Econometric Theory, **0**, 2024, 1–21.
doi:10.1017/S0266466624000033

COINTEGRATING POLYNOMIAL REGRESSIONS: ROBUSTNESS OF FULLY MODIFIED OLS

Oliver Stypka
Flossbach von Storch

Martin Wagner
University of Klagenfurt, Bank of Slovenia, and Institute for Advanced Studies

Peter Grabarczyk
TU Dortmund University

Rafael Kawka
TU Dortmund University

Cointegrating polynomial regressions (CPRs) include deterministic variables, integrated variables, and their powers as explanatory variables. Based on a novel kernel-weighted limit result and a novel functional central limit theorem, this paper shows that the fully modified ordinary least squares (FM-OLS) estimator of Phillips and Hansen (1990, *Review of Economic Studies* 57, 99–125) is robust to being used in CPRs. Being used in CPRs refers to a widespread empirical practice that treats the integrated variables and their powers, incorrectly, as a vector of integrated variables and uses textbook FM-OLS. Robustness means that this "formal" FM-OLS practice leads to a zero mean Gaussian mixture limiting distribution that coincides with the limiting distribution of the Wagner and Hong (2016, *Econometric Theory* 32, 1289–1315) application of the FM estimation principle to the CPR case. The only restriction for this result to hold is that all integrated variables to power one are included as regressors. Even though simulation results indicate performance advantages of the Wagner and Hong (2016, *Econometric Theory* 32, 1289–1315) estimator, partly even in large samples, the results of the paper give an asymptotic foundation to "formal" FM-OLS and thus enlarge the usability of the Phillips and Hansen (1990, *Review of Economic Studies* 57, 99–125) estimator implemented in many software packages.

All authors acknowledge partial financial support from the Collaborative Research Center 823: *Statistical Modelling of Nonlinear Dynamic Processes* supported by the Deutsche Forschungsgemeinschaft (DFG). Martin Wagner furthermore acknowledges research support from the Jubilaumsfonds of the Oesterreichische Nationalbank via several grants. We gratefully acknowledge the helpful comments of, in particular, the editor Peter C. B. Phillips, as well as of a co-editor and four reviewers. We furthermore are grateful for the many comments received from participants in numerous conferences and seminars since 2016. This paper solely reflects the views of the authors and not necessarily those of Flossbach von Storch, the Bank of Slovenia, or the European System of Central Banks. On top of this, the usual disclaimer applies. Address correspondence to Martin Wagner, Department of Economics, University of Klagenfurt, Klagenfurt, Austria; e-mail: martin.wagner@aau.at.

1. INTRODUCTION

In several empirical literatures, including (most voluminously) the environmental Kuznets curve (EKC) literature (for an early survey paper counting more than 100 refereed publications, see Yandle, Bhattarai, and Vijayaraghavan, 2004),[1] the material Kuznets curve literature (see, e. g., Grabarczyk et al., 2018), the intensity-of-use literature (see, e. g., Malenbaum, 1978; Labson and Crompton, 1993), the exchange rate target zone literature (see, e. g., Svensson, 1992; Darvas, 2008) or the fiscal reaction function literature (for a recent contribution, see, e. g., Di Iorio and Fachin, 2022), so-called cointegrating polynomial regression (CPR) models are routinely employed, using here the terminology of Wagner and Hong (2016). These are regression models that include deterministic variables, integrated variables, and their powers as explanatory variables.[2] As in the linear cointegration literature, both regressor endogeneity and error serial correlation are allowed for in CPRs.

The simplest example of a CPR model is the quadratic formulation, which is also the workhorse model in the EKC literature, that is, $y_t = x_t\beta_1 + x_t^2\beta_2 + u_t = X_t'\beta + u_t$, with $x_t = x_{t-1} + v_t$, $X_t = [x_t, x_t^2]'$ and $\beta = [\beta_1, \beta_2]'$. Detailed definitions and assumptions are given below in Section 2.1 and, for the heuristic discussion here in the introduction, we furthermore ignore deterministic components. A CPR model is probably the simplest case of a nonlinear cointegrating relationship that is in fact linear in parameters and for which, hence, closed-form solutions of, for example, modified least squares estimators, that allow for asymptotic standard inference, are available. Wagner and Hong (2016) extend the fully modified OLS (FM-OLS) estimator of Phillips and Hansen (1990) from cointegrating linear relationships to cointegrating polynomial relationships and dub the resultant estimator FM-CPR, for the example at hand given by $\hat{\beta}^+ = \left(\sum_{t=1}^T X_t X_t'\right)^{-1}\left(\sum_{t=1}^T X_t y_t^+ - \hat{\Delta}_{vu}^+\begin{pmatrix} T \\ 2\sum_{t=1}^T x_t \end{pmatrix}\right)$, with $y_t^+ = y_t - \Delta x_t \hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu} = y_t - v_t\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$ and $\hat{\Delta}_{vu}^+ := \hat{\Delta}_{vu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$. The matrices Δ and Ω, partitioned according to u_t and v_t denote the half long-run and long-run covariance matrices of $\{[u_t, v_t]'\}_{t\in\mathbb{Z}}$, respectively, and $\hat{\Delta}$ and $\hat{\Omega}$ denote consistent estimators based on $[\hat{u}_t, v_t]'$, with $\hat{u}_t$ denoting the OLS residuals. FM-type estimation relies upon two transformations: First, the dependent variable y_t is replaced by y_t^+, with this transformation being instrumental for removing endogeneity biases. Second, the subtraction of an additive correction term, $\hat{\Delta}_{vu}^+[T, 2\sum_{t=1}^T x_t]'$ in the example, removes so-called second-order bias terms from the limiting distribution. Both transformations together lead to a zero mean Gaussian mixture limiting distribution of the FM-OLS estimator, which is the basis for asymptotic standard inference. Why the Phillips and Hansen (1990) replacement of y_t by y_t^+ is

[1]The term environmental Kuznets curve (EKC) refers by analogy to the inverted U-shaped relationship between the level of economic development and the degree of inequality postulated by Simon Kuznets (1955) in his 1954 presidential address to the American Economic Association. Since the important early contribution of Grossman and Krueger (1995) at the latest, the EKC literature has been rapidly expanding.

[2]Sadly Stefano Fachin, a very nice colleague and friend, passed away in early 2023. Given that he was an avid user of CPR models and methods, this paper is also dedicated to him.

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

instrumental for removing endogeneity biases becomes clear from considering the transformed regression model $y_t^+ = X_t'\beta + u_t^+$, in which, loosely speaking, $u_t^+ = u_t - v_t\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$ is asymptotically uncorrelated with the regressors X_t. More precisely, it holds (when long-run covariance estimation is performed consistently) that the scaled limit process $B_{u\cdot v}(r) = \lim_{T\to\infty} \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} \left(u_t - v_t\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}\right) = B_u(r) - B_v(r)\Omega_{vv}^{-1}\Omega_{vu}$, $0 \leq r \leq 1$, is a Brownian motion that is uncorrelated with—and thus independent of—the Brownian motion $B_v(r)$ corresponding to x_t.[3] Importantly, the transformed error limit (partial sum) process $B_{u\cdot v}(r)$ is, due to Gaussianity, also independent of the limit process corresponding to the regressor x_t^2, that is, to $B_v^2(r)$; in fact, $B_{u\cdot v}(r)$ is independent of any well-defined function $F(B_v(r))$. Therefore, the Phillips and Hansen (1990) replacement of y_t by y_t^+ asymptotically addresses endogeneity issues not only in linear but also in nonlinear cointegrating regression settings (for early discussions, see, e. g., Phillips, 1989, 1991a). The literature extending the Phillips and Hansen (1990) approach from linear to nonlinear cointegration settings, of course, builds upon this fact; with early contributions including Chang, Park, and Phillips (2001) and Park and Phillips (1999, 2001) and more recent contributions including Chan and Wang (2015), Ibragimov and Phillips (2008), or Liang et al. (2016). Whilst the transformation of the dependent variable is invariant to the specification of the model, the additive correction term, in the example considered $\hat{\Delta}_{vu}^+[T, 2\sum_{t=1}^T x_t]'$, depends on the specification of the model.[4]

In the empirical literature, in particular in the EKC literature, it has been and is common practice to "formally" use the Phillips and Hansen (1990) FM-OLS estimator by "interpreting," in the example considered, the vector X_t not as a vector composed of an integrated process and its square, but as a vector of two (non-cointegrated) I(1) processes (for a discussion of this practice, see Wagner, 2015) also rather than to use the FM-CPR estimator tailor-made for CPR models.[5] This leads to the "formal" FM-OLS estimator of β, given by $\hat{\beta}^{++} =$

[3]The transformation also has a well-known (limit) Hilbert space interpretation discussed already in early work by Peter C. B. Phillips (and co-authors); see Appendix C of the Supplementary Material for details in the CPR context. Consider (either) for a fixed $0 < r \leq 1$ (or analogously, see Appendix C of the Supplementary Material, for the interval $0 \leq r \leq 1$) the population regression of $B_u(r)$ on $B_v(r)$, that is, $B_u(r) = B_v(r)\Theta_{[1]}(r) + B_{u\cdot v}(r)$. Then $\Theta_{[1]}(r) = (\mathbb{E}(B_v^2(r)))^{-1}\mathbb{E}(B_v(r)B_u(r)) = (r\Omega_{vv})^{-1}(r\Omega_{vu}) = \Omega_{vv}^{-1}\Omega_{vu}$, with $B_{u\cdot v}(r)$ by construction uncorrelated with and thus independent of $B_v(r)$. Thus, $\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$ can be interpreted—as is well-known—as a consistent estimator of the population regression coefficient $\Theta_{[1]}(r)$.

[4]Denoting with $G = \mathrm{diag}(T^{-1}, T^{-3/2})$ and with $\mathbf{B}_v(r) = [B_v(r), B_v^2(r)]'$, it is well-known (see, e. g., Wagner and Hong, 2016, Prop. 1) that $G^{-1}\left(\hat{\beta}^+ - \beta\right) \Rightarrow \left(\int_0^1 \mathbf{B}_v(r)\mathbf{B}_v(r)'\right)^{-1} \int_0^1 \mathbf{B}_v(r)dB_{u\cdot v}(r)$, with $\mathbf{B}_v(r)$ and $B_{u\cdot v}(r)$, as discussed, independent of each other. Since $G^{-1}\sum_{t=1}^T X_t u_t^+ \Rightarrow \int_0^1 \mathbf{B}_v(r)dB_{u\cdot v}(r) + \Delta_{vu}^+[1, 2\int_0^1 B_v(r)dr]'$, the second-order bias term $\Delta_{vu}^+[1, 2\int_0^1 B_v(r)dr]'$ has to be asymptotically removed, which is asymptotically achieved by subtracting $\hat{\Delta}_{vu}^+[T, 2\sum_{t=1}^T x_t]'$ in the definition of the estimator, to arrive at a zero mean Gaussian mixture limiting distribution. Details on the mechanics of the additive bias term removal of FM-CPR are contained, for example, in the proof of Wagner and Hong (2016, Prop. 1).

[5]Based on the insights of Phillips and Hansen (1990), the *appropriate* label for FM-CPR should probably be FM-OLS also in the CPR case. However, the label FM-OLS is reserved in this paper for the "formal" application of the Phillips and Hansen (1990) estimator in CPRs. For brevity, we do not always carry the adjective "formal" along.

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

$\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1}\left(\sum_{t=1}^{T} X_t y_t^{++} - T\hat{\Delta}_{wu}^{+}\right)$, with $y_t^{++} = y_t - \Delta X_t \hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}$ based on $w_t = \Delta X_t = [v_t, 2x_t v_t - v_t^2]'$. Here, the second component of w_t is, obviously, a nonstationary process. This implies that $\hat{\Omega}_{ww}$ as well as $\hat{\Omega}_{wu}$ are, of course, not estimators of underlying long-run covariances of stationary processes.

This paper shows that "formal" FM-OLS is robust to being used in CPRs, that is, $\hat{\beta}^{++}$ has the same limiting distribution as the FM-CPR estimator $\hat{\beta}^{+}$. Some aspects of the result are not surprising, in particular from the perspective of Hilbert space geometry of the limiting experiment. However, the formal proof requires two results: First, the asymptotic behavior of (properly scaled) formal long-run covariance matrix estimators like $\hat{\Omega}_{ww}$ and $\hat{\Omega}_{wu}$ needs to be established (Theorem 2.7). Second, a functional central limit theorem involving the products of powers of integrated processes with first differences of powers of integrated processes has to be derived (Theorem 2.9). Note that these underlying theorems may serve as important inputs or blueprints also, for example, for misspecification analysis of nonlinear cointegrating relationships or for understanding the properties of Sieve approximations to nonlinear cointegrating relationships, over and above being the key ingredients for the FM-OLS robustness result presented in this paper.[6]

To illustrate the key ingredient for the robustness result, let us consider, similarly to Footnote 4, the term $\sum_{t=1}^{T} X_t u_t^{++}$, with $u_t^{++} = u_t - w_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}$, in some detail, with $G_W = \mathrm{diag}(1, T^{-1/2})$ and $\dot{\mathbf{B}}_v(r) = [1, 2B_v(r)]'$, but with *omitting* additive bias terms that are present in the limit (and whose asymptotic removal is discussed in detail in Section 2.2) for brevity:

$$
\begin{aligned}
G^{-1}\sum_{t=1}^{T} X_t u_t^{++} &= G^{-1}\sum_{t=1}^{T} X_t u_t - \left(G^{-1}\sum_{t=1}^{T} X_t w_t' G_W\right)\left(G_W \hat{\Omega}_{ww} G_W\right)^{-1} G_W \hat{\Omega}_{wu} \\
&\Rightarrow \int_0^1 \mathbf{B}_v(r) dB_u(r) - \int_0^1 \mathbf{B}_v(r) d\mathbf{B}_v(r)' \left(\Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)' dr\right)^{-1} \Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r) \\
&= \int_0^1 \mathbf{B}_v(r) dB_u(r) - \int_0^1 \mathbf{B}_v(r)[dB_v(r)\, dB_v^2(r)] \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0 \end{bmatrix} \\
&= \int_0^1 \mathbf{B}_v(r) dB_{u\cdot v}(r),
\end{aligned}
$$

with the convergence results derived, for the general case, in Theorems 2.7 and 2.9. The key algebraic property underlying this result is that $(\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)' dr)^{-1}$ $\int_0^1 \dot{\mathbf{B}}_v(r) = [1, 0]'$, since the first element of $\dot{\mathbf{B}}_v(r) = 1$. This implies that the second term in the third line above simplifies to $\int_0^1 \mathbf{B}_v(r) dB_v(r)\Omega_{vv}^{-1}\hat{\Omega}_{vu}$ and thus—up to bias terms that are subtracted—identical limits of $G^{-1}\sum_{t=1}^{T} X_t u_t^{++}$ and $G^{-1}\sum_{t=1}^{T} X_t u_t^{+}$. Given the discussion concerning the FM transformation above (in particular, in Footnote 3), this is not surprising from a limit Hilbert space perspective: Consider, with full details given in Appendix C of the

[6] Another example where these results have already been fruitfully applied is Stypka and Wagner (2019), who analyze the asymptotic behavior of Phillips (1987)-type unit root tests applied to polynomials of integrated processes.

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

Supplementary Material, the population regression of $B_u(r)$ on $\mathbf{B}_v(r)$, that is, $B_u(r) = \mathbf{B}_v(r)\Theta_{[1:2]}(r) + B_{u\cdot v}(r)$. Gaussianity directly implies that $\Theta_{[1:2]}(r) = [\Omega_{vv}^{-1}\Omega_{vu}, 0]' = [\Theta_{[1]}(r), 0]'$. Thus, orthogonalization with respect to $B_v(r)$ suffices to achieve orthogonality (in fact independence) between the vector of powers $\mathbf{B}_v(r)$ and $B_{u\cdot v}(r)$. There is, however, one subtle important point: In the case of polynomials (of degree larger than one), a Hilbert space population regression coefficient interpretation does not fully apply. To be more precise, in the quadratic example, the population regression coefficient (here again considered for fixed $0 < r \leq 1$ and with, by definition, a similar result for the interval $0 \leq r \leq 1$) is defined as $\Theta_{[1:2]}(r) = \left(\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)')\right)^{-1}\mathbb{E}(\mathbf{B}_v(r)B_{u\cdot v}(r))$, with $\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)') = \begin{bmatrix} r\Omega_{vv} & 0 \\ 0 & 3r^2\Omega_{vv}^2 \end{bmatrix}$ and $\mathbb{E}(\mathbf{B}_v(r)B_{u\cdot v}(r)) = [r\Omega_{vu}, 0]'$. Due to the algebraic structure of the result of Theorem 2.7, however, the product of the inverse of $\Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr$ and $\Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)$, which are both random, is identical to the Hilbert space population regression coefficient, albeit neither of the two terms equals the expected value terms defining the Hilbert space population regression coefficient.[7] Thus from a computational perspective, the "formal" FM-OLS estimator can be interpreted as an estimator whose calculation contains several asymptotically vanishing terms that are set to zero from the outset when considering the FM-CPR estimator. Care has to be taken when testing for cointegration based on the "formal" FM-OLS residuals using, for example, a Shin (1994)-type test. Using the critical values of Shin (1994) for, to stick to the example, two integrated regressors leads to asymptotically invalid inference, as discussed also briefly at the end of Section 2.2. Correct critical values are provided, for example, in Wagner (2023).

The simulation results, relegated to Appendix F of the Supplementary Material due to space limitations, indicate that—*as expected* given the above discussion—FM-CPR outperforms ("formal") FM-OLS in finite samples. In the case of large endogeneity and serial correlation, marked performance differences occur even for large samples like $T = 1{,}000$; despite both estimators having the same limiting distribution. In these cases, the presence of asymptotically superfluous quantities in the FM-OLS estimator detrimentally impacts the performance of FM-OLS compared to FM-CPR even in large samples. The performance advantages occur in all considered dimensions, that is, estimator bias and RMSE, performance of parameter hypothesis tests and performance of cointegration tests. In the case of data with little or no endogeneity and serial correlation, the differences between the estimators effectively vanish for the larger sample sizes ($T = 1{,}000$) considered. The performance differences are particularly large for Shin (1994)-type cointegration testing, also when the test statistic based on the FM-OLS residuals is used in conjunction with the correct critical values.

[7] Note for completeness that (in the considered example) $\mathbb{E}(\Omega_{\tilde{w}\tilde{w}}) = \mathrm{diag}(\Omega_{vv}, 2\Omega_{vv}^2)$ and $\mathbb{E}(\Omega_{\tilde{w}u}) = [\Omega_{vu}, 0]'$, leading to $(\mathbb{E}(\Omega_{\tilde{w}\tilde{w}}))^{-1}\mathbb{E}(\Omega_{\tilde{w}u})$ also equal to $[\Omega_{vv}^{-1}\Omega_{vu}, 0]' = \Theta_{[1:2]}(r)$.

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

The paper is organized as follows: Section 2 presents the setting, the assumptions, and the theoretical results. Section 3 briefly summarizes. The proofs of the main theorems are relegated to Appendix A. Five additional appendices are available in the Supplementary Material: Appendix B of the Supplementary Material describes for completeness, and as reference point, the FM-CPR estimator of Wagner and Hong (2016) for the case of one integrated regressor and its powers considered in the main text. Appendix C of the Supplementary Material provides a more detailed discussion of the main ingredients of the robustness result from an asymptotic perspective. Appendix D of the Supplementary Material contains the proofs of two auxiliary lemmas. Appendix E of the Supplementary Material illustrates the necessary modifications of the main arguments of the proofs to cover the case of multiple integrated regressors and their powers. Appendix F of the Supplementary Material, as mentioned above, contains a selection of results from a simulation study assessing the finite sample differences between FM-OLS and FM-CPR and test statistics based upon them. Additional supplementary material available upon request provides further simulation results.

We use the following notation: Definitional equality is signified by $:=$, equality in distribution by $\stackrel{d}{=}$, weak convergence by $\Rightarrow$ and convergence in probability by $\stackrel{\mathbb{P}}{\to}$. With $O_{\mathbb{P}}(1)$, $o_{\mathbb{P}}(1)$, and $o_{a.s.}(1)$, we denote boundedness in probability, convergence to zero in probability and convergence to zero almost surely. The integer part of $x \in \mathbb{R}$ is denoted by $\lfloor x \rfloor$ and a diagonal matrix, with entries specified throughout, by $\mathrm{diag}(\cdot)$. For a vector $x = (x_i)_{i=1,\ldots,n}$, we denote its Euclidean norm with $\|x\| := \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$. For a matrix A, the (i,j) element is denoted with $A_{(i,j)}$, its jth column is labeled as $A_{(\cdot,j)}$ and $0_{m\times n}$ denotes an $(m \times n)$ matrix with all entries equal to zero. We use $\mathbb{E}$ to denote expectation and L to denote the backward-shift operator, that is, $L\{x_t\}_{t\in\mathbb{Z}} = \{x_{t-1}\}_{t\in\mathbb{Z}}$. The first-difference operator is denoted with Δ, that is, $\Delta := 1 - L$. For two vector-valued continuous semi-martingales $X(r)$, $Y(r)$, $r \in [0,1]$, we define the quadratic covariation $\langle X(r), Y(r)\rangle_0^t := X(t)Y(t)' - X(0)Y(0)' - \int_0^t X(r)dY(r)' - (\int_0^t Y(r)dX(r)')'$, $t \in [0,1]$. Brownian motions, with covariance matrices specified in the context, are denoted by $B(r)$ and standard Brownian motions by $W(r)$.

2. THEORY

2.1. Setup and Assumptions

We consider a CPR with only one integrated regressor and its powers:[8]

$$y_t = D_t'\delta + X_t'\beta + u_t, \quad \text{for } t = 1, \ldots, T, \tag{1}$$

[8]Not all consecutive powers of x_t need to be included. In the multiple integrated regressor case, the included powers of the integrated regressors may differ across integrated regressors. See Appendix E of the Supplementary Material for a discussion of the required modifications to the mathematical arguments for extending the results of this paper to the case of multiple integrated regressors and their powers.

What is, however, key for the robustness result for FM-OLS is that the integrated variable x_t is itself included in the regression, see also the discussion and illustrative example in (C.9) in Appendix C of the Supplementary Material. The initial value x_0 is allowed to be any well-defined $O_{\mathbb{P}}(1)$ random variable.

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

$$x_t = x_{t-1} + v_t, \tag{2}$$

where y_t is a scalar process, $D_t \in \mathbb{R}^q$ is a deterministic component, x_t is a scalar $I(1)$ process and $X_t := [x_t, x_t^2, \dots, x_t^p]' \in \mathbb{R}^p$. Denoting with $Z_t := [D_t', X_t']' \in \mathbb{R}^{q+p}$ the stacked regressor vector and with $\theta := [\delta', \beta']' \in \mathbb{R}^{q+p}$ the parameter vector, Equation (1) can be rewritten more compactly as $y_t = Z_t'\theta + u_t$ for $t = 1, \dots, T$. The precise assumptions concerning the deterministic component D_t, the regressor x_t, and the errors u_t are given next.

Assumption 2.1. For the deterministic component, there exists a sequence of $q \times q$ scaling matrices $G_D = G_D(T)$ and a q-dimensional vector of càdlàg functions $D(s)$, with $0 < \int_0^s D(z)D(z)'dz < \infty$ for $0 < s \leq 1$, such that for $0 \leq s \leq 1$ it holds that $\lim_{T\to\infty} T^{1/2} G_D D_{\lfloor sT \rfloor} = D(s)$.

For the leading case of polynomial time trends, that is, $D_t = [1, t, t^2, \dots, t^{q-1}]'$, clearly $G_D = \mathrm{diag}(T^{-1/2}, T^{-3/2}, T^{-5/2}, \dots, T^{-(q-1/2)})$ and $D(s) = [1, s, s^2, \dots, s^{q-1}]'$.[9]

Assumption 2.2. The processes $\{u_t\}_{t\in\mathbb{Z}}$ and $\{\Delta x_t\}_{t\in\mathbb{Z}} = \{v_t\}_{t\in\mathbb{Z}}$ are generated as $u_t = C_u(L)\zeta_t = \sum_{j=0}^{\infty} c_{uj}\zeta_{t-j}$ and $\Delta x_t = v_t = C_v(L)\varepsilon_t = \sum_{j=0}^{\infty} c_{vj}\varepsilon_{t-j}$, with $\sum_{j=0}^{\infty} j|c_{uj}| < \infty$, $\sum_{j=0}^{\infty} j|c_{vj}| < \infty$ and $C_v(1) \neq 0$. Furthermore, we assume that the process $\{\xi_t^0\}_{t\in\mathbb{Z}} := \{[\zeta_t, \varepsilon_t]'\}_{t\in\mathbb{Z}}$ is a sequence of independently and identically distributed random variables with $\mathbb{E}(\|\xi_t^0\|^l) < \infty$ for some $l > \max(8, 4/(1-2b))$ with $0 < b < 1/3$ and positive definite covariance matrix $\Sigma_{\xi^0\xi^0}$.

The moment conditions and i.i.d. assumption stated in Assumption 2.2 are stronger than in the corresponding Assumption 1 in Wagner and Hong (2016), which only requires finite fourth (conditional) moments in a martingale difference sequence framework. The strengthening allows us to draw upon some results of Kasparis (2008).[10] For univariate $\{x_t\}_{t\in\mathbb{Z}}$, the assumption $C_v(1) \neq 0$ excludes stationary $\{x_t\}_{t\in\mathbb{Z}}$ and has to be modified in the multivariate case to $\det(C_v(1)) \neq 0$, that is, in the multivariate case (as, e. g., in the discussion in Appendix E of the Supplementary Material), the vector process $\{x_t\}_{t\in\mathbb{Z}}$ is assumed to be non-cointegrated. For long-run covariance estimation, we posit the following assumptions concerning kernel and bandwidth.

Assumption 2.3. The kernel function $k(\cdot)$ satisfies:

1. $k(0) = 1$, $k(\cdot)$ is continuous at 0 and $\bar{k}(0) := \sup_{x\geq 0} |k(x)| < \infty$.
2. $\int_0^{\infty} \bar{k}(x)dx < \infty$, where $\bar{k}(x) := \sup_{y\geq x} |k(y)|$.

[9] In the EKC literature, the deterministic component typically consists of an intercept and a linear trend, with the latter intended to capture autonomous energy efficiency increases.

[10] Note that Kasparis (2008, Assumption 1(b), p. 1376) posits the condition $l > \min(8, 4/(1-2b))$. In the proof of his Lemma A1, however, at different places moments of order $4/(1-2b)$ (p. 1391) and order 8 (p. 1395) are needed. Thus, we think that the minimum should be replaced by the maximum. Since we rely upon similar arguments in the proof of our Lemma A.3, we require moments of order $\max(8, 4/(1-2b))$.

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

Assumption 2.4. The bandwidth parameter $M_T \to \infty$ fulfills $M_T = O(T^b)$, for the same parameter b as in Assumption 2.2.

The bandwidth Assumption 2.4 implies $\lim_{T\to\infty}(M_T^{-1} + T^{-1/3}M_T) = 0$, whereas Jansson (2002) assumes $\lim_{T\to\infty}(M_T^{-1} + T^{-1/2}M_T) = 0$, which corresponds to $M_T = O(T^b)$, with $0 < b < 1/2$. Thus, we require a tighter upper bound on the bandwidth. This stems from the fact that in the asymptotic analysis of the "formal" FM-OLS estimator kernel "long-run covariance" estimators involving (properly scaled) powers of integrated processes need to be analyzed. To have uniform notation for kernel-weighted sums irrespective of the properties of the sequences considered, we use the following definition.

DEFINITION 2.5. *For two sequences $\{a_t\}_{t=1,\dots,T}$ and $\{b_t\}_{t=1,\dots,T}$, we define*[11]

$$\hat{\Delta}_{ab} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} a_t b'_{t+h}, \tag{3}$$

neglecting the dependence on $k(\cdot)$, M_T and the sample range $1,\dots,T$ for brevity. Furthermore, $\hat{\Omega}_{ab} := \hat{\Delta}_{ab} + \hat{\Delta}'_{ab} - \hat{\Sigma}_{ab}$, with $\hat{\Sigma}_{ab} := \frac{1}{T}\sum_{t=1}^{T} a_t b'_t$. Based on these quantities, we define $\hat{\Delta}^+_{ab} := \hat{\Delta}_{ab} - \hat{\Delta}_{aa}\hat{\Omega}_{aa}^{-1}\hat{\Omega}_{ab}$ and $\hat{\omega}_{a\cdot b} := \hat{\Omega}_{aa} - \hat{\Omega}_{ab}\hat{\Omega}_{bb}^{-1}\hat{\Omega}_{ba}$.

In case $\{a_t\}_{t\in\mathbb{Z}}$ and $\{b_t\}_{t\in\mathbb{Z}}$ are jointly stationary processes with finite half long-run covariance matrix $\Delta_{ab} := \sum_{h=0}^{\infty} \mathbb{E}(a_0 b'_h)$, then under appropriate assumptions $\hat{\Delta}_{ab}$ is a consistent estimator of Δ_{ab}, with a similar result holding for $\Sigma_{ab} := \mathbb{E}(a_0 b'_0)$ and a fortiori for $\Omega_{ab} := \sum_{h=-\infty}^{\infty} \mathbb{E}(a_0 b'_h)$.

Remark 2.6. Note that in our definition of $\hat{\Delta}_{ab}$ in (3) we use the bandwidth M_T (like, e. g., Phillips, 1995) rather than $T-1$ (like, e. g., Jansson, 2002) as upper bound of the summation over the index h. For truncated kernels, with $k(x) = 0$ for $|x| > 1$, this is of course inconsequential. It can also be shown (based on, e. g., Jansson, 2002) that for "standard" long-run covariance estimation problems, consistency is not affected by the summation index choice, M_T or $T-1$, for untruncated kernels like the Quadratic Spectral kernel either. In our setting, however, where we analyze the asymptotic behavior of $\hat{\Delta}$ quantities for (properly scaled) nonstationary processes in Theorem 2.7, the summation bound is important and our proof of Theorem 2.7 hinges upon summation only up to M_T. More specifically, we rely upon the summation bound M_T in the proof of Lemma A.3, which is related to Kasparis (2008, Lem. A1, pp. 1394–1396), who also uses M_T (in a slightly different context).

Assumption 2.2 implies that the process $\{\xi_t\}_{t\in\mathbb{Z}} := \{[u_t, v_t]'\}_{t\in\mathbb{Z}}$ fulfills a functional central limit theorem of the form

[11]The standard notation for half long-run covariance matrices is Δ and, therefore, we also use this letter. We are confident that no confusion with the first difference operator, also labeled Δ, arises.

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

$$\frac{1}{T^{1/2}}\sum_{t=1}^{\lfloor rT\rfloor}\xi_t \Rightarrow B(r)=\begin{bmatrix}B_u(r)\\ B_v(r)\end{bmatrix}=\Omega_{\xi\xi}^{1/2}W(r),\quad r\in[0,1], \tag{4}$$

with the covariance matrix $\Omega_{\xi\xi}>0$ of $B(r)$ given by the long-run covariance matrix of $\{\xi_t\}_{t\in\mathbb{Z}}$, that is,

$$\Omega_{\xi\xi}:=\begin{bmatrix}\Omega_{uu} & \Omega_{uv}\\ \Omega_{vu} & \Omega_{vv}\end{bmatrix}=\sum_{h=-\infty}^{\infty}\mathbb{E}(\xi_0\xi_h'). \tag{5}$$

Later, we will also need the corresponding half long-run covariance matrix $\Delta_{\xi\xi}:=\sum_{h=0}^{\infty}\mathbb{E}(\xi_0\xi_h')$, partitioned similarly to $\Omega_{\xi\xi}$. As is well known, FM-type estimation requires estimates of the half long-run and long-run covariances Δ and Ω. With $\Omega=\Delta+\Delta'-\Sigma$ holding by definition, we focus below on the estimation of Δ and Σ. For actual calculations, the unobserved errors u_t are furthermore replaced by the OLS residuals $\hat{u}_t$ from (1), that is, by $\hat{u}_t:=y_t-Z_t'\hat{\theta}$ with $\hat{\theta}:=(Z'Z)^{-1}Z'y$. This defines $\hat{\xi}_t:=[\hat{u}_t,v_t]'$.

2.2. FM-OLS in Cointegrating Polynomial Regressions

Performing "formal" FM-OLS estimation à la Phillips and Hansen (1990) of θ in (1) amounts to (mis-)treating $X_t=[x_t,\ldots,x_t^p]'$ as p (non-cointegrated) integrated regressors—rather than, as would be correct, as p consecutive powers of a single integrated regressor. This "interpretation" implies that, instead of (1) and (2), one considers:

$$\begin{aligned}y_t&=D_t'\delta+X_t'\beta+u_t,\\ X_t&=X_{t-1}+w_t,\end{aligned}$$

which defines $w_t:=\Delta X_t$ as

$$w_t=\begin{bmatrix}\Delta x_t\\ \Delta x_t^2\\ \Delta x_t^3\\ \vdots\\ \Delta x_t^p\end{bmatrix}=\begin{bmatrix}v_t\\ 2x_tv_t-v_t^2\\ 3x_t^2v_t-3x_tv_t^2+v_t^3\\ \vdots\\ -\sum_{k=1}^{p}\binom{p}{k}x_t^{p-k}(-v_t)^k\end{bmatrix}\simeq\begin{bmatrix}v_t\\ 2x_tv_t-v_t^2\\ 3x_t^2v_t-3x_tv_t^2\\ \vdots\\ px_t^{p-1}v_t-\frac{p(p-1)}{2}x_t^{p-2}v_t^2\end{bmatrix}, \tag{6}$$

with $\simeq$ indicating that, for the subsequent asymptotic analysis, only the first two terms in the (binomial) expansion of Δx_t^i, for $i\geq 2$, are relevant; to be precise, only the first term for Theorem 2.7 and the first two for Theorem 2.9.

The "formal" FM-OLS estimator of θ is defined as

$$\hat{\theta}^{++}:=(Z'Z)^{-1}(Z'y^{++}-A^{**}), \tag{7}$$

with $y^{++}:=[y_1^{++},\ldots,y_T^{++}]'$, where $y_t^{++}:=y_t-\Delta X_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}=y_t-w_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}$, $Z:=[Z_1,\ldots,Z_T]'$, $Z_t:=[D_t',X_t']'$ and

$$A^{**} := \begin{bmatrix} 0_{q\times 1} \\ T\hat{\Delta}^{+}_{wu} \end{bmatrix} = \begin{bmatrix} 0_{q\times 1} \\ T(\hat{\Delta}_{wu} - \hat{\Delta}_{ww}\hat{\Omega}^{-1}_{ww}\hat{\Omega}_{wu}) \end{bmatrix}. \quad \textbf{(8)}$$

With w_t containing products of powers of the integrated process x_t and of v_t, $\hat{\Omega}_{ww}$, $\hat{\Omega}_{wu}$, $\hat{\Delta}_{ww}$, $\hat{\Delta}_{wu}$, and $\hat{\Delta}^{+}_{wu}$ have to be interpreted in the sense of Definition 2.5.

Deriving the asymptotic distribution of $\hat{\theta}^{++}$ defined in (7) crucially rests upon understanding the asymptotic behavior of two quantities: (i) of the properly scaled long-run covariance estimators, for example, $\hat{\Delta}_{ww}$, derived in Theorem 2.7, and (ii) of the properly scaled product $Z'y^{++}$, respectively, after centering the properly scaled product $Z'u^{++}$, with $u^{++} := [u_1^{++},\dots,u_T^{++}]'$ and $u_t^{++} := u_t - w_t'\hat{\Omega}^{-1}_{ww}\hat{\Omega}_{wu}$, derived in Theorem 2.9.

The representation of w_t in (6) indicates that the (re)scaling required to establish convergence will require considering the process $\tilde{w}_t := \left[v_t, \frac{\Delta x_t^2}{T^{1/2}}, \dots, \frac{\Delta x_t^p}{T^{\frac{p-1}{2}}}\right]'$, that is, $\tilde{w}_t = G_W w_t$, with $G_W := \mathrm{diag}(1, T^{-1/2}, \dots, T^{-(p-1)/2})$ instead of w_t. Define $\hat{\eta}_t := [\hat{u}_t, \tilde{w}_t']'$, with $\hat{u}_t$ the OLS residuals from (1). The first theorem establishes the asymptotic properties of the (formal) long-run covariance estimators.

THEOREM 2.7. *Let the data be generated by (1) and (2) under Assumptions 2.1 and 2.2 and let long-run covariance estimation be performed under Assumptions 2.3 and 2.4. Then, it holds for $T \to \infty$ that*

$$\hat{\Delta}_{\eta\eta} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\hat{\eta}_t\hat{\eta}_{t+h}' \Rightarrow \Delta_{\eta\eta} := \begin{bmatrix} \Delta_{uu} & \Delta_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr \\ \Delta_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr & \Delta_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr \end{bmatrix}, \quad \textbf{(9)}$$

with $\dot{\mathbf{B}}_v(r) := \left[1, 2B_v(r), \dots, pB_v^{p-1}(r)\right]'$. Furthermore, it holds for $T \to \infty$ that

$$\hat{\Sigma}_{\eta\eta} := \frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t' \Rightarrow \Sigma_{\eta\eta} := \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr \\ \Sigma_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr & \Sigma_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr \end{bmatrix}. \quad \textbf{(10)}$$

Combining the above two results leads to $\hat{\Omega}_{\eta\eta} := \hat{\Delta}_{\eta\eta} + \hat{\Delta}'_{\eta\eta} - \hat{\Sigma}_{\eta\eta} \Rightarrow \Delta_{\eta\eta} + \Delta'_{\eta\eta} - \Sigma_{\eta\eta} =: \Omega_{\eta\eta}$, with

$$\Omega_{\eta\eta} = \begin{bmatrix} \Omega_{uu} & \Omega_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr \\ \Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr & \Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr \end{bmatrix} = \langle \mathcal{B}(r), \mathcal{B}(r)\rangle_0^1, \quad \textbf{(11)}$$

where $\mathcal{B}(r) := [B_u(r), \mathbf{B}_v(r)']'$ and $\boldsymbol{B}_v(r) := [B_v(r), B_v^2(r), \dots, B_v^p(r)]'$.

Note that, by definition, the upper 2×2 blocks of these limits correspond to the half long-run covariance matrix, the covariance matrix and the long-run covariance matrix of $\{\xi_t\}_{t\in\mathbb{Z}}$.[12]

[12]Considering the re-scaled process $\tilde{w}_t$ implies that the formal (half) long-run covariance estimators, for example, $\hat{\Delta}_{\eta\eta}$, converge without further scaling. If one wants to instead highlight the necessary re-scaling of the formal long-run

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

Remark 2.8. The kernel-weighted sum result established in Theorem 2.7 is related to an early result of Phillips (1991b). Upon some approximations detailed in the proof and considering here the case of several integrated regressors (see Appendix E of the Supplementary Material for details and notation), a typical element of $\hat{\Delta}_{\tilde{w}\tilde{w}}$ is given by $\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\frac{d_k x_{kt}^{d_k-1}}{T^{(d_k-1)/2}}\frac{d_l x_{l,t+h}^{d_l-1}}{T^{(d_l-1)/2}} v_{kt}v_{l,t+h} + o_{\mathbb{P}}(1) \Rightarrow \Delta_{v_k v_l} d_k d_l \int_0^1 B_{v_k}^{d_k-1}(r) B_{v_l}^{d_l-1}(r)dr$, with x_{kt} and x_{lt} two integrated regressors. Phillips (1991b) considers spectral estimation of cointegrating linear relationships. The spectral estimator, based on (formal) spectral density estimators at frequency zero, involves kernel-weighted sums of products of integrated processes as well as of products of integrated and stationary processes, see (A. 11) to (A. 13) in the appendix of Phillips (1991b). Using our notation and considering half long-run covariance estimation, Phillips (1991b) shows that $\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\frac{x_{kt}}{T^{1/2}}\frac{x_{l,t+h}}{T^{1/2}} \Rightarrow \int_0^1 k(r)dr \int_0^1 B_k(r)B_l(r)dr$.

To derive the asymptotic distribution of properly scaled $Z'u^{++}$, define the scaling matrix $G := \operatorname{diag}(G_D, G_X)$ with G_D as in Assumption 2.1 and $G_X := \operatorname{diag}(T^{-1}, T^{-3/2}, \ldots, T^{-(p+1)/2})$:

$$\begin{aligned} GZ'u^{++} &= GZ'(u - W\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}) \\ &= GZ'u - GZ'W\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \\ &= GZ'u - GZ'WG_W G_W^{-1}\hat{\Omega}_{ww}^{-1}G_W^{-1}G_W\hat{\Omega}_{wu} \\ &= GZ'u - GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u}, \end{aligned} \tag{12}$$

with $u := [u_1, \ldots, u_T]'$, $W := [w_1, \ldots, w_T]'$, and $\tilde{W} = [\tilde{w}_1, \ldots, \tilde{w}_T]' := WG_W$. The asymptotic behavior of $GZ'u$ is well understood (see, e. g., (A.3) in the proof of Proposition 1 in Wagner and Hong, 2016). Since the asymptotic behavior of $\hat{\Omega}_{\tilde{w}\tilde{w}}$ and $\hat{\Omega}_{\tilde{w}u}$ has been derived in Theorem 2.7, the only term that remains to be investigated is $GZ'WG_W = GZ'\tilde{W}$.[13]

THEOREM 2.9. *Under Assumptions 2.1 and 2.2, it holds for $T \to \infty$ that*

$$\begin{aligned} GZ'\tilde{W} &\Rightarrow \int_0^1 J(r)d\mathbf{B}_v(r)' + \begin{bmatrix} 0_{q\times p} \\ \Delta_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr \end{bmatrix} \\ &= \int_0^1 J(r)d\mathbf{B}_v(r)' + \begin{bmatrix} 0_{q\times p} \\ \Delta_{vv}\Omega_{vv}^{-1}\langle \mathbf{B}_v(r), \mathbf{B}_v(r)\rangle_0^1 \end{bmatrix}, \end{aligned} \tag{13}$$

with $J(r) := [D(r)', \boldsymbol{B}_v(r)']'$.

covariance estimators by considering w_t itself, the first result of Theorem 2.7, for example, can trivially be rewritten by defining $\hat{\xi}_t^* := [\hat{u}_t, w_t']'$ and $H_W := \operatorname{diag}(1, G_W)$. Then it holds that $H_W\hat{\Delta}_{\xi^*\xi^*}H_W \Rightarrow \Delta_{\eta\eta}$.

[13]Note that the asymptotic behavior of the first column of the result in Theorem 2.9, that is, the limit of $GZ'v$, with $v := [v_1, \ldots, v_T]'$, corresponding to the first component v_t of $\tilde{w}_t$ (similar to the limit of $GZ'u$), is also already known, compare, for example, Wagner and Hong (2016, p. 1312).

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

Remark 2.10. The above result generalizes the usual functional central limit theorem applied in unit root and cointegration analysis, for example, $\frac{1}{T}\sum_{t=1}^{T} x_t v_t \Rightarrow \int_0^1 B_v(r)dB_v(r) + \Delta_{vv}$ in cointegrating linear regressions, in two ways. The first is the consideration of nonlinear functions—in the present paper polynomials—of integrated processes as integrands. Nonlinear functions of integrated processes as integrands are by now, of course, standard in the nonlinear cointegrating regression literature (see, e. g., the references in Wagner and Hong, 2016). The second generalization is much less common in the unit root and cointegration context: The integrator, usually $dB_v(r)$, is replaced (in this paper) by the vector composed of $dB_v^j(r), j = 1, \ldots, p$, that is, by $d\mathbf{B}_v(r) = [dB_v(r), dB_v^2(r), \ldots, dB_v^p(r)]'$. Since $\mathbf{B}_v(r)$ is a continuous semi-martingale, stochastic integration with respect to $d\mathbf{B}_v(r)$ is well defined.[14] The proof of Theorem 2.9 shows that this result follows from the fact that, asymptotically, only the first two terms in the binomial representation of w_t given in (6) matter. These two terms are related asymptotically to the first two formal derivative vectors of $\mathbf{B}_v(r)$, see (A.11)–(A.13) in Appendix A, which allows deriving the result using Itô's Lemma. The arguments are also illustrated in Appendix C of the Supplementary Material in (C.6) and (C.7).

We have now collected the necessary ingredients to establish the robustness result for the "formal" FM-OLS estimator $\hat{\theta}^{++}$.

PROPOSITION 2.11. *Let the data be generated by (1) under Assumptions 2.1 and 2.2. Furthermore, let long-run covariance estimation be performed under Assumptions 2.3 and 2.4. Then, it holds for $T \to \infty$ that*

$$G^{-1}(\hat{\theta}^{++} - \theta) \Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \int_0^1 J(r)dB_{u\cdot v}(r), \tag{14}$$

with $B_{u\cdot v}(r) := B_u(r) - B_v(r)\Omega_{vv}^{-1}\Omega_{vu}$ a Brownian motion with variance $\omega_{u\cdot v} := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$. Therefore, the FM-OLS estimator and the FM-CPR estimator of Wagner and Hong (2016), see (B.3) in Appendix B of the Supplementary Material, have the same limiting distribution.

Remark 2.12. Two aspects are worth mentioning in relation to the above result: First, to establish robustness of FM-OLS, a weaker version of the result in Theorem 2.9 would be sufficient, that is, it would be sufficient to derive the limiting distribution of the first column of $GZ'\tilde{W}$ and show that all other columns are $O_{\mathbb{P}}(1)$. This stems from the fact that $\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \xrightarrow{\mathbb{P}} [\Omega_{vv}^{-1}\Omega_{vu}, 0_{p-1}']'$. The full version of the result of Theorem 2.9 is, for example, needed in Appendix C of the Supplementary Material, where we consider FM-OLS in an example of a CPR where the regressor x_t itself is not included and FM-OLS does not lead to a zero mean Gaussian mixture

[14]Thus, the generalization consists, for polynomials in this paper and ignoring the additive bias term, in arriving at $\int_0^1 F(B_v(r))dF(B_v(r))$ instead of, as usual in the cointegration literature, at $\int_0^1 F(B_v(r))dB_v(r)$, with $F(B_v(r)) = [B_v(r), \ldots, B_v^p(r)]' = \mathbf{B}_v(r)$.

limiting distribution, see (C.9). Second, for the additive correction term A^{**} defined in (8), it holds even when x_t is not included as regressor and thus $\dot{\mathbf{B}}_v(r)$ does not contain the element 1, that $GA^{*+} \Rightarrow \Delta_{vu}^{+}[0_{1\times q}, \int_0^1 \dot{\mathbf{B}}_v(r)'dr]'$.

Upon establishing that the scalar formal long-run variance estimator $\hat{\omega}_{u\cdot w} := \hat{\Omega}_{uu} - \hat{\Omega}_{uw}\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}$ *also* converges to $\omega_{u\cdot v}$, included for brevity in the proof of Proposition 2.11, the result in Proposition 2.11 implies that the Wald-type test statistic (compare Proposition 2 of Wagner and Hong, 2016) for the null hypothesis $R\theta = r$, that is,[15]

$$T_W^{++} := \left(R\hat{\theta}^{++} - r\right)' \left[\hat{\omega}_{u\cdot w} R(Z'Z)^{-1}R'\right]^{-1} \left(R\hat{\theta}^{++} - r\right)', \tag{15}$$

is (under the constraints on R discussed in Wagner and Hong, 2016) asymptotically chi-square distributed under the null hypothesis. The result also implies that the Shin (1994)-type test for the null hypothesis of cointegration, that is,

$$\mathrm{CT}^{++} := \frac{1}{T\hat{\omega}_{u\cdot w}} \sum_{t=1}^{T} \left(\frac{1}{T^{1/2}} \sum_{j=1}^{t} \hat{u}_j^{++} \right)^2, \tag{16}$$

with $\hat{u}_t^{++} := y_t - Z_t'\hat{\theta}^{++}$, converges under the null hypothesis to the limiting distribution given in Wagner and Hong (2016, Prop. 5).

When using the Shin (1994)-type test statistic CT^{++} one has to be careful to use the correct critical values corresponding to the (specification-dependent) limit process (as discussed and tabulated in Wagner, 2023), see, for example, (B.5) and (B.6) in Appendix B of the Supplementary Material. In relation to the quadratic example without deterministic components discussed in the introduction, the differences in the critical values stem from considering either $J^W(r) = [W_1(r), W_2(r)]'$ with two independent standard Brownian motions $W_1(r)$ and $W_2(r)$, leading to the (for the CPR case) incorrect Shin (1994) critical values, or $J^W(r) = [W(r), W(r)^2]'$ with standard Brownian motion $W(r)$, leading to the correct Wagner (2023) critical values. Table 1 (see the upper left block for $D_t = \emptyset$ and $p = 2$ for the example from the introduction) illustrates that these differences become the more pronounced the more complex the CPR model is. Thus, using the FM-OLS residuals in conjunction with the Shin (1994) critical values leads to invalid inference with respect to the null hypothesis of cointegration even asymptotically.

3. SUMMARY

The paper has shown that the textbook FM-OLS estimator of Phillips and Hansen (1990) is robust to being used in CPRs. Using FM-OLS in CPRs refers to a widespread practice in the empirical literature to treat all integrated variables and their powers, incorrectly, as a vector of integrated variables and to use textbook

[15] An analogous result also holds for the LM-type test statistic considered in Wagner and Hong (2016, Prop. 4) in the context of specification testing.

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

Table 1. Critical values for the Shin (1994, Table 1) test for p integrated regressors and for the CT test for cointegration in the single integrated regressor CPR model of degree p from Wagner (2023, Table 6). The three block-columns correspond to the cases without deterministic component ($D_t = \emptyset$), with intercept only ($D_t = 1$) and with intercept and linear trend ($D_t = [1,t]'$).

	$D_t = \emptyset$			$D_t = 1$			$D_t = [1,t]'$		
α	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Two integrated regressors/quadratic specification ($p = 2$)									
Shin	0.624	0.895	1.623	0.163	0.221	0.380	0.081	0.101	0.150
CT	0.664	0.947	1.712	0.213	0.293	0.504	0.086	0.106	0.157
Panel B: Three integrated regressors/cubic specification ($p = 3$)									
Shin	0.475	0.682	1.305	0.121	0.159	0.271	0.069	0.085	0.126
CT	0.561	0.804	1.473	0.204	0.281	0.490	0.081	0.101	0.150

FM-OLS accordingly. Robustness means that this "formal" FM-OLS practice leads to a zero mean Gaussian mixture limiting distribution that coincides with the limiting distribution of the Wagner and Hong (2016) application of the FM estimation principle to the CPR case.

From a limit Hilbert space geometry perspective, this result is not surprising, since in order to arrive at a zero mean Gaussian mixture limiting distribution, all that is required is to ensure a zero long-run covariance between the (modified) errors and the errors generating the integrated regressors. In the words of Phillips (1991a) "*...what is important in estimation and inference in cointegrated systems, at least as far as ensuring the applicability of local asymptotic mixed normality theory, is not the precise form of the specification but the information concerning the presence of unit roots that is employed in estimation*." Nevertheless, to put this observation to work in the CPR context requires to establish two novel limit results. Theorem 2.7 establishes the asymptotic behavior of kernel-weighted sums involving first differences of powers of integrated processes. Theorem 2.9 derives a functional central limit theorem involving products of powers of integrated processes with first differences of powers of integrated processes. The combination of the two results, where in particular the algebraic structure of the result of Theorem 2.7 that mimics the limit Hilbert space projection geometry is key, leads to the robustness result for the "formal" FM-OLS estimator.

APPENDIX A. Proofs of the Main Results

The proofs of Theorems 2.7 and 2.9 rely upon three lemmas that we state first: Lemma A.1 is identical to Kasparis (2008, Lem. A1(i)) and Lemmas A.2 and A.3 draw upon some ideas used in the proof of Kasparis (2008, Lem. A1). Appendix D of the Supplementary Material contains proofs of Lemmas A.2 and A.3.

LEMMA A.1. *Under Assumption 2.2, it holds for* $0 \le b < 1/3$ *that*

$$\sup_{r\in[0,1]} T^{-1/2} \sum_{h=0}^{T^b} |v_{\lfloor rT \rfloor + h}| = o_{a.s.}(1). \tag{A.1}$$

LEMMA A.2. *Under Assumptions 2.2–2.4, it holds for all integers* $0 \le p$ *and* $1 \le q$ *that*

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

LEMMA A.3. *Under Assumptions 2.2–2.4, it holds for all integers* $0 \le p$ *that*

$$\left| \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p \left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right) \right| = o_{\mathbb{P}}(1). \tag{A.3}$$

Proof of Theorem 2.7. The proof proceeds in two steps: First, the results are shown for the infeasible estimators $\tilde{\Delta}_{\eta\eta}$ and $\tilde{\Sigma}_{\eta\eta}$ that are computed from $\eta_t := [u_t, \tilde{w}_t']'$ rather than from $\hat{\eta}_t$. Thereafter, it will be shown that the results continue to hold when u_t is replaced by $\hat{u}_t$. Clearly, it suffices to discuss the result for $\hat{\Delta}_{\eta\eta}$, the result for $\hat{\Sigma}_{\eta\eta}$ follows analogously, but more easily, as it deals with the $h = 0$ term only.

First, the $(1,1)$ element of $\tilde{\Delta}_{\eta\eta}$ is given by $\left(\tilde{\Delta}_{\eta\eta}\right)_{(1,1)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} u_t u_{t+h}$. For the first and second columns (and rows), that is, for:

$$\left(\tilde{\Delta}_{\eta\eta}\right)_{(i+1,1)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \frac{\Delta x_t^i}{T^{\frac{i-1}{2}}} u_{t+h},$$

$$\left(\tilde{\Delta}_{\eta\eta}\right)_{(i+1,2)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \frac{\Delta x_t^i}{T^{\frac{i-1}{2}}} v_{t+h},$$

for $i = 1, \dots, p$, exactly the same arguments apply due to the assumptions on $\{u_t\}_{t\in\mathbb{Z}}$ and $\{v_t\}_{t\in\mathbb{Z}}$. Therefore, it is sufficient in the subsequent discussion to consider the lower right $p \times p$ block of the estimator $\tilde{\Delta}_{\eta\eta}$, which is given by $\tilde{\Delta}_{\tilde{w}\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \tilde{w}_t \tilde{w}_{t+h}'$. Note that

$$\begin{aligned} \frac{\Delta x_t^i}{T^{(i-1)/2}} &= -\frac{1}{T^{(i-1)/2}} \sum_{k=1}^{i} \binom{i}{k} x_t^{i-k} (-v_t)^k \\ &= i\left(\frac{x_t}{T^{1/2}}\right)^{i-1} v_t - \sum_{k=2}^{i} \binom{i}{k} (-1)^k \left(\frac{x_t}{T^{1/2}}\right)^{i-k} \left(\frac{v_t}{T^{1/2}}\right)^{k-2} \frac{v_t^2}{T^{1/2}}. \end{aligned}$$

Lemma A.1 shows that $\sup_{r\in[0,1]} T^{-1/2} v_{\lfloor rT \rfloor} = o_{a.s.}(1)$. Additionally, $\sup_{r\in[0,1]} T^{-1/2} |x_{\lfloor rT \rfloor}| \le C + o_{a.s.}(1)$. Convergence of $\mathbb{E}[T^{-1/2} v_{\lfloor rT \rfloor}^2] = T^{-1/2} \Sigma_{vv} \to 0$ for all $r \in [0,1]$ implies that $\frac{\Delta x_t^i}{T^{(i-1)/2}} = i\left(\frac{x_t}{T^{1/2}}\right)^{i-1} v_t + O_{\mathbb{P}}(T^{-1/2})$. The kernel is bounded and $M_T = o(T^{1/3})$ by assumption, hence $\tilde{\Delta}_{\tilde{w}\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \dot{X}_{t,T} \dot{X}_{t+h,T}' v_t v_{t+h} +$

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

$o_{\mathbb{P}}(1)$, where $\dot{X}_{t,T} := G_W \dot{X}_t$, with $\dot{X}_t := [1,\ldots,px_t^{p-1}]'$. Clearly, the upper left element of the above term converges in probability to Δ_{vv}. Lemma A.2 implies that $\tilde{\Delta}_{\tilde{w}\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \dot{X}_{t,T} \dot{X}_{t,T}' v_t v_{t+h} + o_{\mathbb{P}}(1)$ and Lemma A.3 establishes that $\tilde{\Delta}_{\tilde{w}\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \frac{1}{T} \sum_{t=1}^{T-h} \dot{X}_{t,T} \dot{X}_{t,T}' + o_{\mathbb{P}}(1)$. Next, similar arguments as in the proof of Jansson (2002, Lem. 6) imply that $\frac{1}{T} \sum_{h=0}^{M_T} \left| k\left(\frac{h}{M_T}\right) \right| |\mathbb{E}[v_0 v_h]| h = o(1)$. In combination with (D.1) in Appendix D of the Supplementary Material, $\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \frac{1}{T} \sum_{t=T-h+1}^{T} \dot{X}_{t,T} \dot{X}_{t,T}' = o_{\mathbb{P}}(1)$ can be established. From this, $\tilde{\Delta}_{\tilde{w}\tilde{w}} = \left(\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \right) \left(\frac{1}{T} \sum_{t=1}^{T} \dot{X}_{t,T} \dot{X}_{t,T}' \right) + o_{\mathbb{P}}(1)$ follows. Convergence of the first term follows immediately, that is, $\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \to \Delta_{vv}$. Using Slutsky's Theorem this establishes that $\tilde{\Delta}_{\tilde{w}\tilde{w}} \Rightarrow \Delta_{vv} \int_0^1 \dot{\mathbf{B}}_v(r) \dot{\mathbf{B}}_v(r)' dr$.

It remains to show that the result can be expressed in terms of quadratic covariation. Given that the elements of the vector process $\mathcal{B}(r)$ are powers of Brownian motions, the process is a continuous semi-martingale and thus its quadratic covariation is well defined. We partition its quadratic covariation matrix as follows:

$$\langle \mathcal{B}(r), \mathcal{B}(r) \rangle_0^1 = \begin{bmatrix} \langle B_u(r), B_u(r) \rangle_0^1 & \langle B_u(r), \mathbf{B}_v(r) \rangle_0^1 \\ \langle \mathbf{B}_v(r), B_u(r) \rangle_0^1 & \langle \mathbf{B}_v(r), \mathbf{B}_v(r) \rangle_0^1 \end{bmatrix}. \tag{A.4}$$

Due to symmetry, only three blocks have to be considered. It is well known (almost by definition) that $\langle B_u(r), B_u(r) \rangle_0^1 = \Omega_{uu}$ and it thus remains to consider the blocks involving $\mathbf{B}_v(r)$. We start with the lower diagonal block of $\Omega_{\eta\eta}$, shown above to equal $\Omega_{vv} \int_0^1 \dot{\mathbf{B}}_v(r) \dot{\mathbf{B}}_v(r)' dr$. The elements of this matrix are given by

$$\Omega_{vv} ij \int_0^1 B_v^{i+j-2}(r) dr, \qquad i,j = 1,\ldots,p. \tag{A.5}$$

Now consider the definition of the quadratic covariation given at the end of the introduction, that is,

$$\langle \mathbf{B}_v(r), \mathbf{B}_v(r) \rangle_0^1 = \mathbf{B}_v(1) \mathbf{B}_v(1)' - \int_0^1 \mathbf{B}_v(r) d\mathbf{B}_v(r)' - \left(\int_0^1 \mathbf{B}_v(r) d\mathbf{B}_v(r)' \right)'. \tag{A.6}$$

Using Itô's Lemma, for example, in the formulation given in Le Gall (2016, Thm. 5.10, p. 113), establishes $\mathbf{B}_v(r) = \int_0^r \dot{\mathbf{B}}_v(s) dB_v(s) + \frac{\Omega_{vv}}{2} \int_0^r \ddot{\mathbf{B}}_v(s) ds$, with $\ddot{\mathbf{B}}_v(r) := [0, 2, \ldots, p(p-1) B_v^{p-2}(r)]'$. Substituting this into (A.6) leads to

$$\begin{aligned} \langle \mathbf{B}_v(r), \mathbf{B}_v(r) \rangle_0^1 = {} & \mathbf{B}_v(1)\, \mathbf{B}_v(1)' - \int_0^1 \left(\mathbf{B}_v(r) \dot{\mathbf{B}}_v(r)' + \dot{\mathbf{B}}_v(r) \mathbf{B}_v(r)' \right) dB_v(r) \\ & - \frac{\Omega_{vv}}{2} \int_0^1 \left(\mathbf{B}_v(r)\, \ddot{\mathbf{B}}_v(r)' + \ddot{\mathbf{B}}_v(r) \mathbf{B}_v(r)' \right) dr. \end{aligned} \tag{A.7}$$

The (i,j) element, $i,j = 1,\ldots,p$, of (A.7) is given by

$$B_v^{i+j}(1) - (i+j) \int_0^1 B_v^{i+j-1}(r) dB_v(r) - \frac{\Omega_{vv}}{2} \left(i(i-1) + j(j-1) \right) \int_0^1 B_v^{i+j-2}(r) dr. \tag{A.8}$$

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

Using Itô's Lemma once again, that is, $B_v^{i+j}(1)-(i+j)\int_0^1 B_v^{i+j-1}(r)dB_v(r)=\frac{\Omega_{vv}}{2}(i+j)(i+j-1)\int_0^1 B_v^{i+j-2}(r)dr$, shows, after simplifying terms, equality of the (i,j) elements in (A.5) and (A.8).

Now consider the off-diagonal block to show equality of $\Omega_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr$ and $\langle B_u(r),\mathbf{B}_v(r)\rangle_0^1$. The definition of $B_{u\cdot v}(r)$ implies $\langle B_u(r),\mathbf{B}_v(r)\rangle_0^1=\langle B_{u\cdot v}(r),\mathbf{B}_v(r)\rangle_0^1+\langle B_v(r),\mathbf{B}_v(r)\rangle_0^1\Omega_{vv}^{-1}\Omega_{vu}$, with the first term on the right-hand side above shown to be zero below. For the second term, the result now immediately follows from above since $B_v(r)$ is the first element of $\mathbf{B}_v(r)$, thus $\langle B_v(r),\mathbf{B}_v(r)\rangle_0^1\Omega_{vv}^{-1}\Omega_{vu}=\Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr\Omega_{vv}^{-1}\Omega_{vu}=\Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)'dr$. To complete the proof, it remains to show that $\langle B_{u\cdot v}(r),\mathbf{B}_v(r)\rangle_0^1=0$. Consider the ith element, $i=1,\ldots,p$, using Itô's Lemma to arrive at the second equality below:

$$\begin{aligned}\langle B_{u\cdot v}(r),B_v^i(r)\rangle_0^1 &= B_{u\cdot v}(1)B_v^i(1)-\int_0^1 B_{u\cdot v}(r)dB_v^i(r)-\int_0^1 B_v^i(r)dB_{u\cdot v}(r)\\ &= B_{u\cdot v}(1)B_v^i(1)-i\int_0^1 B_{u\cdot v}(r)B_v^{i-1}(r)dB_v(r)\\ &\quad -\frac{\Omega_{vv}}{2}i(i-1)\int_0^1 B_{u\cdot v}(r)B_v^{i-2}(r)dr-\int_0^1 B_v^i(r)dB_{u\cdot v}(r).\end{aligned} \tag{A.9}$$

Finally, using once again Itô's Lemma in the formulation of Le Gall (2016, Thm. 5.10, p. 113), with $F(x,y)=xy^i$ and the fact that the quadratic covariation of independent Brownian motions is zero (see, e.g., Proposition 4.16, p. 88 in Le Gall, 2016) show that $B_{u\cdot v}(1)B_v^i(1)=\int_0^1 B_v^i(r)dB_{u\cdot v}(r)+i\int_0^1 B_{u\cdot v}(r)B_v^{i-1}(r)dB_v(r)+\frac{\Omega_{vv}}{2}i(i-1)\int_0^1 B_{u\cdot v}(r)B_v^{i-2}(r)dr$, which upon combining terms establishes the postulated zero quadratic covariation.

This finishes the proof for the infeasible estimators and it remains to show that the results continue to hold with $\hat{u}_t$ in place of u_t. The OLS residuals are given by $\hat{u}_t=u_t-Z_t'(\hat{\theta}-\theta)$, with $\hat{\theta}$ denoting the OLS estimator of the parameters in (1). In analogy to the above derivations, consider the term

$$\begin{aligned}\hat{\Delta}_{u\tilde{w}} &= \sum_{h=0}^{M_T}k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\hat{u}_t\tilde{w}_{t+h}\\ &= \sum_{h=0}^{M_T}k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}u_t\tilde{w}_{t+h} - \sum_{h=0}^{M_T}k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}Z_t'(\hat{\theta}-\theta)\tilde{w}_{t+h}.\end{aligned}$$

The first term has already been shown to converge in distribution to $\Delta_{u\tilde{w}}$. Therefore, it remains to show that the second term is $o_{\mathbb{P}}(1)$. Similar arguments as above imply that

$$\begin{aligned}&\sum_{h=0}^{M_T}k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}Z_t'(\hat{\theta}-\theta)\tilde{w}_{t+h}\\ &= (\hat{\theta}-\theta)'G^{-1}\sum_{h=0}^{M_T}k\left(\frac{h}{M_T}\right)\frac{1}{T^{3/2}}\sum_{t=1}^{T-h}T^{1/2}GZ_t\dot{X}_{t+h,T}v_{t+h}+o_{\mathbb{P}}(1)\\ &= (\hat{\theta}-\theta)'G^{-1}\sum_{h=0}^{M_T}k\left(\frac{h}{M_T}\right)\frac{1}{T^{3/2}}\sum_{t=1}^{T}T^{1/2}GZ_t\dot{X}_{t,T}v_t+o_{\mathbb{P}}(1),\end{aligned} \tag{A.10}$$

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

with $G := \mathrm{diag}(G_D, G_X)$, with G_D as in Assumption 2.1 and $G_X := \mathrm{diag}(T^{-1}, T^{-3/2}, \ldots, T^{-(p+1)/2})$. Since $G(\hat{\theta} - \theta) = O_{\mathbb{P}}(1)$ and $\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T^{3/2}} \sum_{t=1}^{T} T^{1/2} G Z_t \dot{X}_{t,T} v_t = O_{\mathbb{P}}(T^{-1})$, the quantity given in (A.10) is $o_{\mathbb{P}}(1)$. This implies that $\hat{\Delta}_{u\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} u_t \tilde{w}_{t+h} + o_{\mathbb{P}}(1)$, from which the claim follows. □

Proof of Theorem 2.9. Consider $G_X \sum_{t=1}^{T} X_t \Delta X_t' G_W$ and define $\ddot{X}_t := [0, 2, \ldots, p(p-1)x_t^{p-2}]'$. Using this notation, it is straightforward to show that

$$G_X \sum_{t=1}^{T} X_t \Delta X_t' G_W = G_X \sum_{t=1}^{T} X_t \dot{X}_t' v_t G_W - G_X \sum_{t=1}^{T} X_t \ddot{X}_t' \frac{v_t^2}{2} G_W + o_{\mathbb{P}}(1). \tag{A.11}$$

Invoking similar arguments as in Wagner and Hong (2016, Prop. 1), formulated slightly differently here to more directly be able to use Itô's Lemma below, shows for the first term on the right-hand side above that

$$\begin{aligned} G_X \sum_{t=1}^{T} X_t \dot{X}_t' v_t G_W \Rightarrow &\int_0^1 \mathbf{B}_v(r) \dot{\mathbf{B}}_v(r)' dB_v(r) \\ &+ \Delta_{vv} \int_0^1 \dot{\mathbf{B}}_v(r) \dot{\mathbf{B}}_v(r)' dr + \Delta_{vv} \int_0^1 \mathbf{B}_v(r) \ddot{\mathbf{B}}_v(r)' dr. \end{aligned} \tag{A.12}$$

Using Lemma A.3 and the continuous mapping theorem, it moreover follows that

$$G_X \sum_{t=1}^{T} X_t \ddot{X}_t' v_t^2 G_W \Rightarrow \Sigma_{vv} \int_0^1 \mathbf{B}_v(r) \ddot{\mathbf{B}}_v(r)' dr. \tag{A.13}$$

Combining (A.12) and (A.13) leads to

$$\begin{aligned} G_X \sum_{t=1}^{T} X_t \Delta X_t' G_W \Rightarrow &\int_0^1 \mathbf{B}_v(r) \dot{\mathbf{B}}_v(r)' dB_v(r) + \frac{\Omega_{vv}}{2} \int_0^1 \mathbf{B}_v(r) \ddot{\mathbf{B}}_v(r)' dr \\ &+ \Delta_{vv} \int_0^1 \dot{\mathbf{B}}_v(r) \dot{\mathbf{B}}_v(r)' dr. \end{aligned} \tag{A.14}$$

Using Itô's Lemma shows that the sum of the first two terms on the right-hand side of (A.14) equals $\int_0^1 \mathbf{B}_v(r) d\mathbf{B}_v(r)'$. With respect to the third term, Theorem 2.7 shows that it is equal to $\langle \mathbf{B}_v(r), \mathbf{B}_v(r) \rangle_0^1$ up to the constant $\Delta_{vv} \Omega_{vv}^{-1}$.

It remains to consider $G_X \sum_{t=1}^{T} D_t \Delta X_t' G_W \Rightarrow \int_0^1 D(r) \dot{\mathbf{B}}_v(r)' dB_v(r) + \frac{\Omega_{vv}}{2} \int_0^1 D(r) \ddot{\mathbf{B}}_v(r)' dr = \int_0^1 D(r) d\mathbf{B}_v(r)'$. The convergence result follows from similar considerations as above and the equality follows again from Itô's Lemma. □

Proof of Proposition 2.11. Consider the two terms given in the last line of (12). The proof of Wagner and Hong (2016, Prop. 1) establishes that

$$GZ'u \Rightarrow \int_0^1 J(r)dB_u(r) + \Delta_{vu}\begin{bmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix}. \tag{A.15}$$

The asymptotic behavior of $GZ'\tilde{W}$ has been established in Theorem 2.9. The first column of this limit, corresponding to the first component v_t of $\tilde{w}_t$, is given by

$$GZ'v \Rightarrow \int_0^1 J(r)dB_v(r) + \Delta_{vv}\begin{bmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix}, \tag{A.16}$$

which is also a well-known result, compare again Wagner and Hong (2016, Prop. 1).

It is important to note that only the first column of (the limit of) $GZ'v$ is required (and it would be sufficient to show that the other columns are $O_{\mathbb{P}}(1)$) to establish the robustness result. This follows from the structure of the limit of $\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u}$, which follows directly from the results obtained in Theorem 2.7, that is, $\hat{\Omega}_{\tilde{w}\tilde{w}} \Rightarrow \Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr$ and $\hat{\Omega}_{\tilde{w}u} \Rightarrow \Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr$. With, as in the considered case, the first element of $\dot{\mathbf{B}}_v(r)$ equal to 1, it follows that the first column of $\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr$ is equal to $\int_0^1 \dot{\mathbf{B}}_v(r)dr$, which immediately implies that

$$\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \xrightarrow{\mathbb{P}} \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p-1)\times 1} \end{bmatrix}. \tag{A.17}$$

Combining Theorem 2.9 with (A.17) yields

$$\begin{aligned} GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} &\Rightarrow \left(\int_0^1 J(r)d\mathbf{B}_v(r)' + \begin{bmatrix} 0_{q\times p} \\ \Delta_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr \end{bmatrix}\right)\Omega_{vv}^{-1}\Omega_{vu}\begin{bmatrix} 1 \\ 0_{(p-1)\times 1} \end{bmatrix} \\ &= \int_0^1 J(r)dB_v(r)\Omega_{vv}^{-1}\Omega_{vu} + \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu}\begin{bmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix}. \end{aligned} \tag{A.18}$$

It remains to consider

$$\begin{aligned} GA^{**} &= \begin{bmatrix} 0_{q\times 1} \\ \hat{\Delta}_{\tilde{w}u}^{+} \end{bmatrix} = \begin{bmatrix} 0_{q\times 1} \\ \hat{\Delta}_{\tilde{w}u} - \hat{\Delta}_{\tilde{w}\tilde{w}}\hat{\Omega}_{\tilde{w}\tilde{w}}^{1}\hat{\Omega}_{\tilde{w}u} \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} 0_{q\times 1} \\ \Delta_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr - \Delta_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr\left(\Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr\right)^{-1}\Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix} \\ &= \begin{bmatrix} 0_{q\times 1} \\ \left(\Delta_{vu} - \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu}\right)\int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix} = \Delta_{vu}^{+}\begin{bmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix}, \end{aligned} \tag{A.19}$$

with this result holding also in the case no (and thus also not the first) element of $\dot{\mathbf{B}}_v(r)$ being equal to 1. Combining the results leads to the following limit of $GZ'u^{++} - GA^{**}$:

$$\begin{aligned} &GZ'u - GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} - \begin{bmatrix} 0_{q\times 1} \\ \hat{\Delta}_{\tilde{w}u}^{+} \end{bmatrix} \\ &\Rightarrow \int_0^1 J(r)dB_u(r) + \Delta_{vu}\begin{bmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix} \end{aligned} \tag{A.20}$$

https://doi.org/10.1017/S0266466624000033 Published online by Cambridge University Press

$$
\begin{aligned}
&- \int_0^1 J(r)dB_v(r)\Omega_{vv}^{-1}\Omega_{vu} - \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu}\begin{bmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix} - \Delta_{vu}^{+}\begin{bmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}_v(r)dr \end{bmatrix} \\
&= \int_0^1 J(r)dB_{u\cdot v}(r),
\end{aligned}
$$

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

To complete the arguments required for the results in this paper, it only remains to show that $\hat{\omega}_{u\cdot w} \xrightarrow{\mathbb{P}} \omega_{u\cdot v}$, which also follows from Theorem 2.7:

$$
\begin{aligned}
\hat{\omega}_{u\cdot w} &= \hat{\Omega}_{uu} - \hat{\Omega}_{uw}\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} = \hat{\Omega}_{uu} - \hat{\Omega}_{u\tilde{w}}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \qquad \textbf{(A.21)} \\
&\Rightarrow \Omega_{uu} - \Omega_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr\left(\Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr\right)^{-1}\Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr \\
&= \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu} = \omega_{u\cdot v},
\end{aligned}
$$

since $\int_0^1 \dot{\mathbf{B}}_v(r)'dr\left(\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr\right)^{-1}\int_0^1 \dot{\mathbf{B}}_v(r)dr = 1$ also follows from the first element of $\dot{\mathbf{B}}_v(r)$ being equal to 1. □

SUPPLEMENTARY MATERIAL

Stypka, O., M. Wagner, P. Grabarczyk and R. Kawka (2024): Supplement to "Cointegrating Polynomial Regressions: Robustness of Fully Modified OLS," *Econometric Theory Supplementary Material*. To view, please visit https://doi.org/10.1017/S0266466624000033.

REFERENCES

Chan, N., & Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics*, 185, 182–195.

Chang, Y., Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal*, 4, 1–36.

Darvas, Z. (2008). Estimation bias and inference in overlapping autoregressions: Implications for the target-zone literature. *Oxford Bulletin of Economics and Statistics*, 70, 1–22.

Di Iorio, F., & Fachin, S. (2022). Fiscal reaction functions for the advanced economies revisited. *Empirical Economics*, 62, 2865–2891.

Grabarczyk, P., Wagner, M., Frondel, M., & Sommer, S. (2018). A cointegrating polynomial regression analysis of the material Kuznets curve hypothesis. *Resources Policy*, 57, 236–245.

Grossman, G. M., & Krueger, A. B. (1995). Economic growth and the environment. *Quarterly Journal of Economics*, 110, 353–377.

Ibragimov, R., & Phillips, P. C. B. (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*, 24, 1373–1403.

Kuznets, S. (1955). Economic growth and income inequality. *American Economic Review*, 45, 1–28.

Labson, B. S., & Crompton, P. L. (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.

Le Gall, J. F. (2016). *Brownian motion, martingales, and stochastic calculus*. Springer International Publishing.

Liang, H., Phillips, P. C. B., Wang, H., & Wang, Q. (2016). Weak convergence to stochastic integrals for econometric applications. *Econometric Theory*, 32, 1349–1375.

Malenbaum, W. (1978). *World demand for raw materials in 1985 and 2000*. McGraw-Hill, EMJ Mining Information Services.

Park, J. Y., & Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, 15, 269–298.

Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, 69, 117–161.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, 55, 277–301.

Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory*, 5, 181–240.

Phillips, P. C. B. (1991a). Optimal inference in cointegrated systems. *Econometrica*, 59, 283–306.

Phillips, P. C. B. (1991b). Spectral regression for cointegrated time series. In W. Barnett, J. Powell, & G. Tauchen (Eds.), *Nonparametric and semiparametric methods in econometrics and statistics* (pp. 413–436). Cambridge University Press.

Phillips, P. C. B. (1995). Fully modified least squares and vector autoregression. *Econometrica*, 63, 1023–1078.

Phillips, P. C. B., & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies*, 57, 99–125.

Shin, Y. (1994). A residual-based test for the null of cointegration against the alternative of no cointegration. *Econometric Theory*, 10, 91–115.

Stypka, O., & Wagner, M. (2019). The Phillips unit root tests for polynomials of integrated processes revisited. *Economics Letters*, 176, 109–113.

Svensson, L. E. O. (1992). An interpretation of recent research on exchange rate target zones. *Journal of Economic Perspectives*, 6, 119–144.

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., & Hong, S. H. (2016). Cointegrating polynomial regressions: Fully modified OLS estimation and inference. *Econometric Theory*, 32, 1289–1315.

Yandle, B., Bhattarai, M., & Vijayaraghavan, M. (2004). Environmental Kuznets curves: A review of findings, methods, and policy implications. *PERC Research Study*, 2(1), 1–38.