IHS Economics Series
Working Paper 264
March 2011

Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference

Seung Hyun Hong
Martin Wagner

Impressum

Author(s):
Seung Hyun Hong, Martin Wagner

Title:
Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference

ISSN: Unspecified

2011 Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
E-Mail: office@ihs.ac.at
Web: www.ihs.ac.at

264

Reihe Ökonomie

Economics Series

Cointegrating Polynomial Regressions:

Fully Modified OLS Estimation and Inference

Seung Hyun Hong, Martin Wagner

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

264
Reihe Ökonomie
Economics Series

Cointegrating Polynomial Regressions:

Fully Modified OLS Estimation and Inference

Seung Hyun Hong, Martin Wagner

March 2011

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Seung Hyun Hong
Korea Institute of Public Finance
Seoul, Korea

Martin Wagner
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56
1060 Vienna, Austria
Vienna, Austria
☎: +43/1/599 91-150
fax: +43/1/599 91-555
email: mawagner@ihs.ac.at
and
Frisch Centre for Economic Research
Oslo, Norway

Abstract

This paper develops a fully modified OLS estimator for cointegrating polynomial regressions, i.e. for regressions including deterministic variables, integrated processes and powers of integrated processes as explanatory variables and stationary errors. The errors are allowed to be serially correlated and the regressors are allowed to be endogenous. The paper thus extends the fully modified approach developed in Phillips and Hansen (1990). The FM-OLS estimator has a zero mean Gaussian mixture limiting distribution, which is the basis for standard asymptotic inference. In addition Wald and LM tests for specification as well as a KPSS-type test for cointegration are derived. The theoretical analysis is complemented by a simulation study which shows that the developed FM-OLS estimator and tests based upon it perform well in the sense that the performance advantages over OLS are by and large similar to the performance advantages of FM-OLS over OLS in cointegrating regressions.

Keywords

Cointegrating polynomial regression, fully modified OLS estimation, integrated process, testing

JEL Classification

C12, C13, C32

Comments

The helpful comments of Peter Boswijk, Robert Kunst, Benedikt M. Pötscher, Martin Stürmer, Timothy J. Vogelsang, of seminar participants at Helmut Schmidt University Hamburg, University of Linz, University of Innsbruck, Universitat Autònoma Barcelona, University of Tübingen, Tinbergen Institute Amsterdam, Korea Institute of Public Finance, Bank of Korea, Korea Institute for International Economic Policy, Korea Securities Research Institute, Korea Insurance Research Institute, Michigan State University, Institute for Advanced Studies and of conference participants at the 2nd International Workshop in Computational and Financial Econometrics in Neuchâtel, an ETSERN Workshop in Copenhagen, the 3rd Italian Congress of Econometrics and Empirical Economics in Ancona, the Econometric Study Group Meeting of the German Economic Association, the Econometric Society European Meeting in Barcelona, the Meeting of the German Economic Association in Magdeburg and the Statistische Woche in Nuernberg are gratefully acknowledged. Furthermore, we would like to thank in particular the co-editor, the associate editor and two anonymous referees for their extremely valuable comments and suggestions. The usual disclaimer applies.

Contents

1 Introduction

This paper develops a fully modified OLS (FM-OLS) estimator for *cointegrating polynomial regressions* (CPRs), i.e. for regressions including deterministic variables, integrated processes and integer powers of integrated processes as explanatory variables and stationary errors. As in the standard cointegration literature the errors are allowed to be serially correlated and the regressors are allowed to be endogenous.[1] Thus, the paper extends the FM-OLS estimator of Phillips and Hansen (1990) from cointegrating (linear) regressions to cointegrating polynomial regressions. A major advantage of considering regressions of the considered form is that they are linear in parameters, which implies that linear least squares based estimation methods can be developed and that it is not necessary to consider nonlinear estimation techniques that require numerical optimization procedures. Clearly, the considered class of functions is restrictive, despite the fact that polynomials can be used to approximate more general nonlinear functions, but clearly this restriction is the price to be paid for having a simple linear least squares based estimation technique available (see Section 2.1 for further discussion). Additionally also specification and cointegration tests are developed. With respect to specification testing amongst other things this paper extends the work of Hong and Phillips (2010), who consider LM-type specification testing based on residuals of cointegrating linear relationships, in several aspects (see Section 2.3). With respect to asymptotic theory our work relies upon important contributions of Chang, Park, and Phillips (2001), Park and Phillips (1999, 2001) and Ibragimov and Phillips (2008).

One motivation for considering CPRs is given by the so-called environmental Kuznets curve (EKC) hypothesis, which postulates an inverse U-shaped relationship between measures of economic activity (typically proxied by per capita GDP) and pollution. The term EKC refers by analogy to the inverted U-shaped relationship between the level of economic development and the degree of income inequality postulated by Kuznets (1955) in his 1954 presidential address to the American Economic Association. Since the seminal work of Grossmann and Krueger (1995) more than one-hundred refereed publications (as counted already several years ago by Yandle, Bjattarai, and Vijayaraghavan, 2004) perform econometric analysis of EKCs.[2] Many of these empirical studies use unit root and cointegration techniques and include as regressors powers of per capita GDP (in

[1] The theory is developed to allow also for predetermined stationary regressors, which are neglected from the discussion here for the sake of brevity. They are included in the discussion in Hong and Wagner (2008).

[2] In addition to the vast empirical literature there is also a large theoretical literature exploring different mechanisms leading to EKC type relationships, see e.g. the survey Brock and Taylor (2005).

order to allow to for U- or inverted U-shaped relationships). This literature neglects throughout that powers, as special cases of nonlinear functions, of integrated processes are not themselves integrated processes, which invalidates the use of standard unit root and cointegration techniques. Such relationships – in our terminology cointegrating polynomial relationships – necessitate the development of appropriate estimation and inference tools, which is done in this paper. For a more detailed discussion of the empirical EKC literature, its problems as well as a detailed analysis using the methods developed in this paper see Hong and Wagner (2008, 2010). A second strand of the empirical literature that can benefit from the theory developed in this paper is the so-called intensity of use literature that investigates the potentially also inverted U-shaped relationship between GDP and energy or metals use (see e.g. Labson and Crompton, 1993).

As in standard cointegrating regressions, the OLS estimator is consistent also in CPRs. Also as in the standard case, its limiting distribution is contaminated by so-called second order bias terms in case of error serial correlation and/or endogeneity of regressors (see the original work of Phillips and Hansen, 1990). This renders valid inference difficult. Consequently, we develop an FM-OLS estimator, which extends the FM-OLS estimator of Phillips and Hansen (1990) to CPRs, that has a zero mean Gaussian mixture limiting distribution and thus allows for standard asymptotic chi-square inference. The zero mean Gaussian mixture limiting distribution of the FM-OLS estimator also forms the basis for specification testing based on augmented (Wald tests) respectively auxiliary (LM tests) regressions. On top of these specification tests we also consider a KPSS-type test as a direct test for nonlinear cointegration of the considered form. The asymptotic distribution of this test depends on the specification of the deterministic components as well as the number and powers of integrated regressors included. This test extends the cointegration test of Shin (1994) from cointegrating to cointegrating polynomial regressions. We furthermore follow Choi and Saikkonen (2010) and consider also a sub-sample test that can be used in conjunction with the Bonferroni bound and which has a limiting distribution independent of the specification.

The theoretical analysis of the paper is complemented by a simulation study to assess the performance of the estimator and tests, with the performance being benchmark against results obtained by applying OLS. Many of the findings with respect to both estimator performance (measured in terms of bias and RMSE) as well as the performance of the coefficient tests (size distortions, size corrected power) are similar as for FM-OLS in standard cointegrating relationships. Summarized in one sentence: the larger the extent of serial correlation and/or regressor endogeneity, the bigger are the performance advantages of the FM-OLS estimator and test statistics based upon it. For sizeable serial correlation and endogeneity the bias can be reduced by about 50% when using the FM-OLS instead of the OLS estimates and the over-rejections that occur for the FM-OLS based tests are less than half as big as for OLS based statistics. The choice of kernel and bandwidth is of relatively minor importance. With respect to the specification tests it turns out that the Wald tests are outperformed by the LM tests, since the latter have essentially the same, or for small sample sizes slightly lower, size corrected power but much smaller over-rejections under the null hypothesis than the former. The simulations also show that using as additional regressors both higher order deterministic trends (originally considered in a unit root and cointegration test context by Park and Choi, 1988; Park, 1990) together with higher polynomial powers of integrated regressors leads to highest power against the variety of alternatives considered. In the simulations the performance of the KPSS-type tests is rather poor. As is well known for KPSS-type tests, their performance is detrimentally affected by the presence of serial correlation, which is also confirmed by our simulations. The sub-sample test suffers additionally from the conservativeness of the Bonferroni bound and performs worse than the full sample test.

The paper is organized as follows. In Section 2 we derive the asymptotic results for the estimators and tests. Section 3 contains a small simulation study to assess the finite sample performance of the proposed methods and Section 4 briefly summarizes and concludes. The proofs of all propositions are relegated to the appendix. Available supplementary material contains further results in relation to the sub-sample KPSS-type test as well as additional simulation results.

We use the following notation: Definitional equality is signified by $:=$ and $\Rightarrow$ denotes weak convergence. Brownian motions, with covariance matrices specified in the context, are denoted with $B(r)$ or B. For integrals of the form $\int_0^1 B(s)ds$ and $\int_0^1 B(s)dB(s)$ we use short-hand notation $\int B$ and $\int BdB$. For notational simplicity we also often drop function arguments. With $\lfloor x \rfloor$ we denote the integer part of $x \in \mathbb{R}$ and $\text{diag}(\cdot)$ denotes a diagonal matrix with the entries specified throughout. For a square matrix A we denote its determinant with $|A|$, for a vector $x = (x_i)$ we denote by $||x||^2 = \sum_i x_i^2$ and for a matrix M we denote by $||M|| = \sup_x \frac{||Mx||}{||x||}$. $\mathbb{E}$ denotes the expected value and L denotes the backward-shift operator, i.e. $L\{x_t\}_{t\in\mathbb{Z}} = \{x_{t-1}\}_{t\in\mathbb{Z}}$.

2 Theory

2.1 Setup and Assumptions

We consider the following equation including a constant and polynomial time trends up to power q (see the discussion below), integer powers of integrated regressors $x_{jt}, j = 1, \dots, m$ up to degrees p_j and a stationary error term u_t:

$$y_t = D_t'\theta_D + \sum_{j=1}^{m} X_{jt}'\theta_{X_j} + u_t \quad , \quad \text{for } t = 1, \dots, T, \tag{1}$$

with $D_t := [1, t, t^2, \dots, t^q]'$, $x_t := [x_{1t}, \dots, x_{mt}]'$, $X_{jt} := [x_{jt}, x_{jt}^2, \dots, x_{jt}^{p_j}]'$ and the parameter vectors $\theta_D \in \mathbb{R}^{q+1}$ and $\theta_{X_j} \in \mathbb{R}^{p_j}$. Furthermore define for later use $X_t := [X_{1t}', \dots, X_{mt}']'$, $Z_t := [D_t', X_t']'$ and $p := \sum_{j=1}^{m} p_j$. In a more compact way we can rewrite (1) as

$$\begin{aligned} y &= D\theta_D + X\theta_X + u \\ &= Z\theta + u, \end{aligned} \tag{2}$$

with $y := [y_1, \dots, y_T]'$, $u := [u_1, \dots, u_T]'$, $Z := [D \;\; X]$ and $\theta = [\theta_D' \;\; \theta_X']' \in \mathbb{R}^{(q+1)+p}$ and

$$D := \begin{bmatrix} D_1' \\ \vdots \\ D_T' \end{bmatrix} \in \mathbb{R}^{T \times (q+1)}, \; X := \begin{bmatrix} X_1' \\ \vdots \\ X_T' \end{bmatrix} \in \mathbb{R}^{T \times p}.$$

Equation (1) is referred to as cointegrating polynomial regression (CPR). Clearly it is a special case of a nonlinear cointegrating relationship as considered in the literature (for recent examples see e.g. Karlsen, Myklebust, and Tjostheim, 2007; Wang and Phillips, 2009) where typically any relationship of the form $y_t = f(x_t, \theta) + u_t$, with x_t an integrated process, u_t stationary and $f(\cdot, \cdot)$ a nonlinear function, is considered to be a nonlinear cointegrating relationship between y_t and x_t.[3] The econometric literature has not yet provided definite answers to the problem of how to extend the concepts of integrated and cointegrated processes, which are concepts inherently related to linear processes, to the nonlinear world. In the formulation just given, e.g. a minimum requirement for a useful extension of the concept clearly is to exclude cointegration in x_t, since otherwise (this example

[3]Any such formulation by construction treats y_t and x_t asymmetrically, with the former being a function (up to u_t) of the latter which is assumed to be integrated in the 'usual' sense of the word. In the linear cointegration framework, under the assumption that there is no cointegration between the components of x_t, this implies that in a triangular system of the form $y_t = x_t'\theta + u_t$, $x_t = x_{t-1} + v_t$ also y_t is integrated and thus this asymmetric formulation is innocuous. In a nonlinear framework the stochastic properties of y_t are in general unclear, when y_t is generated by $y_t = f(x_t, \theta) + u_t$, $x_t = x_{t-1} + v_t$.

is taken from Choi and Saikkonen, 2010) a nonlinear function of the form $f(x_t, \theta) = x_t'\theta + (x_t'\theta)^2$, with θ a cointegrating vector of x_t, would lead to meaningless forms of nonlinear cointegration (some simple examples are also discussed in Granger and Hallman, 1991). The appeal of certain types of nonlinear functions to be used in nonlinear cointegration analysis, stems from the implied stochastic properties of $f(x_t, \theta)$ and consequently of y_t. In this respect the use of integer powers of integrated processes is appealing since due to the simplicity of this formulation the stochastic properties of $f(x_t, \theta)$ can be understood to a certain extent. Polynomial transformations of integrated processes are (see the discussion in Ermini and Granger, 1993, Section 3) in many ways from an empirical perspective similar to random walks with trend components. E.g. their sample autocorrelations decay very slowly, i.e. there is high persistence, which makes it difficult to distinguish them from unit root processes in samples typically available. However, they are not integrated processes according to any of the usual definitions, since no difference of any order is a covariance stationary process.[4] Under the assumption that the elements of x_t are not cointegrated, also $\sum_{j=1}^{m} X_{jt}'\theta_{X_j}$ – and thus y_t as given by (1) – behaves like a polynomial transformation of an integrated process, i.e. is empirically hard to distinguish from a random walk with a trend component.[5] This ability of CPRs to generate variables that appear very similar to random walks with trend components makes CPRs in our view a useful and simple framework for nonlinear cointegration analysis.

Let us now state the assumptions concerning the regressors and the error processes:

Assumption 1 *The processes $\{\Delta x_t\}_{t\in\mathbb{Z}}$ and $\{u_t\}_{t\in\mathbb{Z}}$ are generated as*

$$\begin{aligned}\Delta x_t = v_t &= C_v(L)\varepsilon_t = \sum_{j=0}^{\infty} c_{vj}\varepsilon_{t-j}\\ u_t &= C_u(L)\zeta_t = \sum_{j=0}^{\infty} c_{uj}\zeta_{t-j},\end{aligned}$$

with the conditions

$$det(C_v(1)) \neq 0, \quad \sum_{j=0}^{\infty} j||c_{vj}|| < \infty \ , \quad \sum_{j=0}^{\infty} j|c_{uj}| < \infty.$$

[4]For the simple case of the square of a random walk this is e.g. discussed in Granger (1995, Example 2). Various attempts have been made to generalize the concept of integration beyond the usual framework that is essentially based on sums of linear processes. One of them is the concept of extended memory processes of Granger (1995) and another example is given by the so-called summability index of Berenguer-Rico and Gonzalez (2010), which is defined (for our setup) as the rate of divergence of stochastic processes. It holds that $T^{-(1+\frac{p}{2})}\sum_{t=1}^{T} x_t^p$, for x_t a scalar I(1) process, converges. The summability index of x_t^p, i.e. the divergence order of $T^{-1/2}\sum_t x_t^p$, is therefore $\frac{p+1}{2}$.

[5]In the words of Berenguer-Rico and Gonzalez (2010), the summability index of $\sum_{j=1}^{m} X_{jt}'\theta_{X_j}$ and by construction also of y_t is equal to $\max_{j=1,\ldots,m} \frac{p_j+1}{2}$.

Furthermore we assume that the process $\{\xi_t^0\}_{t\in\mathbb{Z}} = \{[\varepsilon_t', \zeta_t]'\}_{t\in\mathbb{Z}}$ *is a stationary and ergodic martingale difference sequence with natural filtration* $\mathcal{F}_t = \sigma\left(\{\xi_s\}_{-\infty}^t\right)$ *and denote the (conditional) covariance matrix by*

$$\Sigma^0 = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\zeta} \\ \Sigma_{\zeta\varepsilon} & \sigma_\zeta^2 \end{bmatrix} := \mathbb{E}(\xi_t^0(\xi_t^0)'|\mathcal{F}_{t-1}) > 0.$$

In addition we assume that $\sup_{t\geq 1}\mathbb{E}(\|\xi_t^0\|^r|\mathcal{F}_{t-1}) < \infty$ *a.s. for some* $r > 4$.

Assumptions similar to the ones used above have been used in several places in the literature with all of these assumptions geared towards establishing an invariance principle for (in our case of cointegrating polynomial regressions) terms like $T^{-\frac{k+1}{2}}\sum_{t=1}^T x_{jt}^k u_t$. Our assumptions are most closely related to those of Chang, Park, and Phillips (2001), Park and Phillips (1999, 2001) and Hong and Phillips (2010).[6] Alternatively we could refer to the martingale theory framework of Ibragimov and Phillips (2008, Theorem 3.1 and Remark 3.3) to establish convergence of the cross-product just given above. Their assumptions are cast in terms of linear processes with moment conditions related in our context to the order of the polynomial considered. Yet a different route has been taken by de Jong (2002, Assumptions 1 and 2) who resorts in his assumptions on the underlying processes to the concept of near epoch dependent sequences and some moment conditions. For the present paper essentially any set of assumptions that leads to the required invariance principle is fine and it is not the purpose of this paper to provide a new set of conditions. The assumption $\det(C_v(1)) \neq 0$ together with positive definiteness of $\Sigma_{\varepsilon\varepsilon}$ implies that x_t is an integrated but not cointegrated process.

Clearly the stated assumption is strong enough to allow for an invariance principle to hold for $\{\xi_t\}_{t\in\mathbb{Z}} = \{[v_t', u_t]'\}_{t\in\mathbb{Z}}$ using the Beveridge-Nelson decomposition (compare Phillips and Solo, 1992)

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\xi_t \Rightarrow B(r) = \begin{bmatrix} B_v(r) \\ B_u(r) \end{bmatrix}. \tag{3}$$

Note here that it holds that $B(r) = \Omega^{1/2}W(r)$ with the long-run covariance matrix $\Omega := \sum_{h=-\infty}^{\infty}\mathbb{E}\left(\xi_0\xi_h'\right)$. We also define the one-sided long-run covariance $\Delta := \sum_{h=0}^{\infty}\mathbb{E}\left(\xi_0\xi_h'\right)$ and both covariance matrices are partitioned according to the partitioning of ξ_t, i.e.:

$$\Omega = \begin{bmatrix} \Omega_{vv} & \Omega_{vu} \\ \Omega_{uv} & \omega_{uu} \end{bmatrix} \quad , \quad \Delta = \begin{bmatrix} \Delta_{vv} & \Delta_{vu} \\ \Delta_{uv} & \Delta_{uu} \end{bmatrix}.$$

[6] The main difference to the assumptions of Chang, Park, and Phillips (2001) is, using the notation of this paper, that they assume $u_t = \zeta_t$ and they also assume that all regressors are predetermined, i.e. in our notation they assume that $\{[\varepsilon_{t+1}', \zeta_t]'\}_{t\in\mathbb{Z}}$ is a stationary and ergodic martingale difference sequence.

When referring to quantities corresponding to only one of the nonstationary regressors and its powers, e.g. X_{jt}, we use the according notation, e.g. $B_{v_j}(r)$ or $\Delta_{v_j u}$.

To study the asymptotic behavior of the estimators, we next introduce appropriate weighting matrices, whose entries reflect the divergence rates of the corresponding variables. Thus, denote with $G(T) = diag\{G_D(T), G_X(T)\}$, where for notational brevity we often use $G := G(T)$. The two diagonal sub-matrices are given by $G_D(T) := \text{diag}(T^{-1/2}, \ldots, T^{-(q+1/2)}) \in \mathbb{R}^{(q+1)\times(q+1)}$ and $G_X(T) := \text{diag}(G_{X_1}, \ldots, G_{X_m}) \in \mathbb{R}^{p\times p}$ with $G_{X_j} := \text{diag}(T^{-1}, \ldots, T^{-\frac{p_j+1}{2}}) \in \mathbb{R}^{p_j\times p_j}$.

Using these weighting matrices, we can define the following limits of the major building blocks. For t such that $\lim_{T\to\infty} t/T = r$ the following results hold:

$$
\begin{aligned}
\lim_{T\to\infty} \sqrt{T} G_D(T) D_t &= \lim_{T\to\infty} \begin{pmatrix} 1 & & \\ & \ddots & \\ & & T^{-q} \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ t^q \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ r^q \end{pmatrix} =: D(r) \\
\lim_{T\to\infty} \sqrt{T} G_{X_j}(T) X_{jt} &= \lim_{T\to\infty} \begin{pmatrix} T^{-1/2} & & \\ & \ddots & \\ & & T^{-p_j/2} \end{pmatrix} \begin{pmatrix} x_{jt} \\ \vdots \\ x_{jt}^{p_j} \end{pmatrix} = \begin{pmatrix} B_{v_j} \\ \vdots \\ B_{v_j}^{p_j} \end{pmatrix} =: \mathbf{B}_{v_j}(r),
\end{aligned}
$$

separating here the coordinates of $v_t = [v_{1t}, \ldots, v_{mt}]'$ corresponding to the different variables x_{jt}. The first result is immediate and the second follows e.g. from Chang, Park, and Phillips (2001, Lemma 5). The stacked vector of the scaled polynomial transformations of the integrated processes is denoted as $\mathbf{B}_v(r) := [\mathbf{B}_{v_1}(r)', \ldots, \mathbf{B}_{v_m}(r)']'$. We are confident that D as defined in (2) is not confused with $D(r)$ defined above even when the latter is used in abbreviated form D in integrals.

More general deterministic components can be included with the necessary condition being that the correspondingly defined limit quantity satisfies $\int DD' > 0$, i.e. that the considered functions are linearly independent in $L^2[0,1]$. This allows in addition to the polynomial trends on which we focus in this paper e.g. also to include time dummies, broken trends or trigonometric functions of time (compare the discussion in Park, 1992). As has been mentioned, the working paper Hong and Wagner (2008) extends the considered regression model by additionally including predetermined stationary regressors, similarly to the model considered in Chang, Park, and Phillips (2001). For brevity we do not include these components here but refer the reader to the mentioned working paper. Note also that the available code allows for stationary regressors.

Furthermore note that the results in this paper extend to triangular systems with multivariate

y_t, as considered in the linear case in Phillips and Hansen (1990), with the required changes to assumptions, expressions and results being straightforward.

2.2 Fully Modified OLS Estimation

As in a (standard) linear cointegrating regression (see Phillips and Hansen, 1990) also in the considered cointegrating polynomial regression situation the usual OLS estimator $\hat{\theta} := (Z'Z)^{-1}Z'y$ of θ is consistent, but its limiting distribution is contaminated by second order bias terms (as shown in the proof of Proposition 1 in the appendix). The presence of these second order bias terms invalidates standard inference and consequently we consider an appropriate fully modified OLS (FM-OLS) estimator. The principle is like in the linear cointegration case, i.e. the fully modified estimator is based on two modifications to the OLS estimator: (i) the dependent variable y_t is replaced by a suitably constructed variable y_t^+ and (ii) additive correction factors are employed.

The dependent variable is modified in the same way as in Phillips and Hansen (1990), i.e. $y_t^+ := y_t - v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$ and $y^+ := [y_1^+, \ldots, y_T^+]'$.[7] Note that for notational brevity in the remainder of the paper we simply assume here that v_1 is available. In an application, with $x_1, \ldots, x_T$ available, only $v_2, \ldots, v_T$ can be actually computed. Thus, in applications FM-OLS computations are typically performed on the sample $t = 2, \ldots, T$.[8] Assuming for the purpose of this paper that v_1 is available saves us from introducing throughout additional, cumbersome notation for data matrices comprising observations only from $t = 2, \ldots, T$ rather than from $t = 1, \ldots, T$.

The additive correction factors are different than in the linear case and are given by

$$M^* := \begin{bmatrix} M_1^* \\ \vdots \\ M_m^* \end{bmatrix}, \quad M_j^* := \hat{\Delta}_{v_j u}^+ \begin{bmatrix} T \\ 2\sum x_{jt} \\ \vdots \\ p_j \sum x_{jt}^{p_j - 1} \end{bmatrix}, \tag{4}$$

In both the definition of y^+ and the correction factors we rely upon consistent estimators of the required long-run variances, $\hat{\Omega}_{vv}$, $\hat{\Omega}_{vu}$, $\hat{\Delta}_{v_j u}$ and $\hat{\Delta}_{v_j u}^+ := \hat{\Delta}_{v_j u} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Delta}_{vv_j}$.

The OLS estimator is consistent, despite the fact that its limiting distribution is contaminated by second order bias terms. This result is important, given that the OLS residuals are used for long-run variance estimation. For our setup, the results of Jansson (2002, Corollary 3) apply, because

[7] Note that here and throughout we ignore for notational simplicity the dependence of e.g. y^+ upon the specific consistent long-run covariance estimator chosen.

[8] Sometimes also the assumption $x_0 = 0$ is made, which also gives an actual sample of size T. Asymptotically none of these choices has an effect.

the OLS estimator converges sufficiently fast. Thus, the assumptions with respect to kernel (A3) and bandwidth choice (A4) formulated in Jansson have to be taken into account. For more explicit calculations with respect to long-run variance estimation in a context related to this paper see also Hong and Phillips (2010). For the remainder of the paper we assume from now on that long-run variance estimation is performed consistently.

With the necessary notation collected, the following Proposition 1 gives the result for the FM-OLS estimator (where as mentioned the limiting distribution of the OLS estimator is also given in the proof contained in the appendix).

Proposition 1 *Let y_t be generated by* (1) *with the regressors Z_t and errors u_t satisfying Assumption 1. Define the FM-OLS estimator of θ as*

$$\hat{\theta}^+ := (Z'Z)^{-1}\left(Z'y^+ - A^*\right),$$

with

$$A^* := \left[\begin{array}{c} 0_{(q+1)\times 1} \\ M^* \end{array}\right],$$

with M^* as given in (4) *and with consistent estimators of the required long-run (co)variances. Then $\hat{\theta}^+$ is consistent and its asymptotic distribution is given by*

$$G^{-1}\left(\hat{\theta}^+ - \theta\right) \quad \Rightarrow \quad \left(\int JJ'\right)^{-1}\int JdB_{u.v}, \tag{5}$$

with $J(r) := [D(r)' \;\; \mathbf{B}_v(r)']'$ and $B_{u.v}(r) := B_u(r) - B_v(r)'\Omega_{vv}^{-1}\Omega_{vu}$.

This limiting distribution is free of second order bias terms and is a zero mean Gaussian mixture. This stems from the fact that $B_{u.v}$ is by construction independent of the vector $\mathbf{B}_v$, being independent of B_v. This in turn implies that conditional upon $\mathbf{B}_v$, the above limiting distribution is actually a normal distribution with (conditional) covariance matrix

$$V_{FM} = \omega_{u.v}\left(\int JJ'\right)^{-1}. \tag{6}$$

By definition of $B_{u.v}$ it holds that $\omega_{u.v} := \omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$. Clearly, when using a consistent estimator $\hat{\omega}_{u.v} = \hat{\omega}_{uu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$, a consistent estimator of this conditional covariance matrix is given by $\hat{V}_{FM} = \hat{\omega}_{u.v}(GZ'ZG)^{-1}$.

Sometimes it is convenient to have separate explicit expressions for the coefficients corresponding to the deterministic components on the one hand and the stochastic regressors on the other (for details see the derivations in the working paper Hong and Wagner, 2008). Such an expression obviously follows using partitioned matrix inversion of $\left(\int JJ'\right)^{-1}$, and is given by

$$G^{-1}\left(\hat{\theta}^{+}-\theta\right)=\left[\begin{array}{c} G_D^{-1}(\hat{\theta}_D^{+}-\theta_D) \\ G_X^{-1}(\hat{\theta}_X^{+}-\theta_X) \end{array}\right] \Rightarrow \left[\begin{array}{c} \left[\int \tilde{D}\tilde{D}'\right]^{-1}\int \tilde{D}dB_{u.v} \\ \left[\int \tilde{\mathbf{B}}_v\tilde{\mathbf{B}}_v'\right]^{-1}\int \tilde{\mathbf{B}}_v dB_{u.v} \end{array}\right], \tag{7}$$

with

$$\begin{aligned} \tilde{D} &:= D-\int D\mathbf{B}_v'\left(\int \mathbf{B}_v\mathbf{B}_v'\right)^{-1}\mathbf{B}_v\ , \\ \tilde{\mathbf{B}}_v &:= \mathbf{B}_v-\int \mathbf{B}_v D'\left(\int DD'\right)^{-1}D. \end{aligned}$$

The zero mean Gaussian mixture limiting distribution given in (5) forms the basis for asymptotic chi-square inference, as discussed in Phillips and Hansen (1990, Theorem 5.1 and the discussion on p. 106). Since in the considered regression the convergence rates differ across coefficients, not all hypothesis can be tested, as is well known (compare Phillips and Hansen, 1990; Sims, Stock and Watson, 1990; Vogelsang and Wagner, 2010). We here merely state a sufficient condition on the constraint matrix $R \in \mathbb{R}^{s\times q+1+p}$ under which the Wald statistics have chi-square limiting distributions. We assume that there exists a nonsingular scaling matrix $G_R \in \mathbb{R}^{s\times s}$ such that

$$\lim_{T\to\infty} G_R R G = R^*, \tag{8}$$

where $R^* \in \mathbb{R}^{s\times q+1+p}$ has rank s. Clearly, this covers as a special case testing of multiple hypotheses, where in none of the hypotheses coefficients with different convergence rates are mixed (e.g. t-tests or testing the significance of several parameters jointly) but allows for more general hypotheses.

Proposition 2 *Let y_t be generated by* (1) *with the regressors Z_t and errors u_t satisfying Assumption 1. Consider s linearly independent restrictions collected in*

$$H_0 : R\theta = r,$$

with $R \in \mathbb{R}^{s\times q+1+p}$ with row full rank s and $r \in \mathbb{R}^s$ and suppose that there exists a matrix G_R such that (8) *is fulfilled. Furthermore let $\hat{\omega}_{u.v}$ denote a consistent estimator of $\omega_{u.v}$. Then it holds that the Wald statistic*

$$W := \left(R\hat{\theta}^{+}-r\right)'\left[\hat{\omega}_{u.v}R\left(Z'Z\right)^{-1}R'\right]^{-1}\left(R\hat{\theta}^{+}-r\right) \to \chi_s^2 \tag{9}$$

under the null hypothesis.

The above result implies, as mentioned, that for instance the appropriate t-statistic for an individual coefficient θ_i, given by $t_{\theta_i} := \frac{\hat{\theta}_i^+}{\sqrt{\hat{\omega}_{u.v}(Z'Z)^{-1}_{[i,i]}}}$, is asymptotically standard normally distributed.

Note that analogously to the Wald test also a corresponding Lagrange Multiplier (LM) test statistics can be derived. For brevity we consider the LM test only in the following subsection, when dealing with specification testing based on an augmented respectively auxiliary regression (see Proposition 4).

2.3 Specification Testing based on Augmented and Auxiliary Regressions

Testing the correct specification of equation (1) is clearly an important issue. In this respect we are particularly interested in the prevalence of cointegration, i.e. stationarity of u_t. Absence of cointegration can be due to several reasons. First, there is no cointegrating relationship of any functional form between y_t and x_t. Second, y_t and x_t are nonlinearly cointegrated but the functional relationship is different than postulated by equation (1). This case covers the possibilities of missing higher order deterministic components or higher order polynomial terms or of cointegration with an entirely different functional form. Third, the absence of cointegration is due to missing explanatory variables in equation (1).

In a general formulation all the above possibilities can be cast into a testing problem within the augmented regression

$$y_t = Z_t'\theta + F(\bar{D}_t, x_t, q_t, \theta_F) + \phi_t, \tag{10}$$

where F is such that $F(\bar{D}_t, x_t, q_t, 0) = 0$, where $\bar{D}_t$ denotes the set of deterministic variables considered (like e.g. higher order time trends) and q_t denotes additional integrated regressors. If cointegration prevails in (1) then $\theta_F = 0$ and $\phi_t = u_t$ in (10)

In many cases the researcher will not have a specific parametric formulation in mind for the function $F(\cdot)$, which implies that typically the unknown $F(\cdot)$ is replaced by a partial sum approximation. This approach has a long tradition in specification testing in a stationary setup, see Ramsey (1969), Phillips (1983), Lee, White, and Granger (1993) or de Benedictis and Giles (1998). Given our FM-OLS results it appears convenient to replace the unknown $F(\cdot)$ by using the additional deterministic variables and additional powers of the integrated regressors. The latter in the most general case include both higher order powers larger than p_j for the components x_{jt} of x_t as well as powers larger or equal than 1 for the additional integrated regressors q_{it}.

Of course this simple approach is also subject to the discussion in the introduction that a simple functional form is considered. However, for specification analysis the advantage of a parsimonious setup may outweigh the potential disadvantages of considering only univariate polynomials since a test based on such a formulation will also have power against alternatives where e.g. products terms are present. Clearly, the power properties of tests based on univariate polynomials depend upon the unknown alternative $F(\cdot)$ and will be the more favorable the more $F(\cdot)$ 'resembles' univariate polynomials. This trade-off is exactly the same as in the stationary case, as also discussed in Hong and Phillips (2010).

To be concrete denote with $\bar{D}_t := [t^{q+1}, \ldots, t^{q+n}]'$, $\bar{X}_{jt} := [x_{jt}^{p_j+1}, x_{jt}^{p_j+2}, \ldots, x_{jt}^{p_j+r_j}]'$ for $j = 1, \ldots, m$, $Q_{it} := [q_{it}^1, q_{it}^2, \ldots, q_{it}^{s_i}]'$ for $i = 1, \ldots, k$, $F_t := [\bar{D}_t', \bar{X}_{1t}', \ldots, \bar{X}_{mt}', Q_{1t}', \ldots, Q_{kt}']'$ and $F := [F_1', \ldots, F_T']'$. Using this notation the augmented polynomial regression including higher order deterministic trends $\bar{D}_t$, higher order polynomial powers of the regressors x_{jt} and polynomial powers of additional integrated regressors q_{it} can be written as

$$y = Z\theta + F\theta_F + \phi, \tag{11}$$

with $\phi := [\phi_1, \ldots, \phi_T]'$. If equation (11) is well specified the parameters can be estimated consistently by FM-OLS according to Proposition 1 if the additional regressors q_{it} fulfill the necessary assumptions stated in Section 2.1 which are now modified to accommodate the additional regressors.

Assumption 2 *When considering additional regressors q_{it} and their polynomial powers define $\tilde{v}_t := [v_t', (v_t^*)']' = [\Delta x_t', \Delta q_t']'$, with $v_t^* = \Delta q_t$ and $q_t = [q_{1t}, \ldots, q_{kt}]'$. Assumption 1 is extended such that it is fulfilled for the extended process $\tilde{v}_t$ generated by $C_{\tilde{v}}(L)\tilde{\varepsilon}_t$, with $C_{\tilde{v}}(L)$ and $\tilde{\varepsilon}_t$ also extended accordingly.*

Note that equation (11) can be well-specified for different reasons. The first is that (1) is a correctly specified cointegrating relationship, in which case consistently estimated coefficients $\hat{\theta}_F^+$ will converge to their true value equal to 0. The second possibility is that (1) is misspecified, but the extended equation (11) is well-specified. In this case at least some entries of $\hat{\theta}_F^+$ will converge to their non-zero true values. In case that both (11) and (1) are misspecified and ϕ_t is not stationary, spurious regression results similar to the linear case that lead to non-zero limit coefficients apply. Consequently, a specification test based on $H_0 : \theta_F = 0$ is consistent against the three discussed forms of misspecification of (1) discussed in the beginning of the sub-section.

Testing the restriction $\theta_F = 0$ in (11) can be done in several ways. One is given by FM-OLS estimation of the augmented regression (11) and performing a Wald test on the estimated coefficients using Proposition 2. Another possibility is to use the FM-OLS residuals of the original equation (2) and to perform a Lagrange Multiplier RESET type test in an auxiliary regression. Note here that the original RESET test of Ramsey (1969) as well as similar tests by Keenan (1985) and Tsay (1986) use powers of the fitted values, whereas Thursby and Schmidt (1977) use polynomials of the regressors and it is this approach that we also follow since this leads to simpler test statistics. Before turning to the LM specification test we first discuss the Wald specification test.

Proposition 3 *Let y_t be generated by* (1) *with the regressors Z_t, Q_t and errors u_t satisfying Assumptions 1 and 2. Denote with $\hat{\theta}_F^+$ the FM-OLS estimator of θ_F in equation* (11)*, with $\tilde{F} := F - Z(Z'Z)^{-1}Z'F$ and let $\hat{\omega}_{u.\tilde{v}}$ be a consistent estimator of $\omega_{u.\tilde{v}}$. Then it holds that the Wald test statistic for the null hypothesis $H_0 : \theta_F = 0$ in equation* (11)*, given by*

$$T_W := \frac{\hat{\theta}_F^{+\prime}(\tilde{F}'\tilde{F})\hat{\theta}_F^+}{\hat{\omega}_{u.\tilde{v}}}, \tag{12}$$

is under the null hypothesis asymptotically χ^2_b distributed, with $b := n + \sum_{j=1}^m r_j + \sum_{j=1}^k s_j$.

Note that the used variance and covariance estimators in Proposition 3 are all based on the $(m+k)$-dimensional process $\tilde{v}_t$. The result given in Proposition 3 follows as a special case from Propositions 1 and 2 using the specific format of the corresponding restriction matrix R.

The basis of the Lagrange Multiplier (LM) test are the FM-OLS residuals $\hat{u}_t^+$ of (2), which are regressed on $\tilde{F}$ in the auxiliary regression

$$\hat{u}^+ = \tilde{F}\theta_{\tilde{F}} + \psi_t, \tag{13}$$

with $\hat{u}^+ = [\hat{u}_1^+, \ldots, \hat{u}_T^+]'$. To allow for asymptotic standard inference the coefficients $\theta_{\tilde{F}}$ in general have to be estimated with a suitable FM-OLS type estimator to achieve a zero mean Gaussian mixture limiting distribution. This is necessary becasue the limiting distribution of the OLS estimator of $\theta_{\tilde{F}}$ in (13) also depends upon second order bias terms (see the proof of Proposition 4 in the appendix for details). The FM-OLS estimator as well as the test statistic for testing the hypothesis $\theta_{\tilde{F}} = 0$ are presented in the following proposition for the case that (1) is well specified. Consistency of the tests against the above-discussed forms of misspecification of (1) follows from the same arguments as for the Wald test.

Proposition 4 *Let y_t be generated by* (1) *with the regressors X_t, Q_t and errors u_t satisfying Assumptions 1 and 2. Define the fully modified OLS estimator of $\theta_{\tilde{F}}$ in equation* (13) *as*

$$\hat{\theta}^{+}_{\tilde{F}} := \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\hat{u}^{+} - O^{F*} - A^{F*} + k^{F*}A^{*}\right), \tag{14}$$

with

$$O^{F*} := \tilde{F}'\tilde{v}\hat{\Omega}^{-1}_{\tilde{v}\tilde{v}}\hat{\Omega}_{\tilde{v}u} - \tilde{F}'v\hat{\Omega}^{-1}_{vv}\hat{\Omega}_{vu},$$

where $v = [v_1, \ldots, v_T]'$, $\tilde{v} = [\tilde{v}_1, \ldots, \tilde{v}_T]'$ and $A^{F*} := [0'_{n\times 1}, M^{*\prime}_{\tilde{X}_1}, \ldots, M^{*\prime}_{\tilde{X}_m}, M^{*\prime}_{Q_1}, \ldots, M^{*\prime}_{Q_k}]'$, where

$$M_{\tilde{X}_j} = \hat{\Delta}^{+}_{v_j u}\begin{bmatrix} (p_j+1)\sum x_{jt}^{p_j} \\ \vdots \\ (p_j+r_j)\sum x_{jt}^{p_j+r_j-1} \end{bmatrix}, \quad M_{Q_i} = \hat{\Delta}^{+}_{v_i^* u}\begin{bmatrix} T \\ 2\sum q_{it} \\ \vdots \\ s_i \sum q_{it}^{s_i-1} \end{bmatrix},$$

$k^{F*} = F'Z(Z'Z)^{-1}$, $\hat{\Delta}^{+}_{v_j u}$ and A^* as defined above in Proposition 1 and $\hat{\Delta}^{+}_{v_i^* u} := \hat{\Delta}_{v_i^* u} - \hat{\Omega}_{uv}\hat{\Omega}^{-1}_{vv}\hat{\Delta}_{vv_i^*}$. Then it holds that under the null hypothesis that $\theta_{\tilde{F}} = 0$ the FM-OLS estimator defined in (14) *has the limiting distribution*

$$G_F^{-1}\hat{\theta}^{+}_{\tilde{F}} \Rightarrow \left(\int \tilde{J}^F \tilde{J}^{F\prime}\right)^{-1}\int \tilde{J}^F dB_{u.\tilde{v}}, \tag{15}$$

with

$$\tilde{J}^F(r) := J^F - \int J^F J'\left(\int JJ'\right)^{-1} J(r), \tag{16}$$

with $B_{u.\tilde{v}}(r) := B_u(r) - \tilde{B}_v(r)'\Omega^{-1}_{\tilde{v}\tilde{v}}\Omega_{\tilde{v}u}$ and and J^F and G_F defined in the proof in the appendix. Consequently, the LM test statistic for the null hypothesis $H_0 : \theta_{\tilde{F}} = 0$ in (13)

$$T_{LM} := \frac{\hat{\theta}^{+\prime}_{\tilde{F}}(\tilde{F}'\tilde{F})\hat{\theta}^{+}_{\tilde{F}}}{\hat{\omega}_{u.\tilde{v}}}, \tag{17}$$

is under the null hypothesis asymptotically distributed as χ^2_b, with $b = n + \sum_{j=1}^{m} r_j + \sum_{j=1}^{k} s_j$.

Proposition 4 can be seen as a generalization of the modified RESET test considered in Hong and Phillips (2010, Theorem 3), who consider a related test in a bivariate linear cointegrating relationship with only one I(1) regressor and without deterministic variables, i.e. they consider the case $q = 0$, $m = 1$ and $p = 1$. A second difference to our result is that Hong and Phillips use the OLS residuals $\hat{u}_t$ of the linear cointegrating relationship in the auxiliary regression, which leads to different bias correction terms than ours based on the FM-OLS residuals $\hat{u}^{+}_t$. In principle also an extension of the Hong and Phillips (2010) test using the OLS residuals of the original regression is possible.[9] In case that for specification analysis in F_t only higher order polynomial trends are included, we arrive at a test that extends those of Park and Choi (1988) and Park (1990). These authors propose tests for linear cointegration based on adding superfluous higher order deterministic trend terms. This approach is nested within ours.

Clearly, any selection of higher order polynomial terms can be chosen as additional regressors and one need not choose, as done for simplicity in the formulation of the test, a set of consecutive powers ranging from e.g. $p_j + 1$ to $p_j + r_j$. Also, similarly to the discussion at the end of Section 2.1, more general deterministic variables can be included in $\bar{D}_t$. The two above propositions continue to hold with obvious modifications.

2.4 KPSS-Type Test for Cointegration

In this section we discuss a residual based 'direct' test for nonlinear cointegration which prevails in (1) if the error process $\{u_t\}_{t\in\mathbb{Z}}$ is stationary. To test this null hypothesis directly we present a Kwiatkowski, Phillips, Schmidt, and Shin (1992), in short KPSS, type test statistic based on the FM-OLS residuals $\hat{u}_t^+$ of (1). The KPSS test is a variance-ratio test, comparing estimated short- and long-run variances, that converges towards a well defined distribution under stationarity but diverges under the unit root alternative. Note that this as well as other related tests can be interpreted to a certain extent as specification tests as well, since persistent nonstationary errors also prevail if e.g. relevant I(1) regressors are omitted in (1). The test statistic is given by

$$CT := \frac{1}{T\hat{\omega}_{u.v}} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{t} \hat{u}_j^+ \right)^2, \tag{18}$$

with $\hat{\omega}_{u.v}$ a consistent estimator of the long-run variance $\omega_{u.v}$ of $\hat{u}_t^+$. The asymptotic distribution of this test statistic is considered in the following proposition.

Proposition 5 *Let y_t be generated by* (1) *with the regressors Z_t and errors u_t satisfying Assumption 1 and let $\hat{\omega}_{u.v}$ be a consistent estimator of $\omega_{u.v}$, then the asymptotic distribution of the test statistic* (18) *defined above is*

$$CT \Rightarrow \int (W^J)^2,$$

[9] One can also perform the LM test using F rather than $\tilde{F}$. The results are similar in structure to those given in Proposition 4 but of course the precise form of the correction factors is different.

with

$$W^J(r) := W(r) - \int_0^r J^{W\prime} \left(\int J^W J^{W\prime} \right)^{-1} \int J^W dW \tag{19}$$

with $J^W(r) := [D(r)', \mathbf{W}(r)']'$, *where* $\mathbf{W}(r) = [\mathbf{W}_1(r)', \ldots, \mathbf{W}_m(r)']'$, $\mathbf{W}_i(r) = [W_i(r), \ldots, W_i(r)^{p_i}]'$ *for* $i = 1, \ldots, m$ *and with* $W(r), W_1(r), \ldots, W_m(r)$ *independent standard Wiener processes.*

The above limiting distribution (19) only depends upon the specification of the deterministic component and the number and the polynomial degrees of the integrated regressors and therefore critical values can be simulated (and are available upon request). Thus, the test given in Proposition 5 extends the test of Shin (1994) from cointegrating regressions to cointegrating polynomial regressions.

Parallelling Choi and Saikkonen (2010), who consider a similar testing problem in a dynamic OLS estimation framework, we also consider a sub-sample based test statistic whose limiting distribution does not depend upon the specification.

Proposition 6 *Under the same assumptions as in Proposition 5 it holds that*

$$CT_{b,i} := \frac{1}{b\hat{\omega}_{u.v}} \sum_{t=i}^{i+b-1} \left(\sum_{j=i}^{t} \frac{1}{\sqrt{b}} \hat{u}_j^+ \right)^2 \Rightarrow \int W^2,$$

with b *such that for* $T \to \infty$ *it holds that* $b \to \infty$ *and* $b/T \to 0$.

Note that for a given block size b there are $M := \lfloor T/b \rfloor$ sub-samples and corresponding test statistics, $\{CT_{b,i_1}, \ldots, CT_{b,i_M}\}$, that all lead to asymptotically valid statistics for the same null hypothesis. Basing a test on all these statistics might lead to reduced power and increased size (compare again Choi and Saikkonen, 2010). Therefore we consider using this set of statistics in combination with the Bonferroni inequality to modify the critical values using

$$\lim_{T \to \infty} \mathbb{P} \left(CT_{max} \leq c_{\alpha/M} \right) \geq 1 - \alpha,$$

where $CT_{max} := \max(CT_{b,i_1}, \ldots, CT_{b,i_M})$, suppressing the dependence of CT_{max} on b for notational brevity, and $c_{\alpha/M}$ denotes the α/M-percent critical value of the distribution of $\int W^2$. For the computation of the critical values from the distribution function, F_{W^2} say, of $\int W^2$ Choi and Saikkonen (2010) obtain the interesting result that

$$F_{W^2}(z) = \sqrt{2} \sum_{n=0}^{\infty} \frac{\Gamma(n+1/2)}{n!\Gamma(1/2)} (-1)^n \left(1 - f\left(\frac{g_n}{2\sqrt{z}} \right) \right), \quad z \geq 0, \tag{20}$$

with $f(x) = \frac{2}{\sqrt{\pi}} \int_0^x exp(-y^2)dy$ and $g_n = \sqrt{2}/2 + 2n\sqrt{2}$. Using this series representation and truncating the series at $n = 30$ we obtain the critical values for the required distribution. Critical values based on $n = 30$ and for comparison also for $n = 10$ (as used in Choi and Saikkonen, 2010) are available in supplementary material.

Another important practical problem when using the sub-sample based test is the choice of the block-length b. As Choi and Saikkonen (2010) we apply the so called *minimum volatility rule* proposed by Romano and Wolf (2001, p. 1297). To be precise, we choose $b_{min} = 0.5\sqrt{T}$ and $b_{max} = 2.5\sqrt{T}$. For all $b \in [b_{min}, b_{max}]$ we compute the standard deviations of the test statistics over the five neighboring block sizes, i.e. for a block size b^*, we use the test statistics $CT_{b,max}$ for $b = b^* - 2, b^* - 1, b^*, b^* + 1, b^* + 2$ to compute the standard deviation of $CT_{b,max}$ as a function of b. The optimal block-length is then given by the value $b_{opt} \in [b_{min} + 2, b_{max} - 2]$ that leads to the smallest standard deviation. We refer to the test procedure using the Bonferroni bound, i.e. when the null hypothesis is rejected if $CT_{max} \geq c_{\alpha/M}$, and with the block-length chosen as just described as CS test in the simulations.[10]

3 Simulation Performance

In this section we briefly report a small selection of simulation results to investigate the finite sample performance of the proposed estimator and tests. For assessing the performance of the estimator and size of the tests we use data generated according to

$$y_t = c + \delta t + \beta_1 x_t + \beta_2 x_t^2 + u_t, \tag{21}$$

where $\Delta x_t = v_t$ and u_t are generated as

$$\begin{aligned} (1 - \rho_1 L)u_t &= e_{1,t} + \rho_2 e_{2,t} \\ v_t &= e_{2,t} + 0.5 e_{2,t-1}, \end{aligned}$$

with $(e_{1,t}, e_{2,t})' \sim \mathcal{N}(0, I_2)$. The two parameters ρ_1 and ρ_2 control the level of serial correlation in the error term and the level of endogeneity of the regressor, respectively. The parameter values

[10]By construction a test based on the Bonferroni bound is conservative and is known to be particularly conservative when the test statistics used are highly correlated (see Hommel, 1986). The literature has provided several less conservative test procedures based on modified Bonferroni bounds (all of which use all M test statistics rather than only the largest one). In supplementary available material we discuss tests based on the modifications of Hommel (1988), Simes (1986) and Rom (1990). In general, however, using these modifications does not lead to very different behavior of the resulting tests compared to the CS test.

are $c = \delta = 1$, $\beta_1 = 5$ and $\beta_2 = -0.3$. The values for β_1, β_2 are inspired by coefficient estimates obtained when applying the developed FM-OLS estimator to GDP and emissions data (compare Hong and Wagner, 2008).

The FM-OLS estimator and test statistics based upon it are computed for two widely used kernels, the Bartlett and Quadratic Spectral (QS) kernels, and for different bandwidth choices. Two widely used kernels are considered to assess whether the kernel choice has an important impact on the estimator and test performance. The five bandwidth choices are $T^{1/5}$, $T^{1/4}$, $T^{1/3}$, the data dependent rule of Andrews (1991) and the sample size dependent rule of Newey and West (1994), i.e. $\lfloor 4(T/100)^{2/9} \rfloor$. The latter choice is very common and has been suggested by Newey and West (1994) as a simplified, feasible rule especially in conjunction with the Bartlett kernel. Compared to the data dependent rule of Andrews (1991) it is clearly computationally simpler (as it depends only on the sample size), but does not take into account serial correlation in the data (which is captured by Andrews' AR(1) based bandwidth selection rule). We use both rules to see whether the computationally more intensive approach leads to better performance. The three different bandwidths $T^{1/5}$, $T^{1/4}$, $T^{1/3}$ are chosen because Hong and Phillips (2010, Theorems 4 and 5) show that the convergence respectively divergence behavior under the null and alternative of their modified RESET test computed from the residuals of a linear cointegrating relationship depends upon the ratio of the bandwidth to the sample size. In particular they show that in their setup smaller bandwidths lead to slower convergence of their test statistic under the null but to faster divergence (i.e. higher rejection probabilities) under the alternative. In their simulations they, however, find only small effects of the bandwidth choice and we include their choices to see whether similar observations also hold in our more general setup. All results are benchmarked against the OLS estimator, which as discussed is also consistent. Inference is performed in two ways for the OLS estimator, none of which is asymptotically valid in the presence of serial correlation and endogeneity (this being a major reason for developing FM-OLS estimation theory). One way is to perform textbook OLS inference ignoring serial correlation and endogeneity (labeled OLS later) and the other is to use 'HAC-robust' standard errors, as is often done in stationary regression, with the bandwidth chosen according to Newey and West (1994) (labeled HAC later). Rejections for the OLS tests are, as for the FM-OLS based tests, carried out using the standard normal distribution for t-tests and chi-square distribution for Wald tests.

The full set of sample sizes that has been considered in simulations is $T \in \{50, 100, 200, 500, 1000\}$ and the values for ρ_1 and ρ_2 are taken from the set $\{0, 0.3, 0.6, 0.8\}$. Here we only report for brevity representative results for $T \in \{100, 200\}$ and where $\rho_1 = \rho_2$. Full sets of results for all five sample sizes and all combinations of ρ_1 and ρ_2 are available from the authors upon request.

Let us start the discussion of results by briefly considering estimator performance, measured in terms of bias and root mean squared error (RMSE), where for brevity we here only verbally summarize some findings. Bias and RMSE tables are available in supplementary material. In many respects the results for bias and RMSE are similar to results that have been found for the FM-OLS estimator of Phillips and Hansen (1990) in linear cointegrating relationships.[11] In case of absence of both serial correlation and endogeneity, the OLS estimator has, as expected, the best performance in terms of both bias and RMSE. With serial correlation and/or endogeneity increasing, the FM-OLS estimator outperforms the OLS estimator. Bias reductions can amount to about almost 50% in case of $\rho_1, \rho_2 = 0.6$. Also the RMSEs are typically smaller for the FM-OLS estimates, with the performance advantage compared to OLS typically not as big as for bias. These results hold qualitatively very similarly for all coefficients, i.e. for the coefficients to the stochastic as well as deterministic regressors. In this respect it is interesting to note that the different convergence rates for the coefficient to the integrated regressor (rate T) on the one hand and the coefficients to the linear trend and the squared integrated regressor (rate $T^{3/2}$ for both) on the other can be clearly seen from the results already for the smallest sample sizes considered. On average the bias for the former is about 1000 times as large as for the latter two for $T = 100, 200$.

In the simulations performed, the choice of the kernel, i.e. Bartlett or QS, has only minor influence on the performance of the estimator and none of the two kernels leads to a clearly better performance over a wide array of sample sizes and parameters. Similarly, also the choice of the bandwidth has only a moderate impact on the results. Typically, the bias is slightly increasing with increasing bandwidth for β_1, but not for δ and β_2, where the bias is typically decreasing with increasing bandwidth for $T = 200$. RMSEs are not very much affected by bandwidth choice either. The simple bandwidth rule of Newey and West (1994) leads to grosso modo similar performance as the rule of Andrews (1991), with none dominating the other.

We now briefly turn to the coefficient tests, where we present in Table 1 empirical null rejection probabilities for t-tests for $H_0 : \beta_1 = 5$ and $H_0 : \beta_2 = -0.3$ in Panels A and B, respectively, and for the Wald test considering these two coefficients jointly (i.e. for $H_0 : \beta_1 = 5, \beta_2 = -0.3$) the results are presented in Panel C. As discussed in the beginning of the section, for OLS we include both

[11]See e.g. the simulation section in Vogelsang and Wagner (2010) where both OLS and FM-OLS are included in the simulations in a linear cointegration framework.

textbook as well as HAC inference. As for the estimators, the results for the tests are very similar to findings in the linear cointegration case. In case of no serial correlation and/or endogeneity, the OLS textbook statistic (in this case asymptotically valid) has the best performance, with its null rejection probabilities being closest to the 0.05 level. With increasing values of ρ_1, ρ_2 the FM-OLS based test statistics outperform the OLS test statistics, with all tests exhibiting increasing over-rejection problems with increasing values of ρ_1, ρ_2. As expected, the larger are ρ_1, ρ_2 the bigger is the performance advantage of the FM-OLS based test statistics. For $\rho_1, \rho_2 = 0.8$ the differences are very large, which shows that the FM-OLS estimation approach appropriately corrects for the second order bias terms that contaminate the OLS estimators' distribution. Note however that for the Wald test involving both coefficients β_1 and β_2 with $\rho_1, \rho_2 = 0.8$ and $T = 200$, the rejection probabilities are above 45% even for the best performing FM test statistic using the Andrews (1991) bandwidth. It is interesting to note that using (incorrect) HAC robust standard errors in conjunction with the OLS estimator also leads to sizeable reductions in over-rejections compared to textbook OLS inference. HAC inference has over-rejections in the vicinity of the worst performing version of FM based inference. The bandwidth choice for HAC is given by the sample size dependent rule of Newey and West (1994). Thus, the most direct comparison is between the columns labeled HAC and NW. Comparing these two corresponding columns shows that – as one would guess – asymptotically valid inference (NW) outperforms asymptotically invalid inference (HAC), already in small samples.[12] The bandwidth choice has an effect on the empirical null rejection probabilities, and is thus a bit more consequential than it was for the performance of the estimators. Similarly to the findings in Hong and Phillips (2010) discussed above it is seen that larger bandwidths lead to lower null rejection probabilities.

[12] Similar results are obtained when the bandwidth for HAC is chosen with any of the other rules concerning bandwidth choice.

Table 1: Empirical Null Rejection Probabilities, 0.05 Level

Panel A: t-tests for $H_0 : \beta_1 = 5$

$T = 100$

		Bartlett kernel						QS Kernel					
ρ_1, ρ_2	OLS	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0594	.1060	.0754	.0826	.1020	.1058	.0932	.1222	.0842	.0972	.1246	.1234	.1136
0.3	.1542	.1424	.1128	.1092	.1138	.1162	.1092	.1462	.1034	.1038	.1168	.1164	.1118
0.6	.3706	.2678	.2146	.1896	.1662	.1604	.1716	.2498	.1788	.1604	.1474	.1488	.1514
0.8	.5876	.4612	.4270	.3998	.3622	.3108	.3774	.4286	.3940	.3680	.3282	.2984	.3408

$T = 200$

		Bartlett kernel						QS Kernel					
ρ_1, ρ_2	OLS	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0478	.0784	.0588	.0634	.0722	.0738	.0650	.0890	.0658	.0714	.0862	.0784	.0726
0.3	.1472	.1100	.0932	.0874	.0872	.0878	.0866	.1058	.0818	.0796	.0844	.0802	.0790
0.6	.3736	.2306	.1930	.1710	.1458	.1350	.1660	.2010	.1638	.1446	.1278	.1242	.1406
0.8	.6154	.4410	.4228	.3906	.3412	.2930	.3820	.3968	.3880	.3538	.3150	.2846	.3446

Panel B: t-tests for $H_0 : \beta_2 = -0.3$

$T = 100$

		Bartlett kernel						QS Kernel					
ρ_1, ρ_2	OLS	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0570	.1030	.0686	.0736	.0874	.0926	.0822	.1188	.0764	.0848	.1060	.1056	.0964
0.3	.1424	.1360	.1102	.1052	.1070	.1088	.1048	.1352	.1006	.0988	.1056	.1064	.1036
0.6	.2776	.1956	.1772	.1590	.1384	.1340	.1466	.1800	.1498	.1314	.1182	.1216	.1210
0.8	.4202	.2784	.2696	.2378	.1970	.1626	.2116	.2462	.2306	.2022	.1640	.1552	.1770

$T = 200$

		Bartlett kernel						QS Kernel					
ρ_1, ρ_2	OLS	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0528	.0748	.0616	.0674	.0758	.0764	.0684	.0846	.0668	.0712	.0846	.0810	.0728
0.3	.1368	.1000	.0916	.0882	.0870	.0872	.0876	.0962	.0822	.0798	.0808	.0798	.0796
0.6	.2678	.1466	.1558	.1364	.1118	.1042	.1326	.1242	.1296	.1112	.0952	.0938	.1082
0.8	.4368	.2438	.2596	.2196	.1726	.1320	.2128	.2056	.2170	.1844	.1460	.1264	.1790

Panel C: Wald tests for $H_0 : \beta_1 = 5, \beta_2 = -0.3$

$T = 100$

		Bartlett kernel						QS Kernel					
ρ_1, ρ_2	OLS	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0568	.1348	.0832	.0946	.1184	.1238	.1092	.1706	.0972	.1146	.1478	.1466	.1352
0.3	.2002	.1958	.1410	.1382	.1442	.1470	.1402	.2032	.1260	.1294	.1476	.1470	.1372
0.6	.5258	.3878	.3018	.2714	.2364	.2226	.2472	.3550	.2556	.2266	.2038	.2070	.2120
0.8	.8124	.6652	.6356	.5982	.5462	.4800	.5650	.6250	.5894	.5486	.4978	.4654	.5188

$T = 200$

		Bartlett kernel						QS Kernel					
ρ_1, ρ_2	OLS	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	HAC	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0532	.0910	.0652	.0726	.0866	.0880	.0740	.1118	.0746	.0802	.1044	.0960	.0830
0.3	.1924	.1426	.1104	.1040	.1020	.1034	.1038	.1384	.0968	.0932	.1022	.0956	.0922
0.6	.5290	.3140	.2668	.2318	.1990	.1870	.2258	.2720	.2216	.1952	.1718	.1698	.1898
0.8	.8254	.6228	.6228	.5826	.5316	.4602	.5758	.5638	.5782	.5392	.4996	.4514	.5348

Next we consider briefly the power of the coefficient tests. Given that we observe quite substantial null rejection probability differences across tests we focus on size corrected power. This allows to see power differences across tests when holding the null rejection probabilities constant at 0.05. This is useful for theoretical power comparisons, but it has to be kept in mind that such size-corrections are not feasible in practice. In Figure 1 we display the size corrected power curves of the t-test for $\beta_2 = -0.3$ with the values for $\beta_2 \in (-0.3, -0.2]$, displayed on the horizontal axis, generated on a grid with mesh 0.05. In Figure 2 we consider the Wald test.[13] Starting from the null hypothesis $\beta_1 = 5$, $\beta_2 = -0.3$ we consider under the alternative $\beta_1 \in (5, 6]$ and $\beta_2 \in (-0.3, -0.2]$ with in total 21 values generated on a grid with mesh 0.05 for β_1 and 0.005 for β_2. These two figures are for $T = 100$, $\rho_1, \rho_2 = 0.8$ and the Bartlett kernel. Results for other values of ρ_1, ρ_2, other sample sizes and the QS kernel are qualitatively similar. The main message of the figures is twofold, given that all size corrected power curves are very close to each other. First, the test statistics based on the FM-OLS estimator are – and this is the main message since it clearly shows the value of the FM-OLS estimator – strictly preferable to the (even asymptotically invalid) OLS based test statistics. This, since they have (especially for $\rho_1, \rho_2 = 0.8$) much lower size distortions and similar or even slightly higher size corrected power. Second, the choice of the bandwidth becomes relatively unimportant from a size corrected power perspective. This implies that a bandwidth choice that results in low null rejection probabilities should be chosen, since doing so does not result in subsequent (size corrected) power losses. Typically, lowest size corrected power is found for the HAC test statistics and the FM test statistic based on the Andrews (1991) data dependent bandwidth choice. The choice of the bandwidth is more important than the choice of the kernel. In fact size corrected power is virtually equal for both the Bartlett and the QS kernels for any of the bandwidths chosen (see the supplementary material). The simple sample size dependent rule of Newey and West (1994) performs well in terms of resulting size corrected power across a range of parameters and sample sizes.

[13] Supplementary material available upon request displays similar results for β_1.

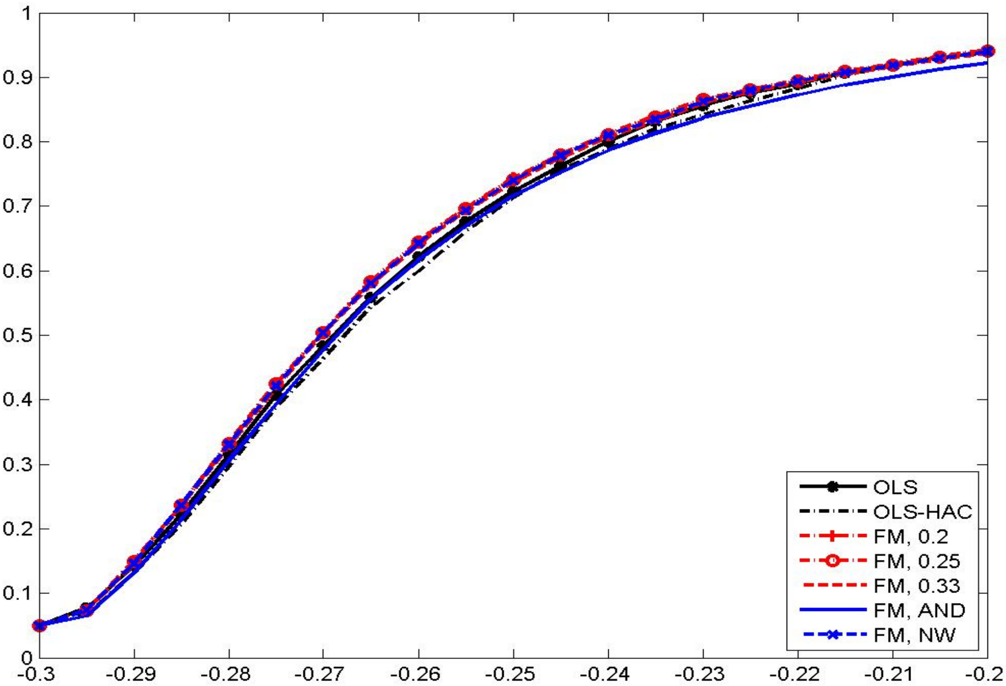

Figure 1: Size Corrected Power, t-test for β_2, $T = 100$, $\rho_1 = \rho_2 = 0.8$, Bartlett Kernel

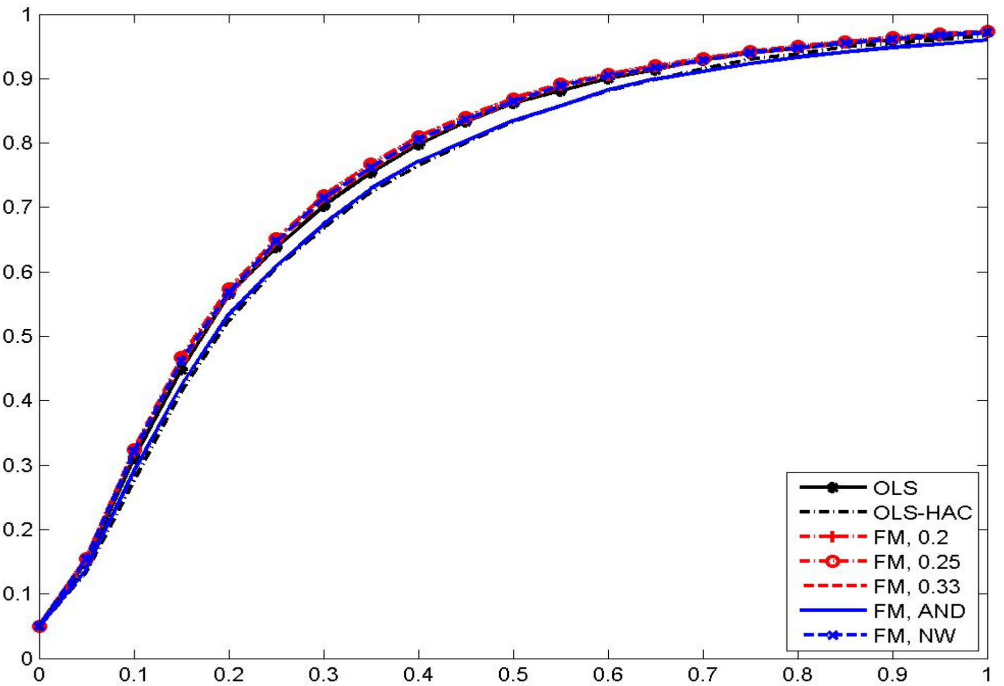

Figure 2: Size Corrected Power, Wald test, $T = 100$, $\rho_1 = \rho_2 = 0.8$, Bartlett Kernel

We close this section by taking a brief look at the performance of the specification (Wald and LM) and cointegration (CT and CS) tests. Data are generated according to three alternative DGPs:

$$(A) \quad : \quad y_t = 1 + t + 5x_t - 0.3x_t^2 + 0.2x_t^3 + u_t$$

$$(B) \quad : \quad y_t = 1 + t + 5x_t - 0.3x_t^2 + e_t, \text{ where } e_t \text{ is an I(1) variable independent of } x_t$$

$$(C) \quad : \quad y_t \text{ and } x_t \text{ are two independent I(1) variables}$$

In case (A) the regressor x_t and error u_t are generated as described above (where as before we only report the results for the cases $\rho_1 = \rho_2$). Also in case (B) x_t is generated as before and $e_t = \sum_{j=1}^{t} \varepsilon_j$, where $\varepsilon_t \sim \mathcal{N}(0,1)$, independent of x_t. Finally, in case (C) both y_t and x_t are independently of each other generated similarly to e_t in case (B). These three DGPs exemplify the main alternatives of interest. Alternative specification (A) covers the case of missing higher order powers of the integrated regressor, alternative (B) corresponds to the case of a missing integrated regressor and alternative (C) corresponds to a spurious regression.

As discussed in the previous section, the performance of the Wald and LM specification tests can be expected to depend upon the unknown alternative DGP as well as the additional regressors included in the augmented respectively auxiliary regression. In this respect we consider four different test specifications. The first set of additional regressors follows the idea of Park and Choi (1988) and Park (1990) to include higher order deterministic trends, where we include $F_t = [t^2,\ t^3]$, labeled I in Table 2 below. The second set of regressors is given by $F_t = [x_t^3,\ x_t^4\ q_t]$, where q_t is an independent I(1) regressor (generated similarly to e_t above), labeled II below. The third choice combines the first two by including both higher order deterministic trends and higher order powers of the integrated regressor, i.e. $F_t = [t^2,\ t^3,\ x_t^3,\ x_t^4]$, labeled III below. The fourth choice is to only include powers of the independent I(1) variable q_t, i.e. $F_t = [q_t,\ q_t^2,\ q_t^3]$, labeled IV below.

In Table 2 we report size corrected power for the specification and cointegration tests. We report results for the Bartlett kernel and the bandwidth chosen according to Newey and West (1994), since we have found before that the coefficient tests' size corrected power is not sensitive with respect to kernel choice and because of the good size corrected power performance resulting from this bandwidth choice.[14] The starting point of all tests is the regression equation (21). Clearly, for the Wald tests the mentioned regressors are added, whereas for the LM tests the FM-OLS residuals

[14]The results are qualitatively very similar with the other bandwidth choices. Since the Bartlett kernel in conjunction with the Newey and West (1994) bandwidth choice is often used in FM estimation, we report the results for this choice of tuning parameters.

Table 2: Size Corrected Power of Specification Tests, 0.05 Level, Bartlett Kernel, Newey-West

	ρ_1, ρ_2	Wald I	Wald II	Wald III	Wald IV	LM I	LM II	LM III	LM IV	CT	CS
					Panel A: T = 100						
(A)	0.0	0.4306	1.0000	1.0000	0.2198	0.3972	1.0000	1.0000	0.1722	0.1388	0.0910
	0.3	0.3806	1.0000	1.0000	0.1954	0.3534	1.0000	1.0000	0.1554	0.0808	0.0442
	0.6	0.2420	1.0000	1.0000	0.1326	0.2666	1.0000	1.0000	0.1256	0.0148	0.0096
	0.8	0.0858	0.9998	0.9996	0.0560	0.1312	1.0000	1.0000	0.0798	0.0000	0.0022
(B)	–	0.7322	0.2850	0.7212	0.3778	0.7250	0.2358	0.6404	0.2936	0.5426	0.3678
(C)	–	0.7236	0.2636	0.7020	0.3728	0.7100	0.2130	0.6088	0.2924	0.5482	0.3764
					Panel B: T = 200						
(A)	0.0	0.5918	1.0000	1.0000	0.3976	0.5628	1.0000	1.0000	0.3764	0.4156	0.3254
	0.3	0.5418	1.0000	1.0000	0.3542	0.5150	1.0000	1.0000	0.3364	0.3066	0.2204
	0.6	0.3948	1.0000	1.0000	0.2644	0.4138	1.0000	1.0000	0.2522	0.1478	0.0772
	0.8	0.1802	1.0000	1.0000	0.1246	0.2226	1.0000	1.0000	0.1540	0.0076	0.0098
(B)	–	0.8488	0.5312	0.8722	0.6196	0.8490	0.5216	0.8592	0.6092	0.8600	0.7918
(C)	–	0.8508	0.5284	0.8714	0.6174	0.8498	0.5114	0.8574	0.6024	0.8618	0.7800

of this equation are the input in the test procedures and the CT and CS tests are also based on the FM-OLS residuals from estimating (21).

It turns out that the null rejection probabilities (available in supplementary material) differ quite substantially between the Wald and LM test versions for any of the chosen regressors F_t. The Wald tests' null rejection probabilities increase strongly with increasing serial correlation and endogeneity in the DGP, whereas the LM tests' null rejection probabilities are much less affected and stay closer to the nominal level. The CS test is the by far most conservative test, which is as expected given the conservativeness of the Bonferroni bound. The effect of sub-sampling and the Bonferroni bound becomes evident by comparing the null rejection probabilities of the CT and CS tests. The CT test, as expected given the performance of KPSS-type tests in standard settings, shows over-rejections that are increasing with increasing serial correlation and endogeneity. The CS test is so conservative that its null rejections, also increasing with increasing serial correlation and endogeneity, are slightly above the nominal level only for the largest considered values of ρ_1, ρ_2 and partly severe under-rejections occur for all other cases. These differences have to be taken into account when discussing size corrected power next.[15] Table 2 shows that test III has highest size corrected power against all considered alternatives. For $T = 100$ the Wald version of this test has higher size corrected power than the LM version, but these differences vanish for larger sample sizes. Given that the over-rejections are much bigger for the Wald than for the LM version leads us to recommend the LM

[15]For completeness we also provide raw power in the supplementary material.

version. It is interesting to compare this with the results for tests I and II. The size corrected power of test I, including higher order deterministic trends only, is strongly deteriorating for alternative (A) with increasing serial correlation and endogeneity. Test II has excellent performance against alternative (A) for all values of ρ_1, ρ_2, which is not a surprise since the additional regressors included when using test II lead to a correctly specified equation (with x_t^4 and q_t being superfluous). On the other hand the inclusion of deterministic trends alone works well against alternatives (B) and (C), which is problematic for test II. Since test III includes both deterministic higher order trends as well as higher order powers of the integrated regressors it combines in a sense the good performance of the first two tests. Test IV, including only powers of an independent random walk cannot be recommended. The above discussion concerning null rejection probabilities already indicates that also the size corrected power of the CT and CS tests is adversely affected by serial correlation and endogeneity (with throughout the CT test outperforming the CS test). Both of these tests and in particular the CT test, however, have size corrected power against alternatives (B) and (C) for $T = 200$ (and larger sample sizes) that is comparable to the power of the best performing versions of the Wald and LM specification tests.

Given these findings it is a good choice to use the LM version of the specification tests rather than the Wald version and to include both deterministic trends as well as higher order powers of the integrated regressor(s) in the auxiliary regression. Using only one or the other type of auxiliary regressors can serve as an indication concerning how to modify the regression model in case of rejection of the null hypothesis of correct specification. In case the sample size is large and the alternative that the researcher has in mind is not one of including higher orders of the regressors already included in the null model, also the CT test may serve as a useful specification respectively cointegration test.

4 Summary and Conclusions

This paper has developed an FM-OLS estimator for cointegrating polynomial regressions (CPRs), by which we refer to regressions with deterministic regressors, integrated regressors, regressors that are powers of integrated regressors and stationary errors. As is common in cointegration analysis the regressors are allowed to be endogenous and the errors are allowed to be serially correlated. The OLS estimator is consistent in this setup, but its limiting distribution is contaminated by second order bias terms in case of regressor endogeneity and serial correlation, which renders inference difficult. Consequently, a fully modified estimator leading to a zero mean Gaussian mixture limiting distribution that allows for standard asymptotic inference is developed. The paper therefore extends the FM-OLS estimator introduced by Phillips and Hansen (1990) from cointegrating regressions to cointegrating polynomial regressions. The theory, as well as the code available from the authors upon request, allows to in addition also include pre-determined stationary regressors. The original motivation to develop estimation and inference theory for this type of relationship stems from the analysis of environmental Kuznets curves (EKCs) that postulate an inverse U-shaped relationship between measures of economic development and measures of pollution. Hong and Wagner (2010) contains a detailed analysis of the EKC for carbon and sulfur dioxide emissions using the methodology developed in this paper.

The zero mean Gaussian mixture limiting distribution of the FM-OLS estimator forms the basis not only for testing hypothesis on the coefficients but also for testing whether the equation is a well-specified CPR using either Wald or LM tests in augmented respectively auxiliary regressions. Asymptotically chi-square distributed Wald and LM specification tests are developed. Additionally also a KPSS-type cointegration test using the FM-OLS residuals is discussed. The limiting distribution depends upon the specification of the equation but is otherwise nuisance parameter free and can thus be simulated. This test is an extension of the test of Shin (1994) from cointegrating to cointegrating polynomial regressions. Additionally, we follow Choi and Saikkonen (2010) and discuss also a sub-sample version of the KPSS-type test that has a limiting distribution independent of the specification. The sub-sample test statistics can be used in conjunction with the Bonferroni bound (or some modified version of it, as discussed in supplementary material).

The theoretical analysis is complemented by a simulation study. The performance advantages of the FM-OLS estimator and tests based on it over the OLS estimator and tests based on it are in many ways similar to the performance advantages found for FM-OLS over OLS in linear cointegrating relationships. In case of no regressor endogeneity and no serial correlation the OLS estimator and tests show, as expected, the best performance. In the presence of endogeneity and/or serial correlation in the errors the FM-OLS estimator and tests outperform the OLS estimator and tests with the performance advantages increasing with increasing endogeneity and serial correlation. With respect to the specification tests it turns out that the LM tests typically outperform the Wald tests and that across the variety of alternatives considered a test using as additional regressors superfluous higher order deterministic trends and higher powers of the integrated regressor performs best. The KPSS-type tests' performance is rather poor in small samples and is, as expected, very negatively affected by serial correlation and endogeneity. In addition the sub-sample version used in conjunction with the Bonferroni bound is very conservative.

Future research will extend the methods developed here to systems of seemingly unrelated cointegrating regressions and also the potential of extending other estimators, in particular the integrated modified OLS estimator of Vogelsang and Wagner (2010), to CPRs will be explored.

Acknowledgements

The helpful comments of Peter Boswijk, Robert Kunst, Benedikt M. Pötscher, Martin Stürmer, Timothy J. Vogelsang, of seminar participants at Helmut Schmidt University Hamburg, University of Linz, University of Innsbruck, Universitat Autònoma Barcelona, University of Tübingen, Tinbergen Institute Amsterdam, Korea Institute of Public Finance, Bank of Korea, Korea Institute for International Economic Policy, Korea Securities Research Institute, Korea Insurance Research Institute, Michigan State University, Institute for Advanced Studies and of conference participants at the 2nd International Workshop in Computational and Financial Econometrics in Neuchâtel, an ETSERN Workshop in Copenhagen, the 3rd Italian Congress of Econometrics and Empirical Economics in Ancona, the Econometric Study Group Meeting of the German Economic Association, the Econometric Society European Meeting in Barcelona, the Meeting of the German Economic Association in Magdeburg and the Statistische Woche in Nuremberg are gratefully acknowledged. Furthermore, we would like to thank in particular the co-editor, the associate editor and two anonymous referees for their extremely valuable comments and suggestions. The usual disclaimer applies.

References

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **59**, 817–858.

Berenguer-Rico, V. and J. Gonzalo (2010). Summability of Stochastic Processes: A Generalization of Integration and Co-Integration Valid for Non-Linear Processes. Mimeo.

Brock, W.A. and M.S. Taylor (2005). Economic Growth and the Environment: A Review of Theory and Empirics. In: Aghion, P. and S. Durlauf (Eds.), *Handbook of Economic Growth.* North-Holland, Amsterdam.

Chang, Y., J.Y. Park, and P.C.B. Phillips (2001). Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors. *Econometrics Journal* **4**, 1–36.

Choi, I. and P. Saikkonen (2010). Tests for Nonlinear Cointegration. *Econometric Theory* **26**, 682–709.

de Benedictis, L.F. and D.E.A. Giles (1998). Diagnostic Testing in Econometrics: Variable Addition, RESET and Fourier Approximations. In Ullah, A. and D.E.A. Giles (Eds.) *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, 383–417.

de Jong, R. (2002). Nonlinear Estimators with Integrated Regressors but without Exogeneity. Mimeo.

Ermini, L. and C.W.J. Granger (1993). Some Generalizations on the Algebra of I(1) Processes. *Journal of Econometrics* **58**, 369–384.

Granger, C.W.J. (1995). Modelling Nonlinear Relationships between Extended-Memory Variables. *Econometrica* **63**, 265–279.

Granger, C.W.J. and J. Hallman (1991). Nonlinear Transformations of Integrated Time Series. *Journal of Time Series Analysis* **12**, 207–224.

Grossmann, G.M. and A.B. Krueger (1995). Economic Growth and the Environment. *Quarterly Journal of Economics* **110**, 353–377.

Hommel, G.A. (1986). Multiple Test Procedures for Arbitrary Dependence Structures. *Metrika* **33**, 321–336.

Hommel, G.A. (1988). A Stagewise Rejective Multiplicative Test Procedure Based on a Modified Bonferroni Test. *Biometrika* **75**, 383–386.

Hong, S.H. and P.C.B. Phillips (2010). Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity. *Journal of Business and Economic Statistics* **28**, 96–114.

Hong, S.H. and M. Wagner (2008). Nonlinear Cointegration Analysis and the Environmental Kuznets Curve. Institute for Advanced Studies, Economics Series No. 224

Hong, S.H. and M. Wagner (2010). A Nonlinear Cointegration Analysis of the Environmental Kuznets Curve. Mimeo.

Ibragimov, R. and P.C.B. Phillips (2008). Regression Asymptotics Using Martingale Convergence Methods. *Econometric Theory* **24**, 888–947.

Jansson, M. (2002). Consistent Covariance Matrix Estimation for Linear Processes. *Econometric Theory* **18**, 1449–1459.

Karlsen, H.A., T. Myklebust, and D. Tjostheim (2007). Nonparametric Estimation in a Nonlinear Cointegration Type Model. *Annals of Statistics* **35**, 252–299.

Keenan, D.M. (1985). A Tukey Nonadditivity Type Test for Time Series Nonlinearity. *Biometrika* **72**, 39–44.

Kuznets, S. (1955). Economic Growth and Income Inequality. *American Economic Review* **45**, 1–28.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How sure are we that Economic Time Series have a Unit Root? *Journal of Econometrics* **54**, 159–178.

Labson, B.S. and P.L. Crompton (1993). Common Trends in Economic Activity and Metals Demand: Cointegration and the Intensity of Use Debate. *Journal of Environmental Economics and Management* **25**, 147–161.

Lee, T.H., H. White, and C.W.J. Granger (1993). Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests. *Journal of Econometrics* **56**, 269–290.

Newey, W. and K. West (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* **61**, 631–654.

Park, J. Y. (1990). Testing for Unit Roots and Cointegration by Variable Addition. In Fomby, T. B. and G. F. Rhodes (Eds.) *Advances in Econometrics, Vol. 8: Co-Integration, Spurious Regression, and Unit Roots.* JAI Press, Greenwich, 107–133.

Park, J.Y. (1992). Canonical Cointegrating Regressions. *Econometrica* **60**, 119–143.

Park, J. Y. and B. Choi (1988). A New Approach to Testing for a Unit Root. CAE Working Paper 88-23, Cornell University.

Park, J.Y. and P.C.B. Phillips (1999). Asymptotics for Nonlinear Transformations of Integrated Time Series. *Econometric Theory* **15**, 269–298.

Park, J.Y. and P.C.B. Phillips (2001). Nonlinear Regressions with Integrated Time Series. *Econometrica* **69**, 117–161.

Phillips, P.C.B. (1983). Best Uniform and Modified Padé Approximants to Probability Densities in Econometrics. In Hildenbrand, W. (Ed.) *Advances in Econometrics*, Cambridge University Press, Cambridge, 123–167.

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

Phillips, P.C.B and V. Solo (1992). Asymptotics for Linear Processes. *The Annals of Statistics* **20**, 971–1001.

Ramsey, J.B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. *Journal of the Royal Statistical Society B* **31**, 350–371.

Rom, D.M. (1990). A Sequentially Rejective Test Procedure Based on a Modified Bonferroni Inequality. *Biometrika* **77**, 663–665.

Romano, J.P. and M. Wolf (2001). Subsampling Intervals in Autoregressive Models with Linear Time Trend. *Econometrica* **69**, 1283–1314.

Shin, Y. (1994). A Residual Based Test for the Null of Cointegration Against the Alternative of No Cointegration, *Econometric Theory* **10**, 91–115.

Sims, C., J.H. Stock and M.W. Watson (1990). Inference in Linear Time Series Models with Some Unit Roots, *Econometrica* **58**, 113–144.

Simes, R.J. (1986). An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika* **73**, 751–754.

Thursby, J. and P. Schmidt (1977). Some Properties of Tests for Specification Error in a Linear Regression Model. *Journal of the American Statistical Association* **72**, 635–641.

Tsay, R.S. (1986). Nonlinearity Tests for Time Series. *Biometrika* **73**, 461–466.

Vogelsang, T.J. and M. Wagner (2010). Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions. Mimeo.

Wang, Q. and P.C.B. Phillips (2009). Structural Nonparametric Cointegrating Regression. *Econometrica* **77**, 1901–1948.

Yandle, B., M. Bjattarai, and M. Vijayaraghavan (2004). Environmental Kuznets Curves: A Review of Findings, Methods, and Policy Implications. Research Study 02.1 update, PERC.

Appendix A: Proofs

Proof of Proposition 1

We start by first establishing the asymptotic behavior of the OLS estimator $\hat{\theta}$,

$$\begin{aligned} G^{-1}(\hat{\theta}-\theta) &= (GZ'ZG)^{-1}GZ'u \\ &\Rightarrow \left(\int JJ'\right)^{-1}\left(\int JdB_u + \begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}\right) \\ &= \left(\int JJ'\right)^{-1}\left(\int JdB_{u.v} + \int JdB_v'\Omega_{vv}^{-1}\Omega_{vu} + \begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}\right). \end{aligned}$$

These results follow by, calculations similar to the ones in Chang, Park, and Phillips (2001), from Assumption 1, with the results for the limit of $GZ'ZG$ being standard in the unit root and cointegration literature. The third follows from the definition of $u_t^+ = u_t - v_t'\Omega_{vv}^{-1}\Omega_{vu}$ and the corresponding limit stochastic process $B_{u.v} = B_u - B_v'\Omega_{vv}^{-1}\Omega_{vu}$.

The stated result for the FM-OLS estimator $\hat{\theta}^+$ follows from considering

$$G^{-1}(\hat{\theta}^+ - \theta) = (GZ'ZG)^{-1}(GZ'u^+ - GA^*),$$

with u^+ denoting the vector of u_t^+ and A^* as given in the main text. By construction it holds for $0 \leq r \leq 1$

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_t^+ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_t - \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu} \\ &\Rightarrow B_u(r) - B_v(r)'\Omega_{vv}^{-1}\Omega_{vu} = B_{u.v}(r). \end{aligned}$$

When considering the asymptotic behavior of $GZ'u^+$ it is convenient to separate the parts corresponding to D and X. For the deterministic components it immediately follows that $G_D D'u^+ \Rightarrow \int DdB_{u.v}$. For a typical cross-product of some power of an integrated regressor and u_t^+ it holds that

$$\begin{aligned} T^{-\frac{k+1}{2}}\sum_{t=1}^{T} x_{jt}^k u_t^+ &= T^{-\frac{k+1}{2}}\sum_{t=1}^{T} x_{jt}^k u_t - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}T^{-\frac{k+1}{2}}\sum_{t=1}^{T} x_{jt}^k v_t \\ &\Rightarrow \int B_{v_j}^k dB_u + k\Delta_{v_j u}\int B_{v_j}^{k-1} - \Omega_{uv}\Omega_{vv}^{-1}\left(\int B_{v_j}^k dB_v + k\Delta_{vv_j}\int B_{v_j}^{k-1}\right) \\ &\Rightarrow \int B_{v_j}^k dB_{u.v} + k\left(\Delta_{v_j u} - \Omega_{uv}\Omega_{vv}^{-1}\Delta_{vv_j}\right)\int B_{v_j}^{k-1}, \end{aligned} \tag{22}$$

where the result concerning $T^{-\frac{k+1}{2}}\sum_{t=1}^{T} x_{jt}^k u_t$ has already been used in Proposition 1 and a result for the terms of the form $T^{-\frac{k+1}{2}}\sum_{t=1}^{T} x_{jt}^k v_t$ can be derived similarly as in the proof of Lemma 4 of

Hong and Phillips (2010). Combining the individual terms this shows that.

$$\begin{aligned} GZ'u^+ &\Rightarrow \int JdB_u + \begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix} - \int JdB_v'\Omega_{vv}^{-1}\Omega_{vu} \\ &= \int JdB_{u.v} + \begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}, \end{aligned}$$

from which the result follows since by construction $GA^* \Rightarrow \begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}$.

Proof of Proposition 2

Under the null hypothesis and the assumption that $\lim_{T\to\infty} G_R R G = R^*$ with G_R invertible and R^* of full rank it holds that

$$\begin{aligned} T_W &= (R\hat{\theta}^+ - r)' \left(\hat{\omega}_{u.v} R(Z'Z)^{-1}R'\right)^{-1} (R\hat{\theta}^+ - r) \\ &= \left(R(\hat{\theta}^+ - \theta)\right)' \left(\hat{\omega}_{u.v} R(Z'Z)^{-1}R'\right)^{-1} \left(R(\hat{\theta}^+ - \theta)\right) \\ &= \left(RGG^{-1}(\hat{\theta}^+ - \theta)\right)' \left(\hat{\omega}_{u.v} RG(GZ'ZG)^{-1}GR'\right)^{-1} \left(RGG^{-1}(\hat{\theta}^+ - \theta)\right) \\ &= \left((G_R RG)G^{-1}(\hat{\theta}^+ - \theta)\right)' \left(\hat{\omega}_{u.v}(G_R RG)(GZ'ZG)^{-1}(GR'G_R')\right)^{-1} \left((G_R RG)G^{-1}(\hat{\theta}^+ - \theta)\right) \\ &\Rightarrow \left[R^* \left(\int JJ'\right)^{-1} \int JdB_{u.v}\right]' \left[\omega_{u.v} R^* \left(\int JJ'\right)^{-1} R^{*\prime}\right]^{-1} \left[R^* \left(\int JJ'\right)^{-1} \int JdB_{u.v}\right], \end{aligned}$$

It remains to show that the above limiting distribution is indeed distributed χ^2_s. Note first that $\int JdB_{u.v} = \omega_{u.v}^{1/2} \int JdW$, with W denoting a standard Brownian motion independent of B_v, is conditional upon B_v (which implies that J is non-random) normally distributed with mean zero and covariance matrix $\omega_{u.v} \int JJ'$. This in turn implies that the conditional distribution of $\left(\int JJ'\right)^{-1} \int JdB_{u.v}$ is given by a normal distribution with mean zero and covariance matrix $\omega_{u.v} \left(\int JJ'\right)^{-1}$, compare also the discussion concerning V_{FM} after Proposition 1 in the main text. Given this, it follows that conditional upon B_v the Wald statistic T_W is asymptotically distributed χ^2_s and since the conditional asymptotic distribution of T_W is independent of B_v it equals the unconditional asymptotic distribution.

Proof of Proposition 3

Clearly the result in this proposition is a special case of a hypothesis covered by Proposition 2 which leads due to the form of the restrictions to a particularly simple form of the test statistic. In the augmented regression (11) the restriction $\theta_F = 0$ corresponds to

$$\begin{bmatrix} 0 & I_b \end{bmatrix} \begin{bmatrix} \theta \\ \theta_F \end{bmatrix} = 0.$$

This immediately implies that $\left(R\begin{bmatrix} Z'Z & Z'F \\ F'Z & F'F \end{bmatrix}^{-1} R'\right)^{-1} = \tilde{F}'\tilde{F}$, with $R = \begin{bmatrix} 0 & I_b \end{bmatrix}$ and $\tilde{F}$ as defined in the main text. Clearly, in this case we can simply take $G_R = G_F^{-1}$ with G_F (defined below in the proof of Proposition 4) denoting the scaling matrix corresponding to F and thus the required condition on R is fulfilled (for all T and not only asymptotically).

Proof of Proposition 4

The proof is in many respects similar to the proof of Proposition 1 in showing that the correction terms given in the proposition (asymptotically) lead to a second order bias free limiting distribution of the proposed FM-OLS estimator. Let us start by defining the corresponding weighting matrix $G_F(T) := \mathrm{diag}(G_{\bar{D}}(T), G_{\bar{X}}(T), G_Q(T))$, with $G_{\bar{D}}(T) := \mathrm{diag}(T^{-(q+\frac{3}{2})}, \ldots, T^{-(q+n+\frac{1}{2})})$,$G_{\bar{X}}(T) := \mathrm{diag}(G_{\bar{X}_1}(T), \ldots, G_{\bar{X}_m}(T))$, $G_Q(T) := \mathrm{diag}(G_{Q_1}(T), \ldots, G_{Q_k}(T))$ with $G_{\bar{X}_j}(T) := \mathrm{diag}(T^{-\frac{p_j+2}{2}}, \ldots, T^{-\frac{p_j+r_j+1}{2}})$ and $G_{Q_i}(T) := \mathrm{diag}(T^{-1}, \ldots, T^{-\frac{s_j+1}{2}})$.

Next define the stacked deterministic and Brownian motion vectors corresponding to the higher order trend terms and higher order polynomial powers of x_{jt} and to the polynomial powers of q_{it}. For t such that $\lim_{T\to\infty} t/T = r$ we have

$$\lim_{T\to\infty} \sqrt{T} G_{\bar{D}(T)} \bar{D}_t = \lim_{T\to\infty} \begin{pmatrix} T^{-(q+1)} & & \\ & \ddots & \\ & & T^{-(q+n)} \end{pmatrix} \begin{pmatrix} t^{q+1} \\ \vdots \\ t^{q+n} \end{pmatrix} = \begin{pmatrix} r^{q+1} \\ \vdots \\ r^{q+n} \end{pmatrix} =: \bar{D}(r)$$

$$\lim_{T\to\infty} \sqrt{T} G_{\bar{X}_j}(T) \bar{X}_{jt} = \lim_{T\to\infty} \begin{pmatrix} T^{-\frac{p_j+1}{2}} & & \\ & \ddots & \\ & & T^{-\frac{p_j+r_j}{2}} \end{pmatrix} \begin{pmatrix} x_{jt}^{p_j} \\ \vdots \\ x_{jt}^{p_j+r_j} \end{pmatrix} = \begin{pmatrix} B_{v_j}^{p_j} \\ \vdots \\ B_{v_j}^{p_j+r_j} \end{pmatrix} =: \mathbf{B}_{v_j}^F(r),$$

$$\lim_{T\to\infty} \sqrt{T} G_{Q_i}(T) Q_{it} = \lim_{T\to\infty} \begin{pmatrix} T^{-\frac{1}{2}} & & \\ & \ddots & \\ & & T^{-\frac{s_i}{2}} \end{pmatrix} \begin{pmatrix} q_{it} \\ \vdots \\ q_{it}^{s_i} \end{pmatrix} = \begin{pmatrix} B_{v_i^*} \\ \vdots \\ B_{v_i^*}^{s_i} \end{pmatrix} =: \mathbf{B}_{v_i^*}^F(r).$$

Stacking all these terms together we define $J^F(r) := [\bar{D}(r)', \mathbf{B}_{v_1}^F(r)', \ldots, \mathbf{B}_{v_m}^F(r)', \mathbf{B}_{v_1^*}^F(r)', \ldots, \mathbf{B}_{v_k^*}^F(r)']'$. The OLS estimator of $\theta_{\tilde{F}}$ of (11) is given by

$$\begin{aligned} \hat{\theta}_{\tilde{F}} &:= (\tilde{F}'\tilde{F})^{-1}\tilde{F}'\hat{u}^+ \\ &= G_F(G_F\tilde{F}'\tilde{F}G_F)^{-1}G_F\tilde{F}'\hat{u}^+. \end{aligned}$$

Using $\tilde{J}$ as defined in (16) it holds that $(G_F\tilde{F}'\tilde{F}G_F)^{-1} \Rightarrow \left(\int \tilde{J}^F\tilde{J}^{F\prime}\right)^{-1}$ and it remains to consider the second term in some detail:

$$\begin{aligned} G_F\tilde{F}'\hat{u}^+ &= G_F\tilde{F}'\left(u^+ - Z(\hat{\theta}^+ - \theta)\right) \\ &= G_F\tilde{F}'u^+ \\ &= G_F F'u^+ - G_F F'ZG(GZ'ZG)^{-1}GZ'u^+, \end{aligned} \tag{23}$$

with the first equality following from $\hat{u}^+ = u^+ - Z(\hat{\theta}^+ - \theta)$, the second from $\tilde{F}'Z = 0$ and the third from the definition of $\tilde{F}$.

The first of the above two terms converges, using similar arguments as in the proof of Proposition 1 to

$$G_F F'u^+ \;\Rightarrow\; \int J^F dB_{u.v} + A^F$$

with $A^F := [0'_{n\times 1}, M^{\bar{X}\prime}, M^{Q\prime}]'$, where $M^{\bar{X}} := [M_1^{\bar{X}\prime}, \ldots, M_m^{\bar{X}\prime}]'$ and $M^Q := [M_1^{Q\prime}, \ldots, M_k^{Q\prime}]'$. The blocks within the latter vectors are given by $M_j^{\bar{X}} := \Delta^+_{v_j u}[(p_j+1)\int B_{v_j}^{p_j}, \ldots, (p_j+r_j)\int B_{v_j}^{p_j+r_j-1}]'$, for $j = 1, \ldots, m$, and $M_j^Q := \Delta^+_{v_j^* u}[1, 2\int B_{v_j^*}, \ldots, s_j\int B_{v_j^*}^{s_j-1}]'$, for $j = 1, \ldots, k$.

For the second term in (23) we obtain $G_F F'ZG(GZ'ZG)^{-1}GZ'u^+ \Rightarrow \int J^F J' \left(\int JJ'\right)^{-1}(JdB_{u.v} + A)$, with A denoting the limit of A^* as used in Proposition 1.

Putting things together we obtain

$$\begin{aligned} G_F^{-1}\hat{\theta}_{\tilde{F}} &\Rightarrow \left(\int \tilde{J}^F\tilde{J}^{F\prime}\right)^{-1}\left(\int J^F dB_{u.v} + A^F - \int J^F J'\left(\int JJ'\right)^{-1}(JdB_{u.v} + A)\right) \\ &= \left(\int \tilde{J}^F\tilde{J}^{F\prime}\right)^{-1}\left(\int \tilde{J}^F dB_{u.v} + A^F - \int J^F J'\left(\int JJ'\right)^{-1} A\right) \\ &= \left(\int \tilde{J}^F\tilde{J}^{F\prime}\right)^{-1}\left(\int \tilde{J}^F dB_{u.v} + A^F + O^F - \int J^F J'\left(\int JJ'\right)^{-1} A\right), \end{aligned} \tag{24}$$

where $O^F := \int \tilde{J}^F dB'_{\tilde{v}}\Omega^{-1}_{\tilde{v}\tilde{v}}\Omega_{\tilde{v}u} - \int \tilde{J}^F dB'_v\Omega^{-1}_{vv}\Omega_{vu}$. Here the second line follows from the first one using the definition of $\tilde{J}^F$ as given in (16) and the third follows from $B_{u.v} = B_{u.\tilde{v}} + B'_{\tilde{v}}\Omega^{-1}_{\tilde{v}\tilde{v}}\Omega_{\tilde{v}u} - B'_v\Omega^{-1}_{vv}\Omega_u$.

The correction factors O^{F*}, A^{F*} and $k^{F*}A^*$ as defined in the formulation of the proposition are such that when scaled by G_F converge to the quantities given above. This implies that the limiting distribution of the FM-OLS estimator of $\theta_{\tilde{F}}$ in the auxiliary regression (13), as defined in (14), is given by $G_F^{-1}\hat{\theta}^+ \Rightarrow \left(\int \tilde{J}^F\tilde{J}^{F\prime}\right)^{-1}\int \tilde{J}^F dB_{u.\tilde{v}}$. Based on this limiting distribution the result for the limiting distribution of the LM test statistic follows in a similar way as shown for the Wald test statistic in Proposition 2.

Proof of Proposition 5

By definition we have $\hat{u}_t^+ = u_t^+ - Z_t'\left(\hat{\theta}^+ - \theta\right)$. From the proof of Proposition 1 we already know that $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_t^+ \Rightarrow B_{u.v}(r)$ and thus we only need to investigate the second term, for which it holds, using the result for the FM-OLS estimator derived in 1 that

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} Z_t' G G^{-1}\left(\hat{\theta}^+ - \theta\right) &= \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} Z_t' G\right) G^{-1}(\hat{\theta}^+ - \theta) \\ &\Rightarrow \int_0^r J'\left(\int JJ'\right)^{-1}\int J dB_{u.v}.\end{aligned}$$

Combining the limits of both terms of $\hat{u}_t^+$ and using the quantities defined in the formulation of the proposition then leads to

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} \hat{u}_t^+ &\Rightarrow B_{u.v}(r) - \int_0^r J'\left(\int JJ'\right)^{-1}\int J dB_{u.v} \\ &= \omega_{u.v}^{1/2}\left(W(r) - \int_0^r J^{W\prime}\left(\int J^W J^{W\prime}\right)^{-1}\int J^W dW\right).\end{aligned}$$

In the above equation the second line follows from the fact that $B_{u.v}(r) = \omega_{u.v}^{1/2}W(r)$ and $\mathbf{B}_v(r) = \mathbf{\Omega_B}^{1/2}\mathbf{W}(r)$ with $\mathbf{\Omega_B} := \mathrm{diag}(\mathbf{\Omega_{B_1}}, \ldots, \mathbf{\Omega_{B_m}})$ where $\mathbf{\Omega_{B_i}} := \mathrm{diag}(\omega_{v_iv_i}, \ldots, \omega_{v_iv_i}^{p_i})$ and $\omega_{v_iv_i}$ the i-th diagonal element of Ω_{vv} for $i = 1, \ldots, m$. This implies (that when a consistent estimator $\hat{\omega}_{u.v}$ is used) that

$$CT \Rightarrow \int (W^J)^2.$$

Proof of Proposition 6

Let $0 \leq r \leq 1$ and $i \leq t = \lfloor br \rfloor + i - 1 \leq i + b - 1$. Similar to the proof of Proposition 5 a functional central limit theorem applies for the sub-sample of residuals and we obtain

$$\frac{1}{\sqrt{b}}\sum_{j=i}^{t} \hat{u}_j^+ = \frac{1}{\sqrt{b}}\sum_{j=i}^{t} u_j^+ + \left(\frac{1}{\sqrt{b}}\sum_{j=i}^{t} Z_j' G(b)\right)\left(G(b)^{-1}G(T)\right)\left(G(T)^{-1}(\hat{\theta}^+ - \theta)\right) \tag{25}$$

Similar to the proof of Proposition 5 one can show, since $b \to \infty$ and $\frac{b}{T} \to 0$, that $\lim_{T\to\infty} \frac{1}{\sqrt{b}}\sum_{j=i}^{t} u_j^+ = B_{u.v}(r)$. The first and the third bracketed terms composing the product on the right hand side above, i.e. $\left(\frac{1}{\sqrt{b}}\sum_{j=i}^{t} Z_j' G(b)\right)$ and $\left(G(T)^{-1}(\hat{\theta}^+ - \theta)\right)$, converge in distribution. The term in the middle is of order $O\left(\sqrt{\frac{b}{T}}\right)$, which implies that the right hand side product term in (25) is $O_p\left(\sqrt{\frac{b}{T}}\right)$. Therefore, since by assumption $\frac{b}{T} \to 0$, we have established that $\frac{1}{\sqrt{b}} \sum_{j=i}^{t} \hat{u}_j^+ \Rightarrow B_{u.v}(r)$. The result then follows from the assumption of consistency of $\hat{\omega}_{u.v}$.

Appendix B: Modified Bonferroni Bound Tests, the Minimum Volatility Rule and Critical Values for CS Test (Supplementary Material)

By construction a test based on the Bonferroni bound is conservative and is known to be particularly conservative when the test statistics used are highly correlated (see Hommel, 1986). In the literature several less conservative modified Bonferroni bound type test procedures have been presented. Some of them are developed in Hommel (1988), Simes (1986) and Rom (1990). Denote the test statistics ordered in magnitude by $CT_b^{(1)} \geq \cdots \geq CT_b^{(M)}$. The modification of Hommel (1988) amounts to rejecting the null hypothesis if at least one of the test statistics $CT_b^{(j)} \geq c_{\alpha^H(j)}$ with $\alpha^H(j) = \frac{j}{C_M}\frac{\alpha}{M}$ and $C_M = 1+1/2+\cdots+1/M$. The modification of Simes (1986) is very similar and almost coincides with the procedure of Hommel with the only difference being that the additional adjustment factor C_M is not included, i.e. $\alpha^S(j) = j\frac{\alpha}{M}$. A further modification of the computation of the levels used in the sequential test procedure has been proposed in Rom (1990). For this modification the levels $\alpha^R(j)$ are computed recursively via $\alpha^R(M) = \alpha$, $\alpha^R(M-1) = \frac{\alpha}{2}$ and for $k = 3, \ldots, M$ they are computed as

$$\alpha^R(M-k+1) = \frac{1}{k}\left[\sum_{j=1}^{k-1}\alpha^j - \sum_{j=1}^{k-1}\binom{k}{j}(\alpha^R(M-j))^{k-j}\right].$$

The null hypothesis is rejected if all test statistics $CT_b^{(j)} \geq c_{\alpha^R(j)}$.

For these modified tests that involve all M test statistics we base the block-length selection on the following procedure. For each block-length $b_i \in [b_{min}, b_{max}]$ we compute the mean and standard deviation of the empirical distribution of the test statistics $\{CT_{b_i,i_1}, \ldots, CT_{b_i,i_M}\}$, which we denote by m_{b_i} and sd_{b_i}. The idea of the minimum volatility principle is now implemented by minimizing (again over five neighboring values of b) the change of the empirical distribution in terms of the first two moments. Hence we choose the block-length to minimize $vm_{b_i} = std(m_{b_i-2}, m_{b_i-1}, m_{b_i}, m_{b_i+1}, m_{b_i+2}) + std(sd_{b_i-2}, sd_{b_i-1}, sd_{b_i}, sd_{b_i+1}, sd_{b_i+2})$, with $std(\cdot)$ denoting the standard deviation.

`MATLAB` code that implements the described test procedures is available from the authors upon request.

Table B1: Critical values $c_{\frac{\alpha}{M}}$ from
$\mathbb{P}\left[\int W^2 \geq c_{\frac{\alpha}{M}}\right] = \frac{\alpha}{M}$ for $\alpha = 5\%$ and 10%

M	5%	10%	M	5%	10%	M	5%	10%
Sum in (20) truncated at 30								
2	2.135	1.656	15	3.588	3.076	28	4.034	3.538
3	2.421	1.934	16	3.635	3.121	29	4.058	3.563
4	2.627	2.135	17	3.680	3.164	30	4.081	3.588
5	2.787	2.292	18	3.721	3.203	31	4.103	3.612
6	2.917	2.421	19	3.760	3.241	32	4.124	3.635
7	3.027	2.531	20	3.797	3.276	33	4.145	3.658
8	3.121	2.627	21	3.832	3.309	34	4.165	3.680
9	3.203	2.711	22	3.865	3.340	35	4.184	3.700
10	3.276	2.787	23	3.897	3.370	36	4.202	3.721
11	3.340	2.855	24	3.927	3.398	37	4.220	3.741
12	3.398	2.917	25	3.955	3.424	38	4.237	3.760
13	3.484	2.974	26	3.983	3.484	39	4.253	3.779
14	3.538	3.027	27	4.009	3.511	40	4.269	3.797
Sum in (20) truncated at 10								
2	2.135	1.656	15	3.582	3.081	28	3.997	3.533
3	2.421	1.934	16	3.627	3.128	29	4.018	3.558
4	2.626	2.135	17	3.669	3.172	30	4.038	3.582
5	2.785	2.292	18	3.709	3.214	31	4.058	3.605
6	2.912	2.421	19	3.746	3.253	32	4.076	3.627
7	3.031	2.531	20	3.781	3.291	33	4.094	3.649
8	3.128	2.626	21	3.813	3.326	34	4.111	3.669
9	3.214	2.710	22	3.844	3.360	35	4.127	3.689
10	3.291	2.785	23	3.873	3.392	36	4.143	3.709
11	3.360	2.852	24	3.900	3.422	37	4.158	3.728
12	3.422	2.912	25	3.926	3.452	38	4.172	3.746
13	3.480	2.977	26	3.951	3.480	39	4.186	3.763
14	3.533	3.031	27	3.974	3.507	40	4.199	3.781

Appendix C: Additional Simulation Results (Supplementary Material)

Table C1: Bias for coefficients β_1 and β_2

Panel A: Bias for coefficient β_1

$T = 100$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	-.0013	-.0019	-.0019	-.0018	-.0019	-.0018	-.0019	-.0018	-.0018	-.0019	-.0018
0.3	.0167	-.0034	-.0039	-.0049	-.0050	-.0044	-.0042	-.0047	-.0060	-.0057	-.0054
0.6	.0743	.0399	.0409	.0419	.0404	.0418	.0410	.0425	.0433	.0411	.0433
0.8	.1952	.1567	.1597	.1655	.1657	.1633	.1595	.1639	.1710	.1655	.1685

$T = 200$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	-.0003	-.0005	-.0005	-.0006	-.0005	-.0005	-.0005	-.0005	-.0006	-.0006	-.0005
0.3	.0087	-.0012	-.0014	-.0018	-.0019	-.0014	-.0015	-.0016	-.0022	-.0020	-.0017
0.6	.0396	.0228	.0239	.0252	.0250	.0241	.0240	.0253	.0265	.0261	.0256
0.8	.1117	.0933	.0967	.1027	.1070	.0974	.0963	.1006	.1078	.1100	.1017

Panel B: Bias ($\times 1000$) for coefficient β_2

$T = 100$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0841	.0895	.0877	.0841	.0852	.0860	.0878	.0842	.0810	.0825	.0832
0.3	.0868	.1162	.1132	.1057	.1067	.1096	.1140	.1086	.0996	.1049	.1046
0.6	.0970	.1611	.1604	.1545	.1425	.1583	.1616	.1600	.1505	.1311	.1571
0.8	.1576	.2340	.2371	.2399	.2197	.2400	.2369	.2421	.2433	.2031	.2459

$T = 200$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	-.0021	-.0017	-.0018	-.0007	-.0005	-.0018	-.0023	-.0023	.0005	-.0015	-.0023
0.3	.0042	-.0000	-.0003	.0008	.0014	-.0003	-.0009	-.0013	.0023	-.0002	-.0015
0.6	.0354	.0245	.0229	.0220	.0271	.0226	.0224	.0202	.0219	.0255	.0197
0.8	.1356	.1213	.1173	.1112	.1150	.1165	.1176	.1122	.1066	.1164	.1108

Table C2: RMSE for coefficients β_1 and β_2

Panel A: RMSE for coefficient β_1

$T = 100$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0670	.0717	.0721	.0728	.0727	.0725	.0723	.0729	.0738	.0736	.0735
0.3	.0938	.0962	.0967	.0977	.0977	.0973	.0967	.0975	.0991	.0987	.0985
0.6	.1725	.1570	.1572	.1580	.1589	.1577	.1572	.1579	.1593	.1606	.1588
0.8	.3285	.3008	.3016	.3038	.3063	.3029	.3015	.3032	.3067	.3093	.3054

$T = 200$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0325	.0336	.0337	.0339	.0339	.0337	.0337	.0339	.0341	.0341	.0339
0.3	.0470	.0464	.0465	.0468	.0468	.0465	.0465	.0467	.0471	.0469	.0468
0.6	.0915	.0812	.0816	.0821	.0823	.0817	.0816	.0822	.0829	.0830	.0823
0.8	.1919	.1752	.1770	.1806	.1846	.1775	.1769	.1794	.1842	.1879	.1801

Panel B: RMSE for coefficient β_2

$T = 100$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0056	.0058	.0058	.0058	.0058	.0058	.0058	.0058	.0059	.0059	.0059
0.3	.0075	.0077	.0077	.0077	.0077	.0077	.0077	.0077	.0078	.0078	.0078
0.6	.0119	.0117	.0117	.0118	.0119	.0118	.0117	.0118	.0119	.0120	.0119
0.8	.0192	.0188	.0188	.0189	.0193	.0189	.0188	.0189	.0190	.0197	.0190

$T = 200$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	.0019	.0019	.0019	.0019	.0019	.0019	.0019	.0019	.0019	.0019	.0019
0.3	.0027	.0027	.0027	.0027	.0027	.0027	.0027	.0027	.0027	.0027	.0027
0.6	.0045	.0043	.0043	.0043	.0043	.0043	.0043	.0043	.0043	.0044	.0043
0.8	.0080	.0077	.0077	.0078	.0079	.0077	.0077	.0078	.0078	.0080	.0078

Table C3: Bias and RMSE for coefficient δ

Panel A: Bias ($\times 1000$) for coefficient δ

$T = 100$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	-0.0220	-0.0305	-0.0337	-0.0422	-0.0443	-0.0373	-0.0303	-0.0390	-0.0365	-0.0398	-0.0411
0.3	-0.1559	-0.0782	-0.0818	-0.0928	-0.0942	-0.0872	-0.0763	-0.0878	-0.0851	-0.0878	-0.0913
0.6	-0.5040	-0.3545	-0.3636	-0.3858	-0.3984	-0.3766	-0.3579	-0.3797	-0.3859	-0.4121	-0.3905
0.8	-1.1533	-0.9913	-1.0030	-1.0309	-1.0289	-1.0198	-0.9971	-1.0246	-1.0354	-0.9970	-1.0396

$T = 200$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	-0.0344	-0.0233	-0.0220	-0.0204	-0.0187	-0.0217	-0.0214	-0.0208	-0.0186	-0.0176	-0.0207
0.3	-0.0635	-0.0417	-0.0393	-0.0359	-0.0335	-0.0389	-0.0389	-0.0373	-0.0327	-0.0318	-0.0369
0.6	-0.1542	-0.1266	-0.1249	-0.1205	-0.1130	-0.1246	-0.1255	-0.1246	-0.1169	-0.1153	-0.1238
0.8	-0.3683	-0.3363	-0.3352	-0.3330	-0.2998	-0.3350	-0.3354	-0.3355	-0.3306	-0.2866	-0.3349

Panel B: RMSE for coefficient δ

$T = 100$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	0.0065	0.0067	0.0068	0.0069	0.0069	0.0069	0.0068	0.0069	0.0071	0.0071	0.0070
0.3	0.0097	0.0092	0.0093	0.0094	0.0095	0.0094	0.0093	0.0094	0.0097	0.0097	0.0096
0.6	0.0206	0.0167	0.0167	0.0167	0.0169	0.0167	0.0167	0.0167	0.0169	0.0171	0.0168
0.8	0.0438	0.0380	0.0381	0.0385	0.0389	0.0384	0.0381	0.0384	0.0390	0.0394	0.0387

$T = 200$

		Bartlett kernel					QS Kernel				
ρ_1, ρ_2	OLS	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW	$T^{1/5}$	$T^{1/4}$	$T^{1/3}$	AND	NW
0.0	0.0022	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023
0.3	0.0034	0.0032	0.0032	0.0032	0.0032	0.0032	0.0032	0.0032	0.0033	0.0032	0.0032
0.6	0.0077	0.0062	0.0062	0.0063	0.0063	0.0062	0.0062	0.0063	0.0064	0.0064	0.0063
0.8	0.0179	0.0157	0.0159	0.0164	0.0168	0.0160	0.0159	0.0162	0.0169	0.0172	0.0163

Table C4: Null Rejection Probabilities of Specification Tests, 0.05 Level, Bartlett Kernel, Newey-West

		Wald				LM				CT	CS
	ρ_1, ρ_2	I	II	III	IV	I	II	III	IV		
						Panel A: T = 100					
(A)	0.0	0.1286	0.1946	0.1954	0.1904	0.0424	0.1158	0.0588	0.1158	0.0540	0.0006
	0.3	0.1734	0.2178	0.2316	0.2142	0.0672	0.1296	0.0638	0.1296	0.0846	0.0034
	0.6	0.2984	0.2582	0.3452	0.2750	0.1420	0.1470	0.0974	0.1470	0.2054	0.0164
	0.8	0.5484	0.3642	0.5936	0.4064	0.3376	0.2028	0.2520	0.2028	0.5120	0.0676
						Panel B: T = 200					
(A)	0.0	0.0858	0.1154	0.1166	0.1100	0.0478	0.0784	0.0548	0.0784	0.0532	0.0016
	0.3	0.1188	0.1476	0.1598	0.1424	0.0754	0.0994	0.0798	0.0994	0.0862	0.0074
	0.6	0.2572	0.2226	0.3082	0.2324	0.1534	0.1504	0.1498	0.1504	0.2310	0.0372
	0.8	0.5556	0.3862	0.6094	0.4170	0.3994	0.2650	0.3896	0.2650	0.6370	0.1814

Table C5: Raw Power of Specification Tests, 0.05 Level, Bartlett Kernel, Newey-West

		Wald				LM				CT	CS
	ρ_1, ρ_2	I	II	III	IV	I	II	III	IV		
		Panel A: T = 100									
(A)	0.0	0.5634	1.0000	1.0000	0.3950	0.3838	1.0000	1.0000	0.2866	0.4612	0.0328
	0.3	0.5624	1.0000	1.0000	0.3944	0.3836	1.0000	1.0000	0.2866	0.4618	0.0324
	0.6	0.5666	1.0000	1.0000	0.3962	0.3836	1.0000	1.0000	0.2854	0.4648	0.0332
	0.8	0.5736	1.0000	1.0000	0.3952	0.3912	1.0000	1.0000	0.2882	0.4750	0.0342
(B)	–	0.8128	0.4844	0.8446	0.5498	0.7122	0.3536	0.6580	0.4412	0.8346	0.1958
(C)	–	0.8024	0.4622	0.8232	0.5524	0.7012	0.3306	0.6314	0.4374	0.8348	0.2000
		Panel B: T = 200									
(A)	0.0	0.6460	1.0000	1.0000	0.4958	0.5564	1.0000	1.0000	0.4314	0.7384	0.2366
	0.3	0.6460	1.0000	1.0000	0.4958	0.5570	1.0000	1.0000	0.4306	0.7378	0.2358
	0.6	0.6468	1.0000	1.0000	0.4958	0.5578	1.0000	1.0000	0.4304	0.7388	0.2356
	0.8	0.6492	1.0000	1.0000	0.4982	0.5596	1.0000	1.0000	0.4314	0.7430	0.2350
(B)	–	0.8728	0.6398	0.9076	0.6998	0.8456	0.5782	0.8640	0.6542	0.9690	0.7008
(C)	–	0.8766	0.6342	0.9086	0.7052	0.8472	0.5610	0.8608	0.6526	0.9750	0.6860

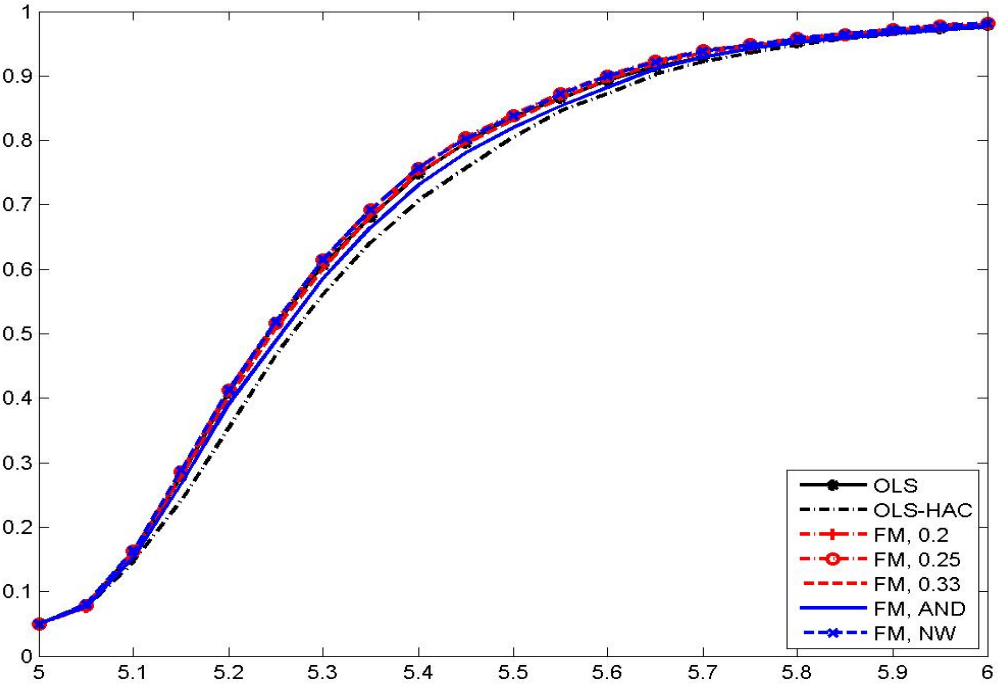

Figure C1: Size Corrected Power, t-test for β_1, $T = 100$, $\rho_1 = \rho_2 = 0.6$, Bartlett Kernel

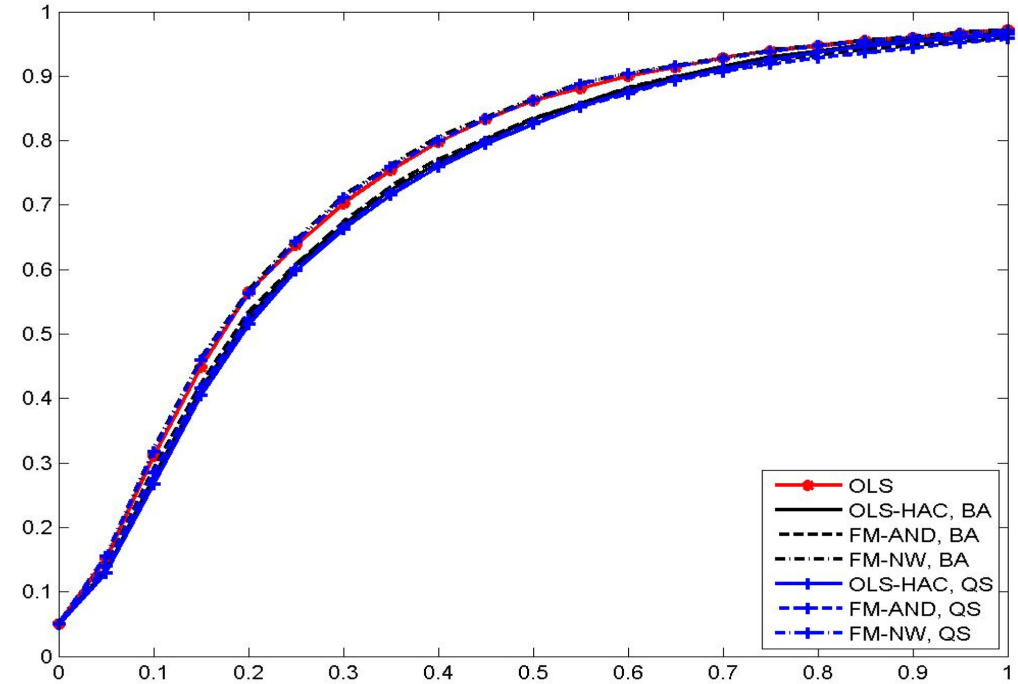

Figure C2: Size Corrected Power, Wald test, $T = 100$, $\rho_1 = \rho_2 = 0.8$, Comparison of Bartlett and Quadratic Spectral Kernels

Authors: Seung Hyun Hong, Martin Wagner

Title: Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference

Reihe Ökonomie / Economics Series 264

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996