Econometric Theory, **32**, 2016, 1289–1315.
doi:10.1017/S0266466615000213

COINTEGRATING POLYNOMIAL REGRESSIONS: FULLY MODIFIED OLS ESTIMATION AND INFERENCE

MARTIN WAGNER
Technical University Dortmund
and
Institute for Advanced Studies
and
Bank of Slovenia

SEUNG HYUN HONG
Korea Institute of Public Finance

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

1. INTRODUCTION

This paper extends the fully modified OLS (FM-OLS) estimator to *cointegrating polynomial regressions* (CPRs), i.e., to regressions that include as explanatory variables deterministic variables, integrated processes, and integer powers of integrated processes. As is common in the linear cointegration literature, the

The authors thank Peter Grabarczyk and Oliver Stypka for excellent research assistance. We also gratefully acknowledge the helpful comments of the editor, the coeditor, two anonymous referees, Peter Boswijk, Robert M. de Jong, Robert Kunst, Anesu N. Mandisodza, Benedikt M. Pötscher, Ulrike Schneider, Martin Stürmer and Timothy J. Vogelsang, as well as the comments of numerous seminar and conference participants. Furthermore the authors gratefully acknowledge financial support by the Oesterreichische Nationalbank (Anniversary fund, project number: 13398). The first author also acknowledges the financial support of the Deutsche Forschungsgemeinschaft via the Collaborative Research Center 823: *Statistical Modelling of Nonlinear Dynamic Processes*. The usual disclaimer applies. Address correspondence to Martin Wagner, Faculty of Statistics, Technical University Dortmund, Vogelpothsweg 87, D-44227 Dortmund, Germany, e-mail: mwagner@statistik.tu-dortmund.de

© Cambridge University Press 2015

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

stationary errors are allowed to be serially correlated and the regressors are allowed to be endogenous. The FM-OLS estimator is shown to have a zero-mean Gaussian mixture limiting distribution that forms the basis for asymptotic standard inference. Wald- and LM-type specification tests are discussed. With respect to specification testing, amongst other things, this paper extends Hong and Phillips (2010) – who consider LM-type specification testing based on the OLS residuals of a linear cointegrating relationship with one integrated regressor – in several ways. The paper also presents a Kwiatkowski *et al.* (1992)-type (or KPSS-type for short) cointegration test. It turns out that the limiting distribution of this test is only nuisance parameter free in the special case where a single integrated regressor occurs with higher-order powers. To overcome this limitation a subsampling version of the test, inspired by Choi and Saikkonen (2010), is considered and shown to have a nuisance parameter free limiting distribution in the general case too. The subsampling test can be used in conjunction with Bonferroni-type bounds. In terms of asymptotic theory our work relies heavily upon important contributions by Chang *et al.* (2001), Park and Phillips (1999, 2001), and Ibragimov and Phillips (2008), with a main difference being that the regressors are allowed to be endogenous rather than only predetermined.

Our original motivation for considering CPRs is the environmental Kuznets curve (EKC) hypothesis, which postulates an inverted U-shaped relationship between measures of economic development (typically proxied by GDP per capita) 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 Grossman and Krueger (1995), more than one hundred refereed publications (a number counted a decade ago by Yandle *et al.*, 2004) have performed econometric analyses of EKCs.[1] Many of these empirical studies use unit root and cointegration techniques and include powers of (the logarithm of) GDP per capita as regressors (in order to allow for U- or inverted U-shaped relationships). This literature neglects that powers of integrated processes, being nonlinear functions of integrated processes, are not themselves integrated processes. This invalidates the use of standard unit root and cointegration techniques (compare, e.g., Wagner, 2012). Such relationships – called cointegrating polynomial relationships in our terminology – necessitate the development of appropriate estimation and inference tools, as is done in this paper.

As an illustrating example consider Figure 1, which displays estimated EKC-type relationships for UK sulfur dioxide (SO_2) emissions over the period 1870–2005. The methods developed in this paper allow us to: (a) test for the presence as well as correct specification of a CPR relationship, and (b) in the case of a CPR relationship, to estimate its parameters with a limiting distribution suitable for asymptotic standard inference. The figure displays the FM-OLS estimates of the EKC-type relationship for the linear, quadratic, and cubic specifications. Only for the cubic specification can the null of a CPR relationship not be rejected

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

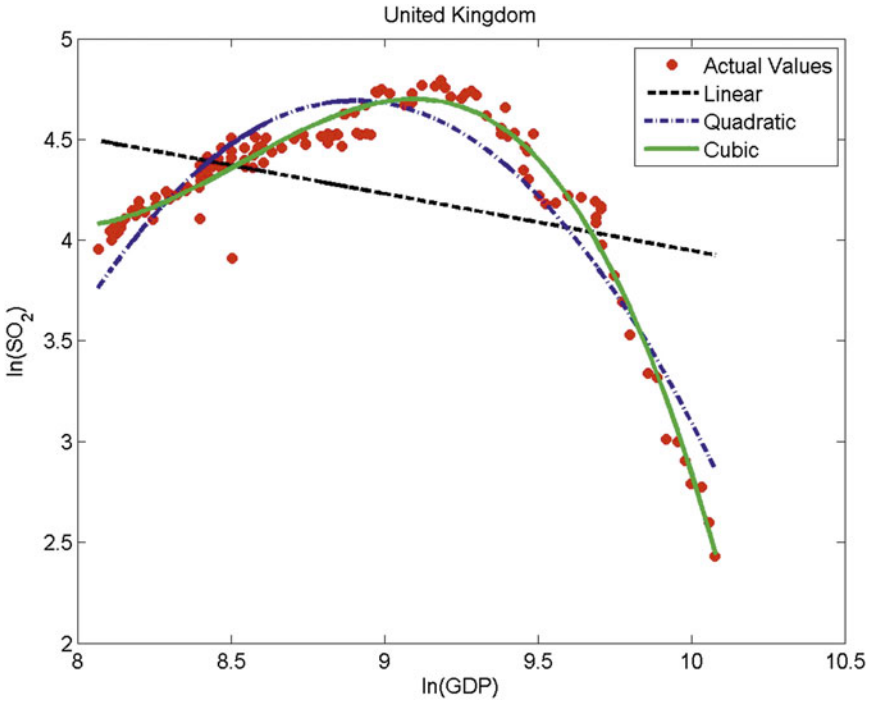


FIGURE 1. Scatter plot and estimated EKC-type relationships for sulfur dioxide emissions for the UK over the period 1870–2005. The three lines display the results of inserting 136 equidistant points from the sample range of $\ln(GDP)$ in the estimated linear, quadratic and cubic relationships.

using the cointegration test presented in Section 2.4. Furthermore, for the cubic formulation, the specification tests discussed in Section 2.3 do not indicate misspecification. A detailed analysis of the EKC hypothesis for both carbon dioxide and sulfur dioxide emissions using the tools developed in this paper is provided in Wagner (2014a). That paper also includes a detailed discussion of the problems arising when treating, as is commonly done, EKC-type relationships as standard, linear cointegrating relationships.

The theoretical analysis of the paper is complemented by a small simulation study to assess the performance of the estimator and tests, benchmarked 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 hypothesis tests (size distortions, size-corrected power), are similar to standard cointegrating relationships. The larger the extent of serial correlation and/or regressor endogeneity, the larger the performance advantages of the FM-OLS estimator – and test statistics based upon it – over OLS. With respect to the specification tests, it turns out that the LM tests outperform the Wald

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

tests since the LM tests have essentially the same size-corrected power as the Wald tests but much smaller over-rejections under the null hypothesis. The specification tests have highest power against the variety of alternatives considered, when the auxiliary regressors comprise both higher-order deterministic trends (originally considered in a unit root and cointegration test context by Park and Choi, 1988; Park, 1990) as well as higher-order powers of the integrated regressors. The simulation performance of the KPSS-type cointegration tests depends strongly upon the data generating process (DGP) under consideration. It is well known that the performance of KPSS-type tests is detrimentally affected by the presence of serial correlation. This is also confirmed by our simulations. The subsample test also suffers from the conservativeness of the Bonferroni-type bounds 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. The journal's website contains online supplementary material that includes critical values for the generally applicable subsampling KPSS-type test, details on the subsampling procedure and the modified Bonferroni bound tests, and more detailed 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 by $B(r)$. By $\lfloor x \rfloor$ we denote the integer part of $x \in \mathbb{R}$ and diag$(\cdot)$ denotes a diagonal matrix with the entries specified throughout. For a vector $x \in \mathbb{R}^n$ 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 is 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 *cointegrating polynomial regression* (CPR) model including the constant and polynomial time trends up to order q, integer powers of I(1) regressors x_{jt}, $j = 1, \ldots, m$ up to degrees p_j, and stationary error term u_t:

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

with $D_t := \left[1, t, t^2, \ldots, t^q\right]'$, $x_t := \left[x_{1t}, \ldots, x_{mt}\right]'$, $X_{jt} := \left[x_{jt}, x_{jt}^2, \ldots, x_{jt}^{p_j}\right]'$ and the parameter vectors $\theta_D \in \mathbb{R}^{q+1}$, and $\theta_{X_j} \in \mathbb{R}^{p_j}$. Furthermore, define $X_t := \left[X_{1t}', \ldots, X_{mt}'\right]'$, $Z_t := \left[D_t', X_t'\right]'$, and $p := \sum_{j=1}^m p_j$. Using these definitions

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

we rewrite (1) in more compact form as:

$$y = D\theta_D + X\theta_X + u \quad (2)$$
$$= Z\theta + u,$$

with $y := [y_1, \ldots, y_T]'$, $u := [u_1, \ldots, u_T]'$, $Z := [D \ X]$, $\theta = \left[\theta_D' \ \theta_X'\right]' \in \mathbb{R}^{(q+1)+p}$, $D = [D_1, \ldots, D_T]' \in \mathbb{R}^{T\times(q+1)}$, and $X = [X_1, \ldots, X_T]' \in \mathbb{R}^{T\times p}$.

The precise assumptions concerning the error process and the regressors are as follows:

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

$$u_t = C_u(L)\zeta_t = \sum_{j=0}^{\infty} c_{uj}\zeta_{t-j}$$

$$\Delta x_t = v_t = C_v(L)\varepsilon_t = \sum_{j=0}^{\infty} c_{vj}\varepsilon_{t-j},$$

with the conditions:

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

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

$$\Sigma^0 = \begin{bmatrix} \Sigma_{\zeta\zeta} & \Sigma_{\zeta\varepsilon} \\ \Sigma_{\varepsilon\zeta} & \Sigma_{\varepsilon\varepsilon} \end{bmatrix} := \mathbb{E}\left(\xi_t^0\xi_t^{0\prime}|\mathcal{F}_{t-1}\right) > 0.$$

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

Assumptions similar to Assumption 1 have been used in a variety of papers dealing with related setups, see, e.g., Chang *et al.* (2001), Park and Phillips (1999, 2001), or Hong and Phillips (2010). The crucial aspect is the asymptotic behavior of cross-product terms, such as $T^{-\frac{k+1}{2}}\sum_{t=1}^{T} x_{jt}^k u_t$. The limit of these terms is needed to derive the limiting distribution of the OLS estimator, which in turn is the basis for developing a FM-OLS-type estimator. The main difference from the assumptions of Chang *et al.* (2001) is, using the notation of this paper, they assume that $u_t = \zeta_t$. Another difference is that they assume that all regressors are predetermined, i.e., in our notation they assume that $\left\{[\zeta_t, \varepsilon_{t+1}']'\right\}_{t\in\mathbb{Z}}$ is a stationary and ergodic martingale difference sequence. The assumption $\det(C_v(1)) \neq 0$ excludes cointegration amongst the components of x_t. This is a typical assumption in the cointegrating regression literature that we also use in our context.

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

For the sake of completeness, note that 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. They use a linear process framework with moment conditions that are in our context related to the order of the polynomial considered. Yet a different route was taken by de Jong (2002, Assumptions 1 and 2) who uses near epoch dependent sequences with appropriate moment conditions. A recent contribution that establishes many useful limit results, and that illustrates their applicability to FM-OLS estimation for a quadratic cointegrating polynomial regression, is Liang *et al.* (2014). Clearly, any framework leading to the desired convergence results is suitable for us.

Remark 1. The key limitation of the above model (1) is, as is common in the nonlinear cointegration literature, additive separability, i.e., cross-product terms (of powers) of the integrated regressors are excluded. The reason for this is that the limiting behavior of terms, such as in the general polynomial case $T^{-\frac{k_1+\cdots+k_m+1}{2}} \sum_{t=1}^{T} x_{1t}^{k_1} x_{2t}^{k_2} \cdots x_{mt}^{k_m} u_t$, needs to be established first, but that is beyond the scope of the present paper. Wagner (2014b) considers cointegrated Translog (cost or production) functions (see, e.g., Christensen *et al.*, 1971), where, in addition to the first two powers of I(1) regressors, i.e., x_{jt} and x_{jt}^2, the simple cross-products $x_{it}x_{jt}$ are also included. Note that the results in Vogelsang and Wagner (2014b) allow for the handling of the general polynomial case, including cross-product terms. That paper uses the IM-OLS estimator of Vogelsang and Wagner (2014a) that can more easily be extended to allow for cross-products than the FM-OLS estimator.[2]

Assumption 1 is sufficient for an invariance principle to hold for $\{\xi_t\}_{t\in\mathbb{Z}} = \left\{[u_t, v_t']'\right\}_{t\in\mathbb{Z}}$:

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_t \Rightarrow B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \Omega^{1/2} W(r), \tag{3}$$

with the long-run covariance matrix $\Omega := \sum_{h=-\infty}^{\infty} \mathbb{E}(\xi_0 \xi_h')$ and $W(r)$ an $m+1$-dimensional standard Wiener process. Partitioning $W(r) =: \left[W_{u\cdot v}(r), W_v(r)'\right]'$ similarly to the partitioning of $B(r)$ and using the Cholesky decomposition of Ω we get:

$$\begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \begin{bmatrix} \omega_{u\cdot v}^{1/2} & \Omega_{uv}(\Omega_{vv}^{-1/2})' \\ 0 & \Omega_{vv}^{1/2} \end{bmatrix} \begin{bmatrix} W_{u\cdot v}(r) \\ W_v(r) \end{bmatrix}, \tag{4}$$

with $\omega_{u\cdot v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$. For later use we also define the one-sided long-run covariance matrix $\Delta := \sum_{h=0}^{\infty} \mathbb{E}(\xi_0 \xi_h')$. Both matrices are partitioned

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

according to the partitioning of ξ_t:

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

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 $\Delta_{v_j u}$. The stacked vector of the powers of the Brownian motions corresponding to the different regressors is denoted as $\mathbf{B}_v(r) := \left[\mathbf{B}_{v_1}(r)', \ldots, \mathbf{B}_{v_m}(r)'\right]'$, with $\mathbf{B}_{v_j}(r) = \left[B_{v_j}(r), B_{v_j}(r)^2, \ldots, B_{v_j}(r)^{p_j}\right]'$ for $j = 1, \ldots, m$. Denoting $G_{X_j}(T) := \operatorname{diag}\left(T^{-1}, \ldots, T^{-\frac{p_j+1}{2}}\right)$ it follows immediately from Assumption 1 that $T^{1/2} G_{X_j}(T) X_{j,[rT]} \Rightarrow \mathbf{B}_{v_j}(r)$ for $0 \le r \le 1$. For the deterministic component, using $G_D(T) := \operatorname{diag}\left(T^{-1/2}, \ldots, T^{-(q+1/2)}\right)$, it immediately follows that $T^{1/2} G_D(T) D_{[rT]} \Rightarrow D(r)$, with $D(r) = \left[1, \ldots, r^q\right]'$, for $0 \le r \le 1$. For later use define $G(T) := \operatorname{diag}(G_D(T), G_X(T))$ with $G_X(T) := \operatorname{diag}(G_{X_1}(T), \ldots, G_{X_m}(T))$. We will typically use the shorthand notation G for $G(T)$.

Some additional remarks are in order: First, Chang *et al.* (2001) include predetermined stationary regressors in the model. Similar to Chang *et al.* (2001), the earlier working paper Hong and Wagner (2008) considers the setup discussed here, including in addition predetermined stationary regressors. For brevity we abstain from including them here. Note that the available code allows also for predetermined stationary regressors. Second, of course, not all consecutive powers of integrated regressors need to be included. Third, the results in the paper extend to triangular systems with multivariate y_t as in the linear case considered in Phillips and Hansen (1990), with the required changes being immediate. Fourth, more general deterministic components can be included, provided that the corresponding (limit) quantities are linearly independent (in $L^2[0, 1]$).

2.2. Fully Modified OLS Estimation

It can be seen from the proof of Proposition 1 that the OLS estimator $\hat{\theta} := \left(Z'Z\right)^{-1} Z'y$ of θ in (1) is consistent with a limiting distribution contaminated by second order bias terms. The presence of these second order bias terms, which depend upon the extent of regressor endogeneity and error serial correlation, renders inference with the OLS estimator difficult. This is very similar to the linear cointegration case and therefore, as in the linear case, the OLS estimator is modified to allow for asymptotic standard inference based on a zero-mean Gaussian mixture limiting distribution. In this paper we consider a straightforward extension of the FM-OLS estimator of Phillips and Hansen (1990), which rests upon a two-part transformation. First, the dependent variable y_t is replaced by $y_t^+ := y_t - \Delta x_t' \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}$, with $y^+ := \left[y_1^+, \ldots, y_T^+\right]'$. This modification is exactly the same as in the linear case. Second, additive correction factors, that depend

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

upon the specification, are required. These are given by:

$$M^* := \begin{bmatrix} M_1^* \\ \vdots \\ M_m^* \end{bmatrix}, \quad M_j^* := \hat{\Delta}_{v_j u}^+ \begin{bmatrix} T \\ 2\sum x_{jt} \\ \vdots \\ p_j \sum x_{jt}^{p_j - 1} \end{bmatrix}, \tag{5}$$

with $\hat{\Delta}_{v_j u}^+ := \hat{\Delta}_{v_j u} - \hat{\mathbf{\Omega}}_{uv} \hat{\mathbf{\Omega}}_{vv}^{-1} \hat{\Delta}_{vv_j}$.

In both the definition of y_t^+ as well as in the definition of the correction factors M_j^* estimates of the long-run variances are required. *Consistent* estimation of the long-run variances, under standard assumptions on the kernel and bandwidth, is based on the ordinary least squares residuals $\hat{u}_t$ from (1) in conjunction with the first differences of the regressors, i.e., it is based upon $\left[\hat{u}_t, \Delta x_t'\right]'$. Detailed discussions and sets of assumptions are given in, e.g., Phillips (1995) or Jansson (2002). To verify fulfillment of the conditions of Jansson (2002), we could formulate our primitive assumptions in terms of linear processes, as in Jansson (2002), combined with Assumption 1. For the purpose of this paper, however, any set of assumptions that allows for consistent long-run variance estimation is feasible and we therefore abstain from a more detailed discussion.

PROPOSITION 1. *Let y_t be generated by (1) with the regressors x_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:

$$A^* := \begin{bmatrix} 0_{(q+1)\times 1} \\ M^* \end{bmatrix},$$

with M^* as given in (5) using consistent estimators of the required long-run variances. Then $\hat{\theta}^+$ is consistent and its asymptotic distribution is given by:

$$G^{-1}\left(\hat{\theta}^+ - \theta\right) \Rightarrow \left(\int_0^1 J(r) J(r)' dr\right)^{-1} \int_0^1 J(r) dB_{u \cdot v}(r), \tag{6}$$

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

The limiting distribution given in (6) is a zero-mean Gaussian mixture since $\mathbf{B}_v(r)$ and $B_{u \cdot v}(r)$ are independent by construction. Therefore, conditional upon $\mathbf{B}_v(r)$, the limiting distribution is normal with zero-mean and covariance matrix:

$$V_{FM} = \omega_{u \cdot v} \left(\int_0^1 J(r) J(r)' dr\right)^{-1}. \tag{7}$$

A consistent estimator of V_{FM} is given by $\hat{V}_{FM} = \hat{\omega}_{u \cdot v} \left(GZ'ZG\right)^{-1}$, with $\hat{\omega}_{u \cdot v}$ denoting a consistent estimator of $\omega_{u \cdot v}$.

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

The limiting distribution of the FM-OLS estimator forms the basis for asymptotic chi-squared inference, as discussed in Phillips and Hansen (1990, Theorem 5.1 and the discussion on p. 106). Because the vector $\hat{\theta}^+$ has elements that converge at different rates, obtaining formal results for the Wald statistics requires a condition on the restriction matrix R that is unnecessary when all estimated coefficients converge at the same rate. As is well known in the literature, for a given constraint (a given row of R), the estimator with the slowest rate of convergence dominates the asymptotic distribution of the linear combination implied by the constraint, see, e.g., Park and Phillips (1988, 1989). When two or more restrictions are tested, it is not necessarily the case that the slowest converging estimator dominates a given restriction. Should another restriction involve that slowest converging estimator, it is usually possible for the restrictions to be rotated so that: (a) the slowest rate estimator only appears in one restriction, and (b) the Wald statistic has the exact same value. Because of this possibility, we do not state conditions on R related to the rates of convergence of the estimators involved in the constraints. We merely state here a sufficient condition on the constraint matrix $R \in \mathbb{R}^{s\times(q+1+p)}$ under which the Wald statistics have chi-squared limiting distributions. We assume that there exists a scaling matrix $G_R = G_R(T) \in \mathbb{R}^{s\times s}$, 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.

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

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

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

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

is asymptotically χ^2_s distributed under the null hypothesis.

Hypothesis testing can also be performed with a Lagrange Multiplier (LM) statistic rather than a Wald statistic. The corresponding LM statistic looks slightly different than the Wald statistic but has, of course, the same limiting distribution. For brevity we only consider the LM statistic in the following subsection when considering specification testing based on augmented regressions (Wald statistic) and auxiliary regressions (LM statistic).

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

2.3. Specification Testing Based on Augmented and Auxiliary Regressions

Misspecification of (1), meaning nonstationarity of u_t, can arise for two reasons: First, the functional form is misspecified but some other form of nonlinear cointegration between y_t and a different function of x_t (and potentially other variables) prevails. Second, there is no cointegration between the variables considered. We use a suitably constructed RESET-type test, in the spirit of Ramsey (1969), as a specification test for the null of correct specification of (1). The RESET test has a long tradition in specification testing in a stationary setup, see Phillips (1983), Lee *et al.* (1993), or de Benedictis and Giles (1998). We consider a version of the RESET test introduced by Thursby and Schmidt (1977), which is based on adding additional regressors to the original equation rather than adding powers of the fitted values. The additional regressors can be added to the original regression and this augmented regression is estimated (Wald test). Alternatively, the residuals of the original regression can be regressed on the additional regressors in an auxiliary regression (LM test). We consider both approaches.

The properties of RESET as well as related tests, such as the CUSUM test (see Xiao and Phillips, 2002), have been analyzed in similar settings in a variety of papers, see, e.g., Arai (2004), Hong and Phillips (2010), and Kasparis (2008). The corresponding results regarding the null and alternative properties of the tests extend (with obvious changes) to our setting. Specification tests like RESET-type tests are consistent against misspecified cointegration when the misspecified component consists of *H-regular* functions. They are not consistent against misspecified cointegration when the misspecified component consists of *I-regular* functions, using the definitions and terminology of Park and Phillips (1999, 2001). Furthermore, such tests are consistent against the alternative of no cointegration, i.e., the alternative of a spurious relation between y_t and x_t. For the case of misspecified additively separable *I-regular* components and pre-determined regressors, Kasparis (2010) presents a specification test based on Bierens (1990). Further investigations in this direction are, however, beyond the scope of the present paper that focuses on polynomial relationships, which are a special case of *H-regular* functions.

For the discussion of the tests we define $\bar{D}_t := \left[t^{q+1}, \ldots, t^{q+n}\right]'$, $\bar{X}_{jt} := \left[x_{jt}^{p_j+1}, x_{jt}^{p_j+2}, \ldots, x_{jt}^{p_j+r_j}\right]'$ for $j = 1, \ldots, m$, $Q_{it} := \left[q_{it}^1, q_{it}^2, \ldots, q_{it}^{s_i}\right]'$ for $i = 1, \ldots, k$, $F_t := \left[\bar{D}_t', \bar{X}_{1t}', \ldots, \bar{X}_{mt}', Q_{1t}', \ldots, Q_{kt}'\right]'$, and $F := \left[F_1, \ldots, F_T\right]'$. 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:

$$y = Z\theta + F\theta_F + \phi, \tag{10}$$

with $\phi := [\phi_1, \ldots, \phi_T]'$. If equation (10) is well specified the parameters can be estimated consistently by FM-OLS, according to Proposition 1, if the additional

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

regressors q_{it} fit into the framework of Assumption 1 (as provided in Section 2.1), which is now modified to accommodate the additional integrated regressors.

Assumption 2. When considering additional integrated regressors q_{it} and their integer powers define $\tilde{v}_t := \left[v_t', v_t^{*\prime}\right]' = \left[\Delta x_t', \Delta q_t'\right]'$, 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 $\left\{\tilde{\xi}_t\right\}_{t\in\mathbb{Z}} := \left[u_t, \tilde{v}_t'\right]'$, with $\tilde{v}_t = C_{\tilde{v}}(L)\tilde{\varepsilon}_t$, and $C_{\tilde{v}}(L)$ and $\tilde{\varepsilon}_t$ extended accordingly.

PROPOSITION 3. *Let y_t be generated by (1) with the regressors x_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 (10), with $\tilde{F} := F - Z(Z'Z)^{-1}Z'F$ and let $\hat{\omega}_{u\cdot\tilde{v}}$ be a consistent estimator of $\omega_{u\cdot\tilde{v}} := \Omega_{uu} - \Omega_{u\tilde{v}}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}$. Then it holds that the Wald test statistic for the null hypothesis $H_0 : \theta_F = 0$ in equation (10), given by:*

$$T_W := \frac{\hat{\theta}_F^{+\prime}\left(\tilde{F}'\tilde{F}\right)\hat{\theta}_F^+}{\hat{\omega}_{u\cdot\tilde{v}}}, \tag{11}$$

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

Arai (2004) considers the Wald version of the RESET test for the null hypothesis of linear cointegration, i.e., he considers the case $p_1 = \ldots = p_m = 1$. Arai (2004) uses the dynamic OLS estimator (see, e.g., Saikkonen, 1991) for the augmented regression rather than the FM-OLS estimator.

The Lagrange Multiplier (LM) test is based on the FM-OLS residuals $\hat{u}_t^+$ of (1), which are regressed on $\tilde{F}$ in the auxiliary regression:

$$\hat{u}^+ = \tilde{F}\theta_{\tilde{F}} + \psi_t, \tag{12}$$

with $\hat{u}_t^+ := y_t^+ - Z_t'\hat{\theta}^+$ and $\hat{u}^+ := \left[\hat{u}_1^+, \ldots, \hat{u}_T^+\right]'$. To allow for asymptotic standard inference, in general, the coefficients $\theta_{\tilde{F}}$ have to be estimated with a suitable FM-OLS-type estimator to achieve a zero-mean Gaussian mixture limiting distribution. This is necessary because the limiting distribution of the OLS estimator of $\theta_{\tilde{F}}$ in (12) 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 when (1) is well specified.

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 (12) as:*

$$\hat{\theta}_{\tilde{F}}^+ := \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\hat{u}^+ - O^{F*} - A^{F*} + k^{F*}A^*\right), \tag{13}$$

with:

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

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

where $\upsilon = [\upsilon_1, \ldots, \upsilon_T]'$, $\tilde{\upsilon} = [\tilde{\upsilon}_1, \ldots, \tilde{\upsilon}_T]'$, *and* $A^{F*} := \left[0'_{n\times 1}, M^{*\prime}_{\tilde{X}_1}, \ldots, M^{*\prime}_{\tilde{X}_m}, M^{*\prime}_{Q_1}, \ldots, M^{*\prime}_{Q_k}\right]'$, *with:*

$$M_{\tilde{X}_j} = \hat{\Delta}^{+}_{\upsilon_j u} \begin{bmatrix} (p_j+1)\sum x_{jt}^{p_j} \\ \vdots \\ (p_j+r_j)\sum x_{jt}^{p_j+r_j-1} \end{bmatrix}, \quad M_{Q_i} = \hat{\Delta}^{+}_{\upsilon_i^* u} \begin{bmatrix} T \\ 2\sum q_{it} \\ \vdots \\ s_i \sum q_{it}^{s_i-1} \end{bmatrix}, \tag{14}$$

$k^{F*} = F'Z(Z'Z)^{-1}$, $\hat{\Delta}^{+}_{\upsilon_j u}$, A^* *as defined above in Proposition 1, and* $\hat{\Delta}^{+}_{\upsilon_i^* u} := \hat{\Delta}_{\upsilon_i^* u} - \hat{\Omega}_{u\upsilon}\hat{\Omega}^{-1}_{\upsilon\upsilon}\hat{\Delta}_{\upsilon\upsilon_i^*}$. *Then, under the null hypothesis* $H_0 : \theta_{\tilde{F}} = 0$, *when using consistent long-run variance estimators for the FM-OLS estimator defined in (13), it holds that:*

$$G_F^{-1}\hat{\theta}^{+}_{\tilde{F}} \Rightarrow \left(\int_0^1 \tilde{J}^F(r)\tilde{J}^F(r)'dr\right)^{-1} \int_0^1 \tilde{J}^F(r) dB_{u\cdot\tilde{\upsilon}}(r), \tag{15}$$

with:

$$\mathbf{B}^F_{\upsilon_j}(r) := \left[B_{\upsilon_j}(r)^{p_j+1}, \ldots, B_{\upsilon_j}(r)^{p_j+r_j}\right]', \quad \text{for } j = 1, \ldots, m \tag{16}$$

$$\mathbf{B}^F_{\upsilon_i^*}(r) := \left[B_{\upsilon_i^*}(r), \ldots, B_{\upsilon_i^*}(r)^{s_i}\right]', \quad \text{for } i = 1, \ldots, k \tag{17}$$

$$J^F(r) := \left[\bar{D}(r)', \mathbf{B}^F_{\upsilon_1}(r)', \ldots, \mathbf{B}^F_{\upsilon_m}(r)', \mathbf{B}^F_{\upsilon_1^*}(r)', \ldots, \mathbf{B}^F_{\upsilon_k^*}(r)'\right]' \tag{18}$$

$$\tilde{J}^F(r) := J^F(r) - \int_0^1 J^F(s)J(s)'ds \left(\int_0^1 J(s)J(s)'ds\right)^{-1} J(r), \tag{19}$$

the corresponding scaling matrix G_F *(defined in the proof in the appendix) and* $B_{u\cdot\tilde{\upsilon}}(r) := B_u(r) - \tilde{B}_\upsilon(r)'\Omega^{-1}_{\tilde{\upsilon}\tilde{\upsilon}}\Omega_{\tilde{\upsilon}u}$. *Consequently, the LM test statistic for the null hypothesis* $H_0 : \theta_{\tilde{F}} = 0$ *in (12), given by:*

$$T_{LM} := \frac{\hat{\theta}^{+\prime}_{\tilde{F}}(\tilde{F}'\tilde{F})\hat{\theta}^{+}_{\tilde{F}}}{\hat{\omega}_{u\cdot\tilde{\upsilon}}}, \tag{20}$$

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

Proposition 4 can be seen as a generalization of the modified RESET test presented in Hong and Phillips (2010, Theorem 3), who derive a related test in a 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 is that Hong and Phillips (2010) use the OLS residuals $\hat{u}_t$ of the linear cointegrating relationship as the dependent variable in the auxiliary regression, which leads to different bias correction terms than ours, which are based on the

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

FM-OLS residuals $\hat{u}_t^+$. If desired, the LM specification test can also be based on the OLS residuals rather than on the FM-OLS residuals in our more general case. The construction of appropriate bias correction terms is straightforward using similar arguments to those in Proposition 4. For specification analysis, if only higher-order polynomial trends are included in F_t, 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 and thus their approach is nested within ours.

2.4. A KPSS-Type Test for Cointegration

In this section we discuss a residual-based test for nonlinear cointegration, which prevails in (1) if the error process $\{u_t\}_{t\in\mathbb{Z}}$ is indeed stationary. To test this null hypothesis directly, we present a KPSS-type test statistic based on the FM-OLS residuals $\hat{u}_t^+$ of (1).[3] 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 also be interpreted as specification tests to a certain extent, since persistent nonstationary errors prevail if, e.g., relevant I(1) regressors are omitted in (1). The test statistic is given by:

$$CT := \frac{1}{T\hat{\omega}_{u\cdot v}} \sum_{t=1}^{T} \left(T^{-1/2} \sum_{j=1}^{t} \hat{u}_j^+ \right)^2, \tag{21}$$

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

PROPOSITION 5. *Let y_t be generated by (1), with the regressors x_t and errors u_t satisfying Assumption 1, and let $\hat{\omega}_{u\cdot v}$ be a consistent estimator of $\omega_{u\cdot v}$. Then, as $T \to \infty$, it holds for the test statistic (21) defined above that:*

$$CT \Rightarrow \int_0^1 \left(W_{u\cdot v}^J(r) \right)^2 dr, \tag{22}$$

with:

$$W_{u\cdot v}^J(r) := W_{u\cdot v}(r) - \int_0^r J(s)' ds \left(\int_0^1 J(s) J(s)' ds \right)^{-1} \int_0^1 J(s) dW_{u\cdot v}(s), \tag{23}$$

where $W_{u\cdot v}(r)$ is a standard Wiener process independent of $W_v(r)$.
In the special case that only one element of x_t appears with higher-order powers in (1), the CT statistic has the following limiting distribution:

$$CT \Rightarrow \int_0^1 \left(W_{u\cdot v}^{J_W}(r) \right)^2 dr, \tag{24}$$

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

with:

$$W_{u\cdot v}^{J_W}(r) := W_{u\cdot v}(r) - \int_0^r J^W(s)'ds\left(\int_0^1 J^W(s)J^W(s)'ds\right)^{-1}\int_0^1 J^W(s)dW_{u\cdot v}(s) \quad \textbf{(25)}$$

and $J^W(r) := \left[D(r)', W_v(r)', W_{v_m}(r)^2, \ldots, W_{v_m}(r)^{p_m}\right]'$, *where* $W_v(r)$ *is a vector of standard Wiener processes independent of* $W_{u\cdot v}(r)$. *Note that we have assumed here, without loss of generality, that the last element of* x_t, *i.e.,* x_{mt}, *is the regressor appearing with higher-order powers.*

In the general case, the limiting distribution given in (22) and (23) depends upon the correlation structure between the variables x_{jt} and their powers and, thus, cannot be simulated because of this nuisance parameter dependency. The limiting distribution given in (24) and (25) only depends upon standard Brownian motions, the specification of the deterministic component and the powers of the integrated regressor x_{mt}. Therefore, in this special case (which covers our original EKC motivation as well as intensity of use curves) critical values can be simulated and are available upon request.

For the general case, a test statistic for which critical values can be tabulated is derived in the following proposition using subsampling techniques. This proposition parallels Choi and Saikkonen (2010), who consider a similar testing problem in a dynamic nonlinear least squares estimation framework. This limiting distribution does not depend on the number of integrated regressors, the powers of the integrated regressors included, or the specification of the deterministic component.

PROPOSITION 6. *Under the same assumptions as in Proposition 5, it holds for* $T \to \infty$ *and* b, *such that* $b \to \infty$ *and* $b/T \to 0$, *that:*

$$CT_{b,i} := \frac{1}{b\hat{\omega}_{u\cdot v}}\sum_{t=i}^{i+b-1}\left(b^{-1/2}\sum_{j=i}^{t}\hat{u}_j^+\right)^2 \Rightarrow \int_0^1 W(r)^2dr, \quad \textbf{(26)}$$

where $W(r)$ *is a standard Wiener process.*

Note that for a given block size b there are $M := \lfloor T/b \rfloor$ subsamples and corresponding test statistics, $\{CT_{b,1}, \ldots, CT_{b,M}\}$, that all lead to asymptotically valid statistics for the same null hypothesis. Using only one of the subsample test statistics can be expected to lead to reduced power (compare to Choi and Saikkonen, 2010, p. 692). Therefore we consider using this set of statistics in combination with the Bonferroni inequality to modify the critical values using:

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

where $CT_{b,\max} := \max(CT_{b,1}, \ldots, CT_{b,M})$ and $c_{\alpha/M}$ denotes the α/M-percent critical value of the distribution of $\int_0^1 W^2(r)dr$. For the computation of the

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

critical values from the distribution function, F_{W^2} say, of $\int_0^1 W^2(r)dr$ Choi and Saikkonen (2010) obtain the interesting result that:

$$F_{W^2}(z) = \sqrt{2}\sum_{n=0}^{\infty} \frac{\Gamma(n+1/2)}{n!\Gamma(1/2)}(-1)^n \left(1 - f\left(\frac{\sqrt{2}/2 + 2n\sqrt{2}}{2\sqrt{z}}\right)\right), \quad z \geq 0, \qquad \textbf{(27)}$$

with $f(x) = \frac{2}{\sqrt{\pi}} \int_0^x exp\left(-y^2\right)dy$. 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 $n = 10$ (as used in Choi and Saikkonen, 2010) are given in Table B1 of the online Appendix B.

Another important practical problem when using the subsample based test is the choice of the block length b. Following Choi and Saikkonen (2010), we apply the *minimum volatility rule* proposed by Romano and Wolf (2001, p. 1297). More details on this are contained in the online Appendix B. We refer to the test procedure using the Bonferroni bound, i.e., when the null hypothesis is rejected if $CT_{b,\max} \geq c_{\alpha/M}$, with the block length chosen with the minimum-volatility rule, as CS test in the simulations.[4]

3. FINITE SAMPLE PERFORMANCE

We generate data for the quadratic cointegrating polynomial regression model:

$$y_t = c + \delta t + \beta_1 x_t + \beta_2 x_t^2 + u_t, \qquad \textbf{(28)}$$

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

$$u_t = \rho_1 u_{t-1} + e_{1,t} + \rho_2 e_{2,t}, \quad u_0 = 0,$$
$$v_t = e_{2,t} + 0.5 e_{2,t-1},$$

with $(e_{1,t}, e_{2,t})' \sim \mathcal{N}(0, I_2)$. The parameter ρ_1 controls the level of serial correlation in the error term u_t, and ρ_2 controls the extent of regressor endogeneity. The parameter values are $c = \delta = 1$, $\beta_1 = 5$, and $\beta_2 = -0.3$. The values for β_1 and β_2 are based on coefficient estimates obtained when applying the developed FM-OLS estimator to GDP and CO_2 emissions data for Austria (see Wagner, 2014a). We present simulation results for $T \in \{100, 200\}$ and for $\rho_1 = \rho_2 \in \{0, 0.3, 0.6, 0.8\}$. The number of replications is 5,000 in all cases and all tests are carried out at the nominal 5% level.

The FM-OLS estimator and test statistics based upon it are computed for the Bartlett and Quadratic Spectral (QS) kernels and two bandwidth choices. These bandwidth choices are 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 Newey and West (1994) suggested it be used as a simplified, feasible rule especially in conjunction with the Bartlett kernel.[5] All results are benchmarked against OLS results. Inference is performed in two ways for the

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

OLS estimator, none of which is asymptotically valid in the presence of serial correlation and endogeneity. One way is to perform textbook OLS inference ignoring serial correlation and endogeneity altogether, labelled OLS later. The other is to use 'HAC-robust' standard errors, as is often done in stationary regression settings, with the bandwidth chosen according to Newey and West (1994), labelled HAC later. Rejections for both the FM-OLS and the OLS tests are carried out using the standard normal distribution for t-tests and the chi-squared distribution for Wald tests.

As mentioned in the introduction, the results for bias and RMSE of the estimated coefficients are similar to results that have been found for the FM-OLS estimator of Phillips and Hansen (1990) in standard cointegration settings.[6] In the absence of serial correlation and endogeneity, the OLS estimator has, as expected, the best performance both with respect to bias and RMSE. With increasing serial correlation and/or endogeneity the FM-OLS estimator increasingly outperforms the OLS estimator. The bias of FM-OLS can be about 50% smaller than the bias of OLS in the case of $\rho_1, \rho_2 = 0.6$. With respect to RMSE, the performance advantages of FM-OLS are not as big. The results are qualitatively similar for coefficients corresponding to both deterministic as well as integrated regressors. In the simulations, it turns out that neither the choice of the kernel function (Bartlett or QS) nor the bandwidth choice has a big impact on estimator performance.

We now briefly turn to hypothesis testing, with null rejection probabilities displayed in Table 1, where Panel A shows t-test results for $H_0 : \beta_1 = 5$, and Panel B shows Wald test results for the joint null hypothesis $H_0 : \beta_1 = 5, \beta_2 = -0.3$. For $\rho_1, \rho_2 = 0$ the textbook OLS statistics show the smallest, in fact almost no, size distortions. Increasing the values of ρ_1, ρ_2 leads to increasing over-rejections for all tests, with the effects being larger for the OLS than the FM-OLS based tests. Thus, the FM-OLS corrections do lead to improved inference. Note that, in this respect, the HAC test statistic shows distinctly smaller over-rejections than the OLS test. For the larger values of ρ_1, ρ_2, the Andrews (1991) bandwidth rule leads to the smallest size distortions, reflecting the fact that data-dependent bandwidth choices – rather than merely sample size dependent bandwidth choices – can lead to improved inference (in contrast to the minimal effects of bandwidth choices on estimator performance). Note, however, that for the Wald test, when $\rho_1, \rho_2 = 0.8$ (even for $T = 200$) size distortions are large even using the Andrews (1991) bandwidth, with actual size at about 45%.

Let us briefly summarize the main findings from simulating (size-corrected) power of the tests against sequences of alternatives. The detailed results are contained in online Appendix C in Figures C1 to C7. The main observation is, as expected, that the FM-OLS tests outperform the OLS-based tests since they have similar or higher size-corrected power as the OLS-based tests but lower size distortions. The bandwidth choice exhibits some impact on size-corrected power. As discussed, under the null the smallest over-rejections are often observed for the Andrews (1991) bandwidth. This, however, comes at the price of a small loss in

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

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	AND	NW	HAC	AND	NW
0.0	.0594	.1060	.1058	.0932	.1222	.1234	.1136
0.3	.1542	.1424	.1162	.1092	.1462	.1164	.1118
0.6	.3706	.2678	.1604	.1716	.2498	.1488	.1514
0.8	.5876	.4612	.3108	.3774	.4286	.2984	.3408

$T = 200$

		Bartlett Kernel			QS Kernel		
ρ_1, ρ_2	OLS	HAC	AND	NW	HAC	AND	NW
0.0	.0478	.0784	.0738	.0650	.0890	.0784	.0726
0.3	.1472	.1100	.0878	.0866	.1058	.0802	.0790
0.6	.3736	.2306	.1350	.1660	.2010	.1242	.1406
0.8	.6154	.4410	.2930	.3820	.3968	.2846	.3446

Panel B: Wald tests for $H_0 : \beta_1 = 5, \beta_2 = -0.3$
$T = 100$

		Bartlett Kernel			QS Kernel		
ρ_1, ρ_2	OLS	HAC	AND	NW	HAC	AND	NW
0.0	.0568	.1348	.1238	.1092	.1706	.1466	.1352
0.3	.2002	.1958	.1470	.1402	.2032	.1470	.1372
0.6	.5258	.3878	.2226	.2472	.3550	.2070	.2120
0.8	.8124	.6652	.4800	.5650	.6250	.4654	.5188

$T = 200$

		Bartlett Kernel			QS Kernel		
ρ_1, ρ_2	OLS	HAC	AND	NW	HAC	AND	NW
0.0	.0532	.0910	.0880	.0740	.1118	.0960	.0830
0.3	.1924	.1426	.1034	.1038	.1384	.0956	.0922
0.6	.5290	.3140	.1870	.2258	.2720	.1698	.1898
0.8	.8254	.6228	.4602	.5758	.5638	.4514	.5348

size-corrected power compared to the simple Newey and West (1994) rule. The kernel choice has virtually no effect on the power properties.

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

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 under the null, given as before in (28), and for the following alternative DGPs:

(A) : $y_t = 1 + t + 5x_t - 0.3x_t^2 + 0.01x_t^3 + u_t$

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

(C) : y_t and x_t are two independent I(1) variables

In case (A) the regressor x_t and error u_t are generated as described above (with the same values of ρ_1, ρ_2 as before). Also in case (B) x_t is generated as before and $e_t = \sum_{j=1}^{t} \varepsilon_j$, with $\varepsilon_t \sim \mathcal{N}(0,1)$ independent of x_t. Finally, in case (C) both y_t and x_t are generated, independently of each other, 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 no cointegration because of a missing integrated regressor, and alternative (C) corresponds to a spurious regression alternative.

For specification testing, we consider four different sets of additional regressors: The first set of additional regressors is based on the idea of Park and Choi (1988) and Park (1990) to include higher-order deterministic trends. We implement this idea by setting $F_t = \left[t^2,\ t^3\right]'$, labelled I in Tables 2 and 3. The second set of regressors is given by $F_t = \left[x_t^3,\ x_t^4,\ q_t\right]'$, where q_t is an independent I(1) regressor (generated similarly to e_t above), labelled II. The third choice combines the first two by including both higher-order deterministic trends and higher-order powers of the integrated regressor, i.e., $F_t = \left[t^2,\ t^3,\ x_t^3,\ x_t^4\right]'$, labelled III. The fourth choice is to only include powers of an independent I(1) variable q_t, i.e., $F_t = \left[q_t,\ q_t^2,\ q_t^3\right]'$, labelled IV. For brevity we only report results for the Bartlett kernel and the Newey and West (1994) bandwidth choice.[7] The starting point of all tests is equation (28). For the Wald tests the regressors F_t are added to (28), and for the LM tests the FM-OLS residuals from estimating (28) are used as the dependent variable in the subsequent auxiliary regression. Also the CT and CS tests are computed from the FM-OLS residuals of (28). The block lengths for the CS test are computed with the minimum volatility rule. Rejections for the specification tests are based on the chi-squared distributions, for the CT test based on simulating the limiting distribution derived in Proposition 5 and for the CS test using an approximation to the analytical expression given in (27) with $n = 30$ (given in Table B1 in the online Appendix B).

Table 2 displays null rejection probabilities. The first main observation that emerges is that for all four variants I to IV the Wald tests have larger over-rejections than the LM tests. The differences between the Wald and the LM test are partly increasing for increasing values of ρ_1, ρ_2. It should also be noted that the over-rejections are partly increasing with T for the larger values of ρ_1, ρ_2. This effect is, however, due to the simple Newey and West (1994) sample size

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

TABLE 2. Empirical 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					
(28)	0.0	.1286	.1946	.1954	.1904	.0424	.1158	.0588	.1240	.0592	.0006
	0.3	.1734	.2178	.2316	.2142	.0672	.1296	.0638	.1418	.0914	.0034
	0.6	.2984	.2582	.3452	.2750	.1420	.1470	.0974	.1728	.2178	.0164
	0.8	.5484	.3642	.5936	.4064	.3376	.2028	.2520	.2596	.5244	.0676
						Panel B: T = 200					
(28)	0.0	.0858	.1154	.1166	.1100	.0478	.0784	.0548	.0784	.0554	.0016
	0.3	.1188	.1476	.1598	.1424	.0754	.0994	.0798	.1046	.0894	.0074
	0.6	.2572	.2226	.3082	.2324	.1534	.1504	.1498	.1678	.2368	.0372
	0.8	.5556	.3862	.6094	.4170	.3994	.2650	.3896	.3156	.6470	.1814

dependent bandwidth choice and is not present when, e.g., the Andrews (1991) bandwidth rule is chosen.[8] The second main observation from Table 2 is that the over-rejections differ strongly between variants I to IV of both the Wald and the LM tests. Over-rejection patterns for the CT test are as expected. The nominal level is maintained for $\rho_1, \rho_2 = 0$ but strong over-rejections occur for $\rho_1, \rho_2 = 0.8$. The subsampling cum Bonferroni bound CS test exhibits substantial under-rejections (except for $\rho_1, \rho_2 = 0.8$), reflecting the conservativeness implied by using the Bonferroni bound.

We close this section by considering size-corrected power of the tests reported in Table 3.[9] We start with the Wald and LM tests, where the table shows that size-corrected power varies substantially across variants and alternatives. For alternative (A) variants II and III exhibit the highest size-corrected power. This reflects the fact that alternative (A) is given by a cubic CPR, and for both tests II and III the third power x_t^3 is included in F_t. For alternatives (B) and (C), variants I and III have the highest size-corrected power. The difference between these two variants and variants II and IV is larger for $T = 100$ than for $T = 200$. Thus, in the case of an omitted I(1) regressor, alternative (B), or a spurious regression, alternative (C), the inclusion of higher-order powers of t leads to well-performing tests. Given that for variant III the set of additional regressors is $F_t = \left[t^2, t^3, x_t^3, x_t^4\right]'$, it is not surprising that this variant has high size-corrected power against all considered alternatives. Thus, if the researcher has no particular alternative in mind, a 'broad' test like variant III appears to be the preferred choice. Table 3 also shows that the differences in size-corrected power between the corresponding variants of the Wald and the LM test are often small. Therefore, given that the LM tests have smaller over-rejections under the null, we recommend using the LM version of the specification tests.

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

TABLE 3. Size Corrected 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	.3840	.9452	.9456	.1978	.3584	.9420	.9500	.1574	.3848	.2898
	0.3	.3188	.9032	.8986	.1702	.3078	.9046	.9170	.1406	.2944	.2174
	0.6	.2110	.8276	.8026	.1150	.2290	.8148	.8390	.1098	.1684	.1118
	0.8	.0936	.6910	.6120	.0632	.1210	.6576	.6826	.0720	.0514	.0452
(B)	–	.7322	.2850	.7212	.3778	.7250	.2358	.6404	.2936	.8258	.6214
(C)	–	.7236	.2636	.7020	.3728	.7100	.2130	.6088	.2924	.8750	.6374
						Panel B: T = 200					
(A)	0.0	.5776	.9999	.9998	.3838	.5498	.9998	.9996	.3684	.7078	.5810
	0.3	.5180	.9982	.9984	.3422	.4968	.9990	.9986	.3270	.6040	.4886
	0.6	.3798	.9926	.9898	.2516	.3910	.9926	.9916	.2380	.4120	.2880
	0.8	.1890	.9632	.9382	.1260	.2232	.9610	.9554	.1512	.1516	.0876
(B)	–	.8488	.5312	.8722	.6196	.8490	.5216	.8592	.6092	.9676	.9106
(C)	–	.8508	.5284	.8714	.6174	.8498	.5114	.8574	.6024	.9728	.9074

The CT and CS tests exhibit very good size-corrected (as well as raw) power properties against alternatives (B) and (C), and good (but less impressive) performance against the four alternatives in (A). In particular, size-corrected power against the cubic CPR alternatives (A) decreases as ρ_1, ρ_2 increase. This is in line with increasing over-rejections under the null for CT and decreasing under-rejections under the null for CS in this case. In this respect, it should be emphasized that the KPSS-type CT and CS tests have been constructed as cointegration rather than as specification tests. The performance in case of alternatives (B) and (C), with a focus on spurious regression type situations rather than misspecification of the polynomial order, appears quite acceptable.

4. SUMMARY AND CONCLUSIONS

This paper has discussed fully modified OLS estimation for additively separable cointegrating polynomial regressions. The zero-mean Gaussian mixture limiting distribution of the FM-OLS estimator forms the basis for asymptotic standard inference. In addition to hypothesis testing, RESET-type specification testing as well as a KPSS-type cointegration test have been discussed. The KPSS-type test has a nuisance parameter dependent limiting distribution in the general case. Therefore, parallelling Choi and Saikkonen (2010), a subsampling test that can be used in conjunction with the Bonferroni bound has been discussed. In the important special case of only one integrated regressor occurring with powers

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

(as in EKC analysis), the KPSS-type test has a nuisance parameter free limiting distribution that consequently can be simulated. The finite sample simulations indicate that the performance of FM-OLS compared to the OLS benchmark is, in many respects, similar to the relative performance of these two estimators in a standard cointegration setting. Increasing the extent of regressor endogeneity and/or error serial correlation leads to larger performance advantages of FM-OLS over OLS, in terms of both estimator and test performance. The simulations also indicate performance advantages of the LM version of the RESET-type specification test over the Wald version. The KPSS-type test works as expected, with the subsampling version suffering from the conservativeness implied by using the Bonferroni bound. Several directions of future research immediately follow: First, the restriction of additive separability needs to be overcome. Second, motivated by the empirical EKC literature, the analysis needs to be extended to multi-country settings by considering systems of seemingly unrelated cointegrating polynomial regressions or panel cointegrating polynomial regressions. Third, the performance of the specification tests needs to be analyzed in more detail, including a detailed comparison with other specification testing approaches. Fourth, the FM-OLS approach needs to be compared to competing approaches like dynamic nonlinear least squares or IM-OLS-type approaches. All these directions are currently under investigation.

NOTES

1. In addition to the vast empirical literature there is also a large theoretical literature exploring different mechanisms that lead to EKC-type relationships, see, e.g., the survey of Brock and Taylor (2005). Another empirical literature – in addition to the original name-giving inequality and income literature – that can benefit from the theory developed in this paper is the intensity of use literature that investigates the potentially inverted U-shaped relationship between GDP and energy or metals use per unit of GDP (see, e.g., Labson and Crompton, 1993).

2. Given that Vogelsang and Wagner (2014a, Proposition 2) show that the IM-OLS estimator is (conditionally) less efficient than the FM-OLS estimator for the linear case, an extension of the FM-OLS estimator to the general polynomial case is currently being investigated despite the availability of some IM-OLS results.

3. Shin (1994) extends the KPSS test from a stationarity to a linear cointegration test and here we consider the corresponding extension to our CPR case.

4. By construction, a test based on the Bonferroni bound is conservative and is known to be particularly conservative when the individual test statistics 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 online Appendix B we discuss some modifications to the Bonferroni bound due to Hommel (1988), Simes (1986), and Rom (1990). Unreported simulations show that using any of these modifications does not lead to very different behavior of the resulting tests compared to the CS test based on the standard Bonferroni bound.

5. Compared to the data-dependent rule of Andrews (1991), the simplified Newey and West (1994) rule is clearly computationally cheaper (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. Additionally, in the online Appendix C, relating to Hong and Phillips (2010, Theorems 4 and 5), the bandwidths $T^{1/5}$, $T^{1/4}$, $T^{1/3}$ are considered to gauge the effect of the ratio of the bandwidth to the sample size on the results.

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

6. See, e.g., the simulation section in Vogelsang and Wagner (2014a) where both OLS and FM-OLS are included in the simulations in a standard cointegration setting. Details with respect to bias and RMSE for our simulation setting are available in the online Appendix C in Tables C1 to C3.

7. Online Appendix C also contains the results for the Andrews (1991) bandwidth in Tables C6 to C8.

8. The material in the online Appendix C shows that using the Andrews (1991) bandwidth rule can lead to substantially smaller over-rejections under the null, see Table C6. However, this often comes at the price of substantially lower size-corrected power, especially against alternatives (B) and (C), see Table C7.

9. Clearly, for the CS test using the empirical critical values simulated under the null does not correspond to (empirical) size correction, as the CS test has a composite null hypothesis (and is conservative by construction). We would like to thank a referee for pointing this out. Please also note that raw power is reported in the online Appendix C in Tables C5 (Newey-West) and C8 (Andrews).

REFERENCES

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

Arai, Y. (2004) Testing Linearity in Regressions with I(1) Processes. Mimeo, University of Tokyo.

Bierens, H. (1990) A Consistent Conditional Moment Test of Functional Form. *Econometrica* 67, 1341–1383.

Brock, W.A. & 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.

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

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

Christensen, L.R., D.W. Jorgenson, & L.J. Lau (1971) Conjugate Duality and The Transcendental Logarithmic Production Function. *Econometrica* 39, 255–256.

de Benedictis, L.F. & 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, 383–417.

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

Grossman, G.M. & 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. & 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. & M. Wagner (2008) Nonlinear Cointegration Analysis and The Environmental Kuznets Curve. Institute for Advanced Studies, Economics Series No. 224.

Ibragimov, R. & 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.

Kasparis, I. (2008) Detection of Functional Form Misspecification in Cointegrating Relations. *Econometric Theory* 28, 1373–1403.

Kasparis, I. (2010) The Bierens Test for Certain Nonstationary Models. *Journal of Econometrics* 158, 221–230.

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

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, & 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. & 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, & 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.

Liang, H., P.C.B. Phillips, H. Wang, & Q. Wang (2014) Weak Convergence to Stochastic Integrals for Econometric Applications. Cowles Foundation, Discussion Paper 1971.

Newey, W. & 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, 107–133.

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

Park, J.Y. & P.C.B. Phillips (1988) Statistical Inference in Regressions with Integrated Processes: Part I. *Econometric Theory* 4, 468–497.

Park, J.Y. & P.C.B. Phillips (1989) Statistical Inference in Regressions with Integrated Processes: Part II. *Econometric Theory* 5, 95–131.

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

Park, J.Y. & 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, 123–167.

Philips, P.C.B. (1995) Fully Modified Least Squares and Vector Autoregression. *Econometrica* 63, 1023–1078.

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

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

Saikkonen, P. (1991) Asymptotically Efficient Estimation of Cointegrating Regressions. *Econometric Theory* 7, 1–21.

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

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

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

Vogelsang, T.J. & M. Wagner (2014a) Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions. *Journal of Econometrics* 148, 741–760.

Vogelsang, T.J. & M. Wagner (2014b) An Integrated Modified OLS RESET Test for Cointegrating Regressions. Technical University Dortmund, SFB823 Discussion Paper 37/14.

Wagner, M. (2012) The Phillips Unit Root Tests for Polynomials of Integrated Processes. *Economics Letters* 114, 299–303.

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

Wagner, M. (2014a) The Environmental Kuznets Curve, Cointegration and Nonlinearity. Forthcoming in *Journal of Applied Econometrics*, DOI: 10.1002/jae.2421

Wagner, M. (2014b) Cointegrated Translog Relationships: Fully Modified OLS Estimation and Inference. Mimeo.

Xiao, Z. & P.C.B. Phillips (2002) A CUSUM Test for Cointegration Using Regression Residuals. *Journal of Econometrics* 108, 43–61.

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

APPENDIX: Proofs

Proof of Proposition 1. Using similar arguments as Chang *et al.* (2001), the asymptotic behavior of the OLS estimator $\hat{\theta} := (Z'Z)^{-1}Z'y$ immediately follows:

$$\begin{aligned} G^{-1}(\hat{\theta}-\theta) &= (GZ'ZG)^{-1}GZ'u \\ &\Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \\ &\quad \times \left(\int_0^1 J(r)dB_{u\cdot v}(r) + \int_0^1 J(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu} + \begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}\right), \end{aligned} \tag{A.1}$$

with $M := \left[M_1', \ldots, M_m'\right]'$, and $M_j := \Delta_{v_j u}\left[1, 2\int_0^1 B_{v_j}(r)dr, \ldots, p_j \int_0^1 B_{v_j}(r)^{p_j-1}dr\right]'$ for $j = 1, \ldots, m$. The stated result for the FM-OLS estimator $\hat{\theta}^+$ follows from considering:

$$G^{-1}(\hat{\theta}^+ - \theta) = \left(GZ'ZG\right)^{-1}\left(GZ'u^+ - GA^*\right), \tag{A.2}$$

with $u_t^+ := u_t - \Delta x_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$, $u^+ := \left[u_1^+, \ldots, u_T^+\right]'$, and A^* as given in the main text. For the deterministic components it immediately follows that $G_D D'u^+ \Rightarrow \int_0^1 D(r)dB_{u\cdot v}(r)$. Using similar arguments as in the proof of Hong and Phillips (2010, Lemma 4) it follows that $T^{-\frac{k+1}{2}}\sum_{t=1}^T x_{jt}^k u_t^+ \Rightarrow \int_0^1 B_{v_j}(r)^k dB_{u\cdot v}(r) + k\left(\Delta_{v_j u} - \Omega_{uv}\Omega_{vv}^{-1}\Delta_{vv_j}\right)\int_0^1 B_{v_j}(r)^{k-1}dr$. Combining all terms we thus have that:

$$\begin{aligned} GZ'u^+ &\Rightarrow \int_0^1 J(r)dB_u(r) + \begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix} - \int_0^1 J(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu} \\ &= \int_0^1 J(r)dB_{u\cdot v}(r) + \begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}, \end{aligned} \tag{A.3}$$

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 restriction (8) with G_R invertible and R^* of full rank it holds that:

$$\begin{aligned} T_W &= (R\hat{\theta}^+ - r)'\left(\hat{\omega}_{u\cdot v}R(Z'Z)^{-1}R'\right)^{-1}\left(R\hat{\theta}^+ - r\right) \\ &= \left(R(\hat{\theta}^+ - \theta)\right)'\left(\hat{\omega}_{u\cdot v}R(Z'Z)^{-1}R'\right)^{-1}\left(R\left(\hat{\theta}^+ - \theta\right)\right) \end{aligned}$$

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

$$
\begin{aligned}
&= \left((G_R RG)G^{-1}\left(\hat{\theta}^+ - \theta\right)\right)'\left(\hat{\omega}_{u\cdot v}(G_R RG)(GZ'ZG)^{-1}\left(GR'G_R'\right)\right)^{-1} \\
&\quad \times \left((G_R RG)G^{-1}(\hat{\theta}^+ - \theta)\right) \\
&\Rightarrow \left[R^*\left(\int_0^1 J(r)J(r)'dr\right)^{-1}\int_0^1 J(r)dB_{u\cdot v}(r)\right]' \\
&\quad \times \left[\omega_{u\cdot v}R^*\left(\int_0^1 J(r)J(r)'dr\right)^{-1}R^{*\prime}\right]^{-1} \times \\
&\quad \times \left[R^*\left(\int_0^1 J(r)J(r)'dr\right)^{-1}\int_0^1 J(r)dB_{u\cdot v}(r)\right], \qquad \textbf{(A.4)}
\end{aligned}
$$

which can be shown to be chi-squared distributed with s degrees of freedom using standard arguments. ■

Proof of Proposition 3. The result in this proposition is a special case of Proposition 2, which leads due to the form of the restrictions to a particularly simple form of the test statistic. In the augmented regression (10) 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. Condition (8) is clearly fulfilled (for all values of T). ■

Proof of Proposition 4. The main arguments of the proof are similar to the arguments used in the proof of Proposition 1. We start by defining the required additional scaling matrix $G_F(T) := \operatorname{diag}\left(G_{\bar{D}}(T), G_{\bar{X}}(T), G_Q(T)\right)$, with $G_{\bar{D}}(T) := \operatorname{diag}\left(T^{-\left(q+\frac{3}{2}\right)}, \ldots, T^{-\left(q+n+\frac{1}{2}\right)}\right)$, $G_{\bar{X}}(T) := \operatorname{diag}\left(G_{\bar{X}_1}(T), \ldots, G_{\bar{X}_m}(T)\right)$, $G_Q(T) := \operatorname{diag}\left(G_{Q_1}(T), \ldots, G_{Q_k}(T)\right)$ with $G_{\bar{X}_j}(T) := \operatorname{diag}\left(T^{-\frac{p_j+2}{2}}, \ldots, T^{-\frac{p_j+r_j+1}{2}}\right)$ and $G_{Q_i}(T) := \operatorname{diag}\left(T^{-1}, \ldots, T^{-\frac{s_j+1}{2}}\right)$. With these matrices defined, it immediately follows for $0 \leq r \leq 1$ that $T^{1/2}G_{\bar{D}}(T)\bar{D}_{[rT]} \Rightarrow \bar{D}(r)$. Analogously to Proposition 1 it also follows for $0 \leq r \leq 1$ that $T^{1/2}G_{\bar{X}_j}(T)\bar{X}_{j,[rT]} \Rightarrow \mathbf{B}^F_{v_j}(r)$, for $j = 1, \ldots, m$ and $T^{1/2}G_{Q_i}(T)Q_{i,[rT]} \Rightarrow \mathbf{B}^F_{v_i^*}(r)$ for $i = 1, \ldots, k$.

The OLS estimator of $\theta_{\tilde{F}}$ of (12) is given by $\hat{\theta}_{\tilde{F}} := (\tilde{F}'\tilde{F})^{-1}\tilde{F}'\hat{u}^+ = G_F(G_F\tilde{F}'\tilde{F}G_F)^{-1}G_F\tilde{F}'\hat{u}^+$. Using $\tilde{J}(r)$ as defined in (16) it holds that $\left(G_F\tilde{F}'\tilde{F}G_F\right)^{-1} \Rightarrow \left(\int_0^1 \tilde{J}^F(r)\tilde{J}^F(r)'dr\right)^{-1}$. The second term can be rewritten as:

$$
\begin{aligned}
G_F\tilde{F}'\hat{u}^+ &= G_F\tilde{F}'\left(u^+ - Z(\hat{\theta}^+ - \theta)\right) \\
&= G_F\tilde{F}'u^+ \\
&= G_F F'u^+ - G_F F'ZG(GZ'ZG)^{-1}GZ'u^+, \qquad \textbf{(A.5)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

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 $\tilde{F}$. Using again similar arguments as in Proposition 1 it can be shown that $G_F F' u^+ \Rightarrow \int_0^1 J^F(r) dB_{u\cdot v}(r) + A^F$, with $A^F := \left[0'_{n\times 1}, M^{\bar{X}\prime}, M^{Q\prime}\right]'$, where $M^{\bar{X}} := \left[M_1^{\bar{X}\prime}, \ldots, M_m^{\bar{X}\prime}\right]'$, and $M^Q := \left[M_1^{Q\prime}, \ldots, M_k^{Q\prime}\right]'$. The blocks within the latter vectors are given by $M_j^{\bar{X}} := \Delta_{v_j u}^+ \left[(p_j+1)\int_0^1 B_{v_j}^{p_j}(r)dr, \ldots, (p_j+r_j)\int_0^1 B_{v_j}^{p_j+r_j-1}(r)dr\right]'$, for $j = 1, \ldots, m$, and $M_i^Q := \Delta_{v_i^* u}^+ \left[1, 2\int_0^1 B_{v_i^*}(r)dr, \ldots, s_i \int_0^1 B_{v_i^*}^{s_i-1}(r)dr\right]'$, for $i = 1, \ldots, k$. For the second term in (A.5) we obtain $G_F F' ZG(GZ'ZG)^{-1} GZ' u^+ \Rightarrow \int_0^1 J^F(r)J(r)'dr \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \left(\int_0^1 J(r)dB_{u\cdot v}(r) + A\right)$, with A as given in Proposition 1. Putting things together we obtain, using $B_{u\cdot v}(r) = B_{u\cdot\tilde{v}}(r) + B_{\tilde{v}}(r)'\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u} - B_v(r)'\Omega_{vv}^{-1}\Omega_u$, that:

$$
\begin{aligned}
G_F^{-1}\hat{\theta}_{\tilde{F}} \Rightarrow & \left(\int_0^1 \tilde{J}^F(r)\tilde{J}^F(r)'dr\right)^{-1} \\
& \times \left(\int_0^1 \tilde{J}^F(r)dB_{u\cdot v}(r) + A^F - \int_0^1 J^F(r)J(r)'dr\left(\int_0^1 J(r)J(r)'dr\right)^{-1} A\right) \\
= & \left(\int_0^1 \tilde{J}^F(r)\tilde{J}^F(r)'dr\right)^{-1} \left(\int_0^1 \tilde{J}^F(r)dB_{u\cdot\tilde{v}}(r) + A^F + O^F\right. \\
& \left. - \int_0^1 J^F(r)J(r)'dr\left(\int_0^1 J(r)J(r)'dr\right)^{-1} A\right), \qquad \textbf{(A.6)}
\end{aligned}
$$

where $O^F := \int_0^1 \tilde{J}^F(r)dB_{\tilde{v}}(r)'\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u} - \int_0^1 \tilde{J}^F(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu}$. The correction factors O^{F*}, A^{F*}, and $k^{F*}A^*$, as given in the formulation of the proposition, then lead to a zero-mean Gaussian mixture limiting distribution for the FM-OLS estimator of $\theta_{\tilde{F}}$ as defined in (13), given by $G_F^{-1}\hat{\theta}_{\tilde{F}}^+ \Rightarrow \left(\int_0^1 \tilde{J}^F(r)\tilde{J}^F(r)'dr\right)^{-1} \int_0^1 \tilde{J}^F(r)dB_{u\cdot\tilde{v}}(r)$. The asymptotic chi-squared null distribution of the LM specification test now follows. ■

Proof of Proposition 5. The starting point is $\hat{u}_t^+ = u_t^+ - Z_t'(\hat{\theta}^+ - \theta)$. From Proposition 1 we already know that $T^{-1/2}\sum_{t=1}^{[rT]} u_t^+ \Rightarrow B_{u\cdot v}(r)$, and $T^{-1/2}\sum_{t=1}^{[rT]} Z_t' GG^{-1}(\hat{\theta}^+ - \theta) \Rightarrow \int_0^r J(s)'ds\left(\int_0^1 J(s)J(s)'ds\right)^{-1}\int_0^1 J(s)dB_{u\cdot v}(s)$ follows for the second term. Combining these terms then leads to:

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor} \hat{u}_t^+ \Rightarrow & B_{u\cdot v}(r) - \int_0^r J(s)'ds\left(\int_0^1 J(s)J(s)'ds\right)^{-1}\int_0^1 J(s)dB_{u\cdot v}(s) \\
= & \omega_{u\cdot v}^{1/2}\left(W_{u\cdot v}(r) - \int_0^r J(s)'ds\left(\int_0^1 J(s)J(s)'ds\right)^{-1}\int_0^1 J(s)dW_{u\cdot v}(s)\right). \qquad \textbf{(A.7)}
\end{aligned}
$$

This establishes the first result when a consistent estimator $\hat{\omega}_{u\cdot v}$ is used.

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press

Let us now consider the special case that only one integrated regressor appears with higher-order powers, without loss of generality the last one, x_{mt}. Denoting with $\omega_m^{1/2}$ the lower right hand corner element of $\mathbf{\Omega}_{vv}^{1/2}$ it follows that $\mathbf{B}_v(r)$ in this case simplifies to:

$$\mathbf{B}_v(r) = \begin{pmatrix} B_v(r) \\ B_{v_m}^2(r) \\ \vdots \\ B_{v_m}^{p_m}(r) \end{pmatrix} = \begin{bmatrix} \Omega_{vv}^{1/2} & 0 & \dots & 0 \\ 0 & \omega_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \omega_m^{p_m/2} \end{bmatrix} \begin{pmatrix} W_v(r) \\ W_{v_m}^2(r) \\ \vdots \\ W_{v_m}^{p_m}(r) \end{pmatrix} \tag{A.8}$$

Inserting the above expression for $\mathbf{B}_v(r)$ into the expression for $J(r)$ then leads to the result that:

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \hat{u}_t^+ &\Rightarrow B_{u\cdot v}(r) - \int_0^r J(s)' ds \left(\int_0^1 J(s) J(s)' ds \right)^{-1} \int_0^1 J(s) dB_{u\cdot v}(s) \\ &= \omega_{u\cdot v}^{1/2} \left(W_{u\cdot v}(r) - \int_0^r J^W(s)' ds \left(\int_0^1 J^W(s) J^W(s)' ds \right)^{-1} \right. \\ &\quad \left. \times \int_0^1 J^W(s) dW_{u\cdot v}(s) \right) \end{aligned} \tag{A.9}$$

and thus to the result (24) when a consistent estimator $\hat{\omega}_{u\cdot v}$ is used, with $J^W(s)$ as given in the formulation of the proposition. ■

Proof of Proposition 6. Let $0 \le r \le 1$ and $i \le t := \lfloor br \rfloor + i - 1 \le i + b - 1$. Similar to the proof of Proposition 5 a functional central limit theorem applies also for sub-samples of residuals and we obtain:

$$\begin{aligned} b^{-1/2} \sum_{j=i}^{t} \hat{u}_j^+ &= b^{-1/2} \sum_{j=i}^{t} u_j^+ \\ &\quad + \left(b^{-1/2} \sum_{j=i}^{t} Z_j' G(b) \right) \left(G(b)^{-1} G(T) \right) \left(G(T)^{-1} (\hat{\theta}^+ - \theta) \right). \end{aligned} \tag{A.10}$$

As in Proposition 5 one can show, since $b \to \infty$ and $\frac{b}{T} \to 0$, that $\lim_{T\to\infty} b^{-1/2} \sum_{j=i}^{t} u_j^+ = B_{u\cdot v}(r)$. The first and the third bracketed terms composing the product on the right hand side above, i.e., $\left(b^{-1/2} \sum_{j=i}^{t} Z_j' G(b) \right)$ 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 (A.10) is $O_p\left(\sqrt{\frac{b}{T}} \right)$. Therefore, since by assumption $\frac{b}{T} \to 0$, we have established that $b^{-1/2} \sum_{j=i}^{t} \hat{u}_j^+ \Rightarrow B_{u\cdot v}(r)$. The result then follows from the assumption of consistency of $\hat{\omega}_{u\cdot v}$. ■

https://doi.org/10.1017/S0266466615000213 Published online by Cambridge University Press