Essays on Stability, Functional Form and Poolability of Nonlinear Cointegrating Regressions

Dissertation

in partial fulfillment of the requirements for the degree of
doctor rerum politicarum (Dr. rer. pol.)

submitted to the Department of Business and Economics
TU Dortmund University

submitted by

Fabian Knorre

Dortmund, February 2025

Veröffentlichung als Dissertation in der Fakultät Wirtschaftswissenschaften
der Technischen Universität Dortmund

Dortmund, Mai 2025

Supervisors:

Prof. Dr. Martin Wagner
Department of Economics, University of Klagenfurt

Prof. Dr. Ludger Linnemann
Department of Business and Economics, TU Dortmund University

Doctoral committee:

Prof. Dr. Martin Wagner
Department of Economics, University of Klagenfurt

Prof. Dr. Ludger Linnemann
Department of Business and Economics, TU Dortmund University

Prof. Dr. Carsten Jentsch
Department of Statistics, TU Dortmund University

Date and location of oral examination:
May 12th, 2025, Dortmund, Germany

Acknowledgements

There are a number of people and institutions that I would like to thank and without whom this thesis would not have been possible. First and foremost, I would like to express my deepest gratitude to my supervisor Martin Wagner. His unwavering support and guidance over the years have been invaluable to me. He was also a continuous sparring partner in discussing all the econometric challenges along the way (and also for matters beyond econometrics) and if the circumstances were demanding, even with out of business hours support. Besides introducing me to the fascinating research field of cointegration, he also gave me some understanding of the word *precision* and continuously challenged me to improve, which improved me. Thank you for your dedication and support throughout this journey! I would also like to express my gratitude to my supervisor Ludger Linnemann, who supported me and this project all the way and had time for my questions whenever I needed them. Furthermore, I want to thank Carsten Jentsch for agreeing to be the third member of the doctoral committee.

I would like to thank the Ruhr Graduate School in Economics for accepting me into their PhD program, their financial support and especially providing a stimulating research environment with many great people. Furthermore, I gratefully acknowledge financial support from Deutsche Forschungsgemeinschaft (DFG) via the Collaborative Research Center SFB823: *Statistical Modelling of Nonlinear Dynamic Processes*. Additionally, I acknowledge the computing time provided on the Linux HPC cluster at TU Dortmund University Dortmund (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the DFG as project 271512359.

This list would not be complete if I did not mention a few colleagues who also enriched my time at the TU Dortmund in both the econometric and non-econometric sense. Thank you, Karsten, Oliver, Rafael, Peter, Lukas, Patrick, Maximilian and Sascha.

Zu guter Letzt möchte ich noch meinen Eltern Tilla und Lothar danken, sowie meiner Schwester Julia, die mich in allen Lebenslangen unterstützt haben und stets für mich da waren. Mein besonderer Dank gilt Ina, für ihre grenzenlose Unterstützung, Motivation und Liebe.

Contents

Introduction **1**

1 Monitoring Cointegrating Polynomial Regressions: Theory and Application to the Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions **5**

1.1 Introduction 5

1.2 Theory 8

1.2.1 Model, Assumptions and Parameter Estimation 8

1.2.2 The Monitoring Statistics 15

1.3 Finite Sample Performance 24

1.4 The Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions 33

1.4.1 Results for Carbon Dioxide Emissions 36

1.4.2 Results for Sulfur Dioxide Emissions 42

1.5 Summary and Conclusions 45

1.6 Appendix 47

1.6.1 Proofs 47

1.6.2 Local Asymptotic Power 50

1.6.3 Additional Finite Sample Performance Results 60

1.6.4 Additional Empirical Results 86

2 Comparison of Nonparametric Estimators and Specification Tests for Nonlinear Cointegrating Regressions: Results from a Large Scale Simulation Study **97**

2.1 Introduction 97

2.2 Model and Assumptions 99

2.3 Estimators 100

2.4 Specification Tests 110

2.5 Large Scale Simulation Study 117

2.5.1 Estimators – Average Root Mean Squared Error . . . 123
2.5.2 Specification Tests – Null Rejection Probabilities and Size-corrected Power . 132
2.6 Summary and Conclusion . . . 141
2.7 Appendix . . . 141
2.7.1 Assumptions . . . 141
2.7.2 Additional Finite Sample Performance Results . . . 150

3 Fully Modified OLS Estimation and Inference for Seemingly Unrelated Cointegrating Polynomial Regressions with Common Integrated Regressors 191

3.1 Introduction . . . 191
3.2 Seemingly Unrelated Cointegrating Polynomial Regressions with Common Integrated Regressors . . . 194
3.2.1 Setup and Assumptions . . . 194
3.2.2 Fully Modified OLS Estimation . . . 196
3.2.3 Specification Testing Based on Augmented and Auxiliary Regressions . . . 198
3.2.4 Group-wise pooling . . . 203
3.3 Finite Sample Performance . . . 211
3.4 Empirical Illustration - MKCs and EKCs . . . 219
3.4.1 Group-wise Pooling . . . 232
3.5 Summary and Conclusion . . . 238
3.6 Appendix . . . 238
3.6.1 Proofs . . . 238
3.6.2 Additional Finite Sample Performance Results . . . 252

List of Figures 256

List of Tables 262

Bibliography 270

Introduction

Cointegration regression techniques and inference based on them provide a useful toolkit when analyzing relationships between integrated time series, as for example in macroeconomics, finance or environmental economics. While many methods estimate linear relationships, this assumption might be too restrictive for certain applications. One way of modeling nonlinearities is to consider transformations such as logarithms or powers of integrated regressors in a linear regression model. Such a model which includes integer powers of integrated processes as explanatory variables is called in the terminology of Wagner and Hong (2016) *cointegrating polynomial regression* (CPR). As pointed out by Wagner (2015), treating powers of integrated processes simply as integrated processes themselves, is conceptually wrong. To appropriately handle such situations, Wagner and Hong (2016) provide theory and extend the fully-modified OLS estimator of Phillips and Hansen (1990) to CPRs. As typical in the cointegrating literature, the regressors are allowed to be endogenous and the stationary errors are allowed to be serially correlated. In such cases, the estimator of Wagner and Hong (2016) has a zero mean Gaussian mixture limiting distribution and thus standard inference can be performed.

This cumulative dissertation contributes to the nonlinear cointegrating regression literature and provides new tools and comparisons of existing procedures. In Chapters 1 and 3 we consider the mentioned CPRs to model nonlinear relationships and resort to similar modifications when developing estimators as Wagner and Hong (2016). Our primary motivation is the analysis of the environmental Kuznets curve (EKC). The EKC hypothesis postulates an inverted U-shaped relationship between the level of economic development and pollution or emissions. 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). Since the seminal contributions of, e. g. , Grossman and Krueger (1995) or Shafik and Bandyopadhyay (1992), the literature has become voluminous and continues to grow rapidly. Surveys discussing the links between economic growth and the environment are provided by Brock and Taylor (2005) or Kijima *et al.* (2010). In our analyses, we consider the relationship between logarithms of per capita GDP, as a proxy for economic development, and carbon and sulfur dioxide (CO_2 and SO_2) emissions, also in logarithms of per capita terms.

In Chapter 1 we develop a residual-based monitoring procedure to detect structural change in CPRs. There are two ways of structural change detectable by our procedure: The relationship

may turn into a spurious relationship and the parameters of the relationship may change. Our developed monitoring statistics extend those of 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 in more detail: Self-normalization and moving window detectors. Second, the approach is extended from cointegrating linear to cointegrating polynomial regressions (CPRs). The starting point for the monitoring statistics is the extension of the Shin (1994) test statistic for the null hypothesis of linear cointegration to the CPR setting (see, e. g. Wagner, 2023). Inspired by Chu *et al.* (1996), the considered monitoring statistics are based on parameter estimation on a calibration period, which is assumed to be free of structural change. These parameter estimates are then used to compute fitted values on the remaining sample. 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. A structural break is then detected if residuals become *too large*, which is evaluated by calculating the monitoring statistics and comparing them with critical values based on their limiting distributions. We perform a simulation study to assess null rejection probabilities, size-corrected power and delay in detection of the monitoring statistics. It turns out, that self-normalization in combination with a moving window approach leads in many cases to the best performance. With self-normalization, where we calculate the ratio of two terms, we avoid the issue of long-run variance estimation. Finally, we use our monitoring procedures to assess for 12 industrialized countries whether – and if so when – the relationship between emissions and economic activity has changed after the first oil price shock in 1973. Chapter 1 is a joint work together with Martin Wagner and Maximilian Grupe and has been published in a slightly modified version in *Econometrics*, Volume 9, Issue 12, pp. 1–35.

Chapter 2 is dedicated to nonparametric cointegrating regressions and specification tests for nonlinear cointegrating relationships. The finding of Wang and Phillips (2009b) that conventional kernel estimators can provide standard limiting distributions in nonparametric cointegrating regressions with the presence of error serial correlation and regressor endogeneity has accelerated the development and attention for nonparametric cointegrating regressions. In addition to the development of asymptotic theory for nonparametric estimators in cointegrating regressions, various specification tests for testing on functional forms in nonlinear cointegrating regressions have been developed. The clear advantage of nonparametric regression compared to the parametric counterpart is that no functional form needs to be specified. On the other hand, parametric esti-

mation exhibits faster convergence rates, which renders specification tests useful to test on certain parametric functional forms and rely on parametric estimation if the null hypothesis of a particular functional form is not rejected. In Chapter 2 we compare the performance of several nonparametric and specification tests for nonlinear cointegrating regressions in a large scale simulation study. As performance measures for the estimators we consider bias and root mean squared error and for the specification tests null rejection probabilities and size-corrected power. Within that study we use several regression functions and consider different combinations of regressor endogeneity and error serial correlation. Furthermore, we assess the impact of necessary user choices, which need to be made for almost all considered estimators and specification tests. For the kernel based estimators and tests, kernel and bandwidth rule need to be chosen. For other estimators or tests truncation parameters or weighting functions need to be selected. We vary those choices to assess their impact on estimation and test performances. With the simulation study, we want to provide practitioners with a guidance to help them select a suitable estimator or test for their situation, taking into account additional parameters to be selected. It turns out, that the Dong *et al.* (2021) estimator without any weighting works best in our simulations. From the specification tests, the subresidual based test of Choi and Saikkonen (2010), using residuals from leads and lags regression, performs overall best. All estimators and specification tests have been implemented by ourselves in `MATLAB` and procedures are available on request. Chapter 2 is a joint work together with Martin Wagner.

In Chapter 3 we develop fully modified-type estimators for seemingly unrelated cointegrating polynomial regressions, including common – across equations – integrated regressors and integer powers of them. The chapter thus extends the results of Wagner *et al.* (2020) to systems with common integrated regressors and furthermore provides the general multiple regressors formulation in contrast to the stylized system with only one integrated regressor. The formulation provides great flexibility in that sense that the degree of the deterministic trend, numbers of integrated regressors and common integrated regressors, as well as their powers, are allowed to differ across equations. This flexibility is also covered by the `MATLAB` code for the proposed estimators. Zero-mean Gaussian mixture limiting distributions are provided for the two estimators and form the basis for asymptotic standard inference. In addition to hypothesis tests, RESET-type specification tests are provided. One major advantage of the proposed system estimators compared to single equation estimation is the possibility to test for group-wise pooling and – if supported by the test results – perform group-wise pooled estimation which reduces the number of estimated parameters and increases estimation precision. We assess the performance of estimators and tests in finite sample simulations and then

apply the procedures for analyzing the EKC and material Kuznets curve (MKC) hypotheses, with the latter postulating an inverted U-shaped relationship between economic development and the use of materials. For the MKC, we consider relationships between GDP and the use of the three metals aluminum, lead and zinc. We estimate systems of equations on both country level and substance level, i. e. , emission or material use. Similarities in the estimated parameters can be found along both levels. Based on those indications and backed by tests on poolability, we perform group-wise pooling for certain materials and countries. Chapter 3 is a joint work together with Martin Wagner.

The bibliographic details of the three chapters are as follows:

1. Knorre, F., M. Wagner and M. Grupe (2021). Monitoring Cointegrating Polynomial Regressions: Theory and Application to the Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions. A slightly modified version has been published in *Econometrics*, Volume 9, Issue 12, pp. 1–35, https://doi.org/10.3390/econometrics9010012.

2. Knorre, F. and M. Wagner (2025). Comparison of Nonparametric Estimators and Specification Tests for Nonlinear Cointegrating Regressions: Results from a Large Scale Simulation Study. *Unpublished manuscript.*

3. Knorre, F. and M. Wagner (2025). Fully Modified OLS Estimation and Inference for Seemingly Unrelated Cointegrating Polynomial Regressions with Common Integrated Regressors. *Unpublished manuscript.*

All simulations and empirical illustrations have been performed using `MATLAB`. The respective code is available upon request.

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

1.1 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

[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.

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 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 powers 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)

[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.

or Kijima *et al.* (2010) provide survey discussions of the links between economic activity or growth 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

[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 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.

period leads to good fit also for the full period until 2016, with obviously even better fit when re-estimating the relationship over the full sample.

The paper is organized as follows: Section 1.2 contains the setting, assumptions, monitoring statistics and asymptotic results. Section 1.3 discusses finite sample simulation results. Section 1.4 presents the monitoring application to CO_2 and SO_2 emissions data. Section 1.5 briefly summarizes and concludes. Appendix 1.6.1 contains all proofs. Appendix 1.6.2 discusses local asymptotic power properties. Appendix 1.6.3 contains additional finite sample simulation results and Appendix 1.6.4 presents additional empirical results. There is a Supplementary Appendix available upon request (of around 400 pages) which contains 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.

1.2 Theory

1.2.1 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:

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

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

with $x_t := [x_{1t}, \dots, x_{kt}]' \in \mathbb{R}^k$ and $X_t := [x_t', x_{kt}^2, \dots, x_{kt}^{p_k}]' \in \mathbb{R}^p$ 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

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

define the combined parameter vectors $\theta := [\theta'_D, \theta'_X]'$, $\theta_1 := [\theta'_{D,1}, \theta'_{X,1}]' \in \mathbb{R}^{q+p}$.

Under the null hypothesis no structural change occurs, that is $\theta_1 = \theta$ and $\{u_t\}_{t\in\mathbb{Z}}$ 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. , $\{u_t\}_{t\in\mathbb{Z}}$ 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:

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

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

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,*

[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.

a sub-sampling based approach cannot be expected to perform well in a monitoring context.

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 $\int_0^s D(z)D(z)'dz < \infty$ for $0 \leq s \leq 1$, such that for $0 \leq s \leq 1$ it holds that:*

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

If, e. g. , $D_t \coloneqq (1, t, t^2, \ldots, t^{q-1})'$, then $G_D(T) \coloneqq \text{diag}(T^{1/2}, T^{3/2}, T^{5/2}, \ldots, T^{q-1/2})$ 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 $G_X(T) \coloneqq \text{diag}(T \times I_k, T^{3/2}, \ldots, T^{\frac{p_k+1}{2}})$ later.

Assumption 2 *The process $\{\eta_t\}_{t\in\mathbb{Z}} \coloneqq \{[u_t, v_t']'\}_{t\in\mathbb{Z}}$ 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:

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

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 $\sup_{t\geq 1} \mathbb{E}(\|\xi_t\|^a|\mathcal{F}_{t-1}) < \infty$ 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 $\{\eta_t\}_{t\in\mathbb{Z}}$:

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

with the positive definite long-run covariance matrix $\Omega \coloneqq \sum_{j=-\infty}^{\infty} \mathbb{E}(\eta_{t-j}\eta_t')$ and $W(s) \coloneqq [W_{u\cdot v}(s), W_v(s)']'$ a $(k+1)$-dimensional vector of standard Brownian motions. We also define the half long-run covariance matrix $\Delta \coloneqq \sum_{j=0}^{\infty} \mathbb{E}(\eta_{t-j}\eta_t')$. 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 $\omega_{u \cdot v}^2 \coloneqq \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and $\lambda_{uv} \coloneqq \Omega_{uv}(\Omega_{vv}^{1/2})^{-1}$. 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* $\bar{k}(0) \coloneqq \sup_{x \geq 0} |k(x)| < \infty$,

(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

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. , $Z_t \coloneqq [D_t', X_t']'$ and $y_{t,m}^+ \coloneqq y_t - \Delta x_t' \hat{\Omega}_{vv,m}^{-1} \hat{\Omega}_{vu,m}$, 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:

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

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

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., $\hat{\eta}_{t,m} \coloneqq [\hat{u}_{t,m}, v_t']'$, with $\hat{u}_{t,m}$ denoting the OLS residuals here:

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

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

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

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 $\hat{\theta}_m^{\mathrm{D}}$ and $\hat{\Theta}_{j,m}^{\mathrm{D}}$, 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:

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

The OLS estimators of θ and φ from (15) estimated over the calibration sample are referred to as $\hat{\theta}_m^{\mathrm{I}}$ and $\hat{\varphi}_m^{\mathrm{I}}$, respectively. Note that endogeneity correction in the IM-OLS estimator does not require any lead-lag 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. , $\mathbf{W}_v(s) \coloneqq [W_{v_1}(s), W_{v_2}(s), \ldots, W_{v_k}(s), W^2_{v_k}(s), \ldots, W^{p_k}_{v_k}(s)]'$, $J(s) \coloneqq [D(s)', \mathbf{W}_v(s)']'$, $f(s) \coloneqq [\int_0^s D(z)'dz,\ \int_0^s \mathbf{W}_v(z)'dz,\ W_v(s)']'$ and $F(s) \coloneqq \int_0^s f(z)dz$.

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 $\hat{\theta}^F_m$, $\hat{\theta}^D_m$ and $\hat{\theta}^I_m$ as discussed before – denote the FM-OLS residuals by:

$$\hat{u}^F_{t,m} \coloneqq y^+_{t,m} - Z_t'\hat{\theta}^F_m, \tag{16}$$

the D-OLS residuals by:

$$\hat{u}^D_{t,m} \coloneqq y_t - Z_t'\hat{\theta}^D_m - \sum_{j=-d_1}^{d_2} \Delta x'_{t-j}\hat{\Theta}^D_{j,m} \tag{17}$$

and the IM-OLS residuals by:

$$\hat{S}^{u,I}_{t,m} \coloneqq S^y_t - S^{Z\prime}_t\hat{\theta}^I_m - x'_t\hat{\varphi}^I_m. \tag{18}$$

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

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

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

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

[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.

1.2.2 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:

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

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:

$$\begin{aligned}\hat{u}^{\mathrm{F}}_{t,m} &\coloneqq y^{+}_{t,m} - Z_t'\hat{\theta}^{\mathrm{F}}_m \qquad (23)\\ &= u_t - v_t'\hat{\Omega}^{-1}_{vv,m}\hat{\Omega}_{vu,m} - D_t'(\hat{\theta}^{\mathrm{F}}_{D,m} - \theta_D) - X_t'(\hat{\theta}^{\mathrm{F}}_{X,m} - \theta_X)\\ \frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor sT\rfloor}\hat{u}^{\mathrm{F}}_{t,m} &= \frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor sT\rfloor}u_t - \frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor sT\rfloor}v_t'\hat{\Omega}^{-1}_{vv,m}\hat{\Omega}_{vu,m} - \frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor sT\rfloor}D_t'(\hat{\theta}^{\mathrm{F}}_{D,m} - \theta_D) \qquad (24)\\ &\quad - \frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor sT\rfloor}X_t'(\hat{\theta}^{\mathrm{F}}_{X,m} - \theta_X)\end{aligned}$$

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 $\hat{\theta}^{\mathrm{F}}_m \to \theta$, 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:

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

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

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

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

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

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 $\hat{H}_{mov}^{m,n}(s)$ 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 $\hat{H}_{mov}^{m,n}(s)$ 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 $H_{mov,sn}^{m,n}(s)$ 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 $c = c(\alpha, \hat{H}, g, m)$ such that for any $0 < \alpha < 1$ it holds that:*

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

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 1.3 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 1.3 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 over-rejections 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*

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

fulfilled.

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

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

(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*

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

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* $\{\gamma_t\}_{t\in\mathbb{Z}}$ *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*

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

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* $0 < c < \infty$ *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*

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

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 Appendix 1.6.2.[19] The LAP results carry over by and large to similar relative performance findings in the finite sample simulations presented in Section 1.3, 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 10 to 16 in Appendix 1.6.2 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 order powers, the estimation method and – as indicated by Proposition 2 – the detector and for the moving window detectors the window size, the weighting function and the calibration fraction. The Supplementary Appendix, which is available on request, 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 1.4 $T_0 = 71$ 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 1.4. 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,

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

the empirical rejection probabilities under both the null and the alternative. This is the reason 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.*

1.3 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:

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

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

[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 $\theta_{X_{1,2}}$.

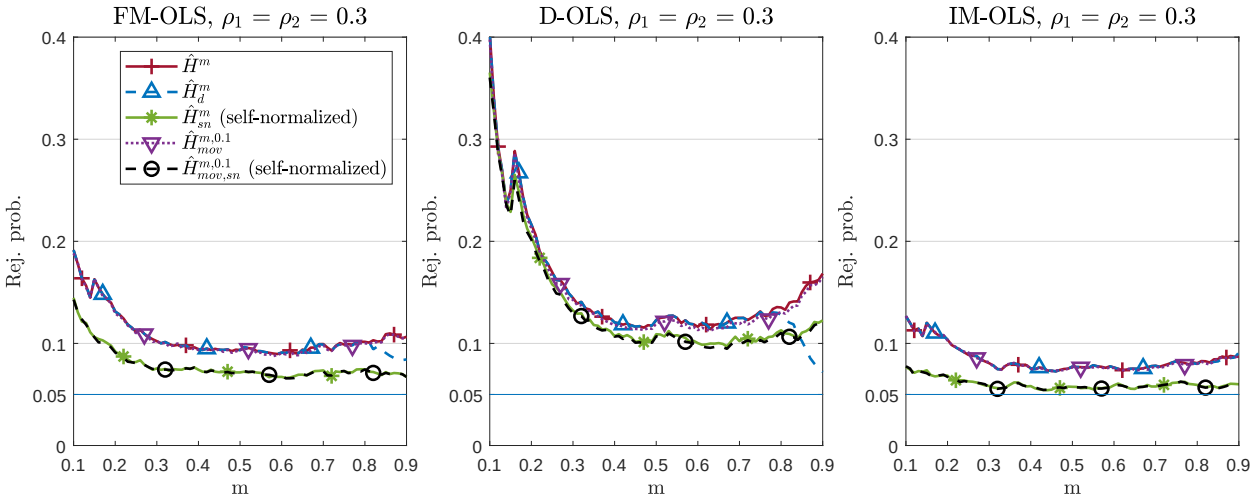

Figure 1: Empirical null rejection probabilities for a grid of values of m, with $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

when using moving window detectors, the choice of the window size has no visible impact on null rejection probabilities.[22]

To assess the performance improvement that can be realized by self-normalization in some detail, Figures 3 and 4 compare $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$ 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

[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 17 and 18 in Appendix 1.6.3. 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.

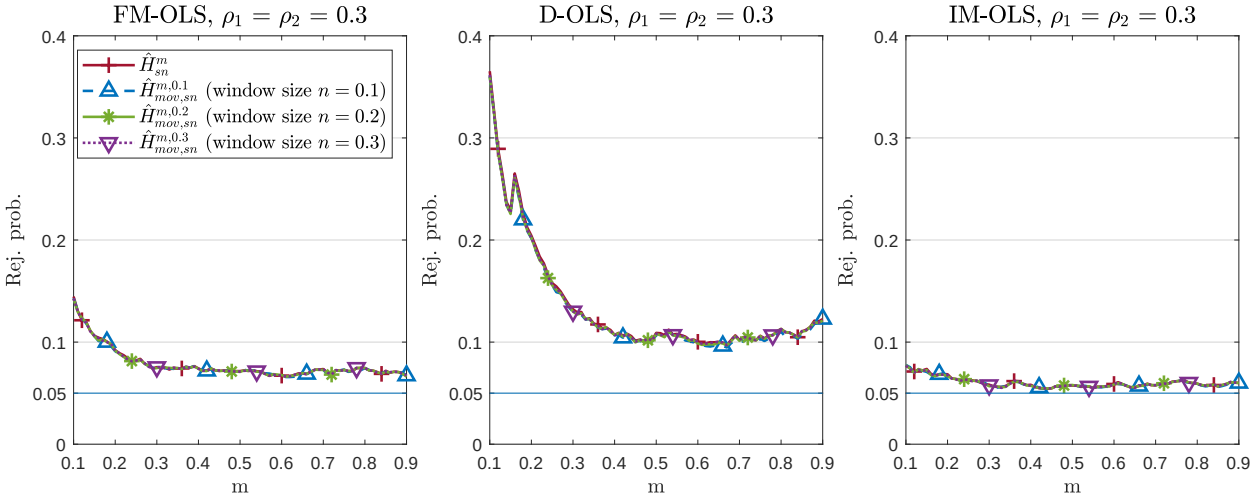

Figure 2: Empirical null rejection probabilities for a grid of values of m, with $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

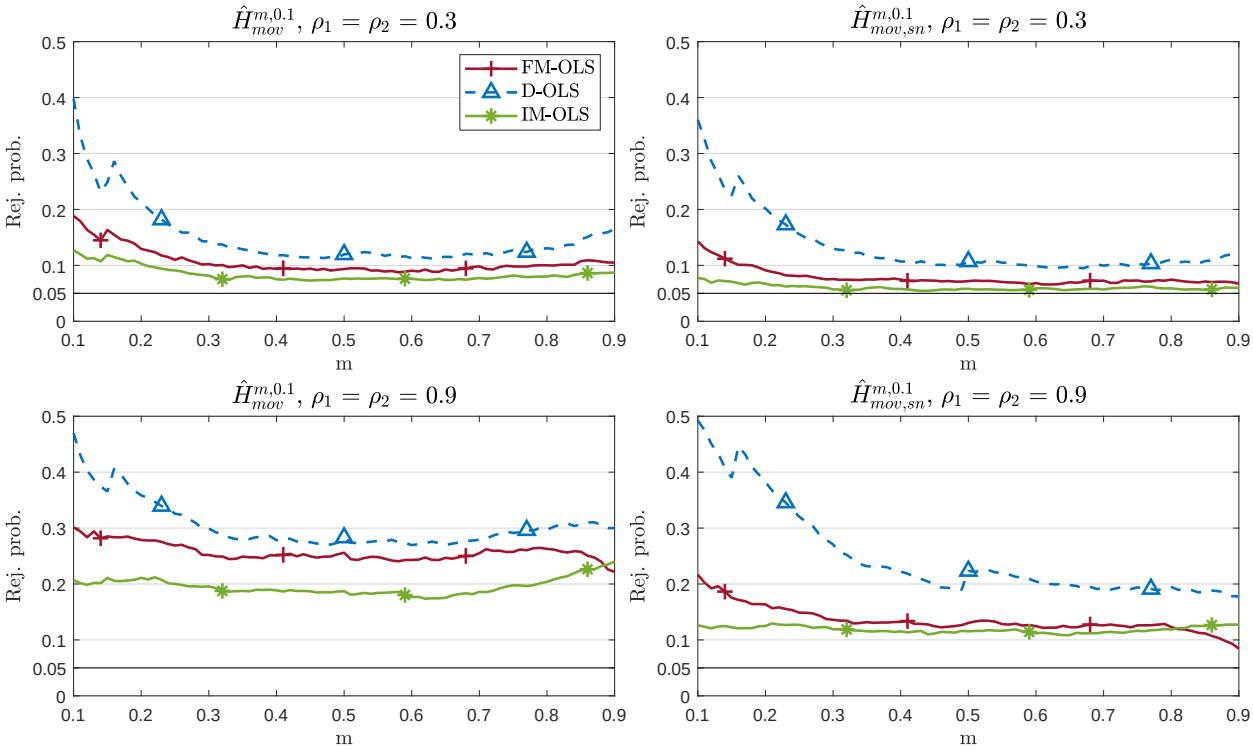

Figure 3: Empirical null rejection probabilities for a grid of values of m and $T = 200$. The left panel displays results for $\hat{H}^{m,0.1}_{mov}$. The right panel displays results for $\hat{H}^{m,0.1}_{mov,sn}$.

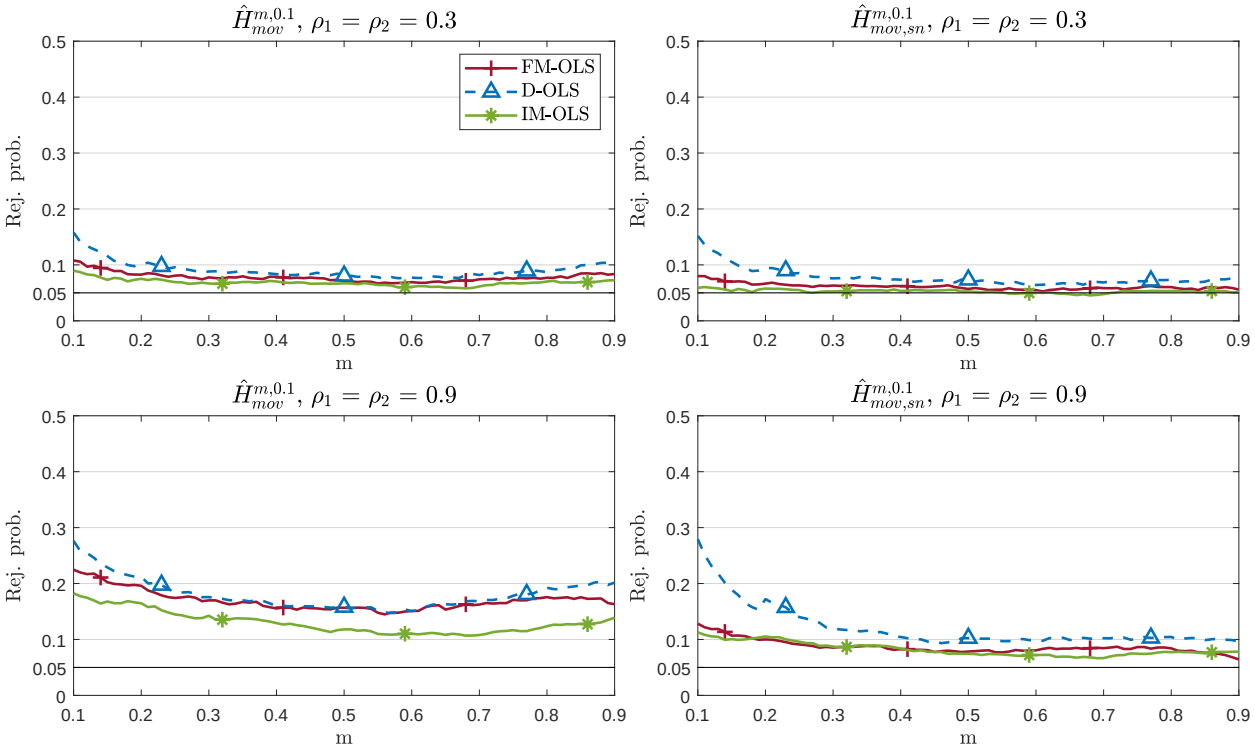

Figure 4: Empirical null rejection probabilities for a grid of values of m and $T = 500$. The left panel displays results for $\hat{H}^{m,0.1}_{mov}$. The right panel displays results for $\hat{H}^{m,0.1}_{mov,sn}$.

in θ_D or θ_X, respectively, are discussed in Appendix 1.6.3.[24] Size-corrected power simulations 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 Appendix 1.6.3 – many observations are qualitatively similar. One difference is that in case of trend breaks the detectors $\hat{H}^m$ or $\hat{H}^m_d$ 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.

[24] See Figures 19 to 27 in Appendix 1.6.3 for size-corrected power results in case of trend and slope breaks.

[25] Tables 9 to 14 in Appendix 1.6.3 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
$\hat{H}_d^m$	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
$\hat{H}_{sn}^m$	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
$\hat{H}_{mov}^{m,0.1}$	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
$\hat{H}_{mov}^{m,0.2}$	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
$\hat{H}_{mov}^{m,0.3}$	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
$\hat{H}_{mov,sn}^{m,0.1}$	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
$\hat{H}_{mov,sn}^{m,0.2}$	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
$\hat{H}_{mov,sn}^{m,0.3}$	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 $\rho_1 = \rho_2 = 0.3$.

It remains to investigate the detection times and delays. We consider again the case of I(1) breaks 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 29 – 34 in Appendix 1.6.3. The figures display the results in the form of box-whiskers plots – for completeness – for all nine considered detectors. The numbers below the detector labels indicate 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 Appendix 1.6.3, 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, when comparing self-normalized monitoring statistics with their standardized versions, self-normalization leads to smaller over-rejections, without having any systematic or sizeable negative impacts on size-corrected power (for $m \leq r$); and without leading to larger delays when compared to standardization. Second, within the group of self-normalized detectors, the moving window detector with $n = 0.1$ leads almost throughout to highest size-corrected power without detrimental effects on null rejection probabilities. For the moving window detectors size-corrected power decreases with increasing window size. For all moving window detectors, size-corrected power is at least as good or higher than for the expanding window detector. Some exceptions occur in case of trend breaks, in which case the expanding window detector sometimes outperforms the moving window detectors in terms of size-corrected power. In addition, the window size exhibits some impact on the delays, with the moving window detector with $n = 0.1$ leading to the shortest delays. Third, with respect to estimator choice the usual trade-off between IM-OLS on the one hand and FM-/D-OLS on the

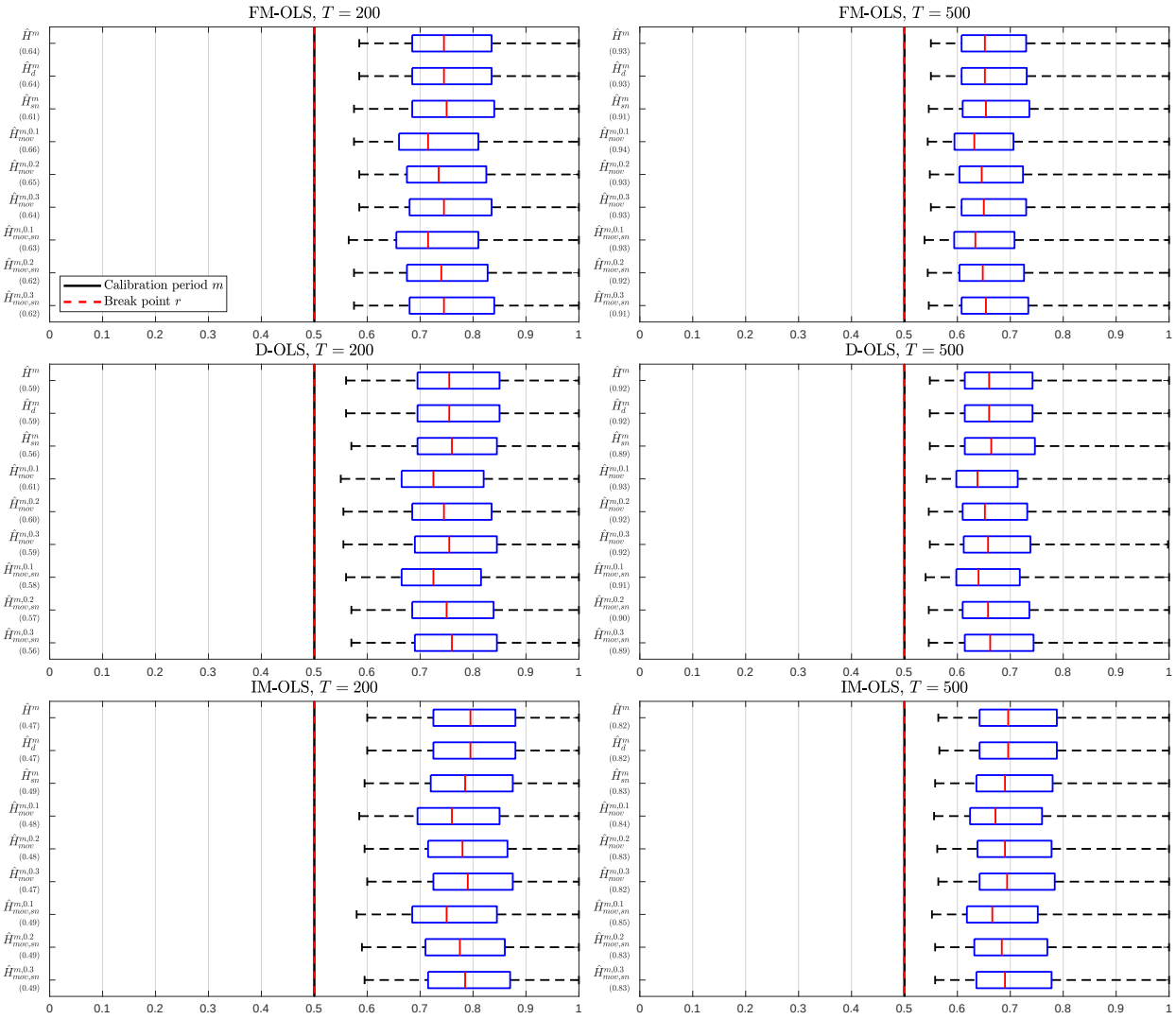

Figure 5: Detection times against $I(1)$ breaks for $m = 0.5$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.3$.

other occurs. IM-OLS leads to lower over-rejections at the cost of lower size-corrected power and slightly larger delays than FM-OLS, with D-OLS outperformed by FM-OLS in terms of larger over-rejections by D-OLS under the null hypothesis. Size-corrected power and detection times based on D-OLS are by and large very similar compared to those based on FM-OLS. 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 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.

Altogether, we recommend to use the self-normalized moving window detector with $n = 0.1$, i. e. , $\hat{H}^{m,0.1}_{mov,sn}$. The choice of the estimator is dictated by the trade-off between null rejection probabilities on target and high size-corrected power.

1.4 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 Appendix 1.6.4 in Table 15 for the calibration period and in Table 16 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 35 and 36 in Appendix 1.6.4. 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* of the two

[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 (2023). The detailed test results are available in Tables 17 and 18 in Appendix 1.6.4.

		$\hat{H}^{m,0.1}_{mov,sn}$	
Country	p	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 $\hat{H}^{m,0.1}_{mov,sn}$ 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%.

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 $\hat{H}^{m,0.1}_{mov,sn}$ with window size $n = 0.1$, i. e. , seven years.[33]

1.4.1 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 Appendix 1.6.4, in Table 22 for CO_2 emissions and in Table 23 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.

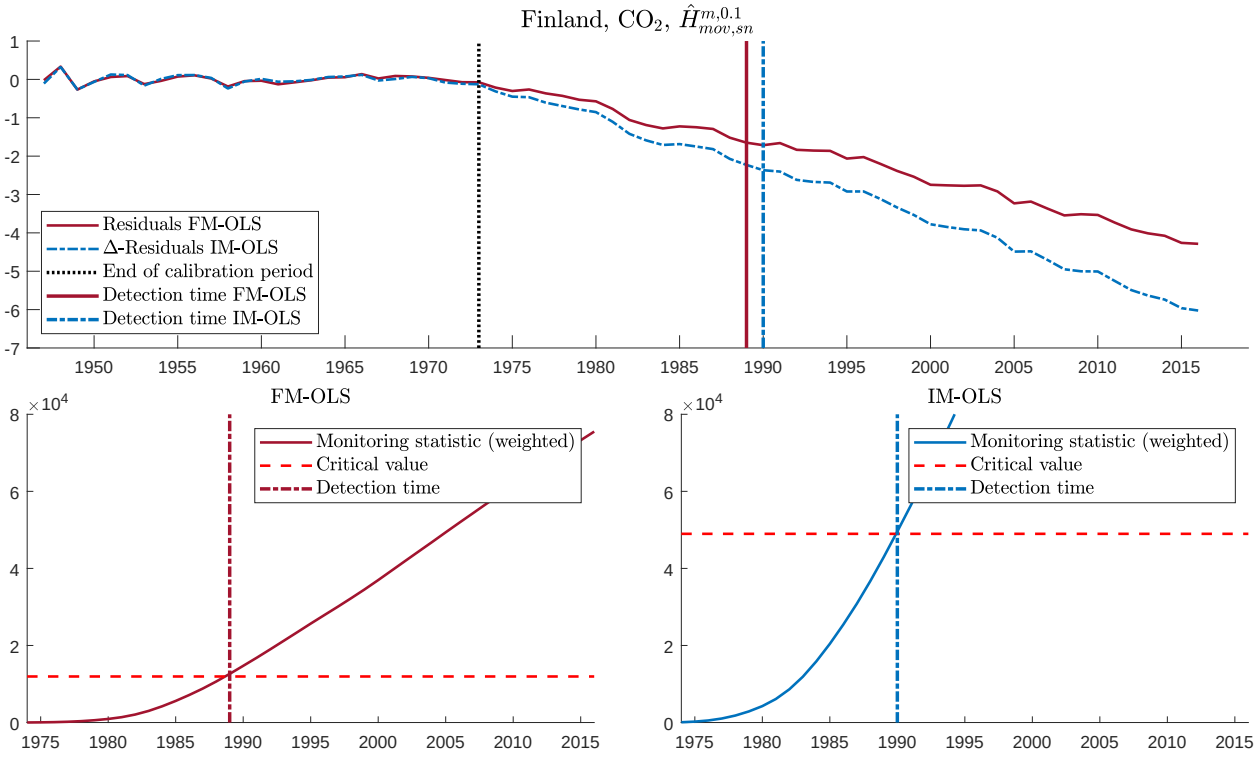

Figure 6: Monitoring results for CO_2 emissions of Finland using $\hat{H}^{m,0.1}_{mov,sn}$ with both FM-OLS and IM-OLS in the quadratic specification. The lower panel shows the detectors, the critical values and the detection times for FM-OLS on the left hand side and for IM-OLS on the right hand side.

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]

As an illustration, Figure 6 displays the mechanics of the monitoring procedure for Finland.The

[34]Considering a quadratic specification over the full sample period, see Table 24 in Appendix 1.6.4 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.

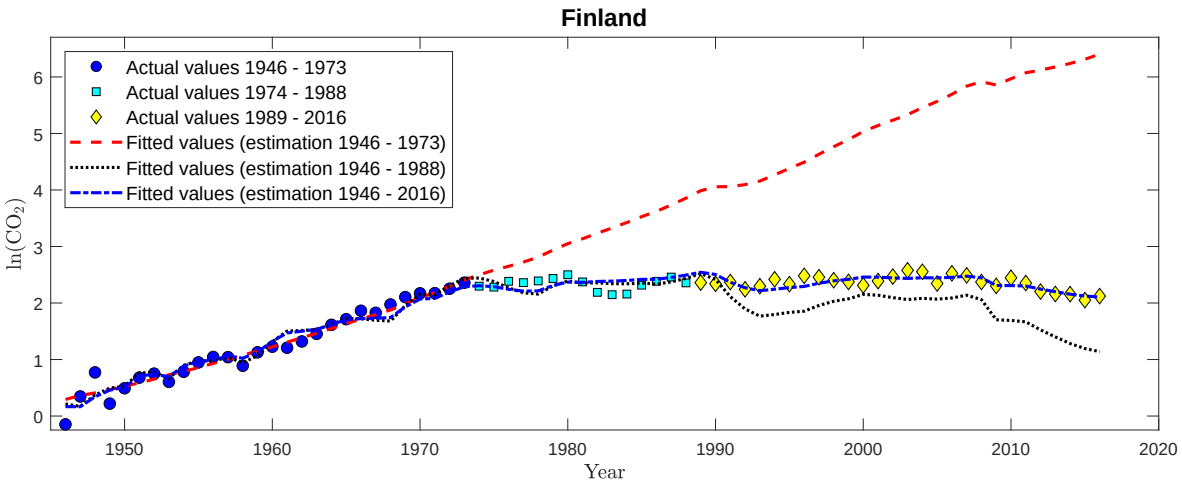

Figure 7: FM-OLS estimation results for a cointegrating quadratic relationship between log per capita GDP and log per capita CO_2 emissions for Finland. The figure shows pairs of observations of log per capita GDP and log per capita CO_2 emissions, for 1946–1973 (blue circles), 1974–1988 (turquoise squares) 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 line 1946–2016.

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 estimation takes place over the full sample.[35] 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

[35]See Figure 37 in Appendix 1.6.4 for analogous – and very similar – IM-OLS results.

	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.

null hypothesis rejected at the 5% significance level.[36] 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 to 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.[37]

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.

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

[36]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.

[37]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.

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.[38] 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.[39] 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.

[38] Figure 38 in Appendix 1.6.4 shows the corresponding results obtained with IM-OLS.

[39] Table 19 in Appendix 1.6.4 contains similar results concerning the minimal polynomial degrees of a CPR relationship for the full sample period as Table 4 for the calibration period.

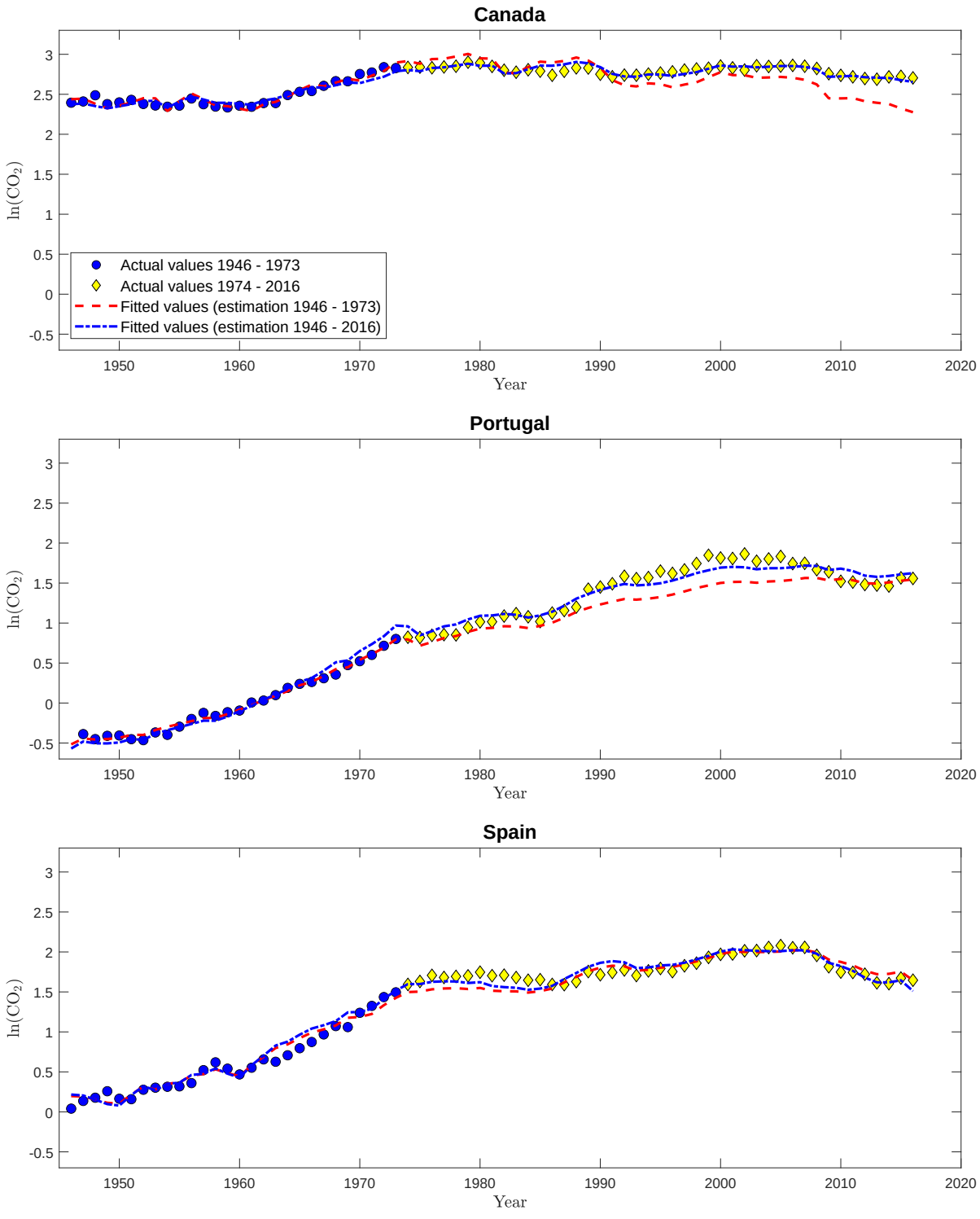

Figure 8: FM-OLS estimation results for a cointegrating linear relationship between log per capita GDP and log per capita CO_2 emissions for Canada, Portugal and Spain. The figure shows pairs of observations of log per capita GDP and log per capita CO_2 emissions, for 1946–1973 (blue circles) and 1974–2016 (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 line 1946–2016.

		$\hat{H}^{m,0.1}_{mov,sn}$	
Country	p	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 $\hat{H}^{m,0.1}_{mov,sn}$ 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%.

1.4.2 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 $\hat{H}^{m,0.1}_{mov,sn}$ for both FM-OLS and IM-OLS.

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.[40]

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 estimation leads to a break being detected. The results lead to some interesting observations.

[40]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.

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.[41] 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 linear relationship for Australia and Portugal or with a cubic relationship for the US.[42]

[41] 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.

[42] Figure 39 in Appendix 1.6.4 shows similar results when using IM-OLS for parameter estimation.

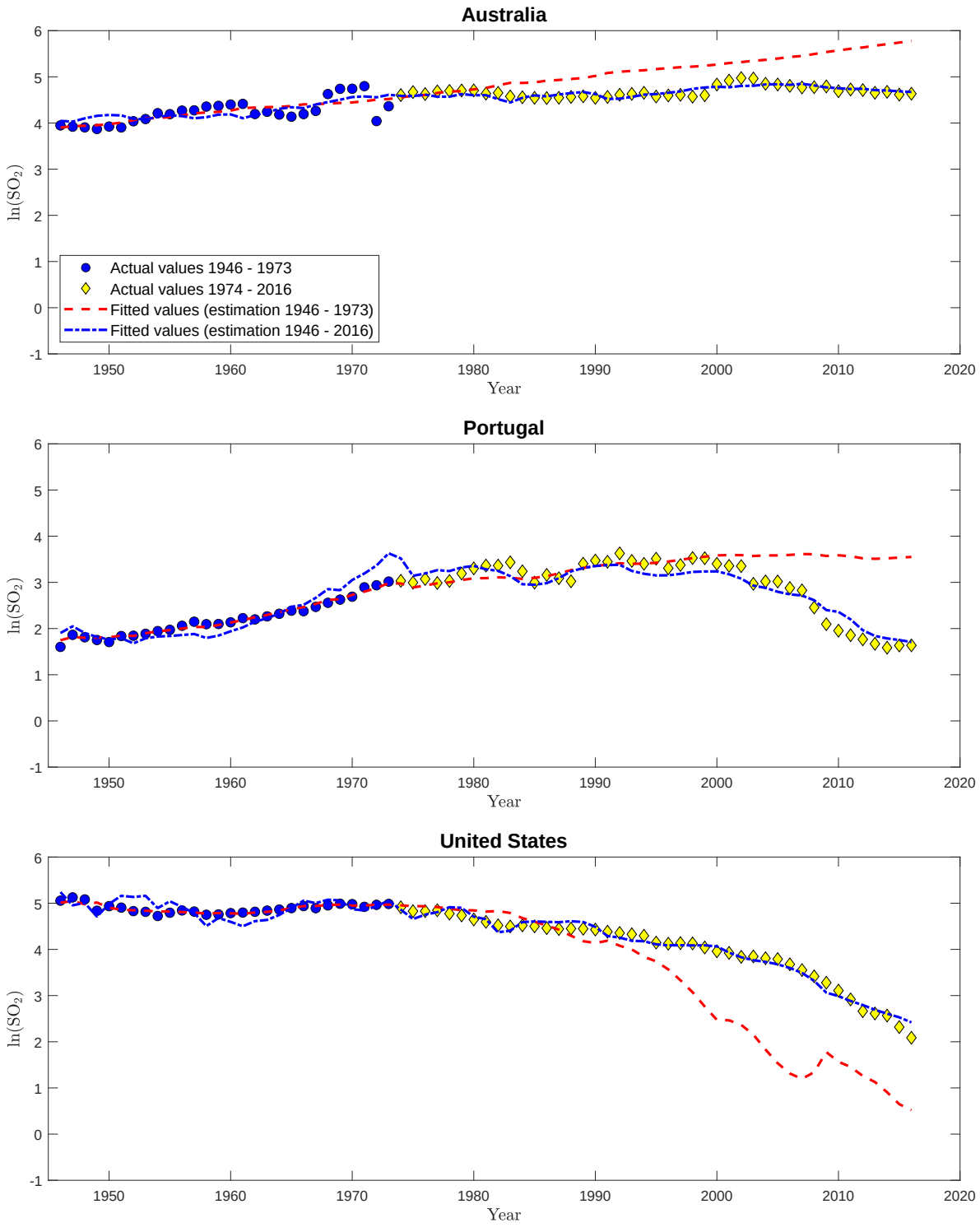

Figure 9: FM-OLS estimation results for a cointegrating linear relationship between log per capita GDP and log per capita CO_2 emissions for Australia and Portugal, and a cointegrating cubic relationship for the United States. The figure shows pairs of observations of log per capita GDP and log per capita 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 line 1946–2016.

Considering the full sample period leads – similar to CO_2 emissions – to a CPR relationship for all nine countries where a structural break 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.[43]

1.5 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 is the extension of the Shin (1994) test statistic for the null hypothesis of linear cointegration to the CPR setting (see, e. g. Wagner, 2023). 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. 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-

[43] See Table 19 in Appendix 1.6.4 for the minimal polynomial degrees on the full sample.

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.

1.6 Appendix

1.6.1 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:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor sT\rfloor}\hat{u}^{\mathrm{F}}_{t,m} = \frac{1}{\sqrt{T}}\sum_{t=2}^{\lfloor rT\rfloor}\hat{u}^{\mathrm{F}}_{t,m} + \frac{1}{\sqrt{T}}\sum_{t=\lfloor rT\rfloor+1}^{\lfloor sT\rfloor}u_t - \frac{1}{\sqrt{T}}\sum_{t=\lfloor rT\rfloor+1}^{\lfloor sT\rfloor}v_t'\hat{\Omega}^{-1}_{vv,m}\hat{\Omega}_{vu,m} \\ - \frac{1}{\sqrt{T}}\sum_{t=\lfloor rT\rfloor+1}^{\lfloor sT\rfloor}D_t'(\hat{\theta}^{\mathrm{F}}_{D,m}-\theta_D) - \frac{1}{\sqrt{T}}\sum_{t=\lfloor rT\rfloor+1}^{\lfloor sT\rfloor}X_t'(\hat{\theta}^{\mathrm{F}}_{X,m}-\theta_X) \\ - \frac{1}{\sqrt{T}}\sum_{t=\lfloor rT\rfloor+1}^{\lfloor sT\rfloor}D_t'(\theta_D-\theta_{D,1}) - \frac{1}{\sqrt{T}}\sum_{t=\lfloor rT\rfloor+1}^{\lfloor sT\rfloor}X_t'(\theta_X-\theta_{X,1}).\end{aligned} \tag{46}$$

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:

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

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

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

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:

$$\begin{aligned}\frac{1}{\sqrt{T}}\hat{S}^{u}_{\lfloor sT\rfloor,m} &\Rightarrow \omega_{u\cdot v}\widetilde{Q}_{u\cdot v,m}(s) + \int_{r}^{s}\mathbf{B}_{v}(z)'dz\Delta_{\theta_X} \\ &= \omega_{u\cdot v}\widetilde{Q}_{u\cdot v,m}(s) + \int_{r}^{s}\mathbf{W}_{v}(z)'dz\mathbf{\Omega}_{vv}^{1/2\prime}\Delta_{\theta_X},\end{aligned} \tag{49}$$

with $\mathbf{\Omega}_{vv}^{1/2} \coloneqq \mathrm{diag}(\Omega_{vv}^{1/2}, \omega_k^2, \ldots, \omega_k^{p_k})$ 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 the case considered equal to $\delta\omega_{\gamma}\int_{r}^{s}\left(W_{\gamma}(z) - W_{\gamma}(r)\right)dz$, $\int_{r}^{s} D(z)'dz\Delta_{\theta_D}$ and $\int_{r}^{s}\mathbf{W}_{v}(z)'dz\mathbf{\Omega}_{vv}^{1/2\prime}\Delta_{\theta_X}$. 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 $\hat{H}^{m,n}_{mov,sn}$ and the case of a local I(1) break, in which case one obtains for $0 < m \leq r < s \leq 1$:

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

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. □

1.6.2 Local Asymptotic Power

In this appendix we consider local asymptotic power (LAP) in models of the form in (1) in the main text with the following specifications: Intercept and intercept plus linear trend, up to four integrated regressors and consecutive powers of up to order three of one of the integrated regressors. Denote as in the main text, w.l.o.g., x_{kt} as the I(1) regressor considered to enter the regression model with powers larger than one. The LAP results are based on 10,000 replications using time series of length 1,000 of i.i.d. standard normals to approximate the limiting distributions. The limiting distributions underlying LAP, as discussed in Proposition 3(b), are based on the functional central limit theorems for the scaled partial sum residual processes under the local alternatives discussed in Appendix 1.6.1. In particular these FCLTs are given in:

- Local I(1) breaks ($\{u_t\}$ changes its behavior from I(0) to I(1)): Equation (47) in Appendix 1.6.1.

- Local trend breaks: Equation (48) in Appendix 1.6.1.
- Local slope breaks: Equation (49) in Appendix 1.6.1.

For local I(1) breaks the limiting distribution depends, in addition to the break magnitude parameter δ, on the ratio of the long-run variance ω_γ^2 of the additional error component $\{\gamma_t\}$ and the conditional long-run variance $\omega_{u\cdot v}^2$.[44] We set this ratio to $\frac{\omega_\gamma^2}{\omega_{u\cdot v}^2} \coloneqq 1$. Similarly, the limiting distribution in case of local slope breaks depends on the long-run covariance matrix $\mathbf{\Omega}_{vv}$ of the regressors. We set $\mathbf{\Omega}_{vv} \coloneqq I_k$. The parameter δ determines the magnitude of a local I(1) break and varies in 1001 equidistant steps over the interval $[0, 10000]$. The value $\delta = 0$ corresponds to the null hypothesis. Similarly, Δ_{θ_D} indicates the magnitude of a local break in the coefficient of the linear trend.[45] The parameter Δ_{θ_D} varies in 751 equidistant steps over the interval $[0, 1500]$, with $\Delta_{\theta_D} = 0$ corresponding to the null hypothesis. In case of local slope breaks, the magnitude of a local break in the slope parameter is given by Δ_{θ_X}.[46] We consider 1001 equally spaced values of Δ_{θ_X} in the interval $[0, 10000]$, with $\Delta_{\theta_X} = 0$ corresponding to the null hypothesis. For local slope breaks we consider a local break either in the coefficient corresponding to x_{kt} or in the coefficient corresponding to x_{kt}^2. All simulations are performed at the 5% significance level.

With respect to the stochastic regressors, LAP depends upon k, the number of I(1) regressors, and p_k, the highest power of x_{kt}. Figure 10 displays for different values of k and p_k LAP against local I(1) breaks in the intercept and linear trend case using $\hat{H}^m$. As expected, LAP decreases when further I(1) regressors or higher powers are included. With this observation in mind, for the sake of brevity all other LAP results are presented only for the case including a single I(1) regressor and its square (i. e. , $k = 1$ and $p_k = 2$), considering both deterministic specifications, intercept and intercept and linear trend. This case also covers the quadratic EKC model, which we use in addition to the linear and cubic model in the empirical application in Section 1.4 in the main text.

In most considered settings the two detectors $\hat{H}^m$ and $\hat{H}_d^m$ lead to nearly identical LAP. In a few cases, however, LAP of $\hat{H}_d^m$ is considerably lower than that of $\hat{H}^m$. Therefore, we neglect $\hat{H}_d^m$ in

[44] Note: The limiting distribution underlying LAP in case of local I(1) breaks in Equation (30) of Wagner and Wied (2017) contains a typo. The long-run variance ω^2 must be replaced by the *conditional* long-run variance $\omega_{u\cdot v}^2$. This also holds for the corresponding discussion in the main text and in Appendix B, where ω_ξ/ω must be replaced by $\omega_\xi/\omega_{u\cdot v}$.

[45] In Proposition 3(b) Δ_{θ_D} refers to a vector defining changes of all elements of $\theta_D \in \mathbb{R}^q$, with $\Delta_{\theta_D} \in \mathbb{R}^q$. In the LAP analysis only a single element of θ_D changes. To avoid overloaded notation, Δ_{θ_D} denotes in the LAP analysis only the change in a single element, thus $\Delta_{\theta_D} \in \mathbb{R}$.

[46] The parameter Δ_{θ_X} refers in the LAP analysis – in a similar fashion as Δ_{θ_D} – to a scalar value, in contrast to Proposition 3(b) where $\Delta_{\theta_X} \in \mathbb{R}^p$.

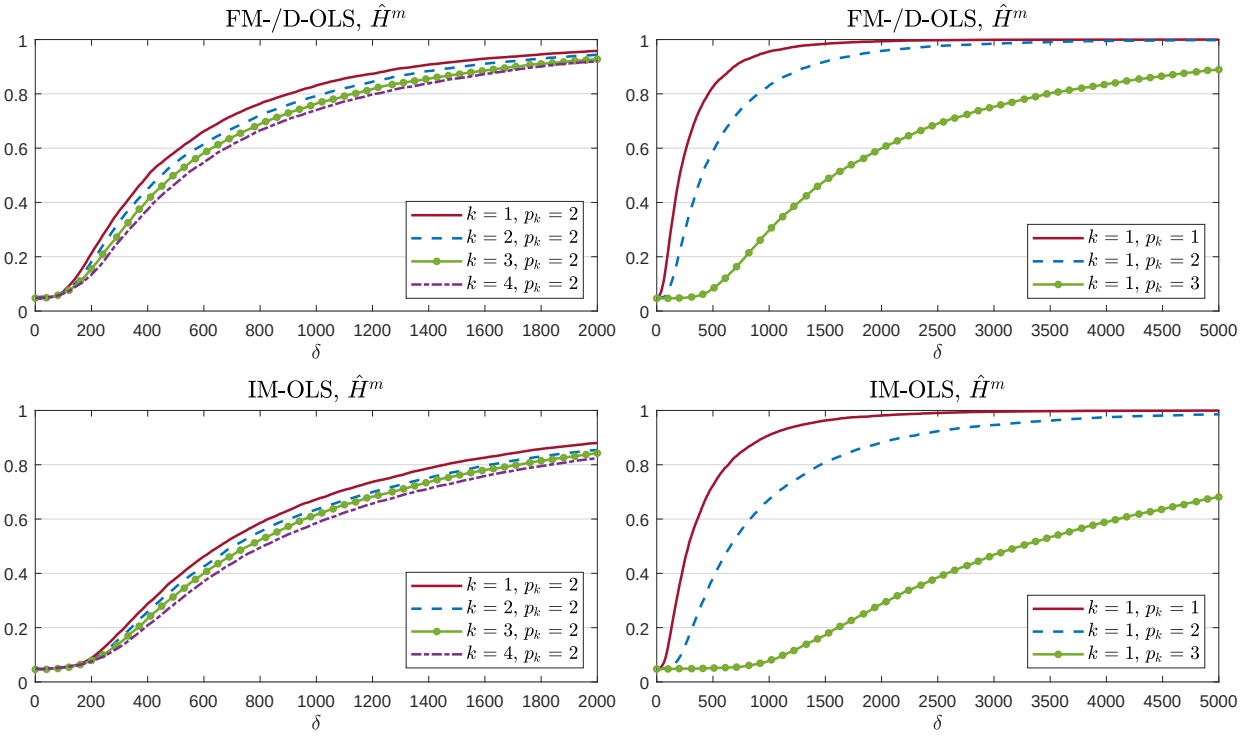

Figure 10: Local asymptotic power against I(1) breaks in the intercept and linear trend case for $\hat{H}^m$. The upper two plots correspond to FM-/D-OLS and the lower plots to IM-OLS. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively.

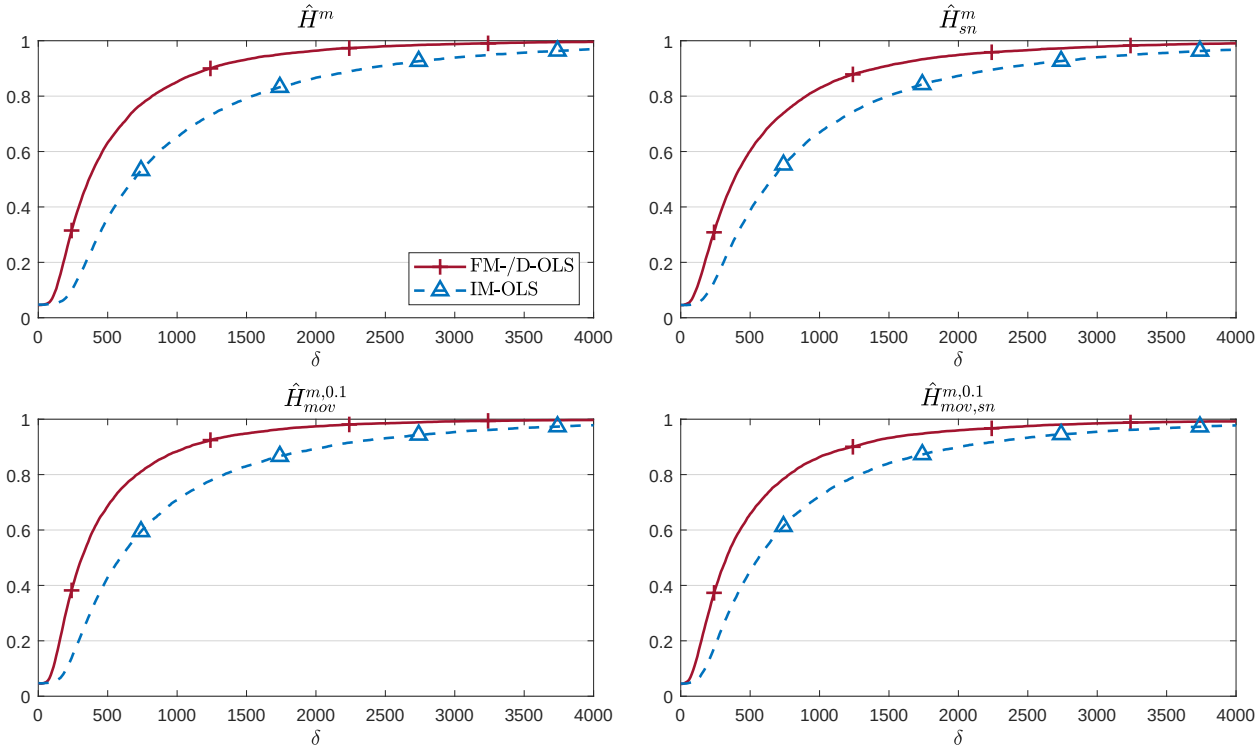

Figure 11: Local asymptotic power against I(1) breaks in the intercept only case for FM-/D-OLS and IM-OLS using $\hat{H}^m$, $\hat{H}^m_{sn}$, $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively.

our LAP analysis and focus on $\hat{H}^m$, $\hat{H}^m_{sn}$, $\hat{H}^{m,n}_{mov}$ and $\hat{H}^{m,n}_{mov,sn}$, for $n = 0.1, 0.2, 0.3$.

We start with local I(1) breaks and compare LAP with respect to the estimation method. Figure 11 displays LAP against local I(1) breaks in the intercept only case for FM-/D-OLS and IM-OLS using the four detectors $\hat{H}^m$, $\hat{H}^m_{sn}$, $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively. For each of the four detectors FM-/D-OLS leads to higher LAP than IM-OLS. This result is independent of the choice of the window size, deterministic specification, calibration and break fraction and is in line with standard asymptotic theory concerning estimator efficiency. The next aspect considered is the impact of the window size on LAP and in addition the comparison to a detector without moving window. Figure 12 displays LAP against local I(1) breaks in the intercept only case for $\hat{H}^m$ and $\hat{H}^{m,n}_{mov}$, and their corresponding self-normalized counterparts $\hat{H}^m_{sn}$ and $\hat{H}^{m,n}_{mov,sn}$, for $n = 0.1, 0.2, 0.3$. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively. LAP is highest for the smallest window size $n = 0.1$, decreases with increasing window size and is lowest for the detector without moving window. This result is independent of estimator choice, self-normalization and also the choices of calibration and

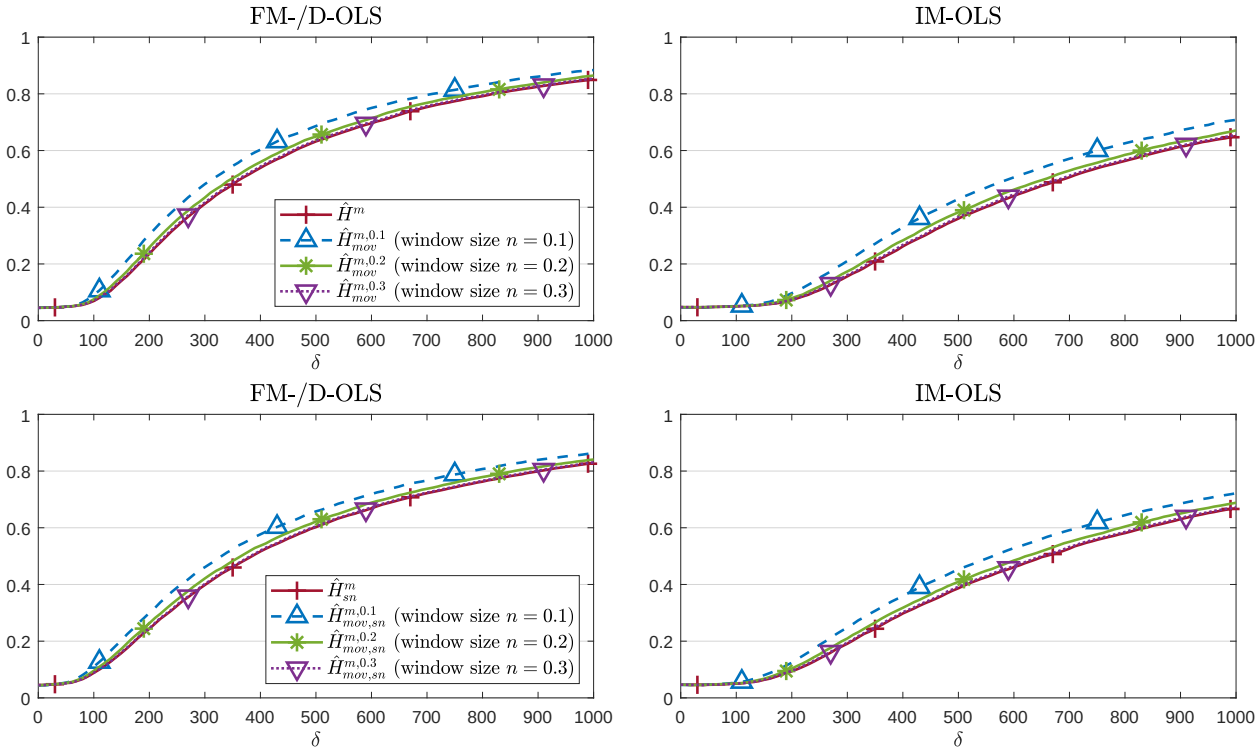

Figure 12: Local asymptotic power against I(1) breaks in the intercept only case for $\hat{H}^m$, $\hat{H}^m_{sn}$, $\hat{H}^{m,n}_{mov}$ and $\hat{H}^{m,n}_{mov,sn}$, for $n = 0.1, 0.2, 0.3$. The upper two plots correspond to standardized detectors, the lower two plots correspond to self-normalized detectors. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively.

break fraction. However, the differences in LAP between the detectors are sometimes very small. Including a linear trend in the model does not cause any qualitative changes in the above results. The last aspect we consider for LAP against local I(1) breaks is the impact of self-normalization in comparison to standardization. Figure 13 shows the results for $\hat{H}^m_{sn}$ and $\hat{H}^{m,0.1}_{mov,sn}$, each compared to its standardized counterpart $\hat{H}^m$ and $\hat{H}^{m,0.1}_{mov}$, respectively, in case of local I(1) breaks with intercept only. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$. Self-normalization by and large reduces LAP for FM-/D-OLS and increases LAP for IM-OLS. Figure 14 displays the corresponding results if additionally a linear trend is included in the model. In this case, when using IM-OLS the advantage of self-normalization over standardization disappears. Changes in the window size or calibration and break fraction have no qualitative influence on the differences in LAP between standardized and self-normalized detectors.

We now turn briefly to LAP against local slope breaks. The results are qualitatively the same as in case of local I(1) breaks with respect to the estimation method, self-normalization, window

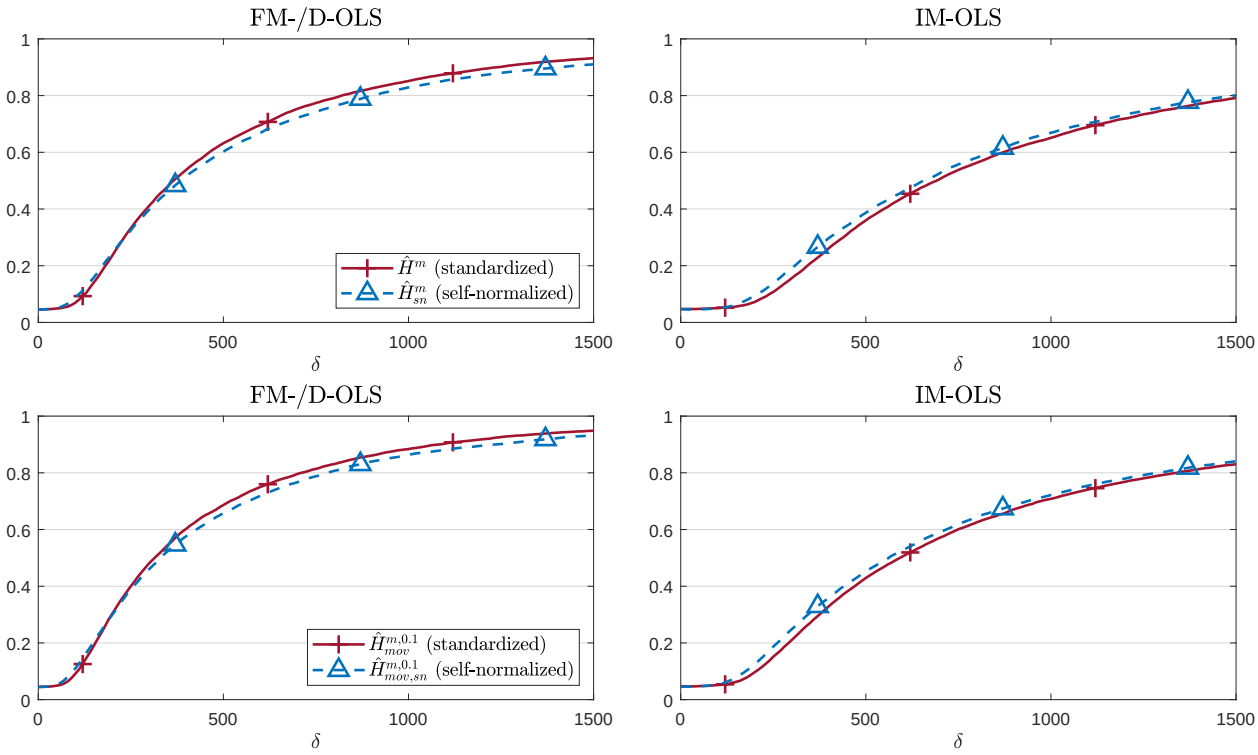

Figure 13: Local asymptotic power against local I(1) breaks in the intercept only case. The upper two plots correspond to $\hat{H}^m$ and $\hat{H}^m_{sn}$. The lower two plots correspond to $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively.

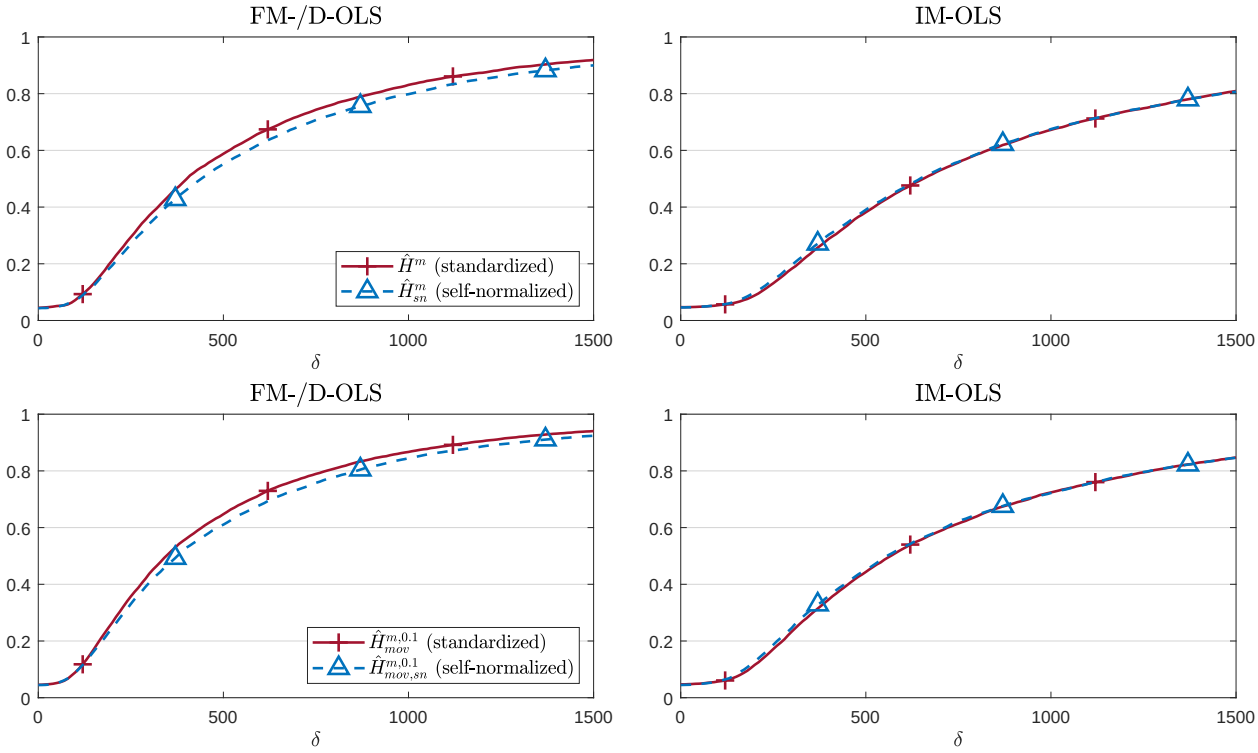

Figure 14: Local asymptotic power against local I(1) breaks in the intercept and linear trend case. For further explanations see notes for Figure 13.

size and deterministic specification and are independent of the calibration and break fraction. Corresponding LAP figures are available upon request.

Finally, we consider LAP in case of local trend breaks. The overall results are very similar to both cases, local I(1) and local slope breaks, but with two major differences: The first difference is that now calibration and break fraction influence which detector leads to the highest LAP. Figure 15 displays LAP for standardized detectors for $m = 0.5$ and $r \in \{0.5, 0.75\}$, with either $\hat{H}^m$ or $\hat{H}^{m,0.1}_{mov}$ leading to the highest LAP, depending on the choice of m and r. We obtain similar results for self-normalized detectors, with either $\hat{H}^m_{sn}$ or $\hat{H}^{m,0.1}_{mov,sn}$ leading to the highest LAP. The second difference in case of local trend breaks is the impact of self-normalization compared to standardization on LAP for IM-OLS. Figure 16 displays LAP in case of local trend breaks for $\hat{H}^m$, $\hat{H}^m_{sn}$, $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$, with $m = 0.5$ and $r = 0.75$. For IM-OLS, the two self-normalized detectors, each compared to its standardized counterpart, lead to higher LAP at smaller break magnitudes at the expense of lower LAP at larger break magnitudes. In contrast, in case of local I(1) or local slope breaks, self-normalization either leads to higher or nearly the same LAP as standardization for IM-OLS, regardless of the break magnitude. For FM-/D-OLS, standardization still leads to higher LAP than self-normalization, as is the case with local I(1) or local slope breaks.

In conclusion, the LAP results can be summarized as follows: First, given a certain detector, break type, calibration and break fraction, LAP is higher for FM-/D-OLS than for IM-OLS. This stems from the fact that for the linear case FM-/D-OLS is asymptotically conditionally more efficient than IM-OLS from a standard asymptotic theory point of view (see Vogelsang and Wagner, 2014a, Proposition 2). A similar result can be obtained for the CPR case. Second, moving window detectors with $n = 0.1$ are the best choice in case of local I(1) and local slope breaks. In case of local trend breaks there is no clear ranking. Third, self-normalization by and large reduces LAP for FM-/D-OLS. In contrast, for IM-OLS self-normalization is by and large advantageous, in case of local trend breaks at least for small break magnitudes. Fourth, local slope breaks lead, as expected, to almost identical LAP patterns as local I(1) breaks. Wagner and Wied (2017, p. 967) explain for the cointegrating linear regression model that the asymptotic behavior in case of slope breaks is similar to the case of I(1) breaks. The same holds for the CPR model when there is only a (local) break in the coefficient corresponding to one I(1) regressor: The slope coefficient after the break, $\theta_{X,1}$, can be separated into θ_X and $\theta_{X,1} - \theta_X$. The additional post break term $X'_t(\theta_{X,1} - \theta_X)$ is an I(1) process – since the coefficient corresponding to one I(1) regressors changes – and thus leads to similar asymptotic behavior as in case of local I(1) breaks, where $\{u_t\}$ changes its behavior from

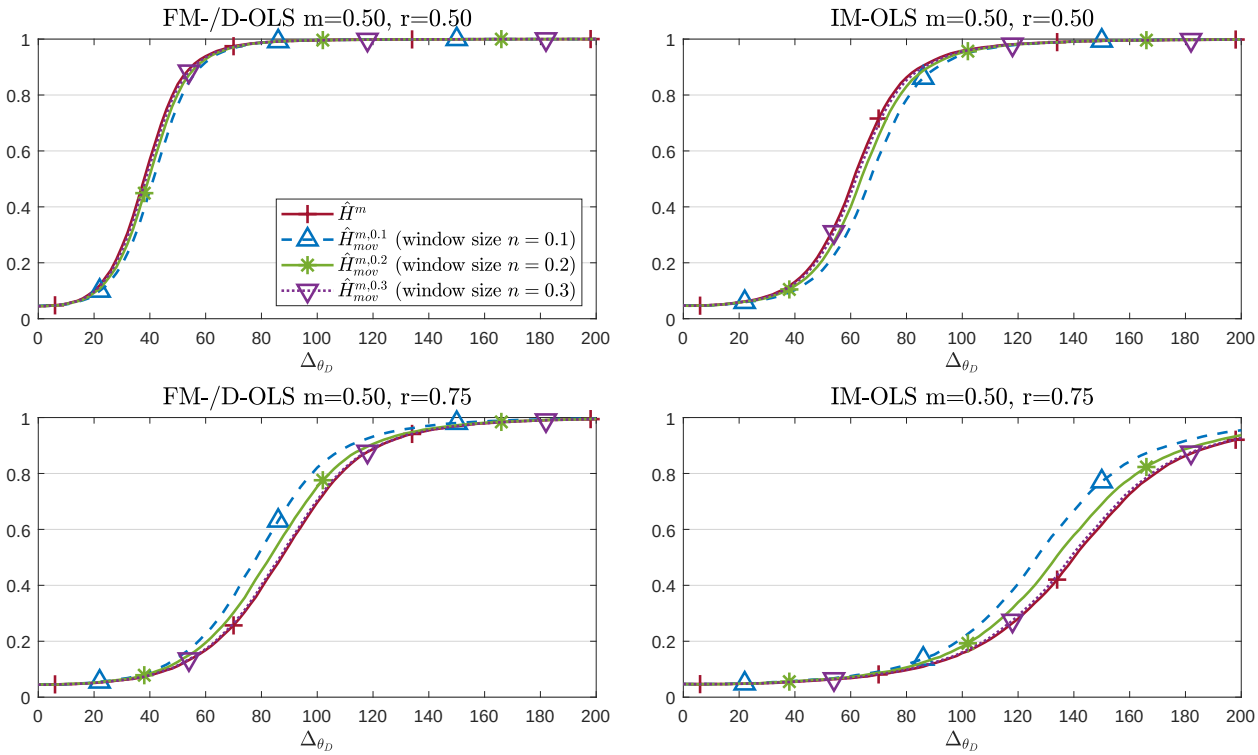

Figure 15: Local asymptotic power against local trend breaks for $\hat{H}^m$ and $\hat{H}^{m,n}_{mov}$ in the intercept and linear trend case. The upper two plots correspond to $m = r = 0.5$. The lower two plots correspond to $m = 0.5$ and $r = 0.75$.

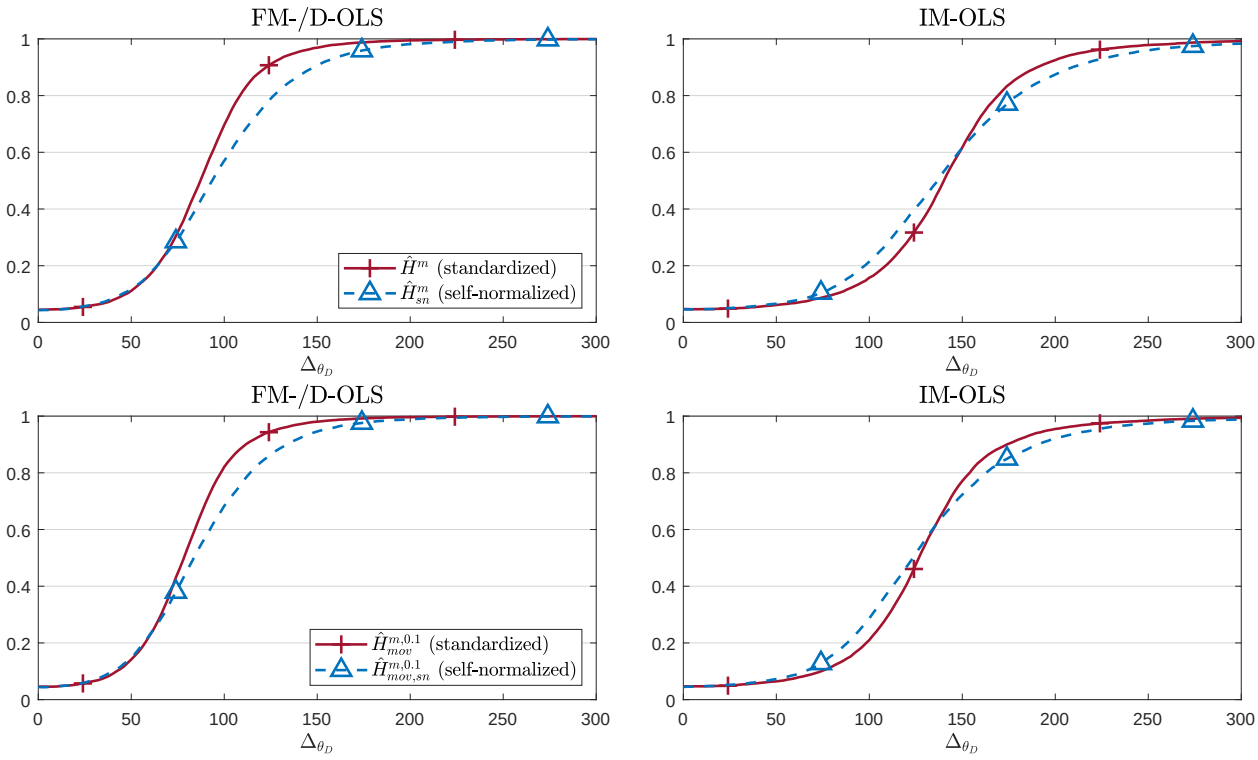

Figure 16: Local asymptotic power against local trend breaks in the model with intercept and linear trend. The upper two plots correspond to $\hat{H}^m$ and $\hat{H}^m_{sn}$. The lower two plots correspond to $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively.

I(0) to I(1). Fifth, $\hat{H}^m$ is preferable to $\hat{H}_d^m$, since $\hat{H}^m$ performs either better or at least as good as $\hat{H}_d^m$ in all considered cases. Finally, regardless of the choice of estimation method and detector, the LAP results suggest to choose the calibration fraction m as large as possible.

The LAP analysis reveals indications about the behavior of the monitoring procedures in finite samples. In fact, we find by and large very similar relative performance results in the finite sample simulations in Section 1.3 in the main document as in the LAP analysis.

1.6.3 Additional Finite Sample Performance Results

Null Rejection Probabilities

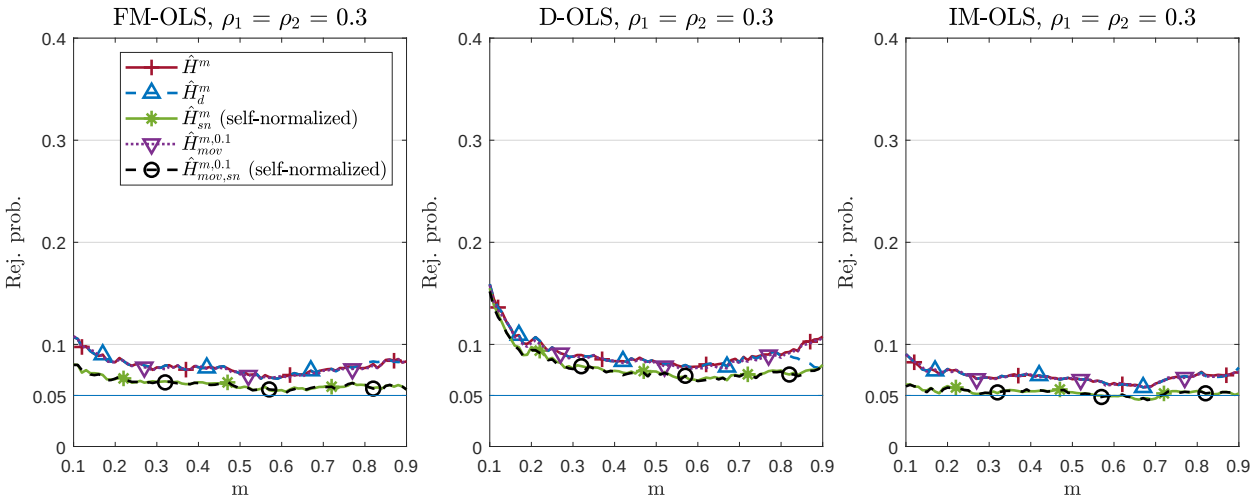

Figure 17: Empirical null rejection probabilities for a grid of values of m, with $T = 500$ and $\rho_1 = \rho_2 = 0.3$.

Size-Corrected Power – I(1) Breaks

In addition to Section 1.3 in the main text, Table 2 displays the size-corrected power results in case of I(1) breaks for $\rho_1 = \rho_2 = 0.3$ and $T = 500$. Furthermore, Tables 9 to 14 display the corresponding results for $\rho_1 = \rho_2 \in \{0, 0.6, 0.9\}$ and $T \in \{200, 500\}$. The findings are similar as for $\rho_1 = \rho_2 = 0.3$ in the main text: Size-corrected power is by and large highest for $\hat{H}_{mov}^{m,0.1}$ or $\hat{H}_{mov,sn}^{m,0.1}$ and lower for IM-OLS than for FM-OLS and D-OLS.

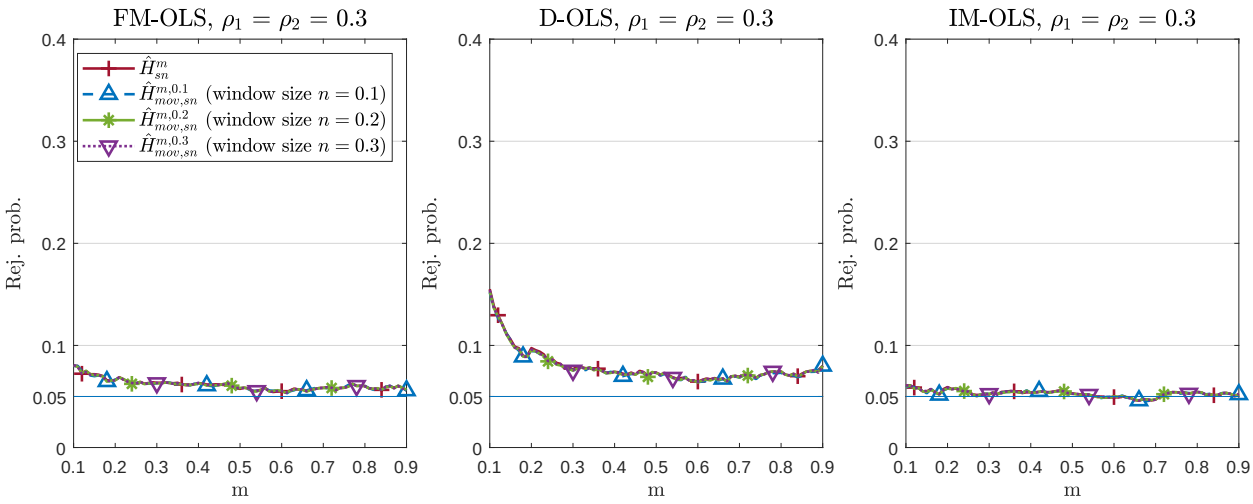

Figure 18: Empirical null rejection probabilities for a grid of values of m, with $T = 500$ and $\rho_1 = \rho_2 = 0.3$.

$\rho_1 = \rho_2 = 0.0$			$m = 0.25$			$m = 0.5$			$m = 0.75$		
		r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\hat{H}^m$	FM		0.26	0.08	0.05	0.37	0.76	0.21	0.38	0.51	0.87
	D		0.19	0.06	0.05	0.37	0.71	0.17	0.39	0.51	0.85
	IM		0.15	0.06	0.05	0.32	0.61	0.11	0.38	0.47	0.77
$\hat{H}_d^m$	FM		0.26	0.08	0.05	0.37	0.76	0.21	0.37	0.50	0.87
	D		0.19	0.06	0.05	0.37	0.71	0.17	0.39	0.51	0.85
	IM		0.15	0.06	0.05	0.32	0.61	0.11	0.38	0.47	0.77
$\hat{H}_{sn}^m$	FM		0.28	0.08	0.05	0.18	0.73	0.21	0.17	0.27	0.85
	D		0.20	0.07	0.05	0.23	0.69	0.17	0.22	0.33	0.83
	IM		0.18	0.06	0.05	0.22	0.62	0.12	0.23	0.34	0.77
$\hat{H}_{mov}^{m,0.1}$	FM		0.26	0.07	0.05	0.37	0.77	0.26	0.38	0.51	0.87
	D		0.18	0.06	0.05	0.36	0.73	0.22	0.39	0.51	0.85
	IM		0.14	0.06	0.05	0.32	0.63	0.14	0.37	0.46	0.78
$\hat{H}_{mov}^{m,0.2}$	FM		0.26	0.08	0.05	0.37	0.76	0.23	0.38	0.51	0.87
	D		0.18	0.07	0.05	0.36	0.72	0.19	0.39	0.51	0.85
	IM		0.15	0.06	0.05	0.32	0.62	0.12	0.38	0.47	0.77
$\hat{H}_{mov}^{m,0.3}$	FM		0.26	0.08	0.05	0.37	0.75	0.21	0.38	0.51	0.87
	D		0.18	0.06	0.05	0.36	0.71	0.17	0.39	0.51	0.85
	IM		0.15	0.06	0.05	0.32	0.61	0.11	0.38	0.47	0.77
$\hat{H}_{mov,sn}^{m,0.1}$	FM		0.29	0.08	0.05	0.18	0.75	0.26	0.17	0.28	0.86
	D		0.21	0.07	0.05	0.23	0.71	0.21	0.22	0.33	0.84
	IM		0.17	0.06	0.05	0.21	0.63	0.15	0.22	0.33	0.77
$\hat{H}_{mov,sn}^{m,0.2}$	FM		0.29	0.08	0.05	0.18	0.74	0.23	0.17	0.27	0.85
	D		0.21	0.07	0.05	0.23	0.69	0.18	0.22	0.33	0.83
	IM		0.17	0.06	0.05	0.22	0.63	0.14	0.23	0.34	0.76
$\hat{H}_{mov,sn}^{m,0.3}$	FM		0.29	0.08	0.05	0.18	0.73	0.22	0.17	0.27	0.85
	D		0.21	0.07	0.05	0.23	0.69	0.17	0.22	0.33	0.83
	IM		0.17	0.06	0.05	0.22	0.62	0.13	0.23	0.34	0.77

Table 9: Size-corrected power against $I(1)$ breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.0$.

$\rho_1 = \rho_2 = 0.6$			$m = 0.25$			$m = 0.5$			$m = 0.75$		
		r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\hat{H}^m$	FM		0.11	0.06	0.05	0.24	0.44	0.08	0.27	0.36	0.64
	D		0.09	0.05	0.05	0.23	0.40	0.07	0.29	0.38	0.61
	IM		0.07	0.05	0.05	0.19	0.27	0.06	0.26	0.32	0.48
$\hat{H}_d^m$	FM		0.11	0.06	0.05	0.24	0.44	0.08	0.27	0.36	0.64
	D		0.09	0.05	0.05	0.23	0.40	0.07	0.28	0.37	0.61
	IM		0.07	0.05	0.05	0.19	0.27	0.06	0.26	0.32	0.48
$\hat{H}_{sn}^m$	FM		0.12	0.06	0.05	0.15	0.42	0.08	0.14	0.23	0.62
	D		0.10	0.06	0.05	0.17	0.38	0.07	0.17	0.25	0.58
	IM		0.07	0.05	0.05	0.15	0.29	0.06	0.18	0.27	0.48
$\hat{H}_{mov}^{m,0.1}$	FM		0.10	0.06	0.05	0.23	0.45	0.09	0.27	0.36	0.64
	D		0.09	0.05	0.05	0.22	0.41	0.08	0.29	0.38	0.62
	IM		0.06	0.05	0.05	0.18	0.26	0.06	0.25	0.31	0.48
$\hat{H}_{mov}^{m,0.2}$	FM		0.10	0.06	0.05	0.24	0.45	0.08	0.27	0.36	0.64
	D		0.09	0.05	0.05	0.22	0.40	0.07	0.29	0.38	0.61
	IM		0.07	0.05	0.05	0.19	0.27	0.06	0.26	0.32	0.48
$\hat{H}_{mov}^{m,0.3}$	FM		0.11	0.06	0.05	0.24	0.44	0.08	0.27	0.36	0.64
	D		0.09	0.05	0.05	0.22	0.40	0.07	0.29	0.38	0.61
	IM		0.07	0.05	0.05	0.19	0.27	0.06	0.26	0.32	0.48
$\hat{H}_{mov,sn}^{m,0.1}$	FM		0.12	0.06	0.05	0.15	0.43	0.10	0.15	0.23	0.62
	D		0.10	0.06	0.05	0.17	0.39	0.08	0.17	0.25	0.59
	IM		0.07	0.05	0.05	0.14	0.27	0.06	0.18	0.26	0.49
$\hat{H}_{mov,sn}^{m,0.2}$	FM		0.12	0.06	0.05	0.15	0.42	0.09	0.14	0.23	0.62
	D		0.09	0.06	0.05	0.17	0.38	0.08	0.17	0.25	0.58
	IM		0.07	0.05	0.05	0.15	0.28	0.06	0.18	0.26	0.48
$\hat{H}_{mov,sn}^{m,0.3}$	FM		0.12	0.06	0.05	0.15	0.42	0.08	0.14	0.23	0.62
	D		0.09	0.06	0.05	0.17	0.38	0.07	0.17	0.25	0.58
	IM		0.07	0.05	0.05	0.15	0.28	0.06	0.18	0.27	0.48

Table 10: Size-corrected power against $I(1)$ breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.6$.

$\rho_1 = \rho_2 = 0.9$			$m = 0.25$			$m = 0.5$			$m = 0.75$		
		r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\hat{H}^m$	FM		0.06	0.05	0.05	0.10	0.10	0.05	0.13	0.13	0.17
	D		0.06	0.05	0.05	0.10	0.10	0.06	0.14	0.15	0.16
	IM		0.05	0.05	0.05	0.08	0.06	0.05	0.12	0.12	0.11
$\hat{H}_d^m$	FM		0.06	0.05	0.05	0.10	0.10	0.05	0.13	0.13	0.17
	D		0.06	0.05	0.05	0.10	0.10	0.06	0.14	0.15	0.16
	IM		0.05	0.05	0.05	0.08	0.06	0.05	0.12	0.12	0.11
$\hat{H}_{sn}^m$	FM		0.06	0.05	0.05	0.08	0.09	0.05	0.08	0.10	0.15
	D		0.06	0.05	0.05	0.08	0.10	0.05	0.10	0.11	0.15
	IM		0.05	0.05	0.05	0.07	0.07	0.05	0.11	0.11	0.12
$\hat{H}_{mov}^{m,0.1}$	FM		0.06	0.05	0.05	0.09	0.09	0.05	0.13	0.13	0.17
	D		0.06	0.05	0.05	0.10	0.10	0.06	0.14	0.15	0.16
	IM		0.05	0.05	0.05	0.08	0.06	0.05	0.12	0.12	0.11
$\hat{H}_{mov}^{m,0.2}$	FM		0.06	0.05	0.05	0.10	0.10	0.05	0.13	0.13	0.17
	D		0.06	0.05	0.05	0.10	0.10	0.06	0.14	0.15	0.16
	IM		0.05	0.05	0.05	0.08	0.06	0.05	0.12	0.12	0.11
$\hat{H}_{mov}^{m,0.3}$	FM		0.06	0.05	0.05	0.10	0.10	0.05	0.13	0.13	0.17
	D		0.06	0.05	0.05	0.10	0.10	0.05	0.14	0.15	0.16
	IM		0.05	0.05	0.05	0.08	0.06	0.05	0.12	0.12	0.11
$\hat{H}_{mov,sn}^{m,0.1}$	FM		0.06	0.05	0.05	0.08	0.10	0.06	0.09	0.10	0.15
	D		0.06	0.05	0.05	0.09	0.10	0.06	0.10	0.11	0.15
	IM		0.05	0.05	0.05	0.07	0.07	0.05	0.11	0.11	0.12
$\hat{H}_{mov,sn}^{m,0.2}$	FM		0.06	0.05	0.05	0.08	0.09	0.05	0.08	0.10	0.15
	D		0.06	0.05	0.05	0.09	0.10	0.06	0.10	0.11	0.15
	IM		0.05	0.05	0.05	0.07	0.07	0.05	0.11	0.11	0.12
$\hat{H}_{mov,sn}^{m,0.3}$	FM		0.06	0.05	0.05	0.08	0.09	0.05	0.08	0.10	0.15
	D		0.06	0.05	0.05	0.08	0.10	0.05	0.10	0.11	0.15
	IM		0.05	0.05	0.05	0.07	0.07	0.05	0.11	0.11	0.12

Table 11: Size-corrected power against $I(1)$ breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.9$.

$\rho_1 = \rho_2 = 0.0$			$m = 0.25$			$m = 0.5$			$m = 0.75$		
		r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\hat{H}^m$	FM		0.78	0.36	0.06	0.48	0.97	0.59	0.49	0.61	0.99
	D		0.75	0.33	0.06	0.47	0.96	0.56	0.50	0.62	0.98
	IM		0.58	0.19	0.05	0.38	0.89	0.38	0.48	0.57	0.95
$\hat{H}_d^m$	FM		0.78	0.36	0.06	0.48	0.97	0.59	0.51	0.65	0.99
	D		0.75	0.33	0.06	0.47	0.96	0.56	0.50	0.62	0.98
	IM		0.58	0.19	0.05	0.38	0.89	0.38	0.47	0.57	0.95
$\hat{H}_{sn}^m$	FM		0.76	0.36	0.06	0.21	0.95	0.57	0.18	0.31	0.98
	D		0.74	0.32	0.06	0.23	0.95	0.54	0.21	0.34	0.98
	IM		0.60	0.20	0.05	0.24	0.90	0.41	0.25	0.38	0.95
$\hat{H}_{mov}^{m,0.1}$	FM		0.82	0.38	0.07	0.49	0.98	0.65	0.49	0.61	0.99
	D		0.79	0.34	0.07	0.48	0.97	0.62	0.50	0.62	0.98
	IM		0.61	0.18	0.06	0.38	0.92	0.45	0.47	0.57	0.96
$\hat{H}_{mov}^{m,0.2}$	FM		0.80	0.37	0.06	0.48	0.97	0.61	0.49	0.61	0.99
	D		0.77	0.34	0.06	0.47	0.96	0.58	0.50	0.62	0.98
	IM		0.60	0.19	0.05	0.38	0.90	0.40	0.47	0.57	0.95
$\hat{H}_{mov}^{m,0.3}$	FM		0.79	0.37	0.06	0.48	0.97	0.59	0.49	0.61	0.99
	D		0.76	0.33	0.06	0.47	0.96	0.57	0.50	0.62	0.98
	IM		0.58	0.19	0.05	0.38	0.89	0.38	0.48	0.57	0.95
$\hat{H}_{mov,sn}^{m,0.1}$	FM		0.80	0.38	0.07	0.21	0.96	0.63	0.19	0.31	0.98
	D		0.78	0.34	0.07	0.22	0.96	0.59	0.21	0.33	0.98
	IM		0.63	0.20	0.05	0.23	0.92	0.47	0.25	0.38	0.96
$\hat{H}_{mov,sn}^{m,0.2}$	FM		0.77	0.37	0.07	0.21	0.96	0.59	0.18	0.31	0.98
	D		0.76	0.33	0.06	0.22	0.95	0.56	0.21	0.34	0.98
	IM		0.61	0.21	0.05	0.23	0.90	0.43	0.25	0.38	0.95
$\hat{H}_{mov,sn}^{m,0.3}$	FM		0.76	0.36	0.06	0.21	0.95	0.58	0.18	0.31	0.98
	D		0.74	0.33	0.06	0.23	0.95	0.55	0.21	0.34	0.98
	IM		0.60	0.20	0.05	0.24	0.90	0.41	0.25	0.38	0.95

Table 12: Size-corrected power against $I(1)$ breaks for $T = 500$ and $\rho_1 = \rho_2 = 0.0$.

$\rho_1 = \rho_2 = 0.6$			$m = 0.25$			$m = 0.5$			$m = 0.75$		
		r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\hat{H}^m$	FM		0.44	0.13	0.05	0.40	0.82	0.27	0.42	0.53	0.91
	D		0.40	0.11	0.05	0.38	0.80	0.24	0.43	0.54	0.89
	IM		0.21	0.07	0.05	0.31	0.66	0.13	0.40	0.49	0.80
$\hat{H}_d^m$	FM		0.44	0.13	0.05	0.40	0.82	0.27	0.41	0.53	0.90
	D		0.40	0.11	0.05	0.38	0.80	0.24	0.41	0.53	0.89
	IM		0.21	0.07	0.05	0.31	0.66	0.13	0.40	0.49	0.80
$\hat{H}_{sn}^m$	FM		0.43	0.13	0.05	0.19	0.79	0.26	0.17	0.28	0.88
	D		0.39	0.12	0.05	0.20	0.78	0.24	0.18	0.29	0.87
	IM		0.23	0.07	0.05	0.21	0.67	0.15	0.23	0.34	0.79
$\hat{H}_{mov}^{m,0.1}$	FM		0.46	0.13	0.05	0.40	0.85	0.34	0.42	0.53	0.91
	D		0.42	0.11	0.05	0.38	0.83	0.30	0.43	0.54	0.90
	IM		0.21	0.07	0.05	0.30	0.67	0.16	0.40	0.48	0.81
$\hat{H}_{mov}^{m,0.2}$	FM		0.45	0.13	0.05	0.40	0.83	0.29	0.42	0.53	0.90
	D		0.42	0.12	0.05	0.38	0.81	0.26	0.43	0.54	0.89
	IM		0.21	0.07	0.05	0.30	0.66	0.14	0.40	0.49	0.80
$\hat{H}_{mov}^{m,0.3}$	FM		0.44	0.13	0.05	0.40	0.82	0.28	0.42	0.53	0.91
	D		0.41	0.11	0.05	0.38	0.80	0.24	0.43	0.54	0.89
	IM		0.21	0.07	0.05	0.31	0.66	0.13	0.41	0.49	0.80
$\hat{H}_{mov,sn}^{m,0.1}$	FM		0.46	0.14	0.06	0.20	0.82	0.32	0.18	0.28	0.89
	D		0.41	0.12	0.06	0.20	0.80	0.29	0.18	0.29	0.88
	IM		0.23	0.07	0.05	0.21	0.69	0.18	0.23	0.34	0.81
$\hat{H}_{mov,sn}^{m,0.2}$	FM		0.45	0.14	0.05	0.20	0.80	0.28	0.17	0.28	0.88
	D		0.40	0.12	0.05	0.20	0.79	0.25	0.18	0.29	0.87
	IM		0.23	0.07	0.05	0.21	0.68	0.16	0.23	0.34	0.79
$\hat{H}_{mov,sn}^{m,0.3}$	FM		0.44	0.14	0.05	0.19	0.80	0.26	0.17	0.28	0.88
	D		0.39	0.12	0.05	0.20	0.78	0.24	0.18	0.29	0.87
	IM		0.23	0.07	0.05	0.21	0.67	0.15	0.23	0.34	0.79

Table 13: Size-corrected power against $I(1)$ breaks for $T = 500$ and $\rho_1 = \rho_2 = 0.6$.

$\rho_1 = \rho_2 = 0.9$			$m = 0.25$			$m = 0.5$			$m = 0.75$		
		r	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\hat{H}^m$	FM		0.09	0.06	0.05	0.23	0.30	0.07	0.27	0.33	0.47
	D		0.09	0.06	0.05	0.24	0.30	0.06	0.29	0.34	0.46
	IM		0.06	0.05	0.05	0.17	0.16	0.05	0.29	0.30	0.33
$\hat{H}_d^m$	FM		0.09	0.06	0.05	0.23	0.30	0.07	0.27	0.32	0.47
	D		0.09	0.06	0.05	0.24	0.30	0.06	0.28	0.33	0.46
	IM		0.06	0.05	0.05	0.17	0.16	0.05	0.29	0.30	0.33
$\hat{H}_{sn}^m$	FM		0.09	0.06	0.05	0.15	0.28	0.07	0.13	0.18	0.42
	D		0.09	0.06	0.05	0.15	0.28	0.07	0.14	0.19	0.41
	IM		0.06	0.05	0.05	0.13	0.17	0.05	0.18	0.23	0.33
$\hat{H}_{mov}^{m,0.1}$	FM		0.09	0.06	0.05	0.22	0.30	0.07	0.28	0.32	0.47
	D		0.09	0.06	0.05	0.23	0.30	0.07	0.29	0.34	0.47
	IM		0.06	0.05	0.05	0.16	0.15	0.05	0.28	0.30	0.33
$\hat{H}_{mov}^{m,0.2}$	FM		0.09	0.06	0.05	0.23	0.30	0.07	0.27	0.33	0.47
	D		0.09	0.06	0.05	0.23	0.30	0.07	0.29	0.34	0.46
	IM		0.06	0.05	0.05	0.16	0.16	0.05	0.29	0.30	0.33
$\hat{H}_{mov}^{m,0.3}$	FM		0.09	0.06	0.05	0.23	0.30	0.07	0.27	0.33	0.47
	D		0.09	0.06	0.05	0.23	0.30	0.06	0.29	0.34	0.46
	IM		0.06	0.05	0.05	0.17	0.16	0.05	0.29	0.30	0.33
$\hat{H}_{mov,sn}^{m,0.1}$	FM		0.09	0.06	0.05	0.14	0.28	0.07	0.13	0.18	0.42
	D		0.09	0.06	0.05	0.15	0.28	0.08	0.14	0.19	0.41
	IM		0.06	0.05	0.05	0.13	0.17	0.06	0.18	0.23	0.33
$\hat{H}_{mov,sn}^{m,0.2}$	FM		0.09	0.06	0.05	0.14	0.28	0.07	0.13	0.18	0.42
	D		0.09	0.06	0.05	0.15	0.28	0.07	0.14	0.19	0.41
	IM		0.06	0.05	0.05	0.13	0.17	0.06	0.18	0.23	0.33
$\hat{H}_{mov,sn}^{m,0.3}$	FM		0.09	0.06	0.05	0.14	0.28	0.07	0.13	0.18	0.42
	D		0.09	0.06	0.05	0.15	0.28	0.07	0.14	0.19	0.41
	IM		0.06	0.05	0.05	0.13	0.17	0.05	0.18	0.23	0.33

Table 14: Size-corrected power against $I(1)$ breaks for $T = 500$ and $\rho_1 = \rho_2 = 0.9$.

Size-Corrected Power – Trend Breaks

The size-corrected power results in case of trend breaks complement the results for the case of I(1) breaks from Section 1.3 in the main text. The parameter θ_{D_1} varies in 101 equidistant steps over the interval $[1, 1.5]$, including the corresponding null hypothesis value $\theta_{D_1} = 1$. Figure 19 displays size-corrected power against trend breaks using $\hat{H}^m$, $\hat{H}^m_{sn}$, $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$. The sample size is $T = 200$, calibration and break fraction are set to $m = 0.5$ and $r = 0.75$ with $\rho_1 = \rho_2 = 0.3$. FM-OLS leads to the highest size-corrected power and IM-OLS to the lowest, with D-OLS in between. Figure 20 displays the corresponding results for $T = 500$. Size-corrected power increases with increasing sample size T. Differences in size-corrected power between FM-OLS and D-OLS become very small for $T = 500$ compared to $T = 200$, reflecting the asymptotic equivalence between FM-OLS and D-OLS. The size-corrected power ranking with respect to FM-OLS, D-OLS and IM-OLS remains unchanged for larger window sizes and other combinations of m and r. Next, consider the impact of different window sizes and in addition the comparison to one expanding window detector. Figure 21 displays size-corrected power for $\hat{H}^m$, $\hat{H}^{m,0.1}_{mov}$, $\hat{H}^{m,0.2}_{mov}$ and $\hat{H}^{m,0.3}_{mov}$ in case of trend breaks. The sample size is $T = 200$ with $\rho_1 = \rho_2 = 0.3$, $m = 0.5$ and $r = 0.75$. Size-corrected power is highest either for $\hat{H}^m$ or $\hat{H}^{m,0.1}_{mov}$. The same holds for other combinations of m and r. Figures 22 and 23 display size-corrected power against trend breaks using $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$ and focus on the effect of self-normalization for several combinations of m and r with $\rho_1 = \rho_2 = 0.3$. Self-normalization leads, compared to standardization, by and large to higher size-corrected power for small breaks in θ_{D_1} and to lower size-corrected power for large breaks in θ_{D_1}. For small breaks in θ_{D_1} the differences in size-corrected power between self-normalization and standardization are larger for IM-OLS than for FM-OLS and D-OLS.

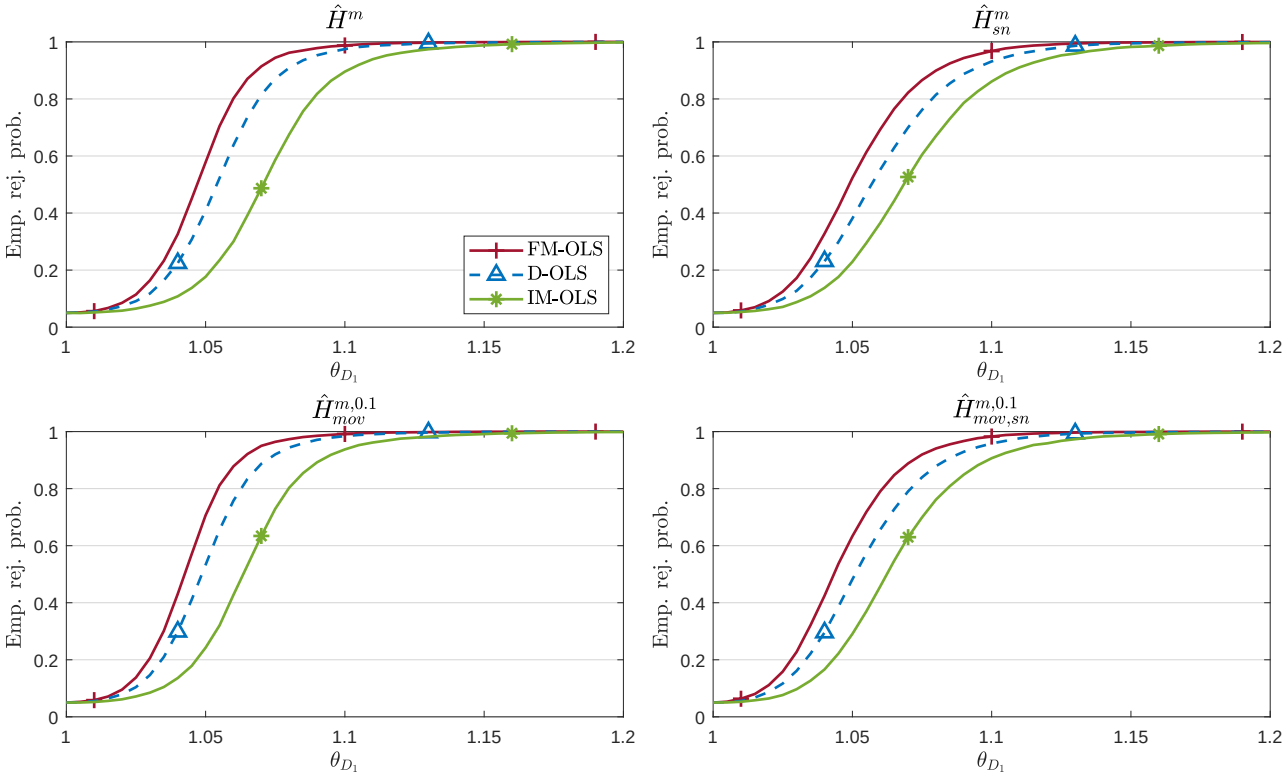

Figure 19: Size-corrected power against trend breaks for $T = 200$, $m = 0.5$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$.

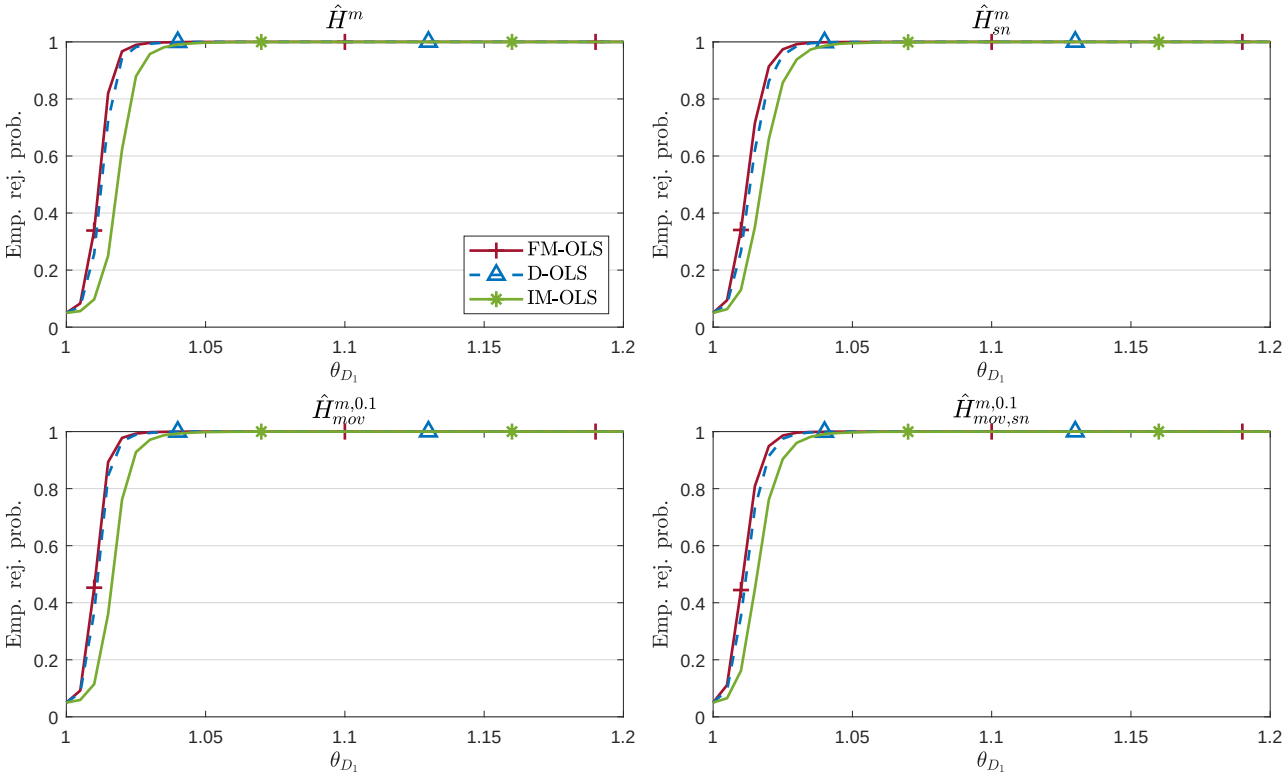

Figure 20: Size-corrected power against trend breaks for $T = 500$, $m = 0.5$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$.

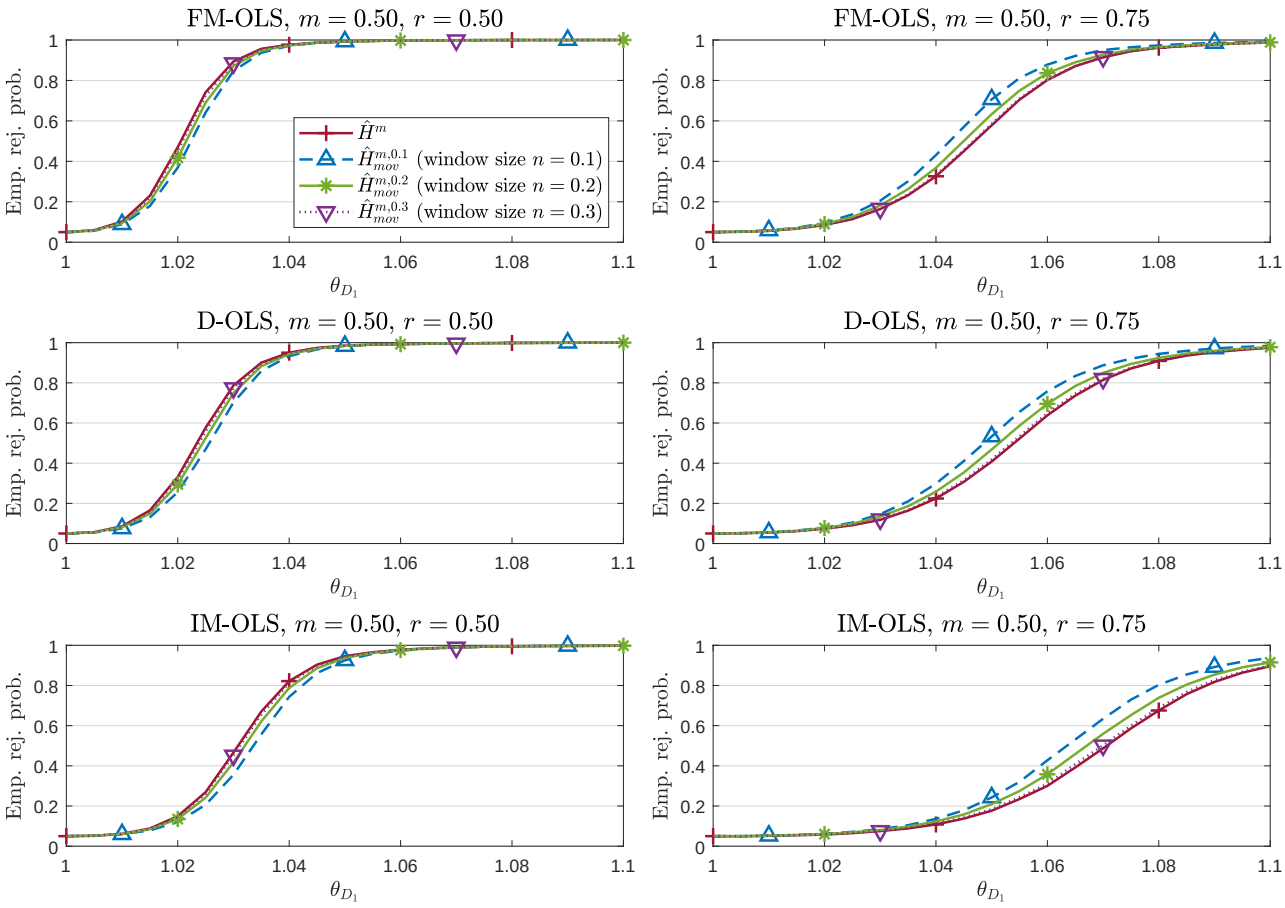

Figure 21: Size-corrected power against trend breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

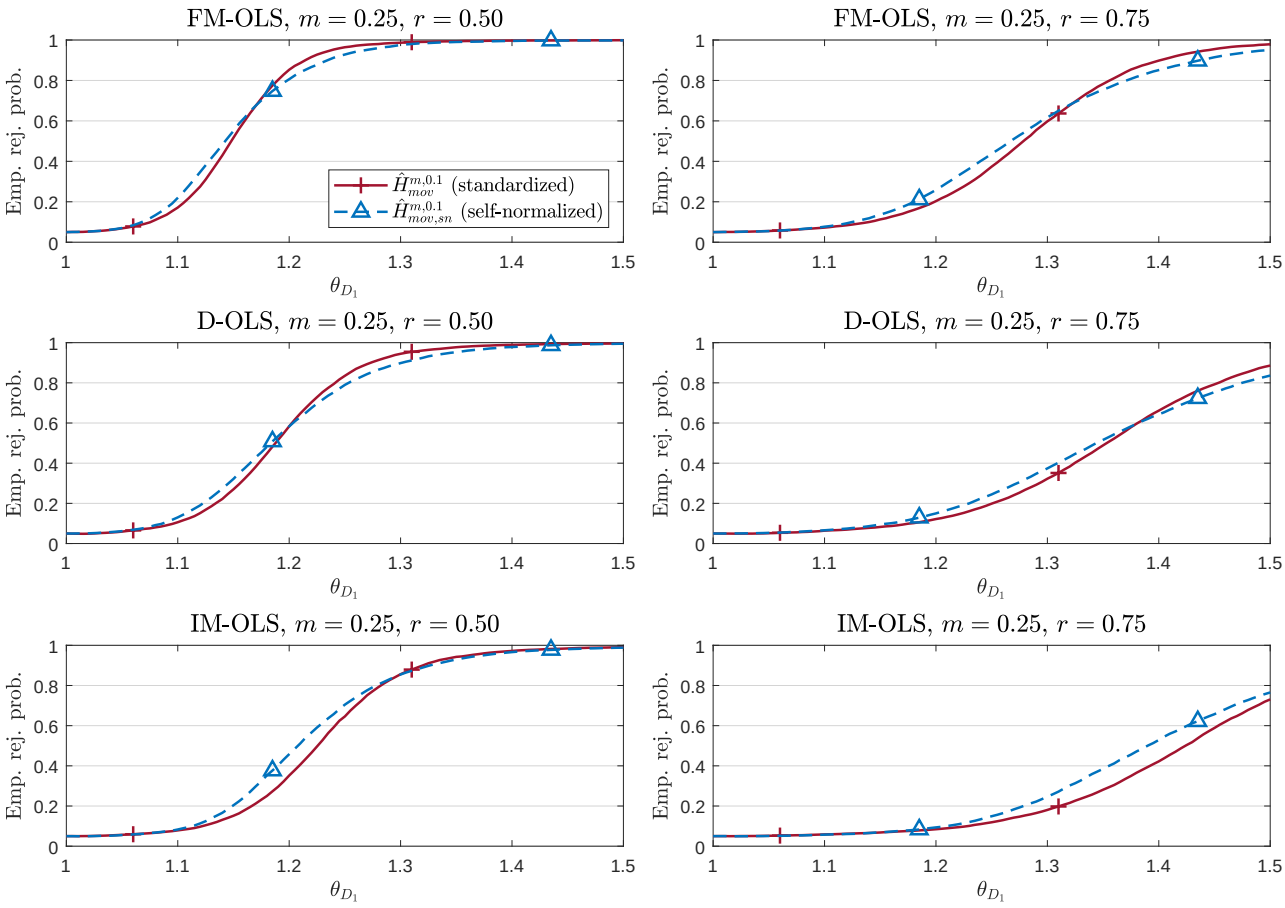

Figure 22: Size-corrected power against trend breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

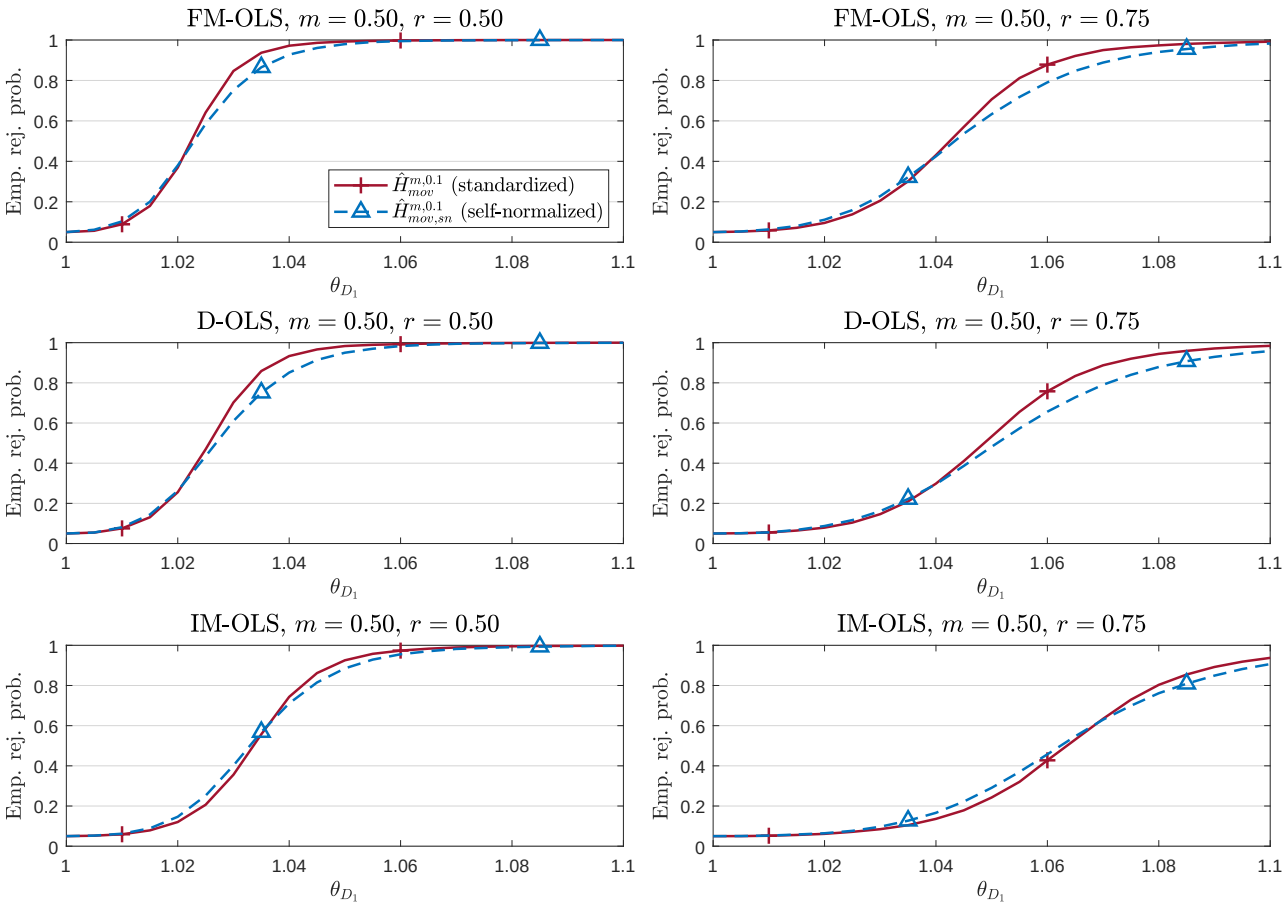

Figure 23: Size-corrected power against trend breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

Size-Corrected Power – Slope Breaks We next consider size-corrected power in case of slope breaks, where either $\theta_{X_{1,1}}$ or $\theta_{X_{1,2}}$ changes. The parameter $\theta_{X_{1,1}}$ varies in 26 equidistant steps over the interval $[5, 10]$ and $\theta_{X_{1,2}}$ varies in 27 equidistant steps over the interval $[-0.3, 1]$. The two intervals each include the corresponding null hypothesis value $\theta_{X_{1,1}} = 5$ and $\theta_{X_{1,2}} = -0.3$, respectively. Figures 24 (slope break in $\theta_{X_{1,1}}$) and 25 (slope break in $\theta_{X_{1,2}}$) display size-corrected power for $\hat{H}^m$, $\hat{H}^{m,0.1}_{mov}$, $\hat{H}^{m,0.2}_{mov}$ and $\hat{H}^{m,0.3}_{mov}$. The sample size is $T = 200$ with $\rho_1 = \rho_2 = 0.3$, $m = 0.5$ and $r = 0.75$. Similar to the case of I(1) breaks, size-corrected power is by and large highest for the moving window detector with $n = 0.1$. To assess the impact of self-normalization, Figures 26 ($m = 0.25$) and 27 ($m = 0.5$) display size-corrected power against slope breaks in $\theta_{X_{1,1}}$ using $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$, for $T = 200$, $\rho_1 = \rho_2 = 0.3$ and $r \in \{0.5, 0.75\}$. In case of $m = 0.25$, self-normalization leads for small breaks in $\theta_{X_{1,1}}$ to slightly higher size-corrected power than standardization, especially for IM-OLS. In case of $m = 0.5$, standardization leads for FM-OLS and D-OLS to slightly higher size-corrected power than self-normalization, with the exception of very small break magnitudes (i. e. , $\theta_{X_{1,1}} < 5.5$) where both standardization and self-normalization lead to equal size-corrected power. For IM-OLS, size-corrected power is almost equal for standardization and self-normalization when $m = 0.5$. The observed results are qualitatively similar for larger sample sizes and other values of ρ_1, ρ_2 and are available upon request.

As expected and already mentioned in the main text, the finite sample size-corrected power results are in line with the results from the LAP analysis in Appendix 1.6.2.

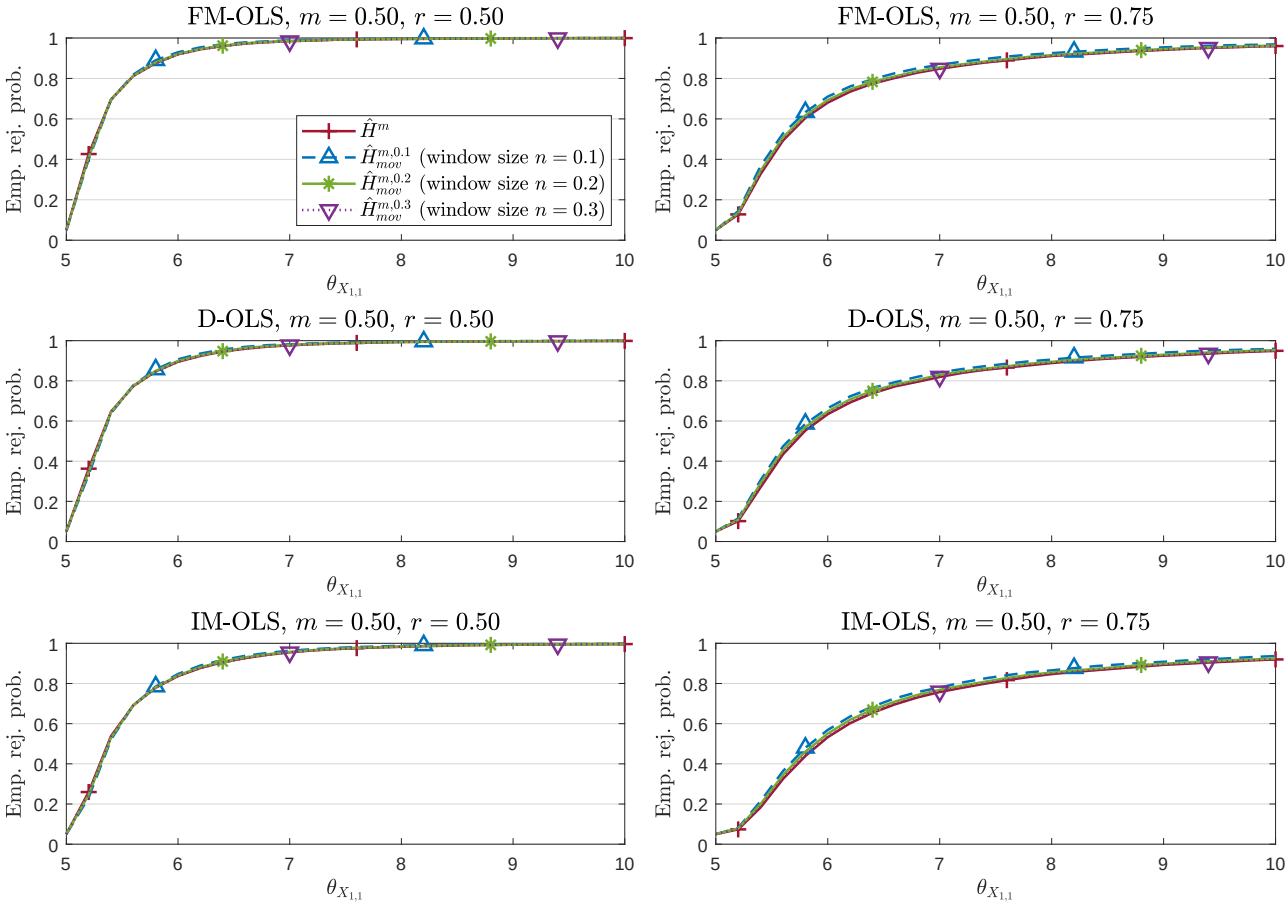

Figure 24: Size-corrected power against slope breaks (in $\theta_{X_{1,1}}$) for $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

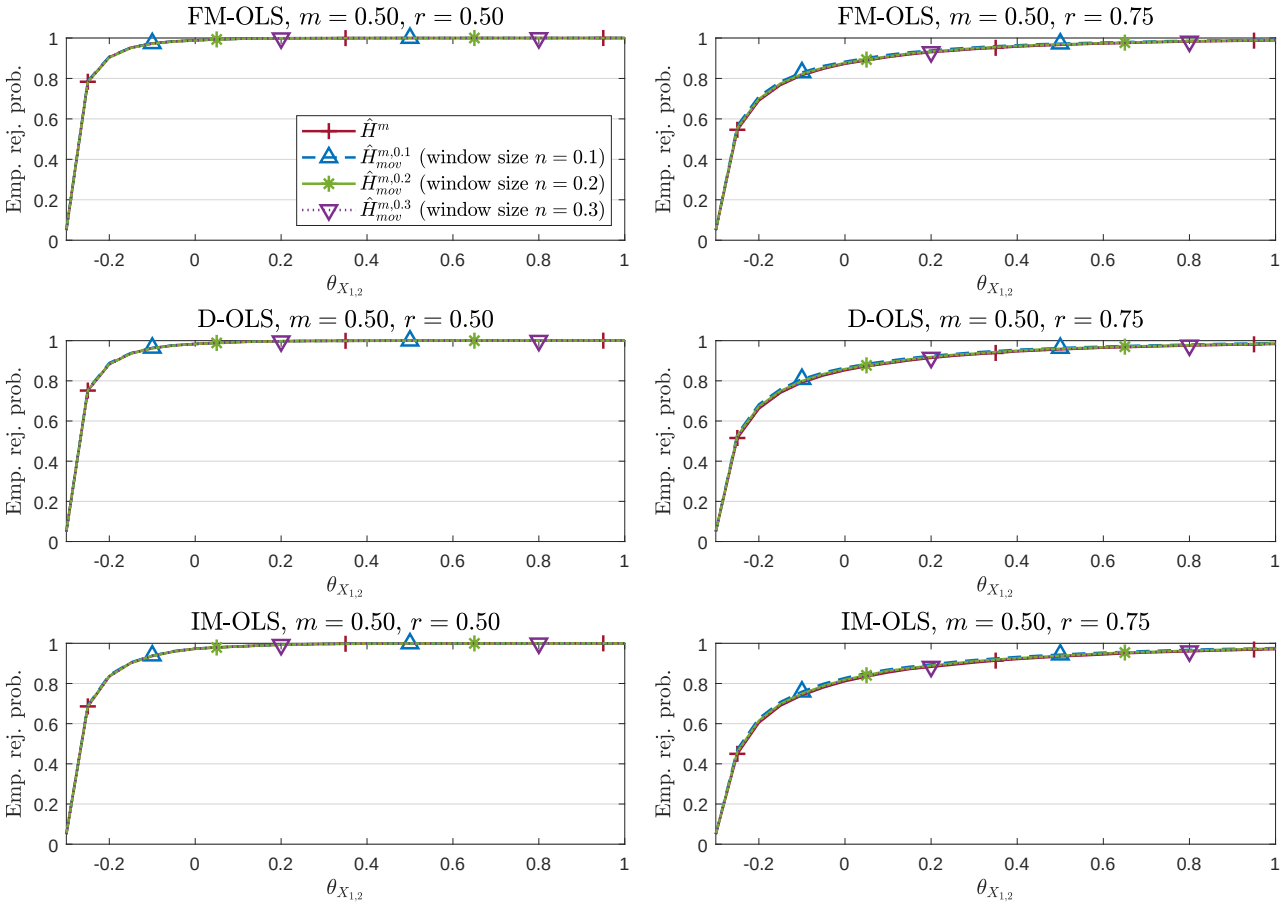

Figure 25: Size-corrected power against slope breaks (in $\theta_{X_{1,2}}$) for $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

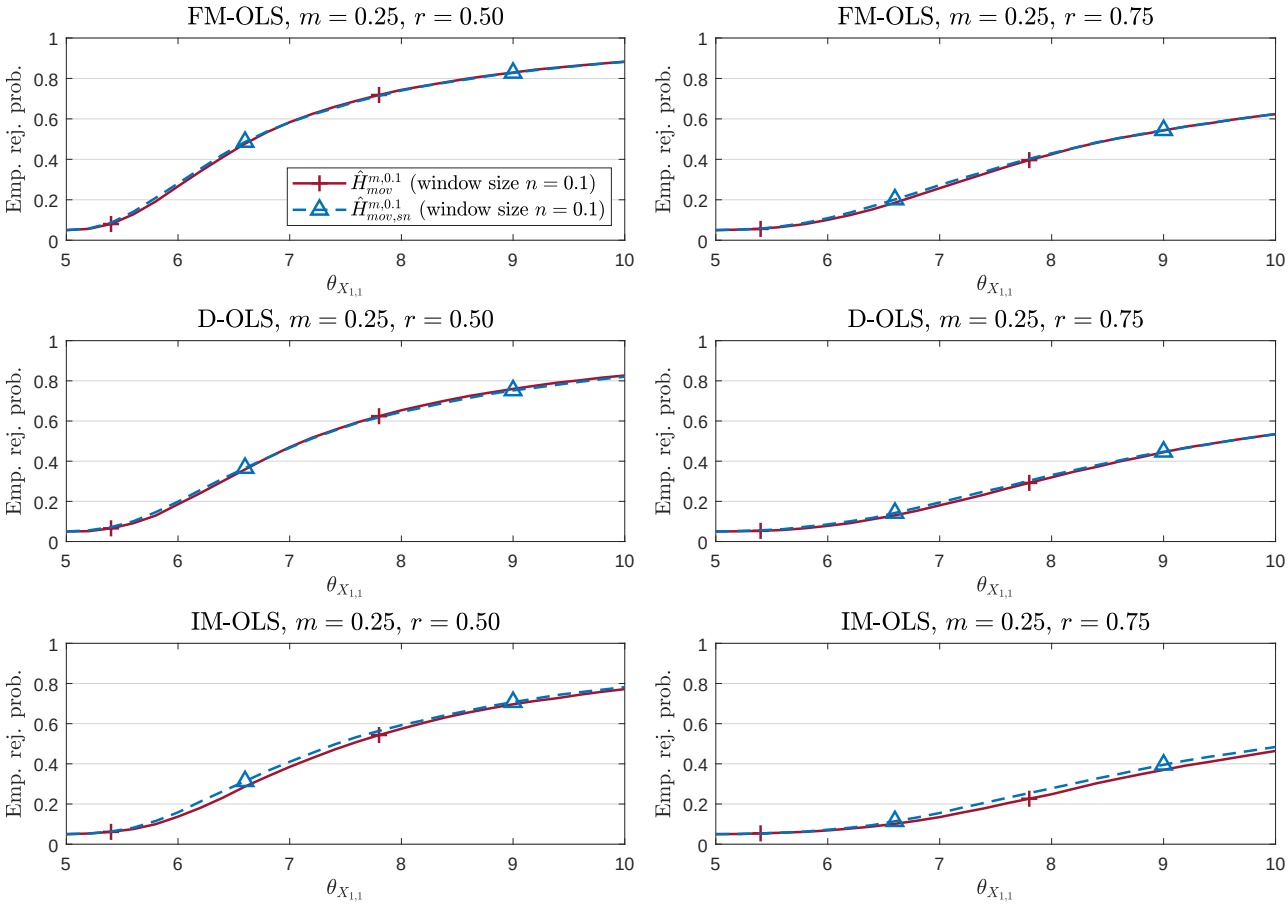

Figure 26: Size-corrected power against slope breaks (in $\theta_{X_{1,1}}$) for $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

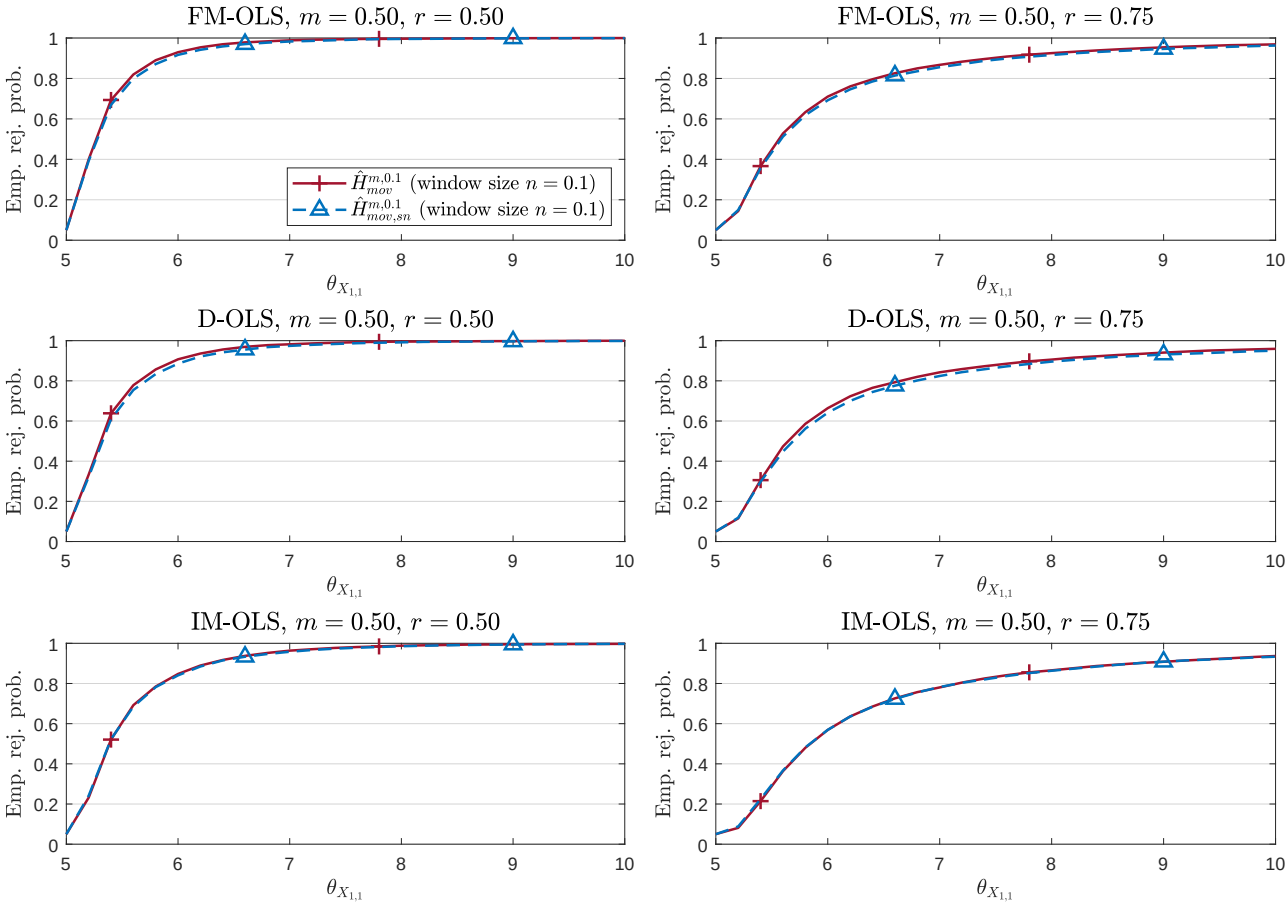

Figure 27: Size-corrected power against slope breaks (in $\theta_{X_{1,1}}$) for $T = 200$ and $\rho_1 = \rho_2 = 0.3$.

Detection Times

We close Appendix 1.6.3 with some additional detection time figures. Figures 29 – 33 here in Appendix 1.6.3 display the results for $\rho_1 = \rho_2 = 0.3$, $T \in \{200, 500\}$ and all other combinations of $m, r \in \{0.25, 0.5, 0.75\}$, with $m \leq r$. In addition, Figure 34 displays detection times for $\rho_1 = \rho_2 = 0.9$, with $m = 0.5$ and $r = 0.5$. As expected, increasing ρ_1, ρ_2 leads to larger delays. Further results, e. g. , for $\rho_1 = \rho_2 = 0$, are available upon request.

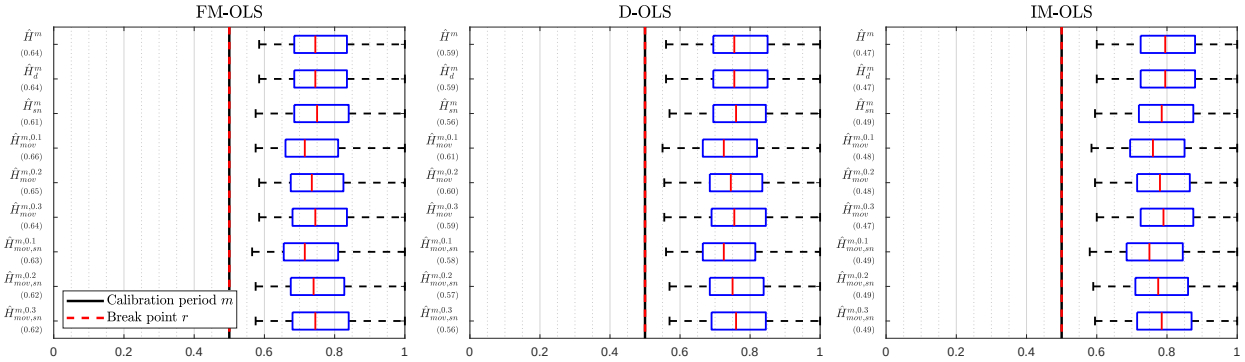

Figure 28: Detection times against $I(1)$ breaks for $T = 500$, $m = 0.5$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.3$.

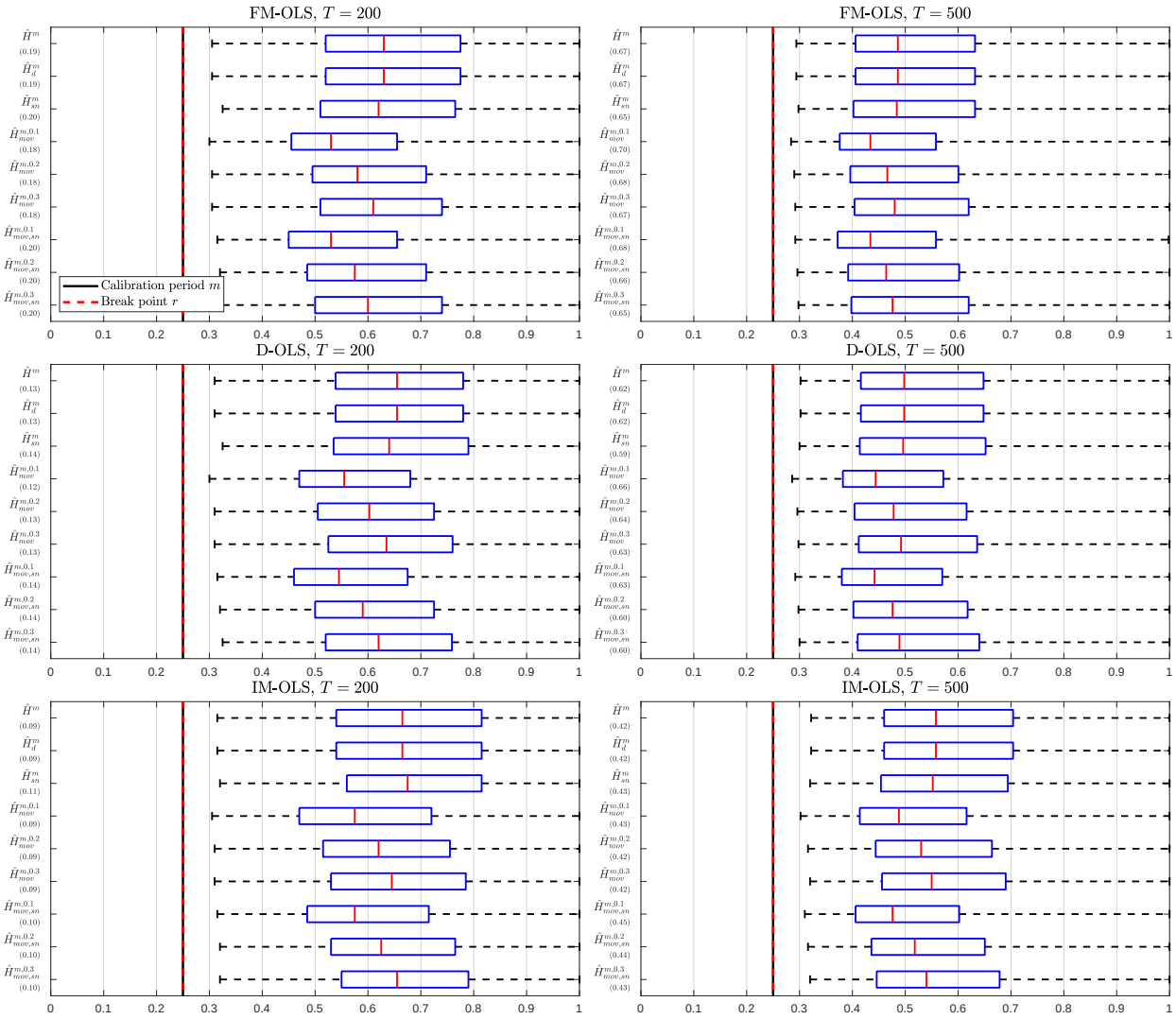

Figure 29: Detection times against $I(1)$ breaks for $m = 0.25$, $r = 0.25$ and $\rho_1 = \rho_2 = 0.3$.

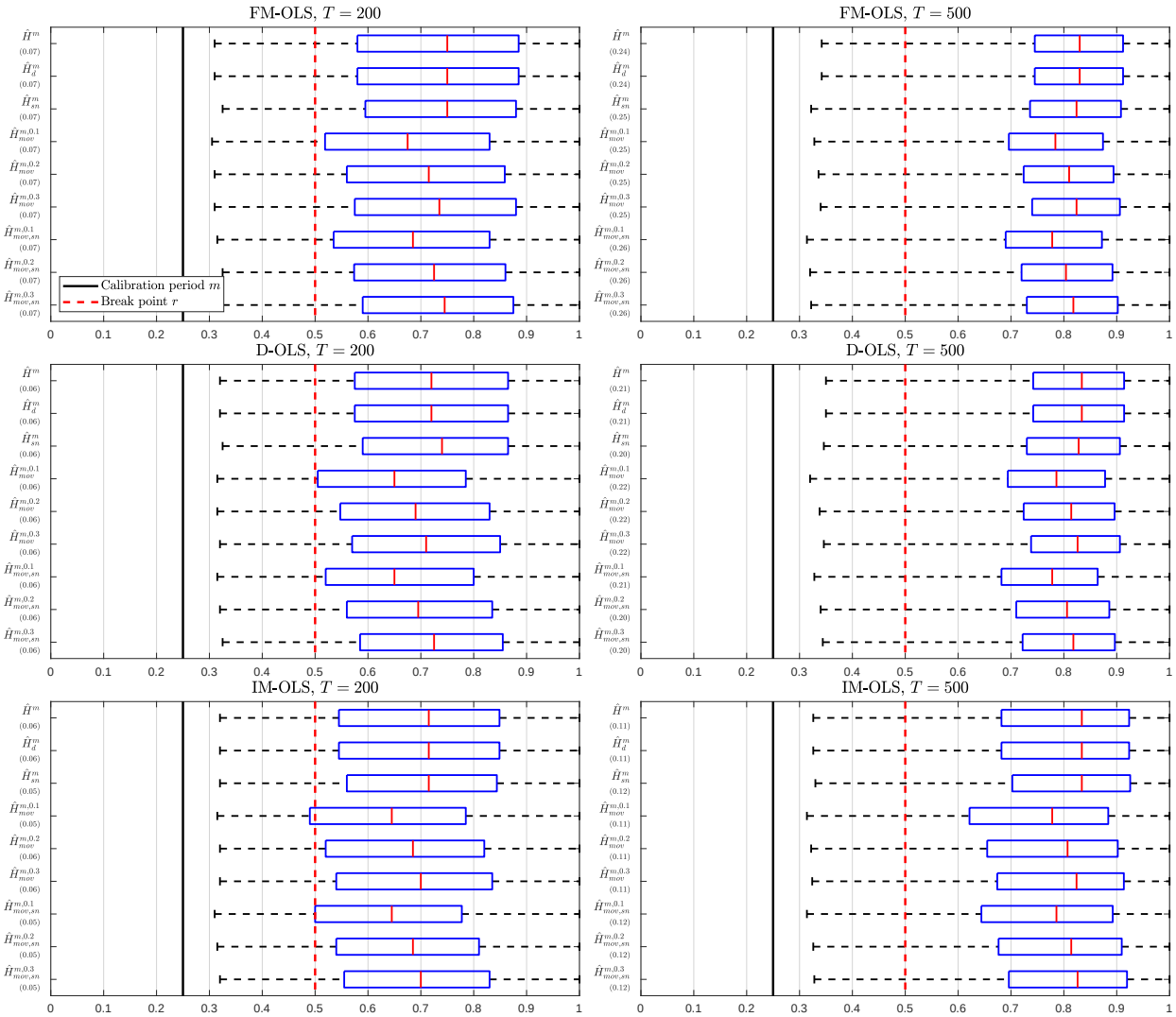

Figure 30: Detection times against $I(1)$ breaks for $m = 0.25$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.3$.

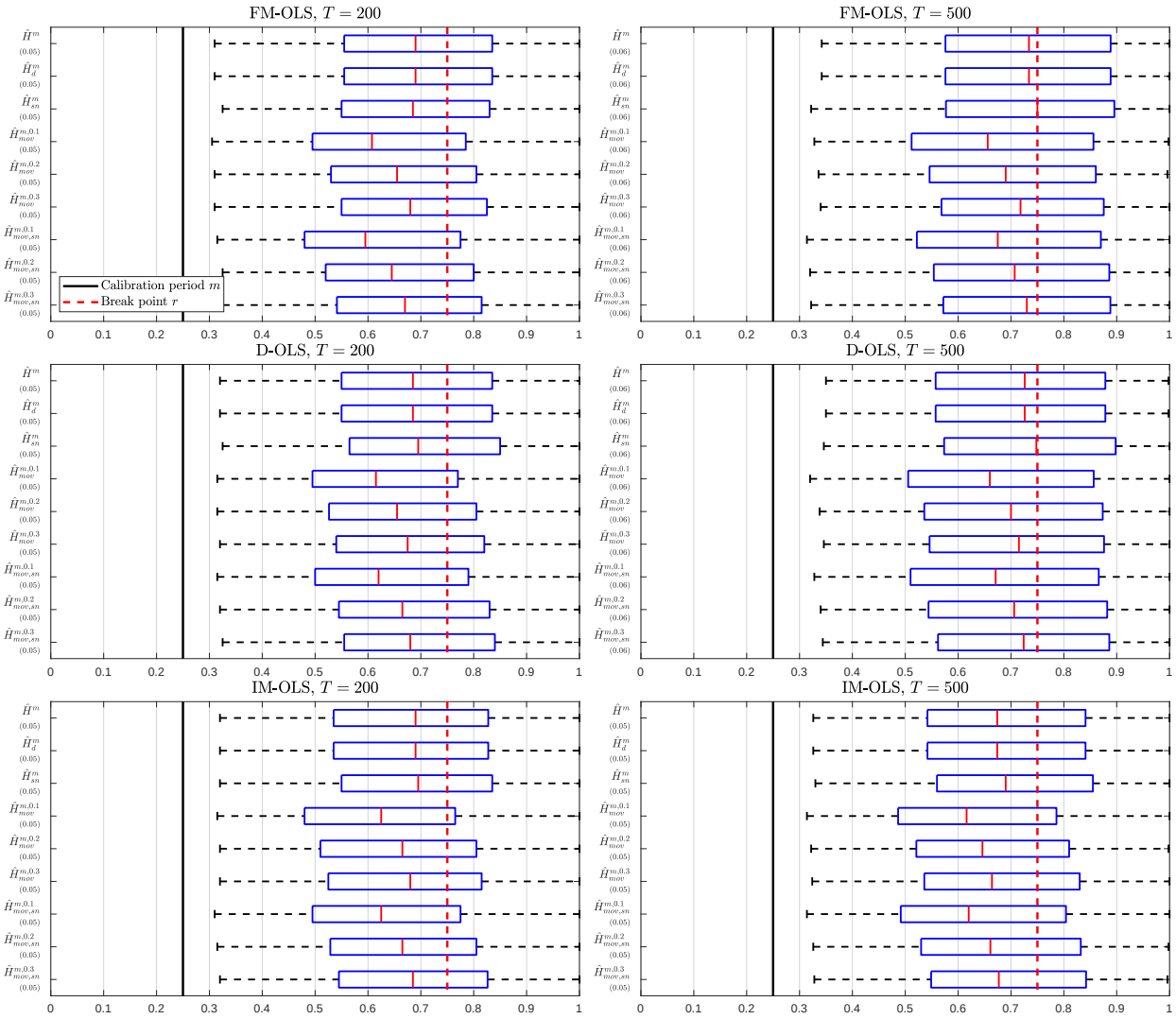

Figure 31: Detection times against $I(1)$ breaks for $m = 0.25$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$.

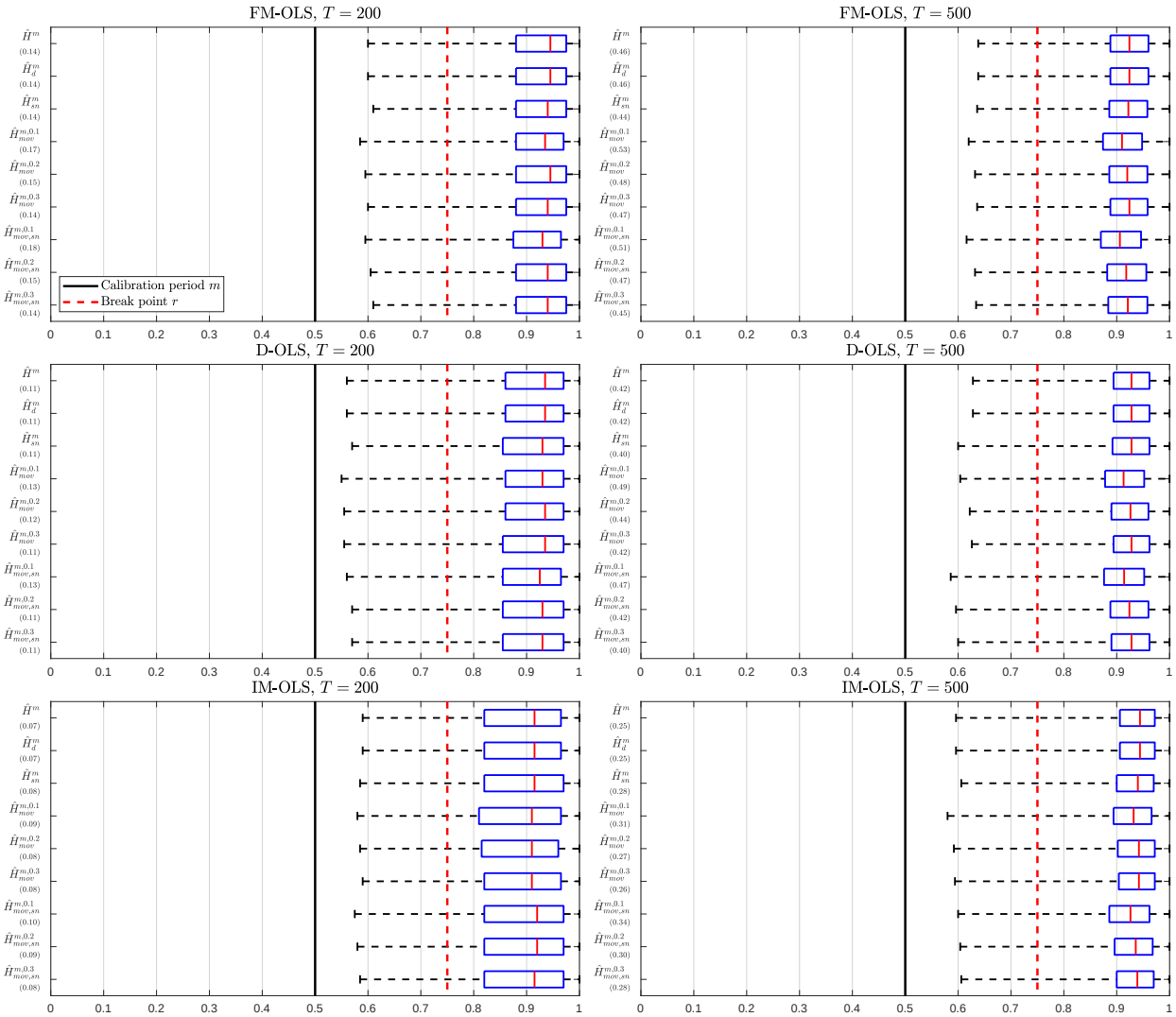

Figure 32: Detection times against $I(1)$ breaks for $m = 0.5$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$.

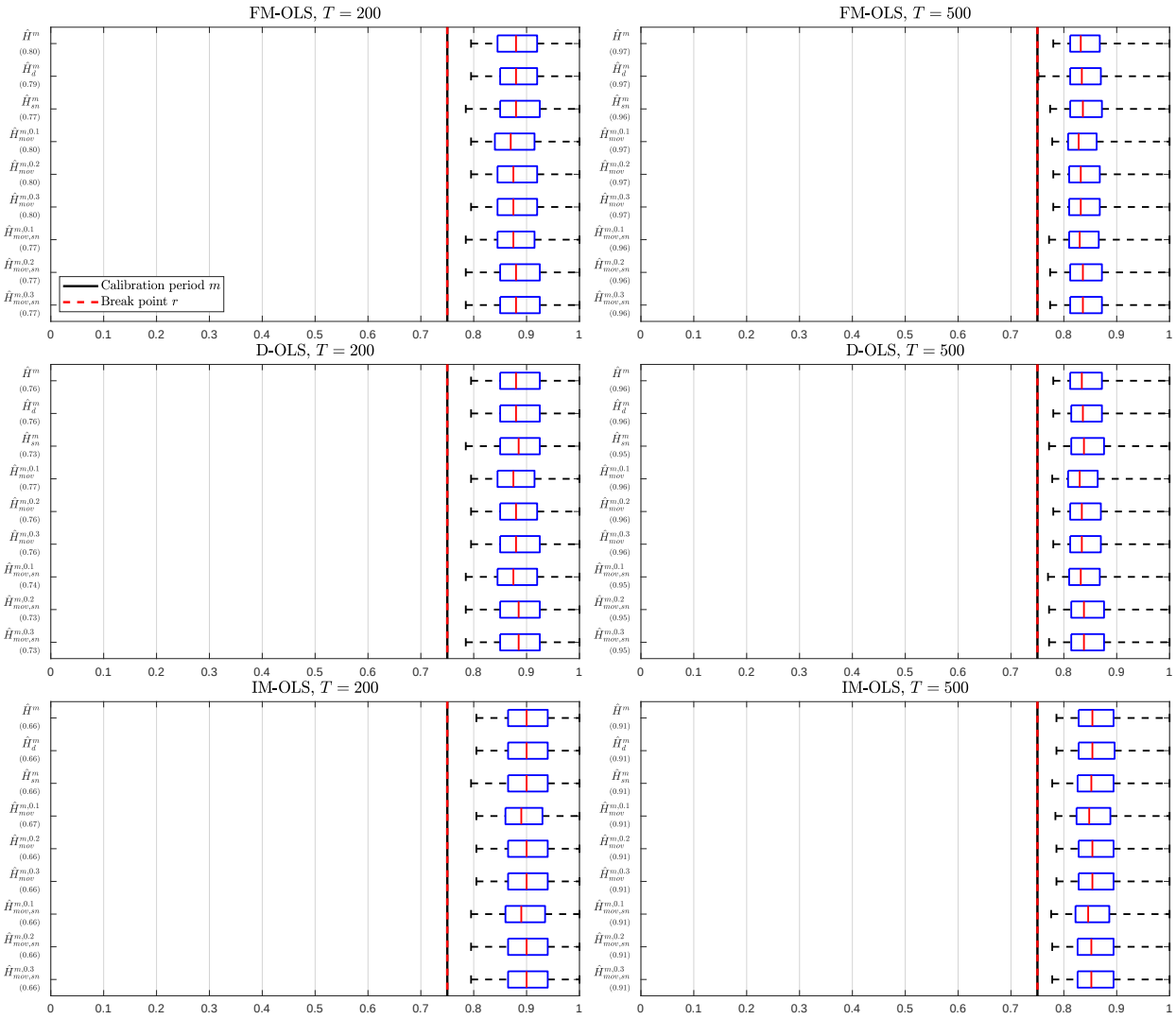

Figure 33: Detection times against $I(1)$ breaks for $m = 0.75$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$.

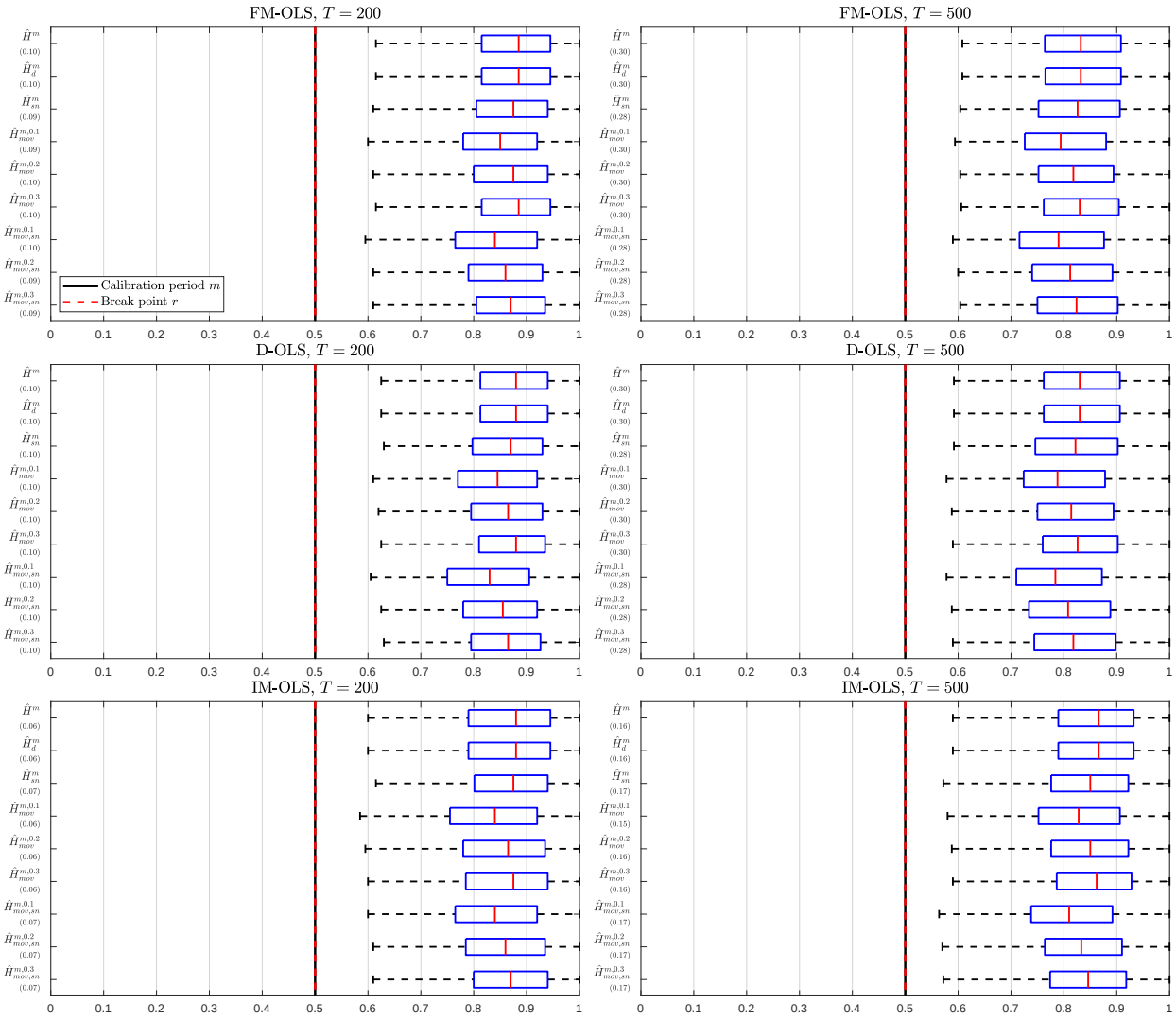

Figure 34: Detection times against $I(1)$ breaks for $m = 0.5$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.9$.

1.6.4 Additional Empirical Results

	ln(GDP)	
Country	ADF	PP
Australia	-1.217	-1.245
Austria	**-7.171**	**-5.324**
Belgium	-0.408	-0.485
Canada	-1.182	-1.044
Denmark	-2.422	-2.629
Finland	-2.178	-2.420
France	**-6.102**	**-3.880**
Germany	**-6.498**	**-5.337**
Italy	*-3.283*	**-4.295**
Japan	-1.724	-1.885
New Zealand	**-5.436**	**-6.237**
Norway	**-5.173**	**-5.221**
Portugal	-0.558	2.010
Spain	-2.415	-2.603
Sweden	-2.460	-2.775
Switzerland	**-5.356**	**-6.322**
United Kingdom	-3.221	*-3.384*
United States	-2.301	-2.547

Table 15: Results of the augmented Dickey and Fuller (1981) and Phillips and Perron (1988) unit root tests for log per capita GDP for the calibration period 1946–1973. The test equation includes intercept and linear trend. For the ADF test, we choose the lag length using the AIC criterion and the maximum lag length is three. *Italic* entries indicate rejection of the null hypothesis at the 10% level and **bold** entries rejection at the 5% level.

	ln(GDP)	
Country	ADF	PP
Australia	-1.146	-1.392
Austria	-2.457	**-5.292**
Belgium	0.123	0.046
Canada	-0.351	-0.377
Denmark	-0.429	-0.440
Finland	-1.151	-0.199
France	-2.069	-2.061
Germany	**-6.960**	**-32.623**
Italy	-0.670	-2.278
Japan	-1.143	-0.878
New Zealand	-2.757	**-3.581**
Norway	1.198	0.733
Portugal	0.089	0.391
Spain	0.505	0.974
Sweden	-1.900	-1.999
Switzerland	-2.154	-1.101
United Kingdom	-1.999	-1.197
United States	-1.303	-1.484

Table 16: Results of the augmented Dickey and Fuller (1981) and Phillips and Perron (1988) unit root tests for log per capita GDP for the full sample period 1946–2016. For further explanations see notes to Table 15.

	$p = 1$			$p = 2$	$p = 3$
	Shin	Johansen		CT	CT
Country		r = 0	r = 1		
Australia	0.075	**73.986**	5.418	0.079	0.055
Belgium	0.035	*23.343*	2.296	0.024	0.028
Canada	0.072	23.197	7.070	*0.091*	0.048
Denmark	0.056	21.761	3.702	0.057	0.056
Finland	**0.147**	12.999	3.362	0.058	0.040
Italy	0.076	*23.399*	5.211	0.065	0.025
Japan	0.045	**37.190**	3.672	0.021	0.025
Portugal	0.052	**43.125**	7.614	0.049	0.036
Spain	*0.099*	*25.045*	10.510	0.061	0.049
Sweden	0.044	*24.473*	6.706	0.034	0.033
United Kingdom	0.072	**30.967**	9.082	0.079	0.053
United States	**0.149**	**26.274**	5.105	0.057	0.049

Table 17: Results of the Shin (1994), Johansen (1995) and CT (Wagner, 2023) tests for log per capita CO_2 emissions and log per capita GDP. The (calibration) period is 1946–1973. For the Johansen test the number of lags is determined using the AIC and the maximum lag length is three. *Italic* entries indicate rejection of the null hypothesis at the 10% level and **bold** entries rejection at the 5% level.

	$p = 1$			$p = 2$	$p = 3$
	Shin	Johansen		CT	CT
Country		r = 0	r = 1		
Australia	0.060	**38.564**	*11.952*	0.059	0.055
Belgium	*0.100*	11.945	2.082	*0.104*	0.022
Canada	0.047	**29.940**	5.195	0.046	0.022
Denmark	**0.137**	17.271	4.976	0.086	0.033
Finland	**0.132**	*23.552*	5.571	0.063	0.031
Italy	*0.111*	**35.603**	5.176	*0.091*	0.028
Japan	**0.145**	14.479	5.003	0.059	0.049
Portugal	0.085	**29.134**	3.959	0.064	0.068
Spain	*0.099*	*25.004*	4.024	0.082	0.076
Sweden	**0.178**	**25.901**	8.170	0.053	0.050
United Kingdom	0.095	**26.260**	4.719	0.054	0.046
United States	**0.142**	**38.707**	**13.683**	**0.113**	0.047

Table 18: Results of the Shin (1994), Johansen (1995) and CT (Wagner, 2023) tests for log per capita SO_2 emissions and log per capita GDP. The (calibration) period is 1946–1973. For further explanations see notes to Table 17.

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

Table 19: Minimal polynomial degrees for cointegrating EKCs over the full sample period 1946–2016. The detailed cointegration test results are available in Tables 20 and 21.

	$p_1 = 1$			$p_1 = 2$	$p_1 = 3$
	Shin	Johansen		CT	CT
Country		r = 0	r = 1		
Australia	**0.165**	**55.183**	9.001	0.047	0.048
Belgium	0.087	22.624	7.421	0.073	0.058
Canada	0.057	22.511	3.539	0.055	0.053
Denmark	0.039	**30.477**	10.229	0.030	0.033
Finland	0.068	**31.796**	**14.518**	0.051	0.056
Italy	**0.121**	**27.923**	4.408	0.082	*0.089*
Japan	*0.111*	**28.798**	7.611	0.073	0.049
Portugal	**0.129**	9.548	3.273	0.057	0.067
Spain	0.055	23.116	6.730	0.056	0.052
Sweden	0.050	**29.520**	*10.818*	0.076	0.054
United Kingdom	**0.200**	**35.167**	7.089	0.068	0.041
United States	*0.102*	20.504	5.966	0.067	0.065

Table 20: Results of the Shin (1994), Johansen (1995) and CT (Wagner, 2023) tests for log per capita CO_2 emissions and log per capita GDP. The (full sample) period is 1946–2016. For further explanations see notes to Table 17.

	$p_1 = 1$			$p_1 = 2$	$p_1 = 3$
	Shin	Johansen		CT	CT
Country		r = 0	r = 1		
Australia	0.064	20.417	3.540	0.047	0.036
Belgium	*0.116*	20.564	4.520	0.062	0.033
Canada	*0.105*	**29.641**	*12.067*	0.060	0.071
Denmark	0.084	**28.038**	4.788	**0.107**	0.057
Finland	*0.109*	14.953	5.029	*0.096*	*0.100*
Italy	*0.101*	**40.453**	3.823	0.070	0.050
Japan	**0.127**	**40.635**	9.440	0.064	0.066
Portugal	*0.100*	13.291	5.239	0.068	0.077
Spain	0.089	18.603	8.650	*0.088*	0.070
Sweden	0.067	*23.999*	8.783	*0.097*	0.071
United Kingdom	**0.213**	**33.910**	9.928	**0.122**	0.065
United States	0.080	**32.551**	**14.114**	0.041	0.041

Table 21: Results of the Shin (1994), Johansen (1995) and CT (Wagner, 2023) tests for log per capita SO_2 emissions and log per capita GDP. The (full sample) period is 1946–2016. For further explanations see notes to Table 17.

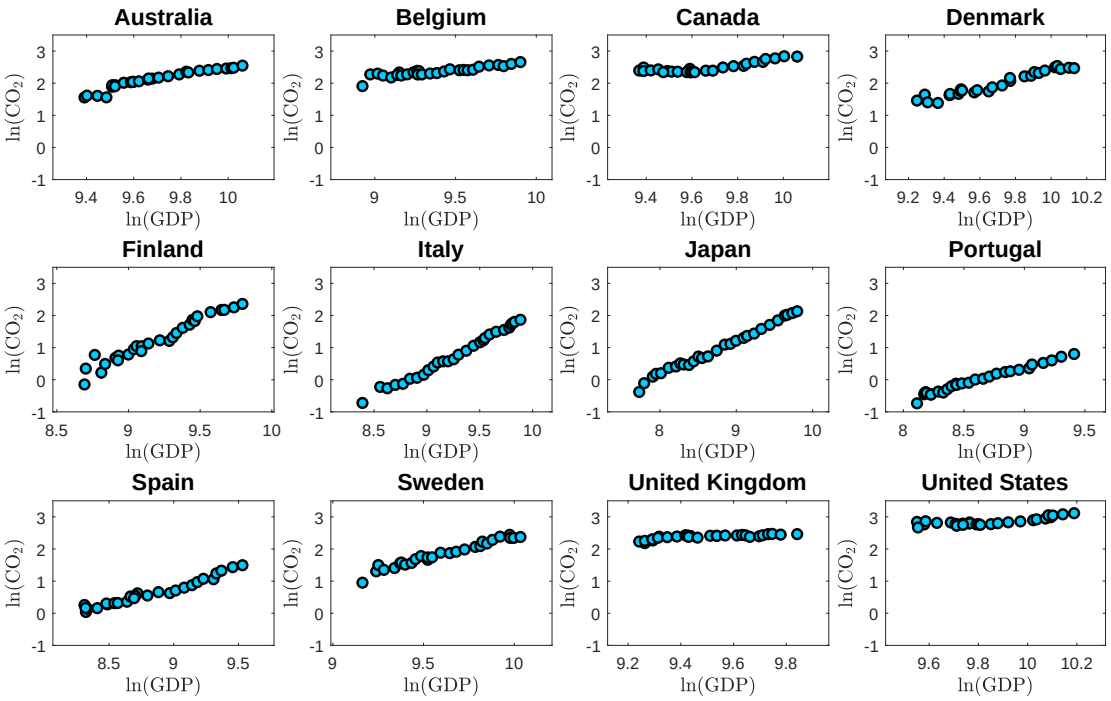

Figure 35: The scatter plots show pairs of observations of log per capita GDP and log per capita CO_2 emissions for the calibration period 1946–1973 for the considered twelve countries.

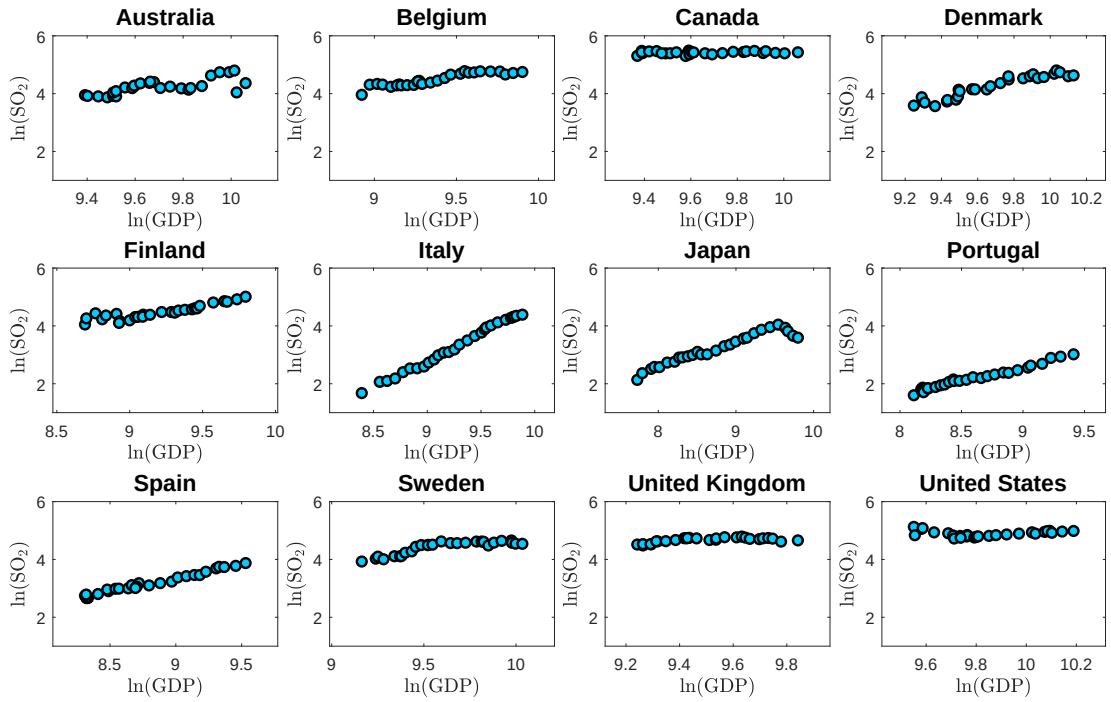

Figure 36: The scatter plots show pairs of observations of log per capita GDP and log per capita SO_2 emissions for the calibration period 1946–1973 for the considered twelve countries.

		$\hat{H}^{m}$			$\hat{H}^{m}_{d}$			$\hat{H}^{m}_{sn}$			$\hat{H}^{m,0.1}_{mov}$			$\hat{H}^{m,0.2}_{mov}$			$\hat{H}^{m,0.3}_{mov}$			$\hat{H}^{m,0.1}_{mov,sn}$			$\hat{H}^{m,0.2}_{mov,sn}$			$\hat{H}^{m,0.3}_{mov,sn}$		
Country	p	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM
Australia	1	1993	1993	2001	1993	1993	2001	1997	2001	2006	1990	1990	1997	1992	1992	1999	1993	1992	2001	1993	1997	2001	1995	1999	2003	1996	2000	2005
Belgium	1	1993	1992	2001	1993	1992	2001	1991	1990	1998	1989	1989	1993	1992	1991	1997	1993	1992	1999	1988	1988	1992	1990	1989	1995	1991	1990	1997
Canada	1	∞	∞	∞	∞	∞	∞	∞	1986	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	1982	∞	∞	1985	∞	∞	1986	∞
Denmark	1	1993	1991	2006	1993	1991	2006	1996	1992	2014	1989	1988	2003	1992	1990	2004	1993	1991	2005	1991	1988	2011	1994	1990	2012	1995	1991	2013
Finland	2	1990	1997	1992	1990	1997	1992	1992	2000	1993	1988	1994	1989	1989	1996	1991	1990	1997	1991	1989	1996	1990	1991	1998	1992	1992	1999	1993
Italy	1	1982	1982	1983	1982	1982	1983	1983	1982	1984	1980	1981	1981	1981	1982	1982	1981	1982	1983	1981	1980	1982	1983	1981	1983	1983	1981	1983
Japan	1	1985	1984	1985	1985	1984	1985	1985	1982	1982	1982	1981	1982	1984	1983	1984	1985	1983	1984	1982	1980	1980	1984	1981	1981	1985	1981	1982
Portugal	1	2000	∞	∞	2000	∞	∞	2002	∞	∞	1996	∞	∞	1998	∞	∞	1999	∞	∞	1998	∞	∞	2000	∞	∞	2001	∞	∞
Spain	1	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
Sweden	1	1983	1982	1985	1984	1982	1985	1983	1983	1985	1982	1981	1983	1983	1982	1984	1983	1982	1985	1982	1981	1983	1983	1982	1985	1983	1982	1985
United Kingdom	1	1984	1988	1988	1984	1988	1988	1986	1985	1991	1982	1986	1985	1983	1987	1987	1984	1988	1988	1984	1983	1987	1985	1985	1989	1986	1985	1991
United States	2	1989	1989	1992	1989	1989	1992	1990	1990	1995	1987	1987	1990	1988	1988	1992	1988	1989	1992	1988	1988	1992	1989	1990	1994	1990	1990	1994

Table 22: Detection times when monitoring CO_2 emissions using all nine detectors and all three estimation methods FM-OLS, D-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%.

		$\hat{H}^{m}$			$\hat{H}^{m}_{d}$			$\hat{H}^{m}_{sn}$			$\hat{H}^{m,0.1}_{mov}$			$\hat{H}^{m,0.2}_{mov}$			$\hat{H}^{m,0.3}_{mov}$			$\hat{H}^{m,0.1}_{mov,sn}$			$\hat{H}^{m,0.2}_{mov,sn}$			$\hat{H}^{m,0.3}_{mov,sn}$		
Country	p	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM	FM	D	IM
Australia	1	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
Belgium	1	1984	1984	1985	1984	1984	1985	1987	1986	1988	1983	1982	1983	1984	1983	1984	1984	1984	1985	1985	1984	1985	1986	1985	1987	1987	1986	1987
Canada	1	1981	1981	1982	1981	1981	1982	1981	1982	1982	1980	1980	1980	1981	1981	1981	1981	1981	1982	1980	1981	1981	1981	1982	1982	1981	1982	1982
Denmark	2	1994	1999	2007	1994	1999	2007	1999	1990	2009	1991	1997	2005	1993	1998	2006	1993	1999	2006	1997	1987	2006	1998	1989	2008	1999	1990	2008
Finland	2	1988	1987	1991	1988	1987	1991	1990	1987	1993	1986	1985	1989	1988	1987	1990	1988	1987	1990	1988	1986	1991	1989	1987	1992	1990	1987	1993
Italy	1	1981	1981	1982	1981	1981	1982	1984	1984	1984	1980	1979	1980	1981	1980	1981	1981	1981	1982	1982	1982	1982	1983	1983	1983	1984	1983	1984
Japan	2	1991	2005	2004	1991	2005	2004	1993	2007	2007	1989	2002	2002	1991	2004	2003	1991	2004	2004	1991	2005	2004	1992	2007	2006	1993	2007	2007
Portugal	1	2015	2013	∞	2015	2013	∞	∞	∞	∞	2013	2012	2015	2014	2013	∞	2015	2013	∞	2015	2015	∞	∞	2016	∞	∞	∞	∞
Spain	1	2003	2002	2011	2003	2002	2011	2007	2007	2013	1983	1983	1984	1986	2002	1987	2003	2002	2011	2005	2005	2012	2007	2007	2013	2007	2007	2014
Sweden	2	1986	1986	1989	1986	1986	1989	1987	1988	1990	1984	1984	1987	1985	1986	1988	1986	1986	1989	1985	1986	1988	1986	1988	1989	1986	1988	1990
United Kingdom	1	1982	1983	1984	1982	1983	1984	1985	1989	1985	1981	1981	1982	1982	1982	1983	1982	1983	1983	1983	1985	1983	1984	1987	1984	1985	1988	1985
United States	3	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞

Table 23: Detection times when monitoring SO_2 emissions using all nine detectors and all three estimation methods FM-OLS, D-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%.

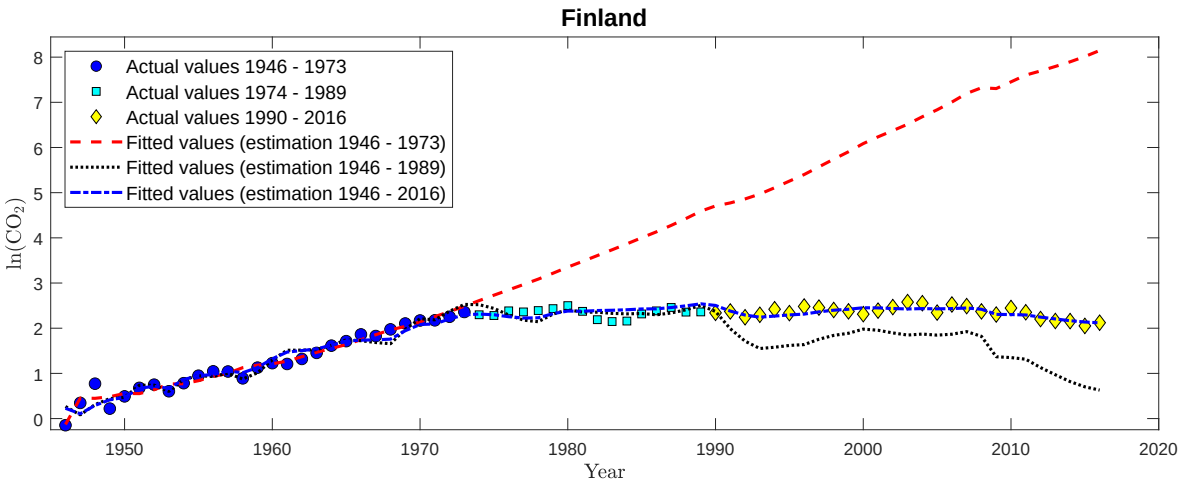

Figure 37: IM-OLS estimation results for a cointegrating quadratic relationship between log per capita GDP and log per capita CO_2 emissions for Finland. The figure shows pairs of observations of log per capita GDP and log per capita CO_2 emissions, for 1946–1973 (blue circles), 1974–1989 (turquoise squares) and 1990–2016 (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 1946–1989 and the dash-dotted blue line 1946–2016.

	1946–2016		
	θ_{D_1}	θ_{X_1}	θ_{X_2}
	Australia		
FM-OLS	0.004	**16.542**	**-0.786**
IM-OLS	0.007	**15.218**	**-0.729**
	Belgium		
FM-OLS	**-0.022**	3.355	-0.125
IM-OLS	**-0.017**	*4.869*	-0.213
	Canada		
FM-OLS	**-0.029**	0.624	0.056
IM-OLS	**-0.027**	2.256	-0.030
	Denmark		
FM-OLS	**-0.053**	**8.091**	*-0.265*
IM-OLS	**-0.051**	**8.548**	*-0.291*
	Finland		
FM-OLS	**-0.032**	**13.660**	**-0.593**
IM-OLS	**-0.029**	**15.157**	**-0.674**
	Italy		
FM-OLS	**-0.022**	**5.892**	**-0.207**
IM-OLS	**-0.014**	**11.367**	**-0.500**

	1946–2016		
	θ_{D_1}	θ_{X_1}	θ_{X_2}
	Japan		
FM-OLS	*-0.009*	**3.323**	**-0.123**
IM-OLS	-0.004	**5.070**	**-0.222**
	Portugal		
FM-OLS	**-0.021**	**-3.283**	**0.272**
IM-OLS	**-0.025**	**-4.487**	**0.343**
	Spain		
FM-OLS	**-0.034**	1.488	0.019
IM-OLS	**-0.036**	1.856	0.003
	Sweden		
FM-OLS	**-0.052**	**14.870**	**-0.621**
IM-OLS	**-0.066**	**10.487**	*-0.373*
	United Kingdom		
FM-OLS	**-0.032**	**9.229**	**-0.399**
IM-OLS	**-0.033**	**7.205**	**-0.295**
	United States		
FM-OLS	**-0.030**	**5.733**	**-0.204**
IM-OLS	**-0.023**	**8.414**	**-0.350**

Table 24: FM-OLS and IM-OLS estimation results for a cointegrating quadratic relationship between log per capita GDP and log per capita CO_2 emissions for the full sample period 1946–2016. *Italic* entries indicate significance of coefficients at the 10% level and **bold** entries significance of coefficients at the 5% level.

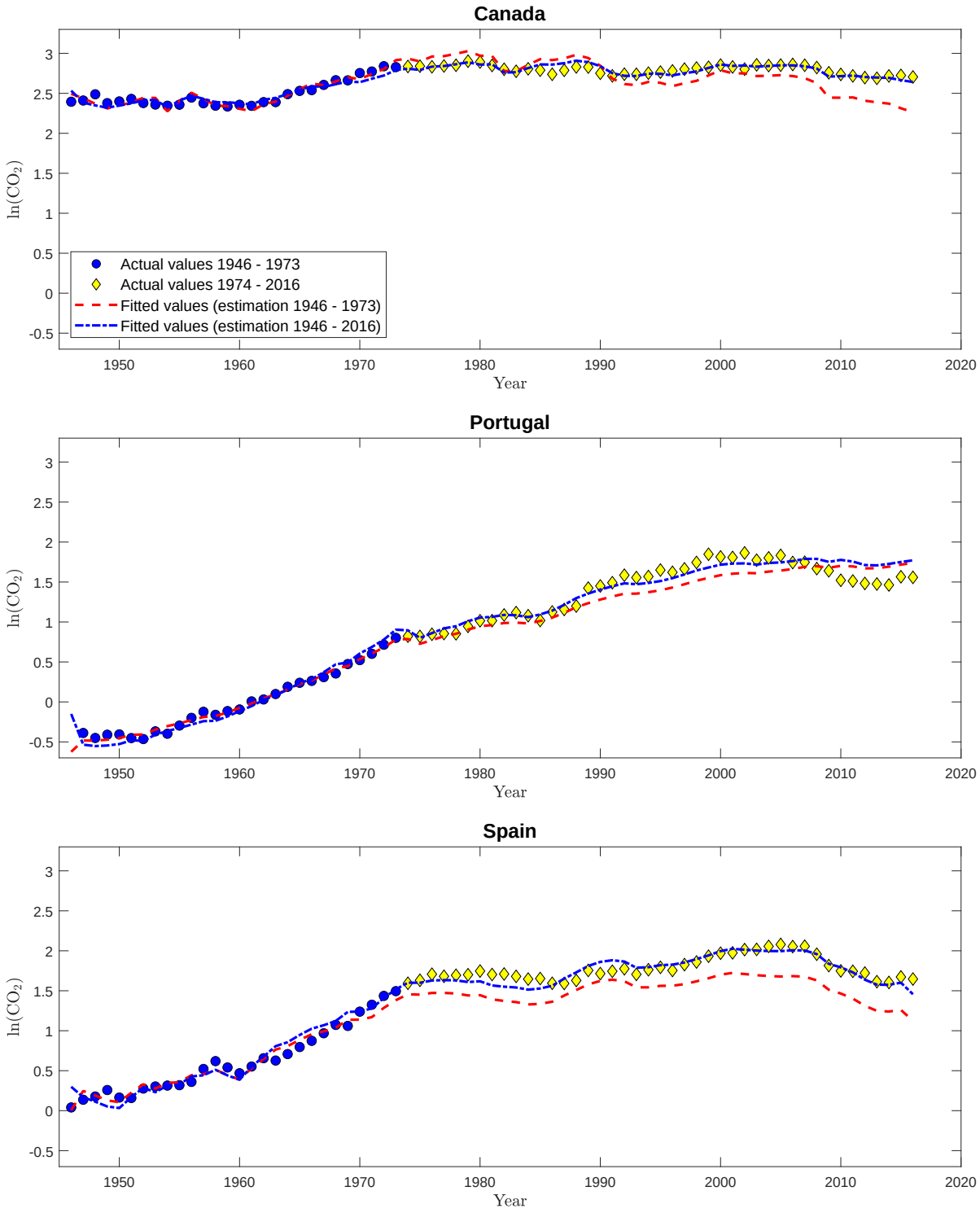

Figure 38: IM-OLS estimation results for a cointegrating linear relationship between log per capita GDP and log per capita CO_2 emissions for Canada, Portugal and Spain. For further explanations see notes to Figure 8 in the main text.

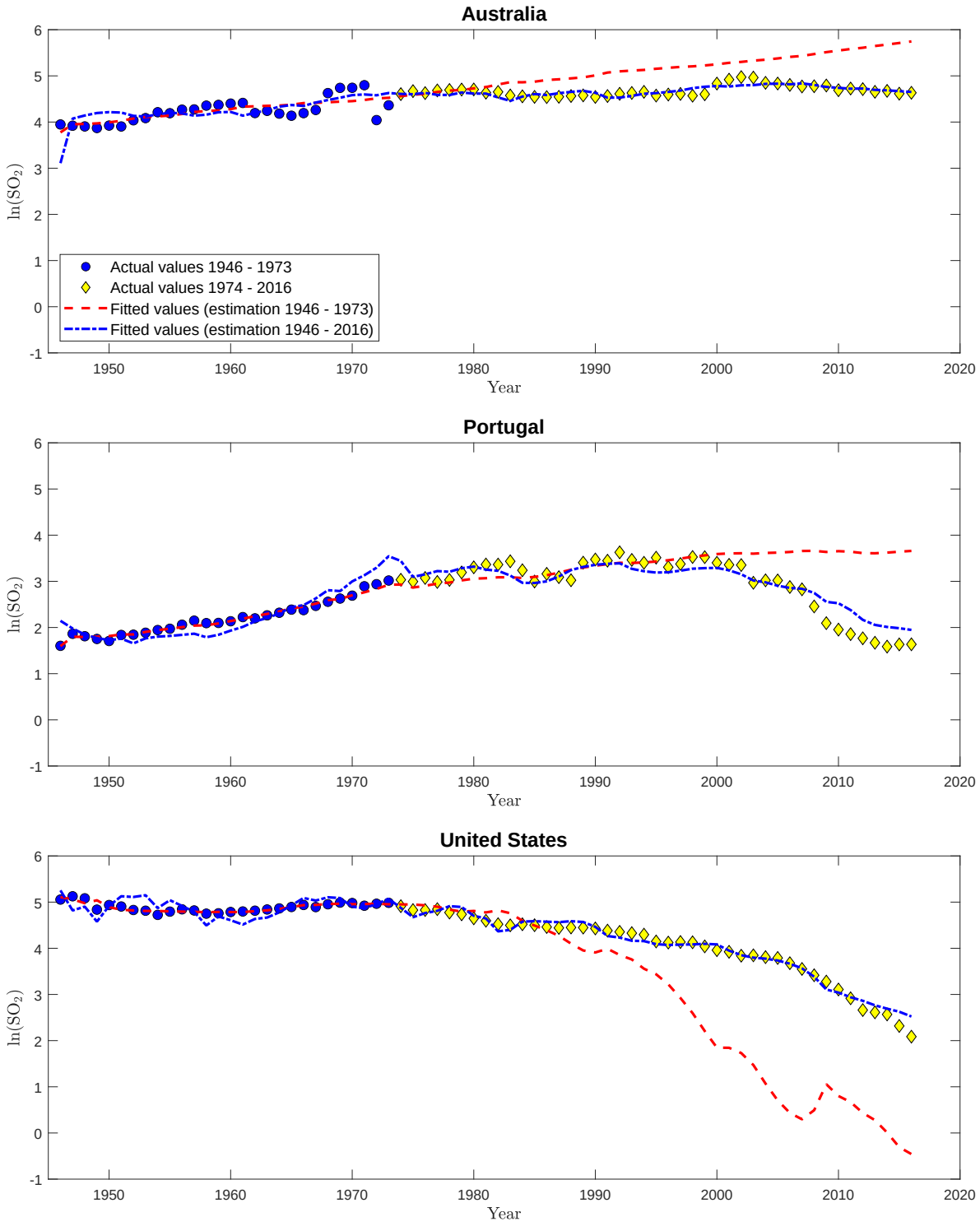

Figure 39: IM-OLS estimation results for a cointegrating linear relationship between log per capita GDP and log per capita SO_2 emissions for Australia and Portugal, and a cointegrating cubic relationship for the United States. For further explanations see notes to Figure 9 in the main text.

2 Comparison of Nonparametric Estimators and Specification Tests for Nonlinear Cointegrating Regressions: Results from a Large Scale Simulation Study

2.1 Introduction

When estimating a nonlinear relationship between integrated time-series the usage of nonparametric estimators seems to be a convenient choice, since the crucial decision of a specific parametric function does not need to be made. Nevertheless, asymptotic theory for nonparametric regression with integrated time-series has only been established recently in contrast to the case with stationary time-series. For nonparametric regression in the stationary case, see, e. g. , Härdle and Linton (1994) and for nonlinear parametric regression with integrated time-series, see, e. g. , Park and Phillips (2001). Going back to nonparametric regression, Karlsen and Tjøstheim (2001) were among the first and developed the theory for nonparametric estimation with kernel based estimators in a nonstationary environment. They considered null recurrent Markov chains as a regressor, to which also random walks belong. Building on that work, Karlsen *et al.* (2007) developed theory for nonparametric estimation of nonlinear cointegrating relationships. Since then, the nonparametric cointegrating literature has grown quite a lot. The finding of Wang and Phillips (2009b) that conventional kernel estimators can provide standard limiting distributions for cointegrating nonparametric regressions in the presence of error serial correlation and regressor endogeneity was quite striking. Wang and Phillips (2016) even more generalized the results for kernel based estimators and Linton and Wang (2016) proposed estimator versions to explicitly handle autoregressive error terms. For a more depth reading, Wang (2015) provides limit theorems, their derivations and far more details for kernel based estimation in nonlinear cointegrating regressions. Dong *et al.* (2016) followed a different route and perform nonparametric estimation using an orthogonal series expansion to approximate the unknown function. Dong and Linton (2018) and Dong *et al.* (2021) further extended and generalized that model

With nonparametric estimation being a convenient choice, parametric estimation has – in case of correct specification – the advantage of faster convergence rates compared to nonparametric estimation. Specification tests for nonlinear cointegrating regressions are of particular interest if, for example, after a nonparametric estimation, you want to check whether a certain functional form correctly describes the cointegrating relationship, which would lead to more precise results in a following parametric estimation, or if you want to conduct hypothesis tests for certain parame-

ters. Specification tests for nonlinear cointegrating regressions would also be of interest if certain functional forms are tested first and then a nonparametric estimation is performed for the cases in which the respective functional form was rejected. Along with the development of estimators for nonlinear and nonparametric cointegrating regressions, there has also been progress on specification tests for nonlinear cointegrating relationships. These tests are usually based on residuals from parametric nonlinear estimation using the parametric form in questions. Choi and Saikkonen (2010) build on residuals from their developed leads and lags estimator, next to residuals from nonlinear least squares estimation, where most of other tests, e. g. , those from Wang and Phillips (2012), Wang and Phillips (2016), Dong and Gao (2018), Wang *et al.* (2018) and Wang and Zhu (2020) only consider the latter residuals. However, this is not intended to homogenize all these tests, how the residuals are used and especially weighted, is based on very different approaches with different assumptions about the nonlinear cointegrating model, with more to less strict requirements on the error term and the integrated regressor.

Given the variety of estimators and specification tests, the question naturally arises which one to choose and how they perform under different circumstances. We attempt to answer these questions in this paper by first reviewing the estimators and tests and comparing whether their underlying assumptions, which we have thoroughly studied, allow for error serial correlation and regressor endogeneity, two typical problem that arise in cointegrating regressions. We then continue with a large scale simulation study to assess the performance of estimators and tests in finite samples, for a variety of functional forms, covering integrable and nonintegrable functions, and with varying degrees of error serial correlation and regressor endogeneity. Most of the considered estimators and tests rely on tuning parameters as bandwidths choices for the kernel based methods or truncation parameters and so on. The impact of varying these choices is also evaluated in this paper. All this gives guidance for practitioners which methods are useful in which situations and how crucial the choices of the specific tuning parameters are.

The paper is organized as follows: In Section 2.2 we present the setting, some assumptions and define further quantities. In Section 2.3 we present the considered estimators and in Section 2.4 the specification tests. Section 2.5 provides the simulation setup and presents the results. Section 2.6 briefly summarizes and concludes.

We use the following notation: Weak convergence is signified by $\Rightarrow$ for $T \to \infty$, the integer part of $x \in \mathbb{R}$ is denoted by $\lfloor x \rfloor$ and $\mathbb{E}(\cdot)$ denotes the expected value. The backward-shift operator is

denoted by L, i. e., $L\{z_t\}_{t\in\mathbb{Z}} \coloneqq \{z_{t-1}\}_{t\in\mathbb{Z}}$ and $\Delta \coloneqq 1 - L$ signifies the first-difference operator. Brownian motions are denoted with $B(r)$, with covariance matrices specified in the context.

2.2 Model and Assumptions

We consider a cointegrating regression model of the form

$$\begin{aligned} y_t &= f(x_t) + u_t \qquad\qquad t = 1, 2, \ldots, T, \\ x_t &= x_{t-1} + v_t, \end{aligned} \tag{51}$$

with x_t a nonstationary one-dimensional regressor and u_t and v_t zero mean stationary error terms. The real function $f(x)$ is unknown and the aim is to estimate $f(x)$ using nonparametric estimators, based on the observation pairs $\{y_t, x_t\}_{t=1}^T$. Nonparametric estimation compared to parametrics estimation has the advantage, that no functional form needs to be specified and thus, there is no risk of misspecification. On the other hand, if correctly specified, parametric estimation has the advantage of faster convergence rates. To test whether more structure ca be put on $f(x)$ and thus to allow for estimating a parametric model instead, we consider specification tests for nonlinear cointegrating regressions. Therefore, we test the regression function on specific parametric functional forms. Given a real function $g(x, \theta)$, with a vector of unknown parameters $\theta = [\theta_1, \ldots, \theta_m]'$ in the compact parameter space $\Theta_0 \subset \mathbb{R}^m$, and a cointegrating regression model as in equation (51), we are interested in testing the parametric null hypothesis

$$H_0 : f(x) = g(x, \theta_0), \tag{52}$$

for $\theta_0 \in \Theta_0$ and $x \in \mathbb{R}$.

Before we get to the considered estimators and specification tests, we introduce a few concepts and quantities. Some of the tests and estimators use *local time* processes in their asymptotic distribution. We denote with $L_W(1, 0)$ the local time process up to time 1, at point 0, of a standard Brownian motion $W(r)$. The local time can be defined as

$$L_W(1, 0) \coloneqq \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^1 \mathbb{1}_{\{|W(r)| \leq \epsilon\}}(r) dr, \tag{53}$$

(see, e. g., Wang and Phillips, 2009a) and can be interpreted as the time spent in a vicinity of 0 by $W(r)$, within the time interval $[0, 1]$. The cumulative distribution function of the local time process $L_W(1, 0)$ is given by $\left(2F_{\mathcal{N}(0,1)}(x) - 1\right) \times \mathbb{1}_{\{x \geq 0\}}(x)$ (see, e. g., Wang and Phillips, 2016), with $F_{\mathcal{N}(0,1)}(x)$ being the cumulative distribution function of the standard normal distribution.

Some estimators and tests allow for error serial correlation and regressor endogeneity and require long-run variance estimation, for which we define some further quantities. Assume that for the stacked process $\{w_t\}_{t\in\mathbb{Z}} \coloneqq \{[u_t, v_t]'\}_{t\in\mathbb{Z}}$ it holds that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor sT \rfloor} w_t \Rightarrow B(s) = \Omega^{1/2} W(s), \quad 0 \leq s \leq 1, \tag{54}$$

with $B(s) = [B_u(s), B_v(s)']'$ a vector of Brownian motions, the positive definite long-run covariance matrix

$$\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} \coloneqq \sum_{j=-\infty}^{\infty} \mathbb{E}\left(w_t w_{t-j}'\right),$$

and $W(s) = [W_{u\cdot v}(s), W_v(s)']'$ a vector of standard Brownian motions. Using, e. g. , the Cholesky decomposition of Ω yields:

$$\Omega^{1/2} = \begin{bmatrix} \omega_{u\cdot v} & \lambda_{uv} \\ 0 & \Omega_{vv}^{1/2} \end{bmatrix}, \tag{55}$$

where $\omega_{u\cdot v}^2 \coloneqq \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and $\lambda_{uv} \coloneqq \Omega_{uv}(\Omega_{vv}^{1/2})^{-1}$. For later usage, we define $\omega_{vv} \coloneqq \Omega_{vv}^{1/2}$.

One estimator is based on Hermite polynomials. Let the Hermite orthogonal polynomial sequence $\{H_\ell(x)\}_{\ell\in\mathbb{N}_0}$ be defined as

$$H_\ell(x) \coloneqq (-1)^\ell \exp(x^2) \frac{d^\ell}{dx^\ell} \exp(-x^2), \quad \ell \geq 0. \tag{56}$$

The inner product is given by $\langle g_1, g_2\rangle = \int_{-\infty}^{\infty} g_1(x) g_2(x) e^{-x^2} dx$ with the induced norm $\|g\|^2 = \langle g, g\rangle$. For the Hermite polynomials it follows that $\int_{-\infty}^{\infty} H_i(x) H_j(x) e^{-x^2} dx = \sqrt{\pi} 2^j j! \delta_{ij}$. Furthermore, define

$$h_\ell(x) \coloneqq (\sqrt{\pi} 2^\ell \ell!)^{-1/2} H_\ell(x), \tag{57}$$

which forms an orthonormal polynomial basis in $L^2(\mathbb{R}, e^{x^2})$ (see, e. g. , Dong *et al.*, 2021).

The individual assumptions of estimators and specification tests are provided in Appendix 2.7.1.

2.3 Estimators

We consider the following estimators for nonparametric cointegrating regressions. The first two sets of estimators consists of kernel based estimators. The first set includes the Nadaraya-Watson estimator and the local linear estimator. The second set includes estimators proposed by Linton and Wang (2016), which originate from the Nadaraya-Watson estimator and adjust for auto regressive error structures. The next estimator is proposed by Dong *et al.* (2021) and based on an orthogonal

expansion using Hermite polynomials. In the finite sample simulations in Section 2.5, we compare the performance of these estimators for different functional forms. Since in the simulations, the underlying regression functions are known by construction, we can evaluate the performance of those estimators against a parametric estimator using the known function $f(x)$. For this benchmark purpose serves the parametric D-OLS estimator of Choi and Saikkonen (2010).

"Conventional" Kernel Estimators

We start with two well known kernel estimators for $f(x)$, the Nadaraya-Watson estimator and the local linear estimator. Both estimators are widely used for nonparametric estimation of regression functions with stationary regressors. For cointegrating regressions, e. g. , Wang and Phillips (2016) and Wang (2015) provide asymptotic results for both estimators and show that even in case of regressor endogeneity and error serial correlation the self-normalized estimators converge to a standard normal distribution.

Define $K_h(s) \coloneqq h^{-1}K(s/h)$, where $K(\cdot)$ is a kernel function and the parameter h a bandwidth parameter with $h \coloneqq h_T \to 0$ for $T \to \infty$. The Nadaraya-Watson estimator approximates $f(x)$ locally by a constant and is given by

$$\hat{f}_{\mathrm{NW}}(x) \coloneqq \frac{\sum_{t=1}^{T} y_t K_h(x_t - x)}{\sum_{t=1}^{T} K_h(x_t - x)}. \tag{58}$$

The local linear estimator models $f(x)$, as the name already indicates, approximates $f(x)$ locally by a linear fit and is given by

$$\hat{f}_{\mathrm{LL}}(x) \coloneqq \frac{\sum_{t=1}^{T} K_h(x_t - x)(S_{T,2} - (x_t - x)S_{T,1})y_t}{\sum_{t=1}^{T} K_h(x_t - x)(S_{T,2} - (x_t - x)S_{T,1})}, \tag{59}$$

with $S_{T,j} = \sum_{t=1}^{T} K_h\left(x_t - x\right)\left(x_t - x\right)^j$. The $S_{T,j}$ terms stem from a locally weighted least squares regression. More details to both estimators can be found in, e. g. , Fan and Gijbels (1996). The Nadaraya-Watson estimator and local linear estimator are both consistent estimators for the unknown function $f(x)$. More precisely, for h satisfying $\sqrt{T}h \to \infty$ and $\sqrt{T}h^{1+2\gamma} \to 0$, with $0 < \gamma \leq 1$ from Assumption 8, for $T \to \infty$ it holds under Assumptions 5–8 that

$$\left(\frac{\sqrt{T}h}{\omega_{vv}}\right)^{1/2} \left(\hat{f}_{\mathrm{NW}}(x) - f(x)\right) \to_d L_W^{-1/2}(1,0)\, \mathcal{N}(0, \tau^2). \tag{60}$$

The self-normalized limit form is given by

$$\left(h\sum_{t=1}^{T}K_h(x_t-x)\right)^{1/2}\left(\hat{f}_{\text{NW}}(x)-f(x)\right)\to_d\mathcal{N}(0,\tau^2),\tag{61}$$

with $\tau^2:=\mathbb{E}(u_0^2)\int_{-\infty}^{\infty}K^2(s)ds$ (see Wang and Phillips, 2016, Theorem 2.1). The stated assumptions allow for regressor endogeneity and error serial correlation. For slightly stronger smoothness conditions on $f(x)$ and the kernel $K(\cdot)$, Wang and Phillips (2016) provide limit theory incorporating an asymptotically vanishing explicit bias term. Under Assumptions 5–8, for any h satisfying $\sqrt{T}h\to\infty$ and $\sqrt{T}h^{1+2(p+1)}\to 0$, the limiting distribution is given by

$$\left(\frac{\sqrt{T}h}{\omega_{vv}}\right)^{1/2}\left(\hat{f}_{\text{NW}}(x)-f(x)-\frac{h^pf^{(p)}(x)}{p!}\int_{-\infty}^{\infty}y^pK(y)dy\right)\to_d L_W^{-1/2}(1,0)\,\mathcal{N}(0,\tau^2),\tag{62}$$

and the self-normalized limit form by

$$\left(h\sum_{t=1}^{T}K_h(x_t-x)\right)^{1/2}\left(\hat{f}_{\text{NW}}(x)-f(x)-\frac{h^pf^{(p)}(x)}{p!}\int_{-\infty}^{\infty}y^pK(y)dy\right)\to_d\mathcal{N}(0,\tau^2).\tag{63}$$

and (for $p=2$) with the local linear estimator $\hat{f}_{\text{LL}}(x)$ leading to the same limiting distribution as the Nadaraya-Watson estimator (see Wang and Phillips, 2016, Theorem 2.2). The Nadaraya-Watson estimator and local linear estimator having the same explicit bias term in case of nonparametric cointegrating regression is of interest, since in the stationary case the Nadaraya-Watson estimator exhibits an additional linear bias term making the the local linear estimator advantageous over the Nadaraya-Watson estimator. This advantage of the local linear estimator is lost in case of nonparametric cointegrating regression with x_t being nonstationary (see, e. g. , Wang and Phillips, 2011).

If we choose $h=T^b$, then it has to hold that $-\frac{1}{2}<b<-\frac{1}{6}$ (the upper bound might be smaller than $-\frac{1}{6}$ depending on γ in Assumption 8).

Linton and Wang (2016)

Linton and Wang (2016) propose three estimators for nonparametric cointegrating regressions for the special case with an autocorrelation structure in the error term, e. g. , $u_t=\rho_1u_{t-1}+\cdots+\rho_pu_{t-p}+\zeta_t$. These estimators account for $AR(1)$ and $AR(p)$ type error serial correlation and are in those cases more efficient – in terms of smaller asymptotic variances – than the Nadaraya-Watson estimator. As a starting point for the Linton and Wang (2016) estimators serve the residuals from

Nadaraya-Watson estimation, i. e. , $\hat{u}_t \coloneqq y_t - \hat{f}_{\text{NW}}(x_t)$, while assuming that $\hat{f}_{\text{NW}}(x)$ is a consistent estimator for $f(x)$. The first estimator of Linton and Wang (2016) is based on

$$\hat{\rho} \coloneqq \frac{\sum\limits_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}}{\sum\limits_{t=2}^{n} \hat{u}_{t-1}^2}, \tag{64}$$

in combination with a Cochrane-Orcutt transformation to obtain an error process without serial correlation and is defined as

$$\hat{f}_{\text{LW1}}(x) \coloneqq \frac{\sum_{t=2}^{n} \left[y_t - \hat{\rho} y_{t-1} + \hat{\rho} \hat{f}_{\text{NW}}(x_{t-1}) \right) \Big] K((x_t - x)/h)}{\sum_{t=2}^{n} K((x_t - x)/h)}. \tag{65}$$

Under Assumptions 10, 11, 12(a), 12(b), 13(a) and $\sum_{i=0}^{\infty} i|\phi_i| < \infty$ the limiting distribution of $f_{\text{LW1}}(x)$ is given by

$$\left(h \sum_{t=1}^{T} K_h(x_t - x) \right)^{1/2} \Big(\hat{f}_{\text{LW1}}(x) - f(x) \Big) \to_d \mathcal{N}(0, \tau_2^2), \tag{66}$$

for any h satisfying $Th^{2+4\beta} \to 0$, $T^{\alpha} h^{2\beta} \to 0$, with $0 < \beta \leq 1$ from Assumption 13(a), and $T^{1-\delta_0} h^2 \to \infty$ for some $\delta_0 > 0$, with $\tau_2^2 \coloneqq \sigma^2 \int_{-\infty}^{\infty} K^2(s) ds$ and $\mathbb{E}(\nu_0^2) = \sigma^2$ (see Linton and Wang, 2016, Theorem 2.2). The assumptions of Linton and Wang (2016) allow for error serial correlation but not for regressor endogeneity. When choosing, e. g. , $h = T^b$, then it has to hold that $-\frac{1-\delta_0}{2} < b < -\frac{1}{6}$ (the upper bound might be smaller than $-\frac{1}{6}$ depending on α and β in Assumption 13). For slightly stronger smoothness conditions on $f(x)$ and the kernel $K(\cdot)$, also Linton and Wang (2016) provide limit theory incorporating an asymptotically vanishing explicit bias term which we omit here. In the described case with AR(1) errors, the Nadaraya-Watson estimator leads to a similar limiting distribution with $\tau_1^2 \coloneqq (1 - \rho^2)^{-1} \sigma^2 \int_{-\infty}^{\infty} K^2(s) ds$ instead of τ_2^2 in (66), and thus $\tau_2^2 < \tau_1^2$ for $0 < |\rho| < 1$. Similar in spirit, Linton and Wang (2016) propose an estimator with a Cochrane-Orcutt type transformation for the case of AR(p) errors, i. e. , $u_t = \rho_1 u_{t-1} + \cdots + \rho_p u_{t-p} + \zeta_t$, with $\zeta_j, j \in \mathbb{Z}$, i.i.d. random variables with $\mathbb{E}(\zeta_0) = 0$ and $\mathbb{E}(\zeta_0^2) = \sigma^2$ and ζ_j are independent of x_t. The estimator for the case of AR(p) errors is given by

$$\hat{f}_{\text{LW2}}(x) \coloneqq \frac{\sum_{t=k+1}^{n} \left[y_t - \sum_{j=1}^{k} \hat{\rho}_j (y_{t-j} - \hat{f}_{\text{NW}}(x_{t-j})) \right] K((x_t - x)/h)}{\sum_{t=k+1}^{n} K((x_t - x)/h)}, \tag{67}$$

with $(\hat{\rho}_1, \ldots, \hat{\rho}_k)'$ the OLS estimator of $(\rho_1, \ldots, \rho_k)'$ in the model $\hat{u}_t = \rho_1 \hat{u}_{t-1} + \cdots + \rho_k \hat{u}_{t-k} + e_t$. The lag length k has to be chosen if p is unknown, e. g. , by using information criteria. The limiting distribution of $f_{\text{LW2}}(x)$ in case of AR(p) errors is the same as the limiting distribution of $f_{\text{LW1}}(x)$

in case of AR(1) errors, and given in (66) with $f_{\text{LW2}}(x)$ replacing $f_{\text{LW1}}(x)$. Linton and Wang (2016) furthermore propose a third estimator, inspired by the method of Linton and Mammen (2008), which accounts for AR(1) errors and is more efficient than $\hat{f}_{\text{LW1}}$. This estimator is given by

$$\hat{f}_{\text{LWeff}}(x) \coloneqq \frac{1}{1+\hat{\rho}^2} \frac{\sum_{t=2}^{n-1} \widehat{y}_t K((x_t - x)/h)}{\sum_{t=2}^{n-1} K((x_t - x)/h)}, \tag{68}$$

with $\widehat{y}_t = \widehat{Z}_t^-(\hat{\rho}) - \hat{\rho}\widehat{Z}_{t+1}^+(\hat{\rho})$, where $\widehat{Z}_t^-(\hat{\rho}) \coloneqq y_t - \hat{\rho}y_{t-1} + \hat{\rho}\hat{f}(x_{t-1})$ and $\widehat{Z}_t^+(\hat{\rho}) \coloneqq y_t - \hat{\rho}y_{t-1} - \hat{f}(x_t)$, with $\hat{f}(x)$ and $\hat{\rho}$ being initial consistent estimators of $f(x)$ and ρ. Under Assumptions 10, 11, 12(a), 12(b), 13(a) and $\sum_{i=0}^{\infty} i|\phi_i|$ it holds that

$$\left(h\sum_{t=1}^{T} K_h(x_t - x)\right)^{1/2} \left(\hat{f}_{\text{LWeff}}(x) - f(x)\right) \to_d \mathcal{N}(0, \tau_3^2), \tag{69}$$

for any h satisfying $Th^{2+4\beta} \to 0$, $T^{\alpha}h^{2\beta} \to 0$ and $T^{1-\delta_0}h^2 \to \infty$ for some $\delta_0 > 0$, with $\tau_3^2 \coloneqq (1+\rho^2)^{-1}\int_{-\infty}^{\infty} K^2(s)ds$. Under the assumption of AR(1) errors it holds that $\tau_3^2 \leq \tau_2^2 \leq \tau_1^2$.

Dong, Linton and Peng (2021)

Dong *et al.* (2021) propose an estimator based on an orthogonal expansion rather than on kernel estimation, a weighted sieve estimator for nonparametric cointegrating regression functions. The general model includes a deterministic component, a stationary process and an integrated process as regressors and is given by

$$y_t = f(r_t, z_t, x_t) + u_t, \tag{70}$$

with deterministic component $r_t \coloneqq t/T$, stationary process z_t and integrated process x_t. Furthermore, τ_t is defined on $[0,1]$, the support of z_t is V_z (either bounded or unbounded support) and x_t has unbounded support. This model is larger than the one-dimensional what we used so far. From this model, we consider a special case with only an integrated process x_t as single regressor.

We start with some technicalities. Similar to Dong *et al.* (2021), we assume that $f(r,z,x) \in L^2([0,1]\times V_z \times \mathbb{R}, \phi(z,x)) = \{g(r,z,x) : \iiint_{[0,1]\times V_z\times\mathbb{R}} g^2(r,z,x) \times\phi(z,x)drdzdx < \infty\}$, with $\phi(z,x)$ a density function defined on $V_z \times \mathbb{R}$. Let $\phi(z,x) \coloneqq m(z)\exp(-x^2)$, with $m(z)$ given by Assumption 15. As a basis in $L^2([0,1]\times V_z\times\mathbb{R}, \phi(z,x))$ we use the tensor product of the bases chosen from $L^2[0,1] = \{u(r) : \int_0^1 u^2(r)dr < \infty\}$, $L^2(V_z, m(z)) = \{p(z) : \int_{V_z} p^2(z)m(z)dz < \infty\}$ and $L^2(\mathbb{R}, e^{-x^2}) = \{g(x) : \int_{-\infty}^{\infty} g^2(x)e^{-x^2}dx < \infty\}$. We start with a basis in the Hilbert space $L^2[0,1]$.

Define $\varphi_0(r) \coloneqq 1$ and $\varphi_i(r) \coloneqq \sqrt{2}\cos(\pi i r)$ for $i \geq 1$. Then, $\{\varphi_i(r)\}_{i\in\mathbb{N}_0}$ is an orthonormal basis in $L^2[0,1]$. The inner product is given by $\langle u_1, u_2\rangle = \int_0^1 u_1(r)u_2(r)dr$ for any $u_1(\cdot), u_2(\cdot) \in L^2[0,1]$, with the induced norm $\|u\|^2 = \langle u, u\rangle$ for any $u(\cdot) \in L^2[0,1]$. It follows that $\langle \varphi_i(r), \varphi_j(r)\rangle = \delta_{ij}$ the Kronecker delta. The orthogonal function sequence $\{p_j(z)\}_{j\in\mathbb{N}_0}$ from Assumption 15 serves as an orthonormal basis of $L^2(V_z, m(z))$.[47] The Hermite orthogonal polynomial sequence $\{H_\ell(x)\}_{\ell\in\mathbb{N}_0}$, defined by

$$H_\ell(x) \coloneqq (-1)^\ell \exp(x^2)\frac{d^\ell}{dx^\ell}\exp(-x^2), \quad \ell \geq 0, \tag{71}$$

forms a basis in $L^2(\mathbb{R}, e^{-x^2})$. The inner product is given by $\langle g_1, g_2\rangle = \int_{-\infty}^{\infty} g_1(x)g_2(x)e^{-x^2}dx$ with the induced norm $\|g\|^2 = \langle g, g\rangle$. For the Hermite polynomials it follows that $\int_{-\infty}^{\infty} H_i(x)H_j(x)e^{-x^2}dx = \sqrt{\pi}2^j j!\delta_{ij}$. Define

$$h_\ell(x) \coloneqq (\sqrt{\pi}2^\ell \ell!)^{-1/2} H_\ell(x), \tag{72}$$

which is an orthonormal polynomial basis in $L^2(\mathbb{R}, e^{-x^2})$. The tensor product $\{\varphi_i(r)\}_{i\in\mathbb{N}_0} \otimes \{p_j(z)\}_{j\in\mathbb{N}_0} \otimes \{h_\ell(r)\}_{\ell\in\mathbb{N}_0}$ is an orthonormal basis in $L^2([0,1] \times V_z \times \mathbb{R}, \phi(z,x))$. Then, the unknown function $f(r,z,x)$ can be represented by an orthogonal expansion

$$f(r,z,x) = \sum_{i,j,l=0}^{\infty} c_{ijl}\varphi_i(r)p_j(z)h_l(x) \tag{73}$$

(see Dong *et al.*, 2021). To obtain a nonparametric estimate for $f(r,z,x)$, the orthogonal expansion is truncated and the parameters $c_{ij\ell}$ are estimated using weighted least squares. The truncation parameters k_1, k_2, k_3 are positive integers, with $k \coloneqq (k_1, k_2, k_3)$ and $K \coloneqq k_1k_2k_3$. Defining the quantities

$$Z_k(\tau_t, z_t, x_t) \coloneqq \begin{pmatrix} \phi_0(\tau_t)p_0(z_t)h_0(x_t) \\ \phi_0(\tau_t)p_0(z_t)h_1(x_t) \\ \vdots \\ \phi_{k_1-1}(\tau_t)p_{k_2-1}(z_t)h_{k_3-1}(x_t) \end{pmatrix} \in \mathbb{R}^K, \quad Z_{T,K} \coloneqq \begin{pmatrix} Z_k(\tau_1, z_1, x_1)' \\ Z_k(\tau_2, z_2, x_2)' \\ \vdots \\ Z_k(\tau_T, z_T, x_T)' \end{pmatrix} \in \mathbb{R}^{T\times K}, \tag{74}$$

$y \coloneqq (y_1, \ldots, y_T)'$, and the weighting matrix $W_T \coloneqq \mathrm{diag}(\phi(z_1, x_1), \ldots, \phi(z_T, x_T))$, then the weighted least squares estimator given by

$$\hat{c} = (Z_{T,K}' W_T Z_{T,K})^{-1} Z_{T,K}' W_T y, \tag{75}$$

leads to an estimation of the nonparametric function $f(r,z,x)$, which is then given by

$$\hat{f}_{\mathrm{DLP}}(r,z,x) = Z_k(r,z,x)'\hat{c}. \tag{76}$$

[47] Since we consider the special case without a stationary regressor in our simulations and application, we do not discuss the choice of the density function $m(z)$ and orthogonal functions $p_j(z)$ further here and refer to Dong *et al.* (2021) for a detailed discussion.

Under Assumptions 14, 15 and 16 the limit of the self-normalized estimator is given by

$$\Sigma_T^{-1}(r,z,x)\left(\hat{f}_{\text{DLP}}(r,z,x) - f(r,z,x)\right) \to_d \mathcal{N}(0,1) \tag{77}$$

with

$$\Sigma_T^2(r,z,x) \coloneqq \sigma_u^2 Z_k(r,z,x)'(Z_{T,K}'W_T Z_{T,K})^{-1} Z_{T,K}' W_T^2 Z_{T,K}(Z_{T,K}'W_T Z_{T,K})^{-1} Z_k(r,z,x). \tag{78}$$

The estimator $\hat{\sigma}^2_{\text{DLP}}$ accounts for the weighted least squares approach and is a consistent estimator for σ_u^2 under Assumptions 14, 15 and 16 and given by

$$\hat{\sigma}^2_{\text{DLP}} = \left(\sum_{t=1}^{T} \phi(z_t, x_t)\right)^{-1} \sum_{t=1}^{T} \hat{u}_t^{\text{DLP}} \phi(z_t, x_t), \tag{79}$$

with $\hat{u}_t^{\text{DLP}} \coloneqq y_t - \hat{f}_{\text{DLP}}(r_t, z_t, x_t)$. The assumptions of Dong *et al.* (2021) do not allow for regressor endogeneity neither for error serial correlation.

In their simulation study, Dong *et al.* (2021) minimize a generalized cross validation function to select the truncation parameters k_1, k_2, k_3:

$$(\hat{k}_1, \hat{k}_2, \hat{k}_3) = \underset{k_1,k_2,k_3}{\arg\min} \frac{(y - Z_{T,K}\hat{c})'W_T(y - Z_{T,K}\hat{c})}{T(1 - \frac{k_1 k_2 k_3}{T})^2}. \tag{80}$$

In our simulations and application, we use a special case of the general model in (70), considering only an integrated process x_t, without any deterministic component r_t and any stationary process z_t, and thus we consider in the weighting matrix W_T only functions $\phi(x)$. Dong *et al.* (2021) use in their simulation study special cases considering only subsets of the three proposed regressors r, z, x as well.

Dynamic Ordinary Least Squares

As last estimator, we consider the estimator of Choi and Saikkonen (2010), who propose a version of their dynamic ordinary least squares (D-OLS) estimator for nonlinear cointegrating regressions. This estimator is in contrast to the other considered estimators a parametric estimator and serves as a benchmark in the finite sample simulations in Section 2.5, where by design the parametric functions behind the data generating processes are known, i. e. , $f(x_t) = g(x_t, \theta)$ for some function $g(\cdot,\cdot)$. The idea behind the D-OLS estimator is to include leads and lags of the first differences of the regressor in the regression equation, which leads to a nuisance parameter free limiting distribution even in case of regressor endogeneity and error serial correlation.

The D-OLS estimator of Choi and Saikkonen (2010) is a two-step estimator, with the first step estimating θ in

$$y_t = g(x_t, \theta) + u_t, \quad t = 1, \ldots, T. \tag{81}$$

using nonlinear least squares (NLLS). Expressing the error term as $u_t \coloneqq \sum_{j=-\infty}^{\infty} \pi_j' v_{t-j} + \epsilon_t$, where ϵ_t is a stationary zero mean process and with defining $e_t \coloneqq \epsilon_t + \sum_{j \in \mathbb{Z}\,:\, j \notin \{-k_2, \ldots, k_1\}} \pi_j' v_{t-j}$, Equation (81) can be written as leads and lags augmented regression by

$$y_t = g(x_t, \theta) + \sum_{j=-k_2}^{k_1} \pi_j' \Delta x_{t-j} + e_t, \quad t = k_1 + 2, \ldots, T - k_2, \tag{82}$$

and in a more compact fashion as

$$y_t = g\left(x_t, \theta\right) + V_t' \pi + e_t, \quad t = k_1 + 2, \ldots, T - k_2, \tag{83}$$

with $V_t \coloneqq \left[\Delta x_{t+k_2}', \ldots, \Delta x_{t-k_1}'\right]'$, and $\pi = \left[\pi_{-k_2}', \ldots, \pi_{k_1}'\right]'$. Using the NLLS estimator $\tilde{\theta}$ as an initial estimator, the D-OLS two-step estimator is defined by

$$\begin{bmatrix} \hat{\theta}^{(1)} \\ \hat{\pi}^{(1)} \end{bmatrix} = \begin{bmatrix} \tilde{\theta} \\ 0 \end{bmatrix} + \left(\sum_{t=k_1+2}^{T-k_2} \tilde{p}_t \tilde{p}_t' \right)^{-1} \sum_{t=k_1+2}^{T-k_2} \tilde{p}_t' \tilde{u}_t, \tag{84}$$

with $\tilde{u}_t = y_t - g\left(x_t, \tilde{\theta}\right)$, $\tilde{p}_t = \left[\tilde{K}\left(x_t, \tilde{\theta}\right)', V_t'\right]'$ and $\tilde{K}\left(x_t, \tilde{\theta}\right) = \left.\frac{\partial g(x_t, \theta)}{\partial \theta}\right|_{\theta = \tilde{\theta}}$. Furthermore, the residuals obtained by using the two-step estimator are defined as $\hat{e}_t = y_t - g\left(x_t, \hat{\theta}^{(1)}\right) - V_t' \hat{\pi}^{(1)}, \quad t = k_1 + 2, \ldots, T - k_2$.

Properties Estimators

The estimators are based on different assumptions with respect to the error term and potential correlation with the regressor, for which their proposed limiting distributions hold. With exception of the DLP estimator, all considered estimators allow for error serial correlation in their assumptions. The estimators' assumptions with respect to regressor endogeneity are more restrictive. Only the Nadaraya-Watson estimator and local-linear estimator allow for regressor endogeneity. Table 25 summarizes whether the estimators' assumptions allow for error serial correlation and for regressor endogeneity. Furthermore, the table shows necessary user choices which have to be made prior to estimation. We did not include the D-OLS estimator in Table 25, which serves as benchmark, and allows for both error serial correlation and regressor endogeneity.

	Correlation	Endogeneity	User Choices
$\hat{f}_{\text{NW}}$	✓	✓	kernel, bandwidth
$\hat{f}_{\text{LocLin}}$	✓	✓	kernel, bandwidth
$\hat{f}_{\text{LW1}}$	✓	✗	kernel, bandwidth
$\hat{f}_{\text{LW2}}$	✓	✗	kernel, bandwidth, AR order
$\hat{f}_{\text{LWeff}}$	✓	✗	kernel, bandwidth
$\hat{f}_{\text{DLP}}$	✗	✗	weighting function, approximation order

Table 25: The checkmarks indicate whether the specific assumptions of the estimators allow for error serial correlation and regressor endogeneity.

Bandwidth Selection

The kernel based estimators and specification tests require the choice of a bandwidth parameter. To overcome the problem of how to choose the bandwidth parameter h, Wang and Phillips (2023) propose two rules for optimal bandwidth selection for kernel based estimators and specification tests in nonlinear cointegrating regressions. One rule provides an optimal pointwise bandwidth, thus returning $h(x)$ for values of x, and the other rule provides an optimal weighted bandwidth, giving a single value h. In our finite sample simulations, we compare for kernel based estimators and tests their performance with respect to these optimal bandwidth selections and deterministic bandwidth selections as, e. g. , $h = T^b$ with b being a constant in the range $-\frac{1}{2} < b < -\frac{1}{6}$. Optimality of the bandwidth selections refers to the conditional MSE criterion, applied on the local linear estimator:

$$\mathbb{E}\left[\left(\hat{f}_{\text{LL}}(x) - f(x)\right)^2 \mid x_1, \ldots, x_T\right], \tag{85}$$

and the conditional weighted average MSE:

$$\int_{-\infty}^{\infty} \mathbb{E}\left[\left(\hat{f}_{\text{LL}}(x) - f(x)\right)^2 \mid x_1, \ldots, x_T\right] \pi(x) dx, \tag{86}$$

where $\pi(x)$ is a weighting function with compact support. Conditional expectations are used here, since unconditional expectations are undefined in the nonstationary case. The optimal pointwise and weighted bandwidths are given by

$$h_{opt}(x) = \left[\frac{\sigma^2(x) \int_{-\infty}^{\infty} K^2(t) dt}{[\tau f''(x)]^2}\right]^{1/5} A_T^{-1/5} \tag{87}$$

and

$$h_{opt} = \left[\frac{\int_{-\infty}^{\infty} \sigma^2(x) W(x) dx \int K^2(x) dx}{\int_{-\infty}^{\infty} \left[f''(x) \right]^2 \pi(x) dx \ \tau^2} \right]^{1/5} A_T^{-1/5}, \tag{88}$$

with $A_T = \sum_{t=1}^{T} K(x_t)$ (see Wang and Phillips, 2023). An estimator for the error variance $\sigma^2(x)$ is given by

$$\hat{\sigma}^2(x) \coloneqq \frac{\sum_{t=1}^{T} (y_t - \hat{f}(x_t))^2 K_h(x_t - x)}{\sum_{t=1}^{T} K_h(x_t - x)}, \tag{89}$$

see, e. g. , Wang and Wang (2013). Both bandwidths can also be calculated using $A_{T,h} = \frac{1}{h} \sum_{t=1}^{T} K($ $(x_t - x)/h)$ (see Wang and Phillips, 2023) and we consider both versions in our simulations, labeling the ones using $A_{T,h}$ as $h_{opt,h}(x)$ and $h_{opt,h}$. For the optimal bandwidth selection, the second derivative of the regression function $f(x)$ is needed. In our simulation study, we calculate the optimal bandwidths once with the true second derivative of the known regression function and once with an estimate of the second derivative of the unknown regression function, which is done using the method of Mack and Müller (1989).

For the specification tests which require bandwidth choices, we do not consider optimal pointwise bandwidths in our simulation study. For one of the later described tests, integration over x from $-\infty$ to $+\infty$ takes place and thus a pointwise bandwidth would need to be calculated at each value of x used in the numerical integration.

2.4 Specification Tests

There has been a lot of development in the area of specification testing for nonlinear cointegrating relationships in recent years. We consider tests for the null hypothesis $H_0 : f(x) = g(x, \theta_0)$ in nonlinear cointegrating regressions. The first one we consider is the test of Choi and Saikkonen (2010) which is based on the residuals of their two-step estimator described above. The test of Dong and Gao (2018) is based on an integral of the difference of a nonparametric estimation of $f(x)$ and a parametric estimation of $g(x, \theta)$. In this case, the nonparametric estimation is based on Hermite polynomials, similar to the Dong *et al.* (2021) estimator. The test of Wang and Phillips (2012) goes into a different direction and uses a normalized, kernel-weighted sum of the cross products of the residuals from parametric estimation. In a similar fashion works the test of Wang and Phillips (2016) but using the integral of a weighting function and the kernel-weighted residuals from parametric estimation. Wang *et al.* (2018) use as test statistic the supremum of a marked empirical process of the residuals from parametric estimation. The test of Wang and Zhu (2020) uses a portmanteau test statistic of autocorrelation-corrected residuals from parametric estimation.

Choi and Saikkonen (2010)

Choi and Saikkonen (2010) propose several tests for testing nonlinear cointegration. The test statistics are based on either the residuals $\tilde{u}_t$ from NLLS estimation of (81) or the residuals $\hat{e}_t$ from the two-step DOLS estimation of (83) (see Section 2.3). The first test statistic we consider from Choi and Saikkonen (2010) is given by

$$C_{LL} \coloneqq (T - k_1 - k_2 - 1)^{-2} \tilde{\omega}_{u \cdot v}^{-2} \sum_{t=k_1+2}^{T-k_2} \left(\sum_{j=k_1+2}^{t} \hat{e}_j \right)^2 ,$$

with $\tilde{\omega}_{u \cdot v}^2$ a consistent estimator of the long-run variance $\omega_{u \cdot v}^2$. Let $k_1, k_2 \to \infty$ such that $k_1^3/T, k_2^3/T \to 0$ and $T^{1/2} \sum_{j \in \mathbb{Z} \,:\, j \notin \{-k_2, \ldots, k_1\}} \|\pi_j\| \to 0$. Under Assumptions 36–41 and with consistent long-run variance estimation, it holds under the null hypothesis that

$$C_{LL} \Rightarrow \omega_{u \cdot v}^{-2} \int_0^1 \left[B_{u \cdot v}(s) - \int_0^s \tilde{K}(B_v^0(r), \theta_0)' dr \; \chi\left(B_v^0, \theta_0\right) \right]^2 ds,$$

with $\chi\left(B_v^0, \theta_0\right) \coloneqq \left(\int_0^1 \tilde{K}\left(B_v^0(z), \theta_0\right) \tilde{K}\left(B_v^0(z), \theta_0\right)' dz \right)^{-1} \int_0^1 \tilde{K}\left(B_v^0(z), \theta_0\right) dB_{u \cdot v}(z)$ and $B_{u \cdot v} \coloneqq \omega_{u \cdot v} W_{u \cdot v}$, with $W_{u \cdot v}$ independent of B_v^0 (see Choi and Saikkonen, 2010). In case of, e. g. , a polynomial regression model, the limiting distribution is nuisance paramter-free up to a scalar variance

parameter, which can be consistently estimated and thus scaled out (see Choi and Saikkonen, 2010, p. 690).

The second and third test statistics we consider from Choi and Saikkonen (2010) are based on subsamples of residuals, denoted as subresiduals, of *block size* b, which is dependent on the sample size. The motivation behind using subresiduals with a certain block size is the following: A block size b, that increases with a slower rate than the sample size, results in a limiting distribution of the test statistic independent of the regressors (see Choi and Saikkonen, 2010, p. 691). That result for the limiting distribution holds for both test statistic versions, one based on subresiduals from nonlinear least squares estimation and the other based on subresiduals from the two-step estimation as described in Section 2.3. The subresidual based test statistics are given by

$$C_{NLLS}^{b,i} \coloneqq b^{-2}\tilde{\omega}_{i,u}^{-2} \sum_{t=i}^{i+b-1} \left(\sum_{j=i}^{t} \tilde{u}_j \right)^2, \tag{90}$$

and

$$C_{LL}^{b,i} \coloneqq (b-k_1-k_2-1)^{-2}\tilde{\omega}_{i,u\cdot v}^{-2} \sum_{t=i+k_1+2}^{i+b-k_2} \left(\sum_{j=i+k_1+2}^{t} \hat{e}_j \right)^2, \tag{91}$$

with i denoting the starting point of the subresiduals and b denoting the block size. Suppose that Assumptions 36–41 hold and let $\tilde{\omega}_{i,u}^2$ be a consistent estimator of $\omega_u^2 \coloneqq \Omega_{uu}$. If $b \to \infty$ and $b/T \to 0$ as $T \to \infty$, for any i with $1 \leq i \leq T-b$, it holds under the null hypothesis that

$$C_{NLLS}^{b,i} \Rightarrow \int_0^1 W^2(s)ds,$$

where $W(s)$ is a standard Brownian motion (see Choi and Saikkonen, 2010). A similar result holds for the test statistic based on residuals from the two-step estimation. Suppose that Assumptions 36–41 hold and that $k_1, k_2 \to \infty$ in such a way that $k_1^3/T, k_2^3/T \to 0$ and $T^{1/2}\sum_{j\notin\{-k_2,\ldots,k_1\}}\|\pi_j\| \to 0$. Furthermore, assume that $\tilde{\omega}_{i,u\cdot v}^2$ is a consistent estimator of $\omega_{u\cdot v}^2$. If $b \to \infty$ and $b/T \to 0$ as $T \to \infty$, for any i with $1 \leq i \leq T-b$, it holds under the null hypothesis that

$$C_{LL}^{b,i} \Rightarrow \int_0^1 W^2(s)ds.$$

Choi and Saikkonen (2010) suggest to jointly consider multiple of one of the above test statistics, using different starting points i, together with a Bonferroni procedure, as the individual tests are likely to have lower power compared tests using the full residuals, i. e. , not subresiduals. For a joint test statistic, select M statistics $C_{NLLS}^{b,i_1}, \ldots, C_{NLLS}^{b,i_M}$ and define

$$C_{NLLS}^{b,\max} \coloneqq \max\left(C_{NLLS}^{b,i_1}, \ldots, C_{NLLS}^{b,i_M}\right), \tag{92}$$

and similarly, for M statistics $C_{LL}^{b,i_1},\ldots,C_{LL}^{b,i_M}$, define

$$C_{LL}^{b,\max} \coloneqq \max\left(C_{LL}^{b,i_1},\ldots,C_{LL}^{b,i_M}\right). \tag{93}$$

For each version, the M test statistics have the same block size but use different starting points, w.l.o.g., $i_1 < i_2 \ldots < i_M$. It holds that

$$\lim_{T\to\infty} P\left[C_{NLLS}^{b,\max} \leq c_{\alpha/M}\right] \geq 1-\alpha, \tag{94}$$

$$\lim_{T\to\infty} P\left[C_{LL}^{b,\max} \leq c_{\alpha/M}\right] \geq 1-\alpha, \tag{95}$$

where $c_{\alpha/M}$ is the α/M-level critical value from the distribution of $\int_0^1 W^2(s)ds$. For detailed suggestions how to choose the number of subresidual tests statistics M, starting points $i_1,\ldots,i_M$, and block size b by the minimum volatility rule, see Choi and Saikkonen (2010). For further usage, we relabel the considered test statistics as $T_{\text{CS,L}} \coloneqq C_{LL}$, $T_{\text{CS,maxL}} \coloneqq C_{LL}^{b,\max}$ and $T_{\text{CS,maxN}} \coloneqq C_{NLLS}^{b,\max}$.

Dong and Gao (2018)

Dong and Gao (2018) propose two specification tests for the null hypothesis in (52) based on an orthogonal series. One version of the test covers integrable regression functions and the other version accounts for a function space of nonintegrable regression functions. Both tests allow for regressor endogeneity. The test statistics are based on Hermite polynomials as defined in (71). Furthermore, define

$$\mathcal{H}_j(x) \coloneqq (\sqrt{\pi}2^j j!)^{-1/2} H_j(x)\exp\left(-\frac{1}{2}x^2\right), \tag{96}$$

and thus $\mathcal{H}_j(x) = h_j(x)\exp(-\frac{1}{2}x^2)$, with $h_j(x)$ as defined in (72). The sequence $\{\mathcal{H}_j(x)\}_{t\in\mathbb{Z}}$ is a complete orthonormal basis in $L^2(\mathbb{R}) = \{g(x) : \int_{-\infty}^{\infty} g^2(x)dx < \infty\}$. Any function $f(x) \in L^2(\mathbb{R})$ can be represented using an orthogonal expansion, i. e. , $f(x) = \sum_{j=0}^{\infty}\beta_j\mathcal{H}_j(x)$ with $\beta_j = \int_{-\infty}^{\infty} f(x)\mathcal{H}_i(x)dx$. The truncated series $f_k(x) = \sum_{j=0}^{k-1}\beta_j\mathcal{H}_j(x)$ with truncation parameter $k \in \mathbb{N}$ approximates the function $f(x)$, with $f_k(x) \to f(x)$ in L^2 sense as $k \to \infty$ (see, e. g. , Dong and Gao, 2018, p. 759). First, we consider the case with an integrable function, i. e. , where $f(x) \in L^2(\mathbb{R})$. Denote $Z_k(x) \coloneqq (\mathcal{H}_0(x),\ldots,\mathcal{H}_{k-1}(x))'$ and $\beta \coloneqq (\beta_0,\ldots,\beta_{k-1})'$, leading to $f_k(x) = Z_k(x)'\beta$. Furthermore, denote $Z \coloneqq (Z_k(x_1),\ldots,Z_k(x_T))'$. Then, the OLS estimator $\hat{\beta} \coloneqq (Z'Z)^{-1}Z'y$ of β leads to a nonparametric estimation of $f(x)$, with $\hat{f}(x) = Z_k(x)'\hat{\beta}$. The underlying idea of the test is to compare the estimation $\hat{f}(x)$, based on Hermite polynomials, with the parametric estimation $g(x,\hat{\theta}_T)$ when considering a specific functional form. To overcome theoretical and computational

difficulties when relying on the inverse of $Z'Z$, Dong and Gao (2018) propose to use $\tilde{\beta} \coloneqq Z'Z\hat{\beta}$ instead of $\hat{\beta}$ and correspondingly replace $\hat{f}(x) = Z_k(x)'\hat{\beta}$ by $\tilde{f}(x) = Z_k(x)'\tilde{\beta}$, which leads to $\tilde{f}(x) = \sum_{t=1}^{T} Z_k(x_t)'Z_k(x)y_t$. A similar transformation is performed for $g(x,\theta)$ and we obtain $\tilde{g}(x,\theta) \coloneqq \sum_{t=1}^{T} Z_k(x_t)'Z_k(x)g(x_t,\theta)$. Now, instead of comparing $\hat{f}(x)$ with $g(x,\hat{\theta}_T)$, the test statistic is based on the distance between $\tilde{f}(x)$ and $\tilde{g}(x,\hat{\theta}_T)$ and given by

$$L_T \coloneqq \int_{-\infty}^{\infty} \left(\tilde{f}(x) - \tilde{g}(x,\hat{\theta}_T)\right)^2 dx = \int_{-\infty}^{\infty} \left(\sum_{t=1}^{T} \left(Z_k(x_t)'Z_k(x)\right)\left(y_t - g(x_t,\hat{\theta}_T)\right)\right)^2 dx, \tag{97}$$

which simplifies by $\int_{-\infty}^{\infty} Z_k(x)Z_k'(x)dx = I_k$ to

$$L_T = \sum_{t=1}^{T}\sum_{s=1}^{T} Z_k(x_t)'Z_k(x_s)\hat{u}_t\hat{u}_s, \tag{98}$$

with $\hat{u}_t \coloneqq y_t - g(x_t,\hat{\theta}_T)$ and $\hat{\theta}_T$ nonlinear least squares estimator of θ_0 in $g(x,\theta_0)$. Under the null hypothesis and Assumptions 20 and 21 it holds that

$$\frac{\psi}{T^{1/2}k\sigma_u^2} L_T \to_d L_W(1,0), \tag{99}$$

with $\psi = \omega_{vv}$ (can be seen from Assumption 20). The unknown parameters can be replaced by consistent estimates and it holds that

$$T_{\text{DG1}} \coloneqq \frac{\hat{\psi}}{T^{1/2}k\hat{\sigma}_u^2} L_T \to_d L_W(1,0), \tag{100}$$

with $\sigma_u^2 \coloneqq \frac{1}{T}\sum_{t=0}^{T} \hat{u}_t^2$ (see Dong and Gao, 2018).

The second test of Dong and Gao (2018) covers the case where $f(x) \in L^2(\mathbb{R}, e^{-x^2}) = \{g(x) : \int_{-\infty}^{\infty} g^2(x)e^{-x^2}dx < \infty\}$ and thus includes a wide range of regression functions, such as polynomials, $\theta\log|x|$ and $\theta e^x/(1+e^x)$. The idea of this second version is, starting from a function $f(x) \in L^2(\mathbb{R}, e^{-x^2})$, to define $\breve{f}(x) \coloneqq f(x)e^{-\frac{1}{2}x^2}$, with $\breve{f}(x) \in L^2(\mathbb{R})$. Defining $\breve{y}_t = y_t e^{-\frac{1}{2}x_t^2}$, $\breve{u}_t = u_t e^{-\frac{1}{2}x_t^2}$ leads to the transformed model

$$\breve{y}_t = \breve{f}(x_t) + \breve{u}_t, \tag{101}$$

which is equivalent to (51). Furthermore, we define $\breve{g}(x,\theta_0) \coloneqq g(x,\theta_0)e^{-\frac{1}{2}x^2}$. The hypothesis $H_0' : \breve{f}(x) = \breve{g}(x,\theta_0)$ is equivalent to (52). This leads to the test statistic

$$\breve{L}_T = \sum_{t=1}^{T}\sum_{s=1}^{T} Z_k(x_t)'Z_k(x_s)\hat{\breve{u}}_t\hat{\breve{u}}_s, \tag{102}$$

with $\check{\hat{u}}_t = \hat{u}_t e^{-\frac{1}{2}x_t^2} = (y_t - g(x_t, \hat{\theta}_T))e^{-\frac{1}{2}x_t^2} = \check{y}_t - \check{g}(x_t, \hat{\theta}_T)$ and $\hat{\theta}_T$ a consistent estimator of θ_0. Under H_0, with $g(x, \theta) \in L^2(\mathbb{R}, e^{-x^2})$, and Assumptions 20 and 22, it holds that

$$\frac{\psi}{T^{1/2} k \sigma_u^2} \check{L}_T \to_d \int \mathcal{F}(x) e^{-x^2} dx \cdot L_W(1, 0), \tag{103}$$

with $\mathcal{F}(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbb{1}_{[-2,2]}(x)$. Again, replacing the unknown parameters by consistent estimators leads to a feasible test statistic and in holds that

$$T_{\text{DG2}} \coloneqq \frac{\hat{\psi}}{T^{1/2} k \hat{\sigma}_u^2} \check{L}_T \to_d \mathcal{F}(x) e^{-x^2} dx \cdot L_W(1, 0) \tag{104}$$

(see Dong and Gao, 2018).

Wang and Phillips (2012)

Wang and Phillips (2012) propose a specification test for nonlinear cointegrating regressions, based on a kernel-weighted U-statistic. Denote with $\hat{u}_t = y_t - g(x_t, \hat{\theta}_T)$

$$S_T = \sum_{s,t=1, s \neq t}^{T} \hat{u}_t \hat{u}_s K((x_t - x_s)/h), \tag{105}$$

with $\hat{\theta}_T$ a consistent parametric estimator of $\theta_0 \in \Theta_0$ under the null hypothesis. Furthermore, define

$$V_T^2 = \sum_{s,t=1, s \neq t}^{T} \hat{u}_t \hat{u}_s K^2((x_t - x_s)/h). \tag{106}$$

Under the null hypothesis and with Assumptions 23–27 in place, it holds that

$$\hat{T}_{\text{WP12}} \coloneqq \frac{S_T}{\sqrt{2} V_T} \to_d \mathcal{N}(0, 1). \tag{107}$$

Under the stated assumptions, the proposed test does not allow for an endogeneous regressor or serially correlated errors.

Wang and Phillips (2016)

Wang and Phillips (2016) propose a specification test based on an integral of kernel weighted residuals from nonlinear least squares estimation. In contrast to the test of Wang and Phillips (2012), this test allows for an endogeneous regressor and serially correlated errors. Having the nonlinear least squares estimator $\hat{\theta}_T$ of θ_0 in $g(x, \theta_0)$, we can test the null hypothesis in (52) with a normalized version of the statistic

$$\mathcal{T}_T = \int_{-\infty}^{\infty} \left(\sum_{t=1}^{T} K\left(\frac{x_t - x}{h} \right) \left(y_t - g(x_t, \hat{\theta}_T) \right) \right)^2 \pi(x) dx, \tag{108}$$

with $\pi(x)$ denoting a positive integrable function (i. e. , $\int_{-\infty}^{\infty}|\pi(x)|dx<\infty$). With Assumptions 5–6 and 17–19 in place, it holds under the null hypothesis that

$$\mathcal{T}_{T,0} := \frac{\omega_{vv}}{T^{1/2}h}\mathcal{T}_T \to_d \tau_0 L_W(1,0), \tag{109}$$

for any h satisfying $\sqrt{T}h^2\log T\to 0$ and $T^{1-\delta_0}h^2\to\infty$, where

$$\tau_0 = \mathbb{E}(u_0^2)\int_{-\infty}^{\infty}K^2(s)ds\int_{-\infty}^{\infty}\pi(x)dx, \tag{110}$$

and δ_0 can be as small as required. The error variance $\mathbb{E}(u_0^2)$ can be estimated by

$$\hat{\sigma}_T^2 = \frac{\sum_{t=1}^T(y_t-g(x_t,\hat{\theta}_T))^2K_h(x_t-x)}{\sum_{t=1}^T K_h(x_t-x)} \tag{111}$$

and the combination with a consistent estimator $\hat{\omega}_{vv}^2\to_p\omega_{vv}^2 := \sum_{j=-\infty}^{\infty}\mathbb{E}\left(v_{t-j}v_t\right)$ delivers a feasible test statistic $T_{\text{WP16}} := \hat{\mathcal{T}}_{T,0}$. If we choose $h=T^b$, then it has to hold $-\frac{1-\delta_0}{2}<b<-\frac{1}{4}$. Gao *et al.* (2012, p. 9), who use the same test for nonlinear cointegrating models but without allowing for endogeneity, suggest to choose the Gaussian kernel function as $\pi(x)$, if the partial derivatives of $g(x,\theta_0)$ with respect to θ_0 are of polynomial forms.

Wang, Wu and Zhu (2018)

Wang *et al.* (2018) propose a test on parametric specification in nonlinear cointegrating regressions, based on a marked empirical process. More precisely, they test the null hypothesis

$$H_0 : \mathbb{E}\left(y_t - g(x_t,\theta_0)|x_t\right) = 0, \tag{112}$$

for some $\theta_0\in\Theta_0$ with $\Theta_0\subset\mathbb{R}$ being a compact space. Denote with $\hat{\theta}_T$ the nonlinear least squares estimator of θ_0. Define $d_T^2 := var(\sum_{k=0}^T v_k)\sim\omega_{vv}^2T$. First, we consider the case of integrable regression functions $g(x,\theta)$. Define the marked empirical process $\alpha_T(x)$ as

$$\alpha_T(x) = \frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-g(x_t,\hat{\theta}_T))\mathbb{1}_{(x_t/d_T\le x)}, \tag{113}$$

where $0<d_T\to\infty$ is a sequence of constants. The test statistic for the null hypothesis in (112) is then defined by

$$S_T = \sup_{x\in\mathbb{R}}|\alpha_T|(x) \tag{114}$$

$$= \sup_{x\in\mathbb{R}}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T(y_t-g(x_t,\hat{\theta}_T))\mathbb{1}_{(x_t\le x)}\right|, \tag{115}$$

where the last equality does not depend anymore on d_T, which simplifies the calculation of S_T in practice. Let $\{B_1(t), B_2(t)\}_{t\geq 0}$ be a 2-dimensional Brownian motion with covariance matrix

$$\Omega_0 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{116}$$

with $\rho \in [-1, 1]$ being the asymptotic correlation coefficient between $T^{-1/2}\sum_{t=1}^{T}\varepsilon_t$ and $(T\sigma^2)^{-1/2}\sum_{t=1}^{T}u_t$ and with $\sigma^2 > 0$ being the asymptotic variance of $T^{-1/2}\sum_{t=1}^{T}u_t$. Under the null hypothesis and Assumptions 28–30 it holds that

$$S_T \to_d \sigma \sup_{x\in\mathbb{R}}|\alpha(x)|, \tag{117}$$

where $\alpha(x) = \int_0^1 \mathbb{1}_{(B_1(t)\leq x)}dB_2(t)$. Under H_0, the variance σ^2 can be consistently estimated by $\hat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2$, with $\hat{u}_t = y_t - g(x_t, \hat{\theta}_T)$ (see Wang *et al.*, 2018).

Next, we consider the case of nonintegrable regression functions. Define $\Psi(t) \coloneqq \dot{h}(B_1(t), \theta_0)$. Under the null hypothesis and Assumptions 28–29 and 31–32, it holds that

$$S_T \to_d \sigma \sup_{x\in\mathbb{R}}|\beta(x)|, \tag{118}$$

where

$$\beta(x) = \int_0^1 \mathbb{1}_{(B_1(t)\leq x)}dB_2(t) - \int_0^1 \mathbb{1}_{(B_1(t)\leq x)}\Psi_1'(s)ds\left(\int_0^1 \Psi_1(u)\Psi_1'(u)du\right)^{-1}\int_0^1 \Psi_1(t)dB_2(t). \tag{119}$$

Wang *et al.* (2018) provide critical values for S_T when considering integrable functions and in addition for $g(\theta, x) = \theta_1 + \theta_2 x$, $g(\theta, x) = \theta_1 + \theta_2 x + \theta_3 x^2$, $g(\theta, x) = \theta_1 + \theta_2 \log x$ and $g(\theta, x) = \theta e^x/(1 + e^x)$, in case of certain error structures. For both cases, integrable and nonintegrable regression functions, denote $T_{\text{WWZ}} \coloneqq S_T$.

Wang and Zhu (2020)

Wang and Zhu (2020) propose a portmanteau test for specification testing in nonlinear cointegrating regression models with serially correlated errors and an endogenous regressor. Wang and Zhu (2020) assume that the errors have the form $u_t = \rho u_{t-1} + \nu_t$. Define $\hat{u}_t = y_t - g(x_t, \hat{\theta}_T)$ with $\hat{\theta}_T$ being a consistent estimator of θ_0. Denote with $\hat{\nu}_t = \hat{u}_t - \widehat{\rho}\hat{u}_{t-1}$ the residual of ν_t. Let

$$\widehat{\rho} = \frac{\sum_{s=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=2}^{T}\hat{u}_{t-1}^2} \tag{120}$$

be the least squares estimator of ρ based on the autoregression $\hat{u}_t = \rho\hat{u}_{t-1} + \nu_t$. If $\rho = 0$ set $\hat{\nu}_t = \hat{u}_t$ for all t. The portmanteau test statistic is then defined as

$$\widehat{U}_T(M) \coloneqq T(T+2)\sum_{k=1}^{M}\frac{\widehat{a}_k^2}{T-k} \tag{121}$$

for some integer $M \geq 1$, where

$$\widehat{a}_k = \frac{\sum_{t=k+1}^{T} \hat{\nu}_t \hat{\nu}_{t-k}}{\sum_{t=1}^{T} \hat{\nu}_t^2} \tag{122}$$

is the sample autocorrelation of $\hat{\nu}_t$ at lag k. The aim of the portmanteau test statistic $\widehat{U}_T(M)$ is to detect the autocorrelation of the residual of ν_t at the first M lags. Under the null hypothesis and Assumptions 33 and 34 it holds for integrable regression functions that the limiting distribution of $\widehat{U}_T(M)$ can be approximated by χ^2_{M-1} for large M (see Wang and Zhu, 2020). In case of nonintegrable regression functions, it holds under the null hypothesis and Assumptions 33 and 35, with $\|D_n(\hat{\theta} - \theta_0\| = O_P(\log^\delta T)$ for some $\delta > 0$ with $D_T = \text{diag}(\sqrt{T}v_1(d_T), \ldots, \sqrt{T}v_m(d_T))$, where $d_T^2 = var(x_T)$, that the limiting distribution of $\widehat{U}_T(M)$ can be approximated by χ^2_{M-1} for large M (see Wang and Zhu, 2020). For both integrable and nonintegrable regression functions denote $\hat{T}_{\text{WZ}} \coloneqq \widehat{U}_T(M)$.

Properties Specification Tests

Similar to the described estimators, the described tests differ in their assumptions under which their test statistics converge under the null hypothesis towards the stated limiting distributions. Again, our main focus is on whether the assumptions allow for error serial correlation and regressor endogeneity, as well as which choices need to be made. Furthermore, we are interested whether long-run variance estimation is required to compute the test statistic. Despite the tests of Wang and Phillips (2012) and Wang *et al.* (2018), the other tests allow for error serial correlation and regressor endogeneity. On the other hand, those two tests and additionally the Wang and Zhu (2020) test do not require long-run variance estimation. Furthermore, the Wang *et al.* (2018) does not require any user choices at all. Table 26 shows which specification test allows in its specific assumptions for error serial correlation and regressor endogeneity. Furthermore, Table 26 lists necessary user choices and whether long-run variance estimation is required.

2.5 Large Scale Simulation Study

In this section we compare the performance of the considered estimators and specification tests for a variety of settings. As performance criteria serve average root mean squared error (ARMSE) for the estimators and null rejection probabilities and *size-corrected* power for the specification tests. We are particularly interested in the effects on performance related to variation in the

[48] Only the subsample-based test requires the choice of a block size.

	Correlation	Endogeneity	User Choices	LR Var
T_{CS}	✓	✓	leads and lags, block size[48]	✓
T_{DG}	✓	✓	approximation order	✓
T_{WP12}	✗	✗	kernel, bandwidth	✗
T_{WP16}	✓	✓	kernel, bandwidth, weighting	✓
T_{WWZ}	✗	✗	✗	✗
T_{WZ}	✓	✓	AR order of errors	✗

Table 26: The checkmarks indicate whether the specific assumptions of the tests allow for error serial correlation and endogeneity, and if long-run variance estimation is required.

following dimensions: Regression function (integrable and nonintegrable, as well as different types), regressor endogeneity and error serial correlation. The degree of endogeneity and error serial correlation (or their presences at all) is naturally unknown in many applications. Therefore, it is of particular interest to which extent endogeneity and error serial correlation influence estimation and test performances. Furthermore we vary estimator and test specific choices, such as kernel, bandwidth and truncation. We generate data according to:

$$\begin{aligned} y_t &= f(x_t) + u_t, \quad t = 1, \dots, T, \\ x_t &= x_{t-1} + v_t, \quad x_0 = 0, \end{aligned} \tag{123}$$

with:

$$\begin{aligned} u_t &= \rho_0 u_{t-1} + e_{1,t} + \rho_1 e_{1,t-1} + \rho_2 e_{2,t}, \\ v_t &= e_{2,t} + 0.5 e_{2,t-1}. \end{aligned} \tag{124}$$

and $[e_{1,t}, e_{2,t}]' \sim \mathcal{N}(0, I_2)$ (and iid) for $t \in \mathbb{N}$. For the simulations, we obtain u_1 and v_1 by using $e_{1,1}, e_{1,0}, e_{2,1}, e_{2,0}, u_0$, whereby $e_{1,0}, e_{2,0}$ and u_0 are drawn from the joint distribution

$$\begin{bmatrix} u_0 \\ e_{1,0} \\ e_{2,0} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{(1+2\rho_0\rho_1+\rho_1^2)\sigma_{e_1}^2 + \rho_2^2\sigma_{e_2}^2}{1-\rho_0^2} & \sigma_{e_1}^2 & \rho_2\sigma_{e_2}^2 \\ \sigma_{e_1}^2 & \sigma_{e_1}^2 & 0 \\ \rho_2\sigma_{e_2}^2 & 0 & \sigma_{e_2}^2 \end{bmatrix} \right). \tag{125}$$

with $\sigma_{e_1} = \sigma_{e_2} = 1$ in our simulations. We do this in order to obtain the desired covariances between u_t and $[e_{1,t}, e_{2,t}]'$ even for very small values of t and to avoid the need for a settling period. Given the DGP above, the long-run variance $\omega_{u \cdot v}^2$ in the Choi-Saikkonen test is given by $\omega_{u \cdot v}^2 = \sigma_{e_1}^2 (1+\rho_1)^2/(1-\rho_0)$.

Tables 27 shows the general parameter choices and Table 28 shows estimator and test specific choices. We perform 5,000 replications and all tests are performed at a nominal level of 5%. For

some regression functions it is necessary, that there are observations in a certain area. For example, for the function f_2, the function value is recognizable above zero only in a domain around zero. Thus, if we have most observations way larger than zero, we only observe function values close to zero and estimating such a function is not what we intend to estimate. Therefore, in each replication, the draw of random numbers for $e_{1,t}$ and $e_{2,t}$ will be repeated until $x_{0.25} < 0 < x_{0.75}$ applies for the resulting observations, with x_p being the p-quantile of $\{x_1, \ldots, x_T\}$. The RMSE is calculated using 100 equidistant grid point between $x_{0.05}$ and $x_{0.95}$. This is to ensure that the grid on which we calculate the RMSE is actually covered by observations, which is not necessarily given with a fixed grid. Additionally, when calculating the function value of kernel based estimators at a specific grid point, the bandwidth for this point is increased, if necessary, in such a way that at least two observations that lead to nonzero kernel values are taken into account. This is to ensure, that each grid point is locally covered by observations. In the finite sample simulations we use for long-run covariance estimation the Bartlett kernel with bandwidth chosen according to Newey and West (1994). To eliminate the detrimental impact of long-run variance estimation, we also run the simulations using the true long-run covariances instead of their estimates estimate. For the optimal bandwidth procedure of Wang and Phillips (2023) we use the Gaussian kernel and for the variance estimation there the bandwidth $h = T^{-1/6}$. We use the BIC to determine the AR order of the error term for calculating the test statistic of Wang and Zhu (2020). For calculating the test statistic of the Choi and Saikkonen (2010) tests, lead and lag length choices are performed using the BIC-type criterion of Choi and Kurozumi (2012). Let us now turn the regression functions from Table 27. We consider two integrable and five nonintegrable functions. The first one is the sum of sinus functions, which we truncate at $j = 4$, similar to Wang and Phillips (2016) in their simulations. The second function is an exponential function. Coming to the nonintegrable functions, we have a quadratic function being a representative of cointegrating polynomial functions in our case with an integrated regressor, furthermore a smooth transition function and an exponential function. These three functions are covered by the assumptions of most of the nonparametric estimators. The next two types of function are not covered by the assumptions of all estimators, namely the power function with one parameter as exponent and the jump function, which is not a real jump in our choice of parameters, but still nondifferentiable. The parameters chosen in our simulations are shown in the third column of Table 27 and the function abbreviations used in our result tables are listed in the fourth column. Figure 40 shows us the shape of the functions given the parameter choices we made. The space of AR and MA coefficients is chosen in such a way, that we do not

obtain $\rho_0 + \rho_1 = 0$. Otherwise, if additionally $\rho_2 = 0$, this would lead to $u_t = e_{1,t}$, a pure iid process, even if $\rho_0 \neq 0$ and $\rho_1 \neq 0$. As an illustration of the scale of the simulation study, we ran with seven regression functions, four MA, AR and endogeneity parameters each and five sample sizes, in total 2,240 simulation setups for the ARMSE simulations with 5,000 iterations each. For size-corrected power, it was even three times as much, due to the three alternative DGPs.

Regression functions			
$f_1(x_t) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \sin(j\pi x_t)}{j^2}$	integrable	*(truncated at $j = 4$)*	(sinc)
$f_2(x_t) = \theta_0 \exp(-\theta_1 x_t^2)$	integrable	$\theta_0 = 1, \theta_1 = 1$	(intexp)
$f_3(x_t) = \theta_0 + \theta_1 x_t + \theta_2 x_t^2$	nonintegrable	$\theta_0 = 1, \theta_1 = 2, \theta_2 = -0.5$	(cpr)
$f_4(x_t) = \theta_0 + \theta_1 x_t + \theta_2 x_t \frac{1}{1+\exp(-(x_t-\theta_3))}$	nonintegrable	$\theta_0 = 1, \theta_1 = 1, \theta_2 = -1.5, \theta_3 = 1$	(stc)
$f_5(x_t) = \frac{\theta_0+\theta_1 \exp(x_t)}{1+\exp(x_t)}$	nonintegrable	$\theta_0 = -1, \theta_1 = 1$	(exp)
$f_6(x_t) = \theta_0 + \theta_1 \lvert x_t \rvert^{\theta_2}$	nonintegrable	$\theta_0 = 1, \theta_1 = -1, \theta_2 = 2$	(pow)
$f_7(x_t) = \theta_0 + \theta_1 x_t 1(x_t \geq 0)$	nonintegrable	$\theta_0 = 1, \theta_1 = 1$	(jump)
Error serial correlation (through AR and MA components)			
AR coefficient: $\rho_0 \in \{0, 0.3, 0.6, 0.9\}$ MA coefficient: $\rho_1 \in \{-0.8, -0.4, 0, 0.4, 0.8\}$.			
Endogeneity			
$\rho_2 \in \{0, 0.3, 0.6, 0.9\}$			
Sample sizes			
$T \in \{50, 100, 200, 500, 1000\}$			

Table 27: General specifications and parameter choices.

Table 28 shows the sets of choices we consider in the estimators and specification tests in the simulation study. As kernel functions, we use three different kernels of which the Gaussian kernel has no compact support. For the related bandwidths, we consider values for b in such a fashion that the requirements on h given in the definition of the estimators are met, i. e. , $h = T^b$ with $b \in \{-1/3, -1/4, -1/5, -1/6\}$. Furthermore, we consider for the choice of h optimal pointwise and weighted bandwidths, determined by the procedures of Wang and Phillips (2023). This is done in two ways, once using A_T and once with $A_{T,h}$, leading in total to the four optimal bandwidth rules $h_{opt}(x)$ (pointwise, with A_T), h_{opt} (weighted, with A_T), $h_{opt,h}(x)$ (pointwise, with $A_{T,h}$) and $h_{opt,h}$ (weighted, with $A_{T,h}$). For the Dong *et al.* (2021) estimator, a weighting function $\phi(x)$ is required. We follow their suggestion and use as one possibility $\phi(x) = \exp(-x^2)$. Since this function leads to very poor results in our simulations for large values of x, we also consider variations in which x is scaled in different ways. Additionally, we also consider no weighting, i. e. , $\phi(x) = 1$. The Wang

and Phillips (2016) specification test requires a weighting function too. In addition to the proposed Gaussian kernel, we also consider a centered version. For the Dong *et al.* (2021) estimator and the Dong and Gao (2018) specification test, a truncation parameter k is necessary, which is defined by Dong and Gao (2018) as $k = \lfloor c \cdot T^{\kappa} \rfloor$. Close to their choices and following the restriction $0 < \kappa < \frac{1}{2}$, we consider $\kappa \in \{1/5, 1/4, 1/3\}$ and $c = 2$. For the generalized cross validation approach to select k, we use $\kappa_{max} = 1/3$ and $c_{max} = 2.5$. For the Wang and Zhu (2020) specification test, the considered maximum amount of lags needs to be specified. Here we consider some fixed values as in Wang and Zhu (2020) and additionally two sample size dependent rules. Two of the described test statistics from Choi and Saikkonen (2010) require a block size, which we choose by the minimum volatility rule as described in their paper. For this, we use a minimum block size of $b_{min} = \lfloor T^{0.7} \rfloor$ and a maximum block size of $b_{max} = \lfloor T^{0.9} \rfloor$, as in Choi and Saikkonen (2010). For the $\hat{f}_{\text{LW2}}$ estimator of Linton and Wang (2016) the AR order needs to be specified. We use the Bayesian information criterion (BIC) with a maximum number of lags of $p_{max} = \lfloor 4(T/100)^{1/4} \rfloor$.

Kernel functions
Bartlett, Epanechnikov, Gaussian *(no compact support)*
Bandwidths
$h = T^b$ and $b \in \{-1/3, -1/4, -1/5, -1/6\}$ $h_{opt}(x)$, h_{opt}, $h_{opt,h}(x)$, $h_{opt,h}$, optimal bandwidth selection of Wang and Phillips (2023),
Weighting functions
$\phi(x) \in \{\exp(-x^2), \exp(-(x/10)^2), \exp(-(x/x_a)^2), \exp(-(x/x_{max})^2), 1\}$ with $x_a \coloneqq \frac{1}{T}\sum_{t=1}^{T} \lvert x_t \rvert$ and $x_{max} \coloneqq max_{t=1,\ldots,T} \lvert x_t \rvert$ $\pi(x)$ (positive integrable function): $\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$ and $\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$
Truncation parameters
$k = \lfloor c \cdot T^{\kappa} \rfloor$ with $c = 2$ and $\kappa \in \{1/5, 1/4, 1/3\}$ $k_{max} = \lfloor c_{max} \cdot T^{\kappa_{max}} \rfloor$ for generalized cross validation, with $c_{max} = 2.5$ and $\kappa_{max} = 1/3$
Number of lags for sample autocorrelation
$M = 4, 6, 12, 18$, as in Wang and Zhu (2020), additionally $\lfloor 4(T/100)^{1/4} \rfloor$ and $\lfloor 12(T/100)^{1/4} \rfloor$
Block size
Using the minimum volatility rule with $b_{min} = \lfloor T^{0.7} \rfloor$ and $b_{max} = \lfloor T^{0.9} \rfloor$
Number of lags for $\hat{f}_{\text{LW2}}$
Lag length selection using BIC, with a maximum number of lags $p_{max} = \lfloor 4(T/100)^{1/4} \rfloor$

Table 28: Estimator and test specific choices.

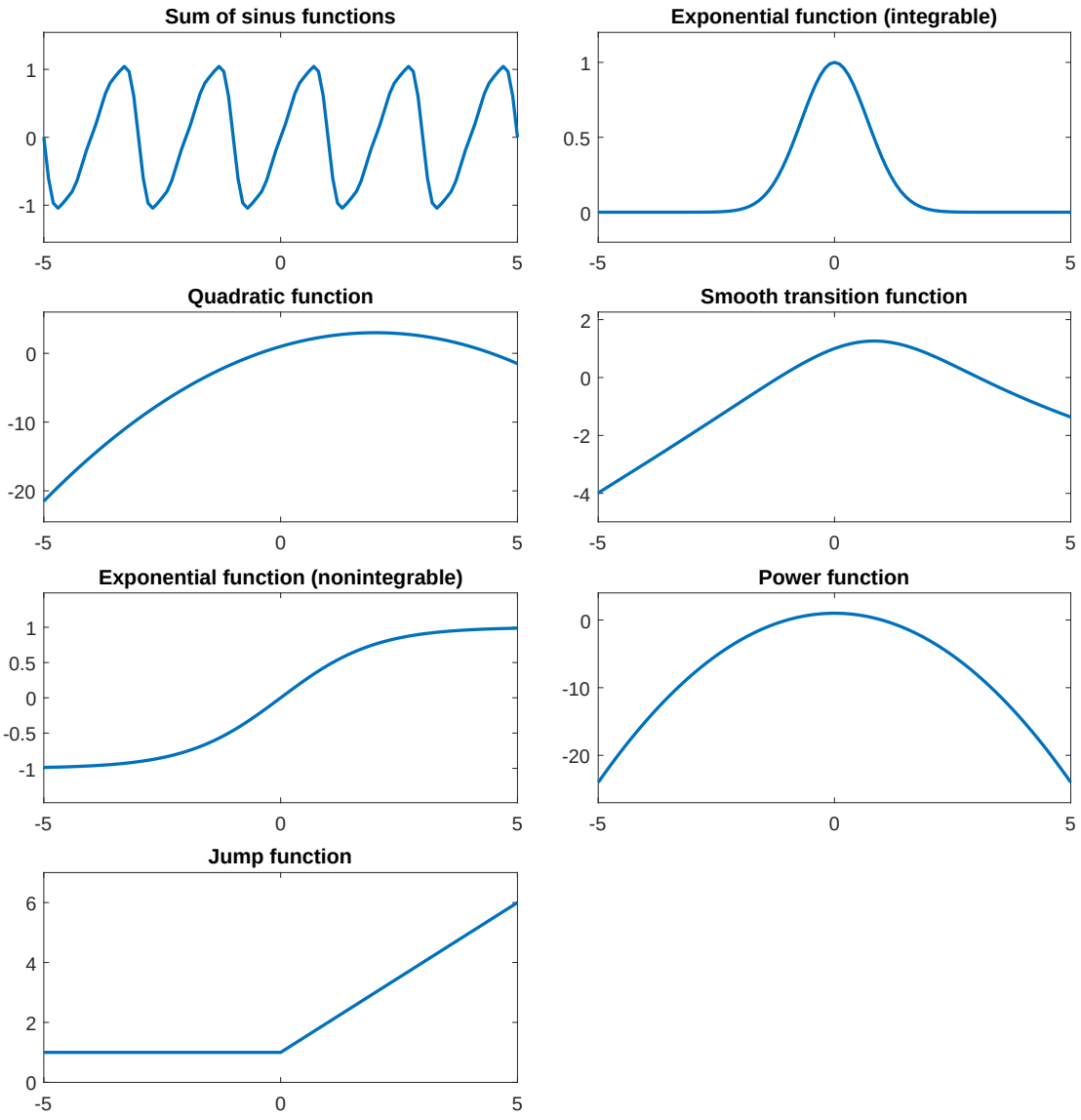

Figure 40: Overview of the used regression functions.

2.5.1 Estimators – Average Root Mean Squared Error

We now turn to the results of the simulation study. For the estimators, we measure their performance for a single simulation run in terms of root mean squared error (RMSE). The RMSE is calculated based on the difference between estimation and function value at 100 equidistant grid points. This is done for each replication. Having 5,000 replications, the resulting 5,000 RMSE values are then averaged, leading to a single average root mean squared error (ARMSE).

Before comparing the different estimators, we look into the results of each individual estimator with their respective parameter choices. The aim is to first identify for each estimator a set of parameters that works best across the different regression functions, since in applications the regression function will be likely unknown. Afterwards, we combine the different estimators each with one set of parameters. In practice, we often have to deal with error serial correlation and regressor endogeneity in cointegrating regressions, or at least, we cannot be sure that they are not present. Thus, we directly compare the ARMSE of estimators in a scenario with non-zero error serial correlation, also including a moving average error component, and regressor endogeneity, i. e. , with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

We start with the the Nadaraya-Watson estimator and display the results for $T = 200$ and $T = 500$ in Table 29.[49] The estimator works best in many scenarios with a bandwidth rule considering the estimated variance of the regressor x_t. Sometimes, however, the performance is very poor compared to other considered rules, which is why we refrain from using this bandwidth rule. In many cases either the weighted optimal bandwidth rule with A_{nh} or the bandwidth rule with the estimated variance of v_t times $T^{-1/6}$ perform best, with the other one being close by. Since the weighted optimal bandwidth rule with A_{nh} often performs better compared to the bandwidth rule with the estimated variance of v_t in cases where the bandwidth rule with the estimated variance of x_t performs best, this is our choice for bandwidth selection. With respect to the kernel choice, there is no clear winner. When choosing the weighted optimal bandwidth rule with A_{nh}, the Gaussian kernel leads in most scenarios to the smallest ARMSE, with the other two kernels being usually only just behind. In cases where the Gaussian kernel performs worse than the Bartlett and Epanechnikov kernel, the differences are substantially larger and the Bartlett kernel performs best. Considering the trade-off between being only slightly behind in many cases, but well ahead in some cases, compared to the exact opposite cases, we select the Bartlett kernel as our favorite,

[49] Results for $T = 50, 100, 1000$ have not been included here for the sake of space and are available on request.

in combination with the weighted optimal bandwidth rule with A_{nh}, i. e. , $h_{opt,h}$.

For the local linear estimator, with results for $T = 200$ and $T = 500$ in Table 30, most of the times a bandwidth rule in combination with the Gaussian kernel leads to the smallest ARMSE. For those cases, in which another kernel leads to the smallest ARMSE, a kernel-bandwidth rule combination with the Gaussian kernel is very close by. Additionally, we observe situations in which, independent from the bandwidth rule, the Bartlett kernel and Epanechnikov kernel lead to much higher ARMSE values. This behavior renders the Gaussian kernel as our favorite kernel choice for the local linear estimator. Focusing on the bandwidth rule with restricting to the Gaussian kernel gives us the following picture: Often, a bandwidth rule considering the estimated variance of the regressor x_t leads to the smallest ARMSE but sometimes, these bandwidth rule choices lead to very high ARMSE values. In those cases, the bandwidth rules with and without including the estimated variance of v_t, i. e. , the first differences of x_t, and also the optimal bandwidth rules lead to very good results, which are also quite good when the rules with the variance of x_t deliver the best results. Since the results are often very close and with the variance of x_t rules sometimes leading to very bad results, our favorite choice is the weighted optimal bandwidth rule with A_{nh}, i. e. , $h_{opt,h}$, as a kind of "uniformly second best" bandwidth choice.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$	$h_{opt}(x)$	$h_{opt,h}(x)$
$T = 200$																	
sinc	bartlett	1.339	1.239	1.172	1.129	1.294	1.185	1.120	1.081	0.837	0.810	0.798	0.792	1.021	1.018	1.043	1.025
	epa	1.311	1.212	1.148	1.107	1.266	1.160	1.099	1.065	0.834	0.806	0.794	0.789	1.025	1.027	1.041	1.030
	gaussian	0.986	0.956	0.951	0.958	0.971	0.951	0.961	0.977	0.794	0.784	0.781	0.780	0.952	0.968	0.956	0.961
intexp	bartlett	1.301	1.197	1.126	1.077	1.254	1.139	1.066	1.018	0.424	0.379	0.358	0.347	0.859	0.832	0.906	0.852
	epa	1.270	1.167	1.097	1.049	1.223	1.110	1.039	0.991	0.416	0.373	0.353	0.344	0.856	0.826	0.892	0.834
	gaussian	0.891	0.840	0.805	0.780	0.868	0.812	0.775	0.749	0.351	0.334	0.330	0.328	0.782	0.743	0.786	0.740
cpr	bartlett	1.897	1.812	1.756	1.720	1.859	1.767	1.714	1.683	17.520	27.613	33.342	36.386	1.659	1.650	1.764	1.724
	epa	1.904	1.824	1.773	1.743	1.868	1.784	1.739	1.716	20.622	31.931	37.716	40.531	1.707	1.706	1.806	1.799
	gaussian	2.138	2.117	2.118	2.134	2.127	2.119	2.145	2.189	35.635	41.451	43.455	44.361	2.118	2.132	2.189	2.181
stc	bartlett	1.305	1.201	1.130	1.081	1.258	1.144	1.071	1.022	1.482	2.265	2.745	3.013	0.880	0.850	0.951	0.881
	epa	1.274	1.172	1.101	1.054	1.228	1.115	1.044	0.997	1.704	2.604	3.105	3.363	0.876	0.844	0.939	0.865
	gaussian	0.903	0.853	0.818	0.793	0.881	0.825	0.788	0.763	2.933	3.455	3.639	3.722	0.804	0.765	0.818	0.778
exp	bartlett	1.300	1.197	1.125	1.076	1.254	1.139	1.066	1.017	0.463	0.500	0.552	0.595	0.860	0.832	0.906	0.850
	epa	1.270	1.167	1.096	1.048	1.223	1.110	1.038	0.991	0.477	0.541	0.610	0.662	0.856	0.825	0.892	0.831
	gaussian	0.888	0.838	0.802	0.776	0.866	0.809	0.771	0.744	0.579	0.697	0.756	0.787	0.780	0.739	0.782	0.735
pow	bartlett	2.872	2.805	2.764	2.743	2.842	2.773	2.741	2.730	34.615	54.341	65.331	71.095	2.734	2.755	2.877	2.901
	epa	2.928	2.869	2.840	2.830	2.901	2.847	2.832	2.836	40.765	62.799	73.827	79.096	2.842	2.882	2.976	3.077
	gaussian	3.778	3.791	3.836	3.900	3.781	3.828	3.928	4.055	69.693	80.598	84.255	85.881	3.788	3.848	3.909	3.933
jump	bartlett	1.304	1.200	1.129	1.080	1.257	1.142	1.069	1.021	1.206	1.921	2.418	2.726	0.894	0.860	0.942	0.873
	epa	1.273	1.170	1.100	1.052	1.226	1.113	1.042	0.994	1.393	2.236	2.779	3.096	0.890	0.853	0.929	0.856
	gaussian	0.897	0.847	0.812	0.786	0.875	0.819	0.781	0.754	2.635	3.276	3.535	3.661	0.807	0.765	0.806	0.762
$T = 500$																	
sinc	bartlett	1.261	1.127	1.041	0.988	1.207	1.066	0.984	0.937	0.775	0.762	0.758	0.756	0.874	0.871	0.901	0.886
	epa	1.230	1.099	1.016	0.966	1.177	1.040	0.963	0.920	0.773	0.761	0.757	0.755	0.877	0.878	0.898	0.892
	gaussian	0.891	0.840	0.824	0.829	0.869	0.826	0.830	0.852	0.758	0.754	0.754	0.753	0.825	0.842	0.835	0.845
intexp	bartlett	1.237	1.100	1.011	0.952	1.182	1.037	0.948	0.892	0.285	0.257	0.247	0.243	0.741	0.711	0.782	0.725
	epa	1.205	1.070	0.982	0.925	1.151	1.008	0.921	0.867	0.281	0.255	0.246	0.242	0.738	0.706	0.770	0.709
	gaussian	0.830	0.761	0.716	0.685	0.803	0.729	0.683	0.653	0.248	0.240	0.239	0.238	0.686	0.643	0.691	0.640
cpr	bartlett	2.062	1.958	1.897	1.863	2.020	1.915	1.861	1.838	51.761	81.461	93.965	99.491	1.836	1.838	1.929	1.921
	epa	2.077	1.981	1.928	1.903	2.037	1.943	1.903	1.892	61.008	92.483	$1.04 \cdot 10^2$	$1.08 \cdot 10^2$	1.896	1.910	1.975	2.020
	gaussian	2.464	2.447	2.463	2.501	2.454	2.456	2.507	2.588	94.673	$1.07 \cdot 10^2$	$1.11 \cdot 10^2$	$1.12 \cdot 10^2$	2.448	2.473	2.496	2.488
stc	bartlett	1.240	1.103	1.013	0.955	1.185	1.040	0.951	0.895	2.634	4.180	4.892	5.217	0.767	0.732	0.843	0.761
	epa	1.208	1.073	0.985	0.928	1.154	1.011	0.925	0.870	3.076	4.751	5.423	5.691	0.764	0.727	0.833	0.747
	gaussian	0.839	0.771	0.725	0.695	0.812	0.739	0.693	0.663	4.908	5.640	5.855	5.942	0.710	0.666	0.730	0.683
exp	bartlett	1.237	1.100	1.010	0.952	1.182	1.037	0.948	0.892	0.423	0.548	0.640	0.697	0.743	0.713	0.783	0.724
	epa	1.205	1.070	0.982	0.925	1.151	1.007	0.921	0.867	0.460	0.613	0.715	0.772	0.740	0.707	0.771	0.708
	gaussian	0.829	0.761	0.715	0.684	0.802	0.729	0.682	0.651	0.640	0.774	0.826	0.849	0.687	0.643	0.690	0.638
pow	bartlett	3.339	3.262	3.226	3.214	3.306	3.236	3.215	3.224	$1.03 \cdot 10^2$	$1.62 \cdot 10^2$	$1.86 \cdot 10^2$	$1.97 \cdot 10^2$	3.220	3.258	3.328	3.439
	epa	3.407	3.346	3.328	3.335	3.380	3.332	3.338	3.371	$1.21 \cdot 10^2$	$1.83 \cdot 10^2$	$2.06 \cdot 10^2$	$2.14 \cdot 10^2$	3.342	3.405	3.427	3.659
	gaussian	4.531	4.565	4.648	4.760	4.537	4.619	4.775	4.972	$1.88 \cdot 10^2$	$2.12 \cdot 10^2$	$2.19 \cdot 10^2$	$2.21 \cdot 10^2$	4.543	4.618	4.619	4.648
jump	bartlett	1.239	1.102	1.013	0.954	1.184	1.039	0.950	0.894	2.196	3.765	4.606	5.035	0.782	0.743	0.829	0.751
	epa	1.207	1.072	0.984	0.927	1.153	1.010	0.924	0.869	2.588	4.338	5.185	5.580	0.779	0.738	0.818	0.735
	gaussian	0.835	0.766	0.720	0.690	0.807	0.734	0.688	0.657	4.614	5.575	5.898	6.038	0.717	0.669	0.718	0.667

Table 29: ARMSE for the Nadaraya-Watson estimator ($\hat{f}_{\text{NW}}$) with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. The columns indicate the bandwidth choices for h.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$	$h_{opt}(x)$	$h_{opt,h}(x)$
$T = 200$																	
sinc	bartlett	33.525	23.679	15.591	9.789	27.694	16.487	9.341	7.387	0.861	0.845	0.839	0.836	3.216	2.456	1.622	1.249
	epa	33.520	23.671	15.580	9.777	27.687	16.477	9.330	7.376	0.860	0.844	0.838	0.835	3.737	2.731	1.861	1.271
	gaussian	1.020	0.976	0.963	0.966	0.998	0.965	0.969	0.983	0.836	0.831	0.830	0.829	0.964	0.975	0.969	0.971
intexp	bartlett	33.517	23.668	15.575	9.769	27.684	16.472	9.319	7.356	0.463	0.438	0.429	0.424	2.437	2.025	1.376	1.135
	epa	33.512	23.660	15.564	9.756	27.678	16.462	9.305	7.342	0.458	0.436	0.428	0.424	2.873	2.185	1.602	1.145
	gaussian	0.940	0.877	0.836	0.808	0.911	0.844	0.802	0.774	0.424	0.418	0.417	0.417	0.810	0.768	0.815	0.768
cpr	bartlett	33.517	23.669	15.576	9.770	27.685	16.473	9.320	7.358	11.722	16.375	18.645	19.949	5.639	3.890	13.713	3.025
	epa	33.512	23.660	15.565	9.757	27.678	16.463	9.307	7.343	13.295	18.304	20.674	22.002	6.225	4.299	14.561	3.373
	gaussian	1.070	1.011	0.972	0.947	1.043	0.980	0.942	0.920	19.790	22.751	24.051	24.720	1.007	0.958	1.075	1.026
stc	bartlett	33.517	23.668	15.575	9.769	27.684	16.473	9.319	7.356	1.023	1.344	1.535	1.657	2.828	2.229	8.939	1.653
	epa	33.512	23.660	15.564	9.756	27.678	16.462	9.305	7.342	1.127	1.489	1.697	1.827	3.147	2.617	9.493	1.833
	gaussian	0.940	0.877	0.836	0.808	0.911	0.844	0.802	0.773	1.645	1.928	2.060	2.129	0.819	0.775	0.830	0.785
exp	bartlett	33.517	23.668	15.575	9.769	27.684	16.473	9.319	7.356	0.467	0.465	0.472	0.478	2.499	2.081	1.640	1.161
	epa	33.512	23.660	15.564	9.756	27.678	16.462	9.305	7.342	0.469	0.475	0.486	0.493	2.931	2.270	2.126	1.189
	gaussian	0.939	0.876	0.834	0.806	0.910	0.842	0.800	0.771	0.478	0.496	0.504	0.508	0.810	0.766	0.812	0.764
pow	bartlett	33.519	23.671	15.579	9.773	27.687	16.476	9.324	7.362	23.390	32.727	37.275	39.886	6.943	4.535	15.849	3.784
	epa	33.514	23.662	15.568	9.760	27.680	16.465	9.310	7.348	26.546	36.589	41.336	43.994	7.841	5.303	16.374	3.878
	gaussian	1.314	1.259	1.225	1.206	1.289	1.232	1.204	1.196	39.567	45.493	48.094	49.432	1.270	1.226	1.420	1.377
jump	bartlett	33.517	23.668	15.576	9.769	27.684	16.473	9.319	7.357	0.748	0.924	1.034	1.106	3.106	2.165	3.978	1.851
	epa	33.512	23.660	15.564	9.756	27.678	16.462	9.306	7.342	0.804	1.008	1.131	1.209	3.445	2.526	4.261	2.003
	gaussian	0.940	0.877	0.836	0.808	0.911	0.844	0.802	0.773	1.099	1.268	1.348	1.390	0.831	0.785	0.829	0.782
$T = 500$																	
sinc	bartlett	27.236	14.093	10.049	8.010	22.412	11.061	7.823	5.624	0.788	0.780	0.778	0.777	3.715	3.159	1.664	1.128
	epa	27.231	14.083	10.037	7.997	22.405	11.049	7.810	5.613	0.787	0.780	0.777	0.776	3.979	3.266	1.764	1.286
	gaussian	0.924	0.857	0.833	0.834	0.894	0.838	0.835	0.854	0.778	0.776	0.776	0.776	0.833	0.844	0.843	0.850
intexp	bartlett	27.232	14.087	10.040	7.996	22.408	11.053	7.809	5.604	0.314	0.299	0.294	0.292	2.999	2.423	1.527	0.971
	epa	27.227	14.077	10.027	7.982	22.400	11.041	7.795	5.589	0.311	0.298	0.294	0.293	3.128	2.838	1.649	1.122
	gaussian	0.874	0.792	0.740	0.706	0.840	0.755	0.704	0.672	0.294	0.292	0.292	0.292	0.707	0.661	0.713	0.659
cpr	bartlett	27.233	14.087	10.040	7.997	22.408	11.053	7.810	5.605	33.632	46.021	51.690	54.839	5.684	4.217	10.692	3.960
	epa	27.227	14.077	10.027	7.983	22.401	11.041	7.796	5.589	38.023	51.172	56.913	59.928	6.468	4.486	11.272	4.134
	gaussian	0.960	0.880	0.831	0.800	0.927	0.845	0.798	0.771	52.476	60.030	62.902	64.217	0.887	0.826	0.957	0.894
stc	bartlett	27.232	14.087	10.040	7.996	22.408	11.053	7.809	5.604	1.602	2.249	2.594	2.797	3.245	2.718	5.479	2.638
	epa	27.227	14.077	10.027	7.982	22.400	11.041	7.795	5.589	1.808	2.510	2.871	3.076	3.494	3.041	5.923	2.757
	gaussian	0.874	0.792	0.740	0.706	0.840	0.755	0.704	0.671	2.640	3.112	3.298	3.384	0.723	0.674	0.742	0.690
exp	bartlett	27.232	14.087	10.040	7.996	22.408	11.053	7.809	5.604	0.379	0.415	0.438	0.451	3.046	2.431	1.384	0.975
	epa	27.227	14.077	10.027	7.982	22.400	11.041	7.795	5.589	0.395	0.437	0.461	0.474	3.245	2.859	1.933	1.053
	gaussian	0.874	0.791	0.739	0.705	0.840	0.754	0.703	0.671	0.443	0.473	0.483	0.487	0.709	0.662	0.712	0.658
pow	bartlett	27.233	14.088	10.041	7.999	22.409	11.055	7.812	5.607	67.255	92.037	$1.03 \cdot 10^2$	$1.10 \cdot 10^2$	7.059	4.908	12.723	4.426
	epa	27.228	14.078	10.029	7.985	22.402	11.043	7.798	5.592	76.037	$1.02 \cdot 10^2$	$1.14 \cdot 10^2$	$1.20 \cdot 10^2$	7.650	5.185	13.040	4.498
	gaussian	1.124	1.048	1.004	0.979	1.092	1.016	0.978	0.962	$1.05 \cdot 10^2$	$1.20 \cdot 10^2$	$1.26 \cdot 10^2$	$1.28 \cdot 10^2$	1.081	1.020	1.197	1.134
jump	bartlett	27.232	14.087	10.040	7.996	22.408	11.053	7.809	5.604	1.063	1.470	1.691	1.822	3.556	3.044	4.873	2.532
	epa	27.227	14.077	10.027	7.982	22.400	11.041	7.795	5.589	1.192	1.638	1.871	2.003	3.972	3.205	5.143	2.600
	gaussian	0.874	0.792	0.740	0.706	0.840	0.755	0.704	0.671	1.721	2.027	2.147	2.203	0.736	0.684	0.737	0.683

Table 30: ARMSE for the local linear estimator ($\hat{f}_{\text{LocLin}}$) with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. The columns indicate the bandwidth choices for h.

The observations with respect to ARMSE are very similar for the estimators $\hat{f}_{\text{LW1}}$, $\hat{f}_{\text{LW2}}$ and $\hat{f}_{\text{LWeff}}$ from Linton and Wang (2016) to those of the Nadaraya-Watson estimator (see Tables 37 to 39 in Appendix 2.7.2). It is no big surprise as these three estimators are modifications of the Nadaraya-Watson estimator to account for error serial correlation. Similarly to our choice for the Nadaraya-Watson estimator, we select the Bartlett kernel together with the weighted optimal bandwidth rule with A_{nh}, i. e. , $h_{opt,h}$, as our favorite combination for the estimators $\hat{f}_{\text{LW1}}$, $\hat{f}_{\text{LW2}}$ and $\hat{f}_{\text{LWeff}}$.

We continue with the results of the estimator $\hat{f}_{\text{DLP}}$ from Dong *et al.* (2021) (see Table 31 in Appendix 2.7.2). Despite Dong *et al.* (2021) propose to use $\phi(x) = \exp(-x^2)$ as weighting function, it works out very poorly in our simulations with leading to very high ARMSE values. That was also the motivation for us to look for other weightings. Weighting functions with a normalization of x, such as $\phi(x) = \exp(-(x/x_a)^2)$ lead to way lower ARMSEs. Nevertheless, estimation without any weighting, i. e. , having a weighting function given by $\phi(x) = 1$, leads in most cases to the best results in our simulation setup, independent of the regression function. The choice of the truncation parameter κ, which determines for a given T the number of used Hermite polynomials, has some but rather minor impact on the ARMSE. In most cases, $\kappa = 1/5$ leads to the smallest ARMSE, with only a few exceptions. Nevertheless, the other choices of κ and the generalized cross-validation approach are often only slightly behind.

	$\phi(x) = \exp(-x^2)$				$\phi(x) = \exp(-(x/10)^2)$				$\phi(x) = \exp(-(x/x_a)^2)$				$\phi(x) = \exp(-(x/x_{max})^2)$				$\phi(x) = 1$			
κ	1/5	1/4	1/3	GCV	1/5	1/4	1/3	GCV	1/5	1/4	1/3	GCV	1/5	1/4	1/3	GCV	1/5	1/4	1/3	GCV
$T = 50$																				
sinc	$1.64 \cdot 10^2$	$1.23 \cdot 10^3$	$1.14 \cdot 10^5$	$1.23 \cdot 10^7$	1.139	1.188	1.265	1.173	1.814	2.160	3.099	1.542	1.132	1.183	1.262	1.167	1.125	1.176	1.255	1.178
intexp	$1.10 \cdot 10^2$	$1.01 \cdot 10^3$	$1.03 \cdot 10^5$	$1.17 \cdot 10^7$	0.852	0.906	0.992	0.912	1.584	2.065	2.347	1.547	0.845	0.901	0.988	0.906	0.836	0.891	0.979	0.913
cpr	$1.09 \cdot 10^2$	$9.54 \cdot 10^2$	$1.03 \cdot 10^5$	$1.19 \cdot 10^7$	0.814	0.877	0.975	0.895	1.484	1.833	2.172	1.439	0.808	0.873	0.971	0.890	0.798	0.863	0.963	0.892
stc	$1.09 \cdot 10^2$	$9.55 \cdot 10^2$	$1.03 \cdot 10^5$	$1.19 \cdot 10^7$	0.886	0.908	0.983	0.966	1.641	1.966	2.275	1.729	0.874	0.902	0.977	0.961	0.865	0.892	0.968	0.959
exp	$1.09 \cdot 10^2$	$9.54 \cdot 10^2$	$1.03 \cdot 10^5$	$1.19 \cdot 10^7$	0.832	0.885	0.976	0.905	1.519	1.849	2.198	1.380	0.825	0.879	0.972	0.902	0.813	0.870	0.963	0.899
pow	$1.09 \cdot 10^2$	$9.54 \cdot 10^2$	$1.03 \cdot 10^5$	$1.19 \cdot 10^7$	0.814	0.877	0.975	0.895	1.484	1.833	2.172	1.428	0.808	0.873	0.971	0.890	0.798	0.863	0.963	0.892
jump	$1.10 \cdot 10^2$	$9.58 \cdot 10^2$	$1.03 \cdot 10^5$	$1.19 \cdot 10^7$	0.846	0.894	0.979	0.939	1.557	1.909	2.187	1.591	0.838	0.889	0.976	0.934	0.830	0.879	0.968	0.932
$T = 100$																				
sinc	$4.10 \cdot 10^3$	$7.77 \cdot 10^4$	$3.88 \cdot 10^8$	$2.07 \cdot 10^{10}$	1.101	1.164	1.279	1.104	1.732	2.025	2.745	1.360	1.014	1.047	1.127	1.046	1.008	1.041	1.124	1.057
intexp	$3.36 \cdot 10^3$	$4.27 \cdot 10^4$	$3.78 \cdot 10^8$	$2.44 \cdot 10^{10}$	0.789	0.871	0.991	0.818	1.647	1.839	2.943	1.229	0.695	0.735	0.827	0.759	0.686	0.726	0.821	0.766
cpr	$3.21 \cdot 10^3$	$4.24 \cdot 10^4$	$3.80 \cdot 10^8$	$2.52 \cdot 10^{10}$	0.750	0.831	0.987	0.814	1.390	1.626	2.648	1.295	0.657	0.703	0.809	0.726	0.647	0.694	0.803	0.734
stc	$3.22 \cdot 10^3$	$4.25 \cdot 10^4$	$3.80 \cdot 10^8$	$2.51 \cdot 10^{10}$	0.873	0.931	0.988	0.917	1.818	1.864	2.633	1.825	0.753	0.752	0.817	0.817	0.746	0.744	0.811	0.815
exp	$3.21 \cdot 10^3$	$4.25 \cdot 10^4$	$3.80 \cdot 10^8$	$2.51 \cdot 10^{10}$	0.789	0.848	0.990	0.843	1.441	1.662	2.641	1.318	0.675	0.712	0.810	0.755	0.666	0.703	0.804	0.756
pow	$3.21 \cdot 10^3$	$4.24 \cdot 10^4$	$3.80 \cdot 10^8$	$2.53 \cdot 10^{10}$	0.750	0.831	0.987	0.814	1.390	1.626	2.648	1.295	0.657	0.703	0.809	0.726	0.647	0.694	0.803	0.734
jump	$3.27 \cdot 10^3$	$4.23 \cdot 10^4$	$3.81 \cdot 10^8$	$2.42 \cdot 10^{10}$	0.807	0.853	1.006	0.899	1.634	1.750	2.635	1.950	0.697	0.726	0.815	0.783	0.688	0.717	0.809	0.784
$T = 200$																				
sinc	$1.25 \cdot 10^4$	$4.61 \cdot 10^6$	$1.03 \cdot 10^{12}$	$1.19 \cdot 10^{15}$	1.472	2.894	14.445	7.115	1.421	1.951	2.906	1.289	0.904	0.948	1.022	0.963	0.900	0.944	1.020	0.976
intexp	$1.08 \cdot 10^4$	$4.34 \cdot 10^6$	$9.69 \cdot 10^{11}$	$1.01 \cdot 10^{15}$	1.147	2.607	28.115	4.207	1.292	1.589	3.112	1.893	0.534	0.596	0.690	0.640	0.527	0.589	0.686	0.648
cpr	$9.95 \cdot 10^3$	$4.23 \cdot 10^6$	$9.61 \cdot 10^{11}$	$1.02 \cdot 10^{15}$	1.080	2.603	18.479	6.524	1.072	1.493	2.461	1.014	0.490	0.563	0.671	0.601	0.483	0.556	0.667	0.612
stc	$9.95 \cdot 10^3$	$4.23 \cdot 10^6$	$9.61 \cdot 10^{11}$	$1.02 \cdot 10^{15}$	2.256	4.343	29.074	10.820	2.184	1.970	2.960	2.139	0.731	0.650	0.686	0.702	0.729	0.645	0.682	0.701
exp	$9.94 \cdot 10^3$	$4.23 \cdot 10^6$	$9.61 \cdot 10^{11}$	$1.02 \cdot 10^{15}$	1.290	3.151	17.961	4.026	1.223	1.647	2.583	1.398	0.533	0.577	0.673	0.646	0.527	0.571	0.669	0.648
pow	$9.95 \cdot 10^3$	$4.23 \cdot 10^6$	$9.61 \cdot 10^{11}$	$1.02 \cdot 10^{15}$	1.080	2.603	18.479	6.524	1.072	1.493	2.461	1.014	0.490	0.563	0.671	0.601	0.483	0.556	0.667	0.612
jump	$1.02 \cdot 10^4$	$4.22 \cdot 10^6$	$9.46 \cdot 10^{11}$	$1.00 \cdot 10^{15}$	1.504	3.600	17.104	28.208	1.600	1.789	2.294	1.769	0.590	0.598	0.679	0.672	0.587	0.592	0.675	0.672
$T = 500$																				
sinc	$3.71 \cdot 10^6$	$3.52 \cdot 10^{10}$	$4.00 \cdot 10^{19}$	$5.23 \cdot 10^{25}$	15.006	$7.54 \cdot 10^2$	$7.33 \cdot 10^5$	$2.15 \cdot 10^7$	1.228	1.228	1.267	1.072	0.824	0.858	0.916	0.879	0.822	0.856	0.916	0.888
intexp	$1.36 \cdot 10^6$	$3.26 \cdot 10^{10}$	$5.05 \cdot 10^{19}$	$1.83 \cdot 10^{25}$	15.784	$7.70 \cdot 10^2$	$5.86 \cdot 10^5$	$8.22 \cdot 10^5$	0.910	0.961	1.043	0.814	0.388	0.445	0.535	0.499	0.384	0.442	0.533	0.506
cpr	$1.35 \cdot 10^6$	$3.18 \cdot 10^{10}$	$4.99 \cdot 10^{19}$	$1.83 \cdot 10^{25}$	14.954	$7.21 \cdot 10^2$	$7.24 \cdot 10^5$	$1.00 \cdot 10^7$	0.804	0.874	0.885	0.732	0.345	0.414	0.517	0.460	0.340	0.411	0.515	0.471
stc	$1.36 \cdot 10^6$	$3.18 \cdot 10^{10}$	$4.99 \cdot 10^{19}$	$1.83 \cdot 10^{25}$	36.509	$1.38 \cdot 10^3$	$1.08 \cdot 10^6$	$1.93 \cdot 10^6$	2.124	1.456	1.055	1.154	0.741	0.560	0.542	0.556	0.742	0.558	0.541	0.554
exp	$1.35 \cdot 10^6$	$3.18 \cdot 10^{10}$	$4.99 \cdot 10^{19}$	$1.83 \cdot 10^{25}$	22.330	$9.10 \cdot 10^2$	$7.26 \cdot 10^5$	$1.25 \cdot 10^7$	1.148	0.919	0.911	0.942	0.411	0.437	0.521	0.509	0.408	0.434	0.520	0.511
pow	$1.35 \cdot 10^6$	$3.18 \cdot 10^{10}$	$4.99 \cdot 10^{19}$	$1.83 \cdot 10^{25}$	14.954	$7.21 \cdot 10^2$	$7.24 \cdot 10^5$	$1.00 \cdot 10^7$	0.804	0.874	0.885	0.732	0.345	0.414	0.517	0.460	0.340	0.411	0.515	0.471
jump	$1.35 \cdot 10^6$	$3.25 \cdot 10^{10}$	$4.76 \cdot 10^{19}$	$1.84 \cdot 10^{25}$	24.178	$9.22 \cdot 10^2$	$9.43 \cdot 10^5$	$1.43 \cdot 10^7$	1.368	1.206	0.924	1.087	0.521	0.470	0.528	0.532	0.520	0.468	0.526	0.532
$T = 1000$																				
sinc	$3.58 \cdot 10^8$	$1.26 \cdot 10^{15}$	$1.32 \cdot 10^{28}$	$4.71 \cdot 10^{35}$	$3.15 \cdot 10^2$	$2.25 \cdot 10^5$	$3.43 \cdot 10^{10}$	$1.05 \cdot 10^{12}$	1.193	1.737	1.398	1.002	0.790	0.816	0.894	0.832	0.789	0.815	1.071	0.838
intexp	$2.81 \cdot 10^8$	$1.13 \cdot 10^{15}$	$1.34 \cdot 10^{28}$	$1.50 \cdot 10^{36}$	$3.35 \cdot 10^2$	$2.13 \cdot 10^5$	$5.32 \cdot 10^{10}$	$5.40 \cdot 10^{12}$	0.825	1.321	1.589	0.954	0.309	0.363	0.480	0.407	0.307	0.361	0.583	0.413
cpr	$2.74 \cdot 10^8$	$1.15 \cdot 10^{15}$	$1.34 \cdot 10^{28}$	$1.50 \cdot 10^{36}$	$3.04 \cdot 10^2$	$2.20 \cdot 10^5$	$2.32 \cdot 10^{10}$	$4.09 \cdot 10^{11}$	0.743	1.181	1.018	0.609	0.269	0.334	0.458	0.372	0.266	0.332	0.623	0.381
stc	$2.74 \cdot 10^8$	$1.15 \cdot 10^{15}$	$1.34 \cdot 10^{28}$	$1.50 \cdot 10^{36}$	$1.39 \cdot 10^3$	$3.48 \cdot 10^5$	$6.89 \cdot 10^{10}$	$7.38 \cdot 10^{12}$	3.339	3.528	3.469	1.336	0.783	0.526	0.537	0.467	0.788	0.525	0.644	0.465
exp	$2.74 \cdot 10^8$	$1.15 \cdot 10^{15}$	$1.34 \cdot 10^{28}$	$1.50 \cdot 10^{36}$	$4.22 \cdot 10^2$	$2.77 \cdot 10^5$	$3.16 \cdot 10^{10}$	$1.04 \cdot 10^{12}$	1.181	1.118	2.014	1.184	0.353	0.364	0.470	0.419	0.351	0.362	0.591	0.418
pow	$2.74 \cdot 10^8$	$1.15 \cdot 10^{15}$	$1.34 \cdot 10^{28}$	$1.50 \cdot 10^{36}$	$3.04 \cdot 10^2$	$2.20 \cdot 10^5$	$2.32 \cdot 10^{10}$	$4.09 \cdot 10^{11}$	0.743	1.181	1.018	0.609	0.269	0.334	0.458	0.372	0.266	0.332	0.623	0.381
jump	$2.86 \cdot 10^8$	$1.18 \cdot 10^{15}$	$1.27 \cdot 10^{28}$	$1.49 \cdot 10^{36}$	$8.20 \cdot 10^2$	$3.94 \cdot 10^5$	$5.67 \cdot 10^{10}$	$3.25 \cdot 10^{11}$	2.115	1.862	1.718	1.236	0.514	0.411	0.501	0.438	0.516	0.410	0.789	0.437

Table 31: ARMSE for the estimator $\hat{f}_{\text{DLP}}$ of with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

Having examined the results of all the considered estimators for a case with a larger degree of error serial correlation and regressor endogeneity, we will now have a look on results in absence of those properties, i. e. , having $\rho_0 = \rho_1 = \rho_2 = 0$. In this case without error serial correlation and regressor endogeneity, we get a very similar picture, with the following differences: For the kernel based estimators, the cases where the bandwidth rule using the estimated variance of x_t performs best have become fewer and the other bandwidth rules have caught up. This behavior supports our choice of bandwith rule with chosing the weighted optimal bandwidth rule (for the Nadaraya-Watson estimator, see Table 40 Tables for the other estimators are available upon request). For the $\hat{f}_{\text{DLP}}$ estimator which makes use of Hermite polynomials, there is a trend to generalized cross-validation for choosing the truncation parameter. The weighting function $\phi(x) = \exp(-(x/x_{max})^2)$ in combination with the generalized cross-validation leads to the smallest ARMSE.

Table 32 shows the user choices from our discussion above, which we propose for each estimator. They are based on the ARMSE results from our simulation study and consider the case under error serial correlation and regressor endogeneity.

Estimator	Proposed user choices
$\hat{f}_{\text{NW}}$	Kernel: Bartlett, Bandwidth rule: weighted optimal with $A_{T,h}$ (i. e. , $h_{opt,h}$),
$\hat{f}_{\text{LocLin}}$	Kernel: Gaussian, Bandwidth rule: weighted optimal with $A_{T,h}$ (i. e. , $h_{opt,h}$),
$\hat{f}_{\text{LW1}}$	Kernel: Bartlett, Bandwidth rule: weighted optimal with $A_{T,h}$ (i. e. , $h_{opt,h}$),
$\hat{f}_{\text{LW2}}$	Kernel: Bartlett, Bandwidth rule: weighted optimal with $A_{T,h}$ (i. e. , $h_{opt,h}$), AR order: using BIC,
$\hat{f}_{\text{LWeff}}$	Kernel: Bartlett, Bandwidth rule: weighted optimal with $A_{T,h}$ (i. e. , $h_{opt,h}$),
$\hat{f}_{\text{DLP}}$	Weighting function: $\phi(x) = 1$, Truncation parameter: $\kappa = 1/5$.

Table 32: Proposed user choices for estimators, based on ARMSE values from our simulation study.

Now we come to the part where we compare the different estimators with each other. Table 33 shows the results of the described estimators including $\hat{f}_{\text{D-OLS}}$, the parametric D-OLS estimator of Choi and Saikkonen (2010), which serves as a benchmark. First of all, in our setup, $\hat{f}_{\text{D-OLS}}$ seems to have issues with the integrable exponential function, and for small T values also for the smooth

transition function and the power function, leading to very high ARMSE values. These high values are due to problems with the nonlinear least squares estimation, which forms the basis for the $\hat{f}_{\text{D-OLS}}$ estimator, and we have not yet been able to resolve them completely. For all functions, with exception of the sinus curves function, in most cases the $\hat{f}_{\text{DLP}}$ estimators works best, leading to the smallest ARMSE values. The $\hat{f}_{\text{DLP}}$ estimator also works surprisingly well in comparison to our benchmark estimator $\hat{f}_{\text{D-OLS}}$, especially for small values of T, of course only looking on the results where $\hat{f}_{\text{D-OLS}}$ produces a reasonable output. For larger values of the, the ARMSE of $\hat{f}_{\text{DLP}}$ compared to $\hat{f}_{\text{D-OLS}}$ is two to three times higher, but still, $\hat{f}_{\text{D-OLS}}$ is calculated knowing the regression function. The kernel based estimators seem to have issues with the cointegrating polynomial regression function and the power function, with average ARMSE values even increasing with increasing sample size T. One exception of the kernel based estimators is the local linear estimator, which is for those two functions worse than for the other functions, but ARMSE is decreasing with sample size and reasonable small. The local linear estimator is from the kernel based estimators in our setup the on average best working estimator. When excluding the cointegrating polynomial regression function and the power function, the LW-efficient estimator works best from the kernel based estimators.

The impact of regressor endogeneity and error serial correlation on ARMSE can be seen when comparing Table 33 with the results from simulations where $\rho_0 = \rho_1 = \rho_2 = 0$, which are shown in Table 41 in Appendix 2.7.2. In general, ARMSE reduces quite substantially in absence of regressor endogeneity and error serial correlation. For the functions intexp, stc, exp and jump, reduction in ARMSE can be up to about two thirds, across estimators. For the kernel based estimators, reduction is smaller for the sinc, cpr and pow functions. The local linear estimator also works best in the setup without regressor endogeneity and error serial correlation among the kernel based estimators. For the Hermite polynomial based estimator $\hat{f}_{\text{DLP}}$, reductions are smaller for the sinc function and larger for the cpr and pow functions, with reductions up to 85%. The observation of such large reductions in ARMSE also reflects that the assumptions of the $\hat{f}_{\text{DLP}}$ do not allow for error serial correlation nor regressor endogeneity. For the scenario without error serial correlation and regressor endogeneity, the $\hat{f}_{\text{DLP}}$ estimator comes for the cpr and pow functions particularly close to the $\hat{f}_{\text{D-OLS}}$ estimator, which highlights the potential of nonparametric estimation.

In summary, the following can be said: The $\hat{f}_{\text{DLP}}$ estimator of Dong *et al.* (2021) with using no weighting performs best in terms of ARMSE. If it is desired to choose a kernel-based estimator, we recommend the $\hat{f}_{\text{LocLin}}$ the local linear estimator in conjunction with the Gaussian kernel and the

weighted optimal bandwidth selection $h_{opt,h}$.

	$\hat{f}_{\text{NW}}$	$\hat{f}_{\text{LocLin}}$	$\hat{f}_{\text{LW1}}$	$\hat{f}_{\text{LW2}}$	$\hat{f}_{\text{LWeff}}$	$\hat{f}_{\text{DLP}}$	$\hat{f}_{\text{D-OLS}}$
$T = 50$							
sinc	1.244	1.176	1.194	1.244	1.179	1.178	–
intexp	1.024	0.949	0.953	1.025	0.936	0.913	$2.53 \cdot 10^{79}$
cpr	1.489	1.178	1.450	1.491	1.455	0.892	0.783
stc	1.041	0.952	0.972	1.041	0.953	0.959	$4.18 \cdot 10^{2}$
exp	1.018	0.938	0.945	1.018	0.925	0.899	0.672
pow	2.144	1.563	2.144	2.155	2.205	0.892	$2.59 \cdot 10^{3}$
jump	1.048	0.957	0.975	1.048	0.953	0.932	–
$T = 100$							
sinc	1.124	1.068	1.061	1.124	1.037	1.057	–
intexp	0.920	0.849	0.839	0.920	0.813	0.766	$1.45 \cdot 10^{140}$
cpr	1.552	1.065	1.509	1.552	1.501	0.734	0.527
stc	0.937	0.855	0.857	0.937	0.830	0.815	78.978
exp	0.917	0.844	0.834	0.917	0.807	0.756	0.453
pow	2.448	1.396	2.440	2.454	2.470	0.734	0.497
jump	0.945	0.861	0.861	0.945	0.829	0.784	–
$T = 200$							
sinc	1.018	0.975	0.949	1.018	0.922	0.976	–
intexp	0.832	0.768	0.749	0.832	0.720	0.648	$8.70 \cdot 10^{27}$
cpr	1.650	0.958	1.609	1.648	1.602	0.612	0.372
stc	0.850	0.775	0.767	0.850	0.736	0.701	0.443
exp	0.832	0.766	0.748	0.832	0.718	0.648	0.321
pow	2.755	1.226	2.742	2.758	2.767	0.612	0.356
jump	0.860	0.785	0.772	0.860	0.737	0.672	–
$T = 500$							
sinc	0.871	0.844	0.801	0.871	0.775	0.888	–
intexp	0.711	0.661	0.630	0.711	0.603	0.506	$1.00 \cdot 10^{112}$
cpr	1.838	0.826	1.800	1.834	1.789	0.471	0.233
stc	0.732	0.674	0.648	0.732	0.618	0.554	0.277
exp	0.713	0.662	0.630	0.713	0.603	0.511	0.200
pow	3.258	1.020	3.241	3.254	3.246	0.471	0.228
jump	0.743	0.684	0.655	0.743	0.622	0.532	–
$T = 1000$							
sinc	0.780	0.758	0.711	0.780	0.687	0.838	–
intexp	0.638	0.594	0.561	0.638	0.536	0.413	$7.48 \cdot 10^{112}$
cpr	2.046	0.758	2.009	2.041	1.997	0.381	0.164
stc	0.664	0.612	0.582	0.664	0.553	0.465	0.195
exp	0.640	0.596	0.562	0.640	0.537	0.418	0.141
pow	3.733	0.925	3.716	3.727	3.718	0.381	0.162
jump	0.674	0.622	0.588	0.674	0.556	0.437	–

Table 33: ARMSE for all considered estimators with their respective parameter choices, with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

2.5.2 Specification Tests – Null Rejection Probabilities and Size-corrected Power

We now turn to comparing the specification tests in terms of null rejection probabilities and size-corrected power. Starting with null rejection probabilities, we test for the regression functions from Table 27 which include θ parameters, i. e. , $f_2, \ldots, f_7$, the null hypothesis $H_0 : f(x) = g(x, \theta_0)$ from (52). The DGP under the null hypothesis is given by

$$\begin{aligned} y_{0,t} &= f_i(x_t) + u_t, \quad t = 1, \ldots, T, \quad i = 2, \ldots, 7 \\ x_t &= x_{t-1} + v_t, \quad x_0 = 0, \end{aligned}$$

as in Equation (123), and we consider three data generating processes (DGPs) under the alternative. Those are given by

$$\begin{aligned} y_{1,t} &= f_i(x_t) + \eta_t, \\ y_{2,t} &= f_i(x_t) + \sqrt{|x_t|} + u_t, \\ y_{3,t} &= f_i(x_t) + 0.1t + u_t, \\ x_t &= x_{t-1} + v_t, \quad x_0 = 0, \end{aligned} \tag{126}$$

with u_t and v_t defined in Equation (124). For $y_{1,t}$, we added with η_t a random walk component, i. e. , an integrated process, which eliminates the cointegration relationship. With $y_{2,t}$, we have a misspecified though still cointegrating model, with including the additional stochastic term $\sqrt{|x_t|}$. In $y_{3,t}$, we also have a misspecified model but now with an additional deterministic component, in contrast to the additional stochastic component in $y_{2,t}$.

Before comparing the different specification tests, we look into the null rejection probabilities and size-corrected power results of each individual specification test with its respective parameter choices. Similar to the estimators, the aim is to first identify for each specification test a test statistic version or a set of parameters which works best across the different regression functions. Again, we mainly focus our comparison on a case with non-zero error serial correlation and regressor endogeneity, i. e. , with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$ but also comparing with the results for the case without error serial correlation and regressor endogeneity.

For the Choi and Saikkonen (2010) specification test, the statistic $C_{LL}^{b,max}$ is our favorite choice. It is leading in all scenarios to the smallest null rejection probabilities with being closest to the nominal 5% level in almost all cases, compared to the C_{LL} and $C_{NLLS}^{b,max}$ statistics (see Table 43 in Appendix 2.7.2). At the same time, size-corrected power of $C_{LL}^{b,max}$ is in some cases highest and in

the other cases most of the time only slightly behind, especially for $T \leq 200$ (compare Table 57 in Appendix 2.7.2). We observe, that size-corrected power is lowest across regression functions and sample sizes for DGP 2, where we added the square root of the absolute value of the regressor. For cases without error serial correlation and regressor endogeneity, $C_{NLLS}^{b,max}$ leads for $T = 50$ to quite high size-corrected power for DGP 1 and DGP 3, while having very small null rejection probabilities (compare Tables 42 and 56 in Appendix 2.7.2).

For the Dong and Gao (2018) specification test, null rejection probabilities using the asymptotic critical values are easily above 30%, even in the absence of error serial correlation and regressor endogeneity (see Table 44 in Appendix 2.7.2) and going above 50% otherwise (see Table 45 in Appendix 2.7.2). Additionally, null rejection probabilities are increasing with sample size for the asymptotic critical values[50]. Dong and Gao (2018) also propose a bootstrap simulation procedure which we use with a block bootstrap and a wild bootstrap. The block boostrap leads – independent of the choice of κ – to the smallest size-distortions in case of error serial correlation and regressor endogeneity. In absence of error serial correlation and regressor endogeneity, both bootstrap methods are about the same in terms of null rejection probabilities. The impact of the truncation parameter κ seems to be rather negligible. When looking on the size-corrected power results in Tables 58 and 59 there is no clear winner, sometimes $\kappa = 1/5$ is ahead, sometimes $\kappa = 1/3$. To make a choice, we select $\kappa = 1/5$ for the performance comparison with the other specification tests. For $\kappa = 1/5$, the truncation parameter k is smallest and thus uses the least amount of Hermite polynomials in the test statistic.

Let us now turn to the Wang and Phillips (2012) specification test. Going through the results, the bandwidth rule $T^{-1/3}$ together with the Bartlett kernel seems to be a reasonable choice across regression functions and extent of error serial correlation and regressor endogeneity, which are both usually unknown in practice (compare for null rejection probabilities Table 46 and for size-corrected power Tables 60 to 65 in Appendix 2.7.2)[51]. In those cases, the null-rejection probabilities are often rather at the lower end while size-corrected power usually being not far behind the leading bandwidth-kernel pairs. In contrast, the Gaussian kernel leads frequently to substantially larger size distortions across bandwidth rules compared to the Bartlett kernel. One exception are the bandwidth rules using the estimated variance of x_t. For those, null rejection probabilities are usually smallest for the Gaussian kernel. For the integrable regression function in our scenario, the

[50]This might point on some flaws in our implementation or the validity of the stated limiting distribution.

[51]For the sake of brevity, we only show the results for $T = 200$. The remaining tables are available on request

bandwidth rule using the estimated variance of x_t with $T^{-1/6}$ performs well in both terms, null rejection probabilities and size-corrected power, for $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, and might be a considered choice if error serial correlation and regressor endogeneity are assumed to be high. The optimal bandwidth selection procedures lead to null rejection probabilities often twice or three times as high as for the other bandwidth rules and is therefore not really an option for us.

We next turn to the simulation results of the Wang and Phillips (2016) specification test. For this, we have as an additional dimension the weighting function, which shall be a positive integrable function. We use the Gaussian kernel $\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$, which is the weighted by $\sqrt{(2\pi)}$ version of the proposed weighting function of Wang and Phillips (2016). Additionally, we also consider a version centered around the mean of the observations x_t, i. e. , $\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$, which performed better in preliminary simulations. We observe, that using the centered weighting function reduces null rejection probabilities in many cases for all bandwidth rules with the exception of those based on the sample variance of x_t. At the same time, size-corrected power is often only slightly reduced compared to the Gaussian kernel as weighting function. Thus, for those bandwidth rules the centered weighting function is our recommended choice. Nevertheless, for the Wang and Phillips (2016) specification test, the Gaussian kernel as weighting function together with the bandwidth rule based on $\hat{\sigma}_x^2 T^{-1/3}$ and the Bartlett kernel is our recommendation. Null rejection probabilities are usually in the same range of the other combination triples and sometimes even way lower with being close to zero (see Tables 48 to 49 in Appendix 2.7.2). At the same time, size-corrected power is often the largest. These observations hold in many cases across regression functions (integrable and nonintegrable), sample sizes, and extent of error serial correlation and regressor endogeneity (compare Tables 66 to 71 in Appendix 2.7.2).

The specification test of Wang *et al.* (2018) has no user choices to be made and will be directly compared with the other specification tests.

Finally, we turn the the Wang and Zhu (2020) specification test. For this test, we need to make a decision about the lag length M. For the simulations, we chose $M = 4, 6, 12, 18$, as in Wang and Zhu (2020), and additionally two sample size dependent choices $\lfloor 4(T/100)^{1/4} \rfloor$ and $\lfloor 12(T/100)^{1/4} \rfloor$. For the case $\rho_0 = \rho_1 = \rho_2 = 0$, null rejection probabilities are for all lag lengths, regression functions and sample sizes close to the nominal level of 5%. For $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, null rejection probabilities are increasing with T and reaching values of 1 or close to 1.

The driver behind seems to be ρ_1, the parameter of the moving average error component in the error term in (124). If we set $\rho_1 = 0$, while keeping $\rho_0 = 0.6$ and $\rho_2 = 0.6$, we observe null rejection probabilities close to 5% across sample sizes, regression functions and lag lengths (see Tables 50 to 52, in Appendix 2.7.2). The impact of the lag lengths on null rejection probabilities is very small and we do not identify a particular favorite. For size-corrected power, in most cases it is either $M = 4$ or $M = \lfloor 4(T/100)^{1/4} \rfloor$ leading to the highest values, with $M = 4$ being slightly ahead for DGP 1 and DGP 3, and $M = \lfloor 4(T/100)^{1/4} \rfloor$ being slightly ahead for DGP 2 (compare Tables 72 to 74 in Appendix 2.7.2). With this slight edge in our setups, we choose the lag length $M = 4$ for comparison across specification tests.

Table 34 shows the user choices from our discussion above, which we propose for each specification test. They are based on the null rejection probabilities and size-corrected power results from our simulation study.

Specification Test	Proposed user choices
T_{CS}	Test statistic: $C_{LL}^{b,max}$
T_{DG}	Truncation parameter: $\kappa = 1/5$, Critical values: Block bootstrap,
T_{WP12}	Kernel: Bartlett, Bandwidth rule: $T^{-1/3}$
T_{WP16}	kernel: Bartlett, Bandwidth rule: $\hat{\sigma}_x^2 T^{-1/3}$, Weighting function: Gaussian kernel,
T_{WWZ}	–
T_{WZ}	Lag length: $M = 4$.

Table 34: Proposed user choices for specification tests based on null rejection probabilities and size-corrected power from our simulation study.

We now turn to the comparison of null rejection probabilities and size-corrected power across the specification tests, each with our preferred parameter choices from our analysis above. First, we look on null rejection probabilities in Table 35 for the case $\rho_0 = 0.6$, $\rho_1 = 0$, and $\rho_2 = 0.6$, to have meaningful results for the Wang and Zhu (2020) test too with choosing $\rho_1 = 0$ instead of $\rho_1 = 0.4$. Across regression functions and sample sizes, it is indeed the specification test of Wang and Zhu (2020) with null rejection probabilities closest to the nominal level of 5%. Not as good but with reasonable results are the tests of Choi and Saikkonen (2010) and Dong and Gao

(2018), with the Choi and Saikkonen (2010) having null rejection probabilities above 5% for the integrable regression function and below 5% for the nonintegrable regression functions. This result holds in our study across sample sizes. For the nonintegrable regression functions, null rejection probabilities are increasing with the sample size and coming closer to 5%. For the Dong and Gao (2018) test, there is not such a clear distinction between integrable and nonintegrable regression functions visible. Null rejection probabilities are a bit too high for $T = 50$, with values between 6% and 13%, and then coming down closer down to 5% with increasing sample size T. The Wang and Phillips (2012) test works quite well for the nonintegrable regression functions for $T = 50$, but null rejection probabilities are then increasing for all regression functions with increasing sample size. with results being quite similar across regression functions for the group of nonintegrable functions. The Wang and Phillips (2016) works worse for the nonintegrable regression functions compared to the Wang and Phillips (2012) test, with null rejection probabilities varying a lot more across regression functions for a fixed sample size, e. g. , with values between 0.5% and 16% for $T = 200$. For the integrable regression function, the null rejection probabilities are quite stable around 13%-14%. The Wang *et al.* (2018) test, which does not allow for power functions and jump functions, has null rejection probabilities close to zero or even zero for $T = 50$ and then increasing especially for both the integrable and nonintegrable exponential regression functions. For the smooth transition cointegration regression function and the cointegrating polynomial regression function, the null rejection probabilities for the Wang *et al.* (2018) test are increasing with sample size but remain below the nominal level of 5%. Let us have a brief look on the null rejection probabilities for the case without error serial correlation and regressor endogeneity, i. e. , $\rho_0 = \rho_1 = \rho_2 = 0$. The specification tests show similar patterns with respect to variability of null rejection probabilities across regression functions but the levels itself are way smaller and barely exceeding the nominal level of 5% or often even below (compare Table 53 in Appendix 2.7.2), which illustrates the detrimental effect of error serial correlation and regressor endogeneity. Introducing a moving average error component by setting $\rho_1 = 0.4$, leads only to slight increases in null rejection probabilities for all but one of the specification tests. For the Wang and Zhu (2020) test null rejection probabilities are heavily increasing and reaching values of around 69% for $T = 200$ and 99% for $T = 500$ (compare Table 55 in Appendix 2.7.2). This is purely driven by the moving average error component and not due to an interaction with the error serial correlation and regressor endogeneity, as we observe those high null rejection probabilities for the Wang and Zhu (2020) test also for the case $\rho_0 = 0$, $\rho_1 = 0.4$, and $\rho_2 = 0$ (compare Table 54 in Appendix 2.7.2). Summarizing our observations from the null

rejection probability simulations, the tests of Choi and Saikkonen (2010) and Dong and Gao (2018) lead to the most satisfactory results across regression functions and sample sizes for the different extents of regressor endogeneity and error serial correlation, potentially including a moving average error component.

	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}
$T = 50$						
intexp	0.064	0.131	0.113	0.111	0.022	0.050
cpr	0.023	0.081	0.045	0.018	0.000	0.059
stc	0.029	0.067	0.040	0.003	0.000	0.061
exp	0.026	0.080	0.060	0.084	0.033	0.056
pow	0.042	0.102	0.063	0.022	–	0.056
jump	–	0.092	0.064	0.111	–	0.055
$T = 100$						
intexp	0.076	0.033	0.138	0.135	0.044	0.050
cpr	0.032	0.072	0.065	0.049	0.001	0.052
stc	0.028	0.067	0.054	0.005	0.002	0.055
exp	0.041	0.075	0.086	0.133	0.075	0.049
pow	0.049	0.076	0.078	0.026	–	0.053
jump	–	0.076	0.089	0.145	–	0.048
$T = 200$						
intexp	0.069	0.033	0.180	0.141	0.075	0.047
cpr	0.029	0.060	0.087	0.059	0.005	0.050
stc	0.027	0.057	0.069	0.005	0.005	0.049
exp	0.037	0.063	0.102	0.160	0.125	0.049
pow	0.046	0.062	0.098	0.019	–	0.050
jump	–	0.067	0.112	0.161	–	0.048
$T = 500$						
intexp	0.068	0.090	0.235	0.145	0.109	0.052
cpr	0.034	0.056	0.125	0.068	0.014	0.048
stc	0.030	0.050	0.099	0.012	0.017	0.048
exp	0.042	0.057	0.141	0.164	0.189	0.047
pow	0.047	0.056	0.129	0.010	–	0.047
jump	–	0.060	0.148	0.156	–	0.048
$T = 1000$						
intexp	0.078	0.036	0.282	0.131	0.135	0.047
cpr	0.045	0.051	0.138	0.065	0.030	0.048
stc	0.042	0.048	0.115	0.011	0.030	0.049
exp	0.047	0.052	0.150	0.144	0.250	0.049
pow	0.051	0.049	0.142	0.010	–	0.048
jump	–	0.055	0.157	0.138	–	0.048

Table 35: Null rejection probabilities for the specification tests with $\rho_0 = 0.6$, $\rho_1 = 0.0$, and $\rho_2 = 0.6$.

Nevertheless, null rejection probabilities have to be seen in conjunction with size-corrected power results, to which we turn now. The term *size-corrected* refers to using empirical critical values instead of theoretical critical values, such that null rejection probabilities are exactly at 5% under the null hypothesis. This is to make null rejection probabilities under the alternative hypothesis, i. e. , power, more comparable. Table 36 shows the results for $\rho_0 = 0.6$, $\rho_1 = 0$, and $\rho_2 = 0.6$. Under alternative DGP 1, which reflects with the random walk error term a spurious regression,

the Choi and Saikkonen (2010) test leads the ranking in terms of size-corrected power, followed the closest but with some distance by the Wang *et al.* (2018) test. All the other tests are in many cases considerably further behind, except in a few cases. For alternative DGP 2, where we included the square root of the absolute value of the regressor in the regression function, size-corrected power is much lower compared to the DGP 1 case, which is not a surprise since there we added a random walk and here we only have a misspecified model. We observe for some tests differences in size-corrected power between the integrable and nonintegrable regression functions. For the integrable regression function, size-corrected power is rather low, going up to 22% for the Wang and Phillips (2016) test, for sample sizes up to $T = 200$. This changes for sample sizes $T = 500$ and $T = 1000$, with size-corrected power reaching values of up to 96% by the Wang and Phillips (2012) test, closely followed by the Wang *et al.* (2018) and Wang and Phillips (2016) tests. The Choi and Saikkonen (2010) test exhibits some size-corrected power, up to 42%, while the Dong and Gao (2018) and Wang and Zhu (2020) tests are far behind, reaching values of only around 18%. For the nonintegrable regression functions, the picture changes slightly, now with the four specification tests Dong and Gao (2018), Wang and Phillips (2012), Wang and Phillips (2016) and Wang *et al.* (2018) leading to the highest size-corrected power with values often quite close to each other. Nevertheless, the Dong and Gao (2018) test stands out somewhat and delivers by far the best results for the smooth transition cointegration function and the power function, for which the size-corrected power is generally significantly lower across the specification tests. The Choi and Saikkonen (2010) and Wang *et al.* (2018) tests are mostly quite far behind. Finally, we examine the size-corrected power results for alternative DGP 3, where we added a deterministic trend. For the integrable regression function, the Choi and Saikkonen (2010) test delivers the highest size-corrected power, leading for $T = 100$ already to 86%. The Wang *et al.* (2018) follows with some distance and then the Wang and Phillips (2012) test slightly behind. The Wang and Zhu (2020) leads to comparable size-corrected power for $T \leq 500$ but not for smaller samples. Again, for the integrable regression function, the Dong and Gao (2018) test is far behind, whereby the Wang and Phillips (2016) reaches at least half the size-corrected power of the other leading tests. For the nonintegrable regression functions, the Choi and Saikkonen (2010) test delivers the highest size-corrected power as well, again followed by the Wang *et al.* (2018) test, but now with the Dong and Gao (2018) test on the third place for $T \leq 200$ with one exemption, for the power function the Wang and Phillips (2012) leads to the third highest size-corrected power for $T \leq 200$. For $T \geq 500$, the Wang and Zhu (2020) often leads to the third highest size-corrected power. The Wang and

Phillips (2016) test is sometimes more and sometimes less far behind the other tests. For the case without error serial correlation and regressor endogeneity, i. e. with $\rho_0 = \rho_1 = \rho_2 = 0$, size-corrected power is substantially elevated across all tests, without changing the orders (compare Table 75 in Appendix 2.7.2). Introducing a moving average error component additionally to the error serial correlation and regressor endogeneity for which we discussed the results in detail above, i. e. having $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, the size-corrected power slightly worsens with exception of the Wang and Zhu (2020) test under alternative DGP 3, where we can observe a more pronounced decline in size-corrected power (compare Table 76 in Appendix 2.7.2).

Now, with all the results we observed, there is no uniformly best specification test in terms of size-corrected power. The Choi and Saikkonen (2010) test performs best for DGP 1 and DGP 3, but for DGP 2 there are others ahead. Nevertheless, our recommended specification test is the Choi and Saikkonen (2010) test with null rejection probabilities very close to the nominal level of 5%, even under error serial correlation and regressor endogeneity, combined with very strong size-corrected power in at least two out of three DGPs in our setup. As a reminder, we used the $C_{LL}^{b,max}$ test statistic from Choi and Saikkonen (2010). The Dong and Gao (2018), which also had very good null rejection probabilities, works quite well for the nonintegrable regression functions under all three DGPs and is especially recommended for large sample sizes. The null rejection probabilities of the Wang and Phillips (2012) and Wang *et al.* (2018) tests were not as good but still reasonable, for the Wang and Phillips (2012) at least for the nonintegrable regression functions, and performed also uniformly quite well in the size-corrected power simulations.

	H_1: DGP 1						H_1: DGP 2						H_1: DGP 3					
M	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}
$T = 50$																		
intexp	0.280	0.060	0.233	0.125	0.253	0.082	0.035	0.016	0.051	0.028	0.028	0.057	0.281	0.027	0.145	0.047	0.167	0.069
cpr	0.263	0.157	0.092	0.118	0.204	0.089	0.047	0.129	0.084	0.132	0.110	0.053	0.286	0.104	0.069	0.090	0.118	0.058
stc	0.253	0.141	0.066	0.105	0.171	0.084	0.053	0.077	0.050	0.078	0.056	0.049	0.279	0.101	0.053	0.086	0.091	0.057
exp	0.289	0.203	0.161	0.210	0.235	0.078	0.064	0.284	0.212	0.387	0.246	0.054	0.311	0.156	0.114	0.168	0.169	0.057
pow	0.304	0.170	0.214	0.151	–	0.083	0.051	0.068	0.060	0.059	–	0.050	0.390	0.129	0.163	0.134	–	0.063
jump	–	0.223	0.205	0.226	–	0.085	–	0.135	0.134	0.200	–	0.048	–	0.147	0.149	0.177	–	0.057
$T = 100$																		
intexp	0.550	0.063	0.376	0.173	0.418	0.127	0.043	0.020	0.077	0.075	0.053	0.049	0.857	0.023	0.441	0.138	0.516	0.156
cpr	0.501	0.226	0.182	0.166	0.313	0.158	0.051	0.210	0.147	0.242	0.221	0.052	0.704	0.224	0.157	0.164	0.280	0.100
stc	0.494	0.194	0.126	0.146	0.265	0.150	0.049	0.085	0.051	0.067	0.065	0.046	0.704	0.179	0.100	0.135	0.187	0.094
exp	0.491	0.304	0.305	0.311	0.409	0.137	0.095	0.503	0.448	0.702	0.561	0.050	0.694	0.358	0.317	0.382	0.449	0.138
pow	0.560	0.199	0.342	0.180	–	0.145	0.053	0.089	0.072	0.063	–	0.049	0.867	0.211	0.406	0.206	–	0.133
jump	–	0.304	0.334	0.318	–	0.156	–	0.292	0.285	0.432	–	0.051	–	0.353	0.359	0.376	–	0.146
$T = 200$																		
intexp	0.748	0.083	0.514	0.213	0.586	0.202	0.082	0.033	0.170	0.217	0.136	0.054	0.993	0.036	0.640	0.232	0.716	0.514
cpr	0.684	0.326	0.306	0.236	0.512	0.274	0.071	0.435	0.293	0.415	0.460	0.050	0.889	0.398	0.331	0.268	0.559	0.285
stc	0.684	0.294	0.236	0.225	0.405	0.294	0.052	0.123	0.065	0.100	0.111	0.050	0.900	0.330	0.197	0.215	0.402	0.289
exp	0.709	0.462	0.481	0.456	0.606	0.249	0.223	0.815	0.781	0.921	0.905	0.069	0.884	0.574	0.541	0.584	0.667	0.424
pow	0.736	0.235	0.483	0.223	–	0.278	0.046	0.150	0.098	0.082	–	0.055	0.982	0.257	0.595	0.292	–	0.403
jump	–	0.434	0.500	0.405	–	0.222	–	0.594	0.534	0.649	–	0.067	–	0.531	0.554	0.494	–	0.440
$T = 500$																		
intexp	0.903	0.135	0.738	0.311	0.767	0.312	0.243	0.122	0.632	0.710	0.524	0.064	1.000	0.151	0.811	0.400	0.852	0.791
cpr	0.872	0.440	0.558	0.299	0.765	0.447	0.161	0.802	0.706	0.619	0.898	0.060	0.980	0.560	0.589	0.353	0.799	0.654
stc	0.876	0.417	0.467	0.261	0.655	0.547	0.083	0.300	0.167	0.194	0.316	0.051	0.977	0.580	0.517	0.409	0.757	0.697
exp	0.880	0.577	0.718	0.597	0.828	0.423	0.497	0.955	0.994	0.982	1.000	0.167	0.983	0.726	0.775	0.723	0.864	0.870
pow	0.897	0.307	0.704	0.299	–	0.459	0.057	0.337	0.215	0.125	–	0.050	0.998	0.358	0.789	0.377	–	0.747
jump	–	0.533	0.726	0.515	–	0.352	–	0.880	0.889	0.820	–	0.153	–	0.651	0.781	0.598	–	0.733
$T = 1000$																		
intexp	0.962	0.149	0.855	0.383	0.876	0.420	0.418	0.176	0.956	0.861	0.912	0.163	1.000	0.162	0.900	0.474	0.920	0.951
cpr	0.953	0.533	0.755	0.356	0.886	0.580	0.295	0.953	0.962	0.714	0.995	0.100	0.997	0.653	0.776	0.406	0.915	0.921
stc	0.953	0.522	0.673	0.317	0.818	0.722	0.116	0.616	0.372	0.309	0.617	0.051	0.995	0.640	0.699	0.448	0.869	0.951
exp	0.961	0.682	0.876	0.685	0.937	0.572	0.717	0.983	1.000	0.993	1.000	0.518	0.998	0.814	0.894	0.779	0.944	0.985
pow	0.961	0.370	0.855	0.361	–	0.586	0.083	0.650	0.457	0.206	–	0.051	0.999	0.421	0.894	0.450	–	0.951
jump	–	0.602	0.864	0.597	–	0.444	–	0.967	0.987	0.882	–	0.417	–	0.722	0.891	0.670	–	0.908

Table 36: Size-corrected power for the specification tests with $\rho_0 = 0.6$, $\rho_1 = 0.0$, and $\rho_2 = 0.6$.

2.6 Summary and Conclusion

This paper compares several nonparametric estimators and nonlinear specification tests for nonlinear cointegrating regressions by means of a simulation study. From the nonparametric estimators, the Dong *et al.* (2021) estimator without any weighting works best in our simulations. Compared to parametric estimation with knowing the functional form, the results are in some cases only slightly worse. The negative impact from error serial correlation and regressor endogeneity does not come as a surprise, since the assumptions of the Dong *et al.* (2021) estimator do not allow for such properties. In that sense, an estimator based on orthogonal series expansion, such as this one, with additional corrections for error serial correlation and regressor endogeneity could be a possible avenue for further research and beneficial for practitioners. From the specification tests, the subresidual based test of Choi and Saikkonen (2010), using residuals from leads and lags regression, performs overall best in our simulations. For two of the three alternative data generation processes, this test performed exceptionally well in terms of null rejection probabilities and size-corrected power, even in case of error serial correlation and regressor endogeneity. Overall, these observations lead us to the conclusion that nonparametric estimation offers a reasonable alternative for nonlinear cointegration relationships when the functional form is unknown, and the knowledge that tests are available which can detect misspecifications with sufficient accuracy.

2.7 Appendix

2.7.1 Assumptions

Here we brought together all the assumptions of the considered estimators and specification tests to give the reader the opportunity to directly see each underlying assumptions without going into the original article. In an earlier version of the paper we used a merged set of assumptions, some of which were much more restrictive than in the individual assumptions (for example, assuming iid errors for all estimators and tests), but this obscures the restrictions of the individual estimators or tests, and so we have refrained from doing so again. The assumptions presented here may differ slightly from the original assumptions in that we have excluded some special cases that are not covered here. For example, if not only integrated but also fractionally integrated processes were covered, we have removed this part of the assumptions to improve comprehensibility. In addition, we have renamed some of the error terms and regressors as well as other parameters or functions in order to standardize them and facilitate the comparison of assumptions between the estimators

or the tests.

Assumption 5 *(short memory case only)* $x_t = \sum_{j=1}^{t} v_j$, *with* $v_j = \sum_{k=0}^{\infty} c_{v,k}\varepsilon_{j-k}$, *where* $c_{v,0} \neq 0$ *and* $\sum_{k=0}^{\infty} |c_{v,k}| < \infty$ *and* $\omega_{vv} \coloneqq \sum_{k=0}^{\infty} c_{v,k} \neq 0$.

Assumption 6 $u_t = \sum_{j=0}^{\infty} C_{u,j}\eta_{t-j}$, *with* $C_{u,j} = (c_{u_1,j}, c_{u_2,j})$ *satisfying* $\sum_{j=0}^{\infty} j^{1/4}\left(|c_{u_1,j}| + |c_{u_2,j}|\right)$ *and* $\sum_{j=0}^{\infty} C_{u,j} \neq 0$.

It follows that

$$\mathbb{E}(u_0^2) = \sum_{j=0}^{\infty} C_{u,j}\Sigma C_{u,j}', \quad \text{where } \Sigma \coloneqq \begin{pmatrix} 1 & \mathbb{E}(\varepsilon_0\zeta_0) \\ \mathbb{E}(\varepsilon_0\zeta_0) & \mathbb{E}(\zeta_0^2) \end{pmatrix}. \tag{127}$$

Assumption 7 $K(x)$ *is a nonnegative bounded continuous function satisfying* $\int_{-\infty}^{\infty} K(x)dx = 1$ *and* $\int_{-\infty}^{\infty} |\hat{K}(x)|dx < \infty$, *where* $\hat{K}(x) = \int_{-\infty}^{\infty} e^{ixt}K(t)dt$.

Assumption 7 is satisfied for, e. g. , the normal kernel and kernels with compact support.

Assumption 8 *For given* x, *there exists a real positive function* $f_1(s,x)$ *and* $\gamma \in (0,1]$ *such that, when* δ *is sufficiently small,* $|f(\delta y + x) - f(x)| \leq \delta^{\gamma} f_1(y,x)$ *for all* $y \in \mathbb{R}$ *and* $\int_{-\infty}^{\infty} K(s)(f_1(s,x) + f_1^2(s,x))ds < \infty$.

Assumption 8 includes for example regression functions $f(x) = |x|^{\beta}$ or $f(x) = 1/(1+|x|^{\beta})$ for some $\beta > 0$.

Assumption 9 *For some* $p \geq 2$

(i) $K(x)$ *satisfies* $\int_{-\infty}^{\infty} K(y)dy = 1$,

$$\int y^p K(y)dy \neq 0, \quad \int y^i K(y)dy = 0, \quad i = 1, 2, \ldots, p-1; \tag{128}$$

(ii) $K(x)$ *has compact support and is twice countinuously differentiable on* $\mathbb{R}$;

(ii) for given fixed x, $f(x)$ *has continiuous* $p+1$ *derivatives in a small neighborhood of* x.

Assumption 10 $x_t = \sum_{j=1}^{t} v_j$, *with* $v_j = \sum_{k=0}^{\infty} c_{v,k}\varepsilon_{j-k}$, *where* $c_{v,0} \neq 0$, $\sum_{k=0}^{\infty}|c_{v,k}| < \infty$ *and* $\omega_{vv} = \sum_{k=0}^{\infty} c_{v,k} \neq 0$ *and* $\{\varepsilon_j\}_{j\in\mathbb{Z}}$ *is a sequence of i.i.d. random variables with* $\mathbb{E}(\varepsilon_0) = 0$, $\mathbb{E}(\varepsilon_0^2) = 1$ *and* $\mathbb{E}(|\varepsilon_0|^{2+\delta}) < \infty$ *for some* $\delta > 0$, *and the characteristic function* $\varphi(t)$ *of* ε_0 *satisfies* $\int_{-\infty}^{\infty}(1 + |t|)|\varphi(t)|dt < \infty$.[52]

[52]We only consider the case where x_t is an integrated process and omit the case x_t is a near integrated process.

Assumption 11 *$u_t = \rho u_{t-1} + \zeta_t$ with $|\rho| < 1$ and $\nu_0 = u_0 = 0$, where $\mathcal{F}_{t,T} = \sigma(\zeta_0, \zeta_1, \ldots, \zeta_t, x_1, \ldots, x_T)$ and $\{\zeta_t, \mathcal{F}_{t,T}\}_{t=1}^{T}$ forms a martingale difference sequence satisfying, as $T \to \infty$ first, and then $m \to \infty$,*

$$\max_{m \le t \le T} |\mathbb{E}(\nu_t^2 \mid \mathcal{F}_{t-1,T}) - \sigma^2| \to_{a.s.} 0, \tag{129}$$

where σ^2 is a given constant, and $\sup_{1 \le t \le T, T \ge 1} \mathbb{E}(|\nu_t|^q \mid \mathcal{F}_{t-1,T}) < \infty$ a.s. for some $q > 2$.

Assumption 12 *(a) $\int_{-\infty}^{\infty} K(s)ds = 1$ and $K(\cdot)$ has compact support;*

(b) For any $x, y \in \mathbb{R}, |(K(x) - K(y)| \le C|x - y|$, where C is a positive constant;

(c) For $p \ge 2$,

$$\int y^p K(y)dy \neq 0, \quad \int y^i K(y)dy = 0, \quad i = 1, 2, \ldots, p-1. \tag{130}$$

Assumption 13 *(a) There exist $0 < \beta \le 1$ and $\alpha \ge 0$ such that*

$$|f(x+y) - f(x)| \le C(1 + |x|^\alpha)|y|^\beta, \tag{131}$$

for any $x \in \mathbb{R}$ and $|y|$ sufficiently small, where C is a positive constant;

(b) For given fixed x, $f(x)$ has continiuous $p+1$ derivatives in a small neighborhood of x, where $p \ge 2$ is defined as in Assumption 12(c).

Dong *et al.* (2021) make use of the following assumptions:

Assumption 14 *(i) Let $\{\varepsilon_j, -\infty < j < \infty\}$ be a scalar sequence of independent and identically distributed random variables having an absolutely continuous distribution with respect to the Lebesgue measure and satisfying $\mathbb{E}(\varepsilon_0) = 0$, $\mathbb{E}(\varepsilon_0^2) = 1$, $\mathbb{E}(|\varepsilon_0|^{q_1}) < \infty$ for some $q_1 \ge 4$. The characteristic function of ε_0 satisfies that $\int_{-\infty}^{\infty} |t||\mathbb{E}(\exp(it\varepsilon_0))|dt < \infty$;*

(ii) Let $v_t = \sum_{k=0}^{\infty} c_{v,k}\varepsilon_{t-k}$, where $\sum_{k=0}^{\infty} k|c_{v,k}| < \infty$ and $\omega_{vv} = \sum_{k=0}^{\infty} c_{v,k} \neq 0$;

(iii) For $t \ge 1$, $x_t = x_{t-1} + v_t$, and $x_0 = O_P(1)$.

Assumption 15 *(i) Suppose that $z_t = \varrho(\varepsilon_t, \ldots, \varepsilon_{t-d+1}; \nu_t)$ with fixed nonnegative integer d and measurable function $\varrho : \mathbb{R}^{d+1} \mapsto \mathbb{R}$, where the sequence $\{\nu_t\}$ is independent of $\{\epsilon_t\}$, and z_t has finite second moment; moreover, suppose that $\{\nu_t\}$ is a strictly stationary α-mixing process with mixing coefficients $\alpha(i)$ such that $\sum_{i=1}^{\infty} \alpha(i) < \infty$.*

(ii) *Suppose that $m(\cdot)$ is a density function on V_z, and if V_z is a bounded interval it is continuous in the interior of V_z and satisfies additionally $m(z) \geq c > 0$ for some constant c. Suppose also that there exists an orthonormal function sequence $\{p_i(z)\}_{i\in\mathbb{N}_0}$ in the space $L^2(V_z, m(z))$ such that $\sup_{z\in V_z} \sup_{i\geq 0} |p_i(z)m^{1/2}(z)| < \infty$.*

(iii) *Suppose that u_t and the filtration $\mathcal{F}_{t,T} = \sigma(z_{j+1}u_j, j \leq t; x_1, \ldots, x_T)$ form a martingale difference sequence such that almost surely $\mathbb{E}(u_t^2 \mid \mathcal{F}_{t-1,T}) = \sigma_u^2$ and $\max_{1\leq t\leq T} \mathbb{E}(|u_t|^{q_2} \mid \mathcal{F}_{t-1,T}) \leq C < \infty$ for some $q_2 \geq 4$ and constant C.*

Define some more quantities for Assumption 16. Denote with Ψ_K and Ξ_K matrices which approximate $Z'_{T,K}W_T Z_{T,K}$ and $Z'_{T,K}W_T^2 Z_{T,K}$, respectively.[53] Let $\lambda^{\Psi}_{\min}$, $\lambda^{\Psi}_{\max}$,$\lambda^{\Xi}_{\min}$ and $\lambda^{\Xi}_{\max}$ denote the minimum and maximum eigenvalues of Ψ_K and Ξ_K, respectively. The relation $A_T \asymp B_T$ indicates, that there exist positive absolute constants c_1, c_2, such that $c_1 < A_T/B_T < c_2$.

Assumption 16 *(i)* *Suppose that $f(r, z, x) \in L^2([0,1] \times V_z \times \mathbb{R}, \phi(z, x))$ and $f(r, z, x)$ is differentiable with respect to r, z and x, respectively, up to the orders of s_1, s_2 and s_3.*

(ii) *Suppose that k_1, k_2 and k_3 are divergent as $T \to \infty$, and that $\lambda^{\Psi}_{\min} \asymp K^{-\varsigma_1}$, $\lambda^{\Psi}_{\max} \asymp K^{\varsigma_2}$, $\lambda^{\Xi}_{\min} \asymp K^{-\iota_1}$ and $\lambda^{\Xi}_{\max} \asymp K^{\iota_2}$, in probability uniformly in $T, \varsigma_i \geq 0$ and $\iota_i \geq 0$, $i = 1, 2$ such that:*

(a) $K^{4+2(\varsigma_1+\varsigma_2)+\iota_1} k_3 = o(T)$.

(b) $K^{\varsigma_2+\iota_1}\sqrt{T}\max(k_1^{-2s_1}\log^2(k_1), k_2^{-s_2}, k_3^{-s_3}) = o(1)$.

(c) $K^{2\varsigma_2+\iota_1}\sqrt{T}\max(k_1^{-2s_1}\log^2(k_1), k_2^{-s_2+1}, k_3^{-s_3+1}) = o(1)$.

Assumption 17 *$K(x)$ has compact support, $\int_{-\infty}^{\infty} K(x)dx = 1$ and $|K(x) - K(y)| \leq C|x - y|$ whenever $|x - y|$ is sufficiently small.*

[53] The two matrices Ψ_K and Ξ_K are square matrices with dimension K. The elements of Ψ_K are given by

$$\mathbb{E}\left(p_j(z_t)p_{j'}(z_t)m(z_t)\right)\frac{d_T}{T}\sum_{t=1}^{T}\phi_i(t/T)\phi_{i'}(t/T)h_l(x_t)h_{l'}(x_t)\exp(-x_t^2),$$

and the elements of Ξ_K are given by

$$\mathbb{E}\left(p_j(z_t)p_{j'}(z_t)m^2(z_t)\right)\frac{d_T}{T}\sum_{t=1}^{T}\phi_i(t/T)\phi_{i'}(t/T)h_l(x_t)h_{l'}(x_t)\exp(-2x_t^2),$$

with $(i, j, l), (i', j', l') \in \mathcal{K} \coloneqq \{0, \ldots, k_1 - 1\} \times \{0, \ldots, k_2 - 1\} \times \{0, \ldots, k_3 - 1\}$.

Assumption 18 *(i) There exist $g_1(x)$ and $g_2(x)$ such that, for each $\theta, \theta_0 \in \Omega_0$,*

$$|g(x,\theta) - g(x,\theta_0)| \leq C||\theta - \theta_0||g_1(x),$$

and for some $0 < \beta \leq 1$,

$$|g_1(x+y) - g_1(x))| \leq C|y|^\beta g_2(x),$$

whenever y is sufficiently small;

(ii) $\int_{-\infty}^{\infty}(1 + g_1^2(x) + g_2^2(x))\pi(x)dx < \infty$.

Assumption 19 *Under H_0,* $||\hat{\theta}_T - \theta_0|| = o_P\left(\left(\sqrt{T}h\right)^{-1/2}\right)$.

Assumption 20 *(i) For $t \geq 1$, let $x_t = x_{t-1} + v_t$ with $x_0 = O_P(1)$.*[54]

(ii) Suppose that $\{v_t\}_{t\in\mathbb{Z}}$ is a linear process defined by $v_t = \Psi(L;\rho_0)\varepsilon_t = \sum_{j=0}^{\infty}\psi_j(\rho_0)\varepsilon_{t-j}$, where $\rho_0 \in \Xi$ is a p-dimensional vector of unknown parameters and Ξ is compact, $\Psi(s;\rho) = \sum_{j=0}^{\infty}\psi_j(\rho)s^j$, in which $\psi_0(\rho) = 1$, $\Psi(s;\rho)$ is continuous in s, and $\psi := \Psi(1;\rho_0) = \sum_{j=0}^{\infty}\psi_j(\rho_0) \neq 0$. Also, $\Psi(s;\rho)$ satisfies

(a) for all $\rho \neq \rho_0$ and $\rho \in \Xi \subset \mathbb{R}^p$, on a set $S \subset \{s : |s| = 1\}$ of positive Lebesgue measure, we have $|\Psi(s;\rho)| \neq |\Psi(s;\rho_0)|$;

(b) for all ρ, $\Psi(e^{i\nu};\rho)$ is differentiable in $\nu \in \mathbb{R}$ with derivative in $\mathrm{Lip}(\varsigma)$ for $\varsigma > 1/2$, that is, $|U(\nu_1;\rho) - U(\nu_2;\rho)| \leq c(\rho)|\nu_1 - \nu_2|^\varsigma$ for any $\nu_1, \nu_2 \in \mathbb{R}$ where $U(\nu;\rho) = \partial\Psi(e^{i\nu};\rho)/\partial\nu$;

(c) for all $\nu \in \mathbb{R}$, $\Psi(e^{i\nu};\rho)$ is continuous in ρ;

(d) for all $\rho \in \Xi$, $|\Psi(s;\rho)| \neq 0$, $|s| \leq 1$.

(iii) The error process $\{u_t\}_{t\in\mathbb{Z}}$ is generated by

(a) $u_t = \sum_{j=0}^{\infty}c_{u,j}\zeta_{t-j}$, where $c_{u,0} = 1$, $\lim_{j\to\infty} j^{\gamma_0}c_{u,j}$ exists with $\gamma_0 > 3/2$, and $\omega_{uu} := \sum_{j=0}^{\infty}c_{u,j} \neq 0$, or

(b) $u_t = \wp(\varepsilon_t, \ldots, \varepsilon_{t-m_0+1}; \zeta_t, \ldots, \zeta_{t-m_1+1})$ where $\min(m_0, m_1) \geq 1$, $\{\zeta_t\}_{t\in\mathbb{Z}}$ is a sequence of i.i.d.$(0,1)$ continuous variables independent of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, and the function $\wp(\ldots)$ is a measurable mapping from $\mathbb{R}^{m_0+m_1} \mapsto \mathbb{R}$ such that $\mathbb{E}(u_t) = 0$ and $\mathbb{E}(u_t^4) < \infty$ for all $t > \max(m_0, m_1)$; we define $u_t = 0$ for $t \leq \max(m_0, m_1)$.

[54] We only consider the case where x_t is an integrated regressor, even though Dong and Gao (2018) allow for fractionally integrated regressors.

Assumption 21 *(i) Let $k \coloneqq \lfloor c \cdot T^{\kappa} \rfloor$ for some $\kappa : 0 < \kappa < \frac{1}{2}$ and constant $c > 0$.*

(ii) Under H_0, there exists a consistent estimator $\hat{\theta}$ of θ_0 such that $\mathbb{P}(\|\hat{\theta} - \theta_0\| > M\zeta_T) < \varepsilon_0$ for any $\varepsilon_0 > 0$ and $M = M(\varepsilon_0)$, where $\zeta_T\sqrt{\frac{T}{kd_T}} = O(1)$. This, moreover, holds as well under H_1 with θ_0 replaced by θ_1.

(iii) Suppose that $g(x,\theta)$ is twice differentiable with respect to θ and $g(x,\theta) \in L^2(\mathbb{R})$ for every fixed $\theta \in \Theta$. Let $l_1(x,\theta) \coloneqq \frac{\partial}{\partial\theta} g(x,\theta)$ and $l_2(x,\theta) \coloneqq \frac{\partial^2}{\partial\theta\partial\theta'} g(x,\theta)$. Suppose further that $\|l_1(x,\cdot)\|, \|l_2(x,\cdot)\| \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and that there exists a positive function $l(x) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ such that $\|l_2(x,\cdot)\| \leq l(x)$.

Assumption 22 *(i) Let k satisfy Assumption 21(i).*

(ii) Under H_0, (with $g(x,\theta) \in L^2(\mathbb{R}, e^{-x^2})$), there exists a consistent estimator $hat\theta$ of θ_0 such that $\mathbb{P}(||\hat{\theta} - \theta_0|| > M\zeta_T) < \varepsilon_0$ for any $\varepsilon_0 > 0$ and $M = M(\varepsilon_0)$, where $\zeta_T\sqrt{\frac{T}{kd_T}} = O(1)$. This, moreover, holds as well under H_1 with θ_0 replaced by θ_1.

(iii) Suppose that $g(x,\theta)$ is twice differentiable with respect to θ and $g(x,\theta) \in L^2(\mathbb{R}, e^{-x^2})$ for every fixed $\theta \in \Theta$. Let $l_1(x,\theta) \coloneqq \frac{\partial}{\partial\theta} g(x,\theta)$ and $l_2(x,\theta) \coloneqq \frac{\partial^2}{\partial\theta\partial\theta'} g(x,\theta)$. Suppose further that $\|l_1(x,\cdot)\|, \|l_2(x,\cdot)\| \in L^2(\mathbb{R}, e^{-x^2})$, while $\|l_2(x,\cdot)\| \leq l(x)$ uniformly over θ with positive function $l(x) \in L^2(\mathbb{R}, e^{-x^2})$.

Assumption 23 *(i) $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed (i.i.d) continuous random variables with $\mathbb{E}(\varepsilon_0) = 0$, $\mathbb{E}(\varepsilon_0^2) = 1$, and with the characteristic function $\varphi(t)$ of ε_0 satisfying $|t||\varphi(t)| \to 0$, as $|t| \to \infty$.*

(ii) $x_t = x_{t-1} + v_t$, $x_0 = 0$, for $1 \leq t \leq T$, and $v_t = \sum_{j=0}^{\infty} c_{v,j}\varepsilon_{t-j}$, where $\omega_{vv} = \sum_{j=0}^{\infty} c_{v,j} \neq 0$ and $\sum_{j=0}^{\infty} j^{1+\delta}|c_{v,j}| < \infty$ for some $\delta > 0$.

Assumption 24 *(i) $\{u_t, \mathcal{F}_t\}_{t\geq 1}$, where $\mathcal{F}_t$ is a sequence of increasing σ-fields which is independent of ε_k, $k \geq t+1$, forms a martingale difference satisfying $\mathbb{E}(u_{t+1}^2 \mid \mathcal{F}_t) \to_{a.s.} \sigma^2 > 0$ as $t \to \infty$ and $\sup_{t\geq 1} \mathbb{E}(|u_{t+1}|^4 \mid \mathcal{F}_t) < \infty$.*

(ii) x_t is adapted to $\mathcal{F}_t$, and there exists a correlated vector Brownian motion (B_1, B_2) such that

$$\left(\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor Tt \rfloor} \varepsilon_j, \frac{1}{\sqrt{T}\sigma} \sum_{j=1}^{\lfloor Tt \rfloor} u_{j+1} \right) \Rightarrow (B_1(t), B_2(t)), \tag{132}$$

on $D[0,1]^2$ as $T \to \infty$.

Assumption 25 *$K(x)$ is a nonnegative real function satisfying $\sup_x K(x) < \infty$ and $\int_{-\infty}^{\infty} K(x)dx < \infty$.*

Assumption 26 *(i) There is a sequence of positive real numbers δ_T satisfying $\delta_T \to 0$ as $T \to \infty$ such that $\sup_{\theta\in\Theta_0}\|\hat{\theta} - \theta_0\| = o_P(\delta_T)$, where $\|\cdot\|$ denotes the Euclidean norm.*

(ii) There exists some $\epsilon_0 > 0$ such that $\frac{\partial^2 g(x,\theta)}{\partial\theta^2}$ is continuous in both $x \in \mathbb{R}$ and $\theta \in \tilde{\Theta}_0$, where $\tilde{\Theta}_0 = \{\theta : \|\theta - \theta_0\| \le \epsilon_0, \theta_0 \in \Theta_0\}$.

(iii) Uniformly for $\theta_0 \in \Theta_0$,

$$\left|\frac{\partial g(x,\theta)}{\partial\theta}\right|_{\theta=\theta_0} + \left|\frac{\partial^2 g(x,\theta)}{\partial\theta^2}\right|_{\theta=\theta_0} \le C(1 + |x|^\beta), \tag{133}$$

for some constants $\beta \ge 0$ and $C > 0$.

(iv) Uniformly for $\theta_0 \in \Theta_0$, there exist $0 < \gamma' \le 1$ and $\max\{0, 3/4 - 2\beta\} < \gamma \le 1$ such that

$$\left|\frac{\partial g(x+y,\theta_0)}{\partial\theta_0} - \frac{\partial g(x,\theta_0)}{\partial\theta_0}\right| \le C|y|^\gamma \begin{cases} 1 + |x|^{\beta-1} + |y|^\beta, & \text{if } \beta > 0, \\ 1 + |x|^{\gamma'-1}, & \text{if } \beta = 0, \end{cases} \tag{134}$$

for any $x, y \in \mathbb{R}$.

Assumption 27 *$Th^2 \to \infty$, $\delta_T^2 T^{1+\beta}\sqrt{h} \to 0$ and $Th^4 log^2(T) \to 0$, where β and δ_T^2 are defined as in Assumption 26. Also, $\int_{-\infty}^{\infty}(1 + |x|^{2\beta+1})K(x)dx < \infty$ and $\mathbb{E}(|\epsilon_0|^{4\beta+2}) < \infty$.*

Assumption 28 *$x_t = x_{t-1} + v_t$, where $x_0 = 0$.*

Assumption 29 *(i) For each $t \ge 1$, $\mathbb{E}(u_t \mid \mathcal{F}_{t-1}) = 0$ and $\sup_{t\ge1}\mathbb{E}(|u_t|^{2+\eta} \mid \mathcal{F}_{t-1}) < \infty$ for some $\eta > 0$, where $\mathcal{F}_t$ is a sequence of increasing σ-fields, so that $x_t \in \mathcal{F}_{t-1}$;*

(ii) we have

$$\left(\frac{1}{\sqrt{T}}\sum_{k=1}^{\lfloor tT\rfloor}\varepsilon_k, \frac{1}{\sqrt{T}}\sum_{k=1}^{\lfloor tT\rfloor}\varepsilon_{-k}, \frac{1}{\sqrt{T}}\sum_{k=1}^{\lfloor tT\rfloor}u_k\right) \Rightarrow (B_1(t), B_2(t), B_3(t)), \quad \text{on } D_{\mathbb{R}^3}[0,\infty), \tag{135}$$

where $\{B_1(t), B_2(t), B_3(t)\}_{t\ge0}$ is a 3-dimensional Brownian motion with covariance matrix Ω.

Assumption 30 *(i) There exist a bounded real function $h(x)$ satisfying $h(x) \downarrow h(0) = 0$, as $x \downarrow 0$, and a bounded and integrable real function $N(x)$ such that, for each $\theta \in \Theta_0$, $|g(x,\theta) - g(x,\theta_0)| \le h(\|\theta - \theta_0\|)N(x)$;*

(ii) $\int_{-\infty}^{\infty}(g(s,\theta)-g(s,\theta_0))^2ds>0$ *for all* $\theta\neq\theta_0$.

Assumption 31 *Let* $p(x,\theta)$ *be one of* $g,\dot{g}_i$ *and* $\ddot{g}_{ij}, 1\leq i,j\leq m$. *There exists a real function* $N_p:\mathbb{R}\to\mathbb{R}$ *such that*

(i) $|p(x,\theta)-p(x,\theta_0)|\leq A_p(\|\theta-\theta_0\|)N_p(x)$, *for each* $\theta\in\Theta_0$, *where* $A_p(x)$ *is a real function satisfying* $A_p(t)\downarrow A_p(0)=0$, *as* $t\downarrow 0$;

(ii) for any bounded x,

$$\sup_{\theta\in\Theta_0}|p(\lambda x,\theta)-v_p(\lambda)h_p(x,\theta)|/N_p(\lambda x)=o(1), \tag{136}$$

as $\lambda\to\infty$, *where, for each* $\theta\in\Theta_0$, $h_p(x,\theta)$ *is a locally bounded function, and* $v_p(\lambda)$ *is a positive real function bounded away from zero as* $\lambda\to\infty$;

(iii) $N_p(\lambda x)\leq C\ v_p(\lambda)(1+|x|^\gamma)$ *as* $|\lambda x|\to\infty$ *for some* $\gamma>0$.

Assumption 32 *We have*

$$\sup_{1\leq i,j\leq m}\left|\frac{v_g(d_T)v_{\ddot{g}_{ij}}(d_T)}{v_{\dot{g}_i}(d_T)v_{\dot{g}_j}(d_T)}\right|<\infty\quad\text{and}\quad\int_{|s|\leq\delta}\dot{h}(s,\theta_0)\dot{h}(s,\theta_0)'ds>0, \tag{137}$$

for some $\delta>0$, *with* $\dot{h}(s,\theta_0)=((h_{\dot{g}_1}(s,\theta_0),\ldots,h_{\dot{g}_m}(s,\theta_0))'$.

Assumption 33 *Let* $x_t=\sum_{j=1}^t\xi_j$, *where* ξ_j, *for* $j\geq 1$, *is a linear process defined by* $\xi_j=\sum_{k=0}^\infty\phi_k\epsilon_{j-k}$, *with coefficients* ϕ_k, *for* $k\geq 0$, *satisfying* $\phi_0\neq 0$ *and* $\sum_{k=0}^\infty|\phi_k|<\infty$ *and* $\phi\equiv\sum_{k=0}^\infty\phi_k\neq 0$.

Assumption 34 *For each* $\theta,\theta_0\in\Omega_0$, *there exists a bounded and integrable real function* $T(x)$ *such that*

$$|g(x,\theta)-g(x,\theta_0)|\leq h(\|\theta-\theta_0\|)T(x) \tag{138}$$

where $h(x)$ *is a bounded real function satisfying* $h(x)\to 0$ *as* $|x|\to 0$.

Assumption 35 *For each* $\theta,\theta_0\in\Omega_0$, *there exist positive real functions* $T(x)$, $v(x)$, *and* $v_j(x)$, *for* $j=1,\ldots,m$, *such that, for any* $\lambda>0$

(i) $T(\lambda x)\leq v(\lambda)\left(1+|x|^\beta\right), |(\partial g(x,\theta_0))/(\partial\theta_j)|\leq T(x)$, *for* $j=1,\ldots,m$, *and*

$$\left|g(x,\theta)-g(x,\theta_0)-\sum_{j=1}^m(\theta_j-\theta_{0j})\frac{\partial g(x,\theta_0)}{\partial\theta_j}\right|\leq\|\theta-\theta_0\|^{1+\alpha}T(x), \tag{139}$$

for some $\alpha>0$ *and* $\beta>0$;

(ii) whenever x and y are in a compact set, for each $1 \leq j \leq m$,

$$\left| \frac{\partial g(\lambda x, \theta_0)}{\partial \theta_j} - \frac{\partial g(\lambda y, \theta_0)}{\partial \theta_j} \right| \leq v_j(\lambda) \left[|x - y| + R_{1j}(\lambda x) + R_{2j}(\lambda y) \right], \tag{140}$$

where $R_{1j}(z)$ and $R_{2j}(z)$ are bounded and integrable functions;

(iii) as $K \to \infty, \sup_{|x| \geq K} \max_{1 \leq j \leq m} v(x)/v_j(x) < \infty$.

Assumption 36 *Assume $x_t = x_{t-1} + v_t, t = 1, 2, \ldots$, where v_t is a zero-mean stationary process and the initial value x_0 may be any random variable with the property $\mathbb{E}\|x_0\|^4 < \infty$.*

Assumption 37 *For some $r > 4$, $w_t = [u_t, v_t']'$ is a stationary, zero-mean, strong mixing sequence with mixing coefficients of size $-4r/(r-4)$ and $\mathbb{E}\|w_t\|^r < \infty$*

Assumption 38 *The spectral density matrix $f_{ww}(\lambda)$ is bounded away from zero:*

$$f_{ww}(\lambda) \geq \varepsilon I_{p+1}, \quad \varepsilon > 0. \tag{141}$$

Assumption 39 *(i) The parameter space Θ of θ is a compact subset of $\mathbb{R}^k$ and the true parameter value $\theta_0 \in \Theta^0$, where Θ^0 denotes the interior of Θ.*

(ii) Here, $g(x, \theta)$ is three times continuously differentiable on $\mathbb{R}^p \times \Theta^*$, where Θ^* is an open set containing Θ.

Assumption 40 *For some $s \in [0, 1]$ and all $\theta \neq \theta_0$,*

$$g\left(B_v^0(s), \theta\right) \neq g\left(B_v^0(s), \theta_0\right) \quad (a.s.), \tag{142}$$

with $B_v^0(s) \coloneqq \sqrt{T} B_v(s)$.

Assumption 41

$$\int_0^1 \tilde{K}\left(B_v^0(s), \theta_0\right) \tilde{K}\left(B_v^0(s), \theta_0\right)' ds > 0 \quad (a.s.), \tag{143}$$

where $\tilde{K}(x, \theta_0) \coloneqq \left. \frac{\partial g(x,\theta)}{\partial \theta} \right|_{\theta=\theta_0}$.

2.7.2 Additional Finite Sample Performance Results

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$	$h_{opt}(x)$	$h_{opt,h}(x)$
$T = 200$																	
sinc	bartlett	1.220	1.114	1.050	1.012	1.170	1.062	1.005	0.974	0.821	0.799	0.789	0.785	0.942	0.949	0.971	0.968
	epa	1.183	1.083	1.025	0.992	1.135	1.036	0.986	0.961	0.819	0.796	0.787	0.784	0.947	0.960	0.973	0.979
	gaussian	0.901	0.881	0.885	0.899	0.890	0.884	0.904	0.926	0.787	0.781	0.780	0.780	0.887	0.914	0.902	0.920
intexp	bartlett	1.177	1.066	0.997	0.952	1.125	1.010	0.943	0.901	0.393	0.356	0.341	0.334	0.771	0.749	0.806	0.767
	epa	1.136	1.031	0.966	0.924	1.087	0.978	0.915	0.876	0.386	0.351	0.338	0.333	0.767	0.743	0.793	0.752
	gaussian	0.789	0.748	0.720	0.700	0.770	0.726	0.696	0.675	0.336	0.329	0.327	0.327	0.702	0.670	0.701	0.667
cpr	bartlett	1.807	1.719	1.667	1.637	1.766	1.677	1.632	1.610	24.680	35.740	40.742	42.964	1.600	1.609	1.707	1.686
	epa	1.814	1.734	1.690	1.667	1.776	1.699	1.665	1.651	27.672	38.576	42.863	44.568	1.654	1.671	1.756	1.771
	gaussian	2.152	2.146	2.166	2.204	2.147	2.162	2.223	2.312	42.398	45.487	46.160	46.382	2.150	2.195	2.230	2.244
stc	bartlett	1.181	1.071	1.002	0.957	1.130	1.015	0.948	0.906	1.931	2.838	3.301	3.524	0.790	0.767	0.849	0.796
	epa	1.141	1.036	0.971	0.929	1.092	0.983	0.921	0.882	2.158	3.086	3.504	3.687	0.787	0.762	0.839	0.783
	gaussian	0.805	0.765	0.737	0.719	0.787	0.743	0.715	0.696	3.463	3.793	3.870	3.896	0.727	0.698	0.740	0.713
exp	bartlett	1.177	1.066	0.996	0.951	1.125	1.009	0.942	0.900	0.478	0.564	0.644	0.698	0.771	0.748	0.805	0.764
	epa	1.136	1.030	0.965	0.923	1.086	0.977	0.914	0.875	0.496	0.606	0.693	0.748	0.767	0.741	0.792	0.748
	gaussian	0.787	0.745	0.716	0.696	0.768	0.722	0.691	0.669	0.682	0.792	0.832	0.848	0.699	0.665	0.696	0.660
pow	bartlett	2.818	2.748	2.712	2.697	2.785	2.720	2.696	2.693	48.682	69.978	79.371	83.464	2.703	2.742	2.851	2.890
	epa	2.879	2.822	2.798	2.795	2.851	2.804	2.798	2.811	54.542	75.388	83.345	86.428	2.819	2.877	2.958	3.075
	gaussian	3.871	3.901	3.973	4.075	3.879	3.960	4.118	4.322	82.414	87.969	89.112	89.475	3.895	3.993	4.026	4.073
jump	bartlett	1.180	1.069	1.000	0.955	1.128	1.013	0.946	0.904	1.636	2.564	3.097	3.381	0.800	0.772	0.837	0.785
	epa	1.139	1.034	0.969	0.927	1.090	0.981	0.919	0.879	1.857	2.838	3.344	3.591	0.795	0.766	0.825	0.770
	gaussian	0.798	0.757	0.729	0.709	0.779	0.735	0.705	0.684	3.308	3.766	3.891	3.938	0.726	0.693	0.722	0.691
$T = 500$																	
sinc	bartlett	1.124	0.987	0.908	0.863	1.066	0.930	0.861	0.824	0.768	0.758	0.755	0.754	0.792	0.801	0.825	0.829
	epa	1.085	0.956	0.884	0.844	1.030	0.904	0.842	0.812	0.767	0.757	0.755	0.754	0.796	0.810	0.827	0.841
	gaussian	0.788	0.753	0.751	0.767	0.771	0.749	0.769	0.801	0.755	0.754	0.753	0.753	0.753	0.786	0.775	0.803
intexp	bartlett	1.097	0.956	0.872	0.821	1.038	0.896	0.818	0.771	0.268	0.247	0.241	0.239	0.653	0.630	0.680	0.640
	epa	1.056	0.922	0.843	0.795	1.000	0.865	0.792	0.748	0.264	0.245	0.241	0.239	0.649	0.625	0.670	0.628
	gaussian	0.714	0.659	0.625	0.602	0.692	0.635	0.601	0.578	0.241	0.238	0.238	0.238	0.603	0.570	0.601	0.565
cpr	bartlett	1.966	1.864	1.812	1.787	1.922	1.826	1.786	1.774	72.155	$1.01 \cdot 10^2$	$1.09 \cdot 10^2$	$1.12 \cdot 10^2$	1.779	1.800	1.867	1.881
	epa	1.981	1.892	1.852	1.837	1.942	1.862	1.837	1.838	80.055	$1.06 \cdot 10^2$	$1.12 \cdot 10^2$	$1.14 \cdot 10^2$	1.843	1.877	1.919	1.988
	gaussian	2.466	2.466	2.505	2.571	2.462	2.490	2.580	2.712	$1.09 \cdot 10^2$	$1.14 \cdot 10^2$	$1.15 \cdot 10^2$	$1.15 \cdot 10^2$	2.467	2.522	2.519	2.530
stc	bartlett	1.100	0.959	0.875	0.824	1.041	0.899	0.821	0.774	3.543	5.136	5.692	5.891	0.674	0.648	0.736	0.674
	epa	1.059	0.925	0.847	0.799	1.003	0.869	0.796	0.752	3.941	5.450	5.889	6.022	0.671	0.644	0.727	0.663
	gaussian	0.725	0.671	0.637	0.615	0.703	0.647	0.614	0.593	5.688	6.040	6.095	6.109	0.627	0.594	0.642	0.610
exp	bartlett	1.097	0.956	0.872	0.821	1.038	0.896	0.818	0.771	0.481	0.650	0.750	0.800	0.654	0.630	0.681	0.639
	epa	1.056	0.922	0.843	0.795	1.000	0.865	0.792	0.748	0.518	0.703	0.798	0.842	0.651	0.625	0.670	0.627
	gaussian	0.713	0.659	0.624	0.601	0.691	0.634	0.599	0.576	0.750	0.851	0.877	0.886	0.603	0.569	0.599	0.563
pow	bartlett	3.282	3.208	3.181	3.177	3.249	3.188	3.178	3.196	$1.43 \cdot 10^2$	$1.99 \cdot 10^2$	$2.16 \cdot 10^2$	$2.21 \cdot 10^2$	3.188	3.241	3.297	3.423
	epa	3.358	3.303	3.294	3.309	3.331	3.295	3.312	3.354	$1.59 \cdot 10^2$	$2.10 \cdot 10^2$	$2.22 \cdot 10^2$	$2.25 \cdot 10^2$	3.318	3.394	3.405	3.652
	gaussian	4.600	4.651	4.761	4.914	4.611	4.721	4.933	5.217	$2.16 \cdot 10^2$	$2.25 \cdot 10^2$	$2.26 \cdot 10^2$	$2.27 \cdot 10^2$	4.621	4.722	4.704	4.748
jump	bartlett	1.099	0.958	0.874	0.823	1.040	0.898	0.820	0.773	3.130	4.919	5.647	5.940	0.685	0.655	0.720	0.662
	epa	1.058	0.924	0.846	0.798	1.002	0.868	0.795	0.751	3.545	5.307	5.919	6.134	0.681	0.650	0.710	0.650
	gaussian	0.720	0.666	0.631	0.609	0.698	0.642	0.607	0.584	5.647	6.179	6.280	6.309	0.629	0.593	0.627	0.591

Table 37: ARMSE for the estimator $\hat{f}_{\text{LW1}}$, with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$	$h_{opt}(x)$	$h_{opt,h}(x)$
$T = 200$																	
sinc	bartlett	1.337	1.239	1.172	1.129	1.293	1.185	1.120	1.081	0.837	0.810	0.798	0.792	1.021	1.018	1.043	1.025
	epa	1.305	1.212	1.148	1.107	1.264	1.160	1.099	1.065	0.834	0.806	0.794	0.789	1.025	1.027	1.041	1.030
	gaussian	0.976	0.954	0.950	0.957	0.966	0.950	0.961	0.977	0.794	0.784	0.781	0.780	0.951	0.968	0.956	0.961
intexp	bartlett	1.298	1.197	1.126	1.077	1.253	1.139	1.066	1.018	0.424	0.379	0.358	0.347	0.859	0.832	0.906	0.852
	epa	1.265	1.167	1.097	1.049	1.222	1.110	1.039	0.991	0.416	0.373	0.353	0.344	0.856	0.826	0.892	0.834
	gaussian	0.886	0.840	0.805	0.780	0.867	0.812	0.775	0.749	0.351	0.334	0.330	0.328	0.782	0.743	0.786	0.740
cpr	bartlett	1.880	1.802	1.748	1.714	1.846	1.759	1.708	1.678	17.520	27.613	33.342	36.386	1.655	1.648	1.760	1.722
	epa	1.878	1.807	1.760	1.732	1.847	1.770	1.729	1.708	20.622	31.931	37.716	40.531	1.702	1.704	1.802	1.797
	gaussian	2.168	2.162	2.178	2.209	2.163	2.176	2.225	2.287	35.635	41.451	43.455	44.361	2.165	2.199	2.240	2.247
stc	bartlett	1.302	1.201	1.130	1.081	1.258	1.144	1.071	1.022	1.482	2.265	2.745	3.013	0.880	0.850	0.951	0.881
	epa	1.269	1.171	1.101	1.054	1.226	1.115	1.044	0.997	1.704	2.604	3.105	3.363	0.876	0.844	0.939	0.865
	gaussian	0.897	0.853	0.818	0.793	0.878	0.825	0.788	0.763	2.933	3.455	3.639	3.722	0.804	0.765	0.818	0.778
exp	bartlett	1.298	1.197	1.125	1.076	1.253	1.139	1.066	1.017	0.463	0.500	0.552	0.595	0.860	0.832	0.906	0.850
	epa	1.265	1.166	1.096	1.048	1.221	1.110	1.038	0.991	0.477	0.541	0.610	0.662	0.856	0.825	0.892	0.831
	gaussian	0.883	0.838	0.802	0.776	0.864	0.809	0.771	0.744	0.579	0.697	0.756	0.787	0.780	0.739	0.782	0.735
pow	bartlett	2.855	2.792	2.754	2.736	2.826	2.763	2.736	2.727	34.615	54.341	65.331	71.095	2.733	2.758	2.877	2.902
	epa	2.907	2.854	2.830	2.825	2.882	2.836	2.828	2.837	40.765	62.799	73.827	79.096	2.843	2.889	2.978	3.081
	gaussian	3.888	3.917	3.987	4.079	3.896	3.974	4.115	4.279	69.693	80.598	84.255	85.881	3.911	4.001	4.036	4.081
jump	bartlett	1.301	1.200	1.129	1.080	1.256	1.142	1.069	1.021	1.206	1.921	2.418	2.726	0.894	0.860	0.942	0.873
	epa	1.268	1.170	1.100	1.052	1.225	1.113	1.042	0.994	1.393	2.236	2.779	3.096	0.890	0.853	0.929	0.856
	gaussian	0.892	0.847	0.811	0.786	0.873	0.818	0.781	0.754	2.635	3.276	3.535	3.661	0.807	0.765	0.806	0.762
$T = 500$																	
sinc	bartlett	1.260	1.127	1.041	0.988	1.207	1.066	0.984	0.937	0.775	0.762	0.758	0.756	0.874	0.871	0.901	0.886
	epa	1.228	1.099	1.016	0.966	1.177	1.040	0.963	0.920	0.773	0.761	0.757	0.755	0.877	0.878	0.898	0.892
	gaussian	0.886	0.840	0.824	0.829	0.867	0.826	0.830	0.852	0.758	0.754	0.754	0.753	0.825	0.842	0.835	0.845
intexp	bartlett	1.237	1.100	1.011	0.952	1.182	1.037	0.948	0.892	0.285	0.257	0.247	0.243	0.741	0.711	0.782	0.725
	epa	1.204	1.070	0.982	0.925	1.151	1.008	0.921	0.867	0.281	0.255	0.246	0.242	0.738	0.706	0.770	0.709
	gaussian	0.828	0.761	0.716	0.685	0.802	0.729	0.683	0.653	0.248	0.240	0.239	0.238	0.686	0.643	0.691	0.640
cpr	bartlett	2.041	1.945	1.887	1.854	2.003	1.904	1.853	1.831	51.761	81.461	93.965	99.491	1.830	1.834	1.926	1.920
	epa	2.040	1.957	1.910	1.889	2.007	1.923	1.889	1.882	61.008	92.477	$1.04 \cdot 10^2$	$1.08 \cdot 10^2$	1.885	1.905	1.971	2.019
	gaussian	2.475	2.475	2.512	2.572	2.471	2.497	2.580	2.688	94.673	$1.07 \cdot 10^2$	$1.11 \cdot 10^2$	$1.12 \cdot 10^2$	2.475	2.526	2.529	2.535
stc	bartlett	1.239	1.103	1.013	0.955	1.185	1.040	0.951	0.895	2.634	4.180	4.892	5.217	0.767	0.732	0.843	0.761
	epa	1.206	1.073	0.985	0.928	1.154	1.011	0.925	0.870	3.076	4.751	5.423	5.691	0.764	0.727	0.833	0.747
	gaussian	0.836	0.771	0.725	0.695	0.811	0.739	0.693	0.663	4.908	5.640	5.855	5.942	0.710	0.666	0.730	0.683
exp	bartlett	1.236	1.100	1.010	0.952	1.182	1.037	0.948	0.892	0.423	0.548	0.640	0.697	0.743	0.713	0.783	0.724
	epa	1.203	1.070	0.982	0.925	1.151	1.007	0.921	0.867	0.460	0.613	0.715	0.772	0.740	0.707	0.771	0.708
	gaussian	0.827	0.761	0.715	0.684	0.802	0.729	0.682	0.651	0.640	0.774	0.826	0.849	0.687	0.643	0.690	0.638
pow	bartlett	3.313	3.242	3.210	3.203	3.282	3.219	3.204	3.217	$1.03 \cdot 10^2$	$1.62 \cdot 10^2$	$1.86 \cdot 10^2$	$1.97 \cdot 10^2$	3.211	3.254	3.328	3.440
	epa	3.378	3.324	3.313	3.326	3.352	3.315	3.329	3.368	$1.21 \cdot 10^2$	$1.83 \cdot 10^2$	$2.06 \cdot 10^2$	$2.14 \cdot 10^2$	3.334	3.405	3.431	3.663
	gaussian	4.609	4.659	4.769	4.915	4.620	4.729	4.933	5.186	$1.88 \cdot 10^2$	$2.12 \cdot 10^2$	$2.19 \cdot 10^2$	$2.21 \cdot 10^2$	4.629	4.729	4.712	4.755
jump	bartlett	1.238	1.102	1.013	0.954	1.184	1.039	0.950	0.894	2.196	3.765	4.606	5.035	0.782	0.743	0.829	0.751
	epa	1.205	1.072	0.984	0.927	1.153	1.010	0.924	0.869	2.588	4.338	5.185	5.580	0.779	0.738	0.818	0.735
	gaussian	0.832	0.766	0.720	0.690	0.807	0.734	0.688	0.657	4.614	5.575	5.898	6.038	0.717	0.669	0.718	0.667

Table 38: ARMSE for the estimator $\hat{f}_{\text{LW2}}$, with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$	$h_{opt}(x)$	$h_{opt,h}(x)$
$T = 200$																	
sinc	bartlett	1.147	1.046	0.989	0.956	1.098	0.999	0.951	0.926	0.820	0.798	0.789	0.785	0.910	0.922	0.943	0.949
	epa	1.106	1.014	0.964	0.937	1.061	0.973	0.933	0.915	0.819	0.796	0.787	0.784	0.914	0.933	0.947	0.962
	gaussian	0.868	0.857	0.865	0.882	0.861	0.863	0.887	0.912	0.787	0.781	0.780	0.779	0.867	0.899	0.889	0.913
intexp	bartlett	1.101	0.995	0.932	0.892	1.050	0.943	0.884	0.848	0.395	0.358	0.343	0.336	0.739	0.720	0.771	0.740
	epa	1.056	0.958	0.900	0.864	1.009	0.911	0.857	0.824	0.388	0.353	0.340	0.334	0.734	0.714	0.760	0.727
	gaussian	0.751	0.718	0.695	0.678	0.735	0.699	0.675	0.657	0.338	0.329	0.328	0.327	0.680	0.653	0.680	0.653
cpr	bartlett	1.764	1.679	1.632	1.606	1.723	1.641	1.603	1.586	25.088	36.105	41.037	43.208	1.585	1.602	1.694	1.680
	epa	1.771	1.697	1.659	1.640	1.735	1.666	1.639	1.631	28.073	38.890	43.084	44.731	1.640	1.668	1.746	1.770
	gaussian	2.161	2.164	2.196	2.249	2.160	2.191	2.272	2.385	42.661	45.629	46.249	46.443	2.170	2.234	2.256	2.283
stc	bartlett	1.106	1.000	0.937	0.897	1.055	0.948	0.889	0.853	2.010	2.916	3.366	3.578	0.756	0.736	0.809	0.764
	epa	1.061	0.964	0.906	0.870	1.014	0.917	0.863	0.830	2.235	3.152	3.552	3.722	0.751	0.730	0.799	0.753
	gaussian	0.767	0.734	0.712	0.696	0.752	0.717	0.693	0.678	3.522	3.825	3.891	3.911	0.703	0.680	0.718	0.697
exp	bartlett	1.101	0.994	0.931	0.891	1.050	0.942	0.883	0.847	0.482	0.574	0.657	0.713	0.738	0.718	0.768	0.735
	epa	1.055	0.957	0.899	0.863	1.008	0.910	0.856	0.822	0.501	0.614	0.704	0.758	0.732	0.711	0.756	0.721
	gaussian	0.747	0.714	0.690	0.673	0.732	0.695	0.669	0.650	0.697	0.806	0.842	0.857	0.675	0.646	0.673	0.643
pow	bartlett	2.813	2.746	2.713	2.701	2.780	2.719	2.701	2.702	49.469	70.684	79.940	83.934	2.717	2.767	2.869	2.912
	epa	2.877	2.824	2.804	2.804	2.850	2.808	2.808	2.826	55.317	75.995	83.769	86.737	2.837	2.905	2.981	3.102
	gaussian	3.915	3.958	4.051	4.179	3.928	4.033	4.230	4.483	82.922	88.238	89.276	89.587	3.949	4.075	4.086	4.150
jump	bartlett	1.104	0.998	0.935	0.895	1.053	0.946	0.887	0.850	1.698	2.635	3.161	3.438	0.761	0.737	0.794	0.750
	epa	1.059	0.961	0.904	0.867	1.012	0.914	0.860	0.826	1.919	2.901	3.396	3.634	0.755	0.730	0.783	0.737
	gaussian	0.757	0.724	0.700	0.683	0.742	0.705	0.679	0.660	3.369	3.804	3.918	3.958	0.697	0.668	0.695	0.667
$T = 500$																	
sinc	bartlett	1.040	0.911	0.842	0.804	0.984	0.861	0.802	0.774	0.768	0.758	0.755	0.754	0.760	0.775	0.796	0.810
	epa	0.997	0.879	0.818	0.787	0.946	0.835	0.785	0.764	0.767	0.757	0.755	0.754	0.763	0.785	0.800	0.825
	gaussian	0.746	0.723	0.728	0.749	0.734	0.725	0.751	0.787	0.755	0.754	0.753	0.753	0.730	0.771	0.759	0.795
intexp	bartlett	1.011	0.878	0.803	0.759	0.954	0.824	0.756	0.717	0.270	0.248	0.242	0.240	0.621	0.603	0.645	0.615
	epa	0.966	0.843	0.774	0.734	0.913	0.793	0.732	0.696	0.266	0.246	0.241	0.239	0.617	0.598	0.636	0.604
	gaussian	0.667	0.624	0.597	0.579	0.649	0.605	0.578	0.560	0.242	0.239	0.238	0.238	0.580	0.553	0.578	0.550
cpr	bartlett	1.912	1.817	1.773	1.754	1.870	1.785	1.754	1.748	72.582	$1.01 \cdot 10^2$	$1.09 \cdot 10^2$	$1.12 \cdot 10^2$	1.757	1.789	1.841	1.868
	epa	1.929	1.848	1.817	1.808	1.892	1.825	1.809	1.818	80.457	$1.06 \cdot 10^2$	$1.12 \cdot 10^2$	$1.14 \cdot 10^2$	1.823	1.869	1.896	1.979
	gaussian	2.474	2.485	2.538	2.622	2.473	2.518	2.633	2.798	$1.09 \cdot 10^2$	$1.14 \cdot 10^2$	$1.15 \cdot 10^2$	$1.15 \cdot 10^2$	2.485	2.559	2.539	2.562
stc	bartlett	1.014	0.881	0.806	0.762	0.957	0.827	0.760	0.721	3.633	5.200	5.734	5.922	0.640	0.618	0.692	0.642
	epa	0.969	0.846	0.778	0.738	0.916	0.797	0.736	0.700	4.024	5.498	5.914	6.038	0.636	0.614	0.682	0.632
	gaussian	0.678	0.636	0.610	0.592	0.661	0.617	0.591	0.574	5.731	6.058	6.105	6.115	0.601	0.575	0.615	0.591
exp	bartlett	1.011	0.877	0.803	0.759	0.954	0.824	0.756	0.717	0.494	0.670	0.770	0.819	0.622	0.603	0.644	0.612
	epa	0.965	0.842	0.774	0.734	0.913	0.793	0.732	0.696	0.530	0.719	0.812	0.853	0.618	0.597	0.635	0.602
	gaussian	0.666	0.623	0.596	0.578	0.649	0.604	0.576	0.557	0.770	0.864	0.886	0.893	0.579	0.551	0.576	0.547
pow	bartlett	3.260	3.191	3.169	3.169	3.228	3.174	3.171	3.195	$1.44 \cdot 10^2$	$2.00 \cdot 10^2$	$2.16 \cdot 10^2$	$2.21 \cdot 10^2$	3.184	3.246	3.295	3.428
	epa	3.340	3.291	3.288	3.307	3.315	3.287	3.310	3.358	$1.60 \cdot 10^2$	$2.10 \cdot 10^2$	$2.22 \cdot 10^2$	$2.25 \cdot 10^2$	3.317	3.403	3.407	3.661
	gaussian	4.642	4.705	4.837	5.021	4.657	4.789	5.043	5.389	$2.16 \cdot 10^2$	$2.25 \cdot 10^2$	$2.26 \cdot 10^2$	$2.27 \cdot 10^2$	4.669	4.790	4.756	4.814
jump	bartlett	1.013	0.880	0.805	0.761	0.956	0.826	0.758	0.719	3.216	4.988	5.697	5.979	0.646	0.622	0.676	0.629
	epa	0.968	0.845	0.777	0.736	0.915	0.796	0.734	0.698	3.626	5.362	5.954	6.158	0.642	0.616	0.666	0.618
	gaussian	0.672	0.629	0.601	0.582	0.654	0.609	0.581	0.561	5.697	6.203	6.295	6.319	0.599	0.569	0.597	0.568

Table 39: ARMSE for the estimator $\hat{f}_{\text{LWeff}}$, with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$	$h_{opt}(x)$	$h_{opt,h}(x)$
$T = 200$																	
sinc	bartlett	0.749	0.687	0.647	0.624	0.721	0.656	0.621	0.605	0.743	0.741	0.740	0.740	0.605	0.626	0.612	0.633
	epa	0.733	0.673	0.636	0.616	0.706	0.644	0.614	0.604	0.747	0.742	0.741	0.741	0.613	0.640	0.617	0.651
	gaussian	0.579	0.577	0.600	0.632	0.575	0.595	0.641	0.683	0.741	0.740	0.740	0.740	0.589	0.638	0.593	0.636
intexp	bartlett	0.676	0.607	0.557	0.523	0.645	0.567	0.516	0.481	0.215	0.222	0.226	0.228	0.366	0.345	0.401	0.360
	epa	0.655	0.586	0.537	0.503	0.625	0.547	0.496	0.462	0.223	0.230	0.233	0.234	0.363	0.340	0.390	0.347
	gaussian	0.396	0.358	0.331	0.312	0.379	0.336	0.308	0.289	0.228	0.233	0.234	0.235	0.315	0.287	0.320	0.287
cpr	bartlett	1.461	1.416	1.388	1.372	1.441	1.395	1.370	1.360	17.277	27.513	33.303	36.374	1.362	1.366	1.446	1.447
	epa	1.487	1.446	1.424	1.415	1.469	1.430	1.415	1.414	20.357	31.848	37.695	40.535	1.416	1.430	1.498	1.537
	gaussian	1.892	1.895	1.917	1.950	1.892	1.913	1.964	2.030	35.598	41.456	43.473	44.384	1.893	1.919	1.963	1.976
stc	bartlett	0.684	0.615	0.567	0.533	0.653	0.577	0.526	0.492	1.414	2.247	2.741	3.013	0.401	0.378	0.465	0.408
	epa	0.664	0.596	0.547	0.514	0.634	0.557	0.507	0.474	1.648	2.592	3.104	3.365	0.400	0.376	0.459	0.400
	gaussian	0.429	0.393	0.368	0.351	0.413	0.373	0.348	0.334	2.932	3.457	3.640	3.722	0.365	0.339	0.387	0.364
exp	bartlett	0.676	0.606	0.557	0.522	0.645	0.567	0.515	0.481	0.310	0.429	0.517	0.577	0.365	0.344	0.403	0.359
	epa	0.655	0.586	0.537	0.503	0.624	0.547	0.496	0.461	0.343	0.484	0.584	0.647	0.362	0.339	0.393	0.345
	gaussian	0.396	0.357	0.330	0.310	0.379	0.335	0.306	0.286	0.558	0.689	0.746	0.775	0.314	0.283	0.318	0.284
pow	bartlett	2.548	2.517	2.501	2.497	2.534	2.505	2.499	2.505	34.123	54.125	65.292	71.137	2.503	2.530	2.618	2.683
	epa	2.621	2.599	2.595	2.603	2.610	2.598	2.608	2.631	40.226	62.632	73.848	79.196	2.613	2.658	2.722	2.868
	gaussian	3.610	3.643	3.704	3.781	3.621	3.693	3.810	3.951	69.678	80.710	84.407	86.052	3.622	3.681	3.727	3.763
jump	bartlett	0.681	0.612	0.564	0.529	0.651	0.574	0.522	0.488	1.155	1.929	2.446	2.759	0.419	0.390	0.450	0.394
	epa	0.661	0.593	0.544	0.510	0.631	0.554	0.504	0.470	1.356	2.251	2.808	3.127	0.417	0.386	0.443	0.384
	gaussian	0.419	0.383	0.357	0.340	0.403	0.362	0.336	0.320	2.667	3.305	3.558	3.680	0.366	0.336	0.366	0.338
$T = 500$																	
sinc	bartlett	0.700	0.618	0.567	0.538	0.667	0.582	0.537	0.516	0.738	0.737	0.737	0.737	0.510	0.530	0.523	0.548
	epa	0.682	0.602	0.555	0.529	0.650	0.568	0.528	0.513	0.739	0.737	0.737	0.737	0.516	0.541	0.526	0.565
	gaussian	0.506	0.489	0.508	0.543	0.496	0.499	0.546	0.600	0.737	0.737	0.737	0.737	0.498	0.548	0.508	0.557
intexp	bartlett	0.654	0.567	0.508	0.468	0.620	0.526	0.466	0.428	0.177	0.182	0.184	0.185	0.320	0.298	0.352	0.309
	epa	0.634	0.547	0.489	0.450	0.600	0.506	0.448	0.410	0.183	0.186	0.188	0.188	0.318	0.294	0.342	0.296
	gaussian	0.387	0.338	0.305	0.282	0.368	0.315	0.281	0.259	0.185	0.187	0.188	0.188	0.283	0.253	0.289	0.252
cpr	bartlett	1.685	1.634	1.608	1.598	1.664	1.616	1.599	1.600	51.858	81.268	93.745	99.291	1.600	1.614	1.663	1.707
	epa	1.716	1.673	1.657	1.657	1.697	1.661	1.658	1.673	61.030	92.194	$1.04 \cdot 10^2$	$1.08 \cdot 10^2$	1.661	1.687	1.713	1.817
	gaussian	2.231	2.244	2.286	2.342	2.232	2.270	2.349	2.449	94.469	$1.07 \cdot 10^2$	$1.11 \cdot 10^2$	$1.12 \cdot 10^2$	2.235	2.268	2.278	2.291
stc	bartlett	0.659	0.572	0.513	0.474	0.625	0.531	0.472	0.434	2.622	4.163	4.871	5.197	0.361	0.334	0.424	0.357
	epa	0.639	0.552	0.495	0.457	0.605	0.512	0.455	0.418	3.064	4.728	5.396	5.665	0.360	0.331	0.418	0.349
	gaussian	0.406	0.359	0.327	0.307	0.388	0.337	0.306	0.287	4.889	5.618	5.831	5.918	0.328	0.296	0.349	0.319
exp	bartlett	0.654	0.567	0.508	0.468	0.620	0.525	0.466	0.427	0.375	0.543	0.645	0.702	0.324	0.301	0.354	0.308
	epa	0.634	0.547	0.489	0.450	0.600	0.506	0.448	0.410	0.421	0.611	0.718	0.774	0.321	0.297	0.345	0.296
	gaussian	0.387	0.338	0.304	0.281	0.368	0.314	0.280	0.257	0.645	0.777	0.825	0.846	0.285	0.253	0.289	0.251
pow	bartlett	3.066	3.036	3.029	3.038	3.052	3.030	3.040	3.069	$1.03 \cdot 10^2$	$1.61 \cdot 10^2$	$1.86 \cdot 10^2$	$1.97 \cdot 10^2$	3.037	3.073	3.129	3.279
	epa	3.148	3.134	3.145	3.172	3.139	3.141	3.176	3.230	$1.21 \cdot 10^2$	$1.83 \cdot 10^2$	$2.05 \cdot 10^2$	$2.14 \cdot 10^2$	3.154	3.215	3.237	3.508
	gaussian	4.346	4.405	4.508	4.632	4.361	4.471	4.646	4.855	$1.87 \cdot 10^2$	$2.12 \cdot 10^2$	$2.18 \cdot 10^2$	$2.21 \cdot 10^2$	4.356	4.425	4.440	4.481
jump	bartlett	0.658	0.571	0.512	0.473	0.624	0.530	0.471	0.433	2.200	3.785	4.632	5.062	0.379	0.347	0.407	0.344
	epa	0.638	0.551	0.494	0.455	0.604	0.511	0.453	0.416	2.592	4.357	5.209	5.605	0.377	0.343	0.400	0.335
	gaussian	0.403	0.355	0.324	0.302	0.384	0.333	0.301	0.282	4.638	5.600	5.921	6.060	0.336	0.301	0.335	0.301

Table 40: ARMSE for the Nadaraya-Watson estimator with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

	$\hat{f}_{\text{NW}}$	$\hat{f}_{\text{LocLin}}$	$\hat{f}_{\text{LW1}}$	$\hat{f}_{\text{LW2}}$	$\hat{f}_{\text{LWeff}}$	$\hat{f}_{\text{DLP}}$	$\hat{f}_{\text{D-OLS}}$
$T = 50$							
sinc	0.774	0.750	0.782	0.774	0.789	0.792	–
intexp	0.421	0.365	0.427	0.422	0.432	0.372	$8.39 \cdot 10^{128}$
cpr	1.061	0.713	1.081	1.070	1.118	0.265	0.230
stc	0.454	0.374	0.460	0.454	0.466	0.377	0.804
exp	0.411	0.351	0.416	0.411	0.421	0.322	0.189
pow	1.796	1.191	1.834	1.817	1.911	0.265	0.225
jump	0.463	0.375	0.469	0.463	0.475	0.363	–
$T = 100$							
sinc	0.700	0.690	0.705	0.700	0.709	0.767	–
intexp	0.381	0.331	0.384	0.381	0.387	0.309	$1.51 \cdot 10^{110}$
cpr	1.233	0.666	1.243	1.240	1.271	0.189	0.153
stc	0.414	0.339	0.417	0.414	0.420	0.319	0.190
exp	0.376	0.324	0.379	0.376	0.382	0.273	0.130
pow	2.214	1.090	2.237	2.235	2.294	0.189	0.154
jump	0.426	0.347	0.429	0.426	0.432	0.306	–
$T = 200$							
sinc	0.626	0.625	0.628	0.626	0.630	0.751	–
intexp	0.345	0.302	0.346	0.345	0.348	0.256	$2.12 \cdot 10^{109}$
cpr	1.366	0.571	1.372	1.371	1.388	0.133	0.106
stc	0.378	0.313	0.380	0.378	0.381	0.269	0.132
exp	0.344	0.299	0.345	0.344	0.347	0.230	0.089
pow	2.530	0.891	2.543	2.546	2.578	0.133	0.106
jump	0.390	0.324	0.392	0.390	0.393	0.256	–
$T = 500$							
sinc	0.530	0.536	0.530	0.530	0.531	0.741	–
intexp	0.298	0.264	0.299	0.298	0.300	0.200	$1.35 \cdot 10^{131}$
cpr	1.614	0.477	1.618	1.619	1.630	0.084	0.067
stc	0.334	0.283	0.335	0.334	0.335	0.217	0.082
exp	0.301	0.265	0.301	0.301	0.302	0.182	0.057
pow	3.073	0.699	3.083	3.087	3.109	0.084	0.067
jump	0.347	0.294	0.347	0.347	0.348	0.202	–
$T = 1000$							
sinc	0.470	0.477	0.470	0.470	0.470	0.738	–
intexp	0.271	0.241	0.271	0.271	0.271	0.165	$1.48 \cdot 10^{127}$
cpr	1.840	0.437	1.844	1.846	1.851	0.059	0.047
stc	0.311	0.266	0.311	0.311	0.311	0.209	0.058
exp	0.274	0.243	0.275	0.274	0.275	0.152	0.040
pow	3.539	0.622	3.550	3.554	3.566	0.059	0.047
jump	0.320	0.275	0.321	0.320	0.321	0.175	–

Table 41: ARMSE for all considered estimators with their respective parameter choices, with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

	C_{LL}	$C_{LL}^{b,max}$	$C_{NLLS}^{b,max}$
$T = 50$			
intexp	0.183	0.010	0.019
cpr	0.055	0.002	0.007
stc	0.082	0.003	0.008
exp	0.052	0.003	0.007
pow	0.081	0.005	0.009
jump	–	–	–
$T = 100$			
intexp	0.185	0.014	0.014
cpr	0.054	0.002	0.002
stc	0.056	0.002	0.003
exp	0.055	0.003	0.004
pow	0.071	0.004	0.006
jump	–	–	–
$T = 200$			
intexp	0.195	0.017	0.017
cpr	0.055	0.003	0.005
stc	0.060	0.002	0.003
exp	0.061	0.004	0.006
pow	0.072	0.005	0.005
jump	–	–	–
$T = 500$			
intexp	0.222	0.024	0.021
cpr	0.056	0.007	0.007
stc	0.061	0.008	0.007
exp	0.066	0.010	0.009
pow	0.076	0.009	0.010
jump	–	–	–
$T = 1000$			
intexp	0.225	0.038	0.034
cpr	0.052	0.011	0.011
stc	0.056	0.010	0.010
exp	0.059	0.013	0.013
pow	0.069	0.014	0.014
jump	–	–	–

Table 42: Null rejection probabilities for the T_{CS} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

	C_{LL}	$C_{LL}^{b,max}$	$C_{NLLS}^{b,max}$
$T = 50$			
intexp	0.293	0.077	0.176
cpr	0.179	0.028	0.140
stc	0.239	0.033	0.132
exp	0.194	0.038	0.134
pow	0.290	0.057	0.164
jump	–	–	–
$T = 100$			
intexp	0.310	0.087	0.141
cpr	0.205	0.042	0.091
stc	0.211	0.040	0.082
exp	0.223	0.051	0.092
pow	0.270	0.064	0.113
jump	–	–	–
$T = 200$			
intexp	0.279	0.074	0.118
cpr	0.194	0.038	0.067
stc	0.198	0.032	0.058
exp	0.211	0.047	0.074
pow	0.241	0.052	0.094
jump	–	–	–
$T = 500$			
intexp	0.253	0.072	0.102
cpr	0.169	0.041	0.061
stc	0.177	0.038	0.057
exp	0.195	0.047	0.063
pow	0.201	0.055	0.080
jump	–	–	–
$T = 1000$			
intexp	0.230	0.081	0.101
cpr	0.146	0.050	0.069
stc	0.147	0.047	0.063
exp	0.160	0.055	0.072
pow	0.176	0.059	0.087
jump	–	–	–

Table 43: Null rejection probabilities for the T_{CS} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

CVs	Asymptotic			Block bootstrap			Wild bootstrap		
κ	1/5	1/4	1/3	1/5	1/4	1/3	1/5	1/4	1/3
$T = 50$									
intexp	0.999	1.000	1.000	0.029	0.021	0.026	0.030	0.020	0.019
cpr	0.179	0.176	0.163	0.061	0.059	0.060	0.048	0.046	0.041
stc	0.093	0.100	0.103	0.061	0.064	0.063	0.055	0.056	0.050
exp	0.189	0.187	0.172	0.056	0.057	0.053	0.048	0.044	0.040
pow	0.168	0.165	0.160	0.061	0.062	0.061	0.049	0.048	0.044
jump	0.201	0.200	0.184	0.057	0.059	0.052	0.046	0.041	0.037
$T = 100$									
intexp	1.000	1.000	1.000	0.013	0.025	0.028	0.019	0.026	0.020
cpr	0.254	0.256	0.251	0.056	0.058	0.058	0.044	0.046	0.045
stc	0.160	0.168	0.181	0.059	0.059	0.060	0.050	0.053	0.051
exp	0.262	0.262	0.257	0.052	0.054	0.054	0.044	0.046	0.043
pow	0.237	0.246	0.249	0.052	0.057	0.056	0.045	0.047	0.044
jump	0.279	0.279	0.273	0.049	0.051	0.053	0.041	0.043	0.044
$T = 200$									
intexp	1.000	1.000	1.000	0.017	0.021	0.031	0.023	0.025	0.030
cpr	0.305	0.311	0.323	0.055	0.054	0.052	0.048	0.047	0.041
stc	0.189	0.215	0.251	0.056	0.054	0.053	0.053	0.049	0.046
exp	0.312	0.312	0.322	0.055	0.055	0.052	0.047	0.044	0.043
pow	0.279	0.289	0.304	0.053	0.053	0.054	0.048	0.047	0.043
jump	0.314	0.322	0.333	0.053	0.055	0.052	0.047	0.046	0.043
$T = 500$									
intexp	1.000	1.000	1.000	0.026	0.026	0.036	0.029	0.030	0.035
cpr	0.365	0.384	0.407	0.050	0.050	0.047	0.046	0.044	0.041
stc	0.264	0.301	0.344	0.052	0.051	0.047	0.048	0.047	0.044
exp	0.368	0.384	0.405	0.049	0.050	0.047	0.046	0.045	0.038
pow	0.340	0.362	0.395	0.055	0.050	0.050	0.048	0.045	0.043
jump	0.371	0.387	0.417	0.050	0.049	0.050	0.046	0.045	0.038
$T = 1000$									
intexp	1.000	1.000	1.000	0.022	0.034	0.044	0.025	0.032	0.038
cpr	0.392	0.410	0.438	0.057	0.053	0.052	0.051	0.049	0.046
stc	0.306	0.346	0.392	0.058	0.053	0.053	0.054	0.049	0.049
exp	0.396	0.413	0.438	0.054	0.050	0.054	0.051	0.048	0.048
pow	0.372	0.399	0.431	0.057	0.054	0.051	0.052	0.048	0.045
jump	0.400	0.419	0.445	0.058	0.054	0.053	0.053	0.047	0.046

Table 44: Null rejection probabilities for the T_{DG} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

CVs	Asymptotic			Block bootstrap			Wild bootstrap		
κ	1/5	1/4	1/3	1/5	1/4	1/3	1/5	1/4	1/3
$T = 50$									
intexp	0.998	0.999	1.000	0.119	0.048	0.053	0.273	0.129	0.101
cpr	0.312	0.307	0.301	0.099	0.097	0.096	0.135	0.130	0.122
stc	0.162	0.167	0.171	0.081	0.082	0.090	0.093	0.092	0.087
exp	0.379	0.362	0.343	0.103	0.097	0.092	0.170	0.157	0.143
pow	0.353	0.337	0.344	0.119	0.114	0.121	0.184	0.171	0.166
jump	0.421	0.399	0.386	0.112	0.110	0.105	0.191	0.170	0.163
$T = 100$									
intexp	1.000	1.000	1.000	0.036	0.079	0.052	0.159	0.242	0.132
cpr	0.465	0.467	0.447	0.090	0.091	0.093	0.176	0.170	0.161
stc	0.277	0.280	0.285	0.075	0.075	0.081	0.119	0.111	0.110
exp	0.502	0.497	0.479	0.090	0.090	0.088	0.211	0.203	0.191
pow	0.428	0.448	0.448	0.090	0.097	0.099	0.184	0.186	0.179
jump	0.522	0.519	0.512	0.092	0.094	0.091	0.220	0.214	0.198
$T = 200$									
intexp	1.000	1.000	1.000	0.031	0.038	0.049	0.181	0.183	0.187
cpr	0.535	0.543	0.535	0.074	0.078	0.072	0.214	0.214	0.200
stc	0.343	0.356	0.389	0.066	0.066	0.070	0.142	0.145	0.133
exp	0.565	0.566	0.549	0.076	0.080	0.073	0.239	0.242	0.221
pow	0.480	0.506	0.526	0.071	0.079	0.078	0.199	0.202	0.204
jump	0.579	0.585	0.574	0.079	0.078	0.074	0.248	0.247	0.228
$T = 500$									
intexp	1.000	1.000	1.000	0.078	0.056	0.071	0.303	0.236	0.246
cpr	0.627	0.627	0.631	0.065	0.067	0.072	0.264	0.257	0.252
stc	0.442	0.468	0.519	0.059	0.059	0.066	0.176	0.172	0.180
exp	0.627	0.633	0.635	0.070	0.068	0.069	0.274	0.263	0.262
pow	0.570	0.581	0.603	0.068	0.067	0.073	0.243	0.239	0.244
jump	0.636	0.640	0.644	0.071	0.070	0.074	0.281	0.272	0.276
$T = 1000$									
intexp	1.000	1.000	1.000	0.038	0.052	0.063	0.229	0.264	0.275
cpr	0.645	0.653	0.647	0.067	0.065	0.061	0.272	0.281	0.276
stc	0.506	0.530	0.557	0.055	0.059	0.059	0.194	0.198	0.211
exp	0.652	0.646	0.645	0.062	0.066	0.063	0.289	0.291	0.283
pow	0.613	0.624	0.619	0.062	0.064	0.062	0.252	0.258	0.260
jump	0.657	0.664	0.655	0.064	0.064	0.061	0.292	0.294	0.293

Table 45: Null rejection probabilities for the T_{DG} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$															
intexp	bartlett	0.071	0.077	0.085	0.086	0.073	0.084	0.088	0.090	0.018	0.007	0.004	0.002	0.103	0.106
	epa	0.072	0.081	0.085	0.088	0.077	0.085	0.089	0.094	0.013	0.004	0.002	0.003	0.100	0.104
	gaussian	0.085	0.096	0.099	0.101	0.090	0.098	0.101	0.101	0.003	0.002	0.002	0.002	0.100	0.103
cpr	bartlett	0.047	0.043	0.041	0.038	0.047	0.042	0.037	0.035	0.000	0.000	0.000	0.000	0.035	0.030
	epa	0.048	0.042	0.039	0.037	0.047	0.040	0.038	0.033	0.000	0.000	0.000	0.000	0.033	0.031
	gaussian	0.039	0.034	0.032	0.028	0.035	0.030	0.027	0.024	0.000	0.000	0.000	0.000	0.035	0.031
stc	bartlett	0.048	0.048	0.046	0.043	0.048	0.046	0.044	0.042	0.000	0.000	0.000	0.000	0.037	0.035
	epa	0.049	0.048	0.045	0.043	0.047	0.046	0.043	0.042	0.000	0.000	0.000	0.000	0.033	0.038
	gaussian	0.044	0.042	0.038	0.037	0.042	0.039	0.036	0.036	0.000	0.000	0.000	0.000	0.038	0.036
exp	bartlett	0.048	0.043	0.042	0.039	0.047	0.042	0.039	0.035	0.001	0.000	0.000	0.000	0.027	0.025
	epa	0.046	0.043	0.038	0.035	0.045	0.039	0.035	0.031	0.001	0.000	0.000	0.000	0.025	0.023
	gaussian	0.041	0.033	0.030	0.029	0.036	0.031	0.026	0.023	0.000	0.000	0.000	0.000	0.027	0.024
pow	bartlett	0.049	0.045	0.043	0.042	0.048	0.044	0.040	0.036	0.001	0.000	0.000	0.000	0.038	0.036
	epa	0.049	0.043	0.042	0.040	0.047	0.042	0.037	0.036	0.001	0.000	0.000	0.000	0.037	0.035
	gaussian	0.041	0.036	0.033	0.029	0.036	0.033	0.028	0.026	0.000	0.000	0.000	0.000	0.038	0.034
jump	bartlett	0.048	0.045	0.043	0.039	0.047	0.043	0.037	0.036	0.002	0.001	0.000	0.000	0.034	0.030
	epa	0.048	0.043	0.039	0.037	0.046	0.040	0.035	0.033	0.001	0.000	0.000	0.000	0.031	0.030
	gaussian	0.040	0.035	0.033	0.028	0.036	0.032	0.028	0.025	0.000	0.000	0.000	0.000	0.033	0.027

Table 46: Null rejection probabilities for the T_{WP12} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$															
intexp	bartlett	0.216	0.274	0.313	0.343	0.246	0.302	0.351	0.377	0.281	0.186	0.144	0.116	0.489	0.505
	epa	0.228	0.288	0.327	0.361	0.258	0.320	0.365	0.397	0.231	0.147	0.105	0.080	0.495	0.511
	gaussian	0.338	0.412	0.456	0.485	0.372	0.445	0.490	0.515	0.119	0.064	0.039	0.029	0.490	0.521
cpr	bartlett	0.120	0.151	0.185	0.203	0.133	0.179	0.212	0.227	0.052	0.010	0.002	0.001	0.249	0.266
	epa	0.124	0.162	0.193	0.214	0.141	0.188	0.218	0.236	0.033	0.005	0.001	0.000	0.251	0.273
	gaussian	0.199	0.249	0.274	0.293	0.222	0.268	0.294	0.307	0.001	0.000	0.000	0.000	0.247	0.283
stc	bartlett	0.090	0.110	0.128	0.137	0.101	0.125	0.140	0.151	0.014	0.002	0.000	0.000	0.190	0.197
	epa	0.093	0.118	0.133	0.144	0.106	0.128	0.146	0.158	0.006	0.001	0.000	0.000	0.194	0.201
	gaussian	0.136	0.163	0.181	0.193	0.147	0.178	0.195	0.203	0.000	0.000	0.000	0.000	0.190	0.203
exp	bartlett	0.145	0.193	0.226	0.249	0.165	0.219	0.255	0.280	0.143	0.054	0.023	0.013	0.371	0.386
	epa	0.156	0.208	0.239	0.264	0.178	0.233	0.272	0.292	0.103	0.031	0.011	0.004	0.376	0.394
	gaussian	0.246	0.310	0.344	0.366	0.274	0.338	0.370	0.391	0.016	0.001	0.000	0.000	0.369	0.397
pow	bartlett	0.132	0.166	0.194	0.211	0.147	0.189	0.218	0.242	0.148	0.102	0.076	0.061	0.250	0.275
	epa	0.141	0.174	0.205	0.228	0.159	0.200	0.234	0.253	0.128	0.080	0.058	0.043	0.253	0.280
	gaussian	0.212	0.264	0.299	0.319	0.237	0.291	0.321	0.335	0.065	0.029	0.015	0.009	0.251	0.296
jump	bartlett	0.156	0.207	0.244	0.267	0.178	0.233	0.271	0.295	0.174	0.088	0.051	0.032	0.371	0.392
	epa	0.164	0.217	0.256	0.281	0.188	0.248	0.288	0.311	0.137	0.056	0.029	0.014	0.376	0.401
	gaussian	0.265	0.326	0.367	0.388	0.293	0.359	0.392	0.415	0.036	0.008	0.002	0.001	0.371	0.408

Table 47: Null rejection probabilities for the T_{WP12} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$	$\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.142	0.165	0.178	0.185	0.153	0.176	0.188	0.195	0.051	0.023	0.018	0.013	0.221	0.222
	epa	0.147	0.171	0.182	0.189	0.157	0.181	0.192	0.199	0.036	0.017	0.014	0.013	0.221	0.223
	gaussian	0.184	0.203	0.212	0.218	0.193	0.214	0.218	0.219	0.014	0.014	0.013	0.011	0.221	0.223
cpr	bartlett	0.086	0.091	0.090	0.089	0.087	0.091	0.090	0.091	0.010	0.001	0.000	0.000	0.093	0.094
	epa	0.087	0.092	0.090	0.091	0.089	0.091	0.090	0.091	0.008	0.001	0.000	0.000	0.093	0.093
	gaussian	0.088	0.091	0.092	0.088	0.089	0.091	0.090	0.085	0.000	0.000	0.000	0.000	0.092	0.092
stc	bartlett	0.072	0.072	0.070	0.065	0.072	0.069	0.066	0.061	0.001	0.000	0.000	0.000	0.052	0.046
	epa	0.072	0.071	0.068	0.064	0.071	0.068	0.062	0.059	0.000	0.000	0.000	0.000	0.052	0.045
	gaussian	0.066	0.055	0.048	0.042	0.062	0.050	0.040	0.034	0.000	0.000	0.000	0.000	0.050	0.041
exp	bartlett	0.085	0.090	0.093	0.093	0.088	0.091	0.094	0.096	0.028	0.011	0.003	0.001	0.104	0.103
	epa	0.086	0.092	0.095	0.097	0.089	0.093	0.097	0.098	0.023	0.007	0.002	0.000	0.104	0.101
	gaussian	0.092	0.098	0.097	0.098	0.095	0.096	0.099	0.094	0.002	0.000	0.000	0.000	0.104	0.098
pow	bartlett	0.078	0.082	0.080	0.079	0.081	0.079	0.078	0.077	0.001	0.000	0.000	0.000	0.080	0.077
	epa	0.079	0.080	0.079	0.079	0.082	0.079	0.078	0.076	0.000	0.000	0.000	0.000	0.081	0.077
	gaussian	0.079	0.074	0.067	0.063	0.076	0.070	0.062	0.055	0.000	0.000	0.000	0.000	0.079	0.075
jump	bartlett	0.087	0.090	0.091	0.092	0.087	0.092	0.093	0.093	0.023	0.008	0.003	0.001	0.099	0.100
	epa	0.087	0.091	0.094	0.094	0.088	0.093	0.093	0.094	0.018	0.005	0.002	0.001	0.100	0.099
	gaussian	0.092	0.095	0.095	0.095	0.093	0.097	0.093	0.092	0.002	0.000	0.000	0.000	0.098	0.097
$T = 200$	$\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.081	0.094	0.098	0.104	0.087	0.095	0.105	0.108	0.052	0.029	0.022	0.019	0.120	0.120
	epa	0.084	0.094	0.103	0.106	0.091	0.101	0.108	0.110	0.043	0.024	0.019	0.019	0.119	0.121
	gaussian	0.103	0.111	0.116	0.117	0.109	0.116	0.117	0.119	0.020	0.019	0.015	0.013	0.119	0.120
cpr	bartlett	0.070	0.073	0.075	0.076	0.070	0.075	0.075	0.074	0.010	0.002	0.000	0.000	0.075	0.074
	epa	0.072	0.074	0.075	0.075	0.071	0.075	0.075	0.072	0.007	0.001	0.000	0.000	0.076	0.074
	gaussian	0.075	0.073	0.073	0.070	0.074	0.073	0.070	0.065	0.000	0.000	0.000	0.000	0.075	0.072
stc	bartlett	0.062	0.063	0.064	0.060	0.060	0.063	0.059	0.055	0.001	0.000	0.000	0.000	0.047	0.044
	epa	0.061	0.063	0.062	0.058	0.060	0.061	0.056	0.053	0.001	0.000	0.000	0.000	0.047	0.042
	gaussian	0.059	0.051	0.046	0.041	0.055	0.046	0.040	0.034	0.000	0.000	0.000	0.000	0.046	0.039
exp	bartlett	0.073	0.077	0.078	0.080	0.074	0.079	0.080	0.079	0.032	0.008	0.002	0.001	0.086	0.083
	epa	0.073	0.078	0.080	0.081	0.075	0.079	0.079	0.080	0.023	0.005	0.001	0.000	0.085	0.082
	gaussian	0.080	0.081	0.084	0.082	0.079	0.085	0.081	0.081	0.001	0.000	0.000	0.000	0.086	0.082
pow	bartlett	0.068	0.070	0.071	0.072	0.066	0.072	0.070	0.071	0.009	0.001	0.000	0.000	0.071	0.069
	epa	0.068	0.072	0.072	0.070	0.068	0.073	0.071	0.069	0.005	0.001	0.000	0.000	0.071	0.069
	gaussian	0.071	0.067	0.065	0.061	0.071	0.066	0.061	0.057	0.000	0.000	0.000	0.000	0.070	0.069
jump	bartlett	0.075	0.077	0.079	0.081	0.076	0.080	0.081	0.082	0.025	0.006	0.002	0.001	0.084	0.085
	epa	0.075	0.078	0.081	0.082	0.076	0.079	0.081	0.083	0.019	0.004	0.001	0.001	0.084	0.084
	gaussian	0.081	0.083	0.083	0.084	0.082	0.083	0.085	0.082	0.001	0.000	0.000	0.000	0.084	0.082

Table 48: Null rejection probabilities for the T_{WP16} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$	$\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.197	0.242	0.273	0.291	0.218	0.266	0.296	0.311	0.159	0.092	0.061	0.042	0.371	0.377
	epa	0.208	0.257	0.287	0.301	0.229	0.280	0.306	0.322	0.131	0.067	0.041	0.029	0.374	0.379
	gaussian	0.289	0.331	0.355	0.365	0.308	0.350	0.367	0.370	0.049	0.024	0.018	0.014	0.370	0.376
cpr	bartlett	0.130	0.154	0.172	0.181	0.139	0.165	0.186	0.196	0.081	0.020	0.005	0.001	0.213	0.229
	epa	0.135	0.158	0.179	0.189	0.145	0.173	0.192	0.203	0.057	0.011	0.002	0.001	0.217	0.233
	gaussian	0.180	0.213	0.235	0.249	0.193	0.230	0.250	0.263	0.002	0.000	0.000	0.000	0.212	0.241
stc	bartlett	0.094	0.102	0.111	0.115	0.098	0.110	0.115	0.122	0.011	0.001	0.000	0.000	0.135	0.136
	epa	0.094	0.107	0.112	0.116	0.099	0.114	0.118	0.124	0.006	0.000	0.000	0.000	0.137	0.137
	gaussian	0.115	0.123	0.129	0.131	0.120	0.128	0.131	0.129	0.000	0.000	0.000	0.000	0.134	0.134
exp	bartlett	0.140	0.167	0.186	0.203	0.153	0.181	0.206	0.222	0.202	0.108	0.056	0.031	0.286	0.296
	epa	0.147	0.174	0.196	0.212	0.161	0.192	0.216	0.230	0.178	0.075	0.036	0.016	0.289	0.299
	gaussian	0.200	0.240	0.266	0.283	0.220	0.261	0.285	0.300	0.040	0.007	0.001	0.000	0.285	0.304
pow	bartlett	0.104	0.122	0.133	0.139	0.113	0.131	0.139	0.152	0.024	0.009	0.004	0.002	0.152	0.162
	epa	0.109	0.125	0.137	0.147	0.117	0.135	0.147	0.155	0.018	0.007	0.003	0.002	0.155	0.164
	gaussian	0.138	0.160	0.169	0.175	0.149	0.167	0.177	0.180	0.003	0.001	0.000	0.000	0.154	0.169
jump	bartlett	0.140	0.172	0.195	0.209	0.155	0.188	0.212	0.227	0.191	0.100	0.058	0.035	0.279	0.296
	epa	0.146	0.182	0.202	0.220	0.161	0.198	0.222	0.236	0.165	0.075	0.035	0.019	0.284	0.301
	gaussian	0.206	0.247	0.279	0.296	0.224	0.271	0.297	0.309	0.042	0.009	0.002	0.001	0.279	0.307
$T = 200$	$\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.133	0.162	0.186	0.203	0.144	0.181	0.207	0.223	0.193	0.114	0.075	0.055	0.285	0.294
	epa	0.137	0.171	0.194	0.213	0.150	0.190	0.218	0.232	0.169	0.087	0.053	0.040	0.288	0.296
	gaussian	0.199	0.240	0.266	0.280	0.219	0.259	0.280	0.295	0.063	0.031	0.020	0.017	0.284	0.301
cpr	bartlett	0.108	0.125	0.142	0.149	0.116	0.137	0.152	0.165	0.068	0.016	0.004	0.002	0.176	0.192
	epa	0.112	0.131	0.146	0.157	0.121	0.143	0.160	0.173	0.050	0.007	0.002	0.000	0.179	0.195
	gaussian	0.147	0.176	0.197	0.209	0.162	0.191	0.206	0.215	0.002	0.000	0.000	0.000	0.175	0.203
stc	bartlett	0.085	0.092	0.099	0.104	0.087	0.096	0.103	0.109	0.013	0.002	0.000	0.000	0.126	0.128
	epa	0.087	0.096	0.100	0.106	0.089	0.100	0.105	0.113	0.008	0.000	0.000	0.000	0.127	0.128
	gaussian	0.101	0.117	0.121	0.123	0.106	0.120	0.124	0.124	0.000	0.000	0.000	0.000	0.124	0.125
exp	bartlett	0.118	0.144	0.161	0.176	0.130	0.159	0.180	0.191	0.207	0.107	0.055	0.031	0.253	0.263
	epa	0.123	0.151	0.169	0.184	0.136	0.167	0.186	0.204	0.182	0.075	0.033	0.015	0.256	0.269
	gaussian	0.173	0.213	0.237	0.250	0.190	0.233	0.253	0.271	0.040	0.007	0.001	0.000	0.251	0.273
pow	bartlett	0.100	0.119	0.134	0.144	0.109	0.130	0.148	0.159	0.099	0.045	0.020	0.013	0.166	0.178
	epa	0.105	0.126	0.138	0.151	0.115	0.138	0.153	0.165	0.081	0.031	0.014	0.007	0.168	0.181
	gaussian	0.142	0.171	0.192	0.203	0.156	0.188	0.204	0.213	0.016	0.002	0.001	0.000	0.165	0.189
jump	bartlett	0.118	0.143	0.165	0.183	0.128	0.161	0.183	0.198	0.204	0.106	0.054	0.032	0.246	0.262
	epa	0.123	0.153	0.176	0.188	0.136	0.170	0.192	0.207	0.179	0.076	0.033	0.018	0.249	0.266
	gaussian	0.181	0.218	0.241	0.262	0.196	0.237	0.263	0.277	0.041	0.007	0.001	0.000	0.246	0.272

Table 49: Null rejection probabilities for the T_{WP16} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

M	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$
$T = 50$						
intexp	0.052	0.051	0.059	0.066	0.056	0.059
cpr	0.071	0.069	0.074	0.080	0.079	0.069
stc	0.072	0.071	0.075	0.081	0.084	0.074
exp	0.059	0.059	0.071	0.075	0.070	0.067
pow	0.061	0.060	0.071	0.075	0.069	0.066
jump	0.058	0.059	0.069	0.073	0.069	0.065
$T = 100$						
intexp	0.049	0.046	0.054	0.056	0.049	0.054
cpr	0.060	0.060	0.064	0.071	0.060	0.064
stc	0.063	0.062	0.069	0.070	0.063	0.069
exp	0.054	0.051	0.059	0.065	0.054	0.059
pow	0.058	0.055	0.060	0.066	0.058	0.060
jump	0.053	0.052	0.059	0.064	0.053	0.059
$T = 200$						
intexp	0.049	0.049	0.052	0.056	0.049	0.051
cpr	0.058	0.057	0.060	0.064	0.058	0.059
stc	0.060	0.057	0.063	0.068	0.060	0.060
exp	0.053	0.052	0.055	0.061	0.053	0.055
pow	0.057	0.053	0.058	0.063	0.057	0.057
jump	0.051	0.052	0.055	0.060	0.051	0.056
$T = 500$						
intexp	0.053	0.047	0.049	0.052	0.047	0.051
cpr	0.052	0.048	0.051	0.053	0.050	0.053
stc	0.054	0.049	0.052	0.055	0.052	0.055
exp	0.051	0.046	0.050	0.053	0.047	0.052
pow	0.053	0.048	0.052	0.053	0.050	0.054
jump	0.052	0.046	0.051	0.052	0.049	0.052
$T = 1000$						
intexp	0.051	0.052	0.050	0.053	0.051	0.053
cpr	0.053	0.051	0.053	0.052	0.051	0.054
stc	0.053	0.052	0.054	0.054	0.053	0.055
exp	0.052	0.050	0.050	0.051	0.051	0.052
pow	0.053	0.052	0.052	0.051	0.051	0.053
jump	0.053	0.050	0.051	0.051	0.051	0.053

Table 50: Null rejection probabilities for the T_{wz} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

M	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$
$T = 50$						
intexp	0.050	0.053	0.065	0.074	0.054	0.061
cpr	0.059	0.059	0.068	0.076	0.059	0.065
stc	0.061	0.055	0.069	0.077	0.058	0.067
exp	0.056	0.056	0.067	0.077	0.053	0.064
pow	0.056	0.055	0.066	0.077	0.053	0.062
jump	0.055	0.057	0.069	0.078	0.055	0.064
$T = 100$						
intexp	0.050	0.049	0.057	0.064	0.050	0.057
cpr	0.052	0.056	0.060	0.066	0.052	0.060
stc	0.055	0.054	0.057	0.067	0.055	0.057
exp	0.049	0.054	0.058	0.066	0.049	0.058
pow	0.053	0.052	0.061	0.067	0.053	0.061
jump	0.048	0.052	0.057	0.066	0.048	0.057
$T = 200$						
intexp	0.047	0.055	0.055	0.055	0.047	0.056
cpr	0.050	0.058	0.059	0.060	0.050	0.059
stc	0.049	0.058	0.057	0.060	0.049	0.058
exp	0.049	0.056	0.058	0.059	0.049	0.058
pow	0.050	0.060	0.056	0.060	0.050	0.060
jump	0.048	0.056	0.056	0.060	0.048	0.058
$T = 500$						
intexp	0.052	0.048	0.048	0.049	0.046	0.051
cpr	0.048	0.050	0.049	0.050	0.050	0.052
stc	0.048	0.049	0.050	0.052	0.048	0.052
exp	0.047	0.048	0.048	0.048	0.049	0.051
pow	0.047	0.050	0.049	0.049	0.049	0.051
jump	0.048	0.049	0.046	0.048	0.049	0.050
$T = 1000$						
intexp	0.047	0.050	0.045	0.047	0.050	0.051
cpr	0.048	0.050	0.045	0.049	0.048	0.052
stc	0.049	0.049	0.044	0.050	0.048	0.052
exp	0.049	0.049	0.045	0.049	0.048	0.051
pow	0.048	0.049	0.046	0.049	0.048	0.050
jump	0.048	0.049	0.044	0.049	0.048	0.049

Table 51: Null rejection probabilities for the T_{WZ} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.0$, and $\rho_2 = 0.6$.

M	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$
$T = 50$						
intexp	0.193	0.181	0.158	0.159	0.205	0.162
cpr	0.239	0.227	0.194	0.196	0.260	0.201
stc	0.224	0.209	0.182	0.187	0.250	0.186
exp	0.226	0.217	0.187	0.189	0.253	0.195
pow	0.213	0.198	0.171	0.182	0.235	0.175
jump	0.224	0.212	0.182	0.186	0.250	0.191
$T = 100$						
intexp	0.354	0.305	0.240	0.222	0.354	0.240
cpr	0.407	0.364	0.282	0.260	0.407	0.282
stc	0.395	0.355	0.272	0.252	0.395	0.272
exp	0.388	0.342	0.268	0.250	0.388	0.268
pow	0.393	0.353	0.268	0.254	0.393	0.268
jump	0.391	0.342	0.269	0.245	0.391	0.269
$T = 200$						
intexp	0.657	0.580	0.449	0.384	0.657	0.425
cpr	0.696	0.624	0.496	0.422	0.696	0.466
stc	0.696	0.619	0.486	0.422	0.696	0.462
exp	0.688	0.611	0.482	0.409	0.688	0.455
pow	0.690	0.621	0.488	0.414	0.690	0.464
jump	0.689	0.612	0.482	0.411	0.689	0.458
$T = 500$						
intexp	0.982	0.961	0.894	0.826	0.970	0.835
cpr	0.986	0.970	0.914	0.847	0.977	0.859
stc	0.986	0.970	0.913	0.851	0.977	0.857
exp	0.985	0.968	0.909	0.844	0.976	0.852
pow	0.986	0.968	0.913	0.847	0.977	0.858
jump	0.985	0.968	0.909	0.843	0.976	0.852
$T = 1000$						
intexp	1.000	1.000	0.999	0.996	1.000	0.994
cpr	1.000	1.000	0.999	0.998	1.000	0.996
stc	1.000	1.000	0.999	0.998	1.000	0.995
exp	1.000	1.000	0.999	0.998	1.000	0.995
pow	1.000	1.000	0.999	0.998	1.000	0.996
jump	1.000	1.000	0.999	0.998	1.000	0.996

Table 52: Null rejection probabilities for the T_{WZ} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}
$T = 50$						
intexp	0.010	0.029	0.044	0.036	0.001	0.052
cpr	0.002	0.061	0.036	0.007	0.000	0.071
stc	0.003	0.061	0.042	0.002	0.000	0.072
exp	0.003	0.056	0.032	0.018	0.001	0.059
pow	0.005	0.061	0.035	0.004	–	0.061
jump	–	0.057	0.032	0.018	–	0.058
$T = 100$						
intexp	0.014	0.013	0.057	0.048	0.001	0.049
cpr	0.002	0.056	0.044	0.011	0.000	0.060
stc	0.002	0.059	0.045	0.001	0.000	0.063
exp	0.003	0.052	0.042	0.031	0.004	0.054
pow	0.004	0.052	0.041	0.003	–	0.058
jump	–	0.049	0.044	0.025	–	0.053
$T = 200$						
intexp	0.017	0.017	0.071	0.051	0.002	0.049
cpr	0.003	0.055	0.047	0.010	0.000	0.058
stc	0.002	0.056	0.048	0.001	0.000	0.060
exp	0.004	0.055	0.048	0.028	0.002	0.053
pow	0.005	0.053	0.049	0.001	–	0.057
jump	–	0.053	0.048	0.023	–	0.051
$T = 500$						
intexp	0.024	0.026	0.081	0.054	0.006	0.053
cpr	0.007	0.050	0.052	0.009	0.000	0.052
stc	0.008	0.052	0.052	0.000	0.000	0.054
exp	0.010	0.049	0.050	0.028	0.006	0.051
pow	0.009	0.055	0.051	0.001	–	0.053
jump	–	0.050	0.049	0.019	–	0.052
$T = 1000$						
intexp	0.038	0.022	0.099	0.066	0.006	0.051
cpr	0.011	0.057	0.051	0.008	0.000	0.053
stc	0.010	0.058	0.050	0.001	0.000	0.053
exp	0.013	0.054	0.053	0.030	0.005	0.052
pow	0.014	0.057	0.051	0.000	–	0.053
jump	–	0.058	0.053	0.023	–	0.053

Table 53: Null rejection probabilities for the specification tests with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}
$T = 50$						
intexp	0.020	0.041	0.062	0.065	0.003	0.132
cpr	0.003	0.075	0.041	0.015	0.000	0.176
stc	0.007	0.079	0.039	0.001	0.000	0.177
exp	0.005	0.071	0.043	0.045	0.007	0.156
pow	0.008	0.074	0.046	0.008	–	0.155
jump	–	0.071	0.043	0.045	–	0.154
$T = 100$						
intexp	0.021	0.018	0.087	0.081	0.006	0.210
cpr	0.006	0.071	0.049	0.025	0.000	0.243
stc	0.005	0.069	0.047	0.002	0.001	0.246
exp	0.007	0.069	0.056	0.076	0.017	0.224
pow	0.009	0.064	0.053	0.009	–	0.235
jump	–	0.065	0.060	0.063	–	0.226
$T = 200$						
intexp	0.023	0.017	0.111	0.090	0.008	0.377
cpr	0.007	0.061	0.057	0.025	0.000	0.411
stc	0.005	0.061	0.053	0.004	0.000	0.422
exp	0.008	0.060	0.063	0.076	0.027	0.394
pow	0.010	0.063	0.058	0.004	–	0.410
jump	–	0.060	0.064	0.060	–	0.395
$T = 500$						
intexp	0.028	0.033	0.139	0.088	0.018	0.802
cpr	0.011	0.058	0.067	0.024	0.001	0.824
stc	0.010	0.054	0.061	0.003	0.001	0.825
exp	0.014	0.057	0.070	0.067	0.039	0.816
pow	0.014	0.056	0.068	0.002	–	0.823
jump	–	0.057	0.072	0.056	–	0.815
$T = 1000$						
intexp	0.041	0.022	0.175	0.102	0.023	0.985
cpr	0.015	0.062	0.076	0.021	0.002	0.988
stc	0.013	0.061	0.069	0.003	0.003	0.988
exp	0.017	0.059	0.081	0.069	0.049	0.987
pow	0.019	0.060	0.076	0.001	–	0.987
jump	–	0.060	0.080	0.054	–	0.986

Table 54: Null rejection probabilities for the specification tests with $\rho_0 = 0.0$, $\rho_1 = 0.4$, and $\rho_2 = 0.0$.

	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}
$T = 50$						
intexp	0.077	0.119	0.125	0.125	0.029	0.193
cpr	0.028	0.099	0.053	0.029	0.000	0.239
stc	0.033	0.081	0.043	0.005	0.001	0.224
exp	0.038	0.103	0.078	0.119	0.053	0.226
pow	0.057	0.119	0.074	0.026	–	0.213
jump	–	0.112	0.085	0.138	–	0.224
$T = 100$						
intexp	0.087	0.036	0.170	0.150	0.052	0.354
cpr	0.042	0.090	0.091	0.071	0.004	0.407
stc	0.040	0.075	0.061	0.008	0.004	0.395
exp	0.051	0.090	0.118	0.175	0.111	0.388
pow	0.064	0.090	0.104	0.032	–	0.393
jump	–	0.092	0.131	0.178	–	0.391
$T = 200$						
intexp	0.074	0.031	0.216	0.159	0.085	0.657
cpr	0.038	0.074	0.120	0.081	0.013	0.696
stc	0.032	0.066	0.090	0.011	0.012	0.696
exp	0.047	0.076	0.145	0.202	0.177	0.688
pow	0.052	0.071	0.132	0.024	–	0.690
jump	–	0.079	0.156	0.191	–	0.689
$T = 500$						
intexp	0.072	0.078	0.296	0.155	0.125	0.982
cpr	0.041	0.065	0.184	0.088	0.036	0.986
stc	0.038	0.059	0.146	0.017	0.032	0.986
exp	0.047	0.070	0.211	0.204	0.275	0.985
pow	0.055	0.068	0.190	0.014	–	0.986
jump	–	0.071	0.221	0.180	–	0.985
$T = 1000$						
intexp	0.081	0.038	0.337	0.136	0.154	1.000
cpr	0.050	0.067	0.211	0.078	0.060	1.000
stc	0.047	0.055	0.177	0.016	0.056	1.000
exp	0.055	0.062	0.232	0.178	0.340	1.000
pow	0.059	0.062	0.218	0.012	–	1.000
jump	–	0.064	0.239	0.160	–	1.000

Table 55: Null rejection probabilities for the specification tests with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

	H_1: DGP 1			H_1: DGP 2			H_1: DGP 3		
	C_{LL}	$C_{LL}^{b,max}$	$C_{NLLS}^{b,max}$	C_{LL}	$C_{LL}^{b,max}$	$C_{NLLS}^{b,max}$	C_{LL}	$C_{LL}^{b,max}$	$C_{NLLS}^{b,max}$
$T = 50$									
intexp	0.366	0.499	0.715	0.004	0.028	0.046	0.600	0.740	0.949
cpr	0.470	0.475	0.726	0.056	0.055	0.060	0.664	0.590	0.853
stc	0.447	0.472	0.711	0.059	0.059	0.052	0.668	0.603	0.809
exp	0.549	0.527	0.718	0.189	0.150	0.160	0.788	0.639	0.830
pow	0.612	0.574	0.696	0.051	0.050	0.050	0.878	0.824	0.886
jump	–	–	–	–	–	–	–	–	–
$T = 100$									
intexp	0.651	0.720	0.827	0.030	0.082	0.102	0.971	0.982	1.000
cpr	0.728	0.696	0.835	0.084	0.089	0.096	0.934	0.870	0.977
stc	0.737	0.690	0.827	0.060	0.057	0.067	0.953	0.879	0.966
exp	0.797	0.716	0.831	0.443	0.315	0.345	0.997	0.870	0.955
pow	0.822	0.767	0.835	0.056	0.055	0.054	0.989	0.975	0.972
jump	–	–	–	–	–	–	–	–	–
$T = 200$									
intexp	0.771	0.855	0.881	0.132	0.229	0.241	0.999	0.998	1.000
cpr	0.884	0.832	0.883	0.191	0.178	0.184	0.993	0.964	0.994
stc	0.879	0.829	0.874	0.075	0.091	0.091	0.995	0.967	0.989
exp	0.917	0.844	0.889	0.679	0.516	0.532	1.000	0.951	0.980
pow	0.913	0.869	0.893	0.061	0.062	0.063	0.999	0.995	0.993
jump	–	–	–	–	–	–	–	–	–
$T = 500$									
intexp	0.901	0.946	0.957	0.351	0.488	0.502	1.000	1.000	1.000
cpr	0.973	0.946	0.957	0.484	0.401	0.409	1.000	0.994	0.999
stc	0.971	0.947	0.949	0.214	0.202	0.208	0.998	0.994	0.964
exp	0.983	0.953	0.964	0.891	0.796	0.793	1.000	0.996	0.999
pow	0.976	0.958	0.964	0.111	0.110	0.105	1.000	0.999	0.994
jump	–	–	–	–	–	–	–	–	–
$T = 1000$									
intexp	0.959	0.979	0.984	0.561	0.658	0.671	1.000	1.000	1.000
cpr	0.994	0.984	0.988	0.725	0.585	0.597	1.000	1.000	1.000
stc	0.993	0.987	0.983	0.422	0.349	0.352	0.998	0.997	0.975
exp	0.997	0.987	0.990	0.971	0.897	0.894	1.000	1.000	1.000
pow	0.994	0.985	0.988	0.219	0.196	0.194	1.000	1.000	0.997
jump	–	–	–	–	–	–	–	–	–

Table 56: Size-corrected power for the T_{CS} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

	H_1: DGP 1			H_1: DGP 2			H_1: DGP 3		
	C_{LL}	$C_{LL}^{b,max}$	$C_{NLLS}^{b,max}$	C_{LL}	$C_{LL}^{b,max}$	$C_{NLLS}^{b,max}$	C_{LL}	$C_{LL}^{b,max}$	$C_{NLLS}^{b,max}$
$T = 50$									
intexp	0.238	0.262	0.339	0.023	0.040	0.049	0.183	0.193	0.279
cpr	0.265	0.240	0.328	0.051	0.050	0.050	0.250	0.195	0.249
stc	0.223	0.239	0.300	0.053	0.060	0.046	0.251	0.201	0.223
exp	0.293	0.257	0.309	0.058	0.058	0.051	0.276	0.210	0.262
pow	0.313	0.266	0.278	0.052	0.051	0.047	0.388	0.250	0.251
jump	–	–	–	–	–	–	–	–	–
$T = 100$									
intexp	0.535	0.532	0.591	0.029	0.043	0.054	0.814	0.777	0.883
cpr	0.503	0.461	0.526	0.048	0.050	0.048	0.675	0.607	0.755
stc	0.492	0.463	0.532	0.054	0.054	0.051	0.693	0.614	0.730
exp	0.554	0.460	0.525	0.101	0.074	0.062	0.778	0.614	0.726
pow	0.589	0.531	0.531	0.053	0.050	0.049	0.866	0.784	0.795
jump	–	–	–	–	–	–	–	–	–
$T = 200$									
intexp	0.720	0.739	0.739	0.044	0.065	0.061	0.995	0.990	0.997
cpr	0.703	0.665	0.684	0.060	0.058	0.063	0.924	0.868	0.914
stc	0.702	0.667	0.687	0.054	0.052	0.048	0.925	0.873	0.914
exp	0.763	0.681	0.685	0.221	0.154	0.129	0.990	0.861	0.889
pow	0.762	0.727	0.710	0.051	0.049	0.047	0.979	0.974	0.965
jump	–	–	–	–	–	–	–	–	–
$T = 500$									
intexp	0.878	0.901	0.887	0.138	0.178	0.160	1.000	1.000	1.000
cpr	0.898	0.864	0.855	0.103	0.114	0.109	0.996	0.975	0.983
stc	0.901	0.865	0.845	0.061	0.068	0.058	0.993	0.974	0.941
exp	0.925	0.871	0.863	0.525	0.398	0.366	1.000	0.980	0.986
pow	0.919	0.889	0.873	0.053	0.050	0.057	0.999	0.998	0.992
jump	–	–	–	–	–	–	–	–	–
$T = 1000$									
intexp	0.955	0.962	0.957	0.321	0.344	0.320	1.000	1.000	1.000
cpr	0.969	0.947	0.943	0.203	0.212	0.192	1.000	0.995	0.996
stc	0.970	0.950	0.936	0.085	0.081	0.081	0.997	0.994	0.975
exp	0.982	0.954	0.948	0.765	0.627	0.578	1.000	0.997	0.998
pow	0.975	0.954	0.942	0.063	0.064	0.063	1.000	0.999	0.995
jump	–	–	–	–	–	–	–	–	–

Table 57: Size-corrected power for the T_{CS} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

	H_1: DGP 1			H_1: DGP 2			H_1: DGP 3		
κ	1/5	1/4	1/3	1/5	1/4	1/3	1/5	1/4	1/3
$T = 50$									
intexp	0.103	0.055	0.066	0.012	0.020	0.053	0.073	0.017	0.054
cpr	0.234	0.229	0.226	0.242	0.220	0.200	0.178	0.173	0.170
stc	0.172	0.169	0.171	0.092	0.089	0.082	0.126	0.122	0.118
exp	0.323	0.317	0.301	0.601	0.552	0.504	0.287	0.279	0.261
pow	0.305	0.297	0.324	0.095	0.095	0.090	0.256	0.259	0.276
jump	0.368	0.362	0.359	0.469	0.435	0.401	0.359	0.346	0.324
$T = 100$									
intexp	0.078	0.092	0.078	0.040	0.078	0.088	0.038	0.091	0.092
cpr	0.340	0.342	0.343	0.465	0.418	0.398	0.384	0.378	0.376
stc	0.259	0.252	0.244	0.132	0.122	0.110	0.258	0.249	0.255
exp	0.470	0.464	0.451	0.882	0.866	0.815	0.546	0.549	0.526
pow	0.336	0.364	0.377	0.169	0.150	0.136	0.361	0.403	0.421
jump	0.494	0.498	0.486	0.736	0.710	0.663	0.568	0.574	0.568
$T = 200$									
intexp	0.098	0.104	0.094	0.052	0.112	0.097	0.048	0.112	0.100
cpr	0.478	0.486	0.483	0.823	0.779	0.698	0.573	0.595	0.584
stc	0.395	0.384	0.353	0.263	0.231	0.175	0.454	0.443	0.432
exp	0.605	0.608	0.598	0.974	0.969	0.959	0.730	0.728	0.734
pow	0.378	0.413	0.438	0.353	0.343	0.269	0.409	0.457	0.498
jump	0.590	0.599	0.592	0.931	0.917	0.877	0.688	0.705	0.710
$T = 500$									
intexp	0.166	0.126	0.091	0.183	0.160	0.108	0.186	0.146	0.105
cpr	0.616	0.635	0.635	0.979	0.975	0.955	0.735	0.754	0.771
stc	0.534	0.519	0.513	0.677	0.579	0.482	0.692	0.685	0.691
exp	0.726	0.744	0.748	0.992	0.993	0.993	0.843	0.854	0.863
pow	0.476	0.491	0.533	0.734	0.684	0.616	0.529	0.542	0.594
jump	0.696	0.710	0.728	0.990	0.990	0.987	0.794	0.812	0.836
$T = 1000$									
intexp	0.153	0.122	0.084	0.219	0.153	0.098	0.175	0.140	0.093
cpr	0.690	0.697	0.700	0.991	0.991	0.987	0.784	0.804	0.818
stc	0.634	0.615	0.589	0.905	0.850	0.753	0.742	0.746	0.744
exp	0.793	0.800	0.798	0.996	0.995	0.994	0.890	0.896	0.899
pow	0.531	0.538	0.553	0.924	0.881	0.817	0.571	0.593	0.613
jump	0.739	0.751	0.760	0.993	0.994	0.994	0.828	0.837	0.854

Table 58: Size-corrected power for the T_{DG} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

	H_1: DGP 1			H_1: DGP 2			H_1: DGP 3		
κ	1/5	1/4	1/3	1/5	1/4	1/3	1/5	1/4	1/3
$T = 50$									
intexp	0.055	0.042	0.055	0.014	0.015	0.039	0.024	0.015	0.038
cpr	0.112	0.121	0.120	0.092	0.091	0.085	0.078	0.082	0.081
stc	0.105	0.108	0.104	0.069	0.069	0.063	0.082	0.084	0.076
exp	0.155	0.150	0.149	0.176	0.165	0.158	0.106	0.107	0.102
pow	0.147	0.145	0.159	0.062	0.064	0.063	0.104	0.105	0.114
jump	0.185	0.179	0.176	0.106	0.097	0.095	0.110	0.106	0.101
$T = 100$									
intexp	0.059	0.072	0.066	0.017	0.045	0.056	0.020	0.061	0.069
cpr	0.173	0.178	0.170	0.141	0.131	0.126	0.159	0.162	0.165
stc	0.154	0.153	0.144	0.069	0.068	0.062	0.133	0.141	0.133
exp	0.254	0.256	0.243	0.342	0.324	0.294	0.280	0.283	0.269
pow	0.167	0.187	0.209	0.071	0.068	0.068	0.163	0.188	0.206
jump	0.258	0.261	0.256	0.200	0.191	0.175	0.286	0.288	0.276
$T = 200$									
intexp	0.077	0.085	0.081	0.024	0.064	0.071	0.030	0.086	0.088
cpr	0.266	0.282	0.288	0.285	0.267	0.225	0.318	0.338	0.342
stc	0.241	0.236	0.234	0.082	0.075	0.073	0.265	0.257	0.274
exp	0.398	0.399	0.408	0.651	0.622	0.578	0.506	0.522	0.530
pow	0.192	0.219	0.255	0.103	0.104	0.091	0.210	0.241	0.293
jump	0.381	0.388	0.387	0.434	0.399	0.350	0.476	0.486	0.488
$T = 500$									
intexp	0.125	0.100	0.082	0.095	0.106	0.086	0.137	0.124	0.097
cpr	0.384	0.400	0.427	0.641	0.596	0.529	0.502	0.530	0.557
stc	0.365	0.363	0.364	0.182	0.157	0.134	0.518	0.533	0.537
exp	0.535	0.561	0.578	0.909	0.909	0.892	0.684	0.710	0.736
pow	0.265	0.276	0.334	0.226	0.206	0.185	0.311	0.338	0.397
jump	0.488	0.512	0.542	0.771	0.754	0.717	0.612	0.642	0.668
$T = 1000$									
intexp	0.141	0.102	0.076	0.152	0.109	0.080	0.152	0.117	0.083
cpr	0.480	0.517	0.520	0.893	0.872	0.812	0.602	0.650	0.663
stc	0.475	0.468	0.441	0.428	0.356	0.258	0.596	0.609	0.616
exp	0.637	0.665	0.679	0.967	0.968	0.964	0.780	0.808	0.820
pow	0.331	0.346	0.381	0.476	0.400	0.329	0.375	0.407	0.449
jump	0.556	0.585	0.613	0.929	0.921	0.902	0.680	0.715	0.741

Table 59: Size-corrected power for the T_{DG} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$															
intexp	bartlett	0.661	0.716	0.740	0.755	0.689	0.737	0.757	0.775	0.851	0.790	0.748	0.721	0.811	0.821
	epa	0.674	0.723	0.746	0.764	0.700	0.745	0.765	0.785	0.808	0.719	0.655	0.604	0.815	0.823
	gaussian	0.753	0.794	0.812	0.821	0.774	0.809	0.824	0.834	0.704	0.593	0.512	0.450	0.813	0.828
cpr	bartlett	0.382	0.468	0.515	0.547	0.421	0.506	0.555	0.583	0.385	0.156	0.070	0.037	0.638	0.660
	epa	0.406	0.489	0.535	0.562	0.442	0.524	0.569	0.597	0.306	0.169	0.138	0.116	0.642	0.666
	gaussian	0.542	0.613	0.653	0.670	0.578	0.646	0.673	0.691	0.041	0.006	0.006	0.007	0.634	0.675
stc	bartlett	0.263	0.327	0.363	0.392	0.293	0.355	0.395	0.420	0.225	0.099	0.057	0.035	0.481	0.493
	epa	0.278	0.340	0.380	0.404	0.310	0.367	0.403	0.434	0.156	0.064	0.041	0.034	0.485	0.497
	gaussian	0.388	0.445	0.475	0.491	0.415	0.472	0.493	0.507	0.032	0.011	0.011	0.012	0.479	0.504
exp	bartlett	0.563	0.640	0.687	0.717	0.598	0.678	0.724	0.746	0.731	0.615	0.515	0.450	0.803	0.815
	epa	0.584	0.661	0.707	0.729	0.621	0.695	0.732	0.756	0.647	0.454	0.320	0.231	0.807	0.820
	gaussian	0.713	0.765	0.793	0.811	0.741	0.791	0.813	0.824	0.422	0.213	0.110	0.061	0.802	0.824
pow	bartlett	0.565	0.630	0.666	0.689	0.593	0.656	0.696	0.721	0.696	0.627	0.590	0.562	0.759	0.781
	epa	0.580	0.646	0.682	0.705	0.612	0.675	0.711	0.733	0.649	0.570	0.508	0.465	0.761	0.785
	gaussian	0.687	0.749	0.775	0.790	0.716	0.771	0.794	0.805	0.545	0.446	0.366	0.314	0.758	0.794
jump	bartlett	0.594	0.671	0.712	0.738	0.637	0.704	0.747	0.771	0.767	0.651	0.569	0.511	0.825	0.834
	epa	0.616	0.689	0.731	0.755	0.655	0.721	0.762	0.784	0.695	0.529	0.414	0.334	0.830	0.840
	gaussian	0.737	0.792	0.821	0.834	0.766	0.815	0.837	0.849	0.488	0.314	0.206	0.143	0.824	0.845

Table 60: Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 1 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$															
intexp	bartlett	0.409	0.485	0.532	0.559	0.447	0.523	0.564	0.593	0.819	0.624	0.450	0.338	0.643	0.662
	epa	0.424	0.503	0.544	0.575	0.464	0.541	0.582	0.611	0.717	0.406	0.225	0.156	0.647	0.666
	gaussian	0.553	0.625	0.650	0.667	0.593	0.647	0.677	0.700	0.327	0.092	0.042	0.031	0.647	0.681
cpr	bartlett	0.537	0.655	0.718	0.762	0.591	0.704	0.771	0.801	0.669	0.317	0.140	0.067	0.819	0.849
	epa	0.567	0.682	0.747	0.783	0.619	0.731	0.787	0.814	0.549	0.341	0.283	0.274	0.821	0.854
	gaussian	0.757	0.831	0.867	0.884	0.797	0.859	0.887	0.902	0.093	0.006	0.002	0.005	0.816	0.869
stc	bartlett	0.071	0.086	0.094	0.103	0.080	0.090	0.103	0.108	0.045	0.022	0.023	0.022	0.122	0.125
	epa	0.076	0.090	0.101	0.106	0.084	0.096	0.105	0.114	0.040	0.044	0.050	0.049	0.125	0.129
	gaussian	0.097	0.115	0.120	0.126	0.107	0.121	0.127	0.132	0.025	0.032	0.037	0.040	0.122	0.129
exp	bartlett	0.983	0.995	0.998	0.998	0.989	0.997	0.999	0.999	1.000	0.999	0.994	0.985	1.000	1.000
	epa	0.986	0.996	0.998	0.999	0.991	0.997	0.999	1.000	0.993	0.905	0.764	0.623	1.000	1.000
	gaussian	0.998	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.946	0.732	0.485	0.295	1.000	1.000
pow	bartlett	0.119	0.148	0.171	0.188	0.133	0.166	0.194	0.214	0.037	0.021	0.020	0.021	0.204	0.232
	epa	0.122	0.162	0.181	0.198	0.141	0.178	0.204	0.220	0.064	0.091	0.085	0.069	0.200	0.230
	gaussian	0.188	0.237	0.255	0.265	0.207	0.251	0.266	0.274	0.029	0.041	0.043	0.046	0.208	0.244
jump	bartlett	0.856	0.906	0.932	0.942	0.880	0.925	0.945	0.954	0.937	0.883	0.847	0.813	0.965	0.975
	epa	0.867	0.916	0.939	0.950	0.890	0.934	0.952	0.958	0.899	0.807	0.719	0.644	0.966	0.975
	gaussian	0.940	0.962	0.974	0.981	0.953	0.971	0.980	0.984	0.801	0.659	0.534	0.437	0.964	0.980

Table 61: Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 2 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$															
intexp	bartlett	0.745	0.783	0.803	0.813	0.763	0.797	0.817	0.828	0.909	0.882	0.850	0.827	0.846	0.853
	epa	0.753	0.791	0.807	0.817	0.771	0.805	0.822	0.834	0.890	0.827	0.769	0.726	0.848	0.857
	gaussian	0.812	0.840	0.851	0.857	0.827	0.849	0.859	0.867	0.817	0.717	0.628	0.536	0.848	0.859
cpr	bartlett	0.401	0.487	0.525	0.558	0.440	0.517	0.567	0.597	0.438	0.190	0.087	0.043	0.655	0.679
	epa	0.420	0.504	0.547	0.574	0.459	0.539	0.582	0.610	0.330	0.192	0.134	0.115	0.657	0.686
	gaussian	0.553	0.628	0.664	0.687	0.593	0.656	0.691	0.706	0.051	0.007	0.006	0.006	0.654	0.695
stc	bartlett	0.238	0.292	0.330	0.361	0.265	0.321	0.364	0.387	0.215	0.088	0.041	0.022	0.444	0.460
	epa	0.246	0.306	0.350	0.374	0.275	0.340	0.377	0.400	0.145	0.050	0.029	0.027	0.447	0.466
	gaussian	0.353	0.410	0.444	0.461	0.382	0.441	0.466	0.482	0.021	0.006	0.007	0.008	0.441	0.476
exp	bartlett	0.630	0.699	0.734	0.756	0.662	0.730	0.761	0.785	0.806	0.716	0.642	0.585	0.829	0.839
	epa	0.646	0.711	0.751	0.769	0.680	0.742	0.775	0.799	0.751	0.575	0.438	0.341	0.832	0.845
	gaussian	0.755	0.809	0.829	0.842	0.784	0.826	0.844	0.854	0.556	0.344	0.207	0.121	0.830	0.848
pow	bartlett	0.659	0.716	0.746	0.767	0.685	0.735	0.771	0.790	0.800	0.750	0.716	0.697	0.830	0.840
	epa	0.672	0.729	0.758	0.776	0.699	0.751	0.784	0.800	0.762	0.691	0.635	0.595	0.831	0.842
	gaussian	0.765	0.811	0.830	0.842	0.789	0.827	0.846	0.856	0.679	0.583	0.500	0.441	0.829	0.851
jump	bartlett	0.647	0.714	0.749	0.767	0.682	0.739	0.776	0.798	0.809	0.721	0.656	0.608	0.835	0.846
	epa	0.665	0.728	0.761	0.784	0.696	0.756	0.792	0.808	0.762	0.630	0.516	0.439	0.839	0.848
	gaussian	0.765	0.817	0.840	0.851	0.794	0.839	0.853	0.863	0.594	0.434	0.319	0.234	0.833	0.855

Table 62: Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 3 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$															
intexp	bartlett	0.486	0.509	0.527	0.536	0.497	0.526	0.534	0.542	0.594	0.557	0.525	0.512	0.551	0.562
	epa	0.488	0.519	0.532	0.540	0.501	0.531	0.540	0.550	0.569	0.498	0.452	0.427	0.554	0.565
	gaussian	0.533	0.557	0.569	0.576	0.544	0.563	0.577	0.582	0.498	0.443	0.395	0.355	0.553	0.570
cpr	bartlett	0.266	0.290	0.314	0.323	0.281	0.309	0.330	0.335	0.297	0.170	0.085	0.047	0.385	0.386
	epa	0.272	0.297	0.319	0.332	0.288	0.318	0.333	0.345	0.244	0.150	0.111	0.087	0.385	0.386
	gaussian	0.320	0.347	0.357	0.359	0.338	0.359	0.360	0.362	0.050	0.013	0.011	0.013	0.381	0.385
stc	bartlett	0.204	0.237	0.250	0.256	0.223	0.249	0.253	0.259	0.234	0.120	0.069	0.046	0.269	0.275
	epa	0.213	0.239	0.250	0.253	0.227	0.248	0.254	0.264	0.162	0.067	0.047	0.035	0.272	0.281
	gaussian	0.256	0.274	0.282	0.288	0.260	0.279	0.287	0.292	0.039	0.018	0.016	0.017	0.274	0.284
exp	bartlett	0.425	0.459	0.477	0.486	0.445	0.479	0.498	0.506	0.521	0.505	0.500	0.460	0.498	0.504
	epa	0.432	0.467	0.481	0.492	0.452	0.480	0.498	0.504	0.481	0.407	0.311	0.229	0.497	0.505
	gaussian	0.486	0.503	0.515	0.516	0.502	0.515	0.520	0.524	0.425	0.230	0.124	0.069	0.500	0.511
pow	bartlett	0.448	0.471	0.490	0.506	0.457	0.487	0.504	0.512	0.508	0.472	0.453	0.444	0.553	0.558
	epa	0.452	0.474	0.495	0.507	0.460	0.492	0.507	0.515	0.482	0.420	0.402	0.385	0.556	0.557
	gaussian	0.504	0.519	0.525	0.530	0.509	0.525	0.534	0.536	0.431	0.406	0.366	0.317	0.551	0.553
jump	bartlett	0.441	0.469	0.489	0.499	0.455	0.485	0.501	0.508	0.521	0.494	0.473	0.466	0.527	0.535
	epa	0.451	0.474	0.494	0.498	0.458	0.492	0.505	0.510	0.492	0.421	0.374	0.317	0.526	0.533
	gaussian	0.497	0.517	0.524	0.526	0.508	0.525	0.526	0.531	0.439	0.333	0.221	0.151	0.529	0.535

Table 63: Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 1 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$															
intexp	bartlett	0.121	0.133	0.143	0.147	0.128	0.145	0.145	0.151	0.162	0.104	0.077	0.066	0.164	0.168
	epa	0.121	0.139	0.145	0.147	0.129	0.146	0.151	0.153	0.135	0.078	0.056	0.051	0.165	0.172
	gaussian	0.146	0.157	0.166	0.173	0.153	0.163	0.172	0.178	0.068	0.047	0.030	0.023	0.162	0.177
cpr	bartlett	0.192	0.218	0.240	0.251	0.212	0.236	0.257	0.263	0.270	0.145	0.069	0.039	0.278	0.283
	epa	0.197	0.224	0.243	0.253	0.219	0.241	0.261	0.267	0.223	0.145	0.119	0.119	0.276	0.286
	gaussian	0.247	0.270	0.276	0.277	0.263	0.277	0.278	0.279	0.049	0.021	0.026	0.032	0.271	0.286
stc	bartlett	0.059	0.067	0.069	0.069	0.063	0.070	0.068	0.070	0.069	0.056	0.046	0.046	0.073	0.073
	epa	0.062	0.066	0.067	0.067	0.065	0.069	0.067	0.075	0.062	0.053	0.053	0.051	0.075	0.075
	gaussian	0.070	0.077	0.077	0.075	0.072	0.074	0.072	0.072	0.046	0.044	0.046	0.046	0.075	0.074
exp	bartlett	0.582	0.645	0.672	0.691	0.618	0.669	0.695	0.709	0.857	0.844	0.829	0.779	0.735	0.737
	epa	0.592	0.655	0.679	0.699	0.628	0.673	0.702	0.713	0.832	0.702	0.520	0.359	0.727	0.736
	gaussian	0.687	0.718	0.734	0.738	0.706	0.732	0.740	0.748	0.743	0.388	0.172	0.077	0.733	0.744
pow	bartlett	0.081	0.081	0.089	0.094	0.082	0.089	0.095	0.097	0.058	0.055	0.051	0.049	0.095	0.097
	epa	0.083	0.083	0.090	0.092	0.079	0.092	0.095	0.096	0.057	0.048	0.048	0.048	0.097	0.096
	gaussian	0.093	0.093	0.092	0.092	0.096	0.096	0.094	0.091	0.048	0.048	0.050	0.048	0.097	0.094
jump	bartlett	0.382	0.416	0.443	0.456	0.397	0.437	0.459	0.465	0.551	0.540	0.530	0.521	0.465	0.473
	epa	0.390	0.426	0.453	0.455	0.403	0.447	0.463	0.467	0.541	0.497	0.447	0.388	0.461	0.472
	gaussian	0.453	0.472	0.482	0.482	0.463	0.481	0.482	0.490	0.511	0.395	0.269	0.182	0.466	0.478

Table 64: Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 2 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$															
intexp	bartlett	0.615	0.632	0.646	0.650	0.622	0.644	0.650	0.656	0.717	0.691	0.669	0.655	0.652	0.661
	epa	0.616	0.641	0.649	0.652	0.624	0.649	0.654	0.658	0.694	0.625	0.556	0.516	0.653	0.663
	gaussian	0.650	0.661	0.669	0.678	0.656	0.668	0.678	0.682	0.632	0.545	0.470	0.416	0.650	0.667
cpr	bartlett	0.278	0.307	0.323	0.334	0.296	0.319	0.338	0.345	0.323	0.186	0.090	0.047	0.397	0.397
	epa	0.284	0.310	0.330	0.342	0.302	0.327	0.343	0.352	0.256	0.148	0.109	0.082	0.398	0.398
	gaussian	0.332	0.352	0.358	0.362	0.345	0.361	0.362	0.364	0.052	0.013	0.013	0.013	0.395	0.398
stc	bartlett	0.165	0.198	0.209	0.219	0.183	0.208	0.219	0.229	0.200	0.090	0.044	0.028	0.229	0.236
	epa	0.176	0.198	0.212	0.219	0.186	0.209	0.221	0.237	0.140	0.053	0.032	0.025	0.233	0.239
	gaussian	0.215	0.242	0.250	0.249	0.228	0.246	0.249	0.251	0.029	0.014	0.014	0.015	0.234	0.246
exp	bartlett	0.483	0.516	0.535	0.545	0.505	0.534	0.548	0.557	0.608	0.603	0.599	0.566	0.548	0.555
	epa	0.485	0.525	0.540	0.551	0.510	0.537	0.553	0.559	0.579	0.507	0.401	0.306	0.542	0.555
	gaussian	0.543	0.563	0.574	0.575	0.554	0.569	0.576	0.579	0.533	0.322	0.184	0.102	0.549	0.561
pow	bartlett	0.556	0.579	0.599	0.612	0.568	0.597	0.609	0.619	0.644	0.610	0.591	0.587	0.674	0.670
	epa	0.561	0.584	0.604	0.614	0.571	0.601	0.616	0.624	0.619	0.559	0.531	0.515	0.674	0.668
	gaussian	0.611	0.626	0.632	0.636	0.617	0.630	0.637	0.639	0.569	0.538	0.497	0.437	0.674	0.666
jump	bartlett	0.498	0.524	0.543	0.551	0.510	0.537	0.555	0.561	0.595	0.571	0.550	0.542	0.558	0.566
	epa	0.503	0.531	0.548	0.551	0.513	0.546	0.558	0.560	0.577	0.515	0.462	0.401	0.556	0.568
	gaussian	0.550	0.565	0.570	0.572	0.558	0.567	0.573	0.580	0.528	0.422	0.303	0.227	0.558	0.569

Table 65: Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 3 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$	$\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.144	0.145	0.146	0.148	0.138	0.137	0.139	0.137	0.432	0.349	0.261	0.205	0.130	0.138
	epa	0.145	0.144	0.148	0.149	0.137	0.137	0.139	0.137	0.433	0.276	0.179	0.133	0.133	0.141
	gaussian	0.147	0.142	0.147	0.149	0.138	0.140	0.149	0.159	0.227	0.116	0.084	0.071	0.131	0.150
cpr	bartlett	0.135	0.154	0.162	0.167	0.144	0.154	0.161	0.166	0.791	0.596	0.390	0.245	0.158	0.185
	epa	0.139	0.154	0.165	0.170	0.145	0.157	0.165	0.169	0.718	0.396	0.181	0.083	0.164	0.189
	gaussian	0.167	0.174	0.196	0.215	0.166	0.189	0.217	0.252	0.299	0.047	0.007	0.001	0.164	0.214
stc	bartlett	0.121	0.130	0.130	0.129	0.117	0.122	0.126	0.124	0.533	0.402	0.264	0.176	0.108	0.121
	epa	0.121	0.129	0.130	0.132	0.118	0.123	0.124	0.125	0.522	0.281	0.138	0.067	0.112	0.125
	gaussian	0.129	0.133	0.142	0.149	0.124	0.134	0.147	0.161	0.206	0.041	0.009	0.002	0.109	0.136
exp	bartlett	0.255	0.296	0.312	0.324	0.270	0.302	0.327	0.346	0.428	0.305	0.254	0.251	0.393	0.415
	epa	0.261	0.303	0.318	0.334	0.277	0.315	0.336	0.355	0.367	0.220	0.156	0.124	0.398	0.420
	gaussian	0.325	0.369	0.407	0.435	0.347	0.399	0.442	0.460	0.228	0.170	0.146	0.136	0.393	0.442
pow	bartlett	0.618	0.717	0.761	0.788	0.663	0.745	0.792	0.823	0.772	0.586	0.523	0.545	0.838	0.870
	epa	0.641	0.738	0.778	0.806	0.686	0.767	0.809	0.836	0.624	0.369	0.246	0.178	0.842	0.876
	gaussian	0.787	0.850	0.883	0.903	0.822	0.875	0.903	0.916	0.423	0.303	0.256	0.238	0.838	0.893
jump	bartlett	0.345	0.395	0.414	0.429	0.365	0.404	0.434	0.453	0.479	0.332	0.274	0.276	0.505	0.528
	epa	0.353	0.406	0.422	0.442	0.376	0.417	0.447	0.462	0.394	0.222	0.155	0.115	0.508	0.529
	gaussian	0.430	0.475	0.515	0.535	0.454	0.502	0.538	0.556	0.237	0.172	0.146	0.138	0.507	0.541
$T = 200$	$\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.187	0.217	0.240	0.250	0.203	0.232	0.252	0.269	0.502	0.397	0.309	0.244	0.300	0.315
	epa	0.196	0.228	0.245	0.259	0.206	0.238	0.261	0.278	0.477	0.329	0.223	0.167	0.304	0.318
	gaussian	0.248	0.289	0.308	0.326	0.270	0.300	0.325	0.336	0.268	0.135	0.088	0.072	0.301	0.322
cpr	bartlett	0.117	0.141	0.162	0.168	0.129	0.154	0.174	0.185	0.466	0.293	0.165	0.092	0.217	0.228
	epa	0.122	0.147	0.165	0.174	0.132	0.159	0.180	0.192	0.417	0.195	0.075	0.032	0.218	0.232
	gaussian	0.166	0.206	0.223	0.242	0.187	0.221	0.243	0.261	0.119	0.009	0.001	0.000	0.216	0.243
stc	bartlett	0.197	0.232	0.260	0.271	0.214	0.254	0.275	0.289	0.578	0.445	0.321	0.223	0.307	0.324
	epa	0.205	0.246	0.265	0.277	0.217	0.259	0.281	0.298	0.545	0.349	0.196	0.118	0.308	0.329
	gaussian	0.268	0.312	0.331	0.353	0.290	0.323	0.352	0.368	0.259	0.072	0.021	0.006	0.307	0.339
exp	bartlett	0.214	0.234	0.253	0.274	0.228	0.253	0.280	0.299	0.369	0.272	0.235	0.236	0.355	0.374
	epa	0.221	0.241	0.264	0.289	0.229	0.263	0.291	0.317	0.331	0.199	0.142	0.109	0.357	0.378
	gaussian	0.272	0.329	0.370	0.393	0.299	0.357	0.400	0.422	0.213	0.157	0.141	0.132	0.355	0.391
pow	bartlett	0.220	0.266	0.295	0.319	0.244	0.292	0.322	0.347	0.528	0.392	0.336	0.345	0.360	0.391
	epa	0.232	0.279	0.310	0.334	0.253	0.303	0.336	0.366	0.471	0.273	0.180	0.131	0.366	0.399
	gaussian	0.316	0.376	0.421	0.447	0.346	0.412	0.453	0.483	0.302	0.223	0.195	0.180	0.363	0.418
jump	bartlett	0.259	0.284	0.307	0.330	0.272	0.306	0.336	0.358	0.403	0.285	0.242	0.248	0.411	0.430
	epa	0.266	0.294	0.320	0.342	0.274	0.317	0.349	0.373	0.346	0.193	0.134	0.099	0.415	0.434
	gaussian	0.328	0.383	0.426	0.444	0.357	0.415	0.452	0.471	0.212	0.155	0.137	0.129	0.413	0.445

Table 66: Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 1 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$	$\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.144	0.145	0.146	0.148	0.138	0.137	0.139	0.137	0.432	0.349	0.261	0.205	0.130	0.138
	epa	0.145	0.144	0.148	0.149	0.137	0.137	0.139	0.137	0.433	0.276	0.179	0.133	0.133	0.141
	gaussian	0.147	0.142	0.147	0.149	0.138	0.140	0.149	0.159	0.227	0.116	0.084	0.071	0.131	0.150
cpr	bartlett	0.135	0.154	0.162	0.167	0.144	0.154	0.161	0.166	0.791	0.596	0.390	0.245	0.158	0.185
	epa	0.139	0.154	0.165	0.170	0.145	0.157	0.165	0.169	0.718	0.396	0.181	0.083	0.164	0.189
	gaussian	0.167	0.174	0.196	0.215	0.166	0.189	0.217	0.252	0.299	0.047	0.007	0.001	0.164	0.214
stc	bartlett	0.121	0.130	0.130	0.129	0.117	0.122	0.126	0.124	0.533	0.402	0.264	0.176	0.108	0.121
	epa	0.121	0.129	0.130	0.132	0.118	0.123	0.124	0.125	0.522	0.281	0.138	0.067	0.112	0.125
	gaussian	0.129	0.133	0.142	0.149	0.124	0.134	0.147	0.161	0.206	0.041	0.009	0.002	0.109	0.136
exp	bartlett	0.255	0.296	0.312	0.324	0.270	0.302	0.327	0.346	0.428	0.305	0.254	0.251	0.393	0.415
	epa	0.261	0.303	0.318	0.334	0.277	0.315	0.336	0.355	0.367	0.220	0.156	0.124	0.398	0.420
	gaussian	0.325	0.369	0.407	0.435	0.347	0.399	0.442	0.460	0.228	0.170	0.146	0.136	0.393	0.442
pow	bartlett	0.618	0.717	0.761	0.788	0.663	0.745	0.792	0.823	0.772	0.586	0.523	0.545	0.838	0.870
	epa	0.641	0.738	0.778	0.806	0.686	0.767	0.809	0.836	0.624	0.369	0.246	0.178	0.842	0.876
	gaussian	0.787	0.850	0.883	0.903	0.822	0.875	0.903	0.916	0.423	0.303	0.256	0.238	0.838	0.893
jump	bartlett	0.345	0.395	0.414	0.429	0.365	0.404	0.434	0.453	0.479	0.332	0.274	0.276	0.505	0.528
	epa	0.353	0.406	0.422	0.442	0.376	0.417	0.447	0.462	0.394	0.222	0.155	0.115	0.508	0.529
	gaussian	0.430	0.475	0.515	0.535	0.454	0.502	0.538	0.556	0.237	0.172	0.146	0.138	0.507	0.541
$T = 200$	$\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.187	0.217	0.240	0.250	0.203	0.232	0.252	0.269	0.502	0.397	0.309	0.244	0.300	0.315
	epa	0.196	0.228	0.245	0.259	0.206	0.238	0.261	0.278	0.477	0.329	0.223	0.167	0.304	0.318
	gaussian	0.248	0.289	0.308	0.326	0.270	0.300	0.325	0.336	0.268	0.135	0.088	0.072	0.301	0.322
cpr	bartlett	0.117	0.141	0.162	0.168	0.129	0.154	0.174	0.185	0.466	0.293	0.165	0.092	0.217	0.228
	epa	0.122	0.147	0.165	0.174	0.132	0.159	0.180	0.192	0.417	0.195	0.075	0.032	0.218	0.232
	gaussian	0.166	0.206	0.223	0.242	0.187	0.221	0.243	0.261	0.119	0.009	0.001	0.000	0.216	0.243
stc	bartlett	0.197	0.232	0.260	0.271	0.214	0.254	0.275	0.289	0.578	0.445	0.321	0.223	0.307	0.324
	epa	0.205	0.246	0.265	0.277	0.217	0.259	0.281	0.298	0.545	0.349	0.196	0.118	0.308	0.329
	gaussian	0.268	0.312	0.331	0.353	0.290	0.323	0.352	0.368	0.259	0.072	0.021	0.006	0.307	0.339
exp	bartlett	0.214	0.234	0.253	0.274	0.228	0.253	0.280	0.299	0.369	0.272	0.235	0.236	0.355	0.374
	epa	0.221	0.241	0.264	0.289	0.229	0.263	0.291	0.317	0.331	0.199	0.142	0.109	0.357	0.378
	gaussian	0.272	0.329	0.370	0.393	0.299	0.357	0.400	0.422	0.213	0.157	0.141	0.132	0.355	0.391
pow	bartlett	0.220	0.266	0.295	0.319	0.244	0.292	0.322	0.347	0.528	0.392	0.336	0.345	0.360	0.391
	epa	0.232	0.279	0.310	0.334	0.253	0.303	0.336	0.366	0.471	0.273	0.180	0.131	0.366	0.399
	gaussian	0.316	0.376	0.421	0.447	0.346	0.412	0.453	0.483	0.302	0.223	0.195	0.180	0.363	0.418
jump	bartlett	0.259	0.284	0.307	0.330	0.272	0.306	0.336	0.358	0.403	0.285	0.242	0.248	0.411	0.430
	epa	0.266	0.294	0.320	0.342	0.274	0.317	0.349	0.373	0.346	0.193	0.134	0.099	0.415	0.434
	gaussian	0.328	0.383	0.426	0.444	0.357	0.415	0.452	0.471	0.212	0.155	0.137	0.129	0.413	0.445

Table 67: Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 2 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$	$\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.144	0.145	0.146	0.148	0.138	0.137	0.139	0.137	0.432	0.349	0.261	0.205	0.130	0.138
	epa	0.145	0.144	0.148	0.149	0.137	0.137	0.139	0.137	0.433	0.276	0.179	0.133	0.133	0.141
	gaussian	0.147	0.142	0.147	0.149	0.138	0.140	0.149	0.159	0.227	0.116	0.084	0.071	0.131	0.150
cpr	bartlett	0.135	0.154	0.162	0.167	0.144	0.154	0.161	0.166	0.791	0.596	0.390	0.245	0.158	0.185
	epa	0.139	0.154	0.165	0.170	0.145	0.157	0.165	0.169	0.718	0.396	0.181	0.083	0.164	0.189
	gaussian	0.167	0.174	0.196	0.215	0.166	0.189	0.217	0.252	0.299	0.047	0.007	0.001	0.164	0.214
stc	bartlett	0.121	0.130	0.130	0.129	0.117	0.122	0.126	0.124	0.533	0.402	0.264	0.176	0.108	0.121
	epa	0.121	0.129	0.130	0.132	0.118	0.123	0.124	0.125	0.522	0.281	0.138	0.067	0.112	0.125
	gaussian	0.129	0.133	0.142	0.149	0.124	0.134	0.147	0.161	0.206	0.041	0.009	0.002	0.109	0.136
exp	bartlett	0.255	0.296	0.312	0.324	0.270	0.302	0.327	0.346	0.428	0.305	0.254	0.251	0.393	0.415
	epa	0.261	0.303	0.318	0.334	0.277	0.315	0.336	0.355	0.367	0.220	0.156	0.124	0.398	0.420
	gaussian	0.325	0.369	0.407	0.435	0.347	0.399	0.442	0.460	0.228	0.170	0.146	0.136	0.393	0.442
pow	bartlett	0.618	0.717	0.761	0.788	0.663	0.745	0.792	0.823	0.772	0.586	0.523	0.545	0.838	0.870
	epa	0.641	0.738	0.778	0.806	0.686	0.767	0.809	0.836	0.624	0.369	0.246	0.178	0.842	0.876
	gaussian	0.787	0.850	0.883	0.903	0.822	0.875	0.903	0.916	0.423	0.303	0.256	0.238	0.838	0.893
jump	bartlett	0.345	0.395	0.414	0.429	0.365	0.404	0.434	0.453	0.479	0.332	0.274	0.276	0.505	0.528
	epa	0.353	0.406	0.422	0.442	0.376	0.417	0.447	0.462	0.394	0.222	0.155	0.115	0.508	0.529
	gaussian	0.430	0.475	0.515	0.535	0.454	0.502	0.538	0.556	0.237	0.172	0.146	0.138	0.507	0.541
$T = 200$	$\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.187	0.217	0.240	0.250	0.203	0.232	0.252	0.269	0.502	0.397	0.309	0.244	0.300	0.315
	epa	0.196	0.228	0.245	0.259	0.206	0.238	0.261	0.278	0.477	0.329	0.223	0.167	0.304	0.318
	gaussian	0.248	0.289	0.308	0.326	0.270	0.300	0.325	0.336	0.268	0.135	0.088	0.072	0.301	0.322
cpr	bartlett	0.117	0.141	0.162	0.168	0.129	0.154	0.174	0.185	0.466	0.293	0.165	0.092	0.217	0.228
	epa	0.122	0.147	0.165	0.174	0.132	0.159	0.180	0.192	0.417	0.195	0.075	0.032	0.218	0.232
	gaussian	0.166	0.206	0.223	0.242	0.187	0.221	0.243	0.261	0.119	0.009	0.001	0.000	0.216	0.243
stc	bartlett	0.197	0.232	0.260	0.271	0.214	0.254	0.275	0.289	0.578	0.445	0.321	0.223	0.307	0.324
	epa	0.205	0.246	0.265	0.277	0.217	0.259	0.281	0.298	0.545	0.349	0.196	0.118	0.308	0.329
	gaussian	0.268	0.312	0.331	0.353	0.290	0.323	0.352	0.368	0.259	0.072	0.021	0.006	0.307	0.339
exp	bartlett	0.214	0.234	0.253	0.274	0.228	0.253	0.280	0.299	0.369	0.272	0.235	0.236	0.355	0.374
	epa	0.221	0.241	0.264	0.289	0.229	0.263	0.291	0.317	0.331	0.199	0.142	0.109	0.357	0.378
	gaussian	0.272	0.329	0.370	0.393	0.299	0.357	0.400	0.422	0.213	0.157	0.141	0.132	0.355	0.391
pow	bartlett	0.220	0.266	0.295	0.319	0.244	0.292	0.322	0.347	0.528	0.392	0.336	0.345	0.360	0.391
	epa	0.232	0.279	0.310	0.334	0.253	0.303	0.336	0.366	0.471	0.273	0.180	0.131	0.366	0.399
	gaussian	0.316	0.376	0.421	0.447	0.346	0.412	0.453	0.483	0.302	0.223	0.195	0.180	0.363	0.418
jump	bartlett	0.259	0.284	0.307	0.330	0.272	0.306	0.336	0.358	0.403	0.285	0.242	0.248	0.411	0.430
	epa	0.266	0.294	0.320	0.342	0.274	0.317	0.349	0.373	0.346	0.193	0.134	0.099	0.415	0.434
	gaussian	0.328	0.383	0.426	0.444	0.357	0.415	0.452	0.471	0.212	0.155	0.137	0.129	0.413	0.445

Table 68: Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 3 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$	$\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.092	0.090	0.090	0.090	0.088	0.087	0.089	0.090	0.205	0.168	0.147	0.134	0.093	0.096
	epa	0.092	0.089	0.089	0.091	0.089	0.088	0.090	0.091	0.205	0.154	0.121	0.101	0.092	0.097
	gaussian	0.091	0.094	0.100	0.105	0.090	0.096	0.102	0.110	0.137	0.090	0.074	0.067	0.091	0.101
cpr	bartlett	0.040	0.039	0.038	0.039	0.040	0.039	0.039	0.040	0.157	0.119	0.088	0.064	0.050	0.055
	epa	0.040	0.039	0.038	0.040	0.040	0.038	0.040	0.041	0.156	0.108	0.061	0.039	0.050	0.056
	gaussian	0.038	0.042	0.051	0.055	0.040	0.048	0.053	0.064	0.074	0.022	0.008	0.004	0.051	0.060
stc	bartlett	0.065	0.065	0.064	0.066	0.063	0.063	0.064	0.066	0.216	0.160	0.120	0.092	0.060	0.069
	epa	0.065	0.063	0.064	0.068	0.063	0.063	0.065	0.067	0.214	0.137	0.089	0.051	0.060	0.069
	gaussian	0.066	0.072	0.079	0.087	0.066	0.075	0.082	0.093	0.108	0.028	0.006	0.002	0.060	0.076
exp	bartlett	0.208	0.211	0.218	0.218	0.202	0.214	0.221	0.221	0.201	0.142	0.124	0.113	0.259	0.263
	epa	0.210	0.217	0.216	0.217	0.209	0.217	0.221	0.224	0.168	0.111	0.088	0.072	0.260	0.264
	gaussian	0.218	0.228	0.238	0.246	0.222	0.236	0.249	0.254	0.114	0.089	0.078	0.075	0.258	0.269
pow	bartlett	0.260	0.301	0.328	0.338	0.273	0.319	0.342	0.355	0.284	0.207	0.180	0.162	0.382	0.399
	epa	0.269	0.315	0.331	0.343	0.290	0.329	0.349	0.362	0.228	0.157	0.111	0.086	0.385	0.399
	gaussian	0.336	0.376	0.393	0.403	0.354	0.389	0.409	0.419	0.165	0.118	0.097	0.092	0.384	0.408
jump	bartlett	0.256	0.268	0.269	0.268	0.255	0.267	0.267	0.270	0.216	0.150	0.134	0.116	0.323	0.324
	epa	0.261	0.270	0.266	0.267	0.264	0.268	0.269	0.272	0.176	0.123	0.088	0.073	0.325	0.325
	gaussian	0.269	0.278	0.286	0.296	0.272	0.285	0.296	0.304	0.123	0.090	0.078	0.074	0.321	0.327
$T = 200$	$\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.128	0.134	0.134	0.137	0.131	0.135	0.135	0.142	0.254	0.204	0.169	0.151	0.144	0.157
	epa	0.129	0.135	0.136	0.139	0.133	0.136	0.140	0.143	0.245	0.181	0.143	0.118	0.147	0.156
	gaussian	0.137	0.154	0.160	0.166	0.141	0.150	0.162	0.179	0.158	0.100	0.073	0.067	0.145	0.162
cpr	bartlett	0.055	0.056	0.057	0.059	0.057	0.056	0.058	0.060	0.098	0.072	0.048	0.033	0.063	0.066
	epa	0.056	0.055	0.059	0.058	0.057	0.057	0.061	0.059	0.091	0.059	0.034	0.021	0.064	0.066
	gaussian	0.059	0.062	0.062	0.065	0.060	0.060	0.067	0.073	0.038	0.013	0.004	0.003	0.064	0.069
stc	bartlett	0.118	0.130	0.132	0.136	0.123	0.130	0.135	0.140	0.299	0.229	0.166	0.128	0.134	0.147
	epa	0.120	0.131	0.134	0.137	0.124	0.132	0.139	0.144	0.288	0.192	0.120	0.076	0.135	0.147
	gaussian	0.135	0.150	0.159	0.168	0.139	0.153	0.169	0.181	0.147	0.053	0.019	0.007	0.134	0.155
exp	bartlett	0.160	0.171	0.178	0.184	0.165	0.176	0.183	0.186	0.173	0.123	0.106	0.105	0.219	0.217
	epa	0.163	0.175	0.179	0.183	0.163	0.178	0.184	0.187	0.152	0.098	0.079	0.067	0.220	0.215
	gaussian	0.184	0.196	0.201	0.207	0.187	0.198	0.210	0.219	0.102	0.082	0.078	0.073	0.220	0.218
pow	bartlett	0.107	0.120	0.130	0.136	0.114	0.126	0.136	0.145	0.167	0.135	0.117	0.114	0.157	0.160
	epa	0.109	0.124	0.133	0.141	0.114	0.128	0.142	0.148	0.156	0.108	0.085	0.071	0.158	0.160
	gaussian	0.135	0.153	0.159	0.165	0.144	0.157	0.165	0.173	0.114	0.091	0.083	0.079	0.158	0.164
jump	bartlett	0.185	0.191	0.196	0.200	0.188	0.195	0.198	0.201	0.182	0.126	0.110	0.106	0.232	0.232
	epa	0.185	0.193	0.195	0.198	0.185	0.193	0.199	0.203	0.158	0.102	0.080	0.067	0.232	0.230
	gaussian	0.199	0.204	0.208	0.215	0.201	0.209	0.215	0.224	0.107	0.084	0.075	0.074	0.232	0.234

Table 69: Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 1 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$	$\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.092	0.090	0.090	0.090	0.088	0.087	0.089	0.090	0.205	0.168	0.147	0.134	0.093	0.096
	epa	0.092	0.089	0.089	0.091	0.089	0.088	0.090	0.091	0.205	0.154	0.121	0.101	0.092	0.097
	gaussian	0.091	0.094	0.100	0.105	0.090	0.096	0.102	0.110	0.137	0.090	0.074	0.067	0.091	0.101
cpr	bartlett	0.040	0.039	0.038	0.039	0.040	0.039	0.039	0.040	0.157	0.119	0.088	0.064	0.050	0.055
	epa	0.040	0.039	0.038	0.040	0.040	0.038	0.040	0.041	0.156	0.108	0.061	0.039	0.050	0.056
	gaussian	0.038	0.042	0.051	0.055	0.040	0.048	0.053	0.064	0.074	0.022	0.008	0.004	0.051	0.060
stc	bartlett	0.065	0.065	0.064	0.066	0.063	0.063	0.064	0.066	0.216	0.160	0.120	0.092	0.060	0.069
	epa	0.065	0.063	0.064	0.068	0.063	0.063	0.065	0.067	0.214	0.137	0.089	0.051	0.060	0.069
	gaussian	0.066	0.072	0.079	0.087	0.066	0.075	0.082	0.093	0.108	0.028	0.006	0.002	0.060	0.076
exp	bartlett	0.208	0.211	0.218	0.218	0.202	0.214	0.221	0.221	0.201	0.142	0.124	0.113	0.259	0.263
	epa	0.210	0.217	0.216	0.217	0.209	0.217	0.221	0.224	0.168	0.111	0.088	0.072	0.260	0.264
	gaussian	0.218	0.228	0.238	0.246	0.222	0.236	0.249	0.254	0.114	0.089	0.078	0.075	0.258	0.269
pow	bartlett	0.260	0.301	0.328	0.338	0.273	0.319	0.342	0.355	0.284	0.207	0.180	0.162	0.382	0.399
	epa	0.269	0.315	0.331	0.343	0.290	0.329	0.349	0.362	0.228	0.157	0.111	0.086	0.385	0.399
	gaussian	0.336	0.376	0.393	0.403	0.354	0.389	0.409	0.419	0.165	0.118	0.097	0.092	0.384	0.408
jump	bartlett	0.256	0.268	0.269	0.268	0.255	0.267	0.267	0.270	0.216	0.150	0.134	0.116	0.323	0.324
	epa	0.261	0.270	0.266	0.267	0.264	0.268	0.269	0.272	0.176	0.123	0.088	0.073	0.325	0.325
	gaussian	0.269	0.278	0.286	0.296	0.272	0.285	0.296	0.304	0.123	0.090	0.078	0.074	0.321	0.327
$T = 200$	$\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.128	0.134	0.134	0.137	0.131	0.135	0.135	0.142	0.254	0.204	0.169	0.151	0.144	0.157
	epa	0.129	0.135	0.136	0.139	0.133	0.136	0.140	0.143	0.245	0.181	0.143	0.118	0.147	0.156
	gaussian	0.137	0.154	0.160	0.166	0.141	0.150	0.162	0.179	0.158	0.100	0.073	0.067	0.145	0.162
cpr	bartlett	0.055	0.056	0.057	0.059	0.057	0.056	0.058	0.060	0.098	0.072	0.048	0.033	0.063	0.066
	epa	0.056	0.055	0.059	0.058	0.057	0.057	0.061	0.059	0.091	0.059	0.034	0.021	0.064	0.066
	gaussian	0.059	0.062	0.062	0.065	0.060	0.060	0.067	0.073	0.038	0.013	0.004	0.003	0.064	0.069
stc	bartlett	0.118	0.130	0.132	0.136	0.123	0.130	0.135	0.140	0.299	0.229	0.166	0.128	0.134	0.147
	epa	0.120	0.131	0.134	0.137	0.124	0.132	0.139	0.144	0.288	0.192	0.120	0.076	0.135	0.147
	gaussian	0.135	0.150	0.159	0.168	0.139	0.153	0.169	0.181	0.147	0.053	0.019	0.007	0.134	0.155
exp	bartlett	0.160	0.171	0.178	0.184	0.165	0.176	0.183	0.186	0.173	0.123	0.106	0.105	0.219	0.217
	epa	0.163	0.175	0.179	0.183	0.163	0.178	0.184	0.187	0.152	0.098	0.079	0.067	0.220	0.215
	gaussian	0.184	0.196	0.201	0.207	0.187	0.198	0.210	0.219	0.102	0.082	0.078	0.073	0.220	0.218
pow	bartlett	0.107	0.120	0.130	0.136	0.114	0.126	0.136	0.145	0.167	0.135	0.117	0.114	0.157	0.160
	epa	0.109	0.124	0.133	0.141	0.114	0.128	0.142	0.148	0.156	0.108	0.085	0.071	0.158	0.160
	gaussian	0.135	0.153	0.159	0.165	0.144	0.157	0.165	0.173	0.114	0.091	0.083	0.079	0.158	0.164
jump	bartlett	0.185	0.191	0.196	0.200	0.188	0.195	0.198	0.201	0.182	0.126	0.110	0.106	0.232	0.232
	epa	0.185	0.193	0.195	0.198	0.185	0.193	0.199	0.203	0.158	0.102	0.080	0.067	0.232	0.230
	gaussian	0.199	0.204	0.208	0.215	0.201	0.209	0.215	0.224	0.107	0.084	0.075	0.074	0.232	0.234

Table 70: Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 2 under the alternative hypothesis.

		$T^{-1/3}$	$T^{-1/4}$	$T^{-1/5}$	$T^{-1/6}$	$\hat{\sigma}_v^2 T^{-1/3}$	$\hat{\sigma}_v^2 T^{-1/4}$	$\hat{\sigma}_v^2 T^{-1/5}$	$\hat{\sigma}_v^2 T^{-1/6}$	$\hat{\sigma}_x^2 T^{-1/3}$	$\hat{\sigma}_x^2 T^{-1/4}$	$\hat{\sigma}_x^2 T^{-1/5}$	$\hat{\sigma}_x^2 T^{-1/6}$	h_{opt}	$h_{opt,h}$
$T = 200$	$\pi(x) = \exp(-x^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.092	0.090	0.090	0.090	0.088	0.087	0.089	0.090	0.205	0.168	0.147	0.134	0.093	0.096
	epa	0.092	0.089	0.089	0.091	0.089	0.088	0.090	0.091	0.205	0.154	0.121	0.101	0.092	0.097
	gaussian	0.091	0.094	0.100	0.105	0.090	0.096	0.102	0.110	0.137	0.090	0.074	0.067	0.091	0.101
cpr	bartlett	0.040	0.039	0.038	0.039	0.040	0.039	0.039	0.040	0.157	0.119	0.088	0.064	0.050	0.055
	epa	0.040	0.039	0.038	0.040	0.040	0.038	0.040	0.041	0.156	0.108	0.061	0.039	0.050	0.056
	gaussian	0.038	0.042	0.051	0.055	0.040	0.048	0.053	0.064	0.074	0.022	0.008	0.004	0.051	0.060
stc	bartlett	0.065	0.065	0.064	0.066	0.063	0.063	0.064	0.066	0.216	0.160	0.120	0.092	0.060	0.069
	epa	0.065	0.063	0.064	0.068	0.063	0.063	0.065	0.067	0.214	0.137	0.089	0.051	0.060	0.069
	gaussian	0.066	0.072	0.079	0.087	0.066	0.075	0.082	0.093	0.108	0.028	0.006	0.002	0.060	0.076
exp	bartlett	0.208	0.211	0.218	0.218	0.202	0.214	0.221	0.221	0.201	0.142	0.124	0.113	0.259	0.263
	epa	0.210	0.217	0.216	0.217	0.209	0.217	0.221	0.224	0.168	0.111	0.088	0.072	0.260	0.264
	gaussian	0.218	0.228	0.238	0.246	0.222	0.236	0.249	0.254	0.114	0.089	0.078	0.075	0.258	0.269
pow	bartlett	0.260	0.301	0.328	0.338	0.273	0.319	0.342	0.355	0.284	0.207	0.180	0.162	0.382	0.399
	epa	0.269	0.315	0.331	0.343	0.290	0.329	0.349	0.362	0.228	0.157	0.111	0.086	0.385	0.399
	gaussian	0.336	0.376	0.393	0.403	0.354	0.389	0.409	0.419	0.165	0.118	0.097	0.092	0.384	0.408
jump	bartlett	0.256	0.268	0.269	0.268	0.255	0.267	0.267	0.270	0.216	0.150	0.134	0.116	0.323	0.324
	epa	0.261	0.270	0.266	0.267	0.264	0.268	0.269	0.272	0.176	0.123	0.088	0.073	0.325	0.325
	gaussian	0.269	0.278	0.286	0.296	0.272	0.285	0.296	0.304	0.123	0.090	0.078	0.074	0.321	0.327
$T = 200$	$\pi(x) = \exp(-(x - \frac{1}{T}\sum_{t=1}^{T} x_t)^2/2)/\sqrt{(2\pi)}$														
intexp	bartlett	0.128	0.134	0.134	0.137	0.131	0.135	0.135	0.142	0.254	0.204	0.169	0.151	0.144	0.157
	epa	0.129	0.135	0.136	0.139	0.133	0.136	0.140	0.143	0.245	0.181	0.143	0.118	0.147	0.156
	gaussian	0.137	0.154	0.160	0.166	0.141	0.150	0.162	0.179	0.158	0.100	0.073	0.067	0.145	0.162
cpr	bartlett	0.055	0.056	0.057	0.059	0.057	0.056	0.058	0.060	0.098	0.072	0.048	0.033	0.063	0.066
	epa	0.056	0.055	0.059	0.058	0.057	0.057	0.061	0.059	0.091	0.059	0.034	0.021	0.064	0.066
	gaussian	0.059	0.062	0.062	0.065	0.060	0.060	0.067	0.073	0.038	0.013	0.004	0.003	0.064	0.069
stc	bartlett	0.118	0.130	0.132	0.136	0.123	0.130	0.135	0.140	0.299	0.229	0.166	0.128	0.134	0.147
	epa	0.120	0.131	0.134	0.137	0.124	0.132	0.139	0.144	0.288	0.192	0.120	0.076	0.135	0.147
	gaussian	0.135	0.150	0.159	0.168	0.139	0.153	0.169	0.181	0.147	0.053	0.019	0.007	0.134	0.155
exp	bartlett	0.160	0.171	0.178	0.184	0.165	0.176	0.183	0.186	0.173	0.123	0.106	0.105	0.219	0.217
	epa	0.163	0.175	0.179	0.183	0.163	0.178	0.184	0.187	0.152	0.098	0.079	0.067	0.220	0.215
	gaussian	0.184	0.196	0.201	0.207	0.187	0.198	0.210	0.219	0.102	0.082	0.078	0.073	0.220	0.218
pow	bartlett	0.107	0.120	0.130	0.136	0.114	0.126	0.136	0.145	0.167	0.135	0.117	0.114	0.157	0.160
	epa	0.109	0.124	0.133	0.141	0.114	0.128	0.142	0.148	0.156	0.108	0.085	0.071	0.158	0.160
	gaussian	0.135	0.153	0.159	0.165	0.144	0.157	0.165	0.173	0.114	0.091	0.083	0.079	0.158	0.164
jump	bartlett	0.185	0.191	0.196	0.200	0.188	0.195	0.198	0.201	0.182	0.126	0.110	0.106	0.232	0.232
	epa	0.185	0.193	0.195	0.198	0.185	0.193	0.199	0.203	0.158	0.102	0.080	0.067	0.232	0.230
	gaussian	0.199	0.204	0.208	0.215	0.201	0.209	0.215	0.224	0.107	0.084	0.075	0.074	0.232	0.234

Table 71: Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 3 under the alternative hypothesis.

	H_1: DGP 1						H_1: DGP 2						H_1: DGP 3					
M	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$
$T = 50$																		
intexp	0.079	0.073	0.058	0.055	0.081	0.065	0.048	0.051	0.050	0.048	0.052	0.051	0.293	0.286	0.208	0.155	0.313	0.238
cpr	0.077	0.074	0.062	0.063	0.073	0.069	0.044	0.045	0.050	0.051	0.041	0.053	0.168	0.144	0.109	0.095	0.161	0.125
stc	0.070	0.068	0.063	0.056	0.065	0.067	0.050	0.050	0.049	0.049	0.047	0.048	0.140	0.130	0.097	0.084	0.139	0.111
exp	0.076	0.073	0.063	0.062	0.080	0.067	0.054	0.043	0.045	0.056	0.053	0.044	0.234	0.212	0.150	0.127	0.249	0.166
pow	0.080	0.074	0.065	0.065	0.078	0.065	0.052	0.049	0.050	0.050	0.048	0.047	0.250	0.235	0.173	0.139	0.251	0.194
jump	0.086	0.081	0.070	0.071	0.085	0.074	0.062	0.050	0.044	0.053	0.055	0.047	0.265	0.238	0.164	0.143	0.262	0.192
$T = 100$																		
intexp	0.121	0.110	0.091	0.085	0.121	0.091	0.069	0.078	0.067	0.062	0.069	0.067	0.855	0.824	0.715	0.636	0.855	0.715
cpr	0.140	0.130	0.109	0.096	0.140	0.109	0.042	0.043	0.043	0.049	0.042	0.043	0.696	0.638	0.520	0.441	0.696	0.520
stc	0.141	0.116	0.098	0.091	0.141	0.098	0.046	0.044	0.048	0.048	0.046	0.048	0.634	0.569	0.458	0.390	0.634	0.458
exp	0.142	0.123	0.104	0.093	0.142	0.104	0.282	0.265	0.194	0.161	0.282	0.194	0.782	0.743	0.629	0.533	0.782	0.629
pow	0.139	0.114	0.098	0.092	0.139	0.098	0.044	0.043	0.044	0.048	0.044	0.044	0.760	0.714	0.612	0.524	0.760	0.612
jump	0.138	0.125	0.110	0.103	0.138	0.110	0.256	0.261	0.216	0.177	0.256	0.216	0.815	0.773	0.672	0.576	0.815	0.672
$T = 200$																		
intexp	0.210	0.191	0.152	0.141	0.210	0.150	0.285	0.294	0.268	0.244	0.285	0.263	0.947	0.931	0.888	0.849	0.947	0.879
cpr	0.262	0.225	0.175	0.160	0.262	0.166	0.188	0.184	0.153	0.137	0.188	0.144	0.901	0.874	0.806	0.753	0.901	0.787
stc	0.273	0.238	0.187	0.164	0.273	0.180	0.047	0.050	0.051	0.051	0.047	0.049	0.849	0.823	0.735	0.676	0.849	0.716
exp	0.236	0.213	0.166	0.145	0.236	0.164	0.820	0.841	0.791	0.729	0.820	0.775	0.877	0.855	0.796	0.754	0.877	0.788
pow	0.263	0.230	0.185	0.168	0.263	0.175	0.052	0.052	0.050	0.054	0.052	0.052	0.893	0.866	0.798	0.758	0.893	0.780
jump	0.218	0.191	0.158	0.144	0.218	0.151	0.688	0.713	0.678	0.636	0.688	0.665	0.924	0.903	0.860	0.828	0.924	0.850
$T = 500$																		
intexp	0.310	0.299	0.260	0.242	0.309	0.243	0.900	0.913	0.900	0.876	0.913	0.879	0.874	0.876	0.865	0.857	0.878	0.859
cpr	0.442	0.407	0.330	0.298	0.424	0.301	0.852	0.863	0.831	0.792	0.865	0.799	0.801	0.785	0.736	0.702	0.794	0.710
stc	0.532	0.493	0.425	0.389	0.512	0.388	0.159	0.177	0.183	0.175	0.168	0.172	0.705	0.684	0.628	0.595	0.688	0.598
exp	0.412	0.398	0.341	0.315	0.399	0.321	1.000	1.000	1.000	1.000	1.000	1.000	0.786	0.795	0.782	0.765	0.789	0.769
pow	0.443	0.422	0.352	0.320	0.431	0.328	0.162	0.176	0.158	0.141	0.176	0.141	0.776	0.769	0.719	0.691	0.771	0.694
jump	0.341	0.320	0.257	0.239	0.323	0.247	0.983	0.987	0.983	0.975	0.986	0.976	0.854	0.849	0.823	0.810	0.848	0.810
$T = 1000$																		
intexp	0.406	0.386	0.354	0.327	0.381	0.321	1.000	0.999	0.999	0.999	1.000	0.998	0.847	0.856	0.878	0.891	0.861	0.894
cpr	0.573	0.542	0.477	0.440	0.533	0.425	0.998	0.999	0.998	0.996	0.999	0.994	0.693	0.671	0.644	0.631	0.666	0.621
stc	0.719	0.682	0.627	0.592	0.673	0.583	0.521	0.562	0.593	0.582	0.569	0.572	0.713	0.688	0.657	0.632	0.683	0.631
exp	0.570	0.548	0.512	0.491	0.541	0.481	1.000	1.000	1.000	1.000	1.000	1.000	0.924	0.939	0.951	0.949	0.938	0.946
pow	0.574	0.545	0.490	0.457	0.536	0.445	0.540	0.564	0.541	0.505	0.563	0.493	0.749	0.736	0.718	0.710	0.731	0.706
jump	0.441	0.421	0.371	0.338	0.409	0.322	1.000	1.000	1.000	1.000	1.000	0.999	0.801	0.789	0.781	0.780	0.784	0.770

Table 72: Size-corrected power for the T_{WZ} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

	H_1: DGP 1						H_1: DGP 2						H_1: DGP 3					
M	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$
$T = 50$																		
intexp	0.082	0.068	0.056	0.048	0.081	0.053	0.057	0.054	0.047	0.049	0.051	0.045	0.069	0.065	0.057	0.052	0.066	0.056
cpr	0.089	0.085	0.075	0.068	0.101	0.074	0.053	0.047	0.050	0.048	0.050	0.044	0.058	0.056	0.054	0.046	0.058	0.052
stc	0.084	0.080	0.070	0.064	0.093	0.069	0.049	0.051	0.052	0.050	0.053	0.050	0.057	0.058	0.048	0.044	0.059	0.049
exp	0.078	0.077	0.071	0.065	0.096	0.066	0.054	0.052	0.051	0.048	0.049	0.044	0.057	0.055	0.054	0.045	0.064	0.049
pow	0.083	0.076	0.060	0.058	0.092	0.062	0.050	0.050	0.050	0.046	0.050	0.048	0.063	0.058	0.050	0.047	0.060	0.050
jump	0.085	0.080	0.074	0.069	0.101	0.071	0.048	0.044	0.042	0.041	0.046	0.042	0.057	0.056	0.051	0.042	0.059	0.049
$T = 100$																		
intexp	0.127	0.113	0.087	0.076	0.127	0.087	0.049	0.054	0.048	0.051	0.049	0.048	0.156	0.147	0.112	0.102	0.156	0.112
cpr	0.158	0.135	0.115	0.106	0.158	0.115	0.052	0.050	0.049	0.052	0.052	0.049	0.100	0.093	0.075	0.070	0.100	0.075
stc	0.150	0.129	0.113	0.093	0.150	0.113	0.046	0.050	0.051	0.044	0.046	0.051	0.094	0.091	0.079	0.066	0.094	0.079
exp	0.137	0.120	0.097	0.094	0.137	0.097	0.050	0.045	0.050	0.050	0.050	0.050	0.138	0.115	0.092	0.084	0.138	0.092
pow	0.145	0.128	0.097	0.086	0.145	0.097	0.049	0.050	0.051	0.050	0.049	0.051	0.133	0.126	0.099	0.092	0.133	0.099
jump	0.156	0.129	0.110	0.100	0.156	0.110	0.051	0.047	0.051	0.046	0.051	0.051	0.146	0.123	0.100	0.091	0.146	0.100
$T = 200$																		
intexp	0.202	0.166	0.150	0.140	0.202	0.141	0.054	0.051	0.055	0.052	0.054	0.052	0.514	0.395	0.315	0.268	0.514	0.291
cpr	0.274	0.218	0.179	0.161	0.274	0.169	0.050	0.048	0.052	0.046	0.050	0.050	0.285	0.225	0.189	0.155	0.285	0.168
stc	0.294	0.235	0.199	0.176	0.294	0.186	0.050	0.049	0.051	0.052	0.050	0.053	0.289	0.226	0.189	0.164	0.289	0.171
exp	0.249	0.188	0.158	0.145	0.249	0.151	0.069	0.062	0.070	0.068	0.069	0.070	0.424	0.321	0.263	0.229	0.424	0.251
pow	0.278	0.220	0.191	0.162	0.278	0.178	0.055	0.050	0.057	0.051	0.055	0.053	0.403	0.298	0.252	0.210	0.403	0.226
jump	0.222	0.176	0.151	0.142	0.222	0.146	0.067	0.064	0.073	0.067	0.067	0.068	0.440	0.338	0.279	0.239	0.440	0.251
$T = 500$																		
intexp	0.312	0.293	0.257	0.244	0.302	0.242	0.064	0.080	0.103	0.112	0.071	0.106	0.791	0.774	0.724	0.696	0.790	0.698
cpr	0.447	0.394	0.335	0.299	0.416	0.305	0.060	0.059	0.071	0.066	0.060	0.065	0.654	0.587	0.504	0.462	0.615	0.460
stc	0.547	0.492	0.425	0.389	0.515	0.392	0.051	0.051	0.047	0.046	0.051	0.047	0.697	0.647	0.550	0.506	0.674	0.512
exp	0.423	0.390	0.349	0.331	0.401	0.323	0.167	0.197	0.315	0.368	0.176	0.351	0.870	0.852	0.801	0.762	0.853	0.762
pow	0.459	0.413	0.353	0.331	0.434	0.333	0.050	0.051	0.049	0.049	0.052	0.047	0.747	0.700	0.621	0.580	0.722	0.583
jump	0.352	0.311	0.273	0.247	0.327	0.246	0.153	0.179	0.282	0.329	0.159	0.321	0.733	0.685	0.622	0.577	0.705	0.583
$T = 1000$																		
intexp	0.420	0.384	0.357	0.336	0.376	0.325	0.163	0.206	0.369	0.425	0.236	0.423	0.951	0.938	0.926	0.916	0.935	0.911
cpr	0.580	0.550	0.497	0.456	0.541	0.440	0.100	0.129	0.242	0.251	0.156	0.250	0.921	0.904	0.874	0.842	0.900	0.830
stc	0.722	0.696	0.643	0.606	0.688	0.593	0.051	0.056	0.065	0.057	0.056	0.060	0.951	0.941	0.914	0.890	0.937	0.877
exp	0.572	0.549	0.521	0.495	0.545	0.492	0.518	0.594	0.848	0.906	0.658	0.913	0.985	0.985	0.981	0.973	0.985	0.968
pow	0.586	0.549	0.510	0.475	0.540	0.462	0.051	0.056	0.063	0.056	0.055	0.058	0.951	0.938	0.909	0.891	0.931	0.884
jump	0.444	0.416	0.376	0.342	0.410	0.334	0.417	0.489	0.721	0.775	0.544	0.785	0.908	0.889	0.864	0.834	0.885	0.831

Table 73: Size-corrected power for the T_{WZ} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.0$, and $\rho_2 = 0.6$.

	H_1: DGP 1						H_1: DGP 2						H_1: DGP 3					
M	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$	4	6	12	18	$\lfloor 4(T/100)^{1/4} \rfloor$	$\lfloor 12(T/100)^{1/4} \rfloor$
$T = 50$																		
intexp	0.012	0.014	0.017	0.019	0.015	0.015	0.061	0.059	0.058	0.060	0.058	0.054	0.042	0.044	0.043	0.046	0.042	0.044
cpr	0.010	0.014	0.022	0.020	0.008	0.020	0.047	0.047	0.049	0.049	0.046	0.046	0.043	0.043	0.046	0.047	0.044	0.045
stc	0.012	0.015	0.021	0.022	0.013	0.021	0.050	0.048	0.048	0.051	0.052	0.052	0.043	0.042	0.045	0.047	0.050	0.046
exp	0.012	0.014	0.020	0.022	0.012	0.018	0.044	0.046	0.049	0.047	0.045	0.046	0.041	0.040	0.045	0.043	0.042	0.042
pow	0.014	0.016	0.018	0.020	0.013	0.017	0.048	0.049	0.047	0.051	0.046	0.043	0.041	0.041	0.040	0.043	0.041	0.036
jump	0.013	0.016	0.025	0.019	0.014	0.021	0.045	0.044	0.048	0.048	0.043	0.042	0.040	0.040	0.040	0.044	0.041	0.041
$T = 100$																		
intexp	0.013	0.014	0.018	0.015	0.013	0.018	0.059	0.055	0.057	0.056	0.059	0.057	0.054	0.051	0.056	0.052	0.054	0.056
cpr	0.014	0.015	0.020	0.022	0.014	0.020	0.049	0.045	0.043	0.047	0.049	0.043	0.048	0.045	0.047	0.049	0.048	0.047
stc	0.014	0.014	0.019	0.021	0.014	0.019	0.050	0.047	0.047	0.045	0.050	0.047	0.048	0.048	0.049	0.043	0.048	0.049
exp	0.013	0.013	0.017	0.020	0.013	0.017	0.043	0.042	0.047	0.047	0.043	0.047	0.048	0.048	0.050	0.045	0.048	0.050
pow	0.013	0.014	0.016	0.016	0.013	0.016	0.047	0.044	0.046	0.050	0.047	0.046	0.053	0.050	0.054	0.052	0.053	0.054
jump	0.013	0.016	0.021	0.022	0.013	0.021	0.043	0.042	0.044	0.044	0.043	0.044	0.047	0.046	0.049	0.047	0.047	0.049
$T = 200$																		
intexp	0.019	0.020	0.024	0.026	0.019	0.026	0.054	0.057	0.053	0.056	0.054	0.057	0.129	0.113	0.120	0.113	0.129	0.120
cpr	0.019	0.017	0.021	0.023	0.019	0.023	0.047	0.043	0.048	0.047	0.047	0.047	0.100	0.087	0.086	0.083	0.100	0.088
stc	0.018	0.020	0.024	0.026	0.018	0.027	0.047	0.045	0.046	0.049	0.047	0.050	0.105	0.095	0.091	0.095	0.105	0.093
exp	0.019	0.021	0.023	0.024	0.019	0.023	0.041	0.037	0.043	0.044	0.041	0.044	0.115	0.104	0.102	0.099	0.115	0.103
pow	0.020	0.019	0.024	0.024	0.020	0.024	0.046	0.043	0.046	0.048	0.046	0.048	0.118	0.106	0.107	0.104	0.118	0.107
jump	0.014	0.017	0.019	0.024	0.014	0.021	0.039	0.040	0.044	0.043	0.039	0.045	0.106	0.099	0.100	0.104	0.106	0.095
$T = 500$																		
intexp	0.031	0.034	0.046	0.053	0.033	0.051	0.046	0.042	0.047	0.049	0.045	0.050	0.348	0.350	0.360	0.356	0.346	0.360
cpr	0.036	0.036	0.038	0.040	0.034	0.039	0.040	0.037	0.040	0.039	0.039	0.040	0.367	0.360	0.342	0.332	0.360	0.336
stc	0.040	0.041	0.041	0.045	0.037	0.045	0.046	0.044	0.045	0.043	0.044	0.046	0.389	0.392	0.368	0.355	0.384	0.362
exp	0.040	0.045	0.056	0.062	0.042	0.060	0.029	0.030	0.033	0.041	0.029	0.041	0.407	0.427	0.430	0.422	0.416	0.426
pow	0.032	0.031	0.037	0.039	0.030	0.040	0.047	0.045	0.045	0.047	0.045	0.047	0.348	0.353	0.345	0.342	0.351	0.341
jump	0.027	0.029	0.033	0.034	0.028	0.036	0.034	0.032	0.037	0.042	0.032	0.043	0.348	0.338	0.326	0.315	0.338	0.321
$T = 1000$																		
intexp	0.045	0.047	0.057	0.065	0.050	0.066	0.036	0.033	0.038	0.042	0.033	0.043	0.608	0.625	0.648	0.658	0.627	0.659
cpr	0.050	0.047	0.049	0.052	0.047	0.054	0.031	0.026	0.031	0.032	0.028	0.032	0.635	0.633	0.643	0.643	0.635	0.640
stc	0.063	0.067	0.077	0.084	0.068	0.086	0.041	0.038	0.039	0.040	0.038	0.040	0.685	0.684	0.705	0.697	0.688	0.700
exp	0.087	0.092	0.114	0.125	0.098	0.130	0.023	0.020	0.039	0.059	0.024	0.072	0.696	0.725	0.763	0.775	0.736	0.778
pow	0.066	0.067	0.075	0.077	0.068	0.079	0.043	0.038	0.041	0.040	0.040	0.041	0.691	0.686	0.708	0.707	0.693	0.703
jump	0.035	0.036	0.042	0.045	0.036	0.046	0.026	0.022	0.035	0.057	0.024	0.067	0.628	0.630	0.643	0.642	0.632	0.643

Table 74: Size-corrected power for the T_{WZ} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

	H_1: DGP 1						H_1: DGP 2						H_1: DGP 3					
M	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}
$T = 50$																		
intexp	0.499	0.103	0.332	0.301	0.515	0.079	0.028	0.012	0.053	0.053	0.029	0.048	0.740	0.073	0.276	0.193	0.512	0.293
cpr	0.475	0.234	0.109	0.195	0.353	0.077	0.055	0.242	0.094	0.263	0.246	0.044	0.590	0.178	0.074	0.151	0.223	0.168
stc	0.472	0.172	0.066	0.134	0.282	0.070	0.059	0.092	0.051	0.092	0.068	0.050	0.603	0.126	0.049	0.104	0.142	0.140
exp	0.527	0.323	0.212	0.378	0.491	0.076	0.150	0.601	0.329	0.836	0.775	0.054	0.639	0.287	0.143	0.362	0.445	0.234
pow	0.574	0.305	0.255	0.284	–	0.080	0.050	0.095	0.060	0.069	–	0.052	0.824	0.256	0.226	0.270	–	0.250
jump	–	0.368	0.257	0.419	–	0.086	–	0.469	0.283	0.664	–	0.062	–	0.359	0.215	0.437	–	0.265
$T = 100$																		
intexp	0.720	0.078	0.493	0.362	0.679	0.121	0.082	0.040	0.122	0.296	0.146	0.069	0.982	0.038	0.584	0.398	0.771	0.855
cpr	0.696	0.340	0.214	0.308	0.590	0.140	0.089	0.465	0.211	0.548	0.627	0.042	0.870	0.384	0.194	0.324	0.582	0.696
stc	0.690	0.259	0.145	0.240	0.485	0.141	0.057	0.132	0.052	0.111	0.147	0.046	0.879	0.258	0.097	0.214	0.423	0.634
exp	0.716	0.470	0.368	0.532	0.700	0.142	0.315	0.882	0.755	0.977	0.992	0.282	0.870	0.546	0.402	0.622	0.723	0.782
pow	0.767	0.336	0.399	0.323	–	0.139	0.055	0.169	0.066	0.092	–	0.044	0.975	0.361	0.480	0.395	–	0.760
jump	–	0.494	0.409	0.545	–	0.138	–	0.736	0.583	0.844	–	0.256	–	0.568	0.449	0.599	–	0.815
$T = 200$																		
intexp	0.855	0.098	0.661	0.432	0.799	0.210	0.229	0.052	0.409	0.791	0.512	0.285	0.998	0.048	0.745	0.533	0.866	0.947
cpr	0.832	0.478	0.382	0.428	0.793	0.262	0.178	0.823	0.537	0.772	0.950	0.188	0.964	0.573	0.401	0.479	0.825	0.901
stc	0.829	0.395	0.263	0.344	0.692	0.273	0.091	0.263	0.071	0.235	0.375	0.047	0.967	0.454	0.238	0.374	0.708	0.849
exp	0.844	0.605	0.563	0.654	0.858	0.236	0.516	0.974	0.983	0.998	1.000	0.820	0.951	0.730	0.630	0.782	0.892	0.877
pow	0.869	0.378	0.565	0.428	–	0.263	0.062	0.353	0.119	0.174	–	0.052	0.995	0.409	0.659	0.512	–	0.893
jump	–	0.590	0.594	0.628	–	0.218	–	0.931	0.856	0.911	–	0.688	–	0.688	0.647	0.697	–	0.924
$T = 500$																		
intexp	0.946	0.166	0.841	0.510	0.923	0.310	0.488	0.183	0.955	0.959	0.979	0.900	1.000	0.186	0.884	0.641	0.940	0.874
cpr	0.946	0.616	0.670	0.517	0.952	0.442	0.401	0.979	0.962	0.875	1.000	0.852	0.994	0.735	0.692	0.591	0.962	0.801
stc	0.947	0.534	0.549	0.437	0.912	0.532	0.202	0.677	0.257	0.480	0.846	0.159	0.994	0.692	0.594	0.586	0.935	0.705
exp	0.953	0.726	0.815	0.777	0.971	0.412	0.796	0.992	1.000	0.999	1.000	1.000	0.996	0.843	0.849	0.857	0.977	0.786
pow	0.958	0.476	0.789	0.534	–	0.443	0.110	0.734	0.396	0.317	–	0.162	0.999	0.529	0.842	0.628	–	0.776
jump	–	0.696	0.829	0.720	–	0.341	–	0.990	0.993	0.949	–	0.983	–	0.794	0.853	0.776	–	0.854
$T = 1000$																		
intexp	0.979	0.153	0.927	0.528	0.965	0.406	0.658	0.219	0.999	0.982	1.000	1.000	1.000	0.175	0.947	0.640	0.976	0.847
cpr	0.984	0.690	0.841	0.583	0.986	0.573	0.585	0.991	0.998	0.939	1.000	0.998	1.000	0.784	0.857	0.647	0.991	0.693
stc	0.987	0.634	0.756	0.477	0.970	0.719	0.349	0.905	0.652	0.535	0.987	0.521	0.997	0.742	0.779	0.620	0.976	0.713
exp	0.987	0.793	0.927	0.822	0.992	0.570	0.897	0.996	1.000	1.000	1.000	1.000	1.000	0.890	0.937	0.884	0.995	0.924
pow	0.985	0.531	0.911	0.572	–	0.574	0.196	0.924	0.765	0.386	–	0.540	1.000	0.571	0.935	0.669	–	0.749
jump	–	0.739	0.924	0.769	–	0.441	–	0.993	1.000	0.965	–	1.000	–	0.828	0.936	0.797	–	0.801

Table 75: Size-corrected power for the specification tests with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$.

	H_1: DGP 1						H_1: DGP 2						H_1: DGP 3					
M	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}	T_{CS}	T_{DG}	T_{WP12}	T_{WP16}	T_{WWZ}	T_{WZ}
$T = 50$																		
intexp	0.262	0.055	0.206	0.115	0.226	0.012	0.040	0.014	0.046	0.033	0.030	0.061	0.193	0.024	0.107	0.038	0.113	0.042
cpr	0.240	0.112	0.084	0.093	0.145	0.010	0.050	0.092	0.076	0.098	0.081	0.047	0.195	0.078	0.063	0.077	0.085	0.043
stc	0.239	0.105	0.064	0.093	0.123	0.012	0.060	0.069	0.048	0.079	0.049	0.050	0.201	0.082	0.052	0.080	0.066	0.043
exp	0.257	0.155	0.132	0.164	0.188	0.012	0.058	0.176	0.151	0.257	0.167	0.044	0.210	0.106	0.095	0.119	0.123	0.041
pow	0.266	0.147	0.196	0.136	–	0.014	0.051	0.062	0.056	0.058	–	0.048	0.250	0.104	0.131	0.110	–	0.041
jump	–	0.185	0.173	0.194	–	0.013	–	0.106	0.099	0.139	–	0.045	–	0.110	0.112	0.129	–	0.040
$T = 100$																		
intexp	0.532	0.059	0.343	0.155	0.398	0.013	0.043	0.017	0.069	0.065	0.048	0.059	0.777	0.020	0.372	0.104	0.460	0.054
cpr	0.461	0.173	0.149	0.126	0.254	0.014	0.050	0.141	0.112	0.157	0.142	0.049	0.607	0.159	0.126	0.123	0.195	0.048
stc	0.463	0.154	0.116	0.119	0.211	0.014	0.054	0.069	0.051	0.064	0.058	0.050	0.614	0.133	0.087	0.110	0.133	0.048
exp	0.460	0.254	0.255	0.256	0.344	0.013	0.074	0.342	0.296	0.474	0.376	0.043	0.614	0.280	0.247	0.278	0.372	0.048
pow	0.531	0.167	0.300	0.160	–	0.013	0.050	0.071	0.057	0.060	–	0.047	0.784	0.163	0.333	0.167	–	0.053
jump	–	0.258	0.293	0.275	–	0.013	–	0.200	0.206	0.294	–	0.043	–	0.286	0.303	0.319	–	0.047
$T = 200$																		
intexp	0.739	0.077	0.486	0.205	0.565	0.019	0.065	0.024	0.121	0.157	0.095	0.054	0.990	0.030	0.615	0.216	0.691	0.129
cpr	0.665	0.266	0.266	0.201	0.433	0.019	0.058	0.285	0.192	0.284	0.293	0.047	0.868	0.318	0.278	0.216	0.462	0.100
stc	0.667	0.241	0.204	0.184	0.348	0.018	0.052	0.082	0.059	0.070	0.084	0.047	0.873	0.265	0.165	0.175	0.326	0.105
exp	0.681	0.398	0.425	0.392	0.539	0.019	0.154	0.651	0.582	0.790	0.716	0.041	0.861	0.506	0.483	0.506	0.605	0.115
pow	0.727	0.192	0.448	0.199	–	0.020	0.049	0.103	0.081	0.071	–	0.046	0.974	0.210	0.556	0.251	–	0.118
jump	–	0.381	0.441	0.375	–	0.014	–	0.434	0.382	0.531	–	0.039	–	0.476	0.498	0.464	–	0.106
$T = 500$																		
intexp	0.901	0.125	0.719	0.292	0.758	0.031	0.178	0.095	0.474	0.571	0.394	0.046	1.000	0.137	0.794	0.373	0.843	0.348
cpr	0.864	0.384	0.504	0.267	0.700	0.036	0.114	0.641	0.523	0.494	0.719	0.040	0.975	0.502	0.535	0.315	0.739	0.367
stc	0.865	0.365	0.417	0.229	0.588	0.040	0.068	0.182	0.118	0.135	0.209	0.046	0.974	0.518	0.454	0.379	0.695	0.389
exp	0.871	0.535	0.674	0.556	0.779	0.040	0.398	0.909	0.962	0.961	0.992	0.029	0.980	0.684	0.734	0.687	0.831	0.407
pow	0.889	0.265	0.661	0.267	–	0.032	0.050	0.226	0.144	0.096	–	0.047	0.998	0.311	0.753	0.345	–	0.348
jump	–	0.488	0.682	0.486	–	0.027	–	0.771	0.781	0.759	–	0.034	–	0.612	0.740	0.575	–	0.348
$T = 1000$																		
intexp	0.962	0.141	0.843	0.376	0.868	0.045	0.344	0.152	0.855	0.799	0.788	0.036	1.000	0.152	0.890	0.463	0.914	0.608
cpr	0.947	0.480	0.700	0.332	0.846	0.050	0.212	0.893	0.858	0.641	0.969	0.031	0.995	0.602	0.725	0.377	0.883	0.635
stc	0.950	0.475	0.616	0.285	0.765	0.063	0.081	0.428	0.232	0.227	0.439	0.041	0.994	0.596	0.641	0.411	0.824	0.685
exp	0.954	0.637	0.839	0.643	0.917	0.087	0.627	0.967	1.000	0.981	1.000	0.023	0.997	0.780	0.870	0.744	0.928	0.696
pow	0.954	0.331	0.818	0.330	–	0.066	0.064	0.476	0.304	0.156	–	0.043	0.999	0.375	0.869	0.414	–	0.691
jump	–	0.556	0.832	0.571	–	0.035	–	0.929	0.951	0.850	–	0.026	–	0.680	0.860	0.647	–	0.628

Table 76: Size-corrected power for the specification tests with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$.

3 Fully Modified OLS Estimation and Inference for Seemingly Unrelated Cointegrating Polynomial Regressions with Common Integrated Regressors

3.1 Introduction

This paper develops two fully modified OLS (FM-OLS) type estimators for systems of seemingly unrelated cointegrating polynomial regressions (SUCPRs), i. e. , systems of regressions systems of regressions that include deterministic variables, integrated processes, integer powers of integrated processes as well as common – across equations – integrated processes and integer powers of the common integrated processes as explanatory variables. The paper thus extends the results of Wagner *et al.* (2020) to systems with common regressors.[55] We consider two estimators: One, FM-SOLS, is based on modifying the system OLS estimator and the other one, FM-SUR, is based on modifying a SUR-GLS-type estimator. With an eye towards applicability of the estimators we put particular focus on *group-wise* pooling, i. e. , estimation where, in a general formulation, subsets of the coefficients are pooled over possibly differing subsets of cross-section members. As is common – in particular in the linear – cointegration literature, the stationary errors are allowed to be serially correlated and the regressors are allowed to be endogenous. We also allow for dynamic cross-sectional correlation of both errors and regressors.

The SUR cointegration literature differs in several respects from the classical SUR literature, initiated by Zellner (1962) and extended to nonlinear equations by, e. g. , Gallant (1975), where the regressors are assumed to be strictly exogenous and stationary, and the stationary errors serially uncorrelated. In our cointegration context, with in general serially correlated errors, thus one change is that estimated long-run variance matrices rather than estimates of contemporaneous variance matrices are used in SUR estimation. The second change is that allowing for regressor endogeneity in a cointegration setting necessitates the usage of modified least squares estimators –

[55] Note furthermore that – for completeness – this paper does not consider a stylized system with only one integrated regressor and its powers per equation, but the general multiple regressors formulation. This, of course, increases notational complexity, but on the other hand paves the way for more complex applications, in particular in conjunction with the MATLAB code available upon request.
FM-OLS has originally been developed for linear cointegrating relations in Phillips and Hansen (1990) and has been extended to single equation cointegrating polynomial regressions (CPRs) in Wagner and Hong (2016). Given that Wagner *et al.* (2020) extend the SUR cointegration literature (see Park and Ogaki, 1991; Moon, 1999; Mark *et al.*, 2005; Moon and Perron, 2005) from the linear to the CPR case, the innovation of this paper is the inclusion of common integrated regressors and their powers; importantly so in a straightforward way that allows to perform fully modified estimation in presence of common regressors based on a (to the best of our knowledge) unnoticed observation.

in our contribution of the FM-OLS type – to allow for asymptotically valid normal or chi-squared inference.

Our primary motivation is the analysis of the environmental Kuznets curve (EKC). The EKC hypothesis postulates an inverted U-shaped relationship between the level of economic development and pollution or emissions.[56] Surveys discussing the links between economic growth and the environment are provided by Brock and Taylor (2005) or Kijima *et al.* (2010). A sizeable part of the empirical EKC literature uses unit root and cointegration techniques, both in (single) time series and panel data settings. For a discussion of the usage of unit root and cointegration techniques and problems emanating from this usage in the time series case see, e. g. , Wagner (2015). In addition to the EKC, we are also investigating the material Kuznets curve (MKC) hypothesis, which postulates an inverted U-shaped relationship between economic development and the use of materials. The MKC literature has not yet grown as large as the EKC literature although the same challenges are faced with respect to unit roots and cointegration techniques. An introduction to the MKC with a cointegrating analysis is provided by, e. g. , Grabarczyk *et al.* (2018). In our analysis we focus on the case of panel data with *small* cross-sectional dimension, which are regularly used in the EKC literature, but also in many other fields using panels of macro time series. It is widely acknowledged in the literature that both cross-sectional heterogeneity (see, e. g. , Vollebergh *et al.*, 2009) and cross-sectional dependencies (see, e. g. , Mazzanti and Musolesi, 2013) are very likely important aspects when using such panels of macro time series. As noted, e. g. , by Pesaran (2006, p. 967–968), the "standard" approach to model systems of cointegrating relationships in case of a small cross-sectional dimension (by which he means smaller than ten) is to consider the relationships jointly as a system of seemingly unrelated relationships. Doing so has the advantage that it allows one to handle both cross-sectional heterogeneity as well as cross-sectional dependencies. Consequently, we follow this "avenue" and develop estimation and inference techniques to combine cointegrating polynomial regressions to a system. Importantly, our approach allows us to test general hypotheses concerning group-wise pooling. Pooled estimation, when appropriate, leads to considerable efficiency gains (as documented in the simulations), but will lead to misleading results when used uncritically or incorrectly, as is often done in, e. g. , the panel EKC literature.

[56]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 (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.

We perform a simulation study to assess the performance of the proposed estimators and tests based on them. The evaluation criteria for the estimators are bias and root mean squared error (RMSE). The test performance is evaluated with both null rejection probabilities as well as with "size-corrected" power against a grid of alternatives. It turns out that the FM-CPR estimator performs best in unrestricted estimation, with often leading to the best or comparable results in all the evaluation criteria. The developed system estimators FM-SUR and FM-SOLS come into play when applying group-wise pooling. Pooling the coefficients for the common integrated regressor and its square across all cross-section members, or even pooling the coefficients for all stochastic regressors, leads to significant improvements in terms of RMSE and size-corrected power for those estimators. There, the FM-SOLS estimator is slightly ahead the FM-SUR estimator in terms of RMSE, null rejection probabilities and size-corrected power.

Finally, we apply the developed methodologies to study EKC-type and MKC-type relationships for eight OECD countries for the period 1927–2006. For the EKCs, we consider relationships between GDP and the two emissions carbon dioxide and sulfur dioxide. For the MKC, we consider relationships between GDP and the use of the three metals aluminum, lead and zinc. We estimate systems of equations on both country level and substance level, i. e. , emission or material use. Similarities in the estimated parameters can be found along both levels. We observe for several countries similar parameter estimates for emissions or materials, especially for lead and zinc, indicating a similar behavior in the material use during the analysis period in the examined countries. Additionally, we find similarities on the substance level, with parameter estimates close to each other for different countries, giving evidence on similar intensity of use paths for certain materials, or emissions across countries. Based on those indications and backed by tests on poolability, we perform group-wise pooling for certain materials and countries. The group-wise pooled estimation leads to very similar results in terms of parameter estimates as the unrestricted individual or system estimation, while having the additional benefits of estimating less parameters and higher estimator precision through the increased number of observations for that parameter.

The paper is organized as follows: In Section 3.2 we introduce the model and underlying assumptions, together with the main theorems. Section 3.3 provides the performance of estimators and tests in finite sample simulations. In Section 3.4 we analize EKCs and MKCs using the proposed estimators and tests. Section 3.5 briefly summarizes and concludes.

We use the following notation: Weak convergence is signified by $\Rightarrow$ for $T \to \infty$, the integer part

of $x \in \mathbb{R}$ is denoted by $\lfloor x \rfloor$ and diag$(\cdot)$ denotes a (block-)diagonal matrix with entries specified throughout. For a vector $x \coloneqq (x_i)_{i=1}^n$ we denote by $\|x\|^2 \coloneqq \sum_{i=1}^n x_i^2$ and for a matrix M we denote by $\|M\| \coloneqq \sup_x \frac{\|Mx\|}{\|x\|}$. For a square matrix A we denote its determinant by $\det(A)$. We denote the m-dimensional identity matrix by I_m and with $0_{m\times n}$ an $(m \times n)$-matrix with all entries equal to zero. Let $\mathbf{1}_s \coloneqq [1, \dots, 1]' \in \mathbb{R}^s$ and with $e_{i,n}$ being the ith unit vector in $\mathbb{R}^n$. For matrices M we denote with $M^{i,j}$ the (i,j)-element, the ith row with $M^{i,\cdot}$ and the jth column with $M^{\cdot,j}$. For a set $\mathcal{S}$, $|\mathcal{S}|$ denotes the number of elements of $\mathcal{S}$. With $\otimes$ we denote the Kronecker product, $\mathbb{E}(\cdot)$ denotes the expected value and L denotes the backward-shift operator, i. e. , $L\{z_t\}_{t\in\mathbb{Z}} \coloneqq \{z_{t-1}\}_{t\in\mathbb{Z}}$. Brownian motions are denoted with $B(r)$, with covariance matrices specified in the context.

3.2 Seemingly Unrelated Cointegrating Polynomial Regressions with Common Integrated Regressors

3.2.1 Setup and Assumptions

We consider the following system of equations:

$$\begin{aligned} y_{i,t} &= D_{i,t}'\theta_{D,i} + \sum_{j=1}^{m_i} X_{i,j,t}'\theta_{X,i,j} + \sum_{j=1}^{m} X_{\mathrm{C},i,j,t}'\theta_{\mathrm{C},i,j} + u_{i,t} \qquad t = 1,\dots,T,\ i = 1,\dots,N, \qquad (144)\\ &= Z_{i,t}'\theta_i + u_{i,t}, \end{aligned}$$

with $D_{i,t} \coloneqq [1, t, \dots, t^{q_i}]'$, $X_{i,j,t} \coloneqq [x_{i,j,t}, x_{i,j,t}^2, \dots, x_{i,j,t}^{p_{i,j}}]'$, $X_{\mathrm{C},i,j,t} \coloneqq [x_{\mathrm{C},j,t}, x_{\mathrm{C},j,t}^2, \dots, x_{\mathrm{C},j,t}^{c_{i,j}}]'$, $Z_{i,t} \coloneqq [D_{i,t}', X_{i,1,t}', \dots, X_{i,m_i,t}', X_{\mathrm{C},i,1,t}', \dots X_{\mathrm{C},i,m,t}']'$, $\theta_i \coloneqq [\theta_{D,i}', \theta_{X,i,1}', \dots, \theta_{X,i,m_i}', \theta_{\mathrm{C},i,1}', \dots, \theta_{\mathrm{C},i,m}']'$. Each equation i contains m_i individual specific integrated regressors and up to m common integrated regressors. All integrated regressors are divided into common integrated regressors and individual specific regressors, depending on whether they appear in only a single equation or in at least two equations. A value of $c_{i,j} = 0$ indicates that the j-th common integrated regressor is not included in equation i and thus $X_{\mathrm{C},i,j,t}$ is omitted from $Z_{i,t}$ and $\theta_{\mathrm{C},i,j}$ is omitted from θ_i. Furthermore, we define $d_i \coloneqq 1 + q_i + \sum_{j=1}^{m_i} p_{i,j} + \sum_{j=1}^{m} c_{i,j}$, $d \coloneqq \sum_{i=1}^N d_i$ and $u_t = [u_{1,t}, \dots, u_{N,t}]'$. The N equations given in (144) can be written as

$$y_t = Z_t'\theta + u_t, \qquad (145)$$

with

$$y_t \coloneqq \begin{bmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{bmatrix} \in \mathbb{R}^N, \quad Z_t \coloneqq \begin{bmatrix} Z_{1,t} & & \\ & \ddots & \\ & & Z_{N,t} \end{bmatrix} \in \mathbb{R}^{d\times N}, \quad u_t \coloneqq \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{N,t} \end{bmatrix} \in \mathbb{R}^N, \qquad (146)$$

and with $\theta \coloneqq [\theta_1', \ldots, \theta_N']'$. Stacking all T observations for the above system, we obtain

$$y = Z\theta + u, \tag{147}$$

with

$$y \coloneqq \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} \in \mathbb{R}^{NT}, \quad Z \coloneqq \begin{bmatrix} Z_1' \\ \vdots \\ Z_T' \end{bmatrix} \in \mathbb{R}^{NT \times d} \quad \text{and } u \coloneqq \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \in \mathbb{R}^{NT}. \tag{148}$$

Before stating the assumptions on the error processes, we furthermore define $x_{i,t} \coloneqq [x_{i,1,t}, \ldots, x_{i,m_i,t}]'$, $x_{\mathrm{C},t} \coloneqq [x_{\mathrm{C},1,t}, \ldots, x_{\mathrm{C},m,t}]'$, $x_t \coloneqq [x_{1,t}', \ldots, x_{N,t}', x_{\mathrm{C},t}']'$ and $v_t \coloneqq \Delta x_t$. The precise assumptions on the error process $\{\xi_t\}_{t\in\mathbb{Z}} \coloneqq \{[u_t', v_t']'\}_{t\in\mathbb{Z}}$ are then as follows:

Assumption 42 *The process $\{\xi_t\}_{t\in\mathbb{Z}}$ is generated as*

$$\xi_t = C(L)\xi_t^0 = \sum_{j=0}^{\infty} C_j \xi_{t-j}^0,$$

with the conditions $\sum_{j=0}^{\infty} j\|C_j\| < \infty$ and $\det(C(1)) \neq 0$. The process $\{\xi_t^0\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic martingale difference sequence with respect to the natural filtration $\mathcal{F}_t = \sigma(\{\xi_t^0\}_{-\infty}^t)$ with positive definite conditional variance matrix $\Sigma^0 \coloneqq \mathbb{E}(\xi_t^0(\xi_t^0)'|\mathcal{F}_{t-1})$ and $\sup_{t\in\mathbb{Z}} \mathbb{E}(\|\xi_t^0\|^r|\mathcal{F}_{t-1}) < \infty$ a.s. for some $r > 4$.

Assumption 42 is sufficient for a central limit theorem to hold for $\{\xi_t\}_{t\in\mathbb{Z}}$, such that

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \xi_t \Rightarrow B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \Omega^{1/2} W(r), \quad 0 \leq r \leq 1, \tag{149}$$

with long-run covariance matrix $\Omega \coloneqq \sum_{j=-\infty}^{\infty} \mathbb{E}(\xi_{t-j}\xi_t')$ and $W(r)$ an $(N+\sum_{i=1}^N m_i+m)$-dimensional standard Brownian motion. Furthermore, the Brownian motion $B_v(r)$ is given by $B_v(r) \coloneqq [B_{v,1}(r)', \ldots, B_{v,N}(r)', B_{v,\mathrm{C}}(r)']'$ with $B_{v,i}(r) \coloneqq [B_{v,i,1}(r), \ldots, B_{v,i,m_i}(r)]'$ and $B_{v,\mathrm{C}}(r) \coloneqq [B_{v,\mathrm{C},1}(r), \ldots, B_{v,\mathrm{C},m}(r)]'$. We define the half long-run covariance matrix $\Delta \coloneqq \sum_{j=0}^{\infty} \mathbb{E}(\xi_{t-j}\xi_t')$, with both covariance matrices partioned according to the partitioning of $B(r)$, i. e. ,

$$\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_{uu} & \Delta_{uv} \\ \Delta_{vu} & \Delta_{vv} \end{bmatrix}. \tag{150}$$

To establish the asymptotic behavior of the estimators, which we will introduce later, scaling matrices are required. Denoting with

$$G_{D,i}(T) \coloneqq \mathrm{diag}(T^{-1/2}, \ldots, T^{-(q_i+1/2)}), \tag{151}$$

$$G_{X,i,j}(T) \coloneqq \mathrm{diag}(T^{-1}, \ldots, T^{-(p_{i,j}+1)/2}), \tag{152}$$

$$G_{\mathrm{C},i,j}(T) \coloneqq \mathrm{diag}(T^{-1}, \ldots, T^{-(c_{i,j}+1)/2}), \tag{153}$$

we obtain

$$T^{1/2}G_{D,i}D_{i,\lfloor rT\rfloor} \Rightarrow D_i(r) \coloneqq [1, r, \ldots, r^{q_i}]', \tag{154}$$

for $0 \leq r \leq 1$ and with Assumption 42 we obtain

$$T^{1/2}G_{X,i,j}X_{i,j,\lfloor rT\rfloor} \Rightarrow \mathbf{B}_{v,i,j}(r) \coloneqq [B_{v,i,j}(r), B^2_{v,i,j}(r), \ldots, B^{p_{i,j}}_{v,i,j}(r)]', \tag{155}$$

$$T^{1/2}G_{\mathrm{C},i,j}X_{\mathrm{C},i,j,\lfloor rT\rfloor} \Rightarrow \mathbf{B}_{v,\mathrm{C},i,j}(r) \coloneqq [B_{v,\mathrm{C},j}(r), B^2_{v,\mathrm{C},j}(r), \ldots, B^{c_{i,j}}_{v,\mathrm{C},j}(r)]', \tag{156}$$

for $0 \leq r \leq 1$. For later usage, we define the the quantities

$$G_i(T) \coloneqq \operatorname{diag}(G_{D,i}(T)G_{X,i,1}(T), \ldots, G_{X,i,m_i}(T), G_{\mathrm{C},i,1}(T), \ldots, G_{\mathrm{C},i,m}(T)), \tag{157}$$

$$G(T) \coloneqq \operatorname{diag}(G_1(T), \ldots, G_N(T)), \tag{158}$$

$$J_i(r) \coloneqq [D_i(r)', \mathbf{B}_{v,i,1}(r)', \ldots, \mathbf{B}_{v,i,m_i}(r)', \mathbf{B}_{v,\mathrm{C},i,1}(r)', \ldots, \mathbf{B}_{v,\mathrm{C},i,m}(r)']', \tag{159}$$

$$J(r) \coloneqq \operatorname{diag}(J_1(r), \ldots, J_N(r)). \tag{160}$$

3.2.2 Fully Modified OLS Estimation

We consider similar estimators as Wagner *et al.* (2020), now extended in such a way they include common integrated regressors in the model. The ordinary least squares (OLS) estimator is given by

$$\hat{\theta}_{\mathrm{OLS}} \coloneqq (Z'Z)^{-1}Z'y, \tag{161}$$

and a feasible seemingly unrelated regression (SUR) estimator, with the estimated long-run covariance matrix $\hat{\Omega}_{uu}$ as weighting matrix, is given by

$$\hat{\theta}_{\mathrm{MSUR}} \coloneqq \left(Z'(I_T \otimes \hat{\Omega}^{-1}_{uu})Z\right)^{-1}\left(Z'(I_T \otimes \hat{\Omega}^{-1}_{uu})y\right). \tag{162}$$

Both estimators are consistent with their limiting distributions contaminated by second order bias terms. To obtain estimators with zero mean Gaussian mixture limiting distributions, which allow for asymptotic standard inference, we propose fully modified type corrected versions of the two estimators defined above. The fully modified type corrections consists of two transformations, similarly to the cointegrating linear case as in Phillips and Hansen (1990). The first transformation is given by $y^+_t \coloneqq y_t - \hat{\Omega}_{uv}\hat{\Omega}^{-1}_{vv}\Delta x_t$. For the second transformation, we define for the OLS based

estimator the correction term $\hat{A} \coloneqq [\hat{A}_1', \ldots, \hat{A}_N']'$, with

$$\hat{A}_i \coloneqq \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \hat{A}_{i,1} \\ \vdots \\ \hat{A}_{i,m_i} \\ \hat{A}_{\mathrm{C},i,1} \\ \vdots \\ \hat{A}_{\mathrm{C},i,m} \end{bmatrix}, \quad \hat{A}_{i,j} \coloneqq (\hat{\Delta}_{vu}^{+})^{k_{i-1}+j,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{i,j,t} \\ \vdots \\ p_{i,j} \sum_{t=1}^{T} x_{i,j,t}^{p_{i,j}-1} \end{bmatrix}, \quad \hat{A}_{\mathrm{C},i,j} \coloneqq (\hat{\Delta}_{vu}^{+})^{k_N+j,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{\mathrm{C},j,t} \\ \vdots \\ c_{i,j} \sum_{t=1}^{T} x_{\mathrm{C},j,t}^{c_{i,j}-1} \end{bmatrix}, \tag{163}$$

where $(\hat{\Delta}_{vu}^{+})^{i,j}$ is a consistent estimator of $(\Delta_{vu}^{+})^{i,j} \coloneqq \Delta_{vu}^{i,j} - \Delta_{vv}^{i,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,j}$ and $k_i \coloneqq \sum_{s=1}^{i} m_s$, with $k_0 = 0$. For the SUR based estimator, we define for the second transformation the correction term $\tilde{A}^* \coloneqq [\tilde{A}_1^{*\prime}, \ldots, \tilde{A}_N^{*\prime}]'$, with

$$\tilde{A}_i^* \coloneqq \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \tilde{A}_{i,1}^* \\ \vdots \\ \tilde{A}_{i,m_i}^* \\ \tilde{A}_{\mathrm{C},i,1}^* \\ \vdots \\ \tilde{A}_{\mathrm{C},i,m}^* \end{bmatrix}, \quad \tilde{A}_{i,j}^* \coloneqq (\hat{\Delta}_{vu}^{+})^{k_{i-1}+j,\cdot} (\hat{\Omega}_{u\cdot v}^{-1})^{\cdot,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{i,j,t} \\ \vdots \\ p_{i,j} \sum_{t=1}^{T} x_{i,j,t}^{p_{i,j}-1} \end{bmatrix}, \tag{164}$$

$$\tilde{A}_{\mathrm{C},i,j}^* \coloneqq (\hat{\Delta}_{vu}^{+})^{k_N+j,\cdot} (\hat{\Omega}_{u\cdot v}^{-1})^{\cdot,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{\mathrm{C},j,t} \\ \vdots \\ c_{i,j} \sum_{t=1}^{T} x_{\mathrm{C},j,t}^{c_{i,j}-1} \end{bmatrix}, \tag{165}$$

where $\hat{\Omega}_{u\cdot v}$ is a consistent estimator of $\Omega_{u\cdot v} \coloneqq \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$. We set $\hat{A}_{i,j} = \tilde{A}_{i,j}^* = \emptyset$ for those i, j, where $p_{i,j} = 0$, and $\hat{A}_{\mathrm{C},i,j} = \tilde{A}_{\mathrm{C},i,j}^* = \emptyset$ for those i, j, where $c_{i,j} = 0$.

Proposition 4 *Let the data be generated according to* (144) *with Assumption 42 in place. Furthermore, assume that based on OLS residuals all long-run covariances are estimated consistently. Using the correction terms $\hat{A}$ and $\tilde{A}^*$, the fully modified systems OLS (FM-SOLS) and the fully modified SUR (FM-SUR) estimators are given by*

$$\hat{\theta} \coloneqq \left(Z'Z\right)^{-1} \left(Z'y^{+} - \hat{A}\right), \tag{166}$$

$$\tilde{\theta} \coloneqq \left(Z'(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1})Z\right)^{-1} \left(Z'(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1})y^{+} - \tilde{A}^*\right), \tag{167}$$

with $y^+ \coloneqq [y_1^{+\prime}, \ldots, y_T^{+\prime}]'$. *For* $T \to \infty$ *it holds that*

$$G^{-1}\left(\hat{\theta} - \theta\right) \Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \int_0^1 J(r)dB_{u\cdot v}(r), \tag{168}$$

$$G^{-1}\left(\tilde{\theta} - \theta\right) \Rightarrow \left(\int_0^1 J(r)\Omega_{u\cdot v}^{-1}J(r)'dr\right)^{-1} \int_0^1 J(r)\Omega_{u\cdot v}^{-1}dB_{u\cdot v}(r), \tag{169}$$

where $B_{u\cdot v} \coloneqq B_u(r) - \Omega_{uv}\Omega_{vv}^{-1}B_v(r)$ *is an* N*-dimensional Brownian motion with covariance matrix* $\Omega_{u\cdot v}$.

The limiting distributions given in (168) and (169) are zero-mean Gaussian mixture, since $B_{u\cdot v}(r)$ is independent of $B_v(r)$ by construction. These limiting distributions form the basis for asymptotic chi-squared inference. We consider Wald-type tests for testing the null hypothesis of linear restrictions on the parameter vector θ.

Proposition 5 *Let the data be generated according to* (144) *with Assumption 42 in place. Furthermore, assume that based on OLS residuals all long-run covariances are estimated consistently. Consider* s *linearly independent restrictions collected in the null hypothesis* $H_0\colon R\theta = r$ *with* $R \in \mathbb{R}^{s\times d}$ *of full row rank* s *and* $r \in \mathbb{R}^s$. *Suppose that there exists a sequence of full rank matrices* $G_R = G_R(T)$, *such that* $\lim_{T\to\infty} G_R R G = R^*$ *with* $R^* \in \mathbb{R}^{s\times d}$ *of full row rank* s. *Then it holds under the null hypothesis that the Wald-type statistics*

$$\hat{T}_W \coloneqq \left(R\hat{\theta} - r\right)' \left[R\left(Z'Z\right)^{-1} Z' \left(I_T \otimes \hat{\Omega}_{u\cdot v}\right) Z \left(Z'Z\right)^{-1} R'\right]^{-1} \left(R\hat{\theta} - r\right), \tag{170}$$

$$\tilde{T}_W \coloneqq \left(R\tilde{\theta} - r\right)' \left[R\left(Z' \left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) Z\right)^{-1} R'\right]^{-1} \left(R\tilde{\theta} - r\right), \tag{171}$$

are asymptotically chi-squared distributed with s *degrees of freedom.*

3.2.3 Specification Testing Based on Augmented and Auxiliary Regressions

When Equation (144) is misspecified, the error process $\{u_t\}_{t\in\mathbb{Z}}$ is nonstationary. This can arise for two reasons: First, the functional form is misspecified but some other form of nonlinear cointegration between y_t and x_t (and potentially other variables) prevails. Second, there is no cointegration between the variables considered. Similar to Wagner and Hong (2016), we consider RESET-type tests as specification tests for testing the null hypothesis of correct specification of (144), using additional regressors. We consider both a Wald-type test based on augmented regression and a Lagrange multiplier (LM)-type test using an auxiliary regression. For

each $i = 1, \ldots, N$, we define $\bar{D}_{i,t} \coloneqq [t^{q_i+1}, \ldots, t^{q_i+\bar{q}_i}]'$, $\bar{X}_{i,j,t} \coloneqq [x_{i,j,t}^{p_{i,j}+1}, \ldots, x_{i,j,t}^{p_{i,j}+\bar{p}_{i,j}}]'$ for $j = 1, \ldots, m_i$, $\bar{X}_{\mathrm{C},i,j,t} \coloneqq [x_{\mathrm{C},j,t}^{c_{i,j}+1}, \ldots, x_{\mathrm{C},j,t}^{c_{i,j}+\bar{c}_{i,j}}]'$ for $j = 1, \ldots, m$, $Q_{i,j,t} \coloneqq [q_{i,j,t}, q_{i,j,t}^2, \ldots, q_{i,j,t}^{s_{i,j}}]'$ for $j = 1, \ldots, \bar{m}_i$, $Q_{\mathrm{C},i,j,t} \coloneqq [q_{\mathrm{C},j,t}, q_{\mathrm{C},j,t}^2, \ldots, q_{\mathrm{C},j,t}^{\varsigma_{i,j}}]'$ for $j = 1, \ldots, \bar{m}$. Furthermore, we define $F_{i,t} \coloneqq [\bar{D}_{i,t}', \bar{X}_{i,1,t}', \ldots, \bar{X}_{i,m_i,t}', \bar{X}_{\mathrm{C},i,1,t}', \ldots, \bar{X}_{\mathrm{C},i,m,t}', Q_{i,1,t}', \ldots, Q_{i,n_i,t}', Q_{\mathrm{C},i,1,t}', \ldots, Q_{\mathrm{C},i,n,t}']'$, $\bar{d}_i \coloneqq \bar{q}_i + \sum_{j=1}^{m_i} \bar{p}_{i,j} + \sum_{j=1}^{m} \bar{c}_{i,j} + \sum_{j=1}^{\bar{m}_i} s_{i,j} + \sum_{j=1}^{\bar{m}} \varsigma_{i,j}$, $\bar{d} \coloneqq \sum_{i=1}^{N} \bar{d}_i$ and

$$F \coloneqq \begin{bmatrix} F_1' \\ \vdots \\ F_T' \end{bmatrix} \in \mathbb{R}^{NT \times \bar{d}}, \text{ with } F_t \coloneqq \begin{bmatrix} F_{1,t} & & \\ & \ddots & \\ & & F_{N,t} \end{bmatrix} \in \mathbb{R}^{\bar{d} \times N}. \tag{172}$$

The augmented regression can be written as follows and includes higher-order deterministic trends $\bar{D}_{i,t}$, higher-order polynomial powers of the individual specific integrated regressors $x_{i,j,t}$ and of the common integrated regressors $x_{\mathrm{C},j,t}$, and polynomial powers of additional individual specific integrated regressors $q_{i,j,t}$ and of additional common integrated regressors $q_{\mathrm{C},j,t}$:

$$y = Z\theta + F\theta_F + \psi, \tag{173}$$

with $\psi \coloneqq [\psi_1, \ldots, \psi_T]'$. We propose a modification of Assumption 42, to account for the additional integrated regressors. For this purpose, we define $q_{i,t} \coloneqq [q_{i,1,t}, \ldots, q_{i,\bar{m}_i,t}]'$, $q_{\mathrm{C},t} \coloneqq [q_{\mathrm{C},1,t}, \ldots, q_{\mathrm{C},\bar{m},t}]'$ and $q_t^* \coloneqq [q_{1,t}', \ldots, q_{N,t}', q_{\mathrm{C},t}']'$.

Assumption 43 *When considering additional individual specific integrated regressors $q_{i,j,t}$ and additional common integrated regressors $q_{C,j,t}$ both with higher order powers, define $\tilde{v}_t \coloneqq [v_t', v_t^{*\prime}]' = [\Delta x_t', \Delta q_t^{*\prime}]'$, with $v_t^* = \Delta q_t^*$. Assumption 42 is extended, such that it is fulfilled for the extended process $\{\tilde{\xi}_t\}_{t\in\mathbb{Z}} \coloneqq \{[u_t', \tilde{v}_t']'\}_{t\in\mathbb{Z}}$, with $\tilde{\xi}_t = \tilde{C}(L)\tilde{\xi}_t^0$, and $\tilde{C}(L)$ and $\tilde{\xi}_t^0$ extended accordingly.*

Assumption 43 is sufficient for a central limit theorem to hold for the extended process $\{\tilde{\xi}_t\}_{t\in\mathbb{Z}}$, such that

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \tilde{\xi}_t \Rightarrow \tilde{B}(r) = \begin{bmatrix} B_u(r) \\ B_{\tilde{v}}(r) \end{bmatrix} = \begin{bmatrix} B_u(r) \\ B_v(r) \\ B_{v^*}(r) \end{bmatrix} = \tilde{\Omega}^{1/2} \tilde{W}(r), \quad 0 \le r \le 1, \tag{174}$$

with extended long-run covariance matrix $\tilde{\Omega} \coloneqq \sum_{j=-\infty}^{\infty} \mathbb{E}(\tilde{\xi}_t \tilde{\xi}_{t-j}')$ and $\tilde{W}(r)$ an $(N + \sum_{i=1}^{N} m_i + m + \sum_{i=1}^{N} \bar{m}_i + \bar{m})$-dimensional standard Brownian motion. The additional Brownian motion $B_{v^*}(r)$, compared to (149), is given by $B_{v^*}(r) \coloneqq [B_{v^*,1}(r)', \ldots, B_{v^*,N}(r)', B_{v^*,\mathrm{C}}(r)']'$ with $B_{v^*,i}(r) \coloneqq [B_{v^*,i,1}(r), \ldots, B_{v^*,i,\bar{m}_i}(r)]'$ and $B_{v^*,\mathrm{C}}(r) \coloneqq [B_{v^*,\mathrm{C},1}(r), \ldots, B_{v^*,\mathrm{C},\bar{m}}(r)]'$. We define the extended half

long-run covariance matrix $\tilde{\Delta} \coloneqq \sum_{j=0}^{\infty} \mathbb{E}(\tilde{\xi}_{t-j}\tilde{\xi}_t')$, with both covariance matrices partioned according to the partitioning of $\tilde{B}(r)$, i. e. ,

$$\tilde{\Omega} = \begin{bmatrix} \Omega_{uu} & \Omega_{u\tilde{v}} \\ \Omega_{\tilde{v}u} & \Omega_{\tilde{v}\tilde{v}} \end{bmatrix}, \quad \tilde{\Delta} = \begin{bmatrix} \Delta_{uu} & \Delta_{u\tilde{v}} \\ \Delta_{\tilde{v}u} & \Delta_{\tilde{v}\tilde{v}} \end{bmatrix}. \tag{175}$$

Proposition 6 *Let the data be generated according to* (144) *with Assumptions* (42) *and* (43) *in place. Denote with* $\hat{\theta}_F$ *the FM-SOLS estimator of* θ_F *in* (173) *and with* $\tilde{\theta}_F$ *the FM-SUR estimator of* θ_F*. Furthermore, define* $\hat{F} \coloneqq F - Z(Z'Z)^{-1}Z'F$ *and* $\tilde{F} \coloneqq F - Z\left(Z'\left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}^{-1}\right)Z\right)^{-1} Z'\left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}^{-1}\right)F$ *and let* $\hat{\Omega}_{u\cdot\tilde{v}}$ *be a consistent estimator of* $\Omega_{u\cdot\tilde{v}} \coloneqq \Omega_{uu} - \Omega_{u\tilde{v}}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}$*. Then it holds that the Wald-type test statistics for the null hypothesis* $H_0 \colon \theta_F = 0$ *in* (173)*, given by*

$$\hat{T}_W^{spec} \coloneqq \hat{\theta}_F' \left(\left(\hat{F}'\hat{F}\right)^{-1} \hat{F}' \left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}\right) \hat{F} \left(\hat{F}'\hat{F}\right)^{-1} \right)^{-1} \hat{\theta}_F, \tag{176}$$

$$\tilde{T}_W^{spec} \coloneqq \tilde{\theta}_F' \left(\tilde{F}' \left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}^{-1}\right) \tilde{F}\right) \tilde{\theta}_F, \tag{177}$$

are under the null hypothesis asymptotically chi-squared distributed with $\bar{d}$ degrees of freedom.

We next consider the Lagrange multiplier (LM)-type specification test based on auxiliary regression. For this purpose, we regress the additional regressors F, orthogonalized with respect to the origin regressors Z, i. e. , $\hat{F}$, on the FM residuals. For this purpose, both can serve as FM residuals, either the FM-SOLS residuals or the FM-SUR residuals of (147), with both leading to the same limiting distribution of the FM-OLS-type estimator in (178), as becomes clear in Proof of Proposition 7. For simplicity, we continue with the FM-SOLS residuals $\hat{u}_t^+ \coloneqq y_t^+ - Z_t'\hat{\theta}$ and $\hat{u}^+ \coloneqq [\hat{u}_1^{+\prime}, \ldots, \hat{u}_T^{+\prime}]'$. This leads to the auxiliary regression given by

$$\hat{u}^+ = \hat{F}\theta_{\hat{F}} + \psi. \tag{178}$$

The null hypothesis of the LM-type test is given by $H_0 \colon \theta_{\hat{F}} = 0$ in the auxiliary regression (178). Prior to the LM-type test statistic and its asymptotic distribution, we define a few correction terms.

Denote with $A^{Z*} \coloneqq [A_1^{Z*\prime}, \dots, A_N^{Z*\prime}]'$ where

$$A_i^{Z*} \coloneqq \begin{bmatrix} 0_{(q_i+1)\times 1} \\ A_{i,1}^{Z*} \\ \vdots \\ A_{i,m_i}^{Z*} \\ A_{\mathrm{C},i,1}^{Z*} \\ \vdots \\ A_{\mathrm{C},i,m}^{Z*} \end{bmatrix}, \quad A_{i,j}^{Z*} \coloneqq \left(\hat{\Delta}_{\tilde{v}u} - \hat{\Delta}_{\tilde{v}\tilde{v}}\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}\right)^{k_{i-1}+j,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{i,j,t} \\ \vdots \\ p_{i,j}\sum_{t=1}^{T} x_{i,j,t}^{p_{i,j}-1} \end{bmatrix}, \tag{179}$$

$$A_{\mathrm{C},i,j}^{Z*} \coloneqq \left(\hat{\Delta}_{\tilde{v}u} - \hat{\Delta}_{\tilde{v}\tilde{v}}\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}\right)^{k_N+j,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{\mathrm{C},j,t} \\ \vdots \\ c_{i,j}\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{c_{i,j}-1} \end{bmatrix}, \tag{180}$$

and furthermore define $A^{F*} \coloneqq [A_1^{F*\prime}, \dots, A_N^{F*\prime}]'$, with

$$A_i^{F*} \coloneqq [0'_{\bar{q}_i\times 1}, A_{\bar{X},i,1}^{F*\prime}, \dots, A_{\bar{X},i,m_i}^{F*\prime}, A_{\bar{X},\mathrm{C},i,1}^{F*\prime}, \dots A_{\bar{X},\mathrm{C},i,m}^{F*\prime}, A_{Q,i,1}^{F*\prime}, \dots, A_{Q,i,\bar{m}_i}^{F*\prime}, A_{Q,\mathrm{C},i,1}^{F*\prime}, \dots, A_{Q,\mathrm{C},i,\bar{m}}^{F*\prime}]', \tag{181}$$

where

$$A_{\bar{X},i,j}^{F*} \coloneqq \left(\hat{\Delta}_{\tilde{v}u} - \hat{\Delta}_{\tilde{v}\tilde{v}}\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}\right)^{k_{i-1}+j,i} \begin{bmatrix} (p_{i,j}+1)\sum_{t=1}^{T} x_{i,j,t}^{p_{i,j}} \\ \vdots \\ (p_{i,j}+\bar{p}_{i,j})\sum_{t=1}^{T} x_{i,j,t}^{p_{i,j}+\bar{p}_{i,j}-1} \end{bmatrix}, \tag{182}$$

$$A_{\bar{X},\mathrm{C},i,j}^{F*} \coloneqq \left(\hat{\Delta}_{\tilde{v}u} - \hat{\Delta}_{\tilde{v}\tilde{v}}\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}\right)^{k_N+j,i} \begin{bmatrix} (c_{i,j}+1)\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{c_{i,j}} \\ \vdots \\ (c_{i,j}+\bar{c}_{i,j})\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{c_{i,j}+\bar{c}_{i,j}-1} \end{bmatrix}, \tag{183}$$

$$A_{Q,i,j}^{F*} \coloneqq \left(\hat{\Delta}_{\tilde{v}u} - \hat{\Delta}_{\tilde{v}\tilde{v}}\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}\right)^{\bar{k}_{i-1}+j,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} q_{i,j,t} \\ \vdots \\ s_{i,j}\sum_{t=1}^{T} q_{i,j,t}^{s_{i,j}-1} \end{bmatrix}, \tag{184}$$

$$A_{Q,\mathrm{C},i,j}^{F*} \coloneqq \left(\hat{\Delta}_{\tilde{v}u} - \hat{\Delta}_{\tilde{v}\tilde{v}}\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}\right)^{\bar{k}_N+j,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} q_{\mathrm{C},j,t} \\ \vdots \\ \varsigma_{i,j}\sum_{t=1}^{T} q_{\mathrm{C},j,t}^{\varsigma_{i,j}-1} \end{bmatrix}, \tag{185}$$

with $\bar{k}_i \coloneqq \sum_{s=1}^{N} m_s + m + \sum_{s=1}^{i} \bar{m}_s = k_N + m + \sum_{s=1}^{i} \bar{m}_s$ and $\bar{k}_0 = k_N + m$. We introduce further

notation for the limiting processes of the additional regressors. Denoting with

$$G_{\bar{D},i}(T) \coloneqq \operatorname{diag}(T^{-(q_i+1+1/2)}, \ldots, T^{-(q_i+\bar{q}_i+1/2)}), \tag{186}$$

$$G_{\bar{X},i,j}(T) \coloneqq \operatorname{diag}(T^{-(p_{i,j}+2)/2}, \ldots, T^{-(p_{i,j}+\bar{p}_{i,j}+1)/2}), \tag{187}$$

$$G_{\bar{X},\mathrm{C},i,j}(T) \coloneqq \operatorname{diag}(T^{-(c_{i,j}+2)/2}, \ldots, T^{-(c_{i,j}+\bar{c}_{i,j}+1)/2}), \tag{188}$$

$$G_{Q,i,j}(T) \coloneqq \operatorname{diag}(T^{-1}, \ldots, T^{-(s_{i,j}+1)/2}), \tag{189}$$

$$G_{Q,\mathrm{C},i,j}(T) \coloneqq \operatorname{diag}(T^{-1}, \ldots, T^{(\varsigma_{i,j}+1)/2}), \tag{190}$$

we obtain

$$T^{1/2} G_{\bar{D},i} D_{i,\lfloor rT \rfloor} \Rightarrow \bar{D}_i(r) \coloneqq [r^{q_i+1}, \ldots, r^{q_i+\bar{q}_i}]', \tag{191}$$

for $0 \leq r \leq 1$ and with Assumption 43 we obtain

$$T^{1/2} G_{\bar{X},i,j} \bar{X}_{i,j,\lfloor rT \rfloor} \Rightarrow \mathbf{B}^F_{v,i,j}(r) \coloneqq [B^{p_{i,j}+1}_{v,i,j}(r), \ldots, B^{p_{i,j}+\bar{p}_{i,j}}_{v,i,j}(r)]', \tag{192}$$

$$T^{1/2} G_{\bar{X},\mathrm{C},i,j} \bar{X}_{\mathrm{C},i,j,\lfloor rT \rfloor} \Rightarrow \mathbf{B}^F_{v,\mathrm{C},i,j}(r) \coloneqq [B^{c_{i,j}+1}_{v,\mathrm{C},j}, \ldots, B^{c_{i,j}+\bar{c}_{i,j}}_{v,\mathrm{C},j}(r)]', \tag{193}$$

$$T^{1/2} G_{Q,i,j} Q_{i,j,\lfloor rT \rfloor} \Rightarrow \mathbf{B}^F_{v^*,i,j}(r) \coloneqq [B_{v^*,i,j}(r), B^2_{v^*,i,j}(r), \ldots, B^{s_{i,j}}_{v^*,i,j}(r)]', \tag{194}$$

$$T^{1/2} G_{Q,\mathrm{C},i,j} Q_{\mathrm{C},i,j,\lfloor rT \rfloor} \Rightarrow \mathbf{B}^F_{v^*,\mathrm{C},i,j}(r) \coloneqq [B_{v^*,\mathrm{C},j}, B^2_{v^*,\mathrm{C},j}, \ldots, B^{\varsigma_{i,j}}_{v^*,\mathrm{C},j}(r)]', \tag{195}$$

for $0 \leq r \leq 1$. Furthermore, we define

$$G_{\bar{X},i}(T) \coloneqq \operatorname{diag}(G_{\bar{X},i,1}(T), \ldots, G_{\bar{X},i,m_i}(T)), \tag{196}$$

$$G_{\bar{X},\mathrm{C},i}(T) \coloneqq \operatorname{diag}(G_{\bar{X},\mathrm{C},i,1}(T), \ldots, G_{\bar{X},\mathrm{C},i,m}(T)), \tag{197}$$

$$G_{Q,i}(T) \coloneqq \operatorname{diag}(G_{Q,i,1}(T), \ldots, G_{Q,i,\bar{m}_i}(T), \tag{198}$$

$$G_{Q,\mathrm{C},i}(T) \coloneqq \operatorname{diag}(G_{Q,\mathrm{C},i,1}(T), \ldots, G_{Q,\mathrm{C},i,\bar{m}}(T)), \tag{199}$$

$$G_{F,i}(T) \coloneqq \operatorname{diag}(G_{\bar{D},i}(T), G_{\bar{X},i}(T), G_{\bar{X},\mathrm{C},i}(T), G_{Q,i}(T), G_{Q,\mathrm{C},i}(T)), \tag{200}$$

$$G_F(T) \coloneqq \operatorname{diag}(G_{F,1}(T), \ldots, G_{F,N}(T)), \tag{201}$$

$$\mathbf{B}^F_{v,i}(r) \coloneqq [\mathbf{B}^F_{v,i,1}(r)', \ldots, \mathbf{B}^F_{v,i,m_i}(r)']', \tag{202}$$

$$\mathbf{B}^F_{v,\mathrm{C},i}(r) \coloneqq [\mathbf{B}^F_{v,\mathrm{C},i,1}(r)', \ldots, \mathbf{B}^F_{v,\mathrm{C},i,m}(r)']', \tag{203}$$

$$\mathbf{B}^F_{v^*,i}(r) \coloneqq [\mathbf{B}^F_{v^*,i,1}(r)', \ldots, \mathbf{B}^F_{v^*_{i,\bar{m}_i}}(r)']', \tag{204}$$

$$\mathbf{B}^F_{v,\mathrm{C},i^*}(r) \coloneqq [\mathbf{B}^F_{v,\mathrm{C},i,1^*}(r), \ldots, \mathbf{B}^F_{v^*_{\mathrm{C},i,\bar{m}}}(r)']', \tag{205}$$

$$J^F_i(r) \coloneqq [\bar{D}_i(r)', \mathbf{B}^F_{v,i}(r)', \mathbf{B}^F_{v,\mathrm{C},i}(r)', \mathbf{B}^F_{v^*,i}(r)', \mathbf{B}^F_{v,\mathrm{C},i^*}(r)']', \tag{206}$$

$$J^F(r) \coloneqq \operatorname{diag}(J^F_1(r), \ldots, J^F_N(r)), \tag{207}$$

$$J^{\hat{F}}(r) \coloneqq J^F(r) - \int_0^1 J^F(s) J(s)' ds \left(J(s) J(s)' ds \right)^{-1} J(r). \tag{208}$$

With the necessary definitions in place, we next propose the FM-OLS estimator for $\theta_{\hat{F}}$ in (178) and the LM-type specification test statistic, both with their corresponding limiting distributions.

Proposition 7 *Let the data be generated according to* (144) *with Assumptions* (42) *and* (43) *in place. Furthermore, assume that based on OLS residuals all long-run covariances are estimated consistently. Define the fully modified OLS estimator of* $\theta_{\hat{F}}$ *in* (178) *as*

$$\hat{\theta}^{+}_{\hat{F}} \coloneqq \left(\hat{F}'\hat{F}\right)^{-1}\left(\hat{F}'\hat{u}^{+} - O^{F*} - A^{F*} + k^{F*}A^{Z*}\right), \tag{209}$$

with $O^{F*} \coloneqq \hat{F}'(I_T \otimes \hat{\Omega}_{u\tilde{v}}\hat{\Omega}^{-1}_{\tilde{v}\tilde{v}})\tilde{v} - \hat{F}'(I_T \otimes \hat{\Omega}_{uv}\hat{\Omega}^{-1}_{vv})v$, *where* $\tilde{v} \coloneqq [\tilde{v}'_1, \ldots, \tilde{v}'_T]'$, *and* $k^{F*} \coloneqq F'Z\,(Z'Z)^{-1}$. *Under the null hypothesis* $H_0\colon \theta_{\hat{F}} = 0$ *and consistent long-run covariance estimation, it holds that*

$$G_F^{-1}\hat{\theta}^{+}_{\hat{F}} \Rightarrow \left(\int_0^1 J^{\hat{F}}(r)J^{\hat{F}}(r)'dr\right)^{-1}\int_0^1 J^{\hat{F}}(r)dB_{u\cdot\tilde{v}}. \tag{210}$$

The LM-type test statistic for the null hypothesis $H_0\colon \theta_{\hat{F}} = 0$ *in* (178), *given by*

$$\hat{T}^{spec}_{LM} \coloneqq \hat{\theta}^{+\prime}_{\hat{F}}\left[\left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'\left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}\right)\hat{F}\left(\hat{F}'\hat{F}\right)^{-1}\right]^{-1}\hat{\theta}^{+}_{\hat{F}} \tag{211}$$

is under the null hypothesis asymptotically chi-squared distributed with $\bar{d}$ *degrees of freedom.*

3.2.4 Group-wise pooling

In case of equal coefficients in the parameter vector θ across equations, pooled estimation of these parameters leads to efficiency gains. Similar to Wagner *et al.* (2020), we perform group-wise pooling for the coefficients of the deterministic regressors, for the coefficients of the individual specific integrated regressors (and their powers) and, additionally, for the coefficients of the common integrated regressors (and their powers).

The formulation here is more general than in Wagner *et al.* (2020), not only due to multiple integrated regressors per equation, we additionally allow for group-wise specific powers of deterministic, integrated and common integrated regressors. For example, Wagner *et al.* (2020) pool the deterministic coefficients in each subset up to power $\breve{q}$, with $\breve{q}$ being less than or equal to the smallest q_i of all equations. In contrast, we pool in each subset l the coefficients up to $\breve{q}_l$, with $\breve{q}_l$ being only smaller than (or equal to) the minimum of q_i in subset l and not with $\breve{q}_l$ necessarily being smaller than (or equal to) the minimum of q_i from all subsets.

Define subsets $\mathcal{I}_{D,j}$ for $j = 1, \ldots, k_D$ such that $\mathcal{I}_D \coloneqq \{1, \ldots, N\} = \bigcup_{j=1}^{k_D}\mathcal{I}_{D,j}$. For $\mathcal{I}_{D,j}$ we test $[t, t^2, \ldots, t^{\breve{q}_j}]$ having equal coefficients for the equations in subset $\mathcal{I}_{D,j}$, with $\breve{q}_j \leq min_{i\in\mathcal{I}_{D,j}}q_i$.

We define $\mathcal{I}_X \coloneqq \{(1,1),\dots,(1,m_1),(2,1),\dots,(2,m_2),\dots,(N,1),\dots,(N,m_N)\}$ and the groups for pooling as $\mathcal{I}_{X,j}$, for $j = 1,\dots,k_X$, such that $\mathcal{I}_X = \bigcup_{j=1}^{k_X} \mathcal{I}_{X,j}$. Let $\check{p}_j \leq \min_{(k,l)\in\mathcal{I}_{X,j}} p_{k,l}$ be the power of the regressors up to which pooling is performed for all regressors in pooling group $\mathcal{I}_{X,j}$. We define $\mathcal{I}_{\mathrm{C}} \coloneqq \{(1,1),\dots,(1,m),(2,1),\dots,(2,m),\dots,(N,1),\dots,(N,m)\}$ an the groups for pooling $\mathcal{I}_{\mathrm{C},j}$, for $j = 1,\dots,k_{\mathrm{C}}$, such that $\mathcal{I}_{\mathrm{C}} = \bigcup_{j=1}^{k_{\mathrm{C}}} \mathcal{I}_{\mathrm{C},j}$. Let $\check{c}_j \leq \min_{(k,l)\in\mathcal{I}_{\mathrm{C},j}} c_{k,l}$ be the power of the regressors up to which pooling is performed for all regressors in pooling group $\mathcal{I}_{\mathrm{C},j}$. Without loss of generality, we order the subsets according to $|\mathcal{I}_{D,1}| \geq \dots, \geq |\mathcal{I}_{D,k_D}|$, $|\mathcal{I}_{X,1}| \geq \dots, \geq |\mathcal{I}_{X,k_X}|$ and $|\mathcal{I}_{\mathrm{C},1}| \geq \dots, \geq |\mathcal{I}_{\mathrm{C},k_{\mathrm{C}}}|$.

Furthermore, let $N_{D,j} \coloneqq |\mathcal{I}_{D,j}|$ and define $\mathcal{I}_{D,j} \coloneqq \{a_{D,1,j}, a_{D,2,j}, \dots, a_{D,N_{D,j},j}\}$, with its elements such that $1 \leq a_{D,1,j} < a_{D,2,j} < \dots < a_{D,M_j,j} \leq N$, for all $j = 1,\dots,k_D$. Similarly, define $N_{X,j} \coloneqq |\mathcal{I}_{X,j}|$ and $\mathcal{I}_{X,j} \coloneqq \{a_{X,1,j}, a_{X,2,j}, \dots, a_{X,N_{X,j},j}\}$, with $1 \leq (a_{X,1,j})^{1,1} < (a_{X,2,j})^{1,1} < \dots < (a_{X,M_j,j})^{1,1} \leq N$, for all $j = 1,\dots,k_X$. Finally, define $N_{\mathrm{C},j} \coloneqq |\mathcal{I}_{\mathrm{C},j}|$ and $\mathcal{I}_{\mathrm{C},j} \coloneqq \{a_{\mathrm{C},1,j}, a_{\mathrm{C},2,j}, \dots, a_{\mathrm{C},N_{\mathrm{C},j},j}\}$ with $1 \leq (a_{\mathrm{C},1,j})^{1,1} < (a_{\mathrm{C},2,j})^{1,1} < \dots < (a_{\mathrm{C},N_{\mathrm{C},j},j})^{1,1} \leq N$, for all $j = 1,\dots,k_{\mathrm{C}}$.

The null hypothesis for testing the described group-wise poolability is given by

$$H_0^{\mathrm{GW}}\colon [\theta_{D,2,i},\dots,\theta_{D,\check{q}_a+1,i}]' = [\theta_{D,2,j},\dots,\theta_{D,\check{q}_a+1,j}]', \tag{212}$$

$$\forall\, i,j \in \mathcal{I}_{D,a}\ \forall\, a \in \{\{1,\dots,k_D\}\colon |\mathcal{I}_{D,a}| > 1\}, \tag{213}$$

$$[\theta_{X,1,i_1,i_2},\dots,\theta_{X,\check{p}_b,i_1,i_2}]' = [\theta_{X,1,j_1,j_2},\dots,\theta_{X,\check{p}_b,j_1,j_2}]', \tag{214}$$

$$\forall\, (i_1,i_2),(j_1,j_2) \in \mathcal{I}_{X,b}\ \forall\, b \in \{\{1,\dots,k_X\}\colon |\mathcal{I}_{X,b}| > 1\}, \tag{215}$$

$$[\theta_{\mathrm{C},1,i_1,i_2},\dots,\theta_{\mathrm{C},\check{c}_c,i_1,i_2}]' = [\theta_{\mathrm{C},1,j_1,j_2},\dots,\theta_{\mathrm{C},\check{c}_c,j_1,j_2}]', \tag{216}$$

$$\forall\, (i_1,i_2),(j_1,j_2) \in \mathcal{I}_{\mathrm{C},c}\ \forall\, c \in \{\{1,\dots,k_{\mathrm{C}}\}\colon |\mathcal{I}_{\mathrm{C},c}| > 1\}, \tag{217}$$

with $\theta_{D,j,i}, \theta_{X,j,i_1,i_2}$ and $\theta_{\mathrm{C},j,i_1,i_2}$ being the j-th element of the parameter vector $\theta_{D,j,i}$, θ_{X,i_1,i_2} and $\theta_{\mathrm{C},i_1,i_2}$, respectively. The restriction matrices for the Wald-type tests for testing the above hypothesis are constructed in the way described below. With the chosen notation, we do not allow for pooling over parameters of deterministic regressors, individual specific integrated regressors and common integrated regressors jointly. Furthermore, we do not allow for more than one individual specific integrated regressor from each equation per subset. The same holds for common integrated regressors. None of the restrictions is particularly severe. These restrictions only serve the clarity of the chosen notation and have no technical reasons.

We define $R_D^{\mathrm{GW}} \coloneqq [R'_{D,1}, R'_{D,2}, \ldots, R'_{D,k_D}]'$ with $R_{D,j} \coloneqq [R_{D,j,1}, \ldots, R_{D,j,N}]$, where

$$R_{D,j,r} \coloneqq \begin{cases} \mathbf{1}_{N_{D,j}-1} \otimes [e_{2,d_r}, \ldots, e_{\check{q}_j+1,d_r}]' & \text{if } r = a_{D,1,j} \\ -e_{(i-1),(N_{D,j}-1)} \otimes [e_{2,d_r}, \ldots, e_{\check{q}_j+1,d_r}]' & \text{if } r = a_{D,i,j} \; \forall \; i = 2, \ldots, N_{D,j} \\ 0_{\check{q}_j(N_{D,j}-1)\times d_r} & \text{else} \end{cases} \tag{218}$$

for j such that $N_{D,j} > 1$ and $R_{D,j} = \emptyset$ otherwise.[57] We define $R_X^{\mathrm{GW}} \coloneqq [R'_{X,1}, R'_{X,2}, \ldots, R'_{X,k_X}]'$ with $R_{X,j} \coloneqq [R_{X,j,1}, \ldots, R_{X,j,N}]$ for j such that $N_{X,j} > 1$ and $R_{X,j} = \emptyset$ otherwise. We furthermore define

$$R_{X,j,r_1} \coloneqq \begin{cases} \mathbf{1}_{N_{X,j}-1} \otimes [e_{l+1,d_{r_1}}, \ldots, e_{l+\check{p}_j,d_{r_1}}]' & \text{if } r_1 = (a_{X,1,j})^{1,1} \\ -e_{(i-1),(N_{X,j}-1)} \otimes [e_{l+1,d_{r_1}}, \ldots, e_{l+\check{p}_j,d_{r_1}}]' & \text{if } r_1 = (a_{X,i,j})^{1,1} \; \forall \; i = 2, \ldots, N_{X,j} \\ 0_{\check{p}_j(N_{X,j}-1)\times d_r} & \text{else} \end{cases} \tag{219}$$

with defining $l \coloneqq 1 + q_{r_1} + \left(\sum_{s=1}^{r_2-1} p_{r_1,s}\right)$ and $r_2 \coloneqq (a_{X,i,j})^{1,2}$ for each $i = 1, \ldots, N_{X,j}$ in each group j. We define $R_{\mathrm{C}}^{\mathrm{GW}} \coloneqq [R'_{\mathrm{C},1}, R'_{\mathrm{C},2}, \ldots, R'_{\mathrm{C},k_{\mathrm{C}}}]'$ with $R_{\mathrm{C},j} \coloneqq [R_{\mathrm{C},j,1}, \ldots, R_{\mathrm{C},j,N}]$ for j such that $N_{\mathrm{C},j} > 1$ and $R_{\mathrm{C},j} = \emptyset$ otherwise. We furthermore define

$$R_{\mathrm{C},j,r_1} \coloneqq \begin{cases} \mathbf{1}_{N_{\mathrm{C},j}-1} \otimes [e_{l+1,d_{r_1}}, \ldots, e_{l+\check{c}_j,d_{r_1}}]' & \text{if } r_1 = (a_{\mathrm{C},1,j})^{1,1} \\ -e_{(i-1),(N_{\mathrm{C},j}-1)} \otimes [e_{l+1,d_{r_1}}, \ldots, e_{l+\check{c}_j,d_{r_1}}]' & \text{if } r_1 = (a_{\mathrm{C},i,j})^{1,1} \; \forall \; i = 2, \ldots, N_{\mathrm{C},j} \\ 0_{\check{c}_j(N_{\mathrm{C},j}-1)\times d_r} & \text{else} \end{cases} \tag{220}$$

with defining $l \coloneqq 1 + q_{r_1} + \left(\sum_{s=1}^{m_{r_1}} p_{r_1,s}\right) + \left(\sum_{s=1}^{r_2-1} c_{r_1,s}\right)$ and $r_2 \coloneqq (a_{\mathrm{C},i,j})^{1,2}$ for each $i = 1, \ldots, N_{\mathrm{C},j}$ in each group j. We then obtain $R^{\mathrm{GW}} \coloneqq [R_D^{\mathrm{GW}\prime}, R_X^{\mathrm{GW}\prime}, R_{\mathrm{C}}^{\mathrm{GW}\prime}]$.

The total number of restrictions is given by

$$s = \sum_{j=1}^{k_D} \check{q}_j(N_{D,j} - 1) + \sum_{j=1}^{k_X} \check{p}_j(N_{X,j} - 1) + \sum_{j=1}^{k_{\mathrm{C}}} \check{c}_j(N_{\mathrm{C},j} - 1) \tag{221}$$

and we have $r = 0_{s\times 1}$ for testing the above null hypothesis using $R^{\mathrm{GW}}\theta = r$ in the Wald-type test.

If the null hypothesis of equal coefficients for specified regressors is not rejected, group-wise pooled estimation can be applied. The parts of the corresponding regression matrix look as follows: For the deterministic regressors, define

$$\check{D}_t \coloneqq [\check{D}'_{1,t}, \ldots, \check{D}'_{k_D,t}]' \in \mathbb{R}^{\left(\sum_{j=1}^{k_D} \check{q}_j\right)\times N}, \tag{222}$$

[57] There is a typo in Equations (21) and (22) of Wagner *et al.* (2020), the subscript r of $e_{r,(N_j-1)}$ and $e_{r,(M_j-1)}$ should be replaced by a subscript defined in a similar fashion as here. Otherwise r may exceed N_j and M_j, which is not allowed by definition.

with

$$\check{D}_{j,t} \coloneqq \sum_{i=1}^{N} \mathbf{1}_{\{i \in \mathcal{I}_{D,j}\}} \cdot (e_{i,N}' \otimes [t, t^2, \ldots, t^{\check{q}_j}]') \in \mathbb{R}^{\check{q}_j \times N}, \tag{223}$$

for $j = 1, \ldots, k_D$. For the individual specific regressors we have

$$\check{X}_t \coloneqq [\check{X}_{1,t}', \ldots, \check{X}_{k_X,t}']' \in \mathbb{R}^{\left(\sum_{j=1}^{k_X} \check{p}_j\right) \times N}, \tag{224}$$

with

$$\check{X}_{j,t} \coloneqq \sum_{i_1=1}^{N} \sum_{i_2=1}^{m_{i_1}} \mathbf{1}_{\{(i_1,i_2) \in \mathcal{I}_{X,j}\}} \cdot (e_{i_1,N}' \otimes [x_{i_1,i_2,t}, x_{i_1,i_2,t}^2, \ldots, x_{i_1,i_2,t}^{\check{p}_j}]') \in \mathbb{R}^{\check{p}_j \times N}, \tag{225}$$

for $j = 1, \ldots, k_X$. In a similar fashion, for the common integrated regressors we have

$$\check{X}_{\mathrm{C},t} \coloneqq [\check{X}_{\mathrm{C},1,t}', \ldots, \check{X}_{\mathrm{C},k_\mathrm{C},t}']' \in \mathbb{R}^{\left(\sum_{j=1}^{k_\mathrm{C}} \check{c}_j\right) \times N}, \tag{226}$$

with

$$\check{X}_{\mathrm{C},j,t} \coloneqq \sum_{i_1=1}^{N} \sum_{i_2=1}^{m} \mathbf{1}_{\{(i_1,i_2) \in \mathcal{I}_{\mathrm{C},j}\}} \cdot (e_{i_1,N}' \otimes [x_{\mathrm{C},i_2,t}, x_{\mathrm{C},i_2,t}^2, \ldots, x_{\mathrm{C},i_2,t}^{\check{c}_j}]') \in \mathbb{R}^{\check{c}_j \times N}, \tag{227}$$

for $j = 1, \ldots, k_\mathrm{C}$. Before defining the regressor matrix for the group-wise pooled estimator, we define some more quantities. Define for all $i = 1, \ldots, N$,

$$\check{q}_{i,1} \coloneqq \sum_{\ell=1}^{k_D} \mathbf{1}_{\{i \in \mathcal{I}_{D,\ell}\}} \cdot \check{q}_\ell, \tag{228}$$

$$\check{p}_{i,j} \coloneqq \sum_{\ell=1}^{k_X} \mathbf{1}_{\{(i,j) \in \mathcal{I}_{X,\ell}\}} \cdot \check{p}_\ell, \ \forall\ j = 1, \ldots, m_i, \tag{229}$$

$$\check{c}_{i,j} \coloneqq \sum_{\ell=1}^{k_\mathrm{C}} \mathbf{1}_{\{(i,j) \in \mathcal{I}_{\mathrm{C},\ell}\}} \cdot \check{c}_\ell, \ \forall\ j = 1, \ldots, m. \tag{230}$$

Furthermore, we define

$$\mathring{D}_{i,t} \coloneqq [t^{\check{q}_{i,1}+1}, \ldots, t^{q_i}]', \tag{231}$$

$$\mathring{X}_{i,j,t} \coloneqq [x_{i,j,t}^{\check{p}_{i,j}+1}, \ldots, x_{i,j,t}^{p_{i,j}}]', \ \forall\ j = 1, \ldots, m_i, \tag{232}$$

$$\mathring{X}_{\mathrm{C},i,j,t} \coloneqq [x_{\mathrm{C},j,t}^{\check{c}_{i,j}+1}, \ldots, x_{\mathrm{C},j,t}^{c_{i,j}}]', \ \forall\ j = 1, \ldots, m, \tag{233}$$

for $i = 1, \ldots, N$. With having

$$\mathring{Z}_{i,t} \coloneqq [\mathring{D}_{i,t}', \mathring{X}_{i,1,t}', \ldots, \mathring{X}_{i,m_i,t}', \mathring{X}_{\mathrm{C},i,1,t}', \ldots \mathring{X}_{\mathrm{C},i,m,t}']', \tag{234}$$

for $i = 1, \ldots, N$, we then define

$$\mathring{Z}_t \coloneqq \operatorname{diag}(\mathring{Z}_{1,t}, \ldots, \mathring{Z}_{N,t}) \in \mathbb{R}^{\mathring{d} \times N}, \tag{235}$$

with $\mathring{d} \coloneqq d - N - \sum_{i=1}^{k_D} \check{q}_i N_{D,i} - \sum_{i=1}^{k_X} \check{p}_i N_{X,i} - \sum_{i=1}^{k_C} \check{c}_i N_{C,i}$. The group-wise pooling model can then be written as

$$y_t = \check{Z}_t' \theta^{\mathrm{GW}} + u_t, \tag{236}$$

with $y_t \coloneqq [y_{1,t}, \ldots, y_{N,t}]'$, $u_t \coloneqq [u_{1,t}, \ldots, u_{N,t}]'$, $\check{Z}_t \coloneqq [I_N, \check{D}_t', \check{X}_t', \check{X}_{\mathrm{C},t}', \mathring{Z}_t']' \in \mathbb{R}^{\check{d} \times N}$ with $\check{d} \coloneqq d - \sum_{i=1}^{k_D} \check{q}_i (N_{D,i} - 1) - \sum_{i=1}^{k_X} \check{p}_i (N_{X,i} - 1) - \sum_{i=1}^{k_C} \check{c}_i (N_{C,i} - 1)$, and the parameter vector is given by

$$\theta^{\mathrm{GW}} \coloneqq [\theta_{D,1,1}, \ldots, \theta_{D,1,N}, \theta_D^{(1)\prime}, \ldots, \theta_D^{(k_D)\prime}, \theta_X^{(1)\prime}, \ldots, \theta_X^{(k_X)\prime}, \theta_{\mathrm{C}}^{(1)\prime}, \ldots, \theta_{\mathrm{C}}^{(k_C)\prime}, \tag{237}$$

$$\theta_{\mathring{D},1}', \theta_{\mathring{X},1,1}', \ldots, \theta_{\mathring{X},1,m_1}', \theta_{\mathring{C},1,1}', \ldots, \theta_{\mathring{C},1,m}', \ldots, \tag{238}$$

$$\theta_{\mathring{D},N}', \theta_{\mathring{X},N,1}', \ldots, \theta_{\mathring{X},N,m_N}', \theta_{\mathring{C},N,1}', \ldots, \theta_{\mathring{C},N,m}']' \in \mathbb{R}^{\mathring{d}}, \tag{239}$$

where $\theta_D^{(j)} \coloneqq [\theta_{D,1}^{(j)}, \ldots, \theta_{D,\check{q}_j}^{(j)}]'$ for $j = 1, \ldots, k_D$, $\theta_X^{(j)} \coloneqq [\theta_{X,1}^{(j)}, \ldots, \theta_{X,\check{p}_j}^{(j)}]'$ for $j = 1, \ldots, k_X$ and $\theta_{\mathrm{C}}^{(j)} \coloneqq [\theta_{\mathrm{C},1}^{(j)}, \ldots, \theta_{\mathrm{C},\check{p}_j}^{(j)}]'$ for $j = 1, \ldots, k_{\mathrm{C}}$, and furthermore for $i = 1, \ldots, N$: $\theta_{\mathring{D},i} \coloneqq [\theta_{\mathring{D},\check{q}_i,1+1,i}, \ldots, \theta_{\mathring{D},q_i,i}]'$, $\theta_{\mathring{X},i,j} \coloneqq [\theta_{\mathring{X},\check{p}_{i,j}+1,i}, \ldots, \theta_{\mathring{X},p_{i,j},i}]'$ for $j = 1, \ldots, m_i$, and $\theta_{\mathring{C},i,j} \coloneqq [\theta_{\mathring{C},\check{c}_{i,j}+1,i}, \ldots, \theta_{\mathring{C},c_{i,j},i}]'$ for $j = 1, \ldots, m$. Stacking all quantities over time leads to

$$y = \check{Z} \theta^{\mathrm{GW}} + u, \tag{240}$$

with $y \coloneqq [y_1, \ldots, y_T]'$, $\check{Z} \coloneqq [\check{Z}_1, \ldots, \check{Z}_T]'$ and $u \coloneqq [u_1, \ldots, u_T]'$.

We next turn to the definitions of the correction factors for the group-wise pooled versions of the FM-SOLS and FM-SUR estimators. Define $\hat{A}^{\mathrm{GW}} \coloneqq [0_{1 \times (N + \sum_{j=1}^{k_D} \check{q}_j)}, \hat{A}_1^{\mathrm{GW}\prime}, \ldots, \hat{A}_{k_X}^{\mathrm{GW}\prime}, \hat{A}_{\mathrm{C},1}^{\mathrm{GW}\prime}, \ldots, \hat{A}_{\mathrm{C},k_{\mathrm{C}}}^{\mathrm{GW}\prime}]'$

and $\tilde{A}^{\mathrm{GW}*} \coloneqq [0_{1\times(N+\sum_{j=1}^{k_D}\check{q}_j)}, \tilde{A}_1^{\mathrm{GW}*\prime}, \ldots, \tilde{A}_{k_X}^{\mathrm{GW}*\prime}, \tilde{A}_{\mathrm{C},1}^{\mathrm{GW}*\prime}, \ldots, \tilde{A}_{\mathrm{C},k_\mathrm{C}}^{\mathrm{GW}*\prime}]'$ with

$$\hat{A}_\ell^{\mathrm{GW}} \coloneqq \sum_{i=1}^{N}\sum_{j=1}^{m_i} \mathbf{1}_{\{(i,j)\in\mathcal{I}_{X,\ell}\}} \cdot (\hat{\Delta}_{vu}^{+})^{k_{i-1}+j,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{i,j,t} \\ \vdots \\ \check{p}_{i,j}\sum_{t=1}^{T} x_{i,j,t}^{\check{p}_{i,j}-1} \end{bmatrix}, \ \ell = 1,\ldots,k_X, \tag{241}$$

$$\hat{A}_{\mathrm{C},\ell}^{\mathrm{GW}} \coloneqq \sum_{i=1}^{N}\sum_{j=1}^{m} \mathbf{1}_{\{(i,j)\in\mathcal{I}_{\mathrm{C},\ell}\}} \cdot (\hat{\Delta}_{vu}^{+})^{k_N+j,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{\mathrm{C},j,t} \\ \vdots \\ \check{c}_{i,j}\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{\check{c}_{i,j}-1} \end{bmatrix}, \ \ell = 1,\ldots,k_\mathrm{C}, \tag{242}$$

$$\tilde{A}_\ell^{\mathrm{GW}*} \coloneqq \sum_{i=1}^{N}\sum_{j=1}^{m_i} \mathbf{1}_{\{(i,j)\in\mathcal{I}_{X,\ell}\}} \cdot (\hat{\Delta}_{vu}^{+})^{k_{i-1}+j,\cdot} (\hat{\Omega}_{u\cdot v}^{-1})^{\cdot,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{i,j,t} \\ \vdots \\ \check{p}_{i,j}\sum_{t=1}^{T} x_{i,j,t}^{\check{p}_{i,j}-1} \end{bmatrix}, \ \ell = 1,\ldots,k_X, \tag{243}$$

$$\tilde{A}_{\mathrm{C},\ell}^{\mathrm{GW}*} \coloneqq \sum_{i=1}^{N}\sum_{j=1}^{m} \mathbf{1}_{\{(i,j)\in\mathcal{I}_{\mathrm{C},\ell}\}} \cdot (\hat{\Delta}_{vu}^{+})^{k_N+j,\cdot} (\hat{\Omega}_{u\cdot v}^{-1})^{\cdot,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{\mathrm{C},j,t} \\ \vdots \\ \check{c}_{i,j}\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{\check{c}_{i,j}-1} \end{bmatrix}, \ \ell = 1,\ldots,k_\mathrm{C}. \tag{244}$$

Furthermore, define $\mathring{\hat{A}}^{\mathrm{GW}} \coloneqq [\mathring{\hat{A}}_1^{\mathrm{GW}\prime}, \ldots, \mathring{\hat{A}}_N^{\mathrm{GW}\prime}]'$, with

$$\mathring{\hat{A}}_i^{\mathrm{GW}} \coloneqq \begin{bmatrix} 0_{(q_i-\check{q}_{i,1})\times 1} \\ \mathring{\hat{A}}_{i,1}^{\mathrm{GW}} \\ \vdots \\ \mathring{\hat{A}}_{i,m_i}^{\mathrm{GW}} \\ \mathring{\hat{A}}_{\mathrm{C},i,1}^{\mathrm{GW}} \\ \vdots \\ \mathring{\hat{A}}_{\mathrm{C},i,m}^{\mathrm{GW}} \end{bmatrix}, \quad \mathring{\hat{A}}_{i,j}^{\mathrm{GW}} \coloneqq (\hat{\Delta}_{vu}^{+})^{k_{i-1}+j,i} \begin{bmatrix} (\check{p}_{i,j}+1)\sum_{t=1}^{T} x_{i,j,t}^{\check{p}_{i,j}} \\ \vdots \\ p_{i,j}\sum_{t=1}^{T} x_{i,j,t}^{p_{i,j}-1} \end{bmatrix}, \tag{245}$$

$$\mathring{\hat{A}}_{\mathrm{C},i,j}^{\mathrm{GW}} \coloneqq (\hat{\Delta}_{vu}^{+})^{k_N+j,i} \begin{bmatrix} (\check{c}_{i,j}+1)\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{\check{c}_{i,j}} \\ \vdots \\ c_{i,j}\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{c_{i,j}-1} \end{bmatrix}, \tag{246}$$

and $\mathring{\tilde{A}}^{\mathrm{GW}*} \coloneqq [\mathring{\tilde{A}}_1^{\mathrm{GW}*\prime}, \ldots, \mathring{\tilde{A}}_N^{\mathrm{GW}*\prime}]'$, with

$$\mathring{\tilde{A}}_i^{\mathrm{GW}*} \coloneqq \begin{bmatrix} 0_{(q_i - \check{q}_{i,1})\times 1} \\ \mathring{\tilde{A}}_{i,1}^{\mathrm{GW}*} \\ \vdots \\ \mathring{\tilde{A}}_{i,m_i}^{\mathrm{GW}*} \\ \mathring{\tilde{A}}_{\mathrm{C},i,1}^{\mathrm{GW}*} \\ \vdots \\ \mathring{\tilde{A}}_{\mathrm{C},i,m}^{\mathrm{GW}*} \end{bmatrix}, \quad \mathring{\tilde{A}}_{i,j}^{\mathrm{GW}*} \coloneqq (\hat{\Delta}_{vu}^{+})^{k_{i-1}+j,\cdot} (\hat{\Omega}_{u\cdot v}^{-1})^{\cdot,i} \begin{bmatrix} (\check{p}_{i,j}+1)\sum_{t=1}^{T} x_{i,j,t}^{\check{p}_{i,j}} \\ \vdots \\ p_{i,j}\sum_{t=1}^{T} x_{i,j,t}^{p_{i,j}-1} \end{bmatrix}, \tag{247}$$

$$\mathring{\tilde{A}}_{\mathrm{C},i,j}^{\mathrm{GW}*} \coloneqq (\hat{\Delta}_{vu}^{+})^{k_N+j,\cdot} (\hat{\Omega}_{u\cdot v}^{-1})^{\cdot,i} \begin{bmatrix} (\check{c}_{i,j}+1)\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{\check{c}_{i,j}} \\ \vdots \\ c_{i,j}\sum_{t=1}^{T} x_{\mathrm{C},j,t}^{c_{i,j}-1} \end{bmatrix}. \tag{248}$$

The weighting matrix for group-wise pooled estimation is then given by

$$\check{G}(T) \coloneqq \mathrm{diag}(\check{G}_1(T), \check{G}_D(T), \check{G}_X(T), \check{G}_{\mathrm{C}}(T), \mathring{G}(T)), \tag{249}$$

with

$$\check{G}_1(T) \coloneqq T^{-1/2} \cdot I_N, \tag{250}$$

$$\check{G}_D(T) \coloneqq \mathrm{diag}(\check{G}_{D,1}(T), \ldots, \check{G}_{D,k_D}(T)), \tag{251}$$

$$\check{G}_{D,i}(T) \coloneqq \mathrm{diag}(T^{-3/2}, \ldots, T^{-(\check{q}_i+1/2)}), \tag{252}$$

$$\check{G}_X(T) \coloneqq \mathrm{diag}(\check{G}_{X,1}(T), \ldots, \check{G}_{X,k_X}(T)), \tag{253}$$

$$\check{G}_{X,i}(T) \coloneqq \mathrm{diag}(T^{-1}, \ldots, T^{-(\check{p}_i+1)/2}), \tag{254}$$

$$\check{G}_{\mathrm{C}}(T) \coloneqq \mathrm{diag}(\check{G}_{\mathrm{C},1}(T), \ldots, \check{G}_{\mathrm{C},k_{\mathrm{C}}}(T)), \tag{255}$$

$$\check{G}_{\mathrm{C},i}(T) \coloneqq \mathrm{diag}(T^{-1}, \ldots, T^{-(\check{c}_i+1)/2}), \tag{256}$$

$$\mathring{G}(T) \coloneqq \mathrm{diag}(\mathring{G}_1(T), \ldots, \mathring{G}_N(T)), \tag{257}$$

$$\mathring{G}_i(T) \coloneqq \mathrm{diag}(\mathring{G}_{D,i}(T), \mathring{G}_{X,i,1}(T), \ldots, \mathring{G}_{X,i,m_i}(T), \tag{258}$$

$$\mathring{G}_{\mathrm{C},i,1}(T), \ldots, \mathring{G}_{\mathrm{C},i,m}(T)), \tag{259}$$

$$\mathring{G}_{D,i}(T) \coloneqq \mathrm{diag}(T^{-(\check{q}_i+3/2)}, \ldots, T^{-(q_i+1/2)}), \tag{260}$$

$$\mathring{G}_{X,i,j}(T) \coloneqq \mathrm{diag}(T^{-(\check{p}_{i,j}+2)/2}, \ldots, T^{-(p_{i,j}+1)/2}), \tag{261}$$

$$\mathring{G}_{\mathrm{C},i,j}(T) \coloneqq \mathrm{diag}(T^{-(\check{c}_{i,j}+2)/2}, \ldots, T^{-(c_{i,j}+1)/2}). \tag{262}$$

Furthermore, define the limit stochastic process as

$$\check{J}(r) \coloneqq [I_N, \check{J}'_D(r), \check{J}'_X(r), \check{J}'_{\mathrm{C}}(r), \mathring{J}(r)']' \tag{263}$$

$$\check{J}'_D(r) \coloneqq [\check{J}_{D,1}(r)', \ldots, \check{J}_{D,k_D}(r)']', \tag{264}$$

$$\check{J}_{D,j}(r) \coloneqq \sum_{i=1}^{N} \mathbf{1}_{\{i \in \mathcal{I}_{D,j}\}} \cdot \left(e'_{i,N} \otimes [r, r^2, \ldots, r^{\check{q}_j}]'\right) \tag{265}$$

$$\check{J}'_X(r) \coloneqq [\check{J}_{X,1}(r)', \ldots, \check{J}_{X,k_X}(r)']', \tag{266}$$

$$\check{J}_{X,j}(r) \coloneqq \sum_{i_1=1}^{N} \sum_{i_2=1}^{m_{i_1}} \mathbf{1}_{\{(i_1,i_2) \in \mathcal{I}_{X,j}\}} \cdot \left(e'_{i_1,N} \otimes [B_{v,i_1,i_2}(r), B^2_{v,i_1,i_2}(r), \ldots, B^{\check{p}_j}_{v,i_1,i_2}(r)]'\right), \tag{267}$$

$$\check{J}'_{\mathrm{C}}(r) \coloneqq [\check{J}_{\mathrm{C},1}(r)', \ldots, \check{J}_{\mathrm{C},k_{\mathrm{C}}}(r)']', \tag{268}$$

$$\check{J}_{\mathrm{C},j}(r) \coloneqq \sum_{i_1=1}^{N} \sum_{i_2=1}^{m} \mathbf{1}_{\{(i_1,i_2) \in \mathcal{I}_{\mathrm{C},j}\}} \cdot \left(e'_{i_1,N} \otimes [B_{v,\mathrm{C},i_2}(r), B^2_{v,\mathrm{C},i_2}(r), \ldots, B^{\check{c}_j}_{v,\mathrm{C},i_2}(r)]'\right), \tag{269}$$

$$\mathring{J}(r) \coloneqq \mathrm{diag}(\mathring{J}_1(r), \ldots, \mathring{J}_N(r)), \tag{270}$$

$$\mathring{J}_i(r) \coloneqq [\mathring{J}_{D,i}(r)', \mathring{J}_{X,i,1}(r)', \ldots, \mathring{J}_{X,i,m_i}(r)', \mathring{J}_{\mathrm{C},i,1}(r)', \ldots, \mathring{J}_{\mathrm{C},i,m}(r)']', \tag{271}$$

$$\mathring{J}_{D,i}(r)' \coloneqq [r^{\check{q}_{i,1}+1}, \ldots, r^{q_i}]', \tag{272}$$

$$\mathring{J}_{X,i,j}(r) \coloneqq [B^{\check{p}_{i,j}+1}_{v,i,j}(r), \ldots, B^{p_{i,j}}_{v,i,j}(r)]', \ \forall\ j = 1, \ldots, m_i, \tag{273}$$

$$\mathring{J}_{\mathrm{C},i,j}(r) \coloneqq [B^{\check{c}_{i,j}+1}_{v,\mathrm{C},j}(r), \ldots, B^{c_{i,j}}_{v,\mathrm{C},j}(r)]', \ \forall\ j = 1, \ldots, m, \tag{274}$$

for $0 \leq r \leq 1$.

Proposition 8 *Let the data be generated according to* (240)*, the restricted version of* (144) *with group-wise pooled parameters, and with Assumption 42 in place. Again, assume that based on OLS residuals all long-run covariances are estimated consistently. Using the correction terms* $\hat{A}^{GW}$*,* $\mathring{\hat{A}}^{GW}$*,* $\tilde{A}^{GW*}$ *and* $\mathring{\tilde{A}}^{GW*}$*, the group-wise FM-SOLS and FM-SUR estimators are given by*

$$\hat{\theta}^{GW} \coloneqq \left(\check{Z}'\check{Z}\right)^{-1} \left(\check{Z}' y^{+} - \begin{bmatrix} \hat{A}^{GW} \\ \mathring{\hat{A}}^{GW} \end{bmatrix}\right), \tag{275}$$

$$\tilde{\theta}^{GW} \coloneqq \left(\check{Z}' \left(I_T \otimes \hat{\Omega}^{-1}_{u \cdot v}\right) \check{Z}\right)^{-1} \left(\check{Z}' \left(I_T \otimes \hat{\Omega}^{-1}_{u \cdot v}\right) y^{+} - \begin{bmatrix} \tilde{A}^{GW*} \\ \mathring{\tilde{A}}^{GW*} \end{bmatrix}\right). \tag{276}$$

For $T \to \infty$ *it holds that*

$$G^{-1}\left(\hat{\theta}^{GW} - \theta^{GW}\right) \Rightarrow \left(\int_0^1 \check{J}(r)\check{J}(r)' dr\right)^{-1} \int_0^1 \check{J}(r) dB_{u \cdot v}(r), \tag{277}$$

$$G^{-1}\left(\tilde{\theta}^{GW} - \theta^{GW}\right) \Rightarrow \left(\int_0^1 \check{J}(r)\Omega^{-1}_{u \cdot v}\check{J}(r)' dr\right)^{-1} \int_0^1 \check{J}(r)\Omega^{-1}_{u \cdot v} dB_{u \cdot v}(r). \tag{278}$$

In case that $\check{q}_{i,1} = q_i$, $\check{p}_{i,j} = p_{i,j}$ or $\check{c}_{i,j} = c_{i,j}$, the corresponding regressor $\mathring{D}_{i,t}$, $\mathring{X}_{i,j,t}$, or $\mathring{X}_{\mathrm{C},i,j,t}$ is empty. Thus, if $\check{q}_{i,1} = q_i$ for all i in $1, \ldots, N$, $\check{p}_{i,j} = p_{i,j}$ for all (i,j) in $\mathcal{I}_X$ and $\check{c}_{i,j} = c_{i,j}$ for all (i,j) in $\mathcal{I}_{\mathrm{C}}$, it follows that $\mathring{\hat{A}}^{\mathrm{GW}} = \emptyset$ and $\mathring{\tilde{A}}^{\mathrm{GW}*} = \emptyset$.

3.3 Finite Sample Performance

We generate data for the quadratic seemingly unrelated cointegrating polynomial regression model, including a single common integrated regressor with powers up to two:

$$y_{i,t} = c_i + \delta_i t + \beta_{1,i} x_{i,t} + \beta_{2,i} x_{i,t}^2 + \beta_{3,i} x_{\mathrm{C},1,t} + \beta_{4,i} x_{\mathrm{C},1,t}^2 + u_{i,t}, \tag{279}$$

where the errors $u_{i,t}$, $\Delta x_{i,t} = v_{i,t}$ and $\Delta x_{\mathrm{C},1,t} = v_{\mathrm{C},1,t}$ are generated as

$$u_{i,t} = \rho_1 u_{i,t-1} + \varepsilon_{i,t} + \rho_2 (v_{i,t} + v_{\mathrm{C},1,t}), \tag{280}$$

$$v_{i,t} = e_{i,t} + 0.5 e_{i,t-1}, \tag{281}$$

$$v_{\mathrm{C},1,t} = e_{\mathrm{C},1,t} + 0.5 e_{\mathrm{C},1,t-1}, \tag{282}$$

with $u_{i,0} = e_{i,0} = e_{\mathrm{C},1,0} = 0$ for all $i = 1, \ldots, N$. The vectors $\varepsilon_t \coloneqq [\varepsilon_{1,t}, \ldots, \varepsilon_{N,t}]'$ and $e_t \coloneqq [e_{1,t}, \ldots, e_{N,t}, e_{\mathrm{C},1,t}]'$ are independently normally distributed with covariance matrices $\Sigma_{\varepsilon\varepsilon}$ and Σ_{ee} given by

$$\Sigma_{\varepsilon\varepsilon} \coloneqq \begin{bmatrix} 1 & \rho_3 & \ldots & \rho_3 \\ \rho_3 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_3 \\ \rho_3 & \ldots & \rho_3 & 1 \end{bmatrix} \in \mathbb{R}^{N \times N}, \; \Sigma_{ee} \coloneqq \begin{bmatrix} 1 & \rho_4 & \ldots & \rho_4 \\ \rho_4 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_4 \\ \rho_4 & \ldots & \rho_4 & 1 \end{bmatrix} \in \mathbb{R}^{(N+1) \times (N+1)}. \tag{283}$$

The parameter ρ_1 controls the level of serial correlation in the error term $u_{i,t}$ and ρ_2 controls the extent of endogeneity between the regressors and the error term. The parameter ρ_3 controls for the correlation between regressors across equations through cross-sectional correlation of $\varepsilon_{i,t}$. Similarly, ρ_4 controls for the correlation between error terms of the N equations through cross-sectional correlation of $e_{i,t}$. For the sake of brevity, we report results only for identical ρ parameters with $\rho \coloneqq \rho_1 = \rho_2 = \rho_3 = \rho_4 \in \{0.0, 0.3, 0.6, 0.9\}$. The model parameters are as follows, $c_i = \delta_i = 1$, $\beta_{1,i} = 5$, $\beta_{2,i} = -0.3$, $\beta_{3,i} = 3$ and $\beta_{4,i} = -0.2$, for $i = 1, \ldots, N$, with being identical across equations to allow for pooling. For pooled estimation, we consider pooling for only the common stochastic regressors $x_{\mathrm{C},1,t}$ and $x_{\mathrm{C},1,t}^2$, labeled (C), and pooling for all stochastic regressors, i. e. , additionally pooling for $x_{i,t}$ and $x_{i,t}^2$, labeled (S). These two cases of pooling still allow for individual specific intercept and trend slopes. We consider sample sizes of $T \in \{100, 200, 500, 1000\}$ and

$N \in \{3, 10\}$. All test decisions are performed at the nominal significance level of 5%. The number of replications is 5000 throughout. We perform estimation and hypothesis testing for $\beta_4 = -0.2$, using estimated long-run variances. Results for β_2 and using the true long-run variances are provided in Appendix 3.6.2. For long-run variance estimation we consider the Bartlett kernel with bandwidths chosen according to Newey and West (1994). All the estimation methods considered here are of fully-modified type. For brevity, we neglect the prefix *FM*.

Table 77 shows results for bias and root mean squared error (RMSE) for the coefficient β_4. Without loss of generality, we consider for unrestricted CPR, SOLS and SUR estimation the results corresponding to equation $i = 1$. Starting with bias in Panel A, for most cases, bias is increasing with increasing ρ and decreasing for larger N. Exceptions are $T = 100$ and $T = 200$, where we often observe the lowest bias for $\rho = 0.6$. Nevertheless, this does not hold for RMSE. For unrestricted estimation, namely CPR, SOLS and SUR, the CPR estimation leads in most cases to the smallest bias, but often only with a very small lead. Pooled estimation substantially reduces bias in most cases. In some instances, the pooled estimation of β_1 and β_2 also leads to a better estimate of β_4. Between SUR and SOLS, regardless of the type of pooling or unrestricted estimation, we do not see a clearly preferable estimator with respect to bias. Turning to RMSE in Panel B, we observe increasing RMSE with increasing ρ and additionally decreasing RMSE with increasing sample size T. Increasing the number of equations has a slightly negative impact on unrestricted SOLS and SUR estimation for $T \leq 200$, and a positive impact on pooled SOLS estimation, and for $\rho \in \{0.0, 0.3\}$ on pooled SUR estimation as well. The unrestricted estimators are generally about the same on, with SUR often slightly ahead, especially for large values of ρ. For pooled estimation, SOLS is often slightly ahead, with exception of $\rho = 0.9$. When comparing the pooling cases (C) and (S), additional pooling of β_1 and β_2 often reduces RMSE also for β_4.

We now turn to empirical null rejection probabilities of the t-test for $H_0 : \beta_4 = -0.2$ against the alternative $H_1 : \beta_4 \neq -0.2$. For unrestricted CPR, SOLS and SUR estimation, we consider the results corresponding to equation $i = 1$, without loss of generality. For the pooled estimation SOLS(C), SUR(C), SOLS(S) and SUR(S), the coefficient β_4 is pooled over all equations $i = 1, \ldots, N$. The results in Table 78 show, that CPR estimation leads to the smallest null rejection probabilities across all estimators, independent of pooling and the choice of N here. For the system estimators SOLS often leads to slightly lower null rejection probabilities in unrestricted estimation than SUR. Pooling increases null rejection probabilities in most cases slightly. Null rejection probabilities also become larger for increasing N. The larger null rejection probabilities of the system estimators

ρ	$N=3$							$N=10$						
	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)
Panel A: Bias(×1000)														
$T=100$														
0.0	-0.0810	-0.1047	-0.1093	-0.0329	-0.0329	-0.0337	-0.0249	-0.0000	-0.0152	0.0030	0.0022	0.0424	-0.0020	0.0433
0.3	-0.2546	-0.2830	-0.2624	-0.1228	-0.1292	-0.1384	-0.1293	0.0844	0.0941	0.0835	0.0286	0.0144	0.0227	0.0200
0.6	-0.0521	-0.0616	-0.0120	-0.0827	-0.0469	-0.0132	0.0209	0.1046	0.1925	0.1643	0.2056	0.2085	0.1871	0.2180
0.9	0.2698	0.5316	0.5197	0.4393	0.4800	0.4383	0.5484	0.0245	0.0077	0.1065	0.5472	0.2623	0.2630	0.1893
$T=200$														
0.0	0.0410	0.0406	0.0449	0.0273	0.0322	0.0169	0.0204	0.0363	0.0452	0.0464	0.0169	0.0206	0.0180	0.0210
0.3	0.0256	0.0222	0.0156	0.0266	0.0196	0.0226	0.0221	0.0112	0.0065	0.0151	0.0154	0.0318	0.0125	0.0294
0.6	0.0073	-0.0039	-0.0015	0.0380	-0.0013	0.0234	-0.0044	0.0032	-0.0128	-0.0224	0.0158	0.0445	-0.0033	0.0129
0.9	-0.0782	-0.1834	-0.0549	0.1763	0.0177	0.1230	-0.0174	-0.3839	-0.3945	-0.3395	-0.5032	-0.6608	-0.4918	-0.5755
$T=500$														
0.0	-0.0108	-0.0122	-0.0123	-0.0052	-0.0044	-0.0052	-0.0047	-0.0061	-0.0084	-0.0093	-0.0030	-0.0041	-0.0027	-0.0037
0.3	-0.0150	-0.0157	-0.0182	-0.0040	-0.0041	-0.0045	-0.0055	-0.0077	-0.0055	-0.0055	0.0019	-0.0023	-0.0003	-0.0041
0.6	-0.0345	-0.0382	-0.0355	-0.0093	-0.0064	-0.0089	-0.0069	-0.0006	-0.0036	-0.0048	0.0048	-0.0068	0.0015	-0.0118
0.9	-0.2102	-0.2363	-0.2008	-0.1041	-0.1207	-0.1122	-0.1201	-0.0458	-0.0774	-0.0786	-0.0806	-0.0327	-0.0390	-0.0513
$T=1000$														
0.0	-0.0042	-0.0043	-0.0045	-0.0011	-0.0015	-0.0003	-0.0007	0.0052	0.0051	0.0052	0.0003	0.0004	0.0002	0.0003
0.3	-0.0038	-0.0041	-0.0040	-0.0014	-0.0011	-0.0002	-0.0000	0.0027	0.0024	0.0022	0.0002	-0.0003	0.0005	-0.0001
0.6	-0.0086	-0.0094	-0.0065	-0.0078	-0.0041	-0.0044	-0.0018	0.0046	0.0037	0.0042	0.0038	0.0024	0.0049	0.0031
0.9	0.0177	0.0200	0.0354	0.0108	0.0289	0.0136	0.0253	0.0228	0.0061	0.0159	0.0045	0.0030	0.0055	0.0004
Panel B: RMSE(×100)														
$T=100$														
0.0	0.6472	0.6683	0.6703	0.3772	0.3859	0.3612	0.3688	0.6496	0.6867	0.6899	0.2159	0.2330	0.2025	0.2238
0.3	0.8799	0.8987	0.8921	0.6380	0.6568	0.6176	0.6363	0.8796	0.9273	0.9158	0.5163	0.5774	0.5062	0.5746
0.6	1.4773	1.4994	1.4275	1.1564	1.1757	1.1155	1.1410	1.4555	1.5552	1.5055	1.0942	1.2826	1.0766	1.2958
0.9	5.4550	5.4823	4.0692	4.0765	3.4735	3.4116	3.1577	5.2595	5.3721	4.4137	3.6732	3.9277	3.1029	3.3283
$T=200$														
0.0	0.2147	0.2176	0.2178	0.1256	0.1272	0.1208	0.1220	0.2243	0.2329	0.2335	0.0733	0.0769	0.0695	0.0736
0.3	0.2959	0.2977	0.2947	0.2114	0.2131	0.2038	0.2064	0.3193	0.3259	0.3235	0.1866	0.2007	0.1834	0.1995
0.6	0.5457	0.5565	0.5273	0.4406	0.4418	0.4261	0.4307	0.5435	0.5717	0.5470	0.4195	0.4674	0.4085	0.4598
0.9	2.7259	2.7552	1.9927	2.0914	1.7540	1.7253	1.5986	2.7107	2.8107	2.2141	1.9168	1.9928	1.5689	1.7284
$T=500$														
0.0	0.0534	0.0537	0.0538	0.0310	0.0315	0.0297	0.0302	0.0509	0.0524	0.0525	0.0172	0.0177	0.0163	0.0170
0.3	0.0777	0.0783	0.0777	0.0544	0.0547	0.0526	0.0531	0.0752	0.0777	0.0766	0.0454	0.0467	0.0444	0.0463
0.6	0.1403	0.1422	0.1351	0.1105	0.1099	0.1064	0.1072	0.1407	0.1478	0.1401	0.1079	0.1141	0.1049	0.1128
0.9	0.8935	0.9078	0.6375	0.6805	0.5547	0.5658	0.5195	0.9012	0.9465	0.7258	0.6489	0.6299	0.5367	0.5786
$T=1000$														
0.0	0.0188	0.0189	0.0189	0.0111	0.0111	0.0107	0.0107	0.0185	0.0189	0.0190	0.0058	0.0060	0.0056	0.0058
0.3	0.0273	0.0272	0.0270	0.0192	0.0192	0.0187	0.0187	0.0267	0.0274	0.0272	0.0157	0.0161	0.0155	0.0159
0.6	0.0499	0.0503	0.0478	0.0400	0.0389	0.0384	0.0380	0.0478	0.0495	0.0459	0.0365	0.0379	0.0353	0.0376
0.9	0.3492	0.3562	0.2402	0.2662	0.2081	0.2192	0.1969	0.3433	0.3673	0.2637	0.2482	0.2239	0.2063	0.2128

Table 77: Bias (×1000) and RMSE (×100) for β_4, with estimated long-run variances.

compared to single equation CPR estimation stem from the impact of long-run variance estimation. Looking at Table 95 in Appendix 3.6.2, which shows the result considering true long-run variances instead of estimated long-run variances as we have in Table 78, we observe very similar null rejection probabilities close to the nominal 5% level across all the considered estimators, with only $\rho = 0.9$ leading to inflated null rejection probabilities.

Since the above performed t-tests for the pooled estimation setups (C) and (S) imply a null hypothesis of $H_0 : \beta_{4,1} = \cdots = \beta_{4,N} = -0.2$, we also compare those t-test results with them of Wald-type tests, as proposed in Proposition 5, for exactly the same null hypothesis (labeled as *Wald*). Additionally, for this null hypothesis we consider the results from t-tests of unrestricted estimation for $\beta_{4,i} = -0.2$ for all N equations simultaneously. To account for the multiple testing

problem in this case and controlling the global significance level, we use the modified Bonferroni correction of Hommel (1988) (labeled as MT). The procedure works as follows: First, order the t-statistics in absolute decreasing order $|t_{(1)}| \geq |t_{(2)}| \geq \cdots \geq |t_{(N)}|$. Define $\alpha(j) \coloneqq \frac{j}{C_N}\frac{\alpha}{N}$, with $C_N \coloneqq 1 + \frac{1}{2} + \cdots + \frac{1}{N}$. The null hypothesis is rejected, if $|t_{(j)}| > q_{1-\frac{\alpha(j)}{2}}$ for at least one j, with q_τ denoting the τ-quantile of the standard normal distribution. Table 79 shows the resulting null rejection probabilities for $H_0 : \beta_{4,1} = \cdots = \beta_{4,N} = -0.2$. For $N = 3$, the results for the multiple testing procedure MT are close to the results of the (S)-pooled single t-test results, with MT having slightly lower size distortions. Both methods lead to considerably smaller size distortions than the Wald-type tests. For $N = 10$, the picture is a bit different, for $T = 100$ and $T = 200$, the (S)-pooled SOLS estimator leads to the smallest null rejection probabilities, whereby for $T = 500$ and $T = 1000$ – with the exception of $\rho = 0.9$ – MT SOLS leads to smallest null rejection probabilities. Size distortions of the Wald-type tests are very large for $N = 10$. Again, size distortions increase with increasing ρ and decrease with increasing T. The bad performance of the Wald-type tests and generally the size distortions are mainly driven by the long-run variance estimation. When looking at Table 96, showing the results using the true instead of the estimated long-run variance, we observe for MT, (S) and Wald null rejection probabilities mainly below the nominal level of 5%, with the exception of $\rho = 0.9$. Also, null rejection probabilities remain quite stable across T, again with the exception of $\rho = 0.9$.

ρ	$N = 3$							$N = 10$						
	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)
$T = 100$														
0.0	0.1032	0.1132	0.1164	0.1070	0.1262	0.1100	0.1250	0.0940	0.1544	0.1618	0.1660	0.2502	0.1640	0.2560
0.3	0.1298	0.1428	0.1472	0.1476	0.1692	0.1444	0.1714	0.1304	0.2076	0.2258	0.2078	0.3234	0.2100	0.3448
0.6	0.1620	0.1810	0.1928	0.1978	0.2340	0.2028	0.2458	0.1648	0.2792	0.3262	0.2962	0.4706	0.2996	0.5086
0.9	0.2528	0.2738	0.2870	0.3044	0.3670	0.3248	0.3902	0.2360	0.3708	0.5240	0.4326	0.6816	0.4398	0.7256
$T = 200$														
0.0	0.0746	0.0826	0.0838	0.0932	0.1026	0.0956	0.1034	0.0878	0.1224	0.1266	0.1200	0.1696	0.1174	0.1788
0.3	0.1054	0.1148	0.1180	0.1226	0.1372	0.1206	0.1314	0.1096	0.1548	0.1628	0.1476	0.2196	0.1504	0.2280
0.6	0.1312	0.1398	0.1428	0.1556	0.1746	0.1500	0.1736	0.1368	0.2050	0.2422	0.2156	0.3442	0.2182	0.3674
0.9	0.2008	0.2242	0.2274	0.2478	0.2936	0.2598	0.3104	0.1998	0.3088	0.4482	0.3658	0.5976	0.3636	0.6278
$T = 500$														
0.0	0.0652	0.0690	0.0700	0.0644	0.0710	0.0696	0.0752	0.0626	0.0798	0.0814	0.0808	0.1002	0.0812	0.1002
0.3	0.0894	0.0916	0.0940	0.0944	0.1008	0.0952	0.0994	0.0906	0.1126	0.1164	0.1082	0.1338	0.1102	0.1394
0.6	0.0944	0.0990	0.1038	0.1000	0.1110	0.1000	0.1142	0.0952	0.1312	0.1504	0.1390	0.2022	0.1446	0.2142
0.9	0.1558	0.1738	0.1638	0.1814	0.1916	0.1826	0.2038	0.1516	0.2276	0.3348	0.2532	0.4060	0.2520	0.4142
$T = 1000$														
0.0	0.0542	0.0566	0.0574	0.0676	0.0706	0.0668	0.0684	0.0554	0.0650	0.0668	0.0608	0.0726	0.0596	0.0738
0.3	0.0820	0.0830	0.0806	0.0810	0.0860	0.0846	0.0862	0.0732	0.0894	0.0884	0.0870	0.1046	0.0874	0.1036
0.6	0.0778	0.0806	0.0818	0.0840	0.0852	0.0844	0.0894	0.0760	0.0980	0.1056	0.0996	0.1294	0.0976	0.1380
0.9	0.1186	0.1338	0.1220	0.1436	0.1372	0.1378	0.1386	0.1138	0.1758	0.2530	0.1860	0.2660	0.1792	0.2794

Table 78: Empirical null rejection probabilities of t-tests for $\beta_4 = -0.2$, with estimated long-run variances.

Finally, we turn to size-corrected power simulations. The term *size-corrected* refers to using

ρ	$N=3$						$N=10$					
	MT		Wald		(S)		MT		Wald		(S)	
	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR
$T=100$												
0.0	0.1130	0.1146	0.1810	0.1854	0.1100	0.1250	0.2624	0.2832	0.5902	0.6150	0.1640	0.2560
0.3	0.1552	0.1580	0.2560	0.2594	0.1444	0.1714	0.3398	0.3842	0.7402	0.7916	0.2100	0.3448
0.6	0.2132	0.2188	0.3302	0.3384	0.2028	0.2458	0.4606	0.5678	0.8666	0.9448	0.2996	0.5086
0.9	0.3504	0.3498	0.4686	0.4712	0.3248	0.3902	0.6722	0.8444	0.9326	0.9970	0.4398	0.7256
$T=200$												
0.0	0.0722	0.0748	0.1276	0.1310	0.0956	0.1034	0.1398	0.1508	0.3964	0.4188	0.1174	0.1788
0.3	0.1112	0.1138	0.1804	0.1824	0.1206	0.1314	0.2004	0.2298	0.5204	0.5742	0.1504	0.2280
0.6	0.1452	0.1468	0.2308	0.2358	0.1500	0.1736	0.2896	0.3714	0.6918	0.8360	0.2182	0.3674
0.9	0.2750	0.2564	0.3748	0.3602	0.2598	0.3104	0.5370	0.7448	0.8448	0.9884	0.3636	0.6278
$T=500$												
0.0	0.0484	0.0488	0.0874	0.0906	0.0696	0.0752	0.0616	0.0644	0.1958	0.2080	0.0812	0.1002
0.3	0.0702	0.0734	0.1306	0.1336	0.0952	0.0994	0.0948	0.1004	0.3122	0.3326	0.1102	0.1394
0.6	0.0786	0.0818	0.1382	0.1500	0.1000	0.1142	0.1346	0.1682	0.4210	0.5462	0.1446	0.2142
0.9	0.1826	0.1554	0.2616	0.2324	0.1826	0.2038	0.3412	0.5502	0.6528	0.9352	0.2520	0.4142
$T=1000$												
0.0	0.0388	0.0384	0.0696	0.0698	0.0668	0.0684	0.0390	0.0414	0.1144	0.1216	0.0596	0.0738
0.3	0.0626	0.0624	0.1110	0.1090	0.0846	0.0862	0.0650	0.0700	0.2090	0.2210	0.0874	0.1036
0.6	0.0596	0.0586	0.1040	0.1110	0.0844	0.0894	0.0772	0.0876	0.2600	0.3414	0.0976	0.1380
0.9	0.1240	0.0934	0.1834	0.1492	0.1378	0.1386	0.2168	0.3554	0.4638	0.8132	0.1792	0.2794

Table 79: Empirical null rejection probabilities of corrected t-tests and Wald-type tests for H_0 : $\beta_{4,1} = \cdots = \beta_{4,N} = -0.2$, with estimated long-run variances.

empirical critical values instead of theoretical critical values, such that null rejection probabilities are exactly at 5% under the null hypothesis. This is to make null rejection probabilities under the alternative hypothesis, i. e. , power, more comparable. Under the null hypothesis, we have $\beta_4 = -0.2$ and vary β_4 under the alternative hypothesis on a grid between -0.2 and -0.16 with steps of size 0.002, while keeping all other parameters constant. Figure 41 shows the results for $T = 200$ and $\rho = 0.3$, for both $N = 3$ and $N = 10$. The estimators are clustered in two groups, with the pooled estimators leading to higher size-corrected power. For unrestricted estimation, the results of the different estimators are very close to each other, with a slight lead of CPR followed by SUR. For $N = 10$, we see the opposite picture for pooled estimation, there the SOLS estimators lead to slightly higher size-corrected power compared to their SUR counterparts.

For unrestricted estimation, CPR estimation leads the ranking slightly, with smallest bias, RMSE and size distortions. Nevertheless, the system estimators give the opportunity to test for - and then estimate with - parameter restrictions across equations, which can be a big advantage. From the system estimators, SUR often leads to smaller RMSE, and slightly higher size-corrected power than SOLS, whilst having little higher null rejection probabilities. Pooled estimation can substantially

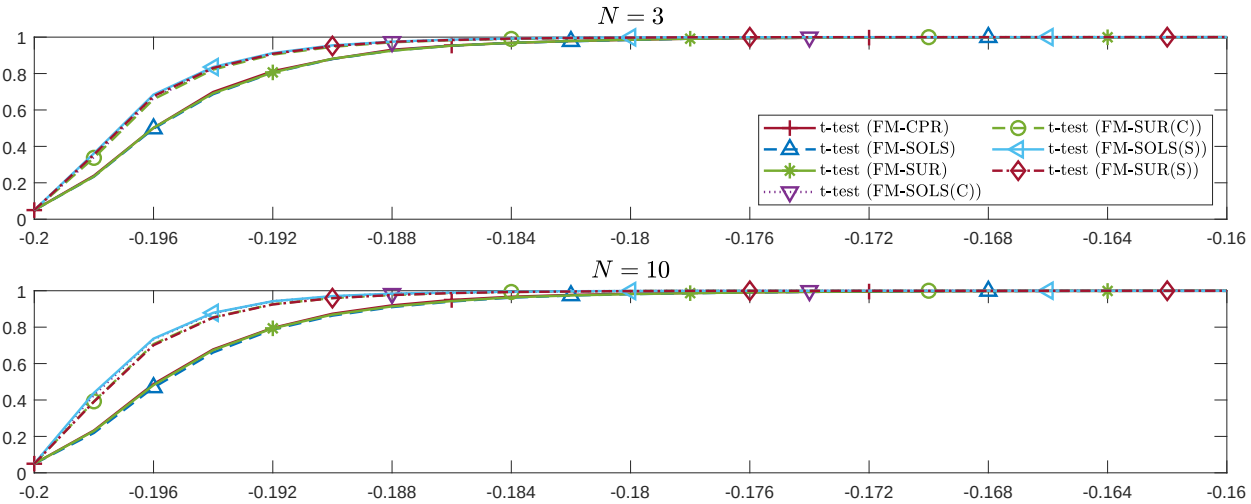

Figure 41: Size-corrected power of t-tests for $H_0 : \beta_4 = -0.2$, with estimated long-run variances, $T = 200$ and $\rho = 0.3$.

improve estimation and size-corrected power, without detrimental impacts on null rejection probabilities. Within the pooled estimators, SOLS is slightly ahead, with lower size distortions and higher size-corrected power.

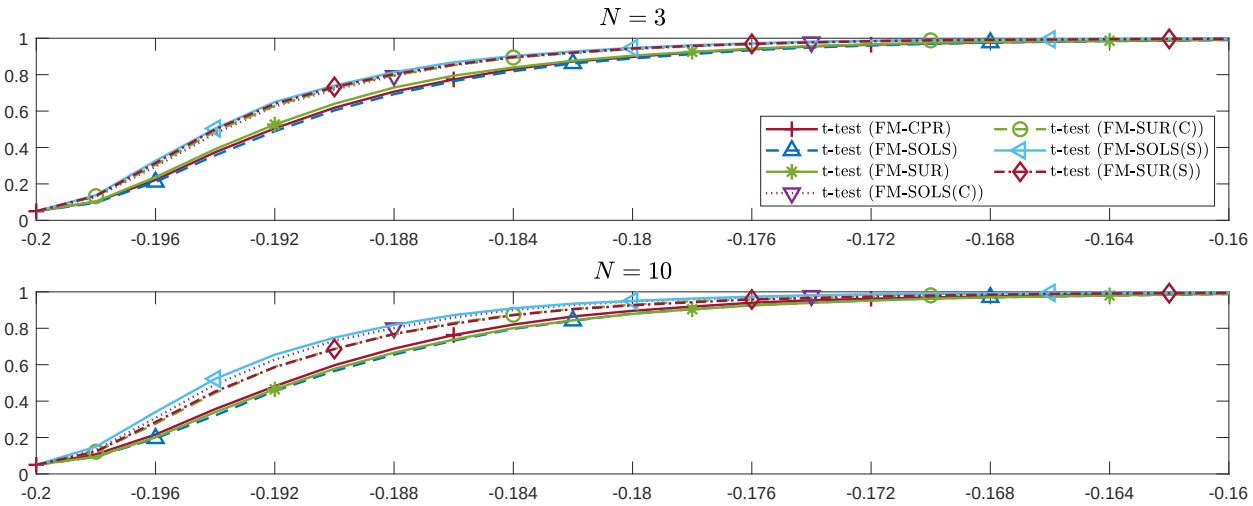

Figure 42: Size-corrected power of t-tests for $H_0 : \beta_4 = -0.2$, with estimated long-run variances, $T = 200$ and $\rho = 0.6$.

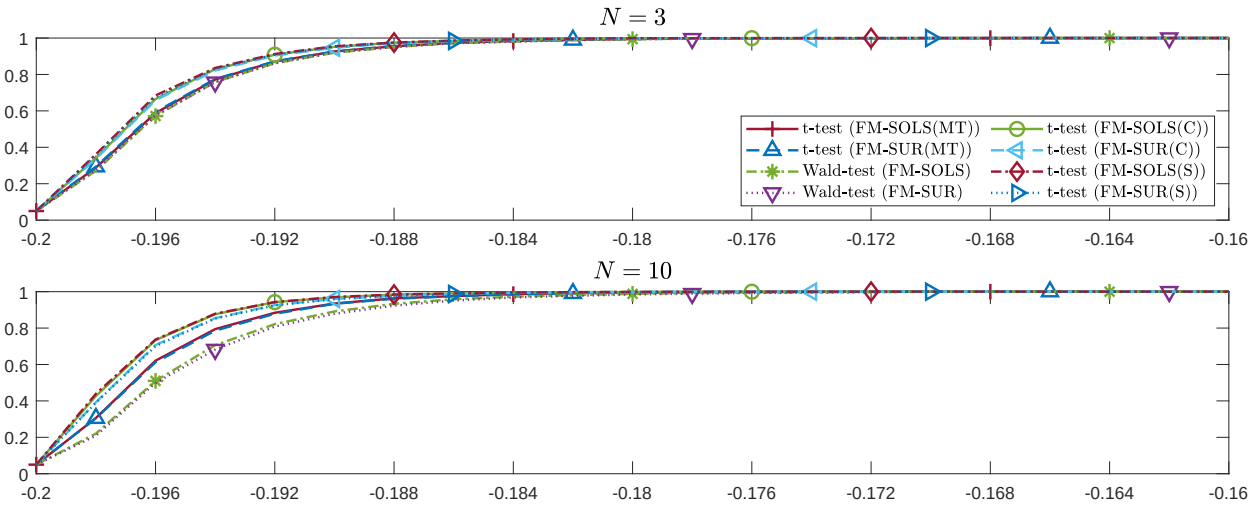

Figure 43: Size-corrected power of tests for $H_0 : \beta_{4,1} = \cdots = \beta_{4,N} = -0.2$, with estimated long-run variances, $T = 200$ and $\rho = 0.3$.

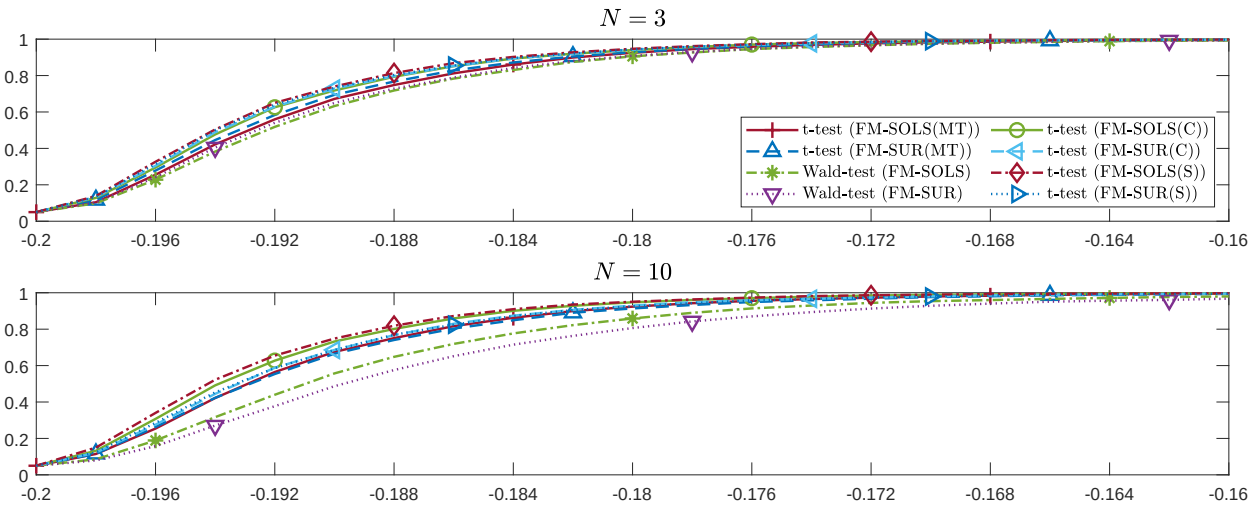

Figure 44: Size-corrected power of tests for $H_0 : \beta_{4,1} = \cdots = \beta_{4,N} = -0.2$, with estimated long-run variances, $T = 200$ and $\rho = 0.6$.

3.4 Empirical Illustration - MKCs and EKCs

We investigate the Material Kuznets Curve (MKC) and Environmental Kuznets Curve (EKC) hypotheses for eight OECD countries and build upon the analyses of Wagner *et al.* (2020) and Grabarczyk *et al.* (2018). We estimate the relationships between a country's level of economic development and once with intensity of material use, and once with pollution, proxied by emissions, in separate equations. This is performed by estimating systems of equations, using the methods from Section 3.2, on a country level and as well on a material and emission level. Table 80 shows the list of countries in our analysis. For the intensity of material use, we consider the metals aluminum

Australia	France	Germany	Italy
Japan	Switzerland	United Kingdom	United States

Table 80: Considered countries for the period 1927–2006.

(Al), lead (Pb) and zinc (Zn), and for the intensity of pollution we consider the emissions of carbon dioxide (CO_2) and sulfur dioxide (SO_2), with GDP as explanatory variable. For each country i, the system of equations looks as follows:

$$y_{i,j,t} = c_{i,j} + \delta_{i,j}t + \beta_{i,j,1}x_{i,\mathrm{C},t} + \beta_{i,j,2}x_{i,\mathrm{C},t}^2 + \beta_{i,j,3}x_{i,\mathrm{C},t}^3 + u_{i,j,t}, \tag{284}$$

with $y_{i,j,t}$ being log *substance* j per capita for country i, with $j \in \{\text{Aluminum, Zinc, Lead, } CO_2, SO_2\}$, and $x_{i,\mathrm{C},t}$ being log GDP per capita for country i, as common integrated regressor.[58] Log GDP per capita serves here - per country - as a common regressor, since it is identical accross substances. Similar to model (284), we estimate systems of equations for each substance, with considering all the countries from Table 80 jointly. For our analysis, we consider annual data from 1927–2006. We do not discuss here dummy variables for specific years nor broken trends, which is out of scope of this paper. Nevertheless, material use and emissions during, e. g. , World War 2, are in many cases already well covered by our model, as you can see in Figure 46 and subsequent figures. We consider for GDP and the metals aluminum, zinc and lead the same underlying data as Grabarczyk *et al.* (2018), restricted to the period 1927–2006, to have for all substances across all countries the same periods. Though it is more restrictive than necessary for some countries or substances, we avoid having different periods for the estimated systems. Due to considering a different period compared to Grabarczyk *et al.* (2018) and Wagner *et al.* (2020), results of unit root tests, (non-)cointegration tests and estimation might also differ. The CO_2 emissions data – which

[58] With *substances*, we refer to the considered materials and emissions jointly.

cover CO_2 emissions from fossil fuel usage – are taken from the Carbon Dioxide Information and Analysis Center (CDIAC), see Boden *et al.* (2018). The SO_2 emissions have been combined from two sources. The data for 1927–2005 are from the NASA Socioeconomic Data and Applications Center (SEDAC), see Smith *et al.* (2011). The data for period 2006 are from the OECD (2020).[59] The CO_2 emissions and SO_2 emissions data are – for a different period – used in Knorre *et al.* (2021).

We perform unit root tests for log GDP per capita and the logs of per-capita substances. For the eight countries investigated, the unit root null hypothesis for log of GDP is at the 5% level only rejected for Australia with the Phillips and Perron (1988) (PP) test and for the United States with the augmented Dickey and Fuller (1981) (ADF) test, as shown in Table 81. For both countries, the respective other test does not reject the null hypothesis at the 5% level (though they reject at the 10% level) and we do not exclude those countries from the analysis. The results of the unit root tests for the substances need to be taken with caution, since in equations using powers of log GDP per capita, the log of the per-capita substance is not an I(1) process anymore.

	ln(GDP)		ln(Al)		ln(Pb)		ln(Zn)		ln(CO_2)		ln(SO_2)	
Country	ADF	PP	ADF	PP	ADF	PP	ADF	PP	ADF	PP	ADF	PP
Australia	*-3.31*	**-3.79**	-0.46	-0.94	-1.16	-0.94	-2.55	-2.81	-1.08	-2.29	**-4.06**	**-4.18**
France	-2.78	-2.32	-0.13	-0.94	*-3.45*	-2.91	*-3.36*	-2.75	-1.73	-1.84	-0.12	-0.45
Germany	*-3.37*	-2.47	*-3.28*	-2.76	-2.47	-2.71	**-3.63**	-3.14	**-3.48**	**-3.47**	-0.57	-0.14
Italy	-2.39	-1.88	*-3.41*	*-3.21*	-2.46	-2.90	-2.77	**-3.50**	-2.37	-2.51	-0.44	0.26
Japan	-1.52	-1.80	-2.20	-1.97	-1.91	-2.26	-1.53	-1.12	-1.93	-1.80	-1.83	-1.57
Switzerland	-0.93	-1.10	-1.40	-3.08	*-3.28*	**-4.17**	-1.34	*-3.36*	-2.62	-2.52	-1.05	-0.97
United Kingdom	*-3.46*	-2.13	-1.96	-1.54	*-3.32*	*-3.40*	-2.66	-2.55	-1.55	-1.60	2.58	2.48
United States	**-4.72**	*-3.21*	-1.21	-1.34	**-3.85**	**-4.03**	-2.46	-2.86	-2.45	-2.67	-0.80	-1.00

Table 81: Results of augmented Dickey-Fuller and Phillips-Perron unit root tests for gross domestic product (GDP), aluminum (Al), lead (Pb), zinc (Zn), carbon dioxide (CO_2) and sulfur dioxide (SO_2). All variables are used in logarithms of per capita quantities. Intercept and linear trend are included in the test equation. The sample period is 1927-2006. *Italic* entries denote rejection of the null hypothesis at the 10% level, and **bold** entries denote rejection at the 5% level

We use the tests for cointegration and non-cointegration for CPRs of Wagner (2023), testing for cubic, quadratic and linear cointegrating MKCs and EKCs, including a constant and a linear trend. Table 82 shows the results of non-cointegration (PU) and cointegration (CT) tests for polynomials of order one to three. For each country-substance combination, we consider in our analysis a polynomial relationship between that country and substance, with using the lowest

[59] 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.

polynomial order for which both conditions apply, first the PU test rejects the null hypothesis of no cointegration and second the CT test does not reject the null hypothesis of cointegration. If for a country-substance combination such a case never occurs for any polynomial order, we neglect that combination from our analysis. For example, in the system of equations for France, we estimate the material use of log lead per capita using log GDP per capita up to order three, and similarly for log zinc per capita as well. No estimation for aluminum, nor carbon dioxide, nor sulfur dioxide takes place. The results of non-cointegration and cointegration tests, and thus the selected set of countries for each substance, may differ to the results from Wagner (2015), Wagner *et al.* (2020) and Grabarczyk *et al.* (2018), due to different considered time periods.

		ln(Al)		ln(Pb)		ln(Zn)		ln(CO_2)		ln(SO_2)	
Country	Order	PU	CT	PU	CT	PU	CT	PU	CT	PU	CT
Australia	1	5.76	**0.30**	8.90	**0.27**	12.71	**0.15**	9.65	**0.28**	**50.87**	**0.15**
	2	44.05	0.09	41.44	0.05	13.97	**0.12**	24.89	*0.11*	**56.85**	0.06
	3	45.26	*0.08*	41.52	0.05	24.10	*0.09*	31.47	0.07	**56.88**	0.06
France	1	10.90	**0.25**	37.75	0.08	18.76	*0.10*	15.85	**0.23**	2.32	**0.25**
	2	22.71	**0.14**	*48.53*	**0.15**	34.62	**0.11**	27.03	**0.17**	4.95	**0.20**
	3	43.95	0.05	**76.53**	0.04	**57.81**	0.05	36.88	0.07	15.91	0.05
Germany	1	31.29	**0.18**	21.22	*0.11*	**48.95**	*0.10*	**57.32**	**0.31**	2.52	**0.25**
	2	*49.84*	**0.11**	25.14	**0.35**	*49.92*	**0.20**	**75.66**	**0.59**	3.92	**0.22**
	3	*51.27*	*0.09*	25.14	**0.26**	*52.87*	**0.15**	**90.05**	**0.21**	10.44	0.08
Italy	1	18.12	**0.16**	38.05	*0.10*	20.95	0.06	25.55	0.07	3.24	**0.26**
	2	29.27	*0.10*	38.14	*0.10*	26.29	**0.19**	30.80	0.06	9.63	**0.17**
	3	31.20	*0.08*	**63.22**	0.06	**79.59**	0.07	33.01	0.04	47.01	0.04
Japan	1	25.72	**0.17**	**49.82**	0.06	7.40	**0.24**	30.85	**0.17**	3.98	**0.16**
	2	*46.85*	0.06	*50.01*	0.04	21.52	**0.21**	*45.92*	0.07	11.04	0.07
	3	46.85	0.06	*50.02*	0.05	30.49	*0.09*	*54.53*	0.06	11.60	0.07
Switzerland	1	**53.66**	0.08	**62.26**	0.08	**52.58**	0.10	27.38	0.08	18.77	0.08
	2	**56.97**	0.07	**62.27**	0.08	**62.42**	0.06	39.33	0.03	21.25	0.08
	3	**57.96**	0.05	**64.00**	0.08	**70.22**	0.04	*48.78*	0.05	40.91	0.06
United Kingdom	1	5.18	**0.14**	35.10	0.06	16.71	**0.17**	8.81	**0.17**	1.59	**0.18**
	2	16.31	**0.13**	35.41	0.07	34.55	**0.12**	18.18	0.08	11.85	**0.11**
	3	30.66	0.06	37.10	0.05	42.93	0.07	18.33	*0.08*	28.66	0.04
United States	1	5.05	**0.29**	**49.68**	0.09	30.84	**0.21**	26.13	**0.22**	7.92	**0.26**
	2	**59.21**	*0.10*	*50.68*	0.04	*52.78*	**0.11**	37.20	**0.11**	20.13	**0.13**
	3	**59.42**	*0.10*	*52.69*	0.06	**67.76**	0.06	38.75	*0.09*	34.40	0.08

Table 82: Results of PU non-cointegration and CT cointegration tests for aluminum (Al), lead (Pb), zinc (Zn), carbon dioxide (CO_2) and sulfur dioxide (SO_2). With GDP serving as common integrated regressor. All variables are used in logarithms of per capita quantities and powers up to order 3. Intercept and linear trend are included in the test equations. The period is 1927-2006.

Table 83 shows for each country the results of estimating MKCs and EKCs. The FM-CPR results are from single equation estimation and the FM-SUR results are from SUCPR estimation, using

all the substances for which we estimated a cointegrating polynomial relationship. The polynomial degrees are determined from the results in Table 82. For system estimation, we only show the FM-SUR results, with the FM-SOLS results being identical to the FM-SUR results by construction, when using only a single common integrated regressor (and potentially powers) and no individual specific integrated regressors. The following results emerge from the table: For quadratic MKCs and EKCs, e. g. , for aluminum and carbon dioxide for Japan, the coefficients for squared log GDP per capita have negative signs, thus indicating indeed an inverse u-shaped relationship. Therefore, after reaching the turning point, material use and emissions are expected to decrease with increasing per-capita GDP. In contrast, all estimated linear cointegrating relationships have a positive slope for log GDP per capita, implying an increase in material use and emissions for increasing GDP per capita. For the estimated cubic cointegrating relationships, with the exception of carbon dioxide in Switzerland, the coefficient of the cubic log GDP per capita is positive, indicating an increase in material use and emissions for increasing GDP per capita, from the second turning point onwards. An encouraging observation is, that for most country-substance pairs the time trend is estimated to be negative, which dampens the effect of a positive slope for log GDP per capita in the estimated linear cointegrating relationships. Differences between single equation FM-CPR estimation and FM-SUR system estimation are rather minor, with the exception of carbon dioxide for Switzerland, where we observe larger differences between the parameter estimates.

Table 84 shows results similar to Table 83 but now with system estimation per substance instead of per country. The single equation FM-CPR results are identical to those in Table 83 but the FM-SUR results for a country-substance pair may now differ to the results in Table 83, since here we collect different countries for a given substance in a system of equations and not different substances for a given country. This also leads to not having one common integrated regressor per system but a set of individual specific integrated regressor, with potentially powers. The idea of substance-wise estimation instead of country-wise is, that we observe similar estimation results for a substance using different countries more often than similar results for different substances within a country. For example, the results for lead are similar within groups of countries with an estimated cubic relationship and countries with an estimated linear relationship. Within those groups, the parameter estimates are quite close across countries. The estimation results for carbon dioxide are not comparable with those from Wagner *et al.* (2020), since we estimate only for Japan a quadratic relationship, which is rejected in Wagner *et al.* (2020).

		$\hat{c}$	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
			Australia			
Sulfur dioxide (SO_2)	FM-CPR	**-20.15**	-0.00	**4.53**	*-0.20*	
	FM-SUR	**-20.15**	-0.00	**4.53**	*-0.20*	
			France			
Lead (Pb)	FM-CPR	**-572.57**	**-0.03**	**185.66**	**-20.11**	**0.73**
	FM-SUR	**-575.88**	**-0.04**	**186.93**	**-20.28**	**0.74**
Zinc (Zn)	FM-CPR	**-386.25**	**-0.02**	**122.71**	**-13.03**	**0.47**
	FM-SUR	**-380.86**	**-0.02**	**120.84**	**-12.81**	**0.46**
			Germany			
Zinc (Zn)	FM-CPR	**-13.86**	**-0.03**	**1.80**		
	FM-SUR	**-13.86**	**-0.03**	**1.80**		
			Italy			
Lead (Pb)	FM-CPR	**-440.53**	**-0.06**	**145.94**	**-16.28**	**0.62**
	FM-SUR	**-438.76**	**-0.06**	**145.42**	**-16.23**	**0.61**
Zinc (Zn)	FM-CPR	**-611.10**	**-0.02**	**201.88**	**-22.32**	**0.83**
	FM-SUR	**-614.67**	**-0.02**	**203.01**	**-22.44**	**0.83**
			Japan			
Aluminum (Al)	FM-CPR	**-83.53**	*0.02*	**17.87**	**-0.94**	
	FM-SUR	**-74.88**	0.02	**15.77**	**-0.81**	
Lead (Pb)	FM-CPR	**-14.37**	**-0.05**	**1.92**		
	FM-SUR	**-14.23**	**-0.05**	**1.90**		
Carbon dioxide (CO_2)	FM-CPR	**-17.24**	**-0.01**	**3.29**	**-0.13**	
	FM-SUR	**-15.77**	**-0.01**	**2.94**	**-0.11**	
			Switzerland			
Aluminum (Al)	FM-CPR	**-14.57**	0.00	**1.75**		
	FM-SUR	**-14.29**	0.00	**1.72**		
Lead (Pb)	FM-CPR	**-19.89**	**-0.06**	**2.44**		
	FM-SUR	**-20.31**	**-0.06**	**2.49**		
Zinc (Zn)	FM-CPR	**-20.98**	**-0.04**	**2.49**		
	FM-SUR	**-20.74**	**-0.04**	**2.46**		
Carbon dioxide (CO_2)	FM-CPR	**2258.66**	*-0.02*	**-724.04**	**77.13**	**-2.73**
	FM-SUR	**1825.07**	**-0.03**	**-580.99**	**61.38**	**-2.15**
			United States			
Aluminum (Al)	FM-CPR	**-201.51**	**0.04**	**42.34**	**-2.21**	
	FM-SUR	**-202.67**	**0.04**	**42.60**	**-2.23**	
Lead (Pb)	FM-CPR	**-10.68**	**-0.02**	**1.38**		
	FM-SUR	**-10.79**	**-0.02**	**1.40**		
Zinc (Zn)	FM-CPR	**-518.06**	*-0.01*	**159.41**	**-16.29**	**0.56**
	FM-SUR	**-490.37**	*-0.01*	**150.55**	**-15.35**	**0.52**

Table 83: Results of FM-CPR, FM-SOLS and FM-SUR estimation. The model includes as deterministics intercept and linear trend. *Italic* entries indicate significance of coefficients at the 10% level, and **bold** entries significance of coefficients at the 5% level. The sample period is 1927-2006.

		$\hat{c}$	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Aluminum (Al)						
Japan	FM-CPR	**-83.53**	*0.02*	**17.87**	**-0.94**	
	FM-SUR	**-74.66**	**0.01**	**15.68**	**-0.80**	
Switzerland	FM-CPR	**-14.57**	0.00	**1.75**		
	FM-SUR	**-16.08**	-0.00	**1.93**		
United States	FM-CPR	**-201.51**	**0.04**	**42.34**	**-2.21**	
	FM-SUR	**-203.53**	**0.04**	**42.76**	**-2.23**	
Lead (Pb)						
France	FM-CPR	**-572.57**	**-0.03**	**185.66**	**-20.11**	**0.73**
	FM-SUR	**-546.59**	**-0.03**	**176.09**	**-18.95**	**0.68**
Italy	FM-CPR	**-440.53**	**-0.06**	**145.94**	**-16.28**	**0.62**
	FM-SUR	**-412.52**	**-0.06**	**135.93**	**-15.09**	**0.57**
Japan	FM-CPR	**-14.37**	**-0.05**	**1.92**		
	FM-SUR	**-14.77**	**-0.05**	**1.98**		
Switzerland	FM-CPR	**-19.89**	**-0.06**	**2.44**		
	FM-SUR	**-16.16**	**-0.05**	**1.99**		
United States	FM-CPR	**-10.68**	**-0.02**	**1.38**		
	FM-SUR	**-11.11**	**-0.03**	**1.44**		
Zinc (Zn)						
France	FM-CPR	**-386.25**	**-0.02**	**122.71**	**-13.03**	**0.47**
	FM-SUR	**-413.77**	**-0.02**	**131.15**	**-13.86**	**0.49**
Germany	FM-CPR	**-13.86**	**-0.03**	**1.80**		
	FM-SUR	**-13.46**	**-0.03**	**1.76**		
Italy	FM-CPR	**-611.10**	**-0.02**	**201.88**	**-22.32**	**0.83**
	FM-SUR	**-645.95**	**-0.02**	**213.99**	**-23.72**	**0.88**
Switzerland	FM-CPR	**-20.98**	**-0.04**	**2.49**		
	FM-SUR	**-21.34**	**-0.04**	**2.52**		
United States	FM-CPR	**-518.06**	*-0.01*	**159.41**	**-16.29**	**0.56**
	FM-SUR	**-446.70**	**-0.01**	**136.75**	**-13.90**	**0.47**
Carbon dioxide (CO_2)						
Japan	FM-CPR	**-17.24**	**-0.01**	**3.29**	**-0.13**	
	FM-SUR	**-17.51**	**-0.01**	**3.35**	**-0.13**	
Switzerland	FM-CPR	**2258.66**	*-0.02*	**-724.04**	**77.13**	**-2.73**
	FM-SUR	**1706.11**	**-0.02**	**-545.37**	**57.90**	**-2.04**
Sulfur dioxide (SO_2)						
Australia	FM-CPR	**-20.15**	-0.00	**4.53**	*-0.20*	
	FM-SUR	**-20.15**	-0.00	**4.53**	*-0.20*	

Table 84: Results of FM-CPR, FM-SOLS and FM-SUR estimation. The model includes as deterministics intercept and linear trend *Italic* entries indicate significance of coefficients at the 10% level, and **bold** entries significance of coefficients at the 5% level. The sample period is 1927-2006.

Country (System)	Substances	Polynomial Degree
Australia	Sulfur dioxide (SO_2)	2
France	Lead (Pb)	3
	Zinc (Zn)	3
Germany	Zinc (Zn)	1
Italy	Lead (Pb)	3
	Zinc (Zn)	3
Japan	Aluminum (Al)	2
	Lead (Pb)	1
	Carbon dioxide (CO_2)	2
Switzerland	Aluminum (Al)	1
	Lead (Pb)	1
	Zinc (Zn)	1
	Carbon dioxide (CO_2)	3
United States	Aluminum (Al)	2
	Lead (Pb)	1
	Zinc (Zn)	3

Table 85: Substances with polynomial degree for each country's system of equation. Intercept and linear trend are included in the equations. The period is 1927-2006.

Substance (System)	Countries	Polynomial Degree
Aluminum (Al)	Japan	2
	Switzerland	1
	United States	2
Lead (Pb)	France	3
	Italy	3
	Japan	1
	Switzerland	1
	United States	1
Zinc (Zn)	France	3
	Germany	1
	Italy	3
	Switzerland	1
	United States	3
Carbon dioxide (CO_2)	Japan	2
	Switzerland	3
Sulfur dioxide (SO_2)	Australia	2

Table 86: Countries with polynomial degree for each substance's system of equation. Intercept and linear trend are included in the equations. The period is 1927-2006.

The country-wise estimation results from Table 83 are displayed graphically in Figures 45 to 51. On the left hand side of each figure, you can see the estimated MKCs and EKCs, constructed by using 80 equidistant values for the explanatory variable from the range of log GDP per capita associated with *fitted* values of the time trend and inserting these values in Equation (284), using the coefficient estimates from both FM-CPR (solid blue line) and FM-SUR (dashed red line with plus-marks). The fitted time trend is determined by the following approach: We regress the time trend from the estimated system of equations, i. e. , $1, 2, \ldots, 80$, on the observed log GDP per capita values together with an intercept. The coefficients from this regression are used to calculate fitted time trend values, based on the equidistant log GDP per capita grid together with an intercept. The reason of this approach using a fitted time trend is that an equidistant time trend leads to distortions in the relationship between trend and observed log GDP per capita and thus to distorted estimated MKCs and EKCs. Additionally, the left hand side graphs of each figure include the scatter plots between log GDP per capita and log per-capita of the respective substance. The very similar coefficient estimates from FM-CPR and FM-SUR translate into very similar, or even indistinguishable, estimated MKCs and EKCs. The right hand side graphs of the figures display the actual values of log per-capita of the respective substance with the fitted values obtained from both FM-CPR and FM-SUR estimation. The two fitted value lines are again very close to each other for all substances in each country, due to the very similar coefficient estimates. The fit is - including periods of the second world war - in general quite good and especially very good for country-substance pairs using quadratic or cubic relationships, such as for the substances in France and Italy.

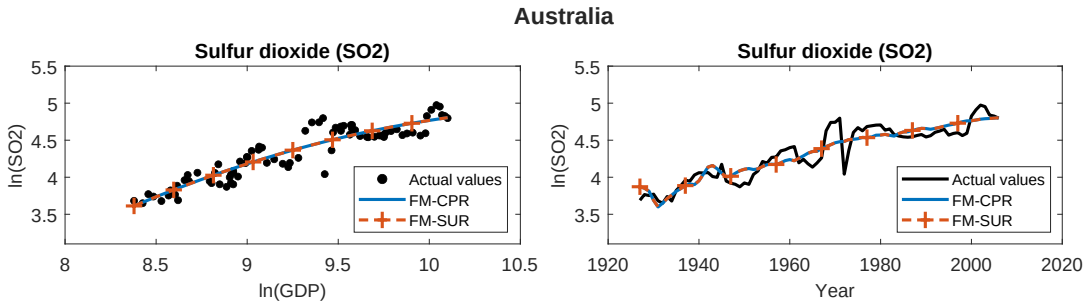

Figure 45: EKC estimation results for Australia for Equation (284). The left hand side shows a scatter plot and an EKC. The dots show the pairs of observations of ln(GDP) per capita and $\ln(SO_2)$ emissions per capita. The lines show results based on inserting 80 equidistant points from the sample range of ln(GDP) per capita, with corresponding fitted values of the linear trend as described in the main text. The solid blue line corresponds to the FM-CPR estimates and the red dashed line to the FM-SUR estimates. The right hand side shows actual and fitted values. The solid black line shows the actual values of $\ln(SO_2)$ per capita emissions, color codes for FM-CPR and FM-SUR estimates remain unchanged.

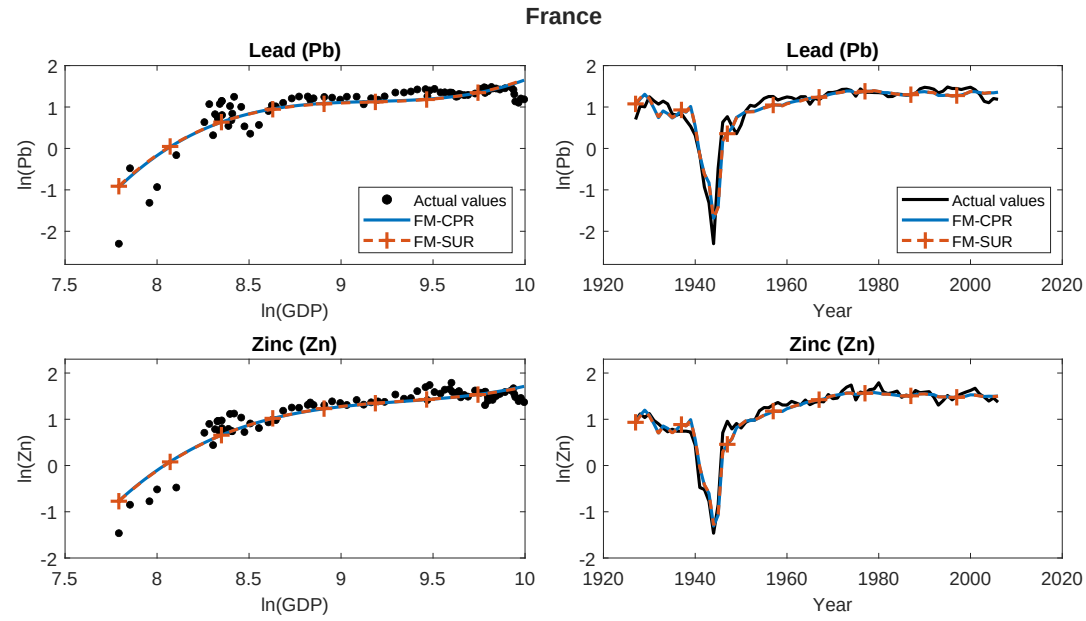

Figure 46: MKC estimation results for France. For further explanations see notes to Figure 45.

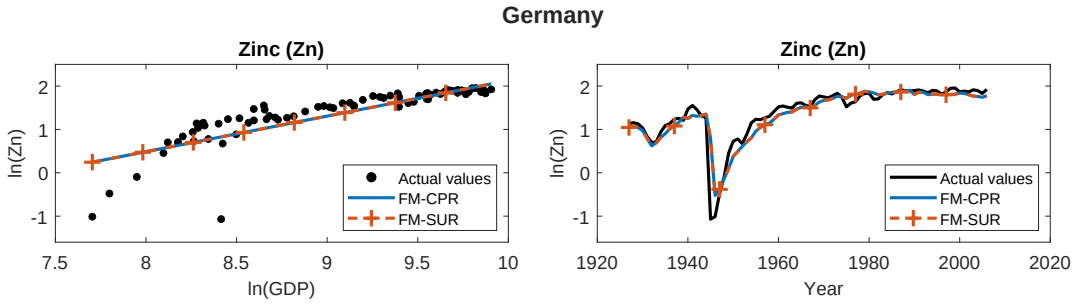

Figure 47: MKC estimation results for Germany. For further explanations see notes to Figure 45.

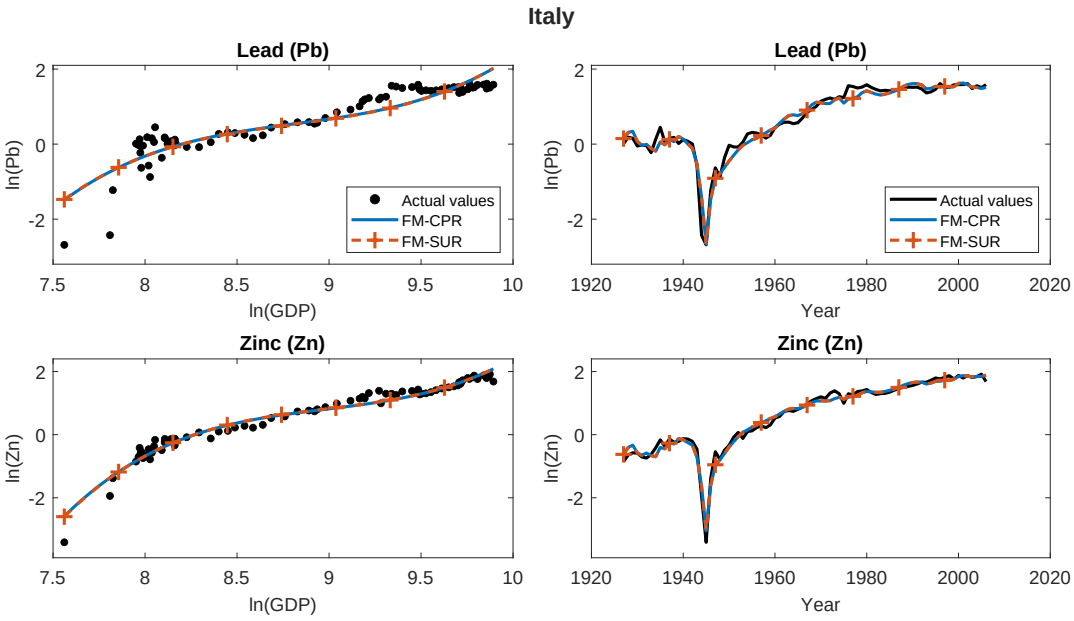

Figure 48: MKC estimation results for Italy. For further explanations see notes to Figure 45.

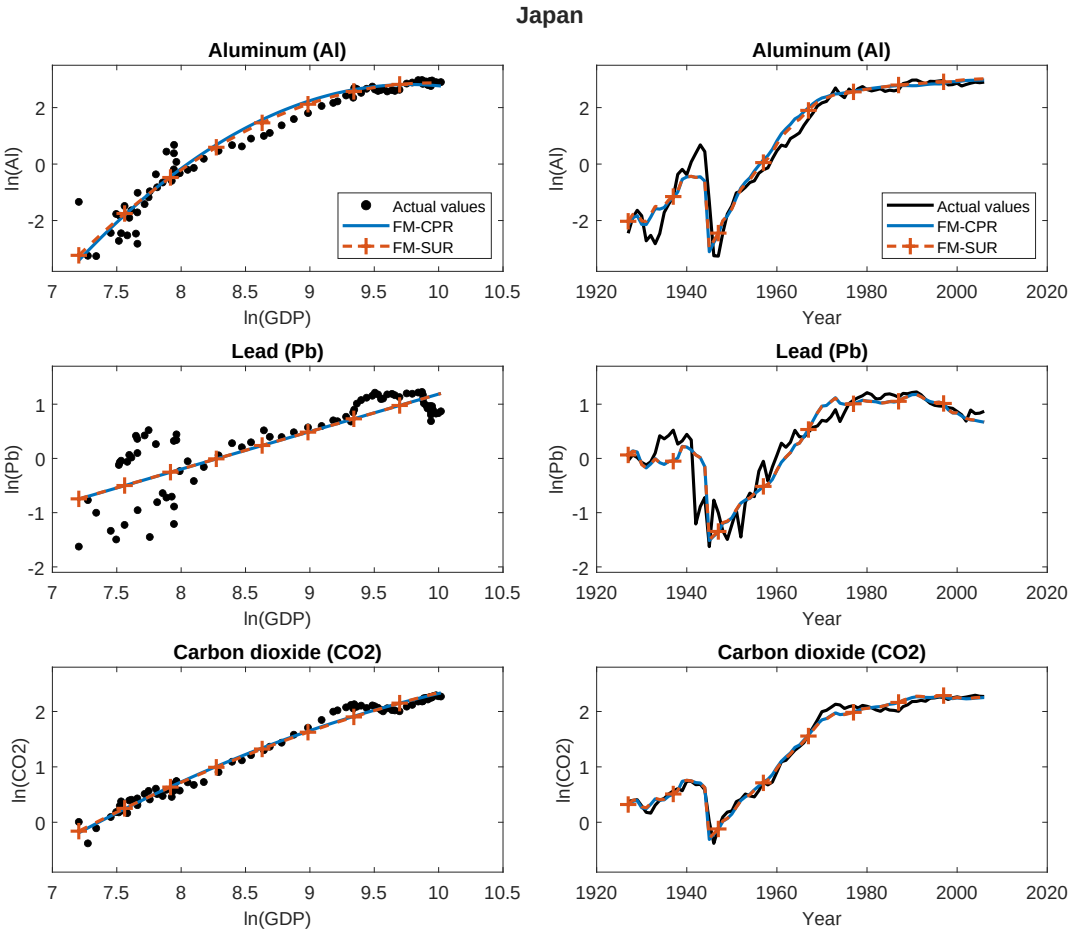

Figure 49: MKC and EKC estimation results for Japan. For further explanations see notes to Figure 45.

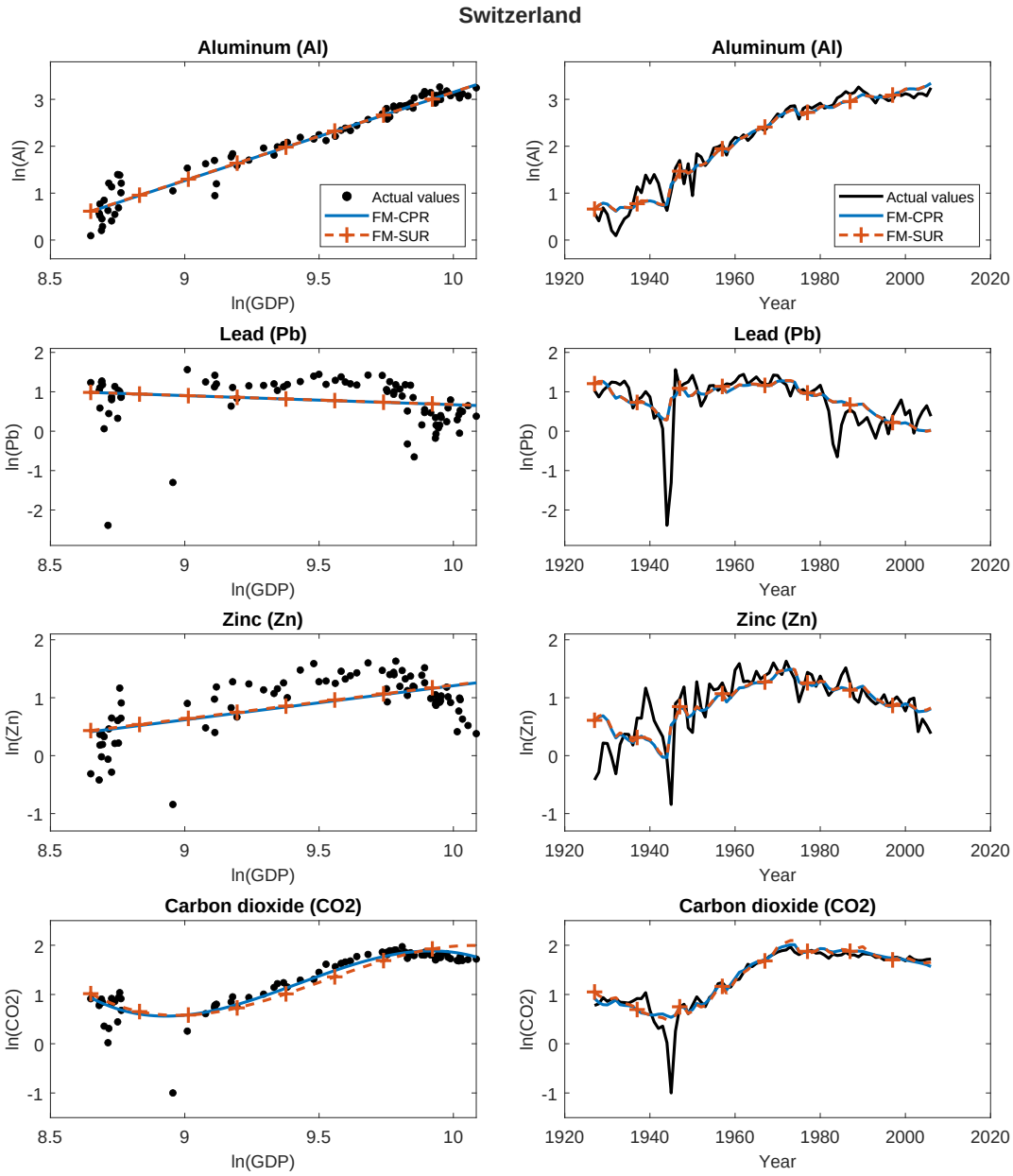

Figure 50: MKC and EKC estimation results for Switzerland. For further explanations see notes to Figure 45.

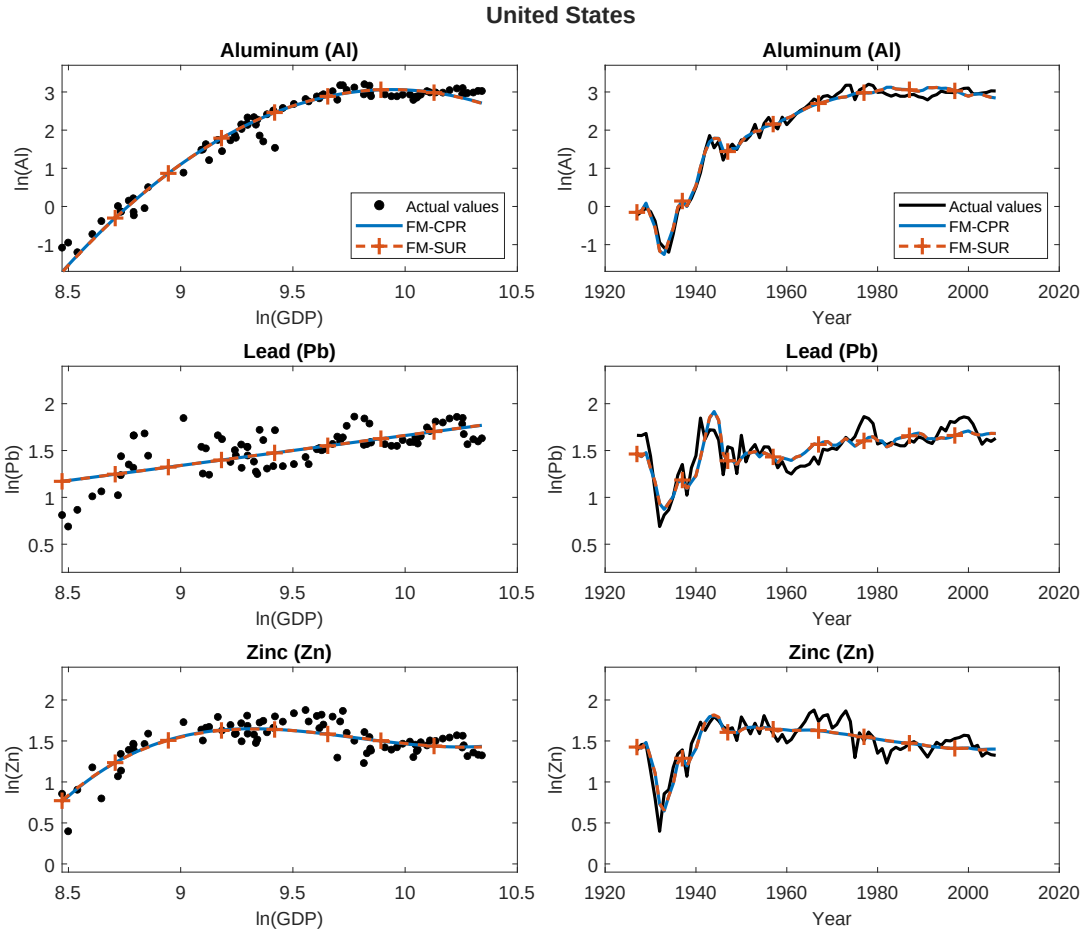

Figure 51: MKC estimation results for United States. For further explanations see notes to Figure 45.

3.4.1 Group-wise Pooling

Based on the FM-SUR estimation results in Tables 83 and 84, we perform tests on group-wise poolability as described in Subsection 3.2.4. For these tests, we consider to pool parameters across substances for a given country, as well as to pool parameters across countries for a given substance. Substances and countries are selected for poolability testing if their estimation results are *close* to each other. We test for pooling the coefficients for the trend (T) and pooling the coefficients for the stochastic (S) or common stochastic (CS) regressor, including potential powers. For the pooling hypotheses (S) and (CS), we only consider groups in which all members have the same polynomial order for that regressor. Then we test for poolability for coefficients up to the highest order. The groups tested for pooling range from sizes two to five, with five being the size of our largest estimated system of equations. All tests are performed at a nominal level of 5%. Tables 87 and 88 show the combinations of materials and countries, for which the null hypothesis of poolability is not rejected. We observe more pooling possibilities for substance-wise estimation compared to country-wise estimation. Thus, material use or emissions are more similar between different countries for a given material than between different materials for a given country. For Switzerland, parameters for the deterministic trend can be pooled for the pairs lead and zinc, or zinc and carbon dioxide, as shown in Table 87. With respect to the common stochastic regressor, parameters can be pooled for lead and zinc. Testing on pooling the trend slope δ across the three substances lead, zinc, and carbon dioxide jointly, leads to a rejection of the null hypothesis. The test to pool both the coefficient for the deterministic trend and the coefficient for the common stochastic regressor, across the materials lead and zinc for Switzerland, also leads to a rejection of the null hypothesis.

	(T)	(CS)
Switzerland	2: Pb-Zn, Zn-CO2	2: Pb-Zn

Table 87: List of group members corresponding to Wald-type tests for poolability.

When looking on the poolability test results for substance-wise system estimation displayed in Table 88, there are several possible ways of pooling for lead for both pooling the trend slope and pooling the slope of the stochastic regressor. The trend parameter can be pooled for four country-pairs and one country-triple (containing IT, JP and CH). The slope parameters β_1 can be pooled for two country-pairs. We also tested whether the trend parameter can be pooled jointly for the two groups Italy, Japan and Switzerland, and France and the United States, together with pooling

the slope parameter β_1 for Japan and Switzerland, which has not been rejected. For zinc, the trend parameter δ can be pooled for two country-pairs and one country-triple (containing FR, DE and IT). For carbon dioxide, the trend parameter δ can be pooled for just one country-pair.

	(T)	(S)
Lead (Pb)	2: IT-JP, IT-CH, JP-CH, FR-USA 3: IT-JP-CH	2: JP-CH, CH-USA
Zinc (Zn)	2: FR-IT, DE-IT 3: FR-DE-IT	
Carbon dioxide (CO_2)	2: JP-CH	

Table 88: List of group members corresponding to Wald-type tests for poolability.

Tables 89 and 90 show the results of group-wise pooled estimation for two systems of equations, one for Switzerland and one for lead. In the first system, for the country Switzerland, we pool the parameter for the common integrated regressor, i. e. , β_1, for lead and zinc. All other parameters remain unchanged. Nevertheless, we show the results for all substances estimated in this system, as the estimates for substances for which no parameter is pooled are also influenced by the pooling of parameters for other substances The results of the group-wise pooled estimation are indicated by "FM-SUR *(gw)*", whereby the FM-CPR and FM-SUR estimates are the results from the previous single equation and system estimations from Table 83. The results change only very slightly when moving from unpooled to pooled FM-SUR estimation, but now with the advantage of estimating one parameter less. Even though the pooling test, which tests for the simultaneous pooling of trend slope δ and slope parameter β_1, was rejected for lead and zinc and thus only one of the two parameters can be pooled (see Table 87), the parameter estimates for lead and zinc are very close, reflecting very similar material use over time in Switzerland. The use of aluminum is also in a similar region, albeit with slightly different values compared to lead and zinc. We make similar observations for the pooled estimation results for lead in Table 90, where the trend slope δ is pooled for Italy, Japan and Switzerland, as well as for France and the United States. Additionally, the slope parameter β_1 is pooled for Japan and Switzerland. Thus, four parameters less are estimated compared to the unpooled FM-SUR estimation. The results from pooled FM-SUR estimation again differ only slightly from unpooled estimation.

The pooled estimation results for Switzerland and lead from Tables 89 and 90 are displayed graphically in Figures 52 and 53. Similarly to Figures 45 to 51, on the left hand side of each figure, you can see the estimated MKCs and EKCs, again constructed by using 80 equidistant

		$\hat{c}$	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
		Switzerland				
Aluminum (Al)	FM-CPR	**-14.57**	0.00	**1.75**		
	FM-SUR	**-14.29**	0.00	**1.72**		
	FM-SUR *(gw)*	**-14.34**	0.00	**1.73**		
Lead (Pb)	FM-CPR	**-19.89**	**-0.06**	**2.44**		
	FM-SUR	**-20.31**	**-0.06**	**2.49**		
	FM-SUR *(gw)*	**-20.17**	**-0.06**	**2.47**		
Zinc (Zn)	FM-CPR	**-20.98**	**-0.04**	**2.49**		
	FM-SUR	**-20.74**	**-0.04**	**2.46**		
	FM-SUR *(gw)*	**-20.78**	**-0.04**	**2.47**		
Carbon dioxide (CO_2)	FM-CPR	**2258.66**	*-0.02*	**-724.04**	**77.13**	**-2.73**
	FM-SUR	**1825.07**	**-0.03**	**-580.99**	**61.38**	**-2.15**
	FM-SUR *(gw)*	**1825.10**	**-0.03**	**-581.00**	**61.38**	**-2.15**

Table 89: Results of FM-CPR, FM-SUR and pooled FM-SUR estimation. The model includes as deterministics intercept and linear trend. The parameter $\hat{\beta}_1$ is pooled for lead and zinc. *Italic* entries indicate significance of coefficients at the 10% level, and **bold** entries significance of coefficients at the 5% level. The sample period is 1927-2006.

values for the explanatory variable from the range of log GDP per capita associated with the fitted values of the time trend and inserting these values in Equation (284), using the coefficient estimates from FM-CPR (solid blue line), FM-SUR (dashed red line with plus-marks) and pooled FM-SUR (dash-dotted orange line) estimation. The right hand side graphs of the figures display the actual values of log per-capita of the respective substance with the fitted values obtained from FM-CPR, FM-SUR and pooled FM-SUR estimation, with the same color coding as on the left hand side. For both Switzerland and lead, for all substances and countries, the differences between FM-SUR and pooled FM-SUR estimation are hardly recognizable visually.

		$\hat{c}$	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
		Lead (Pb)				
France	FM-CPR	**-572.57**	**-0.03**	**185.66**	**-20.11**	**0.73**
	FM-SUR	**-546.59**	**-0.03**	**176.09**	**-18.95**	**0.68**
	FM-SUR *(pooled)*	**-562.59**	**-0.03**	**181.23**	**-19.49**	**0.70**
Italy	FM-CPR	**-440.53**	**-0.06**	**145.94**	**-16.28**	**0.62**
	FM-SUR	**-412.52**	**-0.06**	**135.93**	**-15.09**	**0.57**
	FM-SUR *(pooled)*	**-406.59**	**-0.05**	**133.79**	**-14.82**	**0.56**
Japan	FM-CPR	**-14.37**	**-0.05**	**1.92**		
	FM-SUR	**-14.77**	**-0.05**	**1.98**		
	FM-SUR *(pooled)*	**-14.31**	**-0.05**	**1.91**		
Switzerland	FM-CPR	**-19.89**	**-0.06**	**2.44**		
	FM-SUR	**-16.16**	**-0.05**	**1.99**		
	FM-SUR *(pooled)*	**-15.27**	**-0.05**	**1.91**		
United States	FM-CPR	**-10.68**	**-0.02**	**1.38**		
	FM-SUR	**-11.11**	**-0.03**	**1.44**		
	FM-SUR *(pooled)*	**-11.80**	**-0.03**	**1.52**		

Table 90: Results of FM-CPR, FM-SUR and pooled FM-SUR estimation. The model includes as deterministics intercept and linear trend. *Italic* entries indicate significance of coefficients at the 10% level, and **bold** entries significance of coefficients at the 5% level. The sample period is 1927-2006.

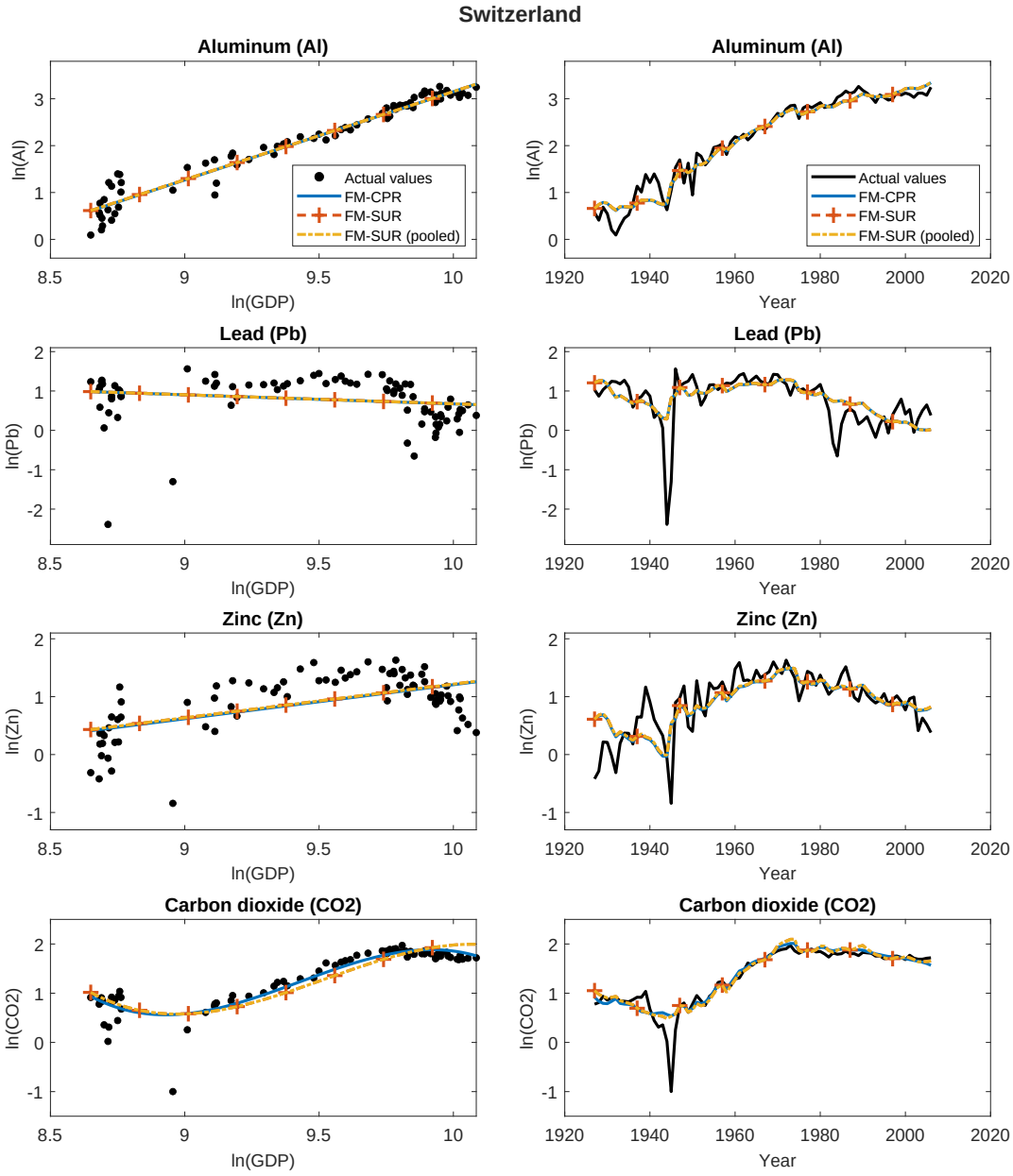

Figure 52: Pooled MKC and EKC estimation results for Switzerland for Equation (284). The parameter $\hat{\beta}_1$ is pooled for lead and zinc. The left hand side shows scatter plots and MKCs and EKCs. The dots show the pairs of observations of log GDP per capita and log substance per capita. The lines show results based on inserting 80 equidistant points from the sample range of ln(GDP) per capita, with corresponding fitted values of the linear trend as described in the main text. The solid blue line corresponds to the FM-CPR estimates, the red dashed line to the FM-SUR estimates and the orange dashed-dotted line to the pooled FM-SUR estimates. The right hand side shows actual and fitted values. The solid black line shows the actual values of log per capita substances, color codes for FM-CPR, FM-SUR and pooled FM-SUR estimates remain unchanged.

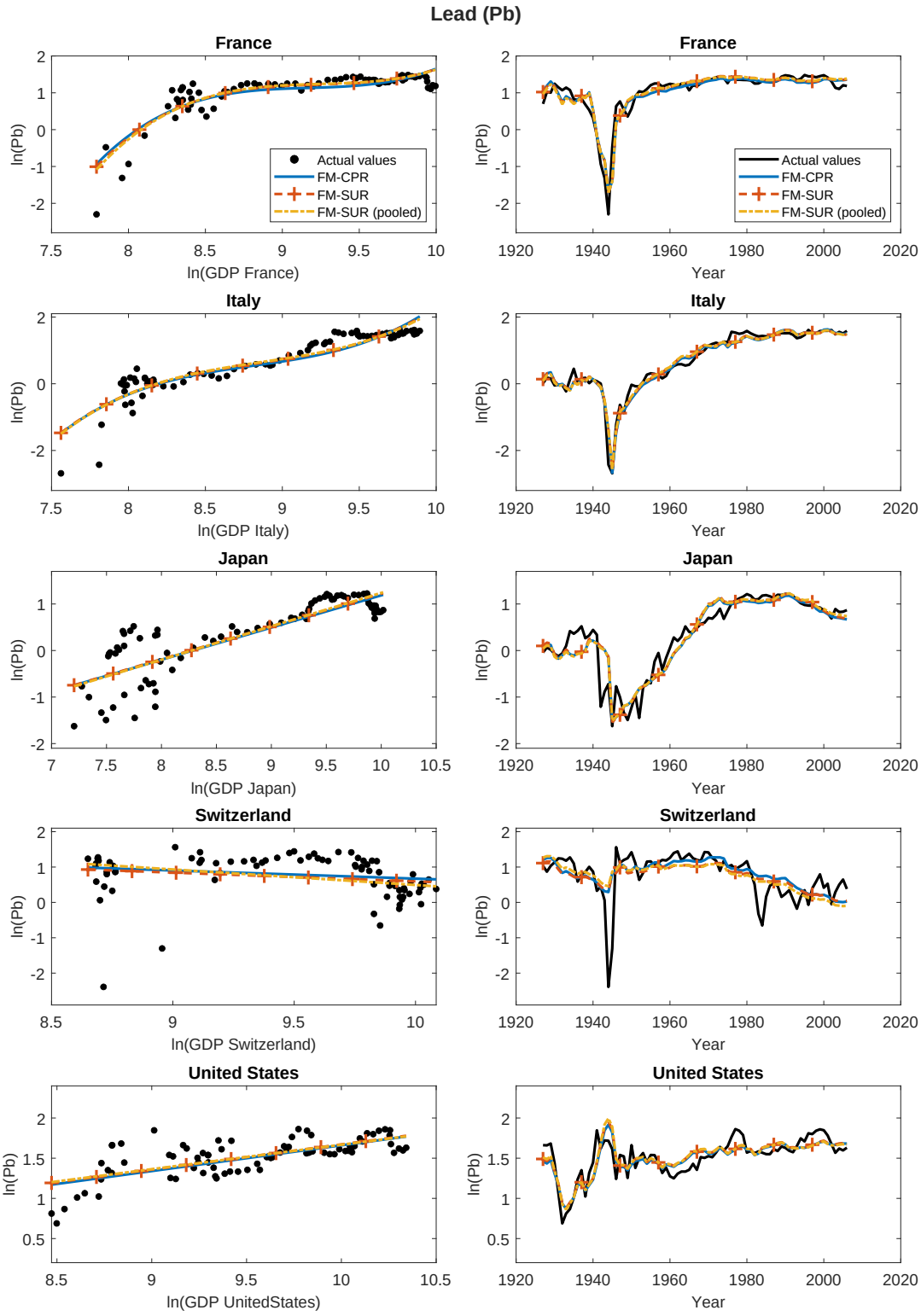

Figure 53: Pooled estimation results for lead. The trend parameter δ is pooled for Italy, Japan and Switzerland, as well as for France and the United States. Additionally, the slope parameter β_1 is pooled for Japan and Switzerland. For further explanations see notes to Figure 52.

3.5 Summary and Conclusion

This paper extends the *fully modified*-SOLS and -SUR estimators of Wagner *et al.* (2020) to systems including common integrated regressors. In addition, the setup – including all the notation – is more generalized to cope with systems including more than regressor per equation. The degree of the deterministic trend, numbers of integrated regressors and common integrated regressors, as well as their powers, are allowed to differ across equations. This flexibility is also covered by the MATLAB code for the proposed estimators. Zero-mean Gaussian mixture limiting distributions are provided for the two estimators and form the basis for asymptotic standard inference. In addition to hypothesis testing, RESET-type specification testing has been discussed. One major advantage of the proposed system estimators compared to single equation estimation is the possibility to test for group-wise pooling and – if supported by the test results – perform group-wise pooled estimation which reduces the number of estimated parameters and increases estimation precision. This is also reflected in the simulation study we performed and discussed. Testing on group-wise pooling in our empirical illustration indicates similarities in the deterministic trend slopes of the material use of lead and zinc, and the emissions of carbon dioxide across several countries for our analysis period 1927–2006. Though the provided estimators allow for great model flexibility, a possible extension would be the addition of stationary stochastic regressors, which however is not a straightforward exercise.

3.6 Appendix

3.6.1 Proofs

Proof of Proposition 4:

The proposition follows from Proof of Proposition 1 in Wagner and Hong (2016) together with straightforward changes of Proof of Proposition 1 in Wagner *et al.* (2020): Starting points for the FM-type bias correction terms are the asymptotic behaviors of the OLS and MSUR estimators, which we derive first.

For the scaled and centered OLS estimator, we have $G^{-1}\left(\hat{\theta}_{\mathrm{OLS}}-\theta\right)=\left(GZ'ZG\right)^{-1}GZ'u$, with

$$GZ'ZG \Rightarrow \int_0^1 J(r)J(r)'dr. \tag{285}$$

Using the results of Wagner and Hong (2016), together with the block-diagonal structure of $J(r)$

and Lemma 1 of Hong and Phillips (2010), we obtain

$$GZ'u = \sum_{t=1}^{T} GZ_t u_t = \begin{pmatrix} \sum_{t=1}^{T} G_1 Z_{1,t} u_{1,t} \\ \vdots \\ \sum_{t=1}^{T} G_N Z_{N,t} u_{N,t} \end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^1 J_1(r) dB_{u_1}(r) + M_1 \\ \vdots \\ \int_0^1 J_N(r) dB_{u_N}(r) + M_N \end{pmatrix} = \int_0^1 J(r) dB_u(r) + M \tag{286}$$

with $M_i \coloneqq [0'_{(q_i+1)\times 1}, M'_{i,1}, \ldots, M'_{i,m_i}, M'_{\mathrm{C},i,1}, \ldots, M'_{\mathrm{C},i,m}]'$, $M_{i,j} \coloneqq \Delta_{vu}^{k_{i-1}+j,i} \int_0^1 \dot{\mathbf{B}}_{v,i,j}(r) dr$, $M_{\mathrm{C},i,j} \coloneqq \Delta_{vu}^{k_N+j,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,j}(r) dr$, where $\dot{\mathbf{B}}_{v,i,j}(r) \coloneqq [1, 2B_{v,i,j}(r), \ldots, p_{i,j} B_{v,i,j}(r)^{p_{i,j}-1}]'$ for $0 \leq r \leq 1$ and $\dot{\mathbf{B}}_{v,\mathrm{C},i,j}(r) \coloneqq [1, 2B_{v,\mathrm{C},j}(r), \ldots, c_{i,j} B_{v,\mathrm{C},j}^{c_{i,j}-1}(r)]'$ for $0 \leq r \leq 1$. Combining all terms leads to

$$G^{-1}\left(\hat{\theta}_{\mathrm{OLS}} - \theta\right) \Rightarrow \left(\int_0^1 J(r) J(r)' dr\right)^{-1} \left(\int_0^1 J(r) dB_u(r) + M\right). \tag{287}$$

We next turn to the MSUR estimator. For the scaled and centered MSUR estimator we have

$$G^{-1}\left(\hat{\theta}_{\mathrm{MSUR}} - \theta\right) = \left(GZ'\left(I_T \otimes \hat{\Omega}_{uu}^{-1}\right) ZG\right)^{-1} \left(GZ'\left(I_T \otimes \hat{\Omega}_{uu}^{-1}\right) u\right). \tag{288}$$

For the first term on the right hand side of Equation (288), using Assumption 42 together with a consistent estimator $\hat{\Omega}_{uu} \to_p \Omega_{uu}$ and the continuous mapping theorem, we obtain

$$GZ'\left(I_T \otimes \hat{\Omega}_{uu}^{-1}\right) ZG = \sum_{t=1}^{T} GZ_t \hat{\Omega}_{uu}^{-1} Z_t' G \Rightarrow \int_0^1 J(r) \Omega_{uu}^{-1} J(r)' dr. \tag{289}$$

The second term of the right hand side of Equation (288) can be rewritten as

$$GZ'\left(I_T \otimes \hat{\Omega}_{uu}^{-1}\right) u = \sum_{t=1}^{T} GZ_t \hat{\Omega}_{uu}^{-1} u_t = \sum_{t=1}^{T} \begin{pmatrix} G_1 Z_{1,t} \left(\hat{\Omega}_{uu}^{-1}\right)^{1,\bullet} u_t \\ \vdots \\ G_N Z_{N,t} \left(\hat{\Omega}_{uu}^{-1}\right)^{N,\bullet} u_t \end{pmatrix} \tag{290}$$

$$= \begin{pmatrix} \sum_{t=1}^{T} G_1 Z_{1,t} \left(\hat{\Omega}_{uu}^{-1}\right)^{1,1} u_{1,t} + \cdots + \sum_{t=1}^{T} G_1 Z_{1,t} \left(\hat{\Omega}_{uu}^{-1}\right)^{1,N} u_{N,t} \\ \vdots \\ \sum_{t=1}^{T} G_N Z_{N,t} \left(\hat{\Omega}_{uu}^{-1}\right)^{N,1} u_{1,t} + \cdots + \sum_{t=1}^{T} G_N Z_{N,t} \left(\hat{\Omega}_{uu}^{-1}\right)^{N,N} u_{N,t} \end{pmatrix}. \tag{291}$$

For the ith block in (291), again using Assumption 42, a consistent estimator $\hat{\Omega}_{uu} \to_p \Omega_{uu}$, Lemma 1 of Hong and Phillips (2010) and the continuous mapping theorem, we obtain

$$\sum_{t=1}^{T} G_i Z_{i,t} \left(\hat{\Omega}_{uu}^{-1}\right)^{i,1} u_{1,t} + \cdots + \sum_{t=1}^{T} G_i Z_{i,t} \left(\hat{\Omega}_{uu}^{-1}\right)^{i,N} u_{N,t} \tag{292}$$

$$\Rightarrow \int_0^1 J_i(r)\left(\Omega_{uu}^{-1}\right)^{i,1} dB_{u_1}(r) + \left(\Omega_{uu}^{-1}\right)^{i,1} \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vu}^{k_{i-1}+1,1}\int_0^1 \dot{\mathbf{B}}_{v,i,1}(r)dr \\ \vdots \\ \Delta_{vu}^{k_{i-1}+m_i,1}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r)dr \\ \Delta_{vu}^{k_N+1,1}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r)dr \\ \vdots \\ \Delta_{vu}^{k_N+m,1}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r)dr \end{bmatrix} + \ldots \tag{293}$$

$$+ \int_0^1 J_i(r)\left(\Omega_{uu}^{-1}\right)^{i,N} dB_{u_N}(r) + \left(\Omega_{uu}^{-1}\right)^{i,N} \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vu}^{k_{i-1}+1,N}\int_0^1 \dot{\mathbf{B}}_{v,i,1}(r)dr \\ \vdots \\ \Delta_{vu}^{k_{i-1}+m_i,N}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r)dr \\ \Delta_{vu}^{k_N+1,N}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r)dr \\ \vdots \\ \Delta_{vu}^{k_N+m,N}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r)dr \end{bmatrix} \tag{294}$$

$$= \int_0^1 J_i(r)\left(\Omega_{uu}^{-1}\right)^{i,\bullet} dB_u(r) + \underbrace{\begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vu}^{k_{i-1}+1,\bullet}\left(\Omega_{uu}^{-1}\right)^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,1}(r)dr \\ \vdots \\ \Delta_{vu}^{k_{i-1}+m_i,\bullet}\left(\Omega_{uu}^{-1}\right)^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r)dr \\ \Delta_{vu}^{k_N+1,\bullet}\left(\Omega_{uu}^{-1}\right)^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r)dr \\ \vdots \\ \Delta_{vu}^{k_N+m,\bullet}\left(\Omega_{uu}^{-1}\right)^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r)dr \end{bmatrix}}_{=:M_i^*} \tag{295}$$

$$= \int_0^1 J_i(r)\left(\Omega_{uu}^{-1}\right)^{i,\bullet} dB_u(r) + M_i^*, \tag{296}$$

using the symmetry of $\left(\Omega_{uu}^{-1}\right)$. Given the block-diagonal form of $J(r)$, we have

$$\begin{bmatrix} \int_0^1 J_1(r)\left(\Omega_{uu}^{-1}\right)^{1,\bullet} dB_u(r) + M_1^* \\ \vdots \\ \int_0^1 J_N(r)\left(\Omega_{uu}^{-1}\right)^{N,\bullet} dB_u(r) + M_N^* \end{bmatrix} = \int_0^1 J(r)\Omega_{uu}^{-1} dB_u(r) + M^* \tag{297}$$

with $M^* \coloneqq [M_1^{*\prime}, \ldots, M_N^{*\prime}]'$. This leads to

$$GZ'\left(I_T \otimes \hat{\Omega}_{uu}^{-1}\right) u \Rightarrow \int_0^1 J(r)\Omega_{uu}^{-1} dB_u(r) + M^*. \tag{298}$$

Combining the results gives the limiting distribution of the MSUR estimator as

$$G^{-1}\left(\hat{\theta}_{\mathrm{MSUR}} - \theta\right) \Rightarrow \left(\int_0^1 J(r)\Omega_{uu}^{-1} J(r)' dr\right)^{-1} \left(\int_0^1 J(r)\Omega_{uu}^{-1} dB_u(r) + M^*\right). \tag{299}$$

Using the results for the OLS and MSUR estimators, we derive the limiting distributions of the

FM-SOLS and FM-SUR estimators. The centered and scaled FM-SOLS estimator is given by

$$G^{-1}\left(\hat{\theta}-\theta\right)=\left(GZ'ZG\right)^{-1}\left(GZ'u^{+}-G\hat{A}\right), \tag{300}$$

with $u^{+}\coloneqq[u_1^{+\prime},\ldots,u_T^{+\prime}]'$ and $u_t^{+}\coloneqq u_t-\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t$ for $t=1,\ldots,T$ and by definition it holds that $u^{+}=u-\left(I_T\otimes\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\right)v$ with $v\coloneqq[v_1',\ldots,v_T']'$. The limit of the first term on the right hand side of (300) is already provided for the OLS estimator in (285). The second term on the right hand side of (300) can be rewritten as

$$GZ'u^{+}-G\hat{A}=GZ'u-GZ'\left(I_T\otimes\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\right)v-G\hat{A}, \tag{301}$$

with the limit of the first term on the right hand side of (301) already provided in (286). We now consider the second term of (301), given by

$$GZ'\left(I_T\otimes\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\right)v=\sum_{t=1}^{T}GZ_t\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t=\begin{pmatrix}\sum_{t=1}^{T}G_1Z_{1,t}v_t'\hat{\Omega}_{vv}^{-1}\left(\hat{\Omega}_{vu}\right)^{\boldsymbol{\cdot},1}\\ \vdots\\ \sum_{t=1}^{T}G_NZ_{N,t}v_t'\hat{\Omega}_{vv}^{-1}\left(\hat{\Omega}_{vu}\right)^{\boldsymbol{\cdot},N}\end{pmatrix}. \tag{302}$$

Using Assumption 42, it follows from, e. g. , Ibragimov and Phillips (2008), consistent long-run covariance estimation and the continuous mapping theorem, that

$$\sum_{t=1}^{T}G_iZ_{i,t}v_t'\hat{\Omega}_{vv}^{-1}\left(\hat{\Omega}_{vu}\right)^{\boldsymbol{\cdot},i}\Rightarrow\int_0^1J_i(r)dB_v(r)'\Omega_{vv}^{-1}\left(\Omega_{vu}\right)^{\boldsymbol{\cdot},i}+\begin{bmatrix}0_{(q_i+1)\times1}\\ \Delta_{vv}^{k_{i-1}+1,\boldsymbol{\cdot}}\Omega_{vv}^{-1}\left(\Omega_{vu}\right)^{\boldsymbol{\cdot},i}\int_0^1\dot{\mathbf{B}}_{v,i,1}(r)dr\\ \vdots\\ \Delta_{vv}^{k_{i-1}+m_i,\boldsymbol{\cdot}}\Omega_{vv}^{-1}\left(\Omega_{vu}\right)^{\boldsymbol{\cdot},i}\int_0^1\dot{\mathbf{B}}_{v,i,m_i}(r)dr\\ \Delta_{vv}^{k_N+1,\boldsymbol{\cdot}}\Omega_{vv}^{-1}\left(\Omega_{vu}\right)^{\boldsymbol{\cdot},i}\int_0^1\dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r)dr\\ \vdots\\ \Delta_{vv}^{k_N+m,\boldsymbol{\cdot}}\Omega_{vv}^{-1}\left(\Omega_{vu}\right)^{\boldsymbol{\cdot},i}\int_0^1\dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r)dr\end{bmatrix}, \tag{303}$$

for the ith block of (302)[60]. With

$$GZ'u^{+}=GZ'u-GZ'\left(I_T\otimes\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\right)v=\sum_{t=1}^{T}GZ_tu_t-\sum_{t=1}^{T}GZ_t\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t \tag{304}$$

$$=\begin{pmatrix}\sum_{t=1}^{T}G_1Z_{1,t}u_{1,t}\\ \vdots\\ \sum_{t=1}^{T}G_NZ_{N,t}u_{N,t}\end{pmatrix}-\begin{pmatrix}\sum_{t=1}^{T}G_1Z_{1,t}v_t'\hat{\Omega}_{vv}^{-1}\left(\hat{\Omega}_{vu}\right)^{\boldsymbol{\cdot},1}\\ \vdots\\ \sum_{t=1}^{T}G_NZ_{N,t}v_t'\hat{\Omega}_{vv}^{-1}\left(\hat{\Omega}_{vu}\right)^{\boldsymbol{\cdot},N}\end{pmatrix}, \tag{305}$$

[60] In the corresponding part of Proof of Proposition 1 in Wagner *et al.* (2020), there are two typos in their Equation (58): On the left hand side the leading sum symbol $\sum_{t=1}^{T}$ is missing and on the right hand side $\Delta_{vu}^{i,\boldsymbol{\cdot}}$ needs to be replaced by $\Delta_{vv}^{i,\boldsymbol{\cdot}}$. Following this, in the second line of (59) $\Delta_{vu}^{i,\boldsymbol{\cdot}}$ needs to be replaced by $\Delta_{vv}^{i,\boldsymbol{\cdot}}$. With this replacement, Equation (60) in Wagner *et al.* (2020) holds with their stated definition of $\left(\Delta_{vu}^{+}\right)^{i,\boldsymbol{\cdot}}$

and the results above, we obtain for the ith block of (305)

$$\sum_{t=1}^{T} G_i Z_{i,t} u_{i,t} - \sum_{t=1}^{T} G_i Z_{i,t} v_t' \hat{\Omega}_{vv}^{-1} \left(\hat{\Omega}_{vu}\right)^{\bullet,i} \Rightarrow \int_0^1 J_i(r) dB_{u_i}(r) + \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vu}^{k_{i-1}+1,i} \int_0^1 \dot{\mathbf{B}}_{v,i,1}(r) dr \\ \vdots \\ \Delta_{vu}^{k_{i-1}+m_i,i} \int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r) dr \\ \Delta_{vu}^{k_N+1,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r) dr \\ \vdots \\ \Delta_{vu}^{k_N+m,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r) dr \end{bmatrix} \tag{306}$$

$$- \int_0^1 J_i(r) dB_v(r)' \Omega_{vv}^{-1} \left(\Omega_{vu}\right)^{\bullet,i} - \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vv}^{k_{i-1}+1,\bullet} \Omega_{vv}^{-1} \left(\Omega_{vu}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,1}(r) dr \\ \vdots \\ \Delta_{vv}^{k_{i-1}+m_i,\bullet} \Omega_{vv}^{-1} \left(\Omega_{vu}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r) dr \\ \Delta_{vv}^{k_N+1,\bullet} \Omega_{vv}^{-1} \left(\Omega_{vu}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r) dr \\ \vdots \\ \Delta_{vv}^{k_N+m,\bullet} \Omega_{vv}^{-1} \left(\Omega_{vu}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r) dr \end{bmatrix} \tag{307}$$

$$= \int_0^1 J_i(r) dB_{u_i \cdot v}(r) + \underbrace{\begin{bmatrix} 0_{(q_i+1)\times 1} \\ \left(\Delta_{vu}^{+}\right)^{k_{i-1}+1,i} \int_0^1 \dot{\mathbf{B}}_{v,i,1}(r) dr \\ \vdots \\ \left(\Delta_{vu}^{+}\right)^{k_{i-1}+m_i,i} \int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r) dr \\ \left(\Delta_{vu}^{+}\right)^{k_N+1,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r) dr \\ \vdots \\ \left(\Delta_{vu}^{+}\right)^{k_N+m,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r) dr \end{bmatrix}}_{=:A_i}, \tag{308}$$

with $B_{u_i \cdot v}(r) \coloneqq B_{u_i} - B_v(r)' \Omega_{vv}^{-1} \Omega_{vu}^{\bullet,i}$ and $\left(\Delta_{vv}^{+}\right)^{i,j}$ as defined in the main text below (163). Given the block-diagonal form of $J(r)$, we have

$$\begin{bmatrix} \int_0^1 J_1(r) dB_{u_1 \cdot v}(r) + A_1 \\ \vdots \\ \int_0^1 J_N(r) dB_{u_N \cdot v}(r) + A_N \end{bmatrix} = \int_0^1 J(r) dB_{u \cdot v}(r) + A \tag{309}$$

with $A \coloneqq [A_1', \ldots, A_N']'$ and obtain

$$GZ'u^+ \Rightarrow \int_0^1 J(r) dB_{u \cdot v}(r) + A. \tag{310}$$

By construction, it follows that

$$G\hat{A} \Rightarrow A. \tag{311}$$

This leads to

$$GZ'u^+ - G\hat{A} \Rightarrow \int_0^1 J(r) dB_{u \cdot v}(r) + A - A = \int_0^1 J(r) dB_{u \cdot v}(r). \tag{312}$$

Combining the results above delivers

$$G^{-1}\left(\hat{\theta}-\theta\right) \Rightarrow \int_0^1 J(r)J(r)'dr \int_0^1 J(r)dB_{u\cdot v}(r). \tag{313}$$

Finally we turn to the limiting distribution of the FM-SUR estimator. The centered and scaled FM-SUR estimator is given by

$$G^{-1}\left(\tilde{\theta}-\theta\right) = \left(GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) ZG\right)^{-1} \left(GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) u^+ - G\tilde{A}^*\right), \tag{314}$$

with the limit of the first term on the right hand side, using Assumption 42 together with a consistent estimator $\hat{\Omega}_{u\cdot v} \to_p \Omega_{u\cdot v}$ and the continuous mapping theorem, provided by

$$GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) ZG = \sum_{t=1}^{T} GZ_t\hat{\Omega}_{u\cdot v}^{-1}Z_t'G \Rightarrow \int_0^1 J(r)\Omega_{u\cdot v}^{-1}J(r)'dr. \tag{315}$$

We can rewrite the second term of the right hand side of (314) as

$$GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) u^+ - G\tilde{A}^* = GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) u - GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right)\left(I_T \otimes \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\right) v - G\tilde{A}^*. \tag{316}$$

We start with considering the first term on the right hand side of (316) and rewrite it as

$$GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) u = \sum_{t=1}^{T} GZ_t\hat{\Omega}_{u\cdot v}^{-1}u_t = \begin{pmatrix} \sum_{t=1}^{T} G_1Z_{1,t}\left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{1,\bullet} u_t \\ \vdots \\ \sum_{t=1}^{T} G_NZ_{N,t}\left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{N,\bullet} u_t \end{pmatrix}. \tag{317}$$

We obtain the limiting distribution of the ith block of (317) in a similar fashion as for the ith block of $GZ'\left(I_T \otimes \hat{\Omega}_{uu}^{-1}\right) u$ in the MSUR estimator above. Using similar arguments as starting from (292) going to (296), with $\hat{\Omega}_{uu}^{-1}$ replaced by $\hat{\Omega}_{u\cdot v}^{-1}$ and Ω_{uu}^{-1} replaced by $\Omega_{u\cdot v}^{-1}$, we obtain

$$\sum_{t=1}^{T} G_iZ_{i,t}\left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{i,\bullet} u_t \Rightarrow \int_0^1 J_i(r)\left(\Omega_{u\cdot v}^{-1}\right)^{i,\bullet} dB_u(r) + \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vu}^{k_{i-1}+1,\bullet}\left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,1}(r)dr \\ \vdots \\ \Delta_{vu}^{k_{i-1}+m_i,\bullet}\left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r)dr \\ \Delta_{vu}^{k_N+1,\bullet}\left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r)dr \\ \vdots \\ \Delta_{vu}^{k_N+m,\bullet}\left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r)dr \end{bmatrix}. \tag{318}$$

For the second term in (316), we have

$$GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right)\left(I_T \otimes \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\right) v = \sum_{t=1}^{T} GZ_t\hat{\Omega}_{u\cdot v}^{-1}\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t = \begin{pmatrix} \sum_{t=1}^{T} G_1Z_{1,t}\left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{1,\bullet}\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t \\ \vdots \\ \sum_{t=1}^{T} G_NZ_{N,t}\left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{N,\bullet}\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t \end{pmatrix}. \tag{319}$$

Considering the ith block of (319), we obtain

$$\sum_{t=1}^{T} G_i Z_{i,t} v_t' \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{\bullet,i} \Rightarrow \int_0^1 J_i(r) dB_v(r)' \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \tag{320}$$

$$+ \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vv}^{k_{i-1}+1,\bullet} \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,1}(r) dr \\ \vdots \\ \Delta_{vv}^{k_{i-1}+m_i,\bullet} \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r) dr \\ \Delta_{vv}^{k_N+1,\bullet} \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r) dr \\ \vdots \\ \Delta_{vv}^{k_N+m,\bullet} \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r) dr \end{bmatrix}. \tag{321}$$

We next consider the ith block of

$$GZ' \left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) u - GZ' \left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) \left(I_T \otimes \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1}\right) v = \sum_{t=1}^{T} GZ_t \hat{\Omega}_{u\cdot v}^{-1} u_t - \sum_{t=1}^{T} GZ_t \hat{\Omega}_{u\cdot v}^{-1} \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1} v_t, \tag{322}$$

Combining the previous results leads to

$$\sum_{t=1}^{T} G_i Z_{i,t} \left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{i,\bullet} u_t - \sum_{t=1}^{T} G_i Z_{i,t} v_t' \hat{\Omega}_{vv}^{-1} \hat{\Omega}_{vu} \left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{\bullet,i} \tag{323}$$

$$\Rightarrow \int_0^1 J_i(r) \left(\Omega_{u\cdot v}^{-1}\right)^{i,\bullet} dB_u(r) + \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vu}^{k_{i-1}+1,\bullet} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,1}(r) dr \\ \vdots \\ \Delta_{vu}^{k_{i-1}+m_i,\bullet} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r) dr \\ \Delta_{vu}^{k_N+1,\bullet} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r) dr \\ \vdots \\ \Delta_{vu}^{k_N+m,\bullet} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r) dr \end{bmatrix} \tag{324}$$

$$- \int_0^1 J_i(r) \left(\Omega_{u\cdot v}^{-1}\right)^{i,\bullet} \Omega_{uv} \Omega_{vv}^{-1} dB_v(r) - \begin{bmatrix} 0_{(q_i+1)\times 1} \\ \Delta_{vv}^{k_{i-1}+1,\bullet} \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,1}(r) dr \\ \vdots \\ \Delta_{vv}^{k_{i-1}+m_i,\bullet} \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r) dr \\ \Delta_{vv}^{k_N+1,\bullet} \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r) dr \\ \vdots \\ \Delta_{vv}^{k_N+m,\bullet} \Omega_{vv}^{-1} \Omega_{vu} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r) dr \end{bmatrix} \tag{325}$$

$$= \int_0^1 J_i(r) \left(\Omega_{u\cdot v}^{-1}\right)^{i,\bullet} dB_{u\cdot v}(r) + \underbrace{\begin{bmatrix} 0_{(q_i+1)\times 1} \\ (\Delta_{vu}^{+})^{k_{i-1}+1,\bullet} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,1}(r)dr \\ \vdots \\ (\Delta_{vu}^{+})^{k_{i-1}+m_i,\bullet} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,i,m_i}(r)dr \\ (\Delta_{vu}^{+})^{k_N+1,\bullet} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}(r)dr \\ \vdots \\ (\Delta_{vu}^{+})^{k_N+m,\bullet} \left(\Omega_{u\cdot v}^{-1}\right)^{\bullet,i} \int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}(r)dr \end{bmatrix}}_{=:A_i^*}. \tag{326}$$

Given the block-diagonal form of $J(r)$, we have

$$\begin{pmatrix} \int_0^1 J_1(r) \left(\Omega_{u\cdot v}^{-1}\right)^{1,\bullet} dB_{u\cdot v}(r) + A_1^* \\ \vdots \\ \int_0^1 J_N(r) \left(\Omega_{u\cdot v}^{-1}\right)^{N,\bullet} dB_{u\cdot v}(r) + A_N^* \end{pmatrix} = \int_0^1 J(r)\Omega_{u\cdot v}^{-1} dB_{u\cdot v}(r) + A^*, \tag{327}$$

with $A^* \coloneqq [A_1^{*\prime}, \ldots, A_N^{*\prime}]'$ and we obtain

$$GZ' \left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) u^+ \Rightarrow \int_0^1 J(r)\Omega_{u\cdot v}^{-1} dB_{u\cdot v}(r) + A^*. \tag{328}$$

By construction, we have

$$G\tilde{A}^* \Rightarrow A^*. \tag{329}$$

Combining the results above leads to the limiting distribution of the FM-SUR estimator, given by

$$G^{-1}\left(\tilde{\theta} - \theta\right) \Rightarrow \left(\int_0^1 J(r)\Omega_{u\cdot v}^{-1} J(r)' dr\right)^{-1} \int_0^1 J(r)\Omega_{u\cdot v}^{-1} dB_{u\cdot v}(r). \tag{330}$$

□

Proof of Proposition 5:

The proof is exactly analogous for both test statistics $\hat{T}_W$ and $\tilde{T}_W$ and we therefore only show the derivation for $\tilde{T}_W$. With Proposition 4 and the constraints from Proposition 5 in place, it follows

that

$$\tilde{T}_W = \left(R\tilde{\theta} - r\right)' \left[R\left(Z'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) Z\right)^{-1} R'\right]^{-1} \left(R\tilde{\theta} - r\right) \tag{331}$$

$$= \left(R\left(\tilde{\theta} - \theta\right)\right)' \left[R\left(Z'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) Z\right)^{-1} R'\right]^{-1} \left(R\left(\tilde{\theta} - \theta\right)\right) \tag{332}$$

$$= \left(G_R^{-1} RGG^{-1}\left(\tilde{\theta} - \theta\right)\right)' \tag{333}$$

$$\times \left[G_R^{-1} RG\left(GZ'\left(I_T \otimes \hat{\Omega}_{u\cdot v}^{-1}\right) ZG\right)^{-1} GR' G_R^{-1}\right]^{-1} \tag{334}$$

$$\times \left(G_R^{-1} RGG^{-1}\left(\tilde{\theta} - \theta\right)\right) \tag{335}$$

$$\Rightarrow \left(R^*\left(\int_0^1 J(r)\Omega_{u\cdot v}^{-1} J(r)' dr\right)^{-1} \int_0^1 J(r)\Omega_{u\cdot v}^{-1} dB_{u\cdot v}(r)\right)' \tag{336}$$

$$\times \left[R^*\left(\int_0^1 J(r)\Omega_{u\cdot v}^{-1} J(r)' dr\right)^{-1} R^{*\prime}\right]^{-1} \tag{337}$$

$$\times \left(R^*\left(\int_0^1 J(r)\Omega_{u\cdot v}^{-1} J(r)' dr\right)^{-1} \int_0^1 J(r)\Omega_{u\cdot v}^{-1} dB_{u\cdot v}(r)\right), \tag{338}$$

which can be shown to be chi-squared distributed with s degrees of freedom using standard arguments involving quadratic forms of functionals of zero mean Gaussian mixture distributions. □

Proof of Proposition 6:

The proposition follows as a special case of Proposition 5. The restriction $\theta_F = 0$ in the augmented regression (173) corresponds to $R = [0_{\bar{d}\times d}, I_{\bar{d}}]$ and $r = 0_{\bar{d}\times 1}$ in Proposition 5. This implies that

$$\left(R\begin{bmatrix} Z'\left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}^{-1}\right) Z & Z'\left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}^{-1}\right) F \\ F'\left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}^{-1}\right) Z & F'\left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}^{-1}\right) F \end{bmatrix}^{-1} R'\right) = \tilde{F}'\left(I_T \otimes \hat{\Omega}_{u\cdot\tilde{v}}^{-1}\right)\tilde{F}, \tag{339}$$

with $\tilde{F}$ as defined in Proposition 6. □

Proof of Proposition 7:

The scaled FM-OLS estimator of $\theta_{\hat{F}}$ in (178) is given by

$$G_F^{-1}\hat{\theta}_{\hat{F}}^{+} := \left(G_F \hat{F}'\hat{F} G_F\right)^{-1}\left(G_F \hat{F}'\hat{u}^{+} - G_F O^{F*} - G_F A^{F*} + G_F k^{F*} A^{Z*}\right), \tag{340}$$

compare (209). For the first term we obtain $\left(G_F \hat{F}'\hat{F} G_F\right)^{-1} \Rightarrow \int_0^1 J^{\hat{F}}(r) J^{\hat{F}}(r)' dr$. We analyze the

remaining terms separately, starting with $G_F\hat{F}'\hat{u}^+$. With

$$\hat{u}_t^+ = y_t^+ - Z_t'\hat{\theta} \tag{341}$$
$$= y_t - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t - Z_t'\hat{\theta} \tag{342}$$
$$= Z_t'\theta + u_t - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t - Z_t'\hat{\theta} \tag{343}$$
$$= u_t - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t - Z_t'\left(\hat{\theta} - \theta\right) \tag{344}$$
$$= u_t^+ - Z_t'\left(\hat{\theta} - \theta\right), \tag{345}$$

and thus $\hat{u}^+ = u^+ - Z\left(\hat{\theta} - \theta\right)$, we can rewrite

$$G_F\hat{F}'\hat{u}^+ = G_F\hat{F}'\left(u^+ - Z\left(\hat{\theta} - \theta\right)\right) \tag{346}$$
$$= G_F\hat{F}'u^+ - G_F\hat{F}'Z\left(\hat{\theta} - \theta\right) \tag{347}$$
$$= G_F\hat{F}'u^+, \tag{348}$$

since $\hat{F}'Z = 0_{\bar{d}\times d}$ by construction. It continuous with

$$G_F\hat{F}'u^+ = G_FF'u^+ - G_FF'Z(Z'Z)^{-1}Z'u^+ \tag{349}$$
$$= G_FF'u^+ - G_FF'ZG(GZ'ZG)^{-1}GZ'u^+. \tag{350}$$

Starting with the first term, we have

$$G_FF'u^+ = \sum_{t=1}^{T} G_FF_tu_t^+ = \begin{bmatrix} \sum_{t=1}^{T} G_{F,1}F_{1,t}u_{1,t}^+ \\ \vdots \\ \sum_{t=1}^{T} G_{F,N}F_{N,t}u_{N,t}^+ \end{bmatrix}, \tag{351}$$

and obtain for the i-th block

$$\sum_{t=1}^{T} G_{F,i}F_{i,t}u_{i,t}^+ = \sum_{t=1}^{T} G_{F,i}F_{i,t}u_{i,t} - \sum_{t=1}^{T} G_{F,i}F_{i,t}\underbrace{\hat{\Omega}_{uv}^{i,\bullet}\hat{\Omega}_{vv}^{-1}v_t}_{=v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{uv}^{\bullet,i}} \Rightarrow \int_0^1 J_i^F(r)dB_{u_i}(r) \tag{352}$$

$$+\underbrace{\begin{bmatrix} 0_{\bar{q}_i\times 1} \\ \Delta_{\tilde{v}u}^{k_{i-1}+1,i}\int_0^1 \dot{\mathbf{B}}_{v,i,1}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}u}^{k_{i-1}+m_i,i}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}^F(r)dr \\ \Delta_{\tilde{v}u}^{k_N+1,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}u}^{k_N+m,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}^F(r)dr \\ \Delta_{\tilde{v}u}^{\bar{k}_{i-1}+1,i}\int_0^1 \dot{\mathbf{B}}_{v_{i,1}^*}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}u}^{\bar{k}_{i-1}+\bar{m}_i,i}\int_0^1 \dot{\mathbf{B}}_{v_{i,\bar{m}_i}^*}^F(r)dr \\ \Delta_{\tilde{v}u}^{\bar{k}_N+1,i}\int_0^1 \dot{\mathbf{B}}_{v_{\mathrm{C},i,1}^*}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}u}^{\bar{k}_N+\bar{m},i}\int_0^1 \dot{\mathbf{B}}_{v_{\mathrm{C},i,\bar{m}}^*}^F(r)dr \end{bmatrix}}_{=:A_i^{Fu}} - \int_0^1 J_i^F(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i} - \underbrace{\begin{bmatrix} 0_{\bar{q}_i\times 1} \\ \Delta_{\tilde{v}v}^{k_{i-1}+1,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}\int_0^1 \dot{\mathbf{B}}_{v,i,1}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}v}^{k_{i-1}+m_i,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}^F(r)dr \\ \Delta_{\tilde{v}v}^{k_N+1,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}v}^{k_N+m,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}^F(r)dr \\ \Delta_{\tilde{v}v}^{\bar{k}_{i-1}+1,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}\int_0^1 \dot{\mathbf{B}}_{v_{i,1}^*}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}v}^{\bar{k}_{i-1}+\bar{m}_i,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}\int_0^1 \dot{\mathbf{B}}_{v_{i,\bar{m}_i}^*}^F(r)dr \\ \Delta_{\tilde{v}v}^{\bar{k}_N+1,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}\int_0^1 \dot{\mathbf{B}}_{v_{\mathrm{C},i,1}^*}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}v}^{\bar{k}_N+\bar{m},\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}\int_0^1 \dot{\mathbf{B}}_{v_{\mathrm{C},i,\bar{m}}^*}^F(r)dr \end{bmatrix}}_{=:A_i^{Fv}} \tag{353}$$

$$= \int_0^1 J_i^F(r)dB_{u_i}(r) + A_i^{Fu} - \int_0^1 J_i^F(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i} - A_i^{Fv}, \tag{354}$$

with

$$\dot{\mathbf{B}}_{v,i,j}^F(r) \coloneqq [(p_{i,j}+1)B_{v,i,j}(r)^{p_{i,j}}, \ldots, (p_{i,j}+\bar{p}_{i,j})B_{v,i,j}(r)^{p_{i,j}+\bar{p}_{i,j}-1}]', \tag{355}$$

$$\dot{\mathbf{B}}_{v,\mathrm{C},i,j}^F(r) \coloneqq [(c_{i,j}+1)B_{v,i,j}(r)^{c_{i,j}}, \ldots, (c_{i,j}+\bar{c}_{i,j})B_{v,i,j}(r)^{c_{i,j}+\bar{c}_{i,j}-1}]', \tag{356}$$

$$\dot{\mathbf{B}}_{v_{i,j}^*}(r) \coloneqq [1, 2B_{v_{i,j}^*}(r), \ldots, s_{i,j}B_{v_{i,j}^*}^{s_{i,j}-1}(r)]', \tag{357}$$

$$\dot{\mathbf{B}}_{v_{\mathrm{C},i,j}^*}(r) \coloneqq [1, 2B_{v_{\mathrm{C},j}^*}(r), \ldots, \varsigma_{i,j}B_{v_{\mathrm{C},j}^*}^{\varsigma_{i,j}-1}(r)]', \tag{358}$$

for $0 \leq r \leq 1$. Given the block-diagonal form of $J^F(r)$, we have

$$G_F F' u^+ \Rightarrow \int_0^1 J^F(r)dB_u - \int_0^1 J^F(r)\Omega_{uv}\Omega_{vv}^{-1}dB_v(r) + A^{Fu} - A^{Fv} \tag{359}$$

$$= \int_0^1 J^F(r)dB_{u\cdot v} + A^{Fu} - A^{Fv} \tag{360}$$

with $A^{Fu} \coloneqq [A_1^{Fu\prime}, \ldots, A_N^{Fu\prime}]'$ and $A^{Fv} \coloneqq [A_1^{Fv\prime}, \ldots, A_N^{Fv\prime}]'$. We next turn to the second term and obtain for the first part

$$G_F F' ZG\left(GZ'ZG\right)^{-1} \Rightarrow \int_0^1 J^F(r)J(r)'dr \left(\int_0^1 J(r)J(r)'dr\right)^{-1}, \tag{361}$$

and from Proof of Proposition 4 we have for the second part

$$GZ'u^+ \Rightarrow \int_0^1 J(r)dB_{u\cdot v}(r) + A, \tag{362}$$

which leads to

$$G_F F' ZG \left(GZ'ZG\right)^{-1} GZ'u^+ \Rightarrow \int_0^1 J^F(r)J(r)'dr \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \left(\int_0^1 J(r)dB_{u\cdot v}(r) + A\right). \tag{363}$$

Combining the results leads to

$$G_F \hat{F}' u^+ \Rightarrow \int_0^1 J^F(r)dB_{u\cdot v} + A^{Fu} - A^{Fv} + \int_0^1 J^F(r)J(r)'dr \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \left(\int_0^1 J(r)dB_{u\cdot v}(r) + A\right) \tag{364}$$

$$= \int_0^1 J^{\hat{F}}(r)dB_{u\cdot v} + A^{Fu} - A^{Fv} - \int_0^1 J^F(r)J(r)'dr \left(\int_0^1 J(r)J(r)'dr\right)^{-1} A. \tag{365}$$

With rewriting $B_{u\cdot v} = B_{u\cdot\tilde{v}} + \Omega_{u\tilde{v}}\Omega_{\tilde{v}\tilde{v}}^{-1} B_{\tilde{v}}(r) - \Omega_{uv}\Omega_{vv}^{-1} B_v(r)$ we obtain

$$G_F \hat{F}' u^+ \Rightarrow \int_0^1 J^{\hat{F}}(r)dB_{u\cdot\tilde{v}}(r) + \int_0^1 J^{\hat{F}}(r)\Omega_{u\tilde{v}}\Omega_{\tilde{v}\tilde{v}}^{-1} dB_{\tilde{v}}(r) - \int_0^1 J^{\hat{F}}(r)\Omega_{uv}\Omega_{vv}^{-1} dB_v(r) \tag{366}$$

$$+ A^{Fu} - A^{Fv} - \int_0^1 J^F(r)J(r)'dr \left(\int_0^1 J(r)J(r)'dr\right)^{-1} A \tag{367}$$

Continuing with the next term of (340) and have

$$G_F O^{F*} = G_F \hat{F}' \left(I_T \otimes \hat{\Omega}_{u\tilde{v}}\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\right)\tilde{v} - G_F \hat{F}' \left(I_T \otimes \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\right) v \tag{368}$$

$$= \sum_{t=1}^{T} G_F \hat{F}_t \hat{\Omega}_{u\tilde{v}}\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\tilde{v}_t - \sum_{t=1}^{T} G_F \hat{F}_t \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1} v_t \tag{369}$$

$$= \begin{bmatrix} \sum_{t=1}^{T} G_{1,F}\hat{F}_{1,t}\tilde{v}_t'\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}^{\bullet,1} \\ \vdots \\ \sum_{t=1}^{T} G_{N,F}\hat{F}_{N,t}\tilde{v}_t'\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}^{\bullet,N} \end{bmatrix} - \begin{bmatrix} \sum_{t=1}^{T} G_{1,F}\hat{F}_{1,t} v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}^{\bullet,1} \\ \vdots \\ \sum_{t=1}^{T} G_{N,F}\hat{F}_{N,t} v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}^{\bullet,N} \end{bmatrix}, \tag{370}$$

with the i-th term leading to

$$\sum_{t=1}^{T} G_{i,F}\hat{F}_{i,t}\tilde{v}_t'\hat{\Omega}_{\tilde{v}\tilde{v}}^{-1}\hat{\Omega}_{\tilde{v}u}^{\bullet,i} - \sum_{t=1}^{T} G_{i,F}\hat{F}_{i,t} v_t'\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}^{\bullet,i} \tag{371}$$

$$\Rightarrow \int_0^1 J_i^{\hat{F}}(r)dB_{\tilde{v}}(r)'\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i} + \underbrace{\begin{bmatrix} 0_{\bar{q}_i\times 1} \\ \Delta_{\tilde{v}\tilde{v}}^{k_{i-1}+1,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,1}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}\tilde{v}}^{k_{i-1}+m_i,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}^F(r)dr \\ \Delta_{\tilde{v}\tilde{v}}^{k_N+1,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}\tilde{v}}^{k_N+m,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}^F(r)dr \\ \Delta_{\tilde{v}\tilde{v}}^{k_{i-1}+1,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v_{i,1}^*}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}\tilde{v}}^{\bar{k}_{i-1}+\bar{m}_i,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v_{i,\bar{m}_i}^*}^F(r)dr \\ \Delta_{\tilde{v}\tilde{v}}^{\bar{k}_N+1,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v_{\mathrm{C},i,1}^*}^F(r)dr \\ \vdots \\ \Delta_{\tilde{v}\tilde{v}}^{\bar{k}_N+\bar{m},\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v_{\mathrm{C},i,\bar{m}}^*}^F(r)dr \end{bmatrix}}_{=:A_i^{F\tilde{v}}} \tag{372}$$

$$-\int_0^1 J_i^F(r)J_i(r)'dr\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\underbrace{\begin{bmatrix} 0_{\bar{q}_i\times 1} \\ \Delta_{v\tilde{v}}^{k_{i-1}+1,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,1}^F(r)dr \\ \vdots \\ \Delta_{v\tilde{v}}^{k_{i-1}+m_i,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}^F(r)dr \\ \Delta_{v\tilde{v}}^{k_N+1,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}^F(r)dr \\ \vdots \\ \Delta_{v\tilde{v}}^{k_N+m,\bullet}\Omega_{\tilde{v}\tilde{v}}^{-1}\Omega_{\tilde{v}u}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}^F(r)dr \end{bmatrix}}_{=:A_i^{Z\tilde{v}}} \tag{373}$$

$$-\int_0^1 J_i^{\hat{F}}(r)dB_v(r)'\Omega_{vv}^{-1}\Omega_{vu}^{\bullet,i} - A_i^{Fv} \tag{374}$$

$$+\int_0^1 J_i^F(r)J_i(r)'dr\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\underbrace{\begin{bmatrix} 0_{\bar{q}_i\times 1} \\ \Delta_{vv}^{k_{i-1}+1,\bullet}\Omega_{vv}^{-1}\Omega_{vu}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,1}^F(r)dr \\ \vdots \\ \Delta_{vv}^{k_{i-1}+m_i,\bullet}\Omega_{vv}^{-1}\Omega_{vu}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,i,m_i}^F(r)dr \\ \Delta_{vv}^{k_N+1,\bullet}\Omega_{vv}^{-1}\Omega_{vu}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,1}^F(r)dr \\ \vdots \\ \Delta_{vv}^{k_N+m,\bullet}\Omega_{vv}^{-1}\Omega_{vu}^{\bullet,i}\int_0^1 \dot{\mathbf{B}}_{v,\mathrm{C},i,m}^F(r)dr \end{bmatrix}}_{=:A_i^{Zv}} \tag{375}$$

Stacking the terms leads to

$$G_F O^{F*} = G_F \hat{F}' \left(I_T \otimes \hat{\Omega}_{u\tilde{v}} \hat{\Omega}_{\tilde{v}\tilde{v}}^{-1} \right) \tilde{v} - G_F \hat{F}' \left(I_T \otimes \hat{\Omega}_{uv} \hat{\Omega}_{vv}^{-1} \right) v \tag{376}$$

$$\Rightarrow \int_0^1 J^{\hat{F}}(r) \Omega_{u\tilde{v}} \Omega_{\tilde{v}\tilde{v}}^{-1} dB_{\tilde{v}}(r) + A^{F\tilde{v}} - \int_0^1 J^F(r) J(r)' dr \left(\int_0^1 J(r) J(r)' dr \right)^{-1} A^{Z\tilde{v}} \tag{377}$$

$$- \int_0^1 J^{\hat{F}}(r) \Omega_{uv} \Omega_{vv}^{-1} dB_v(r) - A^{Fv} + \int_0^1 J^F(r) J(r)' dr \left(\int_0^1 J(r) J(r)' dr \right)^{-1} A^{Zv}, \tag{378}$$

with $A^{F\tilde{v}} \coloneqq [A_1^{F\tilde{v}\prime}, \ldots, A_N^{F\tilde{v}\prime}]'$, $A^{Z\tilde{v}} \coloneqq [A_1^{Z\tilde{v}\prime}, \ldots, A_N^{Z\tilde{v}\prime}]'$, $A^{Zv} \coloneqq [A_1^{Zv\prime}, \ldots, A_N^{Zv\prime}]'$ and A^{Fv} already defined above. For the last two terms of (340) it is easy to verify that

$$G_F A^{F*} \Rightarrow \left(A^{Fu} - A^{F\tilde{v}} \right), \tag{379}$$

and

$$G_F k^{F*} A^{Z*} = G_F F' Z G \left(G Z' Z G \right)^{-1} G A^{Z*} \tag{380}$$

$$\Rightarrow \left(\int_0^1 J^F(r) J(r)' dr \left(J(r) J(r)' dr \right)^{-1} \right) \left(A + A^{Zv} - A^{Z\tilde{v}} \right), \tag{381}$$

with $A^{F*}, A, A^{Zv}, k^{F*}$ and $A^{Z\tilde{v}}$ defined in the main text.

Combining all results leads to the limiting distribution of the scaled estimator, which is given by

$$G_F^{-1} \hat{\theta}_{\hat{F}}^{+} = \left(G_F \hat{F}' \hat{F} G_F \right)^{-1} \left(G_F \hat{F}' \hat{u}^{+} - G_F O^{F*} - G_F A^{F*} + G_F k^{F*} A^{Z*} \right) \tag{382}$$

$$\Rightarrow \left(\int_0^1 J^{\hat{F}}(r) J^{\hat{F}}(r)' dr \right) \int_0^1 J^{\hat{F}}(r) dB_{u\cdot\tilde{v}}(r) \tag{383}$$

The limiting distribution given in (383) together with the continuous mapping theorem and consistent long-run covariance and half long-run covariance estimation implies the asymptotic chi-squared null distribution of the LM-type test statistic in (211). □

Proof of Proposition 8:

The result follows with similar arguments as in Proposition 4, by replacing all pooled regressors with the group-wise stacked pooled regressors. □

3.6.2 Additional Finite Sample Performance Results

ρ	$N=3$							$N=10$						
	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)
Panel A: Bias(×1000)														
$T=100$														
0.0	0.0072	0.0290	0.0424	-0.0223	-0.0117	-0.0113	-0.0166	0.0032	0.0372	-0.0111	-0.0027	-0.0256	0.0064	0.0054
0.3	-0.0136	0.0224	0.0132	-0.0320	-0.0487	-0.0048	-0.0172	-0.0039	0.0245	0.0766	-0.0350	0.0056	0.0107	0.0162
0.6	-0.2515	-0.2139	-0.1977	-0.1523	-0.1322	-0.0199	-0.0397	-0.0992	-0.0026	-0.0052	-0.0879	-0.0979	-0.0192	0.0284
0.9	-0.3415	-0.4852	-0.3273	-0.2771	-0.1016	-0.0472	0.0054	0.2630	0.2904	0.1194	-0.2756	0.1055	-0.0704	0.1146
$T=200$														
0.0	0.0010	0.0130	0.0058	-0.0085	-0.0124	-0.0122	-0.0130	-0.1076	-0.1068	-0.1268	-0.1095	-0.1277	-0.0022	-0.0037
0.3	-0.0024	0.0132	0.0267	-0.0141	0.0015	-0.0100	-0.0062	-0.1279	-0.1340	-0.1565	-0.1398	-0.1466	-0.0026	-0.0032
0.6	-0.0058	0.0152	0.0363	-0.0079	0.0278	-0.0072	0.0085	-0.1811	-0.1856	-0.1828	-0.1921	-0.1662	-0.0051	-0.0048
0.9	0.0983	0.2594	0.1918	0.1491	0.1735	0.0134	0.0632	-0.1054	-0.0625	-0.1418	0.0092	0.1049	0.0057	0.0206
$T=500$														
0.0	-0.0045	-0.0052	-0.0055	-0.0073	-0.0075	0.0004	0.0003	0.0023	0.0041	0.0055	0.0000	-0.0001	0.0002	0.0004
0.3	-0.0022	-0.0055	-0.0014	-0.0083	-0.0046	0.0005	0.0013	0.0050	0.0092	0.0070	0.0003	-0.0038	0.0001	0.0002
0.6	-0.0060	-0.0132	0.0018	-0.0213	-0.0083	0.0013	0.0033	0.0205	0.0341	0.0215	0.0096	-0.0049	-0.0003	0.0003
0.9	0.0611	0.0565	0.0875	-0.0719	0.0043	0.0224	0.0308	0.1353	0.1905	0.0986	0.1363	0.0296	0.0221	0.0041
$T=1000$														
0.0	0.0024	0.0027	0.0025	0.0036	0.0033	0.0016	0.0016	-0.0008	-0.0008	-0.0013	-0.0016	-0.0018	0.0002	0.0001
0.3	0.0029	0.0037	0.0031	0.0040	0.0036	0.0028	0.0020	-0.0008	-0.0013	-0.0003	-0.0010	0.0005	0.0004	0.0005
0.6	0.0023	0.0041	0.0031	0.0046	0.0033	0.0050	0.0026	0.0003	-0.0008	-0.0000	-0.0020	0.0008	0.0001	0.0005
0.9	-0.0327	-0.0337	-0.0173	-0.0092	-0.0088	0.0020	0.0020	-0.0152	-0.0118	-0.0159	-0.0293	-0.0169	-0.0030	0.0002
Panel B: RMSE(×100)														
$T=100$														
0.0	0.6510	0.6622	0.6719	0.6252	0.6366	0.1442	0.1473	0.6272	0.6753	0.6753	0.6203	0.6308	0.0516	0.0536
0.3	0.8628	0.8727	0.8314	0.8133	0.7776	0.2125	0.2077	0.8445	0.8912	0.8063	0.8030	0.7464	0.0806	0.0791
0.6	1.4534	1.4661	1.1622	1.2826	1.0383	0.4379	0.3700	1.4408	1.5428	1.2429	1.2826	1.0698	0.2003	0.1768
0.9	5.3522	5.3509	3.1732	3.9217	2.3985	2.0353	1.3039	5.3170	5.4218	3.8701	3.5654	2.4837	1.2124	0.7385
$T=200$														
0.0	0.2135	0.2153	0.2177	0.2016	0.2044	0.0511	0.0521	0.2220	0.2285	0.2359	0.2126	0.2198	0.0179	0.0186
0.3	0.2927	0.2955	0.2820	0.2753	0.2646	0.0757	0.0729	0.3058	0.3142	0.2867	0.2878	0.2645	0.0280	0.0270
0.6	0.5250	0.5343	0.4173	0.4720	0.3750	0.1590	0.1295	0.5522	0.5828	0.4500	0.4935	0.3791	0.0727	0.0596
0.9	2.7161	2.7459	1.5158	2.0637	1.1447	1.0847	0.6119	2.7668	2.8678	1.8970	1.9010	1.1644	0.6222	0.3432
$T=500$														
0.0	0.0523	0.0528	0.0530	0.0506	0.0511	0.0122	0.0123	0.0528	0.0538	0.0544	0.0509	0.0517	0.0044	0.0045
0.3	0.0746	0.0750	0.0710	0.0705	0.0673	0.0187	0.0176	0.0752	0.0767	0.0700	0.0705	0.0647	0.0070	0.0065
0.6	0.1384	0.1398	0.1082	0.1226	0.0969	0.0396	0.0315	0.1389	0.1455	0.1089	0.1234	0.0924	0.0177	0.0132
0.9	0.8996	0.9094	0.4536	0.6670	0.3304	0.3404	0.1680	0.8942	0.9451	0.5785	0.6213	0.3373	0.2033	0.0896
$T=1000$														
0.0	0.0181	0.0183	0.0184	0.0176	0.0177	0.0044	0.0044	0.0183	0.0187	0.0189	0.0176	0.0178	0.0015	0.0016
0.3	0.0261	0.0262	0.0249	0.0248	0.0236	0.0066	0.0062	0.0264	0.0269	0.0240	0.0247	0.0221	0.0025	0.0023
0.6	0.0486	0.0492	0.0378	0.0437	0.0342	0.0145	0.0111	0.0498	0.0519	0.0372	0.0434	0.0316	0.0063	0.0044
0.9	0.3438	0.3499	0.1639	0.2481	0.1123	0.1296	0.0573	0.3463	0.3710	0.2009	0.2324	0.1111	0.0752	0.0301

Table 91: Bias (×1000) and RMSE (×100) for β_2 with estimated long-run variance.

ρ	$N=3$							$N=10$						
	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)
$T=100$														
0.0	0.1014	0.1132	0.1240	0.1062	0.1222	0.1070	0.1264	0.0970	0.1566	0.2140	0.1592	0.2166	0.1576	0.2382
0.3	0.1256	0.1424	0.1560	0.1482	0.1630	0.1452	0.1774	0.1270	0.2106	0.2982	0.2088	0.2966	0.2156	0.3432
0.6	0.1640	0.1842	0.1988	0.1872	0.2146	0.2054	0.2572	0.1626	0.2758	0.4690	0.2794	0.4632	0.3094	0.5598
0.9	0.2420	0.2664	0.2712	0.2744	0.3042	0.3084	0.3710	0.2434	0.3850	0.7308	0.4024	0.7120	0.4398	0.7846
$T=200$														
0.0	0.0772	0.0844	0.0928	0.0840	0.0910	0.0866	0.0968	0.0770	0.1158	0.1468	0.1146	0.1474	0.1134	0.1606
0.3	0.1042	0.1120	0.1212	0.1116	0.1258	0.1210	0.1384	0.1042	0.1450	0.2010	0.1452	0.2008	0.1468	0.2294
0.6	0.1280	0.1428	0.1548	0.1382	0.1572	0.1562	0.1902	0.1240	0.1940	0.3302	0.1912	0.3282	0.2164	0.4070
0.9	0.2006	0.2218	0.2156	0.2394	0.2362	0.2554	0.2854	0.2036	0.3080	0.6524	0.3324	0.6346	0.3568	0.6962
$T=500$														
0.0	0.0602	0.0674	0.0726	0.0676	0.0734	0.0652	0.0716	0.0576	0.0740	0.0874	0.0740	0.0910	0.0810	0.0978
0.3	0.0852	0.0908	0.0930	0.0888	0.0948	0.0866	0.0974	0.0802	0.0978	0.1280	0.0958	0.1246	0.1046	0.1394
0.6	0.0908	0.0978	0.1036	0.0964	0.1048	0.1000	0.1182	0.0848	0.1258	0.2080	0.1196	0.1886	0.1390	0.2374
0.9	0.1544	0.1686	0.1536	0.1666	0.1432	0.1794	0.1754	0.1476	0.2276	0.5066	0.2412	0.4734	0.2544	0.5000
$T=1000$														
0.0	0.0562	0.0614	0.0646	0.0600	0.0624	0.0600	0.0628	0.0568	0.0688	0.0780	0.0642	0.0750	0.0696	0.0808
0.3	0.0750	0.0760	0.0852	0.0762	0.0834	0.0834	0.0870	0.0796	0.0922	0.1072	0.0888	0.1000	0.0918	0.1122
0.6	0.0776	0.0820	0.0866	0.0804	0.0870	0.0846	0.0926	0.0804	0.1042	0.1554	0.1028	0.1344	0.1088	0.1624
0.9	0.1162	0.1248	0.1076	0.1208	0.0968	0.1322	0.1130	0.1190	0.1834	0.4004	0.1764	0.3400	0.1866	0.3476

Table 92: Empirical null rejection probabilities of t-tests for $\beta_2 = -0.3$ with estimated long-run variance.

ρ	$N=3$						$N=10$					
	MT		Wald		(S)		MT		Wald		(S)	
	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR
$T=100$												
0.0	0.1114	0.1300	0.1782	0.2068	0.1070	0.1264	0.2592	0.4028	0.5130	0.6622	0.1576	0.2382
0.3	0.1618	0.1936	0.2464	0.2758	0.1452	0.1774	0.3708	0.5992	0.6836	0.8460	0.2156	0.3432
0.6	0.2236	0.2660	0.3168	0.3624	0.2054	0.2572	0.5638	0.9260	0.8426	0.9846	0.3094	0.5598
0.9	0.3494	0.3776	0.4654	0.4758	0.3084	0.3710	0.7252	0.9972	0.9364	1.0000	0.4398	0.7846
$T=200$												
0.0	0.0696	0.0804	0.1170	0.1338	0.0866	0.0968	0.1404	0.2176	0.3232	0.4466	0.1134	0.1606
0.3	0.1056	0.1194	0.1732	0.1946	0.1210	0.1384	0.2084	0.3720	0.4634	0.6468	0.1468	0.2294
0.6	0.1526	0.1790	0.2238	0.2536	0.1562	0.1902	0.3518	0.7430	0.6520	0.9216	0.2164	0.4070
0.9	0.2852	0.2808	0.3798	0.3648	0.2554	0.2854	0.5838	0.9906	0.8472	0.9990	0.3568	0.6962
$T=500$												
0.0	0.0494	0.0508	0.0802	0.0852	0.0652	0.0716	0.0570	0.0892	0.1578	0.2284	0.0810	0.0978
0.3	0.0700	0.0778	0.1182	0.1316	0.0866	0.0974	0.1104	0.1594	0.2706	0.3878	0.1046	0.1394
0.6	0.0816	0.0944	0.1352	0.1552	0.1000	0.1182	0.1558	0.4036	0.3804	0.6778	0.1390	0.2374
0.9	0.1878	0.1610	0.2726	0.2294	0.1794	0.1754	0.3764	0.9272	0.6568	0.9856	0.2544	0.5000
$T=1000$												
0.0	0.0370	0.0408	0.0722	0.0772	0.0600	0.0628	0.0362	0.0496	0.1050	0.1394	0.0696	0.0808
0.3	0.0596	0.0620	0.1050	0.1094	0.0834	0.0870	0.0696	0.0956	0.1820	0.2422	0.0918	0.1122
0.6	0.0668	0.0682	0.1106	0.1132	0.0846	0.0926	0.0996	0.2154	0.2440	0.4504	0.1088	0.1624
0.9	0.1190	0.0868	0.1814	0.1348	0.1322	0.1130	0.2354	0.7664	0.4658	0.9208	0.1866	0.3476

Table 93: Empirical null rejection probabilities of corrected t-tests and Wald tests for $\beta_2 = -0.3$ with estimated long-run variance.

ρ	$N=3$							$N=10$						
	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)
Panel A: Bias(×1000)														
$T=100$														
0.0	-0.0658	-0.0658	-0.0658	-0.0118	-0.0118	-0.0093	-0.0093	0.0104	0.0104	0.0104	0.0116	0.0116	0.0051	0.0051
0.3	-0.1928	-0.1928	-0.1886	-0.0569	-0.0646	-0.0842	-0.0852	0.0827	0.0827	0.0636	-0.0121	-0.0099	-0.0181	-0.0196
0.6	0.1741	0.1741	0.1585	0.1520	0.1223	0.1551	0.1232	0.0912	0.0912	0.0868	0.0326	0.0639	0.0551	0.0374
0.9	1.4080	1.4080	-0.6012	0.9456	-0.7935	-0.0162	-0.7956	-0.0386	-0.0386	0.4433	-0.1253	0.3611	-0.0485	0.1832
$T=200$														
0.0	0.0502	0.0502	0.0502	0.0295	0.0295	0.0167	0.0167	0.0385	0.0385	0.0385	0.0103	0.0103	0.0105	0.0105
0.3	0.0262	0.0262	0.0184	0.0252	0.0200	0.0209	0.0219	0.0123	0.0123	0.0163	0.0135	0.0100	0.0104	0.0106
0.6	-0.0278	-0.0278	-0.0296	0.0390	0.0096	0.0348	0.0222	-0.0167	-0.0167	-0.0336	0.0010	-0.0242	-0.0139	-0.0220
0.9	-0.5338	-0.5338	-0.2380	0.1576	-0.1109	0.3154	0.0136	-0.5033	-0.5033	-0.5171	-0.5247	-0.5970	-0.6376	-0.5762
$T=500$														
0.0	-0.0097	-0.0097	-0.0097	-0.0052	-0.0052	-0.0054	-0.0054	-0.0058	-0.0058	-0.0058	-0.0025	-0.0025	-0.0022	-0.0022
0.3	-0.0140	-0.0140	-0.0173	-0.0041	-0.0050	-0.0053	-0.0060	-0.0078	-0.0078	-0.0084	0.0017	0.0007	-0.0003	-0.0012
0.6	-0.0326	-0.0326	-0.0312	-0.0076	-0.0050	-0.0077	-0.0050	-0.0038	-0.0038	-0.0073	0.0080	0.0064	0.0035	0.0031
0.9	-0.0791	-0.0791	-0.0842	-0.0232	-0.0452	-0.0589	-0.0424	-0.1590	-0.1590	-0.1161	-0.0948	-0.0285	-0.0648	-0.0318
$T=1000$														
0.0	-0.0041	-0.0041	-0.0041	-0.0013	-0.0013	-0.0004	-0.0004	0.0051	0.0051	0.0051	0.0003	0.0003	0.0002	0.0002
0.3	-0.0036	-0.0036	-0.0034	-0.0014	-0.0012	-0.0001	0.0000	0.0023	0.0023	0.0023	0.0006	0.0005	0.0007	0.0006
0.6	-0.0090	-0.0090	-0.0063	-0.0061	-0.0034	-0.0027	-0.0005	0.0057	0.0057	0.0062	0.0064	0.0061	0.0073	0.0066
0.9	-0.0132	-0.0132	0.0306	0.0130	0.0477	0.0333	0.0549	-0.0004	-0.0004	-0.0052	-0.0072	-0.0051	0.0021	-0.0024
Panel B: RMSE(×100)														
$T=100$														
0.0	0.6348	0.6348	0.6348	0.3632	0.3632	0.3518	0.3518	0.6375	0.6375	0.6375	0.2028	0.2028	0.1916	0.1916
0.3	0.8579	0.8579	0.8524	0.6147	0.6126	0.5970	0.5963	0.8603	0.8603	0.8517	0.4863	0.4867	0.4796	0.4794
0.6	1.4864	1.4864	1.3861	1.1757	1.1426	1.1225	1.1103	1.4793	1.4793	1.3469	1.0658	1.0384	1.0419	1.0341
0.9	16.4368	16.4368	8.8931	11.7917	8.1434	9.3394	7.8265	16.5605	16.5605	8.6063	10.4472	7.7537	8.2281	7.5840
$T=200$														
0.0	0.2124	0.2124	0.2124	0.1218	0.1218	0.1172	0.1172	0.2188	0.2188	0.2188	0.0691	0.0691	0.0658	0.0658
0.3	0.2897	0.2897	0.2865	0.2053	0.2036	0.1991	0.1978	0.3135	0.3135	0.3098	0.1765	0.1770	0.1740	0.1743
0.6	0.5359	0.5359	0.5031	0.4331	0.4190	0.4160	0.4085	0.5373	0.5373	0.4889	0.3959	0.3869	0.3860	0.3845
0.9	5.1841	5.1841	2.9721	3.8059	2.7388	3.0920	2.6367	5.2402	5.2402	2.7765	3.3317	2.5870	2.7249	2.5520
$T=500$														
0.0	0.0530	0.0530	0.0530	0.0304	0.0304	0.0292	0.0292	0.0504	0.0504	0.0504	0.0166	0.0166	0.0158	0.0158
0.3	0.0770	0.0770	0.0764	0.0536	0.0533	0.0518	0.0517	0.0743	0.0743	0.0732	0.0437	0.0434	0.0427	0.0427
0.6	0.1354	0.1354	0.1285	0.1064	0.1040	0.1028	0.1015	0.1382	0.1382	0.1287	0.1001	0.0980	0.0975	0.0970
0.9	1.0746	1.0746	0.6616	0.8099	0.6160	0.6724	0.5942	1.1192	1.1192	0.6546	0.7525	0.5984	0.6334	0.5923
$T=1000$														
0.0	0.0187	0.0187	0.0187	0.0110	0.0110	0.0106	0.0106	0.0185	0.0185	0.0185	0.0057	0.0057	0.0055	0.0055
0.3	0.0273	0.0273	0.0271	0.0191	0.0190	0.0185	0.0185	0.0266	0.0266	0.0263	0.0153	0.0152	0.0151	0.0151
0.6	0.0495	0.0495	0.0468	0.0394	0.0382	0.0378	0.0373	0.0472	0.0472	0.0428	0.0347	0.0337	0.0336	0.0333
0.9	0.3559	0.3559	0.2201	0.2646	0.2014	0.2215	0.1961	0.3541	0.3541	0.2195	0.2405	0.1988	0.2067	0.1970

Table 94: Bias (×1000) and RMSE (×100) for β_4 using the true long-run variance.

ρ	$N=3$							$N=10$						
	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)	CPR	SOLS	SUR	SOLS(C)	SUR(C)	SOLS(S)	SUR(S)
$T=100$														
0.0	0.0476	0.0476	0.0476	0.0460	0.0460	0.0460	0.0460	0.0510	0.0510	0.0510	0.0472	0.0472	0.0514	0.0514
0.3	0.0330	0.0330	0.0360	0.0380	0.0402	0.0386	0.0396	0.0364	0.0364	0.0416	0.0382	0.0414	0.0436	0.0440
0.6	0.0260	0.0260	0.0254	0.0306	0.0296	0.0300	0.0338	0.0270	0.0270	0.0252	0.0314	0.0330	0.0336	0.0334
0.9	0.2390	0.2390	0.1360	0.2312	0.1596	0.2134	0.1622	0.2442	0.2442	0.1410	0.2266	0.1734	0.1918	0.1658
$T=200$														
0.0	0.0462	0.0462	0.0462	0.0562	0.0562	0.0560	0.0560	0.0528	0.0528	0.0528	0.0506	0.0506	0.0508	0.0508
0.3	0.0458	0.0458	0.0440	0.0470	0.0488	0.0454	0.0460	0.0418	0.0418	0.0396	0.0414	0.0436	0.0440	0.0436
0.6	0.0366	0.0366	0.0348	0.0388	0.0398	0.0396	0.0396	0.0342	0.0342	0.0322	0.0414	0.0428	0.0422	0.0424
0.9	0.1954	0.1954	0.1176	0.1904	0.1336	0.1742	0.1352	0.1970	0.1970	0.1138	0.1838	0.1430	0.1582	0.1370
$T=500$														
0.0	0.0466	0.0466	0.0466	0.0472	0.0472	0.0502	0.0502	0.0470	0.0470	0.0470	0.0530	0.0530	0.0512	0.0512
0.3	0.0454	0.0454	0.0448	0.0496	0.0494	0.0470	0.0466	0.0516	0.0516	0.0500	0.0466	0.0482	0.0464	0.0472
0.6	0.0426	0.0426	0.0446	0.0434	0.0438	0.0438	0.0404	0.0468	0.0468	0.0454	0.0480	0.0464	0.0470	0.0462
0.9	0.1236	0.1236	0.0774	0.1166	0.0952	0.1056	0.0914	0.1286	0.1286	0.0902	0.1300	0.1038	0.1110	0.1018
$T=1000$														
0.0	0.0472	0.0472	0.0472	0.0568	0.0568	0.0546	0.0546	0.0472	0.0472	0.0472	0.0466	0.0466	0.0436	0.0436
0.3	0.0506	0.0506	0.0508	0.0536	0.0526	0.0524	0.0538	0.0492	0.0492	0.0502	0.0458	0.0448	0.0476	0.0478
0.6	0.0516	0.0516	0.0502	0.0488	0.0474	0.0506	0.0524	0.0480	0.0480	0.0452	0.0476	0.0456	0.0458	0.0456
0.9	0.0918	0.0918	0.0628	0.0876	0.0740	0.0826	0.0786	0.0914	0.0914	0.0656	0.0908	0.0786	0.0868	0.0786

Table 95: Empirical null rejection probabilities of t-tests for $\beta_4 = -0.2$ using the true long-run variance.

ρ	$N=3$						$N=10$					
	MT		Wald		(S)		MT		Wald		(S)	
	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR	SOLS	SUR
$T=100$												
0.0	0.0298	0.0298	0.0556	0.0556	0.0460	0.0460	0.0164	0.0164	0.0602	0.0602	0.0514	0.0514
0.3	0.0170	0.0156	0.0322	0.0324	0.0386	0.0396	0.0068	0.0076	0.0242	0.0260	0.0436	0.0440
0.6	0.0098	0.0092	0.0196	0.0136	0.0300	0.0338	0.0044	0.0030	0.0100	0.0052	0.0336	0.0334
0.9	0.2730	0.0994	0.4274	0.1228	0.2134	0.1622	0.3238	0.0796	0.7222	0.0910	0.1918	0.1658
$T=200$												
0.0	0.0260	0.0260	0.0534	0.0534	0.0560	0.0560	0.0130	0.0130	0.0506	0.0506	0.0508	0.0508
0.3	0.0214	0.0202	0.0402	0.0414	0.0454	0.0460	0.0096	0.0116	0.0322	0.0332	0.0440	0.0436
0.6	0.0172	0.0162	0.0316	0.0286	0.0396	0.0396	0.0066	0.0060	0.0184	0.0160	0.0422	0.0424
0.9	0.1992	0.0784	0.3260	0.0960	0.1742	0.1352	0.2238	0.0574	0.5842	0.0612	0.1582	0.1370
$T=500$												
0.0	0.0248	0.0248	0.0538	0.0538	0.0502	0.0502	0.0172	0.0172	0.0484	0.0484	0.0512	0.0512
0.3	0.0260	0.0246	0.0478	0.0480	0.0470	0.0466	0.0130	0.0138	0.0470	0.0468	0.0464	0.0472
0.6	0.0208	0.0194	0.0376	0.0382	0.0438	0.0404	0.0102	0.0116	0.0300	0.0304	0.0470	0.0462
0.9	0.0986	0.0446	0.1770	0.0576	0.1056	0.0914	0.1044	0.0308	0.3360	0.0392	0.1110	0.1018
$T=1000$												
0.0	0.0288	0.0288	0.0496	0.0496	0.0546	0.0546	0.0192	0.0192	0.0544	0.0544	0.0436	0.0436
0.3	0.0276	0.0272	0.0536	0.0534	0.0524	0.0538	0.0164	0.0170	0.0444	0.0436	0.0476	0.0478
0.6	0.0244	0.0248	0.0448	0.0478	0.0506	0.0524	0.0126	0.0134	0.0382	0.0362	0.0458	0.0456
0.9	0.0634	0.0370	0.1198	0.0594	0.0826	0.0786	0.0558	0.0232	0.1918	0.0350	0.0868	0.0786

Table 96: Empirical null rejection probabilities of corrected t-tests and Wald tests for $\beta_4 = -0.2$ using the true long-run variance.

List of Figures

1 Empirical null rejection probabilities for a grid of values of m, with $T = 200$ and $\rho_1 = \rho_2 = 0.3$. 25

2 Empirical null rejection probabilities for a grid of values of m, with $T = 200$ and $\rho_1 = \rho_2 = 0.3$. 26

3 Empirical null rejection probabilities for a grid of values of m and $T = 200$. The left panel displays results for $\hat{H}^{m,0.1}_{mov}$. The right panel displays results for $\hat{H}^{m,0.1}_{mov,sn}$. 26

4 Empirical null rejection probabilities for a grid of values of m and $T = 500$. The left panel displays results for $\hat{H}^{m,0.1}_{mov}$. The right panel displays results for $\hat{H}^{m,0.1}_{mov,sn}$. 27

5 Detection times against $I(1)$ breaks for $m = 0.5$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.3$. 32

6 Monitoring results for CO_2 emissions of Finland using $\hat{H}^{m,0.1}_{mov,sn}$ 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. 37

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 line 1946–2016. 38

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 line 1946–2016. 41

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 line 1946–2016. 44

10 Local asymptotic power against I(1) breaks in the intercept and linear trend case for $\hat{H}^m$. The upper two plots correspond to FM-/D-OLS and the lower plots to IM-OLS. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively. 52

11 Local asymptotic power against I(1) breaks in the intercept only case for FM-/D-OLS and IM-OLS using $\hat{H}^m$, $\hat{H}^m_{sn}$, $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively. 53

12 Local asymptotic power against I(1) breaks in the intercept only case for $\hat{H}^m$, $\hat{H}^m_{sn}$, $\hat{H}^{m,n}_{mov}$ and $\hat{H}^{m,n}_{mov,sn}$, for $n = 0.1, 0.2, 0.3$. The upper two plots correspond to standardized detectors, the lower two plots correspond to self-normalized detectors. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively. 54

13 Local asymptotic power against local I(1) breaks in the intercept only case. The upper two plots correspond to $\hat{H}^m$ and $\hat{H}^m_{sn}$. The lower two plots correspond to $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively. 55

14 Local asymptotic power against local I(1) breaks in the intercept and linear trend case. For further explanations see notes for Figure 13. 56

15 Local asymptotic power against local trend breaks for $\hat{H}^m$ and $\hat{H}^{m,n}_{mov}$ in the intercept and linear trend case. The upper two plots correspond to $m = r = 0.5$. The lower two plots correspond to $m = 0.5$ and $r = 0.75$. 58

16 Local asymptotic power against local trend breaks in the model with intercept and linear trend. The upper two plots correspond to $\hat{H}^m$ and $\hat{H}^m_{sn}$. The lower two plots correspond to $\hat{H}^{m,0.1}_{mov}$ and $\hat{H}^{m,0.1}_{mov,sn}$. Calibration and break fraction are set to $m = 0.5$ and $r = 0.75$, respectively. 59

17 Empirical null rejection probabilities for a grid of values of m, with $T = 500$ and $\rho_1 = \rho_2 = 0.3$. 60

18 Empirical null rejection probabilities for a grid of values of m, with $T = 500$ and $\rho_1 = \rho_2 = 0.3$. 61

19 Size-corrected power against trend breaks for $T = 200$, $m = 0.5$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$. 69

20 Size-corrected power against trend breaks for $T = 500$, $m = 0.5$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$. 70

21 Size-corrected power against trend breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.3$. 71

22 Size-corrected power against trend breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.3$. 72

23 Size-corrected power against trend breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.3$. 73

24 Size-corrected power against slope breaks (in $\theta_{X_{1,1}}$) for $T = 200$ and $\rho_1 = \rho_2 = 0.3$. . 75

25 Size-corrected power against slope breaks (in $\theta_{X_{1,2}}$) for $T = 200$ and $\rho_1 = \rho_2 = 0.3$. . 76

26 Size-corrected power against slope breaks (in $\theta_{X_{1,1}}$) for $T = 200$ and $\rho_1 = \rho_2 = 0.3$. . 77

27 Size-corrected power against slope breaks (in $\theta_{X_{1,1}}$) for $T = 200$ and $\rho_1 = \rho_2 = 0.3$. . 78

28 Detection times against $I(1)$ breaks for $T = 500$, $m = 0.5$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.3$. 79

29 Detection times against $I(1)$ breaks for $m = 0.25$, $r = 0.25$ and $\rho_1 = \rho_2 = 0.3$. 80

30 Detection times against $I(1)$ breaks for $m = 0.25$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.3$. 81

31 Detection times against $I(1)$ breaks for $m = 0.25$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$. 82

32 Detection times against $I(1)$ breaks for $m = 0.5$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$. 83

33 Detection times against $I(1)$ breaks for $m = 0.75$, $r = 0.75$ and $\rho_1 = \rho_2 = 0.3$. 84

34 Detection times against $I(1)$ breaks for $m = 0.5$, $r = 0.5$ and $\rho_1 = \rho_2 = 0.9$. 85

35 The scatter plots show pairs of observations of log per capita GDP and log per capita CO_2 emissions for the calibration period 1946–1973 for the considered twelve countries. 90

36 The scatter plots show pairs of observations of log per capita GDP and log per capita SO_2 emissions for the calibration period 1946–1973 for the considered twelve countries. 90

37 IM-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), 1974–1989 (turquoise squares) and 1990–2016 (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 1946–1989 and the dash-dotted blue line 1946–2016. 93

38 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 further explanations see notes to Figure 8 in the main text. 95

39 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 further explanations see notes to Figure 9 in the main text. 96

40 Overview of the used regression functions. 122

41 Size-corrected power of t-tests for $H_0 : \beta_4 = -0.2$, with estimated long-run variances, $T = 200$ and $\rho = 0.3$. 216

42 Size-corrected power of t-tests for $H_0 : \beta_4 = -0.2$, with estimated long-run variances, $T = 200$ and $\rho = 0.6$. 217

43 Size-corrected power of tests for $H_0 : \beta_{4,1} = \cdots = \beta_{4,N} = -0.2$, with estimated long-run variances, $T = 200$ and $\rho = 0.3$. 217

44 Size-corrected power of tests for $H_0 : \beta_{4,1} = \cdots = \beta_{4,N} = -0.2$, with estimated long-run variances, $T = 200$ and $\rho = 0.6$. 218

45 EKC estimation results for Australia for Equation (284). The left hand side shows a scatter plot and an EKC. The dots show the pairs of observations of ln(GDP) per capita and ln(SO_2) emissions per capita. The lines show results based on inserting 80 equidistant points from the sample range of ln(GDP) per capita, with corresponding fitted values of the linear trend as described in the main text. The solid blue line corresponds to the FM-CPR estimates and the red dashed line to the FM-SUR estimates. The right hand side shows actual and fitted values. The solid black line shows the actual values of ln(SO_2) per capita emissions, color codes for FM-CPR and FM-SUR estimates remain unchanged. 227

46 MKC estimation results for France. For further explanations see notes to Figure 45. 227

47 MKC estimation results for Germany. For further explanations see notes to Figure 45. 228

48 MKC estimation results for Italy. For further explanations see notes to Figure 45. . 228

49 MKC and EKC estimation results for Japan. For further explanations see notes to Figure 45. 229

50 MKC and EKC estimation results for Switzerland. For further explanations see notes to Figure 45. 230

51 MKC estimation results for United States. For further explanations see notes to Figure 45. 231

52 Pooled MKC and EKC estimation results for Switzerland for Equation (284). The parameter $\hat{\beta}_1$ is pooled for lead and zinc. The left hand side shows scatter plots and MKCs and EKCs. The dots show the pairs of observations of log GDP per capita and log substance per capita. The lines show results based on inserting 80 equidistant points from the sample range of ln(GDP) per capita, with corresponding fitted values of the linear trend as described in the main text. The solid blue line corresponds to the FM-CPR estimates, the red dashed line to the FM-SUR estimates and the orange dashed-dotted line to the pooled FM-SUR estimates. The right hand side shows actual and fitted values. The solid black line shows the actual values of log per capita substances, color codes for FM-CPR, FM-SUR and pooled FM-SUR estimates remain unchanged. 236

53 Pooled estimation results for lead. The trend parameter δ is pooled for Italy, Japan and Switzerland, as well as for France and the United States. Additionally, the slope parameter β_1 is pooled for Japan and Switzerland. For further explanations see notes to Figure 52. 237

List of Tables

1 Size-corrected power against I(1) breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.3$. 29

2 Size-corrected power against I(1) breaks for $T = 500$ and $\rho_1 = \rho_2 = 0.3$. 30

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. 34

4 Minimal polynomial degrees for cointegrating EKCs over the calibration period 1946–1973. 35

5 Detection times when monitoring CO_2 emissions using the detector $\hat{H}^{m,0.1}_{mov,sn}$ 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%. 36

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. 39

7 Detection times when monitoring SO_2 emissions using the detector $\hat{H}^{m,0.1}_{mov,sn}$ 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%. 42

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. 43

9 Size-corrected power against $I(1)$ breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.0$. 62

10 Size-corrected power against $I(1)$ breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.6$. 63

11 Size-corrected power against $I(1)$ breaks for $T = 200$ and $\rho_1 = \rho_2 = 0.9$. 64

12 Size-corrected power against $I(1)$ breaks for $T = 500$ and $\rho_1 = \rho_2 = 0.0$. 65

13 Size-corrected power against $I(1)$ breaks for $T = 500$ and $\rho_1 = \rho_2 = 0.6$. 66

14 Size-corrected power against $I(1)$ breaks for $T = 500$ and $\rho_1 = \rho_2 = 0.9$. 67

15 Results of the augmented Dickey and Fuller (1981) and Phillips and Perron (1988) unit root tests for log per capita GDP for the calibration period 1946–1973. The test equation includes intercept and linear trend. For the ADF test, we choose the lag length using the AIC criterion and the maximum lag length is three. *Italic* entries indicate rejection of the null hypothesis at the 10% level and **bold** entries rejection at the 5% level. 86

16 Results of the augmented Dickey and Fuller (1981) and Phillips and Perron (1988) unit root tests for log per capita GDP for the full sample period 1946–2016. For further explanations see notes to Table 15. 87

17 Results of the Shin (1994), Johansen (1995) and CT (Wagner, 2023) tests for log per capita CO_2 emissions and log per capita GDP. The (calibration) period is 1946–1973. For the Johansen test the number of lags is determined using the AIC and the maximum lag length is three. *Italic* entries indicate rejection of the null hypothesis at the 10% level and **bold** entries rejection at the 5% level. 87

18 Results of the Shin (1994), Johansen (1995) and CT (Wagner, 2023) tests for log per capita SO_2 emissions and log per capita GDP. The (calibration) period is 1946–1973. For further explanations see notes to Table 17. 88

19 Minimal polynomial degrees for cointegrating EKCs over the full sample period 1946–2016. The detailed cointegration test results are available in Tables 20 and 21. . . . 88

20 Results of the Shin (1994), Johansen (1995) and CT (Wagner, 2023) tests for log per capita CO_2 emissions and log per capita GDP. The (full sample) period is 1946–2016. For further explanations see notes to Table 17. 89

21 Results of the Shin (1994), Johansen (1995) and CT (Wagner, 2023) tests for log per capita SO_2 emissions and log per capita GDP. The (full sample) period is 1946–2016. For further explanations see notes to Table 17. 89

22 Detection times when monitoring CO_2 emissions using all nine detectors and all three estimation methods FM-OLS, D-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%. 91

23 Detection times when monitoring SO_2 emissions using all nine detectors and all three estimation methods FM-OLS, D-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%. 92

24 FM-OLS and IM-OLS estimation results for a cointegrating quadratic relationship between log per capita GDP and log per capita CO_2 emissions for the full sample period 1946–2016. *Italic* entries indicate significance of coefficients at the 10% level and **bold** entries significance of coefficients at the 5% level. 94

25 The checkmarks indicate whether the specific assumptions of the estimators allow for error serial correlation and regressor endogeneity. 108

26 The checkmarks indicate whether the specific assumptions of the tests allow for error serial correlation and endogeneity, and if long-run variance estimation is required. . . 118

27 General specifications and parameter choices. 120

28 Estimator and test specific choices. 121

29 ARMSE for the Nadaraya-Watson estimator ($\hat{f}_{\mathrm{NW}}$) with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. The columns indicate the bandwidth choices for h. 125

30 ARMSE for the local linear estimator ($\hat{f}_{\mathrm{LocLin}}$) with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. The columns indicate the bandwidth choices for h. 126

31 ARMSE for the estimator $\hat{f}_{\mathrm{DLP}}$ of with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 128

32 Proposed user choices for estimators, based on ARMSE values from our simulation study. 129

33 ARMSE for all considered estimators with their respective parameter choices, with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 131

34 Proposed user choices for specification tests based on null rejection probabilities and size-corrected power from our simulation study. 135

35 Null rejection probabilities for the specification tests with $\rho_0 = 0.6$, $\rho_1 = 0.0$, and $\rho_2 = 0.6$. 137

36 Size-corrected power for the specification tests with $\rho_0 = 0.6$, $\rho_1 = 0.0$, and $\rho_2 = 0.6$. 140

37 ARMSE for the estimator $\hat{f}_{\text{LW1}}$, with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 151

38 ARMSE for the estimator $\hat{f}_{\text{LW2}}$, with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 152

39 ARMSE for the estimator $\hat{f}_{\text{LWeff}}$, with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 153

40 ARMSE for the Nadaraya-Watson estimator with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. . 154

41 ARMSE for all considered estimators with their respective parameter choices, with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 155

42 Null rejection probabilities for the T_{CS} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 156

43 Null rejection probabilities for the T_{CS} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 157

44 Null rejection probabilities for the T_{DG} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 158

45 Null rejection probabilities for the T_{DG} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 159

46 Null rejection probabilities for the T_{WP12} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 160

47 Null rejection probabilities for the T_{WP12} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 161

48 Null rejection probabilities for the T_{WP16} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 162

49 Null rejection probabilities for the T_{WP16} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 163

50 Null rejection probabilities for the T_{WZ} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 164

51 Null rejection probabilities for the T_{WZ} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.0$, and $\rho_2 = 0.6$. 165

52 Null rejection probabilities for the T_{WZ} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 166

53 Null rejection probabilities for the specification tests with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 167

54 Null rejection probabilities for the specification tests with $\rho_0 = 0.0$, $\rho_1 = 0.4$, and $\rho_2 = 0.0$. 168

55 Null rejection probabilities for the specification tests with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 169

56 Size-corrected power for the T_{CS} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 170

57 Size-corrected power for the T_{CS} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 171

58 Size-corrected power for the T_{DG} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 172

59 Size-corrected power for the T_{DG} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 173

60 Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 1 under the alternative hypothesis. 174

61 Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 2 under the alternative hypothesis. 175

62 Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 3 under the alternative hypothesis. 176

63 Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 1 under the alternative hypothesis. 177

64 Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 2 under the alternative hypothesis. 178

65 Size-corrected power for the T_{WP12} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 3 under the alternative hypothesis. 179

66 Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 1 under the alternative hypothesis. 180

67 Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 2 under the alternative hypothesis. 181

68 Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$, for DGP 3 under the alternative hypothesis. 182

69 Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 1 under the alternative hypothesis. 183

70 Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 2 under the alternative hypothesis. 184

71 Size-corrected power for the T_{WP16} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$, for DGP 3 under the alternative hypothesis. 185

72 Size-corrected power for the T_{WZ} specification test with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 186

73 Size-corrected power for the T_{WZ} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.0$, and $\rho_2 = 0.6$. 187

74 Size-corrected power for the T_{WZ} specification test with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 188

75 Size-corrected power for the specification tests with $\rho_0 = 0.0$, $\rho_1 = 0.0$, and $\rho_2 = 0.0$. 189

76 Size-corrected power for the specification tests with $\rho_0 = 0.6$, $\rho_1 = 0.4$, and $\rho_2 = 0.6$. 190

77 Bias ($\times 1000$) and RMSE ($\times 100$) for β_4, with estimated long-run variances. 213

78 Empirical null rejection probabilities of t-tests for $\beta_4 = -0.2$, with estimated long-run variances. 214

79 Empirical null rejection probabilities of corrected t-tests and Wald-type tests for $H_0 : \beta_{4,1} = \cdots = \beta_{4,N} = -0.2$, with estimated long-run variances. 215

80 Considered countries for the period 1927–2006. 219

81 Results of augmented Dickey-Fuller and Phillips-Perron unit root tests for gross domestic product (GDP), aluminum (Al), lead (Pb), zinc (Zn), carbon dioxide (CO_2) and sulfur dioxide (SO_2). All variables are used in logarithms of per capita quantities. Intercept and linear trend are included in the test equation. The sample period is 1927-2006. *Italic* entries denote rejection of the null hypothesis at the 10% level, and **bold** entries denote rejection at the 5% level 220

82 Results of PU non-cointegration and CT cointegration tests for aluminum (Al), lead (Pb), zinc (Zn), carbon dioxide (CO_2) and sulfur dioxide (SO_2). With GDP serving as common integrated regressor. All variables are used in logarithms of per capita quantities and powers up to order 3. Intercept and linear trend are included in the test equations. The period is 1927-2006. 221

83 Results of FM-CPR, FM-SOLS and FM-SUR estimation. The model includes as deterministics intercept and linear trend. *Italic* entries indicate significance of coefficients at the 10% level, and **bold** entries significance of coefficients at the 5% level. The sample period is 1927-2006. 223

84 Results of FM-CPR, FM-SOLS and FM-SUR estimation. The model includes as deterministics intercept and linear trend *Italic* entries indicate significance of coefficients at the 10% level, and **bold** entries significance of coefficients at the 5% level. The sample period is 1927-2006. 224

85 Substances with polynomial degree for each country's system of equation. Intercept and linear trend are included in the equations. The period is 1927-2006. 225

86 Countries with polynomial degree for each substance's system of equation. Intercept and linear trend are included in the equations. The period is 1927-2006. 225

87 List of group members corresponding to Wald-type tests for poolability. 232

88 List of group members corresponding to Wald-type tests for poolability. 233

89 Results of FM-CPR, FM-SUR and pooled FM-SUR estimation. The model includes as deterministics intercept and linear trend. The parameter $\hat{\beta}_1$ is pooled for lead and zinc. *Italic* entries indicate significance of coefficients at the 10% level, and **bold** entries significance of coefficients at the 5% level. The sample period is 1927-2006. . 234

90 Results of FM-CPR, FM-SUR and pooled FM-SUR estimation. The model includes as deterministics intercept and linear trend. *Italic* entries indicate significance of coefficients at the 10% level, and **bold** entries significance of coefficients at the 5% level. The sample period is 1927-2006. 235

91 Bias ($\times 1000$) and RMSE ($\times 100$) for β_2 with estimated long-run variance. 252

92 Empirical null rejection probabilities of t-tests for $\beta_2 = -0.3$ with estimated long-run variance. 253

93 Empirical null rejection probabilities of corrected t-tests and Wald tests for $\beta_2 = -0.3$ with estimated long-run variance. 253

94 Bias ($\times 1000$) and RMSE ($\times 100$) for β_4 using the true long-run variance. 254

95 Empirical null rejection probabilities of t-tests for $\beta_4 = -0.2$ using the true long-run variance. 255

96 Empirical null rejection probabilities of corrected t-tests and Wald tests for $\beta_4 = -0.2$ using the true long-run variance. 255

Bibliography

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.

Dong, C. and J. Gao (2018). Specification Testing Driven by Orthogonal Series for Nonlinear Cointegration with Endogeneity. *Econometric Theory* **34**, 754–789.

Dong, C., J. Gao and D. Tjøstheim (2016). Estimation for Single-Index and Partially Linear Single-Index Integrated Models. *The Annals of Statistics* **44**, 425–453.

Dong, C. and O. Linton (2018). Additive Nonparametric Models with Time Variable and Both Stationary and Nonstationary Regressors. *Journal of Econometrics* **207**, 212–236.

Dong, C., O. Linton and B. Peng (2021). A Weighted Sieve Estimator for Nonparametric Time Series Models with Nonstationary Variables. *Journal of Econometrics* **222**, 909–932.

Fan, J. and I. Gijbels (1996). Local Polynomial Modeling and its Applications. Chapman and Hall, London.

Gallant, A.R. (1975). Seemingly Unrelated Nonlinear Regressions. *Journal of Econometrics* **3**, 35–50.

Gao, J., D. Tjøstheim and J. Yin (2012). Model Specification between Parametric and Nonparametric Cointegration. Working Paper.

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.

Härdle, W. and O. Linton (1994). Chapter 38 Applied Nonparametric Methods. In *Handbook of Econometrics*, volume 4, Engle, R.F. and D.L. McFadden (eds.), Elsevier, 2295–2339.

Hommel, G. (1988). A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test. *Biometrika* **75**, 383–386.

Hong, S.H. and P.C.B. Phillips (2010). Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity. *Journal of Business and Economic Statistics* **28**, 96–114.

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

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

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

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.

Karlsen, H.A. and D. Tjøstheim (2001). Nonparametric Estimation in Null Recurrent Time Series. *The Annals of Statistics* **29**, 372—416.

Karlsen, H.A., T. Myklebust and D. Tjøstheim (2007). Nonparametric Estimation in a Nonlinear Cointegration Type Model. *The Annals of Statistics* **35**, 252—299.

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.

Knorre, F., M. Wagner and M. Grupe (2021). Monitoring Cointegrating Polynomial Regressions: Theory and Application to the Environmental Kuznets Curves for Carbon and Sulfur Dioxide Emissions. *Econometrics* **9**, 12, 1–35.

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.

Linton, O. and E. Mammen (2008). Nonparametric Transformation to White Noise. *Journal of Econometrics* **142**, 241–264.

Linton, O. and Q. Wang (2016). Nonparametric Transformation Regression with Nonstationary Data. *Econometric Theory* **32**, 1–29.

Mack, Y.P. and H.-G. Müller (1989). Derivative Estimation in Nonparametric Regression with Random Predictor Variable. *Sankhyā: The Indian Journal of Statistics* **51**, 59–72.

Mark, N.C., M. Ogaki and D. Sul (2005). Dynamic Seemingly Unrelated Cointegrating Regressions. *Review of Economic Studies* **72**, 797–820.

Mazzanti, M. and A. Musolesi (2013). The Heterogeneity of Carbon Kuznets Curves for Advanced Countries: Comparing Homogenous, Heterogeneous and Shrinkage/Bayesian Estimators. *Applied Economics* **45**, 3827–3842.

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.

Moon, H.R. (1999). A Note on Fully-Modified Estimation of Seemingly Unrelated Regressions Models With Integrated Regressors. *Economics Letters* **65**, 25–31.

Moon, H.R. and B. Perron (2005). Efficient Estimation of the SUR Cointegration Regression Model and Testing for Purchasing Power Parity. *Econometric Reviews* **23**, 293–323.

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).

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

Park, J.Y. and M. Ogaki (1991). Seemingly Unrelated Canonical Cointegrating Regressions. Mimeo.

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

Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica* **74**, 967–1012.

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.

Vollebergh, H. R. J., B. Melenberg and E. Dijkgraaf (2009). Identifying Reduced-Form Relations with Panel Data: The Case of Pollution and Income. *Journal of Environmental Economics and Management* **58**, 27–42.

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

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.

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

Wang, Q. and P.C.B. Phillips (2009a). Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression. *Econometric Theory* **25**, 710–738.

Wang, Q. and P.C.B. Phillips (2009b). Structural Nonparametric Cointegrating Regression. *Econometrica* **77**, 1901–1948.

Wang, Q. and P.C.B. Phillips (2011). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory* **27**, 235–259.

Wang, Q. and P.C.B. Phillips (2012). A Specification Test for Nonlinear Nonstationary Models. *The Annals of Statistics* **40**, 727–758.

Wang, Q. and P.C.B. Phillips (2016). Nonparametric Cointegrating Regression with Endogeneity and Long Memory. *Econometric Theory* **32**, 359–401.

Wang, Q. and P.C.B. Phillips (2023). Optimal Bandwidth Selection in Nonlinear Cointegrating Regression. *Econometric Theory* **39**, 1325–1337.

Wang, Q. and Y.X.R. Wang (2013). Nonparametric Cointegrating Regression with NNH Errors. *Econometric Theory* **29**, 1–27.

Wang, Q., D. Wu and K. Zhu (2018). Model Checks for Nonlinear Cointegrating Regression. *Journal of Econometrics* **207**, 261–284.

Wang, Q. and K. Zhu (2020). On a Measure of Lack of Fit in Nonlinear Cointegrating Regression with Endogeneity. *Statistica Sinica* **30**, 371–396.

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.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* **57**, 348–368.