Supplementary Material:

Cointegrating Polynomial Regressions: Robustness of Fully Modified OLS

Oliver Stypka[1], Martin Wagner[2,3,4], Peter Grabarczyk[5] and Rafael Kawka[5]

[1]Flossbach von Storch, Cologne
[2]Department of Economics, University of Klagenfurt
[3]Bank of Slovenia, Ljubljana
[4]Institute for Advanced Studies, Vienna
[5]Faculty of Statistics, TU Dortmund University

APPENDIX B: The FM-CPR Estimator of Wagner and Hong (2016)

Wagner and Hong (2016) extend the fully modified OLS (FM-OLS) estimator of Phillips and Hansen (1990) from the cointegrating linear to the cointegrating polynomial regression (CPR) case. The discussion here focuses on the case considered in the main text in (1) and thereby defines the estimator referred to as FM-CPR in the main text. As also discussed in the main text, FM-type estimation entails two modifications. The modification of the dependent variable is exactly as proposed by Phillips and Hansen (1990) in the linear cointegration case, with the dependent variable y_t replaced by $y_t^+ := y_t - \Delta x_t \hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$. This transformation dynamically orthogonalizes the limit partial sum process of the modified errors:

$$u_t^+ := u_t - v_t \hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}, \tag{B.1}$$

i. e., $B_{u\cdot v}(r)$ as defined in Proposition 2.11 below (14) in the main text, from the limiting process corresponding to x_t, i. e., $B_v(r)$. In the case of Gaussian limits, uncorrelatedness is equivalent to independence. Thus, $B_{u\cdot v}(r)$ is "automatically" also independent of powers of $B_v(r)$ that appear in the asymptotic distributions in the CPR case. Consequently, the modification to orthogonalize regressors and errors need not be changed when considering FM-OLS estimation in the CPR setting rather than in the linear cointegration setting; orthogonalization with respect to $B_v(r)$ suffices.

The second modification, correcting for additive bias terms, depends upon the precise form of the model considered. For the specification given in (1) in the main text the FM-CPR bias correction term is given by $A^* := \hat{\Delta}_{vu}^+[0_{q\times 1}, T, 2\sum_{t=1}^T x_t, \ldots, p\sum_{t=1}^T x_t^{p-1}]'$, with $\hat{\Delta}_{vu}^+ := \hat{\Delta}_{vu} - \hat{\Delta}_{vv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$. Defining $y^+ := [y_1^+, \ldots, y_T^+]'$ and $Z := [Z_1, \ldots, Z_T]'$, leads to the FM-CPR estimator of θ given by:

$$\hat{\theta}^+ := (Z'Z)^{-1}(Z'y^+ - A^*). \tag{B.2}$$

Wagner and Hong (2016, Proposition 1) show, under slightly weaker assumptions than considered in the main text, that:

$$G^{-1}(\hat{\theta}^+ - \theta) \Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \int_0^1 J(r)dB_{u\cdot v}(r). \tag{B.3}$$

This is exactly the same asymptotic distribution as derived for the FM-OLS estimator in Proposition 2.11 in the main text.

Denoting the FM-CPR residuals with $u_t^+ := y_t - Z_t'\hat{\theta}^+$, Wagner and Hong (2016, Proposition 5) provides the limiting distribution of a Shin (1994)-type test of Wagner and Hong (2016, Proposition 5) test for the null hypothesis of no cointegration defined as:

$$\mathrm{CT}^+ := \frac{1}{T\hat{\omega}_{u\cdot v}} \sum_{t=1}^{T} \left(\frac{1}{T^{1/2}} \sum_{j=1}^{t} \hat{u}_j^+ \right)^2 . \tag{B.4}$$

Under the null hypothesis it holds that:

$$\mathrm{CT}^+ \Rightarrow \int_0^1 \left(W_{u\cdot v}^{J_W}(r) \right)^2 dr, \tag{B.5}$$

with:

$$W_{u\cdot v}^{J_W}(r) := W_{u\cdot v}(r) - \int_0^r J^W(s)'ds \left(\int_0^1 J^W(s) J^W(s)'ds \right)^{-1} \int_0^1 J^W(s) dW_{u\cdot v}(s), \tag{B.6}$$

where $J^W(r) := [D(r)', W_v(r), W_v^2(r), \ldots, W_v^p(r)]'$ and $W_{u\cdot v}(r)$ and $W_v(r)$ are two standard Brownian motions independent of each other.

The limiting distribution given in (B.5) and (B.6) is nuisance parameter free since the single integrated regressor case is, in the words of Vogelsang and Wagner (2014), of *full design*, which allows for a bijection between functionals of Brownian motions and standard Brownian motions and therefore a limiting distribution that can be simulated.

In the multiple integrated regressor CPR case, full design need not necessarily prevail. In this case the result for both CT^+ and CT^{++} still holds true, however, with a nuisance parameter dependent limiting distribution given in Wagner and Hong (2016, eq. (22) and (23)). For this case Wagner and Hong (2016, Proposition 6) propose a sub-sampling approach to achieve a nuisance parameter free limiting distribution.

APPENDIX C: A Discussion of the Robustness Result

Proposition 2.11 in the main text is based on the underlying kernel-weighted sum limit result developed in Theorem 2.7 and the functional central limit theorem developed in

Theorem 2.9, both given in the main text. Some of the results, in particular the structure of the kernel-weighted sum limits, have a similar structure as arising when considering (the scaled partial sum) limit processes of the transformed errors u_t^+, related to the orthogonalization of normally distributed random variables or Brownian motions. This subsection discusses these links and thereby highlights the importance of including x_t itself as a regressor for the robustness result of "formal" FM-OLS, i. e., of the asymptotic equivalence of "formal" FM-OLS and FM-CPR.[1]

To set the stage, consider the "proper" FM-type modification of the dependent variable, i. e., the Phillips and Hansen (1990) or Wagner and Hong (2016) modification of the dependent variable and its centered version already defined in (B.1) in Appendix B:

$$\begin{aligned} u_t^+ &= u_t - \Delta x_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \\ &= u_t - v_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}. \end{aligned} \tag{C.1}$$

When using consistent long-run covariance estimators, the scaled partial sum process of the transformed errors u_t^+ converges to a Brownian motion independent of the scaled partial sum corresponding to $\{v_t\}_{t \in \mathbb{Z}}$, i. e., $\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} u_t^+ = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} u_t - \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} v_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}$ $\Rightarrow B_u(r) - B_v(r) \Omega_{vv}^{-1} \Omega_{vu} = B_{u \cdot v}(r)$. The above result defines $B_{u \cdot v}(r)$ to be by construction independent of $B_v(r)$, however, the result can also be interpreted (or arrived at) from a different (well-known) perspective. Consider the following *population regression* equation for a fixed $0 < r \leq 1$:

$$B_u(r) = B_v(r) \Theta_{[1]}(r) + B_{u \cdot v}(r), \tag{C.2}$$

The population regression coefficient $\Theta_{[1]}(r)$ is given by:

$$\begin{aligned} \Theta_{[1]}(r) &:= \left(\mathbb{E}(B_v(r) B_v(r)) \right)^{-1} \mathbb{E}(B_v(r) B_u(r)) \\ &= (r \Omega_{vv})^{-1} (r \Omega_{vu}) = \Omega_{vv}^{-1} \Omega_{vu} = \Theta_{[1]}, \end{aligned} \tag{C.3}$$

where the notation $\Theta_{[1]}(r)$ indicates that in the regression only the first power of $B_v(r)$ is included.[2]

[1] Note that the analogies only apply to the orthogonalization of the dependent variable steps of the two considered estimators and do not extend to the additive bias terms and their removal. The analogy also does not "explain" Theorems 2.7 and 2.9.

[2] Of course, the result can also be obtained by considering not only a fixed value $0 < r \leq 1$, but the interval $[0, 1]$. In this case the starting point is again (C.2): Pre-multiplying this equation by $B_u(r)$, integrating from 0 to 1 and taking expectations leads – using independence of $B_v(r)$ and $B_{u \cdot v}(r)$ – to $\Theta_{[1]} = \left(\mathbb{E} \left(\int_0^1 B_v^2(s) ds \right) \right)^{-1} \mathbb{E} \left(\int_0^1 B_v(s) B_u(s) ds \right) = \left(\frac{1}{2} \Omega_{vv} \right)^{-1} \left(\frac{1}{2} \Omega_{vu} \right) = \Omega_{vv}^{-1} \Omega_{vu}$. This result also

Consider next in a similar fashion – which can again be done analogously either for fixed $0 < r \leq 1$ or the interval $[0,1]$ – the population regression equation relating $B_u(r)$ and $B_v(r), B_v^2(r), \ldots, B_v^p(r)$:

$$\begin{aligned} B_u(r) &= \left[B_v(r), B_v^2(r), \ldots, B_v^p(r)\right] \Theta_{[1:p]} + B_{u\cdot v}(r) \\ &= \left[B_v(r), B_v^2(r), \ldots, B_v^p(r)\right] \begin{bmatrix} \Theta_{[1]} \\ 0_{(p-1)\times 1} \end{bmatrix} + B_{u\cdot v}(r), \end{aligned} \tag{C.4}$$

with obvious notation for $\Theta_{[1:p]}$. The second equality follows from Gaussianity, which implies that $B_{u\cdot v}(r)$ is not only independent from $B_v(r)$, but also from all powers of $B_v(r)$. For an early precise discussion of the fact that for two Gaussian processes all that needs to be achieved by an appropriate transformation to arrive at independence is zero correlation see, e. g., Phillips (1989, Lemma 3.1). The second equality can also be established by calculating $\Theta_{[1:p]}$ explicitly, analogously to (C.3):

$$\begin{aligned} \Theta_{[1:p]} &:= \left(\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)')\right)^{-1} \mathbb{E}(\mathbf{B}_v(r)B_u(r)) \\ &= \left(\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)')\right)^{-1} \mathbb{E}(\mathbf{B}_v(r)(B_{u\cdot v}(r) + B_v(r)\Omega_{vv}^{-1}\Omega_{vu})) \\ &= \left(\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)')\right)^{-1} \mathbb{E}(\mathbf{B}_v(r)B_v(r))\Omega_{vv}^{-1}\Omega_{vu} \\ &= \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p-1)\times 1} \end{bmatrix} = \begin{bmatrix} \Theta_{[1]} \\ 0_{(p-1)\times 1} \end{bmatrix}, \end{aligned} \tag{C.5}$$

using independence of $\mathbf{B}_v(r)$ and $B_{u\cdot v}(r)$ and "partial inversion".[3] Note that for partial inversion to apply it is necessary that $\mathbf{B}_v(r)$ includes $B_v(r)$ as, e. g., first element.

explains the final equality in (C.3), i. e., $\Theta_{[1]}(r) = \Theta_{[1]}$. Note for completeness that the population regression coefficient $\Theta_{[1]}$ cannot be interpreted as regression coefficient obtained from the continuous time regression of $B_u(r)$ on $B_v(r)$ over the interval $[0,1]$. The corresponding continuous time least squares regression coefficient is given by $\hat{\Theta}_{[1]} = \left(\int_0^1 B_v(s)^2 ds\right)^{-1} \int_0^1 B_v(s)B_u(s)ds$, resulting in a "residual process" of the form $\hat{B}_{u\cdot v}(r) = B_u(r) - B_v(r)\left(\int_0^1 B_v(s)^2 ds\right)^{-1} \int_0^1 B_v(s)B_u(s)ds$.

[3] "Partial inversion" is our label for the obvious fact that for any regular matrix $A \in \mathbb{R}^{n\times n}$ with columns $A_{(\cdot,j)}$ it holds that $A^{-1}A_{(\cdot,j)} = [0,\ldots,0,1,0,\ldots,0]'$, with the 1 occurring as j-th entry. Exactly the same result as shown in (C.5) again also holds for the interval $[0,1]$ version, similarly as discussed for $\Theta_{[1]}$ in Footnote 2.

Now consider the "formal" FM-OLS transformed errors from a similar perspective as u_t^+ in (C.1) above, i. e., consider:

$$
\begin{aligned}
u_t^{++} &= u_t - \Delta X_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \\
&= u_t - w_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \\
&= u_t - \left[v_t, \Delta x_t^2, \ldots, \Delta x_t^p\right]\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \\
&\approx u_t - \left[v_t, \frac{2x_t v_t - v_t^2}{T^{1/2}}, \ldots, \frac{p x_t^{p-1} v_t - \frac{p(p-1)}{2} x_t^{p-2} v_t^2}{T^{\frac{p-1}{2}}}\right]\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u},
\end{aligned} \tag{C.6}
$$

with – and this is a key aspect – v_t included only if x_t is included in the regression and where for Δx_t^j, $j = 2, \ldots, p$ only the two (asymptotically relevant) leading terms are considered, compare (6) in the main text. Convergence of the corresponding scaled partial sum process, $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_t^{++}$, $0 \leq r \leq 1$, follows from a combination of several results, e. g., Liang *et al.* (2016, Theorem 2.1) or Wagner and Hong (2016, Proposition 1), Itô's Lemma (see, e. g., Theorem 3.3., p. 149 in Karatzas and Shreve, 1991) and Theorem 2.7 in the main text, with the limit process given by:[4]

$$
\begin{aligned}
B_{u\cdot v}^{++}(r) := B_u(r) - \Bigg[&B_v(r), 2\int_0^r B_v(s)dB_v(s) + r\Omega_{vv}, \ldots, \\
&p\int_0^r B_v^{p-1}(s)dB_v(s) + \frac{p(p-1)}{2}\Omega_{vv}\int_0^r B_v^{p-2}(s)ds\Bigg]\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u} \\
= B_u(r) - &\left[B_v(r), B_v^2(r), \ldots, B_v^p(r)\right]\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u}.
\end{aligned} \tag{C.7}
$$

Theorem 2.7 shows that $\hat{\Omega}_{\tilde{w}\tilde{w}} \Rightarrow \Omega_{\tilde{w}\tilde{w}} = \Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr$ and that $\hat{\Omega}_{\tilde{w}u} \Rightarrow \Omega_{\tilde{w}u} = \Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr$. In the case that x_t is included (as first element) in X_t, the first element of $\dot{\mathbf{B}}_v(r) = 1$ and it follows again via partial inversion that:

$$
\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u} = \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p-1)\times 1} \end{bmatrix} = \Theta_{[1:p]}, \tag{C.8}
$$

which in turn immediately implies that $B_{u\cdot v}^{++}(r) = B_{u\cdot v}(r)$. Therefore, the robustness result algebraically critically hinges on the finding that the product $\Omega_{\tilde{w}\tilde{w}}^{-1}\tilde{\Omega}_{wu}$ coincides with

[4] We use (C.6) as starting point as it highlights the relevant quantities for the asymptotic results. If one is merely interested in the partial sum process and its limit it is easier to directly consider:

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_t^{++} = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_t - \frac{1}{\sqrt{T}}X_{\lfloor rT \rfloor}' G_W \hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \Rightarrow B_u(r) - \mathbf{B}_v(r)'\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u},
$$

with this approach also relying upon Theorem 2.7 in the main text.

$\Theta_{[1:p]}$, which hinges upon the first element of $\dot{\mathbf{B}}_v(r) = 1$, from which partial inversion leads to the result. In this case, asymptotically, the transformations of the dependent variable y_t, respectively of the errors u_t, undertaken in both FM-OLS and FM-CPR coincide. However, in finite samples the FM-OLS transformation to y_t^{++}, respectively u_t^{++}, involves the subtraction of a number of terms that are asymptotically all zero. These extraneous subtractions are not performed in the FM-CPR transformations to y_t^{+}, respectively u_t^{+}. Loosely speaking, thus FM-OLS is affected by asymptotically vanishing nuisance parameters in finite samples. To conclude the discussion, note that the asymptotic equivalence of the additive bias correction terms subtracted in FM-OLS and FM-CPR follows along similar lines, as detailed in the proof of Proposition 2.11 in the main text.

The above discussion makes clear that the robustness result for "formal" FM-OLS breaks down when x_t itself is not included in the regression. Nevertheless, it may be informative to see this also explicitly. Consider the simplest example $y_t = x_t^2\beta + u_t$, $x_t = x_{t-1} + v_t$. In this case straightforward (given the results of the paper) derivations show that the FM-OLS estimator does not converge to – the corresponding case of – the limiting distribution given in (14) in the main text, but to:[5]

$$
\begin{aligned}
T^{3/2}(\hat{\beta}^{++} - \beta) \Rightarrow \left(\int_0^1 B_v^4(r)dr\right)^{-1} \Bigg(&\int_0^1 B_v^2(r)dB_{u\cdot v}(r) \\
&+ \int_0^1 B_v(r)dr\Omega_{vv}^{-1}\Omega_{vu}\left[\int_0^1 B_v^2(r)dB_v(r)\left(\int_0^1 B_v(r)dr\right)^{-1}\right. \\
&\left.- \int_0^1 B_v^3(r)dB_v(r)\left(\int_0^1 B_v^2(r)dr\right)^{-1} - \frac{\Omega_{vv}}{2}\right]\Bigg).
\end{aligned} \tag{C.9}
$$

The corresponding FM-CPR limit distribution coincides with the expression in the first line of (C.9). The terms in the second and third line of (C.9) comprise the "orthogonalization" error that occurs when $B_u(r)$ is orthogonalized with respect to the non-Gaussian process $B_v^2(r)$ rather than the Gaussian process $B_v(r)$. This step does not lead to independence between the limit partial sum process of u_t^{++}, given here by $B_{u\cdot v}^{++}(r) = B_u(r) - \frac{1}{2}\Omega_{vv}^{-1}\Omega_{vu}\int_0^1 B_v(r)dr$, and $B_v(r)$, hence the extra terms.

[5]The relevant terms for the specific example considered are $GZ'\tilde{W} \Rightarrow 2\int_0^1 B_v^3(r)dB_v(r) + 6\Delta_{vv}\int_0^1 B_v^2(r)dr - \Sigma_{vv}\int_0^1 B_v^2(r)dr$, $\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \Rightarrow \frac{1}{2}\Omega_{vv}^{-1}\Omega_{vu}\left(\int_0^1 B_v^2(r)dr\right)^{-1}\int_0^1 B_v(r)dr$ and $GA^{**} \Rightarrow 2\Delta_{vu}^{+}\int_0^1 B_v(r)dr$.

APPENDIX D: Proofs of Auxiliary Lemmas

Proof of Lemma A.2. Consider $f(x) := x^q$, $x \in \mathbb{R}$. The mean value theorem states that $f(y) - f(x) = f'(\zeta)(y - x)$, i. e., $y^q - x^q = q\zeta^{q-1}(y - x)$, with $x < y$ and $x < \zeta < y$. Therefore, it holds that:

$$\left(\frac{x_{t+h}}{T^{1/2}}\right)^q - \left(\frac{x_t}{T^{1/2}}\right)^q = q\left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \frac{x_{t+h} - x_t}{T^{1/2}} = \frac{q}{T^{1/2}}\left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \sum_{m=1}^{h} v_{t+m},$$

with $\overline{x}_t^h = x_t + \gamma_t \sum_{m=1}^{h} v_{t+m}$ and some $0 < \gamma_t < 1$. Using this representation, it follows that:

$$\begin{aligned}
&\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p \left[\left(\frac{x_{t+h}}{T^{1/2}}\right)^q - \left(\frac{x_t}{T^{1/2}}\right)^q\right] v_t v_{t+h} \\
&\quad = \frac{q}{T^{1/2}}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p \left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \sum_{m=1}^{h} v_t v_{t+m} v_{t+h}.
\end{aligned}$$

The assertion is hence equivalent to showing that:

$$\frac{1}{T^{1/2}}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p \left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \sum_{m=1}^{h} v_t v_{t+m} v_{t+h} = o_{\mathbb{P}}(1).$$

In the course of the proof, it is helpful to resort to strong approximations, obtained from the Skorohod representation theorem, see Pollard (1984, p. 71–72) or Csörgo and Horváth (1993, p. 4). For a discussion of this issue in a nonlinear cointegration context see, e. g., Park and Phillips (1999, Lemma 2.3) and Park and Phillips (2001). Since we are concerned with weak convergence results in this paper, we can w.l.o.g. use a distributionally equivalent version of $T^{-1/2}x_{\lfloor rT \rfloor}$, $X_T^*(r)$ say, that fulfills $\sup_{r\in[0,1]} |X_T^*(r) - B_v(r)| = o_{a.s.}(1)$, with $B_v(r)$ the Brownian motion given in (4) in the main text. For convenience, we continue to use x_t and $T^{-1/2}x_{\lfloor rT \rfloor}$ also when working with the distributionally equivalent version. Setting $\tilde{C} := \sup_{r\in[0,1]} |B_v(r)| + 1/2$, it holds that:

$$\sup_{r\in[0,1]} T^{-1/2}|x_{\lfloor rT \rfloor}| \leq \tilde{C} + o_{a.s.}(1). \tag{D.1}$$

Furthermore, $\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|\sum_{m=1}^{h} v_{\lfloor rT\rfloor+m}| \leq \sup_{r\in[0,1]} T^{-1/2}\sum_{m=1}^{M_T}|v_{\lfloor rT\rfloor+m}|$ and, thus, it follows from Lemma A.1 that:

$$\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT\rfloor+h}-x_{\lfloor rT\rfloor}| = o_{a.s.}(1). \tag{D.2}$$

Via the triangular inequality, this implies $\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT\rfloor+h}|| \leq C + o_{a.s.}(1)$, with $C := \sup_{r\in[0,1]}|B_v(r)|+1$ and it also implies that:

$$\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|\overline{x}^h_{\lfloor rT\rfloor}| \leq C + o_{a.s.}(1). \tag{D.3}$$

Using the triangular inequality and the bounds given in (D.1)–(D.3), the following inequalities hold:

$$\begin{aligned}
&\left|\frac{1}{T^{1/2}}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\left(\frac{\overline{x}^h_t}{T^{1/2}}\right)^{q-1}\sum_{m=1}^{h} v_t v_{t+m} v_{t+h}\right| \qquad \text{(D.4)}\\
&\leq \left(\frac{M_T^3}{T}\right)^{1/2}\frac{1}{M_T}\sum_{h=0}^{M_T}\left|k\left(\frac{h}{M_T}\right)\right|\frac{1}{T}\sum_{t=1}^{T-h}\left|\left(\frac{x_t}{T^{1/2}}\right)^p\left(\frac{\overline{x}^h_t}{T^{1/2}}\right)^{q-1}\right||v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{h} v_{t+m}\right|\\
&\leq \left(\frac{M_T^3}{T}\right)^{1/2}\overline{k}(0)C^{p+q-1}\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{h} v_{t+m}\right| + o_{\mathbb{P}}(1),
\end{aligned}$$

with $\overline{k}(0) = \sup_{x\geq 0}|k(x)|$ as defined in Assumption 2.3 in the main text. Consider next:

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{h} v_{t+m}\right|\right)\\
&= \frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}\mathbb{E}\left(|v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{h} v_{t+m}\right|\right)\\
&\leq \frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}\mathbb{E}(v_t^4)^{1/4}\mathbb{E}(v_{t+h}^4)^{1/4}\mathbb{E}\left[\left(M_T^{-1/2}\sum_{m=1}^{h} v_{t+m}\right)^2\right]^{1/2}.
\end{aligned}$$

It holds that:

$$\mathbb{E}\left[\left(M_T^{-1/2}\sum_{m=1}^{h} v_{t+m}\right)^2\right] = \frac{1}{M_T}\sum_{m_1=1}^{h}\sum_{m_2=1}^{h}\mathbb{E}(v_{t+m_1}v_{t+m_2})$$
$$= \frac{1}{M_T}\sum_{m=1}^{h}\mathbb{E}(v_{t+m}^2) + \frac{2}{M_T}\sum_{m_1=1}^{h}\sum_{m_2=m_1+1}^{h}\mathbb{E}(v_{t+m_1}v_{t+m_2})$$
$$\leq \frac{h}{M_T}\,\Omega_{vv},$$

which implies $\mathbb{E}\left(\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{h} v_{t+m}\right|\right) = O(1)$. The assertion is now an immediate consequence of $M_T^3/T \to 0$ by Assumption 2.4 in the main text and the remaining terms contained in the expression in (D.4) being $O_{\mathbb{P}}(1)$. ■

Proof of Lemma A.3. The proof of Lemma A1(iv) in Kasparis (2008) shows that:

$$\left|\frac{1}{M_T}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\sum_{m=1}^{h}\left(v_t v_{t+m} - \mathbb{E}[v_t v_{t+m}]\right)\right| = o_{\mathbb{P}}(1)$$

by showing that:

$$\sup_{0\leq h\leq M_T}\left|\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\sum_{m=1}^{h}\left(v_t v_{t+m} - \mathbb{E}[v_t v_{t+m}]\right)\right| = o_{\mathbb{P}}(1). \tag{D.5}$$

The left-hand side of (A.3) can be written as:

$$\left|\frac{1}{M_T}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p M_T\left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right)\right|.$$

Using similar arguments as Kasparis (2008, p. 1394–1396) to show (D.5), corresponding to his Equation (A.7), it follows that:

$$\sup_{0\leq h\leq M_T}\left|\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p M_T\left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right)\right| = o_{\mathbb{P}}(1),$$

which implies the claim of this lemma. Since we rely upon arguments of Kasparis (2008), we need the same moment and bandwidth assumptions that are, therefore, incorporated in our Assumptions 2.2 to 2.4 in the main text. ■

APPENDIX E: Multiple Integrated Regressors and Their Powers

This appendix sketches how the underlying main results, i. e., Theorems 2.7 and 2.9 in the main text, can be extended to thereby also extend the robustness result, i. e., Proposition 2.11 in the main text to the case of multiple integrated regressor CPRs as considered in Wagner and Hong (2016). The results hold – with full details available upon request – under exactly the same assumptions as in the main text, with Assumption 2.2 modified to a multivariate version, compare also Wagner and Hong (2016, Assumption 1). Notation is as in the main text with the additionally defined Khatri-Rao product: Let $\otimes$ denote the Kronecker product, let the matrix $\mathbf{A} = (A_{ij})$ be partitioned with A_{ij} denoting its (i,j)-th block sub-matrix and let $\mathbf{B} = (B_{ij})$ be analogously partitioned in block sub-matrices B_{ij}, for some $i, j = 1, 2, \ldots$. Then the Khatri-Rao product is defined and denoted as $\mathbf{A} \odot \mathbf{B} := (A_{ij} \otimes B_{ij})_{ij}$.

The setting is given by:

$$\begin{aligned} y_t &= D_t'\delta + x_t'\beta + \sum_{j=1}^{m} X_{jt}'\beta_{X_j} + u_t, \quad \text{for } t = 1, \ldots, T, \qquad \text{(E.1)} \\ &= D_t'\delta + X_t'\beta_X + u_t \\ &= Z_t'\theta + u_t \\ x_t &= x_{t-1} + v_t, \end{aligned}$$

where y_t is a scalar process, $D_t \in \mathbb{R}^q$, $x_t := [x_{1t}, \ldots, x_{mt}]'$, $X_{jt} := [x_{jt}^2, \ldots, x_{jt}^{p_j}]'$, $X_t := [x_t', X_{1t}', \ldots, X_{mt}']'$, $Z_t := [D_t', X_t']' \in \mathbb{R}^{q+p^*}$, $\beta_X := [\beta', \beta_{X_1}', \ldots, \beta_{X_m}']'$ and $\theta := [\delta', \beta_X']' \in \mathbb{R}^{q+p^*}$ with $p^* := \sum_{j=1}^{m} p_j$. Up to a different ordering of the regressors in X_t – grouping all first powers of the integrated regressors on top of the vector – this setting is similar to Wagner and Hong (2016, eq. (1), p. 1292). This reordering is done to, e. g., collect the components of $\hat{\Delta}_{\tilde{w}\tilde{w}}$ that correspond to long-run covariance estimation proper in the upper left blocks (which consequently leads to a similar structure as in the single integrated regressor case as considered in the main text). Note also that we only consider, as we focus here on highlighting the necessary steps to establish asymptotic equivalence of FM-OLS and FM-CPR, the case where all m-elements of x_t are included as regressors. There are no restrictions on the presence or absence of higher order powers of any of the integrated regressors, compare again Wagner and Hong (2016), but for notational brevity we discuss the case where the full sets of powers of the integrated regressors up to degrees $p_j,\ j=1,\ldots,m$ are included.

The limiting distribution of the FM-CPR estimator of the above equation (E.1) follows – with the reordering already taken into account below in (E.2) – from the corresponding result given in Wagner and Hong (2016, eq. (6), p. 1296), i. e.,:

$$G^{-1}(\hat{\theta}^{+}-\theta) \Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1}\int_0^1 J(r)dB_{u\cdot v}(r), \tag{E.2}$$

with $G := \text{diag}\left(G_D, T^{-1}I_m, \text{diag}\left(T^{-3/2},\ldots,T^{-(p_1+1)/2}\right),\ldots,\text{diag}\left(T^{-3/2},\ldots,T^{-(p_m+1)/2}\right)\right)$, $J(r) := [D(r)', B_v(r)', \mathbf{B}_v^*(r)']'$, with $B_v(r) := [B_{v_1}(r),\ldots,B_{v_m}(r)]'$, $\mathbf{B}_v^*(r) := [\mathbf{B}_{v_1}^*(r)',\ldots,\mathbf{B}_{v_m}^*(r)']' = [B_{v_1}^2(r),\ldots,B_{v_1}^{p_1}(r),B_{v_2}^2(r),\ldots,B_{v_m}^{p_m}(r)]'$ and $B_{u\cdot v}(r) := B_u(r)-B_v(r)'\Omega_{vv}^{-1}\Omega_{vu}$, with $B_v(r)$ now m-dimensional. In the considered setting, with the described up-front collection of all first powers, the multiple integrated regressor version of $w_t := \Delta X_t$ is given by $w_t := \left[v_{1t},\ldots,v_{mt},\Delta x_{1t}^2,\ldots,\Delta x_{1t}^{p_1},\ldots,\Delta x_{mt}^2,\ldots,\Delta x_{mt}^{p_m}\right]'$, with the corresponding scaling matrix G_W to arrive at $\tilde{w}_t := G_W w_t$ now given by $G_W := \text{diag}\left(I_m, \text{diag}\left(T^{-1/2},\ldots,T^{-(p_1-1)/2}\right),\ldots,\text{diag}\left(T^{-1/2},\ldots,T^{-(p_m-1)/2}\right)\right)$. Using similar arguments as in the proofs given in the appendices of the main text, Theorem 2.7 can be generalized to the multiple integrated regressor case. The key difference is the occurrence of cross-products of (first differences of) powers of different integrated regressors. This necessitates to establish the corresponding limits, using similar arguments as in the proof of Theorem 2.7 in the main text, to such cross-products. This is tedious but entails no fundamental additional complexities. Using similar notation as in the main text, i. e., $\{\xi_t\}_{t\in\mathbb{Z}} := \{[u_t, v_t']'\}_{t\in\mathbb{Z}}$ and $\hat{\eta}_t := [\hat{u}_t, \tilde{w}_t']'$ it can be shown that:

$$\hat{\Delta}_{\eta\eta} \Rightarrow \begin{bmatrix} \Delta_{\xi\xi} & \Delta_{\xi v}\odot\int_0^1 \dot{\mathbf{B}}_v^*(r)'dr \\ \Delta_{v\xi}\odot\int_0^1 \dot{\mathbf{B}}_v^*(r)dr & \Delta_{vv}\odot\int_0^1 \dot{\mathbf{B}}_v^*(r)\dot{\mathbf{B}}_v^*(r)'dr \end{bmatrix}, \tag{E.3}$$

with $\dot{\mathbf{B}}_v^*(r) := [\dot{\mathbf{B}}_{v_1}^*(r)',\ldots,\dot{\mathbf{B}}_{v_m}^*(r)']' = [2B_{v_1}(r),\ldots,p_1B_{v_1}^{p_1-1}(r),2B_{v_2}(r),\ldots,p_mB_{v_m}^{p_m-1}(r)]'$, where the Khatri-Rao product is used for a partitioning according to the rows $\Delta_{v_i\xi}$, the columns $\Delta_{\xi v_i}$, the (scalar) elements of Δ_{vv} and the blocks $\dot{\mathbf{B}}_{v_i}^*(r)$, $i=1,\ldots,m$.

As in the single integrated regressor case, $\hat{\Sigma}_{\eta\eta} := \frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t' \Rightarrow \Sigma_{\eta\eta}$, with $\Sigma_{\eta\eta}$ of similar structure as $\Delta_{\eta\eta}$ given above in (E.3). Combining the two results implies that $\hat{\Omega}_{\eta\eta} \Rightarrow \Omega_{\eta\eta}$; of similar structure as $\Delta_{\eta\eta}$ in (E.3).

Parallelling the structure of the proof in Appendix A in the main text one important step is to derive a block-column version of (A.17), which follows from:

$$\Omega_{\tilde{w}\tilde{w}} = \begin{bmatrix} \Omega_{vv} & \Omega_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)'dr \\ \Omega_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr & \Omega_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)\dot{\mathbf{B}}_v^*(r)'dr \end{bmatrix}, \tag{E.4}$$

$$\Omega_{\tilde{w}u} = \begin{bmatrix} \Omega_{vu} \\ \Omega_{vu} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix},$$

and from using again partial inversion in conjunction with the properties of the Khatri-Rao product:[6]

$$\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \xrightarrow{\mathbb{P}} \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p^*-m)\times 1} \end{bmatrix}. \tag{E.5}$$

The extension of Theorem 2.9 in the main text to the multiple integrated regressor case works using similar arguments as the derivation of Theorem 2.9 in Appendix A in the main text, by deriving the limits for cross-products of powers of x_{it} times the first differences of powers of x_{jt} for $i, j = 1, \ldots, m$ and $i \neq j$, with details available upon request:

$$GZ'\tilde{W} \Rightarrow \int_0^1 J(r)d\mathbf{B}_v^*(r)' + \begin{bmatrix} 0_{q\times m} & 0_{q\times(p^*-m)} \\ \Delta_{vv} & \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)'dr \\ \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr & \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)\dot{\mathbf{B}}_v^*(r)'dr \end{bmatrix}, \tag{E.6}$$

which implies the multiple integrated regressors version of (A.18):

$$GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \Rightarrow \int_0^1 J(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu} + \begin{bmatrix} 0_{q\times 1} \\ \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu} \\ \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix}, \tag{E.7}$$

with partitioning according to the rows $\Delta_{v_i v}\Omega_{vv}^{-1}\Omega_{vu}$, $i = 1, \ldots, m$. It now only remains to consider the structure of (the limit of):

$$GA^{**} = \begin{bmatrix} 0_{q\times 1} \\ \hat{\Delta}_{\tilde{w}u}^{+} \end{bmatrix} = \begin{bmatrix} 0_{q\times 1} \\ \hat{\Delta}_{\tilde{w}u} - \hat{\Delta}_{\tilde{w}\tilde{w}}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \end{bmatrix} \Rightarrow \begin{bmatrix} 0_{q\times 1} \\ \Delta_{\tilde{w}u} - \Delta_{\tilde{w}\tilde{w}}\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u} \end{bmatrix}. \tag{E.8}$$

[6] With the Khatri-Rao product being effectively an efficient notation for using the Kronecker product for matrices with block structure, the result stems from the properties of the inverse of a Kronecker product and the product rule for Kronecker products of matrices, in the case all elements of x_t are included in the CPR model.

(E.3) implies:

$$\Delta_{\tilde{w}\tilde{w}} = \begin{bmatrix} \Delta_{vv} & \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)'dr \\ \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr & \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)\dot{\mathbf{B}}_v^*(r)'dr \end{bmatrix}, \tag{E.9}$$

$$\Delta_{\tilde{w}u} = \begin{bmatrix} \Delta_{vu} \\ \Delta_{vu} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix}.$$

Combining (E.9) with (E.5) yields:

$$\begin{aligned} \Delta_{\tilde{w}u}^{+} &= \begin{bmatrix} \Delta_{vu} \\ \Delta_{vu} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix} - \begin{bmatrix} \Delta_{vv} & \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)'dr \\ \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr & \Delta_{vv} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)\dot{\mathbf{B}}_v^*(r)'dr \end{bmatrix} \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p^*-m)\times 1} \end{bmatrix} \\ &= \begin{bmatrix} \Delta_{vu} \\ \Delta_{vu} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix} - \begin{bmatrix} \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu} \\ \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix} \\ &= \begin{bmatrix} \Delta_{vu}^{+} \\ \Delta_{vu}^{+} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix}. \end{aligned}$$

Combining the above results allows to establish the limit of of $GZ'u^{++} - GA^{**}$:

$$\begin{aligned} &GZ'u - GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} - GA^{**} \qquad \text{(E.10)} \\ &\Rightarrow \int_0^1 J(r)dB_u(r) + \begin{bmatrix} 0_{q\times 1} \\ \Delta_{vu} \\ \Delta_{vu} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix} \\ &\quad - \int_0^1 J(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu} - \begin{bmatrix} 0_{q\times 1} \\ \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu} \\ \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix} - \begin{bmatrix} 0_{q\times 1} \\ \Delta_{vu}^{+} \\ \Delta_{vu}^{+} \odot \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix} \\ &= \int_0^1 J(r)dB_{u\cdot v}(r), \end{aligned}$$

with the last line following from the definition of $B_{u\cdot v}(r)$.

APPENDIX F: Finite Sample Performance

The most voluminous empirical literature using cointegrating polynomial regression (CPR) models is the environmental Kuznets curve (EKC) literature, typically relying upon quadratic or cubic CPR specifications. As discussed in the introduction of the main text, this growing literature continues to use, when applying cointegration techniques, to a surprisingly large extent estimators designed for linear cointegrating relationships, i. e., this empirical literature ignores the impacts that polynomial transformations of integrated processes entail for estimation and inference (see, e. g., from a large list of contributions Baek, 2015; Esteve and Tamarit, 2012; Fosten *et al.*, 2012; Friedl and Getzner, 2003; Galeotti *et al.*, 2006; He and Richard, 2010; Jalil and Mahmud, 2009; Lindmark, 2002; Özcan and Öztürk, 2019). This happens despite the fact that some papers such as, e. g., Bradford *et al.* (2005), Müller-Fürstenberger and Wagner (2007), Wagner (2008) or Wagner (2015), have discussed and highlighted the implied issues and complications for a long time.

Given that the paper shows asymptotic equivalence of "formal" FM-OLS to FM-CPR, with the need to be careful with the choice of correct values in cointegration testing, this section assesses the finite sample "price" that a user of "formal" FM-OLS has to pay compared to using, e. g., FM-CPR. This is done for a prototypical EKC-type specification inspired by Wagner (2015) used also for the finite sample performance assessment in Wagner and Hong (2016, Section 3). More specifically, the simulations use data generated from the following quadratic cointegrating polynomial regression model $y_t = c + \delta t + \beta_1 x_t + \beta_2 x_t^2 + u_t$, with the errors u_t and $v_t = \Delta x_t$ generated as $u_t = \rho_1 u_{t-1} + \varepsilon_t + \rho_2 e_t, \quad u_0 = 0$ and $v_t = e_t + 0.5 e_{t-1}$, where $(\varepsilon_t, e_t)' \sim \mathcal{N}(0, I_2)$. The parameter ρ_1 controls the level of serial correlation in the errors u_t and ρ_2 controls the extent of regressor endogeneity. The parameter values are set to $c = \delta = 1$, $\beta_1 = 5$ and $\beta_2 = -0.3$. The values for β_1 and β_2 are based on coefficient estimates obtained by applying the FM-CPR estimator to GDP and CO_2 emissions data for Austria (see Wagner, 2015). We present simulation results for $T \in \{50, 100, 200, 500, 1000\}$ and for $\rho_1 = \rho_2 \in \{0, 0.3, 0.6, 0.8\}$. The number of replications is 10,000 throughout and all tests are carried out at the nominal 5% level.

We only report results for the Bartlett kernel, and merely note that the results for the Quadratic Spectral kernel, available upon request in supplementary material, are qualitatively very similar. With respect to bandwidth choice we report results for three bandwidth selection rules. These are the data-dependent rules of Andrews (1991), labelled And, Newey and West (1994), labelled NW, as well as a "simplified" sample size dependent version of the latter, i. e., $M_T = \lfloor 4(T/100)^{2/9} \rfloor$, labelled $\mathrm{NW_T}$, that is widely-used.[7] The parameter hypothesis test results are "benchmarked" against OLS-based

[7]The usage of these bandwidth rules is purely pragmatic given that these are implemented in many software packages. However, there is no optimality theory available for the situation considered in this paper. Furthermore, from an asymptotic perspective the following has to be taken into account: The chosen data-dependent bandwidths are of the form $\hat{M}_T = \hat{\gamma} T^{\frac{1}{1+2r}}$, where $\hat{\gamma}$ is a parameter

ρ_1, ρ_2	Bias Ratio			RMSE Ratio		
	And	NW	$\mathrm{NW_T}$	And	NW	$\mathrm{NW_T}$
Panel A: $T = 50$						
0.0	0.6475	1.9067	0.4405	0.9798	0.9920	0.9953
0.3	0.9575	1.1808	0.9847	1.0177	1.0207	1.0332
0.6	0.9838	1.0960	1.0272	1.0566	1.0662	1.0787
0.8	0.9952	1.0466	1.0245	1.0666	1.0715	1.0893
Panel B: $T = 100$						
0.0	1.1342	1.1153	1.0193	1.0143	1.0123	1.0149
0.3	1.0410	1.1959	1.0245	1.0466	1.0382	1.0475
0.6	1.0159	1.0756	1.0396	1.0754	1.0689	1.0876
0.8	1.0226	1.0749	1.0268	1.0826	1.0773	1.0940
Panel C: $T = 200$						
0.0	1.8361	1.9520	1.7630	1.0287	1.0226	1.0223
0.3	1.1629	1.3829	1.1087	1.0495	1.0399	1.0405
0.6	1.0504	1.1447	1.0424	1.0741	1.0699	1.0664
0.8	1.0920	1.1718	1.0253	1.0939	1.1044	1.0707
Panel D: $T = 500$						
0.0	-13.7188	35.9936	17.6654	1.0251	1.0150	1.0133
0.3	1.1604	1.3262	1.1487	1.0351	1.0224	1.0208
0.6	1.0829	1.2659	1.0326	1.0500	1.0438	1.0359
0.8	1.2211	1.3725	1.0183	1.0811	1.1060	1.0442
Panel E: $T = 1000$						
0.0	1.0984	1.1024	1.1868	1.0221	1.0153	1.0109
0.3	1.1090	1.2001	1.0678	1.0286	1.0216	1.0164
0.6	1.0979	1.3687	1.0221	1.0369	1.0357	1.0262
0.8	1.3381	1.5752	1.0134	1.0726	1.1110	1.0320

Table 1: Bias and RMSE ratios, FM-OLS/FM-CPR, for β_1.

test results. We use textbook OLS inference ignoring serial correlation and endogeneity altogether, labelled OLS, which is asymptotically invalid in the presence of serial correlation and endogeneity. Rejections for the Wald-type parameter tests are carried out using the chi-squared distribution.[8]

We start the discussion of the results by comparing bias and RMSE of the two estimators. Table 1 presents the results for β_1 in the form of ratios, with the FM-OLS results divided

related to the shape of the spectral density at the origin and r is the characteristic exponent of the kernel function. For the Bartlett kernel $r = 1$ (see, e. g., Section 5 of Parzen, 1957) and thus $\hat{M}_T = O_{\mathbb{P}}(T^{1/3})$. This, at face value, violates the rate restriction given in Assumption 2.4 in the main text, but has no immediate effect on finite sample performance.

[8] A large variety of additional results – as mentioned also for the Quadratic Spectral kernel – including results for the other coefficients or t-tests also for the cubic and quartic specifications are contained in supplementary material available upon request.
One important additional result from the simulations is that $\hat{\omega}_{u\cdot v}$ (based on FM-CPR) exhibits in many circumstances better performance – meaning smaller bias and RMSE – than $\hat{\omega}_{u\cdot w}$ (based on FM-OLS). These differences are, in addition to the different performance of the estimators, an important ingredient for the different performance of parameter hypothesis as well as cointegration tests based on the two estimators.

by the FM-CPR results, since we are primarily interested in the comparison of the two estimators in this paper. The results are very similar also for δ and β_2. By definition, numbers larger than one (in absolute value) indicate that FM-CPR outperforms FM-OLS and with very few exceptions, when $T = 50$ and the Andrews (1991) bandwidth is used, this is what happens.

Before turning to the relative performance of FM-OLS and FM-CPR in more detail some brief comments on absolute performance are in order. The bias resulting from $\mathrm{NW_T}$ is often larger than when using the data-dependent bandwidth rules, especially for the larger values of T and ρ_1, ρ_2. The Andrews (1991) and Newey and West (1994) bandwidth rules lead to very similar biases. For RMSE the differences are very small for all three bandwidth rules with no clear ranking. These observations hold for both FM-OLS and FM-CPR. Given the – to be expected – absolute disadvantage of $\mathrm{NW_T}$, especially in the presence of error serial correlation and regressor endogeneity, we focus below on the two data-dependent rules.

With respect to the bias ratio one key observation is that the performance advantage of FM-CPR over FM-OLS increases with increasing sample size for large values of ρ_1, ρ_2. For small values of ρ_1, ρ_2 the differences tend to get smaller with increasing T.[9] The RMSE ratios increase throughout for any given T with increasing ρ_1, ρ_2. The variability of the RMSE results is, however, less pronounced than for bias. Roughly speaking, the performance disadvantage of FM-OLS relative to FM-CPR is less severe when using the Andrews (1991) bandwidth than when using the Newey and West (1994) bandwidth.

From the estimator results the empirical null rejection results of the Wald-type tests for the null hypothesis $H_0: \ \beta_1 = 5, \ \beta_2 = -0.3$ can to a certain extent already be guessed, see Table 2 and Figure 1 that contain the same information presented in two different ways. For any given bandwidth choice, size distortions are smaller for the test statistics computed from the FM-CPR estimates compared to those calculated from the FM-OLS estimates. Again the differences are sizeable even for $T = 1000$ for the larger values of ρ_1, ρ_2. The table and figure also illustrate the well-known result that OLS based test statistics do not lead to asymptotic chi-squared distributions in the case of regressor endogeneity and/or error serial correlation, see, e. g., Hong and Phillips (2010, Theorem 2). In our setting mostly the Andrews (1991) rule leads to slightly better results than the Newey and West (1994) rule. The sample-size dependent bandwidth

[9]The large negative values for the bias ratio for $T = 500$ and $\rho_1, \rho_2 = 0$ are driven by "base-effects", i. e., both the numerator and the denominator are very small, with the denominator by one order smaller.

ρ_1, ρ_2	OLS	FM-OLS			FM-CPR		
		And	NW	NW_T	And	NW	NW_T
Panel A: $T = 50$							
0.0	0.0757	0.1944	0.2139	0.1638	0.1777	0.1762	0.1472
0.3	0.2184	0.2686	0.2918	0.2397	0.2241	0.2396	0.2036
0.6	0.5141	0.4171	0.4462	0.4037	0.3399	0.3684	0.3418
0.8	0.7853	0.6396	0.6734	0.6468	0.5569	0.5816	0.5927
Panel B: $T = 100$							
0.0	0.0597	0.1370	0.1222	0.1183	0.1231	0.1018	0.1063
0.3	0.2066	0.1807	0.1868	0.1686	0.1545	0.1588	0.1434
0.6	0.5352	0.3067	0.3444	0.3075	0.2436	0.2645	0.2563
0.8	0.8164	0.5353	0.6049	0.5634	0.4272	0.4587	0.5120
Panel C: $T = 200$							
0.0	0.0572	0.1070	0.0987	0.0859	0.0940	0.0836	0.0777
0.3	0.2045	0.1385	0.1450	0.1265	0.1176	0.1255	0.1136
0.6	0.5449	0.2224	0.2663	0.2497	0.1748	0.1941	0.2201
0.8	0.8279	0.4234	0.5102	0.5166	0.2974	0.3253	0.4854
Panel D: $T = 500$							
0.0	0.0517	0.0848	0.0766	0.0673	0.0744	0.0663	0.0630
0.3	0.2022	0.1046	0.1123	0.0985	0.0886	0.0980	0.0882
0.6	0.5498	0.1469	0.1965	0.1803	0.1151	0.1248	0.1649
0.8	0.8380	0.2952	0.4016	0.4175	0.1787	0.1913	0.3974
Panel E: $T = 1000$							
0.0	0.0520	0.0711	0.0641	0.0612	0.0645	0.0600	0.0587
0.3	0.2046	0.0840	0.0911	0.0839	0.0747	0.0866	0.0788
0.6	0.5560	0.1131	0.1611	0.1438	0.0904	0.0962	0.1363
0.8	0.8439	0.2166	0.3340	0.3464	0.1286	0.1400	0.3341

Table 2: Empirical null rejection probabilities of Wald-type tests for $H_0\colon \beta_1 = 5, \beta_2 = -0.3$.

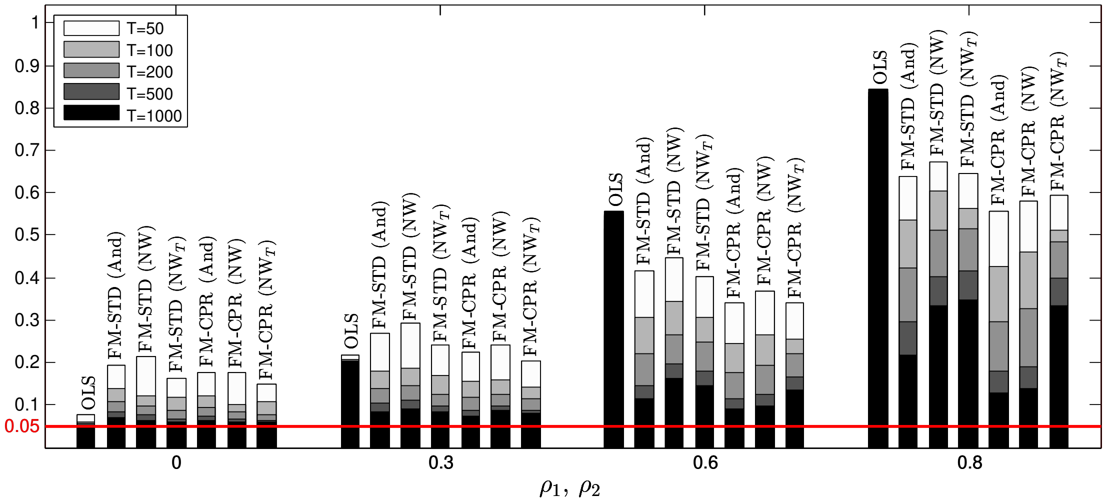

Figure 1: Empirical null rejection probabilities of Wald-type tests for $H_0\colon \beta_1 = 5, \beta_2 = -0.3$.

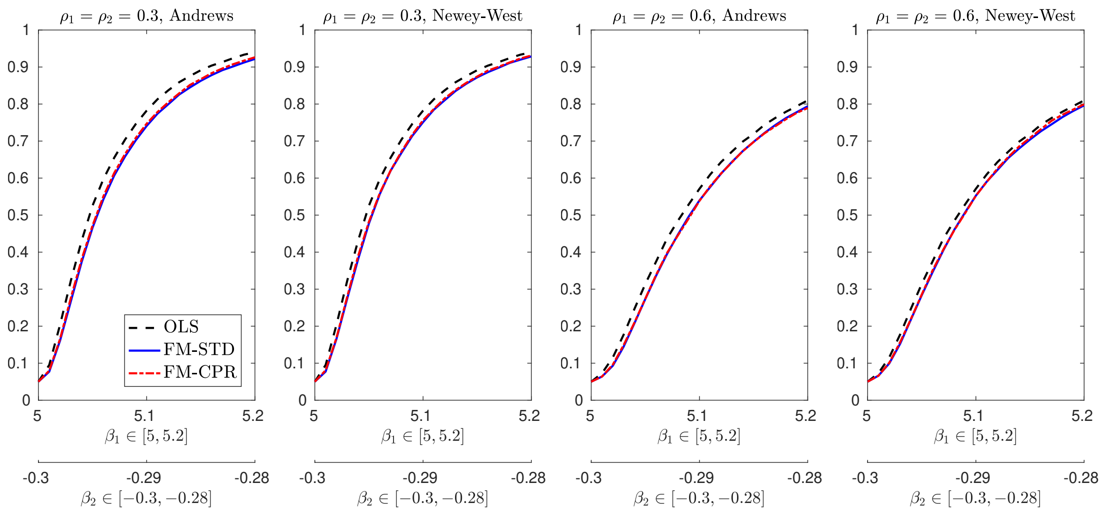

Figure 2: Size-corrected power of Wald-type tests for $H_0 : \beta_1 = 5, \beta_2 = -0.3$ for $T = 100$. The two left graphs correspond to $\rho_1 = \rho_2 = 0.3$ and the two right graphs to $\rho_1 = \rho_2 = 0.6$. Within these pairs the left graph corresponds to the Andrews (1991) bandwidth and the right graph to the Newey and West (1994) bandwidth.

$\mathrm{NW_T}$ performs – as expected – especially poorly in the case of large serial correlation (and large sample sizes). Large correlation cannot be adequately taken into account with the – in such cases – “too small” $\mathrm{NW_T}$ bandwidth that is independent of the second moment structure.

We now turn briefly to size-corrected power of the Wald-type test just considered under the null by considering size-corrected power for a grid of (including the null) 21 points. The values for β_1 are chosen from the interval $[5, 5.2]$ on an equidistant grid with mesh 0.01 and the values for β_2 from the interval $[-0.3, -0.28]$ on an equidistant grid with mesh 0.001. Figure 2 displays results for $T = 100$ for $\rho_1, \rho_2 = 0.3$ in the left two graphs and for $\rho_1, \rho_2 = 0.6$ in the right two graphs. Within these two graphs, the left graph corresponds to the Andrews (1991) bandwidth and the right graph to the Newey and West (1994) bandwidth.

Figure 2 shows some very typical findings: First, size-corrected power is slightly higher for OLS, which, however, has the highest size distortions under the null and leads to invalid inference for $\rho_1, \rho_2 \neq 0$ even asymptotically. Second, size-corrected power is virtually identical for FM-OLS and FM-CPR. Third, the Andrews (1991) bandwidth leads to marginally lower size-corrected power than the Newey and West (1994) bandwidth, which has to be seen, however, in conjunction with the lower size distortions resulting from using the Andrews (1991) bandwidth. Overall, the bandwidth rule of Andrews (1991) leads to the best performance for parameter hypothesis testing.

Let us now turn briefly to cointegration testing. We report in Table 3 the null rejection probabilities for the test variants discussed at the end of Section 2 in the main text. The three-block columns correspond to the following variants: The first column, $\mathrm{CT}^{++}_{\mathrm{Shin}}$, corresponds to the widespread empirical practice of using the FM-OLS residuals in conjunction with the (inappropriate) Shin (1994) critical values. The second column, CT^{++}, is a "hybrid" version, with the test statistic calculated from the FM-OLS residuals and the test decisions based on the asymptotically correct critical values. The third column, CT^{+}, reports the results obtained using the FM-CPR residuals and the critical values corresponding to the limiting distribution given in (B.5) and (B.6); tabulated in Wagner (2023, Table 4); with the critical values required for our setting also available in Table 1 in the main text.

The null performance of the different cointegration test versions can be summarized as follows: The $\mathrm{CT}^{++}_{\mathrm{Shin}}$-test typically exhibits the largest over-rejections. These over-rejections, that stay substantial even for $T = 1000$, reflect the usage of wrong critical values. The hybrid CT^{++}-test exhibits a very similar performance as the $\mathrm{CT}^{++}_{\mathrm{Shin}}$-test. This is partly not surprising, since the same test statistic is used and the critical values only differ by little in the considered specification (compare Table 1). Thus, the results cannot differ too much either. Another reason for the poor performance of CT^{++}

ρ_1, ρ_2	CT^{++}_{Shin}			CT^{++}			CT^{+}		
	And	NW	NW_T	And	NW	NW_T	And	NW	NW_T
Panel A: $T = 50$									
0.0	0.0332	0.1050	0.0321	0.0319	0.1015	0.0303	0.0389	0.0769	0.0400
0.3	0.0640	0.1368	0.0614	0.0614	0.1336	0.0589	0.0600	0.1139	0.0722
0.6	0.1368	0.2265	0.1419	0.1334	0.2210	0.1372	0.0792	0.1928	0.1660
0.8	0.2270	0.3745	0.3249	0.2198	0.3669	0.3178	0.1135	0.2849	0.3734
Panel B: $T = 100$									
0.0	0.0411	0.0518	0.0442	0.0368	0.0447	0.0379	0.0421	0.0472	0.0450
0.3	0.0646	0.0955	0.0717	0.0577	0.0876	0.0646	0.0630	0.0965	0.0728
0.6	0.1280	0.2415	0.1529	0.1151	0.2248	0.1399	0.0768	0.1568	0.1556
0.8	0.2892	0.4932	0.4031	0.2687	0.4756	0.3812	0.0867	0.2449	0.4181
Panel C: $T = 200$									
0.0	0.0480	0.0517	0.0534	0.0413	0.0441	0.0437	0.0465	0.0480	0.0485
0.3	0.0677	0.0968	0.0878	0.0581	0.0865	0.0784	0.0654	0.0926	0.0815
0.6	0.1198	0.2282	0.2073	0.1078	0.2129	0.1886	0.0752	0.1267	0.1952
0.8	0.2928	0.4755	0.5467	0.2673	0.4518	0.5152	0.0712	0.1715	0.5323
Panel D: $T = 500$									
0.0	0.0535	0.0537	0.0570	0.0461	0.0459	0.0487	0.0492	0.0487	0.0493
0.3	0.0679	0.0917	0.0850	0.0581	0.0782	0.0753	0.0625	0.0845	0.0763
0.6	0.1012	0.2035	0.1773	0.0870	0.1842	0.1548	0.0666	0.0850	0.1590
0.8	0.2282	0.4392	0.4859	0.2042	0.4169	0.4530	0.0597	0.1105	0.4597
Panel E: $T = 1000$									
0.0	0.0582	0.0602	0.0604	0.0488	0.0511	0.0514	0.0518	0.0507	0.0530
0.3	0.0705	0.0914	0.0857	0.0599	0.0786	0.0740	0.0621	0.0809	0.0748
0.6	0.0957	0.1847	0.1576	0.0814	0.1669	0.1384	0.0648	0.0760	0.1401
0.8	0.1856	0.3882	0.4258	0.1637	0.3628	0.3905	0.0582	0.0866	0.3959

Table 3: Empirical null rejection probabilities of cointegration tests. The block-column CT^{++}_{Shin} reports the results from using the test statistic (16) and the Shin (1994) critical values. The block-columns CT^{++} and CT^{+} report the results from using (16) based on either the FM-OLS residuals or the FM-CPR residuals and the corresponding critical value tabulated in Wagner (2023, Table 6, p. 17–18). For the considered specification the 5% critical values are 0.101 (Shin) and 0.106 (Wagner) respectively, compare also Table 1.

compared to CT^+ is that it suffers from the poor performance of the estimator $\hat{\omega}_{u \cdot w}$ mentioned in Footnote 8. This leads to poor performance even when comparing the statistic with the correct critical values. The performance of the CT^+-test is substantially better, with a performance advantage that grows for the large values of ρ_1, ρ_2. In these comparisons as before the sample size dependent bandwidth NW_T has to be considered separately, with again poor performance in the case of large ρ_1, ρ_2 and all values of T. From the two data-dependent bandwidths the Andrews (1991) bandwidth leads to – partly substantially – better results.

We close this section by considering the power performance of the cointegration test variants, considering the following three alternative DGPs:

$$
\begin{aligned}
\text{(I)} &: \; y_t = 1 + t + 5x_t - 0.3x_t^2 + 0.01x_t^3 + u_t \\
\text{(II)} &: \; y_t = 1 + t + 5x_t - 0.3x_t^2 + e_t, \text{ with unobserved } e_t \sim I(1) \text{ independent of } x_t \\
\text{(III)} &: \; y_t,\, x_t \text{ are two independent } I(1) \text{ variables}
\end{aligned}
$$

In (I) the regressor x_t and error u_t are generated as described above (with the same values of ρ_1, ρ_2). Also in case (II), x_t is generated as before and $e_t = \sum_{j=1}^{t} \varepsilon_j$, with $\varepsilon_j \sim \mathcal{N}(0,1)$ independent of x_t. Finally, in case (III) y_t and x_t are generated independently of each other, exactly as e_t in case (II). These three DGPs cover some main alternatives of interest. Case (I) covers misspecification of the polynomial degree, alternative (II) corresponds to the case of no cointegration because of a missing integrated regressor, and alternative (III) corresponds to a spurious regression alternative.

Size-corrected power of all variants of the CT-tests depends strongly upon alternatives considered. In particular, size-corrected power is much larger for alternatives (II) and (III) than for alternative (I), especially when considering large values of ρ_1, ρ_2. For alternatives (II) and (III) using the Andrews (1991) bandwidth leads to size-corrected power that decreases with the sample size. Note that Lee (1996) and Xiao and Phillips (2002) report potential problems with the power of stationarity and cointegration tests respectively, when such tests are used in combination with data-dependent bandwidths. The higher size-corrected power often observed with the Newey and West (1994) bandwidth compared to the Andrews (1991) bandwidth has to be seen in conjunction with the larger size distortions obtained when using the Newey and West (1994) bandwidth. This effect is even more pronounced when using the sample-size dependent NW_T bandwidth, especially for the larger sample sizes. For alternative (I) size corrected power is higher for CT^+ than for the two variants of the CT^{++} test, whereas this ordering is,

	ρ_1, ρ_2	CT^{++}_{Shin} & CT^{++}			CT^{+}		
		And	NW	NW_T	And	NW	NW_T
Panel A: $T = 50$							
(I)	0.0	0.1336	0.0302	0.1995	0.1237	0.0934	0.2122
	0.3	0.0705	0.0206	0.1255	0.0901	0.0576	0.1309
	0.6	0.0334	0.0094	0.0385	0.0716	0.0270	0.0417
	0.8	0.0223	0.0035	0.0032	0.0535	0.0168	0.0024
(II)	-	0.3923	0.2938	0.5680	0.2615	0.3206	0.5985
(III)	-	0.5126	0.3247	0.5934	0.2864	0.3223	0.6295
Panel B: $T = 100$							
(I)	0.0	0.1407	0.1587	0.3549	0.1554	0.1964	0.3699
	0.3	0.0912	0.0713	0.2742	0.1105	0.0892	0.2871
	0.6	0.0465	0.0027	0.1467	0.0943	0.0360	0.1471
	0.8	0.0280	0.0000	0.0246	0.0866	0.0150	0.0238
(II)	-	0.5157	0.7670	0.7822	0.3071	0.5856	0.8040
(III)	-	0.4769	0.7533	0.7848	0.2935	0.5806	0.8087
Panel C: $T = 200$							
(I)	0.0	0.1333	0.2465	0.6864	0.1499	0.2828	0.6871
	0.3	0.0964	0.1263	0.5856	0.1072	0.1581	0.5958
	0.6	0.0559	0.0106	0.3868	0.0918	0.1065	0.3972
	0.8	0.0310	0.0000	0.1289	0.1028	0.0749	0.1313
(II)	-	0.5652	0.8913	0.9640	0.2960	0.6906	0.9660
(III)	-	0.3841	0.8862	0.9649	0.2892	0.7001	0.9689
Panel D: $T = 500$							
(I)	0.0	0.1224	0.4333	0.9182	0.1533	0.4562	0.9193
	0.3	0.0941	0.3205	0.8781	0.1152	0.3404	0.8780
	0.6	0.0626	0.1167	0.7777	0.1098	0.3378	0.7791
	0.8	0.0376	0.0002	0.4985	0.1277	0.2965	0.5030
(II)	-	0.4650	0.9803	0.9982	0.2929	0.8586	0.9981
(III)	-	0.2603	0.9797	0.9979	0.2750	0.8624	0.9981
Panel E: $T = 1000$							
(I)	0.0	0.1134	0.5740	0.9814	0.1404	0.5969	0.9817
	0.3	0.0931	0.4722	0.9711	0.1133	0.4960	0.9714
	0.6	0.0715	0.2644	0.9382	0.1073	0.5087	0.9398
	0.8	0.0428	0.0272	0.8015	0.1269	0.4839	0.8031
(II)	-	0.3605	0.9979	0.9999	0.2734	0.9395	0.9999
(III)	-	0.2085	0.9949	0.9999	0.2505	0.9381	0.9999

Table 4: Size-corrected power of cointegration tests. Size correction using the empirical distribution leads by construction to identical results for CT^{++}_{Shin} and CT^{++}.

surprisingly, mostly reversed for alternatives (II) and (III). However, again, this higher size-corrected power has to be seen in conjunction with the partly substantially larger size distortions under the null.

References

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

Baek, J. (2015). Environmental Kuznets Curve for CO_2 Emissions: The Case of Arctic Countries. *Energy Economics* **50**, 13–17.

Bradford, D., R. Fender, S.H. Shore and M. Wagner (2005). The Environmental Kuznets Curve: Exploring a Fresh Specification. *Contributions to Economic Analysis and Policy* **4**, No. 1, Article 5. Berkeley Electronic Press.

Csörgo, M. and L. Horváth (1993). *Weighted Approximations in Probability and Statistics.* New York: John Wiley & Sons.

Esteve, V. and C. Tamarit (2012). Threshold Cointegration and Nonlinear Adjustment Between CO2 and Income: The Environmental Kuznets Curve in Spain, 1857-2007. *Energy Economics* **34**, 2148–2156.

Fosten, J., B. Morley and T. Taylor (2012). Dynamic Misspecification in the Environmental Kuznets Curve: Evidence from CO_2 and SO_2 Emissions in the United Kingdom. *Ecological Economics* **76**, 25–33.

Friedl, B. and M. Getzner (2003). Determinants of CO_2 Emissions in a Small Open Economy. *Ecological Economics* **45**, 133–148.

Galeotti, M., A. Lanza and F. Pauli (2006). Reassessing the Environmental Kuznets Curve for CO_2 Emissions: A Robustness Exercise. *Ecological Economics* **57**, 152–163.

He, J. and P. Richard (2010). Environmental Kuznets Curve for CO_2 in Canada. *Ecological Economics* **69**, 1083–1093.

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.

Jalil, A. and S. Mahmud (2009). Environment Kuznets Curve for CO_2 Emissions: A Cointegration Analysis for China. *Energy Policy* **37**, 5167–5172.

Karatzas, I. and S. Shreve (1991). *Brownian Motion and Stochastic Calculus.* New York: Springer.

Kasparis, I. (2008). Detection of Functional Form Misspecification in Cointegrating Relations. *Econometric Theory* **24**, 1373–1403.

Lee, J. (1996). On the Power of Stationarity Tests Using Optimal Bandwidth Estimates. *Economics Letters* **51**, 131–137.

Liang, H., P.C.B. Phillips, H. Wang and Q. Wang (2016). Weak Convergence to Stochastic Integrals for Econometric Applications. *Econometric Theory* **32**, 1349–1375.

Lindmark, M. (2002). An EKC-Pattern in Historical Perspective: Carbon Dioxide Emissions, Technology, Fuel Prices and Growth in Sweden, 1970–1997. *Ecological Economics* **42**, 333–347.

Müller-Fürstenberger, G. and M. Wagner (2007). Exploring the Environmental Kuznets Hypothesis: Theoretical and Econometric Problems. *Ecological Economics* **62**, 648–660.

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

Özcan, B. and I. Öztürk, Eds. (2019). *Environmental Kuznets Curve: A Manual.* Elsevier: Academic Press.

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.

Parzen, E. (1957). On Consistent Estimates of the Spectrum of a Stationary Time Series. *Annals of Mathematical Statistics* **28**, 329–348.

Phillips, P.C.B. (1989). Partially Identified Econometric Models. *Econometric Theory* **5**, 181–240.

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.

Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer.

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

Vogelsang, T.J. and M. Wagner (2014). An Integrated Modified OLS RESET Test for Cointegrating Regressions. *SFB823 Discussion Paper* 37/14.

Wagner, M. (2008). The Carbon Kuznets Curve: A Cloudy Picture Emitted by Bad Econometrics? *Resource and Energy Economics* **30**, 388–408.

Wagner, M. (2015). The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* **30**, 948–967.

Wagner, M. (2023). Residual-Based Cointegration and Non-Cointegration Tests for Cointegrating Polynomial Regressions. *Empirical Economics* **65**, 1–31.

Wagner, M. and S.H. Hong (2016). Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference. *Econometric Theory* **32**, 1289–1315.

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