A unified framework for testing in the linear regression model under unknown order of fractional integration*

Bent Jesper Christensen[†], Robinson Kruse[‡] and Philipp Sibbertsen[§]

October 22, 2013

Abstract

We consider hypothesis testing in a general linear time series regression framework when the possibly fractional order of integration of the error term is unknown. We show that the approach suggested by Vogelsang (1998a) for the case of integer integration does not apply to the case of fractional integration. We propose a Lagrange Multiplier-type test whose limiting distribution is independent of the order of integration of the errors. Different testing scenarios for the case of deterministic and stochastic regressors are considered. Simulations demonstrate that the proposed test works well for a variety of different cases, thereby emphasizing its generality.

Key words: Long memory; linear time series regression; Lagrange Multiplier test.

JEL classification: C12 (Hypothesis Testing), C22 (Time-Series Models)

*We are grateful to the participants of the Workshop on Long Memory Time Series in Aarhus, the Workshop on Nonlinear and Persistent Time Series in Hannover, the Workshop in Honor of James Davidson in Exeter, the European Meeting of Statisticians in Budapest and the ESEM in Gothenburg for helpful comments. We are especially grateful to Karim Abadir, Richard Baillie, Jörg Breitung, Matei Demetrescu, David Hendry and Peter Phillips for fruitful discussions. Financial support of the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. Bent Jesper Christensen and Robinson Kruse gratefully acknowledges financial support from CREATES funded by the Danish National Research Foundation. Part of this research was carried out while Philipp Sibbertsen was a Guest Professor at CREATES. Their kind hospitality is greatly appreciated.

†CREATES, Aarhus University, School of Business and Social Sciences, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark and Aarhus Institute of Advanced Studies, Høegh-Guldbergs Gade 6B, DK-8000 Aarhus C. E-mail address: `bjchristensen@creates.au.dk`.

‡CREATES, Aarhus University, School of Business and Social Sciences, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. E-mail address: `rkruse@creates.au.dk`. Leibniz University Hannover, School of Economics and Management, Institute for Statistics, Königsworther Platz 1, D-30167 Hannover, Germany. E-mail address: `kruse@statistik.uni-hannover.de`.

§Leibniz University Hannover, School of Economics and Management, Institute for Statistics, Königsworther Platz 1, D-30167 Hannover, Germany. E-mail address: `sibbertsen@statistik.uni-hannover.de`

1 Introduction

This paper considers the problem of hypothesis testing in a linear regression model when the order of possibly fractional integration of the error term is unknown. Specifically, we consider the linear regression model

$$y_t = \beta' z_t + e_t,$$

where z_t is a $p \times 1$ vector of deterministic or stochastic (or both) regressors fulfilling some regularity conditions specified later, β is a $p \times 1$ vector of regression parameters and e_t is an error term which is possibly fractionally integrated of unknown order d_0, i.e. $e_t \sim I(d_0)$. In the case of stochastic regressors, their orders of integration are assumed to be the same (d_1) and unknown as well. We are concerned with testing hypotheses for β. It is well known that the behavior of standard tests for restrictions on β strongly depend on the order of integration of the error term. Neglecting the fractional nature of the errors leads typically to invalid inference.

Vogelsang (1998a) suggests a testing approach which allows valid inference to be drawn when the order of integration of the error term is unknown. The test is embedded in a linear trend testing framework where the errors are allowed to be $I(0)$ or $I(1)$. Other contributions cover the case of testing for location (Zambrano and Vogelsang, 2000), testing against a mean shift (Vogelsang, 1998b) and nonlinear trends (Harvey, Leybourne and Xiao, 2010). Regarding test against trend breaks, Harvey, Leybourne and Taylor (2009) and Perron and Yabu (2009) suggest alternative approaches. All of the aforementioned studies are based on having a choice of two possible orders of integration of which one is the true order. In the fractional case we have a continuum of possible orders of integration which renders methods designed for the $I(0)$ and $I(1)$ case problematic. Another important feature of these tests is that they are dealing with deterministic regressors.

In order to achieve the goal of conducting valid inference on β in linear regression models with deterministic or stochastic regressors with $I(d)$ errors, we suggest a Lagrange multiplier (LM) test. The test is derived under fairly mild regularity assumptions ensuring flexibility of our method. In a similar regression framework Robinson (1994) suggests an LM test for the order of integration d. Thereby, our work complements Robinson (1994). Our test exhibits a standard limiting distribution and is shown to have non-trivial power against local alternatives. These results remain valid, when an LM-type test is carried out where the memory parameter is estimated in a semi-parametric way, by e.g. exact local Whittle, see Shimotsu and Phillips (2005). The reason for applying semi-parametric estimation arises from the need of an estimator for the fractional integration parameter which is consistent under validity of the null and the alternative hypothesis. Biased estimation would lead to size distortions and power losses in finite samples. To this end, estimation is carried out under the alternative hypothesis. Simulation results for various situations, i.e. linear and quadratic trends, mean shifts and trend breaks, as well as fractional cointegration, shed light on the empirical properties of the test. The results demonstrate that the LM-type test performs well in practice.

The problem of testing under an unknown order of possibly fractional integration has recently been investigated elsewhere in several specific testing situations involving deterministic regressors. Iacone, Leybourne and Taylor (2013a) propose a sup-Wald test against a permanent change in the trend slope. A fixed-b perspective is taken in Iacone, Leybourne and Taylor (2013b). Xu and Perron (2013) suggest a quasi GLS-based trend test under fractional integration of the errors. A common feature of these studies is the use of semi-parametric plug-in estimation for the fractional integration parameter.

The rest of the paper is organized as follows. Section 2 focuses on the properties of the Vogelsang (1998a) approach under fractional integration. Section 3 contains our LM-type test for deterministic and stochastic regressors. Section 4 contains some examples including trend tests and fractional cointegration and a Monte Carlo study investigating the finite sample properties of the test. Section 5 concludes. All proofs are gathered in Appendix A. Appendices B and C cover additional technical details.

For the rest of the paper, let $\Rightarrow$ denote weak convergence and $\xrightarrow{P}$ stand for convergence in probability.

2 Testing under an unknown order of integration

We consider the linear regression model

$$y_t = \beta' z_t + e_t \tag{1}$$

for which the following Assumptions 1, 2 and 3 hold. The parameter vector $\beta = (\beta_1, \ldots, \beta_p)'$ is p-dimensional and z_t is a $p \times 1$ vector of exogenous regressors. The error term e_t is assumed to be a type II fractionally integrated process (see Marinucci and Robinson, 1999), i.e.

$$(1-B)^{d_0} e_t = u_t 1(t \geq 0), \tag{2}$$

where u_t is an $I(0)$ process satisfying Assumption 1 below, d_0 is the unknown order of fractional integration and $(1-B)^d$ is defined by its binomial expansion

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} B^j \tag{3}$$

with $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$. B denotes the Backshift operator, $Be_t = e_{t-1}$. Our asymptotic results hold also for type I fractionally integrated processes with the respective modifications in the proofs. The main insights of our paper remain unaffected when working with type II processes. Let us further assume that $d_0 \in [0, 1/2) \cup (1/2, 3/2)$ which covers the relevant values for economic applications.

Assumption 1 *u_t has the Wold representation*

$$u_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}, \tag{4}$$

with $\sum_{j=0}^{\infty} \|\theta_j\|^2 < \infty$ and the innovations ϵ_t are assumed to be a martingale difference sequence satisfying $E(\epsilon_t|\mathfrak{F}_{t-1}) = 0$ and $E(\epsilon_t^2|\mathfrak{F}_{t-1}) < \infty$ with $\mathfrak{F}_t = \sigma(\{\epsilon_s, s \leq t\})$. Furthermore, we assume that $u_t = 0$ for $t \leq 0$.

The last part of the assumption is made in accordance with our choice to work with type II processes. A detailed discussion of initial conditions in non-stationary long memory time series can be found in Johansen and Nielsen (2013).

Assumption 2 *Let $\phi(B; \theta)$ be a prescribed function of the Backshift operator B fulfilling*

$$\phi(B; \theta)e_t = u_t$$

with u_t satisfying Assumption 1, and $\phi(B; \theta)$ is a function of the lag operator B and possible short memory and heteroskedasticity parameterized by θ, satisfying the normalization $\phi(0; \theta) = 1$. The long memory parameter d_0 is included in θ, and the derivative $\partial\phi(.; \theta)/\partial\theta$ exists and is bounded. Furthermore, the derivative fulfills

$$\frac{1}{T}\sum_{t=1}^{T} \frac{\partial\phi(B;\theta)}{\partial\theta} e_t \left(\frac{\partial\phi(B;\theta)}{\partial\theta} e_t\right)' \to A, \tag{5}$$

with A being a non-singular matrix.

As this section covers only deterministic regressors, the next assumption deals with non-stochastic regressors only, while a modified assumption for the stochastic regressor case is stated in Section 3.2 where we treat this situation in detail.

Assumption 3 *The regressors z_t are non-stochastic, with*

$$\frac{1}{T}\sum_{t=1}^{T} \phi(B, \theta)z_t\phi(B, \theta)z_t' \to J, \tag{6}$$

where $J > 0$ is a positive definite $p \times p$ matrix, with trace $tr(J) = \xi^2 > 0$.

Since by implication $\lim_{T\to\infty}(\sum_{t=1}^{T}\phi(B,\theta)z_t\phi(B,\theta)z_t')^{-1}=0$, the condition is sufficient for consistency of the OLS estimator. It is rather weak and allows for instance for polynomial trends, i.e. $z_t=(1,t,t^2,...,t^{p-1})'$.

We are interested in testing some set of restrictions on β which are either linear or non-linear, i.e. $R\beta=r$ or $f(\beta)=r$. For simplicity, we reduce the testing problem to $H_0:\beta=0$ against the alternative $H_1:\beta\neq 0$ without loss of generality (see Appendix C).

In the context of testing for the presence of a linear trend ($p=2$), Vogelsang (1998a) suggested a general testing approach which allows valid inference in the case where the order of integration of the errors is unknown.[1] The order of integration is assumed to be either zero or one.

The basic idea of Vogelsang's approach is the following, see for example Harvey and Leybourne (2007) for an application to linearity tests against smooth transition autoregressive models. Suppose that one is interested in the testing problem of H_0 against the alternative H_1. The relevant test statistic is denoted by W which has a non-degenerate and pivotal limiting distribution, say D_0 if the time series under test is integrated of order null. In typical applications where standard testing problems are considered, D_0 is a standard distribution. The limiting distribution of W generally differs from D_0 if the series is integrated of order one instead. The limiting distribution for this case is denoted by D_1 and typically depends on functionals of Brownian motions. In order to design a test statistic which is independent of the order of integration let H denote a function of a unit root test statistic against the stationary alternative. It is assumed that H has a non-degenerate pivotal limiting distribution, say D_H when the series is non-stationary and that it converges in probability to zero under stationarity of the series. An example for H is the inverse of a Dickey-Fuller statistic in absolute values. A modified test statistic along the line of Vogelsang (1998a) is

$$W^*=\exp(-bH)W$$

for some positive scaling constant b. If the series under test is an $I(0)$ process it follows that $W^*\Rightarrow D_0$ under validity of the null hypothesis, whereas $W^*\Rightarrow\exp(-bD_H)D_1$ if the series is an $I(1)$ process. The specific value b is chosen such that

$$P(D_0>c_\alpha)=P(\exp(-bD_H)D_1>c_\alpha)=\alpha,$$

where c_α denotes the asymptotic critical value at the given nominal significance level α. The limiting distribution of W^* thus coincides with that of W under $I(0)$ irrespective of whether the series is $I(0)$ or $I(1)$. Thus, critical values from standard distributions can be

[1]The nonlinear trend case is investigated in Harvey et al. (2010).

used in cases where the true order of integration is zero or one, but unknown. Especially for economic time series it is difficult establishing clear results for their orders of integration. Therefore, Vogelsang's approach appears to be useful for economic applications.

However, this approach bears some limitations. It only allows for two different orders of integration (zero and one) and it assumes that one of them is the true order. To the best of our knowledge, the behaviour of the Vogelsang (1998a) approach where both possible order of integration are wrong has not been explored so far in the related literature. Such a situation may easily occur under fractional integration with $0 < d < 1$ and when the two possible orders of integration are zero or one. It should be noted that Vogelsang's approach can be modified to deal with a situation where the assumed orders of integration are both fractional, but at least one of the assumed integration orders has to be the true one. In practice, such a restriction may be viewed as too strong and thereby lowering the practical use of such an approach when fractional integration is considered as a possibility.

The following theorem considers the case where $H = |DF|^{-1}$, DF being the standard Dickey-Fuller unit root test statistic. It shows that Vogelsang's approach is not extendable beyond the case where the unknown order of integration d_0 is either 0 or 1. All proofs are collected in Appendix A.

Theorem 1 *Let y_t be generated by $y_t = \beta' z_t + e_t$, $e_t \sim I(d_0)$ with $d_0 \in (0, 1/2) \cup (1/2, 3/2)$ and let Assumptions 1, 2 and 3 hold. Under the validity of the null hypothesis $H_0 : \beta = 0$, the test statistic $W^* = \exp(-b|DF|^{-1})W$ has the same limiting distribution as W, both depending on the unknown order of integration d_0. Moreover, it holds that*

1. *if $d_0 \in (0, 1/2)$ then $\exp(-b|DF|^{-1}) \xrightarrow{P} 1$ and $W^* = O(T^{2d_0})$;*
2. *if $d_0 \in (1/2, 1) \cup (1, 3/2)$ then $\exp(-b|DF|^{-1}) \xrightarrow{P} 1$ and $W^* = O(T^{2d_0+1})$.*

Applying the statistic W^* allows for robustness in the sense of an identical limiting null distribution regardless whether $d_0 = 0$ or $d_0 = 1$. This limiting distribution is that which in the case $d_0 = 0$ governs the unmodified statistic W. Theorem 1 demonstrates that this kind of robustness is lost under fractional integration, i.e. $d_0 \in (0, 1/2) \cup (1/2, 3/2)$. Neither W nor W^* has the same limiting distribution when $d_0 \in (0, 1/2) \cup (1/2, 3/2)$ as the one governing W when $d_0 = 0$. While $W = O(1)$ for $d_0 = 0$, the theorem shows that W and W^* diverge for $d_0 \in (0, 1/2) \cup (1/2, 3/2)$. In short, the Vogelsang approach bears some important difficulties when fractional integration is present in the data generating process. If the unknown order of integration is neither 0 nor 1, the limiting distribution of W^* is no longer the one of W, but dependent on d_0.

Of course, Vogelsang's approach can be applied to any other known value of d_0, i.e., this must not necessarily be the value of zero. In the given setup, a test could be constructed which is valid if the series is either $I(d_0)$ or $I(d_1)$, for some $d_0, d_1 \in [0, 1/2) \cup (1/2, 3/2)$. In this case, the test statistic W shall be modified in such a way that it has a non-degenerate limiting distribution when $e_t \sim I(d_0)$. However, the approach is still restricted to the case of two orders of integration, of which one has to be the true order. These issues pose some important practical limitations. In the next section, we propose a test which is robust to a continuum of possible fractional orders. We propose a testing framework that is applicable for trend testing and more generally, for deterministic regressors. Furthermore, we extend the analysis to the case of stochastic regressors and fractional cointegration.

3 LM-type testing approach

Denote by $\eta' = (\beta', \theta, \sigma^2)'$ the vector of all parameters. We consider testing the hypothesis

$$H_0 : \beta = 0 \qquad \text{vs.} \qquad H_1 : \beta \neq 0. \tag{7}$$

If the hypothesis of interest takes the linear form

$$H_0 : R\beta = r \qquad \text{vs.} \qquad H_1 : R\beta \neq r,$$

with R and r given, of dimension $k \times p$ and $k \times 1$, respectively, $k < p$, then β is decomposed into a free component of dimension $(p-k) \times 1$ and a component of dimension $k \times 1$ that is zero under the null, so the hypothesis tested is on the latter component and of the form (7). Recasting the problem in a suitable form in this case also requires a transformation of the data and consistent estimation of the free component of β, which is assumed to be possible, see Appendix C for a detailed discussion. Denote by $L(\eta)$ the log-likelihood function of y_t. Then the score test statistic (see Rao 1973) is given by

$$\left(\frac{\partial L(\eta)}{\partial \eta}\right)' \left[E\left(\frac{\partial L(\eta)}{\partial \eta}\right)\left(\frac{\partial L(\eta)}{\partial \eta}\right)'\right]^{-1} \left(\frac{\partial L(\eta)}{\partial \eta}\right)\Bigg|_{\beta=0;\theta=\hat{\theta};\sigma^2=\hat{\sigma}^2} \tag{8}$$

The expectation is taken under H_0 and $\hat{\theta}$ and $\hat{\sigma}^2$ are the ML estimates for θ and σ^2, respectively.

3.1 Deterministic Regressors

Let us first consider the case of purely deterministic regressors z_t. We thus impose Assumption 3 in the following. In this situation the LM-type test statistic takes the form

$$LM(\beta) = \frac{1}{\widehat{\sigma}^2} A' B^{-1} A, \tag{9}$$

with $A = \sum_{t=1}^{T} \left(\phi(B;\widehat{\theta})y_t\right) \phi(B;\widehat{\theta})z_t$ and $B = \sum_{t=1}^{T} \left(\phi(B;\widehat{\theta})z_t\right) \left(\phi(B;\widehat{\theta})z_t\right)'$. A detailed derivation of the score test statistic can be found in Appendix B. We first show that the test statistic has a limiting χ^2-distribution independent of d and we derive the local power function of the test. These results hinge on the consistency of the MLEs for θ (including d) and σ^2 which has been proven by Beran (1995). For simplicity we consider the interval $d \in [0, 1/2) \cup (1/2, 3/2)$. Subsequently, we show that it is also possible to draw inference on the basis of standard distributions when other estimators than the MLE are used in (9).

Theorem 2 *Under Assumptions 1, 2, and 3 and for $d \in [0, 1/2) \cup (1/2, 3/2)$, the limiting distribution of the LM test statistic under H_0 is given by*

$$LM(\beta) \Rightarrow \chi^2_p,$$

as $T \to \infty$.

Note that $\widehat{\sigma}^2 \xrightarrow{P} \sigma^2$. Next, we consider a local alternative of the form (i.e. a Pitman drift)

$$\beta_T = \frac{c}{\sqrt{T}},$$

for an arbitrary non-zero p-vector c. The score for β involves y_t that has a non-zero mean.

Theorem 3 *Under the assumptions of Theorem 2 the LM-test $LM(\beta_T)$ has non-trivial power against local alternatives and converges weakly to a non-central χ^2_p-distribution with non-centrality parameter $\frac{\xi^2}{\sigma^2}c'c$.*

Here, ξ^2 is the trace of the matrix J from Assumption 3. Thus, the test enjoys Pitman optimality and has non-trivial power against local alternatives.

We now turn to the use of other estimators than the MLE of d in (9). The main reason for considering different estimators for d is the need for consistency under both, the null and the alternative hypothesis. Biased estimation of d would induce incorrect fractional differencing which ultimately leads to size distortions and power losses in finite samples. When other estimators $\widehat{d}$ are used, the statistic $LM(\beta)$ is no longer an LM-test, and we indicate this by the notation $LM^*(\beta)$, and we continue to refer to it as an LM-type test. We proceed by stating an assumption on the asymptotic behaviour of alternative estimators for d.

Assumption 4 *$\widehat{d}$ is a consistent estimator of d satisfying*

$$(\widehat{d} - d) \log T = o_P(1).$$

Assumption 4 is sufficiently general to include the most commonly used consistent estimators, such as the log-periodogram regression estimator of Geweke and Porter-Hudak (1983), the Gaussian semiparametric or local Whittle estimator of Robinson (1995) as well as the exact local Whittle estimator by Shimotsu and Phillips (2005).

Theorem 4 *Under Assumptions 1, 2, 3, and 4, and for* $d \in [0, 1/2) \cup (1/2, 3/2)$, *the limiting distribution of the LM*-statistic under* H_0 *is given by*

$$LM^*(\beta) \Rightarrow \chi^2_p,$$

as $T \to \infty$.

Thus, the limiting null distribution of LM^* coincides with that of the LM-test in Theorem 2 and is thus independent of the unknown integration order. In this sense, it is also possible to draw valid inference on β under an unknown order of integration even when using other estimators satisfying Assumption 4 than the MLE. The results for the local power function, see Theorem 3, carry over to LM^*.[2]

3.2 Stochastic Regressors

So far, the analysis has focussed on the case of deterministic regressors. However, the new LM^* testing approach can be applied in situations with stochastic regressors, too. As before we consider the model

$$y_t = \beta' z_t + e_t, \tag{10}$$

where now y_t and z_t are type II fractional processes of order d_1. The parameter vector $\beta = (\beta_1, \ldots, \beta_p)'$ and z_t are both p-dimensional. As we focus on fractional cointegration below we assume that all components of z_t are integrated of the same order d_1 and that $E[z'_{it} z_{jt}] = 0$ for $i \neq j$. Furthermore, we assume that $d_1 > 1/2$ holds. The error term e_t is assumed to be a type II fractional process of order $I(d_0)$ with $d_0 < d_1$ and $d_1 - d_0 > 1/2 + \epsilon$ for some $\epsilon > 0$. Consider testing the hypothesis

$$H_0 : \beta = \beta_0, \qquad \text{against} \qquad H_1 : \beta \neq \beta_0, \tag{11}$$

with β and β_0 being p-vectors. Imposing an other restriction than $\beta_0 = 0$ does not change our results. As we are aware that this restriction might have some attraction to keep the notation simple it is of minor practical relevance in this set up and therefore we do not impose it in this section. Regarding exogeneity of the regressors we impose the following assumption:

[2] Although the proof of Theorem 4 is specific to the case of deterministic regressors, it is a general property that when plug-in estimators of nuisance parameters (such as $\widehat{d}$) converge sufficiently fast, then limiting distributions and local power of tests are unaffected (see, e.g., Jansson and Nielsen, 2012).

Assumption 5 *The z_t are strictly exogenous with*

$$\phi(B, \theta_1)z_t = u_{1t}$$

and

$$\phi(B, \theta_0)e_t = u_t,$$

and u_{1t} and u_t fulfill Assumption 1. Here, $\phi(B, \theta)$ is defined as before with θ_1 containing d_1 and θ_0 containing d_0 as orders of integration.

The log-likelihood function in this framework is similar to the case of non-stochastic regressors and can be found in Appendix B where also the derivation of the test statistic is located. By $\hat{\theta}$ and $\hat{\sigma}^2$ we denote the ML-estimator for θ and σ^2, respectively. By evaluating the scores at $\beta = \beta_0$, we obtain by similar arguments as in the case of non-stochastic regressors that the LM-test statistic is given by

$$LM(\beta) = \frac{1}{\hat{\sigma}^2} A' B^{-1} A \tag{12}$$

with $A = \sum_{t=1}^{T} (\phi(B, \hat{\theta})(y_t - \beta_0' z_t)(\phi(B, \hat{\theta})z_t)$ and $B = \sum_{t=1}^{T} (\phi(B, \hat{\theta})z_t)(\phi(B, \hat{\theta})z_t)'$. The test statistic (12) in the case of stochastic exogenous regressors is thus of exactly the same form as the test statistic (9) for non-stochastic regressors. However, it should be mentioned that the term $\phi(B, \hat{\theta})z_t$ is now stochastic and integrated of order $d_1 - d_0$.

Theorem 5 *Under Assumptions 1, 2 and 5, the limiting distribution of the LM test statistic under H_0 is given by*

$$LM(\beta) \Rightarrow \chi_p^2$$

as $T \to \infty$.

Thus, neither the fractional integration order of the regressors (d_1) nor the fractional order of the errors (d_0) influences the distribution of the test statistic which is similar to the case of deterministic regressors, a standard one. By similar arguments as in Theorem 4 we can use other estimators for d_1 and d_0 than the MLE as long as they fulfill Assumption 4. The test has local power also in the case of stochastic exogenous regressors.

Theorem 6 *Under the assumptions of Theorem 5, the LM-test $LM(\beta_T)$ has non-trivial power against local alternatives and converges to a non-central χ_p^2-distribution with non-centrality parameter $\frac{\xi^2}{\sigma^2}c'c$.*

We now relax the assumption of z_t being exogenous and derive the limiting distribution of the LM-type test for this case. Our result shows that the test is biased when no endogeneity correction is applied. As demonstrated in the Monte Carlo section, correcting for endogeniety via fully modified OLS estimation of β results in an unbiased and correctly sized test. Instead of Assumption 5 let us now assume

Assumption 6 *The regressors z_t are defined by*

$$\phi(B,\theta_1)z_t = u_{1t}$$

and

$$\phi(B,\theta_0)e_t = u_t,$$

and u_{1t} and u_t fulfill Assumption 1. Here, $\phi(B,\theta)$ is defined as before with θ_1 containing d_1 and θ_0 containing d_0 as orders of integration. Furthermore, $\epsilon_t = (u_t, u_{1t}^{'})' \sim iid(0,\Sigma)$ with finite fourth moments and Σ is a positive definite matrix partitioned as

$$\Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{21}^{'} \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

The log-likelihood function is in this case different to the exogenous case. We rely on the conditional log-likelihood for this setup, see also Nielsen (2004). It is given by

$$L(\beta,\theta,\Sigma) = -\frac{T}{2}\log\sigma_{1.2}^2 - \frac{1}{2\sigma_{1.2}^2}\sum_{t=1}^{T}\left[\phi(B,\theta)(y_t - \beta^{'}z_t) - \sigma_{21}^{'}\Sigma_{22}^{-1}\phi(B,\theta_1)z_t\right]^2, \qquad (13)$$

where $\sigma_{1.2}^2 = \sigma_{11}^2 - \sigma_{21}^{'}\Sigma_{22}^{-1}\sigma_{21}$. The parameter vector θ contains the order of integration $d_1 - d_0$. Thus, the LM-test is in this setup given by

$$LM = \frac{1}{\hat{\sigma}_{1.2}^2}A^{'}B^{-1}A \qquad (14)$$

with

$$\begin{aligned} A &= \frac{\partial L(\beta,\theta,\Sigma)}{\partial\beta}\Big|_{\beta=\beta_0,\theta=\hat{\theta},\Sigma=\hat{\Sigma}} \qquad (15) \\ &= \frac{1}{\hat{\sigma}_{1.2}^2}\sum_{t=1}^{T}\left[\phi(B,\hat{\theta})(y_t - \beta_0^{'}z_t)\phi(B,\hat{\theta})z_t - \hat{\sigma}_{21}^{'}\hat{\Sigma}_{22}^{-1}\phi(B,\hat{\theta}_1)z_t\phi(B,\hat{\theta})z_t\right] \end{aligned}$$

and the Hessian

$$\begin{aligned} B &= \frac{\partial^2 L(\beta,\theta,\Sigma)}{\partial\beta\partial\beta'}\Big|_{\beta=\beta_0,\theta=\hat{\theta},\Sigma=\hat{\Sigma}} \qquad (16) \\ &= \frac{1}{\hat{\sigma}_{1.2}^2}\sum_{t=1}^{T}(\phi(B,\theta)z_t)(\phi(B,\theta)z_t)^{'}. \end{aligned}$$

Theorem 7 *Under Assumptions 1, 2, and 6, the limiting distribution of the LM test* *statistic under H_0 is given by*

$$LM(\beta) \Rightarrow \chi^2_p + \Lambda,$$

as $T \to \infty$. Λ denotes an endogeneity bias.

Note that $\widehat{\Sigma} \xrightarrow{P} \Sigma$.

4 Examples and Monte Carlo simulations

In this section, we discuss the application of the LM-type testing approach to cases of testing for a deterministic linear and quadratic trend component. We further consider the case of a possible mean shift and trend break in a linear trend function. As an example of stochastic regressors we draw inference on the fractional cointegration vector and, as an extension, also on deterministic components in a cointegrating regression. We combine these examples with Monte Carlo simulations demonstrating the finite sample behavior of the test in different situations.

4.1 Linear and quadratic trend tests under fractional integration

We start by testing for a polynomial trend. Let us assume the data generating process

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_{p-1} t^{p-1} + e_t, \tag{17}$$

where $(1-B)^{d_0} e_t = u_t$, (fractional Gaussian noise) with d_0 the unknown order of integration, and the innovations u_t fulfill Assumptions 1 and 2. The parameters $\beta = (\beta_0, \beta_1, \ldots, \beta_{k-1})'$ and regressors $z_t = (1, t, t^2, \ldots, t^{k-1})'$ are deterministic and fulfill Assumption 4. We focus on the empirically relevant cases of $p = 2, 3$ and consider tests for the hypotheses that a linear or a quadratic trend component is present in the series y_t. The corresponding null hypotheses are given by $H_0 : \beta_1 = 0$ or $H_0 : \beta_2 = 0$, respectively. The requirements of Theorem 2 are fulfilled. Thereby, the critical values for the LM-type tests are taken from the $\chi^2(1)$-distribution.

We treat all parameters, including the order of fractional integration d_0, as unknown. Regarding estimation of the fractional order of integration, we compare the two-step exact local Whittle estimator by Shimoutsu (2010) [2-ELW] to the fully extended local Whittle estimator by Abadir, Distaso and Giraitis (2007) [FELW]. Both estimators are consistent under the null and the alternative hypotheses. By Theorem 4, the limiting null distribution of LM^* applies without modification when using one of these semi-parametric estimators.

In the following, we report the empirical size and power of the LM-type test based on the two different estimators. As a benchmark we include the performance of an infeasible

Table 1: Empirical *size and power* for testing $H_0 : \beta_1 = 0$ or $H_0 : \beta_2 = 0$, $T = 500$, normally distributed errors

Linear Trend

d_0	Stat	Size	avg $\widehat{d}_0$	sd $\widehat{d}_0$	Power	avg $\widehat{d}_0$	sd $\widehat{d}_0$
0	2-ELW	0.057	0.001	0.057	1.000	0.000	0.057
	FELW	0.069	-0.013	0.055	1.000	-0.014	0.055
	Infeasible	0.050	0.000	0.000	1.000	0.000	0.000
0.4	2-ELW	0.059	0.403	0.058	0.836	0.402	0.058
	FELW	0.074	0.380	0.060	0.869	0.379	0.060
	Infeasible	0.048	0.400	0.000	0.872	0.400	0.000
0.8	2-ELW	0.053	0.810	0.057	0.104	0.808	0.057
	FELW	0.075	0.783	0.055	0.140	0.781	0.055
	Infeasible	0.050	0.800	0.000	0.102	0.800	0.000
1.2	2-ELW	0.049	1.212	0.055	0.058	1.211	0.055
	FELW	0.084	1.174	0.054	0.096	1.173	0.054
	Infeasible	0.047	1.200	0.000	0.052	1.200	0.000

Quadratic Trend

d_0	Stat	Size	avg $\widehat{d}_0$	sd $\widehat{d}_0$	Power	avg $\widehat{d}_0$	sd $\widehat{d}_0$
0	2-ELW	0.065	-0.010	0.059	1.000	-0.009	0.059
	FELW	0.075	-0.026	0.057	1.000	-0.025	0.058
	Infeasible	0.051	0.000	0.000	1.000	0.000	0.000
0.4	2-ELW	0.062	0.392	0.060	0.993	0.393	0.060
	FELW	0.079	0.369	0.062	0.996	0.370	0.063
	Infeasible	0.051	0.400	0.000	0.997	0.400	0.000
0.8	2-ELW	0.052	0.800	0.060	0.210	0.801	0.060
	FELW	0.075	0.773	0.057	0.265	0.774	0.057
	Infeasible	0.048	0.800	0.000	0.204	0.800	0.000
1.2	2-ELW	0.055	1.204	0.057	0.070	1.204	0.057
	FELW	0.084	1.164	0.056	0.104	1.164	0.056
	Infeasible	0.050	1.200	0.000	0.066	1.200	0.000

test using the true parameter d_0 instead of an estimate. The nominal significance level equals five percent, the sample size is $T = 500$ and the number of Monte Carlo replications equals 5,000 for each single experiment. All computations are carried out in MATLAB.[3] In addition to the size and power results, we also report the Monte Carlo mean and standard deviation of the 2ELW and the FELW estimator. The importance of normality is investigated by considering non-normal errors. To this end, we generate fat-tailed $t(6)$-distributed and skewed standardized $\chi^2(5)$-distributed errors. In these cases, estimation of β is achieved via quasi-ML assuming normality for constructing the likelihood function.

[3] We would like to thank Karim Abadir, Walter Distaso and Luidas Giraitis for sharing their MATLAB code for the FELW estimator with us. The code for the Shimotsu (2010) estimator can be found on the author's webpage.

Table 2: Empirical *size* for testing $H_0 : \beta_1 = 0$ or $H_0 : \beta_2 = 0$, $T = 500$, non-normally distributed errors

Linear Trend

d_0	Stat	$t(6)$	avg $\widehat{d}_0$	sd $\widehat{d}_0$	$\chi^2(5)$	avg $\widehat{d}_0$	sd $\widehat{d}_0$
0	2-ELW	0.058	0.000	0.057	0.044	0.001	0.056
	FELW	0.070	-0.014	0.056	0.066	-0.013	0.054
	Infeasible	0.050	0.000	0.000	0.041	0.000	0.000
0.4	2-ELW	0.058	0.403	0.058	0.043	0.403	0.057
	FELW	0.079	0.380	0.061	0.080	0.381	0.059
	Infeasible	0.051	0.400	0.000	0.042	0.400	0.000
0.8	2-ELW	0.059	0.808	0.057	0.052	0.809	0.056
	FELW	0.082	0.781	0.055	0.051	0.782	0.054
	Infeasible	0.053	0.800	0.000	0.050	0.800	0.000
1.2	2-ELW	0.053	1.212	0.054	0.057	1.212	0.054
	FELW	0.086	1.174	0.054	0.069	1.174	0.053
	Infeasible	0.050	1.200	0.000	0.050	1.200	0.000

Quadratic Trend

d_0	Stat	$t(6)$	avg $\widehat{d}_0$	sd $\widehat{d}_0$	$\chi^2(5)$	avg $\widehat{d}_0$	sd $\widehat{d}_0$
0	2-ELW	0.067	-0.009	0.058	0.063	-0.010	0.059
	FELW	0.075	-0.024	0.057	0.073	-0.026	0.058
	Infeasible	0.052	0.000	0.000	0.051	0.000	0.000
0.4	2-ELW	0.059	0.393	0.060	0.067	0.392	0.060
	FELW	0.076	0.370	0.062	0.083	0.369	0.061
	Infeasible	0.046	0.400	0.000	0.049	0.400	0.000
0.8	2-ELW	0.056	0.801	0.058	0.055	0.801	0.059
	FELW	0.074	0.775	0.055	0.076	0.774	0.057
	Infeasible	0.045	0.800	0.000	0.053	0.800	0.000
1.2	2-ELW	0.053	1.206	0.056	0.058	1.205	0.057
	FELW	0.085	1.166	0.055	0.092	1.165	0.056
	Infeasible	0.051	1.200	0.000	0.048	1.200	0.000

Results shed light on the role of normality for the finite-sample performance of the LM-type test.

Table 1 presents the size and power results for testing for a linear (upper panel) and quadratic trend component (lower panel). The parameter d_0 takes the values 0, 0.4, 0.8 and 1.2. Thereby, we cover short memory, stationary and non-stationary fractional integration. Regarding the size results, we find the LM-type test based on the 2-ELW estimator provides accurate levels, whereas the variant using the FELW estimator tends to be somewhat oversized. This finding relates to a slight downward bias of the FELW estimator. Consider, e.g. the results for $d = 0.8$: the Monte Carlo average for the FELW estimator is 0.783 and the corresponding size of the test is 0.075. In contrast, the 2-ELW estimator is a bit upward biased (less than the FELW in absolute terms, however) (0.810), but the simulated size is 0.053. Noteworthy, the FELW estimator has a smaller standard deviation (0.055 versus 0.057). An even slight downward bias implies that the series will not be differenced to $I(0)$, but to a small positive order of integration. The LM-type test appears to be sensitive towards such a bias and therefore, we recommend to use the 2-ELW estimator for testing linear trends. This finding also holds for testing quadratic trend components. The infeasible test has very accurate size in all cases. Interestingly, the LM-type test based on the 2-ELW estimator performs very well in comparison to the infeasible test. The results for both estimators show their merits as they perform very well under H_0 and H_1.

The reported results for the empirical power ($\beta_1 = 0.005$ and $\beta_2 = 0.00005$) indicate similar performances amongst the different variants of the test. It shall be noted that the values for β_1 and β_2 are fixed for different values of d_0. The larger the fractional order of integration, the larger is the variance of the stochastic part of the process and the more difficult is the detection of a linear or quadratic trend component. Thus, it is not surprising that the power decreases as d_0 increases.

Table 2 presents size results for fat-tailed and skewed errors. The results are very similar to the ones obtained for normally distributed errors. The main conclusions do not change and the results suggest robustness against a misspecification of the error distribution. The LM-type test also performs well when errors are assumed to be normal, but are fat-tailed or skewed instead.

4.2 Structural change tests for intercept and slope of a linear trend model under fractional integration

We also consider the problem of testing for a break in the intercept and slope parameter of a linear trend function. A slightly different model (without intercept shift) is studied in Iacone et al. (2013a, 2013b). They consider the behavior of fixed-b tests for trend breaks and robustify those to the fractional case. The model of our interest, see also Perron and Yabu (2009) for a quasi-feasible GLS approach in case of $I(0)$ or $I(1)$ errors, is given by

$$y_t = \beta_0 + \delta_0 DU_t + \beta_1 t + \delta_1 DT_t + e_t \tag{18}$$

where the variables DU_t and DT_t are given by

$$\begin{aligned} DU_t &= 1(t > c), \\ DT_t &= 1(t > c)(t - \lfloor cT \rfloor). \end{aligned}$$

The relative break point is denoted by c and $c \in [\tau_L, \tau_U] =: \Lambda \subset [0,1]$ and τ_L and τ_U are the lower and upper bounds of the interval to which the break point is restricted.

Table 3: Empirical *size and power* for testing $H_0 : \delta_0 = 0$ or $H_0 : \delta_1 = 0$, $T = 500$, normally distributed errors

d	Stat	MS Size	TB Size	avg $\widehat{d}$	sd $\widehat{d}$	MS Power	TB Power	avg $\widehat{d}$	sd $\widehat{d}$
0	2-ELW	0.046	0.055	0.005	0.050	1.000	1.000	0.005	0.050
	FELW	0.062	0.076	-0.026	0.048	1.000	1.000	-0.026	0.048
	Infeasible	0.044	0.051	0.000	0.000	1.000	1.000	0.000	0.000
0.4	2-ELW	0.050	0.052	0.407	0.051	0.512	1.000	0.406	0.052
	FELW	0.075	0.085	0.360	0.051	0.575	1.000	0.359	0.052
	Infeasible	0.048	0.050	0.400	0.000	0.516	1.000	0.400	0.000
0.8	2-ELW	0.054	0.045	0.816	0.051	0.263	0.344	0.816	0.051
	FELW	0.058	0.085	0.758	0.048	0.284	0.488	0.759	0.049
	Infeasible	0.051	0.046	0.800	0.000	0.267	0.392	0.800	0.000
1.2	2-ELW	0.051	0.083	1.165	0.066	0.257	0.132	1.164	0.065
	FELW	0.052	0.151	1.091	0.063	0.255	0.214	1.091	0.062
	Infeasible	0.049	0.047	1.200	0.000	0.258	0.080	1.200	0.000

In practice c is usually treated as unknown and consistent breakpoint estimators are readily available from the related literature, see also Perron and Zhu (2005) and Iacone et al. (2013a, 2013b). For simplicity of the exposition, we work with a known value of $c = 0.5$ (middle of the sample). The fractionally integrated error term e_t is generated via $(1 - B)^{d_0} e_t = u_t$ and u_t satisfies Assumption 1.

We are interested in testing the two distinct null hypotheses $H_0 : \delta_0 = 0$ (no mean shift) or $H_0 : \delta_1 = 0$ (no trend break). A main problem in this setup is the estimation of d_0 in order to obtain a feasible test. In line with our previous analysis, we compare the two-step ELW estimator to the fully extended LW estimator and an infeasible version using the true value d_0 instead of an estimate. The data y_t is first de-trended as given in equation (2), i.e. $\widehat{e}_t$ denote the corresponding OLS residuals. We note that this framework is not directly captured in the asymptotic derivations of Shimotsu (2010) and Abadir et al. (2007), but our conjecture is that their procedures remain valid. Results for size and power with normally distributed innovations u_t are reported in Table 3. Our findings suggest similar conclusions as before. The LM-type test based on the 2-ELW estimator by Shimotsu (2010) continues to provide accurate size performance in almost all cases. The only exception is the trend break test for $d_0 = 1.2$. The LM-type test also exhibits non-trivial power against mean shifts ($\delta_0 = 1$) and trend breaks ($\delta_1 = 0.05$). The results for both estimators also underline their usefulness as they are nearly unbiased, both under the null and under the alternative hypothesis.

4.3 Inference on the fractional cointegration vector and deterministic trends

Cointegration is a natural example in the case of stochastic regressors. In the standard Engle and Granger (1987) cointegration model it is assumed that the data series (y_t, z_t) are $I(1)$, and cointegration exists if one or more linear combinations of the series are $I(0)$. Thus, in our regression framework

$$y_t = \beta' z_t + e_t,$$

cointegration is implied if $e_t \sim I(0)$, and spurious regression if $e_t \sim I(1)$. As in Vogelsang's approach to trend testing, it is critical that the order of integration of the errors (d_0) is known to be one of only two possible values (again, 0 or 1). In practical applications, other possible values for d must often be considered, as well. In case of fractional cointegration, z_t may be (possibly fractionally) integrated of order d_1, say, not necessarily restricting d_1 to unity. Fractional cointegration exists if the errors are integrated of lower order, $e_t \sim I(d_0)$, with $d_0 < d_1$. In this case, y_t is integrated of order d_1, too, and the series y_t and z_t move together in the long run, in a more general sense not necessarily implying standard $I(1)-I(0)$ cointegration. In particular, d may take values in a continuum, rather than being restricted to be either 0 or 1. In particular, we consider the simple case of fractional cointegration

$$\begin{aligned} y_t &= \beta z_t + e_t && (19)\\ (1-B)^{d_0} e_t &= u_t && (20)\\ (1-B)^{d_1} z_t &= u_{1t} && (21) \end{aligned}$$

satisfying Assumption 6. y_t and z_t are scalar type II fractional processes of order $d_1 > 1/2$. The errors e_t are integrated of order $d_0 = 0$ for simplicity. This type of triangular system has been considered by Phillips (1991) for the case of standard cointegration, and by Nielsen (2004) for fractional cointegration, and is easily extended to the case of multiple cointegrating relations. Evidently, z_t are stochastic regressors in the cointegration case, whether standard or fractional, and the latter is a natural application of our results in Section 3.2. Estimation of the cointegration system under unknown order of integration has been considered by Robinson and Hualde (2003).

With y_t and z_t moving together in the long run, interest centers on the cointegrating relation, as described by the cointegrating vector β. The linear hypothesis of interest is given by $H_0 : \beta = 1$ which is a restriction on the sub-cointegrating vector and imposes a one-to-one long-run relationship between y_t and z_t. This is a typical example for a hypothesis of long-run unbiasedness in predictive regressions, e.g., testing expectation

Table 4: Stochastic regressors - *size* for LM-type testing $H_0 : \beta = 1$ in $y_t = \beta z_t + e_t$ with $e_t \sim I(d_0)$ and $y_t, z_t \sim I(d_1)$ and $\rho = corr(u_t, v_t)$

T	d_1	$LM^*\|d$	$LM^*\|\widehat{d}$	$\widehat{d}_1$	$\widehat{d}_0$	$\widehat{\rho}\|d$	$\widehat{\rho}\|\widehat{d}$
$\rho = 0$: exogeneity							
250	0.6	0.062	0.103	0.578	-0.011	0.000	0.000
	0.8	0.064	0.094	0.762	-0.012	0.001	0.001
	1	0.065	0.091	0.948	-0.008	-0.003	-0.003
	1.2	0.064	0.091	1.137	-0.012	0.001	0.001
500	0.6	0.056	0.108	0.584	-0.007	0.001	0.001
	0.8	0.061	0.084	0.771	-0.006	-0.001	-0.001
	1	0.054	0.086	0.962	-0.005	-0.001	-0.001
	1.2	0.042	0.068	1.160	-0.006	-0.001	-0.001
1000	0.6	0.059	0.076	0.590	-0.005	0.001	0.001
	0.8	0.043	0.051	0.779	-0.003	0.001	0.001
	1	0.050	0.069	0.974	-0.005	0.001	0.001
	1.2	0.060	0.083	1.170	-0.002	0.000	0.000
$\rho = 0.5$: endogeneity							
250	0.6	0.078	0.251	0.578	0.011	0.395	0.389
	0.8	0.074	0.132	0.761	-0.002	0.473	0.468
	1	0.052	0.109	0.948	-0.003	0.493	0.488
	1.2	0.075	0.124	1.138	-0.010	0.500	0.496
500	0.6	0.068	0.242	0.583	0.017	0.413	0.409
	0.8	0.054	0.088	0.771	0.004	0.478	0.475
	1	0.049	0.096	0.962	-0.002	0.498	0.496
	1.2	0.047	0.076	1.161	-0.004	0.499	0.497
1000	0.6	0.077	0.243	0.590	0.022	0.424	0.422
	0.8	0.050	0.074	0.779	0.005	0.486	0.484
	1	0.053	0.080	0.974	-0.003	0.500	0.498
	1.2	0.059	0.079	1.170	0.000	0.501	0.499

hypotheses, interest rate parity, purchasing power parity, etc., in persistent data.

A typical assumption in the fractional cointegration literature, e.g. Nielsen (2004), is that $d_1 \geq d_1 - d_0 \geq 3/4 + \epsilon$, for some $\epsilon > 0$. Thus, the series are nonstationary, and cointegration reduces the integration order by more than 3/4. However, our Theorem 5 applies more generally, as long as the strength of cointegration $d_1 - d_0 \geq 1/2 + \epsilon$, for some $\epsilon > 0$. Of course, this requires in addition exogeneity of z_t, i.e., u_t and u_{1t} are orthogonal. In general, this would seem a rather restrictive form of cointegration, but it is indeed relevant when testing predictive regressions. In particular, if the theory under test posits that z_t is the conditional expectation of y_t given an information set that includes u_t, then u_t and u_{1t} must be orthogonal, and exogeneity is implied, c.f. Assumption 5. On the other hand, the construction of our LM* test requires substituting estimators for θ and σ^2 in the LM-statistic from Section 3.2 that are consistent both under the null and the

Table 5: Stochastic regressors - *power* for LM-type testing $H_0 : \beta = 1$ in $y_t = 1.005 z_t + e_t$ with $e_t \sim I(d_0)$ and $y_t, z_t \sim I(d_1)$ and $\rho = corr(u_t, v_t) = 0$

T	d_1	$LM^*\|d$	$LM^*\|\widehat{d}$	$\widehat{d_1}$	$\widehat{d_0}$	$\widehat{\rho}\|d$	$\widehat{\rho}\|\widehat{d}$
250	0.6	0.061	0.122	0.577	-0.007	0.000	0.000
	0.8	0.081	0.119	0.757	-0.012	0.002	0.002
	1	0.155	0.199	0.949	-0.010	-0.002	-0.002
	1.2	0.498	0.501	1.138	-0.008	0.002	0.002
500	0.6	0.086	0.114	0.584	-0.005	-0.001	-0.001
	0.8	0.111	0.138	0.770	-0.007	0.002	0.002
	1	0.356	0.382	0.962	-0.007	0.000	0.000
	1.2	0.837	0.839	1.157	-0.007	0.002	0.002
1000	0.6	0.095	0.128	0.589	-0.006	0.000	0.000
	0.8	0.203	0.233	0.779	-0.005	0.000	0.001
	1	0.707	0.718	0.973	-0.003	0.000	0.000
	1.2	0.980	0.982	1.170	-0.003	-0.001	-0.001

alternative.

Some generalizations are possible. Exogeneity may be relaxed by considering fully modified residuals, but the limiting distribution becomes more complicated. It is possible to combine our results and consider, e.g., both deterministic and stochastic trends, and test on either, or both. It is also possible in the fractional cointegration case to test only on a subvector of β, substituting estimators consistent both under the null and the alternative for the other components (when estimating θ and σ^2). For the estimator for β, we may consider those in Kim and Phillips (2001), Nielsen (2004), or Christensen and Nielsen (2006). Other possible choices for estimators of d and σ^2 are those suggested by Nielsen (2004), Christensen and Nielsen (2006), or Nielsen (2007).

In the following, we estimate d_0 and d_1 via the FELW estimator. The innovations u_{1t} which generate z_t are possibly correlated with u_t of strength ρ. As long as $\rho = 0$ holds, z_t is strictly exogenous (c.f. Assumption 5). If this restriction is true, plain OLS estimation of β delivers a well-sized LM-type test. Otherwise, the size of the test behaves badly even for small deviations, i.e. $\rho = 0.1$ (unreported results). Instead, we apply fully modified OLS estimation (see Phillips and Hansen, 1990) for fractional cointegration systems as suggested in Kim and Phillips (2001) and further studied in Nielsen (2004). Our theoretical results in Theorem 7 show the validity of this approach.

Simulation results for the empirical size under exogeneity (upper panel) and endogeneity (lower panel) are reported in Table 4. The sample sizes vary as follows: $T = \{250, 500, 1000\}$. The fractional order of integration of the series takes the values $d_1 = \{0.6, 0.8, 1, 1.2\}$. Thereby, we consider the case where the restriction $d_1 \geq d_1 - d_0 \geq 3/4 + \epsilon$ is not fulfilled, i.e. when $d_1 = 0.6$. As $d_0 = 0$ for simplicity, d_1 directly measures

Table 6: Stochastic regressors - *size* for LM-type testing $H_0 : \beta = 1$ or $H_0 : \gamma = 0$ in $y_t = \eta + \gamma t + \beta z_t + e_t$ with $e_t \sim I(d_0)$ and $y_t, z_t \sim I(d_1)$ and $\rho = corr(u_t, v_t)$

T	d_1	$LM^*(\gamma)\|d$	$LM^*(\beta)\|d$	$LM^*(\gamma)\|\widehat{d}$	$LM^*(\beta)\|\widehat{d}$	$\widehat{d}_1$	$\widehat{d}_0$	$\widehat{\rho}\|d$	$\widehat{\rho}\|\widehat{d}$
$\rho = 0$: exogeneity									
250	0.6	0.062	0.046	0.103	0.099	0.575	-0.024	0.001	0.001
	0.8	0.055	0.057	0.086	0.105	0.759	-0.026	0.000	0.000
	1	0.054	0.055	0.099	0.092	0.950	-0.028	-0.001	-0.001
	1.2	0.046	0.060	0.081	0.097	1.138	-0.030	0.000	0.000
500	0.6	0.055	0.053	0.078	0.090	0.581	-0.011	-0.004	-0.004
	0.8	0.061	0.044	0.083	0.076	0.769	-0.015	0.000	0.000
	1	0.058	0.057	0.079	0.072	0.965	-0.016	0.001	0.001
	1.2	0.053	0.052	0.081	0.077	1.157	-0.017	0.001	0.001
1000	0.6	0.049	0.040	0.078	0.071	0.590	-0.010	0.000	0.000
	0.8	0.051	0.045	0.075	0.078	0.779	-0.011	0.000	0.000
	1	0.046	0.052	0.064	0.059	0.975	-0.008	0.000	0.000
	1.2	0.055	0.047	0.077	0.065	1.172	-0.011	0.000	0.000
$\rho = 0.5$: endogeneity									
250	0.6	0.061	0.053	0.103	0.536	0.578	0.043	0.314	0.304
	0.8	0.061	0.055	0.075	0.081	0.761	0.014	0.438	0.428
	1	0.059	0.057	0.085	0.100	0.948	-0.010	0.485	0.480
	1.2	0.043	0.045	0.086	0.083	1.138	-0.027	0.497	0.493
500	0.6	0.062	0.063	0.126	0.588	0.583	0.054	0.341	0.333
	0.8	0.062	0.051	0.056	0.075	0.771	0.019	0.454	0.448
	1	0.044	0.053	0.061	0.075	0.962	-0.006	0.494	0.491
	1.2	0.067	0.049	0.087	0.078	1.161	-0.016	0.498	0.496
1000	0.6	0.067	0.051	0.106	0.644	0.590	0.063	0.363	0.358
	0.8	0.057	0.051	0.049	0.071	0.779	0.021	0.468	0.465
	1	0.060	0.068	0.067	0.083	0.974	-0.005	0.497	0.496
	1.2	0.072	0.061	0.085	0.079	1.170	-0.009	0.500	0.498

the strength of cointegration in the bivariate system. We report the following statistics: $LM^*|d$ (infeasible, assuming known $d = (d_0, d_1)'$), $LM^*|\widehat{d}$, the Monte Carlo averages of the FELW estimator for d_1 and d_0, $\widehat{\rho}|d$ (infeasible estimator for the correlation between u_t and u_{1t}) and finally $\widehat{\rho}|\widehat{d}$. The null hypothesis is given by $H_0 : \beta = 1$.

The results indicate the following findings: under exogeneity the size results are satisfying. Performance improves along two directions: (i) an increasing sample size and (ii) an increasing strength of cointegration. For large T, differences between the feasible and the infeasible test are reduced, whereas we observe some upward size distortions for a fairly small sample size of $T = 250$. The Monte Carlo averages of the estimator for d_1 and d_0 are reasonably close to the true values, albeit a small downward-bias for d_1 is present for the smallest sample size. The estimator for d_0 is nearly unbiased, although the series from which d_0 is estimated are residuals of the OLS regression $y_t = \beta z_t + e_t$. The estima-

Table 7: Stochastic regressors - *power* for LM-type testing $H_0 : \beta = 1$ or $H_0 : \gamma = 0$ in $y_t = \eta + 0.0005t + 1.005z_t + e_t$ with $e_t \sim I(d_0)$ and $y_t, z_t \sim I(d_1)$ and $\rho = corr(u_t, v_t) = 0$

T	d_1	$LM^*(\gamma)\|d$	$LM^*(\beta)\|d$	$LM^*(\gamma)\|\widehat{d}$	$LM^*(\beta)\|\widehat{d}$	$\widehat{d}_1$	$\widehat{d}_0$	$\widehat{\rho}\|d$	$\widehat{\rho}\|\widehat{d}$
250	0.6	0.203	0.098	0.237	0.155	0.577	-0.023	0.000	0.001
	0.8	0.168	0.179	0.195	0.224	0.755	-0.024	-0.001	-0.001
	1	0.134	0.338	0.185	0.394	0.949	-0.028	-0.004	-0.004
	1.2	0.131	0.730	0.170	0.745	1.139	-0.030	-0.002	-0.002
500	0.6	0.827	0.170	0.815	0.195	0.586	-0.013	0.000	0.000
	0.8	0.779	0.379	0.789	0.400	0.771	-0.017	-0.001	-0.001
	1	0.639	0.801	0.654	0.807	0.964	-0.017	0.000	0.000
	1.2	0.510	0.992	0.530	0.994	1.156	-0.017	0.001	0.001
1000	0.6	1.000	0.303	1.000	0.332	0.589	-0.009	0.000	0.000
	0.8	0.996	0.723	0.997	0.731	0.778	-0.009	-0.001	-0.001
	1	0.982	0.997	0.979	0.996	0.974	-0.010	0.002	0.002
	1.2	0.905	1.000	0.903	1.000	1.169	-0.010	0.001	0.000

tion of ρ is performing well, too. It appears that assuming the true value of $d = (d_1, d_0)'$ leads to similar results as if d is treated as unknown. Under endogeneity, we find that the test has expected problems when the restriction $d_1 \geq d_1 - d_0 \geq 3/4 + \epsilon$ is not met. On the contrary, results are satisfying for $d_1 \geq 0.8$. The test statistic depends on estimation of β, d_1, d_0 and ρ which depends itself on the previously mentioned estimators. Clearly, any estimation error carries over to the LM-type test. Therefore, we conclude that the size results are reasonable and we expect improvements from using efficient estimators. Estimation results for d_1 and d_0 are very similar to the case where $\rho = 0$. Estimation of ρ is less precise, however, for small d_1 and small T. Only small gains can be found when assuming the true value of d. For large d_1 and large T results improve considerably.

Regarding empirical power, we report simulation results under exogeneity in Table 5. The true parameter takes the value $\beta = 1.005$ and is thus only a slight deviation from the null hypothesis of unity. The empirical power enhances with growing T and growing d_1. Importantly, the estimators for d_1 and d_0 perform well under H_1. A similar result is found for the estimator of ρ.

Next, we study the case where additional linear deterministic components are present in the data generating process. To this end, the model is given by

$$y_t = \eta + \gamma t + \beta z_t + e_t.$$

A linear deterministic trend is now present and we are interested in testing for (i) $H_0 : \gamma = 0$ (absence of a linear trend) or (ii) $H_0 : \beta = 1$. Again, endogeneity is allowed for by using fully modified OLS estimation. Results are reported in Tables 6 and 7. The infeasible tests $LM^*(\gamma)|d$ and $LM^*(\beta)|d$ perform very well under exogeneity and endogeneity as well. For the feasible versions of both tests we observe some mild upward size distortions. These can be connected to the slight downward-bias of the estimators for d_1 mainly and partly for d_0. As before, the case of $d_1 = 0.6$ leads to expected problems: The test for $H_0 : \beta = 1$ is heavily affected due to endogeneity and the estimator for d_0 is clearly upward biased which in turn renders the estimator for ρ to be largely downward-biased. Unsurprisingly, even for a large sample size of $T = 1000$ the problem persists. On the other hand, the test for absence of a linear trend in the cointegrating regression is nearly unaffected as the linear trend is deterministic. While point estimation of ρ is very exact under exogeneity, some bias arises when $d_1 = 0.6$. As long as $d \geq 0.8$, estimation of ρ works reasonably well.

5 Conclusion

In this paper we consider testing in a general linear regression framework when the order of possible fractional integration of the error term is unknown. We show that the approach suggested by Vogelsang (1998a) for the $I(0)/I(1)$ case cannot be carried over to the fractional case. Therefore, we suggest an LM-type test which is independent of the order of fractional integration of the error term and allows testing in a general setup. The test statistic exhibits a standard limiting distribution and has non-trivial power against local alternatives. As possible examples covered by our approach we consider testing for polynomial trends, mean shifts and trend breaks. Regarding stochastic regressors, we test restrictions on the fractional cointegrating vector as well as the deterministic components in cointegrating regressions. The LM-type test shows in all considered situations reasonable size and power properties and thus provides a flexible technique for practitioners who are uncertain about the possibly fractional order(s) of integration.

Appendix A

This appendix contains the proofs of our theorems, and some necessary technical lemmas with proofs.

Proof of Theorem 1: Let us assume that y_t follows the data generating process (1)

$$y_t = \beta' z_t + e_t. \tag{22}$$

Under the null hypothesis $H_0 : \beta = 0$ this deduces to

$$y_t = e_t \tag{23}$$

where we now assume that $e_t = \gamma e_{t-1} + \varepsilon_t$ and $\varepsilon_t \sim I(\delta)$ with $\delta \in (-1/2, 1/2)$. Furthermore, we set $e_0 = 0$ as in Assumption 1. In this setup, the process e_t is $I(d)$ with $d = 1+\delta$ if $\gamma = 1$, and $I(\delta)$ if $|\gamma| < 1$. The standard Dickey-Fuller unit root test statistic is given by

$$DF = \frac{\sum_{t=1}^{T} e_{t-1}\varepsilon_t}{\sqrt{\hat{\sigma}_\varepsilon^2 \left(\sum_{t=1}^{T} e_{t-1}^2\right)}}. \tag{24}$$

For the estimator of the residual variance we have

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 &= \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 \\ &= \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2 + o_P(1) \\ &\xrightarrow{P} \sigma_\varepsilon^2. \end{aligned} \tag{25}$$

We furthermore define $\sigma_T^2 = Var(\sum_{t=1}^{T} \varepsilon_t)$. Following along the lines of Sowell (1990), we can write

$$\frac{\sum_{t=1}^{T} e_{t-1}\varepsilon_t}{\hat{\sigma}_\varepsilon \left(\sum_{t=1}^{T} e_{t-1}^2\right)^{1/2}} = \frac{\frac{1}{\sqrt{T}\sigma_T}\sum_{t=1}^{T} e_{t-1}\varepsilon_t}{\frac{1}{\sqrt{T}\sigma_T}\hat{\sigma}_\varepsilon \left(\sum_{t=1}^{T} e_{t-1}^2\right)^{1/2}}. \tag{26}$$

Thus, we have for the denominator

$$\begin{aligned} \frac{1}{\sqrt{T}\sigma_T}\left(\sum_{t=1}^{T} e_{t-1}^2\right)^{1/2} &= \left(\frac{1}{T\sigma_T^2}\sum_{t=1}^{T} e_{t-1}^2\right)^{1/2} \\ &\Rightarrow \left(\int_0^1 B_\delta(s)^2 ds\right)^{1/2}, \end{aligned} \tag{27}$$

where $B_\delta(t)$ denotes fractional Brownian motion. For the numerator, we always have $\varepsilon_t^2 = (e_t - e_{t-1})^2 = e_t^2 + e_{t-1}^2 - 2e_{t-1}e_t$, or $e_{t-1}e_t = (e_t^2 + e_{t-1}^2 - \varepsilon_t^2)/2$, and inserting this in $e_{t-1}\varepsilon_t = e_{t-1}e_t - e_{t-1}^2$ yields

$$\begin{aligned} 2\sum_{t=1}^{T} e_{t-1}\varepsilon_t &= \sum_{t=1}^{T}(e_t^2 - e_{t-1}^2) - \sum_{t=1}^{T}\varepsilon_t^2 \\ &= e_T^2 - \sum_{t=1}^{T}\varepsilon_t^2. \end{aligned} \tag{28}$$

Hence, we have for the numerator

$$\frac{1}{\sqrt{T}\sigma_T}\sum_{t=1}^{T} e_{t-1}\varepsilon_t = \frac{\sigma_T}{2\sqrt{T}}\left(\frac{e_T}{\sigma_T}\right)^2 - \frac{\sqrt{T}}{2\sigma_T}\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right).$$

Since

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{P} E\varepsilon_t^2 < \infty$$

by Assumption 1 and because of $\sigma_T = O_P(T^{1/2+\delta})$, we have if $\delta > 0$ that $\frac{\sigma_T}{2\sqrt{T}} = O_P(T^{\delta})$ and $\frac{\sqrt{T}}{2\sigma_T} \to 0$, and if $\delta < 0$ the vice versa relations. Therefore, in both cases $|DF|^{-1} \xrightarrow{P} 0$ and thus $\exp(-b|DF|^{-1}) \to 1$. Similar arguments hold for $\delta < -1/2$ and thus $d < 1/2$. For the Wald test W we obtain by standard long memory asymptotics

1. $d \in (0, 1/2) : W = O_P(T^{2d})$;
2. $d \in (1/2, 1) : W = O_P(T^{2d+1})$.

This proves the theorem. □

Proof of Theorem 2: Denote by $\hat{\theta}$ the ML-estimator for θ. Under $H_0 : \beta = 0$ we have from Assumption 2 that $\phi(B,\hat{\theta})y_t = \phi(B,\hat{\theta})e_t + o_P(1) = u_t + o_P(1)$. From Assumption 3, we know that $T^{-1}\sum_{t=1}^{T}\phi(B,\theta)z_t\phi(B,\theta)z_t' \to J > 0$. As u_t is a martingale difference sequence by Assumption 1, we have by Davidson (1994), Theorem 24.5 that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t\phi(B,\hat{\theta})z_t \Rightarrow N(0,\sigma^2 J) + o_P(1). \tag{29}$$

Denote by $\hat{\sigma}^2$ a $\sqrt{T}$-consistent estimator for σ^2. Then we have for the LM-test under H_0 that

$$\begin{aligned} LM(\beta) &= \frac{1}{\hat{\sigma}^2}\left(\sum_{t=1}^{T}u_t\phi(B,\hat{\theta})z_t\right)'\left(\sum_{t=1}^{T}(\phi(B,\hat{\theta})z_t)(\phi(B,\hat{\theta})z_t)'\right)^{-1}\left(\sum_{t=1}^{T}u_t\phi(B,\hat{\theta})z_t\right) \\ &\Rightarrow \chi_p^2. \end{aligned}$$

This proves the result in Theorem 2. □

Proof of Theorem 3: The proof of this theorem uses results for the local power of score tests by Harris and Peers (1980). Before applying their results, recall that $\frac{\partial L(0,\theta,\sigma^2)}{\partial\theta}|_{\theta=\hat{\theta}} = 0$ and $\frac{\partial L(0,\theta,\sigma^2)}{\partial\sigma^2}|_{\sigma^2=\hat{\sigma}^2} = 0$, where $\hat{\theta}$ and $\hat{\sigma}^2$ denote the ML-estimators for θ and σ^2, respectively. Following Harris and Peers (1980), the distribution of the score test under the alternative is a non-central χ^2-distribution with non-centrality parameter $\lambda = \beta_T' K|_{\beta=0}\beta_T$, where $K = E\left[\left(\frac{\partial L(\beta,\theta,\sigma^2)}{\partial\beta}\right)'\left(\frac{\partial L(\beta,\theta,\sigma^2)}{\partial\beta}\right)\right]$. Let us first evaluate K. It is given by

$$
\begin{aligned}
K &= \frac{1}{\sigma^4} E\Bigg[\sum_{t=1}^{T} \phi(B,\theta)(y_t-\beta' z_t)(\phi(B,\theta)z_t)'(\phi(B,\theta)z_t)\phi(B,\theta)(y_t-\beta' z_t) \\
&+ \sum_{s,t=1;s\neq t}^{T} \phi(B,\theta)(y_t-\beta' z_t)(\phi(B,\theta)z_t)'(\phi(B,\theta)z_s)\phi(B,\theta)(y_s-\beta' z_s)\Bigg] \\
&\overset{\beta=0}{=} \frac{1}{\sigma^4}\sum_{t=1}^{T} E\Big[\,(\phi(B,\theta)y_t)^2\,(\phi(B,\theta)z_t)'(\phi(B,\theta)z_t)\Big] \\
&+ \sum_{s,t=1;s\neq t}^{T} E\Big[\phi(B,\theta)y_t\phi(B,\theta)y_s(\phi(B,\theta)z_t)'(\phi(B,\theta)z_s)\Big] \\
&= \frac{1}{\sigma^4}\sum_{t=1}^{T} E\Big[u_t^2(\phi(B,\theta)z_t)'(\phi(B,\theta)z_t)\Big] \\
&+ \sum_{s,t=1;s\neq t}^{T} E\Big[u_t u_s(\phi(B,\theta)z_t)'(\phi(B,\theta)z_s)\Big] \\
&= \frac{1}{\sigma^2}\sum_{t=1}^{T}(\phi(B,\theta)z_t)'(\phi(B,\theta)z_t).
\end{aligned}
$$

From Assumption 3, we have that

$$
T^{-1}\sum_{t=1}^{T}(\phi(B,\theta)z_t)'(\phi(B,\theta)z_t) = tr(T^{-1}\sum_{t=1}^{T}(\phi(B,\theta)z_t)(\phi(B,\theta)z_t)') \to \xi^2.
$$

Altogether, we have

$$
\begin{aligned}
\lambda &= \beta_T' K|_{\beta=0}\beta_T \\
&= \frac{1}{\sigma^2}c'\frac{1}{T}\sum_{t=1}^{T}(\phi(B,\theta)z_t)'(\phi(B,\theta)z_t)c \\
&\to \frac{\xi^2}{\sigma^2}c'c,
\end{aligned}
$$

which proves the theorem. □

Before proving Theorem 4, we define a class of admissible filters or possible lag functions $\phi(z,\theta)$, similar to the class H in Robinson (1994). In the strict sense of our approach, this is not really necessary, as we only deal with short-term autocorrelations and possible heteroscedasticity in the error term. However, it basically shows how wide the range of possible filters $\phi(z,\theta)$ can be chosen, allowing for instance also for seasonality.

Assumption 7 *In addition to Assumption 2, we assume furthermore that for real* λ

$$\psi(\lambda) = Re\left(\frac{d}{d\theta}\log\phi(e^{i\lambda}, 0)\right) \tag{30}$$

has finitely many poles ρ_l *on* $(-\pi, \pi]$ *such that* $\|\psi(\lambda)\|$ *is monotonically increasing as* $\lambda \to \rho_{l-}$ *and as* $\lambda \to \rho_{l+}$ *and there exists disjoint intervals* S_l *such that* $\bigcup S_l \in (-\pi, \pi]$, $\rho_l \in S_l$, $\rho_l \notin S_k$ *for* $l \neq k$, *and*

$$\sup_{\lambda \in S_k - (\rho_k - \delta, \rho_k + \delta)} \{|\lambda - \rho_k| \, \|\psi(\lambda) - \psi(\lambda \pm 0.5\delta)\|\} = O(\delta^\eta), \tag{31}$$

as $\delta \to 0$ *and for some* $\eta > 1/2$.

Lemma 1 *Let* $\hat{d}$ *be an estimator of the memory parameter fulfilling Assumption 4. Then, when Assumption 7 is fulfilled,*

$$\frac{1}{\sigma^2}\sum_{t=1}^{T}\left(\frac{\partial}{\partial d}\phi(B, \hat{d})y_t\right)\phi(B, \hat{d})y_t \xrightarrow{P} 0. \tag{32}$$

Proof: We show first similar as in Lemma 1 in Robinson (1994)

$$\lim_{T\to\infty} T^{-1/2} \max_{\lambda_j \in M} \|\psi(\lambda_j, d)\| = 0, \tag{33}$$

where $M = \{\lambda : \pi < \lambda < \pi, \lambda \notin (\rho_l - \lambda_l, \rho_l + \lambda_l), l = 1, \ldots, s\}$. It is

$$\begin{aligned} \max_{\lambda_j \in M} \|\psi(\lambda_j, d)\| &\leq \sum_{l=1}^{r} (\|\psi(\rho_l - \lambda_l, d)\| + \|\psi(\rho_l + \lambda_l, d)\|) \\ &\leq \lambda_l^{-1} \sum_{l=1}^{r} \int_{\rho_l - \lambda_l}^{\rho_l + \lambda_l} \|\psi(\lambda_j, d)\| d\lambda \\ &\leq \left(\frac{2}{\lambda_l}\right)^{1/2} \sum_{l=1}^{r} \left(\int_{\rho_l - \lambda_l}^{\rho_l + \lambda_l} \|\psi(\lambda_j, d)\|^2 d\lambda\right)^{1/2} \\ &= o(T^{1/2}). \end{aligned} \tag{34}$$

From $\log T(\hat{d} - d) = o_P(1)$ we have

$$\lim_{T\to\infty} T^{-1/2} \max_{\lambda_j \in M} \|\psi(\lambda_j, \hat{d})\| = 0. \tag{35}$$

Furthermore, we have

$$E[\phi(B, \hat{d})y_t - \phi(B, d)y_t] = o_P(\log T). \tag{36}$$

This altogether shows that

$$\lim_{T\to\infty} T^{-1/2}\sum_{t=1}^{T}\left(\frac{\partial}{\partial d}\phi(B,\hat{d})y_t\right)\phi(B,\hat{d})y_t \xrightarrow{P} 0 \tag{37}$$

which proves the lemma. □

Lemma 2 *Let $\hat{d}$ be again a consistent estimator for d fulfilling Assumption 4. Then*

$$\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left[(\phi(B,\hat{d})y_t)^2-\sigma^2\right] \xrightarrow{P} 0. \tag{38}$$

Proof: The result follows directly from

$$E[\phi(B,\hat{d})y_t-\phi(B,d)y_t]=o_P(\log T). \tag{39}$$

□

Proof of Theorem 4: From Lemma 1 and Lemma 2 we have that the second and third equation of the score function is zero if a consistent estimator for d is used. Furthermore, we have under H_0 and using Assumption 2,

$$\begin{aligned}\phi(B,\hat{d})y_t &= \phi(B,\hat{d})e_t\\ &= \phi(B,d)e_t+(\hat{d}-d)\frac{\partial}{\partial d}\phi(B,d)e_t\\ &= u_t+o_P(\log T).\end{aligned}$$

Additionally,

$$\begin{aligned}\phi(B,\hat{d})z_t &= \phi(B,d)z_t+(\hat{d}-d)\frac{\partial}{\partial d}\phi(B,d)z_t\\ &= \phi(B,d)z_t+o_P(\log T).\end{aligned}$$

Altogether, we have

$$LM(\beta)^* = LM(\beta)+o_P(\log T), \tag{40}$$

which proves the theorem. □

Proof of Theorem 5: For simplicity of notation denote $b := d_1 - d_0$ in what follows.

For the matrix B we have:

$$T^{-2b}\sum_{t=1}^{T}(\phi(B,\hat{\theta})z_t)(\phi(B,\hat{\theta})z_t)' \Rightarrow \int_0^1 B_b(s)B_b(s)'ds,$$

where $B_b(r) = \frac{1}{\Gamma(b)}\int_0^r (r-s)^{b-1}dB(s)$ is a p-dimensional fractional Brownian motion of order b and $B(s)$ is a standard Brownian motion. From Dolado and Marmol (1999) (see also Davidson and Hashimzade, 2009 or Dolado and Marmol, 2004) we obtain for the matrix A

$$T^{-b}\sum_{t=1}^{T}\phi(B,\hat{\theta})(y_t-\beta_0 z_t)(\phi(B,\hat{\theta})z_t) \Rightarrow \int_0^1 B_b(s)dB(s).$$

Conditional on $\phi(B,\hat{\theta})z_t$ we thus have asymptotically that $A \sim N(0,\int_0^1 B_b(s)B_b(s)'ds)$ which proves the theorem. □

Proof of Theorem 6: The proof follows just the lines as that of Theorem 3 by denoting that

$$E[(\phi(B,\theta)(y_t-\beta_0 z_t))^2\,(\phi(B,\theta)z_t)'(\phi(B,\theta)z_t)] = K < \infty$$

with K being a finite constant by using equation (3.8) in Davidson and Hashimzade (2009). Therefore,

$$\frac{1}{T}\sum_{t=1}^{T}E[(\phi(B,\theta)(y_t-\beta_0 z_t))^2\,(\phi(B,\theta)z_t)'(\phi(B,\theta)z_t)] \rightarrow \xi^2.$$

□

Proof of Theorem 7: The proof is similar to that of Theorem 5. For simplicity of notation denote again $b := d_1 - d_0$ in what follows. For the matrix B we have:

$$T^{-2b}\sum_{t=1}^{T}(\phi(B,\hat{\theta})z_t)(\phi(B,\hat{\theta})z_t)' \Rightarrow \int_0^1 B_b(s)B_b(s)'ds,$$

Analogous to Theorem 5, we obtain

$$T^{-b}\sum_{t=1}^{T}\phi(B,\hat{\theta})(y_t-\beta_0 z_t)(\phi(B,\hat{\theta})z_t) \Rightarrow (1-\sigma_{21}'\Sigma_{22}^{-1})\int_0^1 B_b(s)dB(s).$$

Conditional on $\phi(B,\hat{\theta})z_t$ we obtain the result of the theorem by similar arguments as in the proof of Theorem 5. □

Appendix B

To derive the LM-test statistic, we consider the Gaussian likelihood function of model (1), given as

$$L(\beta, \theta, \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \left(\phi(B;\theta)(y_t - \beta' z_t)\right)^2 . \tag{41}$$

We first consider the score functions

$$\begin{aligned}
\frac{\partial L(\beta, \theta, \sigma^2)}{\partial \beta} &= \frac{1}{\sigma^2} \sum_{t=1}^{T} \left(\phi(B;\theta)(y_t - \beta' z_t)\right) \phi(B;\theta) z_t, \\
\frac{\partial L(\beta, \theta, \sigma^2)}{\partial \theta} &= -\frac{1}{\sigma^2} \sum_{t=1}^{T} \left(\frac{\partial \phi(B;\theta)}{\partial \theta}(y_t - \beta' z_t)\right) \left(\phi(B;\theta)(y_t - \beta' z_t)\right), \\
\frac{\partial L(\beta, \theta, \sigma^2)}{\partial \sigma^2} &= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^{T} \left(\phi(B;\theta)(y_t - \beta' z_t)\right)^2 .
\end{aligned}$$

The portion of the Hessian corresponding to β is

$$\frac{\partial^2 L(\beta, \theta, \sigma^2)}{\partial \beta \partial \beta'} = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \left(\phi(B;\theta) z_t\right) \left(\phi(B;\theta) z_t\right)' .$$

By Assumption 3, this behaves asymptotically as $-TJ/\sigma^2$. For the LM-test of (7), we use the score under the null ($\beta = 0$):

$$\begin{aligned}
\frac{\partial L(0, \theta, \sigma^2)}{\partial \beta} &= \frac{1}{\sigma^2} \sum_{t=1}^{T} \left(\phi(B;\theta) y_t\right) \phi(B;\theta) z_t, \\
\frac{\partial L(0, \theta, \sigma^2)}{\partial \theta} &= -\frac{1}{\sigma^2} \sum_{t=1}^{T} \left(\frac{\partial \phi(B;\theta)}{\partial \theta} y_t\right) \phi(B;\theta) y_t, \\
\frac{\partial L(0, \theta, \sigma^2)}{\partial \sigma^2} &= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^{T} \left(\phi(B;\theta) y_t\right)^2 .
\end{aligned}$$

The maximum likelihood (ML) estimators of θ and σ^2 equate the last two components of the score to zero. The LM-test is constructed as the norm of the remaining score for β, in the appropriate information matrix metric, given from above as

$$\left(-\frac{\partial^2 L(\beta, \theta, \sigma^2)}{\partial \beta \partial \beta'}\right)^{-1} = \sigma^2 \left(\sum_{t=1}^{T} \left(\phi(B;\theta) z_t\right) \left(\phi(B;\theta) z_t\right)'\right)^{-1} ,$$

and substituting estimates for θ and σ^2.

Appendix C

The case where the hypothesis of interest is of the general linear form

$$R\beta = r,$$

with R a given matrix of dimension $k \times p$, for $p > k \geqq 1$, and r a given k-vector, readily reduces to the situation considered in the main text. Thus, select a suitable $(p-k) \times p$ matrix so that the $p \times p$ matrix

$$C = \begin{pmatrix} R \\ S \end{pmatrix}$$

has full rank $p = \dim \beta$. We may without loss of generality reparametrize to

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} R\beta - r \\ S\beta \end{pmatrix},$$

and the hypothesis is recast as

$$H_0 : \mu_1 = 0,$$

with μ_2 free. Writing $s = (r', 0)'$, of dimension $p \times 1$, we have $C\beta = \mu + s$, and hence $\beta = C^{-1}(\mu + s)$, so $\beta' z_t = \left[C^{-1}(\mu + s)\right]' z_t = (\mu + s)' \left[C^{-1}\right]' z_t$. Again without loss of generality, we may transform the regressors to

$$\widetilde{z}_t = \begin{pmatrix} \widetilde{z}_{1t} \\ \widetilde{z}_{2t} \end{pmatrix} = \left[C^{-1}\right]' z_t,$$

with $\widetilde{z}_{1t}$ and $\widetilde{z}_{2t}$ of dimension $k \times 1$ and $(p-k) \times 1$, respectively. Assume there exists an estimator $\widetilde{\beta}$ that is consistent under the null, and let $\widetilde{\mu}_2 = S\widetilde{\beta}$. Then y_t is transformed to $\widetilde{y}_t = y_t - s'\widetilde{z}_t - \widetilde{\mu}_2' \widetilde{z}_{2t}$, or, equivalently,

$$\widetilde{y}_t = y_t - r'\widetilde{z}_{1t} - \widetilde{\mu}_2' \widetilde{z}_{2t}.$$

The transformed regression is

$$\widetilde{y}_t = \mu_1' \widetilde{z}_{1t} + \widetilde{e}_t,$$

in which the zero restriction $\mu_1 = 0$ and hence equivalently the general linear hypothesis $H_0 : R\beta = r$ may be tested in the same manner as the hypothesis $\beta = 0$ in the untransformed regression. Note that since R and r do not depend on parameters, neither does $\widetilde{z}_t$, and for sufficient convergence rate of $\widetilde{\beta}$ under the null, the test of the hypothesis $\mu_1 = 0$ in the general situation may be analyzed along the same lines as that of the hypothesis

$\beta = 0$. Nonlinear hypotheses are handled similarly, writing the hypothesis as

$$f(\beta) = 0,$$

where $f : \boldsymbol{R}^p \to \boldsymbol{R}^k$ may be nonlinear. Let

$$R = \nabla f(\widetilde{\beta}),$$

the $k \times p$ Jacobian of f at $\widetilde{\beta}$, and

$$r = \nabla f(\widetilde{\beta})\widetilde{\beta} - f(\widetilde{\beta}).$$

Now z_t and y_t are transformed as before, and the test carried out on μ_1 (i.e., formally as a test of the linear hypothesis $\mu_1 = 0$).

References

Abadir, K.M., Distaso, W. and Giraitis, L. (2007) : "Nonstationarity-extended local Whittle estimation." *Journal of Econometrics*, 141, 1353 – 1384.

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 B*, 57, 659 – 672.

Christensen, B. J., Nielsen, M. Ø. (2006) : "Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting." *Journal of Econometrics*, 113, 343 – 371.

Davidson, J. (1994) : *Stochastic Limit Theory.* Oxford University Press.

Davidson, J., Hashimzade, N. (2009) : "Representation and weak convergence of stochastic integrals with fractional integrator processes." *Econometric Theory*, 25, 1589 – 1624.

Dolado, J. J., Marmol, F. (1999) : "Asymptotic inference for nonstationary fractionally integrated processes." *Working Paper 99-68, Statistics and Econometrics Series 23, Universidad Carlos III de Madrid.*

Dolado, J. J., Marmol, F. (2004) : "Asymptotic inference results for multivariate long-memory processes." *Econometrics Journal*, 7, 168 – 190.

Engle, R. and Granger, C.W.J. (1987) : "Co-Integration and Error Correction: Representation, Estimation, and Testing." *Econometrica*, 55, 251 – 276.

Geweke, J., Porter-Hudak, S. (1983) : "The estimation and application of long memory time series models." *Journal of Time Series Analysis*, 4, 221 – 238.

Harris, P., Peers, H.W. (1980) : "The local power of the efficient scores test statistic." *Biometrika*, 67, 525 – 529.

Harvey, D.I. and Leybourne, S.J. (2007) : Testing for time series linearity. *Econometrics Journal*, 10, 149 – 165.

Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2009) : "Simple, robust and powerful tests of the breaking trend hypothesis." *Econometric Theory*, 25, 995 – 1029.

Harvey, D.I., Leybourne, S.J. and Xiao, L. (2010) : "Testing for nonlinear deterministic components when the order of integration is unknown. " *Journal of Time Series Analysis*, 31, 379 – 391.

Iacone, F., Leybourne, S.J. and Taylor, A.M.R. (2013a) : "On the behaviour of fixed-b trend break tests under fractional integration." *Econometric Theory*, 29, 393 – 418.

Iacone, F., Leybourne, S.J. and Taylor, A.M.R. (2013b) : "Testing for a break in trend when the order of integration is unknown." *Journal of Econometrics*, 176, 30 – 45.

Jansson, M. and Nielsen, M. Ø. (2012) : "Nearly efficient likelihood ratio tests of the unit root hypothesis." *Econometrica*, 80, 2321 – 2332.

Johansen, M. and Nielsen, M. Ø. (2013) : "The role of initial values in nonstationary fractional time series models." *Working Paper, Queens University.*

Kim, C. S., and Phillips, P. C. B. (2001) : "Fully Modified Estimation of Fractional Cointegration Models. " *Preprint, Yale University.*

Marinucci, D. and Robinson, P. (1999) : "Alternative forms of fractional Brownian motion." *Journal of Statistical Planning and Inference*, 80, 111 – 122.

Nielsen, M. Ø. (2004) : "Optimal residual-based tests for fractional cointegration and exchange rate dynamics." *Journal of Business and Economic Statistics*, 22, 331 – 345.

Nielsen, M. Ø. (2007) : "Local Whittle Analysis of Stationary Fractional Cointegration and the Implied - Realized Volatility Relation." *Journal of Business and Economics Statistics*, 25, 427 – 446.

Perron, P., and Yabu, T. (2009) : ”Testing for Shifts in Trend with an Integrated or Stationary Noise Component.” *Journal of Business and Economics Statistics*, 27, 369 – 396.

Perron, P. and Zhu, X. (2005) : ”Structural breaks with deterministic and stochastic trends.” *Journal of Econometrics*, 129, 65 – 119.

Phillips, P.C.B. (1991) : ”Optimal Inference in Cointegrated Systems.” *Econometrica*, 59, 283 – 306.

Phillips, P.C.B. and Hansen, B.E. (1990) : ”Statistical Inference in Instrumental Variables Regression With I(1) Processes.” *Review of Economic Studies*, 57, 99-125.

Rao, C.R. (1973) : *Linear Statistical Inference and its Applications.* Wiley, New York.

Robinson, P. M. (1994) : ”Efficient Tests of Nonstationary Hypotheses.” *Journal of the American Statistical Association*, 89, 1420 – 1437.

Robinson, P. M. (1995) : ”Gaussian semiparametric estimation of long range dependence.” *Annals of Statistics*, 23, 1048 – 1072.

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

Shimotsu, K. and Phillips, P.C.B. (2005) : ”Exact local Whittle estimation of fractional integration.” *Annals of Statistics*, 33, 1890 – 1933.

Shimotsu, K. (2010) : ”Exact local Whittle estimation of fractional integration with unknown mean and time trend.” *Econometric Theory*, 26, 501 – 540.

Sowell, F. (1990) : ”The fractional unit root distribution.” *Econometrica*, 58, 495 – 505.

Vogelsang, T.J. (1998a) : ”Trend function hypothesis testing in the presence of serial correlation.” *Econometrica*, 66, 123 – 148.

Vogelsang, T.J. (1998b) : ”Testing for a shift in mean without having to estimate serial-correlation parameters.” *Journal of Business and Economic Statistics*, 16, 73 – 80.

Xu, J. and Perron, P. (2013) : ”Robust testing of time trend and mean with unknown integration order errors.” *Working Paper, Boston University.*

Zambrano, E. and Vogelsang, T.J. (2000) : ”A simple test of the law of demand for the United States.” *Econometrica*, 68, 1011 – 1020.