The Asymptotic Validity of "Standard" Fully Modified OLS Estimation and Inference in Cointegrating Polynomial Regressions

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

[1]Faculty of Statistics, Technical University Dortmund
[2]Institute for Advanced Studies, Vienna
[3]Bank of Slovenia, Ljubljana

Proposed running head:

"Standard" FM-OLS in CPRs

Corresponding author:

Martin Wagner
Faculty of Statistics
Technical University Dortmund
Vogelpothsweg 87
44227 Dortmund, Germany
Tel: +49 231 755 3174
Email: mwagner@statistik.tu-dortmund.de

Abstract

The paper considers estimation and inference in cointegrating polynomial regressions, i. e., regressions that include deterministic variables, integrated processes and their powers as explanatory variables. The stationary errors are allowed to be serially correlated and the regressors to be endogenous. We show that estimating such relationships using the Phillips and Hansen (1990) fully modified OLS approach developed for linear cointegrating relationships by *incorrectly considering all integrated regressors and their powers as integrated regressors* leads to the same limiting distribution as the Wagner and Hong (2016) fully modified type estimator developed for cointegrating polynomial regressions. The only restriction for this result to hold is that all integrated variables themselves are included as regressors. Key ingredients for our results are novel limit results for kernel weighted sums of properly scaled nonstationary processes involving powers of integrated processes and a functional central limit theorem involving polynomials of Brownian motions as both integrand and integrator. Even though simulation results indicate performance advantages of the Wagner and Hong (2016) estimator that are partly present even in large samples, the results of the paper drastically enlarge the useability of the Phillips and Hansen (1990) estimator implemented in many software packages.

JEL Classification: C13, C32

Keywords: Cointegrating Polynomial Regression, Cointegration Test, Fully Modified OLS Estimation, Integrated Process, Kernel Weighted Sums

1. Introduction

One motivation to consider cointegrating polynomial regressions (CPRs), using the terminology of Wagner and Hong (2016), is the environmental Kuznets curve (EKC) literature that investigates a potentially inverted U-shaped relationship between measures of economic development (typically proxied by GDP per capita) and pollution. This literature continues to grow at rapid pace since the seminal work of Grossman and Krueger (1995).[1] Early survey papers, like Yandle *et al.* (2004), count more than 100 refereed publications already more than a decade ago. As an example of the relationship considered in this literature consider the scatterplot between the logarithm of GDP per capita and the logarithm of CO_2 emissions per capita in Belgium over the period 1870–2009 in Figure 1.

The estimation results displayed in this figure are obtained from estimating the relationship:

$$\ln(\mathrm{CO}_2)_t = c + \delta t + \beta_1 \ln(\mathrm{GDP})_t + \beta_2 (\ln(\mathrm{GDP})_t)^2 + u_t, \qquad (1)$$

where the logarithm of Belgian GDP per capita is well-described as an integrated process of order one, compare Wagner (2015). When the errors u_t are stationary, the above relationship is a CPR relationship. An integrated process and its square cannot both be integrated processes of order one (see, e. g., Wagner, 2012) and obviously there is an exact deterministic relationship between the logarithm of GDP per capita and its square. These basic observations lead Wagner and Hong (2016) to extend the fully modified OLS (FM-OLS) estimator of Phillips and Hansen (1990) from the linear cointegration setting to the CPR setting.[2] The corresponding estimation results, referred to as FM-CPR in this paper, are displayed as the dashed line in Figure 1.

[1] The term EKC refers by analogy to the inverted U-shaped relationship between the level of economic development and the degree of income inequality postulated by Simon Kuznets (1955) in his 1954 presidential address to the American Economic Association. Inverted U-shaped relationships are also prominent in other areas, including the so-called intensity-of-use literature investigating the relationship between energy or material intensity and GDP per capita, see, e. g., Malenbaum (1978) or Labson and Crompton (1993).

[2] As discussed in Wagner and Hong (2016), similar results are or could also be obtained under alternative assumptions that partly need to be augmented to accommodate powers of integrated regressors, see, e. g., Chan and Wang (2015), Chang *et al.* (2001), de Jong (2002), Ibragimov and Phillips (2008) or Liang *et al.* (2016). A key difference between the results here and those of, e. g., Chang *et al.* (2001) is that we allow $\{u_t\}_{t\in\mathbb{Z}}$ to be serially correlated, in an MDS setting in Wagner and Hong (2016) and in a linear process setting in this paper. Wang (2015) is an excellent monograph on asymptotic theory for nonlinear cointegration in a regression framework.

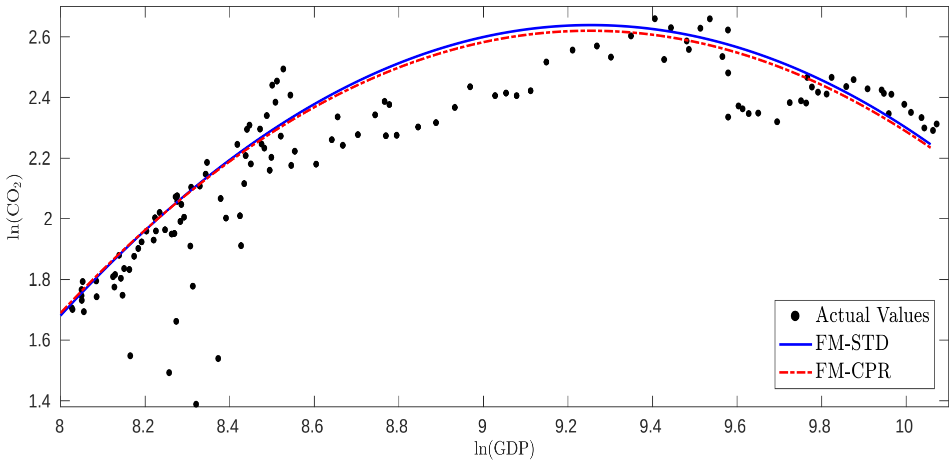

Figure 1: Estimated EKC for CO_2 emissions for Belgium over the period 1870–2009; variables in logarithms of per capita quantities. The curves result from inserting 140 equidistantly spaced observations from the sample range of ln(GDP), with trend values given by $t = 1, \dots, 140$, in the estimated relationship (1). The solid line corresponds to the FM-STD coefficient estimates and the dashed line to the FM-CPR coefficient estimates.

The solid line also displayed in Figure 1 corresponds to how cointegration methods are routinely used in the EKC literature: The estimates are derived from treating (1) *as if it were a linear cointegrating relationship with two integrated regressors*, estimated using, e. g., the FM-OLS estimator of Phillips and Hansen (1990). Both, log GDP per capita and its square are thereby considered as integrated processes of order one that are furthermore assumed to be not cointegrated, which as just discussed is mathematically impossible. This estimator is referred to as FM-STD estimator in this paper.

Given the conceptual differences between a linear cointegration relationship and a cointegrating polynomial relationship it appears to be misguided to use the FM-STD estimator in a cointegrating polynomial regression. However, the figure shows that very similar results are obtained when using either FM-CPR or FM-STD. This observation has also been made with data for 19 countries in Wagner (2015). We provide an asymptotic explanation for such similar findings: The two estimators, FM-CPR and FM-STD, have the same asymptotic distribution in the CPR case. This result holds true for the general CPR case considered in Wagner and Hong (2016), with multiple integrated regressors, arbitrary polynomial powers and general deterministic components. A practical implication of this result is that one can use standard software package implementations of FM-OLS of Phillips and Hansen (1990) for estimation and inference in cointegrating polynomial relationships by "formally" (for the software) considering all integrated regressors and their powers as integrated regressors.[3] The only restriction for the result to hold is that all integrated variables are themselves, i. e., to the power one, included as regressors in the CPR relationship. This restriction is directly related to the following main observation, discussed in detail in Section 2.3: A key step in FM-OLS-type estimation is an asymptotic orthogonalization of two Brownian motions to obtain a zero mean Gaussian mixture limiting distribution. Given that Brownian motions are by definition Gaussian, achieving independence is equivalent to achieving uncorrelatedness. The latter is obtained by the first-step modification of the dependent variable not only – by tailor-made design – of FM-CPR, but also by the first-step modification of FM-STD, if all integrated variables are included as regressors with their first power. In a sense made precise below, FM-STD thus contains and calculates (asymptotically) superfluous quantities in the orthogonalization step (and also in the bias correction step).

A key ingredient, of independent interest also in other contexts (see, e. g., Stypka and Wagner, 2018), are weak convergence results for kernel weighted sums ("long-run covariance estimators") of – properly scaled – processes involving powers of integrated processes. These arise in both transformations that the FM estimation principle is based upon, in the modification of the dependent variable and in the additive bias correction. Turning back to our example equation (1), with full details and all definitions contained in Section 2, the dependent variable, logarithm of CO_2 emission per capita, y_t for brevity, is modified – when using FM-STD – to:

$$
\begin{aligned}
y_t^{++} &= y_t - [\Delta x_t, \Delta x_t^2]\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \\
&= y_t - [\Delta x_t, 2x_t\Delta x_t - (\Delta x_t)^2]\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}
\end{aligned} \tag{2}
$$

with x_t denoting the logarithm of GDP per capita. Using $w_t = [\Delta x_t, 2x_t\Delta x_t - (\Delta x_t)^2]'$, the above transformation involves "long-run covariance" estimators involving (in the quadratic case) not only a stationary process, Δx_t, but also $x_t\Delta x_t$. In this paper we derive the weak limits for (properly scaled) kernel weighted sums involving such quantities. More specifically, we show that these kernel weighted sums converge to the quadratic covariation of the limiting process corresponding to the errors and regressors (including the powers). For our asymptotic equivalence result of FM-STD and FM-CPR it is key that these limits exhibit a specific structure discussed in detail below.

[3] For notational brevity we focus in the main text on the single integrated regressor case, which facilitates reading and suffices to see all elements required for the results "in action". In Appendix D we outline the changes and modifications necessary for the multiple integrated regressor case.

The asymptotic equivalence of the FM-STD and FM-CPR estimators implies asymptotic equivalence also of residual based cointegration tests like, e. g., the Shin (1994) test. This test has been extended to CPRs in Wagner and Hong (2016) and Wagner (2013), with critical values depending, as usual in the cointegration literature, upon the specification of the cointegrating polynomial relationship. The EKC literature uses the FM-STD residuals (asymptotically valid given the results in this paper), but in conjunction with the original Shin (1994) critical values. This combination results in invalid inference even asymptotically, as discussed in Section 2.4.

The simulation results indicate that *as expected* FM-CPR outperforms FM-STD. In case of large endogeneity and serial correlation of the errors this leads to marked performance differences even in large samples like $T = 1000$ despite asymptotic equivalence. In these cases the calculation of superfluous quantities alluded to above and explained in more detail in Section 2.3 impacts the performance of FM-STD detrimentally. The performance advantages occur in all considered dimensions, estimator bias and RMSE, performance of parameter hypothesis tests, and performance of cointegration tests. In case of data with little or no endogeneity and serial correlation the differences between the estimators more or less vanish for the larger sample sizes considered. The differences are large for cointegration testing, even when the cointegration test calculated from the FM-STD residuals is used in conjunction with the correct rather than the Shin (1994) critical values.

The paper is organized as follows: In Section 2 we present the setting, the assumptions and the theoretical results. Section 3 contains a small selection of results from a simulation study assessing the finite sample differences between the two asymptotically equivalent estimators and test statistics based upon them. Section 4 briefly summarizes and concludes. Four appendices follow the main text: Appendix A describes for completeness and as reference point the FM-CPR estimator of Wagner and Hong (2016), Appendix B contains some auxiliary lemmata, Appendix C contains the proofs of the main results and Appendix D illustrates how to modify the main arguments of the proofs to cover the general, multiple integrated regressor case. Supplementary material available upon request contains additional 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 $\stackrel{\mathbb{P}}{\to}$. We use $O_{\mathbb{P}}(1)$ to denote boundedness in probability. With $o_{\mathbb{P}}(1)$ and $o_{a.s.}(1)$ we denote convergence to zero in probability and almost surely respectively. The integer part of

$x \in \mathbb{R}$ is given by $\lfloor x \rfloor$ and a diagonal matrix with entries specified throughout by $\text{diag}(\cdot)$. For a vector $x = (x_i)_{i=1,\dots,n}$ 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 j-th column is labelled by $A_{(\cdot,j)}$, $0_{m\times n}$ denotes an $(m \times n)$-matrix with all entries equal to zero and e_m^n denotes the m-th unit vector in $\mathbb{R}^n$. The Kronecker product is denoted by $\otimes$. Let $\mathbf{A} = (A_{ij})$ be partitioned with $A_{ij} \in \mathbb{R}^{m_i \times n_j}$ as its (i,j)-th block sub-matrix and let $\mathbf{B} = (B_{ij})$ be partitioned with $B_{ij} \in \mathbb{R}^{p_i \times q_j}$ as its (i,j)-th block sub-matrix. Then, we denote their Khatri-Rao product as $\mathbf{A} * \mathbf{B} := (A_{ij} \otimes B_{ij})_{ij}$. We use $\mathbb{E}$ to denote expectation and L is the backward-shift operator, i. e., $L\{x_t\}_{t\in\mathbb{Z}} = \{x_{t-1}\}_{t\in\mathbb{Z}}$. The first-difference operator is denoted with Δ, i. e., $\Delta := 1 - L$. 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 $B(r)$. Standard Brownian motion is denoted by $W(r)$.

2. Theory

2.1. Setup and Assumptions

As mentioned in the introduction, it suffices to consider a cointegrating polynomial regression with only one integrated regressor and its powers:[4]

$$\begin{aligned} y_t &= D_t'\delta + X_t'\beta + u_t, \quad \text{for } t = 1,\dots,T, \\ x_t &= x_{t-1} + v_t, \end{aligned} \tag{3}$$

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 (3) can be rewritten more compactly as:

$$y_t = Z_t'\theta + u_t, \quad \text{for } t = 1,\dots,T. \tag{4}$$

[4]Clearly, not all consecutive powers of x_t need to be included and in the multiple integrated regressor case the included powers may differ across integrated variables. These changes only complicate "book-keeping". What is, however, important for asymptotic equivalence is that the integrated variable x_t itself is included in the regression, as discussed in detail at the end of Section 2.3. The initial value x_0 is allowed to be any well-defined $O_{\mathbb{P}}(1)$ random variable.

Assumption 1. For the deterministic component we assume that 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:

$$\lim_{T\to\infty} T^{1/2} G_D D_{\lfloor sT \rfloor} = D(s). \tag{5}$$

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

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

Assumption 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:

$$u_t = C_u(L)\zeta_t = \sum_{j=0}^{\infty} c_{uj}\zeta_{t-j}, \tag{6}$$

$$\Delta x_t = v_t = C_v(L)\varepsilon_t = \sum_{j=0}^{\infty} c_{vj}\varepsilon_{t-j}, \tag{7}$$

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 $\Sigma_{\xi^0\xi^0}$.

Assumption 2 is stronger than the corresponding assumption in Wagner and Hong (2016), which allows us to directly draw upon some results of Kasparis (2008).[6] 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 $\det(C_v(1)) \neq 0$, i. e., in the multivariate case (e. g. in the discussion in Appendix D) 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:

[5] 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.

[6] 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 believe that the minimum should be replaced by the maximum. Since we rely upon similar arguments in the proof of Lemma 4 we require moments of order $\max(8, 4/(1-2b))$.

Assumption 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 $\bar{k}(x) = \sup_{y \geq x} |k(y)|$

Assumption 4. The bandwidth parameter $M_T \to \infty$ fulfills $M_T = O(T^b)$, for the same parameter b as in Assumption 2.

The bandwidth Assumption 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 FM-STD estimator kernel "long-run covariance" estimators involving (properly scaled) powers of integrated processes need to be analyzed. In order to have uniform notation for kernel weighted sums irrespective of the properties of the sequences considered we *formally* define:

Definition 1. *For two sequences $\{a_t\}_{t=1,\ldots,T}$ and $\{b_t\}_{t=1,\ldots,T}$ we define:*[7]

$$\hat{\Delta}_{ab} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} a_t b'_{t+h}, \tag{8}$$

neglecting the dependence on $k(\cdot)$, M_T and the sample range $1, \ldots, T$ for brevity. Furthermore,

$$\hat{\Omega}_{ab} := \hat{\Delta}_{ab} + \hat{\Delta}'_{ab} - \hat{\Sigma}_{ab}, \tag{9}$$

with $\hat{\Sigma}_{ab} := \frac{1}{T}\sum_{t=1}^{T} a_t b'_t$. Based on these quantities we furthermore 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 that $\{a_t\}_{t\in\mathbb{Z}}$ and $\{b_t\}_{t\in\mathbb{Z}}$ are jointly stationary processes with finite half long-run covariance $\Delta_{ab} := \sum_{h=0}^{\infty} \mathbb{E}(a_0 b'_h)$, 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 1. Note that in our definition of $\hat{\Delta}_{ab}$ in (8) 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

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

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, where we analyze the asymptotic behavior of $\hat{\Delta}$-quantities for (properly scaled) nonstationary processes (in Theorem 1 and Corollary 1), the summation bound is important. A key result of this paper, Theorem 1 below, hinges upon summation only up to M_T. More specifically, we rely upon the summation bound M_T in the proof of Lemma 3, which is related to Kasparis (2008, Lemma A1, p. 1394–1396), who also uses M_T (in a slightly different context).

Assumption 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:

$$\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor rT \rfloor} \xi_t \Rightarrow B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \Omega_{\xi\xi}^{1/2} W(r), \quad r \in [0,1], \tag{10}$$

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

$$\Omega_{\xi\xi} := \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} = \sum_{h=-\infty}^{\infty} \mathbb{E}(\xi_0 \xi_h'). \tag{11}$$

Later we also need the corresponding half (or one-sided) long-run covariance matrix $\Delta_{\xi\xi} := \sum_{h=0}^{\infty} \mathbb{E}(\xi_0 \xi_h')$, partitioned similarly as $\Omega_{\xi\xi}$. 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 furthermore the unobserved errors u_t are replaced by the OLS residuals $\hat{u}_t$ from (3), defining $\hat{\xi}_t := [\hat{u}_t, v_t]'$.[8]

[8] We keep using, e. g., $\hat{\Omega}_{\xi\xi}$ when using the observable $\hat{u}_t$ instead of u_t in long-run covariance estimation. Infeasible estimation involving the unobserved errors u_t is labelled with a tilde-symbol, e. g., $\tilde{\Omega}_{\xi\xi}$, see Theorem 1 below.

2.2. "Standard" Fully Modified OLS Estimation

Performing what we call "standard" FM-OLS estimation amounts to considering (3) "formally" as a standard linear cointegrating regression with p integrated regressors that can consequently be written as:

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

which defines

$$w_t := \Delta X_t = \begin{bmatrix} \Delta x_t \\ \Delta x_t^2 \\ \vdots \\ \Delta x_t^p \end{bmatrix} = \begin{bmatrix} v_t \\ 2x_t v_t - v_t^2 \\ \vdots \\ -\sum_{k=1}^{p} \binom{p}{k} x_t^{p-k}(-v_t)^k \end{bmatrix}, \tag{12}$$

i. e., the j-th component of the vector w_t is given by $w_{t,j} = -\sum_{k=1}^{j} \binom{j}{k} x_t^{j-k}(-v_t)^k$. The correspondingly modified dependent variable is given by:

$$\begin{aligned} y_t^{++} &:= y_t - \Delta X_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}, \\ &= y_t - w_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \end{aligned} \tag{13}$$

with $\hat{\Omega}_{ww}$ and $\hat{\Omega}_{wu}$ to be interpreted in the sense of Definition 1. The additive correction term for FM-STD is given by:

$$A^{**} := \begin{bmatrix} 0_{q\times 1} \\ T\hat{\Delta}_{wu}^{+} \end{bmatrix} = \begin{bmatrix} 0_{q\times 1} \\ T(\hat{\Delta}_{wu} - \hat{\Delta}_{ww}\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}) \end{bmatrix} \tag{14}$$

with $\hat{\Delta}_{ww}$, $\hat{\Delta}_{wu}$ and $\hat{\Delta}_{wu}^{+}$ also interpreted in the sense of Definition 1. Given that the process $\{w_t\}_{t\in\mathbb{Z}}$ contains powers of integrated processes, see again (12), and is therefore clearly nonstationary it is to be expected that "long-run covariance" estimators involving this process will exhibit nonstandard asymptotic behavior. That this is indeed the case is shown in Theorem 1 below and the ensuing result is a key building block for showing estimator equivalency.

With all quantities required defined, the FM-STD estimator is given by:

$$\hat{\theta}^{++} := (Z'Z)^{-1}(Z'y^{++} - A^{**}), \tag{15}$$

with $y^{++} := [y_1^{++}, \dots, y_T^{++}]'$, $Z := [Z_1, \dots, Z_T]'$, $Z_t := [D_t', X_t']'$. Denoting with $\hat{u}^{++} := [\hat{u}_1^{++}, \dots, \hat{u}_T^{++}]'$, where

$$\hat{u}_t^{++} := u_t - w_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}, \tag{16}$$

the centered and scaled FM-STD estimator can be written as:

$$G^{-1}(\hat{\theta}^{++} - \theta) = \left(GZ'ZG\right)^{-1}\left(GZ'u^{++} - GA^{**}\right), \tag{17}$$

with the scaling matrix G defined as:

$$G = G(T) := \text{diag}(G_D(T), G_X(T)), \tag{18}$$

where $G_X(T) := \text{diag}(T^{-1}, T^{-3/2}, \dots, T^{-(p+1)/2})$.

The asymptotic behavior of the first term, $(GZ'ZG)^{-1}$, is easy to establish:

$$(GZ'ZG)^{-1} \Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1}, \tag{19}$$

with $J(r) := [D(r)', \mathbf{B}_v(r)']'$, where $\mathbf{B}_v(r) := [B_v(r), B_v^2(r), \dots, B_v^p(r)]'$.

Thus, establishing the asymptotic behavior of FM-STD requires to understand the quantities composing the second term in (17). Defining $G_W := G_W(T) = \text{diag}(1, T^{-1/2}, \dots, T^{-(p-1)/2})$ leads to:

$$\begin{aligned} GZ'u^{++} &= GZ'(u - W\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu}) \\ &= GZ'u - GZ'W\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \\ &= GZ'u - GZ'WG_WG_W^{-1}\hat{\Omega}_{ww}^{-1}G_W^{-1}G_W\hat{\Omega}_{wu} \\ &= GZ'u - GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u}, \end{aligned} \tag{20}$$

with $W := [w_1, \dots, w_T]'$, $\tilde{W} = [\tilde{w}_1, \dots, \tilde{w}_T]' := WG_W$, where $\tilde{w}_t := [v_t, \frac{\Delta x_t^2}{T^{1/2}}, \dots, \frac{\Delta x_t^p}{T^{\frac{p-1}{2}}}]'$, $\hat{\Omega}_{\tilde{w}\tilde{w}} := G_W\hat{\Omega}_{ww}G_W$ and $\hat{\Omega}_{\tilde{w}u} := G_W\hat{\Omega}_{wu}$. In the above expression the first term, $GZ'u$, is well-understood (see, e. g., (A.3) in the proof of Proposition 1 in Wagner and Hong, 2016). The re-scaling with G_W allows to arrive at well defined limits of $GZ'\tilde{W}$, $\hat{\Omega}_{\tilde{w}\tilde{w}}$ and $\hat{\Omega}_{\tilde{w}u}$.

The final term, GA^{**}, can be rewritten as:

$$GA^{**} = \begin{bmatrix} G_D & 0 \\ 0 & G_X \end{bmatrix} \begin{bmatrix} 0_{q\times 1} \\ T\hat{\Delta}^{+}_{wu} \end{bmatrix} = \begin{bmatrix} 0_{q\times 1} \\ G_W\hat{\Delta}^{+}_{wu} \end{bmatrix} = \begin{bmatrix} 0_{q\times 1} \\ \hat{\Delta}^{+}_{\tilde{w}u} \end{bmatrix}, \tag{21}$$

using $G_X T = G_W$.

A key result for deriving the asymptotic behavior of the FM-STD estimator is the asymptotic behavior of the *properly scaled* "long-run covariance" estimators $\hat{\Omega}_{\tilde{w}\tilde{w}}$, $\hat{\Omega}_{\tilde{w}u}$ and their half counterparts $\hat{\Delta}_{\tilde{w}\tilde{w}}$, $\hat{\Delta}_{\tilde{w}u}$. The analysis proceeds in two steps. First, we show the result for $\eta_t := [u_t, \tilde{w}_t']'$ (Theorem 1) and then we show that the same limits also hold for $\hat{\eta}_t := [\hat{u}_t, \tilde{w}_t']'$ (Corollary 1), with $\hat{u}_t$ the OLS residuals from (3).

Theorem 1. *Under Assumptions 2 to 4 it holds for $T \to \infty$ that*

$$\tilde{\Delta}_{\eta\eta} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\eta_t\eta_{t+h}' \Rightarrow \Delta_{\eta\eta} := \begin{bmatrix} \Delta_{uu} & \Delta_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr \\ \Delta_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr & \Delta_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr \end{bmatrix}, \tag{22}$$

with

$$\dot{\mathbf{B}}_v(r) := \left[1, 2B_v(r), \ldots, pB_v^{p-1}(r)\right]'. \tag{23}$$

Furthermore, it holds for $T \to \infty$ that:

$$\tilde{\Sigma}_{\eta\eta} := \frac{1}{T}\sum_{t=1}^{T}\eta_t\eta_t' \Rightarrow \Sigma_{\eta\eta} := \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr \\ \Sigma_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr & \Sigma_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr \end{bmatrix}. \tag{24}$$

Combining the above two results lead to:

$$\tilde{\Omega}_{\eta\eta} := \tilde{\Delta}_{\eta\eta} + \tilde{\Delta}_{\eta\eta}' - \tilde{\Sigma}_{\eta\eta} \Rightarrow \Delta_{\eta\eta} + \Delta_{\eta\eta}' - \Sigma_{\eta\eta} =: \Omega_{\eta\eta}, \tag{25}$$

with

$$\Omega_{\eta\eta} = \begin{bmatrix} \Omega_{uu} & \Omega_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr \\ \Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)dr & \Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr \end{bmatrix} = \langle \mathcal{B}(r), \mathcal{B}(r)\rangle_0^1, \tag{26}$$

with $\mathcal{B}(r) := [B_u(r), \mathbf{B}_v(r)']'$.

Corollary 1. *Let the data be generated by* (3) *under Assumptions 1 and 2 and let long-run covariance estimation be performed under Assumptions 3 and 4. Then the results of Theorem 1 also hold for* $\hat{\eta}_t$ *in place of* η_t*, i. e., as* $T \to \infty$*:*

$$\hat{\Delta}_{\eta\eta} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \hat{\eta}_t \hat{\eta}_{t+h}' \Rightarrow \Delta_{\eta\eta} \tag{27}$$

$$\hat{\Sigma}_{\eta\eta} := \frac{1}{T} \sum_{t=1}^{T} \hat{\eta}_t \hat{\eta}_t' \Rightarrow \Sigma_{\eta\eta} \tag{28}$$

$$\hat{\Omega}_{\eta\eta} := \hat{\Delta}_{\eta\eta} + \hat{\Delta}_{\eta\eta}' - \hat{\Sigma}_{\eta\eta} \Rightarrow \Omega_{\eta\eta} \tag{29}$$

Remark 2. In light of Remark 1 we continue to use standard notation for the limits, i. e., $\Delta_{\eta\eta}$, $\Sigma_{\eta\eta}$ and $\Omega_{\eta\eta}$, but these are not covariance matrices of underlying stationary processes. Only, by construction, 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 $\{\xi_t\}_{t\in\mathbb{Z}}$.

It remains to characterize the asymptotic behavior of $GZ'\tilde{W}$, which is derived in the following result.[9]

Proposition 1. *Under Assumptions 1 and 2 it holds for* $T \to \infty$ *that:*

$$GZ'\tilde{W} \Rightarrow \int_0^1 J(r) d\mathbf{B}_v(r)' + \begin{pmatrix} 0 \\ \Delta_{vv}\Omega_{vv}^{-1}\langle \mathbf{B}(r), \mathbf{B}(r)\rangle_0^1 \end{pmatrix} \tag{30}$$

The above result is more general than the usual type of functional central limit theorem used in the unit root and cointegration literature in two ways. The simplest form arising, e. g., in unit root testing is given by $\frac{1}{T}\sum_{t=1}^{T} x_t \Delta x_t = \frac{1}{T}\sum_{t=1}^{T} x_t v_t \Rightarrow \int_0^1 B_v(r) dB_v(r) + \Lambda_{vv}$. The first dimension along which the result is more general is to consider nonlinear functions – here polynomials – as integrands. This is by now a widely-studied issue in the nonlinear cointegration literature (compare, e. g., the references in Wagner and Hong, 2016); over and beyond polynomials in many ways. The second generalization appears less studied in the unit root and cointegration context: the integrator is generalized from $dB_v(r)$, corresponding to v_t, to $d\mathbf{B}_v(r)$. Since $\mathbf{B}_v(r)$ is a semi-martingale, stochastic integration is well-defined.

[9] Note that the first column, corresponding to the component v_t of $\tilde{w}_t$, of the limiting expression derived in this proposition is known, compare Wagner and Hong (2016, p. 1312).

Combining the results of Theorem 1, Corollary 1 and Proposition 1 allows to establish the main result of this paper by exploiting the structure of the "long-run covariance" limits as discussed in detail in Section 2.3 (and in Appendix D for the general case):

Theorem 2. *Let the data be generated by* (3) *under Assumptions 1 and 2. Furthermore, let long-run covariance estimation be performed under Assumptions 3 and 4. Then it holds for* $T \to \infty$ *that:*

$$G^{-1}(\hat{\theta}^{++} - \theta) \Rightarrow \left(\int_0^1 J(r)J(r)'dr \right)^{-1} \int_0^1 J(r)dB_{u \cdot v}(r), \tag{31}$$

Thus, the FM-STD and the FM-CPR estimator, compare (A.5) *in Appendix A, have the same limiting distribution.*

The above result implies that all hypothesis test statistics based on either of the two estimators have the same asymptotic null distribution as well. This includes, of course, Wald-type parameter hypothesis tests, but also the Wald- and LM-type specification tests considered in Wagner and Hong (2016, Propositions 3 and 4).

2.3. A Discussion of the Equivalence Result

To better understand the equivalence result given in Theorem 2 it is helpful to discuss both key elements, the modification of the dependent variable and the structure of the kernel weighted sum limits.

We start with considering the Phillips and Hansen (1990) modification of the dependent variable, see (A.1) and its centered counterpart (A.2) in Appendix A:

$$\begin{aligned} u_t^+ &= u_t - \Delta x_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \\ &= u_t - v_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \end{aligned} \tag{32}$$

With consistent long-run covariance estimation, the partial sum process on the left hand side converges to a Gaussian process independent of the Gaussian limit partial process corresponding to $\{v_t\}_{t \in \mathbb{Z}}$, i. e.:

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} u_t^+ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} u_t - \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} v_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \\ &\Rightarrow B_u(r) - B_v(r) \Omega_{vv}^{-1} \Omega_{vu} = B_{u \cdot v}(r). \end{aligned} \tag{33}$$

The above relationship can also be interpreted differently, as the problem of orthogonalizing two Gaussian random variables (for fixed argument r or for two Gaussian random processes with $0 \leq r \leq 1$). This problem has by definition a regression flavor, with $B_u(r)$ being the dependent variable and $B_v(r)$ the regressor:

$$B_u(r) = B_v(r)\Theta[1] + B_{u\cdot v}(r), \tag{34}$$

with the population regression coefficient $\Theta[1]$ given by:

$$\begin{aligned}\Theta[1] &:= \left(\mathbb{E}(B_v(r)B_v(r))\right)^{-1}\mathbb{E}(B_v(r)B_u(r)) \\ &= \Omega_{vv}^{-1}\Omega_{vu},\end{aligned} \tag{35}$$

where the notation $\Theta[1]$ indicates that in the regression on only the first power of $B_v(r)$ is included.

Consider next in a similar fashion the population regression of $B_u(r)$ on $B_v(r), B_v^2(r), \ldots, B_v^p(r)$. Due to Gaussianity, $B_{u\cdot v}(r)$ is independent of $B_v(r)$ and thus also of powers of $B_v(r)$, consequently:

$$\begin{aligned}B_u(r) &= \left[B_v(r), B_v^2(r), \ldots, B_v^p(r)\right]\Theta[1:p] + B_{u\cdot v}(r) \\ &= \left[B_v(r), B_v^2(r), \ldots, B_v^p(r)\right]\begin{bmatrix}\Theta[1] \\ 0_{(p-1)\times 1}\end{bmatrix} + B_{u\cdot v}(r),\end{aligned} \tag{36}$$

with obvious notation for $\Theta[1:p]$.

Calculating the regression coefficient as in (35) also leads to this result:

$$\begin{aligned}\Theta[1:p] &:= \left(\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)')\right)^{-1}\mathbb{E}(\mathbf{B}_v(r)B_u(r)) \\ &= \left(\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)')\right)^{-1}\mathbb{E}(\mathbf{B}_v(r)(B_{u\cdot v}(r) + B_v(r)\Omega_{vv}^{-1}\Omega_{vu})) \\ &= \left(\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)')\right)^{-1}\mathbb{E}(\mathbf{B}_v(r)B_v(r))\Omega_{vv}^{-1}\Omega_{vu} \\ &= \begin{bmatrix}\Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p-1)\times 1}\end{bmatrix} = \begin{bmatrix}\Theta[1] \\ 0_{(p-1)\times 1}\end{bmatrix},\end{aligned} \tag{37}$$

using independence of $\mathbf{B}_v(r)$ and $B_{u\cdot v}(r)$ and "partial inversion".[10] Note that for partial inversion to apply here it is necessary that $\mathbf{B}_v(r)$ includes $B_v(r)$ as an element.

[10]With partial inversion we simply denote the obvious fact that for a regular matrix $A \in \mathbb{R}^{n\times n}$ with the columns denoted by $A_{(\cdot,j)}$, it holds that $A^{-1}A_{(\cdot,j)} = e_j^n$.

Now consider the FM-STD transformed errors, compare (13) and its centered counterpart (16), from a similar perspective as for u_t^+ in (32) above:

$$\begin{aligned} u_t^{++} &= u_t - \Delta X_t' \hat{\Omega}_{ww}^{-1} \hat{\Omega}_{wu} \\ &= u_t - w_t' \hat{\Omega}_{ww}^{-1} \hat{\Omega}_{wu} \\ &= u_t - \left[v_t, \Delta x_t^2, \ldots, \Delta x_t^p \right] \hat{\Omega}_{ww}^{-1} \hat{\Omega}_{wu} \\ &\approx u_t - \left[v_t, \frac{2x_t v_t - v_t^2}{T^{1/2}}, \ldots, \frac{p x_t^{p-1} v_t - \frac{p(p-1)}{2} x_t^{p-2} v_t^2}{T^{\frac{p-1}{2}}} \right] \hat{\Omega}_{\tilde{w}\tilde{w}}^{-1} \hat{\Omega}_{\tilde{w}u}, \end{aligned} \tag{38}$$

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 (12).

It can be shown, with the details given in the proofs and by using Itô's Lemma (see, e. g., Theorem 3.3., p. 149 in Karatzas and Shreve, 1991), that the corresponding partial sum limit process converges to:[11]

$$\begin{aligned} B_{u \cdot v}^{++}(r) := B_u(r) - \Bigg[& B_v(r), 2\int_0^r B_v(s) dB_v(s) + r\Omega_{vv}, \ldots, \\ & p\int_0^r B_v^{p-1}(s) dB_v(s) + \frac{p(p-1)}{2} \Omega_{vv} \int_0^r B_v^{p-2}(s) ds \Bigg] \Omega_{\tilde{w}\tilde{w}}^{-1} \Omega_{\tilde{w}u} \\ = B_u(r) - & \left[B_v(r), B_v^2(r), \ldots, B_v^p(r) \right] \Omega_{\tilde{w}\tilde{w}}^{-1} \Omega_{\tilde{w}u}. \end{aligned} \tag{39}$$

We know from Theorem 1 and Corollary 1 that:

$$\Omega_{\tilde{w}\tilde{w}} = \Omega_{vv} \int_0^1 \dot{\mathbf{B}}(r) \dot{\mathbf{B}}(r)' dr, \qquad \Omega_{\tilde{w}u} = \Omega_{vu} \int_0^1 \dot{\mathbf{B}}(r) dr, \tag{40}$$

which implies that

$$\Omega_{\tilde{w}\tilde{w}}^{-1} \Omega_{\tilde{w}u} = \begin{bmatrix} \Omega_{vv}^{-1} \Omega_{vu} \\ 0_{(p-1)\times 1} \end{bmatrix} = \Theta[1:p], \tag{41}$$

[11] We use (38) 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:

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

which in turn implies that $B^{++}_{u\cdot v}(r) = B_{u\cdot v}(r)$. For a partial inversion result to hold for $\Omega_{\tilde{w}\tilde{w}}$ and $\Omega_{\tilde{w}u}$ it is necessary that x_t itself is included as regressor and thus v_t in w_t. Considering the results given in (40) it is clear that for partial inversion to apply, one element of $\dot{\mathbf{B}}(r)$ needs to equal one. This is the case if and only if the integrated process x_t itself is a regressor. In case that x_t itself is not included as a regressor and thus v_t is not included in w_t, partial inversion does not apply as is illustrated in more detail below. It is also interesting to note that $\Omega_{\tilde{w}\tilde{w}}$ and $\Omega_{\tilde{w}u}$ are not equal to $\mathbb{E}(\mathbf{B}_v(r)\mathbf{B}_v(r)')$ and $\mathbb{E}(\mathbf{B}_v(r)B_u(r))$, but the limit features the same "partial inversion" structure as used above in (37).

Hence, as long as x_t itself is included as regressor in the cointegrating polynomial regression, the modification of the dependent variable performed by FM-STD coincides asymptotically with the transformation performed by FM-CPR. In finite samples the FM-STD transformation subtracts in this transformation a number of terms – involving the first differences of powers of x_t and kernel weighted sums converging to random limits – that are asymptotically all equal to zero. Loosely speaking, FM-STD suffers from something like a "degrees of freedom" loss.

The result that also the additive bias correction of FM-STD leads – again under the assumption that x_t itself is included in the regression – to the same result as the FM-CPR bias correction stems again from the structure of the kernel weighted sum limits that allow for "partial inversion", see (C.35) in Appendix C.

The above discussion has already made clear that the equivalence of FM-STD and FM-CPR breaks down when x_t itself is not included in the regression. Nevertheless, it may be informative to see this explicitly. We consider the simple 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-STD estimator does not converge to the limiting distribution given in (31), but to:[12]

$$T^{3/2}(\hat{\beta}^{++}-\beta) \Rightarrow \left(\int_0^1 B_v^4(r)dr\right)^{-1}\left(\int_0^1 B_v^2(r)dB_{u\cdot v}(r)\right. \tag{42}$$
$$+\int_0^1 B_v(r)dr\Omega_{vv}^{-1}\Omega_{vu}\left[\int_0^1 B_v^2(r)dB_v(r)\left(\int_0^1 B_v(r)dr\right)^{-1}\right.$$
$$\left.\left.-\int_0^1 B_v^3(r)dB_v(r)\left(\int_0^1 B_v^2(r)dr\right)^{-1}-\frac{\Omega_{vv}}{2}\right]\right).$$

For this example, the FM-CPR limit distribution (A.5) coincides with the expression in the first line of (42). The terms in the second and third line of (42) 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 therefore 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 $B_v(r)$.

2.4. Shin-Type Cointegration Testing

The asymptotic equivalence result established in Theorem 2 immediately implies that the Shin (1994)-type test of Wagner and Hong (2016, Proposition 5) for the null hypothesis of cointegration in the CPR setting can be based not only on the residuals of FM-CPR but also on the residuals of FM-STD estimation. Both test statistics have the same asymptotic null distribution given in the following corollary.

Corollary 2. *Let the data be generated by* (3) *under Assumptions 1 and 2 and let long-run covariance estimation be carried out under Assumptions 3 and 4. Denote, as in* (16), *the FM-STD residuals with* $\hat{u}_t^{++}$ *and the FM-CPR residuals, defined in* (A.2), *by* $\hat{u}_t^{+}$. *Then it holds that both:*

$$CT^{++} := \frac{1}{T\hat{\omega}_{u\cdot w}}\sum_{t=1}^{T}\left(\frac{1}{T^{1/2}}\sum_{j=1}^{t}\hat{u}_j^{++}\right)^2, \tag{43}$$

[12] The relevant terms for the specific case of (20) and (21) for the 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 $GA^{**} \Rightarrow 2\Delta_{vu}^{+}\int_0^1 B_v(r)dr$.

and

$$CT^{+} := \frac{1}{T\hat{\omega}_{u\cdot v}} \sum_{t=1}^{T} \left(\frac{1}{T^{1/2}} \sum_{j=1}^{t} \hat{u}_j^{+} \right)^2 \tag{44}$$

converge for $T \to \infty$ to

$$\int_0^1 \left(W_{u\cdot v}^{J_W}(r) \right)^2 dr, \tag{45}$$

with

$$W_{u\cdot v}^{J_W}(r) := W_{u\cdot v}(r) - \int_0^r J^W(s)'ds \left(\int_0^1 J^W(s)J^W(s)'ds \right)^{-1} \int_0^1 J^W(s)dW_{u\cdot v}(s), \tag{46}$$

where $J^W(r) := [D(r)', W_v(r), W_v^2(r), \ldots, W_v^p(r)]'$. Under the stated assumptions both $\hat{\omega}_{u\cdot v}$ and $\hat{\omega}_{u\cdot w}$ are consistent estimators of $\omega_{u\cdot v} := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$, the variance of $B_{u\cdot v}(r)$.

The limiting distribution given in (45) and (46) 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.

In the multiple integrated regressor CPR case, full design need not necessarily prevail. In this case the result of Corollary 2 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. Their Proposition 6, formulated for the FM-CPR residuals, extends to the FM-STD residuals as well.

As outlined in the introduction, the EKC literature using the Shin (1994) test uses the critical values corresponding to a specification with p integrated regressors, i. e., quantiles corresponding to a limiting distribution similar to (45) and (46) in format, but with $W_{u\cdot v}^{J_{W_p}}(r)$ and $J^{W_p}(r) := [D(r)', W_1(r), \ldots, W_p(r)]'$, where $W_i(r)$ are independent standard Brownian motions for $i = 1, \ldots, p$, in place of $W_{u\cdot v}^{J_W}(r)$ and $J^W(r)$. In other words the limiting distribution used is a function of p independent standard Brownian motions rather than of p powers of a single standard Brownian motion. Clearly, this makes a difference, as seen in Table 1. The table illustrates that the differences become

	$D_t = \emptyset$			$D_t = 1$			$D_t = [1, t]'$		
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

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 (2013, Table 4). 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 ($D_t = [1, t]'$).

bigger when the regression model becomes more complicated, i. e., when more powers of the integrated regressor are included. Using the FM-STD residuals in conjunction with the Shin (1994) critical values therefore leads to invalid inference even asymptotically.

3. Finite Sample Performance

In our simulations we use exactly the same data generating processes (DGPs) as Wagner and Hong (2016, Section 3), i. e., we generate data for the quadratic cointegrating polynomial regression model:

$$y_t = c + \delta t + \beta_1 x_t + \beta_2 x_t^2 + u_t, \tag{47}$$

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

$$\begin{aligned} u_t &= \rho_1 u_{t-1} + \varepsilon_t + \rho_2 e_t, \quad u_0 = 0, \\ v_t &= e_t + 0.5 e_{t-1}, \end{aligned}$$

with $(\varepsilon_t, e_t)' \sim \mathcal{N}(0, I_2)$. The parameter ρ_1 controls the level of serial correlation in the error term u_t and ρ_2 controls the extent of regressor endogeneity. The parameter values are 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 the bandwidth we report results for three choices. These are the data-dependent rules of Andrews (1991), labelled And, 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.[13] The parameter hypothesis test results are "benchmarked" against OLS-based 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 performed are carried out using the chi-squared distribution.[14]

We start the discussion of the results by comparing bias and RMSE of the two estimators. The results for β_1 are given in Table 2 as ratios, with FM-STD divided by FM-CPR, since we are primarily interested in the comparison of the two 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-STD 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-STD and FM-CPR 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

[13] 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 to be estimated 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 4, but has no immediate effect on finite sample performance.

[14] 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-STD). 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.

ρ_1, ρ_2	Bias Ratio			RMSE Ratio		
	And	NW	NW_T	And	NW	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 2: Bias and RMSE ratios, FM-STD/FM-CPR, for β_1.

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-STD and FM-CPR. Given the absolute disadvantage of NW_T, especially in cases 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-STD 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.[15] 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-STD relative to FM-CPR is less severe when using the Andrews (1991) bandwidth than when using the Newey and West (1994) bandwidth.

[15]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-STD			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 3: Empirical null rejection probabilities of Wald-type tests for $H_0\colon \beta_1 = 5, \beta_2 = -0.3$.

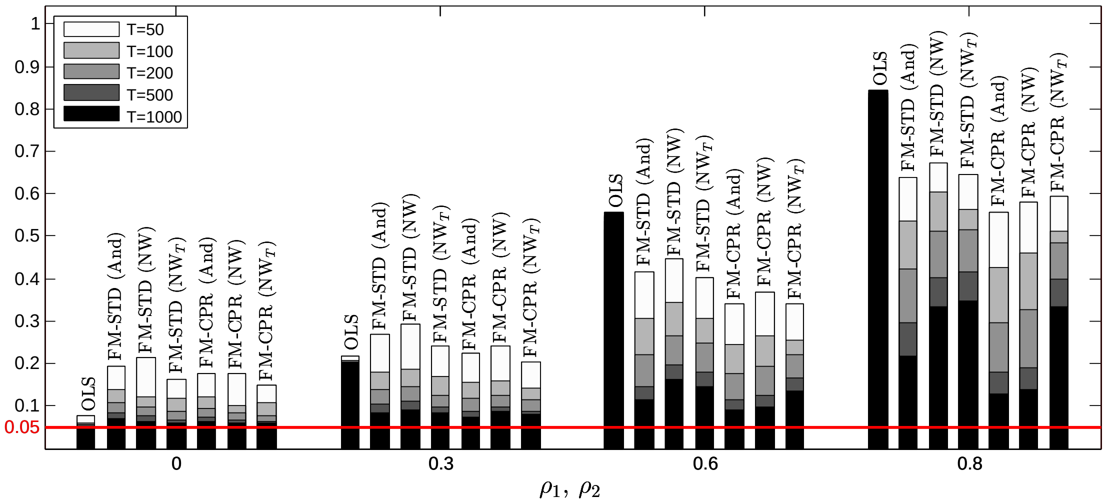

Figure 2: Empirical null rejection probabilities of Wald-type tests for $H_0\colon \beta_1 = 5, \beta_2 = -0.3$.

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 3 and Figure 2 that contain the same information presented differently. 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-STD 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 case of regressor endogeneity and/or error serial correlation, see, e. g., Hong and Phillips (2010, Theorem 2). In our setting the Andrews (1991) rule leads mostly to slightly better results than the Newey and West (1994) rule. The sample-size dependent bandwidth $\mathrm{NW_T}$ performs – as expected – especially poor in 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 features.

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 3 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

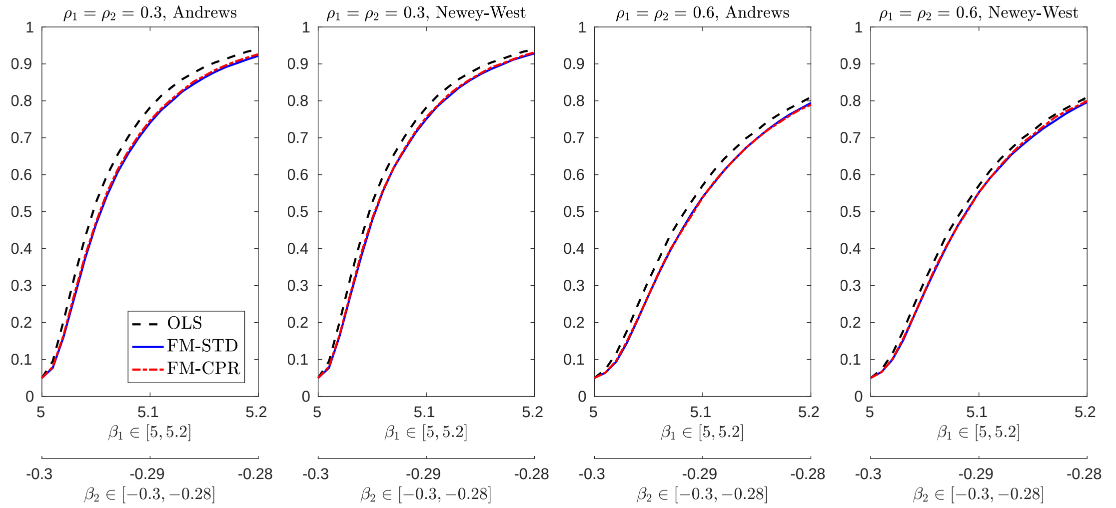

Figure 3: Size-corrected power of Wald-type tests for $H_0 : \beta_1 = 5, \beta_2 = -0.3$ for $T = 100$. The two left graphs correspond to $\rho_1 = \rho_2 = 0.3$ and the two right graphs to $\rho_1 = \rho_2 = 0.6$. Within these pairs the left graph corresponds to the Andrews (1991) bandwidth and the right one to the Newey and West (1994) bandwidth.

corresponds to the Andrews (1991) bandwidth and the right one to the Newey and West (1994) bandwidth.

Figure 3 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-STD 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 best performance for parameter hypothesis testing is obtained with the bandwidth rule of Andrews (1991).

Let us now turn briefly to cointegration testing. We report in Table 4 the null rejection probabilities for the test discussed in Section 2.4. 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-STD residuals in conjunction with the (inappropriate) Shin (1994) 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 (45) and (46); tabulated in Wagner (2013, Table 4); with the critical values required for our setting also available in Table 1 in this paper. The second column,

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 4: Empirical null rejection probabilities of cointegration tests. The block-column CT^{++}_{Shin} reports the results from using the test statistic (43) and the Shin (1994) critical values. The block-columns CT^{++} and CT^{+} report the results from using (43) based on either the FM-STD residuals or the FM-CPR residuals and the corresponding critical value tabulated in Wagner (2013, Table 4). For the considered specification the 5% critical values are 0.101 (Shin) and 0.106 (Wagner) respectively, compare also Table 1.

CT^{++}, is a "hybrid" version based on the asymptotic result given in Corollary 2. This test statistic is calculated from the FM-STD residuals, but the decisions are based on the asymptotically correct critical values.

The null performance of the different cointegration test versions can be summarized as follows: The $\text{CT}^{++}_{\text{Shin}}$-test typically exhibits the largest over-rejections. These over-rejections, that stay substantial even for $T = 1000$, reflect that wrong critical values are used. The hybrid CT^{++}-test exhibits a performance very similar to the $\text{CT}^{++}_{\text{Shin}}$-test. This is partly not surprising, since the same test statistic is used and the critical values differ only marginally in the considered specification (compare Table 1). Thus, the findings cannot differ too much. Another reason for the poor performance of CT^{++} is that it suffers from the poor performance of the estimator $\hat{\omega}_{u\cdot w}$ mentioned in Footnote 14. This effect results in poor performance even when comparing the statistic with the correct critical values. The performance of the CT^{+}-test is substantially better, with a performance margin that widens 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 case of large ρ_1, ρ_2 and all values of T. From the two data-dependent bandwidths better – partly substantially better - results are obtained with the Andrews (1991) bandwidth.

We close this section by considering the power performance of the cointegration test variants, where we consider 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 alternative considered. In particular, size-corrected power is much larger for alternatives (II) and

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

(III) than for alternative (I), especially when considering large values of ρ_1, ρ_2. Using the Andrews (1991) bandwidth leads for alternatives (II) and (III) 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 respectively cointegration tests when 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 much 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, 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.

4. Summary and Conclusions

This paper shows that the "standard" Phillips-Hansen FM-OLS estimator – when used in the way described in cointegrating polynomial regressions – leads to the same asymptotic distribution as the FM-CPR estimator of Wagner and Hong (2016). A convenient implication of this result is, from an asymptotic perspective, that the standard FM-OLS estimator of Phillips and Hansen (1990), implemented in many software packages, can be used for estimation and inference not only for cointegrating linear regressions but also for cointegrating polynomial regressions. The asymptotic equivalence of the two estimators immediately implies also asymptotic equivalence not only of parameter hypothesis tests but also of the Shin (1994)-type cointegration tests based on either the FM-STD or FM-CPR residuals. In this respect, however, it is important to use appropriate critical values that differ from those of Shin (1994). The usage of the latter leads to invalid inference even asymptotically. Furthermore, it may be potentially comforting for practitioners that our result asymptotically "rehabilitates" widespread practice in the environmental Kuznets curve and related literatures, which routinely performs FM-STD estimation in quadratic or cubic CPRs.

A very important restriction for the equivalence result of FM-STD and FM-CPR to hold is that the integrated regressor x_t itself is – or in the general case all components of the integrated regressor vector x_t are – included in the regression. This stems from the fact that only in this case orthogonalization between $B_u(r)$ and $B_v(r)$ can be performed by the first stage modifications of *both* fully modified type estimators, as discussed in Section 2.3.[16]

The other key ingredient, in addition to the orthogonalization aspects, for deriving asymptotic equivalence of the FM-STD and FM-CPR estimators are weak convergence results for kernel weighted sums ("long-run covariance" estimators) for processes involving properly scaled powers of integrated regressors. In particular we show that these converge to the quadratic covariation of the limiting process corresponding to the errors and regressors. Such results prove useful also elsewhere, e. g., in the analysis of Phillips (1987)-type unit root tests applied to polynomials of integrated processes, see Stypka and Wagner (2018).

The finite sample simulations indicate, as expected, performance advantages of FM-CPR over FM-STD along all considered dimensions. These occur in the case of large endogeneity and error serial correlation even for $T = 1000$. Smaller levels of endogeneity and error serial correlation lead to smaller performance differences throughout.

The results and observations of this paper naturally lead to the questions whether similar results extend to other modified least squares estimators used in the cointegration literature and whether similar results extend to more general nonlinear cointegration settings. With respect to the first question let us note that preliminary back-of-the-envelope calculations indicate that it might be easier to obtain similar results for the IM-OLS estimator of Vogelsang and Wagner (2014) than for the D-OLS estimator of Phillips and Loretan (1991), Saikkonen (1991) or Stock and Watson (1993). With respect to the types of nonlinearities to which such results may potentially extend let us note that linearity in parameters, albeit of course a key aspect, need not be sufficient. In case of, e. g., I-regular rather H-regular (including polynomials) nonlinear functions (in the terminology of Park and Phillips, 2001), the limiting distributions involve local times and the type of asymptotic orthogonality results used here may not hold.

[16]From a practical perspective, if one is inclined to use FM-STD this is, of course, not a real limitation as one can always include an integrated regressor itself even if it is not necessary.

Acknowledgments

All authors acknowledge financial support from the Collaborative Research Center 823: *Statistical Modelling of Nonlinear Dynamic Processes* supported by Deutsche Forschungsgemeinschaft (DFG). Martin Wagner furthermore acknowledges research support from the Jubiläumsfonds of the Oesterreichische Nationalbank via several grants. The helpful comments of participants at the German Probability and Statistics Days 2016 in Bochum, the Computational and Financial Econometrics Conference 2016 in Sevilla, the European Meeting of Statisticians 2017 in Helsinki, the 2018 NBER/NSF Time Series Conference in San Diego and of seminar participants at the University of Nuremberg are gratefully acknowledged. At the time of the completion of this manuscript Martin Wagner serves as Chief Economist of the Bank of Slovenia, however, this paper solely reflects the views of the authors and not necessarily those of the Bank of Slovenia or the European System of Central Banks. On top of this the usual disclaimer applies.

References

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

Chan, N. and Q. Wang (2015). Nonlinear Regressions with Nonstationary Time Series. *Journal of Econometrics* **185**, 182–195.

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

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

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

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

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

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.

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

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

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.

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

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

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

Le Gall, J.F. (2016). *Brownian Motion, Martingales, and Stochastic Calculus.* Springer International Publishing Switzerland.

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.

Malenbaum, W. (1978). *World Demand for Raw Materials in 1985 and 2000.* New York: McGraw-Hill, EMJ Mining Information Services.

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

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

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

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

Phillips, P.C.B. (1987). Time Series Regression with a Unit Root. *Econometrica* **55**, 277–301.

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

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

Phillips, P.C.B. and M. Loretan (1991). Estimating Long Run Economic Equilibria. *Review of Economic Studies* **58**, 407–436.

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

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

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

Stock, J.H. and M.W. Watson (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* **76**, 175–194.

Stypka, O. and M. Wagner (2018). The Phillips Unit Root Tests for Polynomials of Integrated Processes. Mimeo.

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

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

Wagner, M. (2013). Residual Based Cointegration and Non-Cointegration Tests for Cointegrating Polynomial Regressions. *Mimeo.*

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

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

Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression.* Singapore: World Scientific.

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

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

APPENDIX A: 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 linear cointegration to the cointegrating polynomial regression (CPR) case. This estimator, briefly described next, is referred to as FM-CPR in this paper. As 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

$$y_t^+ := y_t - \Delta x_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}. \tag{A.1}$$

This transformation dynamically orthogonalizes the limit partial sum process of the modified errors

$$u_t^+ := u_t - v_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu}, \tag{A.2}$$

i. e., $B_{u \cdot v}(r)$ as defined in Theorem 2 below (31), from the limiting process corresponding to x_t, i. e., $B_v(r)$. In 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 specification (3) the bias correction term is given by:

$$A^* := \hat{\Delta}_{vu}^+ \begin{bmatrix} 0_{q \times 1} \\ T \\ 2\sum_{t=1}^T x_t \\ \vdots \\ p\sum_{t=1}^T x_t^{p-1} \end{bmatrix}, \tag{A.3}$$

with $\hat{\Delta}^+_{vu} := \hat{\Delta}_{vu} - \hat{\Delta}_{vv}\hat{\Omega}^{-1}_{vv}\hat{\Omega}_{vu}$. Defining $y^+ := [y^+_1, \ldots, y^+_T]'$ and $Z := [Z_1, \ldots, Z_T]'$, leads to the FM-CPR estimator of θ given by:

$$\hat{\theta}^+ := (Z'Z)^{-1}(Z'y^+ - A^*). \tag{A.4}$$

Wagner and Hong (2016, Proposition 1) show, under slightly weaker assumptions than considered in this paper, that:

$$G^{-1}(\hat{\theta}^+ - \theta) \Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \int_0^1 J(r)dB_{u\cdot v}(r), \tag{A.5}$$

This is exactly the same distribution as derived for the FM-STD estimator in Theorem 2.

APPENDIX B: Auxiliary Lemmata

This appendix contains some auxiliary lemmata, required for showing the main results of the paper. The following Lemmata 2 and 3 draw upon some ideas used in the proofs of Kasparis (2008, Lemma A1). Lemma 1 is identical to Kasparis (2008, Lemma A1(i)).

Lemma 1. *Under Assumption 2 it holds for* $0 \leq b < 1/3$ *that:*

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

Lemma 2. *Under Assumptions 2 to 4 it holds for all integers* $0 \leq p$ *and* $1 \leq q$ *that:*

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

Proof. 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 (10). 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

$$\sup_{r\in[0,1]} T^{-1/2}|x_{\lfloor rT \rfloor}| \leq \tilde{C} + o_{a.s.}(1). \tag{B.3}$$

Furthermore, it holds that

$$\begin{aligned}
&\sup_{r\in[0,1]} \sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT \rfloor+h} - x_{\lfloor rT \rfloor}| \\
&\quad = \sup_{r\in[0,1]} \sup_{0\leq h\leq M_T} T^{-1/2}\Big|\sum_{m=1}^{h} v_{\lfloor rT \rfloor+m}\Big| \leq \sup_{r\in[0,1]} T^{-1/2}\sum_{m=1}^{M_T} |v_{\lfloor rT \rfloor+m}|
\end{aligned}$$

and thus it follows from Lemma 1 that

$$\sup_{r\in[0,1]} \sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT \rfloor+h} - x_{\lfloor rT \rfloor}| = o_{a.s.}(1). \tag{B.4}$$

This implies

$$\begin{aligned}
&\sup_{r\in[0,1]} \sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT \rfloor+h}| \\
&\quad \leq \sup_{r\in[0,1]} \sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT \rfloor+h} - x_{\lfloor rT \rfloor}| + \sup_{r\in[0,1]} T^{-1/2}|x_{\lfloor rT \rfloor}| \leq C + o_{a.s.}(1),
\end{aligned}$$

with $C := \sup_{r\in[0,1]} |B_v(r)| + 1$ and also

$$\sup_{r\in[0,1]} \sup_{0\leq h\leq M_T} T^{-1/2}|\overline{x}^h_{\lfloor rT \rfloor}| \leq C + o_{a.s.}(1). \tag{B.5}$$

Using the triangular inequality and the bounds given in (B.3)–(B.5), the following inequalities hold:

$$\begin{aligned}
&\left|\frac{1}{T^{1/2}}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1}\sum_{m=1}^{h} v_t v_{t+m} v_{t+h}\right| \qquad \text{(B.6)}\\
&\quad\le \left(\frac{M_T^3}{T}\right)^{1/2}\frac{1}{M_T}\sum_{h=0}^{M_T}\left|k\left(\frac{h}{M_T}\right)\right|\frac{1}{T}\sum_{t=1}^{T-h}\left|\left(\frac{x_t}{T^{1/2}}\right)^p\left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1}\right||v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{h} v_{t+m}\right|\\
&\quad\le \left(\frac{M_T^3}{T}\right)^{1/2}\overline{k}(0)C^{p+q-1}\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{h} v_{t+m}\right| + o_{\mathbb{P}}(1),
\end{aligned}$$

with $\overline{k}(0) = \sup_{x\ge 0}|k(x)|$ as defined in Assumption 3. Consider next

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

There, it holds that

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

and thus we have

$$\mathbb{E}\left(\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{h} v_{t+m}\right|\right) = O_{\mathbb{P}}(1).$$

Finally, the assertion is an immediate consequence of $M_T^3/T \to 0$ by Assumption 4, and the remaining terms contained in the expression in (B.6) being $O_{\mathbb{P}}(1)$. ■

Lemma 3. *With Assumptions 2 to 4 in place it holds for all integers* $0 \leq p$ *that:*

$$\left| \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p (v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]) \right| = o_{\mathbb{P}}(1). \tag{B.7}$$

Proof. In the proof of Lemma A1(iv) in Kasparis (2008) it is shown that

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

by showing that

$$\sup_{0 \leq h \leq M_T} \left| \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p \sum_{m=1}^{h} (v_t v_{t+m} - \mathbb{E}[v_t v_{t+m}]) \right| = o_{\mathbb{P}}(1). \tag{B.8}$$

The left-hand side of (B.7) can be written as

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

Using similar arguments as Kasparis (2008, p. 1394–1396) to show (B.8), corresponding to his Equation (A.7), it follows that

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

which implies the claim of this lemma, since

$$\begin{aligned} &\left| \frac{1}{M_T} \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p M_T (v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]) \right| \\ &\qquad \leq \tilde{k} \left| \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p M_T (v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]) \right|, \end{aligned}$$

with $\tilde{k} := \overline{k}(0) + 1$. Since we use arguments of Kasparis (2008), the same moment and bandwidth assumptions are required and are therefore contained in our Assumptions 2 to 4. ■

APPENDIX C: Proofs of the Main Results

Proof of Theorem 1. First, the $(1,1)$-element of $\tilde{\Delta}_{\eta\eta}$ is given by

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

cf. Remark 2. For $i \in \{1,\dots,p\}$ it holds that

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

i. e., for the first and second columns (and rows) 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

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

From Lemma 1 we know that $T^{-1/2}v_{\lfloor rT\rfloor} = o_{a.s.}(1)$. Additionally, it holds that $T^{-1/2}|x_{\lfloor rT\rfloor}| \leq C + o_{a.s.}(1)$. From $\mathbb{E}[T^{-1/2}v_{\lfloor rT\rfloor}^2] = T^{-1/2}\Sigma_{vv} \to 0$ for all $r \in [0,1]$ we conclude that

$$\frac{\Delta x_t^i}{T^{(i-1)/2}} = i\left(\frac{x_t}{T^{1/2}}\right)^{i-1} v_t + O_{\mathbb{P}}(T^{-1/2}). \tag{C.1}$$

The kernel is bounded and $M_T = o(T^{1/3})$ by assumption, hence it follows that

$$\tilde{\Delta}_{\tilde{w}\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h} \dot{X}_{t,T}\dot{X}_{t+h,T}' v_t v_{t+h} + o_{\mathbb{P}}(1), \tag{C.2}$$

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}, cf. Remark 2 again. From Lemma 2 we get

$$\tilde{\Delta}_{\tilde{w}\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \dot{X}_{t,T} \dot{X}_{t,T}' v_t v_{t+h} + o_{\mathbb{P}}(1) \tag{C.3}$$

and from Lemma 3 it follows that

$$\tilde{\Delta}_{\tilde{w}\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \frac{1}{T} \sum_{t=1}^{T-h} \dot{X}_{t,T} \dot{X}_{t,T}' + o_{\mathbb{P}}(1). \tag{C.4}$$

Now we consider

$$\begin{aligned} &\left\| \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \frac{1}{T} \sum_{t=T-h+1}^{T} \dot{X}_{t,T} \dot{X}_{t,T}' \right\| \qquad \text{(C.5)} \\ &\quad \leq \sum_{h=0}^{M_T} \left| k\left(\frac{h}{M_T}\right) \right| |\mathbb{E}[v_0 v_h]| \frac{1}{T} \sum_{t=T-h+1}^{T} \left\| \dot{X}_{t,T} \dot{X}_{t,T}' \right\| \\ &\quad \leq C^* \frac{1}{T} \sum_{h=0}^{M_T} \left| k\left(\frac{h}{M_T}\right) \right| |\mathbb{E}[v_0 v_h]| h. \end{aligned}$$

Similar arguments as in the proof of Jansson (2002, Lemma 6) imply that $\frac{1}{T} \sum_{h=0}^{M_T} \left| k\left(\frac{h}{M_T}\right) \right| |\mathbb{E}[v_0 v_h]| h$ is $o(1)$, while $C^* = O_{\mathbb{P}}(1)$. Thus, it follows that

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \frac{1}{T} \sum_{t=T-h+1}^{T} \dot{X}_{t,T} \dot{X}_{t,T}' = o_{\mathbb{P}}(1). \tag{C.6}$$

Therefore, we obtain

$$\tilde{\Delta}_{\tilde{w}\tilde{w}} = \left(\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \right) \left(\frac{1}{T} \sum_{t=1}^{T} \dot{X}_{t,T} \dot{X}_{t,T}' \right) + o_{\mathbb{P}}(1). \tag{C.7}$$

For the first term it holds that

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \to \Delta_{vv}. \tag{C.8}$$

Hence, using Slutsky's Theorem, cf. e. g., Davidson (1994, Theorem 18.10, p. 286), we obtain

$$\tilde{\Delta}_{\tilde{w}\tilde{w}} \Rightarrow \Delta_{vv} \int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr. \tag{C.9}$$

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

$$\langle \mathcal{B}(r), \mathcal{B}(r)\rangle_0^1 = \begin{bmatrix} \langle B_u(r), B_u(r)\rangle_0^1 & \langle B_u(r), \mathbf{B}_v(r)\rangle_0^1 \\ \langle \mathbf{B}_v(r), B_u(r)\rangle_0^1 & \langle \mathbf{B}_v(r), \mathbf{B}_v(r)\rangle_0^1 \end{bmatrix}. \tag{C.10}$$

Due to symmetry 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 $\Omega_{\eta\eta}$, for which we have shown above that it is equal to $\Omega_{vv} \int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr$. The elements of this matrix are given by

$$\Omega_{vv} ij \int_0^1 B_v^{i+j-2}(r)dr, \qquad i, j = 1, \ldots, p. \tag{C.11}$$

Now consider the definition of the quadratic covariation given at the end of the introduction, i. e.,

$$\langle \mathbf{B}_v(r), \mathbf{B}_v(r)\rangle_0^1 = \mathbf{B}_v(1)\mathbf{B}_v(1)' - \int_0^1 \mathbf{B}_v(r)d\mathbf{B}_v(r)' - \left(\int_0^1 \mathbf{B}_v(r)d\mathbf{B}_v(r)'\right)'. \tag{C.12}$$

Next, by using Itô's Lemma, e. g., in the formulation given in Le Gall (2016, Theorem 5.10, p. 113) we know that

$$\mathbf{B}_v(r) = \int_0^r \dot{\mathbf{B}}_v(s)dB_v(s) + \frac{\Omega_{vv}}{2}\int_0^r \ddot{\mathbf{B}}_v(s)ds, \tag{C.13}$$

with $\ddot{\mathbf{B}}_v(r) := [0, 2, \ldots, p(p-1)B_v^{p-2}(r)]'$. Substituting this into (C.12) leads to

$$\begin{aligned}\langle \mathbf{B}_v(r), \mathbf{B}_v(r)\rangle_0^1 = \mathbf{B}_v(1)\mathbf{B}_v(1)' &- \int_0^1 \left(\mathbf{B}_v(r)\dot{\mathbf{B}}_v(r)' + \dot{\mathbf{B}}_v(r)\mathbf{B}_v(r)'\right) dB_v(r) \\ &- \frac{\Omega_{vv}}{2}\int_0^1 \left(\mathbf{B}_v(r)\ddot{\mathbf{B}}_v(r)' + \ddot{\mathbf{B}}_v(r)\mathbf{B}_v(r)'\right) dr.\end{aligned} \tag{C.14}$$

The (i,j)-element, $i,j = 1,\ldots,p$, of (C.14) is given by

$$B_v^{i+j}(1) - (i+j)\int_0^1 B_v^{i+j-1}(r)dB_v(r) - \frac{\Omega_{vv}}{2}\left(i(i-1)+j(j-1)\right)\int_0^1 B_v^{i+j-2}(r)dr. \tag{C.15}$$

Using Itô's Lemma once again, i. e.,

$$B_v^{i+j}(1) - (i+j)\int_0^1 B_v^{i+j-1}(r)dB_v(r) = \frac{\Omega_{vv}}{2}(i+j)(i+j-1)\int_0^1 B_v^{i+j-2}(r)dr, \tag{C.16}$$

shows upon simplifying terms equality of the (i,j)-elements in (C.11) and (C.15).

Now consider the off diagonal block, to show equality of $\Omega_{uv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr$ and $\langle B_u(r), \mathbf{B}_v(r)\rangle_0^1$. By definition of $B_{u\cdot v}(r)$ it holds that

$$\langle B_u(r), \mathbf{B}_v(r)\rangle_0^1 = \langle B_{u\cdot v}(r), \mathbf{B}_v(r)\rangle_0^1 + \langle B_v(r), \mathbf{B}_v(r)\rangle_0^1 \Omega_{vv}^{-1}\Omega_{vu}, \tag{C.17}$$

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

$$\begin{aligned}\langle B_v(r), \mathbf{B}_v(r)\rangle_0^1 \Omega_{vv}^{-1}\Omega_{vu} &= \Omega_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)'dr\Omega_{vv}^{-1}\Omega_{vu} \\ &= \Omega_{vu}\int_0^1 \dot{\mathbf{B}}_v(r)'dr.\end{aligned} \tag{C.18}$$

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 i-th element, $i = 1,\ldots,p$, using Itô's Lemma to arrive at the second equality below

$$\begin{aligned}\langle B_{u\cdot v}(r), B_v^i(r)\rangle_0^1 &= B_{u\cdot v}(1)B_v^i(1) - \int_0^1 B_{u\cdot v}(r)dB_v^i(r) - \int_0^1 B_v^i(r)dB_{u\cdot v}(r) \\ &= B_{u\cdot v}(1)B_v^i(1) - i\int_0^1 B_{u\cdot v}(r)B_v^{i-1}(r)dB_v(r) \\ &\quad - \frac{\Omega_{vv}}{2}i(i-1)\int_0^1 B_{u\cdot v}(r)B_v^{i-2}(r)dr - \int_0^1 B_v^i(r)dB_{u\cdot v}(r).\end{aligned} \tag{C.19}$$

Finally, using once again Itô's Lemma in the formulation of Le Gall (2016, Theorem 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) shows that

$$B_{u\cdot v}(1)B_v^i(1) = \int_0^1 B_v^i(r)dB_{u\cdot v}(r) + i\int_0^1 B_{u\cdot v}(r)B_v^{i-1}(r)dB_v(r) + \frac{\Omega_{vv}}{2}i(i-1)\int_0^1 B_{u\cdot v}(r)B_v^{i-2}(r)dr, \tag{C.20}$$

which upon combining terms establishes the required zero quadratic covariation. ■

Proof of Corollary 1. 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 (3). Similar to the proof of Theorem 1 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 converges in distribution to $\Delta_{u\tilde{w}}$ by Theorem 1. Therefore, it remains to show that the second term is $o_{\mathbb{P}}(1)$. Similar arguments as in the proof of Theorem 1 imply that

$$\begin{aligned}&\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}Z_t'(\hat{\theta}-\theta)\tilde{w}_{t+h} \\ &= (\hat{\theta}-\theta)'G^{-1}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T^{3/2}}\sum_{t=1}^{T-h}T^{1/2}GZ_t\dot{X}_{t+h,T}v_{t+h} + o_{\mathbb{P}}(1) \\ &= (\hat{\theta}-\theta)'G^{-1}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T^{3/2}}\sum_{t=1}^{T-h}T^{1/2}GZ_t\dot{X}_{t+h,T}v_{t+h} + o_{\mathbb{P}}(1) \\ &= (\hat{\theta}-\theta)'G^{-1}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T^{3/2}}\sum_{t=1}^{T}T^{1/2}GZ_t\dot{X}_{t,T}v_{t} + o_{\mathbb{P}}(1),\end{aligned} \tag{C.21}$$

with G defined in (18). As $G(\hat{\theta}-\theta) = O_{\mathbb{P}}(1)$ and

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T^{3/2}}\sum_{t=1}^{T}T^{1/2}GZ_t\dot{X}_{t,T}v_t = O_{\mathbb{P}}(T^{-1}) \tag{C.22}$$

the expressions (C.21) is $o_{\mathbb{P}}(1)$. This implies that

$$\hat{\Delta}_{u\tilde{w}} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h} u_t\tilde{w}_{t+h} + o_{\mathbb{P}}(1), \tag{C.23}$$

from which the claim follows. ■

Proof of Proposition 1. We start with considering $G_X \sum_{t=1}^{T} X_t \Delta X_t' G_W$ and define

$$\ddot{X}_t := [0, 2, \ldots, p(p-1)x_t^{p-2}]'.$$

Using this notation it is straightforward to show that

$$G_X \sum_{t=1}^{T} X_t \Delta X_t' G_W = G_X \sum_{t=1}^{T} X_t \dot{X}_t' v_t G_W + G_X \sum_{t=1}^{T} X_t \ddot{X}_t' \frac{v_t^2}{2} G_W + o_{\mathbb{P}}(1). \tag{C.24}$$

Similar as in Wagner and Hong (2016, Proposition 1), but written slightly differently to more directly be able to use Itô's Lemma below, it holds for the first term on the right hand side above that:

$$\begin{aligned} G_X \sum_{t=1}^{T} X_t \dot{X}_t' v_t G_W \Rightarrow & \int_0^1 \mathbf{B}_v(r)\dot{\mathbf{B}}_v(r)' dB_v(r) \\ & + \Delta_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr + \Delta_{vv}\int_0^1 \mathbf{B}_v(r)\ddot{\mathbf{B}}_v(r)'dr. \end{aligned} \tag{C.25}$$

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

$$G_X \sum_{t=1}^{T} X_t \ddot{X}_t' v_t^2 G_W \Rightarrow \Sigma_{vv}\int_0^1 \mathbf{B}_v(r)\ddot{\mathbf{B}}_v(r)'dr. \tag{C.26}$$

Combining (C.25) and (C.26) leads to

$$\begin{aligned} G_X \sum_{t=1}^{T} X_t \Delta X_t' G_W \Rightarrow & \int_0^1 \mathbf{B}_v(r)\dot{\mathbf{B}}_v(r)' dB_v(r) + \frac{\Omega_{vv}}{2}\int_0^1 \mathbf{B}_v(r)\ddot{\mathbf{B}}_v(r)'dr \\ & + \Delta_{vv}\int_0^1 \dot{\mathbf{B}}_v(r)\dot{\mathbf{B}}_v(r)'dr. \end{aligned} \tag{C.27}$$

Using Itô's Lemma shows that the first two terms on the right hand side of (C.27) are together equal to $\int_0^1 \mathbf{B}_v(r)d\mathbf{B}_v(r)'$. With respect to the third term, Theorem 1 shows that it is up to the constant $\Delta_{vv}\Omega_{vv}^{-1}$ equal to $\langle \mathbf{B}(r), \mathbf{B}(r)\rangle_0^1$.

It remains to consider

$$G_X \sum_{t=1}^{T} D_t \Delta X_t' G_W \Rightarrow \int_0^1 D(r)\dot{\mathbf{B}}_v(r)' dB_v(r) + \frac{\Omega_{vv}}{2}\int_0^1 \mathbf{B}_v(r)\ddot{\mathbf{B}}_v(r)'dr \tag{C.28}$$
$$= \int_0^1 D(r)d\mathbf{B}_v(r)'.$$

The convergence result follows from similar considerations as above and the second equality follows from Itô's Lemma. ■

Proof of Theorem 2. Consider the two terms given in the last line of (20). From the proof of Wagner and Hong (2016, Proposition 1) it is known that

$$GZ'u \Rightarrow \int_0^1 J(r)dB_u(r) + \Delta_{vu}\begin{pmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}(r)dr \end{pmatrix}. \tag{C.29}$$

The asymptotic behavior of $GZ'\tilde{W}$ has been established in Proposition 1. The first column, corresponding to the first component v_t of $\tilde{w}_t$, of this limit is given by

$$GZ'v \Rightarrow \int_0^1 J(r)dB_v(r) + \Delta_{vv}\begin{pmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}(r)dr \end{pmatrix}, \tag{C.30}$$

which is also a well-known result, compare again Wagner and Hong (2016, Proposition 1). The reason that only the first column is needed is the following result concerning the limit of $\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u}$. In the single integrated regressor case with Ω_{vv} scalar, it is clear that $\Omega_{\tilde{w}\tilde{w}} = \Omega_{vv}\Pi_v$, with

$$\Pi_v := \int_0^1 \dot{\mathbf{B}}(r)\dot{\mathbf{B}}(r)'dr. \tag{C.31}$$

From Theorem 1 and Corollary 1 we know that $\hat{\Omega}_{\tilde{w}\tilde{w}} \Rightarrow \Omega_{vv}\Pi_v$ and $\hat{\Omega}_{\tilde{w}u} \Rightarrow \Omega_{vu}\Pi_v e_1^p$, which implies

$$\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \xrightarrow{\mathbb{P}} \Omega_{vv}^{-1}\Omega_{vu}e_1^p. \tag{C.32}$$

Proposition 1 in combination with (C.32) gives:

$$GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \Rightarrow \int_0^1 J(r)dB_v(r)\Omega_{vv}^{-1}\Omega_{vu} + \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu}\begin{pmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}(r)dr \end{pmatrix}. \tag{C.33}$$

It remains to consider GA^{**}, for which we find

$$GA^{**} = \begin{bmatrix} 0_{q\times 1} \\ \hat{\Delta}_{\tilde{w}u}^{+} \end{bmatrix} \Rightarrow \Delta_{vu}^{+}\begin{bmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}(r)dr \end{bmatrix}, \tag{C.34}$$

which follows from

$$\begin{aligned} \hat{\Delta}_{\tilde{w}u}^{+} =& \hat{\Delta}_{\tilde{w}u} - \hat{\Delta}_{\tilde{w}\tilde{w}}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \\ \Rightarrow& \Delta_{vu}\int_0^1 \dot{\mathbf{B}}(r)dr - \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu}\Pi_v\begin{bmatrix} 1 \\ 0_{(p-1)\times 1} \end{bmatrix} = \Delta_{vu}^{+}\int_0^1 \dot{\mathbf{B}}(r)dr. \end{aligned} \tag{C.35}$$

Combining all terms from (20) we arrive at

$$\begin{aligned} &GZ'u - GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} - \hat{\Delta}_{\tilde{w}u}^{+} \\ &\Rightarrow \int_0^1 J(r)dB_u(r) + \Delta_{vu}\begin{pmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}(r)dr \end{pmatrix} \\ &- \int_0^1 J(r)dB_v(r)\Omega_{vv}^{-1}\Omega_{vu} - \Delta_{vv}\Omega_{vv}^{-1}\Omega_{vu}\begin{pmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}(r)dr \end{pmatrix} - \Delta_{vu}^{+}\begin{pmatrix} 0_{q\times 1} \\ \int_0^1 \dot{\mathbf{B}}(r)dr \end{pmatrix} \\ &= \int_0^1 J(r)dB_{u\cdot v}(r), \end{aligned} \tag{C.36}$$

from which the result follows by rearranging terms and using the definition of $B_{u\cdot v}(r)$. ■

Proof of Corollary 2. That the limiting distributions of the test statistics given in (43) and (44) coincide follows directly from the asymptotic equivalence of the estimators in turn implying the same limit partial sum processes for both residual processes. It therefore only remains to show that $\hat{\omega}_{u\cdot w}$ is also a consistent estimator of $\omega_{u\cdot v}$, which follows directly from Theorem 1 and Corollary 1:

$$\begin{aligned}\hat{\omega}_{u\cdot w} &= \hat{\Omega}_{uu} - \hat{\Omega}_{uw}\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \\ &= \hat{\Omega}_{uu} - \hat{\Omega}_{u\tilde{w}}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \\ &\Rightarrow \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}e_1^{p\prime}\Pi_v\Pi_v^{-1}\Pi_v e_1^p \\ &= \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu} = \omega_{u\cdot v}.\end{aligned} \tag{C.37}$$

■

APPENDIX D: The Multiple Integrated Regressor Case

We now briefly sketch how the proofs and results of Theorems 1 and 2 and Corollary 1 have to be modified when considering a multiple integrated regressor CPR. All assumptions are exactly as in the main text, with Assumption 2 in its multivariate version commented upon in the main text (with cointegration in the now m-dimensional x_t excluded). We also use the same notation as in the main text with most (implicit) changes immediate and the non-trivial changes explained.

To be precise, the considered setting is given by:

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

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 $p^* := \sum_{j=1}^m p_j$.

The above equation is similar to Wagner and Hong (2016, eq. (1), p. 1292), with the only difference being a different ordering of the regressors. Wagner and Hong (2016) order the variables in groups that include all powers of the different integrated regressors, whereas here we consider all first powers separately in x_t. This is to collect the components of, e. g., $\hat{\Delta}_{\tilde{w}\tilde{w}}$ with standard limits in in the upper left blocks (with therefore a similar structure as in the single integrated regressor case considered in the main text).

As discussed at the end of Section 2.2, we need all elements of x_t included in the CPR to have asymptotic equivalence of FM-CPR and FM-STD. However, clearly not all consecutive powers need to be included, compare again Wagner and Hong (2016).

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

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

with $G := \text{diag}(G_D, T^{-1}I_m, I_m \otimes \text{diag}(T^{-3/2}, \ldots, T^{-\frac{p+1}{2}}))$, $J(r) := [D(r)', B_v(r)', \mathbf{B}_v^*(r)']'$, with $B_v(r) := [B_{v_1}(r), \ldots, B_{v_m}(r)]'$, $\mathbf{B}_v^*(r) := [B_{v_1}^2(r), \ldots, B_{v_1}^{p_1}(r), B_{v_2}^2(r), \ldots, B_{v_m}^{p_m}(r)]'$ and $B_{u\cdot v}(r) := B_u(r) - B_v(r)'\Omega_{vv}^{-1}\Omega_{vu}$, where $B_v(r)$ is now m-dimensional.

In the considered setting the multiple integrated regressor version of $w_t := \Delta X_t$ is given by

$$w_t := \left[v_{1t}, \ldots, v_{mt}, \Delta x_{1t}^2, \ldots, \Delta x_{1t}^{p_1}, \ldots, \Delta x_{mt}^2, \ldots, \Delta x_{mt}^{p_m}\right]' \tag{D.3}$$

and the corresponding scaling matrix G_W to arrive at $\tilde{w}_t := G_W w_t$ is now given by:

$$G_W := \text{diag}\left(I_m, \text{diag}\left(T^{-1/2}, \ldots, T^{-(p_1-1)/2}\right), \ldots, \text{diag}\left(T^{-1/2}, \ldots, T^{-(p_m-1)/2}\right)\right). \tag{D.4}$$

The results of Theorem 1 and Corollary 1 can be generalized to the multiple integrated regressor case using similar arguments as detailed in the earlier proofs. The main difference is, of course, the occurrence of cross-products of first differences of powers of different integrated regressors occur. This entails, e. g., to modify the results given in Appendix B to deal with cross-products. Essentially, this mostly complicates notation and book-keeping but does not lead to additional mathematical complexities. More precisely, for $\hat{\eta}_t := [\hat{u}_t, \tilde{w}_t']'$ it can be shown that:

$$\hat{\Delta}_{\eta\eta} \Rightarrow \begin{bmatrix} \Delta_{\xi\xi} & \Delta_{\xi v} * \int_0^1 \dot{\mathbf{B}}_v^*(r)'dr \\ \Delta_{v\xi} * \int_0^1 \dot{\mathbf{B}}_v^*(r)dr & \Delta_{vv} * \int_0^1 \dot{\mathbf{B}}_v^*(r)\dot{\mathbf{B}}_v^*(r)'dr \end{bmatrix} =: \begin{bmatrix} \Delta_{\xi\xi} & \Delta_{\xi v}^{\mathbf{B}} \\ \Delta_{v\xi}^{\mathbf{B}} & \Delta_{vv}^{\mathbf{B}} \end{bmatrix}, \tag{D.5}$$

where $\dot{\mathbf{B}}_v^*(r) := [2B_{v_1}(r), \ldots, p_1 B_{v_1}^{p_1-1}(r), 2B_{v_2}(r), \ldots, p_m B_{v_m}^{p_m-1}(r)]'$ and the Khatri-Rao product used in a partitioning according "powers of regressors" blocks. As in the single integrated regressor case it holds that $\hat{\Sigma}_{\eta\eta} := \frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t' \Rightarrow \Sigma_{\eta\eta}$, with $\Sigma_{\eta\eta}$ of similar structure as $\Delta_{\eta\eta}$ given just above in (D.5). Combining the two results then leads again to $\hat{\Omega}_{\eta\eta} \Rightarrow \Omega_{\eta\eta}$.

Based upon these results, a crucial step is to show that a "partial inversion" result of the form (C.32) holds again, with the first column now being given by a block-column composed of m rows. Specifically it holds that

$$\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \xrightarrow{\mathbb{P}} \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p*-m)\times 1} \end{bmatrix}, \tag{D.6}$$

which follows from showing that

$$\Omega_{\tilde{w}u} = \Omega_{\tilde{w}\tilde{w}} \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p*-m)\times 1} \end{bmatrix}. \tag{D.7}$$

By definition it holds that

$$\Omega_{\tilde{w}u} = \begin{bmatrix} \Omega_{uv} & \Omega_{uv} * \int_0^1 \dot{\mathbf{B}}_v^*(r)'dr \end{bmatrix}', \tag{D.8}$$

and the result follows from (D.7):

$$\begin{bmatrix} \Omega_{vv} & \Omega_{vv} * \int_0^1 \dot{\mathbf{B}}_v^*(r)'dr \\ \Omega_{vv} * \int_0^1 \dot{\mathbf{B}}_v^*(r)dr & \Omega_{vv} * \int_0^1 \dot{\mathbf{B}}_v^*(r)\dot{\mathbf{B}}_v^*(r)'dr \end{bmatrix} \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{(p^*-m)\times 1} \end{bmatrix} = \begin{bmatrix} \Omega_{vu} \\ \Omega_{vu} * \int_0^1 \dot{\mathbf{B}}_v^*(r)dr \end{bmatrix}. \tag{D.9}$$

This shows (D.6) and leads together with a well-defined limit of $GZ'\tilde{W}$ to the multiple integrated regressor version of (C.33). The result for the limit of the first block-column in $GZ'\tilde{W}$ is already contained in the proof of Wagner and Hong (2016, Proposition 1) for the multiple integrated regressor case (without the reordering considered here). It thus has to be shown, extending the result of Proposition 1, that the other block-columns have well-defined limits as well; the details are available upon request. Also, to arrive at the multivariate version of the results using the quadratic covariation relies upon the same tools as used in the proof of Theorem 1.

To arrive at the multiple integrated regressor version of (C.36) – to show asymptotic equivalence of FM-STD and FM-CPR – the limit of GA^{**} remains to be analyzed, which extends (C.35). Here we get, using similar arguments as just above, that:

$$\begin{aligned}\Delta^+_{\tilde{w}u} &= \Delta_{\tilde{w}u} - \Delta_{\tilde{w}\tilde{w}}\Omega^{-1}_{\tilde{w}\tilde{w}}\Omega_{\tilde{w}u} \\ &= \Delta_{\tilde{w}u} - \Delta_{\tilde{w}v}\Omega^{-1}_{vv}\Omega_{vu} \\ &= \begin{bmatrix} \Delta_{vu} \\ \Delta_{vu} * \int_0^1 \dot{\mathbf{B}}^*_v(r)dr \end{bmatrix} - \begin{bmatrix} \Delta_{vv} \\ \Delta_{vv} * \int_0^1 \dot{\mathbf{B}}^*_v(r)dr \end{bmatrix} \Omega^{-1}_{vv}\Omega_{vu},\end{aligned} \tag{D.10}$$

which corresponds up to the reordering with the term $\Delta^+_{vu}M$ given below Equation (A.1) in Wagner and Hong (2016, p. 1312).

As in the main text, with estimator equivalence established, the subsequent results concerning the parameter hypothesis and cointegration tests all follow.