SFB 823

Integrated modified OLS estimation for cointegrating polynomial regressions - with an application to the environmental Kuznets curve for CO_2 emissions

Discussion Paper

Manuel Frondel, Peter Grabarczyk, Martin Wagner

Nr. 74/2016

SFB 823

Integrated Modified OLS Estimation for Cointegrating Polynomial Regressions – With an Application to the Environmental Kuznets Curve for CO_2 Emissions

Manuel Frondel
RWI
Leibniz Institute for Economic Research
Essen, Germany
&
Ruhr Universität Bochum
Bochum, Germany

Peter Grabarczyk
Faculty of Statistics
Technical University Dortmund
Dortmund, Germany

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

Abstract

This paper considers the integrated modified OLS (IM-OLS) estimator for cointegrating polynomial regressions recently developed in Vogelsang and Wagner (2014a; 2014b). Cointegrating polynomial regressions include deterministic variables, integrated processes and integer powers of integrated processes as explanatory variables. The stochastic regressors are allowed to be endogenous and the stationary errors are allowed to be serially correlated. The IM-OLS estimator allows for asymptotically standard inference in this framework when using consistent estimators of the long run variance. Additionally, we also provide fixed-b asymptotic theory for the case of full design to capture the impact of kernel and bandwidth choice on the sampling distributions of estimators and test statistics. We investigate the properties of the IM-OLS estimator and hypothesis tests based upon it by means of a simulation study to compare its performance with fully modified OLS (FM-OLS) and dynamic OLS (D-OLS). Finally, we apply the method to estimate the environmental Kuznets curve for CO_2 emissions over the period 1870–2009.

JEL Classification: C12, C13, C32, Q20

Keywords: Bandwidth, Cointegrating Polynomial Regression, Environmental Kuznets Curve, Fixed-b asymptotics, Integrated Modified OLS Estimation, Kernel

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 variance has to be estimated.

This paper 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. The IM-OLS estimator is a tuning parameter free estimator. Also in the CPR framework this 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 variance parameter. Since asymptotic standard inference does not capture the impact of kernel and bandwidth choices on the limiting 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 variance estimation. These adjusted IM-OLS residuals are obtained in exactly the same way as in the linear case. This leads 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.

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 Kwiatkowski et al. (1992)-type (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 small 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.

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

[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.

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 to estimate the EKC based on a data set containing CO_2 emissions and GDP for 18 early industrialized countries over the time period 1870–2010 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 cubic EKC relationship for six countries (thereof four countries with evidence for a quadratic EKC relationship) with similar coefficient estimates across the different methods for most of the countries.

The paper is organized as follows. In Section 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. Section 3 contains a small simulation study to evaluate the finite sample performance of the proposed methods. In Section 4 we apply these methods to analyze the EKC hypothesis. Section 5 briefly summarizes and concludes. All proofs are given in Appendix A, whereas Appendix B contains additional results of the empirical analysis and additional simulation results are given in Appendix C.

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. 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 for notational simplicity. We denote the m-dimensional identity matrix by I_m and $\mathbb{E}(\cdot)$ denotes the expected value.

2 Theory

2.1 Setup and Assumptions

We consider the following cointegrating polynomial regression (CPR) model

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

$$x_t = x_{t-1} + v_t, \quad (2)$$

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

Remark 1. *We can include more general deterministic components D_t, with the assumption*

$$\lim_{T\to\infty} \sqrt{T} G_D D_{[rT]} = D(r) \text{ with } 0 < \int_0^r D(z)D(z)'\,\mathrm{d}z < \infty, \; r \in [0,1], \quad (3)$$

where $G_D = G_D(T) \in \mathbb{R}^{d+1)\times(d+1)}$. For the leading case of polynomial time trends given in (1)*, we have*

$$G_D := diag\left(T^{-1/2}, T^{-3/2}, \dots, T^{-(d+1/2)}\right)$$

and $D(r) := [1, r, \dots, r^d]'$.

For the simulation study in Section 3 we restrict ourselves to one of the following cases: (a) no deterministics, (b) intercept only or (c) intercept and linear time trend.

Remark 2. *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* (1)*.*

Next we define $\eta_t := [u_t, v_t']'$ 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

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \eta_t \Rightarrow B(r) = \Omega^{1/2} W(r), \quad r \in [0,1], \quad (4)$$

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

$$\Omega := \sum_{j=-\infty}^{\infty} \mathbb{E}(\eta_{t-j}\eta_t') = \begin{pmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{pmatrix} > 0, \tag{5}$$

is the long run covariance matrix of the vector error process. Since we want to exclude cointegration within x_t, we assume $\Omega_{vv} > 0$.

We partition the Brownian motion processes according to

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

so that we can write the limit process in (4) by means of the Cholesky decomposition of $\Omega^{1/2}$ as

$$B(r) = \begin{pmatrix} B_u(r) \\ B_v(r) \end{pmatrix} = \begin{pmatrix} \omega_{u\cdot v}^{1/2} & \Omega_{uv}(\Omega_{vv}^{-1/2})' \\ 0 & \Omega_{vv}^{1/2} \end{pmatrix} \begin{pmatrix} w_{u\cdot v}(r) \\ W_v(r) \end{pmatrix}, \tag{6}$$

where $\omega_{u\cdot v} := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$.

Unless otherwise stated we denote the OLS residuals from (1) by $\hat{u}_t$ such that a nonparametric kernel estimator of Ω is given by

$$\hat{\Omega} := T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{T} k\left(\frac{|i-j|}{M}\right)\hat{\eta}_i\hat{\eta}_j', \tag{7}$$

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 (7) provide consistent estimates of the long run variance. For later purposes we also define the one-sided long run covariance matrix $\Lambda := \sum_{j=1}^{\infty}\mathbb{E}[\eta_{t-j}\eta_t']$ and the contemporaneous covariance matrix $\Sigma := \mathbb{E}[\eta_t\eta_t']$, which are partitioned in the same way as Ω. This notation allows us to write the long run covariance matrix as $\Omega = \Sigma + \Lambda + \Lambda'$ and to define $\Delta := \Sigma + \Lambda$.

For the asymptotic behavior of the vector process X_{jt} we 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, Park, and Phillips, 2001)

$$\lim_{T\to\infty}\sqrt{T}G_{X_j}X_{jt} = \lim_{T\to\infty}\begin{pmatrix} T^{-1/2} & & \\ & \ddots & \\ & & T^{-p_j/2} \end{pmatrix}\begin{pmatrix} x_{jt} \\ \vdots \\ x_{jt}^{p_j} \end{pmatrix} = \begin{pmatrix} B_{v_j} \\ \vdots \\ B_{v_j}^{p_j} \end{pmatrix} =: \mathbf{B}_{v_j}(r), \tag{8}$$

with $v_t := [v_{1t}, \dots, v_{mt}]'$. Next denote the stacked vector Brownian motion process as $\mathbf{B}_v(r) := [\mathbf{B}_{v_1}(r)', \dots, \mathbf{B}_{v_m}(r)']'$.

2.2 IM-OLS Estimation in the CPR Framework

In order to establish the IM-OLS estimator compute the partial sums in model (1) as

$$\begin{aligned} S_t^y &= S_t^{D\prime}\delta + \sum_{j=1}^{m} S_t^{X_j\prime}\beta_{X_j} + S_t^u, \\ S_t^y &= S_t^{D\prime}\delta + S_t^{X\prime}\beta + S_t^u, \end{aligned} \tag{9}$$

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 $\beta := [\beta_{X_1}', \dots, \beta_{X_m}']'$. We stack the vectors in the following form $S_t^X := [S_t^{X_1\prime}, \dots, S_t^{X_m\prime}]'$ and $S_t^{\tilde{X}} := [S_t^{D\prime}, S_t^{X\prime}]$. Now we can write equation (9) in matrix form as

$$S^y = S^{\tilde{X}}\theta + S^u, \tag{10}$$

with $\theta := [\delta', \beta']'$. Concerning endogeneity we also add the vector x_t as a regressor (i.e. only the integrated regressors to the power one) into equation (9)

$$S_t^y = S_t^{D\prime}\delta + S_t^{X\prime}\beta + x_t'\gamma + S_t^u \tag{11}$$

and redefine $\theta := [\delta', \beta', \gamma']'$.
Estimating equation (11) via OLS leads to residuals which we denote by

$$\tilde{S}_t^u = S_t^y - S_t^{D\prime}\tilde{\delta} - S_t^{X\prime}\tilde{\beta} - x_t'\tilde{\gamma}, \tag{12}$$

where we label $\tilde{\delta}, \tilde{\beta}, \tilde{\gamma}$ the IM-OLS estimators. The following proposition gives the asymptotic distribution of the IM-OLS estimators $\tilde{\delta}, \tilde{\beta}, \tilde{\gamma}$. The IM-OLS estimator is given by the OLS estimator of the model (11) with $S_t^{\xi} := [S_t^{D\prime}, S_t^{X\prime}, x_t']$ replacing $S_t^{\tilde{X}}$. Based on the discussion at the beginning of this section concerning the asymptotic behavior of the deterministic components as well as the processes η_t and X_t, 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 (1) and (2), the deterministic components satisfy (3) and the error process satisfies a FCLT of the form (4). With* $\theta := [\delta', \beta', (\Omega_{vv}^{-1}\Omega_{vu})']'$ *it holds for* $T \to \infty$ *that*

$$A_{IM}^{-1}(\tilde{\theta} - \theta) = \begin{pmatrix} G_D(\tilde{\delta} - \delta) \\ G_X\left(\tilde{\beta} - \beta\right) \\ (\tilde{\gamma} - \Omega_{vv}^{-1}\Omega_{vu}) \end{pmatrix} = \left(T^{-2}A_{IM}S^{\xi\prime}S^{\xi}A_{IM}\right)^{-1}\left(T^{-2}A_{IM}S^{\xi\prime}S^{u}\right) - \begin{pmatrix} 0 \\ 0 \\ \Omega_{vv}^{-1}\Omega_{vu} \end{pmatrix} \tag{13}$$

$$\begin{aligned} &\Rightarrow \quad \omega_{u\cdot v}^{1/2}\left(\int f(s)f(s)'ds\right)^{-1}\int f(s)w_{u\cdot v}(s)ds \\ &= \quad \omega_{u\cdot v}^{1/2}\left(\int f(s)f(s)'ds\right)^{-1}\int [F(1) - F(s)]dw_{u\cdot v}(s), \end{aligned} \tag{14}$$

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 (14) is, conditional on $W_v(r)$*, normally distributed with zero mean and covariance matrix*

$$V_{IM} := \omega_{u\cdot v}\left(\int f(s)f(s)'ds\right)^{-1}\left(\int [F(1) - F(s)][F(1) - F(s)]'ds\right)\left(\int f(s)f(s)'ds\right)^{-1}. \tag{15}$$

Full Design

In this section we briefly introduce *full design* that allows to perform fixed-b inference in CPR models based on the IM-OLS estimator. Full design always prevails when only one of the integrated regressors enters with powers larger than one. In more general cases, full design can always be achieved by including additional regressors appropriately into the model. However, this is costly in terms that more parameters have to be estimated.

Consider for simplicity the following data generating process

$$y_t = \beta_1 x_{1t} + \beta_2 x_{1t}^2 + \beta_3 x_{2t} + \beta_4 x_{2t}^2 + u_t, \tag{16}$$

where the error process $\Delta x_t = v_t$ satisfies a FCLT similar to (6)

$$T^{-1/2}\sum_{t=1}^{[rT]} v_t \Rightarrow \begin{pmatrix} B_{v_1}(r) \\ B_{v_2}(r) \end{pmatrix} = \Omega_{vv}^{1/2}W_v(r) = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ 0 & \lambda_{22} \end{pmatrix}\begin{pmatrix} W_{v_1}(r) \\ W_{v_2}(r) \end{pmatrix}. \tag{17}$$

In order to establish the asymptotic distribution of the IM-OLS estimator in this setup, one needs to consider 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 (17) 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

$$\begin{pmatrix} B_{v_1}(r) \\ B_{v_1}^2(r) \\ B_{v_2}(r) \\ B_{v_2}^2(r) \end{pmatrix} = \underbrace{\begin{pmatrix} \lambda_{11} & 0 & \lambda_{12} & 0 & 0 \\ 0 & \lambda_{11}^2 & 0 & \lambda_{12}^2 & 2\lambda_{11}\lambda_{12} \\ 0 & 0 & \lambda_{22} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{22}^2 & 0 \end{pmatrix}}_{=:F(\Omega_{vv})} \underbrace{\begin{pmatrix} W_{v_1}(r) \\ W_{v_1}^2(r) \\ W_{v_2}(r) \\ W_{v_2}^2(r) \\ W_{v_1}(r)W_{v_2}(r) \end{pmatrix}}_{=:\mathbf{W}_v(r)}. \tag{18}$$

If λ_{12} is not equal to zero, the transformation matrix $F(\Omega_{vv})$ does not define a bijective mapping and thus 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 (16) 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*. The benefit is that in the limiting distribution of the IM-OLS estimator (14) certain terms factor out and therefore the limiting distribution can be expressed as 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.

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

$$\begin{aligned} A_{IM}^{-1}(\tilde{\theta} - \theta) &\Rightarrow \omega_{u\cdot v}^{1/2} \left(\Pi \int g(s)g(s)'ds\Pi' \right)^{-1} \Pi \int g(s)w_{u\cdot v}(s)ds \\ &= \omega_{u\cdot v}^{1/2} (\Pi')^{-1} \left(g(s)g(s)'ds \right)^{-1} \int [G(1) - G(s)]dw_{u\cdot v}(s), \end{aligned} \tag{19}$$

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 (17) are not equal to zero, which is excluded by the assumption $\Omega_{vv} > 0$.

Remark 3. *(i) It is obvious that full design always prevails when only one of the integrated regressors enters with powers larger than one.*

(ii) If one wants to perform fixed-b inference, it is possible to achieve full design in the augmented regression by simply including all necessary regressors. Based on the representation (6), where we assume that $\Omega_{vv}^{1/2}$ is an upper-triangular matrix, we sort the integrated regressors $x_{1t}, \ldots, x_{mt}$ such that $p_1 \leq \ldots \leq p_m$. This guarantees the most parsimonious way to achieve full design.[3]

2.3 IM-OLS Based Inference in the CPR Framework

Next we want to consider Wald tests for testing q linear hypotheses of the form $H_0 : R\theta = r$, where we assume that there exists a nonsingular $q \times q$ scaling matrix A_R such that

$$\lim_{T\to\infty} A_R^{-1} R A_{IM} = R^*, \tag{20}$$

where R^* has rank q and A_{IM} is the scaling matrix for the asymptotic distribution of the IM-OLS estimator in equation (11). The condition on the matrix R given in equation (20) 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 (11) and S^{ξ} as the stacked matrix across time. The covariance matrix V_{IM} of this asymptotic distribution immediately suggests estimators of the form

$$\begin{aligned} \check{V}_{IM} &:= \check{\omega}_{u\cdot v} A_{IM}^{-1} \left(S^{\xi\prime} S^{\xi}\right)^{-1} (C'C) \left(S^{\xi\prime} S^{\xi}\right)^{-1} A_{IM}^{-1}, \\ &= \check{\omega}_{u\cdot v} \left(T^{-2} A_{IM} S^{\xi\prime} S^{\xi} A_{IM}\right)^{-1} \left(T^{-4} A_{IM} C'C A_{IM}\right) \left(T^{-2} A_{IM} S^{\xi\prime} S^{\xi} A_{IM}\right)^{-1} \end{aligned} \tag{21}$$

where $C := [c_1, \ldots, c_T]'$ with $c_t := S_T^{S^{\xi}} - S_{t-1}^{S^{\xi}}$ and $S_t^{S^{\xi}} := \sum_{j=1}^{t} S_j^{\xi}$. $\check{\omega}_{u\cdot v}$ denotes an estimator for $\omega_{u\cdot v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ and we have three different candidates for such an estimator. First, $\hat{\omega}_{u\cdot v}$ based on the OLS residuals from model (1), so that we use the estimator for Ω given in equation (7). Second, we can use the first differences of the OLS residuals of the

[3] For example, if we consider a model with two integrated regressors, where the first enters together with its second power and the second regressor enters with powers up to power three, we simply have to include the cross-product $x_{1t}x_{2t}$ to achieve full design. Otherwise, if we would sort both regressors the other way round, then we would need to include $x_{1t}x_{2t}$ and also $x_{1t}^2 x_{2t}, x_{1t}x_{2t}^2, x_{2t}^3$ to achieve full design, because of the assumed Cholesky decomposition of $\Omega_{vv}^{1/2}$ in upper-triangular form in (6).

regression in equation (11) 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, so that we do not consider these tests in the simulation study in Section 3.

Following the discussion in Vogelsang and Wagner (2014a) Section 5, correlation between these residuals and the OLS estimators of equation (11) causes problems for fixed-b inference regarding δ, β and γ, so that we have to adjust the residuals in a similar way. Define the vector z_t as

$$z_t := t\sum_{j=1}^{T}\xi_j - \sum_{j=1}^{t-1}\sum_{s=1}^{j}\xi_s, \quad \xi_t := [S_t^{D\prime}, S_t^{X\prime}, x_t^{\prime}]^{\prime} \tag{22}$$

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

$$\tilde{S}_t^{u*} := \tilde{S}_t^u - z_t^{\perp\prime}\hat{\pi}, \tag{23}$$

where $\hat{\pi} := \left(\sum_{t=1}^{T} z_t^{\perp} z_t^{\perp\prime}\right)^{-1}\sum_{t=1}^{T} z_t^{\perp}\tilde{S}_t^u$. We obtain asymptotically pivotal test statistics under fixed-b asymptotics based on the residuals (23), see Lemma 1 below. As a third option for estimating $\omega_{u.v}$ we therefore use the first differences of the adjusted residuals given in equation (23):

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

This estimator of the long run variance $\omega_{u\cdot v}$ has the required properties to deliver a pivotal fixed-b limit for the Wald statistics.

Lemma 1. *(i) Consider the OLS estimator of (35) 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

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

(ii) With this result it follows that the asymptotic behavior of the partial sum process of the adjusted residuals is given by

$$T^{-1/2}\sum_{t=2}^{[rT]}\Delta\tilde{S}_t^{u*} \Rightarrow \omega_{u\cdot v}^{1/2}\left(\int_0^r dw_{u\cdot v}(s) - h(r)'\left(\int_0^1 h(s)h(s)'ds\right)^{-1}\int_0^1 [H(1)-H(s)]dw_{u\cdot v}(s)\right)$$
$$=: \omega_{u\cdot v}^{1/2}\widetilde{P}^*(r), \tag{24}$$

where, conditional on $W_v(r)$, $\widetilde{P}^*(r)$ is uncorrelated with the scaled and centered limit of $\tilde{\theta}$ given in equation (14) of Proposition 1.

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 1 provides the basis for pivotal fixed-b limit for Wald statistics.

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

$$\tilde{W}^* \Rightarrow \frac{\chi_q^2}{Q_b(\widetilde{P^*}, \widetilde{P^*})}, \tag{25}$$

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

The expression $Q_b(\widetilde{P^*}, \widetilde{P^*})$ is the fixed-b limit of the long run covariance estimator of the form (7) using $\Delta\tilde{S}_t^{u*}$ instead of $\hat{\eta}_t$. Therefore 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.[4] Furthermore, t- as well as Wald-type tests can be performed based on long run covariance estimation with $\Delta\tilde{S}_t^{u*}$. On the other hand 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 $T \to \infty$

$$\hat{W} \Rightarrow \chi_q^2. \tag{26}$$

[4]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.

3 Simulation Study

In this section we compare the performance of the 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 by means of a simulation study in a CPR framework. We label these estimators D-CPR, FM-CPR and IM-CPR, respectively. Furthermore, we include the OLS estimator as a benchmark. We assess 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. We use data generated according to

$$\begin{aligned} y_t &= \delta_1 + \delta_2 t + \beta_1 x_t + \beta_2 x_t^2 + u_t && (27) \\ x_t &= x_{t-1} + v_t, \quad x_0 = 0, && (28) \end{aligned}$$

with

$$\begin{aligned} u_t &= \rho_1 u_{t-1} + e_{1,t} + \rho_2 e_{2,t}, \quad u_0 = 0 && (29) \\ v_t &= e_{2,t} + 0.5 e_{2,t-1}, && (30) \end{aligned}$$

where $e_{1,t}, e_{2,t}$ are independent identically distributed standard normal random variables independent of each other. In addition, we also consider a cubic data generation process, where we include the regressor x_t to the power three in equation (27). The parameter values chosen are $\delta_1 = \delta_2 = 1, \beta_1 = 5, \beta_2 = -0.3$ for the quadratic specification and $\delta_1 = \delta_2 = 1, \beta_1 = -50, \beta_2 = 10, \beta_3 = -0.4$ for the cubic specification, respectively. These values are based on prior estimation results in conjunction with the environmental Kuznets curve (EKC) hypothesis with FM-CPR and D-CPR. Please note that the values of δ_1, δ_2 have no effect on the β-estimators. 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, for the time being, we focus on the case $\rho_1 = \rho_2$. For the FM-CPR estimator we choose the Bartlett and quadratic spectral 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 $\lfloor 4(T/100)^{2/9} \rfloor$, labelled NW_T. For the D-CPR estimator we use the Akaike information criterion based lead and lag length choice of Choi and Kurozumi (2012). We consider 5000 replications for the sample sizes $T = 100, 200, 500, 1000$.

Bias and RMSE

Let us briefly summarize the main findings of the Bias and RMSE simulations. First of all, only little difference is found between the results for the different kernels and bandwidths used for FM-CPR estimation. Additionally, the results are qualitatively equal for the quadratic and the cubic specification. Consequently, we report only the results for the quadratic specification together with the Bartlett kernel and the bandwidth according to Andrews (1991).

We start the discussion for the bias of β_1, where we find that the estimators are virtually unbiased in the case of no correlation. With increasing $\rho = \rho_1 = \rho_2$ the bias of the OLS and the FM-CPR estimator increases, whereas D-CPR and IM-CPR estimators appear to be less sensitive to the increasing correlation parameters. Specially when the value of ρ switches from 0.6 to 0.9 there is a jump in the bias for all estimators. When the sample size T increases all of the estimators have reduced biases, as expected. The results for β_2 are the same qualitatively: For smaller sample sizes the D-CPR estimator shows the best performance, whereas for greater sample sizes IM-CPR has the best outcome and FM-CPR becomes inferior.

In the case of RMSE, inverted patterns emerge for both coefficients β_1 and β_2. For small sample sizes, OLS and FM-CPR have the smallest root mean squared errors. This also holds true for larger sample sizes, but the differences between the estimators become smaller. As mentioned in Vogelsang and Wagner (2014a) this is not unforeseen because IM-CPR uses a regression with an $I(1)$ error S_t^u, whereas OLS and FM-CPR uses an $I(0)$ error u_t. The RMSE of the IM-CPR remains the largest of all estimators even for large sample sizes T and vast correlation parameters ρ, but the difference compared to the other estimators becomes negligible.

The simulation study shows that the IM-CPR estimator is more effective in reducing bias than the FM-CPR estimator. In smaller sample sizes D-CPR is less biased than IM-CPR. Conversely, IM-CPR and D-CPR generally have larger RMSE than OLS and FM-CPR, nonetheless the difference decreases for larger sample sizes.

$\rho_1 = \rho_2$	OLS	D-CPR	FM-CPR			IM-CPR
			And91	NW	NW_T	
			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
			T=500			
0.0	-0.000	0.000	-0.000	-0.000	-0.000	-0.000
0.3	0.004	0.000	0.000	0.000	0.000	-0.000
0.6	0.017	0.001	0.004	0.004	0.005	0.001
0.9	0.111	0.022	0.061	0.065	0.085	0.027
			T=1000			
0.0	0.000	0.000	0.000	0.000	0.000	0.000
0.3	0.002	0.000	0.000	0.000	0.000	0.000
0.6	0.009	0.000	0.002	0.002	0.002	0.000
0.9	0.060	0.007	0.026	0.028	0.044	0.009

Table 1: Finite sample bias of four estimators for coefficient β_1, Bartlett kernel.

Empirical Null Rejection Probabilities

Now we investigate the finite sample behavior of the t-type and $Wald$-type tests introduced in Section 2, where again we restrict ourselves to the case $\rho_1 = \rho_2$ for the present. In this case we test the hypotheses $H_0 : \beta_1 = 5$ and $H_0 : \beta_2 = -0.3$ separately (t-test) as well as jointly ($Wald$-test). First of all, we consider standard asymptotic results based on traditional bandwidth and kernel assumptions that are in the spirit of the $\hat{W}$ test statistic in (26). For this type of asymptotics we study OLS, OLS-HAC, D-CPR, FM-CPR and IM-CPR test statistics, where once again we incorporate the Bartlett and quadratic spectral kernels and the bandwidths according to the rules of Andrews (1991), Newey and West (1994) and the simplified NW_T rule. 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 test statistics for the IM-CPR estimator are implemented in two ways: the first

$\rho_1 = \rho_2$	OLS	D-CPR	FM-CPR			IM-CPR
			And91	NW	NW_T	
			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
			T=500			
0.0	0.013	0.017	0.013	0.013	0.013	0.021
0.3	0.019	0.024	0.018	0.018	0.018	0.030
0.6	0.038	0.040	0.032	0.032	0.032	0.052
0.9	0.173	0.139	0.137	0.137	0.151	0.204
			T=1000			
0.0	0.006	0.008	0.007	0.007	0.007	0.010
0.3	0.010	0.011	0.009	0.009	0.009	0.015
0.6	0.020	0.019	0.016	0.016	0.016	0.026
0.9	0.097	0.069	0.070	0.071	0.082	0.101

Table 2: Finite sample RMSE of four estimators for coefficient β_1, Bartlett kernel.

one rests upon a grid $b = 0.02, 0.04, \ldots, 1.00$ and uses a specific bandwidth $M = bT$, whereas for the second version we compute a bandwidth M^* according to the above mentioned rules. Subsequently, we determine $b^* = M^*/T$ and choose b as the largest multiple of 0.02 smaller or equal to b^* for fixed-b inference [5]. The simulated critical values for this type of inference depend on both kernel and bandwidth. Again, the number of replications is 5000 and the nominal level is 0.05 in all cases.

Table 3 and Table 4 show empirical null rejection probabilities using data dependent bandwidth choices in conjunction with the Bartlett and the quadratic spectral kernel. They are ordered as follows: OLS, OLS-HAC, D-CPR, FM-CPR and the test statistic using the $\hat{\omega}_{u\cdot v}$ estimator for standard asymptotic inference based on the IM-CPR estimator, labelled IM-CPR(O). In the last three columns we add the results for fixed-b inference using one of

[5] This is labelled "Data-Dep" later in the text.

the data dependent bandwidth rules, as described in the last paragraph (we refer to this as IM-CPR(Fb) "Data-Dep") and we also perform fixed-b inference using the fixed-b values $b = 0.1$ and $b = 0.2$, respectively.

$\rho_1 = \rho_2$	OLS	OLS-HAC	D-CPR	FM-CPR	IM-CPR(O)	IM-CPR(Fb)		
						Data-Dep	b=0.1	b=0.2
				T=100				
0.0	0.059	0.128	0.164	0.077	0.100	0.049	0.053	0.069
0.3	0.154	0.149	0.206	0.110	0.109	0.068	0.060	0.077
0.6	0.371	0.260	0.270	0.170	0.137	0.127	0.077	0.092
0.9	0.725	0.556	0.425	0.411	0.356	0.452	0.300	0.249
				T=200				
0.0	0.048	0.097	0.089	0.059	0.071	0.045	0.048	0.055
0.3	0.147	0.108	0.118	0.079	0.077	0.060	0.051	0.055
0.6	0.374	0.196	0.145	0.126	0.089	0.097	0.056	0.064
0.9	0.746	0.486	0.269	0.314	0.242	0.438	0.161	0.136
				T=500				
0.0	0.054	0.077	0.067	0.057	0.066	0.054	0.050	0.057
0.3	0.156	0.081	0.082	0.070	0.070	0.055	0.050	0.057
0.6	0.375	0.141	0.091	0.086	0.079	0.064	0.050	0.058
0.9	0.762	0.366	0.142	0.223	0.117	0.191	0.067	0.068
				T=1000				
0.0	0.053	0.068	0.062	0.055	0.061	0.055	0.052	0.054
0.3	0.163	0.078	0.072	0.064	0.066	0.055	0.051	0.055
0.6	0.400	0.131	0.079	0.081	0.070	0.057	0.052	0.055
0.9	0.787	0.314	0.114	0.176	0.087	0.097	0.055	0.063

Table 3: Empirical null rejection probabilities for $H_0 : \beta_1 = 5$, quadratic specification, Andrews (1991) bandwidth, quadratic spectral kernel, 0.05 level. The last three columns show results of fixed-b inference based on the IM-CPR estimator: in the column "Data-Dep" we have applied a data dependent bandwidth rule in order to determine a value b^*, in the latter two columns we have used fixed values for b.

Below is a summary of the main findings of this simulation starting with the coefficient β_1. As expected OLS textbook tests show the best performance in case of no correlation, but have severe size distortions at least for $\rho > 0.3$. The OLS-HAC version has slight size distortions for $\rho = 0.0$, but outperforms the OLS textbook test for increasing values of ρ_1 and ρ_2. For small sample sizes of $T = 100$ the D-CPR tests are very size distorted even in the non-correlated case, but improve for $T \geq 200$. The FM-CPR and IM-CPR(O) tests show a similar performance, where the latter has some slight advantages in case of increased sample size especially for the Wald-type test with multiple hypotheses, see Table 4. The IM-

$\rho_1 = \rho_2$	OLS	OLS-HAC	D-CPR	FM-CPR	IM-CPR(O)	IM-CPR(Fb)		
						Data-Dep	b=0.1	b=0.2
				T=100				
0.0	0.057	0.184	0.212	0.086	0.117	0.047	0.058	0.069
0.3	0.200	0.211	0.282	0.143	0.133	0.071	0.061	0.082
0.6	0.526	0.379	0.382	0.242	0.180	0.164	0.082	0.099
0.9	0.922	0.831	0.612	0.601	0.554	0.626	0.419	0.311
				T=200				
0.0	0.053	0.125	0.099	0.066	0.079	0.047	0.053	0.053
0.3	0.192	0.146	0.145	0.105	0.087	0.063	0.056	0.056
0.6	0.529	0.267	0.187	0.166	0.119	0.124	0.060	0.064
0.9	0.932	0.716	0.391	0.471	0.351	0.623	0.204	0.156
				T=500				
0.0	0.046	0.088	0.065	0.054	0.060	0.047	0.052	0.059
0.3	0.195	0.100	0.089	0.075	0.069	0.051	0.051	0.056
0.6	0.558	0.193	0.108	0.104	0.082	0.063	0.054	0.059
0.9	0.940	0.556	0.208	0.338	0.167	0.283	0.075	0.076
				T=1000				
0.0	0.053	0.075	0.065	0.057	0.063	0.058	0.052	0.053
0.3	0.214	0.082	0.076	0.073	0.067	0.059	0.053	0.057
0.6	0.563	0.151	0.090	0.091	0.076	0.063	0.054	0.056
0.9	0.949	0.447	0.149	0.250	0.107	0.120	0.059	0.066

Table 4: Empirical null rejection probabilities for $H_0 : \beta_1 = 5,\ \beta_2 = -0.3$, quadratic specification, Andrews (1991) bandwidth, quadratic spectral kernel, 0.05 level. The last three columns show results of fixed-b inference based on the IM-CPR estimator: in the column "Data-Dep" we have applied a data dependent bandwidth rule in order to determine a value b^*, in the latter two columns we have used fixed values for b.

CPR based fixed-b tests behave decently compared to the standard asymptotic tests. The empirical null rejection probabilities of these tests exceed the 10% threshold rarely and only in high-correlation cases in conjunction with small sample sizes. The fixed-b tests outperform the standard asymptotic tests throughout and are barely size distorted also for the Wald-type tests with multiple hypotheses, where the standard asymptotic tests behave very poorly. Please note that the data dependent bandwidth rules for the fixed-b tests typically lead to a b value of 0.02 or 0.04, for which the fixed-b tests show the worst performance. 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 1 and Figure 2. The figures show that the tests for $b \leq 0.04$ have the highest rejection probabilities and the best results are given for b around 0.1 and/or

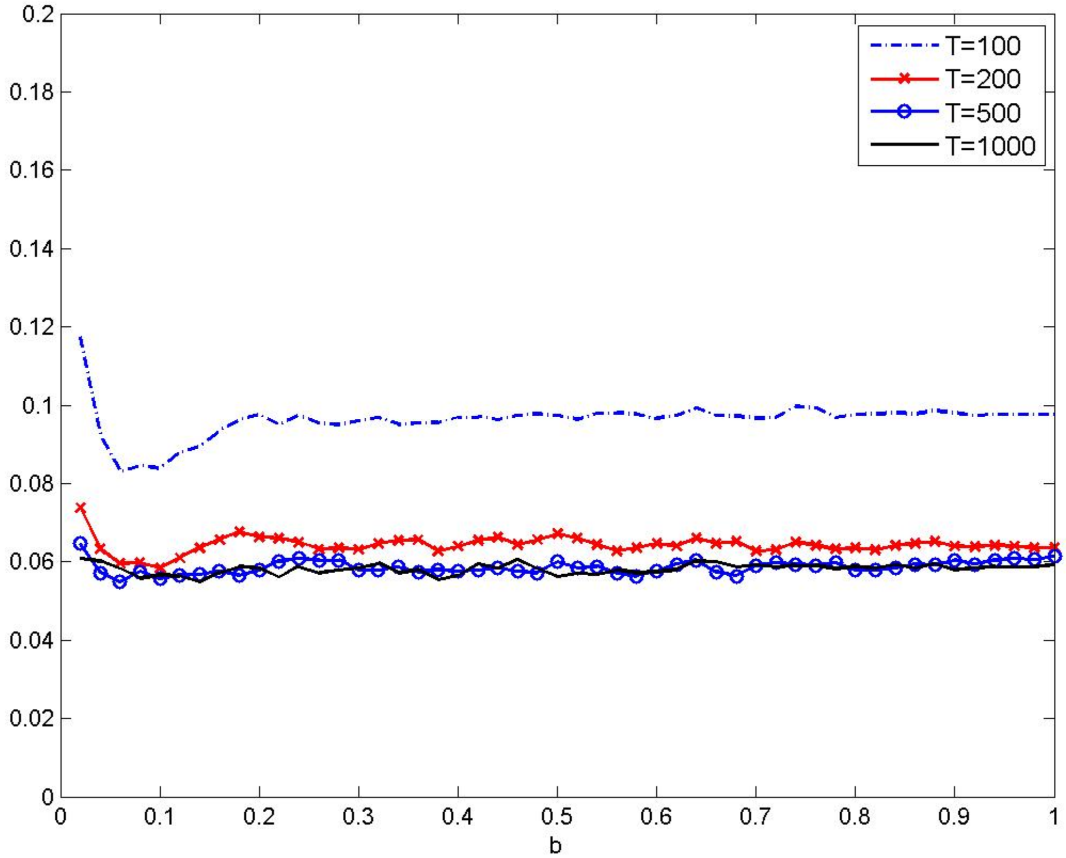


Figure 1: Empirical null rejections, IM-CPR(Fb) inference: t-test for β_1, $\rho_1 = \rho_2 = 0.3$, Bartlett kernel, Andrews (1991) bandwidth

around $b = 1$.

Regarding bandwidth and kernel choice, it should be noted that overall the Andrews (1991) bandwidth choice is better than Newey and West (1994) and that the quadratic spectral kernel dominates the Bartlett kernel with respect to statistical inference. The results for β_2 and the Wald tests are qualitatively very similar. We observe that the sizes for the Wald-type test statistics are slightly larger than the ones of the t-tests due increased degrees of freedom.

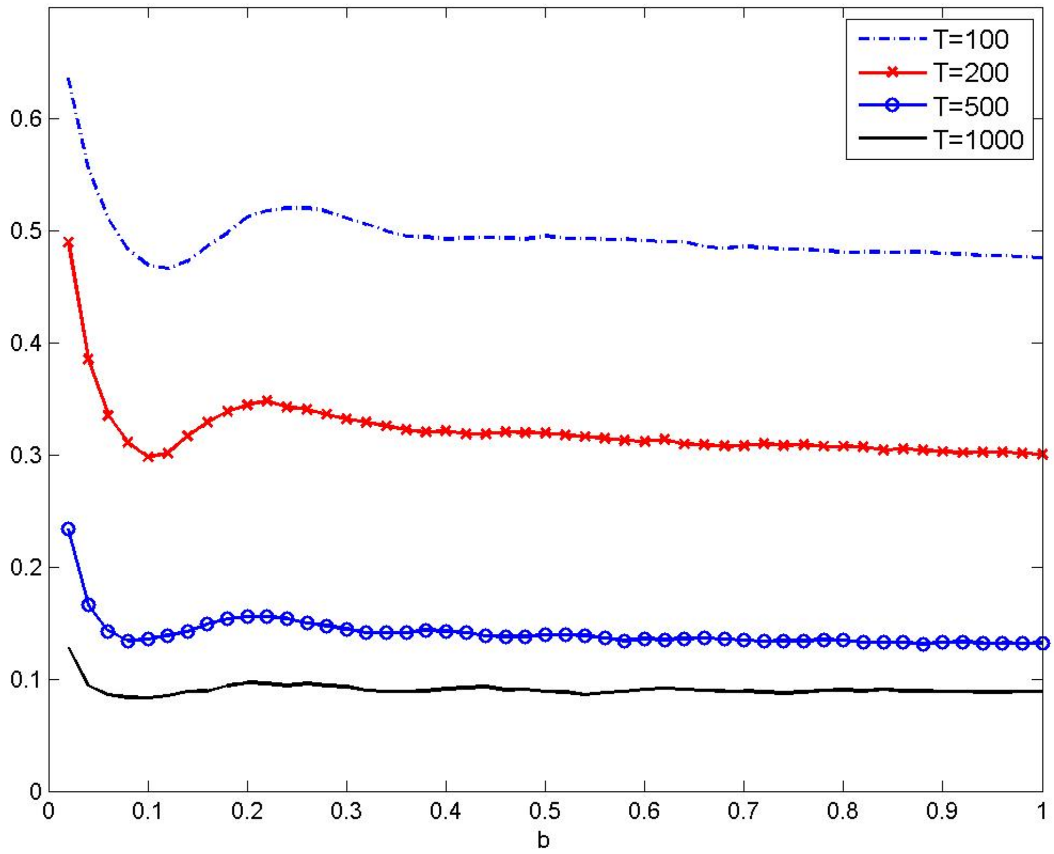


Figure 2: Empirical null rejections, IM-CPR(Fb) inference: t-test for β_1, $\rho_1 = \rho_2 = 0.9$, Bartlett kernel, Andrews (1991) bandwidth

Size-Corrected Power Analysis

We complete this section by considering size-corrected power properties of the tests. Although size-corrections are not feasible in practice, they are a useful tool for theoretical comparisons since they overcome potential over-rejection problems under the null hypothesis. Therefore, we use empirical critical values, which were computed simultaneously in the above empirical null rejection simulation study, in order to hold them constant at 0.05 under the null. Starting from the null 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 of 21 values generated on a grid with mesh 0.05 for β_1 and 0.005 for β_2. The figures are qualitatively similar for the t-test and the Wald-test. For the t-test the curves are very close together, making it difficult to identify any differences.

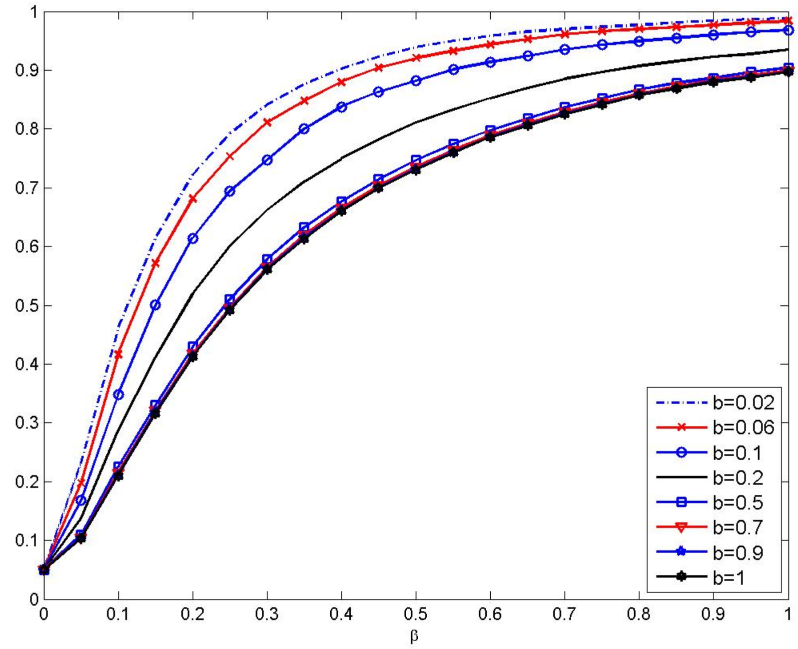


Figure 3: Size Corrected Power, $Wald$-test, T=100, $\rho_1 = \rho_2 = 0.6$, quadratic spectral kernel

As a result, it appears to be more suitable to look at the power curves of the Wald-test.

We consider two types of size-corrected power comparisons: In the first step we compare the IM-CPR(Fb) tests for different values of b. Figure 3 shows that increasing the value of b leads to some power losses. However, these power losses are very small in most cases. In Figures 1 and 2 we have seen that, especially in high-correlated cases, empirical null-rejection tend to be lower with increasing b. The minimal power losses in the size-corrected power study seem to be the price to be paid for less finite sample size distortions.

Comparing both kernels investigated for this simulation study, we see that the quadratic spectral kernel is much more sensitive to the bandwidth choice than the Bartlett kernel. For increasing values of b power, using the quadratic spectral kernel becomes much lower than for the Bartlett kernel. Whereas, as described above, tests using the quadratic spectral 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).

Finally, Figure 4 shows power comparisons for different tests, namely OLS, OLS-HAC,

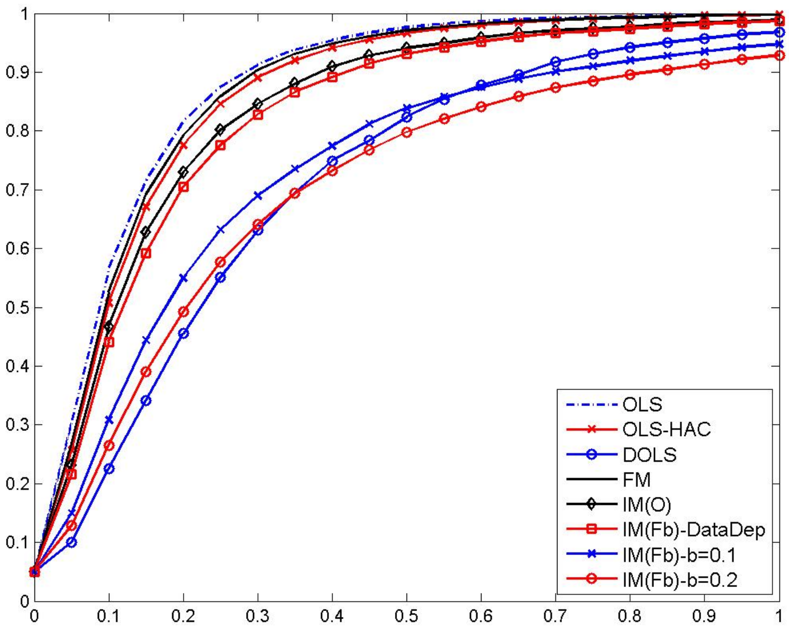


Figure 4: Size Corrected Power, $Wald$-test, T=100, $\rho_1 = \rho_2 = 0.6$, quadratic spectral kernel, Andrews (1991) bandwidth

D-CPR, FM-CPR, IM(O) and IM(Fb) using a data dependent bandwidth rule. We see that the D-CPR based test has by far the smallest power for $T = 100$, but has slightly higher power than the IM-CPR tests for $T = 200$. Throughout OLS and FM-CPR have had the highest-size corrected power, whereas both IM-CPR based tests have shown a small but non-trivial reduction in power. Once again, we see that the use of $\tilde{\omega}^*_{u\cdot v}$ to obtain asymptotically fixed-b inference and less finite sample size distortions comes at the cost of some power loss.

4 Application: EKC Analysis

For the empirical analysis of the environmental Kuznets curve (EKC) hypothesis we consider data for 18 early industrialized countries over the time period 1870-2009 for carbon dioxide (CO_2) emissions and real GDP. All of these quantities are used in per capita terms and transformed to logarithms for the empirical analysis.

The CO_2 emissions data is from the homepage of the Carbon Dioxide Information Analysis

List of Countries

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

Table 5: The sample range is 1870-2009 for GDP and CO_2 with the exception of New Zealand which has 1878 as its starting point.

Center of the US Department of Energy (http://cdiac.ornl.gov), the GPD data was downloaded from the homepage of the Maddison Project (http://www.ggdc.net/maddison/maddison-project/home.htm). CO_2 emissions data for New Zealand starts in 1878. The required long run covariance estimates for the EKC estimation are based on the quadratic spectral kernel and the data dependent bandwidth rule of Andrews (1991).

We consider the quadratic formulation

$$\begin{aligned} e_t &= c + \delta t + \beta_1 y_t + \beta_2 y_t^2 + u_t, \\ y_t &= y_{t-1} + v_t, \end{aligned} \tag{31}$$

as well as the cubic formulation

$$\begin{aligned} e_t &= c + \delta t + \beta_1 y_t + \beta_2 y_t^2 + \beta_3 y_t^3 + u_t, \\ y_t &= y_{t-1} + v_t, \end{aligned} \tag{32}$$

where e_t denotes log per capita CO_2 emissions and y_t denotes log per capita GDP. In terms of model (1) we have $d = 2, m = 1, p_1 = 2$ for the quadratic specification and $d = 2, m = 1, p_1 = 3$ for the cubic specification, respectively.

Prior to estimation of models (31) and (32), two steps need to be performed: the first step is to test the unit root hypothesis for the variable on the right-hand side (i.e. for log per capita GDP).[6] To be more specific, we make use of the Phillips and Perron (1988) t-test and the fixed-b Phillips-Perron unit root test introduced by Vogelsang and Wagner (2013) with the null hypothesis of an unit root for the specification with an intercept and a linear trend. The results are reported in Table 6. The unit root null hypothesis based on the

[6]Following the arguments in Section 2 we, of course, do not have to test the unit root hypothesis for the second and third power of log per capita GDP.

standard Phillips-Perron test is rejected for none of the countries, which forms the basis for our analysis. The PP(fb) unit root test rejects the null hypothesis for log GDP per capita only for the USA.

The second step is to carry out CPR based (non-)cointegration tests for the specifications (31) and (32). Therefore, we employ the extension of the FM-OLS residual based Shin (1994) cointegration test (CT_{FM}) and the FM-OLS residual based extension of the Phillips and Ouliaris (1990) non-cointegration test (PU). Both tests are introduced in Wagner (2013) and have already been investigated for a similar data set with different kernel and bandwidth choices in Grabarczyk and Wagner (2016).We briefly summarize the results given in Table 7: The EKC cointegration hypothesis for the quadratic specification is supported for Austria, Belgium, Denmark, Finland and Switzerland. For Denmark mixed evidence prevails, i.e. the null of cointegration is not rejected for the CT tests, but rejection of the PU non-cointegration test only occurs at the 10% level. When we take the cubic specification into account, we additionally have evidence for Germany. Given the results of the cointegration tests, we consider the following countries for the CPR based estimation of the EKC: Austria, Belgium, Denmark, Finland, Germany and Switzerland.

We turn to the estimation results for the specifications (31) and (32), where we include the estimators considered in Section 3, i.e. OLS, D-CPR, FM-CPR and IM-CPR (for significance tests we include standard t-values as well as fixed-b t-values). The results for the quadratic specification are given in Table 8 and for the cubic specification in Table 9, respectively. Given that the estimated coefficient $\hat{\beta}_3$ in the cubic specification is not significant for Austria, Denmark and Switzerland, it is sufficient to consider the quadratic specification, whereas we consider the cubic specification for the remaining countries. For Finland the estimated coefficient for β_3 is significant, but the quadratic specification shows a better fit (see below).

The results for the quadratic specification for Austria, Denmark, Finland and Switzerland show that the coefficient to squared GDP is negative for all countries indicating an inversed U-shape. For Austria this coefficient is significantly different from zero only for the IM-CPR estimator. The CO_2 emissions data for Austria shows a flat course over the past few decades, which makes this coefficient difficult to estimate. All estimators show similar results for Denmark and Switzerland. The results for the cubic specification are very similar to each other for Belgium and Germany with the exception of the D-CPR estimator, which shows a poor performance in terms of Bias and RMSE for sample sizes about $T = 100$, as shown in the simulation study.

Figures 5-8 show the actual and fitted values as well as the estimated EKCs for the considered countries using the estimated coefficients via IM-CPR from models (31) and (32). The fits are very good for all considered countries especially for the time period after the Second World War. With the exception of some time periods for Austria and Belgium, the fits perform well for the time before and between the two world wars.

In order to estimate the EKCs we use for the explanatory variable $T = 140$ 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 $1, \dots, 140$ and insert these values together with our coefficient estimates. For the cubic specification it is shown that the estimated EKC for Belgium has an inverted U-shape. For Germany we find a N-shape rather than an inverted U-shape.

5 Summary and Conclusions

This paper 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 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 higher 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 power losses. Overall, the simulation results support the findings of Vogelsang and Wagner (2014a) in the linear cointegration case.

We apply the developed methods for the estimation of the EKC using a data set of GDP and CO_2 emissions for 18 early industrialized countries over the period 1870–2009. We find evidence for the existence of a cubic EKC for six countries. Four out of these six countries show evidence for a quadratic EKC. The coefficient estimates are similar across the considered methods for most of the countries.

Future research will move in the following directions: First, the choice of an optimal b value is an interesting but non-trivial problem. Until now we choose the fixed-b values according to one of the data dependent bandwidth rules designed for long run covariance estimation or choose $b = 0.1$ or $b = 0.2$. Second, 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. 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).

Acknowledgments

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

References

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

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

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.

Grabarczyk, P. and M. Wagner (2016). The Environmental Kuznets Curve for Carbon Dioxide Emissions: A Seemingly Unrelated Cointegrating Polynomial Regression Approach. Mimeo.

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

Hong, S.H. and P.C.B. Phillips (2010). Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity. *Journal of Business and Economic Statistics* **28**, 96–114.

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

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

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.

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

Phillips, P.C.B. (1995). Fully Modified Least Squares and Vector Autoregression. *Econometrica* **59**, 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 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.

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.

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

Vogelsang, T.J., and M. Wagner (2013). A Fixed-b Perspective on the Phillips-Perron Tests. *Econometric Theory* **29**, no. 3, 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**, 1–29.

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**, no. 6, 948–967.

Wagner, M., and S.H. Hong (2016). Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference. *Econometric Theory* **32**, no. 5, 1289–1315.

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.

A Proofs

Proof of Proposition 1. We examine the asymptotic behavior of the elements of the last term in (13). 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 (8). This result leads to

$$\left(T^{-2}A_{IM}S^{\xi\prime}S^{\xi}A_{IM}\right)^{-1} = \left(\frac{1}{T}\left(T^{-1/2}A_{IM}S^{\xi\prime}\right)\left(T^{-1/2}A_{IM}S^{\xi}\right)\right)^{-1} \Rightarrow \left(\int f(s)f(s)'ds\right)^{-1}. \tag{33}$$

For the second factor in (13) 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

$$T^{-2}A_{IM}S^{\xi\prime}S^{u} \Rightarrow \int f(s)B_u(s)ds = \omega_{u\cdot v}^{1/2}\int f(s)w_{u\cdot v}ds + \int f(s)W_v(s)'\,\Omega_{vv}^{-1/2}\Omega_{vu}ds, \tag{34}$$

using $B_u(r) = \omega_{u\cdot v}^{1/2}w_{u\cdot v} + \Omega_{uv}(\Omega_{vv}^{-1/2})'W_v(r)$. Multiplying (33) and the second term of (34) 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 (14) follows using integration by parts. The expression for the (conditional) covariance matrix (15) 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 1. 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 1. For part (i) we can use the results already established in Proposition 1 and Corollary 1, 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 (ii).

Note that the adjusted residuals $\tilde{S}_t^{u*}$ defined in (23) coincide with the OLS residuals from the regression

$$S_t^y = S_t^{D\prime} \delta^* + S_t^{X\prime} \beta^* + x_t' \gamma^* + z_t' \kappa^* + S_t^u, \tag{35}$$

which follows immediately using standard projection arguments. For part (ii) we consider the OLS residuals from (35),

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

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 (14) 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 (14), 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 last two results and once again the fact that $[w_{u\cdot v}, w_{u\cdot v}]_s = s$, now gives us

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

□

Proof of Proposition 2. First, we have to make sure using standard calculations that the expression given in (21) (up to $\breve{\omega}_{u\cdot v}$) converges to (15) (up to $\omega_{u\cdot v}$).
Now the assumption given in (20) 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) Proposition 1 that the fixed-b limit of $\tilde{\omega}^*_{u\cdot v}$ is given by $Q_b(\widetilde{P^*}, \widetilde{P^*})$ and therefore $\tilde{V^*} \Rightarrow Q_b(\widetilde{P^*}, \widetilde{P^*})V_{IM}$. (Unconditional) Independence of χ^2_q and $Q_b(\widetilde{P^*}, \widetilde{P^*})$ follows using the same arguments as Vogelsang and Wagner (2014a) in the proof of Theorem 3. $\square$

B EKC Analysis: Estimation Results and Figures

	PP	PP(fb)–One Step	PP(fb)–Two Step
Australia	-1.208	-1.212	-1.255
Austria	-1.826	-1.909	-1.923
Belgium	-1.257	-1.406	-1.443
Canada	-2.482	*-3.124*	*-3.125*
Denmark	-2.366	-2.426	-2.433
Finland	-2.291	-2.415	-2.421
France	-1.931	-2.145	-2.153
Germany	-2.316	-2.692	-2.696
Italy	-1.676	-1.844	-1.860
Japan	-1.701	-1.785	-1.786
New Zealand	-2.588	-2.674	-2.678
Norway	-2.083	-2.119	-2.098
Portugal	-1.728	-1.752	-1.718
Spain	-0.774	-0.944	-0.990
Sweden	-2.462	-2.539	-2.540
Switzerland	-2.758	-2.848	-2.850
United Kingdom	-1.283	-1.553	-1.631
United States	-2.961	**-3.642**	**-3.642**

Table 6: 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, quadratic spectral kernel, Andrews (1991) bandwidth. 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_{FM}	PU	CT_{FM}	PU
Australia	0.061	15.372	0.071	16.076
Austria	0.060	**71.894**	0.051	**72.945**
Belgium	0.044	**91.126**	0.031	**100.652**
Canada	0.081	10.016	0.043	23.476
Denmark	0.050	*49.615*	0.038	*50.256*
Finland	0.037	**92.147**	0.032	**101.698**
France	0.049	42.268	0.044	42.291
Germany	*0.098*	**68.369**	0.049	**74.618**
Italy	*0.089*	26.148	0.075	39.708
Japan	0.083	8.659	0.051	13.015
New Zealand	0.077	11.226	0.069	12.307
Norway	0.051	21.897	0.045	22.385
Portugal	**0.121**	15.815	**0.125**	16.063
Spain	0.063	38.798	0.054	39.527
Sweden	0.061	34.590	0.064	35.799
Switzerland	0.067	**87.659**	0.055	**101.107**
UK	**0.107**	**124.480**	**1.618**	**133.025**
US	0.069	9.665	0.054	19.64
Critical Values ($\alpha = 10\%$)	0.086	45.237	0.081	47.925
Critical Values ($\alpha = 5\%$)	0.106	52.952	0.101	55.926

Table 7: Results for the CT cointegration test using the FM-OLS residuals as well as the FM-OLS residual based $P_{\hat{u}}$ non-cointegration test for the quadratic and cubic specification in conjunction with the Andrews (1991) data dependent bandwidth rule and the quadratic spectral kernel. Numbers in italics denote rejection of the null hypothesis at the 10% level and bold numbers indicate rejection at the 5% level.

	δ	β_1	β_2	Turning Point
Austria				
OLS	-0.014	6.124	-0.277	62618.781
(t-values)	-2.194	1.619	-1.370	
D-CPR	-0.015	6.265	-0.284	62698.940
(t-values)	-3.177	2.212	-1.822	
FM-CPR	-0.016	6.457	-0.290	67805.827
(t-values)	-3.675	2.538	**-2.063**	
IM-CPR	-0.023	10.197	-0.488	34151.260
(t-values)	-4.111	2.883	**-2.530**	
(fixed-b t-values)	-13.389	9.387	**-8.240**	
Denmark				
OLS	-0.007	11.827	-0.590	22415.852
(t-values)	-1.096	9.811	**-10.260**	
D-CPR	-0.022	13.168	-0.623	38848.957
(t-values)	-2.642	6.677	**-6.950**	
FM-CPR	-0.011	12.500	-0.616	25393.955
(t-values)	-1.965	9.918	**-10.302**	
IM-CPR	-0.017	12.776	-0.616	31982.617
(t-values)	-2.316	8.154	**-8.512**	
(fixed-b t-values)	-4.356	15.336	**-16.009**	
Finland				
OLS	-0.034	16.745	-0.792	38762.325
(t-values)	-2.546	8.457	**-8.691**	
D-CPR	-0.001	8.437	-0.385	57166.366
(t-values)	-0.067	2.672	**-2.431**	
FM-CPR	-0.025	15.246	-0.724	37323.312
(t-values)	-2.495	7.571	**-7.125**	
IM-CPR	-0.021	13.982	-0.661	39338.330
(t-values)	-1.522	5.159	**-4.930**	
(fixed-b t-values)	-3.414	11.569	**-11.056**	
Switzerland				
OLS	-0.020	6.910	-0.241	1.7×10^6
(t-values)	-2.871	5.110	**-2.894**	
D-CPR	-0.021	6.888	-0.234	2.5×10^6
(t-values)	-5.432	5.260	**-3.260**	
FM-CPR	-0.024	6.997	-0.231	3.9×10^6
(t-values)	-6.203	5.266	**-3.152**	
IM-CPR	-0.025	9.507	-0.363	4.9×10^5
(t-values)	-5.375	5.831	**-4.157**	
(fixed-b t-values)	-6.767	7.341	**-5.234**	

Table 8: Estimation results for equation (31) with quadratic spectral kernel and Andrews (1991) data dependent bandwidth for long run variance estimation. **Bold** t-values indicate significance at the 5% level and *italic* t-values significance at the 10% level (only considered for the coefficient β_2 here). We consider those countries for which at least a mixed evidence for cointegration is present. The turning points are computed as $\exp(-\hat{\beta}_1/(2\hat{\beta}_2))$.

	δ	β_1	β_2	β_3	Turning Point(s)	
			Belgium			
OLS	-0.007	60.762	-6.062	0.201	17729.300	29013.122
(t-values)	-4.237	2.586	**-2.351**	**2.137**		
D-CPR	-0.001	24.454	-2.195	0.063	11705.299	932135.570
(t-values)	-0.348	0.851	-0.693	0.543		
FM-CPR	-0.006	55.151	-5.463	0.180	15953.742	38163.199
(t-values)	-3.652	2.735	**-2.450**	**2.191**		
IM-CPR	-0.007	79.763	-8.189	0.280	–	
(t-values)	-2.435	2.530	**-2.362**	**2.204**		
(fixed-b t-values)	-4.864	5.053	**-4.718**	**4.403**		
			Germany			
OLS	-0.029	136.576	-14.647	0.528	–	
(t-values)	-2.793	2.988	**-2.840**	**2.710**		
D-CPR	-0.024	134.076	-14.588	0.533	–	
(t-values)	-3.187	2.361	**-2.252**	**2.165**		
FM-CPR	-0.032	180.844	-19.801	0.728	–	
(t-values)	-5.862	3.683	**-3.521**	**3.395**		
IM-CPR	-0.036	173.097	-18.714	0.680	–	
(t-values)	-4.355	2.749	**-2.610**	**2.499**		
(fixed-b t-values)	-5.790	3.655	**-3.470**	**3.322**		

Table 9: Estimation results for equation (32) with quadratic spectral kernel and Andrews (1991) data dependent bandwidth for long run variance estimation. **Bold** t-values indicate significance at the 5% level and *italic* t-values significance at the 10% level (only considered for the coefficients β_2 and β_3 here). We consider those countries for which at least a mixed evidence for cointegration is present. The turning points are computed as $\exp\left(-\frac{\hat{\beta}_2}{3\hat{\beta}_3} \pm \sqrt{\frac{\hat{\beta}_2^2-3\hat{\beta}_1\hat{\beta}_3}{9\hat{\beta}_3^2}}\right)$ provided $\frac{\hat{\beta}_2^2-3\hat{\beta}_1\hat{\beta}_3}{9\hat{\beta}_3^2} \geq 0$.

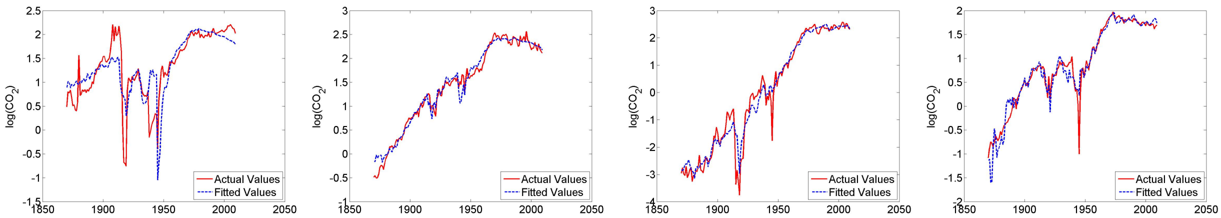
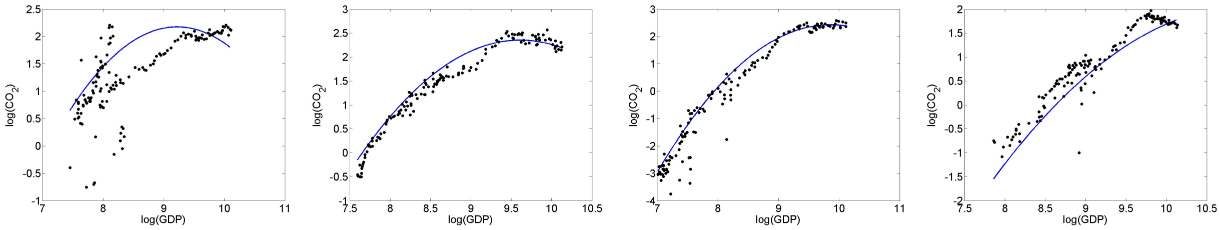


Figure 5: Actual and Fitted Values of log per capita CO_2 emissions estimating the *quadratic* EKC regression equation via IM-CPR.



Austria Denmark Finland Switzerland

Figure 6: EKC estimation for CO_2 using coefficient estimates obtained by IM-CPR in the *quadratic* EKC regression equation.

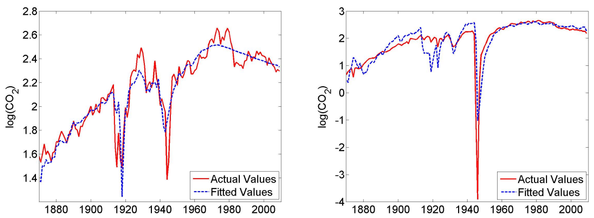
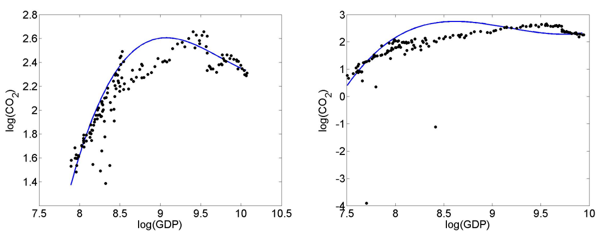


Figure 7: Actual and Fitted Values of log per capita CO_2 emissions estimating the *cubic* EKC regression equation via IM-CPR.



Belgium Germany

Figure 8: EKC estimation for CO_2 using coefficient estimates obtained by IM-CPR in the *cubic* EKC regression equation.

C 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 10: Finite sample bias $(\times 1000)$ and RMSE of four estimators for coefficient β_2, Bartlett kernel.

$\rho_1 = \rho_2$	OLS	OLS-HAC	D-CPR	FM-CPR	IM-CPR(O)	IM-CPR(Fb)		
						Data-Dep	b=0.1	b=0.2
				T=100				
0.0	0.057	0.094	0.282	0.119	0.078	0.055	0.055	0.060
0.3	0.142	0.142	0.264	0.123	0.121	0.100	0.080	0.093
0.6	0.278	0.199	0.270	0.151	0.175	0.204	0.134	0.166
0.9	0.505	0.324	0.333	0.230	0.347	0.602	0.413	0.461
				T=200				
0.0	0.053	0.079	0.131	0.086	0.062	0.051	0.046	0.049
0.3	0.137	0.106	0.141	0.098	0.091	0.094	0.060	0.065
0.6	0.268	0.139	0.152	0.110	0.121	0.188	0.087	0.099
0.9	0.547	0.254	0.224	0.191	0.244	0.516	0.267	0.318
				T=500				
0.0	0.048	0.063	0.074	0.062	0.058	0.055	0.054	0.054
0.3	0.139	0.086	0.090	0.076	0.083	0.065	0.058	0.062
0.6	0.303	0.100	0.098	0.086	0.096	0.086	0.069	0.076
0.9	0.602	0.214	0.166	0.161	0.158	0.229	0.129	0.151
				T=1000				
0.0	0.054	0.064	0.069	0.064	0.060	0.055	0.054	0.054
0.3	0.143	0.078	0.083	0.076	0.078	0.059	0.056	0.058
0.6	0.314	0.086	0.090	0.083	0.082	0.069	0.060	0.062
0.9	0.632	0.161	0.142	0.136	0.117	0.129	0.084	0.095

Table 11: Empirical null rejection probabilities for $H_0 : \beta_2 = -0.3$, quadratic specification, Newey-West bandwidth, Bartlett kernel, 0.05 level. The last three columns show results of fixed-b inference based on the IM-CPR estimator: in the column "Data-Dep" we have applied a data dependent bandwidth rule in order to determine a value b^*, in the latter two columns we have used fixed values for b.

$\rho_1 = \rho_2$	OLS	OLS-HAC	D-CPR	FM-CPR	IM-CPR(O)	IM-CPR(Fb)		
						Data-Dep	b=0.1	b=0.2
				T=100				
0.0	0.062	0.135	0.181	0.079	0.102	0.051	0.045	0.043
0.3	0.134	0.146	0.193	0.101	0.116	0.069	0.058	0.054
0.6	0.246	0.186	0.217	0.119	0.142	0.131	0.084	0.086
0.9	0.412	0.258	0.240	0.139	0.275	0.447	0.244	0.251
				T=200				
0.0	0.052	0.108	0.085	0.060	0.072	0.057	0.050	0.039
0.3	0.134	0.115	0.103	0.079	0.081	0.067	0.054	0.055
0.6	0.283	0.160	0.114	0.095	0.103	0.112	0.068	0.082
0.9	0.503	0.234	0.130	0.111	0.198	0.419	0.152	0.195
				T=500				
0.0	0.052	0.084	0.066	0.057	0.068	0.053	0.053	0.048
0.3	0.146	0.091	0.075	0.071	0.073	0.054	0.056	0.056
0.6	0.306	0.121	0.087	0.082	0.080	0.067	0.060	0.072
0.9	0.613	0.213	0.096	0.110	0.116	0.202	0.076	0.100
				T=1000				
0.0	0.053	0.076	0.061	0.055	0.053	0.047	0.050	0.052
0.3	0.145	0.077	0.070	0.065	0.056	0.048	0.052	0.064
0.6	0.328	0.104	0.075	0.073	0.064	0.050	0.053	0.070
0.9	0.662	0.195	0.081	0.111	0.080	0.094	0.054	0.075

Table 12: Empirical null rejection probabilities for $H_0 : \beta_3 = -0.4$, cubic specification, Andrews (1991) bandwidth, quadratic spectral kernel, 0.05 level. The last three columns show results of fixed-b inference based on the IM-CPR estimator: in the column "Data-Dep" we have applied a data dependent bandwidth rule in order to determine a value b^*, in the latter two columns we have used fixed values for b.

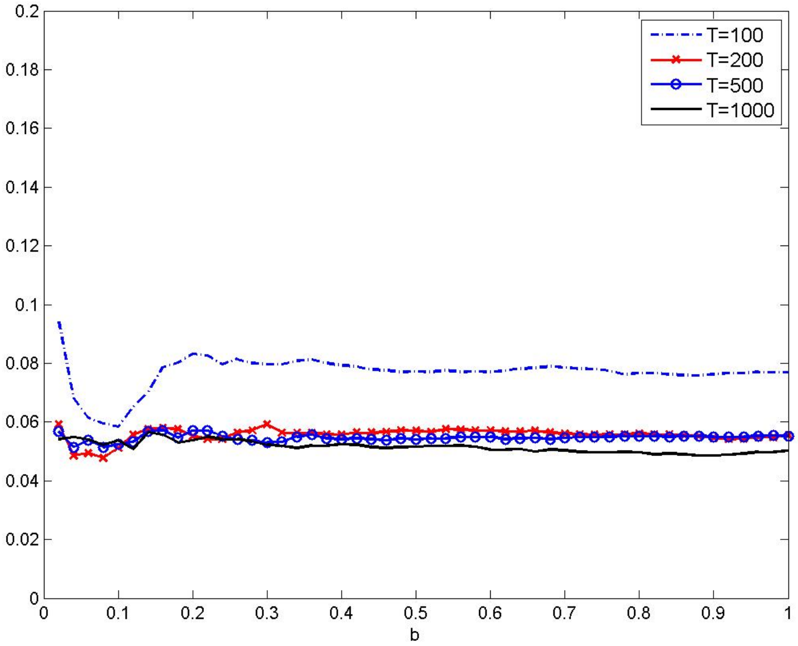


Figure 9: Empirical null rejections, IM-CPR(Fb) inference: t-test for β_2, $\rho_1 = \rho_2 = 0.3$, quadratic spectral kernel, Andrews (1991) bandwidth

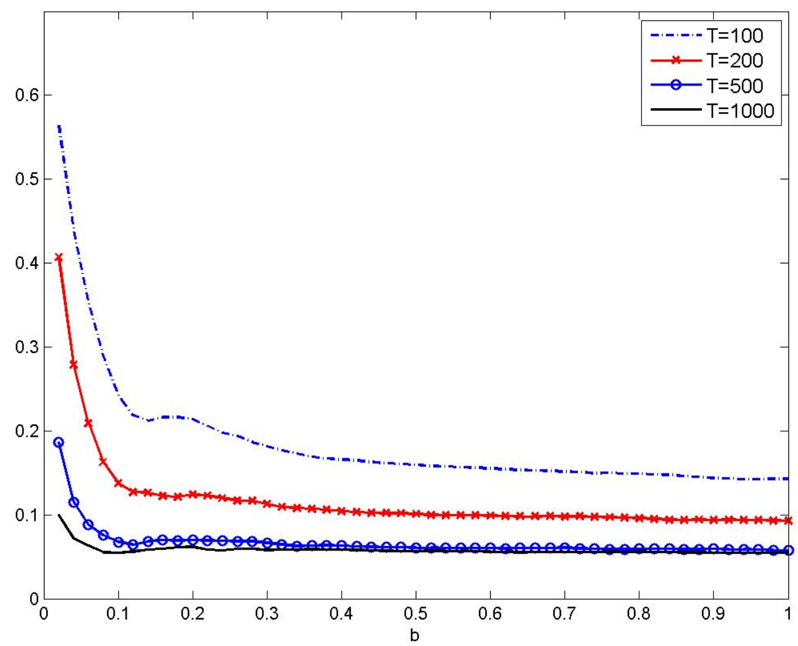


Figure 10: Empirical null rejections, IM-CPR(Fb) inference: t-test for β_2, $\rho_1 = \rho_2 = 0.9$, quadratic spectral kernel, Andrews (1991) bandwidth

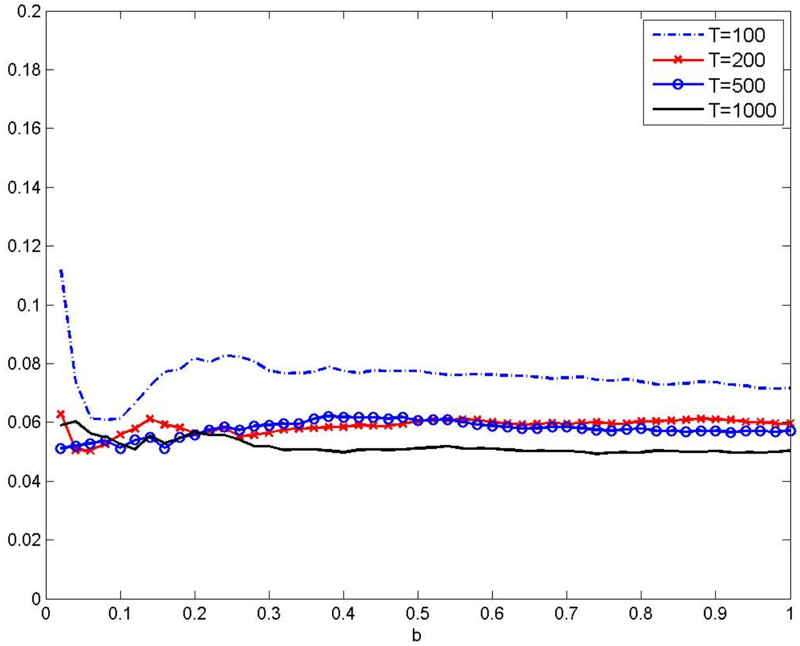


Figure 11: Empirical null rejections, IM-CPR(Fb) inference: $Wald$-test for β_1 and β_2, $\rho_1 = \rho_2 = 0.3$, quadratic spectral kernel, Andrews (1991) bandwidth

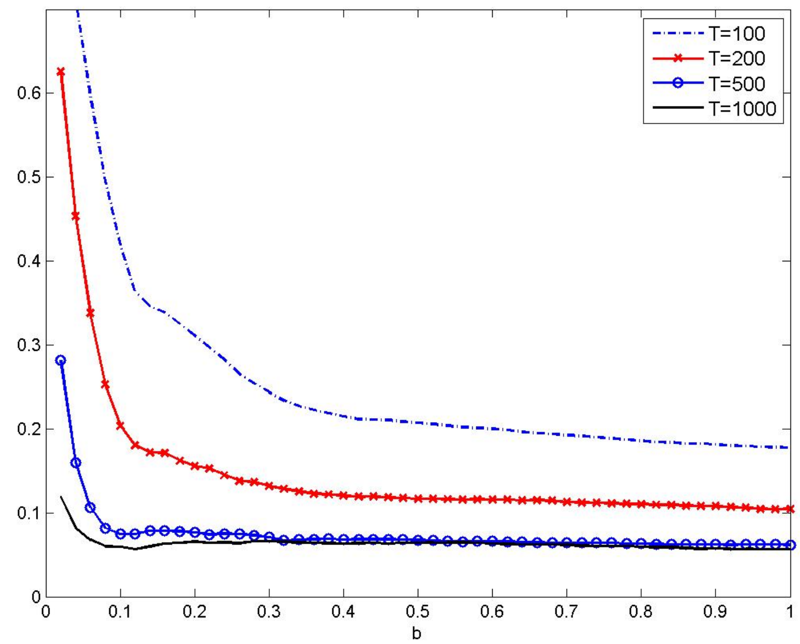


Figure 12: Empirical null rejections, IM-CPR(Fb) inference: $Wald$-test for β_1 and β_2, $\rho_1 = \rho_2 = 0.9$, quadratic spectral kernel, Andrews (1991) bandwidth

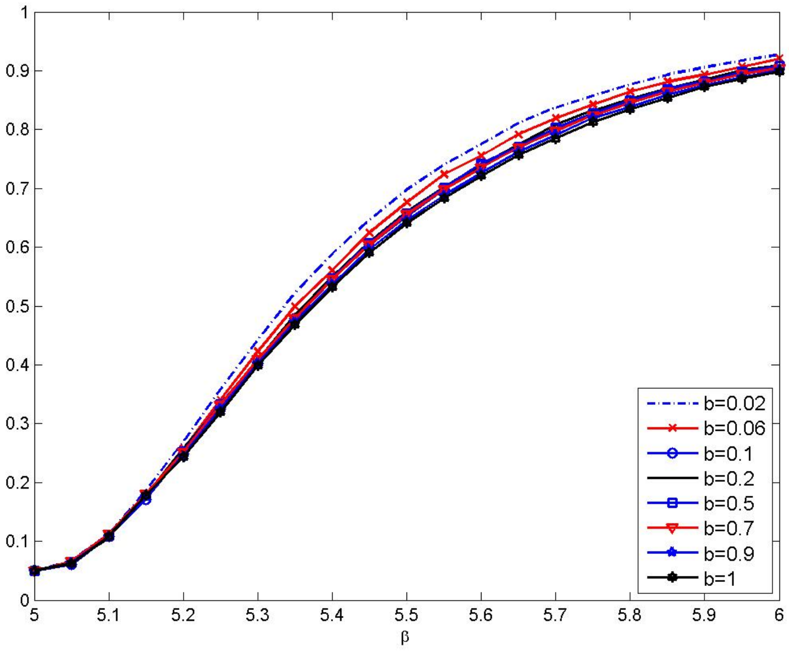


Figure 13: Size Corrected Power, t-test for β_1, T=100, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

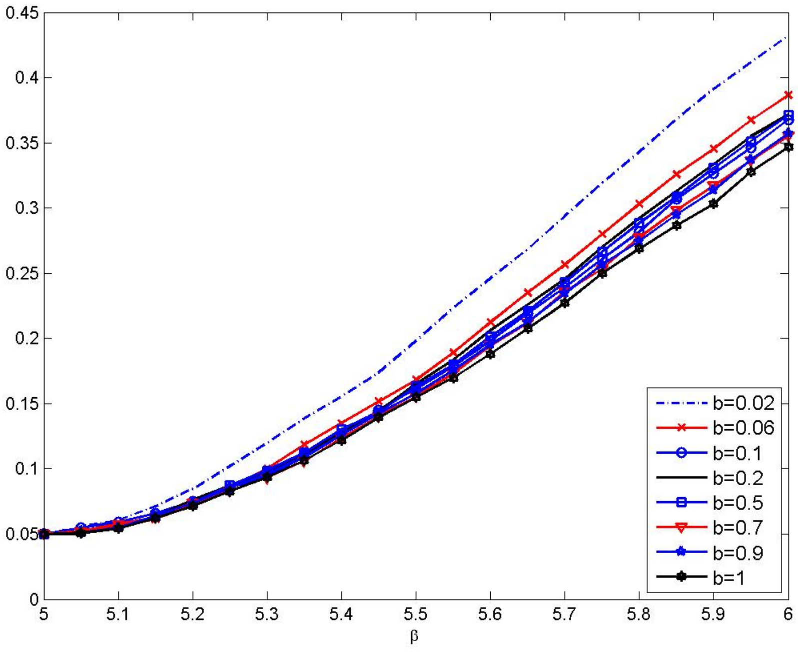


Figure 14: Size Corrected Power, t-test for β_1, T=100, $\rho_1 = \rho_2 = 0.9$, Bartlett kernel

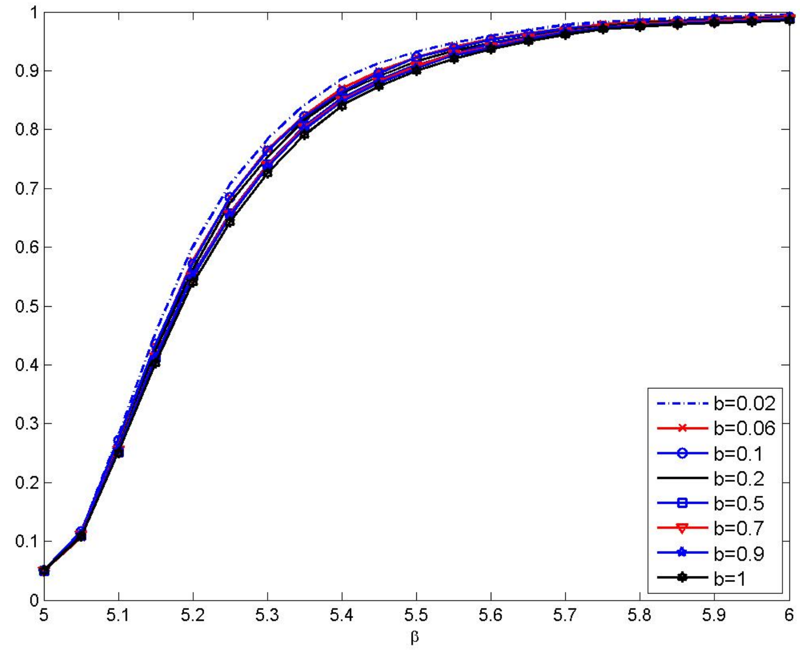


Figure 15: Size Corrected Power, t-test for β_1, T=200, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

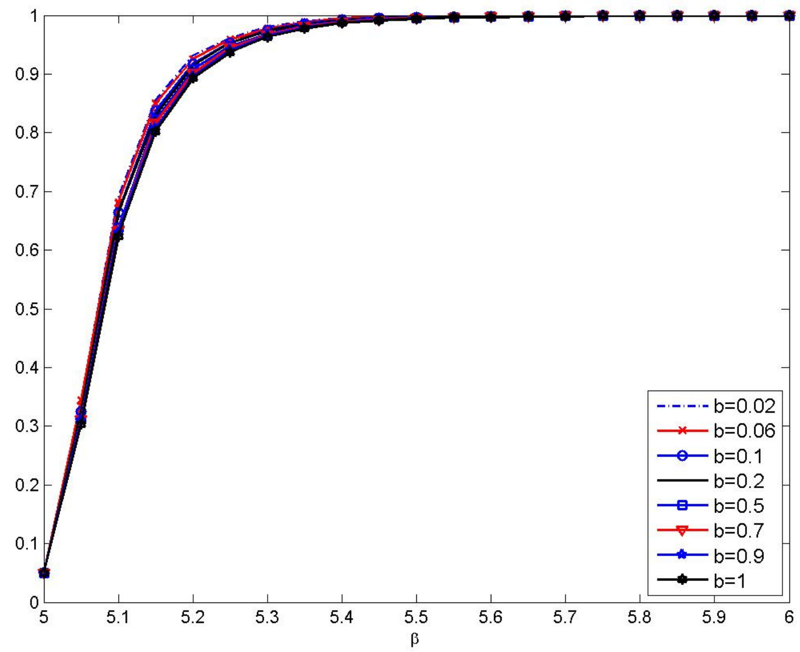


Figure 16: Size Corrected Power, t-test for β_1, T=500, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

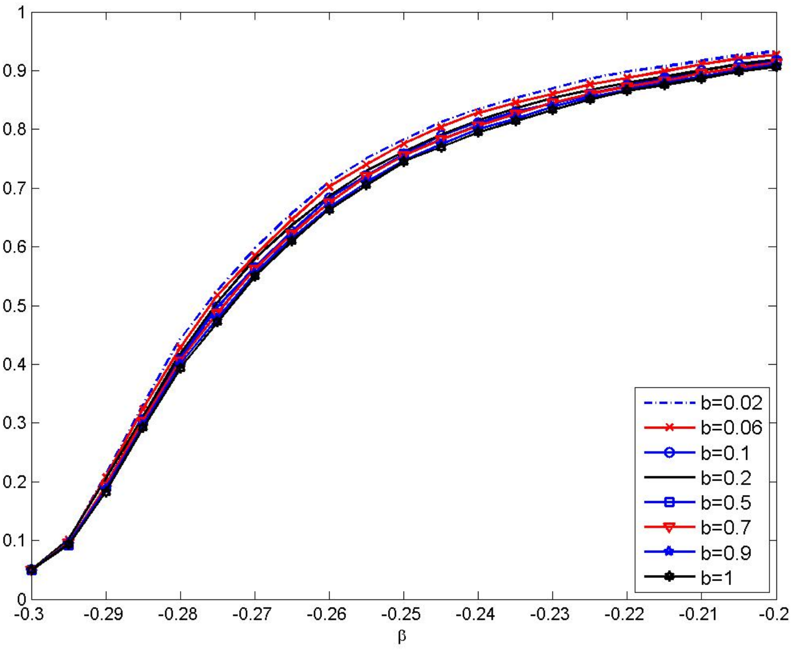


Figure 17: Size Corrected Power, t-test for β_2, T=100, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

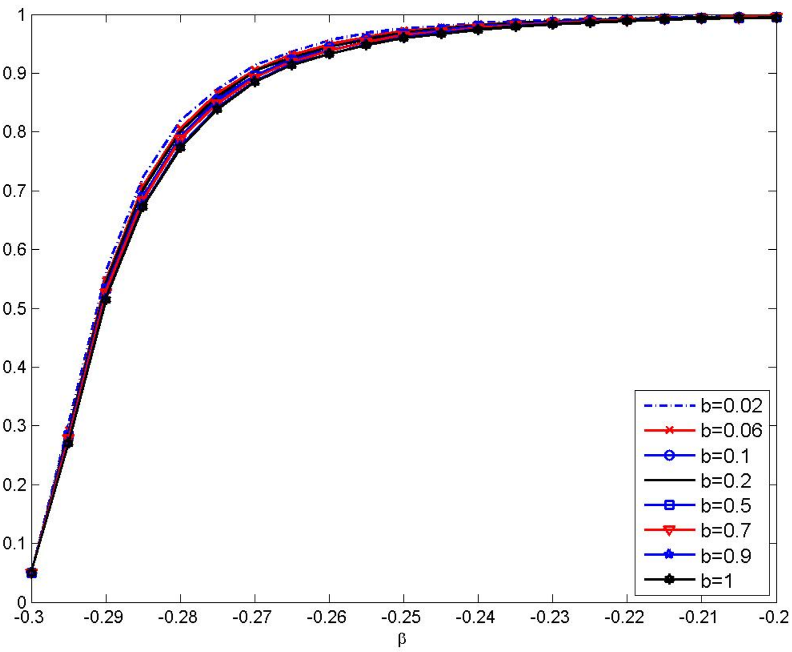


Figure 18: Size Corrected Power, t-test for β_2, T=200, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

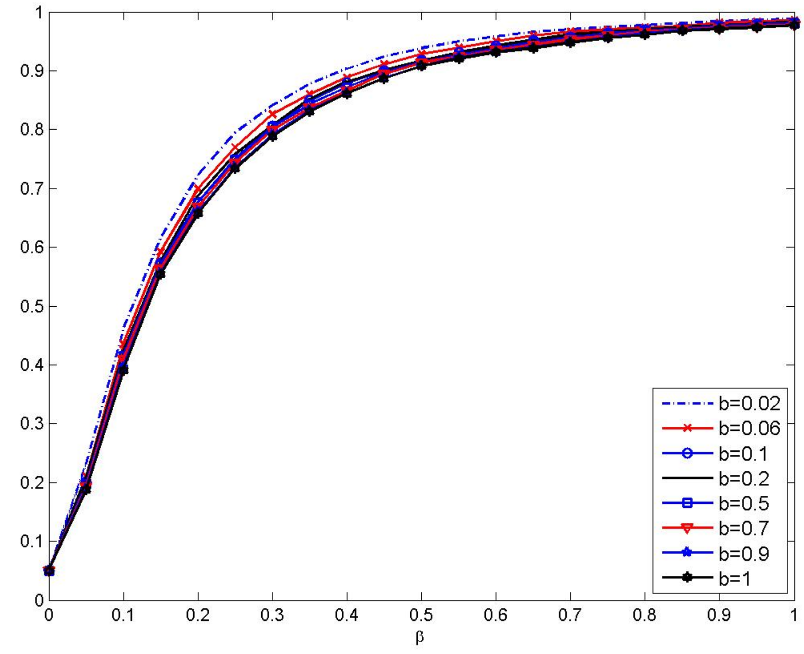


Figure 19: Size Corrected Power, $Wald$-test, T=100, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

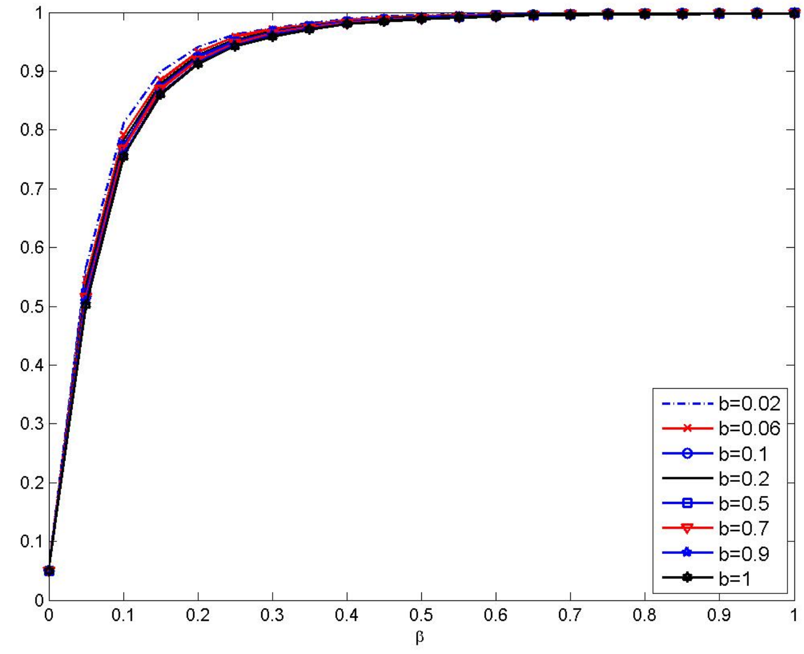


Figure 20: Size Corrected Power, $Wald$-test, T=200, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

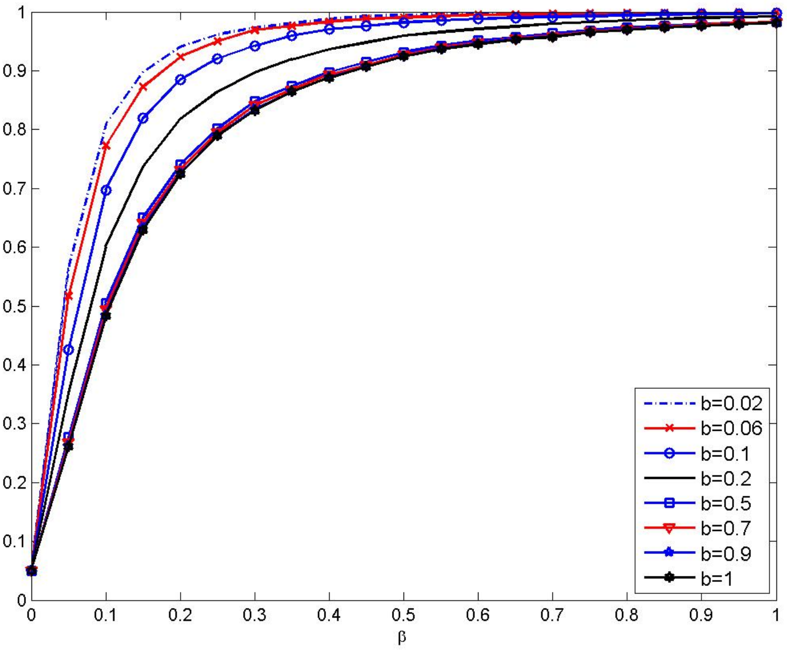


Figure 21: Size Corrected Power, $Wald$-test, T=200, $\rho_1 = \rho_2 = 0.6$, quadratic spectral kernel

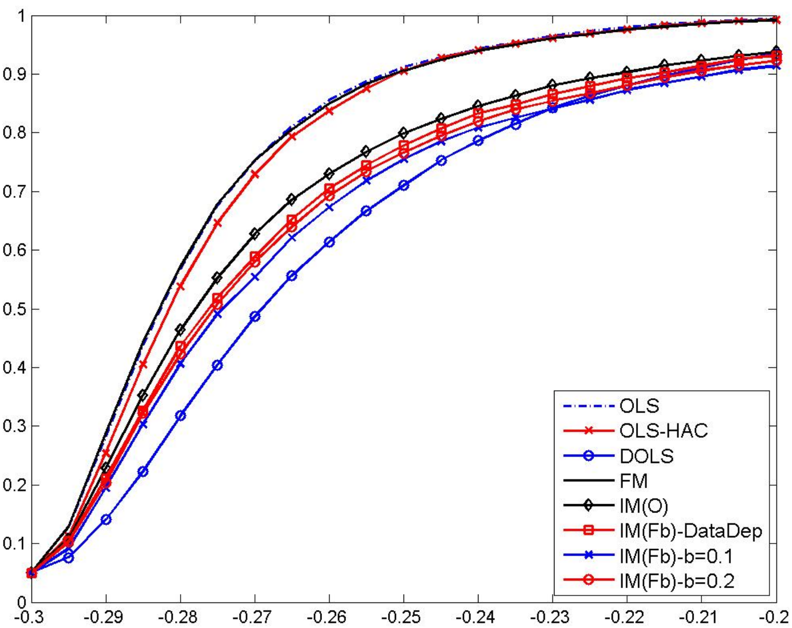


Figure 22: Size Corrected Power, t-test for β_2, T=100, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

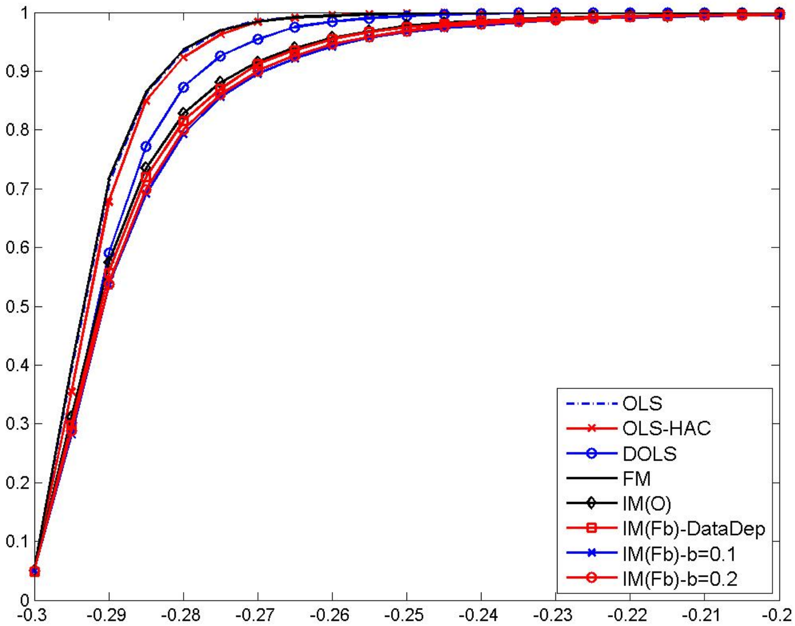


Figure 23: Size Corrected Power, t-test for β_2, T=200, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel

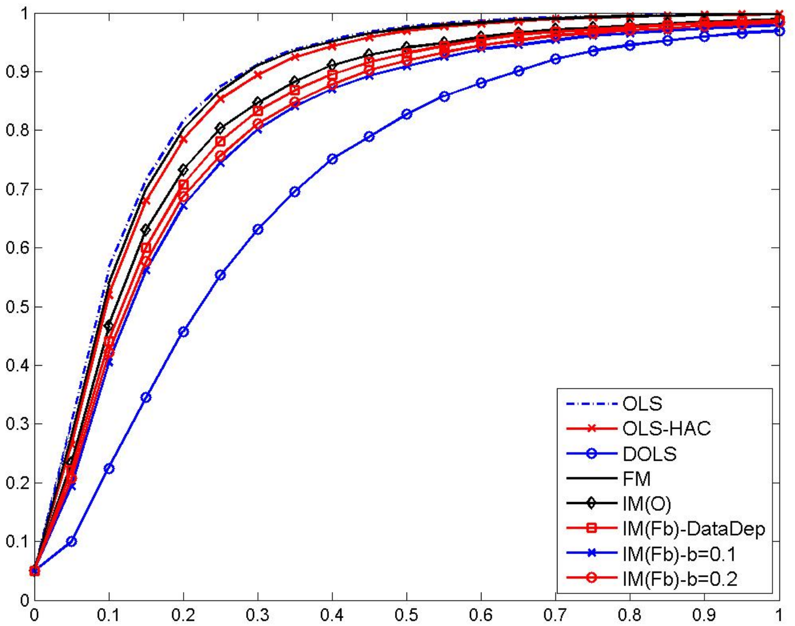


Figure 24: Size Corrected Power, $Wald$-test, T=100, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel, Andrews (1991) bandwidth

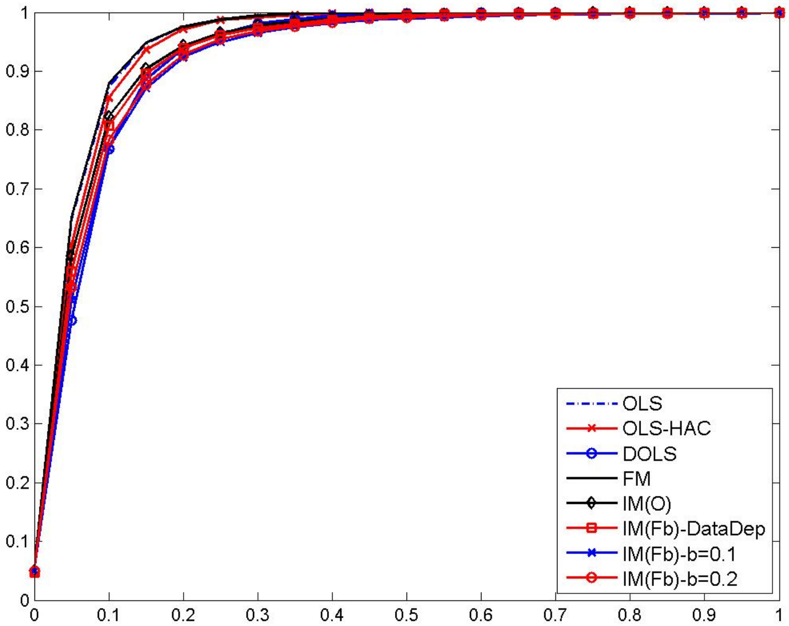


Figure 25: Size Corrected Power, $Wald$-test, T=200, $\rho_1 = \rho_2 = 0.6$, Bartlett kernel, Andrews (1991) bandwidth

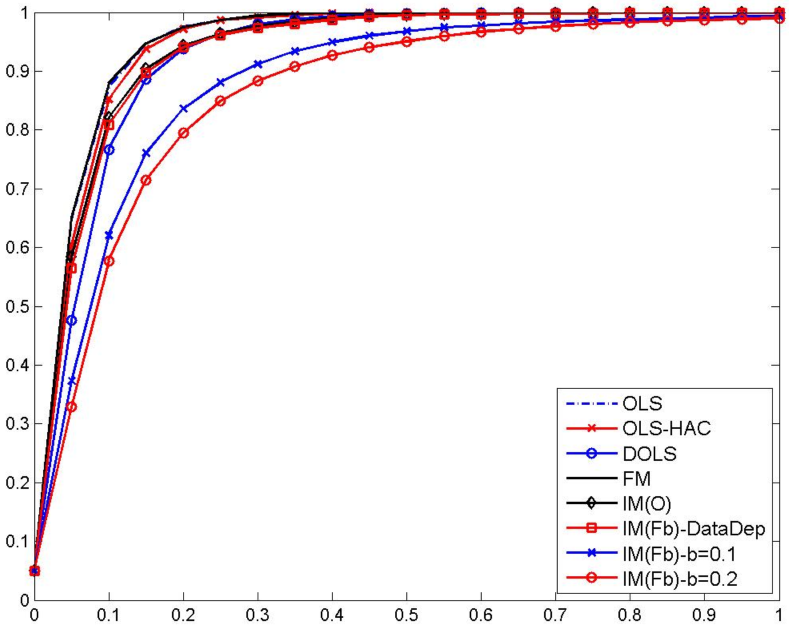


Figure 26: Size Corrected Power, $Wald$-test, T=200, $\rho_1 = \rho_2 = 0.6$, quadratic spectral kernel, Andrews (1991) bandwidth