Monitoring Cointegrating Polynomial Regressions: Theory and Application to the Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions

Fabian Knorre

Martin Wagner

Maximilian Grupe

INSTITUTE FOR
ADVANCED STUDIES
VIENNA

Author(s)

Fabian Knorre, Martin Wagner, Maximilian Grupe

Editor(s)

Robert M. Kunst

Title

Monitoring Cointegrating Polynomial Regressions: Theory and Application to the Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions

Monitoring Cointegrating Polynomial Regressions: Theory and Application to the Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions

Fabian Knorre
Department of Statistics
TU Dortmund University
Dortmund, Germany
&
Ruhr Graduate School in Economics
Essen, Germany

Martin Wagner
Department of Economics
University of Klagenfurt
Klagenfurt, Austria
&
Bank of Slovenia
Ljubljana, Slovenia
&
Institute for Advanced Studies
Vienna, Austria

Maximilian Grupe
Department of Statistics
TU Dortmund University
Dortmund, Germany

December 29, 2020

Abstract

This paper develops residual-based monitoring procedures for cointegrating polynomial regressions, i. e. , regression models including deterministic variables, integrated processes as well as integer powers of integrated processes as regressors. The regressors are allowed to be endogenous and the stationary errors are allowed to be serially correlated. We consider five variants of monitoring statistics and develop the results for three modified least squares estimators for the parameters of the CPRs. The simulations show that using the combination of self-normalization and a moving window leads to the best performance. We use the developed monitoring statistics to assess the structural stability of environmental Kuznets curves (EKCs) for both CO_2 and SO_2 emissions for twelve industrialized country since the first oil price shock.

JEL Classification: C22, C52, Q56

Keywords: Cointegrating Polynomial Regression, Environmental Kuznets Curve, Monitoring, Structural Change

1 Introduction

This paper develops residual-based monitoring procedures for structural change in cointegrating polynomial regressions (CPRs), using the terminology of Wagner and Hong (2016). CPRs are regression models that include as explanatory variables deterministic terms, integrated processes and integer powers of integrated processes. The regressors are allowed to be endogenous and the stationary errors are allowed to be serially correlated. Structural change – at an unknown point in time – can occur in two facets: First, the relationship may turn into a spurious relationship.[1] Second, the parameters of the relationship may change. The developed monitoring statistics extend those of Wagner and Wied (2017) in two dimensions. First, a variety of monitoring statistics is considered, including self-normalized versions and moving window detectors.[2] Second, the approach is extended from cointegrating linear to cointegrating polynomial regressions.

All considered monitoring statistics are based on parameter estimation for the CPR relationship over a *calibration period* known to be – or at least assumed to be – free of structural change, an approach to monitoring inspired by Chu *et al.* (1996). With regressors that are potentially endogenous and errors that are potentially serially correlated, appropriately modified least squares estimators have to be employed to allow for the construction of nuisance parameter free limiting distributions of the detectors obtained by scaling out a scalar long-run variance parameter. We consider the CPR-versions of three well-known estimators: Fully Modified OLS (FM-OLS) considered in the CPR context in, e. g. , Wagner and Hong (2016), Dynamic OLS (D-OLS) considered for more general functions in Choi and Saikkonen (2010), and IM-OLS considered in Vogelsang and Wagner (2014b).[3] In the *general* CPR case, however, even the usage of the mentioned modified least squares estimators is not sufficient for nuisance parameter free limiting distributions and the assumption of *full design*, using the terminology of Vogelsang and Wagner (2014b), is required. Full design means that the limiting distributions of the modified estimators can be written as the product of a regular non-random matrix and a functional of standard Brownian motions and

---

[1]In the CPR setting the concept of spurious regression has to be interpreted a bit wider than in cointegrating linear regression settings. If, e. g. , the polynomial degree of a CPR relationship increases at a certain point in time but one continues to consider a CPR relationship with an unchanged polynomial degree, then the error term of this *spurious relationship* contains higher order powers of an integrated process and is thus not integrated, as in the usual form of spuriousness considered in linear cointegration.

[2]Some of these possibilities have been mentioned in Wagner and Wied (2017, Footnote 4), but have not been explored in full detail and systematically.

[3]The settings considered in Choi and Saikkonen (2010) and Vogelsang and Wagner (2014b) are discussed in a bit more detail in Footnote 11.

deterministic components. This allows to construct monitoring statistics, based on the residual limit processes, which are proportional up to a scalar long-run variance to functionals of standard Brownian motions and deterministic components. Scaling out the long-run variance, either by self-normalization or by standardization, then leads to nuisance parameter free limiting distributions of the monitoring statistics. Note that full design, albeit not generally prevalent in the CPR case, holds in a variety of empirically relevant settings, including cointegrating linear regressions, cointegrating polynomial regressions where only one of the integrated regressors occurs as regressor also with higher order powers, and Translog-type relationships (see, e.g. Christensen *et al.*, 1971). Environmental Kuznets curves (EKCs), which are the focus of our paper, typically include only one integrated regressor and its power and are, therefore, of full design.

We perform a detailed simulation study of the five considered variants of the monitoring procedure in combination with the three mentioned parameter estimation methods. The performance dimensions considered include null rejection probabilities, (empirical) size-corrected power as well as detection *delays*. It turns out that the combination of self-normalization and a moving window leads almost throughout to the best performance in terms of lowest over-rejections, whilst exhibiting very favorable size-corrected power properties and short delays. The choice of the estimator, FM-, D-, or IM-OLS, affects the results in particular for small samples. IM-OLS mostly leads to smaller over-rejections under the null hypothesis than FM-OLS, which in turn outperforms D-OLS. With respect to size-corrected power IM-OLS is outperformed, as expected, by both FM-OLS and D-OLS. Also, the delays are often a bit smaller for FM-OLS than for IM-OLS. Since the differences are often relatively small, there is no clear choice between IM-OLS and FM-OLS.

We use the developed monitoring tools to assess the stability of environmental Kuznets curves (EKCs) for CO_2 and SO_2 emissions for twelve countries using a calibration period 1946–1973 and a monitoring period 1974–2016. The EKC hypothesis postulates an inverted U-shaped relationship between the level of economic development and pollution or emissions.[4] Brock and Taylor (2005) or Kijima *et al.* (2010) provide survey discussions of the links between economic activity or growth

---

[4]The term EKC refers by analogy to the inverted U-shaped relationship between the level of economic development and the degree of income inequality postulated by Kuznets (1955) in his 1954 presidential address to the American Economic Association. Since the seminal contributions of, e.g., Grossman and Krueger (1991; 1993; 1995) or Shafik and Bandyopadhyay (1992), the literature – both theoretical as well as empirical – has become voluminous and continues to grow rapidly. Already early survey papers like Yandle *et al.* (2004) count more than 100 refereed publications on the subject.

and the environment.[5] We thus use our monitoring tools to assess whether and if so when the relationship between emissions and economic activity has changed after the first oil price shock, which has led to fundamental changes in economic activity triggered not least by changing energy prices, but also by changes in environmental legislation that has been put in place in the 1970s.[6] Taking into account that the polynomial functional form should probably be interpreted more as an approximation to an underlying relationship of unknown form rather than as a true relationship, of course, implies a trade-off (at least in-sample) between finding structural breaks and approximation with a higher polynomial degree.[7] Against this background it may be interesting to use monitoring tools to study whether and at which point in time, the EKC relationship needs to be modelled with more *curvature*: Over the calibration period 1946–1973 for most countries (as minimum polynomial degree) a cointegrating linear relationship prevails.[8] This is not a too big surprise, since about until the mid 1970s, economic activity expanded roughly in line with emissions in many countries. Only thereafter and triggered – as mentioned – by price, technical and legislative changes the economic activity and pollution start to be decoupled to a certain extent. From a CPR perspective this could mean either a structural change in the parameters of a relationship of given degree or a change to a relationship with a higher polynomial degree, e.g., from a linear to a quadratic relationship. For both CO_2 and SO_2 emissions for nine of the twelve countries structural breaks are detected. The detected break points are in some cases quite late, which most likely reflects the delays inherent in monitoring procedures. The evidence, when considering also the full sample, is mixed with respect to structural change in the parameters but unchanged polynomial degree or structural change also with respect to the polynomial degree. The monitoring decisions lead to, as a simple empirical cross-check, good results in the following sense: For those country-pollutant combinations where no structural break is detected, using the specification and parameter values from the calibration period leads to good fit also for the full period until 2016, with obviously even better fit when

---

[5]The long list of theory contributions presenting specific models that lead to EKC-type behavior under certain assumptions includes Andreoni and Levinson (2001), Arrow *et al.* (1995), Brock and Taylor (2010), Cropper and Griffiths (1994), Jones and Manuelli (2001), Selden and Song (1995) or Stokey (1998).

[6]This means that we use our monitoring tools for an ex-post analysis rather than "true" online monitoring.

[7]This is obvious, since one can achieve perfect fit with a polynomial of degree sample size minus one. There is an ongoing discussion in the EKC literature concerning appropriate functional form and estimation strategies (see, e.g. Bertinelli and Strobl, 2005; Millimet *et al.*, 2003; Schmalensee *et al.*, 1998). Inverted U-shaped relationships are also considered, e.g., in the intensity-of-use intensity of use or material Kuznets curve (MKC) literature that investigates the potentially inverted U-shaped relationship between GDP and energy or metals use per unit of GDP (see, e.g. Grabarczyk *et al.*, 2018; Guzmán *et al.*, 2005; Labson and Crompton, 1993; Stuermer, 2018) for which the tools developed in this paper may also be useful.

[8]A cointegrating linear relationship implies tautologically that CPRs with higher polynomial degrees are also present, albeit with (theoretically) zero coefficients to the higher order powers.

re-estimating the relationship over the full sample.

The paper is organized as follows: Section 2 contains the setting, assumptions, monitoring statistics and asymptotic results. Section 3 discusses finite sample simulation results. Section 4 presents the monitoring application to CO_2 and SO_2 emissions data. Section 5 briefly summarizes and concludes. Appendix A contains all proofs. In addition, four supplementary appendices are available: Supplementary Appendix B discusses local asymptotic power properties. Supplementary Appendix C contains additional finite sample simulation results and Supplementary Appendix D presents additional empirical results. Supplementary Appendix E includes tables with critical values for the detectors for a broad variety of specifications relevant for EKC-type analysis. `MATLAB` code – including the critical values for the mentioned variety of specifications – for the monitoring statistics developed in this paper is available upon request.[9]

We use the following notation: Definitional equality is signified by $:=$ and $\Rightarrow$ denotes weak convergence. $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$ and $\text{diag}(\cdot)$ denotes a diagonal matrix. For a vector $x \in \mathbb{R}^n$ we use $\|x\|^2 = \sum_{i=1}^n x_i^2$ and for a matrix $M \in \mathbb{R}^{m \times n}$ we use $\|M\| = \sup_x \frac{\|Mx\|}{\|x\|}$. $\mathbb{E}(\cdot)$ denotes the expected value and L is the backward-shift operator, i.e., $L\{x_t\}_{t \in \mathbb{Z}} = \{x_{t-1}\}_{t \in \mathbb{Z}}$. The first-difference operator is denoted with $\Delta := 1 - L$. We denote the k-dimensional identity matrix with I_k. A Brownian motion with covariance matrix specified in the context is denoted by $B(r)$ and $W(r)$ denotes a standard Brownian motion.

2 Theory

2.1 Model, Assumptions and Parameter Estimation

We consider monitoring – using the terminology of Wagner and Hong (2016) – a cointegrating polynomial regression (CPR), i.e., a regression of the form:

$$y_t = \begin{cases} D_t'\theta_D + X_t'\theta_X + u_t, & t = 1, \ldots, \lfloor rT \rfloor, \\ D_t'\theta_{D,1} + X_t'\theta_{X,1} + u_t, & t = \lfloor rT \rfloor + 1, \ldots, T, \end{cases} \tag{1}$$

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

with $x_t := [x_{1t}, \ldots, x_{kt}]' \in \mathbb{R}^k$ and $X_t := [x_t', x_{kt}^2, \ldots, x_{kt}^{p_k}]' \in \mathbb{R}^p$ with $p = k-1+p_k$, the deterministic trend function $D_t \in \mathbb{R}^q$, the parameter vectors $\theta_D, \theta_{D,1} \in \mathbb{R}^q$ and $\theta_X, \theta_{X,1} \in \mathbb{R}^p$. Furthermore we define the combined parameter vectors $\theta := [\theta_D', \theta_X']'$, $\theta_1 := [\theta_{D,1}', \theta_{X,1}']' \in \mathbb{R}^{q+p}$.

---

[9]The `MATLAB` code can be straightforwardly modified to other specifications to obtain additional critical values; under the assumption of full design.

Under the null hypothesis no structural change occurs, that is $\theta_1 = \theta$ and $\{u_t\}_{t \in mZ}$ is an I(0) process, with detailed assumptions specified below, throughout. Under the alternative hypothesis, either some parameters change or the relation turns spurious, i.e., $\{u_t\}_{t \in mZ}$ turns into an I(1) process for *every* $\theta \in \mathbb{R}^{q+p}$, or both at a sample fraction $\lfloor rT \rfloor$ that has to be – as discussed in the introduction – larger than $\lfloor mT \rfloor$ for some $0 < m < 1$.[10] In formal terms:

$$H_0 \colon \begin{cases} \theta_1 = \theta \ \forall \ r \in [m, 1) \text{ and} \\ u_t \text{ is } I(0) \text{ for } t = 1, \ldots, T \end{cases}$$

$$H_1 \colon \begin{cases} \exists \ r \in [m, 1) \colon \theta_1 \neq \theta \text{ or} \\ \exists \ r \in [m, 1) \colon u_t, t = 1, \ldots, \lfloor rT \rfloor \text{ is } I(0) \text{ and} \\ u_t, t = \lfloor rT \rfloor + 1, \ldots, T, \text{ is } I(1). \end{cases}$$

Remark 1 *Note that the regression model given in* (1) *is a special case of the CPR model considered in Wagner and Hong (2016), since only one of the integrated regressors, w.l.o.g. x_{kt}, is allowed to enter the regression model with powers larger than one. Whilst this is, obviously, restrictive compared to the case where higher order powers of all elements of x_t can be present as regressors, this special case covers environmental and material Kuznets curves and similar applications. The mathematical reason for considering this special case is that it allows for – potentially up to a scalar nuisance parameter that can be consistently estimated and scaled out – nuisance parameter free limiting distributions of the considered detectors that can be simulated. In the terminology of Vogelsang and Wagner (2014b) a cointegrating polynomial regression needs to exhibit full design to allow for asymptotic standard inference by scaling out a scalar long-run variance. This is the case for EKC-type relationships with only one integrated regressor present with powers larger than one, but also for some other economically relevant more general cases of CPRs, e.g., Translog functions (see, e.g. Christensen et al., 1971). Given our focus on EKCs we abstain from formulating the results here in the most general form; the required extensions are straightforward. For even more general specifications that do not exhibit full design, a sub-sampling approach may be considered relying upon similar arguments as discussed in Wagner and Hong (2016, Proposition 6). The performance of sub-sampling based procedures, however, suffers particularly from short sample periods; as also illustrated by the simulations reported in Wagner and Hong (2016). Consequently, a sub-sampling based approach cannot be expected to perform well in a monitoring context.*

---

[10]Effectively, $\{u_t\}_{t \in \mathbb{Z}}$ being an I(0) process in this paper means that it satisfies Assumption 2. An I(1) process is a process that does not fulfill Assumption 2, but where the first difference does.

The results developed below rest upon the following assumptions:

Assumption 1 *There exists a sequence of $q \times q$ scaling matrices $G_D(T)$ and a q-dimensional vector of functions $D(z)$, with $\int_0^s D(z)D(z)'ds < \infty$ for $0 \le s \le 1$, such that for $0 \le s \le 1$ it holds that:*

$$\lim_{T\to\infty} \sqrt{T}G_D^{-1}(T)D_{\lfloor sT \rfloor} = D(s). \tag{3}$$

If, e. g., $D_t := (1, t, t^2, \ldots, t^{q-1})'$, then $G_D(T) := \operatorname{diag}(T^{1/2}, T^{3/2}, T^{5/2}, \ldots, T^{q-1/2})$ and $D(z) = (1, z, z^2, \ldots, z^{q-1})'$. In relation to the integrated regressors and the powers we need a scaling matrix $G_X(T) := \operatorname{diag}(T \times I_k, T^{3/2}, \ldots, T^{\frac{p_k+1}{2}})$ later.

Assumption 2 *The process $\{\eta_t\}_{t\in\mathbb{Z}} := \{[u_t, v_t']'\}_{t\in\mathbb{Z}}$ is generated under the null hypothesis as:*

$$\eta_t = C(L)\xi_t = \sum_{j=0}^{\infty} C_j \xi_{t-j}, \quad t \in \mathbb{Z}, \tag{4}$$

with $C_j \in \mathbb{R}^{(k+1)\times(k+1)}, j = 0, 1, \ldots,$ and the conditions:

$$\det(C(1)) \ne 0, \ \sum_{j=0}^{\infty} j\|C_j\| < \infty \ \text{and} \ \|C_0\| < \infty. \tag{5}$$

Furthermore, we assume that the process $\{\xi_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic martingale difference sequence with natural filtration $\mathcal{F}_t = \sigma(\{\xi_s\}_{-\infty}^t)$, $\mathbb{E}(\xi_t\xi_t'|\mathcal{F}_{t-1}) = \Sigma_{\xi\xi} > 0$ with in addition $\sup_{t\ge1}\mathbb{E}(\|\xi_t\|^a|\mathcal{F}_{t-1}) < \infty$ a.s. for some $a > 4$.

The assumptions on the deterministic components, the regressors and error terms are similar to the assumptions used in Wagner and Hong (2016) and, more implicitly, in Wagner and Wied (2017). In particular Assumption 2 is sufficient for a functional central limit theorem to hold for $\{\eta_t\}_{t\in\mathbb{Z}}$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor sT \rfloor} \eta_t \Rightarrow B(s) = \begin{bmatrix} B_u(s) \\ B_v(s) \end{bmatrix} = \Omega^{1/2}W(s), \quad 0 \le s \le 1, \tag{6}$$

with the positive definite long-run covariance matrix $\Omega := \sum_{j=-\infty}^{\infty} \mathbb{E}(\eta_{t-j}\eta_t')$ and $W(s) := [W_{u\cdot v}(s), W_v(s)']'$ a $(k+1)$-dimensional vector of standard Brownian motions. We also define the half long-run co-variance matrix $\Delta := \sum_{j=0}^{\infty} \mathbb{E}(\eta_{t-j}\eta_t')$. The matrices Ω and Δ are partitioned according to the partitioning of $B(s)$, i. e.,

$$\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_{uu} & \Delta_{uv} \\ \Delta_{vu} & \Delta_{vv} \end{bmatrix}. \tag{7}$$

Using, e. g. , the Cholesky decomposition of Ω yields:

$$\Omega^{1/2} = \begin{bmatrix} \omega_{u \cdot v} & \lambda_{uv} \\ 0 & \Omega_{vv}^{1/2} \end{bmatrix}, \tag{8}$$

where $\omega_{u \cdot v}^2 := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and $\lambda_{uv} := \Omega_{uv}(\Omega_{vv}^{1/2})^{-1}$. The conditional long-run variance $\omega_{u \cdot v}^2$ is a key quantity that needs to be estimated for all but the two *self-normalized* detectors.

Where needed, consistent long-run covariance estimation is performed non-parametrically, requiring the choice of both a kernel function and a bandwidth parameter. The inputs in the non-parametric estimation are the OLS residuals from estimating (1) over the calibration period and the first difference of x_t over the same period. For consistent long-run covariance estimation it suffices to assume (following, e. g. , Jansson, 2002):

Assumption 3 *The kernel function $k(\cdot)$ satisfies:*

(i) $k(0) = 1$, $k(\cdot)$ *is continuous at 0 and $\bar{k}(0) := \sup_{x \geq 0} |k(x)| < \infty$,*

(ii) $\int_0^\infty \bar{k}(x)dx < \infty$, *where $\bar{k}(x) = \sup_{y \geq x} |k(y)|$.*

Assumption 4 *The bandwidth parameter $M_T \subseteq (0, \infty)$ satisfies $\lim_{T \to \infty}(M_T^{-1} + T^{-1/2}M_T) = 0$.*

All our monitoring statistics discussed in the following subsection are based on consistent parameter estimators that are required to lead to limiting distributions that are nuisance parameter free up to a scalar parameter, the conditional long-run variance $\omega_{u \cdot v}^2$, that can (asymptotically) be scaled out, either by scaling by a consistent estimator, which we refer to later as standardized, or by self-normalization.

As indicated, estimation takes place on the calibration sample $t = 1, \ldots, \lfloor mT \rfloor$ for some $0 < m < 1$ that is known to be generated under the null hypothesis. This approach to monitoring, based on parameter estimation on a calibration sample known to be – or at least assumed to be – free of structural change has been popularized in the econometrics community by the seminal work of Chu *et al.* (1996).

The cointegration literature provides a variety of *modified* ordinary least squares estimators of θ with the required asymptotic properties, see, e. g. , Wagner (2018) for a survey. All these estimators commence from the fact that the OLS estimator of θ is consistent with – in case of regressor endogeneity and error serial correlation – a limiting distribution that is contaminated by second

8

order bias terms. These second order bias terms are removed, one way or another, by the various modifications of OLS. In this paper we consider three modified OLS estimators: Fully Modified OLS (FM-OLS), Dynamic OLS (D-OLS) and Integrated Modified OLS (IM-OLS). These three estimators have originally been developed for cointegrating linear regressions: FM-OLS in Phillips and Hansen (1990), D-OLS in Saikkonen (1991), Phillips and Loretan (1991) or Stock and Watson (1993) and IM-OLS in Vogelsang and Wagner (2014a). The extensions to the CPR setting are discussed for FM-OLS in Wagner and Hong (2016), for D-OLS in Choi and Saikkonen (2010) and for IM-OLS in Vogelsang and Wagner (2014b).[11]

Our brief discussion of the three estimators first necessitates the definition of a few more quantities, i. e., $Z_t := [D'_t, X'_t]'$ and $y^+_{t,m} := y_t - \Delta x'_t \hat{\Omega}^{-1}_{vv,m} \hat{\Omega}_{vu,m}$, with the second subscript m indicating that estimation of the long-run covariances is – as mentioned – also based on the calibration sample $t = 1, \ldots, \lfloor mT \rfloor$. Furthermore, define:

$$
A^*_m := \begin{bmatrix} 0_{q \times 1} \\ \lfloor mT \rfloor \hat{\Delta}^+_{vu,m} \\ M^*_m \end{bmatrix}, \qquad M^*_m := \hat{\Delta}^+_{v_k u,m} \begin{bmatrix} 2 \sum_{t=1}^{\lfloor mT \rfloor} x_{kt} \\ \vdots \\ p_k \sum_{t=1}^{\lfloor mT \rfloor} x^{p_k-1}_{kt} \end{bmatrix}, \tag{9}
$$

with $\hat{\Delta}^+_{vu,m} := \hat{\Delta}_{vu,m} \hat{\Omega}^{-1}_{vv,m} \hat{\Delta}_{vv,m}$ and $\hat{\Delta}^+_{v_k u,m} := \hat{\Delta}_{v_k u,m} \hat{\Omega}^{-1}_{vv,m} \hat{\Delta}_{vv_k,m}$.

Long-run covariance estimation uses the OLS residuals of (1) from estimation over the calibration period $t = 1, \ldots, \lfloor mT \rfloor$ in conjunction with the first differences of the integrated regressors, i.e., $\hat{\eta}_{t,m} := [\hat{u}_{t,m}, v'_t]'$, with $\hat{u}_{t,m}$ denoting the OLS residuals here:

$$
\hat{\Delta}_m := \frac{1}{\lfloor mT \rfloor} \sum_{h=0}^{\lfloor mT \rfloor - 1} k \left(\frac{h}{M_T} \right) \sum_{t=1}^{\lfloor mT \rfloor - h} \hat{\eta}_{t,m} \hat{\eta}'_{t+h,m}, \tag{10}
$$

$$
\hat{\Sigma}_m := \frac{1}{\lfloor mT \rfloor} \sum_{t=1}^{\lfloor mT \rfloor} \hat{\eta}_{t,m} \hat{\eta}'_{t,m}, \tag{11}
$$

$$
\hat{\Omega}_m := \hat{\Delta}_m + \hat{\Delta}'_m - \hat{\Sigma}_m. \tag{12}
$$

In both the finite sample simulations as well as the application we use for long-run covariance estimation, in line with Assumptions 3 and 4, the Bartlett kernel with bandwidth chosen according to Newey and West (1994).

---

[11]To be precise, Choi and Saikkonen (2010) propose an extension of the dynamic regression approach, adding leads and lags of the first differences of the integrated regressors, to a more general setting than CPRs. Given that the CPR model is linear in parameters, D-OLS can be extended relatively straightforwardly, without the need to resort to modified nonlinear least squares type estimators. Vogelsang and Wagner (2014b) consider an extension of IM-OLS to general multivariate polynomials allowing also for arbitrary cross-products of powers of integrated regressors. Stypka and Wagner (2020) extend the FM-OLS estimation principle to this more general polynomial-type setting.

With all required quantities defined, the FM-OLS estimator computed over the calibration sample is given by:

$$\hat{\theta}_m^F := \left(\sum_{t=1}^{\lfloor mT \rfloor} Z_t Z_t' \right)^{-1} \left(\sum_{t=1}^{\lfloor mT \rfloor} Z_t y_{t,m}^+ - A_m^* \right). \tag{13}$$

While FM-OLS is based on a two-part nonparametric transformation to remove endogeneity and serial correlation related bias terms from the limiting distribution of the OLS estimator, D-OLS is based on a more "classical projection and orthogonalization" argument by performing OLS estimation in an augmented version of the CPR regression (1), with leads and lags of the first differences of x_t added as regressors to "clean the limiting distribution". The D-OLS regression – estimated by OLS over the calibration sample – is given by:

$$y_t = Z_t'\theta + \sum_{j=-d_1}^{d_2} \Delta x_{t-j}' \Theta_j + u_t, \tag{14}$$

with the number of leads (d_1) and lags (d_2) chosen to ensure consistent parameter estimation of θ with a limiting distribution that is – up to a scalar – nuisance parameter free. In general this requires that $d_1, d_2 \to \infty$ at suitable rates. More specifically, in our finite sample simulations and the application, we choose leads and lags using the AIC-type criterion of Choi and Kurozumi (2012). The resultant OLS estimators of θ and $\hat{\Theta}_j$ from (14) estimated over the calibration sample are referred to as $\hat{\theta}_m^D$ and $\hat{\Theta}_{j,m}^D$, respectively.

The third estimation principle addresses endogeneity correction by partial summation. Define for a sequence z_t, $t = 1, \ldots, T$ the partial summed variable by $S_t^z := \sum_{j=1}^{t} z_j, t = 1, \ldots, T$. Then the IM-OLS regression – estimated by OLS over the calibration sample – is given by:

$$S_t^y = S_t^{Z'}\theta + x_t'\varphi + S_t^u. \tag{15}$$

The OLS estimators of θ and φ from (15) estimated over the calibration sample are referred to as $\hat{\theta}_m^I$ and $\hat{\varphi}_m^I$, respectively. Note that endogeneity correction in the IM-OLS estimator does not require any leads-lags or kernel-bandwidth choices, as it suffices to simply add the original integrated regressor vector x_t to the partial summed regression.

The key input for the monitoring statistics discussed in the following subsection are the residual (processes) obtained with these three estimators. In particular, the asymptotic null behavior of the residual (partial sum) processes is the key ingredient to derive asymptotic properties

of any of our variance-ratio type monitoring statistics. This result is formalized in the following lemma, whose formulation requires to define some additional (asymptotic) quantities first, i.e., $\mathbf{W}_v(s) := [W_{v_1}(s), W_{v_2}(s), \ldots, W_{v_k}(s), W_{v_k}^2(s), \ldots, W_{v_k}^{p_k}(s)]'$, $J(s) := [D(s)', \mathbf{W}_v(s)']'$, $f(s) := [\int_0^s D(z)'dz, \int_0^s \mathbf{W}_v(z)'dz, W_v(s)']'$ and $F(s) := \int_0^s f(z)dz$.

Lemma 1 *Let the data be generated according to* (1) *and* (2) *with Assumptions 1 and 2 in place. Furthermore, let long-run covariance estimation be performed under Assumptions 3 and 4 and let lead-lag choices be made as discussed in Choi and Kurozumi (2012).*

Based on estimation over the calibration period $t = 1, \ldots, \lfloor mT \rfloor$ – with the estimators $\hat{\theta}_m^F$, $\hat{\theta}_m^D$ and $\hat{\theta}_m^I$ as discussed before – denote the FM-OLS residuals by:

$$\hat{u}_{t,m}^F := y_{t,m}^+ - Z_t'\hat{\theta}_m^F, \tag{16}$$

the D-OLS residuals by:

$$\hat{u}_{t,m}^D := y_t - Z_t'\hat{\theta}_m^D - \sum_{j=-d_1}^{d_2} \Delta x_{t-j}'\hat{\Theta}_{j,m}^D \tag{17}$$

and the IM-OLS residuals by:

$$\hat{S}_{t,m}^{u,I} := S_t^y - S_t^{Z'}\hat{\theta}_m^I - x_t'\hat{\varphi}_m^I. \tag{18}$$

For $T \to \infty$ it holds under the null hypothesis for $m \le s \le 1$ that:[12]

$$\frac{1}{\sqrt{T}}\hat{S}_{\lfloor sT \rfloor,m}^{u,F} := \frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor sT \rfloor} \hat{u}_{t,m}^F \Rightarrow \omega_{u\cdot v}\left(W_{u\cdot v}(s) - \int_0^s J(z)'dz \left(\int_0^m J(z)J(z)'dz\right)^{-1} \int_0^m J(z)dW_{u\cdot v}(z)\right)$$
$$=: \omega_{u\cdot v}\widetilde{W}_{u\cdot v,m}(s) \tag{19}$$

$$\frac{1}{\sqrt{T}}\hat{S}_{\lfloor sT \rfloor,m}^{u,D} := \frac{1}{\sqrt{T}}\sum_{t=d_2+2}^{\min\{\lfloor sT \rfloor, T-d_1\}} \hat{u}_{t,m}^D \Rightarrow \omega_{u\cdot v}\widetilde{W}_{u\cdot v,m}(s) \tag{20}$$

$$\frac{1}{\sqrt{T}}\hat{S}_{\lfloor sT \rfloor,m}^{u,I} := \frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor sT \rfloor} \Delta\hat{S}_{t,m}^{u,I} \Rightarrow \omega_{u\cdot v}\left(W_{u\cdot v}(s) - f(s)'\left(\int_0^m f(z)f(z)'dz\right)^{-1} \int_0^m [F(m) - F(z)]dW_{u\cdot v}(z)\right)$$
$$=: \omega_{u\cdot v}\widetilde{P}_{u\cdot v,m}(s). \tag{21}$$

---

[12]For the asymptotic results the lower bounds of the summations could all be set equal to $t = 1$. We, however, start the sums with the first residual actually available for computations, at the expense of potentially making matters appear overly complicated, at least in terms of notation, but replicable for implementation by the reader.

The lemma shows that indeed all three partial sum processes of the residuals converge to processes that are (i) functionals of standard Brownian motions, $W_v(r)$ and $W_{u \cdot v}(r)$, and (ii) proportional to $\omega_{u \cdot v}$, a scalar nuisance parameter that can be consistently estimated and hence scaled out from the limit processes or that can be eliminated by self-normalization. The limiting null distributions of test statistics based on the (normalized) limit processes consequently can be obtained by simulating the corresponding functionals of standard Brownian motions. Note that the FM-OLS and D-OLS residual partial sum processes converge to the same limiting process, which is a consequence of these two estimators having identical limiting distributions.

2.2 The Monitoring Statistics

Similarly to Wagner and Wied (2017) the starting point of our monitoring statistics is to combine the approach of Chu *et al.* (1996) with variance-ratio statistics that diverge under the alternative. More specifically, the underlying variance-ratio (full sample) statistic motivating the construction of our monitoring statistics is the Kwiatkowski *et al.* (1992) stationarity test, respectively the related Shin (1994) cointegration test.[13] Using our notation, the (full sample) Shin-statistic is given by:

$$
\begin{aligned}
T_{\mathrm{Shin}} &:= \frac{1}{\hat{\omega}_{u \cdot v}^2} \left(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{T}} \sum_{i=1}^{t} \hat{u}_i \right)^2 \right) \\
&= \frac{1}{\hat{\omega}_{u \cdot v}^2} \left(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{T}} \hat{S}_i^u \right)^2 \right),
\end{aligned}
\tag{22}
$$

with $\hat{u}_t$ denoting the residuals from (full sample) estimation with, e. g. , FM-OLS or D-OLS. In case that IM-OLS is used for estimation, the resultant residuals are already partial summed quantities, i. e. , one immediately obtains (by construction) quantities to insert into the expression in the second line of (22). The test statistic given above converges under the null hypothesis to a functional of standard Brownian motions, which is as expected when considering (22), where convergence to the squared integral of a *standard* Brownian motion follows immediately from our assumptions if instead of $\hat{u}_t$ the errors u_t were used (and scaling would take place by a consistent estimator of ω_u^2). Using $\hat{u}_t$ instead of u_t leads to a similar result, but with a different (specification dependent) functional of standard Brownian motions after scaling out $\omega_{u \cdot v}^2$ rather than ω_u^2. To be precise, when

---

[13]For completeness note that the Shin (1994) test has been considered in the CPR setting in Wagner and Hong (2016). In principle, of course, also other variance-ratio type statistics for the null hypothesis of stationarity – or cointegration – could serve as building blocks, e. g. , the test statistic of Busetti and Taylor (2004) or Kim (2000) more or less directly leads, when extended and considered for monitoring, to a self-normalized detector similar to $\hat{H}_{sn}^m$.

using FM-OLS or D-OLS for parameter estimation the limiting null distribution will be a function of $\widetilde{W}_{u\cdot v,m}(r)$. When using IM-OLS for parameter estimation the limiting null distribution will be a function of $\widetilde{P}_{u\cdot v,m}(r)$, see Proposition 1 below.

The above test statistic (22) can be easily seen to diverge under the alternative hypothesis of a structural break *occurring after the calibration period*. Consider, e. g., the FM-OLS residuals (with the argument entirely analogous for all estimators) using our already established notation:

$$\hat{u}_{t,m}^{\mathrm{F}} := y_{t,m}^+ - Z_t' \hat{\theta}_m^{\mathrm{F}} \tag{23}$$

$$= u_t - v_t' \hat{\Omega}_{vv,m}^{-1} \hat{\Omega}_{vu,m} - D_t'(\hat{\theta}_{D,m}^{\mathrm{F}} - \theta_D) - X_t'(\hat{\theta}_{X,m}^{\mathrm{F}} - \theta_X)$$

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \hat{u}_{t,m}^{\mathrm{F}} = \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} u_t - \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} v_t' \hat{\Omega}_{vv,m}^{-1} \hat{\Omega}_{vu,m} - \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} D_t'(\hat{\theta}_{D,m}^{\mathrm{F}} - \theta_D) \tag{24}$$

$$- \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} X_t'(\hat{\theta}_{X,m}^{\mathrm{F}} - \theta_X)$$

Now, suppose that at some time point $\lfloor rT \rfloor > \lfloor mT \rfloor$ a structural change occurs. If, e. g., $\{u_t\}_{t\in\mathbb{Z}}$ turns from being I(0) to I(1), then the first term in (24) diverges for $s > r$. Similarly, the third or fourth term (or both) diverge in case of change in the parameter vector, i. e., when $\theta_1 \neq \theta$, as of course $\hat{\theta}_m^{\mathrm{F}} \to \theta$, because of parameter estimation on the calibration sample.

We consider for each of the three considered estimators – neglecting for notational brevity the dependence of the residuals and thus the test statistics on the estimation method – five monitoring statistics:

$$\hat{H}^m(s) := \frac{1}{\hat{\omega}_{u\cdot v,m}^2} \left(\frac{1}{T} \sum_{i=\lfloor mT \rfloor+1}^{\lfloor sT \rfloor} \left(\frac{1}{\sqrt{T}} \hat{S}_{i,m}^u \right)^2 \right) \tag{25}$$

$$\hat{H}_d^m(s) := \frac{1}{\hat{\omega}_{u\cdot v,m}^2} \left(\frac{1}{T} \sum_{i=\lfloor mT \rfloor+1}^{\lfloor sT \rfloor} \left(\frac{1}{\sqrt{T}} \hat{S}_{i,m}^u \right)^2 - \frac{1}{T} \sum_{i=1}^{\lfloor mT \rfloor} \left(\frac{1}{\sqrt{T}} \hat{S}_{i,m}^u \right)^2 \right) \tag{26}$$

$$\hat{H}_{sn}^m(s) := \frac{\sum_{i=\lfloor mT \rfloor+1}^{\lfloor sT \rfloor} \left(\hat{S}_{i,m}^u \right)^2}{\sum_{i=1}^{\lfloor mT \rfloor} \left(\hat{S}_{i,m}^u \right)^2} \tag{27}$$

$$\hat{H}_{mov}^{m,n}(s) := \frac{1}{\hat{\omega}_{u\cdot v,m}^2} \left(\frac{1}{T} \sum_{i=\max\{1,\lfloor sT \rfloor-\lfloor nT \rfloor+1\}}^{\lfloor sT \rfloor} \left(\frac{1}{\sqrt{T}} \hat{S}_{i,m}^u \right)^2 \right) \tag{28}$$

$$\hat{H}_{mov,sn}^{m,n}(s) := \frac{\sum_{i=\max\{1,\lfloor sT \rfloor-\lfloor nT \rfloor+1\}}^{\lfloor sT \rfloor} \left(\hat{S}_{i,m}^u \right)^2}{\sum_{i=1}^{\lfloor mT \rfloor} \left(\hat{S}_{i,m}^u \right)^2} \tag{29}$$

The monitoring statistic $\hat{H}^m(s)$ given in (25) is of the same form as the monitoring statistic used in Wagner and Wied (2017) considered here in the CPR context. The monitoring statistic $\hat{H}_d^m(s)$ given in (26) – with a term calculated only over the calibration sample subtracted – is of a similar form as used in Chu *et al.* (1996). The third variant $\hat{H}_{sn}^m(s)$ given in (27) is a *self-normalized* statistic, for which under the null hypothesis both the numerator and denominator converge (appropriately scaled) to functionals of standard Brownian motions proportional to $\omega_{u \cdot v}$, which is hence scaled out in the ratio. Long-run covariance estimation is known to be a notoriously problematic aspect in unit root and cointegration analysis and therefore test statistics that do not require this step may exhibit better performance. The fourth considered variant is a *moving window* statistic $\hat{H}_{mov}^{m,n}$ given in (28) with n denoting the moving window (sample fraction or) length. The key difference between the moving window detector and the *expanding* window detectors is that $\hat{H}_{mov}^{m,n}$ is based on a constant number of residual partial sums for all values of s. This construction increases, under the alternative hypothesis, the impact of post-break residuals on the test statistic, which is ex ante expected to lead to faster detection of structural breaks. The performance of the fourth variant will depend on the length of the moving window, to be chosen in applications. Finally, the fifth monitoring statistic $H_{mov,sn}^{m,n}$ given in (29) combines self-normalization and moving window estimation, with the performance as for the fourth variant expected to depend upon the moving window length.[14] The following proposition summarizes the asymptotic behavior of the monitoring statistics under the null hypothesis.

Proposition 1 *Let the data be generated according to* (1) *and* (2) *with Assumptions 1 and 2 in place. Furthermore, let long-run covariance estimation be performed under Assumptions 3 and 4 and let lead-lag choices be made as discussed in Choi and Kurozumi (2012).*
In case parameter estimation is performed with FM-OLS or D-OLS the limiting process $\widetilde{Q}_{u \cdot v,m}(s)$ below equals $\widetilde{W}_{u \cdot v,m}(s)$ and in case parameter estimation is performed by IM-OLS it equals $\widetilde{P}_{u \cdot v,m}(s)$. The defined monitoring statistics converge under the null hypothesis for $T \to \infty$, in particular it

---

[14]To be precise, only when using $\hat{H}_{sn}^m(s)$ or $\hat{H}_{mov,sn}^{m,n}(s)$ in conjunction with D-OLS or IM-OLS no long-run covariance estimators are required, whereas estimated long-run covariances are required for FM-OLS estimation. For D-OLS still lead-lag length choices have to be made and only when using $\hat{H}_{sn}^m(s)$ or $\hat{H}_{mov,sn}^{m,n}(s)$ in conjunction with the IM-OLS estimator no kernel/bandwidth or lead-lag choices have to be made. In this case, the only choice to still be made when using $\hat{H}_{sn}^m(s)$ is the length of the calibration sample, a choice required throughout. In case of $\hat{H}_{mov,sn}^{m,n}(s)$ both the calibration sample length m and the moving window length n have to be chosen.

holds that:

$$\hat{H}^m(s) \Rightarrow \int_m^s \widetilde{Q}^2_{u \cdot v, m}(z) dz =: \mathcal{H}^m(\widetilde{Q}_{u \cdot v, m}, s) \tag{30}$$

$$\hat{H}^m_d(s) \Rightarrow \int_m^s \widetilde{Q}^2_{u \cdot v, m}(z) dz - \int_0^m \widetilde{Q}^2_{u \cdot v, m}(z) dz =: \mathcal{H}^m_d(\widetilde{Q}_{u \cdot v, m}, s) \tag{31}$$

$$\hat{H}^m_{sn}(s) \Rightarrow \frac{\int_m^s \widetilde{Q}^2_{u \cdot v, m}(z) dz}{\int_0^m \widetilde{Q}^2_{u \cdot v, m}(z) dz} =: \mathcal{H}^m_{sn}(\widetilde{Q}_{u \cdot v, m}, s) \tag{32}$$

$$\hat{H}^{m,n}_{mov}(s) \Rightarrow \int_{\max\{0, s-n\}}^s \widetilde{Q}^2_{u \cdot v, m}(z) dz =: \mathcal{H}^{m,n}_{mov}(\widetilde{Q}_{u \cdot v, m}, s) \tag{33}$$

$$\hat{H}^{m,n}_{mov,sn}(s) \Rightarrow \frac{\int_{\max\{0, s-n\}}^s \widetilde{Q}^2_{u \cdot v, m}(z) dz}{\int_0^m \widetilde{Q}^2_{u \cdot v, m}(z) dz} =: \mathcal{H}^{m,n}_{mov,sn}(\widetilde{Q}_{u \cdot v, m}, s) \tag{34}$$

It is widely-used practice in monitoring to base the decision not on monitoring statistics as just defined, but on monitoring statistics divided by a weighting function, $g(s)$ say. For chosen weighting function $g(s)$ – with $0 < g(s) < \infty$ – the null hypothesis is rejected, if the weighted monitoring statistic $\left| \frac{\hat{H}(s)}{g(s)} \right|$ is larger than a critical value c for the first time. We denote this point in time as *detection time τ_m*, i.e.:

$$\tau_m := \min_{s : \lfloor mT \rfloor + 1 \leq \lfloor sT \rfloor \leq T} \left\{ \left| \frac{\hat{H}(s)}{g(s)} \right| > c \right\}, \tag{35}$$

with $\hat{H}(s)$ short-hand notation for any of the considered detectors. In case no structural change is detected, i. e. , $\left| \frac{\hat{H}(s)}{g(s)} \right| \leq c$ for all $m \leq s \leq 1$, we set $\tau_m = \infty$. A finite value of τ_m not only indicates a structural break but also contains information about the location of the potential break point.

Weighting function and critical value have to be chosen so that under the null hypothesis it holds that:

$$
\begin{aligned}
\lim_{T \to \infty} \mathbb{P}(\tau_m < \infty) &= \lim_{T \to \infty} \mathbb{P} \left(\min_{s : [mT]+1 \leq [sT] \leq T} \left\{ \left| \frac{\hat{H}(s)}{g(s)} \right| > c \right\} < \infty \right) \\
&= \lim_{T \to \infty} \mathbb{P} \left(\sup_{s : [mT]+1 \leq [sT] \leq T} \left| \frac{\hat{H}(s)}{g(s)} \right| > c \right) \\
&= \mathbb{P} \left(\sup_{m \leq s \leq 1} \left| \frac{\mathcal{H}(s)}{g(s)} \right| > c \right) = \alpha,
\end{aligned}
\tag{36}
$$

with α denoting the chosen significance level, and $\mathcal{H}(s)$ short-hand notation for the limit corresponding to the considered monitoring statistic. Considering only positive and bounded weighting functions, see also Aue *et al.* (2012, Assumption 3.6), allows to derive the required result given above based on the developed asymptotic null behavior of the monitoring statistics and the continuous mapping theorem.

Proposition 2 *Let the data be generated according to* (1) *and* (2) *with Assumptions 1 and 2 in place. Furthermore, let long-run covariance estimation be performed under Assumptions 3 and 4 and let lead-lag choices be made as discussed in Choi and Kurozumi (2012). In addition assume that the weighting function $g(s)$ is continuous and bounded. Then there exist critical values $c = c(\alpha, g, m, n)$ such that for any $0 < \alpha < 1$ it holds that:*

$$\lim_{T \to \infty} \mathbb{P}\left(\tau_m(\hat{H}(s), g(s), c) < \infty\right) = \alpha \tag{37}$$

Clearly, the choice of a weighting function $g(s)$ impacts the performance of monitoring procedures and has to combine two opposing goals: (i) small size distortions under the null hypothesis and (ii) small delays under the alternative hypothesis, that is, detection of a break as soon as possible after the break. The discussion in Chu *et al.* (1996, Section 3) makes clear that it is in general, even in more standard regression models, impossible to derive analytically tractable optimal weighting functions (from a certain class of functions), e. g., with respect to minimal expected delay.[15]

Given the lack of analytical results concerning *optimal* choices of weighting functions we have performed a large number of preliminary simulations using a range of candidate weighting functions.[16] The starting point of these considerations is Wagner and Wied (2017), who choose the weighting function in relation to the expected value of the monitoring statistics, resulting in $g(s) = s^3$ in case $D_t = 1$ (intercept only) and $g(s) = s^5$ in case $D_t = [1, t]'$ (intercept and linear trend). In case of a linear trend we have in addition experimented with $g(s) \in \{1, s^{10}, s^{5(0.5+m)}, s^{5(0.85+m)^2}, \sqrt{m}(1 + \frac{s-m}{m}), \frac{s}{\sqrt{m}}\left(\frac{s-m}{s}\right)^{1/2}\}$, with the last two functions inspired by Horváth *et al.* (2004).[17] It turns out that no weighting, i. e., $g(s) = 1$ does not lead to favorable performance compared to $g(s) = s^3$ or s^5. The function s^{10} is chosen by "extrapolation" of the fact that s^5 works better than $s^0 = 1$. The idea of the third and fourth functions is – merely the result of some experimentation and heuristics – to make the detector more sensitive by increasing the value of the statistic whilst at the same leaving the critical values effectively unchanged. The effects are

---

[15]Aue *et al.* (2009) derive the limiting distributions of the delay time for a one-time parameter change in a linear regression model with stationary errors for a simple class of weighting functions depending only upon a single (tuning) parameter. The situation is much more involved in our context and any result concerning asymptotic distributions of delay times will depend upon intricate crossing-probability calculations involving complicated functions of Brownian motions. Results in this direction therefore appear to be very hard to obtain, for us at least.

[16]We have performed the type of simulations reported in Section 3 investigating the performance with respect to null rejection probabilities, size corrected power and detection times for all weighing functions discussed here. The simulations in Section 3 report the results based on the overall best performing weighting function, s^3 (intercept only) and s^5 (intercept and linear trend).

[17]In the intercept only specification the set of functions considered is given by $\{1, s^6, s^{3(0.5+m)}, s^{3(0.85+m)^2}, \sqrt{m}(1 + \frac{s-m}{m}), \frac{s}{\sqrt{m}}\left(\frac{s-m}{s}\right)^{1/2}\}$. The observations are similar for both specifications of the deterministic component.

to a certain extent as expected, without, however, leading to overall better performance. Taking $s^6 = (s^3)^2$ or $s^{10} = (s^5)^2$ as weighting functions does indeed lead, e. g., to earlier detection times, however, often also to detections in cases when there is no structural change, i. e., these functions lead to larger over-rejections under the null hypothesis. The two functions $s^{5(0.5+m)}, s^{5(0.85+m)^2}$, where we have also experimented with other powers and values, try to strike a balance between earlier rejections and size distortions. Altogether, however, the simple functions s^3 and s^5 perform most stably over a variety of configurations, in terms of comparably low size distortions under the null as first priority and short delays in the detection times.[18] Finally, the two functions inspired by Horváth *et al.* (2004), where we have also experimented with different powers, lead to essentially the same null rejection probabilities and size corrected power as, e. g., s^3 or s^5, but lead to partly substantially bigger delays than the other weighting functions. Therefore, we stick to the weighting functions already used also in Wagner and Wied (2017), i. e., the end point of the considerations is the starting point. As mentioned, it remains an open challenge to make progress on finding *optimal* weighting functions for the monitoring problem and detectors considered in this paper, or more generally when monitoring cointegrating relationships.

It remains to characterize the asymptotic behavior of the proposed monitoring procedures under the relevant alternatives in our setting: First, the error process $\{u_t\}_{t\in\mathbb{Z}}$ changes its behavior from I(0) to I(1), i. e., it changes to being an integrated process and second, there are breaks in (some of) the parameter values. For both cases we consider the asymptotic behavior against fixed and local alternatives, with the local alternatives having to be specified, as always, in line with the convergence rates of parameter estimation.

Proposition 3 *Let the data be generated for $t = 1, \ldots, \lfloor rT \rfloor$ according to (1) and (2) with Assumptions 1 and 2 in place. Furthermore, let long-run covariance estimation be performed under Assumptions 3 and 4 and let lead-lag choices be made as discussed in Choi and Kurozumi (2012). In addition assume that the weighting function $g(s)$ is continuous, positive and bounded. Furthermore, $\hat{H}(s)$ denotes again any of the considered monitoring statistics.*

(a) Let

(i) $\{u_t\}_{t\in\mathbb{Z}}$ be an I(1) process from $\lfloor rT \rfloor + 1$ onwards, or

---

[18]More specifically, the simpler functions lead to the lowest over-rejections almost throughout, the "race" in terms of size-corrected power is relatively even, and in some cases the more complicated weighting functions, in particular $s^{3(0.85+m)^2}$ or $s^{5(0.85+m)^2}$ lead to slightly smaller delays.

(ii) $\theta_1 \neq \theta$, with the condition

$$\lim_{T \to \infty} \frac{1}{\sqrt{T}} \sum_{t=\lfloor rT \rfloor + 1}^{T} D_t'(\theta_D - \theta_{D,1}) = \pm \infty \qquad (38)$$

fulfilled.

Then the monitoring procedures are consistent, i. e. , for any $0 < c < \infty$ it holds that

$$\lim_{T \to \infty} \mathbb{P}(\tau_m(\hat{H}(s), g(s), c) < \infty) = 1. \qquad (39)$$

(b) Let

(i) $\{u_t\}_{t \in \mathbb{Z}} = \{u_t^0\}_{t \in \mathbb{Z}}$ for all $t \leq \lfloor rT \rfloor$, with $\{u_t^0\}_{t \in \mathbb{Z}}$ satisfying Assumption 2 and

$$u_t = u_t^0 + \frac{\delta}{T} \sum_{i=\lfloor rT \rfloor + 1}^{t} \gamma_i, \qquad (40)$$

for all $t > \lfloor rT \rfloor$, where $\{u_t^0\}_{t \in \mathbb{Z}}$ and $\{\gamma_t\}_{t \in \mathbb{Z}}$ are independent processes and where $\{\gamma_t\}$ fulfills an invariance principle with long-run variance $\omega_\gamma^2 > 0$ and $\delta > 0$;

(ii) $\theta_{D,1} = \theta_D + G_D^{-1}(T)\Delta_{\theta_D}$ from $\lfloor rT \rfloor + 1$ onwards with $G_D(T)$ as in Assumption 1 and Δ_{θ_D} fulfilling

$$\int_r^1 D(z)'dz\Delta_{\theta_D} \neq 0; \qquad (41)$$

or

(iii) $\theta_{X,1} = \theta_X + G_X^{-1}(T)\Delta_{\theta_X}$ from $\lfloor rT \rfloor + 1$ onwards with $\Delta_{\theta_X} \neq 0$.

Then for any $0 < \varepsilon \leq 1 - \alpha$ and the critical value $0 < c(\alpha) < \infty$ from Proposition 2 there exists a $\delta = \delta(c, g)$, $\Delta_{\theta_X} = \Delta_{\theta_X}(c, g)$ or $\Delta_{\theta_D} = \Delta_{\theta_D}(c, g)$ such that

$$\lim_{T \to \infty} \mathbb{P}(\tau_m(\hat{H}(s), g(s), c) < \infty) \geq 1 - \varepsilon. \qquad (42)$$

The local asymptotic power (LAP) properties of the procedures with respect to break type, estimation method, self-normalization and window size of the moving window detectors, are discussed in some detail in Supplementary Appendix B.[19] The LAP results carry over by and large to similar relative performance findings in the finite sample simulations presented in Section 3, e. g. , in case of I(1) breaks a small moving window leads to both highest LAP as well as highest size-corrected finite

---

[19] Figures 1 to 7 in Supplementary Appendix B display corresponding LAP results.

sample power. Another example are power differences across estimation methods, with IM-OLS dominated by FM-/D-OLS in terms of LAP and also in finite sample size-corrected power for large sample sizes.

The critical values depend on the specification in several ways: on the specification of the deterministic component, the number of I(1) regressors, the highest power of the single integrated regressor that enters the CPR with higher powers, the detector, the estimation method and – as indicated by Proposition 2 – the weighting function, the calibration fraction and, for the moving window detectors, the window size. Supplementary Appendix E provides tables containing critical values for the usual deterministic components, i.e., an intercept only and intercept and linear trend, and up to four integrated regressors, of which one regressor enters the model with up to power three. Furthermore, the critical values are available for all considered detectors, FM-/D-OLS and IM-OLS estimation and significance levels of $0.01, 0.025, 0.05, 0.1$. The weighting function $g(s)$ is set to s^3 and s^5, depending on the deterministic specification. Critical values are available for a fine grid of the calibration fraction m ranging from $0.1, 0.11, \ldots, 0.9$. For the moving window detectors, critical values are available for the window sizes equal to $10\%, 20\%$ and 30% of the sample size.[20]

It remains to clarify the "meaning" of m and T for the monitoring procedures (see also p. 967 in Wagner and Wied, 2017). It is convenient to interpret T as the sample size including the out-of-sample monitoring period. Let T_0 denote the length of the actually available sample – in our application in Section 4 $T_0 = 71$ with annual data from 1946–2016. Then, denote $T = T_0 + H$, with $H > 0$ indicating that out of sample monitoring is intended and $H = 0$ indicating that monitoring takes place on a historic data set, as in Section 4. The fact that the critical values depend on m, and the moving detectors on n, means that a decision has to be made about the length of the calibration period, potentially the length of the moving window and about the out-of-sample monitoring period H prior to the analysis. The latter necessity renders our procedure a closed-end monitoring procedure. The calibration period will be chosen as large as possible (as a sub-sample $1, \ldots, T_C$ of $1, \ldots, T_0$) to increase the precision of the parameter estimates while avoiding the risk of having a structural break in the calibration period. Now, m is given by $m = \frac{T_C}{T_0 + H}$. Thus, choosing H larger implies that m is smaller, which in turn implies that the critical values are larger (since they are decreasing in m). This decreases ceteris paribus, despite asymptotic size control, the empirical rejection probabilities under both the null and the alternative. This is the reason

---

[20]Altogether this makes for more than 400 pages of tables. However, of course these are *embedded* in the available `MATLAB` code.

why one should choose the monitoring period as short as possible, a calibration period as large as possible and an out-of-sample monitoring period as short as possible.

Finally, similar as in Wagner and Wied (2017), the procedures are consistent also against a variety of other forms of structural changes, with all results following more or less straightforwardly from the construction principle. Of course, the finite sample performance may be relatively poor in some cases.

Remark 2 *The developed monitoring procedures are, in addition to the results provided in Proposition 3, also consistent against the following types of structural change:*

(i) *The process $\{u_t\}_{t\in\mathbb{Z}}$ changes its behavior from I(0) to being a near-integrated process, compare Phillips (1987), from $\lfloor rT\rfloor + 1$ onwards. In this case, effectively, under the alternative functionals of Wiener processes will be replaced by functionals of Ornstein-Uhlenbeck processes. The rates of divergence are the same as for the I(1) alternative.*

(ii) *Similarly, consistency also prevails in case $\{u_t\}_{t\in\mathbb{Z}}$ changes its behavior from $\lfloor rT\rfloor + 1$ onwards to being fractionally integrated, compare Davidson and de Jong (2000), with fractional integration parameter $0 < f < 1/2$. In this case, contrary to item (i) the divergence rate under the alternative changes and depends upon f, since under this alternative $\frac{1}{T^{1/2+f}}\sum_{t=\lfloor rT\rfloor+1}^{\lfloor sT\rfloor} u_t$ converges to a fractional Brownian motion. Thus, the smaller f, the more difficult it will be be to detect this form of structural change.*

(iii) *The approach can also be employed for detecting bubbles. In the recent literature, a bubble is often characterized as a period where the behavior of a time series has switched to explosive behavior, compare, e. g., Phillips et al. (2011), Phillips et al. (2015a; 2015b) and Phillips and Shi (2018). Thus, our procedure allows to detect (the beginning of) a bubble by considering the first difference of the series, since in the absence of a bubble the first differences are stationary, whereas in case of explosive behavior also the first differences exhibit explosive behavior.*

(iv) *In relation to the previous item, with bubbles typically considered to be temporary rather than permanent phenomena, it has to be noted that our procedures will be consistent in detecting episodes of I(1) or explosive behavior, as long as these episodes have asymptotically positive length. E. g., in the case of only one period under the alternative it has to hold that this period occurs over a sub-sample of the form $\lfloor r_1 T\rfloor,\ldots,\lfloor r_2 T\rfloor$ with $r_1 < r_2$. It is immediate that consistency generalizes to multiple periods of this form.*

3 Finite Sample Performance

Under the null hypothesis of no structural change we consider the same data generating process as Wagner and Hong (2016, Section 3), i.e., we consider a quadratic cointegrating polynomial regression model:

$$y_t = \theta_{D_0} + \theta_{D_1} t + \theta_{X_{1,1}} x_t + \theta_{X_{1,2}} x_t^2 + u_t, \tag{43}$$

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

$$u_t = \rho_1 u_{t-1} + e_{1,t} + \rho_2 e_{2,t}, \quad u_0 = 0,$$
$$v_t = e_{2,t} + 0.5 e_{2,t-1}, \tag{44}$$

where $(e_{1,t}, e_{2,t})' \sim \mathcal{N}(0, I_2)$. The parameter ρ_1 controls the extent of serial correlation in $\{u_t\}_{t \in \mathbb{Z}}$ and is set to $\rho_1 = 1$ after $\lfloor rT \rfloor$ under the alternative of I(1) errors, whereas ρ_2 controls the extent of regressor endogeneity. The parameter values are $\theta_{D_0} = \theta_{D_1} = 1$, $\theta_{X_{1,1}} = 5$ and $\theta_{X_{1,2}} = -0.3$, with the values for $\theta_{X_{1,1}}$ and $\theta_{X_{1,2}}$ inspired by the FM-OLS EKC coefficient estimates for Austria (see Wagner, 2015).[21]

We present simulation results for $T \in \{200, 500\}$ and $\rho_1 = \rho_2 \in \{0, 0.3, 0.6, 0.9\}$. For the moving window and self-normalized moving window detectors we use window sizes $n \in \{0.1, 0.2, 0.3\}$. As indicated in the previous section, long-run covariance estimation is performed with the Bartlett kernel with bandwidth chosen according to Newey and West (1994). For D-OLS estimation leads and lags choices are performed using the AIC-type criterion of Choi and Kurozumi (2012). The number of replications is 10,000 throughout. All monitoring decisions are performed at the nominal 5% significance level.

We start the analysis by considering empirical null rejection probabilities. In doing so, we vary the calibration fraction over a grid of 81 values in the range $m = 0.1, 0.11, \ldots, 0.9$. Two main observations that allow to zoom in subsequently mostly on the self-normalized detectors emerge. Figure 1 clearly shows that with respect to null rejection probabilities, the detectors are separated in two groups, with the better performance offered by self-normalization. Figure 2 shows that when using moving window detectors, the choice of the window size has no visible impact on null

---

[21]The results are, of course, invariant with respect to the values chosen for the parameters θ_{D_0}, θ_{D_1}, $\theta_{X_{1,1}}$ and $\theta_{X_{1,2}}$.
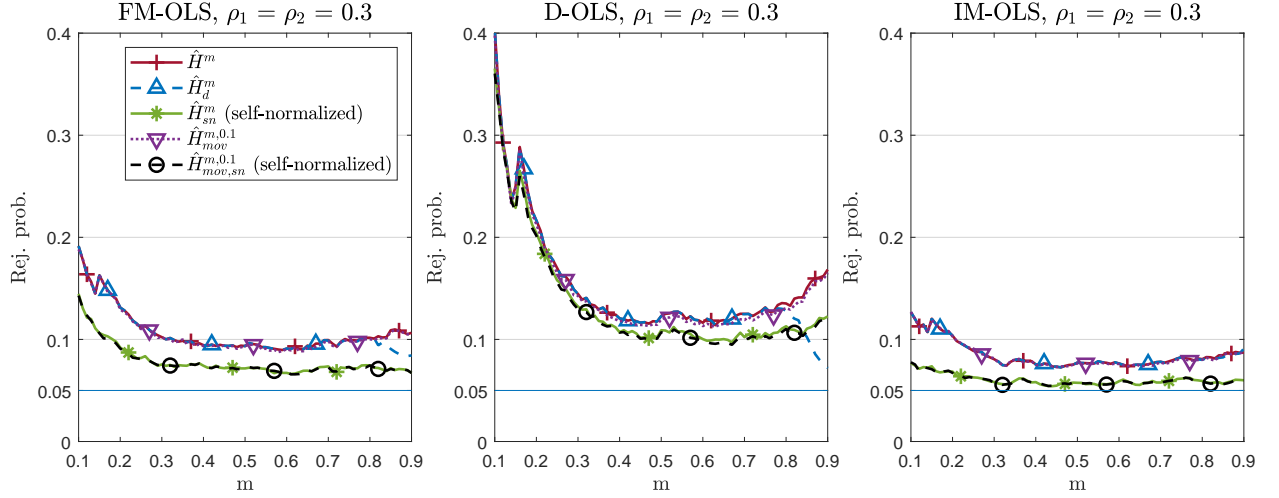
Figure 1: Empirical null rejection probabilities for a grid of values of m, with $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

rejection probabilities.[22]

To assess the performance improvement that can be realized by self-normalization in some detail, Figures 3 and 4 compare $\hat{H}_{mov}^{m,0.1}$ and $\hat{H}_{mov,sn}^{m,0.1}$ for $T = 200$ and $T = 500$, respectively. These two figures illustrate also other more generally observed patterns: First, using IM-OLS for parameter estimation leads to the smallest over-rejections under the null hypothesis. Often, and in particular for small values of m, D-OLS estimation leads to the largest over-rejections. The differences across estimation methods widen for increasing ρ_1, ρ_2, with D-OLS most strongly negatively affected. Also, with the exception of D-OLS, self-normalization attenuates the detrimental impact of increasing ρ_1, ρ_2. An increasing sample size, of course, reduces over-rejections by and large.

We turn to *size-corrected* power.[23] For brevity the main text focuses on size-corrected power against I(1) breaks, i.e., the situation where $\{u_t\}_{t \in \mathbb{Z}}$ changes its behavior from I(0) to I(1) after $\lfloor rT \rfloor$. The other two breaks dealt with in Proposition 3, trend and slope breaks, with changes in θ_D or θ_X, respectively, are discussed in Supplementary Appendix C.[24] Size-corrected power simulations

---

[22]The null rejection probability differences between the standardized and self-normalized detectors increase with increasing ρ_1, ρ_2. The null rejection probability results for $T = 500$ are contained in Figures 8 and 9 in Supplementary Appendix C. As expected, over-rejections are smaller than for $T = 200$, especially for small values of m, and also the differences between the detectors decrease.

[23]We focus on size-corrected power because of the potential over-rejection problems under the null hypothesis. This allows us to see power differences across detectors while holding null rejection probabilities constant at 0.05. Clearly, this is useful for theoretical power comparisons, but it has to be kept in mind that such size-corrections are not feasible in practice.

[24]See Figures 10 to 18 in Supplementary Appendix C for size-corrected power results in case of trend and slope breaks.
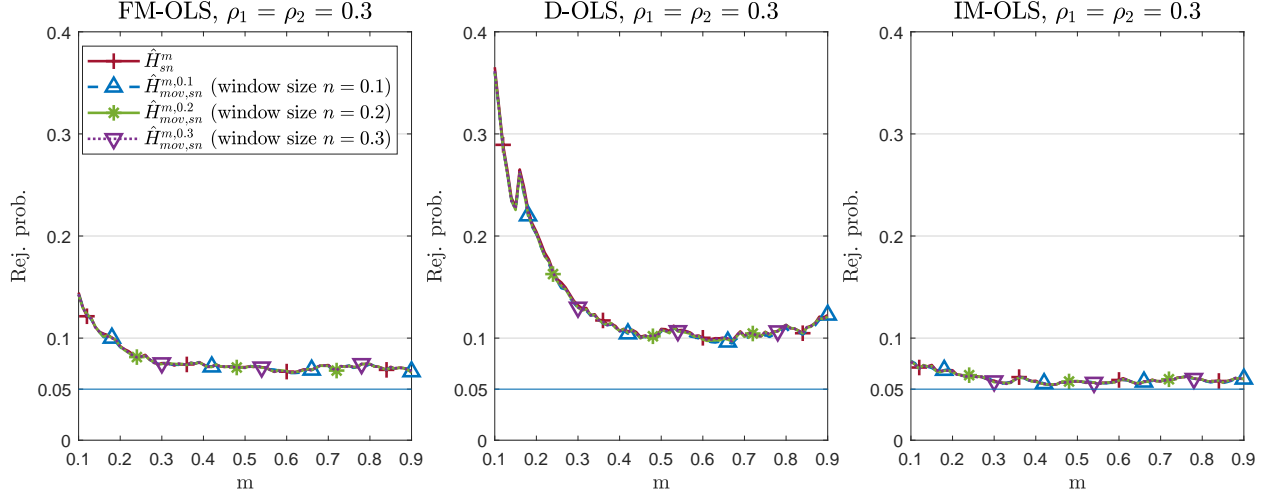
Figure 2: Empirical null rejection probabilities for a grid of values of m, with $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

Figure 3: Empirical null rejection probabilities for a grid of values of m and $T = 200$. The left panel displays results for $\hat{H}_{mov}^{m,0.1}$. The right panel displays results for $\hat{H}_{mov,sn}^{m,0.1}$.

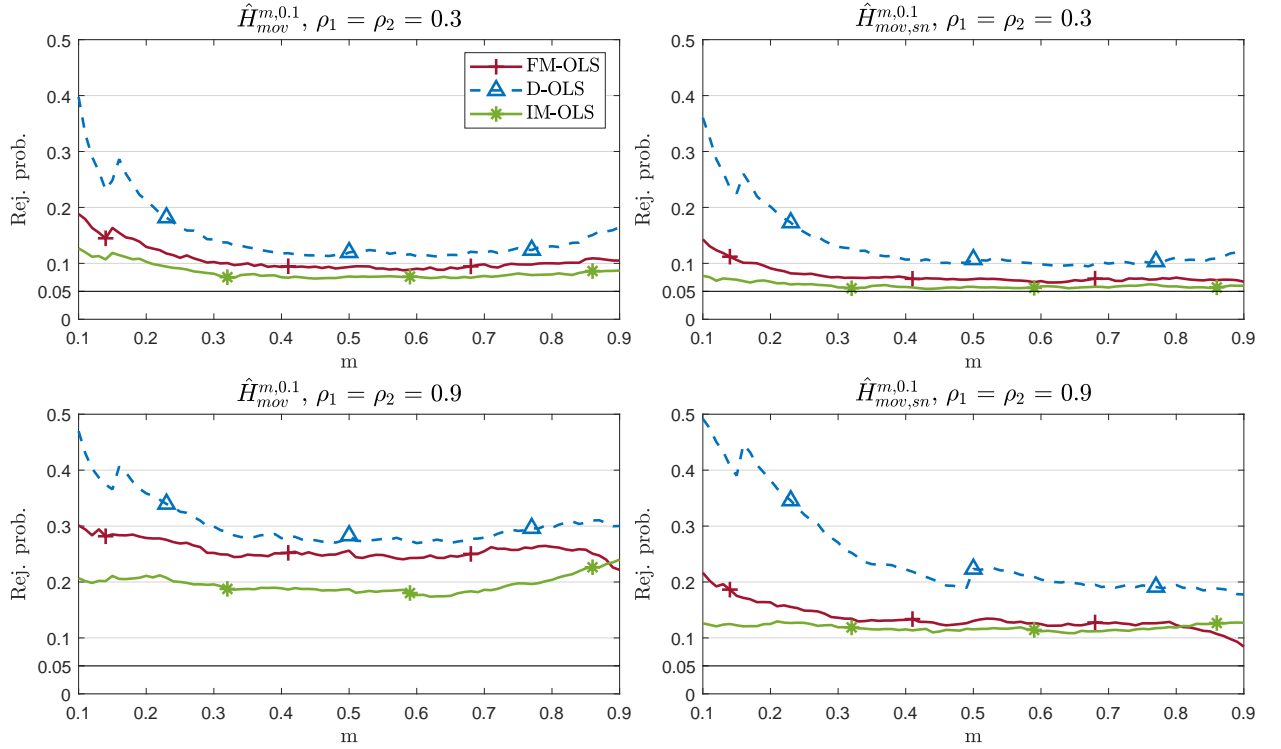Figure 4: Empirical null rejection probabilities for a grid of values of m and $T = 500$. The left panel displays results for $\hat{H}_{mov}^{m,0.1}$. The right panel displays results for $\hat{H}_{mov,sn}^{m,0.1}$.
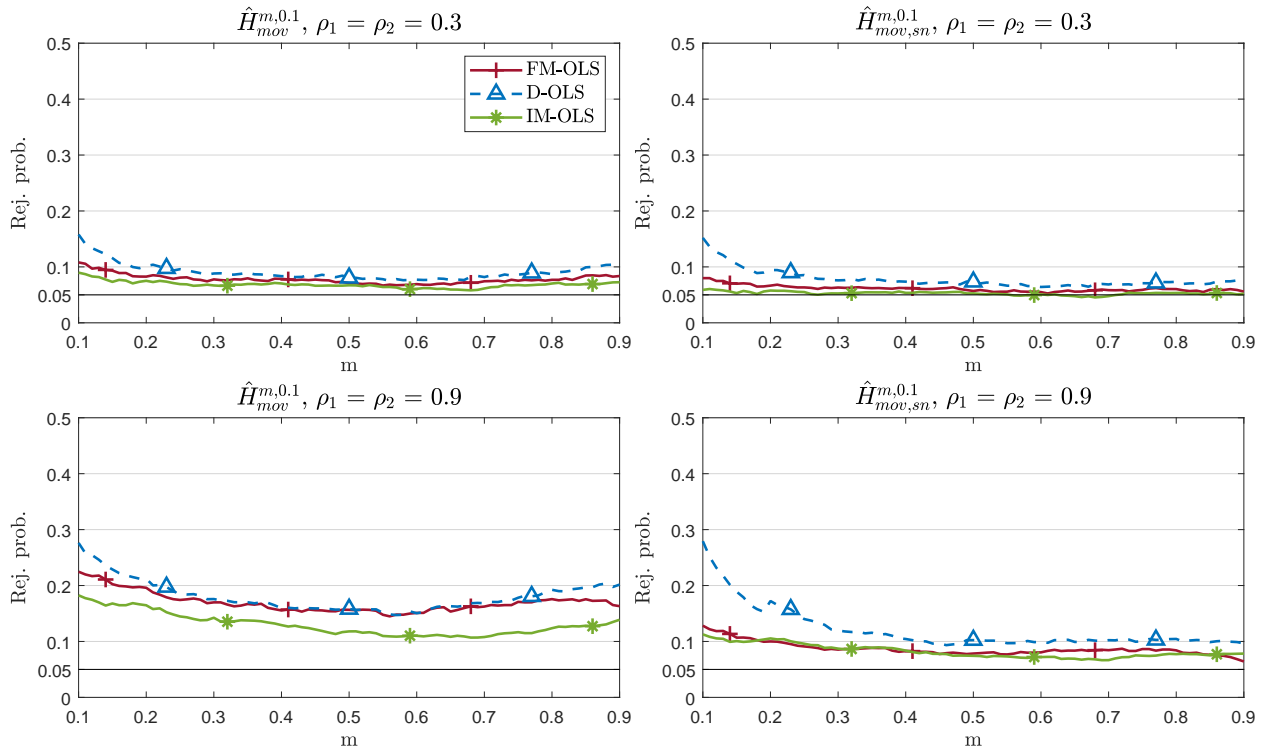
are performed for $m, r \in \{0.25, 0.5, 0.75\}$, which includes, therefore, cases where $r < m$, i.e., where a break occurs in the calibration period. We report results for all nine detectors considered, for all three estimation methods and $\rho_1 = \rho_2 = 0.3$ until $\lfloor rT \rfloor$ in Table 1 for $T = 200$ and Table 2 for $T = 500$.[25] The following observations emerge: *Grosso modo* highest size-corrected power is achieved by the moving window detector with $n = 0.1$, either with or without self-normalization; with as reported above self-normalization leading to smaller over-rejections under the null hypothesis. The differences in size-corrected power are typically relatively or even very small. This, in conjunction with the performance improvement of self-normalization under the null hypothesis, makes the case in favor of self-normalization clear. In line with standard asymptotic theory concerning estimator efficiency, size-corrected power is typically lower for IM-OLS than for FM-/D-OLS, with the difference between the latter two often rather small, also in line with asymptotic theory.[26] These results, of course, have to be seen in conjunction with the smaller over-rejections of IM-OLS under the null hypothesis.

Next, consider the impact of m and r on size-corrected power. In case $r < m$, i.e., when a break occurs in the calibration period, size-corrected power is often low, which is related to inconsistency of parameter estimation due to the structural break in the calibration period. It turns out that in this case the self-normalized detectors lead to lower size-corrected power than the standardized detectors, see Tables 1 ($T = 200$) and 2 ($T = 500$). For $m \leq r$ size-corrected power behavior is as expected, i.e., for fixed m size-corrected power decreases with increasing r and for fixed r size-corrected power increases with increasing calibration period m. The case $m = r$ leads to the highest size-corrected power. Also, as expected increasing T leads to higher size-corrected power, whereas increasing ρ_1, ρ_2 leads to lower size-corrected power. For trend breaks – investigated in some more detail in Supplementary Appendix C – many observations are qualitatively similar. One difference is that in case of trend breaks the detectors $\hat{H}^m$ or $\hat{H}_d^m$ lead to highest size-corrected power in some configurations. This finding, however, has to be considered in light of the larger over-rejections exhibited by these two detectors under the null. Slope breaks lead to very similar results as I(1) breaks.

It remains to investigate the detection times and delays. We consider again the case of I(1) breaks

---

[25]Tables 1 to 6 in Supplementary Appendix C display corresponding size-corrected power results for $\rho_1 = \rho_2 \in \{0, 0.6, 0.9\}$.

[26]The main motivation for developing IM-OLS in Vogelsang and Wagner (2014a) was to develop an estimator that allows to perform fixed-b inference, which is an alternative asymptotic theory that captures the impact of kernel and bandwidth choices. These aspects are not covered by standard asymptotic theory.

$\rho_1 = \rho_2 = 0.3$			$m = 0.25$			$m = 0.5$			$m = 0.75$	
	r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\hat{H}^m$	FM	0.19	0.07	0.05	0.32	0.64	0.14	0.34	0.46	0.80
	D	0.13	0.06	0.05	0.30	0.59	0.11	0.34	0.46	0.76
	IM	0.09	0.06	0.05	0.26	0.47	0.07	0.32	0.41	0.66
$\hat{H}_d^m$	FM	0.19	0.07	0.05	0.32	0.64	0.14	0.33	0.45	0.79
	D	0.13	0.06	0.05	0.30	0.59	0.11	0.33	0.45	0.76
	IM	0.09	0.06	0.05	0.26	0.47	0.07	0.32	0.41	0.66
$\hat{H}_{sn}^m$	FM	0.20	0.07	0.05	0.17	0.61	0.14	0.16	0.26	0.77
	D	0.14	0.06	0.05	0.20	0.56	0.11	0.19	0.30	0.73
	IM	0.11	0.05	0.05	0.19	0.49	0.08	0.21	0.31	0.66
$\hat{H}_{mov}^{m,0.1}$	FM	0.18	0.07	0.05	0.32	0.66	0.17	0.34	0.46	0.80
	D	0.12	0.06	0.05	0.29	0.61	0.13	0.34	0.46	0.77
	IM	0.09	0.05	0.05	0.25	0.48	0.09	0.31	0.40	0.67
$\hat{H}_{mov}^{m,0.2}$	FM	0.18	0.07	0.05	0.32	0.65	0.15	0.34	0.46	0.80
	D	0.13	0.06	0.05	0.30	0.60	0.12	0.34	0.46	0.76
	IM	0.09	0.06	0.05	0.26	0.48	0.08	0.32	0.41	0.66
$\hat{H}_{mov}^{m,0.3}$	FM	0.18	0.07	0.05	0.32	0.64	0.14	0.34	0.46	0.80
	D	0.13	0.06	0.05	0.30	0.59	0.11	0.34	0.46	0.76
	IM	0.09	0.06	0.05	0.26	0.47	0.08	0.32	0.41	0.66
$\hat{H}_{mov,sn}^{m,0.1}$	FM	0.20	0.07	0.05	0.17	0.63	0.18	0.16	0.26	0.77
	D	0.14	0.06	0.05	0.20	0.58	0.13	0.20	0.30	0.74
	IM	0.10	0.05	0.05	0.18	0.49	0.10	0.20	0.30	0.66
$\hat{H}_{mov,sn}^{m,0.2}$	FM	0.20	0.07	0.05	0.17	0.62	0.15	0.16	0.26	0.77
	D	0.14	0.06	0.05	0.20	0.57	0.11	0.19	0.30	0.73
	IM	0.10	0.05	0.05	0.19	0.49	0.09	0.21	0.31	0.66
$\hat{H}_{mov,sn}^{m,0.3}$	FM	0.20	0.07	0.05	0.17	0.62	0.14	0.16	0.26	0.77
	D	0.14	0.06	0.05	0.20	0.56	0.11	0.19	0.30	0.73
	IM	0.10	0.05	0.05	0.19	0.49	0.08	0.21	0.31	0.66

Table 1: Size-corrected power against I(1) breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

$\rho_1 = \rho_2 = 0.3$		$m = 0.25$			$m = 0.5$			$m = 0.75$		
	r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\hat{H}^m$	FM	0.67	0.24	0.06	0.44	0.93	0.46	0.45	0.57	0.97
	D	0.62	0.21	0.06	0.42	0.92	0.42	0.46	0.58	0.96
	IM	0.42	0.11	0.05	0.34	0.82	0.25	0.43	0.52	0.91
$\hat{H}^m_d$	FM	0.67	0.24	0.06	0.44	0.93	0.46	0.45	0.58	0.97
	D	0.62	0.21	0.06	0.42	0.92	0.42	0.45	0.57	0.96
	IM	0.42	0.11	0.05	0.34	0.82	0.25	0.43	0.52	0.91
$\hat{H}^m_{sn}$	FM	0.65	0.25	0.06	0.21	0.91	0.44	0.18	0.30	0.96
	D	0.59	0.20	0.06	0.21	0.89	0.40	0.20	0.32	0.95
	IM	0.43	0.12	0.05	0.23	0.83	0.28	0.24	0.36	0.91
$\hat{H}^{m,0.1}_{mov}$	FM	0.70	0.25	0.06	0.44	0.94	0.53	0.45	0.57	0.97
	D	0.66	0.22	0.06	0.42	0.93	0.49	0.47	0.58	0.96
	IM	0.43	0.11	0.05	0.33	0.84	0.31	0.43	0.52	0.91
$\hat{H}^{m,0.2}_{mov}$	FM	0.68	0.25	0.06	0.44	0.93	0.48	0.45	0.57	0.97
	D	0.64	0.22	0.06	0.42	0.92	0.44	0.46	0.58	0.96
	IM	0.42	0.11	0.05	0.34	0.83	0.27	0.43	0.52	0.91
$\hat{H}^{m,0.3}_{mov}$	FM	0.67	0.25	0.06	0.44	0.93	0.47	0.45	0.57	0.97
	D	0.63	0.22	0.06	0.42	0.92	0.42	0.46	0.58	0.96
	IM	0.42	0.11	0.05	0.34	0.82	0.26	0.43	0.52	0.91
$\hat{H}^{m,0.1}_{mov,sn}$	FM	0.68	0.26	0.06	0.21	0.93	0.51	0.18	0.30	0.96
	D	0.63	0.21	0.06	0.21	0.91	0.47	0.20	0.32	0.95
	IM	0.45	0.12	0.05	0.22	0.85	0.34	0.24	0.36	0.91
$\hat{H}^{m,0.2}_{mov,sn}$	FM	0.66	0.26	0.06	0.21	0.92	0.47	0.18	0.30	0.96
	D	0.60	0.20	0.06	0.21	0.90	0.42	0.20	0.32	0.95
	IM	0.44	0.12	0.05	0.23	0.83	0.30	0.24	0.36	0.91
$\hat{H}^{m,0.3}_{mov,sn}$	FM	0.65	0.26	0.06	0.21	0.91	0.45	0.18	0.30	0.96
	D	0.60	0.20	0.06	0.21	0.89	0.40	0.20	0.32	0.95
	IM	0.43	0.12	0.05	0.23	0.83	0.28	0.24	0.36	0.91

Table 2: Size-corrected power against I(1) breaks for $T = 500$ and $\rho_1 = \rho_2 = 0.3$.

in the main text; for $T = 200$ and $T = 500$ for $\rho_1 = \rho_2 = 0.3$ in Figure 5. With respect to m and r we only display results for $m = 0.5$ and $r = 0.5$; additional configurations leading to qualitatively very similar results are displayed in Figures 19 – 24 in Supplementary Appendix C. The figures display the results in the form of box-whiskers plots – for completeness – for all nine considered detectors. The numbers below the detector label indicates the corresponding null rejection probabilities, i.e., size-corrected power, given in Tables 1 and 2. Thus, the different box-whiskers plots are based on different *numbers of replications* because of different numbers of rejections across different detectors, sample sizes and ρ values.

One relatively clear observation that emerges from the figures, both in the main text and in Supplementary Appendix C, is that the choice of the estimation method does not exhibit major impacts on detection times and delays. Given the standard asymptotic properties of the estimators, as expected IM-OLS leads to slightly larger delays than FM-OLS and D-OLS in many cases. The moving window detectors lead to the shortest delays, both standardized and self-normalized, with the best performance achieved with $n = 0.1$. One interesting observation is that for the delay, the choice of the window size does matter and in fact exerts bigger influence on the results than the choice of standardizing or self-normalizing. Furthermore, an increasing sample size leads to a – ceteris paribus – more concentrated distribution of the estimated detection times (based on a larger number of observations), but does not throughout lead to smaller average delays.

The results of the finite sample simulations can essentially be summarized as follows: First, self-normalization leads to smaller over-rejections, without having any systematic or sizeable negative impact on size-corrected power (for $m \leq r$); and without larger delays when compared to standardization. Second, when considering only moving window detectors – either self-normalized or standardized – choosing $n = 0.1$ leads almost throughout to highest size-corrected power without detrimental effects on null rejection probabilities; with some exceptions in case of trend breaks where $n = 0.3$ leads to marginally better size-corrected power performance. The window size does have some impact on the delays. With respect to estimator choice the usual trade-off between IM-OLS on the one hand and FM-/D-OLS on the other occurs, with D-OLS outperformed by FM-OLS in terms of larger over-rejections by D-OLS under the null hypothesis. Self-normalization in conjunction with IM-OLS is particularly beneficial for small samples, as in this case no long-run covariance estimates are required. The poor performance of FM-OLS compared to IM-OLS for small samples can be traced back to long-run covariance estimation required for FM-OLS parameter estimation, even when using self-normalized detectors. For larger sample sizes, with better
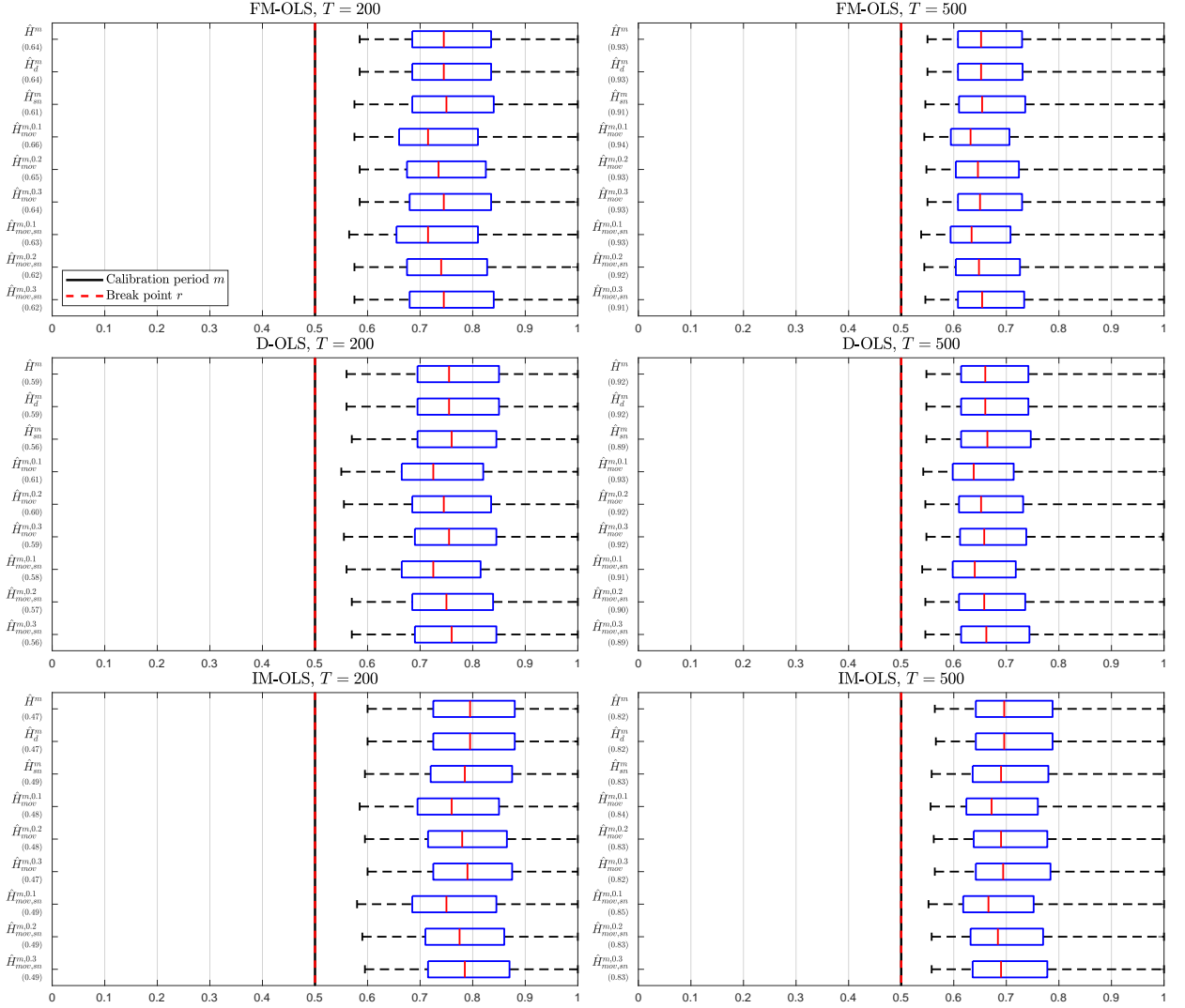
Figure 5: Detection times against $I(1)$ breaks for $m = 0.5$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.3$.

properties of long-run covariance estimation the asymptotic efficiency advantage of FM-OLS over IM-OLS becomes visible, in particular with respect to higher size-corrected power.

Second, with respect to the window size, $n = 0.1$ leads to the highest size-corrected power against I(1) and slope breaks, and for some combinations of m and r also against trend breaks. At the same time, choosing $n = 0.1$ has no negative effect with respect to over-rejections and leads in addition to the shortest delays. Therefore we recommend to use the self-normalized moving window detector with $n = 0.1$, i.e., $\hat{H}_{mov,sn}^{m,0.1}$. The choice of the estimator is dictated by the trade-off between null rejection probabilities on target and high size-corrected power. IM-OLS leads to lower over-rejections at the cost of lower size-corrected power and slightly larger delays than FM-OLS. Furthermore, FM-OLS outperforms D-OLS, since size-corrected power and detection times based on D-OLS are by and large very similar compared to those based on FM-OLS, but the null rejection probabilities are worse.

4   The Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions

We now apply the developed monitoring procedures to EKCs for carbon and sulfur dioxide (CO_2 and SO_2) emissions. We commence from a sample of 18 industrialized countries, listed in Table 3, over the period 1946–2016.[27] The highest polynomial degree we consider is the cubic cointegrating polynomial relationship, i.e.:

$$y_t = \theta_{D_0} + \theta_{D_1} t + \theta_{X_{1,1}} x_t + \theta_{X_{1,2}} x_t^2 + \theta_{X_{1,3}} x_t^3 + u_t, \tag{45}$$

with x_t the logarithm of real per capita GDP and y_t the logarithm of per capita CO_2 or SO_2 emissions. As will be seen below, over the calibration period 1946–1973 for many countries in fact a cointegrating quadratic or even cointegrating linear relationship prevails, i.e., is not rejected.

The GDP and CO_2 emissions data are similar to those used in Wagner *et al.* (2020). The main difference is the extension of the sample period from 2013 to 2016, additionally there are some marginal changes in the GDP data in the newer vintage. More precisely, the GDP (in 2011 prices) and population data stem from the Maddison project database (in the 2018 version of Bolt *et al.*, 2018). The CO_2 emissions data – which cover CO_2 emissions from fossil fuel usage – are taken

---

[27]Only for twelve of these 18 countries, as discussed below, is the unit root null hypothesis not rejected for the logarithm of real per capita GDP. This, of course, reduces the number of countries for which monitoring is performed in this section to twelve countries.

Australia	**Austria**	Belgium	Canada	Denmark	Finland
France	**Germany**	*Italy*	Japan	**New Zealand**	**Norway**
Portugal	Spain	Sweden	**Switzerland**	United Kingdom	United States

Table 3: List of countries included in the empirical analysis. The sample range is 1946–2016. *Italic* country names indicate that the augmented Dickey-Fuller test rejects the unit root null hypothesis for log GDP per capita on the calibration period (1946–1973) at the 10% level and **bold** country names indicate rejections at the 5% level. Intercept and linear trend are included.

from the Carbon Dioxide Information and Analysis Center (CDIAC) of the US Department of Energy, see Boden *et al.* (2018). The SO_2 emissions have been combined from two sources. The data for 1946–2005 are from the NASA Socioeconomic Data and Applications Center (SEDAC), see Smith *et al.* (2011). The data for the period 2006–2016 are from the OECD (2020).[28]

With the developed methods resting upon a calibration period, a corresponding choice has to be made. We choose the calibration period 1946–1973, reflecting that the first oil price shock of 1974 is considered a major event for changes in energy consumption patterns.[29] In our notation this amounts to $m = 28/71 \approx 0.4$. Given this choice, the first step is to perform the CPR modelling cycle for the calibration period.

The augmented Dickey and Fuller (1981) test results for the calibration period for the null hypothesis of a unit root in log real per capita GDP against the alternative of trend stationarity with a linear trend are contained in the country list Table 3.[30] These results lead to the following twelve countries being included in the subsequent analysis (based on a 5% significance level): Australia, Belgium, Canada, Denmark, Finland, Italy, Japan, Portugal, Spain, Sweden, the United

---

[28]Note that the combination of these two data sources using growth rates rests upon the assumption that the share of SO_2 in SO_x is constant at about 98% also over the period 2006 onwards, as the OECD data comprise all SO_x emissions and not only SO_2 emissions.

[29]In addition and preceding the oil price shock, many countries have put more stringent environmental legislation in place in the late 1960s or early 1970s, e. g., the United States introduced Clean Air Acts in 1963 and 1970, Canada introduced a similarly named law 1971 and Sweden introduced its Environmental Protection Act in 1969.

[30]The detailed unit root test results using both the augmented Dickey-Fuller and the Phillips and Perron (1988) tests are contained in Supplementary Appendix D in Table 7 for the calibration period and in Table 8 for the full sample period. With the exception of Germany and the augmented Dickey-Fuller test and Austria, Germany and New Zealand and the Phillips-Perron test, the unit root null hypothesis is not rejected over the full sample period. Thus, for the full sample period the evidence for I(1) behavior of log real per capita GDP is, as expected, much stronger. We could, in principle, also consider a larger set of countries in the subsequent analysis, based on the probably more precise unit root test results obtained from a longer period.

Zooming in a bit more by using a modified Phillips-Perron test of Perron and Vogelsang (1993) leads, e. g., for Austria to a non-rejection of the unit root null hypothesis when allowing for breaks in the intercept and trend slopes. Investigating such issues further, i. e., allowing for breaks in the regressors is despite its importance beyond the scope of the present paper.

| | Polynomial degree | |
Country	CO_2	SO_2
Australia	1	1
Belgium	1	1
Canada	1	1
Denmark	1	2
Finland	2	2
Italy	1	1
Japan	1	2
Portugal	1	1
Spain	1	1
Sweden	1	2
United Kingdom	1	1
United States	2	3

Table 4: Minimal polynomial degrees for cointegrating EKCs over the calibration period 1946–1973.

Kingdom (UK) and the United States (US).

The next step is to test for the prevalence of a CPR relationship over the monitoring period – for both CO_2 and SO_2 emissions for the countries with I(1) log real per capita GDP. Given that the polynomial degree is ex ante unclear, we perform a testing sequence with polynomial degrees ranging from three (the cubic specification) to one (linear cointegration).[31] The deterministic specification consists of intercept and linear trend throughout. The lowest degree polynomial for which a CPR relationship is not rejected is considered as starting point for monitoring in the following subsections, see Table 4.

The most striking feature of Table 4 is that in most countries – for CO_2 even more than for SO_2 – a cointegrating linear relationship appears to be present over the period 1946–1973. This seems to be at odds at first sight with the EKC hypothesis of an inverted U-shaped relationship. However, these results reflect the fact that until the early 1970s per capita GDP and per capita emissions developed very similarly on a "log-linear extension path", see the scatter plots in Figures 25 and 26 in Supplementary Appendix D. Only starting in the mid 1970s the oil price shock as well as environmental legislation – in particular with respect to SO_2 and "acid rain" – lead to a *decoupling*

---

[31]For linear cointegration we consider also the Johansen (1995) test in addition to the Shin (1994) test. For the higher order polynomial degrees we use the extension discussed in Wagner and Hong (2016) or Wagner (2020). The detailed test results are available in Tables 9 and 10 in Supplementary Appendix D.

Country	p	$\hat{H}_{mov,sn}^{m,0.1}$ FM-OLS	IM-OLS
Australia	1	1993	2001
Belgium	1	1988	1992
Canada	1	∞	∞
Denmark	1	1991	2011
Finland	2	1989	1990
Italy	1	1981	1982
Japan	1	1982	1980
Portugal	1	1998	∞
Spain	1	∞	∞
Sweden	1	1982	1983
United Kingdom	1	1984	1987
United States	2	1988	1992

Table 5: Detection times when monitoring CO_2 emissions using the detector $\hat{H}_{mov,sn}^{m,0.1}$ and both, FM-OLS and IM-OLS. The column p indicates the polynomial degree, the calibration period is 1946–1973, the monitoring period is 1974–2016. The nominal significance level is 5%.

of the two quantities, to a certain extent, in many countries.[32]

Given these preliminary results we turn in the following two subsections to monitoring, using the moving window detector $\hat{H}_{mov,sn}^{m,0.1}$ with window size $n = 0.1$, i. e. , seven years.[33]

4.1 Results for Carbon Dioxide Emissions

The detection times for the twelve considered countries using, with the exceptions of Finland and the United States where the quadratic specification is estimated, the linear specification are displayed in Table 5.

With the exception of Canada and Spain for all countries the linear relationship found over the calibration period appears to break down. Furthermore, Portugal is a "mixed" case, with unsurprisingly similar behavior to Spain in many ways, as discussed below. In line with the simulation

---

[32]It is probably a *philosophical* question whether we observe in these cases a "straight-looking" line segment of an actually inverted U-type relationship or "really" a linear relationship. From an econometric perspective the evidence is in favor of a linear relationship in a number of countries and it thus is an interesting question to detect and date structural breaks based on monitoring a cointegrating linear relationship. We could alternatively also monitor for these countries, e. g. , over-specified cubic relationships, where we would be also bound to find structural change for those countries where the data after the calibration period are changing from a linear towards an inverted U-shaped CPR relationship with a polynomial degree larger than one.

[33]The full sets of results including detection times for all detectors and estimators are available in Supplementary Appendix D, in Table 14 for CO_2 emissions and in Table 15 for SO_2 emissions. Given the performance advantages of FM-OLS over D-OLS we exclude the D-OLS results from the discussion in the main text and discuss only FM-OLS and IM-OLS results here.
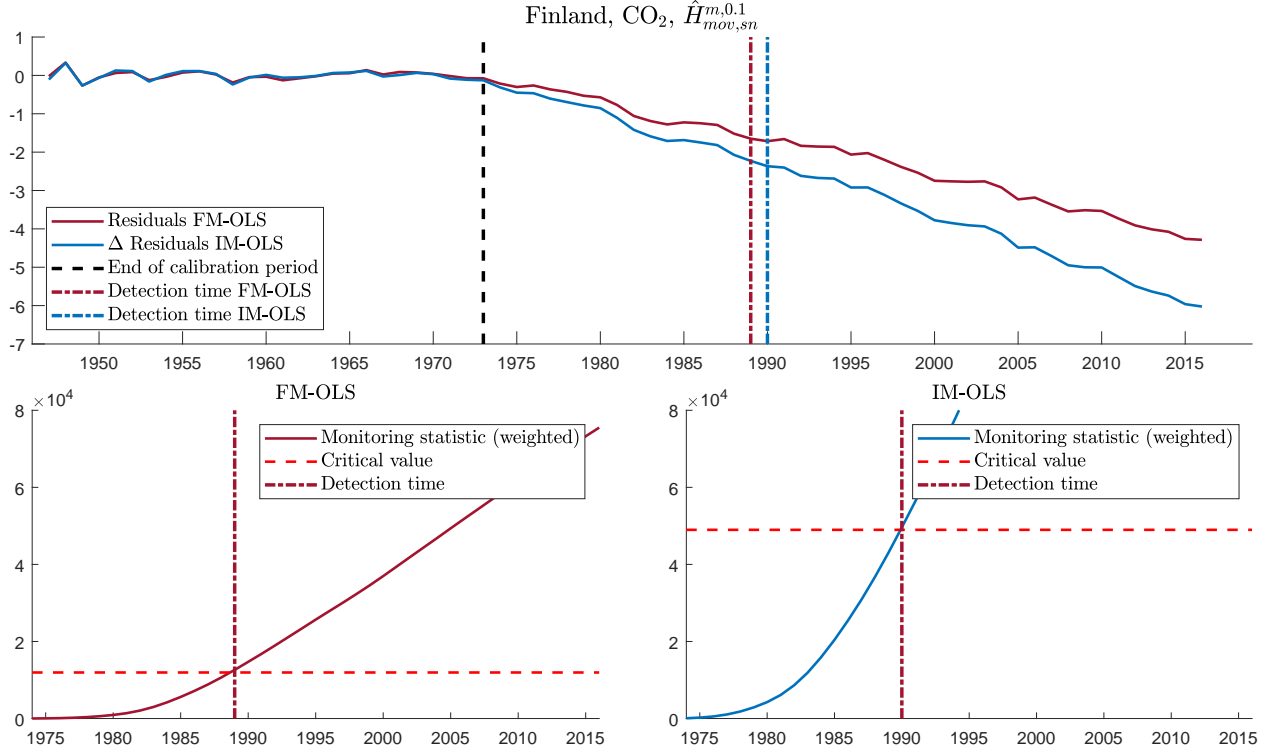
Figure 6: Monitoring results for CO_2 emissions of Finland using $\hat{H}_{mov,sn}^{m,0.1}$ with both, FM-OLS and IM-OLS, in the quadratic specification. The lower panel shows the detectors, the critical values and the detection times for FM-OLS on the left hand side and for IM-OLS on the right hand side.

results, the detection times are – with the exception of Japan – earlier with FM-OLS than with IM-OLS, with the differences often just a few years. Mostly, breaks are dated in the 1980s or early 1990s, with a few exceptions: Australia (IM-OLS: 2001), Denmark (IM-OLS: 2011), Portugal (FM-OLS: 1998; IM-OLS: no break, which is the only difference in finding a break point or not between the two methods). This can be interpreted, admittedly unsurprisingly, as strong evidence against a continued log-linear co-movement, i.e., aligned expansion, of log per capita GDP and CO_2 emissions from the – given the delays seen in the simulation section *at latest* – early 1990s onwards. Or, the other way round, these findings are indicative of *curvature picking up* in CO_2 EKC-type relationships in the 1980s in most of the considered countries, with the exception of Canada, potentially Portugal and Spain, discussed further below.[34]

---

[34]Considering a quadratic specification over the full sample period, see Table 16 in Supplementary Appendix D for details, confirms this. For all countries except Canada, Portugal and Spain, the coefficients to log per capita GDP squared are (significantly) negative, whereas they show a positive sign for these three countries. The coefficient is, however, only significantly positive for Portugal, underlining the "borderline case" behavior of Portugal.
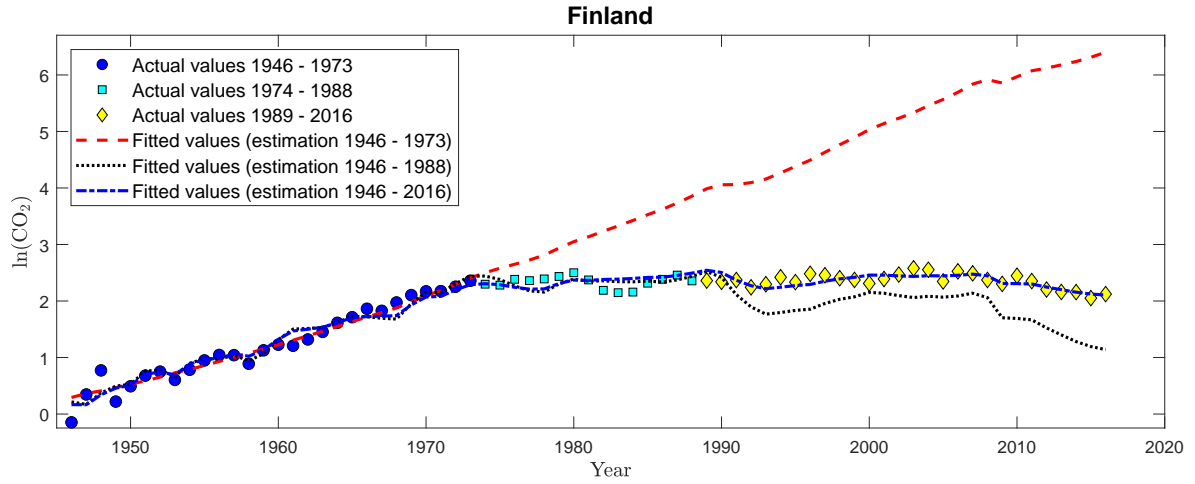
Finland

Figure 7: FM-OLS estimation results for a cointegrating quadratic relationship between log per capita GDP and log per capita CO_2 emissions for Finland. The figure shows pairs of observations of log per capita GDP and log per capita CO_2 emissions, for 1946–1973 (blue circles), 1974–1988 (turquoise squares) and 1989–2016 (yellow diamonds). The lines display fitted values over time obtained using different samples for parameter estimation: the dashed red line 1946–1973, the dotted black line 1946–1988 and the dash-dotted blue 1946–2016.

As an illustration, Figure 6 displays the mechanics of the monitoring procedure for Finland.[35] The upper graph displays the FM- and IM-OLS residuals over both the calibration and the monitoring period. By definition, since an intercept is included, the residuals have zero mean over the calibration period, and then turn systematically (more and more) negative thereafter on the monitoring period. This, by definition, means that the estimated – in case of Finland quadratic – relationship systematically *over-predicts* actual log per capita CO_2 emissions, i.e., the slope of an actual relationship between output and emissions is estimated as too high, see also Figure 7. The statistical monitoring procedures need to *collect enough signals* until a break is detected, in this example this takes until 1989 for FM-OLS and until 1990 for IM-OLS, which is clearly too late. Nevertheless, it may be interesting to note that Finland was severely adversely affected by the collapse of the Soviet Union in the early 1990s, which in fact for Finland led to a deeper recession than the Great Depression in the 1930s.

Figure 7 shows the fitted values obtained when estimating the quadratic EKC with FM-OLS over the calibration sample, the sample ending prior to the detected break-point 1988 and when

---

[35]Figures 30 to 40 in Supplementary Appendix D display monitoring results for CO_2 emissions for the other eleven countries in exactly the same format as Figure 6 for Finland.

	1946–1973		1946–2016	
	θ_{D_1}	$\theta_{X_{1,1}}$	θ_{D_1}	$\theta_{X_{1,1}}$
Canada				
FM-OLS	**-0.056**	**2.841**	**-0.027**	**1.661**
IM-OLS	**-0.059**	**2.990**	**-0.028**	**1.714**
Portugal				
FM-OLS	0.000	**1.003**	-0.008	**1.341**
IM-OLS	0.008	**0.869**	0.001	**1.156**
Spain				
FM-OLS	-0.023	**1.519**	**-0.033**	**1.818**
IM-OLS	-0.039	**1.789**	**-0.036**	**1.911**

Table 6: FM-OLS and IM-OLS estimation results for a cointegrating linear relationship between log per capita GDP and log per capita CO_2 emissions for Canada, Portugal and Spain for the calibration period 1946–1973 (left panel) and the full sample period 1946–2016 (right panel). *Italic* entries indicate significance of coefficients at the 10% level and **bold** entries significance of coefficients at the 5% level.

estimation takes place over the full sample.[36] The dashed red line shows that in fact more or less immediately after the oil price shock the relationship between GDP and CO_2 emissions appears to have changed, with the detector taking about 15 years to collect enough signals to declare the null hypothesis rejected at the 5% significance level.[37] Parameter estimation until 1988 "catches the turn" of the mid 1970s and leads to good fit. However, using the full sample for parameter estimation of the quadratic EKC leads to the best fit, in particular for the period after 1989. This leads some support for the interpretation that maybe the relationship is after all quadratic and one needs to have a sample that is (i) larger and (ii) covers the inverted U behavior to be able to estimate the relationship with sufficient precision, compare Footnote 32.[38]

To complete the analysis for CO_2 emissions we now turn to the countries for which no structural change is indicated by the monitoring procedure, Canada, the mixed case Portugal and Spain. Table 6 displays the estimated coefficients for these three countries when the cointegrating (linear) relationships between log per capita GDP and log per capita CO_2 emissions are estimated over the calibration sample and over the whole sample period.

[36]See Figure 27 in Supplementary Appendix D for analogous – and very similar – IM-OLS results.

[37]Clearly, this is a long delay – in particular when looking at Figure 7, but it has to be taken into account that the estimation sample comprises only 28 observations, which is an rather small sample for cointegration analysis.

[38]This appears to be the case indeed. The coefficients to squared log per capita GDP are very small but positive and borderline significant when estimated over the calibration sample and are significantly negative when estimated over the full sample.

The results reported in Table 6 are a bit mixed and no uniform pattern across all three countries emerges, with some of the smoke clearing when considering Figure 8 below. For Canada, the slope coefficient corresponding to log per capita GDP becomes much smaller when estimated over the full sample, but also the trend coefficient becomes less negative. For Portugal the trend coefficient is not significant in both periods, but turns from non-negative point estimates in the calibration period to smaller, and for FM-OLS negative, values in the full period. The slope coefficient, however, is bigger, when estimated over the full sample. For Spain, the trend coefficient does become more negative and significant when estimated over the full sample period and the slope coefficient becomes bigger. Thus, qualitatively the results are similar for the two countries on the Iberian peninsula, which not only share many economic similarities but also political similarities, with the end of the Franco and Salazar regimes in these countries in the mid 1970s.

Probably more informative, Figure 8 displays similar results for these three countries as displayed for Finland in Figure 7.[39] The message is clear, in particular in comparison with Figure 7. In case no structural break is detected, estimation over the larger sample does lead – unsurprisingly – to better fit, but the differences in fit are, compared to the differences observed in Figure 7, minor. From a standard inverted U quadratic EKC perspective the excellent fit obtained for these three countries with the linear specification is to a certain extent surprising – and, which might be bad news, of course there is no turning point.

Note that when considering the full sample period 1946–2016 we find a cointegrating EKC relationship for all nine countries for which monitoring detects a structural break.[40] For about a third of the countries the polynomial degree of the EKC appears to be larger for the full sample than for the calibration period, supporting the notion of more *curvature*. For most countries the evidence, with an unchanged minimal polynomial degree, points, however, more towards structural change of the parameters for a given form of the relationship. For Finland and the US it appears that the polynomial degree is lower over the full sample than for the calibration period, albeit the coefficients to squared log per capita GDP are significant when estimated over the full sample, which leads to an unclear picture. Altogether, the results in this CO_2 subsection indicate that the procedures *work* even on such small samples, albeit of course for many applications the sizeable delay is most certainly problematic.

---

[39]Figure 28 in Supplementary Appendix D shows the corresponding results obtained with IM-OLS.

[40]Table 11 in Supplementary Appendix D contains similar results concerning the minimal polynomial degrees of a CPR relationship for the full sample period as Table 4 for the calibration period.
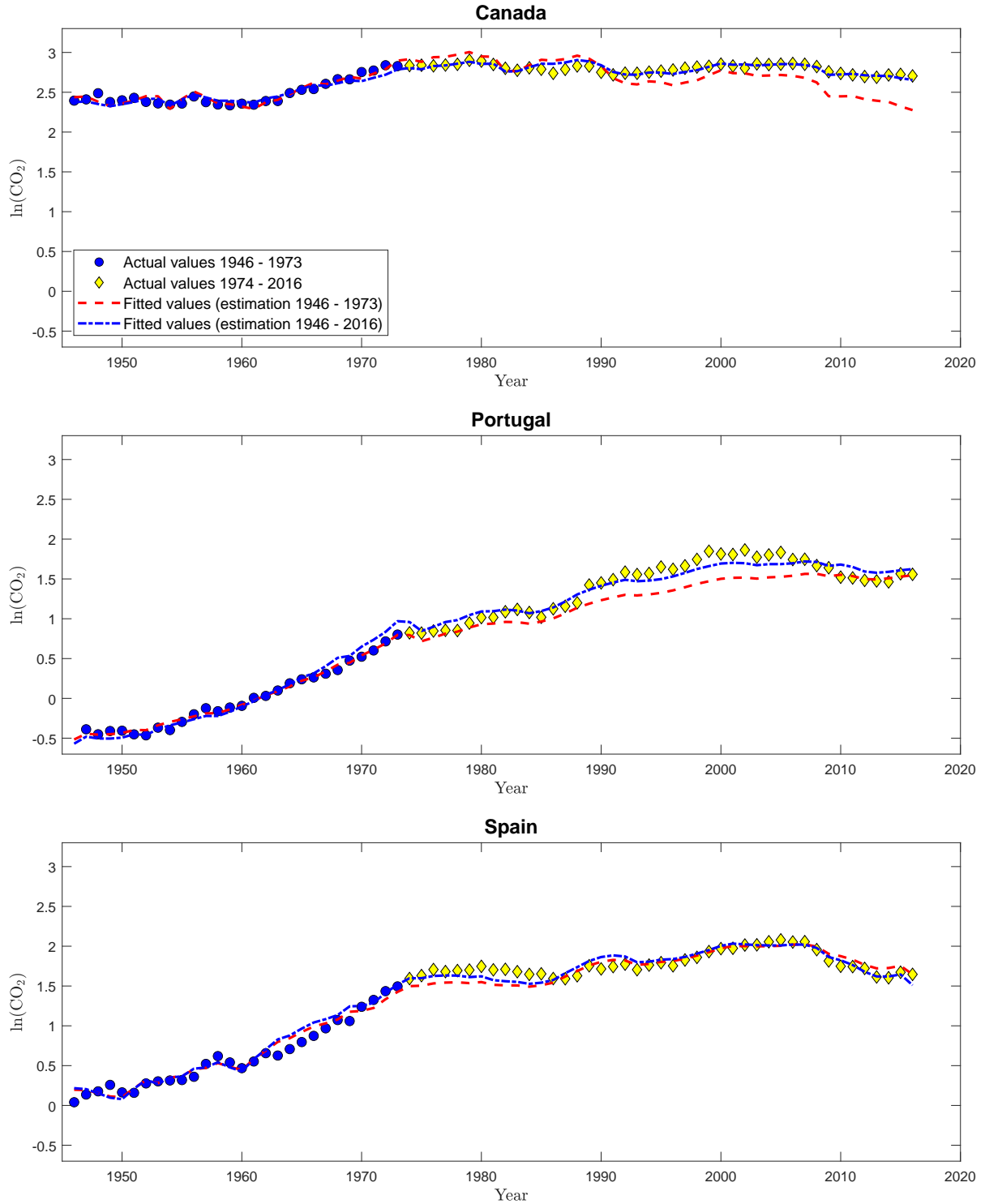
Figure 8: FM-OLS estimation results for a cointegrating linear relationship between log per capita GDP and log per capita CO_2 emissions for Canada, Portugal and Spain. The figure shows pairs of observations of log per capita GDP and log per capita CO_2 emissions, for 1946–1973 (blue circles) and 1974–2016 (yellow diamonds). The lines display fitted values over time obtained using different samples for parameter estimation: the dashed red line 1946–1973 and the dash-dotted blue 1946–2016.

| Country | p | $\hat{H}_{mov,sn}^{m,0.1}$ | |
		FM-OLS	IM-OLS
Australia	1	∞	∞
Belgium	1	1985	1985
Canada	1	1980	1981
Denmark	2	1997	2006
Finland	2	1988	1991
Italy	1	1982	1982
Japan	2	1991	2004
Portugal	1	2015	∞
Spain	1	2005	2012
Sweden	2	1985	1988
United Kingdom	1	1983	1983
United States	3	∞	∞

Table 7: Detection times when monitoring SO_2 emissions using the detector $\hat{H}_{mov,sn}^{m,0.1}$ and both, FM-OLS and IM-OLS. The column p indicates the polynomial degree, the calibration period is 1946–1973, the monitoring period is 1974–2016. The nominal significance level is 5%.

4.2 Results for Sulfur Dioxide Emissions

We now turn to monitoring the EKC for SO_2 emissions, with the results displayed in Table 7 for the detector $\hat{H}_{mov,sn}^{m,0.1}$ for both FM-OLS and IM-OLS.[41]

Qualitatively, the results for SO_2 are similar to the results for CO_2 discussed in the previous subsection. With the exception of Australia, the United States and the mixed case (as for CO_2 emissions) Portugal monitoring indicates structural breaks. Again, in line with the simulation findings, the FM-OLS break points are in no case later than the IM-OLS break points, with equal break dates from both estimators for three countries. This similarity of detected break points can be tentatively interpreted as evidence for underlying structural changes in the relation between economic activity and emissions in - or due to the delays – prior to the 1980s, potentially as a consequence of tighter environmental legislation.[42]

Table 8 displays the estimation results for Australia, Portugal and the US, i. e. , the three countries where no structural break was detected, or – to be precise – where for Portugal only FM-OLS

---

[41]Figures 41 to 52 in Supplementary Appendix D display monitoring results for SO_2 emissions, with residuals and monitoring statistics analogous to Figure 6 for CO_2 emissions for Finland, for all twelve countries.

[42]This finding again indicates the potential gains to be reaped by considering pooling, across countries, or pollutants or both. Compare Wagner *et al.* (2020) for a discussion of pooling issues and options in the context of EKC analysis. This suggests that a worthwhile extension to be considered could be combining Wagner *et al.* (2020) with the monitoring ideas pursued in this paper. Such an extension, beyond the scope of this paper, could potentially lead to smaller delays.

	1946–1973				1946–2016			
	θ_{D_1}	$\theta_{X_{1,1}}$	$\theta_{X_{1,2}}$	$\theta_{X_{1,3}}$	θ_{D_1}	$\theta_{X_{1,1}}$	$\theta_{X_{1,2}}$	$\theta_{X_{1,3}}$
				Australia				
FM-OLS	0.043	-0.819			**-0.042**	**2.541**		
IM-OLS	0.042	-0.770			**-0.043**	**2.550**		
				Portugal				
FM-OLS	-0.005	**1.047**			**-0.104**	**3.491**		
IM-OLS	0.002	**0.873**			**-0.092**	**3.215**		
				United States				
FM-OLS	*-0.013*	**-1831.120**	**183.982**	**-6.158**	**-0.133**	-305.981	31.361	-1.052
IM-OLS	-0.005	**-2190.198**	**220.246**	**-7.380**	**-0.133**	-61.963	7.512	-0.275

Table 8: FM-OLS and IM-OLS estimation results for a cointegrating linear relationship between log per capita GDP and log per capita SO_2 emissions for Australia and Portugal, and a cointegrating cubic relationship for the United States, for the calibration period 1946–1973 (left panel) and the full sample period 1946–2016 (right panel). *Italic* entries indicate significance of coefficients at the 10% level and **bold** entries significance of coefficients at the 5% level.

estimation leads to a break being detected. The results lead to some interesting observations. For Australia, estimation on the calibration period leads to (with all coefficients not significantly different from zero) positive trend coefficient and negative slope coefficient in a cointegrating linear relationship.[43] This is surprising to a certain extent, as a negative trend slope is typically considered to capture *autonomous energy efficiency increases*, admittedly more important for CO_2 emissions rather than SO_2 emissions. The expected signs, with the trend coefficient negative and the slope coefficient positive (and significant), emerge only over the full sample. For Portugal, the coefficient signs are as expected, with the trend coefficient only significant for the full sample. The slope coefficient to log per capita GDP becomes much larger when estimated over the full sample. The results for the cubic specification estimated for the US are hard to interpret with the enormous fluctuations in the coefficient values. However, the implied turning points do change in expected ways, meaning that the first one becomes smaller and the second one larger, e. g. , for FM-OLS and estimation over the short sample the two turning points are at about 16,000 and 27,000 US dollars, whereas full sample estimation leads to turning points at about 5,400 and 79,000 US dollars.

The results in Figure 9, however, indicate that indeed the relationship *specified* over the calibration period, when estimated over the full sample, leads to very good fit, either with a cointegrating

---

[43]It is in all likelihood mere coincidence that the trend coefficients are of more or less same magnitude over both sample periods with, effectively, only the sign changing.
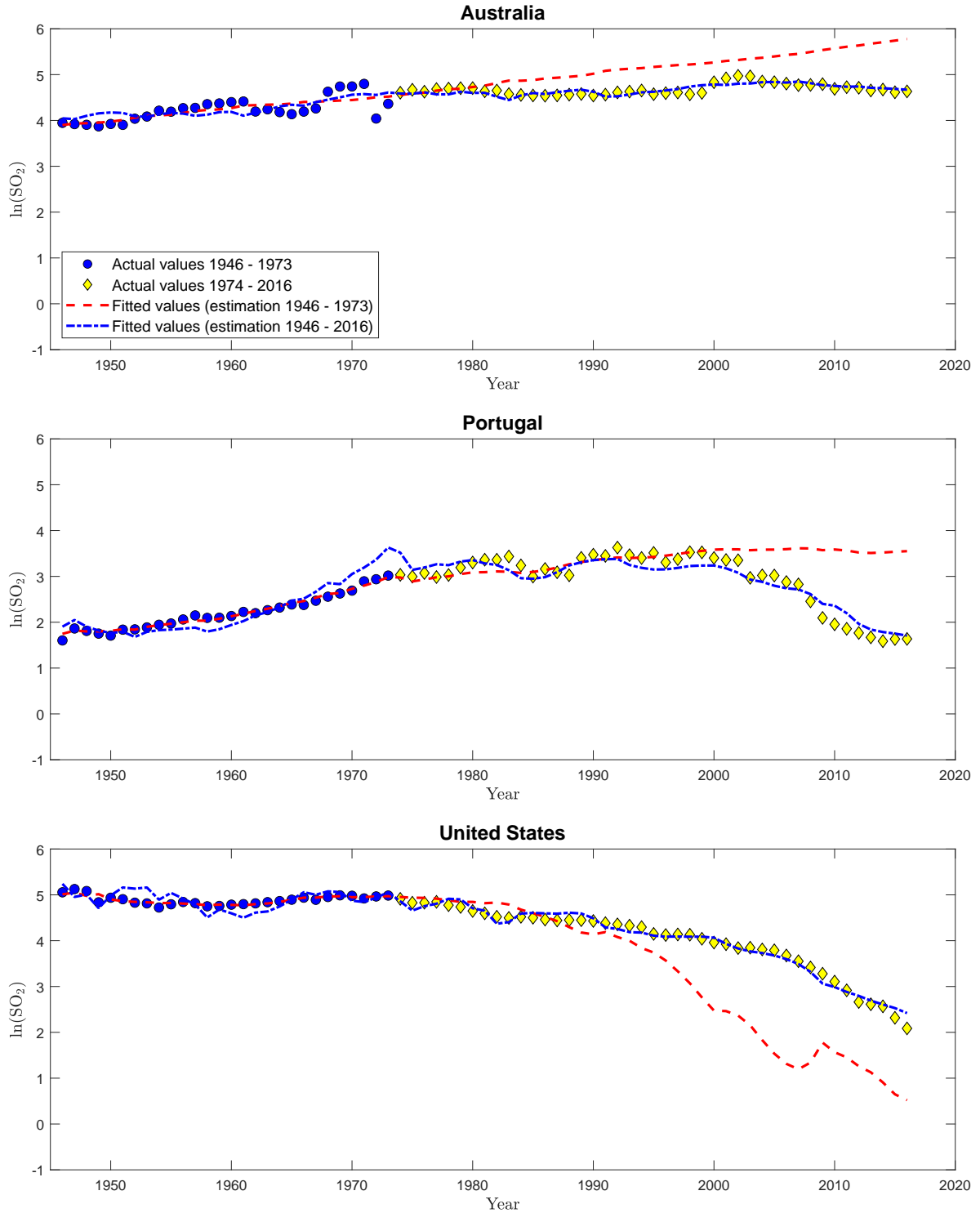
Figure 9: FM-OLS estimation results for a cointegrating linear relationship between log per capita GDP and log per capita CO_2 emissions for Australia and Portugal, and a cointegrating cubic relationship for the United States. The figure shows pairs of observations of log per capita GDP and log per capita SO_2 emissions, for 1946–1973 (blue circles) and 1974–2016 (yellow diamonds). The lines display fitted values over time obtained using different samples for parameter estimation: the dashed red line 1946–1973 and the dash-dotted blue 1946–2016.

41

linear relationship for Australia and Portugal or with a cubic relationship for the US.[44]

Considering the full sample period leads – similar to CO_2 emissions – to a CPR relationship for all nine countries where a structural has been detected. Also similar to CO_2 the evidence is mixed with respect to either change in the parameters or a change in the minimal polynomial degree of a cointegrating EKC for SO_2 emissions.[45]

5   Summary and Conclusions

The paper extends the residual-based monitoring procedure for cointegrating relationship developed in Wagner and Wied (2017) in two dimensions. First, in addition to the detector studied in that paper, we consider a number of detectors that consider two aspects not considered in detail in the earlier paper: self-normalization and moving window detectors. Second, the approach is extended from cointegrating linear to cointegrating polynomial regressions (CPRs).

The starting point of the considerations is the – corresponding extension to the CPR case (see, e. g. Wagner, 2020) of the – Shin (1994) test statistic for the null hypothesis of cointegration. The full sample test statistic is then used as the basis for developing monitoring procedures, similar to Wagner and Wied (2017) for the case of cointegrating linear regressions. As also usual in the cointegrating regression literature, the regressors are allowed to be endogenous and the stationary errors are allowed to be serially correlated. Consequently, parameter estimation is based on modified OLS estimators, to be precise, FM-OLS, D-OLS and IM-OLS, in their versions adapted to the CPR setting to allow for the construction of nuisance parameter free monitoring statistics by scaling out a scalar long-run variance. However, even the usage of suitably modified least squares regressors is not, in the general case, sufficient to arrive at nuisance parameter free monitoring statistics. In the CPR case, as mentioned in Remark 1, additionally *full design* is required to arrive at nuisance parameter free test statistics; either by standardization or by self-normalization. By construction, and as in the cointegrating linear case, the limiting distributions of the monitoring statistics coincide for FM-OLS and D-OLS, whilst IM-OLS leads to different limiting distributions for the monitoring statistics.

Both, self-normalization as well as the consideration of moving window detectors turns out to be beneficial. The combination of these two new elements leads to the *grosso modo* best performance.

---

[44]Figure 29 in Supplementary Appendix D shows similar results when using IM-OLS for parameter estimation.

[45]See Table 11 in Supplementary Appendix D for the minimal polynomial degrees on the full sample.

More precisely, this means that self-normalization leads to smaller over-rejections under the null, without leading to strong disadvantages in terms of either size-corrected power or detection delays. The performance differences between a standardized and a self-normalized detector reflect the (well-known small-sample) problems associated with long-run covariance estimation. Using a moving window rather than an expanding window also contributes positively to performance. The idea behind moving window detectors is to reduce the impact of pre-break observations on the monitoring statistics in the post-break period and the simulation evidence indicates that this is indeed what happens. The finite sample performance differs to a certain extent across the three estimators. IM-OLS leads to, with the differences more pronounced for small samples, smaller over-rejections under the null hypothesis than FM-OLS, which in turn leads to lower over-rejections than D-OLS. This, however, comes in conjunction with slightly smaller size-corrected power and slightly bigger delays of IM-OLS compared to FM-OLS. The evidence is a bit mixed and the differences are often a bit too small to give an unequivocal recommendation concerning the usage of either FM-OLS or IM-OLS.

The monitoring statistics, to be precise (in the main text only) the best performing variant using self-normalization in conjunction with a short moving window, are then used to investigate the stability of cointegrating (polynomial) environmental Kuznets curves for both carbon and sulfur dioxide emissions for a sample of twelve countries over the period 1946–2016, with a calibration period ranging from 1946–1973, i.e., until prior to the first oil price shock. One interesting observation is that over the calibration period especially for CO_2 for the majority of countries in fact a cointegrating *linear* relationship is present, in line with effectively "balanced" growth of economic activity and emissions (from burning fossil fuel) until the mid 1970s. The monitoring procedures indicate, for both CO_2 and SO_2, structural changes in nine out of twelve countries. Of course, especially for the cases where a linear cointegrating relationship prevails over the calibration period this may not be too big a surprise. No structural break is detected for Canada, Portugal and Spain for CO_2 emissions and Australia, Portugal and the US for SO_2 emissions. For the country-pollutant combinations where no structural break is detected, the *type of* cointegrating relationship specified over the calibration period leads to very good fit also over the full sample period, and of course even better fit when the parameters are estimated over the full sample. This is not the case for countries where breaks are detected, indicating that structural breaks are indeed detected when present. Altogether, with the caveat of in some cases late detection times, remember the delays illustrated in the simulation study, the monitoring procedures lead to "plausible" findings in our

application.

Future work needs to address inter alia the following issues left open in this paper: First, of course, large delays are problematic, even in an ex-post exercise one would want to date the breaks as precisely as possible, e.g., to gauge the time it takes until behavioral or legislative changes translate into changes in the economic activity-emissions nexus. For certain applications (when, e.g., legislation comes into place in a group of countries at the same time) it may be possible to rely also upon the cross-sectional dimension to date break points. The seemingly unrelated cointegrating polynomial regression analysis put forward in Wagner *et al.* (2020) or the classical panel EKC analysis of de Jong and Wagner (2018) might serve as starting points for developing monitoring statistics for small and large cross-sectional dimension, respectively. Second, if one wants to use the monitoring procedures in a real time manner – rather than for a historical analysis as in this paper – it may be important to consider extending end-of-sample cointegrating break point tests to the setting considered here (one such test for the cointegrating linear case is discussed in Andrews and Kim, 2006). Extensions to more general settings are, however, not obvious. Third, it may be interesting to monitor and detect structural breaks *in the other direction*, i.e., from a spurious regression to a cointegrating polynomial regression. This could be achieved, e.g., by suitably extending the results of Sakarya *et al.* (2015) from cointegrating linear to cointegrating polynomial regressions.

Acknowledgments

References

Andreoni, J. and A. Levinson (2001). The Simple Analytics of the Environmental Kuznets Curve. *Journal of Public Economics* **80**, 269–286.

Andrews, D.W.K. and J.Y. Kim (2006). Tests for Cointegration Breakdown Over a Short Time Period. *Journal of Business and Economic Statistics* **24**, 379–394.

Arrow, K.J., B. Polin, R. Costanza, P. Dasgupta, C. Folke, C.S. Holling, B.O. Jansson, S. Levin, K.G. Maler, C. Perrings and D. Pimentel (1995). Economic Growth, Carrying Capacity, and the Environment. *Science* **268**, 520–521.

Aue, A., S. Hörmann, L. Horváth, M. Hušková and J. Steinebach (2012). Sequential Testing for the Stability of High-Frequency Portfolio Betas. *Econometric Theory* **28**, 804–837.

Aue, A., L. Horváth and M.L. Reimherr (2009). Delay Times of Sequential Procedures for Multiple Time Series Regression Models. *Journal of Econometrics* **149**, 174–190.

Bertinelli, L. and E. Strobl (2005). The Environmental Kuznets Curve Semi-Parametrically Revisited. *Economics Letters* **88**, 350–357.

Boden, T.A., G. Marland and R.J. Andres (2018). Global, Regional, and National Fossil-Fuel CO_2 Emissions. Carbon Dioxide Information Analysis Center at Appalachian State University, Boone North Carolina.

Bolt, J., R. Inklaar, H. de Jong and J.L. van Zanden (2018). Rebasing Maddison: New Income Comparisons and the Shape of Long-Run Economic Development. *Maddison Project Working paper* 10. Maddison Project Database, version 2018.

Brock, W.A. and M.S. Taylor (2005). Economic Growth and the Environment: A Review of Theory and Empirics. In: Aghion, P. and S. Durlauf (Eds.), *Handbook of Economic Growth*, Ch. 28, 1749–1821. North-Holland, Amsterdam.

Brock, W.A. and M.S. Taylor (2010). The Green Solow Model. *Journal of Economic Growth* **15**, 127–153.

Busetti, F. and A.M.R. Taylor (2004). Tests of Stationarity Against a Change in Persistence. *Journal of Econometrics* **123**, 33–66.

Choi, I. and E. Kurozumi (2012). Model Selection Criteria for the Leads-and-Lags Cointegrating Regressions. *Journal of Econometrics* **169**, 224–238.

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

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

Chu, C.S.J., M. Stinchcombe and H. White (1996). Monitoring Structural Change. *Econometrica* **64**, 1045–1065.

Cropper, M. and C. Griffiths (1994). The Interaction of Population Growth and Environmental Quality. *American Economic Review* **84**, 250–254.

Davidson, J. and R.M. de Jong (2000). The Functional Central Limit Theorem and Convergence to Stochastic Integrals II: Fractionally Integrated Processes. *Econometric Theory* **16**, 643–666.

de Jong, R.M. and M. Wagner (2018). Panel Nonlinear Cointegration Analysis and the Environmental Kuznets Curve. *SFB823 Discussion Paper* 22/18.

Dickey, D.A. and W.A. Fuller (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica* **49**, 1057–1072.

Grabarczyk, P., M. Wagner, M. Frondel and M. Sommer (2018). A Cointegrating Polynomial Regression Analysis of the Material Kuznets Curve Hypothesis. *Resources Policy* **57**, 236–245.

Grossman, G.M. and A.B. Krueger (1991). Environmental Impacts of a North American Free Trade Agreement. *NBER Working Paper* 3914.

Grossman, G.M. and A.B. Krueger (1993). Environmental Impacts of a North American Free Trade Agreement. In Garber, P. (Ed.) *The Mexico-US Free Trade Agreement*, 13–56. MIT Press, Cambridge.

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

Guzmán, J.I., T. Nishiyama and J.E. Tilton (2005). Trends in the Intensity of Copper Use in Japan Since 1960. *Resources Policy* **30**, 21–27.

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Horváth, L., M. Hušková, P. Kokoszka and J. Steinebach (2004). Monitoring Changes in Linear Models. *Journal of Statistical Planning and Inference* **126**, 225–251.

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

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford.

Jones, L.E. and R.E. Manuelli (2001). Endogenous Policy Choice: The Case of Pollution and Growth. *Review of Economic Dynamics* **4**, 369–405.

Kijima, M., K. Nishide and A. Oyama (2010). Economic Models for the Environmental Kuznets Curve: A Survey. *Journal of Economic Dynamics and Control* **34**, 1187–1201.

Kim, J.Y. (2000). Detection of Change in Persistence of a Linear Time Series. *Journal of Econometrics* **95**, 97–116.

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

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How sure are we that Economic Time Series have a Unit Root? *Journal of Econometrics* **54**, 159–178.

Labson, B.S. 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.

Millimet, D.L., J.A. List and T. Stengos (2003). The Environmental Kuznets Curve: Real Progress or Misspecified Models? *Review of Economics and Statistics* **85**, 1038–1047.

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

OECD (2020). Air and GHG emissions (indicator). https://doi.org/10.1787/93d10cf7-en (accessed on 05 March 2020).

Perron, P. and T.J. Vogelsang (1993). The Great Crash, the Oil Price Shock and the Unit Root Hypothesis: Erratum. *Econometrica* **61**, 248–249.

Phillips, P.C.B. (1987). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika* **74**, 535–547.

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.

Phillips, P.C.B. and P. Perron (1988). Testing for a Unit Root in Time Series Regression. *Biometrika* **75**, 335–346.

Phillips, P.C.B. and S. Shi (2018). Financial Bubble Implosion and Reverse Regression. *Econometric Theory* **34**, 705–753.

Phillips, P.C.B., S. Shi and J. Yu (2015a). Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500. *International Economic Review* **56**, 1043–1078.

Phillips, P.C.B., S. Shi and J. Yu (2015b). Testing for Multiple Bubbles: Limit Theory of Real Time Detectors. *International Economic Review* **56**, 1079–1134.

Phillips, P.C.B., Y. Wu and J. Yu (2011). Explosive Behavior in the 1990s NASDAQ: When Did Exuberance Escalate Asset Values? *International Economic Review* **52**, 201–226.

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

Sakarya, N., M. Wagner and D. Wied (2015). Monitoring a Change from Spurious Regression to Cointegration. *Mimeo*.

Schmalensee, R., T.M. Stoker and R.A. Judson (1998). World Carbon Dioxide Emissions: 1950–2050. *Review of Economics and Statistics* **80**, 15–27.

Selden, T.M. and D. Song (1995). Neoclassical Growth, the J Curve for Abatement, and the Inverted U Curve for Pollution. *Journal of Environmental Economics and Management* **29**, 162–168.

Shafik, N. and S. Bandyopadhyay (1992). Economic Growth and Environmental Quality: Time-Series and Cross-Country Evidence. *World Bank Policy Research Working Paper* WPS 904.

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

Smith, S.J., J. van Aardenne, Z. Klimont, R.J. Andres, A. Volke and S. Delgado Arias (2011). Anthropogenic Sulfur Dioxide Emissions, 1850–2005: National and Regional Data Set by Source Category, Version 2.86. Data distributed by the NASA Socioeconomic Data and Applications Center (SEDAC), CIESIN, Columbia University, Palisades, New York.

Stock, J.H. and M.W. Watson (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* **61**, 783–820.

Stokey, N.L. (1998). Are There Limits to Growth? *International Economic Review* **39**, 1–31.

Stuermer, M. (2018). 150 Years of Boom and Bust: What Drives Mineral Commodity Prices? *Macroeconomic Dynamics* **22**, 702–717.

Stypka, O. and M. Wagner (2020). Cointegrating Multivariate Polynomial Regressions: Fully Modified OLS Estimation and Inference. *Mimeo.*

Vogelsang, T.J. and M. Wagner (2014a). Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions. *Journal of Econometrics* **178**, 741–760.

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

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

Wagner, M. (2018). Estimation and Inference for Cointegrating Regressions. In: *Oxford Research Encyclopedia of Economics and Finance.*

Wagner, M. (2020). Residual Based Cointegration and Non-Cointegration Tests for Cointegrating Polynomial Regressions. Revised and resubmitted to *Empirical Economics*.

Wagner, M., P. Grabarczyk and S.H. Hong (2020). Fully Modified OLS Estimation and Inference for Cointegrating Polynomial Regressions and the Environmental Kuznets Curve for Carbon Dioxide Emissions. *Journal of Econometrics* **214**, 216–255.

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

Wagner, M. and D. Wied (2017). Consistent Monitoring of Cointegrating Relationships: The US Housing Market and the Subprime Crisis. *Journal of Time Series Analysis* **38**, 960–980.

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.

A Appendix A: Proofs

Proof of Lemma 1:

The lemma follows with straightforward changes essentially from available *full sample* results: For FM-OLS see Wagner and Hong (2016, Proof of Proposition 5) and for D-OLS see Choi and Saikkonen (2010, Lemma A.2). For IM-OLS the result follows with two changes from the results in Vogelsang and Wagner (2014b). First, again a change from a full sample result to a calibration (sub-)sample result and second, the results in that paper are all formulated only under the null hypothesis of a true linear relationship nested in a CPR-type relationship; a few results have to be correspondingly – and straightforwardly – modified. □

Proof of Proposition 1:

Using the results from Lemma 1, the continuous mapping theorem (see, e.g., Hall and Heyde, 1980, Theorem A.3.) and the assumption of consistent long-run covariance estimation leads to the stated asymptotic distributions under the null hypothesis. □

Proof of Proposition 2:

The result follows from Proposition 1 in conjunction with the continuous mapping theorem, since the limit $\mathcal{H}(s)$ of $\hat{H}(s)$, using short-hand notation also for the limit quantities, is well defined for all considered monitoring statistics. The same holds true for $|\frac{\hat{H}(s)}{g(s)}|$, since $0 < g(s) < \infty$ is assumed to be continuous. □

Proof of Proposition 3:

The proof of the proposition is similar in spirit and follows from similar arguments as the proof of Proposition 2 of Wagner and Wied (2017, Supplementary Appendix A) for monitoring in the linear cointegration case.

To show consistency against fixed alternatives for part (a) of the proposition, the limiting behavior of the residual partial sum processes is key. Consider here the FM-OLS residual partial sum process,

with the results being similar for both D-OLS and IM-OLS (without partial summing due to the partial summed regression). For $0 < m \leq r < s \leq 1$ it holds that:

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor sT \rfloor} \hat{u}_{t,m}^{\mathrm{F}} = \frac{1}{\sqrt{T}} \sum_{t=2}^{\lfloor rT \rfloor} \hat{u}_{t,m}^{\mathrm{F}} + \frac{1}{\sqrt{T}} \sum_{t=\lfloor rT \rfloor+1}^{\lfloor sT \rfloor} u_t - \frac{1}{\sqrt{T}} \sum_{t=\lfloor rT \rfloor+1}^{\lfloor sT \rfloor} v_t' \hat{\Omega}_{vv,m}^{-1} \hat{\Omega}_{vu,m}$$

$$- \frac{1}{\sqrt{T}} \sum_{t=\lfloor rT \rfloor+1}^{\lfloor sT \rfloor} D_t'(\hat{\theta}_{D,m}^{\mathrm{F}} - \theta_D) - \frac{1}{\sqrt{T}} \sum_{t=\lfloor rT \rfloor+1}^{\lfloor sT \rfloor} X_t'(\hat{\theta}_{X,m}^{\mathrm{F}} - \theta_X) \quad (46)$$

$$- \frac{1}{\sqrt{T}} \sum_{t=\lfloor rT \rfloor+1}^{\lfloor sT \rfloor} D_t'(\theta_D - \theta_{D,1}) - \frac{1}{\sqrt{T}} \sum_{t=\lfloor rT \rfloor+1}^{\lfloor sT \rfloor} X_t'(\theta_X - \theta_{X,1}).$$

The first term converges to $\omega_{u \cdot v} \widetilde{W}_{u \cdot v,m}(r)$ according to Lemma 1. Depending upon which case considered, at least one of the other terms diverges in case of a fixed alternative. In case (i) the second term diverges at rate T, since in this case $\{u_t\}$ is an I(1) process for $t \geq \lfloor rT \rfloor + 1$. Similarly, in cases (ii) or (iii) either the next to last or last term diverges – or both in case breaks occur in both θ_D and θ_X.

Therefore, for any of the considered fixed alternatives the partial summed residual process or – in case of IM-OLS – the residual process diverges. The result stated in (a) follows immediately from the definitions of the considered monitoring statistics.

We now turn to part (b) of Proposition 3 and consider local asymptotic power. For item (i) it follows for the residual partial sum processes of FM-OLS, D-OLS and IM-OLS residuals that:

$$\frac{1}{\sqrt{T}} \hat{S}_{\lfloor sT \rfloor,m}^u \Rightarrow \omega_{u \cdot v} \widetilde{Q}_{u \cdot v,m}(s) + \delta \omega_\gamma \int_r^s (W_\gamma(z) - W_\gamma(r)) \, dz \quad (47)$$

with $\widetilde{Q}_{u \cdot v,m}(s)$ denoting the corresponding limiting process as in Proposition 1. For item (ii) it follows that:

$$\frac{1}{\sqrt{T}} \hat{S}_{\lfloor sT \rfloor,m}^u \Rightarrow \omega_{u \cdot v} \widetilde{Q}_{u \cdot v,m}(s) + \int_r^s D(z)' dz \Delta_{\theta_D}. \quad (48)$$

It remains to consider the asymptotic behavior of the residual partial sum process appearing in the third item of part (b) of the proposition. The partial sum process of the residuals converges to the following limiting process:

$$\frac{1}{\sqrt{T}} \hat{S}_{\lfloor sT \rfloor,m}^u \Rightarrow \omega_{u \cdot v} \widetilde{Q}_{u \cdot v,m}(s) + \int_r^s \mathbf{B}_v(s)' dz \Delta_{\theta_X}$$

$$= \omega_{u \cdot v} \widetilde{Q}_{u \cdot v,m}(s) + \int_r^s \mathbf{W}_v(s)' dz \Omega_{vv}^{1/2'} \Delta_{\theta_X}, \quad (49)$$

with $\mathbf{\Omega}_{vv}^{1/2} := \mathrm{diag}(\Omega_{vv}^{1/2}, \omega_k^2, \ldots, \omega_k^{p_k})$ and ω_k the lower right hand corner element of $\Omega_{vv}^{1/2}$ corresponding to the k-the element of v_t. Equations (47) to (49) show that in case of local alternatives the limiting processes differ from the limiting processes derived under the null hypothesis by an extra term, depending upon case considered equal to $\delta\omega_\gamma \int_r^s (W_\gamma(z) - W_\gamma(r))\, dz$, $\int_r^s D(z)' dz \Delta_{\theta_D}$ and $\int_r^s \mathbf{W}_v(s)' dz \mathbf{\Omega}_{vv}^{1/2\prime} \Delta_{\theta_X}$. These terms can be made arbitrarily big (in mean square sense) by choosing δ, Δ_{θ_D} or Δ_{θ_X} which translates in the detectors also becoming arbitrarily large. To illustrate the mechanics consider $\hat{H}_{mov,sn}^{m,n}$ and the case of a local I(1) break, in which case one obtains for $0 < m \leq r < s \leq 1$:

$$
\begin{aligned}
\hat{H}_{mov,sn}^{m,n}(s) \Rightarrow\ & \frac{\int_{\max\{0,s-n\}}^s \left(\omega_{u\cdot v}\widetilde{Q}_{u\cdot v,m}(z) + \delta\omega_\gamma \int_r^z W_\gamma(\zeta) - W_\gamma(r)d\zeta\right)^2 dz}{\omega_{u\cdot v}^2 \int_0^m \widetilde{Q}_{u\cdot v,m}^2(z)dz} \\[2mm]
=\ & \frac{\int_{\max\{0,s-n\}}^s \widetilde{Q}_{u\cdot v,m}^2(z)dz}{\int_0^m \widetilde{Q}_{u\cdot v,m}^2(z)dz} \\[2mm]
& + \frac{2\delta\omega_\gamma \int_{\max\{0,s-n\}}^s \widetilde{Q}_{u\cdot v,m}(z) \int_r^z W_\gamma(\zeta) - W_\gamma(r)d\zeta dz}{\omega_{u\cdot v} \int_0^m \widetilde{Q}_{u\cdot v,m}^2(z)dz} \\[2mm]
& + \frac{(\delta\omega_\gamma)^2 \int_{\max\{0,s-n\}}^s \left(\int_r^z W_\gamma(\zeta) - W_\gamma(r)d\zeta\right)^2 dz}{\omega_{u\cdot v}^2 \int_0^m \widetilde{Q}_{u\cdot v,m}^2(z)dz} \\[2mm]
=\ & \mathcal{H}_{mov,sn}^{m,n}(\widetilde{Q}_{u\cdot v,m}, s) + \frac{2\delta\omega_\gamma \int_{\max\{0,s-n\}}^s \widetilde{Q}_{u\cdot v,m}(z) \int_r^z W_\gamma(\zeta) - W_\gamma(r)d\zeta dz}{\omega_{u\cdot v} \int_0^m \widetilde{Q}_{u\cdot v,m}^2(z)dz} \\[2mm]
& + \frac{(\delta\omega_\gamma)^2 \int_{\max\{0,s-n\}}^s \left(\int_r^z W_\gamma(\zeta) - W_\gamma(r)d\zeta\right)^2 dz}{\omega_{u\cdot v}^2 \int_0^m \widetilde{Q}_{u\cdot v,m}^2(z)dz},
\end{aligned}
\tag{50}
$$

with $\int_r^z W_\gamma(\zeta) - W_\gamma(r)d\zeta = 0$ for $r > z$. Equation (50) shows that, as expected, the extra term in the residual process translates (upon squaring) into two additional terms in the limiting process in comparison to the behavior under the null hypothesis. The third term, being a squared expression, can be made arbitrarily large and positive by choosing δ large enough and is of larger order than the second (non-squared) term. The result follows analogously for all detectors and all three considered local alternatives. $\qquad\square$