Essays on Cointegrating Polynomial Regressions with Applications to the EKC

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften

eingereicht an der Fakultät Statistik
der Technischen Universität Dortmund

vorgelegt von
Peter Grabarczyk

Dortmund, April 2017

Dissertation
eingereicht an der Fakultät Statistik
der Technischen Universität Dortmund

Dortmund, 27. April 2017

Gutachter:
Prof. Dr. Martin Wagner, TU Dortmund
Prof. Dr. Walter Krämer, TU Dortmund
Prof. Dr. Eric Hillebrand, Aarhus University
Tag der mündlichen Prüfung:
1. August 2017

To Emilia

Contents

Introduction

The *environmental Kuznets curve* (EKC) hypothesis postulates an inverted U-shaped relationship between measures of economic development, typically the logarithm of gross domestic product (GDP) per capita, and the logarithm of measures of pollution or emissions per capita, such as carbon dioxide (CO_2) or sulfur dioxide (SO_2). By analogy, the term refers 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 (Bradford *et al.*, 2005). From its inception with the pioneering work of Grossman and Krueger (1991, 1993, 1995) hundreds of refereed publications, both theoretical as well as empirical, have contributed to the still steadily growing EKC literature, see, e. g., Stern (2017) for a recent literature review. The EKC hypothesis is most commonly analyzed in a regression of log emissions per capita on log GDP per capita and its square, or even higher order powers. From an econometric perspective, this approach has been criticized, e. g., with respect to the use of appropriate unit root and cointegration methods. For instance, the logarithm of GDP per capita is often found to be integrated of order one. A large part of the EKC literature ignores the fact that powers of integrated processes are not integrated themselves (Wagner, 2012) and uses standard, i. e., linear, cointegration techniques. In fact, a regression including log GDP per capita and its powers as regressors is a *cointegrating polynomial regression* (CPR), a term coined by Wagner and Hong (2016). CPRs include deterministic variables and polynomially transformed integrated variables as explanatory variables and stationary errors. In this thesis, we address problems related to the use of standard cointegration techniques applied to CPRs for the empirical analysis of EKC-type relationships from an analytical point of view. Furthermore, we provide suitable estimation and inference as well as cointegration testing techniques for CPRs in single equations and also perform multi-country analysis of the EKC including cross-sectional dependencies and parameter heterogeneity.

Alternative approaches to analyze EKC-type cointegrating relationships have been put forward recently in, e. g., Chan and Wang (2015) or Liang *et al.* (2016), who consider nonlinear least squares estimation in a parametric cointegrating regression model involving a known nonlinear regression function. These papers provide limit theory for a wide class of nonlinear regression functions including polynomials, but are restricted to univariate regressors. This may be sufficient for analyzing the EKC hypothesis, but is a limitation for the analysis of related problems involving multiple integrated regressors. Another virtue of considering nonlinear cointegrating relationships in a CPR framework is the preservation of linearity in parameters, which allows for closed form least squares based estimation methods. In presence of endogeneity the limiting distribution of the OLS estimator is contaminated by so-called second order bias terms rendering OLS based inference difficult. To overcome this limitation, several modified OLS estimators have been proposed in the literature, such as the fully modified OLS (FM-OLS) estimator (Phillips and Hansen, 1990), the dynamic OLS (D-OLS) estimator (Saikkonen, 1991; Stock and Watson, 1993), and the integrated modified OLS (IM-OLS) estimator (Vogelsang and Wagner, 2014a).

Chapter 1 analyzes the asymptotics of the standard FM-OLS estimator of Phillips and Hansen (1990) for cointegrating polynomial regressions, i.e., treating not only the stochastic regressor, but also its powers incorrectly as integrated regressors, as is common practice in the EKC literature. The analysis of linear cointegrating relationships dominates a large part of the literature due to its conceptual simplicity and convenience in use. The deployment of these tools in several software packages makes the standard methods tempting to use for the EKC analysis. The empirical analysis in Wagner (2015) illustrates different conclusions with respect to identifying countries in which a cointegrating EKC relationship is present. In this chapter, we show that the asymptotic distribution of the standard FM-OLS estimator turns out to coincide for CPRs with the tailor-made CPR extension of the FM-OLS estimator introduced in Wagner and Hong (2016). In addition, some intermediate results of independent interest are derived. In particular, we show the asymptotic behavior of nonparametric covariance-type estimators involving (scaled) first differences of polynomially transformed integrated processes. The use of linear cointegration tests in CPRs, e. g., Shin (1994)-type cointegration tests based on standard FM-OLS residuals, in conjunction with the Shin (1994) critical values is invalid even asymptotically. In CPRs the limiting distribution of Shin (1994)-type cointegration test statistics depends, apart from the deterministic component and the number of integrated regressors, also on the powers of the integrated regressors included. This is neglected when conducting cointegration tests in CPRs in conjunction with the Shin (1994) critical values. A simulation study is conducted to assess the estimator performance in finite samples. The results illustrate that both, the standard FM-OLS as well as the CPR extension of the FM-OLS estimator, perform similar in CPR models in terms of bias and root mean squared error (RMSE). However, tests based upon the latter show a better performance in terms of lower over-rejections under the null and larger (size-corrected) power for hypothesis testing as well as cointegration testing.

Chapter 2 provides an extension of the integrated modified OLS (IM-OLS) estimator for cointegrating polynomial regressions recently developed in Vogelsang and Wagner (2014a) for the linear cointegration case and extended for a RESET-type test for the null hypothesis of linearity of a cointegrating relationship in Vogelsang and Wagner (2014b). This estimator is based on a partial sum transformation and an augmentation by including all integrated regressors. Unlike other common OLS modifications, such as the FM-OLS estimator or the D-OLS estimator, no tuning parameter is required for estimation. However, for inference a scalar long-run covariance has to be estimated based on suitable choice of kernel and bandwidth. It is shown that the IM-OLS estimator adjusted to CPRs has a zero mean Gaussian mixture limiting distribution that forms the basis for asymptotic standard inference. Since asymptotic standard inference does not capture the impact of kernel and bandwidth choices on the sampling distributions, fixed-b asymptotic theory has been developed in the stationary framework in Kiefer and Vogelsang (2005). We provide fixed-b asymptotic theory for the IM-OLS estimator in the CPR framework, which is asymptotically nuisance parameter free under suitable conditions on the design of the regression equation, referred to as *full design*. In this case, critical values can be tabulated, which depend upon the kernel function, the bandwidth choice, the specification of the deterministic components, the number of integrated regressors and the powers included. Furthermore, an IM-OLS residual based Kwiatkowski *et al.* (1992)-type (KPSS-type) cointegration test is provided with a nuisance parameter free limiting distribution of the test statistic in the full design case. A simulation study suggest that tests based on the IM-OLS estimator in CPRs, both standard asymptotic as well as fixed-b tests, can lead to substantially smaller size distortions for hypothesis testing at the cost of some minor losses in (size-corrected) power compared to FM-OLS and D-OLS based tests, especially for larger extents of serial correlation and endogeneity. The IM-OLS residual based cointegration test performs similar to the FM-OLS residual based test and has good power properties against the variety of alternatives considered in this simulation study. We also apply the established estimation and testing techniques to the EKC hypothesis based on a data set containing CO_2 emissions and GDP for 19 early industrialized countries over the time period 1870–2013. We find evidence for the existence of a quadratic EKC relationship for six countries and in one additional country for a cubic EKC relationship. The results of the FM-OLS and IM-OLS based cointegration tests are well in line with each other. The findings in this chapter indicate that the extension of the IM-OLS estimator to CPRs adds another concept into the toolkit for analyzing CPR relationships, which in particular is robust to serial correlation and endogeneity.

Finally, Chapter 3 analyzes the EKC hypothesis in a multi-country system of equations approach. In addition to the above-mentioned problems of linear cointegration methods for the analysis of EKC-type relationships, a large part of the EKC literature uses panel data cointegration techniques, which are plagued by the restrictive assumptions of cross-sectional independence and parameter homogeneity. The EKC analysis based on multi-country data involves GDP series for countries of geographic contiguity, e.g., Belgium and the Netherlands, which are not expected to be independent. On the other hand, going through different stages of development as well as the absence of coordinated policies against CO_2 or SO_2 emissions in the past may imply different trajectories of country-individual EKCs and in turn cross-sectional parameter heterogeneity (see, e.g., Dijkgraaf and Vollebergh, 2005). Therefore, building upon Hong and Wagner (2014) we consider fully modified OLS estimation for systems of *seemingly unrelated cointegrating polynomial regressions* (SUCPRs). In addition to single-equation cointegrating polynomial regression analysis, this setting allows for the consideration of cross-sectional dependence of the regressors as well as the errors and does not impose parameter homogeneity. Instead, we provide Wald-type tests for *poolability*, i.e. equality of parameters, for subsets of coefficients over potentially different subsets of cross-sections. Non-rejection of the null hypothesis for these tests allows for fully flexible estimation of the system of equations, which turns out to be very useful in the EKC application. We refer to this as *group-wise* pooled settings and consider group-wise pooled estimation of the EKC for CO_2 emissions for six early industrialized countries over the period 1870–2013. The estimation results are similar to those obtained in unrestricted individual CPRs despite the reduction of the number of estimated parameters by about one third. Conversely, we show that estimation in a classical panel approach including cross-sectional parameter homogeneity – except for the intercepts – is rejected by poolability testing and performs severely worse in this application. In case that the cross-sectional dimension is small compared to the time series dimension, a problem-specific approach to pooling that the SUCPR methodology provides is a helpful tool for analyzing multi-country EKC-type relationships.

All simulations and computations for empirical applications have been performed in `MATLAB`. The code containing the respective procedures can be obtained from the author upon request.

Acknowledgments

First of all, I would like to thank my advisor and co-author Professor Martin Wagner, who introduced me into this interesting topic. His constant guidance and support as well as the constructive discussions helped me throughout the development of this thesis and motivated me to keep improving my work.

The financial support of the Deutsche Forschungsgemeinschaft (SFB 823, "Statistical modeling of nonlinear dynamical processes") is gratefully acknowledged. I would like to thank the director of this research project and my dissertation reviewer, Professor Walter Krämer, for his effort in establishing this fruitful research environment. I am also grateful to my dissertation reviewer Professor Eric Hillebrand for spending his valuable time during the review process.

Special thanks go to my colleagues, in particular Rafael Kawka and Oliver Stypka, for the pleasant collaboration and the countless hours of econometric (and non-econometric) discussions.

Last but not least, I would like to thank my friends and my family for their endless support and love.

1. “Standard” Fully Modified OLS Estimation of Cointegrating Polynomial Regressions

1.1. Introduction

The development of asymptotic estimation and inference theory for unit root and cointegration analysis has experienced rapid progress over the past few decades. Most models employed in empirical research are linear in variables and convenient in use for applied work as several software packages give access to these tools of econometric analysis to many fields of empirical research. Given the particular application at hand, linear models may be too restrictive to capture the features of long-run relationships adequately. Extensions to nonlinear cointegrating relationships have been put forward recently. However, the nonlinear cutting-edge techniques are still in its infancy relative to the linear counterparts – especially with respect to applied work. From this point of view, it is worth investigating the impact of applying linear cointegration estimation and inference techniques in nonlinear cointegration models.

The present chapter analyzes analytically the asymptotics of the fully modified OLS (FM-OLS) estimator of Phillips and Hansen (1990) for cointegrating polynomial regressions (CPRs), i. e., regressions including deterministic variables, integrated processes as well as integer powers of integrated processes as explanatory variables and stationary errors. The CPR framework allows to develop linear least squares based estimation methods and is applicable in the contexts of, e. g., purchasing power parity (PPP) or the environmental Kuznets curve (EKC) hypothesis.[1] The former is considered in Hong and Phillips (2010), who present a specification test for more general nonlinear cointegration regressions based

[1]The term EKC, coined by Grossman and Krueger (1995), refers by analogy to the inverted U-shaped relationship between the level of economic development and the degree of income inequality postulated by Simon Kuznets (1955) in his 1954 presidential address to the American Economic Association. Already early survey papers like Stern (2004) or Yandle *et al.* (2004) find more than 100 refereed publications; with many more written since then. See also the discussions in Wagner (2015) and Wagner and Grabarczyk (2017) for additional references and some background.

on approximations by polynomial basis functions. On the other hand, the EKC hypothesis, which postulates an inverted U-shaped relationship between GDP and emissions, is the original motivation for considering CPRs in Wagner and Hong (2016). The hypothesized inverted U-shape suggests the inclusion of GDP and at least its square as explanatory variables. It is known that integer powers of an integrated process are not integrated processes (see, e. g., Wagner, 2012). Nevertheless, the empirical EKC literature that uses unit root and cointegration techniques employs standard estimation methods for linear cointegrating relationships, with few exceptions, e. g., Chan and Wang (2015) and Wagner (2015). This means that, e. g., the FM-OLS estimator is applied treating not only the stochastic regressor, but also its integer powers *incorrectly* as integrated regressors. This approach is referred to as FM-LIN in this chapter (defined in (1.10) in Section 1.2). Wagner and Hong (2016) adapt the FM-OLS estimator to the CPR case (defined in (1.6) in Section 1.2), labeled FM-CPR hereafter. The main result of this chapter shows that the asymptotic distributions of the FM-LIN and the FM-CPR estimators coincide for CPRs, thereby developing some intermediate results related to nonparametric long-run covariance estimation that are of independent interest.

An immediate implication of the main result is that the asymptotic distributions of the Shin (1994)-type cointegration test statistic, as discussed in Wagner and Hong (2016) for CPRs, coincide for both the FM-LIN and the FM-CPR residuals. The critical values for this test depend upon the specification of the equation (Wagner, 2013), i. e., upon the deterministic component as well as the number and powers of integrated regressors included. Consequently, testing for cointegration using the FM-LIN residuals in conjunction with the Shin (1994) critical values, is invalid even asymptotically. Thus, in contrast to estimation for cointegration testing, no surprising asymptotic result rescues the "linear approach". The discussion in Section 1.2 is for the CPR case with only one integrated process and powers thereof as regressors, which is also the most relevant case for the applications we are aware of. The result, however, extends, with only additional notational complexity, to the more general situation considered in Wagner and Hong (2016).[2] Details for the general case are given in Appendix A.2.

The scatter plot shown in Figure 1.1 displays the relationship between log GDP per capita and log CO_2 emissions per capita for Belgium over the period 1870–2009. In addition to the scatter plot, the figure displays estimates obtained by FM-LIN (dashed) and FM-CPR (solid). If log GDP per capita is an integrated process, the results in the figure are derived from a regression involving a unit root process and its square, an intercept and a linear trend as regressors and log CO_2 emissions per capita as dependent variable. Details including definitions and precise assumptions are given in Section 1.2. The results are

[2] The detailed discussion in Section 1.2 shows that the asymptotic equivalence result requires stricter assumptions than used in, e. g., Wagner and Hong (2016).

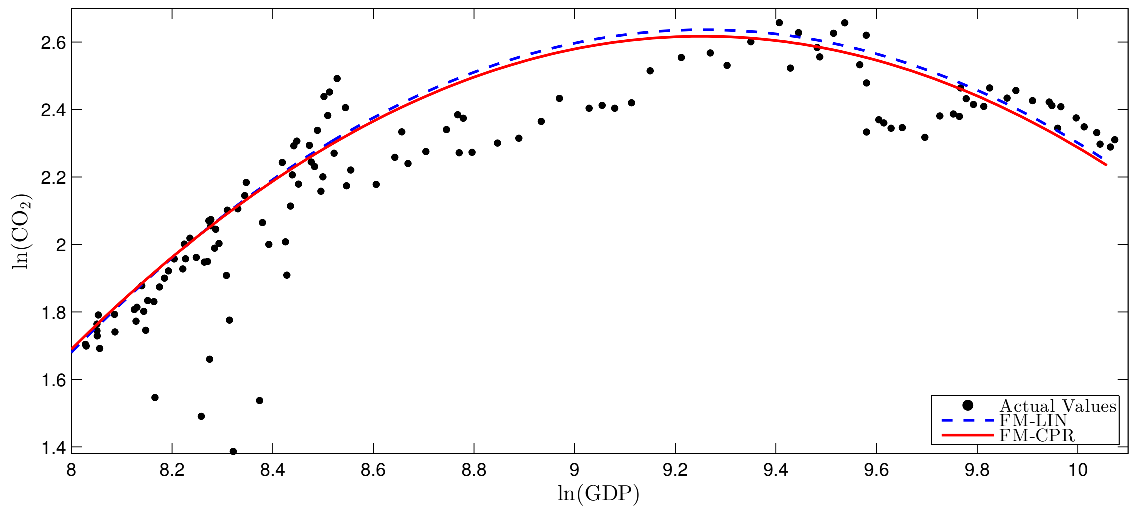

Figure 1.1.: EKC estimation results. The dots show the pairs of observations of log GDP and log CO_2 in per capita terms for Belgium for the years 1870–2009. The curves result from inserting 140 equidistantly spaced points based on the sample range of log GDP per capita and the corresponding values of the trend given in the estimated relationship $\ln(\mathrm{CO}_2)_t = c + \delta t + \beta_1 \ln(\mathrm{GDP})_t + \beta_2 \ln(\mathrm{GDP})_t^2$. Thereby, the coefficient estimates are obtained by FM-LIN (dashed) and FM-CPR (solid).

very similar, despite the fact that the FM-LIN estimator is used in a setting for which it has not been designed.

The theoretical analysis is complemented by a simulation study, that confirms the main result of this chapter and assesses the performance of the FM-LIN estimator for CPRs in small samples. The FM-LIN estimator performs qualitatively similar to the FM-CPR estimator in terms of bias and root mean squared error, but the simulation study indicates a better hypothesis test as well as cointegration test performance for the latter.

The chapter is organized as follows: In Section 1.2 we present the model and assumptions as well as the theoretical results. Section 1.3 is devoted to a brief simulation study and Section 1.4 summarizes and concludes. Two appendices follow the main text: Appendix A.1 contains some auxiliary lemmata and proofs of the main results. Appendix A.2 illustrates the main arguments of the proofs for the case with more than one integrated regressor. Available additional material contains more detailed simulation results.

We use the following notation: Definitional equality is signified by $:=$, equality in distribution by $\overset{d}{=}$ and weak convergence by $\Rightarrow$. We use $O_{\mathbb{P}}(\cdot)$ to denote boundedness in probability, whereas $o_{\mathbb{P}}(\cdot)$ and $o_{a.s.}(\cdot)$ denote convergence in probability and almost sure convergence. The integer part of $x \in \mathbb{R}$ is given by $\lfloor x \rfloor$ and a diagonal matrix by $\mathrm{diag}(\cdot)$ with entries specified throughout. For a vector $x = (x_i)_{i=1,\dots,n}$ we consider the Euclidean norm $\|x\|^2 := \sum_{i=1}^n x_i^2$ and for a matrix A the j-th column is labeled by $A_{(\cdot,j)}$. We denote with $0_{m\times n}$ an $(m \times n)$-matrix with all entries equal to zero and e_l^k defines the l-th unit vector in $\mathbb{R}^k$. The expectation operator and the first difference operator are labeled by $\mathbb{E}$ and Δ, respectively. Brownian motions are denoted by $B(r)$, with covariance matrix specified in the context and standard Brownian motions by $W(r)$.

1.2. Theory

1.2.1. Model and Assumptions

As mentioned in the introduction, to understand the arguments leading to the results it suffices to consider a cointegrating polynomial regression with one integrated regressor and its powers[3], i. e.,

$$\begin{aligned} y_t &= D_t'\delta + X_t{}'\beta + u_t, \quad \text{for } t = 1, \dots, T, \\ x_t &= x_{t-1} + v_t, \end{aligned} \tag{1.1}$$

where y_t is a scalar process, $D_t \in \mathbb{R}^q$ is a deterministic component, x_t is a scalar $I(1)$ process and $X_t := [x_t, x_t^2, \dots, x_t^p]' \in \mathbb{R}^p$. Denoting with $Z_t := [D_t', X_t']' \in \mathbb{R}^{q+p}$ the stacked regressor matrix and with $\theta := [\delta', \beta']' \in \mathbb{R}^{(q+p)}$ the parameter vector, equation (1.1) can be rewritten more compactly as:

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

Assumption 1. For the deterministic components it suffices to assume that there exists a sequence of $q \times q$ scaling matrices $G_D = G_D(T)$ and a q-dimensional vector of càdlàg functions $D(s)$, with $0 < \int_0^s D(z)D(z)'dz < \infty$ for $0 < s \leq 1$, such that for $0 \leq s \leq 1$ it holds that:

$$\lim_{T\to\infty} T^{1/2} G_D D_{[sT]} = D(s).$$

For the leading case of polynomial time trends[4], the deterministic component has the form $D_t = [1, t, t^2, \dots, t^{q-1}]'$ with $G_D = \text{diag}(T^{-1/2}, T^{-3/2}, T^{-5/2}, \dots, T^{-(q-1/2)})$ and $D(s) = [1, s, s^2, \dots, s^{q-1}]'$.

[3] Note that, of course, not all consecutive powers of x_t need to be included and in case of more than one integrated regressor the included powers can differ across integrated regressors. These changes lead to notational complications only. The initial value x_0 can be any well-defined $O_{\mathbb{P}}(1)$ random variable.

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

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

Assumption 2. The processes $\{u_t\}_{t\in\mathbb{Z}}$ and $\{\Delta x_t\}_{t\in\mathbb{Z}} = \{v_t\}_{t\in\mathbb{Z}}$ are generated as:

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

with $\sum_{j=0}^{\infty} j|c_{uj}| < \infty$, $\sum_{j=0}^{\infty} j|c_{vj}| < \infty$ and $C_v(1) \neq 0$. Furthermore, we assume that the process $\{\xi_t^0\}_{t\in\mathbb{Z}} := \{[\zeta_t, \varepsilon_t]'\}_{t\in\mathbb{Z}}$ is independently and identically distributed with $\mathbb{E}(\|\xi_t^0\|^l) < \infty$ for some $l > \max(8, 4/(1-2b))$ with $0 < b < 1/3$.

The above Assumption 2 is stronger than the corresponding assumption used in Wagner and Hong (2016). To be able to draw upon some of the results of Kasparis (2008) we replace the martingale difference sequence assumptions used in Wagner and Hong (2016) with a linear process assumption and the moment assumption of Kasparis (2008).[5] For univariate $\{x_t\}_{t\in\mathbb{Z}}$ the assumption $C_v(1) \neq 0$ excludes stationary $\{x_t\}_{t\in\mathbb{Z}}$, and has to be modified in the multivariate case to $\det(C_v(1)) \neq 0$, i. e., in the multivariate case the vector process $\{x_t\}_{t\in\mathbb{Z}}$ is assumed to be non-cointegrated.

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

Assumption 3. For the kernel function $k(\cdot)$ we assume that:

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

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

Assumption 4. For the bandwidth parameter M_T we assume that $M_T \subseteq (0, \infty)$ and $M_T = O(T^b)$, with the same parameter b as in Assumption 2.

[5]Note that in Kasparis (2008, Assumption 1(b), p. 1376) a condition of the form $l > \min(8, 4/(1-2b))$ is posited. In the proof of his Lemma A1, however, at different places moments of order $4/(1-2b)$ (p. 1391) and order 8 (p. 1395) are needed. Thus, we believe that the minimum should be replaced by the maximum. Since we use similar arguments in the proofs of our Lemmata 3 and 4 we require moments of order $\max(8, 4/(1-2b))$. As discussed in Wagner and Hong (2016) similar results could also be established under alternative assumptions in the spirit of, e. g., Ibragimov and Phillips (2008) or de Jong (2002), augmented correspondingly to accommodate the powers of the integrated regressor. A key difference to, e. g., Chang *et al.* (2001) is that $\{u_t\}_{t\in\mathbb{Z}}$ is allowed to be serially correlated, in an MDS setting in Wagner and Hong (2016) and in a linear process setting here.

Our Assumption 4 on the bandwidth implies $\lim_{T\to\infty}(M_T^{-1} + T^{-1/3}M_T) = 0$, whereas Jansson (2002) assumes $\lim_{T\to\infty}(M_T^{-1} + T^{-1/2}M_T) = 0$, which corresponds to $M_T = O(T^b)$, with $0 < b < 1/2$. Clearly, our assumption here is stronger. This tightening of the upper bound stems from the fact that for the asymptotic analysis of the FM-LIN estimator defined in (1.10) we need to consider “long-run covariance” estimators involving nonstationary processes. Establishing weak convergence of these terms requires smaller bandwidths. In order to have uniform notation we *formally* define:

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

$$\hat{\Delta}_{ab} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} a_t b'_{t+h}, \tag{1.2}$$

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

$$\hat{\Omega}_{ab} := \hat{\Delta}_{ab} + \hat{\Delta}'_{ab} - \hat{\Sigma}_{ab}, \tag{1.3}$$

with $\hat{\Sigma}_{ab} := T^{-1}\sum_{t=1}^{T} a_t b'_t$.

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

Remark 1. Note also that in our definition of $\hat{\Delta}_{ab}$ we use (like, e. g., Phillips, 1995) the bandwidth M_T rather than $T - 1$ as upper bound of the summation over the index h (like, e. g., Jansson, 2002). For truncated kernels with $k(x) = 0$ for $|x| > 1$ this is of course inconsequential. It can also be shown (see, e. g., Phillips, 1995) that for standard long-run covariance estimation problems, consistency is not affected by either summation index choice also for untruncated kernels like the Quadratic Spectral kernel. In our setting, where the asymptotic behavior of $\hat{\Delta}$-quantities is analyzed for a (properly scaled but) nonstationary process (see Theorem 1 and Corollary 1), the summation bound is important. The key result in Theorem 1 below hinges upon summation only up to M_T. The tighter summation bounds are related to the smaller bandwidths needed postulated in Assumption 4. More specifically, we need this in the proof of Lemma 5. This lemma is related to Kasparis (2008, Lemma A1, p. 1394–1396), where this summation bound is also used (in a slightly different context).

Assumption 2 implies that the process $\{\xi_t\}_{t\in\mathbb{Z}} := \{[u_t, v_t]'\}_{t\in\mathbb{Z}}$ fulfills a functional central limit theorem of the form:

$$\frac{1}{T^{1/2}}\sum_{t=1}^{[rT]}\xi_t \Rightarrow B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \Omega_{\xi\xi}^{1/2} W(r), \quad r \in [0,1], \tag{1.4}$$

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

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

The half (or one-sided) long-run covariance matrix $\Delta_{\xi\xi} := \sum_{h=0}^{\infty}\mathbb{E}(\xi_0\xi_h')$ is also needed below and partitioned similarly as $\Omega_{\xi\xi}$. For FM-type estimation, estimates of the above long-run covariance matrices are required. Below we focus on the estimation of $\Delta_{\xi\xi}$, from which an estimator of $\Omega_{\xi\xi}$ follows using (1.3) and an estimator of $\Sigma_{\xi\xi}$, since the asymptotic behavior of estimators of Δ-type quantities is one of the key elements for the result in Theorem 1.

Unless otherwise stated, in long-run covariance estimation the unobserved errors u_t are replaced by the OLS residuals from (1.1), $\hat{u}_t$. This defines $\hat{\xi}_t := [\hat{u}_t, v_t]'$ and the effects of this replacement are analyzed below.

1.2.2. Fully Modified OLS Estimation

A fully modified OLS (FM-OLS) type estimator for the parameters in (1.1) is presented in Wagner and Hong (2016) by extending the FM-OLS estimation principle from the linear cointegration case considered in Phillips and Hansen (1990) to the CPR setting.[6] We briefly describe the two-part transformation required for FM-CPR estimation. First, the dependent variable y_t is replaced by:

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

[6]Note again that related work has also been undertaken by other authors, including – as already mentioned – Chang *et al.* (2001), de Jong (2002), Ibragimov and Phillips (2008) or Liang *et al.* (2016).

with the long-run covariances estimated from $\hat{\xi}_t$. The second transformation consists of a bias-correction term that is for specification (1.1) given by:

$$A^* := \hat{\Delta}^+_{v\hat{u}} \begin{bmatrix} 0_{q\times 1} \\ T \\ 2\sum_{t=1}^T x_t \\ \vdots \\ p\sum_{t=1}^T x_t^{p-1} \end{bmatrix}, \tag{1.5}$$

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

$$\hat{\theta}^+ := (Z'Z)^{-1}(Z'y^+ - A^*). \tag{1.6}$$

Define

$$G = G(T) := \operatorname{diag}(G_D(T), G_X(T)), \tag{1.7}$$

with $G_X(T) := \operatorname{diag}(T^{-1}, T^{-3/2}, \dots, T^{-(p+1)/2})$ and $J(r) := [D(r)', \mathbf{B}_v(r)']'$, where $\mathbf{B}_v(r) := [B_v(r), B_v^2(r), \dots, B_v^p(r)]'$. Wagner and Hong (2016, Proposition 1) show under slightly weaker assumptions that:

$$G^{-1}(\hat{\theta}^+ - \theta) \Rightarrow \left(\int_0^1 J(r)J(r)'dr\right)^{-1} \int_0^1 J(r)dB_{u\cdot v}(r), \tag{1.8}$$

with $B_{u\cdot v}(r) := B_u(r) - B_v(r)\Omega_{vv}^{-1}\Omega_{vu}$. The zero-mean Gaussian mixture limiting distribution given in (1.8) forms the basis for asymptotically valid standard (chi-squared) inference.

1.2.3. "Standard" Fully Modified OLS Estimation

We now consider the "wrong" approach outlined in the introduction and show that it is asymptotically equivalent to the FM-CPR estimator discussed in the previous subsection, i. e., it is in fact not "wrong" asymptotically. We refer to this estimator, defined formally in (1.10), for brevity as FM-LIN estimator.

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

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

with

$$w_t := \begin{bmatrix} \Delta x_t \\ \Delta x_t^2 \\ \vdots \\ \Delta x_t^p \end{bmatrix} = \begin{bmatrix} v_t \\ 2x_t v_t - v_t^2 \\ \vdots \\ -\sum_{k=1}^{p} \binom{p}{k} x_t^{p-k}(-v_t)^k \end{bmatrix}, \tag{1.9}$$

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

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

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

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

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

$$\hat{\theta}^{++} := (Z'Z)^{-1}(Z'y^{++} - A^{**}), \tag{1.10}$$

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

$$G^{-1}(\hat{\theta}^{++} - \theta) = \left(GZ'ZG\right)^{-1}\left(GZ'u^{++} - GA^{**}\right), \tag{1.11}$$

with the first term, obviously, unchanged compared to the FM-CPR estimator. Thus, consider the two parts of the second expression in (1.11) in more detail using $W := [w_1', \ldots, w_T']'$ and $G_W := G_W(T) = \mathrm{diag}(1, T^{-1/2}, \ldots, T^{-(p-1)/2})$:

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

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

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

where $TG_X = G_W$. Combining the above expressions we can rewrite the centered and scaled FM-LIN estimator as:

$$G^{-1}(\hat{\theta}^{++} - \theta) = \left(GZ'ZG\right)^{-1} \left(GZ'u - GZ'\tilde{W}\hat{\Omega}^{-1}_{\tilde{w}\tilde{w}}\hat{\Omega}_{\tilde{w}u} - GA^{**}\right). \tag{1.12}$$

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

Theorem 1. *Under Assumptions 2 to 4 it holds that*

$$\hat{\Delta}_{\eta\eta} := \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\eta_t\eta'_{t+h} \Rightarrow \Delta_{\eta\eta} := \begin{bmatrix} \Delta_{uu} & \Delta_{uv} & \Delta_{uv}\mathcal{B}' \\ \Delta_{vu} & \Delta_{vv} & \Delta_{vv}\mathcal{B}' \\ \Delta_{vu}\mathcal{B} & \Delta_{vv}\mathcal{B} & \Delta_{vv}\widetilde{\mathcal{B}} \end{bmatrix}, \tag{1.13}$$

as $T \to \infty$, with

$$\mathcal{B} := \left[2\int_0^1 B_v(r)dr, \ldots, p\int_0^1 B_v^{p-1}(r)dr\right]' \tag{1.14}$$

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

$$\widetilde{\mathcal{B}}_{(i,j)} := (1+i)(1+j)\int_0^1 B_v^{i+j}(r)dr. \tag{1.15}$$

Furthermore, as $T \to \infty$, it holds that

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

The above two results lead to:

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

Remark 2. By construction the upper 2×2-blocks in the above results coincides with the long-run and half long-run covariances of the process $\{\xi_t\}_{t\in\mathbb{Z}}$. For all other terms involving an integrated process or some powers of an integrated process we observe weak convergence to functionals of Brownian motions. This is not unexpected, since these terms are the limits of continuous functions (continuous kernel weighted sums) of scaled powers of integrated processes. In particular these terms are not long-run covariances of some underlying stationary processes, but we continue to use the "symbolic notation" $\Delta_{\eta\eta}$, $\Sigma_{\eta\eta}$ and $\Omega_{\eta\eta}$, compare Remark 1. Note again, only the upper left 2×2 blocks are (long-run) covariances.

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

Corollary 1. *Under Assumptions 1 to 4 the same results as above also hold for $\{\tilde{\eta}_t\}$, i. e. as $T \to \infty$:*

$$\begin{aligned}
\hat{\Delta}_{\tilde{\eta}\tilde{\eta}} &:= \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \tilde{\eta}_t \tilde{\eta}_{t+h}' \Rightarrow \Delta_{\eta\eta} \\
\hat{\Sigma}_{\tilde{\eta}\tilde{\eta}} &:= \frac{1}{T} \sum_{t=1}^{T} \tilde{\eta}_t \tilde{\eta}_t' \Rightarrow \Sigma_{\eta\eta} \\
\hat{\Omega}_{\tilde{\eta}\tilde{\eta}} &:= \hat{\Delta}_{\tilde{\eta}\tilde{\eta}} + \hat{\Delta}_{\tilde{\eta}\tilde{\eta}}' - \hat{\Sigma}_{\tilde{\eta}\tilde{\eta}} \Rightarrow \Omega_{\eta\eta}
\end{aligned}$$

It remains to characterize the asymptotic behavior of the remaining component on the right hand side of (1.12).

Lemma 1. *With the data given by* (1.1) *under Assumptions 1 and 2 it holds for*

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

as $T \to \infty$ that:

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

and

$$\left(G_D \sum_{t=1}^{T} D_t w_t' G_w \right)_{(\cdot,j)} \Rightarrow j \int_0^1 D(r) B_v^{j-1}(r) dB_v(r) + j(j-1)\Delta_{vv} \int_0^1 D(r) B_v^{j-2}(r) dr - \binom{j}{2} \Sigma_{vv} \int_0^1 D(r) B_v^{j-2}(r) dr, \tag{1.16}$$

for $j = 2, \dots, p$ *and*

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

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

Combining the results of Theorem 1, Corollary 1 and Lemma 1 allows to establish the main result of this chapter.

Theorem 2. *Let the data be given by* (1.1) *with Assumptions 1 and 2 in place. Furthermore, let long-run covariance estimation be performed with Assumptions 3 and 4 in place. Then it holds for* $T \to \infty$ *that:*

$$G^{-1}(\hat{\theta}^{++} - \theta) \Rightarrow \left(\int_0^1 J(r)J(r)'dr \right)^{-1} \int_0^1 J(r)dB_{u \cdot v}(r). \tag{1.17}$$

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

1.2.4. Testing for Cointegration

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

Corollary 2. *Consider again the cointegrating polynomial regression given in* (1.1)*, Assumptions 2 to 4 in place and denote as before with* $\hat{u}_t^+$ *the FM-CPR and by* $\hat{u}_t^{++}$ *the FM-LIN residuals. Then it holds that both*

$$CT^+ := \frac{1}{T\hat{\omega}_{\hat{u} \cdot v}} \sum_{t=1}^{T} \left(\frac{1}{T^{1/2}} \sum_{j=1}^{t} \hat{u}_j^+ \right)^2 \tag{1.18}$$

and

$$CT^{++} := \frac{1}{T\hat{\omega}_{\hat{u} \cdot w}} \sum_{t=1}^{T} \left(\frac{1}{T^{1/2}} \sum_{j=1}^{t} \hat{u}_j^{++} \right)^2 \tag{1.19}$$

with $\hat{\omega}_{\hat{u}\cdot v} := \hat{\Omega}_{\hat{u}\hat{u}} - \hat{\Omega}_{\hat{u}v}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{v\hat{u}}$ *and* $\hat{\omega}_{\hat{u}\cdot w} := \hat{\Omega}_{\hat{u}\hat{u}} - \hat{\Omega}_{\hat{u}w}\hat{\Omega}_{ww}^{-1}\hat{\Omega}_{w\hat{u}}$ *converge under the null hypothesis as* $T \to \infty$ *to*

$$\int_0^1 \left(W_{u\cdot v}^{J_W}(r)\right)^2 dr, \tag{1.20}$$

with

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

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

Remark 3. Note that in more general CPR models the above test statistic does not necessarily have a nuisance parameter free limiting distribution. The key requirement for this is, using the terminology of Vogelsang and Wagner (2014b), *full design*. In case of only one integrated regressor full design automatically prevails.

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

1.3. Finite Sample Performance

We assess the performance of the FM-LIN and FM-CPR estimators and hypothesis tests based upon them, benchmarked against results obtained with OLS, as well as FM-LIN and FM-CPR based cointegration tests.

We consider the following data generating process:

$$y_t = c + \delta t + \beta_1 x_t + \beta_2 x_t^2 + u_t, \tag{1.21}$$

where the error processes $\{u_t\}_{t\in\mathbb{Z}}$ and $\{\Delta x_t\}_{t\in\mathbb{Z}} = \{v_t\}_{t\in\mathbb{Z}}$ are generated as:

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

ρ_1, ρ_2	OLS	FM-LIN			FM-CPR		
		And	NW	NW_T	And	NW	NW_T
Panel A: Bias							
0.0	0.0003	0.0005	0.0001	0.0005	0.0004	0.0003	0.0005
0.3	0.0191	0.0079	0.0071	0.0074	0.0076	0.0064	0.0071
0.6	0.0785	0.0417	0.0416	0.0403	0.0412	0.0391	0.0382
0.8	0.2020	0.1445	0.1464	0.1447	0.1451	0.1378	0.1397
Panel B: RMSE							
0.0	0.0668	0.0734	0.0728	0.0735	0.0714	0.0712	0.0714
0.3	0.0938	0.1033	0.1020	0.1035	0.0973	0.0965	0.0971
0.6	0.1721	0.1780	0.1748	0.1775	0.1635	0.1592	0.1597
0.8	0.3284	0.3377	0.3290	0.3337	0.3092	0.2982	0.2981

Table 1.1.: Bias and RMSE for coefficient β_1, QS kernel, $T = 100$.

with $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ and $\{e_t\}_{t\in\mathbb{Z}}$ i.i.d. standard normally distributed. The parameter values chosen are $c = \delta = 1$, $\beta_1 = 5$ and $\beta_2 = -0.3$, motivated by the empirical results for EKC estimation in Wagner (2015). The parameter ρ_1 controls the level of serial correlation in the regression error and the parameter ρ_2 controls the level of endogeneity. Both parameters are chosen equally from the set $\{0, 0.3, 0.6, 0.8\}$. We consider the sample sizes $T \in \{50, 100, 200, 500, 1000\}$ and all test decisions are carried out at the nominal 5% significance level. The number of replications is $10,000$ throughout.

For long-run covariance estimation we use the Quadratic Spectral (QS) kernel[7] and bandwidths chosen according to the following rules: the data dependent bandwidth rules of Andrews (1991) (labeled And) and Newey and West (1994) (labeled NW), as well as a simplified sample size dependent version of the latter, i. e., $M_T = \lfloor 4(T/100)^{2/9} \rfloor$ (labeled NW_T).

1.3.1. Bias and Root Mean Squared Error

We start the analysis by considering the performance of the estimators in terms of bias and root mean squared error (RMSE). Given that the results for the coefficients β_1 and β_2 are qualitatively similar throughout, we focus on the results for β_1, displayed in Table 1.1 for sample size $T = 100$. Results for different sample sizes as well as for the coefficient β_2 are available upon request. The FM-LIN and FM-CPR estimators of β_1 are less biased than the OLS estimator in case of serial correlation and endogeneity. Comparable effects also emerge for the RMSE of the estimators. However, the RMSE indicates a slightly better performance of the FM-CPR estimator compared to the FM-LIN estimator. The results

[7]Simulations have also been performed using the Bartlett kernel with results given in additional material. Regarding the kernel choice the estimators and tests based upon them perform similar in case of low level of serial correlation and endogeneity, while the performance is slightly better using the QS kernel in case of high level of correlation.

ρ_1, ρ_2	OLS	FM-LIN			FM-CPR		
		And	NW	$\mathrm{NW_T}$	And	NW	$\mathrm{NW_T}$
Panel A: $T = 50$							
0.0	0.0757	0.2599	0.2798	0.2146	0.2385	0.2308	0.1854
0.3	0.2184	0.2993	0.3096	0.2544	0.2508	0.2558	0.2082
0.6	0.5141	0.4129	0.4253	0.3890	0.3417	0.3428	0.3178
0.8	0.7853	0.6263	0.6363	0.6235	0.5615	0.5541	0.5563
Panel B: $T = 100$							
0.0	0.0597	0.1645	0.1509	0.1493	0.1431	0.1233	0.1285
0.3	0.2066	0.1883	0.1874	0.1738	0.1528	0.1494	0.1415
0.6	0.5352	0.2901	0.3072	0.2813	0.2307	0.2440	0.2253
0.8	0.8164	0.5033	0.5484	0.5196	0.4192	0.4374	0.4564
Panel C: $T = 200$							
0.0	0.0572	0.1184	0.1201	0.1021	0.1027	0.0998	0.0868
0.3	0.2045	0.1338	0.1411	0.1210	0.1101	0.1159	0.1033
0.6	0.5449	0.2011	0.2237	0.2117	0.1574	0.1738	0.1754
0.8	0.8279	0.3885	0.4336	0.4531	0.2776	0.3060	0.4093
Panel D: $T = 500$							
0.0	0.0517	0.0844	0.0785	0.0758	0.0730	0.0672	0.0681
0.3	0.2022	0.0945	0.0977	0.0879	0.0794	0.0851	0.0781
0.6	0.5498	0.1322	0.1650	0.1428	0.0999	0.1136	0.1250
0.8	0.8380	0.2800	0.3583	0.3375	0.1542	0.1964	0.3098
Panel E: $T = 1000$							
0.0	0.0520	0.0691	0.0627	0.0665	0.0648	0.0597	0.0626
0.3	0.2046	0.0748	0.0804	0.0731	0.0684	0.0751	0.0684
0.6	0.5560	0.1026	0.1421	0.1107	0.0792	0.0888	0.0989
0.8	0.8439	0.2185	0.3363	0.2643	0.1078	0.1549	0.2465

Table 1.2.: Empirical null rejection probabilities, Wald test for $H_0 : \beta_1 = 5, \beta_2 = -0.3$, QS kernel, 0.05 level.

for different sample sizes are qualitatively similar. Bias and RMSE become substantially smaller with increasing sample size reflecting (super-) consistency of all estimators.

1.3.2. Finite Sample Performance of Hypothesis Test Statistics

Next, we analyze the performance of the estimators in terms of empirical null rejection probabilities by considering Wald tests[8] for the joint hypothesis $H_0 : \beta_1 = 5, \beta_2 = -0.3$. Following Wagner and Hong (2016, Proposition 2), rejections for all tests are carried out using the chi-squared distribution. Additionally, we consider size-corrected power for a sequence of alternatives on a grid of 21 steps. The values for β_1 are chosen from the set $[5, 5.2]$ on an equidistant grid with mesh 0.01 and for β_2 from $[-0.3, -0.28]$ on an equidistant grid with mesh 0.001.

The simulation results concerning the empirical null rejection probabilities are given in Table 1.2. In case of no serial correlation and endogeneity tests based upon the OLS estimator show only minor size distortions already for small sample sizes. For increasing level

[8]Empirical null rejection probabilities for t-tests for the simple hypotheses $H_0 : \beta_1 = 5$ as well as $H_0 : \beta_2 = -0.3$ are similar to the results for the Wald tests, but substantially closer to the nominal size.

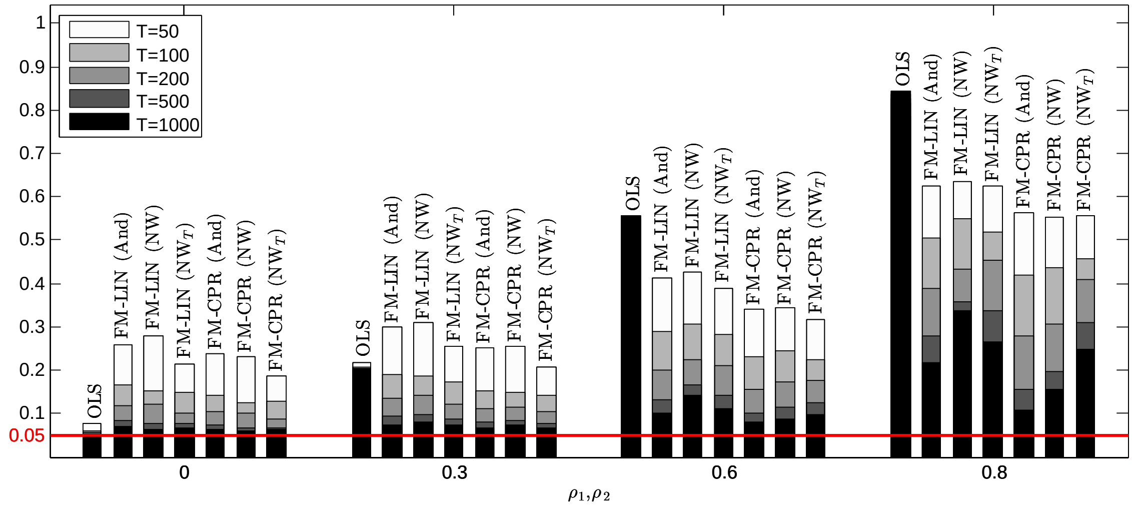

Figure 1.2.: Empirical null rejection probabilities for $H_0 : \beta_1 = 5, \beta_2 = -0.3$, QS kernel.

of serial correlation and endogeneity tests based on the FM-LIN and FM-CPR estimator increasingly outperform OLS based tests. Hong and Phillips (2010, Theorem 2) show that Wald test statistics based on the OLS estimator converge to a noncentral chi-squared distribution with a random noncentrality parameter depending on the extent of correlation. Consequently, we observe increasing empirical null rejection probabilities with increasing sample size in case of positive correlation. Furthermore, the results confirm the theoretical findings from Section 1.2 that not only the tailor-made FM-CPR estimator, but also the FM-LIN estimator corrects for the second-order bias suitably. However, tests based on the FM-CPR estimator exhibit lower size distortions than tests based on the FM-LIN estimator throughout, illustrated in Figure 1.2.

In contrast to the kernel choice, the bandwidth choice has bigger impact on the empirical null rejection probabilities. In particular, the Andrews (1991) data dependent bandwidth choice leads to the smallest size distortions, whereas the data dependent rule of Newey and West (1994) shows poor performance in case of high correlation for FM-LIN based tests. Note that the data dependent bandwidth rules (And and NW) lead to different bandwidths for the FM-LIN and FM-CPR estimators as different inputs enter the procedures. More precisely, for the FM-LIN estimator data dependent bandwidth computation (and consequently long-run covariance estimation) is based on $[\hat{u}_t, w_t']'$, whereas for the FM-CPR estimator it is based on $[\hat{u}_t, v_t]'$. Although the empirical null rejection probabilities decrease for both estimators as the sample size increases, FM-CPR based tests approach the 0.05 level faster in case of high correlation.

Conversely, only the simplified sample size dependent rule $\mathrm{NW_T}$ leads to identical bandwidths, for which the difference between FM-LIN and FM-CPR based tests in terms

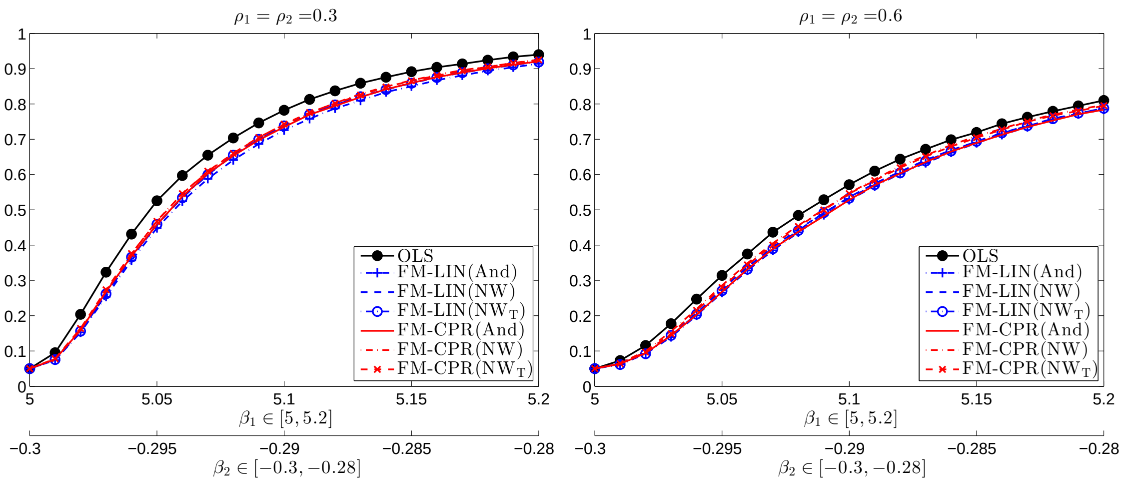

Figure 1.3.: Size-corrected power, Wald test for $H_0 : \beta_1 = 5, \beta_2 = -0.3$, $T = 100$, QS kernel. Left: $\rho_1 = \rho_2 = 0.3$, right: $\rho_1 = \rho_2 = 0.6$.

of empirical null rejection probabilities is smaller and decreasing with increasing sample size.

In terms of size-corrected power, tests based upon the FM-LIN and FM-CPR estimators perform similar irrespective of the bandwidth choice, illustrated in Figure 1.3 for $T = 100$ and moderately large level of correlation. In fact, tests based on the OLS estimator exhibit slightly larger size-corrected power, but suffer from severe size distortions in case of serial correlation and/or endogeneity. In sum, FM-CPR based tests outperform FM-LIN based tests, as substantially lower size distortions are accompanied by no loss in size-corrected power.

1.3.3. Finite Sample Performance of Cointegration Tests

We now consider the performance of the different cointegration tests described below Remark 3. Wagner (2015) highlights empirically the different conclusions from cointegration testing in the EKC analysis. We investigate empirical null rejection probabilities and size-corrected power of the classical cointegration test of Shin (1994) together with the FM-LIN and FM-CPR residual based CT tests given in Corollary 2. The classical Shin (1994) test and the cointegration test based on the FM-LIN residuals share the same test statistic given in (1.19). However, test decisions are based on different limiting distributions and, consequently, on different critical values[9]. We label tests based on the critical values given

[9]Note that the Shin (1994) test is designed for linear cointegrating relationships without polynomial transformations of integrated regressors. Thus, critical values are obtained from a limiting distribution, which does not consist of powers of standard Brownian motions.

ρ_1, ρ_2	CT(Shin)			CT(FM-LIN)			CT(FM-CPR)		
	And	NW	NW_T	And	NW	NW_T	And	NW	NW_T
Panel A: $T = 50$									
0.0	0.0872	0.1404	0.0454	0.0834	0.1372	0.0436	0.0770	0.1086	0.0529
0.3	0.0701	0.1264	0.0475	0.0679	0.1236	0.0448	0.0587	0.0980	0.0578
0.6	0.0870	0.1493	0.0900	0.0836	0.1458	0.0860	0.0533	0.1212	0.1072
0.8	0.1304	0.2265	0.2047	0.1267	0.2222	0.1992	0.1189	0.1967	0.2519
Panel B: $T = 100$									
0.0	0.0480	0.0548	0.0491	0.0430	0.0492	0.0426	0.0498	0.0515	0.0507
0.3	0.0541	0.0701	0.0530	0.0482	0.0636	0.0472	0.0518	0.0721	0.0559
0.6	0.0869	0.1475	0.0933	0.0800	0.1342	0.0847	0.0523	0.1232	0.0954
0.8	0.1706	0.3453	0.2616	0.1520	0.3252	0.2446	0.0634	0.2086	0.2777
Panel C: $T = 200$									
0.0	0.0526	0.0536	0.0554	0.0451	0.0467	0.0479	0.0498	0.0516	0.0518
0.3	0.0596	0.0679	0.0689	0.0521	0.0593	0.0589	0.0554	0.0694	0.0633
0.6	0.0979	0.1331	0.1340	0.0852	0.1196	0.1184	0.0560	0.1082	0.1267
0.8	0.2157	0.3262	0.3797	0.1920	0.3052	0.3526	0.0454	0.1576	0.3724
Panel D: $T = 500$									
0.0	0.0571	0.0559	0.0580	0.0489	0.0488	0.0491	0.0517	0.0511	0.0507
0.3	0.0634	0.0731	0.0684	0.0543	0.0634	0.0580	0.0567	0.0722	0.0604
0.6	0.0903	0.1499	0.1139	0.0795	0.1322	0.1014	0.0571	0.0824	0.1013
0.8	0.2154	0.3797	0.3291	0.1924	0.3527	0.2998	0.0466	0.1342	0.3054
Panel E: $T = 1000$									
0.0	0.0612	0.0605	0.0611	0.0507	0.0521	0.0514	0.0534	0.0527	0.0532
0.3	0.0656	0.0821	0.0683	0.0561	0.0700	0.0586	0.0579	0.0695	0.0592
0.6	0.0899	0.1629	0.1063	0.0769	0.1458	0.0920	0.0584	0.0761	0.0937
0.8	0.2004	0.4000	0.2755	0.1758	0.3717	0.2465	0.0499	0.1222	0.2501

Table 1.3.: Empirical null rejection probabilities of cointegration tests, QS kernel, 0.05 level.

in Shin (1994) by CT(Shin) and tests based on the limiting distribution (1.20) by CT(FM-LIN). Cointegration tests based on the FM-CPR residuals are labeled CT(FM-CPR). We use the data generating process (1.21) under the null. Following Wagner and Hong (2016), we consider three alternative DGPs:

$$\begin{aligned}
\text{(I)} &: \; y_t = 1 + t + 5x_t - 0.3x_t^2 + 0.01x_t^3 + u_t \\
\text{(II)} &: \; y_t = 1 + t + 5x_t - 0.3x_t^2 + z_t, \; \text{ where } z_t \sim I(1) \text{ independent of } x_t \\
\text{(III)} &: \; y_t, \, x_t \text{ are two independent } I(1) \text{ variables}
\end{aligned}$$

These DGPs cover the main alternatives of interest, i. e., (I) missing higher order polynomials of the integrated regressor, (II) no cointegration because of a missing integrated regressor, and (III) spurious regression.

Let us briefly summarize the simulation results concerning the empirical null rejection probabilities. The results displayed in Table 1.3 reveal larger size distortions of the CT(Shin) and CT(FM-LIN) tests compared to the CT(FM-CPR) tests. The CT(FM-LIN) tests show slightly lower size distortions than the CT(Shin) tests throughout. However, the differences between both tests are not severe since the critical values of the corresponding limiting distributions are virtually indistinguishable in the quadratic specification.[10] With respect to the bandwidth choice, the data dependent bandwidth rules, And and NW, lead to empirical sizes of the CT(FM-CPR) tests close to the nominal size. In particular, the former rule leads to almost no size distortions already in small samples. Conversely, the data dependent bandwidth rules lead to substantial size distortions for the CT(Shin) and CT(FM-LIN) tests in case of high correlation also in large samples. The simplified sample size dependent bandwidth rule $\mathrm{NW_T}$, however, leads to size distortions for all three tests considered. The similar empirical null rejection probabilities of the three tests in conjunction with the $\mathrm{NW_T}$ bandwidth rule reflect the empirical findings in Wagner (2015), who shows almost identical cointegration test results in the EKC analysis based on the CT(Shin) and CT(FM-CPR) tests using this particular bandwidth choice.

We complete this section by considering size-corrected power of the cointegration tests against the alternatives (I)–(III). Note that the CT(Shin) and the CT(FM-LIN) tests are based on the same test statistic given in (1.19) and consequently have identical size-corrected power.[11] Therefore, we consider size-corrected power of the CT(Shin) tests and the CT(FM-CPR) tests in Table 1.4. The results indicate that the And bandwidth rule, which leads to the smallest over-rejections under the null, leads to substantially lower size-corrected power than the NW and $\mathrm{NW_T}$ bandwidth rules against alternatives (II) and (III), even for fairly large sample sizes. Against the cubic alternatives (I) size-corrected power is smaller and decreasing for increasing correlation parameters ρ_1, ρ_2. However, the CT(FM-CPR) tests are less sensitive to increasing ρ_1, ρ_2. To summarize, the CT(FM-CPR) tests have more power than the CT(Shin) tests against the three considered alternatives, especially when the NW bandwidth rule is used.

1.4. Summary and Conclusions

The present chapter shows that the asymptotic distribution of the FM-OLS estimator of Phillips and Hansen (1990) when – seemingly unjustified – applied to CPRs coincides with the asymptotic distribution established for the FM-CPR estimator of Wagner and Hong (2016), an estimator tailor-made for CPRs. This result is in turn driven by some results of independent interest for long-run covariance estimation, in the sense of Definition 1, collected in Theorem 1. In contrast to hypothesis testing, FM-LIN residual based

[10] We also consider data generating processes including the third (and fourth) power of the integrated regressor x_t in (1.21). The results, available upon request, indicate that the CT(Shin) tests have more pronounced size distortions in case that higher order powers of integrated regressors are included.

[11] Raw power, reported in additional material, shows similar results for both tests with slight advantages for the CT(Shin) tests.

	ρ_1, ρ_2	CT(Shin)			CT(FM-CPR)		
		And	NW	NW_T	And	NW	NW_T
Panel A: $T = 50$							
(I)	0.0	0.1719	0.0182	0.0852	0.1351	0.0431	0.0985
	0.3	0.1867	0.0221	0.0834	0.1602	0.0520	0.0908
	0.6	0.1713	0.0251	0.0364	0.1670	0.0408	0.0391
	0.8	0.1339	0.0156	0.0048	0.0943	0.0249	0.0040
(II)	-	0.1444	0.0916	0.3747	0.2351	0.1567	0.4249
(III)	-	0.2473	0.0929	0.3965	0.2404	0.1548	0.4508
Panel B: $T = 100$							
(I)	0.0	0.2712	0.1215	0.2114	0.2127	0.1496	0.2282
	0.3	0.2546	0.0892	0.2000	0.2060	0.1045	0.2147
	0.6	0.1972	0.0189	0.1221	0.2048	0.0416	0.1296
	0.8	0.1531	0.0006	0.0254	0.1842	0.0154	0.0254
(II)	-	0.3127	0.6478	0.6305	0.3514	0.5235	0.6696
(III)	-	0.3617	0.6195	0.6306	0.3181	0.5169	0.6705
Panel C: $T = 200$							
(I)	0.0	0.2847	0.2282	0.5372	0.2186	0.2571	0.5444
	0.3	0.2612	0.1939	0.4977	0.2038	0.2057	0.5098
	0.6	0.1960	0.0748	0.3505	0.2014	0.1288	0.3679
	0.8	0.1300	0.0013	0.1246	0.2304	0.0854	0.1288
(II)	-	0.3614	0.8277	0.9142	0.3604	0.6840	0.9217
(III)	-	0.3683	0.8129	0.9148	0.3380	0.6926	0.9241
Panel D: $T = 500$							
(I)	0.0	0.2674	0.5594	0.8446	0.2144	0.5658	0.8471
	0.3	0.2567	0.4940	0.8245	0.2012	0.4947	0.8241
	0.6	0.2036	0.3226	0.7416	0.2016	0.4642	0.7421
	0.8	0.1351	0.0659	0.4942	0.2354	0.3773	0.5008
(II)	-	0.3636	0.9850	0.9935	0.3738	0.9294	0.9939
(III)	-	0.3823	0.9837	0.9941	0.3229	0.9298	0.9940
Panel E: $T = 1000$							
(I)	0.0	0.2775	0.7950	0.9550	0.2128	0.8035	0.9552
	0.3	0.2665	0.7438	0.9494	0.2039	0.7555	0.9502
	0.6	0.2246	0.5843	0.9214	0.2023	0.7455	0.9216
	0.8	0.1477	0.2430	0.7968	0.2230	0.6604	0.7991
(II)	-	0.3799	0.9997	0.9998	0.3680	0.9921	0.9998
(III)	-	0.3854	0.9986	0.9997	0.3138	0.9881	0.9997

Table 1.4.: Size-corrected power of cointegration tests, QS kernel, 0.05 level.

cointegration tests are valid only in conjunction with critical values depending upon the correct model specification, particularly the number and powers of integrated regressors included. The results of a simulation study indicate that both estimators perform similarly in terms of bias and RMSE in finite samples, an observation in line with empirical findings in Wagner (2015). However, the tailor-made FM-CPR estimator has finite sample performance advantages compared to FM-LIN with respect to hypothesis and cointegration testing. Therefore, this chapter justifies *ex post* the usage of standard cointegration methods in the EKC literature, at least for estimation and hypothesis testing. In particular, higher convenience, constituted by ready-to-use software packages, may outweigh the performance loss. The results of this chapter, obviously, raise the question whether such an asymptotic equivalence result between FM-LIN and extensions of the FM-OLS estimator can also be established in more general nonlinear cointegration settings. This intriguing question will be explored in detail in future research.

2. Integrated Modified OLS Estimation for Cointegrating Polynomial Regressions

2.1. Introduction

Cointegration methods are commonly used for modeling empirical financial and macroeconomic relationships. While the largest part of the literature deals with linear cointegrating relationships, which may be sufficient or serve as an adequate approximation in many applications, nonlinear cointegrating relationships have become much more prominent in the last decade. Recent examples are given by empirical analyses in the contexts of purchasing power parity (Hong and Phillips, 2010), money demand functions (Choi and Saikkonen, 2010) or the environmental Kuznets curve hypothesis (Wagner, 2015).

The ordinary least squares (OLS) estimator is super-consistent in cointegrating regression models. In presence of endogeneity and serial correlation its limiting distribution is contaminated by second order bias terms, which renders inference difficult. To overcome this limitation, several modifications of the OLS estimator have been proposed in the linear case, such as the fully modified OLS (FM-OLS) estimator (Phillips and Hansen, 1990), the dynamic OLS (D-OLS) estimator (Saikkonen, 1991) and the integrated modified OLS (IM-OLS) estimator (Vogelsang and Wagner, 2014a). FM-OLS and D-OLS both require the choice of tuning parameters for estimation. FM-OLS is based on a two-step transformation to remove the second order bias terms. These transformations necessitate choices of kernel and bandwidth for long-run covariance estimation. In D-OLS estimation the number of leads and lags included in an augmented regression have to be selected prior to estimation. This augmented regression asymptotically corrects for endogeneity. In contrast to these two OLS modifications, the IM-OLS estimator does not require the choice of tuning parameters. However, for inference a scalar long-run covariance has to be estimated.

This chapter considers the IM-OLS estimator introduced by Vogelsang and Wagner (2014a, 2014b) for cointegrating polynomial regressions (CPRs). Cointegrating polynomial regressions include deterministic variables, integrated processes and integer powers of integrated processes as explanatory variables and stationary errors. Furthermore, the stochastic regressors are allowed to be endogenous and the errors are allowed to be serially correlated. In the CPR framework the IM-OLS estimator is, exactly as in the linear case, based on a partial sum transformation and an augmentation by including all integrated regressors. It is shown that the IM-OLS estimator adjusted to CPRs has a zero mean Gaussian mixture limiting distribution that forms the basis for asymptotic standard inference using a consistent estimator for a long-run covariance parameter. Since asymptotic standard inference does not capture the impact of kernel and bandwidth choices on the sampling distributions, fixed-b asymptotic theory has been developed in the stationary framework in Kiefer and Vogelsang (2005), for the linear cointegration case in Vogelsang and Wagner (2014a) and for a RESET-type test for the null hypothesis of linearity of a cointegrating relationship in Vogelsang and Wagner (2014b). Given full design, defined in the following section, it is shown that the fixed-b limiting distribution of the IM-OLS estimator in the CPR framework is asymptotically nuisance parameter free when using suitably adjusted IM-OLS residuals for long-run covariance estimation. These adjusted IM-OLS residuals are obtained in exactly the same way as in the linear case and lead to fixed-b test statistics with pivotal asymptotic distributions. Thus, critical values can be tabulated in the full design case. They depend upon the kernel function, the bandwidth choice, the specification of the deterministic components, the number of integrated regressors and the powers included. Additionally, an IM-OLS residual based Kwiatkowski *et al.* (1992)-type (KPSS-type) test for cointegration is provided. Again, the limiting distribution is nuisance parameter free in the full design case and can therefore be tabulated.

Extensions of the other mentioned modified OLS estimators to the CPR framework have also been put forward in two recent publications in the literature: Wagner and Hong (2016) develop the FM-OLS estimator for CPRs. They show that this estimator has a zero mean Gaussian mixture limiting distribution and derive Wald- and LM-type specification tests with asymptotic chi-square limiting distributions as well as KPSS-type cointegration tests. Saikkonen and Choi (2004) consider an extension of the D-OLS estimator to more general nonlinear cointegrating regressions, including CPRs.

The theoretical analysis is complemented by a simulation study to assess the finite sample performance of the estimators in terms of bias and root mean squared error (RMSE) as well as the test performance in terms of empirical null rejection probabilities and size-corrected power. For the IM-OLS estimator we consider both, standard asymptotic inference as well as fixed-b inference. Apart from the above mentioned extensions of the FM-OLS and D-OLS estimator, we also benchmark the results against the standard OLS estimator with an in general nuisance parameter dependent limiting distribution. We find that the D-OLS and IM-OLS estimator show slightly lower bias relative to FM-OLS, but the IM-OLS estimator shows weaker performance in terms of finite sample RMSE than D-OLS and FM-OLS. For the hypothesis tests, we observe partly substantially smaller size distortions for tests based on the IM-OLS estimator especially for a larger extent of serial correlation and endogeneity. This holds for both versions of IM-OLS based inference, standard asymptotic inference and fixed-b inference. Comparing both versions directly, the fixed-b version shows overall the smallest size distortions. However, these smaller size distortions come at the cost of some minor losses in (size-adjusted) power. Compared to the FM-OLS residual based cointegration test the introduced IM-OLS residual based cointegration test shows slightly higher over-rejections under the null of cointegration, but has higher size-corrected power against the variety of alternatives considered.

Finally, we use our theoretical findings to estimate the environmental Kuznets curve (EKC)[1], our prime motivation for developing estimation and inference techniques for CPRs. The EKC hypothesis postulates an inverted U-shaped relationship between economic development (measured here by GDP per capita) and pollution (measured here by CO_2 emissions per capita). In order to estimate an inverted U-shape, in addition to GDP per capita also the square and maybe higher integer powers have to be included as explanatory variables in a regression. Starting with the seminal work of Grossman and Krueger (1995), a large part of the empirical EKC literature does not use unit root and cointegration techniques at all. The part of the empirical EKC literature that uses such techniques, however, neglects the fact that powers of integrated regressors are not integrated themselves and applies linear cointegration estimation techniques for the empirical EKC analysis. Wagner (2015) illustrates the different implications of linear versus CPR based cointegration techniques. Thus, building upon the empirical analysis in Wagner (2015), we use the IM-OLS based methods from Section 2.2 to estimate the EKC based on a data set containing CO_2 emissions and GDP for 19 early industrialized countries over the time period 1870–2013 and compare the findings with those obtained by the CPR based extensions of the D-OLS and FM-OLS estimator. We find evidence for the existence of a quadratic EKC relationship for six countries and in one additional country for a cubic EKC relationship. The coefficient estimates are quite similar across the different methods for most of the countries.

The chapter is organized as follows: In Section 2.2 we present the extension of the IM-OLS estimator to the CPR framework and derive its limiting distribution. With respect to inference, we discuss both standard and fixed-b asymptotics for hypothesis tests. We also suggest a KPSS-type test for cointegration based on the IM-OLS residuals. Section 2.3 contains a small simulation study to evaluate the finite sample performance of the proposed

[1]The term refers by analogy to the inverted U-shaped relationship between the level of economic development and the degree of income inequality postulated by Simon Kuznets (1955) in his 1954 presidential address to the American Economic Association.

methods. In Section 2.4 we apply these methods to analyze the EKC hypothesis. Section 2.5 briefly summarizes and concludes. All proofs are given in Appendix B.1, whereas Appendix B.2 contains results of the empirical EKC analysis. Appendix B.3 contains the critical values for the IM-OLS residual based cointegration test and additional simulation results are given in Appendix B.4.

We use the following notation: $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$ and diag$(\cdot)$ denotes a diagonal matrix with entries specified throughout. We denote the m-dimensional identity matrix by I_m and $\mathbb{E}(\cdot)$ denotes the expected value. Definitional equality is signified by $:=$ and $\Rightarrow$ denotes weak convergence. Brownian motions are denoted $B(r)$ or short-hand by B, with covariance matrices specified in the context. For integrals of the form $\int_0^1 B(s)ds$ or $\int_0^1 B(s)dB(s)$, we often use the short-hand notation $\int B$ or $\int BdB$.

2.2. Theory

2.2.1. Setup and Assumptions

We consider the following cointegrating polynomial regression (CPR) model

$$y_t = D_t'\delta + \sum_{j=1}^{m} X_{jt}'\beta_{X_j} + u_t, \tag{2.1}$$

$$x_t = x_{t-1} + v_t, \tag{2.2}$$

where y_t is a scalar time series, $D_t \in \mathbb{R}^d$ a deterministic component, $x_t := [x_{1t}, \ldots, x_{mt}]'$ a non-cointegrating vector of $I(1)$ processes and $X_{jt} := [x_{jt}, x_{jt}^2, \ldots, x_{jt}^{p_j}]'$ a vector including the j-th integrated regressor together with its powers up to power p_j with corresponding parameter vector $\beta_{X_j} := [\beta_{1,j}, \ldots, \beta_{p_j,j}]'$. Furthermore, $X_t := [X_{1t}', \ldots, X_{mt}']'$ and $p := \sum_{j=1}^{m} p_j$.

Assumption 5. For the deterministic component D_t we assume that there exists a d-dimensional vector of càdlàg functions $D(r)$ with $0 < \int_0^r D(z)D(z)'\,\mathrm{d}z < \infty$ for $r \in (0,1]$, such that

$$\lim_{T\to\infty} \sqrt{T} G_D D_{[rT]} = D(r), \; r \in [0,1], \tag{2.3}$$

where $G_D = G_D(T) \in \mathbb{R}^{d\times d}$.

For the leading case of polynomial time trends of the form $D_t = [1, t, t^2, \ldots, t^{d-1}]'$, we have $G_D := \mathrm{diag}\left(T^{-1/2}, T^{-3/2}, \ldots, T^{-(d-1/2)}\right)$ and $D(r) := [1, r, r^2, \ldots, r^{d-1}]'$.

Remark 4. Using consecutive sets of powers for all integrated regressors is merely for ease of notation and any selection of powers can be included in equation (2.1).

Remark 5. CPR models are additively separable, i. e. each nonlinear transformation involves only one integrated regressor, and therefore cross-product terms of integrated regressors are excluded. Vogelsang and Wagner (2014b) consider an integrated modified OLS RESET specification test with an augmented regression including cross-products of powers of the integrated regressors. Such a model is referred to as *multivariate* CPR model and CPR models of the form (2.1) are a special case thereof. However, CPR models cover the most relevant case for applications we are aware of, but exclude, e.g., translog production functions (Christensen *et al.*, 1971), where simple cross-products of integrated regressors are included.

Assumption 6. Define $\{\eta_t\}_{t\in\mathbb{Z}} := \{[u_t, v_t']'\}_{t\in\mathbb{Z}}$ by stacking the error processes and assume that this is a vector of $I(0)$ processes, which satisfies a functional central limit theorem (FCLT) of the form

$$T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor}\eta_t \Rightarrow B(r) = \Omega^{1/2}W(r), \quad r\in[0,1], \tag{2.4}$$

where $W(r)$ is a $(1+m)-$dimensional vector of independent standard Brownian motions and

$$\Omega := \sum_{j=-\infty}^{\infty}\mathbb{E}(\eta_{t-j}\eta_t') = \begin{pmatrix}\Omega_{uu} & \Omega_{uv}\\ \Omega_{vu} & \Omega_{vv}\end{pmatrix} > 0 \tag{2.5}$$

is the long-run covariance matrix of the vector error process. Since we want to exclude cointegration in the $I(1)$ vector process $\{x_t\}_{t\in\mathbb{Z}}$, we assume $\Omega_{vv} > 0$.

We partition the Brownian motion processes according to

$$B(r) = \begin{pmatrix}B_u(r)\\ B_v(r)\end{pmatrix}, \quad W(r) = \begin{pmatrix}w_{u\cdot v}(r)\\ W_v(r)\end{pmatrix},$$

and write the limit process in (2.4) by means of the Cholesky decomposition of $\Omega^{1/2}$ as

$$B(r) = \begin{pmatrix}B_u(r)\\ B_v(r)\end{pmatrix} = \begin{pmatrix}\omega_{u\cdot v}^{1/2} & \Omega_{uv}(\Omega_{vv}^{-1/2})'\\ 0 & \Omega_{vv}^{1/2}\end{pmatrix}\begin{pmatrix}w_{u\cdot v}(r)\\ W_v(r)\end{pmatrix}, \tag{2.6}$$

where $\omega_{u\cdot v} := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$.
Unless otherwise stated we denote the OLS residuals from (2.1) by $\hat{u}_t$ such that a non-parametric kernel estimator of Ω is given by

$$\hat{\Omega} := T^{-1} \sum_{i=1}^{T} \sum_{j=1}^{T} k\left(\frac{|i-j|}{M}\right) \hat{\eta}_i \hat{\eta}_j', \tag{2.7}$$

where $\hat{\eta}_t := [\hat{u}_t, v_t']'$, $k(\cdot)$ is the kernel weighting function and M is the bandwidth. Under standard assumptions on kernel and bandwidth (see e.g. Jansson, 2002, Phillips, 1995) estimators of the form (2.7) provide consistent estimates of the long-run covariance.

For the asymptotics of the powers of the integrated regressors define the weighting matrix $G_X(T) := \mathrm{diag}(G_{X_1}(T), \ldots, G_{X_m}(T))$ with $G_{X_j}(T) := \mathrm{diag}(T^{-1}, T^{-3/2}, \ldots, T^{-\frac{p_j+1}{2}})$, for notational brevity we often drop the argument and simply write $G_X = G_X(T)$. Under the assumptions stated, for t such that $\lim_{T\to\infty} t/T = r$ the following result holds (compare Chang *et al.*, 2001)

$$\lim_{T\to\infty} \sqrt{T} G_{X_j} X_{jt} = \lim_{T\to\infty} \begin{pmatrix} T^{-1/2} & & \\ & \ddots & \\ & & T^{-p_j/2} \end{pmatrix} \begin{pmatrix} x_{jt} \\ \vdots \\ x_{jt}^{p_j} \end{pmatrix} = \begin{pmatrix} B_{v_j} \\ \vdots \\ B_{v_j}^{p_j} \end{pmatrix} =: \mathbf{B}_{v_j}(r), \tag{2.8}$$

with $v_t := [v_{1t}, \ldots, v_{mt}]'$ and denote the stacked vector of powers of Brownian motions as $\mathbf{B}_v(r) := [\mathbf{B}_{v_1}(r)', \ldots, \mathbf{B}_{v_m}(r)']'$.

2.2.2. IM-OLS Estimation in the CPR Framework

In order to establish the IM-OLS estimator compute the partial sums in model (2.1) as

$$\begin{aligned} S_t^y &= S_t^{D\prime}\delta + \sum_{j=1}^{m} S_t^{X_j\prime}\beta_{X_j} + S_t^u, \\ S_t^y &= S_t^{D\prime}\delta + S_t^{X\prime}\beta + S_t^u, \end{aligned} \tag{2.9}$$

where $S_t^y := \sum_{i=1}^{t} y_i$ and $S_t^D, S_t^{X_j}, S_t^X$ and S_t^u defined analogously. The parameter vector β_{X_j} belongs to the j-th integrated regressors and its powers, thus $\beta := [\beta_{X_1}', \ldots, \beta_{X_m}']'$. We stack the vectors in the following form $S_t^X := [S_t^{X_1\prime}, \ldots, S_t^{X_m\prime}]'$ and $S_t^{\tilde{X}} := [S_t^{D\prime}, S_t^{X\prime}]$, such that equation (2.9) is given in compact form as

$$S^y = S^{\tilde{X}}\theta + S^u, \tag{2.10}$$

with $\theta := [\delta', \beta']'$. To correct for endogeneity the partial summed regression is augmented by the vector of integrated regressors x_t, which leads to

$$S_t^y = S_t^{D\prime}\delta + S_t^{X\prime}\beta + x_t'\gamma + S_t^u. \tag{2.11}$$

Setting $S_t^\xi := [S_t^{D\prime}, S_t^{X\prime}, x_t']$ and redefining $\theta := [\delta', \beta', \gamma']'$ gives the more compact form

$$S_t^y = S_t^\xi\theta + S_t^u. \tag{2.12}$$

The IM-OLS estimator is defined as the OLS estimator of the model (2.12). Estimating equation (2.12) via OLS leads to residuals, which we denote by

$$\tilde{S}_t^u = S_t^y - S_t^{D\prime}\tilde{\delta} - S_t^{X\prime}\tilde{\beta} - x_t'\tilde{\gamma}, \tag{2.13}$$

where $\tilde{\theta} = [\tilde{\delta}', \tilde{\beta}', \tilde{\gamma}']'$ denotes the IM-OLS estimator.

The following proposition gives the asymptotic distribution of the IM-OLS estimator and is a special case of Proposition 1 in Vogelsang and Wagner (2014b) for CPR models, for which we define the scaling matrix

$$A_{IM} := \begin{pmatrix} G_D & 0 & 0 \\ 0 & G_X & 0 \\ 0 & 0 & I_m \end{pmatrix}.$$

Proposition 1. *Assume that the data generating process is given by (2.1) and (2.2), the deterministic components satisfy (2.3) and the error process satisfies a FCLT of the form (2.4). With $\theta := [\delta', \beta', (\Omega_{vv}^{-1}\Omega_{vu})']'$ and S^ξ the stacked matrix of S_t^ξ across time, it holds for $T \to \infty$ that*

$$A_{IM}^{-1}(\tilde{\theta} - \theta) = \begin{pmatrix} G_D(\tilde{\delta} - \delta) \\ G_X\left(\tilde{\beta} - \beta\right) \\ (\tilde{\gamma} - \Omega_{vv}^{-1}\Omega_{vu}) \end{pmatrix} = \left(T^{-2}A_{IM}S^{\xi\prime}S^\xi A_{IM}\right)^{-1}\left(T^{-2}A_{IM}S^{\xi\prime}S^u\right) - \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1}\Omega_{vu} \end{pmatrix} \tag{2.14}$$

$$\begin{aligned} \Rightarrow \quad & \omega_{u\cdot v}^{1/2}\left(\int f(s)f(s)'ds\right)^{-1}\int f(s)w_{u\cdot v}(s)ds \\ = \quad & \omega_{u\cdot v}^{1/2}\left(\int f(s)f(s)'ds\right)^{-1}\int [F(1) - F(s)]dw_{u\cdot v}(s), \end{aligned} \tag{2.15}$$

where

$$f(r) := \begin{pmatrix} \int_0^r D(s)ds \\ \int_0^r \mathbf{B}_v(s)ds \\ B_v(r) \end{pmatrix}, F(r) := \int_0^r f(s)ds.$$

The expression (2.15) is, conditional on $W_v(r)$, normally distributed with zero mean and covariance matrix

$$V_{IM} := \omega_{u\cdot v} \left(\int f(s)f(s)'ds \right)^{-1} \left(\int [F(1) - F(s)][F(1) - F(s)]'ds \right) \left(\int f(s)f(s)'ds \right)^{-1}. \tag{2.16}$$

Full Design

In order to perform fixed-b inference in CPR models based on the IM-OLS estimator a necessary condition on the design of the regression equation needs to be ensured, which we refer to as *full design*. In this case the limiting distribution given in Proposition 1 involves only powers of *standard* Brownian motions and is thus nuisance parameter free up to the scalar long-run covariance $\omega_{u\cdot v}$. Full design prevails when only one of the integrated regressors enters with powers larger than one, which is the most relevant case for empirical applications. In more general cases, full design can always be achieved by including additional regressors appropriately into the model. However, this is costly in terms of more parameters to be estimated.

Consider for simplicity the following data generating process

$$y_t = \beta_1 x_{1t} + \beta_2 x_{1t}^2 + \beta_3 x_{2t} + \beta_4 x_{2t}^2 + u_t, \tag{2.17}$$

where under Assumption 6 we have

$$T^{-1/2} \sum_{t=1}^{[rT]} v_t \Rightarrow \begin{pmatrix} B_{v_1}(r) \\ B_{v_2}(r) \end{pmatrix} = \Omega_{vv}^{1/2} W_v(r) = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ 0 & \lambda_{22} \end{pmatrix} \begin{pmatrix} W_{v_1}(r) \\ W_{v_2}(r) \end{pmatrix}. \tag{2.18}$$

The asymptotic distribution of the IM-OLS estimator in this setup involves the vector $\mathbf{B}_v(r) := [B_{v_1}(r), B_{v_1}^2(r), B_{v_2}(r), B_{v_2}^2(r)]'$. It follows from (2.18) that

$$\begin{aligned} B_{v_1}^2(r) &= (\lambda_{11} W_{v_1}(r) + \lambda_{12} W_{v_2}(r))^2 = \lambda_{11}^2 W_{v_1}^2(r) + \lambda_{12}^2 W_{v_2}^2(r) + 2\lambda_{11}\lambda_{12} W_{v_1}(r) W_{v_2}(r) \\ B_{v_2}^2(r) &= \lambda_{22}^2 W_{v_2}^2(r). \end{aligned}$$

Therefore, we have

$$\begin{pmatrix} B_{v_1}(r) \\ B_{v_1}^2(r) \\ B_{v_2}(r) \\ B_{v_2}^2(r) \end{pmatrix} = \underbrace{\begin{pmatrix} \lambda_{11} & 0 & \lambda_{12} & 0 & 0 \\ 0 & \lambda_{11}^2 & 0 & \lambda_{12}^2 & 2\lambda_{11}\lambda_{12} \\ 0 & 0 & \lambda_{22} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{22}^2 & 0 \end{pmatrix}}_{=:F(\Omega_{vv})} \underbrace{\begin{pmatrix} W_{v_1}(r) \\ W_{v_1}^2(r) \\ W_{v_2}(r) \\ W_{v_2}^2(r) \\ W_{v_1}(r)W_{v_2}(r) \end{pmatrix}}_{=:\mathbf{W}_v(r)}. \tag{2.19}$$

If λ_{12} is not equal to zero, the transformation matrix $F(\Omega_{vv})$ does not define a bijective mapping. Consequently, there is no bijective relation between $\mathbf{B}_v(r)$ and $\mathbf{W}_v(r)$. Including the cross-product $x_{1t}x_{2t}$ as an additional regressor in equation (2.17) leads to a transformation matrix $F(\Omega_{vv})$ which is symmetric and of full rank, p^* say, resulting in a bijection between $\mathbf{W}_v(r)$ and $\mathbf{B}_v(r)$, which is now augmented by $B_{v_1}(r)B_{v_2}(r)$.[2] We refer to situations with such a bijection between $\mathbf{W}_v(r)$ and $\mathbf{B}_v(r)$ as *full design*. Given full design, the limiting distribution of the IM-OLS estimator in (2.15) is a function of standard Brownian motions $W(r)$. This allows for asymptotically pivotal fixed-b inference, which we discuss in the next subsection in more detail. Beforehand, we state the limiting distribution of the IM-OLS estimator in the full design case, which is a special case of Corollary 1 in Vogelsang and Wagner (2014b) for CPR models.

Corollary 3. *Suppose that full design prevails and the assumptions of Proposition 1 hold, then for $T \to \infty$*

$$\begin{aligned} A_{IM}^{-1}(\tilde{\theta} - \theta) &\Rightarrow \omega_{u\cdot v}^{1/2} \left(\Pi \int g(s)g(s)' ds \Pi' \right)^{-1} \Pi \int g(s) w_{u\cdot v}(s) ds \\ &= \omega_{u\cdot v}^{1/2} (\Pi')^{-1} \left(g(s)g(s)' ds \right)^{-1} \int [G(1) - G(s)] dw_{u\cdot v}(s), \end{aligned} \tag{2.20}$$

where

$$\Pi := \begin{pmatrix} I_d & 0 & 0 \\ 0 & F(\Omega_{vv}) & 0 \\ 0 & 0 & \Omega_{vv}^{1/2} \end{pmatrix}, \; g(r) := \begin{pmatrix} \int_0^r D(s)ds \\ \int_0^r \mathbf{W}_v(s)ds \\ W_v(r) \end{pmatrix}, \; G(r) := \int_0^r g(s)ds.$$

[2] Clearly, in this case $F(\Omega_{vv})$ is of full rank as long as λ_{11} and λ_{22} in (2.18) are not equal to zero, which is excluded by the assumption $\Omega_{vv} > 0$.

2.2.3. IM-OLS Based Inference in the CPR Framework

We discuss Wald tests for q linear hypotheses of the form $H_0 : R\theta = r$, where we assume the existence of a nonsingular $q \times q$ scaling matrix A_R such that

$$\lim_{T\to\infty} A_R^{-1} R A_{IM} = R^*, \tag{2.21}$$

where R^* has rank q. The condition on the matrix R given in equation (2.21) is sufficient for the Wald statistics to have chi-squared limiting distributions. Recall the definition $S_t^\xi = [S_t^{D\prime}, S_t^{X\prime}, x_t']$ from equation (2.12) and S^ξ as the stacked matrix across time. The covariance matrix V_{IM} of this asymptotic distribution immediately suggests estimators of the form

$$\begin{aligned}\check{V}_{IM} &:= \check{\omega}_{u\cdot v} A_{IM}^{-1} \left(S^{\xi\prime} S^\xi\right)^{-1} (C'C) \left(S^{\xi\prime} S^\xi\right)^{-1} A_{IM}^{-1}, \\ &= \check{\omega}_{u\cdot v} \left(T^{-2} A_{IM} S^{\xi\prime} S^\xi A_{IM}\right)^{-1} \left(T^{-4} A_{IM} C'C A_{IM}\right) \left(T^{-2} A_{IM} S^{\xi\prime} S^\xi A_{IM}\right)^{-1} \end{aligned} \tag{2.22}$$

where $\check{\omega}_{u\cdot v}$ denotes an estimator for $\omega_{u\cdot v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$, $C := [c_1, \ldots, c_T]'$ with $c_t := S_T^{S^\xi} - S_{t-1}^{S^\xi}$ and $S_t^{S^\xi} := \sum_{j=1}^t S_j^\xi$. There are several different candidates for an estimator $\check{\omega}_{u\cdot v}$: First, $\hat{\omega}_{u\cdot v}$ based on the OLS residuals from model (2.1), i.e. the estimator for Ω given in equation (2.7). Second, use the first differences of the OLS residuals of the regression in equation (2.12) to estimate $\omega_{u\cdot v}$ as

$$\tilde{\omega}_{u\cdot v} := T^{-1} \sum_{i=2}^T \sum_{j=2}^T k\left(\frac{|i-j|}{M}\right) \Delta\tilde{S}_i^u \Delta\tilde{S}_j^u.$$

Tests using this estimator are shown to be asymptotically conservative under standard asymptotics, because this estimator is inconsistent under traditional bandwidth assumptions. Therefore, we abstain to consider the asymptotics for tests based on this estimator. Following the discussion in Vogelsang and Wagner (2014a) Section 5, correlation between these residuals and the OLS estimator of equation (2.12) causes problems for fixed-b inference for θ. Consequently, residuals need to be adjusted in a similar way. Define the vector z_t as

$$z_t := t\sum_{j=1}^T \xi_j - \sum_{j=1}^{t-1}\sum_{s=1}^j \xi_s, \quad \xi_t := [S_t^{D\prime}, S_t^{X\prime}, x_t']' \tag{2.23}$$

and let $z_t^\perp$ denote the vector of residuals from individually regressing each element of z_t on the regressors S_t^D, S_t^X, x_t. The adjusted residuals obtained as the OLS residuals from the regression of $\tilde{S}_t^u$ on $z_t^\perp$ are

$$\tilde{S}_t^{u*} := \tilde{S}_t^u - z_t^{\perp\prime}\hat{\pi}, \tag{2.24}$$

where $\hat{\pi} := \left(\sum_{t=1}^{T} z_t^{\perp} z_t^{\perp\prime}\right)^{-1} \sum_{t=1}^{T} z_t^{\perp} \tilde{S}_t^{u}$. As a third option for estimating $\omega_{u.v}$, we use the first differences of the adjusted residuals given in equation (2.24):

$$\tilde{\omega}_{u\cdot v}^{*} := T^{-1} \sum_{i=2}^{T} \sum_{j=2}^{T} k\left(\frac{|i-j|}{M}\right) \Delta\tilde{S}_i^{u*} \Delta\tilde{S}_j^{u*}.$$

This estimator of the long-run covariance $\omega_{u\cdot v}$ has the required properties to deliver a pivotal fixed-b limit for the Wald statistics, see Proposition 2 below. Beforehand, we present the asymptotic behavior of the partial sum process $\Delta\tilde{S}_t^{u*}$.

Lemma 2. *(i) Consider the OLS estimator from the regression*

$$S_t^{y} = S_t^{D\prime}\delta^{*} + S_t^{X\prime}\beta^{*} + x_t^{\prime}\gamma^{*} + z_t^{\prime}\kappa^{*} + S_t^{u}, \tag{2.25}$$

denoted by $\tilde{\theta}^{*}$ *with* $\theta^{*} := [\delta', \beta', (\Omega_{vv}^{-1}\Omega_{vu})', 0]'$. *Under full design it holds that*

$$\begin{pmatrix} A_{IM} & 0 \\ 0 & T^{-2}A_{IM} \end{pmatrix}^{-1} \left(\tilde{\theta}^{*} - \theta^{*}\right) \Rightarrow \omega_{u\cdot v}^{1/2} \begin{pmatrix} (\Pi')^{-1} & 0 \\ 0 & (\Pi')^{-1} \end{pmatrix} \left(\int h(s)h(s)'ds\right)^{-1} \int [H(1) - H(s)]\, dw_{u\cdot v}(s),$$

with

$$h(r) := \begin{pmatrix} g(r) \\ \int_0^r [G(1) - G(s)]ds \end{pmatrix}, \quad H(r) =: \int_0^r h(s)ds.$$

(ii) The limiting distribution of the scaled partial sum process of the adjusted residuals is given by

$$\begin{aligned} T^{-1/2} \sum_{t=2}^{[rT]} \Delta\tilde{S}_t^{u*} &\Rightarrow \omega_{u\cdot v}^{1/2} \left(\int_0^r dw_{u\cdot v}(s) - h(r)' \left(\int_0^1 h(s)h(s)'ds\right)^{-1} \int_0^1 [H(1) - H(s)]dw_{u\cdot v}(s)\right) \\ &=: \omega_{u\cdot v}^{1/2} \widetilde{P}^{*}(r), \end{aligned} \tag{2.26}$$

where, conditional on $W_v(r)$, $\widetilde{P}^{*}(r)$ *is uncorrelated with the scaled and centered limit of* $\tilde{\theta}$ *given in equation (2.15) of Proposition 1. Given that both quantities are conditionally Gaussian implies independence.*

The Wald statistic is defined as

$$\check{W} := \left(R\tilde{\theta} - r\right)' \left(RA_{IM}\check{V}_{IM}A_{IM}R'\right)^{-1} \left(R\tilde{\theta} - r\right),$$

where the superscript of W and V_{IM} indicates which estimator is used for $\omega_{u\cdot v}$. The asymptotic behavior of the partial sum process of the first differences $\Delta\tilde{S}_t^{u*}$ given in Lemma 2 provides the basis for pivotal fixed-b limit for Wald statistics. The following proposition is a special case of Proposition 3 in Vogelsang and Wagner (2014b) for CPR models.

Proposition 2. *If $M := bT$ with $b \in [0,1]$ is held fixed as $T \to \infty$, then*

$$\tilde{W}^* \Rightarrow \frac{\chi_q^2}{Q_b(\widetilde{P^*}, \widetilde{P^*})}, \tag{2.27}$$

where $Q_b(\widetilde{P^*}, \widetilde{P^*})$ is independent of χ_q^2.

Standard asymptotic results are given for $\hat{W}$ based on conditions on M and $k(\cdot)$ that lead to consistency of $\hat{\omega}_{u\cdot v}$, as $T \to \infty$

$$\hat{W} \Rightarrow \chi_q^2. \tag{2.28}$$

The expression $Q_b(\widetilde{P^*}, \widetilde{P^*})$ is the fixed-b limit of the long-run covariance estimator of the form (2.7) using $\Delta\tilde{S}_t^{u*}$ instead of $\hat{\eta}_t$. Therefore, t- as well as Wald-type tests can be performed based on long-run covariance estimation with $\Delta\tilde{S}_t^{u*}$. Critical values can be tabulated depending on the specification of the deterministic components, the number of integrated regressors and its powers included, the kernel function and the bandwidth choice.[3]

2.2.4. An IM-OLS Residual Based Cointegration Test for CPRs

Lemma 2 provides the asymptotic limiting distribution of the scaled partial sum process of the adjusted IM-OLS residuals $\tilde{S}_t^{u*}$. The following result for the scaled partial sum process of the (non-adjusted) IM-OLS residuals follows straightforwardly replacing $h(r)$ by $g(r)$.

Corollary 4. *Suppose that full design prevails and the assumptions of Proposition 1 hold, then for $T \to \infty$*

$$\begin{aligned} T^{-1/2} \sum_{t=2}^{[rT]} \Delta\tilde{S}_t^u &\Rightarrow \omega_{u\cdot v}^{1/2} \left(\int_0^r dw_{u\cdot v}(s) - g(r)' \left(\int_0^1 g(s)g(s)'ds \right)^{-1} \int_0^1 [G(1) - G(s)] dw_{u\cdot v}(s) \right) \\ &=: \omega_{u\cdot v}^{1/2} \widetilde{P}(r). \end{aligned} \tag{2.29}$$

Note that $\widetilde{P}(r)$ consists of independent standard Brownian motions in case of full design.

The result given in Corollary 4 immediately suggests a Kwiatkowski *et al.* (1992)-type (KPSS-type) test with the null hypothesis of cointegration. Shin (1994) extends the KPSS test from a stationarity to a linear cointegration test and here we consider the corresponding extension to the CPR framework based on the IM-OLS residuals.

[3] Tables with fixed-b critical values for IM-OLS based tests in the CPR case for different specifications of deterministics (intercept, intercept and linear trend), up to four integrated regressors and the last integrated regressor entering with integer powers up to power four as well as for different kernel functions are available upon request.

Proposition 3. *Suppose that full design prevails and the assumptions of Proposition 1 hold. Then the limit of the IM-OLS residual based KPSS-type test statistic under the null hypothesis for $T \to \infty$ is given by*

$$CT_{IM} := \frac{1}{T^2 \hat{\omega}_{u \cdot v}} \sum_{t=2}^{T} \left(\sum_{i=2}^{t} \Delta \tilde{S}_i^u \right)^2 \Rightarrow \int_0^1 \left(\widetilde{P}(r) \right)^2 dr, \tag{2.30}$$

where $\hat{\omega}_{u \cdot v}$ denotes a consistent estimator of $\omega_{u \cdot v}$.

Since $\widetilde{P}(r)$ consists of independent standard Brownian motions, critical values for the CT_{IM} test statistic can be tabulated which depend upon the specification of the deterministic components, the number of integrated regressors and its powers included. Critical values are given in Table B.5 for up to four regressors, the integrated regressor entering with powers up to power four and three specifications of the deterministic components: (i) no deterministics, (ii) intercept only, and (iii) intercept and linear trend.

Remark 6. Following the discussion before Lemma 2 the estimator $\tilde{\omega}_{u \cdot v}$ based on the IM-OLS residuals $\tilde{S}_t^u$ is inconsistent for $\omega_{u \cdot v}$. Therefore, we use the OLS residuals from model (2.1) in order to obtain a consistent estimator $\hat{\omega}_{u \cdot v}$.

2.3. Simulation Study

In this section we assess the performance of the CPR extensions of the D-OLS estimator by Saikkonen and Choi (2004), the FM-OLS estimator by Wagner and Hong (2016) and the IM-OLS estimator introduced in Section 2.2 by means of a simulation study benchmarked against the OLS estimator. The estimators are labeled D-CPR, FM-CPR and IM-CPR, respectively. We evaluate the performance in terms of bias and root mean squared error (RMSE) as well as in terms of empirical null rejection probabilities and size-corrected power of tests based on these estimators. Data is generated according to

$$y_t = \delta_1 + \delta_2 t + \beta_1 x_t + \beta_2 x_t^2 + u_t \tag{2.31}$$

$$x_t = x_{t-1} + v_t, \tag{2.32}$$

with

$$u_t = \rho_1 u_{t-1} + e_{1,t} + \rho_2 e_{2,t}, \tag{2.33}$$

$$v_t = e_{2,t} + 0.5 e_{2,t-1}, \tag{2.34}$$

$\rho_1 = \rho_2$	OLS	D-CPR		FM-CPR		IM-CPR
			And91	NW	NW_T	
Bias						
T=100						
0.0	-0.001	-0.004	-0.002	-0.002	-0.002	-0.001
0.3	0.017	-0.005	0.004	0.002	0.003	-0.001
0.6	0.074	-0.003	0.038	0.036	0.035	0.013
0.9	0.362	0.164	0.305	0.302	0.308	0.243
T=200						
0.0	-0.000	0.000	-0.001	-0.001	-0.001	-0.000
0.3	0.009	0.001	0.001	0.001	0.001	0.000
0.6	0.040	0.001	0.015	0.015	0.016	0.004
0.9	0.227	0.073	0.166	0.168	0.188	0.107
RMSE						
T=100						
0.0	0.067	0.216	0.071	0.071	0.071	0.100
0.3	0.094	0.283	0.097	0.096	0.097	0.142
0.6	0.173	0.423	0.162	0.159	0.159	0.239
0.9	0.521	0.888	0.501	0.494	0.495	0.748
T=200						
0.0	0.033	0.060	0.034	0.034	0.033	0.049
0.3	0.047	0.083	0.047	0.047	0.047	0.070
0.6	0.092	0.133	0.081	0.080	0.080	0.121
0.9	0.340	0.370	0.302	0.300	0.311	0.451

Table 2.1.: Finite sample bias and RMSE for coefficient β_1, Bartlett kernel.

where $e_{1,t}, e_{2,t}$ are independent identically distributed standard normal random variables. The parameter values chosen are $\delta_1 = \delta_2 = 1, \beta_1 = 5, \beta_2 = -0.3$ motivated by estimation results for the environmental Kuznets curve (EKC) hypothesis with FM-CPR and D-CPR in Wagner (2015). The parameter ρ_1 controls serial correlation in the regression error u_t and the parameter ρ_2 controls the level of endogeneity of the regressor x_t. The values for the correlation parameters are chosen from the set $\{0.0, 0.3, 0.6, 0.9\}$, where we focus on the case $\rho_1 = \rho_2$. For long-run covariance estimation we choose the Bartlett and Qaudratic Spectral (QS) kernels with bandwidths being chosen according to the data dependent rules of Andrews (1991) and Newey and West (1994) as well as the sample size dependent Newey-West bandwidth, i.e., $\lfloor 4(T/100)^{2/9} \rfloor$, labeled NW_T. Furthermore, for the D-CPR estimator we use the Akaike information criterion (AIC) based lead and lag length choice of Choi and Kurozumi (2012). We consider 5,000 replications for the sample sizes $T \in \{100, 200, 500, 1000\}$.

Bias and RMSE

Let us briefly summarize the simulation results for bias and RMSE given in Table 2.1. We only report the results for the Bartlett kernel and the bandwidth according to the data dependent rule of Andrews (1991), since the results for the different kernels and bandwidths are quite similar. In case of no correlation none of the estimators for β_1 shows much bias. With increasing $\rho_1 = \rho_2$ the bias increases for all estimators, where D-CPR and IM-CPR estimators appear to be less sensitive to increasing level of correlation than FM-CPR and especially OLS. For increasing sample size T all estimators have reduced bias. With respect to RMSE, OLS and FM-CPR have the smallest root mean squared errors for small sample sizes. As already pointed out in Vogelsang and Wagner (2014a) for the linear cointegration case, the larger RMSE for IM-CPR is not surprising, because IM-CPR uses a regression with an $I(1)$ error S_t^u, whereas OLS and FM-CPR use an $I(0)$ error u_t. However, for larger sample sizes the difference between the estimators is decreasing.

Empirical Null Rejection Probabilities

We turn to the finite sample results for the hypothesis tests introduced in Section 2.2. We consider t-tests for the hypotheses $H_0 : \beta_1 = 5$ and $H_0 : \beta_2 = -0.3$ as well as Wald tests for the joint hypothesis $H_0 : \beta_1 = 5, \beta_2 = -0.3$. For standard asymptotic tests based on traditional bandwidth and kernel assumptions we provide results corresponding to the test statistic $\hat{W}$ and the chi-squared limiting distribution in (2.28). Rejections for these test statistics are carried out using $N(0,1)$ critical values for the t-tests and χ_2^2 critical values for the Wald test, respectively. The fixed-b tests for the IM-CPR estimator introduced in Proposition 2 are implemented in two ways: (i) consider a grid $b \in \{0.02, 0.04, \ldots, 1.00\}$ and choose bandwidth according to $M = bT$, (ii) compute a bandwidth M^* according to one of the data dependent rules and subsequently determine the largest multiple of 0.02 smaller or equal to $b^* := M^*/T$. The latter method is labeled Data-Dep below. Simulated critical values for fixed-b inference depend particularly on both kernel and bandwidth. The nominal level is 0.05 throughout.

Table 2.2 shows empirical null rejection probabilities for the t-tests for β_1 and Table 2.3 reports the results for the Wald tests. Furthermore, results for the t-tests for β_2, which are qualitatively similar, are given in Table B.7 in the appendix. The tables contain the following columns: OLS, D-CPR, FM-CPR and the test statistic using $\hat{\omega}_{u \cdot v}$ for standard asymptotic inference based on the IM-CPR estimator, labeled IM-CPR(O). The last three columns show the results for fixed-b inference using one of the data dependent bandwidth rules as well as using fixed values $b \in \{0.1, 0.2\}$. As expected OLS based tests show the best performance in case of no correlation, but have severe size distortions in case of positive correlation. D-CPR based tests are very size distorted even in the non-correlation case for $T = 100$, but improve with increasing sample size. The FM-CPR and IM-CPR(O) tests show a similar performance, where the latter has slightly lower over-rejections in case of increased level of correlation especially for the Wald-type test with multiple hypotheses. The IM-CPR based fixed-b tests behave properly compared to the standard asymptotic

$\rho_1 = \rho_2$	OLS	D-CPR	FM-CPR	IM-CPR(O)	IM-CPR(Fb)		
					Data-Dep	b=0.1	b=0.2
T=100							
0.0	0.059	0.164	0.077	0.100	0.049	0.053	0.069
0.3	0.154	0.206	0.110	0.109	0.068	0.060	0.077
0.6	0.371	0.270	0.170	0.137	0.127	0.077	0.092
0.9	0.725	0.425	0.411	0.356	0.452	0.300	0.249
T=200							
0.0	0.048	0.089	0.059	0.071	0.045	0.048	0.055
0.3	0.147	0.118	0.079	0.077	0.060	0.051	0.055
0.6	0.374	0.145	0.126	0.089	0.097	0.056	0.064
0.9	0.746	0.269	0.314	0.242	0.438	0.161	0.136
T=500							
0.0	0.054	0.067	0.057	0.066	0.054	0.050	0.057
0.3	0.156	0.082	0.070	0.070	0.055	0.050	0.057
0.6	0.375	0.091	0.086	0.079	0.064	0.050	0.058
0.9	0.762	0.142	0.223	0.117	0.191	0.067	0.068
T=1000							
0.0	0.053	0.062	0.055	0.061	0.055	0.052	0.054
0.3	0.163	0.072	0.064	0.066	0.055	0.051	0.055
0.6	0.400	0.079	0.081	0.070	0.057	0.052	0.055
0.9	0.787	0.114	0.176	0.087	0.097	0.055	0.063

Table 2.2.: Empirical null rejection probabilities for $H_0 : \beta_1 = 5$, Andrews (1991) bandwidth, QS kernel, 0.05 level.

tests. The empirical null rejection probabilities are close to the nominal size and large over-rejections occur only in high-correlation cases in conjunction with small sample sizes. Fixed-b tests outperform standard asymptotic tests throughout and are only moderately size distorted especially for $b \in \{0.1, 0.2\}$. The data dependent bandwidth rules for the fixed-b tests in the fifth column typically lead to b equal to 0.02 or 0.04. In order to illustrate the impact of the choice of b on the test performance, we plot empirical size rejections for different sample sizes and different correlation parameters as a function of b. The results are given in Figure 2.1, which shows that the tests for b smaller or equal to 0.04 have the highest rejection probabilities. Regarding bandwidth and kernel choice, typically the Andrews (1991) bandwidth choice leads to lower size distortions than the Newey and West (1994) bandwidth choice and the QS kernel dominates the Bartlett kernel in terms of empirical null rejection probabilities. The results for different kernel and bandwidth choice are available upon request.

Size-Corrected Power Analysis

We close this section on hypothesis testing considering size-corrected power properties of the tests. Although size-corrections are not feasible in practice, they are a useful tool for

$\rho_1 = \rho_2$	OLS	D-CPR	FM-CPR	IM-CPR(O)	IM-CPR(Fb)		
					Data-Dep	b=0.1	b=0.2
T=100							
0.0	0.057	0.212	0.086	0.117	0.047	0.058	0.069
0.3	0.200	0.282	0.143	0.133	0.071	0.061	0.082
0.6	0.526	0.382	0.242	0.180	0.164	0.082	0.099
0.9	0.922	0.612	0.601	0.554	0.626	0.419	0.311
T=200							
0.0	0.053	0.099	0.066	0.079	0.047	0.053	0.053
0.3	0.192	0.145	0.105	0.087	0.063	0.056	0.056
0.6	0.529	0.187	0.166	0.119	0.124	0.060	0.064
0.9	0.932	0.391	0.471	0.351	0.623	0.204	0.156
T=500							
0.0	0.046	0.065	0.054	0.060	0.047	0.052	0.059
0.3	0.195	0.089	0.075	0.069	0.051	0.051	0.056
0.6	0.558	0.108	0.104	0.082	0.063	0.054	0.059
0.9	0.940	0.208	0.338	0.167	0.283	0.075	0.076
T=1000							
0.0	0.053	0.065	0.057	0.063	0.058	0.052	0.053
0.3	0.214	0.076	0.073	0.067	0.059	0.053	0.057
0.6	0.563	0.090	0.091	0.076	0.063	0.054	0.056
0.9	0.949	0.149	0.250	0.107	0.120	0.059	0.066

Table 2.3.: Empirical null rejection probabilities for H_0 : $\beta_1 = 5$, $\beta_2 = -0.3$, Andrews (1991) bandwidth, QS kernel, 0.05 level.

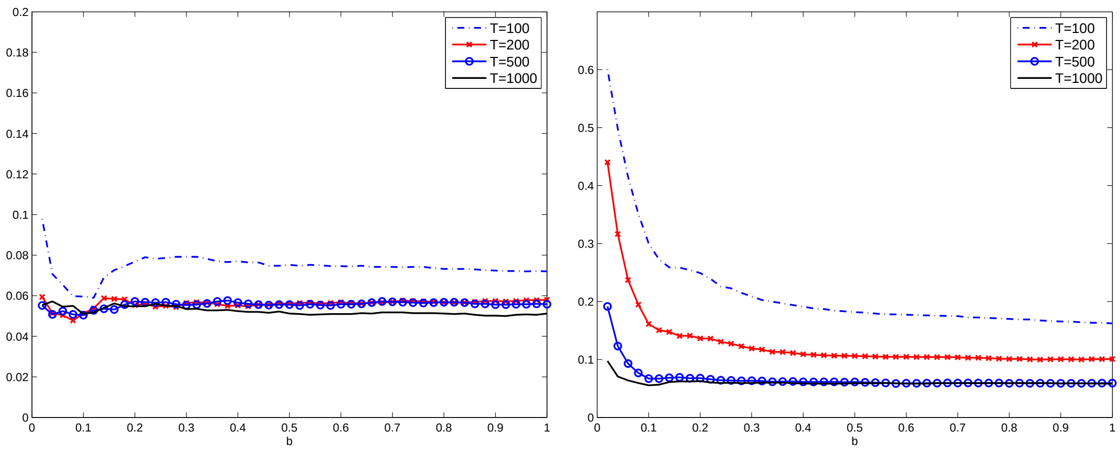

Figure 2.1.: Empirical null rejections, IM-CPR(Fb) inference: t-test for β_1, QS kernel, Andrews (1991) bandwidth, $\rho_1 = \rho_2 = 0.3$ (left panel), $\rho_1 = \rho_2 = 0.9$ (right panel).

theoretical comparisons since they overcome potential over-rejection problems under the null hypothesis. Therefore, we use empirical critical values in order to hold the empirical null rejection probabilities constant at 0.05 under the null. Starting from the true values of β_1 and β_2 we consider under the alternative $\beta_1 \in (5, 6]$ and $\beta_2 \in (-0.3, 0.2]$ with a total

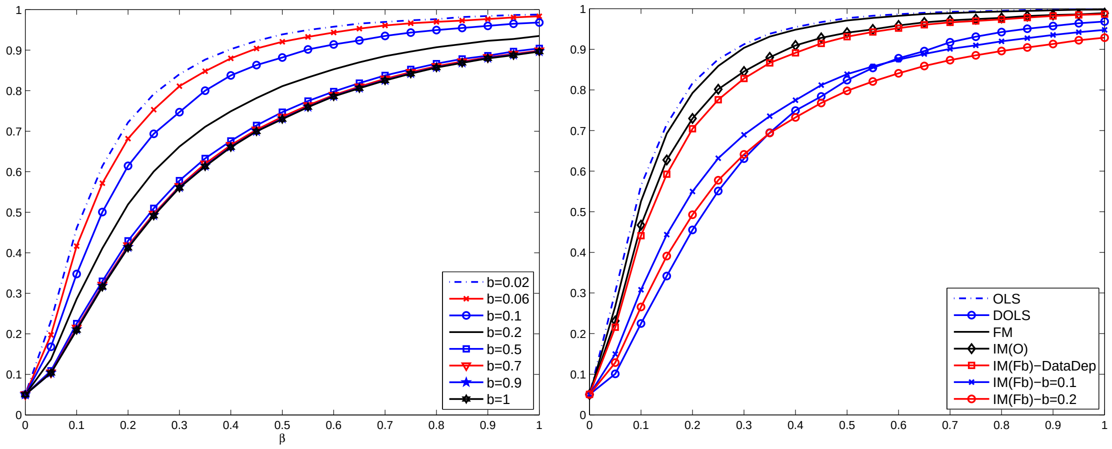

Figure 2.2.: Size-corrected Power, $Wald$-test, T=100, $\rho_1 = \rho_2 = 0.6$, QS kernel.

of 20 values generated on a grid with mesh 0.05 for β_1 and 0.005 for β_2. The figures for the t-tests and the Wald tests are qualitatively similar, thus we focus on the latter. The left panel in Figure 2.2 shows that increasing value of b leads to some power loss. However, this power loss is minor in most cases, whereas Figure 2.1 shows that empirical null-rejection tend to be lower with increasing b. The minimal power losses in size-corrected power seems to be the price to be paid for less size distortions. Comparing both kernels, we observe that the QS kernel is much more sensitive to the bandwidth choice than the Bartlett kernel. For increasing value of b size-corrected power becomes much lower using the QS kernel than the Bartlett kernel. Results for size-corrected power using the Bartlett kernel are given in Appendix B.4. As described above, tests using the QS kernel exhibit much fewer over-rejection problems under the null especially for larger bandwidths. This size-power trade-off for kernel and bandwidth choice has already been observed by Kiefer and Vogelsang (2005) as well as by Vogelsang and Wagner (2014a). The right panel in Figure 2.2 shows power comparisons for different tests, namely OLS, D-CPR, FM-CPR, IM-CPR(O) and IM-CPR(Fb) using the Andrews (1991) data dependent bandwidth rule. The D-CPR based test has the lowest power when the true parameter values are close to those under the null, but slightly higher power than the IM-CPR(Fb) tests when the difference between true parameter values and the values under null increases. Throughout OLS and FM-CPR based tests have the highest size-corrected power, whereas both IM-CPR based tests show small but non-trivial reduction in power.

Cointegration Testing

We also assess the performance of the IM-CPR residual based cointegration test CT_{IM} in terms of empirical null rejections probabilities and size-corrected power. We use the data

ρ	CT_{Shin}	CT_{FM}	CT_{IM}	CT_{Shin}	CT_{FM}	CT_{IM}
		And			NW	
T=100						
0.0	0.052	0.050	0.089	0.057	0.053	0.083
0.3	0.058	0.056	0.079	0.073	0.074	0.107
0.6	0.094	0.056	0.071	0.149	0.126	0.159
0.8	0.174	0.070	0.107	0.352	0.214	0.215
T=200						
0.0	0.049	0.047	0.061	0.051	0.051	0.065
0.3	0.056	0.051	0.063	0.062	0.067	0.083
0.6	0.094	0.053	0.065	0.128	0.103	0.117
0.8	0.212	0.042	0.055	0.326	0.152	0.146
T=500						
0.0	0.062	0.056	0.058	0.059	0.054	0.055
0.3	0.068	0.060	0.062	0.077	0.074	0.080
0.6	0.095	0.061	0.063	0.155	0.085	0.095
0.8	0.218	0.048	0.049	0.387	0.138	0.127
T=1000						
0.0	0.061	0.051	0.055	0.060	0.051	0.053
0.3	0.064	0.056	0.060	0.081	0.069	0.077
0.6	0.088	0.057	0.062	0.161	0.075	0.082
0.8	0.204	0.051	0.056	0.400	0.122	0.116

Table 2.4.: Empirical null rejection probabilities of cointegration tests, QS kernel, 0.05 level.

generating process (2.31) under the null and similar to Wagner and Hong (2016) three alternative DGPs:

$$
\begin{aligned}
&(\text{A}): \ y_t = 1 + t + 5x_t - 0.3x_t^2 + 0.01x_t^3 + u_t \\
&(\text{B}): \ y_t = 1 + t + 5x_t - 0.3x_t^2 + z_t, \text{ where } z_t \sim I(1) \text{ independent of } x_t \\
&(\text{C}): \ y_t, x_t \text{ are two independent } I(1) \text{ variables}
\end{aligned}
$$

These DGPs cover the main alternatives of interest, i.e., (A) missing higher order polynomials of the integrated regressor, (B) no cointegration because of a missing integrated regressor, and (C) spurious regression. We compare the performance of the IM-CPR residual based CT_{IM} test and the FM-CPR residual based cointegration test for CPRs proposed by Wagner and Hong (2015), labeled CT_{FM}. The results are benchmarked against the Shin (1994) cointegration test (CT_{Shin}) for linear cointegrating relationships, because linear cointegration techniques instead of CPR techniques are commonly used in, e.g., the EKC literature. Results for a D-CPR residual based cointegrationg test, which has the same limiting distribution under the null as the CT_{FM} test, are available upon request. Comparable to the observations made in hypothesis testing section the D-CPR residual based cointegration test performs worse in small samples, but similar to the CT_{FM} test for moderately large sample sizes.

Table 2.4 presents the simulation results for the empirical null rejection probabilities for the CT_{Shin}, CT_{FM} and CT_{IM} test, respectively. The CT_{Shin} test is based on the *standard* FM-OLS residuals, i.e., residuals obtained from cointegrating regression estimation techniques treating x_t and its powers as separate integrated regressors. Stypka *et al.* (2017) show that *standard* FM-OLS techniques applied to CPRs do not have an asymptotic effect for parameter estimation but for cointegration testing. Consequently, the test shows poor performance in case of high level of correlation. The CT_{FM} and CT_{IM} tests show good performance in the lower correlation case already for small sample sizes. The CT_{IM} test exhibits some over-rejections for $T = 100$, but for increasing sample size both tests are close to the nominal level also in high correlation cases. Throughout the CT_{FM} test has slightly lower over-rejections compared to the CT_{IM} test. The impact of the bandwidth choice on the cointegration tests' performance is more pronounced than it is the case for parameter tests. The data dependent bandwidth rule according to Andrews (1991) outperforms the Newey and West (1994) bandwidth rule for all considered tests, which has already been observed in Wagner and Hong (2016) for the CT_{FM} test. Result tables for the other specifications are available upon request. We observe that the performance of the considered tests is poorer when the number of integrated regressors and/or the number of powers included increases reflecting computational difficulties associated with estimating models with more parameters.

Table 2.5 reports size-corrected power of the cointegration tests against the three alternatives. The CT_{Shin} test has low power against the cubic CPR alternatives (A) especially in conjunction with the Newey and West (1994) bandwidth. The CT_{IM} test has higher size-corrected power compared to the CT_{FM} test, which has slightly lower over-rejections under the null. Against the cubic alternatives (A) size-corrected power decreases with increasing level of correlation. All considered tests show good performance in terms of size-corrected power against alternatives (B) and (C). As already pointed out in Wagner and Hong (2016) and Stypka *et al.* (2017), using the Andrews (1991) bandwidth rule leads to substantially smaller size distortions for cointegration testing under the null, but this comes at the cost of substantially reduced size-corrected power. In sum, the CT_{FM} and CT_{IM} tests perform fairly similar. The CT_{FM} test exhibits slightly lower over-rejections under the null compared to the CT_{IM} test, while the latter performs better in terms of size-corrected power against the three alternatives considered in this section. Both tests outperform the CT_{Shin} test for linear cointegrating relationship in this CPR setting.

	ρ	CT_{Shin}	CT_{FM}	CT_{IM}	CT_{Shin}	CT_{FM}	CT_{IM}
			And			NW	
T=100							
(A)	0.0	0.311	0.237	0.278	0.148	0.151	0.130
	0.3	0.294	0.223	0.292	0.105	0.110	0.094
	0.6	0.237	0.225	0.302	0.027	0.043	0.045
	0.8	0.187	0.209	0.238	0.001	0.014	0.030
(B)	-	0.325	0.372	0.359	0.653	0.522	0.405
(C)	-	0.387	0.323	0.310	0.616	0.513	0.396
T=200							
(A)	0.0	0.332	0.253	0.314	0.290	0.267	0.242
	0.3	0.313	0.239	0.301	0.242	0.214	0.193
	0.6	0.240	0.234	0.311	0.106	0.148	0.120
	0.8	0.161	0.263	0.335	0.006	0.095	0.109
(B)	-	0.390	0.377	0.381	0.836	0.685	0.575
(C)	-	0.406	0.364	0.356	0.822	0.693	0.594
T=500							
(A)	0.0	0.310	0.234	0.324	0.599	0.576	0.597
	0.3	0.293	0.218	0.302	0.539	0.497	0.527
	0.6	0.244	0.219	0.302	0.367	0.466	0.500
	0.8	0.163	0.255	0.351	0.085	0.389	0.427
(B)	-	0.403	0.387	0.425	0.984	0.933	0.919
(C)	-	0.423	0.326	0.369	0.984	0.932	0.922
T=1000							
(A)	0.0	0.317	0.241	0.331	0.828	0.820	0.848
	0.3	0.306	0.228	0.308	0.782	0.772	0.804
	0.6	0.263	0.225	0.293	0.619	0.760	0.793
	0.8	0.177	0.241	0.325	0.270	0.677	0.736
(B)	-	0.421	0.388	0.425	1.000	0.994	0.993
(C)	-	0.432	0.340	0.373	0.999	0.990	0.992

Table 2.5.: Size-corrected power of cointegration tests, QS kernel, 0.05 level.

2.4. Application: EKC Analysis

For the empirical analysis of the environmental Kuznets curve (EKC) hypothesis we consider annual data for 19 early industrialized countries over the time period 1870-2013 for carbon dioxide (CO_2) emissions and real GDP. All of these quantities are used in per capita terms and transformed to logarithms.

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

Table 2.6.: The sample range is 1870-2013 for GDP and CO_2 with the exception of New Zealand which has 1878 as its starting point.

The CO_2 emissions data is from the homepage of the Carbon Dioxide Information Analysis Center of the US Department of Energy[4], the GDP data was downloaded from the homepage of the Maddison Project[5] and from The Conference Board Total Economy Database[6]. The required long-run covariance estimates for the EKC estimation are based on the Bartlett kernel and the data dependent bandwidth rule of Newey and West (1994) similar to the analysis in Wagner (2015) and in Wagner and Grabarczyk (2017). We consider the quadratic formulation

$$\begin{aligned} e_t &= c + \delta t + \beta_1 y_t + \beta_2 y_t^2 + u_t, \qquad (2.35) \\ y_t &= y_{t-1} + v_t, \end{aligned}$$

as well as the cubic formulation

$$\begin{aligned} e_t &= c + \delta t + \beta_1 y_t + \beta_2 y_t^2 + \beta_3 y_t^3 + u_t, \qquad (2.36) \\ y_t &= y_{t-1} + v_t, \end{aligned}$$

where e_t denotes log per capita CO_2 emissions and y_t denotes log per capita GDP.

Prior to estimation, we test the unit root hypothesis for the variable on the right-hand side, i.e. log per capita GDP, investigating the Phillips and Perron (1988) t-test and the fixed-b Phillips-Perron unit root test introduced by Vogelsang and Wagner (2013) for the specification with an intercept and a linear trend. The results are reported in Table B.1.

[4] http://cdiac.ornl.gov
[5] http://www.ggdc.net/maddison/maddison-project/home.htm
[6] http://www.conference-board.org/data/economydatabase

The unit root null hypothesis based on the standard Phillips-Perron test is rejected for none of the countries. The PP(fb) unit root test rejects the null hypothesis for log GDP per capita for Canada and the USA only at the 10% level.

We also carry out CPR based (non-)cointegration tests for the specifications (2.35) and (2.36). In addition to the IM-OLS residual based cointegration test CT_{IM} from Section 2.2, we employ the CPR extension of the FM-OLS residual based Shin (1994) cointegration test (CT_{FM}) and the OLS residual based extension of the Phillips and Ouliaris (1990) non-cointegration test ($P_{\hat{u}}$) introduced in Wagner (2013). The $P_{\hat{u}}$ test is based on the assumption that the dependent variable, i.e. e_t, is an integrated process under the null. Thus, the null corresponds to the spurious regression alternative (C) in Section 2.3. Since the simulation study has revealed that the CT tests do not have impressive power against the alternative of missing higher order polynomials, it appears prudent to only take those countries into account for the EKC analysis, where the results between the $P_{\hat{u}}$ test and the CT tests are not contradictory, i.e. rejection of the former and non-rejection of the latter. We briefly summarize the results given in Table B.2. Based on the $P_{\hat{u}}$ test the null hypothesis of non-cointegration for both specifications is rejected for Austria, Belgium, Finland, Germany, the Netherlands, Switzerland and the UK. Considering these countries, the results for the CT_{IM} and the CT_{FM} tests are well in line with rejection at the 0.05 level only occurring for Germany in the quadratic specification. Conflicting results between both cointegration tests only appear for the UK in the cubic specification, where the null of cointegration is rejected for the CT_{IM}, but not rejected for the CT_{FM} test. Summarizing the results of the cointegration tests, we consider the following countries for the CPR based estimation of the EKC relationship: Austria, Belgium, Finland, the Netherlands, Switzerland and the UK. Furthermore, for Germany we consider the cubic specification only.

We briefly turn to the estimation results for the specifications (2.35) and (2.36), where we include the estimators considered in Section 2.3, i.e. OLS, D-CPR, FM-CPR and IM-CPR. For significance tests based on the IM-CPR estimator we include standard t-values as well as t-values obtained from fixed-b inference. The results for the quadratic specification in Table B.3 show that the coefficient to squared GDP is significant and has the expected negative sign indicating an inverted U-shape for all countries. Wagner (2015) analyzes the EKC hypothesis for a similar data set including D-CPR and FM-CPR among others and finds relatively small differences across the results of both estimation methods. Consequently, it is expected that the CPR based estimators in this analysis also lead to similar results. Strongly different coefficient estimates, and consequently different turning points, across the methods occur only for Austria and Switzerland. The results for the cubic specification in Table B.4 indicate that the coefficient β_3 is not significantly different from zero at the 0.05 level for Austria, the Netherlands, Switzerland and the UK. For Belgium and Finland, all coefficients are not significant based on the D-CPR estimator, where entirely different estimation results are obtained compared to FM-CPR and IM-CPR. The simulation study in Section 2.3 reveals that D-CPR can perform poorly for such sample sizes considered in this empirical analysis. Figures B.1-B.4 show actual and fitted values as well as estimated EKCs using the coefficients estimated by IM-CPR from models (2.35) and (2.36). The fits are very good for all considered countries especially after the Second World War. With the exception of some time periods for Austria and the UK, the fits are also good for the time before and between the two world wars. Comparing the fits for the quadratic and cubic specifications directly, we find minor differences for the majority of the countries. Merely for the most recent decades for Austria the fitted values obtained for the cubic specification are closer to the actual values than those from the quadratic specification. In order to estimate the EKCs we use for the explanatory variable $T = 144$ equidistant values ranging from the minimal value of log per capita GDP up to the maximal value. For the linear time trend t we use values $1, \ldots, 144$ and insert these values together with the coefficient estimates. Focusing on the estimated EKCs of Belgium and Finland, where the coefficients to the third power of GDP are significantly different from zero, we find an inverted U-shaped EKC for the former in both specifications. In case of Finland the estimated EKC does not seem to describe the income-emissions relationship adequately for the cubic specification, see Figure B.4. For Germany a cointegrating polynomial relationship is supported only for the cubic specification. However, the estimated EKC has also an inverted U-shape rather than an N-shape. Here we observe huge difference between the FM-CPR and IM-CPR estimates, where the coefficients are not significant throughout based on the FM-CPR estimator. In sum, we find that the quadratic specification appears to be sufficient in describing the income-emissions relationship especially for Belgium, Finland, the Netherlands and the UK, where we find inverted U-shaped EKCs. As expected, the estimation results do not differ strongly for the FM-CPR and IM-CPR estimators for most of the countries. Furthermore, the results of the FM-CPR and IM-CPR based cointegration tests are well in line. With the exception of the UK in the cubic specification, both tests identify the same countries in which a cointegrating polynomial relationship between income and emissions is present.

2.5. Summary and Conclusions

This chapter considers the extension of the integrated modified OLS estimator from linear cointegrating regressions to cointegrating polynomial regressions. The zero mean Gaussian mixture distribution of the obtained estimator forms the basis for standard asymptotic inference. For the case of full design, we additionally perform fixed-b asymptotic inference. Full design prevails, e.g., when only one integrated regressor enters the regression equation with powers larger than one. This is the case in, e.g., the EKC analysis. The chapter also presents an IM-OLS residual based cointegration test, which has a nuisance parameter free limiting distribution in case of full design.

The theoretical results are complemented by a small simulation study to compare the IM-CPR estimator with OLS, FM-CPR and D-CPR. We find that the IM-CPR estimator has a slightly lower bias relative to FM-CPR and D-CPR, but marginally larger RMSE. In terms of empirical null rejection probabilities, hypothesis tests based on the IM-CPR estimator outperform FM-CPR and D-CPR based tests, especially the fixed-b version for small sample sizes and a high level of correlation. This comes at the cost of minor loss in size-corrected power.

We apply the developed methods for the estimation of the EKC using a data set of GDP and CO_2 emissions for 19 early industrialized countries over the period 1870–2013. We find evidence for the existence of a quadratic EKC for six countries and one additional country for a cubic EKC. The results of the FM-CPR and IM-CPR based cointegration tests are well in line with each other. The coefficient estimates are similar across the considered methods for most of the countries.

Future research will move in the following directions: First, in respect of the EKC analysis also integrated modified OLS estimators for multi-equation systems of CPRs are worth considering. This includes CPR extensions of seemingly unrelated regression (SUR) models (Zellner, 1962) or panel data models. Second, the choice of an optimal b value is an interesting but non-trivial problem. In this chapter the fixed-b values are chosen according to one of the data dependent bandwidth rules designed for long-run covariance estimation or set to a specific value. Third, the developed methods can also be applied to other economic questions such as the intensity-of-use debate, which postulates an inverted U-shaped relationship between GDP and intensity of metal use (Labson and Crompton, 1993).

3. The EKC for CO_2 Emissions: A Seemingly Unrelated Cointegrating Polynomial Regressions Approach

3.1. Introduction

The environmental Kuznets curve (EKC) hypothesis postulates an inverted U-shaped relationship between measures of economic development, typically GDP per capita, and measures of per capita 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) in his 1954 presidential address to the American Economic Association.

Starting with the pioneering work of Grossman and Krueger (1991, 1993, 1995) and Shafik and Bandyopadhyay (1992) a large and still growing body of research, both theoretical and empirical, is devoted to the EKC hypothesis. Theoretical contributions include Andreoni and Levinson (2001), Arrow *et al.* (1995), Brock and Taylor (2005, 2010), Cropper and Griffiths (1994), Dinda (2005), Jones and Manuelli (2001), Selden and Song (1995) or Stokey (1998).[1] Müller-Fürstenberger and Wagner (2007) discuss problems that arise at the intersection of theoretical and empirical EKC analysis. Additional early empirical contributions on top of the mentioned seminal papers include Agras and Chapman (1999), Antweiler *et al.* (2001), Hilton and Levinson (1998), Holtz-Eakin and Selden (1995),[2] Kahn (1998), List and Gallet (1999) or Torras and Boyce (1998).

Criticism of the EKC is as old as the EKC itself, both on theoretical as well as on econometric grounds. In this chapter we focus on discussing two problems related to (i) using

[1] A relatively recent survey of economic models for analyzing the EKC is given by Kijima *et al.* (2010). Uchiyama (2016, Chapter 2) contains a detailed discussion of the model of Stokey (1998) as well as an overview discussion of empirical work on the EKC. Already early survey papers like Stern (2004) or Yandle *et al.* (2004) find more than 100 refereed publications; and many more have been written since then.

[2] The quadratic formulation, i.e., the functional form that can literally lead to an inverted U-shape has first been used in this paper, whereas Grossman and Krueger used a third order polynomial.

unit root and cointegration methods for (ii) multi-country (or multi-regional) data in a parametric approach to the EKC. The problems addressed also impact – if unit root nonstationary behavior of explanatory variables is indeed present – the validity of other estimation approaches to the EKC, including nonparametric approaches (see, e.g., Millimet *et al.*, 2003), semiparametric approaches (see, e.g., Bertinelli and Strobl, 2005) or specifications based on spline interpolations (see, e.g., Schmalensee *et al.*, 1998).

Given that a significant part of the empirical literature uses unit root and cointegration techniques, understanding the implications of (i) and (ii) is important for empirical practice. Papers that use time series unit root and cointegration methods for EKC analysis include Esteve and Tamarit (2012), Fosten *et al.* (2012), Friedl and Getzner (2003), He and Richard (2010), Jalil and Mahmud (2009) and Lindmark (2002). Panel data EKC studies using unit root and cointegration techniques include Apergis (2016), Auffhammer and Carson (2008), Baek (2015), Bernard *et al.* (2015), Dijkgraaf and Vollebergh (2005), Dinda and Coondoo (2006), Galeotti *et al.* (2006), Perman and Stern (2003) or Romero-Avila (2008). As pointed out by Wagner (2015), based on Wagner and Hong (2016), these papers ignore the fact that powers of integrated processes are not themselves integrated processes (see also Wagner, 2012). Therefore, a regression of (the logarithm of) emissions per capita on (the logarithm of) GDP per capita and its powers is not a *standard* cointegrating regression, but in the terminology of Wagner and Hong (2016, Eq. (1)) a *cointegrating polynomial regression* (CPR); if this specific form of nonlinear cointegration prevails and the regression is not spurious.[3]

In the presence of powers of integrated regressors in cointegrating regressions, estimators like the fully modified OLS (FM-OLS) estimator (introduced for the linear cointegration case in Phillips and Hansen, 1990) can be adapted by using *appropriately constructed* additive correction terms. The precise form of these correction terms depends upon the specification of the relationship. They differ from the correction terms in the linear case, see Wagner and Hong (2016).[4] The implications of this difference for EKC analysis for time series data are illustrated in Wagner (2015). The asymptotic behavior of using standard FM-OLS treating unit root processes and their powers all as unit root processes is discussed in Stypka *et al.* (2017).[5]

[3] Prior to the estimation of these relationships, testing for nonlinear cointegration in EKC-type relationships need to be performed, see, e.g., Choi and Saikkonen (2010), Wagner (2013) or Wagner and Hong (2016).

[4] Important earlier work in this respect has been undertaken by Park and Phillips (1999, 2001), Chang *et al.* (2001) or Ibragimov and Phillips (2008). The difference between the work of Wagner and Hong (2016) and, e.g., Chang *et al.* (2001) is that the latter assume that the regressors are pre-determined and the errors serially uncorrelated. Wagner and Hong (2016) remove these two assumptions and consider the "standard" setting in cointegration analysis with endogenous regressors and serially correlated errors.

[5] In the example of a quadratic EKC this means that log GDP per capita and its square are treated as two integrated regressors and standard FM-OLS is performed in the two regressor case. The above-listed

The part of the empirical EKC literature that uses panel unit root and cointegration techniques relies entirely upon methods for *linear* cointegration developed for cross-sectionally independent panels. Thus, a fortiori the above-mentioned problems continue to be present. Importantly, additionally the assumption of cross-sectional independence that is employed in these studies, utilizing standard panel cointegration techniques like Kao and Chiang (2000), Phillips and Moon (1999) or Pedroni (2000), is clearly often unrealistic.[6] Also, the tacit assumption of these studies that all coefficients (except for, usually, the intercepts) are indeed identical, i.e., can be *pooled*, for all cross-section members may be too restrictive in many applications. In case that the cross-sectional dimension is small (compared to the time series dimension) a *seemingly unrelated regressions* (SUR) approach allows to relax both the cross-sectional independence as well as the poolability assumption. Based on Hong and Wagner (2014) we present in Section 3.2 fully modified OLS SUR estimators for systems of *seemingly unrelated cointegrating polynomial regressions* (SUCPR) formulated here for the quadratic EKC specification as used in the application.[7] In the SUCPR setting we allow for cross-sectional dependence of both the regressors and the errors and do not impose any poolability assumptions on the coefficients. Instead of having to impose poolability of the coefficients, we can test for any form of pooling and then, if the corresponding restrictions are not rejected, estimate the parameters pooled correspondingly. Some basic forms of pooling related to panel analysis are reviewed and stated in Appendix C.1: (P) all coefficients but the intercepts are pooled, (S) only the coefficients corresponding to log GDP per capita and its square are pooled, and (T) only the coefficient corresponding to the linear time trend is pooled. More generally, however, it may be the case that only some coefficients can be pooled over (potentially) different subsets of cross-section members. This turns out to be the case in the application in Section 3.3. Therefore we discuss estimation in *group-wise* pooled settings of a form relevant for our application in detail in Section 3.2.2.

The application of our methodology to study the EKC for CO_2 emissions for six early industrialized countries over the period 1870–2013 highlights the usefulness of the SUCPR approach. Group-wise pooled estimation of the EKC leads to almost the same results (estimated parameters, turning points, and fitted values) as those obtained with unrestricted individual or SUCPR estimation. This happens despite the reduction of the number of

papers employing cointegration methods all use cointegration techniques this way, as also discussed in Wagner (2015).

[6] Apergis (2016) and Romero-Avila (2008) acknowledge the potential of cross-sectional dependencies in time series panels by considering some form of cross-sectional dependence testing. That alone, however, does not solve the associated problems.

[7] In terms of econometric methodology Hong and Wagner (2014) discuss an extension of SUR cointegration analysis from the linear cointegration SUR case (see, e.g., Park and Ogaki, 1991; Mark *et al.*, 2005; Moon, 1999; Moon and Perron, 2005) to the SUCPR case. This is similar in scope – now for the SUR case – to the extension of FM-OLS from the linear cointegration to the CPR case presented in Wagner and Hong (2016).

parameters to be estimated by about one third. Fully pooled estimation, rejected by poolability testing, on the other hand, performs drastically worse. This shows that a situation- or problem-specific approach to pooling that our methodology provides is a helpful addition to the EKC analysis toolkit. The flexibility of the approach will allow for fruitful applicability also when modeling other relationships for data sets with a small cross-sectional dimension compared to a large time series dimension.

The chapter is organized as follows: In Section 3.2 we present the econometric methodology, i.e., two fully modified least squares estimators for systems of seemingly unrelated cointegrating polynomial regressions including a discussion of group-wise pooling – both with respect to testing for poolability as well estimation imposing the corresponding pooling restrictions – of a form relevant for our application. Section 3.3 presents and discusses the empirical findings and Section 3.4 briefly summarizes and concludes. Appendix C.1 is divided in two subsections. The first contains some additional material and results concerning the three variants (P), (S) and (T) of pooled estimation and the second provides the derivation of the limiting distributions of the group-wise pooled estimators. Appendix C.2 contains additional empirical results.

We use the following notation: $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$ and $\text{diag}(\cdot)$ denotes a diagonal matrix with entries specified throughout. For a vector $x = (x_i)_{i=1,\ldots,n}$ we denote by $\|x\|^2 = \sum_{i=1}^n x_i^2$ and for a matrix M we denote by $\|M\| = \sup_x \frac{\|Mx\|}{\|x\|}$. For a square matrix A we denote its determinant by $|A|$. We denote the m-dimensional identity matrix by I_m, with $0_{m\times n}$ a $(m \times n)$-matrix with all entries equal to zero, with $\mathbf{1}_s = [1, \ldots, 1]' \in \mathbb{R}^s$ and with $e_{i,n}$ the i-th unit vector in $\mathbb{R}^n$. For (block-)matrices M we denote the (i,j)-(block-)element with $M^{i,j}$, the i-th (block-)row with $M^{i,\cdot}$ and the j-th (block-)column with $M^{\cdot,j}$. With $\mathbb{1}_{\{\cdot\}}$ we denote the indicator function. Furthermore, $\otimes$ denotes the Kronecker product, $\mathbb{E}(\cdot)$ denotes the expected value and L denotes the backward-shift operator, i.e., $L\{z_t\}_{t\in\mathbb{Z}} = \{z_{t-1}\}_{t\in\mathbb{Z}}$. Definitional equality is signified by $:=$ and $\Rightarrow$ denotes weak convergence. Brownian motions are denoted $B(r)$ or short-hand by B, with covariance matrices specified in the context. For integrals of the form $\int_0^1 B(s)ds$ or $\int_0^1 B(s)dB(s)$, we often use the short-hand notation $\int B$ or $\int BdB$ and drop function arguments and integration bounds for notational simplicity.

3.2. Seemingly Unrelated Cointegrating Polynomial Regressions

For the discussion in this chapter it suffices to consider the special case of a system of seemingly unrelated quadratic polynomial regressions, where in the application in the following section $y_{i,t}$ denotes log CO_2 emissions per capita and $x_{i,t}$ log GDP per capita in country i in year t:

$$\begin{aligned} y_{i,t} &= c_i + \delta_i t + \beta_{1,i} x_{i,t} + \beta_{2,i} x_{i,t}^2 + u_{i,t}, \quad i = 1, \dots, N, \ t = 1, \dots, T, \qquad (3.1) \\ &= [D_{i,t}', X_{i,t}']\theta_i + u_{i,t}, \\ &= Z_{i,t}'\theta_i + u_{i,t}, \\ x_{i,t} &= x_{i,t-1} + v_{i,t}, \end{aligned}$$

with $Z_{i,t} := [D_{i,t}', X_{i,t}']'$, where $D_{i,t} := [1, t]'$ and $X_{i,t} := [x_{i,t}, x_{i,t}^2]'$, and $\theta_i := [c_i, \delta_i, \beta_{1,i}, \beta_{2,i}]'$. Denoting with $x_t := [x_{1,t}, \dots, x_{N,t}]'$, with $u_t := [u_{1,t}, \dots, u_{N,t}]'$ and with $v_t := [v_{1,t}, \dots, v_{N,t}]'$, we assume for $\xi_t := [u_t', v_t']'$ that

$$\begin{aligned} u_t &:= \Psi(L)\zeta_t = \sum_{j=0}^{\infty} \Psi_j \zeta_{t-j}, \qquad (3.2) \\ \Delta x_t = v_t &:= \Phi(L)\epsilon_t = \sum_{j=0}^{\infty} \Phi_j \epsilon_{t-j}, \end{aligned}$$

with $\sum_{j=0}^{\infty} j\|\Phi_j\| < \infty$ and $\sum_{j=0}^{\infty} j\|\Psi_j\| < \infty$. Furthermore, we assume $|\Phi(1)| \neq 0$, which excludes cointegration in the I(1) vector process $\{x_t\}$, and $|\Psi(1)| \neq 0$, since we need regularity of this matrix for the construction of the *modified SUR* estimator, a term coined by Park and Ogaki (1991) in the linear SUR cointegration setting. The stacked process $\{\xi_t^0\}_{t\in\mathbb{Z}} := \{[\zeta_t', \epsilon_t']'\}_{t\in\mathbb{Z}}$ is assumed to be a strictly stationary and ergodic martingale difference sequence with respect to the natural filtration $\mathcal{F}_t$ with positive definite conditional variance matrix $\Sigma := \mathbb{E}\left(\xi_t^0(\xi_t^0)'|\mathcal{F}_{t-1}\right)$ and $\sup_{t\geq 1} \mathbb{E}(\|\xi_t^0\|^r|\mathcal{F}_{t-1}) < \infty$ a.s. for some $r > 4$.

Remark 7. The above setting in (3.1) can be generalized in several ways: First, several integrated regressors and their powers can be included, with the specifications allowed to be equation specific. In the above example this means that different powers can be included in the different equations. Second, more general (equation-specific) deterministic components can be included. Third, pre-determined (or even more easily strictly exogenous) stationary regressors can be included as well. Fourth, common non-cointegrated nonstationary regressors can also be included in the equation system, which may be of particular relevance in, e.g., regional applications where country-wide variables may be important determinants for all regions. For more details in these respects see Hong and Wagner (2014).

The above assumptions are sufficient for a functional central limit theorem to hold, i.e.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\xi_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\begin{bmatrix}u_t\\ v_t\end{bmatrix} \Rightarrow B(r) = \begin{bmatrix}B_u(r)\\ B_v(r)\end{bmatrix} := \Omega^{1/2}W(r), \;\; 0 \leq r \leq 1, \tag{3.3}$$

with $W(r)$ a $2N$-dimensional standard Wiener process and $\Omega := \sum_{h=-\infty}^{\infty}\mathbb{E}(\xi_0\xi_h')$ the so-called long run variance of $\{\xi_t\}_{t\in\mathbb{Z}}$. For later usage we define also the one-sided long run variance given by $\Delta := \sum_{h=0}^{\infty}\mathbb{E}(\xi_0\xi_h')$ and both matrices are partitioned according to the partitioning of ξ_t:

$$\Omega := \begin{bmatrix}\Omega_{uu} & \Omega_{uv}\\ \Omega_{vu} & \Omega_{vv}\end{bmatrix}, \;\Delta := \begin{bmatrix}\Delta_{uu} & \Delta_{uv}\\ \Delta_{vu} & \Delta_{vv}\end{bmatrix}. \tag{3.4}$$

The above set of N equations (3.1) can be jointly written as

$$y_t = Z_t'\theta + u_t, \;\; t = 1, \ldots, T \tag{3.5}$$

with

$$y_t := \begin{bmatrix}y_{1,t}\\ \vdots\\ y_{N,t}\end{bmatrix} \in \mathbb{R}^N, \; Z_t := \begin{bmatrix}Z_{1,t} & & \\ & \ddots & \\ & & Z_{N,t}\end{bmatrix} \in \mathbb{R}^{4N\times N}, \; u_t := \begin{bmatrix}u_{1,t}\\ \vdots\\ u_{N,t}\end{bmatrix} \in \mathbb{R}^N,$$

and with $\theta := [\theta_1', \ldots, \theta_N']'$. Stacking all T observations for the above equation (3.5) we arrive at

$$y = Z\theta + u, \tag{3.6}$$

with

$$y := \begin{bmatrix}y_1\\ \vdots\\ y_T\end{bmatrix} \in \mathbb{R}^{NT}, \;\; Z := \begin{bmatrix}Z_1'\\ \vdots\\ Z_T'\end{bmatrix} \in \mathbb{R}^{NT\times 4N}.$$

A few basic observations concerning parameter estimation in (3.6) can already be made: First, it is straightforward to show that the OLS estimator of θ in (3.6) is consistent with a limiting distribution contaminated by second order bias terms, just as in the linear seemingly unrelated cointegration case studied in Park and Ogaki (1991) or Moon (1999). Alternatively, the results for the OLS estimator given in Wagner and Hong (2016) for the single equation case, of course, generalize to the SUCPR case. Second, in the classical SUR approach of Zellner (1962) the errors are typically assumed to be serially uncorrelated (and the regressors nonstochastic). Correspondingly, the weighting matrix used in "classical" SUR estimation, i.e., in GLS estimation, is an estimate of the contemporaneous error variance matrix. In the cointegration setting we allow for both error serial correlation and endogenous regressors. To take these two generalizations into account, Park and Ogaki (1991) define a *modified SUR* (MSUR) estimator using an estimate of the long run variance matrix of the errors as weighting matrix. The asymptotic behavior of the OLS and MSUR estimators is derived in Hong and Wagner (2014, Proposition 1) for the SUCPR case. The nuisance parameter dependent limiting distributions of these two estimators provide guidance for the construction of appropriate two-part FM-type corrections.[8] One of the corrections is as in the linear case, i.e., the dependent variable y_t is replaced by $y_t^+ := y_t - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t$, with consistent estimators of the long run variances.[9] The second transformation consists of subtracting an appropriately constructed correction term. In the SUR setting we need two sets of correction terms, depending upon estimator considered as starting point (OLS or MSUR). For our specification (3.1) these are given by $A^* := [A_1^{*\prime}, \ldots, A_N^{*\prime}]'$ and $\tilde{A}^* := [\tilde{A}_1^{*\prime}, \ldots, \tilde{A}_N^{*\prime}]'$, with

$$A_i^* := (\hat{\Delta}_{vu}^+)^{i,i} \begin{bmatrix} 0_{2\times 1} \\ T \\ 2\sum_{t=1}^{T} x_{i,t} \end{bmatrix}, \quad \tilde{A}_i^* := (\hat{\Delta}_{vu}^+)^{i,\cdot}(\hat{\Omega}_{u.v}^{-1})^{\cdot,i} \begin{bmatrix} 0_{2\times 1} \\ T \\ 2\sum_{t=1}^{T} x_{i,t} \end{bmatrix}, \tag{3.7}$$

where $(\hat{\Delta}_{vu}^+)^{i,i}$ is a consistent estimator of $(\Delta_{vu}^+)^{i,i} := \Delta_{vu}^{i,i} - \Delta_{vv}^{i,\cdot}\Omega_{vv}^{-1}\Omega_{vu}^{\cdot,i}$ and $\hat{\Omega}_{u.v}$ is a consistent estimator of $\Omega_{u.v} := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$.

In order to finally define the two fully modified estimators and to state their asymptotic distributions we still need some additional quantities. We define, again for our special case, the weighting matrix $G = G(T) := I_N \otimes G_\bullet(T)$, with $G_\bullet(T) := \mathrm{diag}(T^{-1/2}, T^{-3/2}, T^{-1}, T^{-3/2})$ and a stochastic process $J(r) := \mathrm{diag}\,(J_1(r), \ldots, J_N(r))$ with $J_i(r) := [1, r, B_{v_i}(r), B_{v_i}^2(r)]'$, where $B_{v_i}(r)$ denotes the i-th coordinate of $B_v(r)$.

Proposition 4 (Hong and Wagner 2014, Proposition 2)**.** *Let y_t be generated by* (3.1) *with the assumptions given in place. Assume furthermore that, based on the OLS residuals, all required long run variances are estimated consistently. Using the correction factors*

[8]For completeness, the OLS estimator is (as always) given by $\hat{\theta}_{\mathrm{OLS}} := (Z'Z)^{-1} Z'y$ and the MSUR estimator is defined as $\tilde{\theta}_{\mathrm{MSUR}} := \left(Z'\left(I_T \otimes \hat{\Omega}_{uu}^{-1}\right) Z\right)^{-1} \left(Z'\left(I_T \otimes \hat{\Omega}_{uu}^{-1}\right) y\right)$. A more detailed discussion concerning possibilities to construct FM-type estimators in the SUR case is given in Hong and Wagner (2014) and Moon (1999).

[9]The results of, e.g., Jansson (2002) apply in our setting and provide conditions on kernels and bandwidths that allow for consistent long run variance estimation. Throughout the chapter we assume these conditions on bandwidth and kernel to be in place.

defined in (3.7) *the fully modified systems OLS (FM-SOLS) and the fully modified SUR (FM-SUR) estimators are given by:*

$$\hat{\theta} := \left(Z'Z\right)^{-1}\left(Z'y^{+} - A^{*}\right), \tag{3.8}$$

$$\tilde{\theta} := \left(Z'\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)Z\right)^{-1}\left(Z'\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)y^{+} - \tilde{A}^{*}\right), \tag{3.9}$$

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

$$G^{-1}\left(\hat{\theta} - \theta\right) \Rightarrow \left(\int JJ'\right)^{-1}\int J dB_{u.v}, \tag{3.10}$$

$$G^{-1}\left(\tilde{\theta} - \theta\right) \Rightarrow \left(\int J\Omega_{u.v}^{-1}J'\right)^{-1}\int J\Omega_{u.v}^{-1} dB_{u.v}, \tag{3.11}$$

where $B_{u.v}(r) := B_u(r) - \Omega_{uv}\Omega_{vv}^{-1}B_v(r)$ *is a Brownian motion with variance matrix* $\Omega_{u.v}$.

By construction $B_{u.v}(r)$ is independent of $B_v(r)$ and consequently the above zero mean Gaussian mixture limiting distributions given in (3.10) and (3.11) form the basis for asymptotic chi-squared inference using, e.g., Wald-type tests. Because the vectors $\hat{\theta}$ and $\tilde{\theta}$ contain elements that converge at different rates, obtaining formal results for the Wald-type test statistics requires a condition on the restriction matrix (in case of linear hypotheses) that is unnecessary when all estimated coefficients converge at the same rate (see, e.g., Park and Phillips, 1988, 1989). We posit in the following proposition a sufficient (asymptotic) rank condition that ensures that the Wald-type test statistics have asymptotic chi-squared null distributions. Note that if none of the hypotheses mixes coefficients with different convergence rates no additional complications compared to a standard situation with all estimated coefficients converging at the same rate arise.

Proposition 5 (Hong and Wagner 2014, Proposition 3)**.** *Let* y_t *be generated by* (3.1) *with the given assumptions in place. Consider* s *linearly independent restrictions collected in* $H_0 : R\theta = r$ *with* $R \in \mathbb{R}^{s \times 4N}$ *of full row rank* s, $r \in \mathbb{R}^s$ *and suppose that there exists a (matrix sequence)* $G_R = G_R(T)$ *such that* $\lim_{T\to\infty} G_R R G = R^*$ *with* $R^* \in \mathbb{R}^{s \times 4N}$ *of full row rank* s.

Then it holds under H_0 *that the Wald-type statistics:*

$$\hat{W} := \left(R\hat{\theta} - r\right)'\left[R\left(Z'Z\right)^{-1}Z'\left(I_T \otimes \hat{\Omega}_{u.v}\right)Z\left(Z'Z\right)^{-1}R'\right]^{-1}\left(R\hat{\theta} - r\right), \tag{3.12}$$

$$\tilde{W} := \left(R\tilde{\theta} - r\right)'\left[R\left(Z'\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)Z\right)^{-1}R'\right]^{-1}\left(R\tilde{\theta} - r\right) \tag{3.13}$$

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

3.2.1. Testing for Poolability and Pooled Estimation

As outlined in the introduction a key advantage of the SUR setting is that it allows to test for *in principle arbitrary forms* of poolability rather than assuming poolability from the outset as in panel analysis. Clearly, the results from Propositions 4 and 5 allow to test for poolability of the coefficients. In Appendix C.1 we briefly present the test statistics and the correspondingly pooled estimators for three "standard" pooling tests involving all cross-section members. These are labelled as: (P), where all coefficients except for the intercepts are pooled; (S), where only the coefficients to $x_{i,t}$ and $x_{i,t}^2$ are pooled and (T), where only the linear trend coefficient is pooled.

The first variant of pooling corresponds closely to a fixed-effects panel model, with individual specific fixed effects. Note, however, that the literature does not yet provide the theory for panel estimation methods (with $N \to \infty$) for cross-sectionally dependent panels of cointegrating polynomial regressions. de Jong and Wagner (2016), based on the seminal work of Phillips and Moon (1999), provide theory for the cross-sectionally independent case for the cubic formulation with one- and two-way fixed effects.[10]

If the considered null hypothesis is not rejected, then pooled estimation, as described for these three cases in Appendix C.1, of a smaller number of parameters allows to lift some efficiency gains in estimation. For our data, the above-given three "global" hypotheses (P), (S) and (T) are rejected.[11] A more detailed analysis, see Section 3.3, of the FM-SUR results reveals that the coefficient corresponding to the linear time trend can be pooled in three subgroups (of sizes three, two and one). For the coefficients to GDP and its square, the stochastic regressors, group-wise pooling analysis identifies one group of size three for which pooling is not rejected.

Exploiting the possibilities of group-wise pooling just indicated necessitates formulating the corresponding Wald-type statistics as well as the corresponding group-wise pooled estimators. This is discussed in the following subsection for the setting relevant in our application. Along similar lines any form of group-wise pooling can be considered in more general SUCPR settings.

[10] Note again that the part of the empirical EKC literature that uses panel cointegration methods, estimates a system of equations like (3.1) with methods for linear cointegration developed for panels of cross-sectionally independent units. The SUCPR approach overcomes these two limitations, allowing for cross-sectional dependence and taking into account the specific form of nonlinear cointegration.

[11] As will be seen in Section 3.3, for the 19 countries considered, (non-)cointegration tests lead to evidence for a CPR relationship in six countries. The CPR and SUCPR analysis is consequently performed with the data for these six countries.

3.2.2. Group-Wise Pooling

In this subsection we consider testing the null hypothesis that the coefficients for the linear time trend are group-wise pooled over a partition of k subsets $I_{n_j},\ j = 1, \ldots, k$ with $I := \{1, \ldots, N\} = \bigcup_{j=1}^{k} I_{n_j}$. Similarly, we consider a partition over l subsets $I_{m_j},\ j = 1, \ldots, l$ for the regressors $[x_{i,t}, x_{i,t}^2]'$, i.e., $I = \bigcup_{j=1}^{l} I_{m_j}$. Without loss of generality we order the subsets according to decreasing cardinality, i.e., $|I_{n_1}| \geq \ldots \geq |I_{n_k}|$ and $|I_{m_1}| \geq \ldots \geq |I_{m_l}|$, denoting with $|S|$ here the number of elements of a set S.

The null hypothesis corresponding to group-wise poolability of the coefficients corresponding to the above partitioning is given by:

$$\begin{aligned} H_0^{\mathrm{GW}}: \qquad \delta_i &= \delta_j \qquad \forall\ i, j \in I_{n_d}\ \ \forall\ d \in \{\{1, \ldots, k\} : |I_{n_d}| > 1\} \\ \begin{pmatrix} \beta_{1,i} \\ \beta_{2,i} \end{pmatrix} &= \begin{pmatrix} \beta_{1,j} \\ \beta_{2,j} \end{pmatrix} \ \ \forall\ i, j \in I_{m_p}\ \ \forall\ p \in \{\{1, \ldots, l\} : |I_{m_p}| > 1\}. \end{aligned} \tag{3.14}$$

To construct the Wald-type test statistics discussed in Proposition 5 for this specific situation it is convenient to define a few more quantities. First, denote with $N_j = |I_{n_j}|,\ j = 1, \ldots, k$ and $M_j = |I_{m_j}|,\ j = 1, \ldots, l$. Furthermore, the elements of the index set I_{n_j}, a_{j,n_j} say, are considered sorted, i.e., $I_{n_j} = (a_{1,n_j}, a_{2,n_j}, \ldots, a_{N_j,n_j})$ with $1 \leq a_{1,n_j} < a_{2,n_j} < \cdots < a_{N_j,n_j} \leq N$ for $j = 1, \ldots, k$ and similarly for the sets $I_{m_j},\ j = 1, \ldots, l$. Using this notation and setting the restriction matrix to test for (the considered form of) group-wise poolability can be written as:

$$R^{\mathrm{GW}} := [R'_{n_1}, \ldots, R'_{n_k}, R'_{m_1}, \ldots, R'_{m_l}]' \in \mathbb{R}^{s \times 4N} \tag{3.15}$$

with

$$R_{n_j} := \left(\left(\mathbf{1}_{(N_j-1)} \otimes e'_{a_{1,n_j},N} \right) - \begin{pmatrix} e'_{a_{2,n_j},N} \\ \vdots \\ e'_{a_{N_j,n_j},N} \end{pmatrix} \right) \otimes e'_{2,4} \in \mathbb{R}^{(N_j-1) \times 4N} \tag{3.16}$$

for j such that $N_j > 1$ and $R_{n_j} = \emptyset$ otherwise; and

$$R_{m_j} := \left(\left(\mathbf{1}_{(M_j-1)} \otimes e'_{a_{1,m_j},N} \right) - \begin{pmatrix} e'_{a_{2,m_j},N} \\ \vdots \\ e'_{a_{M_j,m_j},N} \end{pmatrix} \right) \otimes \begin{pmatrix} e'_{3,4} \\ e'_{4,4} \end{pmatrix} \in \mathbb{R}^{2(M_j-1) \times 4N} \tag{3.17}$$

for j such that $M_j > 1$ and $R_{m_j} = \emptyset$ otherwise. The total number of restrictions is

$$s = \sum_{j=1}^{k} (N_j - 1) + 2 \sum_{j=1}^{l} (M_j - 1) \tag{3.18}$$

and, clearly, $r = 0$ (in $R\theta = r$) here. Using either the FM-SOLS estimates or the FM-SUR estimates, the two test statistics (3.12) and (3.13) can be calculated to test the considered null hypothesis H_0^{GW}.

Remark 8. In the above definition of the blocks of the restriction matrix, setting, e.g., $R_{n_j} = \emptyset$ for $N_j = 1$, merely states that for groups of size one, of course, no poolability hypothesis testing is performed. Equivalently, including only the subsets of size larger than one in the restrictions matrix R^{GW} would require to define another index, n_k^* say, until which the groups – ordered according to non-increasing size – comprise more than one element.

In case that the null hypothesis discussed above is not rejected, the corresponding group-wise pooled estimators can be (defined and) employed. To this end consider:

$$\check{D}_t := [\check{D}_{1,t}', \ldots, \check{D}_{k,t}']' \in \mathbb{R}^{k \times N}, \tag{3.19}$$

where

$$\check{D}_{j,t} := \sum_{i=1}^{N} \mathbb{1}_{\{i \in I_{n_j}\}} \cdot t \cdot e_{i,N}', \quad j = 1, \ldots, k. \tag{3.20}$$

For the stochastic regressors we similarly have

$$\check{X}_t := [\check{X}_{1,t}', \ldots, \check{X}_{l,t}']' \in \mathbb{R}^{2l \times N}, \tag{3.21}$$

with

$$\check{X}_{j,t} := \sum_{i=1}^{N} \mathbb{1}_{\{i \in I_{m_j}\}} \cdot \left(e_{i,N}' \otimes X_{i,t}\right), \quad j = 1, \ldots, l. \tag{3.22}$$

With these quantities the group-wise pooled model can be compactly written as

$$y_t = \check{Z}_t' \theta^{\mathrm{GW}} + u_t, \tag{3.23}$$

with $y_t := [y_{1,t}, \ldots, y_{N,t}]'$, $u_t := [u_{1,t}, \ldots, u_{N,t}]'$, $\check{Z}_t := [I_N, \check{D}_t', \check{X}_t']' \in \mathbb{R}^{(N+k+2l)\times N}$ and the parameter vector $\theta^{\text{GW}} := [c_1, \ldots, c_N, \delta_1, \ldots, \delta_k, \beta_1', \ldots, \beta_l']' \in \mathbb{R}^{N+k+2l}$, where $\beta_j = [\beta_{1,j}, \beta_{2,j}]'$ for $j = 1, \ldots, l$. Finally, stacking the quantities over time gives

$$y = \check{Z}\theta^{\text{GW}} + u, \tag{3.24}$$

with $y = [y_1, \ldots, y_T]'$, $\check{Z} = [\check{Z}_1, \ldots, \check{Z}_T]'$ and $u = [u_1, \ldots, u_T]'$.

The correction terms for the group-wise pooled FM-SOLS and FM-SUR estimators are defined as $A^{\text{GW}*} := [0_{1\times(N+n_k)}, A_1^{\text{GW}*\prime}, \ldots, A_l^{\text{GW}*\prime}]'$, $\tilde{A}^{\text{GW}*} := [0_{1\times(N+n_k)}, \tilde{A}_1^{\text{GW}*\prime}, \ldots, \tilde{A}_l^{\text{GW}*\prime}]'$, with

$$A_j^{\text{GW}*} := \sum_{i=1}^{N} \mathbb{1}_{\{i\in I_{m_j}\}} \cdot \left(\hat{\Delta}_{vu}^{+}\right)^{i,i} \cdot \begin{pmatrix} T \\ 2\sum_{t=1}^{T} x_{i,t} \end{pmatrix}, \quad j = 1, \ldots, l, \tag{3.25}$$

$$\tilde{A}_j^{\text{GW}*} := \sum_{i=1}^{N} \mathbb{1}_{\{i\in I_{m_j}\}} \cdot \left(\hat{\Delta}_{vu}^{+}\right)^{i,\cdot} \left(\hat{\Omega}_{u\cdot v}^{-1}\right)^{\cdot,i} \cdot \begin{pmatrix} T \\ 2\sum_{t=1}^{T} x_{i,t} \end{pmatrix}, \quad j = 1, \ldots, l. \tag{3.26}$$

For group-wise pooled estimation the weighting matrix is given by $\check{G} := \operatorname{diag}(\check{G}_c, \check{G}_D, \check{G}_X) = \operatorname{diag}(T^{-1/2}\cdot I_N, T^{-3/2}\cdot I_k, I_l \otimes \operatorname{diag}(T^{-1}, T^{-3/2}))$. The limit stochastic process is now given by $\check{J}(r) := [I_N, \check{J}_D', \check{J}_X']'$, with $\check{J}_D(r) := [\check{J}_{D_1}(r)', \ldots, \check{J}_{D_k}(r)']'$ and $\check{J}_{D_j}(r) := \sum_{i=1}^{N} \mathbb{1}_{\{i\in I_{n_j}\}} \cdot r \cdot e_{i,N}'$ for $j = 1, \ldots, k$. The process $\check{J}_X(r) := [\check{J}_{X_1}(r)', \ldots, \check{J}_{X_l}(r)']'$ is composed of $\check{J}_{X_j}(r) := \sum_{i=1}^{N} \mathbb{1}_{\{i\in I_{m_j}\}} \cdot \left(e_{i,N}' \otimes \begin{pmatrix} B_{v_i}(r) \\ B_{v_i}^2(r) \end{pmatrix}\right)$ for $j = 1, \ldots, l$.

Proposition 6. *Let y_t be generated by* (3.24)*, the discussed restricted version of* (3.1) *with group-wise pooled parameters, with the assumptions given in place. Assume again that, based on the OLS residuals, all required long run variances are estimated consistently. Using the correction factors defined in* (3.25) *and* (3.26)*, the* group-wise *FM-SOLS and FM-SUR estimators are given by:*

$$\hat{\theta}^{GW} := \left(\check{Z}'\check{Z}\right)^{-1}\left(\check{Z}'y^{+} - A^{GW*}\right), \tag{3.27}$$

$$\tilde{\theta}^{GW} := \left(\check{Z}'\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)\check{Z}\right)^{-1}\left(\check{Z}'\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)y^{+} - \tilde{A}^{GW*}\right). \tag{3.28}$$

As $T \to \infty$ it holds that:

$$\check{G}^{-1}\left(\hat{\theta}^{GW} - \theta^{GW}\right) \Rightarrow \left(\int \check{J}\check{J}'\right)^{-1} \int \check{J}dB_{u\cdot v}, \tag{3.29}$$

$$\check{G}^{-1}\left(\tilde{\theta}^{GW} - \theta^{GW}\right) \Rightarrow \left(\int \check{J}\Omega_{u\cdot v}^{-1}\check{J}'\right)^{-1} \int \check{J}\Omega_{u\cdot v}^{-1}dB_{u\cdot v}. \tag{3.30}$$

In the following empirical analysis we discuss and compare unrestricted, pooled and group-wise pooled estimation results.

3.3. Empirical Analysis

The empirical analysis builds upon Wagner (2015), who performs individual country FM-CPR analysis of the EKC for CO_2 emissions for 19 early industrialized countries. The first step, prior to the SUR analysis performed here, is to reassess the findings of the earlier paper, since we now have data ranging from 1870–2013 rather than only until 2000. The CO_2 emissions data are from the Carbon Dioxide Information Analysis Center of the US Department of Energy and comprise total CO_2 emissions from fossil fuel usage.[12] The GDP data, measured in 1990 Geary-Khamis Dollars, are from the Maddison project at the University of Groningen and from The Conference Board Total Economy Database.[13] The data are used in logarithms of per capita quantities. Throughout, for all estimators and all tests we use the Bartlett kernel and the bandwidth chosen according to Newey and West (1994).

For all 19 early industrialized countries investigated, the unit root null hypothesis is not rejected for log GDP per capita using the unit root tests of Phillips and Perron (1988) as well as the fixed-b versions of this test developed by Vogelsang and Wagner (2013).[14] Using the tests for cointegration in CPRs of Wagner (2013) and Wagner and Hong (2016) leads to evidence for a quadratic cointegrating EKC for CO_2 emissions for the following six countries: Austria (AT), Belgium (BE), Finland (FI), the Netherlands (NL), Switzerland (CH) and the UK.[15]

Table 3.1 shows the results of estimating the quadratic EKC (3.1) using both individual country FM-CPR (as used in Wagner, 2015) and the two SUR estimators discussed in Section 3.2, FM-SOLS and FM-SUR, for the six countries listed above. In addition, the lower left block of the table contains the results when estimating the EKC "fully" pooled, allowing only for country specific intercepts (the form of pooling referred to as (P) in Section 3.2.1).[16] The following messages emerge from the table: First, the estimated

[12]See Boden *et al.* (2016) and `http://cdiac.ornl.gov`.
[13]See Bolt and van Zanden (2014), `http://www.ggdc.net/maddison/maddison-project/home.htm` and `http://www.conference-board.org/data/economydatabase`.
[14]The results are given in Table C.1 in Appendix C.2.
[15]This is slightly different from Wagner (2015) who finds evidence for a quadratic EKC for CO_2 emissions for only four out of the six countries above: Austria, Belgium, Finland and the UK. These differences may stem from the different sample range and/or the fact that the CO_2 emissions data have been updated.
[16]In formal terms, estimation of (3.1) is performed under the restrictions $\delta_i = \delta$, $\beta_{1,i} = \beta_1$ and $\beta_{2,1} = \beta_2$ for $i = 1, \ldots, 6$. Note also that we obtain very similar results for the cubic specification, both with

coefficients (all significant with "correct" signs) and a fortiori the estimated turning points do usually not differ strongly across the three methods for each country. The exception here is Austria where the FM-CPR turning point is more than twice as large as the FM-SOLS and FM-SUR turning points. For Switzerland, the turning point is estimated far outside the sample range, with values ranging from 1.3 to 3.1 millions, by all three estimators. This finding is related to the fact that, see Figure 3.2, per capita CO_2 emissions are essentially constant in Switzerland since about 1980. Second, with respect to the two SUR estimators the differences are mostly very minor, with the one exception being Finland. For this reason we focus on the FM-SUR estimator in the discussion from now on.[17] Third, the estimated coefficients and consequently the estimated turning points differ substantially across countries and this heterogeneity can – by construction – not be captured by the pooled, i.e., almost panel-type, estimation results in the lower left block. This finding highlights that commonly used panel methods need to be considered very carefully, or maybe not used at all for situations as considered here.[18]

The results from Table 3.1 are displayed graphically in Figures 3.1 and 3.2. The first figure displays the estimated EKCs, given by using 144 equidistant values for the explanatory variable from the range of log GDP per capita associated with values of the time trend ranging from 1,...,144 and inserting these values in Equation (3.1) using the coefficient estimates obtained from both FM-CPR (solid with x-marks) and FM-SUR (solid). Additionally the graphs include the scatter plots between log GDP per capita and log CO_2 emissions per capita. The similar coefficient estimates translate, as expected or in fact necessary, into very similar estimated EKCs. Figure 3.2 displays the actual values of log per capita CO_2 emissions with the fitted values obtained from both FM-CPR and FM-SUR estimation. Clearly, the two fitted value lines corresponding to FM-CPR and FM-SUR are very close to each other for all countries, with the still small but relatively largest differences for Austria (for which also the estimated turning point differs most between the two methods). In general the fit is very good, especially for the period since the second world war.

Performing the poolability tests (P), (S) and (T) described in Section 3.2.1 and in more detail in Appendix C.1 for the six considered countries leads throughout to rejections of the respective null hypotheses for both tests, i.e., the tests based on the FM-SOLS

respect to cointegration testing and estimation results. The coefficient to the third power of GDP is not significant throughout and it thus suffices to consider the quadratic specification.

[17]The similarity of the findings with both the FM-SOLS and the FM-SUR estimators is made clearly visible in Figures C.1 and C.2 in Appendix C.2.

[18]As already mentioned, de Jong and Wagner (2016) consider a panel version of FM-type estimators for panels of cointegrating polynomial regressions under the assumption of cross-sectional independence. Under appropriate assumptions it may be the case that the pooled estimates converge to "average coefficients", see Phillips and Moon (1999) for details. These issues remain to be studied for the cointegrating polynomial regression case.

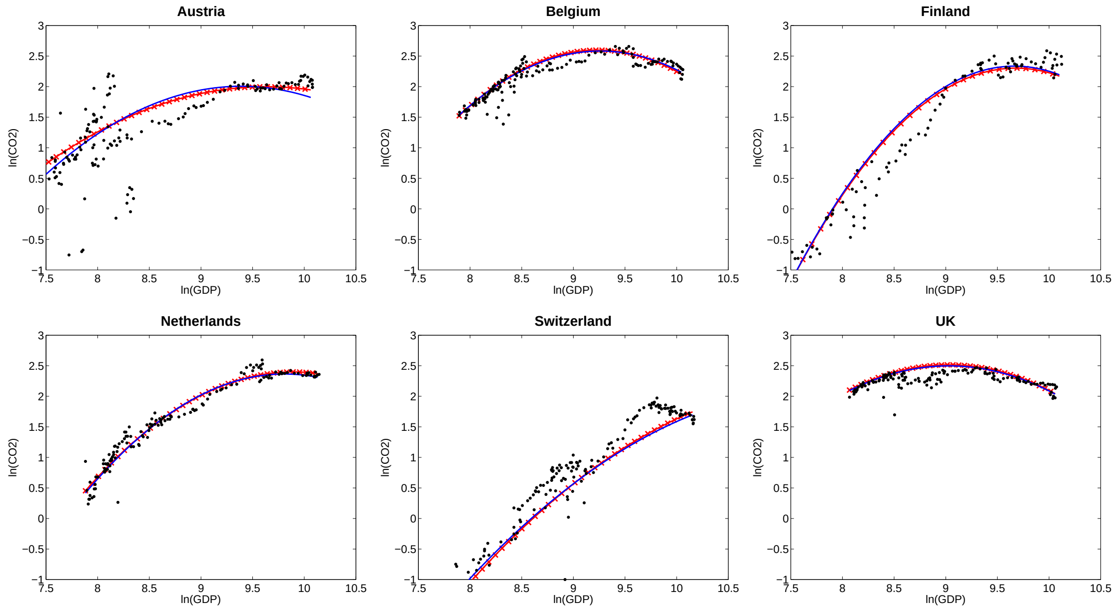

Figure 3.1.: EKC estimation results for Equation (3.1): scatter plot and EKC. The dots show the pairs of observations of ln(GDP) per capita and ln(CO_2) emissions per capita. The lines show results based on inserting 144 equidistant points from the sample range of ln(GDP) per capita, with corresponding values of the linear trend given by $t = 1, \dots, 144$ in the estimated relationship (3.1). The solid lines with x-marks correspond to the FM-CPR estimates and the solid lines to the FM-SUR estimates.

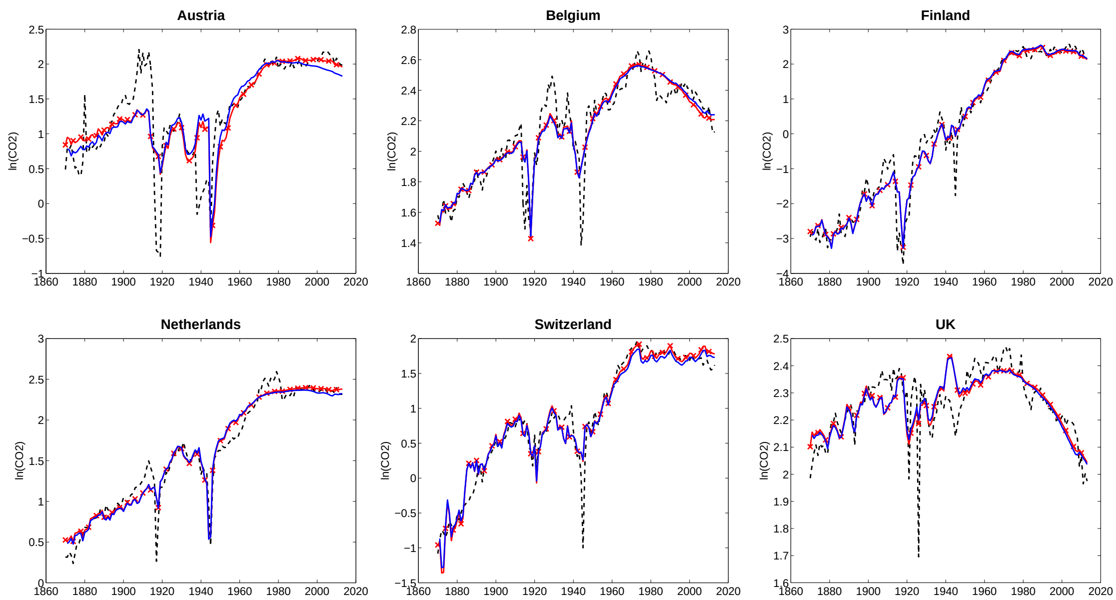

Figure 3.2.: EKC estimation results for Equation (3.1): actual and fitted values. The dashed lines show the actual values of ln(CO_2) per capita emissions, the solid lines with x-marks the FM-CPR fitted values and the solid lines the FM-SUR fitted values.

estimator (3.12) and the FM-SUR estimator (3.13). For the hypothesis (P) this is already expected, given the cross-country heterogeneity of the unrestricted estimates, compare again the results in Table 3.1. The prize to be paid when applying pooled estimation, allowing only for country specific intercepts, despite this restriction being rejected, is clearly visible when looking at Figures 3.3 and 3.4, which are similar in structure to Figures 3.1 and 3.2. For all six countries the differences are quite huge, both with respect to slope and shape. These differences translate directly into partly drastic reductions of fit, when considering the fitted value graphs in Figure 3.4. Thus, testing for group-wise poolability and potentially group-wise pooled estimation, as outlined in Section 3.2.2, are the logical next steps.

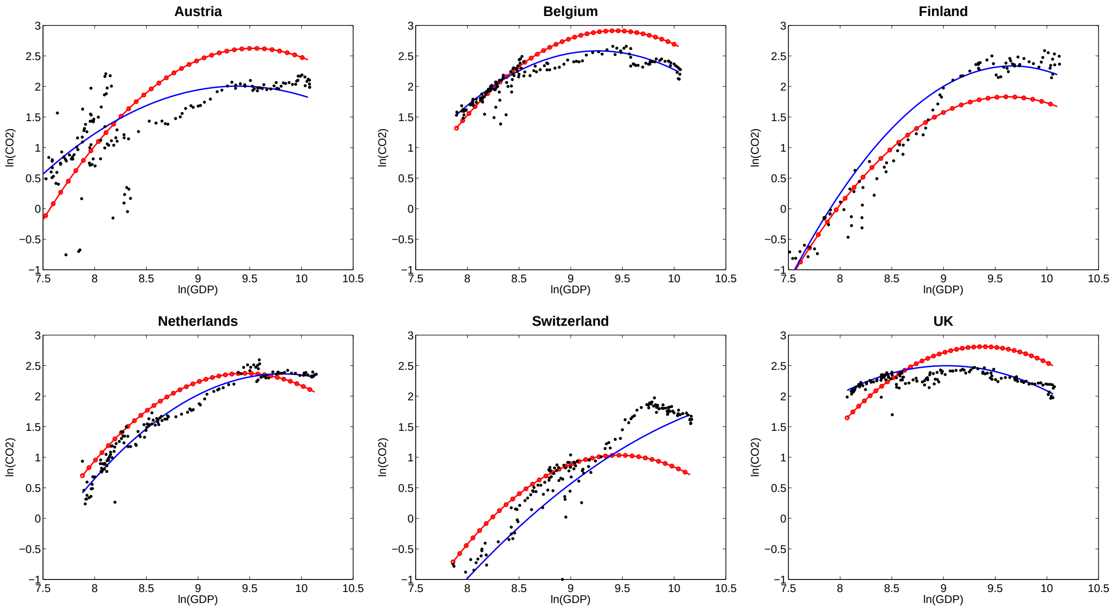

Figure 3.3.: EKC estimation results for Equation (3.1): scatter plot and EKC. The solid lines correspond to the FM-SUR estimates and the solid lines with o-marks to the pooled FM-SUR estimates. For further explanations see notes to Figure 3.1.

In many applications the researcher may have some prior knowledge concerning candidates for group-wise pooling. To a certain extent this is also the case here, as one expects that very similar countries, e.g., Belgium and the Netherlands, may have very similar EKCs. Here, however, we pursue a more exploratory approach. We start by testing for the discussed three forms of pooling – (P), (S) and (T) – in all possible sub-groups. This means that we test for these forms of poolability in all 15 possible country-pairs, 20 country-triples and so on.[19] The results are given in Table 3.2 and Table C.3 in Appendix C.2.

[19]Note that we test for the three forms of poolability using only data for the subset of countries under investigation. We do not perform all possible tests of group-wise poolability in all possible partitions into multiple subgroups using the data for all six countries. Doing that would entail a rather large number of tests to be performed. Let us stress also that the approach is to be understood exploratory,

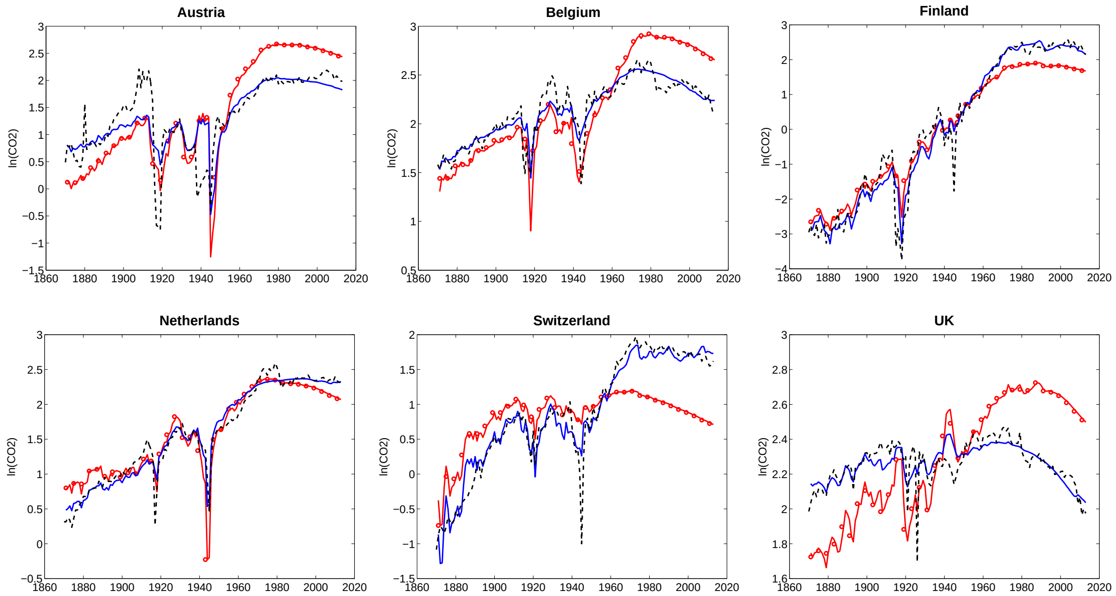

Figure 3.4.: EKC estimation results for Equation (3.1): actual and fitted values. The dashed lines show the actual values, the solid lines the FM-SUR fitted values and the solid lines with o-marks the pooled FM-SUR fitted values.

Table 3.2 contains the numbers of groups of the respective sizes for which the corresponding poolability hypothesis cannot be rejected, with the group members displayed in Table C.3. As for the coefficients, also for the tests the differences are minor between the FM-SOLS and FM-SUR results and thus we focus again on the results obtained with FM-SUR. The full pooling hypothesis (P) is rejected throughout, even for all pairs. With FM-SUR, the slope parameters β_1 and β_2 can be pooled for (i.e., the pooling hypothesis (S) is not rejected for) four country-pairs, two country-triples and one group of size four (containing AT, BE, NL and UK). With respect to the trend parameters there are three country groups of size three, for which the trend slope can be pooled. Austria, Finland and the UK are each present in two of the three groups.

We take the above results as starting point to estimate the EKC for the six countries in a group-wise pooled fashion. In particular we consider: the trend slope pooled in three groups, comprising Austria, Finland and Switzerland; Belgium and the UK; and the Netherlands (as group of size one) respectively. The slope parameters are pooled in four groups, given by Belgium, the Netherlands and the UK; and the three single member

since neither of the complications resulting from multiple testing is even addressed, let alone solved. Note that there is a recent literature to identify (coefficient) structure in panel data, see Ke *et al.* (2016) or Su *et al.* (2016). However, our problem does not fit that literature either, since we have small (to medium) N and cointegration in the SUCPR setting, whereas this literature is to date concerned with standard stationary settings.

groups Austria; Finland; and Switzerland.[20] Table 3.3 displays the estimation results. As observed up to now, the estimates are also very similar for the now group-wise pooled FM-SOLS and FM-SUR estimates. Looking at the coefficients in the individual groups clearly shows that the group-wise pooled estimates are – almost by construction when using group-wise pooled least squares estimation – close to the averages of the country specific estimates given in Table 3.1. Of course, group-wise pooled estimation is not simply *mean-group* estimation, and thus the group-wise pooled coefficients estimates do not simply coincide with the averages. The same observations as for the coefficients hold, of course again by implication, for the estimated turning points.

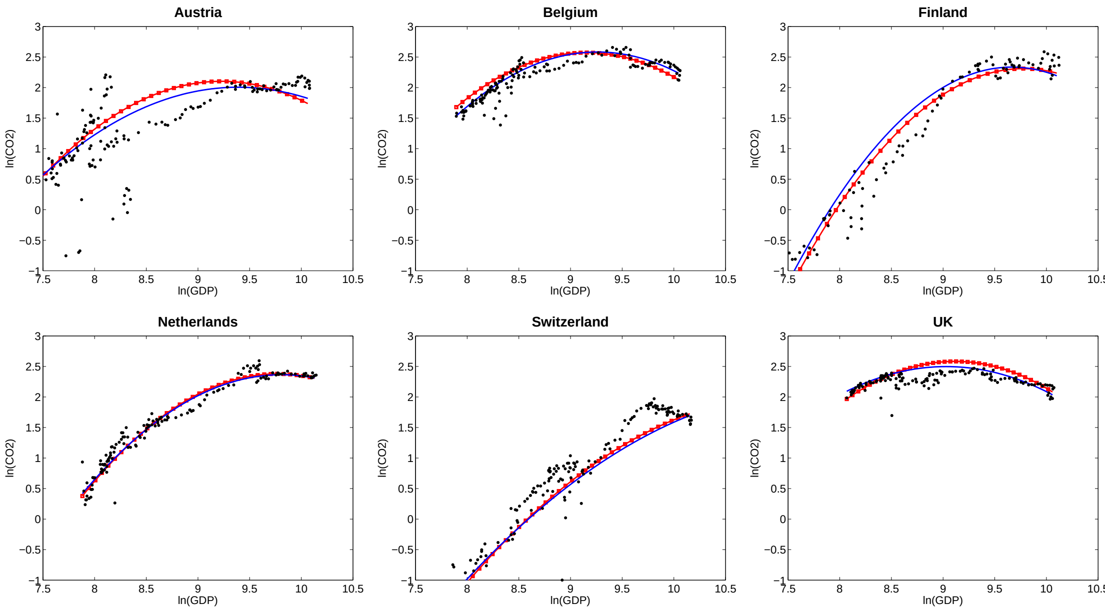

Figure 3.5.: EKC estimation results for Equation (3.1): scatter plot and EKC. The solid lines correspond to the FM-SUR estimates and the solid lines with the square symbols to the group-wise pooled FM-SUR estimates. For further explanations see notes to Figure 3.1.

The benefit of group-wise pooling becomes clearly visible when considering the results graphically in Figures 3.5 and 3.6. These two figures, again similar in structure to Figures 3.1 and 3.2, show clearly that imposing group-wise poolability restrictions supported by hypothesis testing in group-wise pooled FM-SUR estimation (solid lines with square symbols) leads to very similar estimates of the EKCs compared to non-pooled FM-SUR estimation (solid lines). Importantly, also the (unavoidable) reduction in fit is negligible (see Figure 3.6), with the exception of the UK to some extent. Recall for comparison the drastic reduction in fit when pooling all slope and trend coefficients over all countries

[20]We take this group of three countries for pooling the trend slope, since for this group the poolability hypothesis is not rejected also for all subgroups of two of these three countries. The choice is made using similar arguments also for the slope parameters: Poolability of the slope parameters is not rejected for the three pairs of countries of the triple Belgium, the Netherlands and the UK.

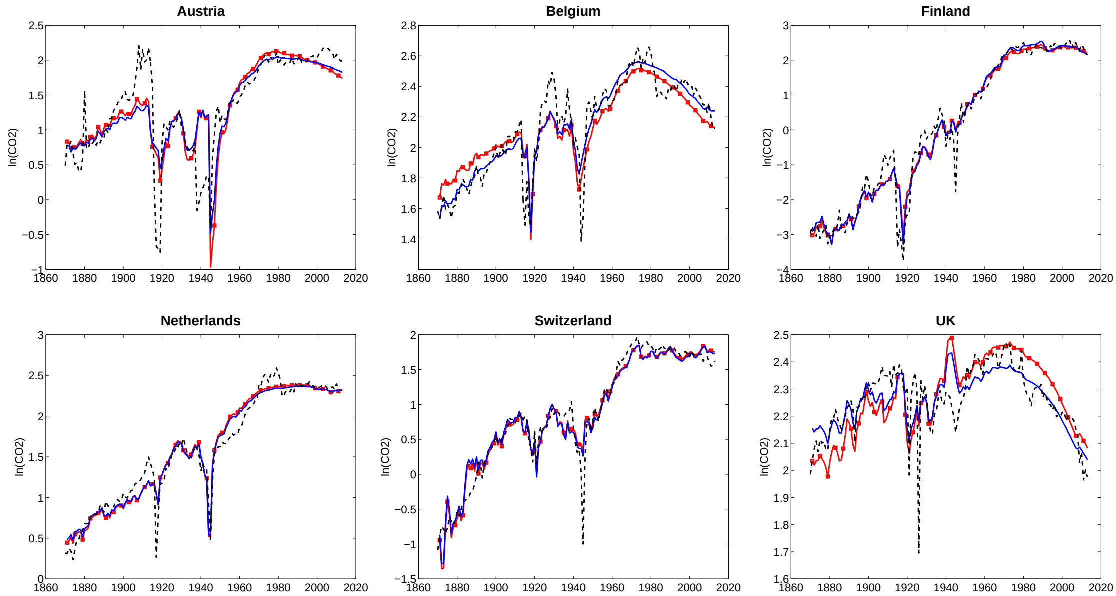

Figure 3.6.: EKC estimation results for Equation (3.1): actual and fitted values. The dashed lines show the actual values, the solid lines the FM-SUR fitted values and the solid lines with square symbols the group-wise pooled FM-SUR fitted values.

displayed in Figures 3.3 and 3.4.[21] Group-wise pooling of a form adapted to the situation leads to a sizeable reduction of the number of parameters to be estimated, in our case from 28 to 18, without any clearly visible losses in terms of approximation quality. Unthoughtful global pooling, i.e., panel-type estimation, leads to drastically worse results. These findings illustrate that a seemingly unrelated CPR approach is indeed very useful for the analysis of the EKC and similar relationships in situations with multi-country or multi-regional data where the cross-sectional dimension is small.

3.4. Summary and Conclusions

We provide tools for multi-country (or multi-regional) cointegration analysis of the environmental Kuznets curve (EKC) by pursuing a seemingly unrelated cointegrating polynomial regressions (SUCPR) approach advocated by Hong and Wagner (2014). The approach can also be applied in other contexts in which inverted U-shaped relationships are studied, such as the intensity of use (IOU) relationship between GDP and energy or material intensity (see, e.g., Guzmán *et al.*, 2005; Labson and Crompton, 1993).

[21] Figures C.3 and Figure C.4 in Appendix C.2 compare the group-wise pooled and pooled FM-SUR results. These two figures clearly make the same point as the figures in the main text, but contrasting group-wise pooled and pooled estimation results in the same figure highlights the benefits of group-wise pooling compared to pooling nicely.

The SUCPR approach addresses three of the main challenges of the existing literature: First, it takes into account that powers of integrated processes are themselves not integrated processes and that consequently cointegration analysis of the EKC needs to resort to methods designed for this specific form of nonlinear relationship, labelled cointegrating polynomial regression by Wagner and Hong (2016). The implications of this fact for single country EKC analysis have been pointed out earlier in Wagner (2015); the present chapter translates and extends this discussion to the multi-country data case. Second, it is not necessarily the case that, e.g., emissions and GDP data for different countries are independent of each other, an assumption typically made in the panel EKC literature. Third, furthermore the EKC relationship, if present, need not be identical (potentially up to country specific individual effects) across countries. This, however, is the the key assumption underlying pooling which panel data analysis rests upon. Our SUCPR approach addresses these three issues and provides new tools for group-wise poolability testing and, in case the restrictions are not rejected, corresponding group-wise pooled estimation.

Developing poolability tests and correspondingly pooled estimators for general sets of restrictions is shown to be extremely useful in our application to CO_2 emissions data for six early industrialized countries over the period 1870–2013. It turns out that the trend respectively slope parameters can be pooled over different country sub-groups, a situation that we label group-wise pooling. The results show that group-wise pooled estimation provides fits that are close to the fits from either individual country or unrestricted SUCPR estimation, whilst the number of parameters to be estimated is substantially reduced. Altogether, the simple reduced form SUCPR EKC analysis leads to very good fit, especially since the second world war, and meaningful estimates of the turning points. Performing SUCPR estimation in a fully pooled fashion with only country specific intercepts, by comparison leads to substantial losses in terms of fit. A major limitation of any SUR approach is the limitation to situations with a relatively small cross-sectional dimension. For data sets with large cross-sectional dimension panel data approaches will need to be pursued, with all advantages and disadvantages. For a first step in this direction see de Jong and Wagner (2016).

The empirical results of this chapter illustrate the usefulness of SUCPR analysis of the EKC, but the reduced form character of the analysis presented here dictates the necessary next steps of the research agenda: First, for certain applications it may be necessary to extend the methodology to allow for the inclusion of stationary regressors.[22] This is a pertinent issue in, e.g., IOU analysis. In case of substitution possibilities between

[22] Pre-determined stationary regressors can be accommodated more easily than endogenous stationary regressors. Endogenous stationary regressors will require to construct an instrumental variables-type extension of the estimators discussed here. Even if an IV-type estimator is developed, the availability of valid and relevant instruments will, as always, be a challenge in actual applications.

different metals (see, e.g., Stuermer, 2016) or energetic resources, the inclusion of *relative prices* is of key importance to capture substitution elasticities. Note in this respect that the SUR approach also can be used to study EKC or IOU relationships for a set of different emissions variables or resource intensities for a given country or a small number of countries. This allows to study the interrelationships in a system of cointegrating polynomial regressions. Second, in particular for regional data it may be important to allow for the inclusion of common *aggregate* variables, i.e., technically speaking for the inclusion of common (nonstationary) regressors.[23] Third, it is always important to strive for extending the discussed methods to allow for a more structural analysis of EKC- or IOU-type relationships by considering more general specifications. Extensions along all three dimensions are or will be investigated in ongoing and planned research.

[23] This may on a bigger scheme also be relevant for multi-country data, e.g., EU data with common EU-wide variables to be included. These could be related to common policies or regulations.

	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	TP	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	TP
	Austria				**Belgium**			
FM-CPR	-0.017	6.247	-0.277	78,059	-0.004	11.358	-0.599	13,142
(t-values)	(-3.713)	(2.510)	(-2.019)		(-2.727)	(10.121)	(-10.159)	
FM-SOLS	-0.018	10.033	-0.486	30,515	-0.005	12.313	-0.649	13,230
(t-values)	(-4.750)	(4.634)	(-4.073)		(-3.629)	(13.325)	(-13.384)	
FM-SUR	-0.013	8.278	-0.403	28,699	-0.004	10.687	-0.562	13,556
(t-values)	(-4.095)	(4.891)	(-4.182)		(-3.935)	(14.073)	(-13.856)	
	Finland				**Netherlands**			
FM-CPR	-0.029	15.610	-0.737	39,523	0.001	9.437	-0.481	18,280
(t-values)	(-3.260)	(9.356)	(-8.796)		(0.614)	(8.438)	(-8.076)	
FM-SOLS	-0.039	16.162	-0.746	50,845	0.001	9.823	-0.502	17,783
(t-values)	(-4.974)	(10.600)	(-9.721)		(0.585)	(9.334)	(-8.970)	
FM-SUR	-0.029	15.892	-0.752	38,892	0.002	10.185	-0.524	16,524
(t-values)	(-5.863)	(14.140)	(-12.276)		(1.053)	(11.511)	(-10.878)	
	Switzerland				**UK**			
FM-CPR	-0.024	7.755	-0.273	1.5×10^6	-0.008	8.657	-0.446	16,287
(t-values)	(-6.421)	(6.312)	(-4.031)		(-3.406)	(6.532)	(-6.794)	
FM-SOLS	-0.024	6.933	-0.232	3.1×10^6	-0.007	9.887	-0.516	14,496
(t-values)	(-7.981)	(7.399)	(-4.463)		(-3.448)	(8.539)	(-9.001)	
FM-SUR	-0.022	7.441	-0.265	1.3×10^6	-0.007	8.402	-0.437	15,068
(t-values)	(-7.743)	(7.665)	(-4.941)		(-4.035)	(8.667)	(-9.001)	
	Pooled							
FM-SOLS	-0.015	13.572	-0.667	26,053				
(t-values)	(-8.344)	(20.474)	(-18.326)					
FM-SUR	-0.013	13.594	-0.677	23,002				
(t-values)	(-15.293)	(35.246)	(-32.226)					

Table 3.1.: FM-CPR, FM-SOLS, FM-SUR and pooled FM-SOLS and FM-SUR estimation results for Equation (3.1). The estimated turning points TP are computed as $\exp\left(-\frac{\hat{\beta}_1}{2\hat{\beta}_2}\right)$.

Group size k	2	3	4	5	6	2	3	4	5	6
Total nr. of groups of size k	15	20	15	6	1	15	20	15	6	1
	FM-SOLS					**FM-SUR**				
Linear Trend & Stochastic Regressors (P)										
Stochastic Regressors (S)	3	2				4	2	1		
Linear Trend (T)	6	2				7	3			

Table 3.2.: Testing for group-wise poolability of subsets of coefficients. The numbers indicate the number of groups of size k for which the indicated null hypothesis of group-wise poolability is not rejected. The members of the groups for which the respective null hypotheses are not rejected are given in Table C.3 in Appendix C.2. Empty entries correspond to zeros. The left column displays the results for the FM-SOLS test statistic (3.12) and the right column displays the results for the FM-SUR test statistic (3.13). Individual test decisions are performed at the 1% significance level.

	$\hat{\delta}_{n_1}$	$\hat{\delta}_{n_2}$	$\hat{\delta}_{n_3}$
Countries	**AT-FI-CH**	**BE-UK**	**NL**
FM-SOLS	-0.022	-0.009	0.001
(t-values)	(-6.825)	(-7.827)	(0.883)
FM-SUR	-0.019	-0.009	0.002
(t-values)	(-9.443)	(-11.122)	(2.017)

	$\hat{\beta}_{1,m_1}$	$\hat{\beta}_{2,m_1}$	$\hat{\beta}_{1,m_2}$	$\hat{\beta}_{2,m_2}$	$\hat{\beta}_{1,m_3}$	$\hat{\beta}_{2,m_3}$	$\hat{\beta}_{1,m_4}$	$\hat{\beta}_{2,m_4}$
Countries	**BE-NL-UK**		**AT**		**FI**		**CH**	
FM-SOLS	11.580	-0.600	11.054	-0.534	13.907	-0.654	6.991	-0.242
(t-values)	(16.445)	(-16.355)	(5.372)	(-4.691)	(12.850)	(-10.435)	(6.980)	(-4.310)
TP	15,514		31,304		41,480		1.9×10^6	
FM-SUR	10.852	-0.562	10.656	-0.521	14.649	-0.704	8.261	-0.319
(t-values)	(21.370)	(-20.896)	(6.370)	(-5.524)	(15.528)	(-12.645)	(8.781)	(-6.109)
TP	15,677		27,646		32,942		4.2×10^5	

Table 3.3.: Group-wise pooled estimation results for Equation (3.1) using FM-SOLS and FM-SUR. The trend parameter δ is pooled in three groups (of sizes three, two and one) and the slope parameters β_1, β_2 are pooled in four groups (of sizes three and thrice one). The estimated turning points TP are computed as $\exp\left(-\frac{\hat{\beta}_1}{2\hat{\beta}_2}\right)$.

A. Appendix to Chapter 1

A.1. Auxiliary Lemmata and Proofs

Auxiliary Lemmata

This subsection contains some auxiliary lemmata which are required to prove the main results of Chapter 1. The following lemma is proven in Kasparis (2008, Lemma A1(i))).

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

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

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

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

Proof. Consider $f(x) := x^q$, $x \in \mathbb{R}$. From the mean value theorem it follows that $f(y) - f(x) = f'(\zeta)(y - x)$, i.e., $y^q - x^q = q\zeta^{q-1}(y - x)$, with $x < y$ and $\zeta \in (x, y)$. Therefore, it holds

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

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

$$\begin{aligned}
&\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p \left[\left(\frac{x_{t+h}}{T^{1/2}}\right)^q - \left(\frac{x_t}{T^{1/2}}\right)^q \right] v_t v_{t+h} \\
&\quad = \frac{q}{T^{1/2}} \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \left(\frac{x_t}{T^{1/2}}\right)^p \left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1} \sum_{m=1}^{h} v_t v_{t+m} v_{t+h}.
\end{aligned}$$

The assertion is hence equivalent to showing that:

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

In the course of the proof it is helpful to resort to strong approximations, which we get from the Skorohod representation theorem, see Pollard (1984, p. 71–72) or Csörgo and Horváth (1993, p.4). For a discussion of this issue in a nonlinear cointegration context see, e. g., Park and Phillips (1999, Lemma 2.3) and Park and Phillips (2001). Since we are concerned with weak convergence results in this chapter, we can w.l.o.g. use a distributionally equivalent version of $T^{-1/2}x_{\lfloor rT \rfloor}$, $X_T^*(r)$ say, that fulfills $\sup_{r\in[0,1]}|(X_T^*(r)) - B_v(r)| = o_{a.s.}(1)$, with $B_v(r)$ the Brownian motion given in (1.4). Setting $\tilde{C} := \sup_{r\in[0,1]}|B_v(r)| + 1/2$, it holds that

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

Furthermore, it holds that

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

and thus it follows from Lemma 3 that

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

This implies

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

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

$$\sup_{r\in[0,1]}\sup_{0\leq h\leq M_T} T^{-1/2}|\overline{x}_{\lfloor rT \rfloor}^h| \leq C + o_{a.s.}(1). \tag{A.3}$$

Using the triangular inequality and the bounds given in (A.1)–(A.3) the following inequalities hold:

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

with $\overline{k}(0) = \sup_{x\geq 0}|k(x)|$ as defined in Assumption 3. By similar arguments as given above it holds due to strict stationarity of $\{v_t\}$ that

$$\sup_{s\in[0,1]}\sup_{t=1,\ldots,T}\left|\frac{1}{M_T^{1/2}}\sum_{m=1}^{\lfloor sM_T\rfloor} v_{t+m}\right| \leq C^* + o_{a.s.}(1),$$

where $C^* \overset{d}{=} \tilde{C}$. Consequently,

$$\begin{aligned}
&\left|\frac{1}{T^{1/2}}\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\left(\frac{\overline{x}_t^h}{T^{1/2}}\right)^{q-1}\sum_{m=1}^{h} v_t v_{t+m} v_{t+h}\right| \\
&\quad\leq \left(\frac{M_T^3}{T}\right)^{1/2}\overline{k}(0)C^{p+q-1}C^*\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}\left|v_t v_{t+h}\right| + o_{\mathbb{P}}(1). \qquad \text{(A.4)}
\end{aligned}$$

Assumption 2 implies that:

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

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

$$\frac{1}{M_T}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}\left|v_t v_{t+h}\right| = O_{\mathbb{P}}(1). \qquad \text{(A.5)}$$

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

Lemma 5. *With assumptions 2 to 4 in place, it holds for all integers $0 \leq p$ that:*

$$\left|\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^p\left(v_t v_{t+h} - \mathbb{E}[v_t v_{t+h}]\right)\right| = o_{\mathbb{P}}(1). \qquad \text{(A.6)}$$

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

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

by showing

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

The left-hand side of (A.6) can be written as

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

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

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

which shows the claim of this lemma, since

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

with $\tilde{k} := \overline{k}(0) + 1$. It is the fact that our proof of this lemma uses some of the arguments of Kasparis (2008) that the same moment and bandwidth assumptions are required. These are consequently contained in our Assumptions 2 to 4. ∎

Proofs of Chapter 1

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

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

which is already well known, cf. Remark 2. For $i \in \{1, \dots, p\}$ it holds that

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

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

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

Note that

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

From Lemma 3 we know that $T^{-1/2} v_t = o_{a.s.}(1)$ for $t = 1, \dots, T$. Additionally, it holds that $T^{-1/2}|x_t| \le C + o_{a.s.}(1)$ for $t = 1, \dots, T$. From $\mathbb{E}[T^{-1/2} v_{\lfloor rT \rfloor}^2] = T^{-1/2}\Sigma_{vv} \to 0$ for all $r \in [0, 1]$, we conclude that

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

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

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

In the linear case, i. e. $i = j = 1$, the above term converges in probability to Δ_{vv}, cf. Remark 2 again. Next, consider $i > 1$ and $j = 1$, i. e.,

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

From Lemma 5 it follows that

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

Now, we show that

$$\left|\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\frac{1}{T}\sum_{t=1}^{T-h}\left(\frac{x_t}{T^{1/2}}\right)^{i-1} - \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i-1}\right|$$
$$= \left|\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\frac{1}{T}\sum_{t=T-h+1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i-1}\right| \tag{A.8}$$

is $o_{\mathbb{P}}(1)$. By Assumption 2, we get

$$\left|\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\frac{1}{T}\sum_{t=T-h+1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i-1}\right|$$
$$\le C^{i-1}\frac{1}{T}\sum_{h=0}^{M_T}\left|k\left(\frac{h}{M_T}\right)\right| |\mathbb{E}[v_0 v_h]| h + o_{\mathbb{P}}(1)$$
$$\le \overline{k}(0)|\Sigma_{\varepsilon\varepsilon}|C^{i-1}\frac{1}{T}\sum_{h=0}^{M_T} h \sum_{j=0}^{\infty}|c_{v,j}c_{v,j+h}| + o_{\mathbb{P}}(1)$$
$$\le \overline{k}(0)|\Sigma_{\varepsilon\varepsilon}|C^{i-1}\frac{1}{T}\sum_{j=0}^{\infty}|c_{v,j}|\sum_{h=0}^{\infty} h|c_{v,h}| + o_{\mathbb{P}}(1).$$

Moreover, it holds that

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

and thus

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

which implies that

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

Therefore, we obtain

$$\begin{aligned}
&\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{1}{T}\sum_{t=1}^{T-h}\frac{\Delta x_t^i}{T^{(i-1)/2}}v_{t+h} \\
&\quad = i\left(\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i-1}\right) + o_{\mathbb{P}}(1).
\end{aligned}$$

For the first term it holds that

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

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

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

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

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

Using Lemma 4 we obtain

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

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

$$\begin{aligned}
&\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\frac{\Delta x_t^i}{T^{\frac{i-1}{2}}}\frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} \\
&\quad = ij\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right)\mathbb{E}[v_0 v_h]\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^{i+j-2} + o_{\mathbb{P}}(1) \\
&\quad \Rightarrow ij\Delta_{vv}\int_0^1 B_v^{i+j-2}(r)dr.
\end{aligned}$$

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

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

$$\begin{aligned}\left(\hat{\Delta}_{\hat{\eta}\hat{\eta}}\right)_{(1,j+1)} &= \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} \hat{u}_t \frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} \\ &= \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} u_t \frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} - \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} Z_t'(\hat{\theta} - \theta) \frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}}.\end{aligned}$$

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

$$\begin{aligned}&\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} Z_t'(\hat{\theta} - \theta) \frac{\Delta x_{t+h}^j}{T^{\frac{j-1}{2}}} \\ &\quad = j \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) \frac{1}{T} \sum_{t=1}^{T-h} Z_t' G G^{-1}(\hat{\theta} - \theta) \left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h} + o_{\mathbb{P}}(1)\end{aligned} \tag{A.9}$$

by similar arguments as in the proof of Theorem 1 with G defined in (1.7). Expression (A.9) can be further rewritten as

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

Finally, we show that

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

Using the notation from Lemma 4 it holds that

$$\begin{aligned}&\left\| \sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) j \, \frac{1}{T^{3/2}} \sum_{t=1}^{T-h} \left(T^{1/2} Z_t' G\right) \left(\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h}\right) \right\| \\ &\quad \leq j \bar{k}(0) \sum_{h=0}^{M_T} \frac{1}{T^{3/2}} \sum_{t=1}^{T-h} \left\| \left(T^{1/2} Z_t' G\right) \left(\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h}\right) \right\| \\ &\quad \leq j \bar{k}(0) C^{j-1} \sum_{h=0}^{M_T} \frac{1}{T^{3/2}} \sum_{t=1}^{T-h} \left\| T^{1/2} Z_t' G \right\| |v_{t+h}| + o_{\mathbb{P}}(1).\end{aligned}$$

Observe that $\left\|\left(T^{1/2}D_t'G_D\right)\right\|^2 \leq C_D + o(1)$ for a finite constant C_D by Assumption 1 and thus

$$\left\|\left(T^{1/2}Z_t'G\right)\right\|^2 = \left\|\left(T^{1/2}D_t'G_D\right)\right\|^2 + \sum_{l=1}^{p}\left(\frac{x_t}{T^{1/2}}\right)^{2l} \leq K + o_{a.s.}(1),$$

with $K := C_D + \sum_{l=1}^{p} C^{2l}$, such that

$$\begin{aligned}&\left\|\sum_{h=0}^{M_T} k\left(\frac{h}{M_T}\right) j \,\frac{1}{T^{3/2}}\sum_{t=1}^{T-h}\left(T^{1/2}Z_t'G\right)\left(\left(\frac{x_{t+h}}{T^{1/2}}\right)^{j-1} v_{t+h}\right)\right\| \\ &\quad\leq\; j\bar{k}(0)C^{j-1}K^{1/2}\frac{1}{T^{1/2}}\sum_{h=0}^{M_T}\frac{1}{T}\sum_{t=1}^{T-h}|v_{t+h}| + o_{\mathbb{P}}(1) \qquad \text{(A.10)}\end{aligned}$$

follows. Similar to the discussion of (A.5) one can show

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

Hence, the expressions (A.10) and, consequently, (A.9) are $o_{\mathbb{P}}(1)$ such that

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

and the claim follows. ■

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

$$\begin{aligned}\left(G_X \sum_{t=1}^{T} X_t w_t' G_w\right)_{(i,1)} &= \frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(\frac{x_t}{T^{1/2}}\right)^i v_t \\ &\Rightarrow \int_0^1 B_v^i(r)dB_v(r) + i\Delta_{vv}\int_0^1 B_v^{i-1}(r)dr. \qquad \text{(A.11)}\end{aligned}$$

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

$$\begin{aligned}
\left(G_X \sum_{t=1}^{T} X_t w_t' G_w \right)_{(i,j)} &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left(\frac{x_t}{T^{1/2}} \right)^i \left(- \sum_{k=1}^{j} \binom{j}{k} \frac{x_t^{j-k} (-v_t)^k}{T^{(j-1)/2}} \right) \\
&= \frac{1}{T^{1/2}} \sum_{t=1}^{T} j \left(\frac{x_t}{T^{1/2}} \right)^{i+j-1} v_t \\
&\quad - \frac{1}{T^{1/2}} \sum_{t=1}^{T} \binom{j}{2} \left(\frac{x_t}{T^{1/2}} \right)^{i+j-2} \frac{v_t^2}{T^{1/2}} \\
&\quad - \frac{1}{T^{1/2}} \sum_{t=1}^{T} \sum_{k=3}^{j} \binom{j}{k} \left(\frac{x_t}{T^{1/2}} \right)^{i+j-k} \frac{(-v_t)^k}{T^{(k-1)/2}}. \qquad \text{(A.12)}
\end{aligned}$$

The first term on the right-hand side converges similarly to (A.11) to

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

For the second term in (A.12) we write $v_t^2 = \Sigma_{vv} + (v_t^2 - \Sigma_{vv})$ and consider both terms separately. First,

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

Second, using Lemma 5 it holds for the remaining term that

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

All additional terms in (A.12) converge to zero being $O_{\mathbb{P}}(T^{-(k-2)/2})$. The result for $G_D \sum_{t=1}^{T} D_t w_t' G_w$ follows analogously. ∎

Proof of Theorem 2. Beforehand, note that we can use the decomposition $\Omega_{\tilde{w}\tilde{w}} = \Omega_{vv}\Pi_v$ with

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

$\mathcal{B}$ and $\tilde{\mathcal{B}}$ defined in (1.14) and (1.15), respectively. From Theorem 1 we know, that $\hat{\Omega}_{\tilde{w}\tilde{w}} \Rightarrow \Omega_{vv}\Pi_v$ and $\hat{\Omega}_{\tilde{w}u} \Rightarrow \Omega_{vu}\Pi_v e_1^p$. Therefore, it follows $\hat{\Omega}_{\tilde{w}\tilde{w}}^{-1}\hat{\Omega}_{\tilde{w}u} \xrightarrow{\mathbb{P}} \Omega_{vv}^{-1}\Omega_{vu}e_1^p$. In (1.12) we have noted that

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

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

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

with $M = [1, \mathcal{B}']'$. From Theorem 1 it follows immediately that A^* and A^{**} have the same limiting distribution, i. e.,

$$A^* \Rightarrow \Delta_{vu}^+\begin{pmatrix}0_{q\times 1}\\ M\end{pmatrix} \text{ and } A^{**} \Rightarrow \Delta_{vu}^+\begin{pmatrix}0_{q\times 1}\\ M\end{pmatrix}.$$

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

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

Therefore, we arrive at:

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

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

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

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

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

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

A.2. A Brief Discussion of the Main Arguments in Case of More Than One Integrated Regressor

In this section we present the main changes for the multiple integrated regressors case, i. e., we consider a cointegrating polynomial regression including integer powers of $I(1)$ regressors x_{jt}, $j = 1, \dots, m$, up to degree p:

$$y_t = D_t'\delta + x_t'\beta + \sum_{j=1}^{m} X_{jt}'\beta_{X_j} + u_t, \quad \text{for } t = 1, \dots, T, \tag{A.13}$$
$$x_t = x_{t-1} + v_t,$$

where y_t is a scalar process, $D_t \in \mathbb{R}^q$, $x_t := [x_{1t}, \dots, x_{mt}]'$, $X_{jt} := [x_{jt}^2, \dots, x_{jt}^p]'$ and $\theta := [\delta', \beta', \beta_{X_1}', \dots, \beta_{X_m}']' \in \mathbb{R}^{(q+mp)}$.

Remark 9. The theory allows for more general setups concerning the integrated regressors, i. e.:

1. The highest powers included need not be equal for each integrated regressor.
2. Not all consecutive powers of integrated regressors need to be included.

The assumptions concerning the error process and the regressors are similar to Assumption 2 given by:

Assumption 7. The processes $\{u_t\}_{t\in\mathbb{Z}}$ and $\{\Delta x_t\}_{t\in\mathbb{Z}} = \{v_t\}_{t\in\mathbb{Z}}$ are generated as:

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

with $\sum_{j=0}^{\infty} j|c_{uj}| < \infty$, $\sum_{j=0}^{\infty} j\|C_{vj}\| < \infty$ and $\det(C_v(1)) \neq 0$. Furthermore, we assume that the process $\{\xi_t^0\}_{t\in\mathbb{Z}} := \{[\zeta_t, \varepsilon_t']'\}_{t\in\mathbb{Z}}$ is independent and identically distributed with $\mathbb{E}(\|\xi_t^0\|^l) < \infty$ for some $l > \max(8, 4/(1-2b))$ with $0 < b < 1/3$.

As already mentioned in the main text, the condition $\det(C_v(1)) \neq 0$ excludes cointegration among the components of the vector process $\{x_t\}_{t\in\mathbb{Z}}$ in the multivariate case. Given the similar assumptions between the processes $\{u_t\}_{t\in\mathbb{Z}}$ and $\{v_t\}_{t\in\mathbb{Z}}$ compared to Assumption 2 as well as between the different components of the vector process $\{v_t\}_{t\in\mathbb{Z}}$, the results presented in Section 1.2 also hold for the case of $m > 1$ integrated regressors. While the extension of Theorem 1 becomes more complicated only from a notational point of view, there are some technical changes in the proof of the multivariate extension of Theorem 2 discussed below.

We define the multiple integrated regressors version of w_t given in (1.9):

$$w_t = \left[v_{1t}, \dots, v_{mt}, \Delta x_{1t}^2, \dots, \Delta x_{1t}^p, \dots, \Delta x_{mt}^2, \dots, \Delta x_{mt}^p\right]$$

and the "properly scaled" version $\tilde{w}_t = G_W w_t$ with

$$G_W := G_W(T) = \operatorname{diag}\left(I_m, I_m \otimes \operatorname{diag}\left(T^{-1/2}, \dots, T^{-(p-1)/2}\right)\right).$$

With the assumptions listed in place it is straightforward to state the multiple integrated regressors extension of Theorem 1.

Corollary 5. *Let the data be generated by* (A.13). *Under Assumptions 3, 4 and 7 it holds for* $\{\eta_t\}_{t\in\mathbb{Z}} = \{[u_t, \tilde{w}_t']'\}_{t\in\mathbb{Z}} \in \mathbb{R}^{(1+mp)}$ *that*

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

where

$$\Delta_{\eta\eta} := \begin{bmatrix} \Delta_{uu} & \Delta_{uv_1} & \dots & \Delta_{uv_m} & \Delta_{uv_1}\mathcal{B}_1' & \dots & \Delta_{uv_m}\mathcal{B}_m' \\ \Delta_{v_1u} & \Delta_{v_1v_1} & \dots & \Delta_{v_1v_m} & \Delta_{v_1v_1}\mathcal{B}_1' & \dots & \Delta_{v_1v_m}\mathcal{B}_m' \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Delta_{v_mu} & \Delta_{v_mv_1} & \dots & \Delta_{v_mv_m} & \Delta_{v_mv_1}\mathcal{B}_1' & \dots & \Delta_{v_mv_m}\mathcal{B}_m' \\ \Delta_{v_1u}\mathcal{B}_1 & \Delta_{v_1v_1}\mathcal{B}_1 & \dots & \Delta_{v_1v_m}\mathcal{B}_1 & \Delta_{v_1v_1}\widetilde{\mathcal{B}}_{11} & \dots & \Delta_{v_1v_m}\widetilde{\mathcal{B}}_{1m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Delta_{v_mu}\mathcal{B}_m & \Delta_{v_mv_1}\mathcal{B}_m & \dots & \Delta_{v_mv_m}\mathcal{B}_m & \Delta_{v_mv_1}\widetilde{\mathcal{B}}_{m1} & \dots & \Delta_{v_mv_m}\widetilde{\mathcal{B}}_{mm} \end{bmatrix}, \quad \text{(A.14)}$$

with

$$\mathcal{B}_i := \left[2\int_0^1 B_{v_i}(r)dr, \dots, p\int_0^1 B_{v_i}^{p-1}(r)dr\right]', \quad i = 1, \dots, m,$$

$$\left(\widetilde{\mathcal{B}}_{ij}\right)_{k,l} := (1+k)(1+l)\int_0^1 B_{v_i}^k(r) B_{v_j}^l(r)dr, \quad i,j = 1, \dots, m, \quad k,l = 1, \dots, p-1.$$

Furthermore, it holds that

$$\hat{\Sigma}_{\eta\eta} := \frac{1}{T}\sum_{t=1}^{T} \eta_t \eta_t' \Rightarrow \Sigma_{\eta\eta},$$

where $\Sigma_{\eta\eta}$ *has similar structure as* $\Delta_{\eta\eta}$ *given in* (A.14), *which leads to:*

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

For the extension of Theorem 2 slightly more complications appear due to the fact that the long-run covariance Ω_{vv} is not scalar in the multiple integrated regressors case. Therefore, it requires a more general approach in order to show that the fully modified transformations are asymptotically equivalent.

Corollary 6. *Let the data be generated by* (A.13) *with Assumptions 1 and 7 in place. Furthermore, let long-run covariance estimation be performed with Assumptions 3 and 4 in place. Then it holds for* $T \to \infty$ *that:*

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

Thus, the FM-LIN and the FM-CPR estimator have the same limiting distribution. Here, G, $J(r)$ *and* $\hat{\theta}^{++}$ *denote the multivariate extensions of the corresponding quantities defined in the main text.*

Proof of Corollary 6. Similar to the beginning of the proof of Theorem 2 the key is to show that $\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u} = e_1^p \otimes \Omega_{vv}^{-1}\Omega_{vu}$. Given that the term on the right-hand side, i. e. $\Omega_{vv}^{-1}\Omega_{vu}$, is not a scalar term in the multiple integrated regressors case, but an $(m \times 1)$-vector, modified arguments are required. Therefore, we partition the matrix of interest into the following blocks

$$\Omega_{\tilde{w}\tilde{w}} := \left[\begin{array}{ccc|ccc} \Omega_{v_1v_1} & \cdots & \Omega_{v_1v_m} & \Omega_{v_1v_1}\mathcal{B}_1' & \cdots & \Omega_{v_1v_m}\mathcal{B}_m' \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Omega_{v_mv_1} & \cdots & \Omega_{v_mv_m} & \Omega_{v_mv_1}\mathcal{B}_1' & \cdots & \Omega_{v_mv_m}\mathcal{B}_m' \\ \hline \Omega_{v_1v_1}\mathcal{B}_1 & \cdots & \Omega_{v_1v_m}\mathcal{B}_1 & \Omega_{v_1v_1}\widetilde{\mathcal{B}}_{11} & \cdots & \Omega_{v_1v_m}\widetilde{\mathcal{B}}_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Omega_{v_mv_1}\mathcal{B}_m & \cdots & \Omega_{v_mv_m}\mathcal{B}_m & \Omega_{v_mv_1}\widetilde{\mathcal{B}}_{m1} & \cdots & \Omega_{v_mv_m}\widetilde{\mathcal{B}}_{mm} \end{array} \right] = \begin{bmatrix} \Omega_{vv} & \Omega_{\mathcal{B}}' \\ \Omega_{\mathcal{B}} & \Omega_{\widetilde{\mathcal{B}}} \end{bmatrix}.$$

Using this representation we show that the term

$$\Omega_{\tilde{w}u} := \left[\Omega_{vu}', \Omega_{v_1u}\mathcal{B}_1', \ldots, \Omega_{v_mu}\mathcal{B}_m' \right]'$$

can be written as $\Omega_{\tilde{w}u} = \Omega_{\tilde{w}\tilde{w}} \left(e_1^p \otimes \Omega_{vv}^{-1}\Omega_{vu} \right)$. Considering the first m rows of $\Omega_{\tilde{w}u}$ we have

$$\begin{bmatrix} \Omega_{vv} & \Omega_{\mathcal{B}}' \end{bmatrix} \begin{bmatrix} \Omega_{vv}^{-1}\Omega_{vu} \\ 0_{m(p-1)\times 1} \end{bmatrix} = \Omega_{vu}.$$

Since for $j = 1, \ldots, m$

$$\left((e_j^{m\,\prime}\,\Omega_{vv}) \otimes \mathcal{B}_j\right)\Omega_{vv}^{-1}\Omega_{vu} = \Omega_{v_j u}\mathcal{B}_j',$$

we get for the remaining $m(p-1)$ rows

$$\begin{bmatrix}\Omega_{\mathcal{B}} & \Omega_{\widetilde{\mathcal{B}}}\end{bmatrix}\begin{bmatrix}\Omega_{vv}^{-1}\Omega_{vu} \\ 0_{m(p-1)\times 1}\end{bmatrix} = \begin{bmatrix}\Omega_{v_1 u}\mathcal{B}_1', \ldots, \Omega_{v_m u}\mathcal{B}_m'\end{bmatrix}'.$$

Thus, we have shown that $\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u} = e_1^p \otimes \Omega_{vv}^{-1}\Omega_{vu}$, which implies

$$\tilde{w}_t'\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u} = v_t'\Omega_{vv}^{-1}\Omega_{vu},$$

i. e., the first-step transformations are asymptotically identical.

For the second-step transformations we obtain by exactly the same arguments

$$\begin{aligned}\Delta_{\tilde{w}u}^{+} &= \Delta_{\tilde{w}u} - \Delta_{\tilde{w}\tilde{w}}\Omega_{\tilde{w}\tilde{w}}^{-1}\Omega_{\tilde{w}u} \\ &= \Delta_{\tilde{w}u} - \Delta_{\tilde{w}v}\Omega_{vv}^{-1}\Omega_{vu} \\ &= \begin{bmatrix}\Delta_{vu} \\ \Delta_{v_1 u}\mathcal{B}_1 \\ \vdots \\ \Delta_{v_m u}\mathcal{B}_m\end{bmatrix} - \begin{bmatrix}\Delta_{vu}\Omega_{vv}^{-1}\Omega_{vu} \\ \Delta_{v_1 u}\Omega_{vv}^{-1}\Omega_{vu}\mathcal{B}_1 \\ \vdots \\ \Delta_{v_m u}\Omega_{vv}^{-1}\Omega_{vu}\mathcal{B}_m\end{bmatrix},\end{aligned}$$

which coincides with the corresponding expression M for the FM-CPR estimator given in the proof of Proposition 1 in Wagner and Hong (2016). ■

B. Appendix to Chapter 2

B.1. Proofs

Proof of Proposition 1. We examine the asymptotic behavior of the elements of the last term in (2.14). We begin with the term $T^{-1/2}A_{IM}S^{\xi}_{[rT]}$ for $T \to \infty$,

$$\begin{pmatrix} T^{-1}\sum_{t=1}^{[rT]}\sqrt{T}G_D D_t \\ T^{-1}\sum_{t=1}^{[rT]}\sqrt{T}G_X X_t \\ T^{-1/2}x_{[rT]} \end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^r D(s)ds \\ \int_0^r \mathbf{B}_{v_m}(s)ds \\ B_v(r) \end{pmatrix} = f(r),$$

here the convergence in the second row holds because of (2.8). This result leads to

$$\left(T^{-2}A_{IM}S^{\xi\prime}S^{\xi}A_{IM}\right)^{-1} = \left(\frac{1}{T}\left(T^{-1/2}A_{IM}S^{\xi\prime}\right)\left(T^{-1/2}A_{IM}S^{\xi}\right)\right)^{-1} \Rightarrow \left(\int f(s)f(s)'ds\right)^{-1}. \tag{B.1}$$

For the second factor in (2.14) we use

$$T^{-1/2}A_{IM}S^{\xi\prime}_{[rT]}T^{-1/2}S^{u}_{[rT]} \Rightarrow f(r)B_u(r)$$

such that

$$T^{-2}A_{IM}S^{\xi\prime}S^{u} \Rightarrow \int f(s)B_u(s)ds = \omega_{u\cdot v}^{1/2}\int f(s)w_{u\cdot v}ds + \int f(s)W_v(s)'\,\Omega_{vv}^{-1/2}\Omega_{vu}ds, \tag{B.2}$$

using $B_u(r) = \omega_{u\cdot v}^{1/2} w_{u\cdot v} + \Omega_{uv}(\Omega_{vv}^{-1/2})' W_v(r)$. Multiplying (B.1) and the second term of (B.2) leads to

$$\begin{aligned} & \left(\int f(s)f(s)'ds\right)^{-1} \int f(s)W_v(s)'ds\, \Omega_{vv}^{-1/2}\Omega_{vu} \\ = & \left(\int f(s)f(s)'ds\right)^{-1} \int f(s)B_v(s)'ds\, \Omega_{vv}^{-1}\Omega_{vu} \\ = & \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1}\Omega_{vu} \end{pmatrix}, \end{aligned}$$

note that $\left(\int f(s)f(s)'ds\right)^{-1} \int f(s)B_v(s)'ds = [0, 0, I_m]'$, since $B_v(r)$ is the last block-component in $f(r)$. Similarly equation (2.15) follows using integration by parts. The expression for the (conditional) covariance matrix (2.16) holds, because the quadratic variation process of a standard Brownian motion $w_{u\cdot v}$ is given by $[w_{u\cdot v}, w_{u\cdot v}]_s = s$. ■

Proof of Corollary 3. In case of full design simply rewrite $f(r)$ as

$$f(r) = \begin{pmatrix} \int_0^r D(s)ds \\ \int_0^r \mathbf{B}_v(s)ds \\ B_v(r) \end{pmatrix} = \begin{pmatrix} \int_0^r D(s)ds \\ F(\Omega_{vv}) \int_0^r \mathbf{W}_v(s)ds \\ \Omega_{vv}^{1/2} W_v(r) \end{pmatrix} = \Pi g(r).$$

■

Proof of Lemma 2. For part (i) we can use the results already established in Proposition 1 and Corollary 3, so that we only have to focus on the additional regressors $z_t = [z_t^{D\prime}, z_t^{S^X\prime}, z_t^{x\prime}]'$. For the limit of z_t^D, z_t^x and the regressors of $z_t^{S^X\prime}$, which do not contain powers, we can one-to-one follow the arguments of Vogelsang and Wagner (2014a) given in the proof of Lemma 1. For the limit of the non-linear parts we define $S_t^{x_j^k} := \sum_{i=1}^t x_{ji}^k$ for $k = 1, \ldots, p$ and $z_t^{S^{x_j^k}}$ as the corresponding part in z_t, then scaled by $T^{-1/2} A_{IM}$ we get

$$\begin{aligned} T^{-5/2} T^{-(k+1)/2} z_{[rT]}^{S^{x_j^k}} &= T^{-5/2} T^{-(k+1)/2} [rT] \sum_{t=1}^T S_t^{x_j^k} - T^{-5/2} T^{-(k+1)/2} \sum_{t=1}^{[rT]} \sum_{l=1}^t S_l^{x_j^k} \\ &= \frac{[rT]}{T} \frac{1}{T} \sum_{t=1}^T T^{-(k+2)/2} S_t^{x_j^k} - \frac{1}{T} \sum_{t=1}^{[rT]} \frac{1}{T} \sum_{l=1}^t S_l^{x_j^k} \\ &\Rightarrow r \int_0^1 \left(\int_0^m B_{v_j}^k(s)ds \right) dm - \int_0^r \left(\int_0^m \left(\int_0^n B_{v_j}^k(s)ds \right) dn \right) dm. \end{aligned}$$

Combining the single parts leads to the asymptotic behavior in (i). Note that the adjusted residuals $\tilde{S}_t^{u*}$ defined in (2.24) coincide with the OLS residuals from the regression

$$S_t^y = S_t^{D\prime}\delta^* + S_t^{X\prime}\beta^* + x_t'\gamma^* + z_t'\kappa^* + S_t^u, \tag{B.3}$$

which follows immediately using standard projection arguments.

For part (ii) we consider the OLS residuals from (B.3),

$$\tilde{S}_t^{u*} = S_t^y - S_t^{\xi*\prime}\tilde{\theta}^* = S_t^u - x_t'\Omega_{vv}^{-1}\Omega_{vu} - S_t^{\xi*\prime}\left(\tilde{\theta}^* - \theta^*\right),$$

with $S_t^{\xi*} := [S_t^{\xi\prime}, z_t']'$. Defining $\xi_t^* := [\xi_t', z_t']'$ we get for the scaled partial sum of the first differences

$$\begin{aligned}
&T^{-1/2}\sum_{t=2}^{[rT]}\Delta\tilde{S}_t^{u*}\\
&= T^{-1/2}\sum_{t=2}^{[rT]}u_t - \Delta x_{[rT]}'\Omega_{vv}^{-1}\Omega_{vu} - T^{-1/2}\sum_{t=2}^{[rT]}\xi_t^{*\prime}\begin{pmatrix}A_{IM} & 0\\ 0 & T^{-2}A_{IM}\end{pmatrix}\begin{pmatrix}A_{IM} & 0\\ 0 & T^{-2}A_{IM}\end{pmatrix}^{-1}\left(\tilde{\theta}^* - \theta^*\right)\\
&\Rightarrow \omega_{u\cdot v}^{1/2}\left(\int_0^r dw_{u\cdot v}(s) - h(r)'\left(\int_0^1 h(s)h(s)'ds\right)^{-1}\int_0^1 [H(1)-H(s)]dw_{u\cdot v}(s)\right) = \omega_{u\cdot v}^{1/2}\widetilde{P}^*(r)
\end{aligned}$$

Finally, we have to show independence of $\widetilde{P}^*(r)$ and the limiting distribution in (2.15) conditional on $W_v(r)$ and since both processes are Gaussian, it suffices to show conditional uncorrelation between $\widetilde{P}^*(r)$ and the relevant quantity in (2.15), namely $\int[G(1) - G(s)]dw_{u\cdot v}(s)$.

First note that integration by parts leads to

$$\int_0^1 [H(1)-H(s)][G(1)-G(s)]'ds = \underbrace{[H(1)-H(s)]\,h_2(s)'\Big|_0^1}_{=0} + \int_0^1 h(s)h_2(s)'ds,$$

where $h_2(\cdot)$ is the second block of $h(\cdot)$. Now it follows that

$$\left(\int_0^1 h(s)h(s)'ds\right)^{-1}\int_0^1 [H(1)-H(s)][G(1)-G(s)]'ds = \begin{pmatrix}0\\ I\end{pmatrix}.$$

Using the previous two results and again the fact that $[w_{u\cdot v}, w_{u\cdot v}]_s = s$, leads to

$$
\begin{aligned}
& Cov\left(\widetilde{P}^*(r), \int [G(1) - G(s)] dw_{u\cdot v}(s)\right) \\
&= \int_0^r [G(1) - G(s)]' ds - h(r)' \left(\int_0^1 h(s)h(s)' ds\right)^{-1} \int_0^1 [H(1) - H(s)][G(1) - G(s)]' ds \\
&= \int_0^r [G(1) - G(s)]' ds - \int_0^r [G(1) - G(s)]' ds \\
&= 0.
\end{aligned}
$$

■

Proof of Proposition 2. First, we have to make sure using standard calculations that the expression given in (2.22) (up to $\breve{\omega}_{u\cdot v}$) converges to (2.16) (up to $\omega_{u\cdot v}$).
The assumption given in (2.21) and the result from Proposition 1 imply that under the null hypothesis

$$
\tilde{W}^* \Rightarrow (R^*\Phi(V_{IM})' \left(Q_b(\widetilde{P^*}, \widetilde{P^*}) R^* V_{IM} R^{*\prime}\right)^{-1} (R^*\Phi(V_{IM})) \sim \frac{\chi_q^2}{Q_b(\widetilde{P^*}, \widetilde{P^*})},
$$

where it follows from Vogelsang and Wagner (2014a) Theorem 1 that the fixed-b limit of $\tilde{\omega}^*_{u\cdot v}$ is given by $Q_b(\widetilde{P^*}, \widetilde{P^*})$ and therefore $\tilde{V}^* \Rightarrow Q_b(\widetilde{P^*}, \widetilde{P^*}) V_{IM}$. Independence of χ_q^2 and $Q_b(\widetilde{P^*}, \widetilde{P^*})$ conditional on $W_v(r)$ follows from Lemma 2. Given the fact that the numerator is χ_q^2-distributed and independent of $W_v(r)$, this implies also unconditional independence. ■

Proof of Proposition 3. Using the result from Corollary 4 we obtain under the null hypothesis

$$
\begin{aligned}
CT_{IM} &= \frac{1}{T^2 \hat{\omega}_{u\cdot v}} \sum_{t=2}^{T} \left(\sum_{i=2}^{t} \Delta \tilde{S}_i^u \right)^2 \\
&= \frac{1}{T \hat{\omega}_{u\cdot v}} \sum_{t=2}^{T} \left(\frac{1}{\sqrt{T}} \sum_{i=2}^{t} \Delta \tilde{S}_i^u \right)^2 \\
&\Rightarrow \frac{1}{\omega_{u\cdot v}} \int_0^1 \left(\omega_{u\cdot v}^{1/2} \widetilde{P}(r) \right)^2 dr \\
&= \int_0^1 \left(\widetilde{P}(r) \right)^2 dr.
\end{aligned}
$$

■

B.2. EKC Analysis: Estimation Results and Figures

	PP	PP(fb)–One Step	PP(fb)–Two Step
Australia	-1.280	-1.304	-1.378
Austria	-1.908	-1.813	-1.878
Belgium	-1.419	-1.554	-1.587
Canada	-2.496	*-3.057*	*-3.056*
Denmark	-2.293	-2.270	-2.273
Finland	-2.321	-2.422	-2.412
France	-1.958	-2.204	-2.214
Germany	-2.356	-2.582	-2.598
Italy	-1.665	-1.825	-1.835
Japan	-1.719	-1.880	-1.893
Netherlands	-2.724	-2.310	-2.334
New Zealand	-2.606	-2.686	-2.688
Norway	-2.142	-2.117	-2.179
Portugal	-1.872	-1.879	-1.869
Spain	-1.005	-1.192	-1.235
Sweden	-2.339	-2.401	-2.407
Switzerland	-2.807	-2.449	-2.481
United Kingdom	-1.567	-1.418	-1.625
United States	-2.981	*-2.911*	*-2.915*

Table B.1.: Standard Phillips-Perron unit-root test (PP) and fixed-b Phillips-Perron unit-root test (PP(fb)) of Vogelsang and Wagner (2013) results with one-step and two-step detrending. Intercept and linear trend for per capita GDP, Bartlett kernel, bandwidth choice according to Newey and West (1994). Per capita GDP is measured in (international) GK-\$. All variables are transformed to logarithms. *Italic* entries denote rejection of the null hypothesis at the 10% level and **bold** entries indicate rejection at the 5% level.

	Quadratic			Cubic		
	CT_{IM}	CT_{FM}	$P_{\hat{u}}$	CT_{IM}	CT_{FM}	$P_{\hat{u}}$
Australia	**0.061**	**0.108**	11.330	**0.052**	**0.107**	11.404
Austria	0.035	0.056	**55.997**	0.024	0.042	**56.830**
Belgium	0.037	0.062	*50.269*	0.036	0.066	*54.423*
Canada	**0.059**	**0.145**	12.420	0.029	0.057	26.109
Denmark	0.022	0.052	40.613	0.022	0.051	40.619
Finland	0.029	0.050	**75.016**	0.026	0.035	**83.088**
France	0.038	0.066	28.847	0.032	0.061	28.916
Germany	**0.056**	**0.111**	**68.343**	*0.039*	*0.090*	**68.549**
Italy	**0.055**	**0.146**	34.141	0.033	*0.095*	*51.661*
Japan	0.042	**0.152**	8.127	0.022	0.066	12.222
Netherlands	0.040	0.074	**96.172**	*0.040*	0.075	**96.183**
New Zealand	0.043	**0.115**	13.337	0.029	*0.100*	14.071
Norway	**0.059**	*0.095*	20.644	**0.052**	*0.093*	20.967
Portugal	0.032	**0.111**	19.959	0.034	**0.113**	20.357
Spain	*0.047*	0.086	42.578	**0.047**	*0.082*	42.790
Sweden	0.044	0.085	28.679	0.035	*0.085*	29.661
Switzerland	0.026	0.084	**85.979**	0.032	0.057	**105.175**
UK	*0.051*	0.073	**97.169**	**0.049**	0.070	**98.034**
US	**0.089**	**0.156**	13.920	**0.052**	0.079	27.017
Critical values ($\alpha = 10\%$)	0.045	0.086	45.237	0.039	0.081	47.925
Critical values ($\alpha = 5\%$)	0.054	0.106	52.952	0.046	0.101	55.926

Table B.2.: Results for the cointegration tests using both the IM-CPR and FM-CPR residuals as well as the OLS residual based $P_{\hat{u}}$ non-cointegration test for the quadratic and cubic specification in conjunction with the Newey and West (1994) data dependent bandwidth rule and the Bartlett kernel. *Italic* numbers denote rejection of the null hypothesis at the 10% level and **bold** numbers indicate rejection at the 5% level.

	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	TP	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	TP
	Austria				**Belgium**			
OLS	-0.014	6.112	-0.277	61982.813	-0.005	11.772	-0.619	13460.033
(t-values)	**-2.461**	**2.344**	**-1.990**		**-3.815**	**12.121**	**-12.194**	
D-CPR	-0.015	6.263	-0.284	61786.430	0.000	8.787	-0.471	11161.979
(t-values)	**-3.050**	**2.203**	*-1.821*		0.197	**6.597**	**-6.757**	
FM-CPR	-0.017	6.247	-0.277	78058.655	-0.004	11.358	-0.599	13142.216
(t-values)	**-3.681**	**2.488**	**-2.001**		**-2.694**	**10.000**	**-10.038**	
IM-CPR	-0.022	9.316	-0.440	39752.329	-0.002	9.380	-0.499	12161.303
(t-values)	**-3.922**	**2.714**	**-2.355**		-0.766	**6.242**	**-6.352**	
(fixed-b t-values)	**-4.692**	**3.247**	**-2.818**		-1.497	**12.200**	**-12.415**	
	Finland				**Netherlands**			
OLS	-0.033	16.502	-0.780	39286.848	0.002	8.970	-0.456	18503.069
(t-values)	**-2.896**	**10.638**	**-11.218**		0.981	**8.726**	**-8.257**	
D-CPR	-0.021	13.779	-0.654	37831.903	-0.003	10.277	-0.515	21471.698
(t-values)	*-1.646*	**4.611**	**-4.375**		-1.288	**7.367**	**-7.113**	
FM-CPR	-0.029	15.610	-0.737	39523.015	0.001	9.437	-0.481	18280.060
(t-values)	**-3.231**	**9.271**	**-8.717**		0.605	**8.305**	**-7.949**	
IM-CPR	-0.021	14.004	-0.662	39213.221	-0.001	9.156	-0.459	21560.316
(t-values)	*-1.758*	**6.320**	**-6.099**		-0.468	**5.886**	**-5.653**	
(fixed-b t-values)	**-2.631**	**9.459**	**-9.127**		-0.705	**8.860**	**-8.509**	
	Switzerland				**UK**			
OLS	-0.020	7.685	-0.283	7.9×10^5	-0.008	8.419	-0.436	15656.196
(t-values)	**-2.978**	**5.401**	**-3.321**		**-3.744**	**6.235**	**-6.389**	
D-CPR	-0.022	7.676	-0.276	1.1×10^6	-0.007	7.655	-0.397	15492.766
(t-values)	**-5.570**	**6.152**	**-4.041**		**-2.282**	**4.754**	**-5.005**	
FM-CPR	-0.024	7.755	-0.273	1.5×10^6	-0.008	8.657	-0.446	16286.963
(t-values)	**-6.229**	**6.122**	**-3.910**		**-3.454**	**6.624**	**-6.889**	
IM-CPR	-0.025	10.108	-0.393	3.8×10^5	-0.012	10.288	-0.522	19188.954
(t-values)	**-5.562**	**6.483**	**-4.706**		**-3.347**	**5.076**	**-5.246**	
(fixed-b t-values)	**-7.279**	**8.484**	**-6.159**		**-4.847**	**7.351**	**-7.597**	

Table B.3.: Estimation results for equation (2.35) with Bartlett kernel and data dependent bandwidth rule according to Newey and West (1994). The turning points are computed as $\exp(-\hat{\beta}_1/(2\hat{\beta}_2))$. **Bold** t-values indicate significance at the 5% level and *italic* t-values significance at the 10% level.

	$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	Turning Points		$\hat{\delta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	Turning Points	
			Austria						**Belgium**			
OLS	-0.015	39.500	-4.087	0.144			-0.007	55.533	-5.454	0.178	15389.624	48471.738
(t-values)	**-2.292**	0.915	-0.856	0.821			**-3.864**	**2.620**	**-2.350**	**2.099**		
D-CPR	-0.015	42.643	-4.412	0.156			-0.001	21.116	-1.823	0.049	11446.615	4.1×10^6
(t-values)	**-3.259**	0.966	-0.882	0.825			-0.223	0.762	-0.600	0.445		
FM-CPR	-0.018	69.209	-7.450	0.271			-0.007	59.293	-5.873	0.194	15693.314	39002.980
(t-values)	**-4.114**	*1.684*	-1.591	1.531			**-4.619**	**3.257**	**-2.925**	**2.620**		
IM-CPR	-0.027	119.121	-12.757	0.460			-0.006	74.792	-7.623	0.259	16119.171	20691.362
(t-values)	**-4.525**	**2.043**	*-1.937*	*1.857*			**-2.578**	**2.705**	**-2.516**	**2.337**		
(fixed-b t-values)	**-6.108**	**2.758**	**-2.615**	*2.507*			**-3.111**	**3.264**	**-3.036**	**2.820**		
			Finland						**Germany**			
OLS	-0.051	71.950	-7.144	0.245			0.001	18.965	-1.575	0.039	10044.038	3.5×10^7
(t-values)	**-5.386**	**4.238**	**-3.692**	**3.314**			0.302	0.955	-0.701	0.468		
D-CPR	-0.031	33.232	-2.848	0.084			-0.010	87.088	-9.289	0.332		
(t-values)	-1.193	0.762	-0.580	0.447			*-1.797*	**2.383**	**-2.232**	**2.097**		
FM-CPR	-0.057	83.931	-8.513	0.298			-0.002	40.166	-3.986	0.131	11939.460	52774.774
(t-values)	**-5.795**	**4.034**	**-3.571**	**3.247**			-0.794	1.571	-1.365	1.181		
IM-CPR	-0.067	104.755	-10.884	0.389			-0.013	74.502	-7.692	0.266		
(t-values)	**-4.470**	**3.834**	**-3.507**	**3.273**			**-2.838**	**2.225**	**-2.023**	*1.848*		
(fixed-b t-values)	**-5.789**	**4.965**	**-4.541**	**4.238**			**-5.347**	**4.192**	**-3.812**	**3.480**		
			Netherlands						**Switzerland**			
OLS	0.002	7.422	-0.285	-0.006	0.000	18325.835	-0.022	-87.695	10.309	-0.390	1320.349	33655.189
(t-values)	1.084	0.353	-0.122	-0.074			**-3.566**	**-4.448**	**4.754**	**-4.949**		
D-CPR	-0.002	-8.030	1.490	-0.073	40.653	19335.092	-0.024	-95.129	11.136	-0.420	1389.356	33605.692
(t-values)	-0.822	-0.342	0.579	-0.780			**-7.215**	**-3.825**	**4.036**	**-4.137**		
FM-CPR	0.001	7.504	-0.267	-0.008	0.000	18065.136	-0.026	-93.449	10.957	-0.414	1337.358	34951.823
(t-values)	0.621	0.330	-0.106	-0.085			**-7.925**	**-3.750**	**3.965**	**-4.066**		
IM-CPR	-0.002	23.099	-1.982	0.056	26774.200	8.0×10^5	-0.026	-23.636	3.328	-0.136	187.548	62624.781
(t-values)	-0.742	0.674	-0.525	0.401			**-6.892**	-0.682	0.870	-0.971		
(fixed-b t-values)	-1.500	1.362	-1.061	0.810			**-8.538**	-0.845	1.077	-1.203		
			UK									
OLS	-0.008	21.422	-1.851	0.051	16869.623	1.6×10^6						
(t-values)	**-3.135**	1.091	-0.882	0.685								
D-CPR	-0.008	17.580	-1.470	0.039	16607.226	5.6×10^6						
(t-values)	**-2.154**	0.736	-0.570	0.416								
FM-CPR	-0.010	27.825	-2.521	0.075	19026.062	2.8×10^5						
(t-values)	**-3.617**	1.308	-1.091	0.894								
IM-CPR	-0.015	34.271	-3.077	0.091	25969.237	2.2×10^5						
(t-values)	**-3.632**	1.305	-1.086	0.892								
(fixed-b t-values)	**-5.378**	1.932	-1.608	1.320								

Table B.4.: Estimation results for equation (2.36) with Bartlett kernel and data dependent bandwidth rule according to Newey and West (1994). The turning points are computed as $\exp\left(-\frac{\hat{\beta}_2}{3\hat{\beta}_3} \pm \sqrt{\frac{\hat{\beta}_2^2-3\hat{\beta}_1\hat{\beta}_3}{9\hat{\beta}_3^2}}\right)$ in case that $\frac{\hat{\beta}_2^2-3\hat{\beta}_1\hat{\beta}_3}{9\hat{\beta}_3^2} \geq 0$. **Bold** t-values indicate significance at the 5% level and *italic* t-values significance at the 10% level.

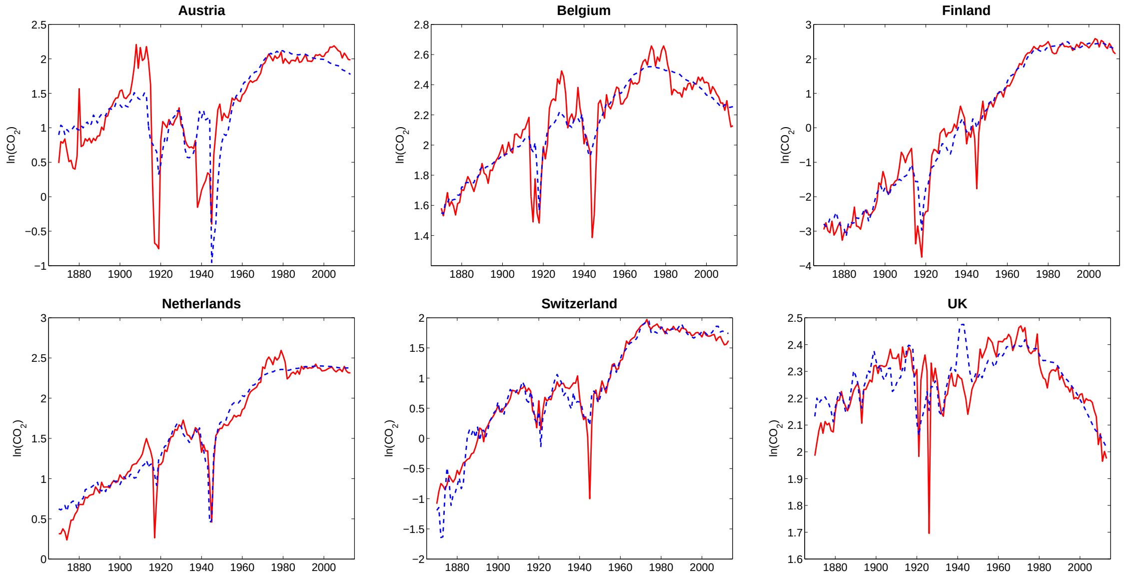

Figure B.1.: Actual values (solid) and fitted values (dashed) of log per capita CO_2 emissions estimating the *quadratic* EKC regression equation with IM-CPR.

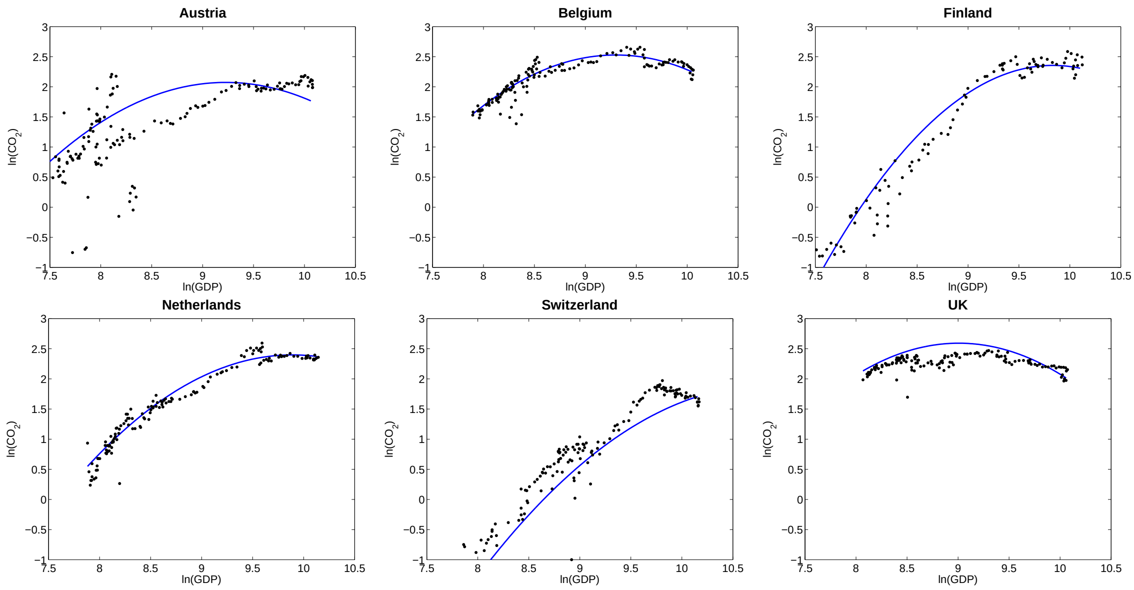

Figure B.2.: EKC estimation for CO_2 using coefficient estimates obtained by IM-CPR in the *quadratic* EKC regression equation.

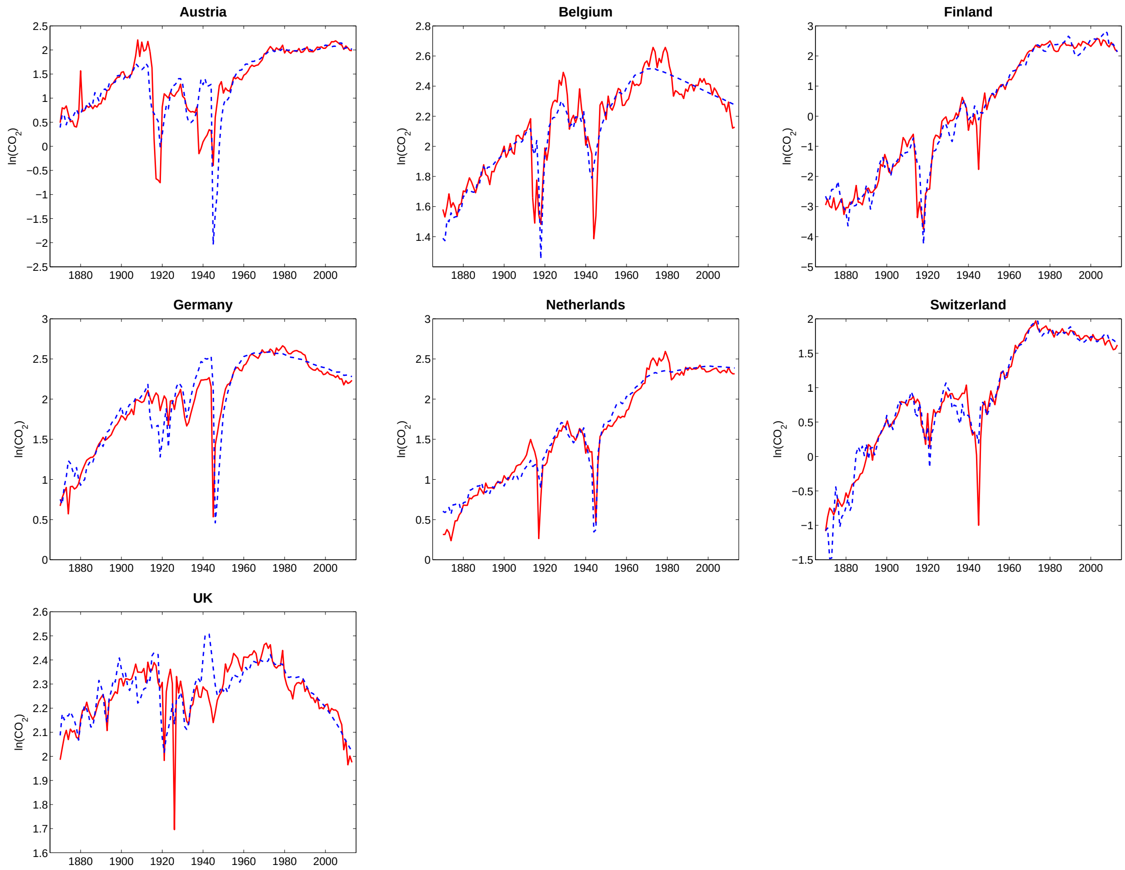

Figure B.3.: Actual values (solid) and fitted values (dashed) of log per capita CO_2 emissions estimating the *cubic* EKC regression equation with IM-CPR.

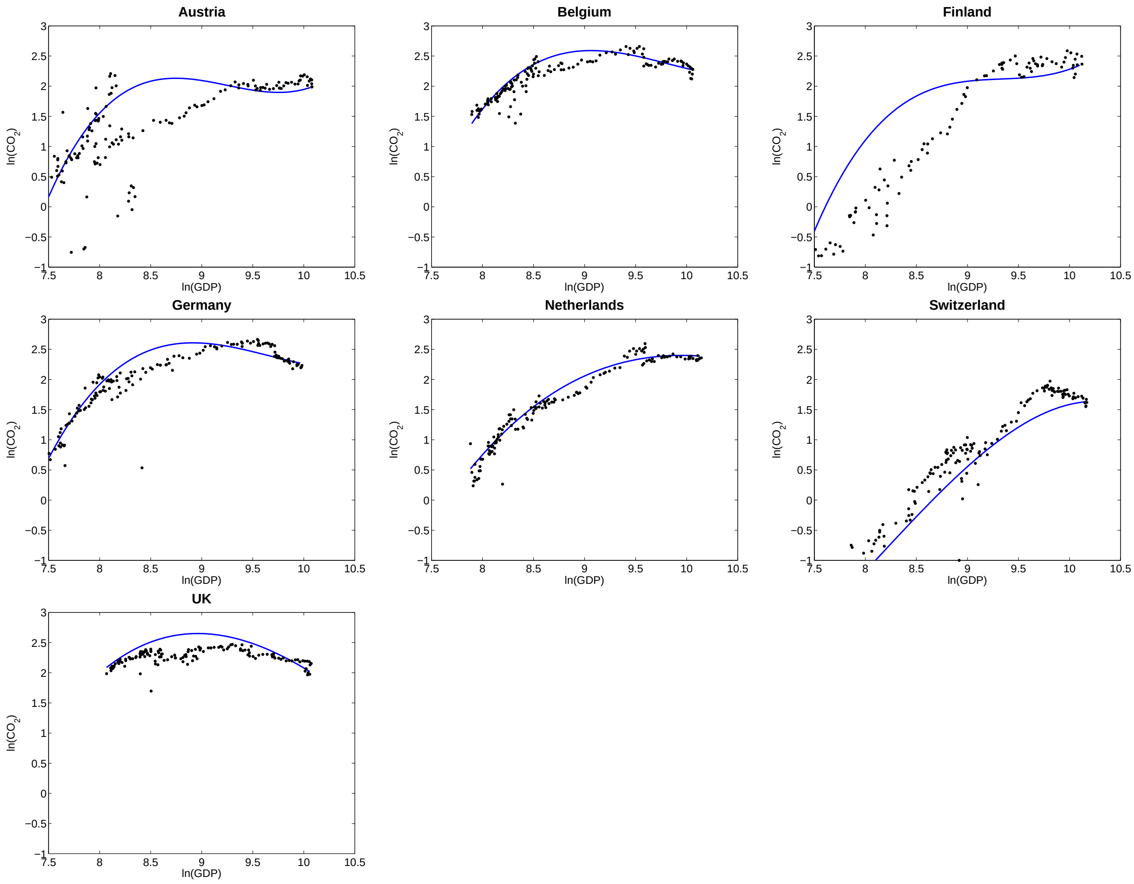

Figure B.4.: EKC estimation for CO_2 using coefficient estimates obtained by IM-CPR in the *cubic* EKC regression equation.

B.3. Critical Values for the CT_{IM} Test

	0.005	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990	0.995
						$m = 1, p = 1$					
	0.0167	0.0187	0.0221	0.0259	0.0316	0.0726	0.2435	0.3767	0.5472	0.8365	1.0940
c	0.0128	0.0141	0.0164	0.0187	0.0219	0.0410	0.0871	0.1102	0.1371	0.1784	0.2156
c,t	0.0108	0.0118	0.0135	0.0151	0.0174	0.0300	0.0563	0.0684	0.0814	0.0988	0.1135
						$m = 1, p = 2$					
	0.0136	0.0150	0.0175	0.0200	0.0238	0.0482	0.1202	0.1642	0.2185	0.3108	0.3989
c	0.0112	0.0122	0.0140	0.0158	0.0182	0.0327	0.0662	0.0828	0.1016	0.1314	0.1569
c,t	0.0096	0.0104	0.0118	0.0132	0.0150	0.0249	0.0450	0.0540	0.0635	0.0766	0.0875
						$m = 1, p = 3$					
	0.0118	0.0129	0.0149	0.0168	0.0197	0.0370	0.0819	0.1071	0.1360	0.1840	0.2250
c	0.0101	0.0110	0.0125	0.0140	0.0161	0.0280	0.0550	0.0684	0.0834	0.1064	0.1269
c,t	0.0088	0.0095	0.0107	0.0119	0.0134	0.0218	0.0386	0.0460	0.0538	0.0654	0.0741
						$m = 1, p = 4$					
	0.0106	0.0115	0.0131	0.0148	0.0171	0.0306	0.0632	0.0800	0.0992	0.1312	0.1570
c	0.0093	0.0101	0.0114	0.0127	0.0145	0.0248	0.0477	0.0589	0.0708	0.0889	0.1059
c,t	0.0082	0.0088	0.0099	0.0109	0.0123	0.0197	0.0343	0.0409	0.0475	0.0574	0.0650
						$m = 2, p = 1$					
	0.0122	0.0133	0.0154	0.0175	0.0204	0.0386	0.0897	0.1217	0.1645	0.2381	0.3085
c	0.0102	0.0112	0.0126	0.0141	0.0162	0.0275	0.0516	0.0627	0.0747	0.0926	0.1088
c,t	0.0091	0.0098	0.0109	0.0121	0.0137	0.0223	0.0390	0.0467	0.0545	0.0650	0.0742
						$m = 2, p = 2$					
	0.0105	0.0115	0.0131	0.0148	0.0172	0.0306	0.0637	0.0814	0.1028	0.1396	0.1724
c	0.0092	0.0099	0.0112	0.0124	0.0141	0.0233	0.0419	0.0505	0.0597	0.0729	0.0846
c,t	0.0082	0.0088	0.0098	0.0108	0.0122	0.0193	0.0328	0.0387	0.0449	0.0530	0.0596
						$m = 2, p = 3$					
	0.0094	0.0103	0.0116	0.0130	0.0149	0.0255	0.0493	0.0612	0.0749	0.0966	0.1167
c	0.0084	0.0090	0.0101	0.0112	0.0127	0.0206	0.0362	0.0433	0.0510	0.0612	0.0706
c,t	0.0076	0.0081	0.0090	0.0099	0.0111	0.0173	0.0288	0.0337	0.0390	0.0461	0.0515
						$m = 2, p = 4$					
	0.0086	0.0093	0.0105	0.0117	0.0134	0.0222	0.0411	0.0502	0.0603	0.0757	0.0900
c	0.0078	0.0084	0.0094	0.0104	0.0117	0.0186	0.0323	0.0384	0.0451	0.0546	0.0620
c,t	0.0071	0.0076	0.0084	0.0092	0.0103	0.0158	0.0260	0.0303	0.0348	0.0413	0.0464
						$m = 3, p = 1$					
	0.0097	0.0106	0.0120	0.0135	0.0154	0.0263	0.0513	0.0644	0.0791	0.1048	0.1291
c	0.0085	0.0092	0.0103	0.0115	0.0129	0.0208	0.0359	0.0426	0.0498	0.0599	0.0685
c,t	0.0077	0.0083	0.0092	0.0102	0.0114	0.0177	0.0291	0.0341	0.0393	0.0467	0.0526
						$m = 3, p = 2$					
	0.0089	0.0095	0.0107	0.0119	0.0135	0.0222	0.0409	0.0501	0.0608	0.0768	0.0926
c	0.0078	0.0084	0.0094	0.0103	0.0116	0.0181	0.0303	0.0358	0.0414	0.0499	0.0565
c,t	0.0071	0.0076	0.0085	0.0093	0.0103	0.0157	0.0252	0.0292	0.0335	0.0395	0.0444
						$m = 3, p = 3$					
	0.0081	0.0087	0.0097	0.0107	0.0121	0.0194	0.0343	0.0412	0.0491	0.0603	0.0711
c	0.0073	0.0078	0.0086	0.0095	0.0106	0.0163	0.0269	0.0316	0.0363	0.0433	0.0491
c,t	0.0067	0.0071	0.0079	0.0086	0.0095	0.0142	0.0226	0.0261	0.0298	0.0348	0.0390
						$m = 3, p = 4$					
	0.0075	0.0080	0.0090	0.0099	0.0111	0.0174	0.0299	0.0355	0.0415	0.0502	0.0579
c	0.0069	0.0073	0.0081	0.0089	0.0099	0.0150	0.0244	0.0286	0.0328	0.0389	0.0438
c,t	0.0063	0.0067	0.0074	0.0081	0.0089	0.0132	0.0207	0.0238	0.0271	0.0317	0.0352
						$m = 4, p = 1$					
	0.0082	0.0089	0.0100	0.0110	0.0125	0.0200	0.0353	0.0424	0.0503	0.0627	0.0742
c	0.0074	0.0080	0.0088	0.0097	0.0108	0.0166	0.0270	0.0315	0.0362	0.0429	0.0480
c,t	0.0068	0.0073	0.0081	0.0088	0.0098	0.0146	0.0230	0.0265	0.0301	0.0353	0.0394
						$m = 4, p = 2$					
	0.0076	0.0081	0.0091	0.0100	0.0112	0.0175	0.0297	0.0352	0.0411	0.0506	0.0587
c	0.0069	0.0074	0.0081	0.0089	0.0099	0.0148	0.0236	0.0273	0.0310	0.0366	0.0409
c,t	0.0064	0.0068	0.0075	0.0081	0.0090	0.0132	0.0204	0.0233	0.0264	0.0305	0.0338
						$m = 4, p = 3$					
	0.0070	0.0075	0.0084	0.0092	0.0102	0.0156	0.0259	0.0304	0.0353	0.0424	0.0486
c	0.0065	0.0069	0.0076	0.0083	0.0092	0.0135	0.0212	0.0245	0.0278	0.0325	0.0363
c,t	0.0060	0.0064	0.0070	0.0076	0.0084	0.0121	0.0185	0.0211	0.0238	0.0276	0.0303
						$m = 4, p = 4$					
	0.0066	0.0071	0.0078	0.0085	0.0095	0.0143	0.0231	0.0270	0.0310	0.0366	0.0417
c	0.0061	0.0065	0.0072	0.0078	0.0086	0.0126	0.0195	0.0223	0.0254	0.0295	0.0328
c,t	0.0057	0.0060	0.0066	0.0072	0.0079	0.0113	0.0171	0.0194	0.0219	0.0252	0.0277

Table B.5.: Critical values for the IM-OLS residual based CT test for the case of only one regressor entering the CPR with powers. The symbols in the first column indicate the deterministic component: none (empty), intercept only (c) and intercept and linear trend (c, t), m indicates the number of integrated regressors and p indicates the highest power included of the regressor entering with powers.

B.4. Additional Simulation Results

$\rho_1 = \rho_2$	OLS	D-CPR	FM-CPR			IM-CPR
			And91	NW	NW_T	
			Bias ($\times 1000$)			
T=100						
0.0	0.084	0.845	0.098	0.096	0.097	0.022
0.3	0.087	0.983	0.129	0.113	0.128	0.044
0.6	0.097	1.294	0.207	0.188	0.180	0.110
0.9	0.269	1.894	0.474	0.424	0.368	0.346
T=200						
0.0	-0.002	-0.008	0.000	-0.000	-0.000	0.072
0.3	0.004	-0.010	0.004	0.003	0.003	0.102
0.6	0.035	0.000	0.027	0.028	0.032	0.178
0.9	0.272	0.222	0.269	0.282	0.281	0.626
T=500						
0.0	0.004	0.001	0.005	0.005	0.005	0.001
0.3	0.004	-0.001	0.006	0.006	0.006	0.001
0.6	0.006	-0.002	0.010	0.010	0.010	0.001
0.9	0.016	0.008	0.031	0.030	0.024	-0.001
T=1000						
0.0	-0.004	-0.004	-0.004	-0.004	-0.004	-0.002
0.3	-0.006	-0.005	-0.005	-0.005	-0.005	-0.002
0.6	-0.011	-0.009	-0.009	-0.009	-0.009	-0.004
0.9	-0.039	-0.030	-0.030	-0.031	-0.034	-0.010
			RMSE			
T=100						
0.0	0.006	0.012	0.006	0.006	0.006	0.010
0.3	0.007	0.015	0.008	0.008	0.008	0.014
0.6	0.012	0.021	0.012	0.012	0.012	0.023
0.9	0.027	0.037	0.027	0.026	0.026	0.069
T=200						
0.0	0.002	0.003	0.002	0.002	0.002	0.003
0.3	0.003	0.004	0.003	0.003	0.003	0.005
0.6	0.004	0.006	0.004	0.004	0.004	0.008
0.9	0.013	0.014	0.012	0.012	0.012	0.028
T=500						
0.0	0.000	0.001	0.000	0.000	0.000	0.001
0.3	0.001	0.001	0.001	0.001	0.001	0.001
0.6	0.001	0.001	0.001	0.001	0.001	0.002
0.9	0.005	0.004	0.004	0.004	0.004	0.009
T=1000						
0.0	0.000	0.000	0.000	0.000	0.000	0.000
0.3	0.000	0.000	0.000	0.000	0.000	0.000
0.6	0.000	0.000	0.000	0.000	0.000	0.001
0.9	0.002	0.002	0.002	0.002	0.002	0.003

Table B.6.: Finite sample bias ($\times 1000$) and RMSE for coefficient β_2, Bartlett kernel.

$\rho_1 = \rho_2$	OLS	D-CPR	FM-CPR	IM-CPR(O)	IM-CPR(Fb)		
					Data-Dep	b=0.1	b=0.2
T=100							
0.0	0.057	0.152	0.071	0.093	0.049	0.054	0.069
0.3	0.142	0.190	0.101	0.103	0.066	0.058	0.083
0.6	0.278	0.237	0.134	0.131	0.130	0.075	0.096
0.9	0.505	0.325	0.194	0.301	0.390	0.242	0.214
T=200							
0.0	0.053	0.087	0.060	0.069	0.048	0.050	0.054
0.3	0.137	0.114	0.079	0.077	0.060	0.051	0.055
0.6	0.268	0.132	0.097	0.092	0.099	0.055	0.063
0.9	0.547	0.195	0.153	0.191	0.411	0.138	0.124
T=500							
0.0	0.048	0.061	0.052	0.063	0.055	0.054	0.057
0.3	0.139	0.075	0.064	0.067	0.057	0.052	0.057
0.6	0.303	0.084	0.077	0.080	0.068	0.054	0.055
0.9	0.602	0.124	0.132	0.113	0.188	0.067	0.070
T=1000							
0.0	0.054	0.064	0.057	0.062	0.054	0.054	0.054
0.3	0.143	0.073	0.066	0.067	0.054	0.054	0.054
0.6	0.314	0.082	0.077	0.072	0.057	0.055	0.054
0.9	0.632	0.103	0.118	0.085	0.100	0.055	0.062

Table B.7.: Empirical null rejection probabilities for $H_0 : \beta_2 = -0.3$, Andrews (1991) bandwidth, QS kernel, 0.05 level.

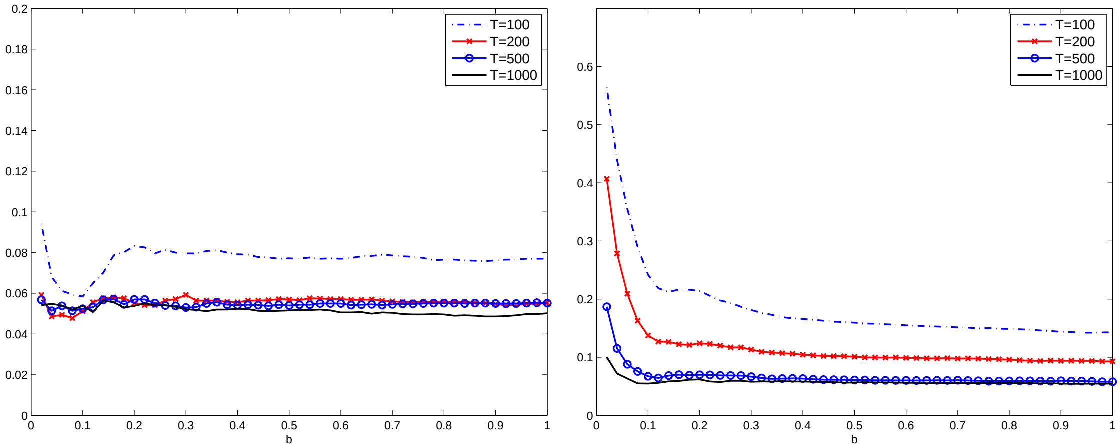

Figure B.5.: Empirical null rejections, IM-CPR(Fb) inference: t-test for β_2, Andrews (1991) bandwidth, QS kernel, $\rho_1 = \rho_2 = 0.3$ (left panel), $\rho_1 = \rho_2 = 0.9$ (right panel).

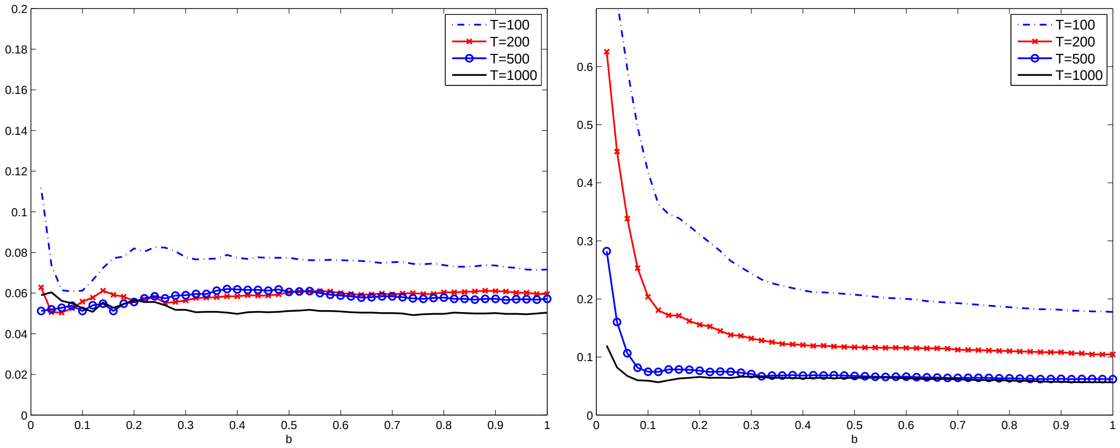

Figure B.6.: Empirical null rejections, IM-CPR(Fb) inference: $Wald$-test for β_1 and β_2, Andrews (1991) bandwidth, QS kernel, $\rho_1 = \rho_2 = 0.3$ (left panel), $\rho_1 = \rho_2 = 0.9$ (right panel).

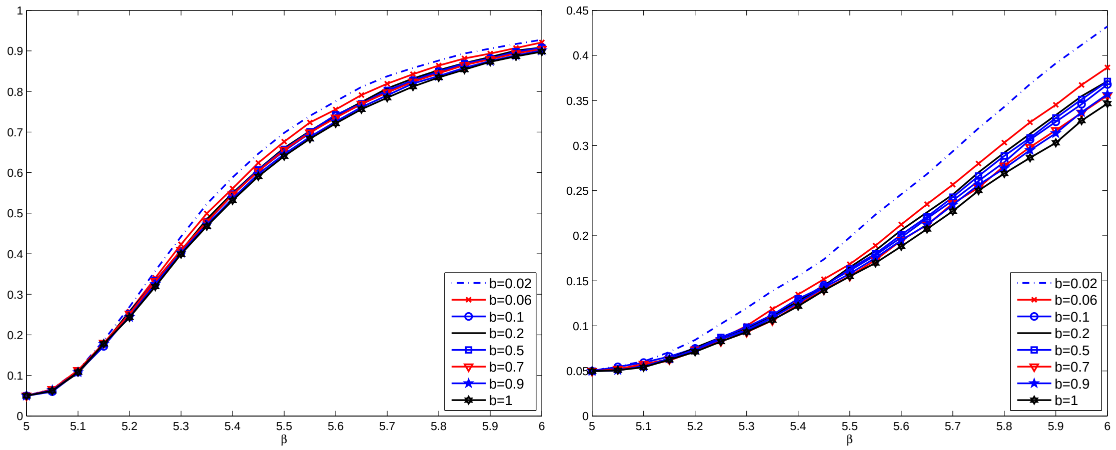

Figure B.7.: Size-corrected Power, t-test for β_1, T=100, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel, $\rho_1 = \rho_2 = 0.6$ (left panel), $\rho_1 = \rho_2 = 0.9$ (right panel).

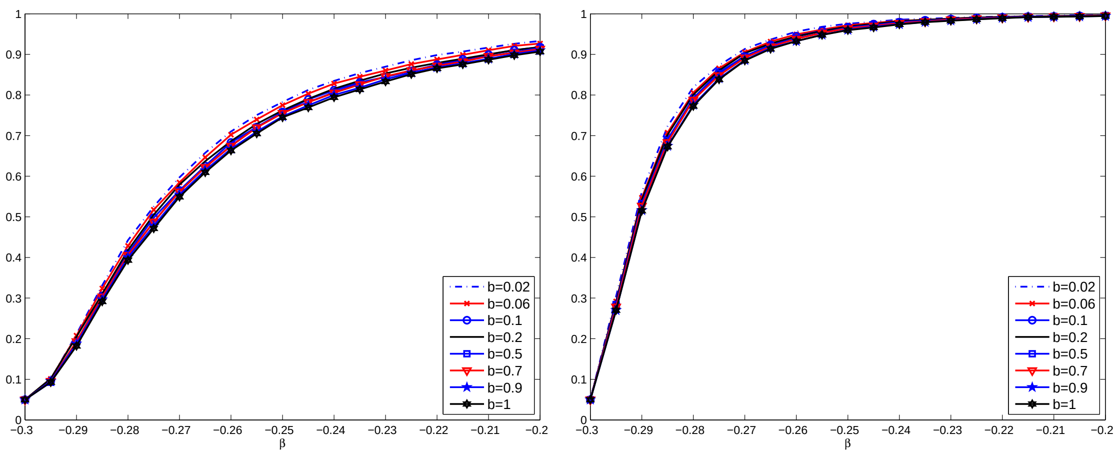

Figure B.8.: Size-corrected Power, t-test for β_2, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel, $T = 100$ (left panel), $T = 200$ (right panel).

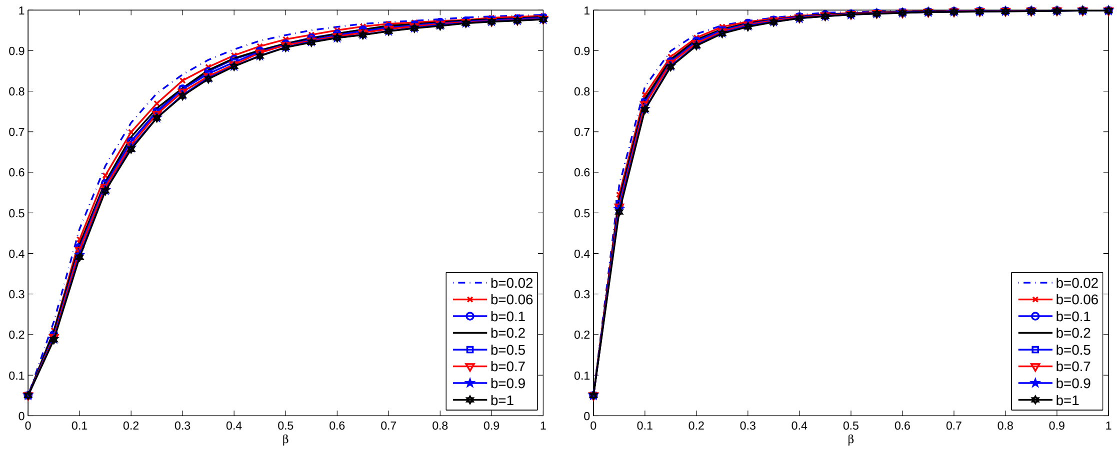

Figure B.9.: Size-corrected Power, $Wald$-test, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel, $T = 100$ (left panel), $T = 200$ (right panel).

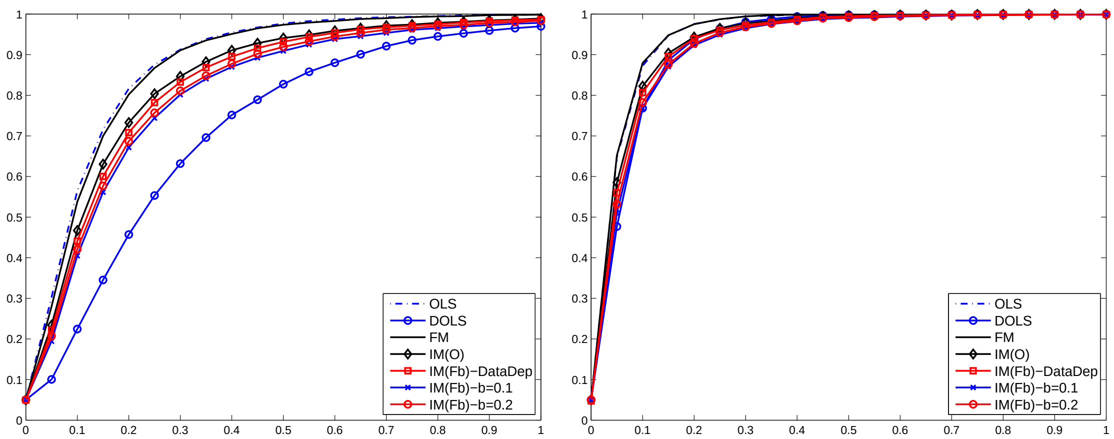

Figure B.10.: Size-corrected Power, $Wald$-test, $\rho_1 = \rho_2 = 0.6$, Andrews (1991) bandwidth, Bartlett kernel, $T = 100$ (left panel), $T = 200$ (right panel).

C. Appendix to Chapter 3

C.1. More Details on Pooling

Details for Pooling Cases (P), (S) and (T)

We consider the three cases of pooling mentioned in the main text and start with defining the quantities corresponding to the three cases. First, we define the three restriction matrices and then we present the correspondingly pooled estimators and their asymptotic distributions.

The Restriction Matrices

$$H_0^{(\mathrm{P})}: \begin{bmatrix} \delta_1 \\ \beta_{1,1} \\ \beta_{2,1} \end{bmatrix} = \begin{bmatrix} \delta_2 \\ \beta_{1,2} \\ \beta_{2,2} \end{bmatrix} = \cdots = \begin{bmatrix} \delta_N \\ \beta_{1,N} \\ \beta_{2,N} \end{bmatrix} \tag{C.1}$$

$$H_0^{(\mathrm{S})}: \begin{bmatrix} \beta_{1,1} \\ \beta_{2,1} \end{bmatrix} = \begin{bmatrix} \beta_{1,2} \\ \beta_{2,2} \end{bmatrix} = \cdots = \begin{bmatrix} \beta_{1,N} \\ \beta_{2,N} \end{bmatrix} \tag{C.2}$$

$$H_0^{(\mathrm{T})}: \delta_1 = \delta_2 = \cdots = \delta_N. \tag{C.3}$$

The corresponding restriction matrices for the Wald-type test are given by:

$$R^{(\mathrm{P})} = \begin{bmatrix} (0_{3\times1}, I_3) & (0_{3\times1}, -I_3) & 0_{3\times4} & \dots & 0_{3\times4} \\ \vdots & 0_{3\times4} & (0_{3\times1}, -I_3) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0_{3\times4} \\ (0_{3\times1}, I_3) & 0_{3\times4} & \dots & 0_{3\times4} & (0_{3\times1}, -I_3) \end{bmatrix} \in \mathbb{R}^{3(N-1)\times4N}, \ r = 0_{3(N-1)\times1}$$

$$R^{(\mathrm{S})} = \begin{bmatrix} (0_{2\times2}, I_2) & (0_{2\times2}, -I_2) & 0_{2\times4} & \dots & 0_{2\times4} \\ \vdots & 0_{2\times4} & (0_{2\times2}, -I_2) & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 0_{2\times4} \\ (0_{2\times2}, I_2) & 0_{2\times4} & \dots & 0_{2\times4} & (0_{2\times2}, -I_2) \end{bmatrix} \in \mathbb{R}^{2(N-1)\times4N}, \ r = 0_{2(N-1)\times1}$$

$$R^{(\mathrm{T})} = \begin{bmatrix} (0, 1, 0_{1\times2}) & (0, -1, 0_{1\times2}) & 0_{1\times4} & \dots & 0_{1\times4} \\ \vdots & 0_{1\times4} & (0, -1, 0_{1\times2}) & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 0_{1\times4} \\ (0, 1, 0_{1\times2}) & 0_{1\times4} & \dots & 0_{1\times4} & (0, -1, 0_{1\times2}) \end{bmatrix} \in \mathbb{R}^{(N-1)\times4N}, \ r = 0_{(N-1)\times1}$$

Pooled Estimation

In case the respective null hypotheses are not rejected, correspondingly pooled estimation is the next step to reap the possible efficiency gains from reducing the number of parameters to be estimated. This basically entails a corresponding redefinition of the regressor matrices, the parameter vectors; and for the asymptotic analysis the weighting matrices and limit processes. We discuss the three given cases in turn and start by defining the necessary adapted quantities:

(P) :

$$Z^{(\mathrm{P})} := \begin{bmatrix} Z_1^{(\mathrm{P})\prime} \\ Z_2^{(\mathrm{P})\prime} \\ \vdots \\ Z_T^{(\mathrm{P})\prime} \end{bmatrix}, \quad Z_t^{(\mathrm{P})} := \begin{bmatrix} I_N \\ X_t^{(\mathrm{P})} \end{bmatrix} \quad \theta^{(\mathrm{P})} := \begin{bmatrix} c_1 \\ \vdots \\ c_N \\ \delta \\ \beta_1 \\ \beta_2 \end{bmatrix},$$

$$X_t^{(\mathrm{P})} := \begin{bmatrix} t & t & \dots & t \\ x_{1,t} & x_{2,t} & \dots & x_{N,t} \\ x_{1,t}^2 & x_{2,t}^2 & \dots & x_{N,t}^2 \end{bmatrix}$$

$$G^{(\mathrm{P})} := \operatorname{diag}\left(T^{-1/2} \cdot I_N, T^{-3/2}, T^{-1}, T^{-3/2}\right)$$

$$J^{(\mathrm{P})}(r) := \begin{bmatrix} I_N \\ \mathbf{B}_N^{(\mathrm{P})}(r) \end{bmatrix}, \quad \mathbf{B}_N^{(\mathrm{P})}(r) := \begin{bmatrix} r & \dots & r \\ B_{v_1}(r) & \dots & B_{v_N}(r) \\ B_{v_1}^2(r) & \dots & B_{v_N}^2(r) \end{bmatrix}$$

(S) :

$$Z^{(\mathrm{S})} := \begin{bmatrix} Z_1^{(\mathrm{S})\prime} \\ Z_2^{(\mathrm{S})\prime} \\ \vdots \\ Z_T^{(\mathrm{S})\prime} \end{bmatrix}, \quad Z_t^{(\mathrm{S})} := \begin{bmatrix} D_{1,t} & 0_{2\times1} & \dots & 0_{2\times1} \\ 0_{2\times1} & D_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{2\times1} \\ 0_{2\times1} & \dots & 0_{2\times1} & D_{N,t} \\ x_{1,t} & x_{2,t} & \dots & x_{N,t} \\ x_{1,t}^2 & x_{2,t}^2 & \dots & x_{N,t}^2 \end{bmatrix}, \quad \theta^{(\mathrm{S})} := \begin{bmatrix} c_1 \\ \delta_1 \\ \vdots \\ c_N \\ \delta_N \\ \beta_1 \\ \beta_2 \end{bmatrix},$$

$$A_i^{(\mathrm{S})*} := \left(\hat{\Delta}_{vu}^{+}\right)^{i,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{i,t} \end{bmatrix}, \quad \tilde{A}_i^{(\mathrm{S})*} := \left(\hat{\Delta}_{vu}^{+}\right)^{i,\cdot} \left(\hat{\Omega}_{u.v}^{-1}\right)^{\cdot,i} \begin{bmatrix} T \\ 2\sum_{t=1}^{T} x_{i,t} \end{bmatrix}$$

$$G^{(\mathrm{S})} := \operatorname{diag}\left(I_N \otimes G_D^{(\mathrm{S})}, G_X^{(\mathrm{S})}\right), \quad G_D^{(\mathrm{S})} := \operatorname{diag}\left(T^{-1/2}, T^{-3/2}\right), \quad G_X^{(\mathrm{S})} := \operatorname{diag}\left(T^{-1}, T^{-3/2}\right)$$

$$J^{(\mathrm{S})}(r) := \begin{bmatrix} D_N^{(\mathrm{S})}(r) \\ \mathbf{B}_N^{(\mathrm{S})}(r) \end{bmatrix}, \quad D_N^{(\mathrm{S})}(r) := I_N \otimes \begin{bmatrix} 1 \\ r \end{bmatrix}, \quad \mathbf{B}_N^{(\mathrm{S})}(r) := \begin{bmatrix} B_{v_1}(r) & \dots & B_{v_N}(r) \\ B_{v_1}^2(r) & \dots & B_{v_N}^2(r) \end{bmatrix}$$

(T) :

$$Z^{(\mathrm{T})} := \begin{bmatrix} Z_1^{(\mathrm{T})\prime} \\ Z_2^{(\mathrm{T})\prime} \\ \vdots \\ Z_T^{(\mathrm{T})\prime} \end{bmatrix}, \quad Z_t^{(\mathrm{T})} := \begin{bmatrix} X_{1,t}^{(\mathrm{T})} & 0_{4\times 1} & \dots & 0_{4\times 1} \\ 0_{4\times 1} & X_{2,t}^{(\mathrm{T})} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{4\times 1} \\ 0_{4\times 1} & \dots & 0_{4\times 1} & X_{N,t}^{(\mathrm{T})} \\ t & t & \dots & t \end{bmatrix}, \quad \theta^{(\mathrm{T})} := \begin{bmatrix} c_1 \\ \beta_{1,1} \\ \beta_{2,1} \\ \vdots \\ c_N \\ \beta_{1,N} \\ \beta_{2,N} \\ \delta \end{bmatrix},$$

$$X_{i,t}^{(\mathrm{T})} := \begin{bmatrix} 1 \\ x_{i,t} \\ x_{i,t}^2 \end{bmatrix}, \quad A^{(\mathrm{T})*} := \begin{bmatrix} A_1^{(\mathrm{T})*} \\ A_2^{(\mathrm{T})*} \\ \vdots \\ A_N^{(\mathrm{T})*} \end{bmatrix}, \quad \tilde{A}^{(\mathrm{T})*} := \begin{bmatrix} \tilde{A}_1^{(\mathrm{T})*} \\ \tilde{A}_2^{(\mathrm{T})*} \\ \vdots \\ \tilde{A}_N^{(\mathrm{T})*} \end{bmatrix},$$

$$A_i^{(\mathrm{T})*} := \left(\hat{\Delta}_{vu}^{+}\right)^{i,i} \begin{bmatrix} 0 \\ T \\ 2\sum_{t=1}^{T} x_{i,t} \end{bmatrix}, \quad \tilde{A}_i^{(\mathrm{T})*} := \left(\hat{\Delta}_{vu}^{+}\right)^{i,\cdot} \left(\hat{\Omega}_{u.v}^{-1}\right)^{\cdot,i} \begin{bmatrix} 0 \\ T \\ 2\sum_{t=1}^{T} x_{i,t} \end{bmatrix}$$

$$G^{(\mathrm{T})} := \mathrm{diag}\left(I_N \otimes G_1^{(\mathrm{T})}, T^{-3/2}\right), \quad G_1^{(\mathrm{T})} := \mathrm{diag}\left(T^{-1/2}, T^{-1}, T^{-3/2}\right)$$

$$J^{(\mathrm{T})}(r) := \begin{bmatrix} \mathbf{B}_{v_1}^{(\mathrm{T})}(r) & & \\ & \ddots & \\ & & \mathbf{B}_{v_N}^{(\mathrm{T})}(r) \\ r & \dots & r \end{bmatrix}, \quad \mathbf{B}_{v_i}^{(\mathrm{T})}(r) := \begin{bmatrix} 1 \\ B_{v_i}(r) \\ B_{v_i}^2(r) \end{bmatrix}$$

Corollary 7 (Based on Hong and Wagner 2014, Corollaries 1 and 2)**.** *Let y_t be generated by* (3.1) *with the assumptions listed in place and where the pooling restrictions considered in either* (P)*,* (S) *or* (T) *are valid. Furthermore, assume again that long run variance estimation is performed consistently. Then, for the three considered cases, the correspondingly pooled FM-SOLS and FM-SUR estimators are, using the quantities defined above, given by:*

$$\hat{\theta}^{(P)} := \left(Z^{(P)\prime}Z^{(P)}\right)^{-1}\left(Z^{(P)\prime}y^{+} - \begin{bmatrix} 0_{N\times 1} \\ \sum_{i=1}^{N} A_i^{*} \end{bmatrix}\right), \tag{C.4}$$

$$\tilde{\theta}^{(P)} := \left(Z^{(P)\prime}\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)Z^{(P)}\right)^{-1}\left(Z^{(P)\prime}\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)y^{+} - \begin{bmatrix} 0_{N\times 1} \\ \sum_{i=1}^{N} \tilde{A}_i^{*} \end{bmatrix}\right), \tag{C.5}$$

$$\hat{\theta}^{(S)} := \left(Z^{(S)\prime}Z^{(S)}\right)^{-1}\left(Z^{(S)\prime}y^{+} - \begin{bmatrix} 0_{2N\times 1} \\ \sum_{i=1}^{N} A_i^{(S)*} \end{bmatrix}\right), \tag{C.6}$$

$$\tilde{\theta}^{(S)} := \left(Z^{(S)\prime}\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)Z^{(S)}\right)^{-1}\left(Z^{(S)\prime}\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)y^{+} - \begin{bmatrix} 0_{2N\times 1} \\ \sum_{i=1}^{N} \tilde{A}_i^{(S)*} \end{bmatrix}\right), \tag{C.7}$$

$$\hat{\theta}^{(T)} := \left(Z^{(T)\prime}Z^{(T)}\right)^{-1}\left(Z^{(T)\prime}y^{+} - \begin{bmatrix} A^{(T)*} \\ 0 \end{bmatrix}\right), \tag{C.8}$$

$$\tilde{\theta}^{(T)} := \left(Z^{(T)\prime}\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)Z^{(T)}\right)^{-1}\left(Z^{(T)\prime}\left(I_T \otimes \hat{\Omega}_{u.v}^{-1}\right)y^{+} - \begin{bmatrix} \tilde{A}^{(T)*} \\ 0 \end{bmatrix}\right). \tag{C.9}$$

For $T \to \infty$ the estimators are consistent with the following limiting distributions:

$$\left(G^{(P)}\right)^{-1}\left(\hat{\theta}^{(P)} - \theta^{(P)}\right) \Rightarrow \left(\int J^{(P)}J^{(P)\prime}\right)^{-1}\int J^{(P)}dB_{u.v}, \tag{C.10}$$

$$\left(G^{(P)}\right)^{-1}\left(\tilde{\theta}^{(P)} - \theta^{(P)}\right) \Rightarrow \left(\int J^{(P)}\Omega_{u.v}^{-1}J^{(P)\prime}\right)^{-1}\int J^{(P)}\Omega_{u.v}^{-1}dB_{u.v}, \tag{C.11}$$

$$\left(G^{(S)}\right)^{-1}\left(\hat{\theta}^{(S)} - \theta^{(S)}\right) \Rightarrow \left(\int J^{(S)}J^{(S)\prime}\right)^{-1}\int J^{(S)}dB_{u.v}, \tag{C.12}$$

$$\left(G^{(S)}\right)^{-1}\left(\tilde{\theta}^{(S)} - \theta^{(S)}\right) \Rightarrow \left(\int J^{(S)}\Omega_{u.v}^{-1}J^{(S)\prime}\right)^{-1}\int J^{(S)}\Omega_{u.v}^{-1}dB_{u.v}, \tag{C.13}$$

$$\left(G^{(T)}\right)^{-1}\left(\hat{\theta}^{(T)} - \theta^{(T)}\right) \Rightarrow \left(\int J^{(T)}J^{(T)\prime}\right)^{-1}\int J^{(T)}dB_{u.v}, \tag{C.14}$$

$$\left(G^{(T)}\right)^{-1}\left(\tilde{\theta}^{(T)} - \theta^{(T)}\right) \Rightarrow \left(\int J^{(T)}\Omega_{u.v}^{-1}J^{(T)\prime}\right)^{-1}\int J^{(T)}\Omega_{u.v}^{-1}dB_{u.v}. \tag{C.15}$$

Group-Wise Pooling (Pooling the Trend Coefficient and the Coefficients of the Stochastic Regressors Over Different Subsets)

Proof of Proposition 6. Deriving the limiting distribution of FM-type estimators commences from the limiting distribution of the underlying OLS and – in the SUR case additionally – the MSUR estimator. We start with the group-wise pooled OLS estimator, which is defined as:

$$\hat{\theta}^{\mathrm{GW}}_{\mathrm{OLS}} \;:=\; \left(\check{Z}'\check{Z}\right)^{-1}\check{Z}'y, \tag{C.16}$$

After centering around the true value θ^{GW} and pre-multiplying with the scaling matrix $\check{G}$ defined in the main text we arrive at:

$$\begin{aligned}\check{G}^{-1}\left(\hat{\theta}^{\mathrm{GW}}_{\mathrm{OLS}}-\theta^{\mathrm{GW}}\right) &= \left(\check{G}\check{Z}'\check{Z}\check{G}\right)^{-1}\left(\check{G}\check{Z}'u\right) \\ &= \left(\sum_{t=1}^{T}\check{G}\check{Z}_t\check{Z}_t'\check{G}\right)^{-1}\left(\sum_{t=1}^{T}\check{G}\check{Z}_t u_t\right)\end{aligned} \tag{C.17}$$

It follows that $\sum_{t=1}^{T}\check{G}\check{Z}_t\check{Z}_t'\check{G} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{T}\check{G}\check{Z}_t\check{Z}_t'\check{G}\sqrt{T} \Rightarrow \int \check{J}\check{J}'$, using $\lim_{T\to\infty}\sqrt{T}\check{G}\check{Z}_{\lfloor rT\rfloor} = \check{J}(r)$ for $0 \le r \le 1$ and the continuous mapping theorem. For the second term similar arguments as in Hong and Wagner (2014, Propositions 1 and 4) – without group-wise pooling in that paper – can be used to establish:

$$\sum_{t=1}^{T}\check{G}\check{Z}_t u_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{T}\check{G}\check{Z}_t u_t \;=\; \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix}\sqrt{T}\check{G}_c u_t\\ \sqrt{T}\check{G}_D\check{D}_t u_t\\ \sqrt{T}\check{G}_X\check{X}_t u_t\end{bmatrix} \Rightarrow \int \check{J}dB_u + F_u^{\mathrm{GW}}, \tag{C.18}$$

with $F_u^{\mathrm{GW}} := [0_{1\times(N+n_k)}, F_{u,1}^{\mathrm{GW}\,\prime}, \ldots, F_{u,l}^{\mathrm{GW}\,\prime}]'$, where

$$F_{u,j}^{\mathrm{GW}} := \sum_{i=1}^{N}\mathbb{1}_{\{i\in I_{m_j}\}}\cdot\Delta_{vu}^{i,i}\cdot\begin{pmatrix}1\\ 2\int B_{v_i}\end{pmatrix}, \quad j=1,\ldots,l. \tag{C.19}$$

Altogether this implies that

$$\begin{aligned}\check{G}^{-1}\left(\hat{\theta}^{\mathrm{GW}}_{\mathrm{OLS}}-\theta^{\mathrm{GW}}\right) &\Rightarrow \left(\check{J}\check{J}'\right)^{-1}\left(\int\check{J}dB_u + F_u^{\mathrm{GW}}\right) \\ &= \left(\check{J}\check{J}'\right)^{-1}\left(\int\check{J}dB_{u\cdot v} + \int\check{J}\Omega_{uv}\Omega_{vv}^{-1}dB_v + F_u^{\mathrm{GW}}\right).\end{aligned} \tag{C.20}$$

We now turn to the group-wise pooled MSUR estimator:

$$\tilde{\theta}^{\mathrm{GW}}_{\mathrm{MSUR}} := \left(\check{Z}'\left(I_T\otimes\hat{\Omega}_{uu}^{-1}\right)\check{Z}\right)^{-1}\check{Z}'\left(I_T\otimes\hat{\Omega}_{uu}^{-1}\right)u, \tag{C.21}$$

which after centering and scaling can be written as:

$$\check{G}^{-1}\left(\tilde{\theta}^{\mathrm{GW}}_{\mathrm{MSUR}} - \theta^{\mathrm{GW}}\right) = \left(\sum_{t=1}^{T} \check{G}\check{Z}_t\hat{\Omega}_{uu}^{-1}\check{Z}_t'\check{G}\right)^{-1}\left(\sum_{t=1}^{T} \check{G}\check{Z}_t\hat{\Omega}_{uu}^{-1}u_t\right). \tag{C.22}$$

Since $\hat{\Omega}_{uu} \to \Omega_{uu}$ by assumption, it immediately follows that $\sum_{t=1}^{T}\check{G}\check{Z}_t\hat{\Omega}_{uu}^{-1}\check{Z}_t\check{G}' \Rightarrow \int \check{J}\Omega_{uu}^{-1}\check{J}'$ and for the second term we now obtain using again similar arguments as in Hong and Wagner (2014, Propositions 1 and 4):

$$\sum_{t=1}^{T} \check{G}\check{Z}_t\hat{\Omega}_{uu}^{-1}u_t \Rightarrow \int \check{J}\Omega_{uu}^{-1}dB_u + \tilde{F}_u^{\mathrm{GW}}, \tag{C.23}$$

with $\tilde{F}_u^{\mathrm{GW}} := [0_{1\times(N+n_k)}, \tilde{F}_{u,1}^{\mathrm{GW}\prime}, \ldots, \tilde{F}_{u,l}^{\mathrm{GW}\prime}]'$, where

$$\tilde{F}_{u,j}^{\mathrm{GW}} := \sum_{i=1}^{N} \mathbb{1}_{\{i\in I_{m_j}\}} \cdot \Delta_{vu}^{i,\cdot} \cdot \left(\Omega_{uu}^{-1}\right)^{\cdot,i} \cdot \begin{pmatrix} 1 \\ 2\int B_{v_i}\end{pmatrix}, \quad j = 1, \ldots, l. \tag{C.24}$$

Combining the two terms gives the asymptotic distribution of the MSUR estimator:

$$\begin{aligned}\check{G}^{-1}\left(\tilde{\theta}^{\mathrm{GW}}_{\mathrm{MSUR}} - \theta^{\mathrm{GW}}\right) &\Rightarrow \left(\int \check{J}\Omega_{uu}^{-1}\check{J}'\right)^{-1}\left(\int \check{J}\Omega_{uu}^{-1}dB_u + \tilde{F}_u^{\mathrm{GW}}\right) \\ &= \left(\int \check{J}\Omega_{uu}^{-1}\check{J}'\right)^{-1}\left(\int \check{J}\Omega_{uu}^{-1}dB_{u\cdot v} + \int \check{J}\Omega_{uu}^{-1}\Omega_{uv}\Omega_{vv}^{-1}dB_v + \tilde{F}_u^{\mathrm{GW}}\right)\end{aligned} \tag{C.25}$$

Having the asymptotic distributions of the OLS and MSUR estimators at hand allows to next derive the asymptotic distribution of the FM-OLS and FM-SUR estimators for the group-wise pooled case.

We start with FM-SOLS. From the definition of the estimator in (3.27) and the definition of y^+ it is clear that the essential term to be understood, after centering and scaling, is

$$\sum_{t=1}^{T} \check{G}\check{Z}_t\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t \Rightarrow \int \check{J}\Omega_{uv}\Omega_{vv}^{-1}dB_v + F_v^{\mathrm{GW}}, \tag{C.26}$$

with $F_v^{\mathrm{GW}} := [0_{1\times(N+n_k)}, F_{v,1}^{\mathrm{GW}\prime}, \ldots, F_{v,l}^{\mathrm{GW}\prime}]'$, where

$$F_{v,j}^{\mathrm{GW}} := \sum_{i=1}^{N} \mathbb{1}_{\{i\in I_{m_j}\}} \cdot \Omega_{uv}^{i,\cdot} \cdot \Omega_{vv}^{-1} \cdot \Delta_{vv}^{\cdot,i} \cdot \begin{pmatrix} 1 \\ 2\int B_{v_i}\end{pmatrix}, \quad j = 1, \ldots, l. \tag{C.27}$$

The above result together with (C.20) implies that

$$\begin{aligned}\sum_{t=1}^{T}\check{G}\check{Z}_t u_t^{+} &= \sum_{t=1}^{T}\check{G}\check{Z}_t u_t - \sum_{t=1}^{T}\check{G}\check{Z}_t\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t \\ &\Rightarrow \int \check{J}dB_{u\cdot v} + \int \check{J}\Omega_{uv}\Omega_{vv}^{-1}dB_v + F_u^{\mathrm{GW}} - \int \check{J}\Omega_{uv}\Omega_{vv}^{-1}dB_v - F_v^{\mathrm{GW}} \\ &= \int \check{J}dB_{u\cdot v} + F_u^{\mathrm{GW}} - F_v^{\mathrm{GW}} = \int \check{J}dB_{u\cdot v} + A^{\mathrm{GW}}.\end{aligned} \tag{C.28}$$

Observing now that the non-zero blocks of A^{GW} are given by

$$\begin{aligned}A_j^{\mathrm{GW}} &= \sum_{i=1}^{N}\mathbb{1}_{\{i\in I_{m_j}\}}\cdot\left(\Delta_{vu}^{i,i} - \Omega_{uv}^{i,\cdot}\cdot\Omega_{vv}^{-1}\cdot\Delta_{vv}^{\cdot,i}\right)\cdot\begin{pmatrix}1\\ 2\int B_{v_i}\end{pmatrix}, \quad j=1,\ldots,l, \\ &= \sum_{i=1}^{N}\mathbb{1}_{\{i\in I_{m_j}\}}\cdot\left(\Delta_{vu}^{+}\right)^{i,i}\cdot\begin{pmatrix}1\\ 2\int B_{v_i}\end{pmatrix}, \quad j=1,\ldots,l,\end{aligned} \tag{C.29}$$

shows that $\check{G}A^{\mathrm{GW}*} \Rightarrow A^{\mathrm{GW}}$, which establishes the result (3.29) for the group-wise FM-SOLS estimator.

A similar reasoning also applies to the group-wise pooled FM-SUR estimator defined in (3.28). Here the relevant term to consider is

$$\sum_{t=1}^{T}\check{G}\check{Z}_t\hat{\Omega}_{u\cdot v}^{-1}\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t \Rightarrow \int \check{J}\Omega_{u\cdot v}^{-1}\Omega_{uv}\Omega_{vv}^{-1}dB_v + \tilde{F}_v^{\mathrm{GW}}, \tag{C.30}$$

with $\tilde{F}_v^{\mathrm{GW}} := [0_{1\times(N+n_k)}, \tilde{F}_{v,1}^{\mathrm{GW}\prime}, \ldots, \tilde{F}_{v,l}^{\mathrm{GW}\prime}]'$, where

$$\tilde{F}_{v,j}^{\mathrm{GW}} := \sum_{i=1}^{N}\mathbb{1}_{\{i\in I_{m_j}\}}\cdot\left(\Omega_{u\cdot v}^{-1}\right)^{i,\cdot}\cdot\Omega_{uv}\cdot\Omega_{vv}^{-1}\cdot\Delta_{vv}^{\cdot,i}\cdot\begin{pmatrix}1\\ 2\int B_{v_i}\end{pmatrix}, \quad j=1,\ldots,l. \tag{C.31}$$

The above result together with (C.25) implies that

$$\begin{aligned}\sum_{t=1}^{T}\check{G}\check{Z}_t\hat{\Omega}_{u\cdot v}^{-1}u_t^{+} &= \sum_{t=1}^{T}\check{G}\check{Z}_t\Omega_{u\cdot v}^{-1}u_t - \sum_{t=1}^{T}\check{G}\check{Z}_t\Omega_{u\cdot v}^{-1}\hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}v_t \\ &\Rightarrow \int \check{J}\Omega_{u\cdot v}^{-1}dB_{u\cdot v} + \int \check{J}\Omega_{u\cdot v}^{-1}\Omega_{uv}\Omega_{vv}^{-1}dB_v + \tilde{F}_u^{\mathrm{GW}} - \int \check{J}\Omega_{u\cdot v}^{-1}\Omega_{uv}\Omega_{vv}^{-1}dB_v - \tilde{F}_v^{\mathrm{GW}} \\ &= \int \check{J}dB_{u\cdot v} + \tilde{F}_u^{\mathrm{GW}} - \tilde{F}_v^{\mathrm{GW}} = \int \check{J}\Omega_{u\cdot v}^{-1}dB_{u\cdot v} + \tilde{A}^{\mathrm{GW}}.\end{aligned} \tag{C.32}$$

The non-zero blocks of $\tilde{A}^{\text{GW}}$ are given by

$$\tilde{A}_j^{\text{GW}} = \sum_{i=1}^{N} \mathbb{1}_{\{i \in I_{m_j}\}} \cdot \left(\Delta_{vu}^{i,\cdot} \cdot \left(\Omega_{u \cdot v}^{-1}\right)^{\cdot,i} - \left(\Omega_{u \cdot v}^{-1}\right)^{i,\cdot} \cdot \Omega_{uv} \cdot \Omega_{vv}^{-1} \cdot \Delta_{vv}^{\cdot,i}\right) \cdot \begin{pmatrix} 1 \\ 2\int B_{v_i} \end{pmatrix}, \quad j = 1, \ldots, l, \tag{C.33}$$

$$= \sum_{i=1}^{N} \mathbb{1}_{\{i \in I_{m_j}\}} \cdot \left(\Delta_{vu}^{+}\right)^{i,\cdot} \cdot \left(\Omega_{u \cdot v}^{-1}\right)^{\cdot,i} \cdot \begin{pmatrix} 1 \\ 2\int B_{v_i} \end{pmatrix}, \quad j = 1, \ldots, l,$$

which implies that $\check{G}\tilde{A}^{\text{GW}*} \Rightarrow \tilde{A}^{\text{GW}}$. ■

C.2. Additional Empirical Results

	Intercept			Intercept and Linear Trend		
	PP	PP(fb)$_1$	PP(fb)$_2$	PP	PP(fb)$_1$	PP(fb)$_2$
Australia	0.945	0.951	0.380	-1.280	-1.304	-1.378
Austria	0.004	0.098	-0.115	-1.908	-1.813	-1.878
Belgium	0.795	0.620	0.225	-1.419	-1.554	-1.587
Canada	-0.255	-0.348	-0.458	-2.496	*-3.057*	*-3.056*
Denmark	0.038	0.062	-0.270	-2.293	-2.270	-2.273
Finland	0.727	0.586	0.169	-2.321	-2.422	-2.412
France	-0.044	-0.159	-0.286	-1.958	-2.204	-2.214
Germany	-0.256	-0.338	-0.317	-2.356	-2.582	-2.598
Italy	0.524	0.234	0.004	-1.665	-1.825	-1.835
Japan	0.222	0.105	-0.134	-1.719	-1.880	-1.893
Netherlands	0.163	0.092	-0.007	-2.184	-2.310	-2.334
New Zealand	-0.100	-0.101	-0.139	-2.606	-2.686	-2.688
Norway	0.909	0.838	0.116	-2.142	-2.177	-2.179
Portugal	1.271	0.997	0.352	-1.872	-1.879	-1.869
Spain	0.704	0.428	0.050	-1.005	-1.192	-1.235
Sweden	0.137	0.131	-0.276	-2.339	-2.401	-2.407
Switzerland	-1.004	-1.053	-1.001	-2.807	-2.449	-2.481
UK	1.118	1.383	0.425	-1.567	-1.418	-1.625
USA	-0.336	-0.308	-0.447	-2.981	*-2.911*	*-2.915*

Table C.1.: Unit root test results for log GDP per capita. The tests employed are the Phillips-Perron (1988) test, PP, as well as the one- and two-step detrended fixed-b versions, PP(fb)$_1$ and PP(fb)$_2$, of this test developed in Vogelsang and Wagner (2013). The specifications of the deterministic components are intercept only and intercept and linear trend. The results are based on the Bartlett kernel with bandwidth chosen according to Newey and West (1994). *Italic* entries denote rejection of the null hypothesis at the 10% level and **bold** entries indicate rejection at the 5% level.

	PO_t	Shin	$P_{\hat{u}}$	CT
Australia	-2.727	**0.129**	11.330	**0.108**
Austria	-3.785	0.077	**55.997**	0.056
Belgium	**-5.586**	0.054	*50.269*	0.062
Canada	-3.349	**0.189**	12.420	**0.145**
Denmark	**-4.937**	0.053	40.613	0.052
Finland	**-5.656**	0.045	**75.016**	0.050
France	**-4.948**	0.060	28.847	0.066
Germany	**-7.895**	**0.411**	**68.343**	**0.111**
Italy	**-4.163**	**0.182**	34.141	**0.146**
Japan	**-5.829**	**0.155**	8.127	**0.152**
Netherlands	**-5.688**	**0.108**	**96.172**	0.074
New Zealand	**-5.375**	**0.132**	13.337	**0.115**
Norway	-3.530	*0.097*	20.644	*0.095*
Portugal	**-9.111**	**0.101**	19.959	**0.111**
Spain	-3.343	*0.091*	42.578	0.086
Sweden	**-4.268**	*0.085*	28.679	0.085
Switzerland	**-6.282**	0.065	**85.979**	0.084
UK	**-7.673**	*0.085*	**97.169**	0.073
USA	-2.448	**0.576**	13.920	**0.156**
Critical Values ($\alpha = 0.1$)	-4.1567	0.081	45.237	0.086
Critical Values ($\alpha = 0.05$)	-3.8429	0.101	52.952	0.106

Table C.2.: Cointegration and non-cointegration test results for (3.1). The left block-column presents the results for the "linear" non-cointegration test PO_t of Phillips and Ouliaris (1990) and the "linear" cointegration test of Shin (1994). Linear here refers to an application of these tests treating log GDP per capita and its square as two integrated processes. The right block-column presents the results for the modifications of these two tests to the CPR setting discussed in Wagner (2013, 2015). These are labelled $P_{\hat{u}}$ (non-cointegration test) and CT (cointegration test). *Italic* entries denote rejection of the null hypothesis at the 10% level and **bold** entries indicate rejection at the 5% level.

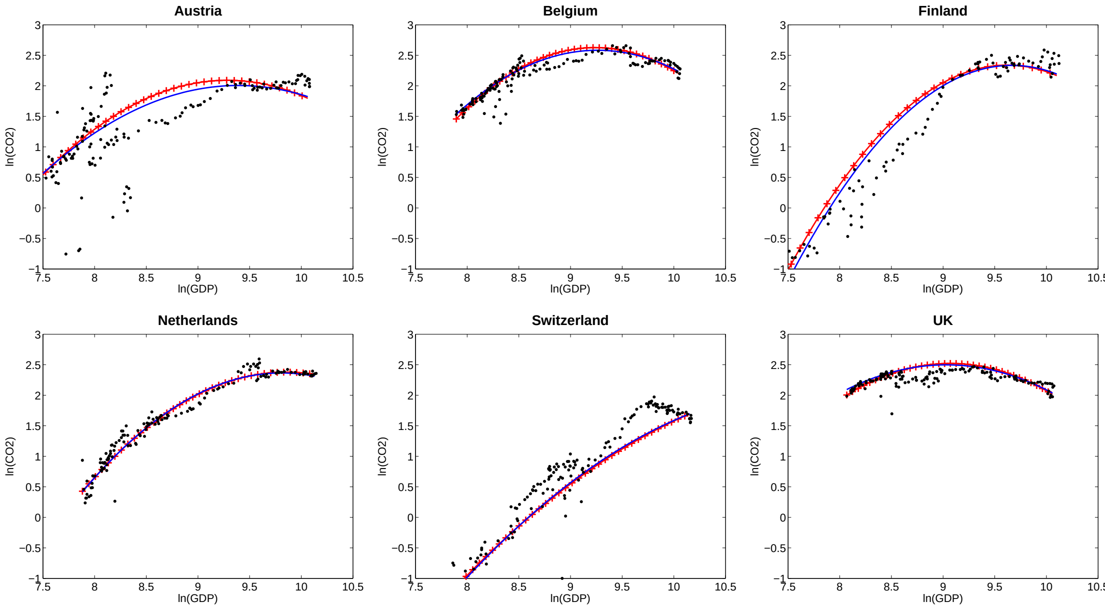

Figure C.1.: EKC estimation results for Equation (3.1): scatter plot and EKC. The solid lines correspond to the FM-SUR estimates and the solid lines with +-marks to the FM-SOLS estimates. For further explanations see notes to Figure 3.1.

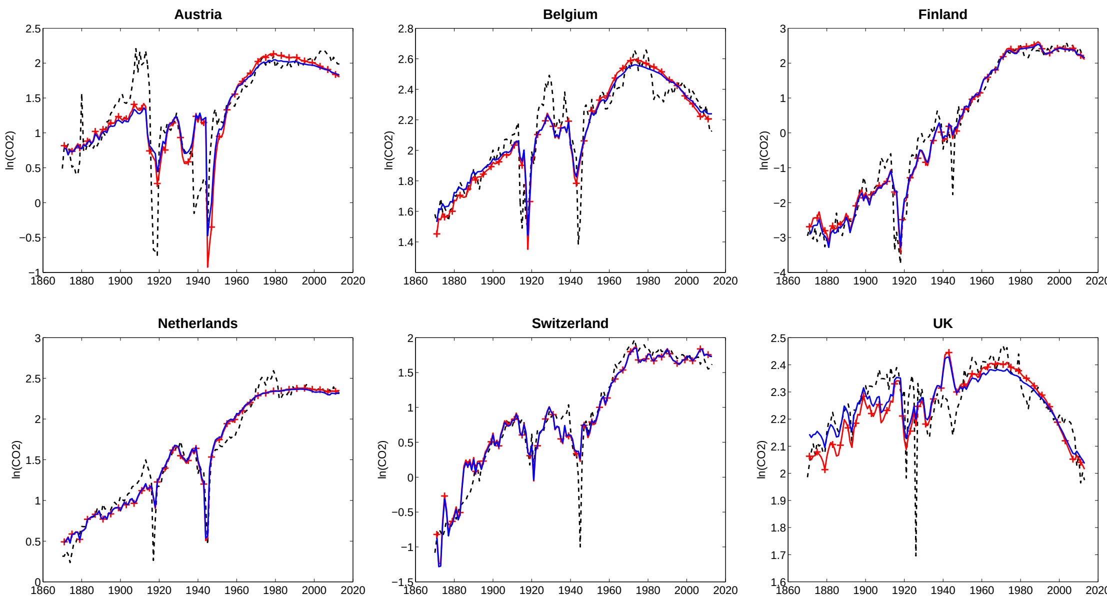

Figure C.2.: EKC estimation results for Equation (3.1): actual and fitted values. The dashed lines show the actual values, the solid lines the FM-SUR fitted values and the solid lines with +-marks the FM-SOLS fitted values.

		FM-SOLS
Linear Trend & Stochastic Regressors (P)		
Stochastic Regressors (S)	2 3	AT-NL, BE-UK, NL-UK, AT-NL-UK, BE-NL-UK
Linear Trend (T)	2 3	AT-FI, AT-CH, AT-UK, BE-NL, BE-UK, FI-CH, AT-BE-UK, AT-FI-CH
		FM-SUR
Linear Trend & Stochastic Regressors (P)		
Stochastic Regressors (S)	2 3 4	AT-NL, BE-NL, BE-UK, NL-UK, AT-NL-UK, BE-NL-UK, AT-BE-NL-UK
Linear Trend (T)	2 3	AT-FI, AT-CH, AT-UK, BE-NL, BE-UK, FI-CH, FI-UK, AT-BE-UK, AT-FI-CH, AT-FI-UK,

Table C.3.: List of group members corresponding to the tests described in Table 3.2. For more details see notes to Table 3.2.

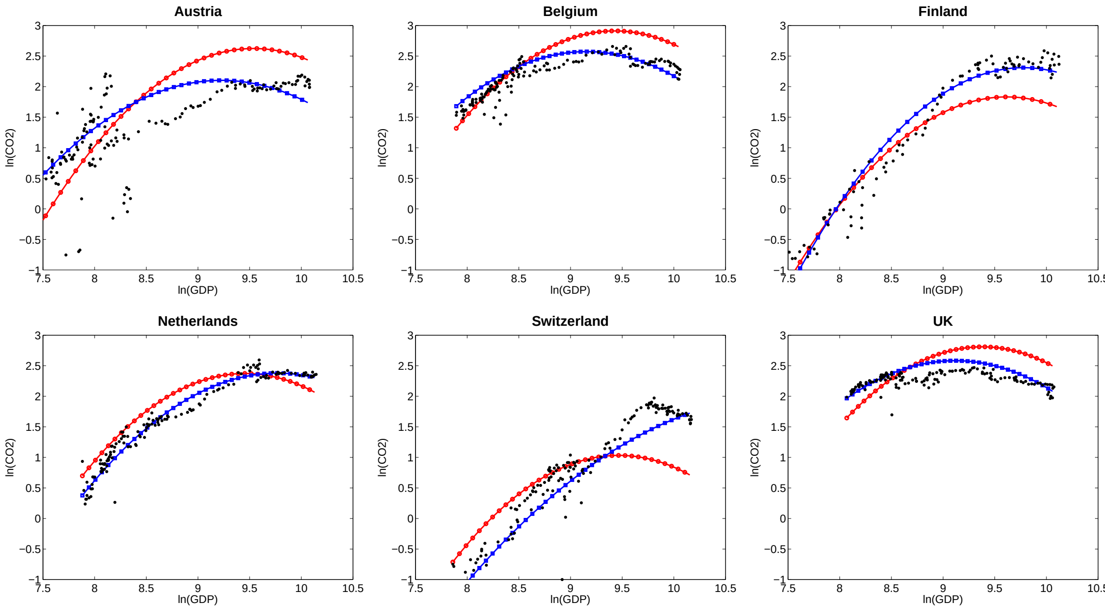

Figure C.3.: EKC estimation results for Equation (3.1): scatter plot and EKC. The solid lines with square symbols correspond to the group-wise pooled FM-SUR estimates and the solid lines with o-marks to the pooled FM-SUR estimates. For further explanations see notes to Figure 3.1.

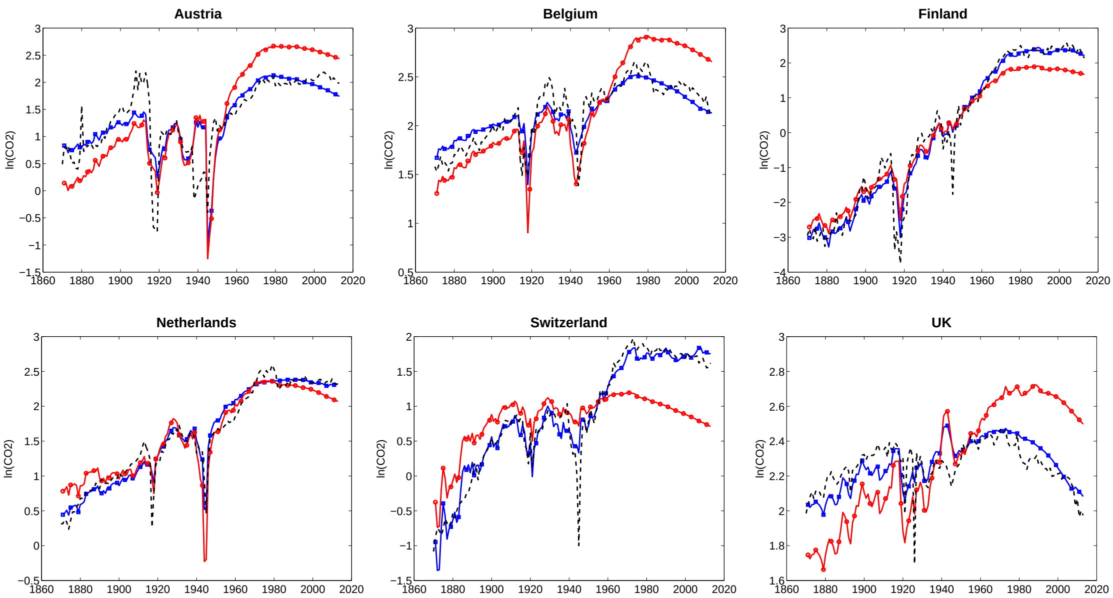

Figure C.4.: EKC estimation results for Equation (3.1): actual and fitted values. The dashed lines show the actual values, the solid lines with square symbols the group-wise pooled FM-SUR fitted values and the solid lines with o-marks the pooled FM-SUR fitted values.

List of Tables

List of Figures

Bibliography

Agras, J. and D. Chapman (1999). A Dynamic Approach to the Environmental Kuznets Curve Hypothesis. *Ecological Economics* **28**, 267–277.

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

Antweiler, W., B.R. Copeland, and M.S. Taylor (2001). Is Free Trade Good for the Environment? *American Economic Review* **91**, 877–908.

Apergis, N. (2016). Environmental Kuznets Curves: New Evidence on Both Panel and Country-Level CO_2 Emissions. *Energy Economics* **54**, 263–271.

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.

Auffhammer, M. and R.T. Carson (2008). Forecasting the Path of China's CO_2 Emissions Using Province-Level Information. *Journal of Environmental Economics and Management* **55**, 229–247.

Baek, J. (2015). Environmental Kuznets Curve for CO_2 Emissions: The Case of Arctic Countries. *Energy Economics* **50**, 13–17.

Bernard, J. T., M. Gavin, L. Khalaf, and M. Voia (2015). Environmental Kuznets Curve: Tipping Points, Uncertainty and Weak Identification. *Environmental and Resource Economics* **60**, 285–315.

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

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

Boden, T.A., G. Marland and R.J. Andres (2016). Global, Regional, and National Fossil-Fuel CO_2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Bolt, J. and J.L. van Zanden (2014). The Maddison Project: Collaborative Research on Historical National Accounts. *The Economic History Review* **67**, 627–651.

Bradford, D., R. Fender, S.H. Shore, and M. Wagner (2005). The Environmental Kuznets Curve: Exploring a Fresh Specification. *Contributions to Economic Analysis and Policy* **4**, Article 5.

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.

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

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

Choi, I. and E. Kurozumi, (2012). Model Selection Criteria for the Leads-and-Lags Cointegrating Regression. *J. 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 Function. *Econometrica* **39**, 255–256.

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

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

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

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

de Jong, R.M. and M. Wagner (2016). Panel Nonlinear Cointegration Analysis of the Environmental Kuznets Curve. *Mimeo.*

Dijkgraaf, E., and H.R. Vollebergh (2005). A Test for Parameter Homogeneity in CO_2 Panel EKC Estimations. *Environmental and Resource Economics* **32**, 229–239.

Dinda, S. (2005). A Theoretical Basis for the Environmental Kuznets Curve. *Ecological Economics* **53**, 403–413.

Dinda, S. and D. Coondoo (2006). Income and Emission: A Panel Data-Based Cointegration Analysis. *Ecological Economics* **57**, 167-181.

Engle, R.F. and C.W.J. Granger (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* **55**, 251-276.

Esteve, V. and C. Tamarit (2012). Threshold Cointegration and Nonlinear Adjustment Between CO2 and Income: The Environmental Kuznets Curve in Spain, 1857-2007. *Energy Economics* **34**, 2148–2156.

Fosten, J., B. Morley, and T. Taylor (2012). Dynamic Misspecification in the Environmental Kuznets Curve: Evidence From CO_2 and SO_2 Emissions in the United Kingdom. *Ecological Economics* **76**, 25–33.

Friedl, B. and M. Getzner (2003). Determinants of CO_2 Emissions in a Small Open Economy. *Ecological Economics* **45**, 133–148.

Galeotti, M., A. Lanza, and F. Pauli (2006). Reassessing the Environmental Kuznets Curve for CO_2 Emissions: A Robustness Exercise. *Ecological Economics* **57**, 152–163.

Grabarczyk, P. and M. Wagner (2017). Integrated Modified OLS Estimation for Cointegrating Polynomial Regressions – With an Application to the Environmental Kuznets Curve for CO_2 Emissions. *Mimeo.*

Grossman, G.M. and A.B. Krueger (1991). Environmental Impacts of a North American Free Trade Agreement. NBER Working Paper No. 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.

He, J. and P. Richard (2010). Environmental Kuznets Curve for CO_2 in Canada. *Ecological Economics* **69**, 1083–1093.

Hilton, F.G.H. and A. Levinson (1998). Factoring the Environmental Kuznets Curve: Evidence From Automobile Lead Emissions. *Journal of Environmental Economics and Management* **35**, 126–141.

Holtz-Eakin, D. and T.M. Selden (1995). Stoking the Fires? CO_2 Emissions and Economic Growth. *Journal of Public Economics* **57**, 85–101.

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.

Hong, S.H. and M. Wagner (2014). Seemingly Unrelated Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference. Mimeo.

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

Jalil, A. and S.F. Mahmud (2009). Environment Kuznets Curve for CO_2 Emissions: A Cointegration Analysis for China. *Energy Policy* **37**, 5167–5172.

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

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

Kahn, M.E. (1998). A Household Level Environmental Kuznets Curve. *Economics Letters* **59**, 269–273.

Kao, C. and M.-H. Chiang (2000). On the Estimation and Inference of a Cointegrated Regression in Panel Data. In Baltagi, B.H. (Ed.) *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, 179–222. Elsevier, Amsterdam.

Kasparis, I. (2008). Detection of Functional Form Misspecification in Cointegrating Relations. *Econometric Theory* **24**, 1373–1403.

Ke, Y., J. Li, and W. Zhang (2016). Structure Identification in Panel Data Analysis. *The Annals of Statistics* **44**, 1193–1233.

Kiefer, N. M. and T. J. Vogelsang (2005). A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests, *Econometric Theory* **21**, 1130–1164.

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.

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.

Liang, H., P.C.B. Phillips, H. Wang, and Q. Wang (2016). Weak Convergence to Stochastic Integrals for Econometric Applications. *Econometric Theory* **32**, 1349–1375.

Lindmark, M. (2002). An EKC-Pattern in Historical Perspective: Carbon Dioxide Emissions, Technology, Fuel Prices and Growth in Sweden 1870–1997. *Ecological Economics* **42**, 333–347.

List, J.A. and C.A. Gallet (1999). The Environmental Kuznets Curve: Does One Size Fit All? *Ecological Economics* **31**, 409–423.

Mark, N.C., M. Ogaki, and D. Sul (2005). Dynamic Seemingly Unrelated Cointegrating Regressions. *Review of Economic Studies* **72**, 797–820.

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.

Müller-Fürstenberger G. and M. Wagner (2007). Exploring the Environmental Kuznets Hypothesis: Theoretical and Econometric Problems. *Ecological Economics* **62**, 648–660.

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

Park, J.Y. and M. Ogaki (1991). Seemingly Unrelated Canonical Cointegrating Regressions. Mimeo.

Park, J.Y. and P.C.B. Phillips (1988). Statistical Inference in Regressions With Integrated Processes: Part I. *Econometric Theory* **4**, 468–497.

Park, J.Y. and P.C.B. Phillips (1989). Statistical Inference in Regressions With Integrated Processes: Part II. *Econometric Theory* **5**, 95–131.

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

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

Pedroni, P. (2000). Fully Modified OLS for Heterogeneous Cointegrated Panels. In Baltagi, B.H. (Ed.) *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, 97–130. Elsevier, Amsterdam.

Perman, R. and D.I. Stern (2003). Evidence from Panel Unit Root and Cointegration Tests That the Environmental Kuznets Curve Does Not Exist. *The Australian Journal of Agricultural and Resource Economics* **47**, 325–347.

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

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

Phillips, P.C.B. and H.R. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* **67**, 1057–1111.

Phillips, P.C.B. and S. Ouliaris (1990). Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica* **58**, 165–193.

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

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

Romero-Avila, D. (2008). Questioning the Empirical Basis of the Environmental Kuznets Curve for CO_2: New Evidence From a Panel Stationarity Test Robust to Multiple Breaks and Cross-Dependence. *Ecological Economics* **64**, 559–574.

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

Saikkonen, P. and I. Choi (2004). Cointegrating Smooth Transition Regressions. *Econometric Theory* **20**, 301-340.

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, D.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.

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

Stern, D.I. (2017). The Environmental Kuznets Curve after 25 Years. *Journal of Bioeconomics* **19**, 7–28.

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. (2016). 150 Years of Boom and Bust: What Drives Mineral Commodity Prices? *Macroeconomic Dynamics*. Forthcoming.

Stypka, O., P. Grabarczyk, R. Kawka, and M. Wagner (2017). “Linear” Fully Modified OLS Estimation of Cointegrating Polynomial Regressions. *Mimeo.*

Su, L., Z. Shi, and P.C.B. Phillips (2016). Identifying Latent Structures in Panel Data. *Econometrica* **84**, 2215–2264.

Torras, M. and J.K. Boyce (1998). Income, Inequality, and Pollution: A Reassessment of the Environmental Kuznets Curve. *Ecological Economics* **25**, 147–160.

Uchiyama, K. (2016). Environmental Kuznets Curve Hypothesis and Carbon Dioxide Emissions. Springer Japan, Tokyo.

Vogelsang, T.J., and M. Wagner (2013). A Fixed-b Perspective on the Phillips-Perron Tests. *Econometric Theory* **29**, 609–628.

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 RESET Test for Cointegrating Regressions. *SFB 823 Discussion Paper 37/2014*, 1–29.

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

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

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

Wagner, M. and P. Grabarczyk (2017). The Environmental Kuznets Curve for Carbon Dioxide Emissions: A Seemingly Unrelated Cointegrating Polynomial Regression Approach. *Mimeo.*

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

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

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.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Dissertation ist bisher keiner anderen Fakultät vorgelegt worden. Ich erkläre, dass ich bisher kein Promotionsverfahren erfolglos beendet habe und dass keine Aberkennung eines bereits erworbenen Doktorgrades vorliegt.

Peter Grabarczyk