymptotic distribution theory in the previous sections.

## 7 Conclusion

We have proposed a maximum likelihood inference technique for multivariate fractionally integrated time series models. This generalizes recent work for univariate fractionally integrated time series models by Robinson (1994), Tanaka (1999), and Nielsen (2001), among others. In our multivariate framework, we show that inference can be drawn from the standard normal and chi-squared distributions and that the local MLE and related tests (Wald, LR, LM) are asymptotically efficient under Gaussianity. Our results have a wide range of potential applications, but in particular, we can easily test if two or more time series are integrated of the same possibly fractional order, with or without specifying this common integration order.

When an autoregressive specification of the errors is estimated, the LR test has the computational advantage that it avoids a cumbersome calculation of covariance matrices. In addition, previous Monte Carlo evidence in Nielsen (2001) shows that, in the univariate model, the LR test is superior to the LM and Wald tests in finite samples, and hence we favor the LR test for practical purposes.

To establish the empirical relevance of the LR test we have evaluated its finite sample properties by Monte Carlo experiments, which show that the size of the test is close to the nominal level and that

rejection frequencies are close to the asymptotic local power for samples as small as  $n = 100$ .

## Appendix A: Proofs

**Proof of Theorem 2.1.** The denominator in  $g(\theta)$  obviously converges in probability to  $|\Sigma|$  under the stated assumptions. By the Mean Value Theorem we have that, as in Tanaka (1999, p. 579),

$$(1-L)^{d+\theta} u_t = e_t - \frac{\delta}{\sqrt{n}} \sum_{j=1}^{t-1} j^{-1} e_{t-j} + o_p(n^{-1/2}) \quad (36)$$

uniformly in  $t$ . Consider each term in the expansion of the numerator in  $g(\theta)$ ,

$$\begin{aligned} n|\Sigma(\theta)| - n|\Sigma(0)| &= \delta\sqrt{n}|\Sigma(0)| \operatorname{tr} \left( \Sigma(0)^{-1} \frac{\partial \Sigma(0)}{\partial \theta} \right) + \frac{\delta^2}{2} |\Sigma(0)| \operatorname{tr} \left( \Sigma(0)^{-1} \frac{\partial^2 \Sigma(0)}{\partial \theta^2} \right) \\ &\quad - \frac{\delta^2}{2} |\Sigma(0)| \operatorname{tr} \left( \Sigma(0)^{-1} \frac{\partial \Sigma(0)}{\partial \theta} \Sigma(0)^{-1} \frac{\partial \Sigma(0)}{\partial \theta} \right) \\ &\quad + \frac{\delta^2}{2} |\Sigma(0)| \operatorname{tr} \left( \Sigma(0)^{-1} \frac{\partial \Sigma(0)}{\partial \theta} \right)^2 + o_p(1). \end{aligned}$$

It follows, using  $\Sigma(0) \rightarrow_p \Sigma$  and (36), that the first term is

$$\begin{aligned} \delta\sqrt{n}|\Sigma| \operatorname{tr} \left( \Sigma^{-1} \frac{2}{n} \sum_{t=1}^n (\ln(1-L) e_t) e_t' \right) + o_p(1) &= -2\sqrt{n}\delta |\Sigma| \operatorname{tr} \left( \Sigma^{-1} \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} e_{t-j} e_t' \right) + o_p(1) \\ &\rightarrow -2\delta |\Sigma| \sqrt{\frac{\pi^2 K}{6}} Z \end{aligned}$$

by Lemma 1 and the relation  $(\operatorname{vec} A)' \operatorname{vec} B = \operatorname{tr}(A'B)$ . The next term is

$$\begin{aligned} &\delta^2 |\Sigma| \operatorname{tr} \left( \Sigma^{-1} \frac{1}{n} \sum_{t=1}^n [(\ln(1-L) \ln(1-L) e_t) e_t' + (\ln(1-L) e_t) (\ln(1-L) e_t)'] \right) + o_p(1) \\ &= \delta^2 |\Sigma| \operatorname{tr} \left( \Sigma^{-1} \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} \left[ \sum_{j'=1}^{t-j-1} j'^{-1} e_{t-j-j'} e_t' + \sum_{j'=1}^{t-1} j'^{-1} e_{t-j} e_{t-j'}' \right] \right) + o_p(1) \\ &\rightarrow_p \delta^2 |\Sigma| \frac{\pi^2 K}{6} \end{aligned}$$

by application of a law of large numbers and using the uncorrelatedness of  $\{e_t\}$ . The last two terms,

$$\begin{aligned} &-2\delta^2 |\Sigma| \operatorname{tr} \left( \Sigma^{-1} \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{j=1}^{t-1} \sum_{j'=1}^{t-1} j^{-1} j'^{-1} e_{t-j} e_t' \Sigma^{-1} e_{s-j'} e_s' \right) + o_p(1), \\ &2\delta^2 |\Sigma| \operatorname{tr} \left( \Sigma^{-1} \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} e_{t-j} e_t' \right)^2 + o_p(1), \end{aligned}$$

are both negligible by Lemma 1. It follows that the numerator in  $g(\theta)$  converges in distribution to  $2\delta\sqrt{\pi^2 K/6}Z - \delta^2\pi^2 K/6$ , and because

$$g(\theta) = -\frac{n}{2} \ln \left( 1 - \frac{1}{n} (2W_s(\delta) + o_p(1)) \right) = W_s(\delta) + o_p(1) \quad (37)$$

we have established (i).

Next, we examine

$$\frac{\partial g(\theta)}{\partial \delta} = \frac{1}{\sqrt{n}} \operatorname{tr} \left( \Sigma(\theta)^{-1} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} \left( (1-L)^{d+\theta} u_{t-j} \right) \left( (1-L)^{d+\theta} u_t \right)' \right),$$

which equals, using (36) and that  $\Sigma(\theta) \rightarrow_p \Sigma$  (also follows from (36)),

$$\begin{aligned} & \frac{1}{\sqrt{n}} \operatorname{tr} \left( \Sigma^{-1} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} \left( e_{t-j} - \frac{\delta}{\sqrt{n}} \sum_{j'=1}^{t-j-1} j'^{-1} e_{t-j-j'} \right) \left( e'_t - \frac{\delta}{\sqrt{n}} \sum_{j'=1}^{t-1} j'^{-1} e'_{t-j'} \right) \right) \\ &= \frac{1}{\sqrt{n}} \operatorname{tr} \left( \Sigma^{-1} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} e_{t-j} e'_t \right) - \frac{\delta}{n} \operatorname{tr} \left( \Sigma^{-1} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-2} e_{t-j} e'_{t-j} \right) + o_p(1) \end{aligned}$$

by uncorrelatedness of  $\{e_t\}$ . As we have already seen above this converges in distribution to  $\sqrt{\pi^2 K/6}Z - \delta\pi^2 K/6$ , establishing (ii).

The second derivative is

$$\begin{aligned} \frac{\partial^2 g(\theta)}{\partial \delta^2} &= -\frac{1}{n} \operatorname{tr} \left( \Sigma(\theta)^{-1} \sum_{t=1}^n \sum_{j=1}^{t-1} \sum_{j'=1}^{t-j-1} j^{-1} j'^{-1} \left( (1-L)^{d+\theta} u_{t-j-j'} \right) \left( (1-L)^{d+\theta} u_t \right)' \right) \\ &\quad - \frac{1}{n} \operatorname{tr} \left( \Sigma(\theta)^{-1} \sum_{t=1}^n \sum_{j=1}^{t-1} \sum_{j'=1}^{t-1} j^{-1} j'^{-1} \left( (1-L)^{d+\theta} u_{t-j} \right) \left( (1-L)^{d+\theta} u_{t-j'} \right)' \right) \\ &\quad - \frac{1}{n} \operatorname{tr} \left( \Sigma(\theta)^{-1} \frac{\partial \Sigma(\theta)}{\partial \theta} \Sigma(\theta)^{-1} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} \left( (1-L)^{d+\theta} u_{t-j} \right) \left( (1-L)^{d+\theta} u_t \right)' \right), \end{aligned}$$

which equals  $-n^{-1} \operatorname{tr}(\Sigma^{-1} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-2} e_{t-j} e'_{t-j}) + o_p(1)$  by uncorrelatedness of  $\{e_t\}$  and (36). Thus, by a law of large numbers we establish (iii), completing the proof. ■

**Proof of Theorem 2.2.** By Theorem 2.1(iii),  $g(\theta)$  is asymptotically a concave function of  $\delta = \sqrt{n}\theta$  in a neighborhood of size  $O(n^{-1/2})$  of  $\theta_0$ . Hence, by Theorem 2.1 and the subsequent analysis,  $\hat{\delta} = \sqrt{n}\hat{\theta}_n$  is asymptotically the unique maximizer of  $W_s(\delta)$  in this neighborhood and its asymptotic distribution is given by (8) by the usual expansion. Under Gaussianity of  $\{e_t\}$ , (4) is the true likelihood. The limiting Fisher information is then given in Theorem 2.1(iii) as

$$\lim_{n \rightarrow \infty} \frac{1}{n} E \left( - \frac{\partial^2 g(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=0} \right) = \mathcal{I}_s,$$

which is the inverse of the asymptotic variance as required. ■

**Proof of Theorem 2.3.** The asymptotic equivalence of the three test statistics follows in the usual way since the information matrix equality holds, c.f. Theorem 2.1.

The distribution of the test statistics under local alternatives follows from Theorems 2.1 and 2.2 and the Continuous Mapping Theorem. Furthermore, in view of Theorem 2.2 the noncentrality parameter is maximal and thus the asymptotic relative efficiency is maximal. ■

**Proof of Theorem 3.1.** As before the denominator in  $g(\theta)$  converges in probability to  $|\Sigma|$ . Corresponding to (36) we now have

$$(1-L)^{d_k+\theta_k} u_{kt} = e_{kt} - \frac{\delta_k}{\sqrt{n}} \sum_{j=1}^{t-1} j^{-1} e_{k,t-j} + o_p(n^{-1/2}), \quad k = 1, \dots, K, \quad (38)$$

uniformly in  $t$ . Again we consider each term in the expansion of the numerator in  $g(\theta)$ ,

$$n |\Sigma(\theta)| - n |\Sigma(0)| = \sqrt{n} |\Sigma(0)| (\text{vec } \Sigma(0)^{-1})' \frac{\partial \text{vec } \Sigma(0)}{\partial \theta'} \delta \quad (39)$$

$$+ \frac{\delta'}{2} |\Sigma(0)| \left( I_K \otimes (\text{vec } \Sigma(0)^{-1})' \right) \frac{\partial}{\partial \theta'} \text{vec } \frac{\partial \text{vec } \Sigma(0)}{\partial \theta'} \delta \quad (40)$$

$$- \frac{\delta'}{2} |\Sigma(0)| \left( \frac{\partial \text{vec } \Sigma(0)}{\partial \theta'} \right)' \left( \Sigma(0)^{-1} \otimes \Sigma(0)^{-1} \right) \frac{\partial \text{vec } \Sigma(0)}{\partial \theta'} \delta \quad (41)$$

$$+ \frac{1}{2} \left( |\Sigma(0)| (\text{vec } \Sigma(0)^{-1})' \frac{\partial \text{vec } \Sigma(0)}{\partial \theta'} \delta \right)^2 + o_p(1). \quad (42)$$

It follows, using that  $\Sigma(0) \rightarrow_p \Sigma$  by (38), that (39) is

$$\begin{aligned} & - \frac{2}{\sqrt{n}} |\Sigma| (\text{vec } \Sigma^{-1})' \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} (\text{diag}(e_{t-j}) \otimes e_t) \delta + o_p(1) \\ &= - \frac{2}{\sqrt{n}} |\Sigma| \delta' \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} \text{diag}(e_{t-j}) \Sigma^{-1} e_t + o_p(1) \\ &= - \frac{2}{\sqrt{n}} |\Sigma| \delta' \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} J'_K (I_K \otimes \Sigma^{-1}) (e_{t-j} \otimes e_t) + o_p(1) \\ &= -2 |\Sigma| \delta' \sqrt{n} \sum_{j=1}^{n-1} j^{-1} J'_K (I_K \otimes \Sigma^{-1}) \text{vec } C(j) + o_p(1) \end{aligned}$$

by use of  $\text{vec}(ABC) = (C' \otimes A) \text{vec } B$  and property 1 of Lemma 2. By Lemma 1 and property 2 of Lemma 2 this converges in distribution to  $-2 |\Sigma| \delta' \sqrt{\pi^2/6} (\Sigma \odot \Sigma^{-1})^{1/2} Z$ . Next, (40) is

$$\frac{\delta'}{n} |\Sigma| \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} \left( \sum_{j'=1}^{t-j-1} j'^{-1} (e_t' \Sigma^{-1} \otimes I_K) J(e_{t-j-j'}) + \sum_{j'=1}^{t-1} j'^{-1} \text{diag}(e_{t-j}) \Sigma^{-1} \text{diag}(e_{t-j'}) \right) \delta + o_p(1),$$

where  $J(e_t)$  is a  $K^2 \times K$  matrix whose  $k$ 'th column has  $((k-1)K+k)$ 'th element  $e_{kt}$  and zeros otherwise. By uncorrelatedness of  $\{e_t\}$  we are left with

$$\delta' |\Sigma| \sum_{j=1}^{n-1} j^{-2} \frac{1}{n} \sum_{t=j+1}^n \text{diag}(e_{t-j}) \Sigma^{-1} \text{diag}(e_{t-j}) \delta + o_p(1) \rightarrow_p \delta' |\Sigma| \frac{\pi^2}{6} (\Sigma \odot \Sigma^{-1}) \delta$$

by property 3 of Lemma 2 and a law of large numbers. Because (41) and (42) can be shown to be negligible as in Theorem 2.1, we have shown (i) by using (37).

Next, we examine the first derivative of  $g(\theta)$  which equals

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} \text{diag} \left( e_{t-j} + \sum_{j'=1}^{t-j-1} j'^{-1} \text{diag}(e_{t-j-j'}) \frac{\delta}{\sqrt{n}} \right) \Sigma^{-1} \left( e_t + \sum_{j'=1}^{t-1} j'^{-1} \text{diag}(e_{t-j'}) \frac{\delta}{\sqrt{n}} \right) + o_p(1),$$

by (38). By uncorrelatedness of  $\{e_t\}$  this is just

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} \text{diag}(e_{t-j}) \Sigma^{-1} e_t + \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-2} \text{diag}(e_{t-j}) \Sigma^{-1} \text{diag}(e_{t-j}) \delta + o_p(1),$$

which establishes (ii) in view of the above.

The last statement follows in the same way. ■

**Proof of Theorem 3.2.** By Theorem 3.1 and the subsequent analysis,  $g(\theta)$  is asymptotically a concave function of  $\delta = \sqrt{n}\theta$  in a neighborhood of  $\theta_0$ , and  $\hat{\delta} = \sqrt{n}\hat{\theta}_n$  is asymptotically the unique maximizer of  $W_d(\delta)$  and is given by (14) in this neighborhood. Under Gaussianity the limiting Fisher information is given by

$$\lim_{n \rightarrow \infty} \frac{1}{n} E \left( - \frac{\partial^2 g(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=0} \right) = \mathcal{I}_d$$

by Theorem 3.1(iii). ■

**Proof of Theorem 3.3.** As before the asymptotic equivalence follows since the information matrix equality holds, c.f. Theorem 3.1. The distribution under local alternatives follows from Theorems 3.1 and 3.2, and in view of Theorem 3.2 the noncentrality parameter is maximal. ■

**Proof of Theorem 4.1.** Suppose  $\theta = \delta/\sqrt{n}$  and  $\psi - \psi_0 = \gamma/\sqrt{n}$ . Then, using that

$$\frac{\partial A(z)}{\partial A_{r,jk}} = -E_{jk} z^r, \quad (43)$$

where  $E_{jk}$  is a  $K \times K$  matrix with  $(j, k)$ 'th element one and all other elements zero, we get that

$$A(L)(1-L)^{d+\theta} u_t = \varepsilon_t - \frac{\delta}{\sqrt{n}} \sum_{j=1}^{t-1} j^{-1} \varepsilon_{t-j} - ((e'_{t-1}, \dots, e'_{t-p}) \otimes I_K) \frac{\gamma}{\sqrt{n}} + o_p(n^{-1/2}) \quad (44)$$

uniformly in  $t$ . This shows, in particular, that  $\Sigma(\theta, \psi) \rightarrow_p \Sigma$ .

Expand the numerator in  $g(\theta, \psi)$  as

$$n |\Sigma(\theta, \psi)| - n |\Sigma(0, \psi_0)| = \sqrt{n} |\Sigma(0, \psi_0)| (\text{vec } \Sigma(0, \psi_0)^{-1})' \frac{\partial \text{vec } \Sigma(0, \psi_0)}{\partial \tau'} \tau \quad (45)$$

$$+ \frac{\delta^2}{2} |\Sigma(0, \psi_0)| \text{tr} \left( \Sigma(0, \psi_0)^{-1} \frac{\partial^2 \Sigma(0, \psi_0)}{\partial \theta^2} \right) \quad (46)$$

$$+ \frac{\gamma'}{2} |\Sigma(0, \psi_0)| (I_K \otimes (\text{vec } \Sigma(0, \psi_0)^{-1})') \frac{\partial}{\partial \gamma'} \text{vec} \frac{\partial \text{vec } \Sigma(0, \psi_0)}{\partial \gamma'} \gamma \quad (47)$$

$$+ \delta |\Sigma(0, \psi_0)| (\text{vec } \Sigma(0, \psi_0)^{-1})' \frac{\partial^2 \text{vec } \Sigma(0, \psi_0)}{\partial \theta \partial \gamma'} \gamma + o_p(1). \quad (48)$$

The terms corresponding to (41) and (42) have been merged into the  $o_p(1)$  term as they continue to be asymptotically negligible. First, (45) is

$$\begin{aligned} & -\sqrt{n} |\Sigma| \frac{2}{n} \sum_{t=1}^n \left( \sum_{j=1}^{t-1} j^{-1} (\text{vec } \Sigma^{-1})' (\varepsilon_{t-j} \otimes \varepsilon_t), (\text{vec } \Sigma^{-1})' ((e'_{t-1}, \dots, e'_{t-p}) \otimes I_K \otimes \varepsilon_t) \right) \tau + o_p(1) \\ &= -|\Sigma| \tau' \frac{2}{\sqrt{n}} \sum_{t=1}^n \left( \begin{array}{c} \sum_{j=1}^{t-1} j^{-1} (\varepsilon'_{t-j} \otimes \varepsilon'_t) \text{vec } \Sigma^{-1} \\ ((e'_{t-1}, \dots, e'_{t-p})' \otimes I_K \otimes \varepsilon'_t) \text{vec } \Sigma^{-1} \end{array} \right) + o_p(1) \\ &= -|\Sigma| \tau' \frac{2}{\sqrt{n}} \sum_{t=1}^n \left( \begin{array}{c} \sum_{j=1}^{t-1} j^{-1} \varepsilon'_t \Sigma^{-1} \varepsilon_{t-j} \\ (e'_{t-1}, \dots, e'_{t-p})' \otimes \Sigma^{-1} \varepsilon_t \end{array} \right) + o_p(1) \\ &= -|\Sigma| \tau' \frac{2}{\sqrt{n}} \sum_{t=1}^n \left( \begin{array}{c} \sum_{j=1}^{t-1} j^{-1} (\text{vec } \Sigma^{-1})' \text{vec}(\varepsilon_t \varepsilon'_{t-j}) \\ \text{vec}(\Sigma^{-1} \varepsilon_t (e'_{t-1}, \dots, e'_{t-p})) \end{array} \right) + o_p(1), \end{aligned}$$

which converges in distribution to  $-2|\Sigma| \tau' \Xi^{1/2} Z$  using that  $e_{t-i} = \sum_{j=i}^{t-i-1} a_{j-i} \varepsilon_{t-j}$  and applying Lemma 1. Next, (46) converges in probability to  $\delta^2 |\Sigma| \pi^2 K/6$  as in the proof of Theorem 2.1, and (47) is

$$\begin{aligned} & \gamma' |\Sigma| \frac{1}{n} \sum_{t=1}^n \left( (e'_{t-1}, \dots, e'_{t-p})' \otimes \Sigma^{-1} \right) \left( (e'_{t-1}, \dots, e'_{t-p}) \otimes I_K \right) \gamma \\ & \quad + \gamma' |\Sigma| \frac{1}{n} \sum_{t=1}^n \left( \begin{array}{c} (e'_{t-2}, \dots, e'_{t-p-1}) \otimes I_K \otimes \Sigma^{-1} \varepsilon_t \\ \vdots \\ (e'_{t-p-1}, \dots, e'_{t-2p}) \otimes I_K \otimes \Sigma^{-1} \varepsilon_t \end{array} \right) \gamma + o_p(1), \end{aligned}$$

where the first term converges in probability to  $\gamma' |\Sigma| (\Gamma \otimes \Sigma^{-1}) \gamma$  by definition of  $\Gamma$  and a law of large numbers and the second term is negligible by uncorrelatedness of  $\{\varepsilon_t\}$ . Lastly, (48) is

$$\begin{aligned} & 2\delta |\Sigma| (\text{vec } \Sigma^{-1})' \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} ((e'_{t-j-1}, \dots, e'_{t-j-p}) \otimes I_K \otimes \varepsilon_t) \gamma \\ & \quad + 2\delta |\Sigma| (\text{vec } \Sigma^{-1})' \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^{t-1} j^{-1} (\varepsilon_{t-j} \otimes (e'_{t-1}, \dots, e'_{t-p}) \otimes I_K) \gamma + o_p(1), \end{aligned}$$

where the first term is asymptotically negligible by uncorrelatedness of  $\{\varepsilon_t\}$  and the second term is

$$2\delta\gamma'|\Sigma|\frac{1}{n}\sum_{t=1}^n\sum_{j=1}^{t-1}j^{-1}\text{vec}\left(\Sigma^{-1}\varepsilon_{t-j}\left(e'_{t-1},\dots,e'_{t-p}\right)\right)\rightarrow_p 2\delta\gamma'|\Sigma|\text{vec}\Phi'$$

using  $e_{t-i}=\sum_{j=i}^{t-i-1}a_{j-i}\varepsilon_{t-j}$ .

Thus, we have shown that  $g(\theta,\psi)$  is asymptotically a concave function of  $\delta$  and  $\gamma$  in a neighborhood of  $\tau_0$ , and  $\hat{\theta}$  and  $\hat{\psi}$  are asymptotically given by (21). The remainder of the theorem follows as in Theorems 2.1-2.3 in view of the above. ■

**Proof of Theorem 4.2.** Combine the arguments of the previous theorems using the properties of Lemma 2 as in the proof of Theorem 3.1. ■

**Proof of Theorem 5.1.** Follows straightforwardly from Theorems 3.3 and 4.2. ■

## Appendix B: Technical Lemmas

Define the (white noise) autocovariance function

$$C(j)=\frac{1}{n}\sum_{t=j+1}^n\varepsilon_t\varepsilon'_{t-j}$$

where  $\{\varepsilon_t\}$  is a mean zero *i.i.d.* sequence with covariance matrix  $\Sigma$  and finite fourth moments. We consider the asymptotic distribution of a particular linear combination of the autocovariances in the following lemma.

**Lemma 1** *Let  $\{\varepsilon_t\}$  be *i.i.d.*  $(0,\Sigma)$  with finite fourth moments. Then*

$$\sqrt{n}\sum_{j=1}^{n-1}j^{-1}\text{vec}C(j)\rightarrow_d N\left(0,\frac{\pi^2}{6}\Sigma\otimes\Sigma\right)$$

as  $n\rightarrow\infty$ .

**Proof.** For a fixed  $m>0$  define the  $K^2m$ -vector  $C_m=((\text{vec}C(1))',\dots,(\text{vec}C(m))')'$ . It is well known that

$$\sqrt{n}C_m\rightarrow_d N(0,I_m\otimes\Sigma\otimes\Sigma),$$

and thus

$$\sqrt{n}\sum_{j=1}^mj^{-1}\text{vec}C(j)\rightarrow_d N\left(0,\sum_{j=1}^mj^{-2}\Sigma\otimes\Sigma\right).$$



The desired convergence now follows by application of Bernstein's Lemma, see e.g. Hall & Heyde (1980, pp. 191-192). ■

The next lemma derives some properties of the Hadamard product, which is defined for two  $m \times n$  matrices  $A = (a_{ij})$  and  $B = (b_{ij})$  as

$$A \odot B = (a_{ij}b_{ij}),$$

see e.g. Magnus & Neudecker (1999, Chapter 3.6) for more details. The proof of the lemma is easy and is omitted.

**Lemma 2** *Property 1. There exists a  $K^2 \times K$  matrix  $J_K := (\text{vec } E_{11}, \dots, \text{vec } E_{KK})$ ,  $E_{ii} = e_i e_i'$  where  $e_i$  is the  $i$ 'th unit  $K$ -vector, such that for any  $K \times K$  matrix  $A$*

$$J_K' \text{vec } A = a,$$

where  $a$  is the  $K$ -vector holding the diagonal of  $A$ . If  $A_d := I_K \odot A$  is the diagonal matrix obtained from  $A$  then

$$\text{vec } A_d = J_K a.$$

*Property 2. Connection with the Kronecker product. For all  $K \times K$  matrices  $A$  and  $B$ ,*

$$J_K' (A \otimes B) J_K = A \odot B$$

where  $J_K$  is defined as in property 1.

*Property 3. Let  $A$  and  $B$  be  $K \times K$  matrices such that  $A$  is diagonal and  $B$  is symmetric. Then*

$$ABA = aa' \odot B$$

where  $a$  is defined as in property 1.

## References

- Agiakloglou, C. & Newbold, P. (1994), 'Lagrange multiplier tests for fractional difference', *Journal of Time Series Analysis* **15**, 253–262.
- Beran, J. (1995), 'Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models', *Journal of the Royal Statistical Society Series B* **57**, 659–672.

- Breitung, J. & Hassler, U. (2002), ‘Inference on the cointegration rank in fractionally integrated processes’, *Journal of Econometrics* **110**, 167–185.
- Choi, I. & Ahn, B. C. (1999), ‘Testing the null of stationarity for multiple time series’, *Journal of Econometrics* **88**, 41–77.
- Dolado, J. J., Gonzalo, J. & Mayoral, L. (2002), ‘A fractional dickey-fuller test for unit roots’, *Econometrica* **70**, 1963–2006.
- Doornik, J. A. (2001), *Ox: An Object-Oriented Matrix Language*, 4th edn, Timberlake Consultants Press, London.
- Doornik, J. A. & Ooms, M. (2001), ‘A package for estimating, forecasting and simulating arfima models: Arfima package 1.01 for Ox’, *Working Paper, Nuffield College, Oxford*.
- Hall, P. & Heyde, C. C. (1980), *Martingale Limit Theory and its Application*, Academic Press, New York.
- Hosking, J. R. M. (1980), ‘The multivariate portmanteau statistic’, *Journal of the American Statistical Association* **75**, 602–608.
- Hosoya, Y. (1996), ‘The quasi-likelihood approach to statistical inference on multiple time-series with long-range dependence’, *Journal of Econometrics* **73**, 217–236.
- Jeganathan, P. (1999), ‘On asymptotic inference in cointegrated time series with fractionally integrated errors’, *Econometric Theory* **15**, 583–621.
- Johansen, S. (1988), ‘Statistical analysis of cointegration vectors’, *Journal of Economic Dynamics and Control* **12**, 231–254.
- Ling, S. & Li, W. K. (2001), ‘Asymptotic inference for nonstationary fractionally integrated autoregressive moving-average models’, *Econometric Theory* **17**, 738–764.
- Lobato, I. N. (1997), ‘Consistency of the averaged cross-periodogram in long memory series’, *Journal of Time Series Analysis* **18**, 137–155.
- Magnus, J. R. & Neudecker, H. (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edn, John Wiley and Sons, New York.
- Marinucci, D. & Robinson, P. M. (1999), ‘Alternative forms of fractional Brownian motion’, *Journal of statistical planning and inference* **80**, 111–122.

- Nielsen, M. Ø. (2001), ‘Efficient likelihood inference in nonstationary univariate models’, *Department of Economics Working Paper 2001-08 (revised 2002), University of Aarhus* .
- Nielsen, M. Ø. (2002), ‘Optimal residual based tests for fractional cointegration and exchange rate dynamics’, *Department of Economics Working Paper 2002-7, University of Aarhus* .
- Phillips, P. C. B. & Durlauf, S. N. (1986), ‘Multiple time series regression with integrated processes’, *Review of Economic Studies* **53**, 473–495.
- Robinson, P. M. (1991), ‘Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regressions’, *Journal of Econometrics* **47**, 67–84.
- Robinson, P. M. (1994), ‘Efficient tests of nonstationary hypotheses’, *Journal of the American Statistical Association* **89**, 1420–1437.
- Sargan, J. D. & Bhargava, A. (1983), ‘Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle’, *Econometrica* **51**, 799–820.
- Tanaka, K. (1999), ‘The nonstationary fractional unit root’, *Econometric Theory* **15**, 549–582.
- Watson, M. W. (1994), Vector autoregressions and cointegration, in R. F. Engle & D. L. McFadden, eds, ‘Handbook of Econometrics, Vol. IV’, North-Holland, Amsterdam, chapter 47, pp. 2843–2915.

Table 1: Simulation results for Model A.

		$n = 100$													
		$\rho = 0$					$\rho = 0.6$								
$\theta_1 \setminus \theta_2$		-0.3	-0.2	-0.1	0	0.1	0.2	0.3	-0.3	-0.2	-0.1	0	0.1	0.2	0.3
-0.3		0.9969 (0.9991)	0.9785 (0.9900)	0.9240 (0.9615)	0.8910 (0.9412)	0.9268 (0.9615)	0.9752 (0.9900)	0.9961 (0.9991)	0.9970 (0.9991)	0.9790 (0.9923)	0.9634 (0.9858)	0.9814 (0.9939)	0.9971 (0.9995)	0.9998 (1.0000)	1.0000 (1.0000)
-0.2		0.9772 (0.9900)	0.8626 (0.9115)	0.6410 (0.7303)	0.5305 (0.6266)	0.6654 (0.7303)	0.8766 (0.9115)	0.9766 (0.9900)	0.9834 (0.9923)	0.8621 (0.9115)	0.7041 (0.7779)	0.7472 (0.8266)	0.9299 (0.9625)	0.9931 (0.9986)	0.9996 (1.0000)
-0.1		0.9247 (0.9615)	0.6554 (0.7303)	0.3030 (0.3491)	0.1665 (0.1919)	0.3257 (0.3491)	0.7144 (0.7303)	0.9393 (0.9615)	0.9674 (0.9858)	0.6909 (0.7779)	0.3080 (0.3491)	0.2425 (0.2803)	0.6166 (0.6548)	0.9322 (0.9625)	0.9927 (0.9995)
0		0.8877 (0.9412)	0.5221 (0.6266)	0.1622 (0.1919)	<b>0.0522</b> (0.0500)	0.2135 (0.1919)	0.6360 (0.6266)	0.9093 (0.9412)	0.9799 (0.9939)	0.7523 (0.8266)	0.2435 (0.2803)	<b>0.0511</b> (0.0500)	0.2981 (0.2803)	0.8083 (0.8266)	0.9806 (0.9939)
0.1		0.9268 (0.9615)	0.6582 (0.7303)	0.3317 (0.3491)	0.2043 (0.1919)	0.3710 (0.3491)	0.7154 (0.7303)	0.9352 (0.9615)	0.9970 (0.9995)	0.9273 (0.9625)	0.6150 (0.6548)	0.2976 (0.2803)	0.3622 (0.3491)	0.7616 (0.7779)	0.9671 (0.9858)
0.2		0.9779 (0.9900)	0.8704 (0.9115)	0.7154 (0.7303)	0.6327 (0.6266)	0.7230 (0.7303)	0.8881 (0.9115)	0.9781 (0.9900)	0.9997 (1.0000)	0.9930 (0.9986)	0.9358 (0.9625)	0.8061 (0.8266)	0.7557 (0.7779)	0.8938 (0.9115)	0.9809 (0.9923)
0.3		0.9971 (0.9991)	0.9798 (0.9900)	0.9388 (0.9615)	0.9141 (0.9412)	0.9387 (0.9615)	0.9765 (0.9900)	0.9958 (0.9991)	1.0000 (1.0000)	0.9999 (1.0000)	0.9958 (0.9995)	0.9788 (0.9939)	0.9693 (0.9858)	0.9802 (0.9923)	0.9960 (0.9991)

  

		$n = 250$													
		$\rho = 0$					$\rho = 0.6$								
$\theta_1 \setminus \theta_2$		-0.3	-0.2	-0.1	0	0.1	0.2	0.3	-0.3	-0.2	-0.1	0	0.1	0.2	0.3
-0.3		1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	0.9999 (0.9999)	0.9996 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)
-0.2		1.0000 (1.0000)	0.9994 (0.9997)	0.9800 (0.9874)	0.9435 (0.9615)	0.9830 (0.9874)	0.9993 (0.9997)	1.0000 (1.0000)	1.0000 (1.0000)	0.9996 (0.9997)	0.9902 (0.9934)	0.9942 (0.9971)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)
-0.1		0.9999 (1.0000)	0.9804 (0.9874)	0.6921 (0.7303)	0.3701 (0.4257)	0.7079 (0.7303)	0.9806 (0.9874)	0.9998 (1.0000)	1.0000 (1.0000)	0.9900 (0.9934)	0.6917 (0.7303)	0.5731 (0.6156)	0.9586 (0.9707)	0.9996 (1.0000)	1.0000 (1.0000)
0		0.9999 (0.9999)	0.9411 (0.9615)	0.3704 (0.4257)	<b>0.0521</b> (0.0500)	0.4599 (0.4257)	0.9468 (0.9615)	0.9997 (0.9999)	1.0000 (1.0000)	0.9936 (0.9971)	0.5683 (0.6156)	<b>0.0535</b> (0.0500)	0.6284 (0.6156)	0.9920 (0.9971)	1.0000 (1.0000)
0.1		1.0000 (1.0000)	0.9804 (0.9874)	0.7118 (0.7303)	0.4402 (0.4257)	0.7323 (0.7303)	0.9800 (0.9874)	0.9999 (1.0000)	1.0000 (1.0000)	0.9997 (1.0000)	0.9570 (0.9706)	0.6269 (0.6156)	0.7360 (0.7303)	0.9871 (0.9934)	1.0000 (1.0000)
0.2		1.0000 (1.0000)	0.9993 (0.9997)	0.9802 (0.9874)	0.9499 (0.9615)	0.9811 (0.9874)	0.9993 (0.9997)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	0.9997 (1.0000)	0.9921 (0.9971)	0.9886 (0.9934)	0.9994 (0.9997)	1.0000 (1.0000)
0.3		1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	0.9998 (0.9999)	0.9999 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)	0.9998 (1.0000)	1.0000 (1.0000)	1.0000 (1.0000)

Table 2: Simulation results for Model B with  $a = 0.5$ .

		$n = 100$													
		$\rho = 0$					$\rho = 0.6$								
$\theta_1 \setminus \theta_2$		-0.3	-0.2	-0.1	0	0.1	0.2	0.3	-0.3	-0.2	-0.1	0	0.1	0.2	0.3
-0.3		0.3777 (0.3844)	0.2555 (0.2876)	0.1977 (0.2296)	0.1872 (0.2096)	0.2061 (0.2290)	0.2394 (0.2876)	0.2858 (0.3844)	0.4178 (0.3844)	0.3064 (0.2986)	0.2871 (0.2729)	0.3199 (0.3084)	0.3520 (0.4033)	0.3809 (0.5457)	0.4076 (0.7051)
-0.2		0.2612 (0.2876)	0.1462 (0.1904)	0.1050 (0.1343)	0.0973 (0.1164)	0.1124 (0.1343)	0.1452 (0.1904)	0.1788 (0.2876)	0.3081 (0.2986)	0.1859 (0.1904)	0.1460 (0.1446)	0.1635 (0.1573)	0.2105 (0.2302)	0.2645 (0.3653)	0.2937 (0.5457)
-0.1		0.1919 (0.2290)	0.1000 (0.1343)	0.0647 (0.0819)	0.0586 (0.0656)	0.0741 (0.0819)	0.1015 (0.1343)	0.1237 (0.2290)	0.2844 (0.2729)	0.1446 (0.1446)	0.0914 (0.0819)	0.0900 (0.0747)	0.1247 (0.1208)	0.1717 (0.2302)	0.2237 (0.4033)
0		0.1870 (0.2096)	0.0990 (0.1164)	0.0636 (0.0656)	<b>0.0570</b> (0.0500)	0.0635 (0.0656)	0.0873 (0.1164)	0.1039 (0.2096)	0.3123 (0.3084)	0.1644 (0.1573)	0.0881 (0.0747)	<b>0.0741</b> (0.0500)	0.0891 (0.0747)	0.1289 (0.1573)	0.1749 (0.3084)
0.1		0.2095 (0.2290)	0.1073 (0.1343)	0.0694 (0.0819)	0.0672 (0.0656)	0.0729 (0.0819)	0.0933 (0.1343)	0.1052 (0.2290)	0.3430 (0.4033)	0.2117 (0.2302)	0.1258 (0.1208)	0.0876 (0.0747)	0.0954 (0.0819)	0.1187 (0.1446)	0.1540 (0.2729)
0.2		0.2382 (0.2876)	0.1439 (0.1904)	0.0994 (0.1343)	0.0907 (0.1164)	0.0884 (0.1343)	0.0983 (0.1904)	0.1230 (0.2876)	0.4008 (0.5457)	0.2558 (0.3653)	0.1796 (0.2302)	0.1354 (0.1573)	0.1183 (0.1446)	0.1235 (0.1904)	0.1487 (0.2986)
0.3		0.2750 (0.3844)	0.1813 (0.2876)	0.1296 (0.2290)	0.1108 (0.2096)	0.1133 (0.2290)	0.1198 (0.2876)	0.1358 (0.3844)	0.4677 (0.7051)	0.3282 (0.5457)	0.2303 (0.4033)	0.1752 (0.3084)	0.1557 (0.2729)	0.1545 (0.2986)	0.1585 (0.3844)

  

		$n = 250$													
		$\rho = 0$					$\rho = 0.6$								
$\theta_1 \setminus \theta_2$		-0.3	-0.2	-0.1	0	0.1	0.2	0.3	-0.3	-0.2	-0.1	0	0.1	0.2	0.3
-0.3		0.8779 (0.7784)	0.7321 (0.6291)	0.6157 (0.5112)	0.5596 (0.4675)	0.5981 (0.5112)	0.6634 (0.6291)	0.7330 (0.7784)	0.8752 (0.7784)	0.7558 (0.6489)	0.7024 (0.6016)	0.7407 (0.6658)	0.7947 (0.8015)	0.8369 (0.9239)	0.8565 (0.9829)
-0.2		0.7282 (0.6291)	0.4796 (0.4220)	0.3130 (0.2778)	0.2603 (0.2290)	0.2990 (0.2778)	0.4012 (0.4220)	0.5107 (0.6291)	0.7537 (0.6489)	0.5032 (0.4220)	0.3616 (0.3053)	0.3819 (0.3388)	0.5030 (0.5139)	0.6109 (0.7532)	0.6932 (0.9239)
-0.1		0.6019 (0.5112)	0.3155 (0.2778)	0.1507 (0.1343)	0.1059 (0.0903)	0.1466 (0.1343)	0.2475 (0.2778)	0.3594 (0.5112)	0.6987 (0.6016)	0.3560 (0.3053)	0.1781 (0.1343)	0.1489 (0.1147)	0.2462 (0.2411)	0.4009 (0.5139)	0.5407 (0.8015)
0		0.5560 (0.4675)	0.2583 (0.2290)	0.1058 (0.0903)	<b>0.0694</b> (0.0500)	0.1007 (0.0903)	0.1931 (0.2290)	0.3021 (0.4675)	0.7401 (0.6658)	0.3880 (0.3388)	0.1433 (0.1147)	<b>0.0838</b> (0.0500)	0.1346 (0.1147)	0.2693 (0.3388)	0.4350 (0.6658)
0.1		0.5901 (0.5112)	0.3055 (0.2778)	0.1427 (0.1343)	0.1061 (0.0903)	0.1287 (0.1343)	0.2103 (0.2778)	0.3129 (0.5112)	0.7834 (0.8015)	0.4985 (0.5139)	0.2534 (0.2411)	0.1342 (0.1147)	0.1414 (0.1343)	0.2481 (0.3053)	0.3906 (0.6016)
0.2		0.6622 (0.6291)	0.4040 (0.4220)	0.2451 (0.2778)	0.1918 (0.2290)	0.2112 (0.2778)	0.2977 (0.4220)	0.3865 (0.6291)	0.8348 (0.9239)	0.6152 (0.7532)	0.3992 (0.5139)	0.2749 (0.3388)	0.2451 (0.3053)	0.3075 (0.4220)	0.4150 (0.6489)
0.3		0.7389 (0.7784)	0.5177 (0.6291)	0.3496 (0.5112)	0.3078 (0.4675)	0.3334 (0.5112)	0.4125 (0.6291)	0.4962 (0.7784)	0.8753 (0.9829)	0.6977 (0.9239)	0.5442 (0.8015)	0.4327 (0.6658)	0.3946 (0.6016)	0.4243 (0.6489)	0.4925 (0.7784)

Table 3: Simulated size of nominal 5% test for Model B.

$n$	$\rho \setminus a$	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
100	0	0.0591	0.0613	0.0571	0.0672	0.0759	0.0825	0.0644	0.0483	0.0565
	0.6	0.0558	0.0564	0.0629	0.0622	0.0736	0.0899	0.0814	0.0677	0.0779
250	0	0.0544	0.0583	0.0519	0.0538	0.0578	0.0771	0.0804	0.0570	0.0599
	0.6	0.0546	0.0526	0.0561	0.0599	0.0564	0.0707	0.1023	0.0684	0.0692
500	0	0.0513	0.0533	0.0498	0.0530	0.0551	0.0601	0.0778	0.570	0.0637
	0.6	0.0523	0.0512	0.0526	0.0537	0.0557	0.0566	0.0822	0.0707	0.0623

## Working Paper

- 2001-5: Bo Sandemann Rasmussen: Efficiency Wages and the Long-Run Incidence of Progressive Taxation.
- 2001-6: Boriss Siliverstovs: Multicointegration in US consumption data.
- 2001-7: Jakob Roland Munch and Michael Svarer: Rent Control and Tenancy Duration.
- 2001-8: Morten Ø. Nielsen: Efficient Likelihood Inference in Non-stationary Univariate Models.
- 2001-9: Effrosyni Diamantoudi: Stable Cartels Revisited.
- 2001-16: Bjarne Brendstrup, Svend Hylleberg, Morten Nielsen, Lars Skipper and Lars Stentoft: Seasonality in Economic Models.
- 2001-17: Martin Paldam: The Economic Freedom of Asian Tigers - an essay on controversy.
- 2001-18: Celso Brunetti and Peter Lildholt: Range-based covariance estimation with a view to foreign exchange rates.
- 2002-1: Peter Jensen, Michael Rosholm and Mette Verner: A Comparison of Different Estimators for Panel Data Sample Selection Models.
- 2002-2: Torben M. Andersen: International Integration and the Welfare State.
- 2002-3: Bo Sandemann Rasmussen: Credibility, Cost of Reneging and the Choice of Fixed Exchange Rate Regime.
- 2002-4: Bo William Hansen and Lars Mayland Nielsen: Can Nominal Wage and Price Rigidities Be Equivalent Propagation Mechanisms? The Case of Open Economies.
- 2002-5: Anna Christina D'Addio and Michael Rosholm: Left-Censoring in Duration Data: Theory and Applications.
- 2002-6: Morten Ørregaard Nielsen: Efficient Inference in Multivariate Fractionally Integration Models.