SFB 823

"Linear" fully modified OLS estimation of cointegrating polynomial regressions

Oliver Stypka, Peter Grabarczyk, Rafael Kawka, Martin Wagner

Discussion Paper

Nr. 77/2016

SFB 823

"Linear" Fully Modified OLS Estimation of Cointegrating Polynomial Regressions

Oliver Stypka
Faculty of Statistics
Technical University Dortmund
Dortmund, Germany

Peter Grabarczyk
Faculty of Statistics
Technical University Dortmund
Dortmund, Germany

Rafael Kawka
Faculty of Statistics
Technical University Dortmund
Dortmund, Germany

Martin Wagner
Faculty of Statistics
Technical University Dortmund
Dortmund, Germany
&
Institute for Advanced Studies
Vienna, Austria
&
Bank of Slovenia
Ljubljana, Slovenia

Abstract

A large part of the empirical environmental Kuznets curve literature uses cointegrating regressions involving a unit root process and its powers as regressors. In this literature the unit root process and its powers are, incorrectly, all treated as integrated processes and modified least squares estimation methods for linear cointegrating regressions are routinely employed. We show that this approach to estimation leads for the Fully Modified OLS estimator surprisingly to the same limiting distribution as obtained for the version of the Fully Modified OLS estimator adapted to the cointegrating polynomial regression setting of Wagner and Hong (2016).

JEL Classification: C13, C32, Q20

Keywords: Cointegrating Polynomial Regression, Environmental Kuznets Curve, Cointegration Test, Fully Modified OLS Estimation, Integrated Process, Nonlinearity

1 Introduction

The scatter plot shown in Figure 1 displays the relationship between log GDP per capita and log CO_2 emissions per capita for Belgium over the period 1870–2009. In addition to the scatter plot, two estimates discussed in detail below are displayed.

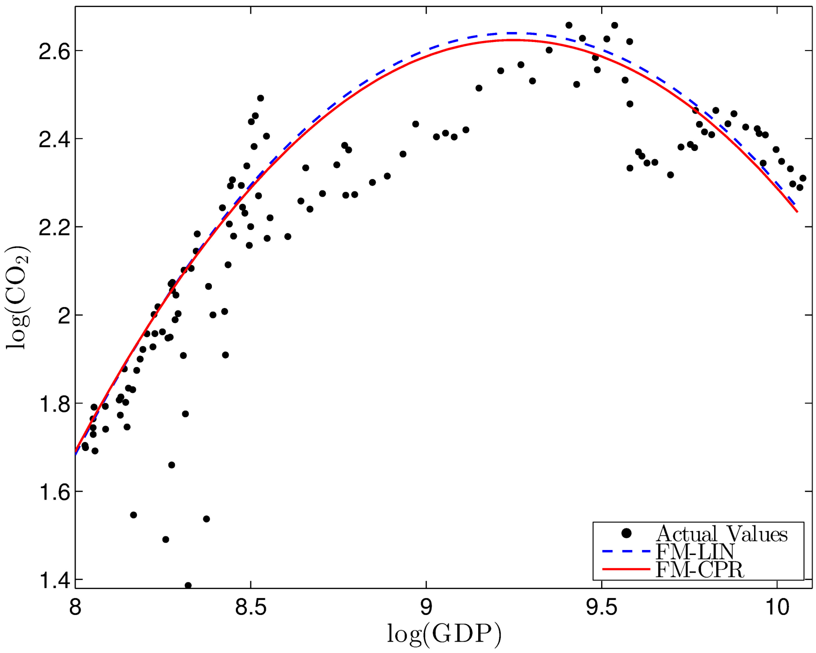

Figure 1: EKC estimation results for the period 1870–2009 for Belgium for CO_2 emissions: Scatter plot and EKC.

[Notes]: The dots show the pairs of observations of log(GDP) and log(CO_2) in per capita terms. The curves show a line that is the result of inserting 140 equidistantly spaced points from the sample range of log(GDP) per capita, with corresponding values of the trend given by $t = 1, \ldots, 140$, in the estimated relationship $y_t = c + \delta t + \beta_1 x_t + \beta_2 x_t^2 + u_t$ using the coefficient estimates obtained by FM-LIN (dashed) and FM-CPR (solid).

An inverted U-shaped relationship between GDP and emissions is known as environmental Kuznets curve (EKC), a phrase coined by Grossman and Krueger (1995).[1] If log GDP per capita, x_t say, is an integrated process, the results in the figure are derived from a regression involving a unit root process and its square, an intercept and a linear trend as regressors and log CO_2 emissions per

[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. Already early survey papers like Stern (2004) or Yandle *et al.* (2004) find more than 100 refereed publications; with many more written since then. See also the discussions in Wagner (2015) and Wagner and Grabarczyk (2016) for additional references and some background.

capita, y_t say, as dependent variable, i.e., from:[2]

$$y_t = c + \delta t + \beta_1 x_t + \beta_2 x_t^2 + u_t, \tag{1}$$
$$x_t = x_{t-1} + v_t.$$

If the errors in (1) are stationary, this is an example of what Wagner and Hong (2016) call *cointegrating polynomial regression* (CPR).

It is known that the square of an integrated process is not an integrated process (see, e.g., Wagner, 2012). Nevertheless, the empirical EKC literature that uses unit root and cointegration techniques employs cointegration estimation methods for linear cointegrating relationships, with the sole exception of Wagner (2015) who applies the methods of Wagner and Hong (2016). This means that, e.g., the Fully Modified OLS (FM-OLS) estimator of Phillips and Hansen (1990) is applied treating x_t and x_t^2 *incorrectly* as two integrated regressors. This approach is referred to as FM-LIN in this paper (defined in (18) in Section 2.3). The results of performing estimation this way are displayed using the label FM-LIN (dashed) in the figure. As mentioned already, Wagner and Hong (2016) adapt the FM-OLS estimator to the CPR case (defined in (11) in Section 2.2). Applying this estimator yields the results labelled FM-CPR (solid) in the figure. The two results are very similar, despite the fact that the FM-LIN estimator, i.e., the standard FM-OLS estimator, is used in a setting for which it has not been designed.

The main result of this paper shows that this similarity is not a coincidence. The asymptotic distributions of the FM-LIN and the FM-CPR estimators coincide for cointegrating polynomial regressions. This main result is shown by developing some intermediate results that are of independent interest. The discussion in Section 2 is for the CPR case with only one integrated process and powers thereof as regressors. The result, however, extends, with only additional notational complexity, to the more general situation considered in Wagner and Hong (2016).[3] Details for the general case are available upon request, whereas for brevity we focus in this paper on the case of only one integrated process and its powers as regressors. This is also the most relevant case for the applications we are aware of.

An immediate implication of the result concerning the equivalence of the asymptotic distributions

[2] All details including definitions and precise assumptions are given in Section 2. Here we only want to set the stage.

[3] The detailed discussion in Section 2 shows that the asymptotic equivalence result requires stricter assumptions than used in, e.g., Wagner and Hong (2016).

is that also the asymptotic distributions of the Shin (1994)-type cointegration test as discussed in Wagner and Hong (2016) for CPRs coincide for both the FM-LIN and the FM-CPR residuals. The critical values for this test depend upon the specification of the equation (see Wagner, 2013), i.e., upon the deterministic component as well as the number and powers of integrated regressors included. Consequently, testing for cointegration using the FM-LIN residuals in conjunction with the Shin (1994) critical values, is invalid even asymptotically. Thus, in contrast to estimation for cointegration testing, no surprising asymptotic result rescues the "linear approach".

The paper is organized as follows: In Section 2 we present the model and assumptions as well as the theoretical results. Section 3 briefly summarizes and concludes. All proofs are relegated to the appendix, including some auxiliary lemmata in the first subsection of the appendix.

We use the following notation: Definitional equality is signified by $:=$ and $\overset{d}{=}$ to denote equality in distribution. Weak convergence is denoted by $\Rightarrow$, $\overset{\mathbb{P}}{\to}$ denotes convergence in probability and $\overset{a.s.}{\to}$ convergence almost surely. $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$ and $\operatorname{diag}(\cdot)$ denotes a diagonal matrix with entries specified throughout. For a vector $x = (x_i)_{i=1,\ldots,n}$ we denote by $\|x\|^2 := \sum_{i=1}^{n} x_i^2$ the Euclidean norm. We denote with $0_{m \times n}$ an $(m \times n)$-matrix with all entries equal to zero. The expected value is denoted by $\mathbb{E}$, L denotes the backward-shift operator, i.e. $L\{z_t\}_{t\in\mathbb{Z}} := \{z_{t-1}\}_{t\in\mathbb{Z}}$, and $\Delta := 1 - L$ denotes the first-difference operator. Brownian motions are denoted as $B(r)$, with covariance matrix specified in the context, and $W(r)$ denotes standard Brownian motion.

2 Theory

2.1 Model and Assumptions

As mentioned in the introduction, to understand the arguments leading to the results it suffices to consider a cointegrating polynomial regression with only one integrated regressor and its powers, i.e.,

$$
\begin{aligned}
y_t &= D_t'\delta + X_t{}'\beta + u_t, \quad \text{for } t = 1, \ldots, T, && (2)\\
x_t &= x_{t-1} + v_t, \quad x_0 = 0,
\end{aligned}
$$

where y_t is a scalar process, $D_t := [1, t, t^2, \ldots, t^q]'$, x_t is a scalar $I(1)$ process and $X_t := [x_t, x_t^2, \ldots, x_t^p]'$.[4] Denoting with $Z_t := [D_t', X_t']'$ the stacked regressor matrix and with $\theta := [\delta', \beta']' \in \mathbb{R}^{(q+1+p)}$ the

[4] Note that, of course, not all consecutive powers of x_t need to be included and in case of more than one integrated regressor the included powers can differ across integrated regressors. These changes lead to notational complications

parameter vector, equation (2) can be rewritten more compactly as:

$$y_t = Z_t'\theta + u_t, \quad \text{for } t = 1, \dots, T. \tag{3}$$

The above example of a polynomial time trend is considered for simplicity only and can be easily relaxed without adding additional complications:[5]

Remark 1 *The results of this paper also hold for more general deterministic components: There exists a sequence of $(q+1) \times (q+1)$ scaling matrices $G_D = G_D(T)$ and a $(q+1)$-dimensional vector of functions $D(z)$, with $0 < \int_0^s D(z)D(z)'ds < \infty$ for $0 < s \leq 1$, such that for $0 \leq s \leq 1$ it holds that:*

$$\lim_{T\to\infty} \sqrt{T} G_D D_{[sT]} = D(s). \tag{4}$$

If, as in (2), $D_t = (1, t, t^2, \dots, t^q)$, then $G_D := \mathrm{diag}(T^{-1/2}, T^{-3/2}, T^{-5/2}, \dots, T^{-(q+1/2)})$ and $D(s) := (1, s, s^2, \dots, s^q)'$.

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

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

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

with $\sum_{j=0}^{\infty} j|c_{uj}| < \infty$, $\sum_{j=0}^{\infty} j|c_{vj}| < \infty$ 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 independent and identically distributed with $\mathbb{E}(\|\xi_t^0\|^l) < \infty$ for some $l > \max(8, 4/(1-2b))$ for some $0 < b < 1/3$.

The above Assumption 1 is stronger than the corresponding assumption used in Wagner and Hong (2016). To be able to draw upon some of the results of Kasparis (2008) we replace the martingale difference sequence assumptions used in Wagner and Hong (2016) with linear process assumptions

only. Clearly, also setting $x_0 = 0$ is only for notational simplicity, the results are unchanged for any well-defined $O_{\mathbb{P}}(1)$ random x_0.

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

and the moment assumption of Kasparis (2008).[6] For univariate $\{x_t\}$ the assumption $C_v(1) \neq 0$ excludes stationary $\{x_t\}$, and has to be modified in the multivariate case to $\det(C_v(1)) \neq 0$, i.e., in the multivariate case the vector process $\{x_t\}$ is assumed to be non-cointegrated.

For long-run covariance estimation we impose the following assumptions with respect to kernel and bandwidth choices, which are closely related to the corresponding assumptions of Jansson (2002):

Assumption 2 *For the kernel function $k(\cdot)$ we assume that:*

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 3 *For the bandwidth parameter M_T we assume that $M_T \subseteq (0, \infty)$ and $M_T = O(T^b)$, with the same $0 < b < 1/3$ as specified in Assumption 1.*

Our Assumption 3 on the bandwidth 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 $0 < b < 1/2$. Clearly, our assumption here is stronger, this tightening of the upper bound stems from the fact that for the asymptotic analysis of the FM-LIN estimator defined in (18) we need to consider "long-run covariance" estimators involving powers of a (scaled) integrated process. Establishing weak convergence of these terms requires smaller bandwidths.

As will be seen in detail below, the FM-LIN estimator involves the usage of kernel estimates of "long-run covariances" and "half long-run covariances" also for nonstationary processes. In order to have uniform notation we *formally* define:

Definition 1 *For two sequences $\{a_t\}$ and $\{b_t\}$ with sample $t = 1, \ldots, T$ we define:*

$$\hat{\Delta}_{ab} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} a_t b_{t+h}', \tag{5}$$

[6] Note that in Kasparis (2008, Assumption 1(b), p. 1376) a condition of the form $l > \min(8, 4/(1-2b))$ is posited. 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 use similar arguments in the proofs of our Lemmata 3 and 4 we require moments of order $\max(8, 4/(1-2b))$.
As discussed in Wagner and Hong (2016) similar results could also be established under alternative assumptions in the spirit of, e.g., Ibragimov and Phillips (2008) or de Jong (2002), augmented correspondingly to accommodate the powers of the integrated regressor. A key difference to, e.g., Chang *et al.* (2001) is that $\{u_t\}_{t\in\mathbb{Z}}$ is allowed to be serially correlated, in an MDS setting in Wagner and Hong (2016) and in a linear process setting here.

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

$$\hat{\Omega}_{ab} := \hat{\Delta}_{ab} + \hat{\Delta}'_{ab} - \hat{\Sigma}_{ab}, \tag{6}$$

with $\hat{\Sigma}_{ab} := \frac{1}{T}\sum_{t=1}^{T} a_t b_t'$.

Clearly, 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 – as usual – a consistent estimator of Δ_{ab}, with similar results holding a fortiori for $\Omega_{ab} := \sum_{h=-\infty}^{\infty} \mathbb{E}(a_0 b_h')$ and $\Sigma_{ab} := \mathbb{E}(a_0 b_0')$.

Remark 2 *Note also that in our definition of $\hat{\Delta}_{ab}$ we use (like, e.g., Phillips, 1995) the bandwidth M_T rather than $T-1$ as upper bound of the summation over the index h (like, e.g., Jansson, 2002). For truncated kernels with $k(x) = 0$ for $|x| > 1$ this is of course inconsequential. It can also be shown (see, e.g., Phillips, 1995) that for standard long-run covariance estimation problems, consistency is not affected by either summation index choice also for untruncated kernels like the Quadratic Spectral kernel.*

In our setting, where the asymptotic behavior of $\hat{\Delta}$-quantities is analyzed for a (properly scaled but) nonstationary process (see Proposition 1 and Corollary 1), the summation bound is important. The key result in Proposition 1 below hinges upon summation only up to M_T. The tighter summation bounds are related to the smaller bandwidths needed postulated in Assumption 3. More specifically, we need this in the proof of Lemma 4. This lemma is related to Kasparis (2008, Lemma A1, p. 1394–1396) where exactly this summation bound is used (in a slightly different context).

Assumption 1 implies that the process $\{\xi_t\}_{t\in\mathbb{Z}} := \{[u_t, v_t]'\}_{t\in\mathbb{Z}}$ fulfills a central limit theorem of the form:

$$T^{-1/2}\sum_{t=1}^{[rT]} \xi_t \Rightarrow B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \Omega_{\xi\xi}^{1/2} W(r), \quad r \in [0,1], \tag{7}$$

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

$$\Omega_{\xi\xi} := \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} = \sum_{h=-\infty}^{\infty} \mathbb{E}(\xi_0 \xi_h') \tag{8}$$

The half (or one-sided) long-run covariance matrix $\Delta_{\xi\xi} := \sum_{h=0}^{\infty} \mathbb{E}(\xi_0 \xi_h')$ is also needed below and partitioned similarly as $\Omega_{\xi\xi}$. For FM-type estimation, estimates of the above long-run covariance matrices are required. Below we focus on the estimation of Δ, from which an estimator of Ω follows using (6) and an estimator of Σ, since the asymptotic behavior of estimators of Δ-type quantities is one of the key elements for the result in Proposition 1.

Unless otherwise stated, in long-run covariance estimation the unobserved errors u_t are replaced by the OLS residuals from (2), $\hat{u}_t$. This defines $\hat{\xi}_t := [\hat{u}_t, v_t]'$ and the effects of this replacement have to be analyzed.

2.2 Fully Modified OLS Estimation

A fully modified OLS (FM-OLS) type estimator for the parameters in (2) is presented in Wagner and Hong (2016) by extending the FM-OLS estimation principle from the linear cointegration case considered in Phillips and Hansen (1990) to the CPR setting.[7]

We briefly describe the two-part transformation required for FM-CPR estimation next: First, the dependent variable y_t is replaced by:

$$y_t^+ := y_t - \Delta x_t \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{v\hat{u}}, \tag{9}$$

with the long-run covariances estimated from $\hat{\xi}_t$. The second transformation consists of adding a bias-correction term that is for specification (2) given by:

$$A^* := \hat{\Delta}_{v\hat{u}}^+ \begin{bmatrix} 0_{(q+1)\times 1} \\ T \\ 2\sum_{t=1}^T x_t \\ \vdots \\ p\sum_{t=1}^T x_t^{p-1} \end{bmatrix}, \tag{10}$$

with $\hat{\Delta}_{v\hat{u}}^+ := \hat{\Delta}_{v\hat{u}} - \hat{\Delta}_{vv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{v\hat{u}}$. Finally, defining $y^+ := [y_1^+, \ldots, y_T^+]'$ and $Z := [Z_1, \ldots, Z_T]'$, the FM-CPR estimator of θ is defined as:

$$\hat{\theta}^+ := (Z'Z)^{-1}(Z'y^+ - A^*). \tag{11}$$

Denoting

$$G = G(T) := \operatorname{diag}(G_D(T), G_X(T)) \tag{12}$$

[7]Note again that related work has also been undertaken by other authors, including – as already mentioned – Chang *et al.* (2001), de Jong (2002), Ibragimov and Phillips (2008) or Liang *et al.* (2015).

with $G_X(T) := \text{diag}(T^{-1}, T^{-3/2}, \ldots, T^{-(p+1)/2})$ and with $J(r) := [D(r)', \mathbf{B}_v(r)']'$ with $\mathbf{B}_v(r) := [B_v(r), B_v(r)^2, \ldots, B_v(r)^p]'$, Wagner and Hong (2016, Proposition 1) show that:[8]

$$G^{-1}(\hat{\theta}^+ - \theta) \Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \int_0^1 J(r)dB_{u\cdot v}(r), \tag{13}$$

with $B_{u\cdot v}(r) := B_u(r) - B_v(r)\Omega_{vv}^{-1}\Omega_{vu}$. The zero-mean Gaussian mixture limiting distribution given in (13) forms the basis for asymptotically valid standard (chi-squared) inference.

2.3 "Linear" Fully Modified OLS Estimation

We now consider the "wrong" approach outlined in the introduction and show that it is asymptotically equivalent to the FM-CPR estimator discussed in the previous subsection, i.e., is in fact asymptotically not "wrong". We refer to this estimator, defined formally in (18), for brevity as FM-LIN estimator.

Considering the CPR "formally" as a standard, linear cointegrating regression problem we rewrite the model as follows:

$$\begin{aligned} y_t &= D_t'\delta + X_t'\beta + u_t \\ X_t &= X_{t-1} + w_t, \end{aligned} \tag{14}$$

with

$$w_t := \begin{bmatrix} \Delta x_t \\ \Delta x_t^2 \\ \vdots \\ \Delta x_t^p \end{bmatrix} = \begin{bmatrix} v_t \\ 2x_{t-1}v_t + v_t^2 \\ \vdots \\ \sum_{k=1}^p \binom{p}{k} x_{t-1}^{p-k} v_t^k \end{bmatrix}, \tag{15}$$

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

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

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

$$A^{**} := \begin{bmatrix} 0_{(q+1)\times 1} \\ T(\hat{\Delta}_{w\hat{u}} - \hat{\Delta}_{ww}\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{w\hat{u}}) \end{bmatrix} = \begin{bmatrix} 0_{(q+1)\times 1} \\ T\hat{\Delta}_{w\hat{u}}^+ \end{bmatrix} \tag{17}$$

[8] The result in Wagner and Hong (2016, Proposition 1) holds with slightly weaker assumptions than Assumptions 1 to 3 used in this paper.

with $\hat{\Delta}_w$ and $\hat{\Delta}_{w\hat{u}}$ also to be interpreted in the sense of Definition 1. This allows to define the FM-LIN estimator as:

$$\hat{\theta}^{++} := (Z'Z)^{-1}(Z'y^{++} - A^{**}), \tag{18}$$

with $y^{++} := [y_1^{++}, \ldots, y_T^{++}]'$. Denoting with $\hat{u}^{++} := [\hat{u}_1^{++}, \ldots, \hat{u}_T^{++}]'$ where $\hat{u}_t^{++} := u_t - w_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{w\hat{u}}$, the centered and scaled estimator can be written as:

$$G^{-1}(\hat{\theta}^{++} - \theta) = \left(GZ'ZG\right)^{-1}\left(GZ'u^{++} - GA^{**}\right), \tag{19}$$

with the first term, obviously, unchanged compared to the FM-CPR estimator. Thus, consider the two parts of the second expression in (19) in more detail using $W := [w_1', \ldots, w_T']'$ and $G_W := G_W(T) = \text{diag}(1, T^{-1/2}, \ldots, T^{-(p-1)/2})$:

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

with $\tilde{W} := WG_W$ a "properly scaled" version of W such that the three terms $GZ'\tilde{W}$, $\hat{\Omega}_{\tilde{w}\tilde{w}}$ and $\hat{\Omega}_{\tilde{w}\hat{u}}$, have well-defined limits established below. Next consider:

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

Combining the above we can rewrite the centered and scaled FM-LIN estimator as:

$$G^{-1}(\hat{\theta}^{++} - \theta) = \left(GZ'ZG\right)^{-1}\left(GZ'u - GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} - \hat{\Delta}_{\tilde{w}u}^{+}\right), \tag{22}$$

Clearly, the asymptotic behavior of the "formal" long-run and half long-run covariance estimators is of key importance and is thus investigated next in two steps. We first consider the process $\{\eta_t\} := \{[u_t, \tilde{w}_t']'\}$ and then show in the second step that the same asymptotic behavior prevails also for $\{\tilde{\eta}_t\} := \{[\hat{u}_t, \tilde{w}_t']'\}$, when using the OLS residuals $\hat{u}_t$ for actual calculations.

Proposition 1 *Under Assumptions 1 to 3 it holds that*

$$\hat{\Delta}_{\eta\eta} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\eta_t\eta_{t+h}' \Rightarrow \Delta_{\eta\eta} := \begin{bmatrix} \Delta_{uu} & \Delta_{uv} & \Delta_{uv}\mathcal{B}' \\ \Delta_{vu} & \Delta_{vv} & \Delta_{vv}\mathcal{B}' \\ \Delta_{vu}\mathcal{B} & \Delta_{vv}\mathcal{B} & \Delta_{vv}\widetilde{\mathcal{B}} \end{bmatrix}, \tag{23}$$

with

$$\mathcal{B} := [2\int_0^1 B_v(r)dr, \ldots, p\int_0^1 B_v^{p-1}(r)dr]' \tag{24}$$

and for $i, j = 1, \ldots, p-1$

$$\widetilde{\mathcal{B}}_{(i,j)} := (1+i)(1+j)\int_0^1 B_v^{i+j}(r)dr. \tag{25}$$

Furthermore, it holds that

$$\hat{\Sigma}_{\eta\eta} := \frac{1}{T}\sum_{t=1}^{T} \eta_t\eta_t' \Rightarrow \Sigma_{\eta\eta} := \begin{bmatrix} \Sigma_{uu} & \Sigma_{uv} & \Sigma_{uv}\mathcal{B}' \\ \Sigma_{vu} & \Sigma_{vv} & \Sigma_{vv}\mathcal{B}' \\ \Sigma_{vu}\mathcal{B} & \Sigma_{vv}\mathcal{B} & \Sigma_{vv}\widetilde{\mathcal{B}} \end{bmatrix}. \tag{26}$$

The above two results lead to:

$$\hat{\Omega}_{\eta\eta} := \hat{\Delta}_{\eta\eta} + \hat{\Delta}_{\eta\eta}' - \hat{\Sigma}_{\eta\eta} \Rightarrow \Delta_{\eta\eta} + \Delta_{\eta\eta}' - \Sigma_{\eta\eta} =: \Omega_{\eta\eta}. \tag{27}$$

Remark 3 *By construction the upper* 2×2*-blocks in the above results coincides with the long-run and half long-run covariances of the process* $\{\xi_t\}_{t\in\mathbb{Z}}$*. For all other terms involving an integrated process or some powers of an integrated process we observe weak convergence to functionals of Brownian motions. This is not unexpected, since these terms are the limits of continuous functions (continuous kernel weighted sums) of scaled powers of integrated processes. In particular these terms are not long-run covariances of some underlying stationary processes but we continue to use the "symbolic notation"* $\Delta_{\eta\eta}$*,* $\Sigma_{\eta\eta}$ *and* $\Omega_{\eta\eta}$*. Note again, only the upper left* 2×2 *blocks are (long-run) covariances.*

As indicated above, replacing u_t by the OLS residuals $\hat{u}_t$ does not change the asymptotic behavior:

Corollary 1 *Under Assumptions 1 to 3 the same results as above also hold for* $\{\tilde{\eta}_t\}$*, i.e.:*

$$\hat{\Delta}_{\tilde{\eta}\tilde{\eta}} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h} \tilde{\eta}_t\tilde{\eta}_{t+h}' \Rightarrow \Delta_{\eta\eta} \tag{28}$$

$$\hat{\Sigma}_{\tilde{\eta}\tilde{\eta}} := \frac{1}{T}\sum_{t=1}^{T} \tilde{\eta}_t\tilde{\eta}_t' \Rightarrow \Sigma_{\eta\eta} \tag{29}$$

$$\hat{\Omega}_{\tilde{\eta}\tilde{\eta}} := \hat{\Delta}_{\tilde{\eta}\tilde{\eta}} + \hat{\Delta}_{\tilde{\eta}\tilde{\eta}}' - \hat{\Sigma}_{\tilde{\eta}\tilde{\eta}} \Rightarrow \Omega_{\eta\eta} \tag{30}$$

It remains to characterize the asymptotic behavior of the remaining component on the right hand side of (22).

Lemma 1 *With the data given by* (2) *under Assumption 1 it holds for*

$$GZ'\tilde{W} = \begin{pmatrix} G_D D'\tilde{W} \\ G_X X'\tilde{W} \end{pmatrix} \tag{31}$$

as $T \to \infty$ *that:*

$$\left(G_D \sum_{t=1}^{T} D_t w_t' G_w \right)_{(i,1)} \Rightarrow \int_0^1 r^i dB_v(r) \tag{32}$$

for $i = 1, \ldots, q+1$ *and*

$$\left(G_D \sum_{t=1}^{T} D_t w_t' G_w \right)_{(i,j)} \Rightarrow j \int_0^1 r^i B_v^{j-1}(r) dB_v(r) + j(j-1)\Delta_{vv} \int_0^1 r^i B_v^{j-2}(r) dr \tag{33}$$
$$- \binom{j}{2} \Sigma_{vv} \int_0^1 r^i B_v^{j-2}(r) dr,$$

for $i = 1, \ldots, q+1;\ j = 2, \ldots, p$ *and*

$$\left(G_X \sum_{t=1}^{T} X_t w_t' G_w \right)_{(i,j)} \Rightarrow j \int_0^1 B_v^{i+j-1}(r) dB_v(r) + j(i+j-1)\Delta_{vv} \int_0^1 B_v^{i+j-2}(r) dr \tag{34}$$
$$- \binom{j}{2} \Sigma_{vv} \int_0^1 B_v^{i+j-2}(r) dr,$$

for $i, j = 1, \ldots, p$.

Combining the results of Proposition 1, Corollary 1 and Lemma 1 allows to establish the main result of this paper given next.

Proposition 2 *Let the data be given by* (2)*, with Assumption 1 in place. Furthermore, let long-run covariance estimation be performed with Assumptions 2 and 3 in place. Then it holds for* $T \to \infty$ *that*

$$G^{-1}(\hat{\theta}^{++} - \theta) \Rightarrow \left(\int_0^1 J(r) J(r)' dr \right)^{-1} \int_0^1 J(r) dB_{u \cdot v}(r). \tag{35}$$

Thus, the FM-LIN and the FM-CPR estimator have the same limiting distribution.

2.4 Testing for Cointegration

The asymptotic equivalence result established in Proposition 2 also implies that the Shin (1994) type test of Wagner and Hong (2016, Proposition 5) for cointegration in the CPR setting can be based on the residuals of both FM-CPR and FM-LIN estimation. Both test statistics have the same asymptotic null distribution as shown in the following corollary.

Corollary 2 *Consider again the cointegrating polynomial regression given in* (2)*, Assumptions 1 to 3 in place and denote as before with* $\hat{u}_t^+$ *the FM-CPR and by* $\hat{u}_t^{++}$ *the FM-LIN residuals. Then it holds that both*

$$CT^+ := \frac{1}{T\hat{\omega}_{\hat{u}\cdot v}} \sum_{t=1}^{T} \left(T^{-1/2} \sum_{j=1}^{t} \hat{u}_j^+ \right)^2 \tag{36}$$

and

$$CT^{++} := \frac{1}{T\hat{\omega}_{\hat{u}\cdot w}} \sum_{t=1}^{T} \left(T^{-1/2} \sum_{j=1}^{t} \hat{u}_j^{++} \right)^2 \tag{37}$$

converge under the null hypothesis as $T \to \infty$ *to*

$$\int_0^1 \left(W_{u\cdot v}^{J_W}(r) \right)^2 dr, \tag{38}$$

with $W_{u\cdot v}^{J_W}(r) := W_{u\cdot v}(r) - \int_0^r J(s)'ds \left(\int_0^1 J^W(s) J^W(s)'ds \right)^{-1} \int_0^1 J^W(s) dW_{u\cdot v}(s)$ *with* $J^W(r) := [D(r)', W_v(r), W_v(r)^2, \ldots, W_v(r)^p]'$, $\hat{\omega}_{\hat{u}\cdot v} := \hat{\Omega}_{\hat{u}\hat{u}} - \hat{\Omega}_{\hat{u}v}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{v\hat{u}}$ *and* $\hat{\Omega}_{\hat{u}\cdot w} := \hat{\Omega}_{\hat{u}\hat{u}} - \hat{\Omega}_{\hat{u}w}\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{w\hat{u}}$. *Under the stated assumptions both* $\hat{\omega}_{\hat{u}\cdot v}$ *and* $\hat{\omega}_{\hat{u}\cdot w}$ *are consistent estimators of* $\omega_{u\cdot v} := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$*, the covariance of* $B_{u\cdot v}(r)$*.*

Remark 4 *Note that in more general CPR models the above test statistic does not necessarily have a nuisance parameter free limiting distribution. The key requirement for this is, using the terminology of Vogelsang and Wagner (2014), full design. In case of only one integrated regressor full design automatically prevails.*

The result of Corollary 2 is in line with the cointegration test findings alluded to in the introduction. Using the FM-LIN residuals to calculate the CT^{++} test statistic, but the Shin (1994) critical values is not mutually consistent. Instead of the Shin (1994) critical values the critical values corresponding to the above limiting distribution need to be used (given in Wagner, 2013). Therefore, using "linear" methods does have an asymptotic effect, not for parameter estimation but for cointegration testing.

3 Summary and Conclusions

We have established asymptotic equivalence of FM-LIN and FM-CPR for cointegrating polynomial regressions (CPRs). It is a surprising feature that the asymptotic distribution of the FM-OLS estimator of Phillips and Hansen (1990) when applied, seemingly unjustified, to CPRs coincides with the asymptotic distribution established for the FM-CPR estimator; an estimator tailor-made for CPRs. This result is in turn driven by some interesting results for long-run covariance estimation, in the sense of Definition 1, collected in Proposition 1. In future research, the asymptotic results will be complemented by finite sample simulation studies to investigate whether the tailor-made FM-CPR estimator of Wagner and Hong (2016) has finite sample performance advantages compared to FM-LIN.

The results of this paper, obviously, raise the question whether such an asymptotic equivalence result between FM-LIN and extensions of the FM-OLS estimator can also be established in more general nonlinear cointegration settings. This intriguing question will be explored in detail in future research.

Acknowledgments

All authors acknowledge financial support from the Collaborative Research Center 823: *Statistical Modelling of Nonlinear Dynamic Processes*. Martin Wagner author furthermore acknowledges research support from the Jubiläumsfonds of the Oesterreichische Nationalbank via several grants. The usual disclaimer applies.

References

Billingsley, P. (2012). Probability and Measure. John Wiley & Sons.

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

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

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.

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.

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.

Liang, H., P.C.B. Phillips, H. Wang, and Q. Wang (2015). Weak Convergence to Stochastic Integrals for Econometric Applications. Forthcoming in *Econometric Theory.*

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. (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.

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

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

Stern, D.I. (2004). The Rise and Fall of the Environmental Kuznets Curve. *World Development* **32**, 1419–1439.

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 P. Grabarczyk (2016). The Environmental Kuznets Curve for Carbon Dioxide Emissions: A Seemingly Unrelated Cointegrating Polynomial Regressions Approach. *Mimeo.*

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

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: Auxiliary Lemmata and Proofs

Auxiliary Lemmata

Lemma 2 [Kasparis (2008), Lemma A1(i)]

Let the data be generated by (2) *with Assumption 1 in place. Then it holds for* $0 \leq b < 1/3$

$$\sup_{r\in[0,1]} T^{-1/2} \sum_{h=0}^{T^b} |v_{\lfloor rT \rfloor + h}| = o_{a.s.}(1). \tag{39}$$

Lemma 3 *Let the data be generated by* (2) *and let Assumptions 1 to 3 be fulfilled. Then it holds for all integers* $0 \leq p$ *and* $1 \leq q$ *that:*

$$\left| \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p \left[\left(\frac{x_{t+h}}{T^{1/2}}\right)^q - \left(\frac{x_t}{T^{1/2}}\right)^q \right] v_t v_{t+h} \right| = o_{\mathbb{P}}(1). \tag{40}$$

Proof of Lemma 3:

Consider $f(x) := x^q$, $x \in \mathbb{R}$. The function f is differentiable on the whole domain. From the mean value theorem it follows 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 $\zeta \in (x, y)$. Therefore, it holds

$$\left(\frac{x_{t+h}}{T^{1/2}}\right)^q - \left(\frac{x_t}{T^{1/2}}\right)^q = q \left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \frac{x_{t+h} - x_t}{T^{1/2}} = \frac{q}{T^{1/2}} \left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \sum_{\nu=1}^{h} v_{t+\nu}, \tag{41}$$

with $\overline{x}_t^h = x_t + \gamma_t \sum_{\nu=1}^h v_{t+\nu}$, $\gamma_t \in (0, 1)$. Using this representation it follows that:

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p \left[\left(\frac{x_{t+h}}{T^{1/2}}\right)^q - \left(\frac{x_t}{T^{1/2}}\right)^q \right] v_t v_{t+h} \tag{42}$$

$$= \frac{q}{T^{1/2}} \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p \left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \sum_{\nu=1}^{h} v_t v_{t+\nu} v_{t+h}. \tag{43}$$

The assertion is hence equivalent to showing that:

$$\frac{1}{T^{1/2}} \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p \left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \sum_{\nu=1}^{h} v_t v_{t+\nu} v_{t+h} = o_{\mathbb{P}}(1). \tag{44}$$

In the course of the proof it is helpful to resort to strong approximations, which we get from the Skorohod representation theorem, see Pollard (1984, p. 71–72) or Csörgo and Horváth (1993, p.4).[9] Since we are concerned with weak convergence results in this paper, we can w.l.o.g. use a

[9] 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).

distributionally equivalent version of $T^{-1/2}x_{\lfloor rT\rfloor}$, X_T^* 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 (7). Setting $\tilde{C} := \sup_{r\in[0,1]}|B_v(r)| + 1/2$, it holds for sufficiently large T that

$$\sup_{r\in[0,1]} T^{-1/2}|x_{\lfloor rT\rfloor}| \leq \tilde{C} \quad \text{a.s.} \tag{45}$$

Furthermore, it holds that

$$\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT\rfloor+h} - x_{\lfloor rT\rfloor}| \tag{46}$$

$$= \sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}\Big|\sum_{\nu=1}^{h} v_{\lfloor rT\rfloor+\nu}\Big| \leq \sup_{r\in[0,1]} T^{-1/2}\sum_{\nu=1}^{M_T}|v_{\lfloor rT\rfloor+\nu}| \tag{47}$$

and thus it follows from Lemma 2 that

$$\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT\rfloor+h} - x_{\lfloor rT\rfloor}| = o_{a.s.}(1). \tag{48}$$

This implies

$$\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT\rfloor+h}| \tag{49}$$

$$\leq \sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|x_{\lfloor rT\rfloor+h} - x_{\lfloor rT\rfloor}| + \sup_{r\in[0,1]} T^{-1/2}|x_{\lfloor rT\rfloor}| \leq C \quad \text{a.s.} \tag{50}$$

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

$$\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|\overline{x}^h_{\lfloor rT\rfloor}| \leq C \quad \text{a.s.} \tag{51}$$

Using the triangular inequality and the bounds given in (45)–(51) the following inequalities hold:

$$\left|\frac{1}{T^{1/2}}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\left(\frac{\overline{x}^h_t}{T^{1/2}}\right)^{q-1}\sum_{\nu=1}^{h} v_t v_{t+\nu} v_{t+h}\right| \tag{52}$$

$$\leq \left(\frac{M_T^3}{T}\right)^{1/2}\frac{1}{M_T}\sum_{h=0}^{M_T}\left|k\left(\frac{h}{M_T}\right)\right|\frac{1}{T}\sum_{t=1}^{T-h}\left|\left(\frac{x_t}{T^{1/2}}\right)^p\left(\frac{\overline{x}^h_t}{T^{1/2}}\right)^{q-1}\right||v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{\nu=1}^{h} v_{t+\nu}\right| \tag{53}$$

$$\leq \left(\frac{M_T^3}{T}\right)^{1/2}\overline{k}(0)C^{p+q-1}\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}|\left|\frac{1}{M_T^{1/2}}\sum_{\nu=1}^{h} v_{t+\nu}\right| \quad \text{a.s.,} \tag{54}$$

with $\overline{k}(0) = \sup_{x\geq 0}|k(x)|$ as defined in Assumption 2. Furthermore, observe that:

$$\sup_{s\in[0,1]}\sup_{t=1,\ldots,T}\left|\frac{1}{M_T^{1/2}}\sum_{\nu=1}^{\lfloor sM_T\rfloor} v_{t+\nu}\right| \leq C \quad \text{a.s.,} \tag{55}$$

due to strict stationarity of $\{v_t\}$. Consequently,

$$\left|\frac{1}{T^{1/2}}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1}\sum_{\nu=1}^{h} v_t v_{t+\nu} v_{t+h}\right| \tag{56}$$

$$\leq \left(\frac{M_T^3}{T}\right)^{1/2}\overline{k}(0)C^{p+q}\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}| \quad \text{a.s.} \tag{57}$$

Assumption 1 implies that:

$$\mathbb{E}\left(\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}|\right) \leq \frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}\left(\mathbb{E}[v_t^2]\mathbb{E}[v_{t+h}^2]\right)^{1/2} \leq \Sigma_{vv} < \infty. \tag{58}$$

From the Markov inequality, see e. g., Billingsley (2012, p.294), it follows that:

$$\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_t v_{t+h}| = O_{\mathbb{P}}(1). \tag{59}$$

Finally, the assertion is an immediate consequence of $M_T^3/T \to 0$ by Assumption 3, and the remaining terms in (57) being $O_{\mathbb{P}}(1)$.

Lemma 4 *Let the data be generated by* (2) *with Assumptions 1 to 3 in place. Then it holds for all integers* $0 \leq p$ *that:*

$$\left|\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right)\right| = o_{\mathbb{P}}(1). \tag{60}$$

Proof of Lemma 4:

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

$$\left|\frac{1}{M_T}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\sum_{\nu=1}^{h}\left(v_t v_{t+\nu} - \mathbb{E}[v_t v_{t+\nu}]\right)\right| = o_{\mathbb{P}}(1) \tag{61}$$

by showing

$$\sup_{0\leq h\leq M_T}\left|\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\sum_{\nu=1}^{h}\left(v_t v_{t+\nu} - \mathbb{E}[v_t v_{t+\nu}]\right)\right| = o_{\mathbb{P}}(1). \tag{62}$$

The left-hand side of (60) can be written as

$$\left|\frac{1}{M_T}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p M_T\left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right)\right| = o_{\mathbb{P}}(1). \tag{63}$$

Using a similar argument as used by Kasparis (2008, p. 1394–1396) to show (62), corresponding to his Equation A.7, it can be shown that

$$\left|\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p M_T\left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right)\right| = o_{\mathbb{P}}(1), \tag{64}$$

which shows the claim of this lemma, since

$$\left|\frac{1}{M_T}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p M_T\left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right)\right| \tag{65}$$

$$\leq \overline{k}(0)\left|\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p M_T\left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right)\right|. \tag{66}$$

It is the fact that our proof of this lemma uses some of the arguments of Kasparis (2008) that the same moment and bandwidth assumptions as in this paper are required. These are consequently contained in our Assumptions 1 to 3.

Proofs of the Results from the Main Text

Proof of Proposition 1:

First, the $(1,1)$-element of $\hat{\Delta}_{\eta\eta}$ is given by

$$\left(\hat{\Delta}_{\eta\eta}\right)_{(1,1)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h} u_t u_{t+h}, \tag{67}$$

which is already well known, cf. Remark 3. For $i \in \{1,\ldots,p\}$ it holds

$$\left(\hat{\Delta}_{\eta\eta}\right)_{(i+1,1)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\frac{\Delta x_t^i}{T^{\frac{i-1}{2}}}u_{t+h} \tag{68}$$

$$\left(\hat{\Delta}_{\eta\eta}\right)_{(i+1,2)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\frac{\Delta x_t^i}{T^{\frac{i-1}{2}}}v_{t+h}, \tag{69}$$

i. e., for the first and second columns (and rows) exactly the same arguments apply due to the similar assumptions on $\{u_t\}$ and $\{v_t\}$. Therefore, it is sufficient in the subsequent discussion to consider the $(i+1, j+1)$-element for $i, j \in \{1,\ldots,p\}$ of the estimator $\hat{\Delta}_{\eta\eta}$, which is given by

$$\left(\hat{\Delta}_{\eta\eta}\right)_{(i+1,j+1)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\frac{\Delta x_t^i}{T^{\frac{i-1}{2}}}\frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}}. \tag{70}$$

Note that

$$\frac{\Delta x_t^i}{T^{(i-1)/2}} = \frac{1}{T^{(i-1)/2}} \sum_{k=1}^{i} \binom{i}{k} x_t^{i-k}(-v_t)^k \tag{71}$$

$$= i\left(\frac{x_t}{T^{1/2}}\right)^{i-1} v_t - \sum_{k=2}^{i} \binom{i}{k} (-1)^k \left(\frac{x_t}{T^{1/2}}\right)^{i-k} \left(\frac{v_t}{T^{1/2}}\right)^{k-2} \frac{v_t^2}{T^{1/2}}. \tag{72}$$

From Lemma 2 we know that $T^{-1/2}v_t = o_{a.s.}(1)$ for all $t = 1,\dots,T$. Additionally, it holds that $T^{-1/2}|x_t| \leq C$ a.s. for $t = 1,\dots,T$. 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

$$\frac{\Delta x_t^i}{T^{(i-1)/2}} = i\left(\frac{x_t}{T^{1/2}}\right)^{i-1} v_t + O_{\mathbb{P}}(T^{-1/2}). \tag{73}$$

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

$$\left(\hat{\Delta}_{\eta\eta}\right)_{(i+1,j+1)} = ij \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^{i-1} \left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_t v_{t+h} + o_{\mathbb{P}}(1). \tag{74}$$

In the linear case, i. e. $i = j = 1$, the above term converges in probability to Δ_{vv}, cf. Remark 3 again. Next, consider $i > 1$ and $j = 1$, i. e.,

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^{(i-1)} v_t v_{t+h}. \tag{75}$$

From Lemma 4 it follows that

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^{(i-1)} v_t v_{t+h} \tag{76}$$

$$= \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^{(i-1)} \mathbb{E}[v_t v_{t+h}] + o_{\mathbb{P}}(1). \tag{77}$$

Now, we show that

$$\left| \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^{i-1} - \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \frac{1}{T} \sum_{t=1}^{T} \left(\frac{x_t}{T^{1/2}}\right)^{i-1} \right| = o_{a.s.}(1). \tag{78}$$

The left-hand side of (78) corresponds to

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \mathbb{E}[v_0 v_h] \frac{1}{T} \sum_{t=T-h+1}^{T} \left(\frac{x_t}{T^{1/2}}\right)^{i-1} \tag{79}$$

and by Assumption 1

$$\left|\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\frac{1}{T}\sum_{t=T-h+1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i-1}\right| \tag{80}$$

$$\leq C^{i-1}\frac{1}{T}\sum_{h=0}^{M_T}\left|k\left(\frac{h}{M_T}\right)\right||\mathbb{E}[v_0 v_h]|h \quad \text{a.s.} \tag{81}$$

$$\leq \overline{k}(0)|\Sigma_{\varepsilon\varepsilon}|C^{i-1}\frac{1}{T}\sum_{h=0}^{M_T} h\sum_{j=0}^{\infty}|c_{v,j}c_{v,j+h}| \tag{82}$$

$$\leq \overline{k}(0)|\Sigma_{\varepsilon\varepsilon}|C^{i-1}\frac{1}{T}\sum_{j=0}^{\infty}|c_{v,j}|\sum_{h=0}^{\infty}h|c_{v,h}| \xrightarrow{\mathbb{P}} 0, \tag{83}$$

which implies that

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\ \frac{1}{T}\sum_{t=T-h+1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i-1} = o_{a.s.}(1). \tag{84}$$

Therefore, we obtain

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\frac{\Delta x_t^i}{T^{(i-1)/2}}v_{t+h} \tag{85}$$

$$= i\left(\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i-1}\right) + o_{\mathbb{P}}(1). \tag{86}$$

Thus, two separate terms need to be considered. For the first it holds that

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h] \to \Delta_{vv}. \tag{87}$$

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

$$i\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i-1} \Rightarrow i\Delta_{vv}\int_0^1 B_v^{i-1}(r)dr. \tag{88}$$

We turn to the case $i > 1$ and $j > 1$, i. e.

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^{(i-1)}\left(\frac{x_{t+h}}{T^{1/2}}\right)^{(j-1)} v_t v_{t+h}. \tag{89}$$

Using Lemma 3 we obtain

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^{(i-1)}\left(\frac{x_{t+h}}{T^{1/2}}\right)^{(j-1)} v_t v_{t+h} \tag{90}$$

$$= \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^{(i+j-2)} v_t v_{t+h} + o_{\mathbb{P}}(1). \tag{91}$$

Now we are in the same setting as for $j = 1$, such that we can immediately conclude

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{\Delta x_t^i}{T^{\frac{i-1}{2}}}\frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} \tag{92}$$

$$= ij\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i+j-2} + o_{\mathbb{P}}(1) \tag{93}$$

$$\Rightarrow ij\Delta_{vv}\int_0^1 B_v^{i+j-2}(r)dr. \tag{94}$$

Joint convergence of the elements in $\hat{\Delta}_{\eta\eta}$, follows by the continuous mapping theorem. □

Proof of Corollary 1:

The OLS residuals are given by $\hat{u}_t = u_t - Z_t'(\hat{\theta} - \theta)$. Similar to the proof of Proposition 1 consider for $j \in \{1, \dots, p\}$ the term

$$\left(\hat{\Delta}_{\hat{\eta}\hat{\eta}}\right)_{(1,j+1)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\hat{u}_t\frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} \tag{95}$$

$$= \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}u_t\frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} - \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}Z_t'(\hat{\theta}-\theta)\frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}}. \tag{96}$$

The first term in (96) converges in distribution to $(\Delta_{\eta\eta})_{(1,j+1)}$ by Proposition 1. Therefore it remains to show that the second term is $o_{\mathbb{P}}(1)$. It follows that

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}Z_t'(\hat{\theta}-\theta)\frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} \tag{97}$$

$$= \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}Z_t'GG^{-1}(\hat{\theta}-\theta)\cdot j\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h} + o_{\mathbb{P}}(1) \tag{98}$$

by similar arguments as in the proof of Proposition 1 with G defined in (12). Expression (98) can be further rewritten as

$$\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) j\,\frac{1}{T^{3/2}}\sum_{t=1}^{T-h}\left(T^{1/2}Z_t'G\right)\left(\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h}\right)\left(G^{-1}(\hat{\theta}-\theta)\right) + o_{\mathbb{P}}(1). \tag{99}$$

Finally, show that

$$\left\|\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) j\,\frac{1}{T^{3/2}}\sum_{t=1}^{T-h}\left(T^{1/2}Z_t'G\right)\left(\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h}\right)\right\| = o_{\mathbb{P}}(1). \tag{100}$$

Using the notation from Lemma 3 it holds that

$$\left\| \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) j \, \frac{1}{T^{3/2}} \sum_{t=1}^{T-h} \left(T^{1/2} Z_t' G\right) \left(\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h}\right) \right\| \tag{101}$$

$$\leq \; j\bar{k}(0) \sum_{h=0}^{M_T} \frac{1}{T^{3/2}} \sum_{t=1}^{T-h} \left\| \left(T^{1/2} Z_t' G\right) \left(\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h}\right) \right\| \tag{102}$$

$$\leq \; j\bar{k}(0) C^{j-1} \sum_{h=0}^{M_T} \frac{1}{T^{3/2}} \sum_{t=1}^{T-h} \left\| T^{1/2} Z_t' G \right\| |v_{t+h}|. \tag{103}$$

In addition, observe that

$$\left\| \left(T^{1/2} Z_t' G\right) \right\|^2 = \sum_{k=0}^{q} \left(\frac{t}{T}\right)^k + \sum_{l=1}^{p} \left(\frac{x_t}{T^{1/2}}\right)^l \leq (q+1) + \sum_{l=1}^{p} C^l =: K \quad \text{a.s.} \tag{104}$$

such that

$$\left\| \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) j \, \frac{1}{T^{3/2}} \sum_{t=1}^{T-h} \left(T^{1/2} Z_t' G\right) \left(\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h}\right) \right\| \tag{105}$$

$$\leq \; j\bar{k}(0) C^{j-1} K^{1/2} \frac{1}{T^{1/2}} \sum_{h=0}^{M_T} \frac{1}{T} \sum_{t=1}^{T-h} |v_{t+h}| \quad \text{a.s.} \tag{106}$$

follows. Similar to the discussion of (59) one can show

$$\frac{1}{T^{1/2}} \sum_{h=0}^{M_T} \frac{1}{T} \sum_{t=1}^{T-h} |v_{t+h}| = o_{\mathbb{P}}(1). \tag{107}$$

Hence, the expressions (106) and, consequently, (97) are $o_{\mathbb{P}}(1)$ such that

$$\left(\hat{\Delta}_{\hat{\eta}\hat{\eta}}\right)_{(1,j+1)} = \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} u_t \frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} + o_{\mathbb{P}}(1) \tag{108}$$

and the claim follows. □

Proof of Lemma 1:

We start with considering the first column of $G_X \sum_{t=1}^{T} X_t w_t' G_w$. According to Wagner and Hong (2016, Proposition 1) the limit of this term for $i = 1, \ldots, p$ and $j = 1$ is given by:

$$\left(G_X \sum_{t=1}^{T} X_t w_t' G_w \right)_{(i,1)} = \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(\frac{x_t}{T^{1/2}} \right)^i v_t \Rightarrow \int_0^1 B_v^i(r) dB_v(r) + i\Delta_{vv} \int_0^1 B_v^{i-1}(r) dr. \quad (109)$$

Consider now again $i = 1, \ldots, p$, but $j > 1$:

$$\begin{aligned}
\left(G_X \sum_{t=1}^{T} X_t w_t' G_w \right)_{(i,j)} &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(\frac{x_t}{T^{1/2}} \right)^i \left(-\sum_{k=1}^{j} \binom{j}{k} \frac{x_t^{j-k}(-v_t)^k}{T^{(j-1)/2}} \right) &(110)\\
&= \frac{1}{T^{1/2}} \sum_{t=1}^{T} j \left(\frac{x_t}{T^{1/2}} \right)^{i+j-1} v_t &(111)\\
&\quad - \frac{1}{T^{1/2}} \sum_{t=1}^{T} \binom{j}{2} \left(\frac{x_t}{T^{1/2}} \right)^{i+j-2} \frac{v_t^2}{T^{1/2}} &(112)\\
&\quad - \frac{1}{T^{1/2}} \sum_{t=1}^{T} \sum_{k=3}^{j} \binom{j}{k} \left(\frac{x_t}{T^{1/2}} \right)^{i+j-k} \frac{(-v_t)^k}{T^{(k-1)/2}}. &(113)
\end{aligned}$$

The first term on the right-hand side converges similarly to (109) to

$$j \int_0^1 B_v^{i+j-1}(r) dB_v(r) + j(i+j-1)\Delta_{vv} \int_0^1 B_v^{i+j-2}(r) dr.$$

For the second term (112) we write $v_t^2 = \Sigma_{vv} + (v_t^2 - \Sigma_{vv})$ and consider both terms separately. First,

$$\binom{j}{2} \frac{\Sigma_{vv}}{T} \sum_{t=1}^{T} \left(\frac{x_t}{T^{1/2}} \right)^{i+j-2} \Rightarrow \binom{j}{2} \Sigma_{vv} \int_0^1 B_v^{i+j-2}(r) dr. \quad (114)$$

Second, using Lemma 4 it holds for the remaining term that

$$\binom{j}{2} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{x_t}{T^{1/2}} \right)^{i+j-2} \left(v_t^2 - \Sigma_{vv} \right) = o_{\mathbb{P}}(1). \quad (115)$$

All additional terms in (113) converge to zero being $O_{\mathbb{P}}(T^{-(k-2)/2})$. The result for $G_D \sum_{t=1}^{T} D_t w_t' G_w$ follows analogously. □

Proof of Proposition 2:

Beforehand, note that we can use the decomposition $\Omega_{\tilde{w}\tilde{w}} = \Omega_{vv}\Pi_v$ with

$$\Pi_v := \begin{bmatrix} 1 & \mathcal{B}' \\ \mathcal{B} & \tilde{\mathcal{B}} \end{bmatrix}$$

and $\mathcal{B}$ and $\tilde{\mathcal{B}}$ defined in (24) and (25), respectively. From Proposition 1 we know, that $\hat{\Omega}_{\tilde{w}\tilde{w}} \Rightarrow \Omega_{vv}\Pi_v$ and $\hat{\Omega}_{\tilde{w}u} \Rightarrow \Omega_{vu}\Pi_v[1,0,\ldots,0]'$. Therefore, it follows $\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \xrightarrow{\mathbb{P}} \Omega_{vv}^{-1}\Omega_{vu}[1,0,\ldots,0]'$. In (22) we have noted that

$$G^{-1}(\hat{\theta}^{++} - \theta) = \left(GZ'ZG\right)^{-1}\left(GZ'u - GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} - \hat{\Delta}_{\tilde{w}u}^{+}\right). \tag{116}$$

Using the same arguments as in Wagner and Hong (2016) it holds that:

$$GZ'u \Rightarrow \int_0^1 J(r)dB_u(r) + \Delta_{vu}\begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}, \tag{117}$$

with $M = [1, \mathcal{B}']'$. From Proposition 1 it follows immediately that A^* and $\hat{\Delta}_{\tilde{w}u}^{+}$ have the same limiting distribution, i. e.,

$$A^* \Rightarrow \Delta_{vu}^{+}\begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix} \text{ and } \hat{\Delta}_{\tilde{w}u}^{+} \Rightarrow \Delta_{vu}^{+}\begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}.$$

Lemma 1 provides the limiting distribution of $GZ'\tilde{W}$, of which we only need the first column due to the structure of the limit of $\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u}$. The first term of $GZ'\tilde{W}$ is given by $GZ'v$ and it holds that:

$$GZ'v \Rightarrow \int_0^1 J(r)dB_v(r) + \Delta_{vv}\begin{pmatrix} 0_{(q+1)\times 1} \\ M \end{pmatrix}. \tag{118}$$

Therefore, arrive at:

$$GZ'u - GZ'\tilde{W}\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} - \hat{\Delta}_{\tilde{w}u}^{+} \Rightarrow \int_0^1 J(r)dB_u(r) - \int_0^1 J(r)dB_v(r)\Omega_{vv}^{-1}\Omega_{vu}. \tag{119}$$

Noting that $B_{u\cdot v}(r) := B_u(r) - B_v(r)\Omega_{vv}^{-1}\Omega_{vu}$ completes the proof. □

Proof of Corollary 2:

The result for CT^{+} is given in Wagner and Hong (2016, Proposition 5) and for the CT^{++} test statistic the corresponding proof for the numerator of the test statistic, i.e., for $\frac{1}{T}\sum_{t=1}^{T}\left(T^{-1/2}\sum_{j=1}^{t}\hat{u}_j^{++}\right)$ follows analogously from considering $\hat{u}_t^{++} = u_t^{++} - Z_t'(\hat{\theta}^{++} - \theta)$ with $u_t^{++} = u_t - w_t'\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{w\hat{u}}$. From the proof of Proposition 1 we know that $\frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]} u_t^{++} \Rightarrow B_{u\cdot v}(r)$ for $0 \leq r \leq 1$. The result for the second part immediately follows as in Wagner and Hong (2016) from the asymptotic equivalence of the FM-CPR and FM-LIN estimators established in Proposition 2.

It thus remains to consider the asymptotic behavior $\hat{\omega}_{\hat{u}\cdot w}$, which from the asymptotic behavior of the “long-run” covariance estimators established in Proposition 1:

$$\begin{aligned}\hat{\omega}_{\hat{u}\cdot v} &= \hat{\Omega}_{uu} - \hat{\Omega}_{uw}\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{wu} \\ &\Rightarrow \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}[1,0,\ldots,0]\Pi_v\Pi_v^{-1}\Pi_v[1,0,\ldots,0]' \\ &= \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu} = \omega_{u\cdot v},\end{aligned} \tag{120}$$

with convergence in probability, i.e. consistency, following from the fact that the limit is non-stochastic.