SFB 823

Panel cointegrating polynomial regressions: Group-mean fully modified OLS estimation and inference

Discussion Paper

Martin Wagner, Karsten Reichold

Nr. 27/2018

SFB 823

Panel Cointegrating Polynomial Regressions: Group-Mean Fully Modified OLS Estimation and Inference

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

Karsten Reichold
Faculty of Statistics
Technical University Dortmund
Dortmund, Germany

November 8, 2018

Abstract

This paper considers group-mean fully modified OLS estimation for a panel of cointegrating polynomial regressions, i. e., regressions that include an integrated process and its powers as explanatory variables. The stationary errors are allowed to be serially correlated, the regressor to be endogenous and – as usual in the nonstationary panel literature – we include individual specific fixed effects. We consider a fixed cross-section dimension, asymptotics in the time dimension only and show that the estimator allows for standard asymptotic inference in this setting. In both the simulations as well as an illustrative application estimating environmental Kuznets curves for carbon dioxide emissions we compare our group-mean estimator with the pooled fully modified OLS estimator of de Jong and Wagner (2018).

JEL Classification: C13, C23

Keywords: Cointegration, Fully Modified OLS, Group-Mean Estimation, Panel Data, Polynomial Transformation

1 Introduction

This paper considers *group-mean* estimation for a panel of cointegrating polynomial regressions (CPRs) in a large time and finite cross-section dimension framework. Cointegrating polynomial regressions include deterministic variables as well as integrated processes and their powers as regressors. The regressors are allowed to be endogenous and the stationary errors are allowed to be serially correlated. For brevity we focus on a simple cubic specification with only one integrated regressor and its square and cube – see (1) and (2) – in this paper. All results generalize, with additional notational complexity only, to the case of multiple integrated regressors and their powers as well as more general deterministic components like time trends. As is commonly done in the nonstationary panel cointegration literature we focus on the case of individual specific fixed effects.

Another reason to confine ourselves here to this simple specification is to facilitate comparison with de Jong and Wagner (2018), who develop pooled modified and fully modified OLS estimators in a large time and large cross-section dimension setting; also exemplified for the cubic specification. The group-mean estimator, i. e., the cross-sectional average of the individual specific FM-OLS coefficient estimators (studied in detail, e. g., in Wagner and Hong, 2016) is, clearly, a natural complement to pooled estimation.

In our simulations we compare the group-mean estimator and hypothesis tests based upon it with the pooled FM-OLS estimator of de Jong and Wagner (2018) and tests based upon it as well as with results obtained when not using the cross-section dimension and using data only from the, without loss of generality, first cross-section member. The results are to a certain extent as expected in that the pooled estimator has the smallest bias and root mean squared error (RMSE), followed by the group-mean estimator and first equation estimator. Tests based upon the group-mean estimator suffer in a variety of scenarios less from severe size distortions than tests based upon the pooled estimator. This happens in particular in situations with a large degree of error serial correlation and regressor endogeneity.

We also briefly illustrate the developed methodology by estimating the environmental Kuznets curve (EKC) for carbon dioxide emissions using the same – long and wide – data sets as de Jong and Wagner (2018). The EKC hypothesis postulates an inverse U-shaped relationship between measures of economic development, typically GDP per capita, and measures of pollution or emissions. The term EKC refers by analogy to the inverted U-shaped relationship between the level of economic development and income inequality postulated by Simon Kuznets (1955) in his 1954 presidential address to the American Economic Association.[1]

The paper is organized as follows: Section 2 presents the setting, the assumptions and the theoretical results. Section 3 contains illustrative results from a simulation study undertaken. Section 4 briefly illustrates the method by estimating EKCs for carbon dioxide emissions using the same data sets as de Jong and Wagner (2018) and Section 5 briefly summarizes and concludes. The proofs are relegated to the appendix. Additional simulation results and information is available in Supplementary Material.

We use the following notation: For sequences we define $\{\cdot\} := \{\cdot\}_{t\in\mathbb{Z}}$, $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$ and $\mathrm{diag}(\cdot)$ denotes a diagonal matrix. With $\Rightarrow$, $\xrightarrow{p}$ and $\xrightarrow{d}$ we denote weak convergence, convergence in probability and convergence in distribution as $T \to \infty$. Brownian motion with variance specified in the context is denoted by $B(r)$ and $W(r)$ denotes a standard Wiener process.

2 Theory

As indicated in the introduction, for sake of brevity we only consider the simple case of a cubic specification with a single unit root regressor, its square and cube as well individual fixed effects in this paper, i.e.,

$$y_{it} = \alpha_i + x_{it}\beta_1 + x_{it}^2\beta_2 + x_{it}^3\beta_3 + u_{it}, \tag{1}$$

$$x_{it} = x_{i,t-1} + v_{it}, \tag{2}$$

where for brevity we assume $x_{i0} = 0$.

The cross-sectionally independent error processes $\{\eta_{it}\} := \{(u_{it}, v_{it})'\}$ are assumed to be random linear processes fulfilling a functional central limit theorem similar to Phillips and Moon (1999,

[1]The empirical EKC literature started in the first half of the 1990s, with early important contributions including Grossman and Krueger (1993) or Holtz-Eakin and Selden (1995). Early survey papers like Stern (2004) or Yandle *et al.* (2004) already count more than 100 refereed publications, with the number growing steadily since then. For more discussion on the empirical literature and theoretical underpinnings of the EKC see, e. g., Wagner (2015). Inverted U-shaped relationships also feature prominently in modelling the relationship between energy or material intensity and GDP per capita (see, e. g., Labson and Crompton, 1993; Malenbaum, 1978). In the exchange rate target zone literature predictive regressions involving an exchange rate and its powers as explanatory variables are widely used (see, e. g., Darvas, 2008; Svensson, 1992). In either of these literatures typically only quadratic or cubic polynomials are considered. Thus, also from this perspective it suffices to describe the estimator in this paper for the cubic specification.

Lemma 3), i.e.,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \eta_{it} \Rightarrow B_i(r) =: \Omega_i^{1/2} W_i(r), \quad 0 \leq r \leq 1, \tag{3}$$

where $W_i(r) := (W_{u_i}(r), W_{v_i}(r))'$, with $B_i(r)$ partitioned analogously, is a bivariate standard Wiener process. The random long run covariance matrices are partitioned as

$$\Omega_i := \begin{pmatrix} \Omega_{u_i u_i} & \Omega_{u_i v_i} \\ \Omega_{v_i u_i} & \Omega_{v_i v_i} \end{pmatrix}. \tag{4}$$

For later usage we also define the half long run covariance matrices partitioned analogously, i. e.,

$$\Delta_i := \begin{pmatrix} \Delta_{u_i u_i} & \Delta_{u_i v_i} \\ \Delta_{v_i u_i} & \Delta_{v_i v_i} \end{pmatrix}, \tag{5}$$

with consequently $\Omega_i = \Delta_i + \Delta_i' - \Sigma_i$, where Σ_i is the random contemporaneous covariance matrix. We denote the time-demeaned variables and the averages over time by, e.g., $\tilde{y}_{it}$ and $\bar{y}_{i.}$, i.e.,

$$\tilde{y}_{it} := y_{it} - \bar{y}_{i.} = y_{it} - \frac{1}{T} \sum_{t=1}^{T} y_{it}, \tag{6}$$

with analogous quantities defined for x_{it} (and its powers), u_{it} and v_{it}. In addition, we write

$$\tilde{X}_{it} := \begin{pmatrix} x_{it} - \overline{x}_{i.} \\ x_{it}^2 - \overline{x_{i.}^2} \\ x_{it}^3 - \overline{x_{i.}^3} \end{pmatrix}. \tag{7}$$

To state our assumptions and results we need to define some additional quantities, i. e., $G_T := \mathrm{diag}(T^{-1}, T^{-3/2}, T^{-2})$, $D_i := \mathrm{diag}(\Omega_{v_i v_i}^{1/2}, \Omega_{v_i v_i}, \Omega_{v_i v_i}^{3/2})$ and $A_i := \left(1, 2\int_0^1 B_{v_i}(r)dr, 3\int_0^1 B_{v_i}^2(r)dr\right)'$.

Assumption 1 *The random processes $\{\eta_{it}\}$ are independent across $i = 1, \ldots, N$, the random matrices (Δ_i, Σ_i) are independent of the Wiener processes $W_i(r)$ for $i = 1, \ldots, N$ and Ω_i is positive definite almost surely for $i = 1, \ldots, N$. Furthermore, it holds for $i = 1, \ldots, N$ and $0 \leq r \leq 1$ that:*

(a) $T^{1/2} G_T \tilde{X}_{i\lfloor rT \rfloor} \Rightarrow \begin{pmatrix} B_{v_i}(r) - \int_0^1 B_{v_i}(s)ds \\ B_{v_i}^2(r) - \int_0^1 B_{v_i}^2(s)ds \\ B_{v_i}^3(r) - \int_0^1 B_{v_i}^3(s)ds \end{pmatrix} = D_i \left(\mathbf{W}_{v_i}(r) - \int_0^1 \mathbf{W}_{v_i}(s)ds\right) =: \tilde{\mathbf{B}}_{v_i}(r),$

(b) $G_T \sum_{t=1}^{T} \tilde{X}_{it} \tilde{u}_{it} \xrightarrow{d} \int_0^1 \tilde{\mathbf{B}}_{v_i}(r) dB_{u_i}(r) + \Delta_{v_i u_i} A_i,$

(c) $G_T \sum_{t=1}^{T} \tilde{X}_{it} v_{it} \xrightarrow{d} \int_0^1 \tilde{\mathbf{B}}_{v_i}(r) dB_{v_i}(r) + \Delta_{v_i v_i} A_i,$

with all quantities converging jointly.

As always in fully modified type estimation, consistent estimators of long run covariances and half long run covariances – based on the OLS residuals $\hat{u}_{it}$ and $v_{it} = \Delta x_{it}$ – are required. This implies restrictions on kernels and bandwidths used. For brevity we simply formulate:

Assumption 2 *The estimators $\hat{\Delta}_i$ and $\hat{\Sigma}_i$ satisfy $\hat{\Delta}_i \xrightarrow{p} \Delta_i$ and $\hat{\Sigma}_i \xrightarrow{p} \Sigma_i$ for $i = 1, \ldots, N$. By definition this implies consistency of $\hat{\Omega}_i := \hat{\Delta}_i + \hat{\Delta}_i' - \hat{\Sigma}_i$ for $i = 1, \ldots, N$.*

We abstain from formulating primitive assumptions that generate our Assumptions 1 and 2. The literature provides several by now well-understood routes to derive these results from primitive assumptions using near epoch dependence concepts, martingale difference sequences or linear processes (see, e. g., de Jong, 2002; Ibragimov and Phillips, 2008; Park and Phillips, 2001). Our formulations and assumptions are similar to de Jong and Wagner (2018) who in turn build upon Phillips and Moon (1999). However, in a finite N setting one can replace the random linear process framework without any substantial loss with more classical assumptions as used, e. g., in Wagner and Hong (2016) in a pure time series setting. As discussed below in Remark 3, the random linear process framework provides fundamental value added only in case of $N \to \infty$. See also the discussion in de Jong and Wagner (2018).

We are now ready to define the group-mean fully modified OLS estimator, i. e., the cross-sectional average of individual specific fully modified OLS estimators (as developed in Wagner and Hong, 2016) of the coefficient vector $\beta := (\beta_1, \beta_2, \beta_3)'$. More precisely, we define for $i = 1, \ldots, N$ the FM-OLS estimator of β from the i-th cross-section member – computed from individual specifically demeaned data – as

$$\hat{\beta}^+(i) := \left(\sum_{t=1}^{T} \tilde{X}_{it} \tilde{X}_{it}' \right)^{-1} \left(\sum_{t=1}^{T} \tilde{X}_{it} \tilde{y}_{it}^+ - C_i \right), \tag{8}$$

with $\tilde{y}_{it}^+ := \tilde{y}_{it} - \Delta x_{it} \hat{\Omega}_{v_i v_i}^{-1} \hat{\Omega}_{v_i u_i}$, $C_i := \hat{\Delta}_{v_i u_i}^+ \left(T, 2\sum_{t=1}^{T} x_{it}, 3\sum_{t=1}^{T} x_{it}^2 \right)'$ and $\hat{\Delta}_{v_i u_i}^+ := \hat{\Delta}_{v_i u_i} - \hat{\Delta}_{v_i v_i} \hat{\Omega}_{v_i v_i}^{-1} \hat{\Omega}_{v_i u_i}$. This allows to define the *group-mean* fully modified OLS estimator as

$$\hat{\beta}_{\mathrm{GM}}^+ := \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}^+(i). \tag{9}$$

Proposition 1 *Let the data be generated by* (1) *and (2) and let Assumptions 1 and 2 be in place. Then it holds for $T \to \infty$,* conditional upon *Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, \ldots, N$, that*

$$G_T^{-1}(\hat{\beta}_{GM}^+ - \beta) \xrightarrow{d} \mathcal{Z} \sim \mathcal{N}(0, V_{GM}), \tag{10}$$

with $\mathcal{N}(0, V_{GM})$ denoting a normal distribution with expectation zero and covariance matrix

$$V_{GM} := \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i\cdot v_i}\left(\int_0^1 \tilde{\mathbf{B}}_{vi}(r)\tilde{\mathbf{B}}_{vi}(r)'dr\right)^{-1} = \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i\cdot v_i}\tilde{M}_{ii}^{-1}, \tag{11}$$

with $\Omega_{u_i\cdot v_i} := \Omega_{u_iu_i} - \Omega_{u_iv_i}\Omega_{v_iv_i}^{-1}\Omega_{v_iu_i}$ the conditional variance of $B_{u_i\cdot v_i}(r) := B_{u_i}(r) - \Omega_{u_iv_i}\Omega_{v_iv_i}^{-1}B_{v_i}(r)$ and $\tilde{M}_{ii}$ defined by the last equality.

Under our assumptions, a consistent estimator of V_{GM} is naturally given by

$$\begin{aligned}\hat{V}_{\mathrm{GM}} &:= \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i\cdot v_i}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T\right)^{-1}, \\ &= G_T^{-1}\hat{S}_{\mathrm{GM}}G_T^{-1},\end{aligned} \tag{12}$$

with $\hat{\Omega}_{u_i\cdot v_i} := \hat{\Omega}_{u_iu_i} - \hat{\Omega}_{u_iv_i}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}$ and $\hat{S}_{\mathrm{GM}}$ defined by the last equality.

The conditional normal limit in conjunction with the availability of a consistent estimator of the covariance matrix as given in (12) leads to standard asymptotic inference. To obtain standard asymptotic behavior of hypothesis tests, we have to take into account that the components of the vector $\hat{\beta}_{\mathrm{GM}}^{+}$ converge at different rates, an issue discussed, e. g., in Sims *et al.* (1990, Section 4) or Wagner and Hong (2016, Section 2.2, p. 1297). A sufficient condition to ensure standard asymptotic behavior is to assume that the constraint matrix fulfills the (asymptotic) restriction posited in the following lemma.

Lemma 1 *Let the data be generated by* (1) *and (2) and let Assumptions 1 and 2 be in place. Consider s linearly independent restrictions collected in*

$$H_0 : R\beta = r, \tag{13}$$

with $R \in \mathbb{R}^{s\times 3}$, $r \in \mathbb{R}^s$ and assume that there exists a non-singular matrix $G_R \in \mathbb{R}^{s\times s}$ such that $\lim_{T\to\infty} G_RRG_T = R^*$, with $R^* \in \mathbb{R}^{s\times 3}$ of rank s. Then it holds under the null hypothesis that the Wald-type statistic

$$W := \left(R\hat{\beta}_{GM}^{+} - r\right)'\left(R\hat{S}_{GM}R'\right)^{-1}\left(R\hat{\beta}_{GM}^{+} - r\right) \tag{14}$$

is asymptotically chi-squared distributed with s degrees of freedom.

Remark 1 *The group-mean estimator is robust to cross-section dependence in the sense that it remains consistent with a zero mean Gaussian mixture limiting distribution even in the presence of* *cross-section dependence. However, the covariance matrix of the asymptotic distribution is different, reflecting the cross-section dependence. Modifying the assumptions to ensure joint convergence for $i = 1, \ldots, N$ of all quantities including cross-products that appear in the limiting covariance matrix leads straightforwardly to "robust" inference in case of cross-section dependence. Denote with $\tilde{M}_{ij} := \int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_j}(r)'dr$ and with $\Omega_{u_i \cdot v_i; u_j \cdot v_j}$ the quadratic covariation (over the interval zero to one) of the processes $B_{u_i \cdot v_i}(r)$ and $B_{u_j \cdot v_j}(r)$. Then the asymptotic covariance matrix of the group-mean estimator given in* (9) *changes from the expression given in* (11) *to*[2]

$$V_{GM}^{rob} := \frac{1}{N^2} \sum_{i,j=1}^{N} \hat{\Omega}_{u_i \cdot v_i; u_j \cdot v_j} \tilde{M}_{ii}^{-1} \tilde{M}_{ij} \tilde{M}_{jj}^{-1}. \tag{15}$$

Remark 2 *Time effects, either put in place instead of individual effects or – more commonly used – in a two-way effects specification with individual and time effects also do not invalidate consistency of the group-mean estimator. However, the limiting distribution is in this case contaminated by second order bias terms related to the presence of cross-section averages of time series limits. In the two-way case the transformed regressor vector, e. g., is given by $\check{X}_{it} := \tilde{X}_{it} - \frac{1}{N}\sum_{j=1}^{N} \tilde{X}_{jt}$, which leads to a partial sum limit (compare Assumption 1) of the form $T^{1/2} G_T \check{X}_{i\lfloor rT \rfloor} \Rightarrow \tilde{\mathbf{B}}_{v_i}(r) - \frac{1}{N}\sum_{j=1}^{N} \tilde{\mathbf{B}}_{v_j}(r) =: \check{\mathbf{B}}_{v_i}(r)$. Thus, the cross-section dependence induced by two-way demeaning shows up in the limit partial sum processes, which in turn leads to second order bias terms also in the limit of $G_T \sum_{t=1}^{T} \check{X}_{it}\check{u}_{it}^{+}$, with $\check{u}_{it}^{+} := \check{u}_{it} - \Delta x_{it}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i}$ and $\check{u}_{it} := \tilde{u}_{it} - \frac{1}{N}\sum_{j=1}^{N} \tilde{u}_{jt}$. With a large cross-section dimension, under appropriate assumptions, $\frac{1}{N}\sum_{j=1}^{N} \tilde{\mathbf{B}}_{v_j}(r)$ fulfills a law of large numbers. This is exploited in the derivation of the large N and large T asymptotic distribution of the pooled estimator in de Jong and Wagner (2018).*

In the cross-sectionally homogenous *case, with $\Delta_i = \Delta$ and $\Sigma_i = \Sigma$ for $i = 1, \ldots, N$, the group-mean estimator can be modified to lead to a nuisance parameter free limiting distribution by using $\check{y}_{it}^{+} := \check{y}_{it} - \Delta \check{x}_{it}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$, with $\check{y}_{it} := \tilde{y}_{it} - \frac{1}{N}\sum_{j=1}^{N} \tilde{y}_{jt}$, as transformed dependent variable and*

$$\check{C}_i := \hat{\Delta}_{vu}^{+} \left(\left(\frac{N-1}{N}\right)^2 \left(T, 2\sum_{t=1}^{T} x_{it}, 3\sum_{t=1}^{T} x_{it}^2\right)' + \frac{1}{N^2} \sum_{j \neq i} \left(T, 2\sum_{t=1}^{T} x_{jt}, 3\sum_{t=1}^{T} x_{jt}^2\right)' \right)$$

as additive correction term when estimating the parameters of the i-th equation with FM-OLS. In this case, e. g., the homogenous long run covariance matrix Ω can be estimated by the cross-section

[2] The quadratic covariation $\Omega_{u_i \cdot v_i; u_j \cdot v_j}$ is equal to the covariance between $B_{u_i \cdot v_i}(r)$ and $B_{u_j \cdot v_j}(r)$ in our setting, i. e.,

$$\Omega_{u_i \cdot v_i; u_j \cdot v_j} = \Omega_{u_i u_j} - \Omega_{u_i v_i}\Omega_{v_i v_i}^{-1}\Omega_{v_i u_j} - \Omega_{u_j v_j}\Omega_{v_j v_j}^{-1}\Omega_{v_j u_i} + \Omega_{u_i v_i}\Omega_{v_i v_i}^{-1}\Omega_{v_i v_j}\Omega_{v_j v_j}^{-1}\Omega_{v_j u_j}.$$

A robust covariance matrix estimator, $\hat{V}_{\mathrm{GM}}^{rob}$ say, is thus immediately available.

average of individual specific long run covariance matrix estimators, i. e., $\hat{\Omega} := \frac{1}{N}\sum_{i=1}^{N}\hat{\Omega}_i$*; and similarly for the other required matrices.*

Remark 3 *To close the theory section note that under (additional) assumptions that ensure the existence of required moments, in particular of* $\mathbb{E}(\Omega_{u_i \cdot v_i}\tilde{M}_{ii}^{-1})$*, it follows that*

$$\sqrt{N}G_T^{-1}\left(\hat{\beta}_{GM}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}(\Omega_{u_i \cdot v_i}\tilde{M}_{ii}^{-1})\right),$$

as $N \to \infty$ *after* $T \to \infty$*. An estimator of the covariance matrix of this limiting distribution is given by*

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\Omega}_{u_i \cdot v_i}\left(G_T\sum_{i=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T\right)^{-1},$$

which – by definition – is equal to $N\hat{V}_{GM}$*, with* $\hat{V}_{GM}$ *the "finite N" covariance matrix estimator given below Proposition 1 in* (12)*.*

3 Finite Sample Performance

We generate data using exactly the same setting as de Jong and Wagner (2018), i. e., data are generated according to the cubic model with fixed effects as given in (1) and (2), with the errors generated from

$$u_{it} = \rho_{1i}u_{i,t-1} + \varepsilon_{it} + \rho_{2i}\nu_{it}, \tag{16}$$

$$v_{it} = \nu_{it} + 0.5\nu_{it,t-1}, \tag{17}$$

with $(\varepsilon_{it}, \nu_{it})' \sim \mathcal{N}(0, I_2)$ cross-sectionally independent. The parameters ρ_{1i} control the level of serial correlation in the error terms u_{it}, and ρ_{2i} control the extent of regressor endogeneity. The parameters ρ_{i1}, ρ_{i2} are cross-sectionally i.i.d. and independent of $(\varepsilon_{it}, \nu_{it})'$. In particular we consider $\rho_{1i} = \rho_1 + \mathcal{U}_{1i}$ and $\rho_{2i} = \rho_2 + \mathcal{U}_{2i}$ with $\mathcal{U}_{1i}, \mathcal{U}_{2i}$ i.i.d. uniform random variables over the interval $[-0.05, 0.05]$, with $\rho_1, \rho_2 \in \{0, 0.3, 0.6, 0.8\}$.[3] The slope parameters are chosen as $\beta_1 = 5$, $\beta_2 = -3$ and $\beta_3 = 0.3$. The individual effects α_i are i.i.d. $\mathcal{N}(0, 1)$. Long-run covariance estimation is performed using the Bartlett kernel in conjunction with the bandwidth selection rule of Andrews (1991). The sample sizes considered are all combinations of $T = 50, 100, 200$ and $N = 5, 10, 25, 50$.

[3] The addition of cross-sectionally i.i.d. random variables to the coefficients ρ_1 and ρ_2 is obviously a simple way of generating data in a random linear process fashion. Considering non-random ρ_{1i} and ρ_{2i} leads to very similar results in the simulations.

For each setting the number of replications is 5,000 and all test decisions are performed at the nominal 5% level. The results obtained with the group-mean estimator, labelled $\hat{\beta}^{+}_{\mathrm{GM}}$ below, are compared with the results obtained from estimating β from the first equation or cross-section member using the FM-CPR estimator of Wagner and Hong (2016), labelled $\hat{\beta}^{+}(1)$ below, and with the pooled estimator of de Jong and Wagner (2018), labelled $\hat{\beta}^{+}_{\mathrm{P}}$ below.

We start with assessing estimator performance measured by bias and RMSE and display the results for β_2 in Table 1, with qualitatively similar results for the other coefficients available in Supplementary Material. Almost throughout the pooled estimator has the smallest bias and RMSE, followed by the group-mean estimator and – as expected – the first equation estimator coming in third place. Clearly, more data are beneficial for estimator performance and this advantage is more pronounced for pooled estimation, where only three slope parameters are estimated, than for group-mean estimation, where three times N slope coefficients are estimated first and then averaged. The relative performance of the group-mean estimator compared to the other two estimators depends upon the exact setting considered. In some cases, when $N = 5$, the first equation estimator even has a slightly smaller bias than the group-mean estimator. Its RMSE is, however, always larger than the RMSE of the group-mean estimator.

We next turn to null rejection probabilities, where we display the results for two null hypotheses. The first is given by $H_0 : \beta_2 = -3$, with the results given in Table 2, and the second is given by $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$, with the results given in Table 3. The single null hypothesis concerning β_2 is tested via a two-sided t-type test using standard normal critical values and the second one via a Wald-type test using chi-squared critical values with three degrees of freedom.

The following main observations emerge: First, the smallest size distortions occur for either the tests based on the pooled or the group-mean estimator. By and large the former leads to the smallest size distortions in case $\rho_1, \rho_2 = 0$ or 0.3 and the latter to the best performance for the larger values of ρ_1, ρ_2. Second, for some configurations the tests based on the pooled estimator lead to the (by far) biggest size distortions in case of large values of ρ_1, ρ_2, with the relative performance disadvantage aggravating with increasing cross-section dimension. This phenomenon of *size divergence* has been found to be widespread for panel unit root and cointegration tests, see, e. g., Hlouskova and Wagner (2006) and Wagner and Hlouskova (2009). As expected, with the exception of the cases where the size distortions of the tests based on the pooled estimator are very large, the tests based on the first equation estimator lead to the largest size distortions. The group-mean estimator based tests

Table 1: Bias and RMSE for the three considered estimators of β_2

		$N=5$			$N=10$			$N=25$			$N=50$		
	$\rho_1=\rho_2$	$\hat\beta_2^+(1)$	$\hat\beta_{\mathrm{P},2}^+$	$\hat\beta_{\mathrm{GM},2}^+$	$\hat\beta_2^+(1)$	$\hat\beta_{\mathrm{P},2}^+$	$\hat\beta_{\mathrm{GM},2}^+$	$\hat\beta_2^+(1)$	$\hat\beta_{\mathrm{P},2}^+$	$\hat\beta_{\mathrm{GM},2}^+$	$\hat\beta_2^+(1)$	$\hat\beta_{\mathrm{P},2}^+$	$\hat\beta_{\mathrm{GM},2}^+$
						Bias ($\times 1,000$)							
$T=50$	0	0.581	-0.061	0.084	-1.207	-0.012	0.178	1.587	0.009	0.081	-0.274	0.011	0.027
	0.3	0.310	-0.071	-0.018	-1.492	-0.009	0.357	2.299	0.013	0.072	-0.132	0.016	0.014
	0.6	-0.880	-0.082	-0.316	-2.605	0.007	0.589	3.256	0.026	0.064	1.376	0.027	-0.059
	0.8	-3.603	-0.093	-1.098	-4.542	0.051	0.485	3.631	0.054	-0.147	5.393	0.050	0.061
$T=100$	0	0.053	-0.018	0.135	-0.187	-0.004	-0.088	0.325	0.001	0.096	0.299	0.003	-0.011
	0.3	0.048	-0.024	0.129	-0.406	-0.004	-0.141	0.198	0.001	0.110	0.342	0.003	-0.032
	0.6	0.032	-0.031	0.143	-0.558	-0.000	-0.209	-0.044	-0.000	0.115	0.364	0.005	-0.062
	0.8	0.023	-0.007	0.394	-0.268	0.019	-0.357	-0.723	-0.010	0.098	0.273	0.010	-0.133
$T=200$	0	0.083	-0.002	0.117	-0.007	-0.001	0.014	0.188	0.001	-0.023	-0.070	-0.000	0.008
	0.3	0.104	-0.002	0.140	0.013	-0.002	0.012	0.240	0.001	-0.021	-0.112	-0.000	0.005
	0.6	0.074	-0.001	0.119	0.027	-0.002	0.008	0.326	0.002	-0.006	-0.154	-0.001	-0.002
	0.8	-0.050	0.004	0.055	-0.155	-0.005	0.071	0.369	0.005	0.003	-0.164	-0.001	-0.001
						RMSE ($\times 10$)							
$T=50$	0	0.747	0.034	0.341	0.810	0.013	0.249	0.727	0.005	0.153	0.746	0.003	0.108
	0.3	0.898	0.046	0.402	0.940	0.018	0.292	0.860	0.008	0.179	0.854	0.005	0.128
	0.6	1.243	0.072	0.536	1.277	0.030	0.391	1.164	0.013	0.241	1.218	0.008	0.171
	0.8	1.936	0.118	0.797	2.122	0.052	0.577	1.782	0.023	0.358	2.017	0.014	0.252
$T=100$	0	0.242	0.011	0.101	0.217	0.005	0.071	0.221	0.002	0.044	0.208	0.001	0.032
	0.3	0.303	0.016	0.128	0.276	0.007	0.089	0.281	0.003	0.056	0.266	0.002	0.040
	0.6	0.441	0.026	0.186	0.395	0.011	0.128	0.415	0.005	0.082	0.402	0.003	0.057
	0.8	0.699	0.048	0.296	0.622	0.021	0.205	0.680	0.009	0.132	0.648	0.006	0.091
$T=200$	0	0.070	0.004	0.032	0.071	0.002	0.023	0.071	0.001	0.014	0.072	0.000	0.010
	0.3	0.093	0.006	0.043	0.094	0.002	0.031	0.095	0.001	0.019	0.095	0.001	0.014
	0.6	0.145	0.010	0.066	0.149	0.004	0.047	0.149	0.002	0.030	0.149	0.001	0.021
	0.8	0.252	0.019	0.115	0.261	0.008	0.082	0.253	0.003	0.051	0.260	0.002	0.037

Table 2: Empirical null rejection probabilities of t-type tests for $H_0 : \beta_2 = -3$ based on the three considered estimators

		$N = 5$			$N = 10$			$N = 25$			$N = 50$		
	$\rho_1 = \rho_2$	$\hat{\beta}_2^+(1)$	$\hat{\beta}_{\mathrm{P},2}^+$	$\hat{\beta}_{\mathrm{GM},2}^+$	$\hat{\beta}_2^+(1)$	$\hat{\beta}_{\mathrm{P},2}^+$	$\hat{\beta}_{\mathrm{GM},2}^+$	$\hat{\beta}_2^+(1)$	$\hat{\beta}_{\mathrm{P},2}^+$	$\hat{\beta}_{\mathrm{GM},2}^+$	$\hat{\beta}_2^+(1)$	$\hat{\beta}_{\mathrm{P},2}^+$	$\hat{\beta}_{\mathrm{GM},2}^+$
$T = 50$	0	0.146	0.085	0.126	0.143	0.072	0.126	0.138	0.066	0.121	0.148	0.075	0.116
	0.3	0.145	0.100	0.112	0.131	0.087	0.102	0.136	0.087	0.097	0.135	0.092	0.091
	0.6	0.140	0.116	0.095	0.123	0.113	0.079	0.134	0.110	0.071	0.132	0.109	0.068
	0.8	0.139	0.140	0.088	0.135	0.136	0.067	0.143	0.142	0.062	0.143	0.137	0.052
$T = 100$	0	0.106	0.066	0.097	0.096	0.067	0.094	0.109	0.066	0.094	0.097	0.059	0.093
	0.3	0.107	0.080	0.091	0.100	0.081	0.083	0.109	0.079	0.084	0.100	0.075	0.079
	0.6	0.110	0.097	0.087	0.104	0.094	0.067	0.118	0.096	0.062	0.110	0.093	0.057
	0.8	0.117	0.112	0.071	0.116	0.108	0.054	0.117	0.113	0.051	0.116	0.104	0.038
$T = 200$	0	0.084	0.060	0.076	0.084	0.058	0.075	0.083	0.061	0.067	0.083	0.057	0.073
	0.3	0.088	0.073	0.075	0.085	0.069	0.074	0.085	0.074	0.066	0.089	0.069	0.071
	0.6	0.098	0.086	0.074	0.093	0.082	0.067	0.093	0.086	0.059	0.097	0.081	0.061
	0.8	0.107	0.098	0.067	0.096	0.092	0.051	0.101	0.091	0.046	0.108	0.085	0.047

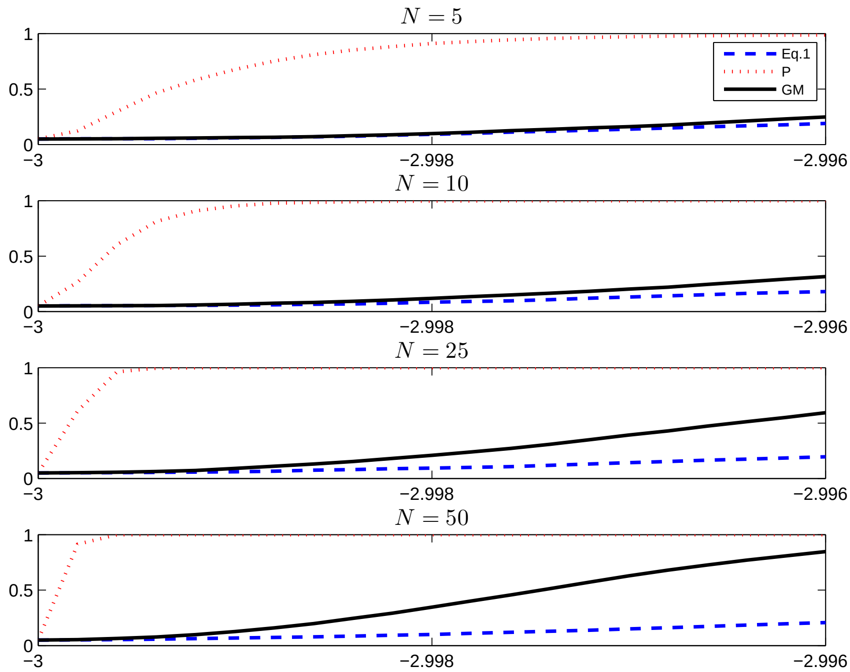
Note: For the pooled FM-OLS estimator, the so-called "standard" covariance matrix estimator is used, see de Jong and Wagner (2018) for details.

Table 3: Empirical null rejection probabilities of Wald-type tests for $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$ based on the three considered estimators

		$N = 5$			$N = 10$			$N = 25$			$N = 50$		
	$\rho_1 = \rho_2$	$\hat{\beta}^+(1)$	$\hat{\beta}_{\mathrm{P}}^+$	$\hat{\beta}_{\mathrm{GM}}^+$	$\hat{\beta}^+(1)$	$\hat{\beta}_{\mathrm{P}}^+$	$\hat{\beta}_{\mathrm{GM}}^+$	$\hat{\beta}^+(1)$	$\hat{\beta}_{\mathrm{P}}^+$	$\hat{\beta}_{\mathrm{GM}}^+$	$\hat{\beta}^+(1)$	$\hat{\beta}_{\mathrm{P}}^+$	$\hat{\beta}_{\mathrm{GM}}^+$
$T = 50$	0	0.220	0.114	0.209	0.205	0.097	0.202	0.218	0.090	0.187	0.224	0.093	0.177
	0.3	0.244	0.156	0.192	0.238	0.137	0.168	0.248	0.127	0.156	0.245	0.141	0.158
	0.6	0.316	0.259	0.203	0.314	0.267	0.167	0.325	0.365	0.157	0.320	0.511	0.184
	0.8	0.451	0.488	0.255	0.451	0.606	0.212	0.468	0.844	0.203	0.466	0.976	0.256
$T = 100$	0	0.142	0.083	0.134	0.139	0.076	0.133	0.145	0.079	0.116	0.136	0.070	0.124
	0.3	0.168	0.107	0.131	0.164	0.109	0.120	0.171	0.113	0.106	0.165	0.101	0.111
	0.6	0.225	0.168	0.146	0.227	0.186	0.128	0.228	0.249	0.118	0.223	0.318	0.134
	0.8	0.336	0.332	0.182	0.356	0.408	0.147	0.356	0.627	0.147	0.338	0.854	0.188
$T = 200$	0	0.110	0.068	0.100	0.110	0.067	0.092	0.106	0.068	0.098	0.100	0.065	0.085
	0.3	0.133	0.093	0.105	0.127	0.090	0.094	0.130	0.089	0.092	0.123	0.083	0.088
	0.6	0.174	0.131	0.122	0.176	0.135	0.111	0.170	0.164	0.104	0.165	0.195	0.121
	0.8	0.257	0.224	0.152	0.249	0.262	0.130	0.246	0.413	0.115	0.237	0.615	0.172

Note: See note to Table 2.

Figure 1: Size-corrected power of t-type tests for $H_0 : \beta_2 = -3$ for $T = 200$, $\rho_1, \rho_2 = 0.3$ and all values of N



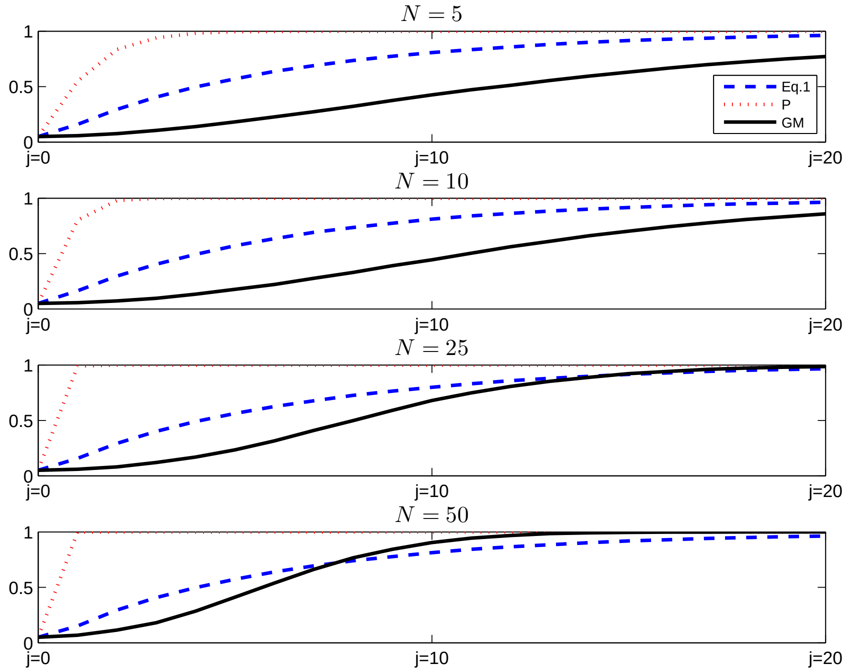
Note: The caption "Eq.1" refers to the test based on the estimator from the first equation, "P" refers to the test based on the pooled estimator and "GM" refers to the test based on the group-mean estimator.

are not prone to size divergence and their size distortions are smaller than those of the tests based on the first equation estimator.

We close this section with a brief look on size-corrected power, considering again the single hypothesis on the coefficient β_2 as well as the multiple hypothesis involving all three coefficients. The rejection probabilities under the alternatives are calculated using an equidistant grid of 21 points for the parameter values (including also the null parameter values) based on the empirical critical values from the null hypothesis simulations. Reflecting the different convergence rates we consider for β_1 the interval $[5, 5.08]$, for β_2 the interval $[-3, -2.996]$ and for β_3 the interval $[0.3, 0.3002]$. Figure 1 shows the results for the t-type tests for β_2 for $T = 200$, $\rho_1, \rho_2 = 0.3$ and all values of N. Figure 2 shows the results for the same configurations for the Wald-type tests involving all three parameters.

Size-corrected power is throughout highest for the tests based on the pooled estimator, which is in line with the superior bias and RMSE performance of this estimator. The second rank depends upon test considered. For the t-type tests typically the tests based on the group-mean estimator

Figure 2: Size-corrected power of Wald-type tests for $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$ for $T = 200$, $\rho_1, \rho_2 = 0.3$ and all values of N



Note: See note to Figure 1.

have second highest power, whereas for the Wald-type tests the tests based on the estimator for the first equation have second highest power. In many configurations the power of the tests based on the group-mean and the first equation estimators is very similar, and distinctly lower than the power of the tests based on the pooled estimator.

Altogether, the simulation evidence is mixed. The pooled estimator exhibits the best performance in terms of bias and RMSE. The group-mean estimator partly leads to tests with the smallest size distortions (in particular in case of large ρ_1, ρ_2) that are – unlike the tests based on the pooled estimator – not susceptible to size divergence as the cross-section dimension increases. However, (size-corrected) power performance of the group-mean based tests is low when compared to the size-corrected power of the tests based on the pooled estimator of de Jong and Wagner (2018).

Table 4: Group-mean EKC estimation results

	Quadratic			Cubic		
	$N = 6$	$N = 19$	$N = 89$	$N = 6$	$N = 19$	$N = 89$
β_1	7.510	7.763	8.609	-24.199	4.439	1043.039
	(9.474)	(14.178)	(2.999)	(-1.513)	(0.375)	(2.485)
β_2	-0.377	-0.390	-0.409	3.190	-0.033	-146.152
	(-8.475)	(-12.782)	(-2.256)	(1.772)	(-0.025)	(-2.669)
β_3				-0.133	-0.013	6.685
				(-1.980)	(-0.261)	(2.800)
TP	21,225	21,024	36,912	17,087	21,595	4,174
				497	0	512
de J&W	22,771	20,240	531,250	$--$	127,784	43,231
					28,054	443

Note: The turning points (TP) are computed as $\exp\left(-\frac{\hat{\beta}_1}{2\hat{\beta}_2}\right)$ in the quadratic case and if real valued as $\exp\left((\pm 1)(-\frac{\hat{\beta}_1}{3\hat{\beta}_3} + (\frac{\hat{\beta}_2}{3\hat{\beta}_3})^2)^{1/2} - \frac{\hat{\beta}_2}{3\hat{\beta}_3}\right)$ in the cubic case. The row labelled de J&W presents the turning points obtained in de Jong and Wagner (2018, Tables 7 and 8) using pooled FM-OLS estimation for the same specifications and data sets. The numbers in brackets are t-statistics.

4 The Environmental Kuznets Curve for Carbon Dioxide Emissions

In this section we briefly illustrate the developed estimator by estimating the EKC for carbon dioxide (CO_2) emissions. The dependent variable is the logarithm of CO_2 emissions per capita and the explanatory variables are the logarithm of GDP per capita and its powers. We consider both the quadratic and the cubic specification and include country fixed effects. As in the simulation section we use the Bartlett kernel and the Andrews (1991) bandwidth selection rule for long run covariance estimation.

We use the same data sets as de Jong and Wagner (2018). These are the *long data set* with $N = 19$ countries for $T = 135$ years and the *wide data set* with $N = 89$ countries and $T = 54$. The long data set has originally been used in Wagner *et al.* (2018) and ranges from $1878 - 2013$ for 19 early industrialized countries.[4] We also consider the subset of six of these 19 countries with data

[4] The 19 countries are given by Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and USA. Note that the data are in fact available from 1870 onwards, with the exception of CO_2 emissions for New Zealand. Considering all 19 countries with 1878 as starting point is merely done to use exactly the same balanced panel data set as de Jong and Wagner (2018). Of course, whether the panel is balanced or not is irrelevant even from a computational perspective for group-mean estimation. A detailed description of the data including the sources is contained in Wagner *et al.* (2018).

available for $1870-2013$ analyzed in more detail in a seemingly unrelated regressions (SUR) setting in Wagner *et al.* (2018). These six countries are Austria (AT), Belgium (BE), Finland (FI), the Netherlands (NL), Switzerland (CH) and the United Kingdom (UK). The country list for the wide data set, with time span $1960 - 2013$ is available as Table 5 in the Supplementary Material.

Table 4 displays the estimation results including the estimated turning points. For comparison the last row of the table displays the turning points obtained in de Jong and Wagner (2018) using pooled FM-OLS estimation. Let us start with the quadratic specification, which strictly speaking may be too simple given that the third order coefficient is significant (marginally) for $N = 6$ and $N = 89$. For all three cross-section dimensions, the coefficient to squared ln(GDP) is significantly different from zero and negative, implying an inverted U-shape. The turning points are well within the sample range for all three data sets. For the long data set the estimated turning points are quite similar, for $N = 6$ and $N = 19$, for both group-mean and pooled estimation. However, whereas de Jong and Wagner (2018) find an extremely large out of sample turning point at about 531,000 for the wide data set in the quadratic specification, group-mean estimation leads to a more standard (in-sample) estimate of the turning point at about 37,000 also when $N = 89$.[5]

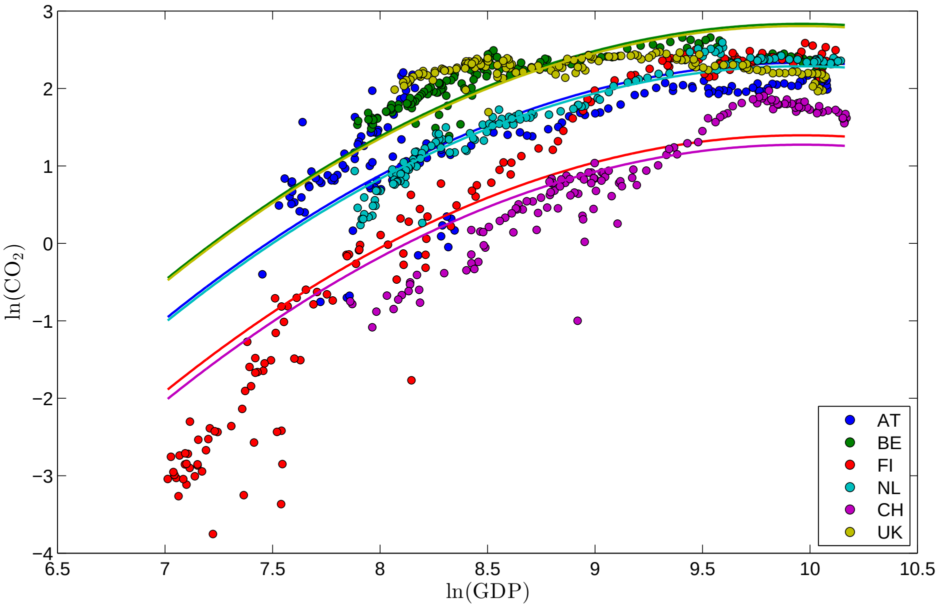
For the long data set, group-mean estimation leads to very similar results as for the quadratic specification also for the cubic specification. This is not the case for the pooled estimator, where the estimated turning points (or even the presence of a turning point) are sensitive to the polynomial degree considered – despite the fact that the coefficient to the third power is insignificant for the long data set throughout when using the pooled estimator. The results obtained with the group-mean estimator for the wide data set and the cubic specification are less satisfactory. The third order coefficient is significant *but* positive, indicating a U-shaped rather than an inverted U-shaped behavior around the larger turning point. The larger turning point is estimated at the very low level of $4,174$ when using the group-mean estimator. For the wide $N = 89$ data set the results obtained with the pooled estimator are more plausible than those obtained with the group-mean estimator, with an inverted U-shaped behavior around the larger turning point estimated at about 43,000.

We close this section with a graphical inspection of the estimation results obtained for the quadratic specification with country fixed effects for $N = 6$ in Figure 3.[6] The figure shows both scatterplots of

[5]The sample ranges are 794 to $31,393$ (measured in 1990 Geary-Khamis Dollars) for the long data set and 340 to $100,959$ (in 2015 US-Dollars) for the wide data set.

[6]We plot the results for the simpler quadratic specification since the third order coefficient is only marginally

Figure 3: Scatter plot and estimated EKC relationship for CO_2 emissions for the $N = 6$ countries over the period 1870–2013



Note: The lines display the results of inserting equidistant points from the sample range of ln(GDP) in the estimated quadratic relationship with the individual effects estimated by $\hat{\alpha}_i := \bar{y}_{i.} - (\bar{x}_{i.}, \overline{x^2_{i.}})\hat{\beta}^+_{\text{GM}}$.

the data as well as the fitted lines (using equidistant points from the sample range of the logarithm of GDP per capita). The country fixed effects are found to be essentially identical for Austria and the Netherlands, for Belgium and the United Kingdom, and very similar for Finland and Switzerland. Apart from the vertical differences due to the country fixed effects, the figure also displays that group-mean estimation results into an "average shape" of the fitted lines that is potentially a bit too steep for Belgium and the UK, and maybe too flat for Finland and Switzerland. For Austria and the Netherlands the shape of the estimated EKC fits the data quite nicely.[7]

5 Summary and Conclusions

This paper extends the set of fully modified OLS estimators available for panel cointegrating polynomial regressions by considering a group-mean fully modified OLS estimator that complements the

significant with a t-statistic of 1.98. The results are very similar for the cubic specification, with the corresponding figures available upon request.

[7]Note for completeness that the poolability tests of Wagner *et al.* (2018), commencing from a seemingly unrelated regressions framework, reject the null hypothesis that the EKC can be fully pooled across these six countries. Figures 7 and 8 in the Supplementary Material display the results obtained when estimating the EKCs separately for Austria, Belgium, the Netherlands and the UK on the one hand and for Finland and Switzerland on the other.

pooled modified and fully modified OLS estimators of de Jong and Wagner (2018). For brevity the results are illustrated for the simple case of a cubic specification with only one integrated regressor and individual fixed effects, which is the most widely-used specification in applications. The results extend with notational complications only to the more general specifications considered in the time series case in Wagner and Hong (2016).

Our assumptions are similar to Phillips and Moon (1999), i. e., we consider a random linear process framework, which also allows for comparability with de Jong and Wagner (2018) who use similar assumptions too. One difference to these two papers is that we consider asymptotics only in the time dimension and consider a fixed cross-section dimension. We show that the group-mean estimator in this setting not only allows for standard asymptotic inference, but is also robust to cross-section dependence. In the presence of cross-section dependence, as expected, only the covariance matrix of the group-mean estimator changes. Using a cross-section dependence robust estimator of the covariance matrix leads to valid inference also in this case. Considering a finite cross-section dimension simplifies the analysis in case of cross-section dependence substantially, as one does not need to specify a more or less restrictive precise form of cross-section dependence that allows to utilize cross-section limit theory. In the presence of time effects the group-mean estimator remains consistent, however, its asymptotic distribution is contaminated by second order bias terms rendering valid inference in general difficult. It is an open question for future research to investigate the asymptotic behavior of the group-mean estimator in the two-way fixed effects case in a large cross-section setting.

The simulation results are to a certain extent as expected with the pooled estimator leading to smaller bias and RMSE than the group-mean estimator. However, hypothesis tests based upon the group-mean estimator often lead to the smallest size distortions in case of large error serial correlation and regressor endogeneity. Furthermore, they are not susceptible to *size divergence*, which is partly found to be a problem for the pooled estimator of de Jong and Wagner (2018).

In our small illustrative application the group-mean estimator leads to plausible results for all three data sets considered and both the quadratic and the cubic specification with only one exception. The exception occurs for the wide data set and the cubic specification, where the estimated larger turning point is around a U-shape and unreasonably small. For the data sets used the pooled estimator appears to be a bit more sensitive, leading to unreasonably large turning points in two cases and no turning points at all in one case.

Acknowledgements

Financial support from Deutsche Forschungsgemeinschaft via the Collaborative Research Center SFB823 *Statistical Modelling of Nonlinear Dynamic Processes* (Projects A3 and A4) is gratefully acknowledged. The authors thank Fabian Knorre for suggesting to present the empirical results as given in Figure 3. At the time of the preparation of this manuscript the first author served as Chief Economist of the Bank of Slovenia. The views expressed in this paper are, however, solely those of the authors and not necessarily those of the Bank of Slovenia or the European System of Central Banks. On top of this the usual disclaimer applies.

References

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

Darvas, Z. (2008). Estimation Bias and Inference in Overlapping Autoregressions: Implications for the Target-Zone Literature. *Oxford Bulletin of Economics and Statistics* **70**, 1–22.

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

de Jong, R.M., Wagner, M. (2018). Panel Cointegrating Polynomial Regression Analysis and the Environmental Kuznets Curve. Dortmund: SFB823 Discussion Paper 22/18.

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

Grossman, G.M., Krueger, A.B. (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.

Hlouskova, J., Wagner, M. (2006). The Performance of Panel Unit Root and Stationarity Tests: Results From a Large Scale Simulation Study. *Econometric Reviews* **25**, 85–116.

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

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

Labson, B.S., Crompton, P.L. (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.

Malenbaum, W. (1978). *World Demand for Raw Materials in 1985 and 2000.* New York: McGraw-Hill, EMJ Mining Information Services.

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

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

Phillips, P.C.B., Moon, H.R. (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* **67**, 1057–1111.

Sims, C.A., Stock, J.H., Watson, M.W. (1990). Inference in Linear Time Series Models with some Unit Roots. *Econometrica* **58**, 113-144.

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

Svensson, L.E.O. (1992). An Interpretation of Recent Research on Exchange Rate Target Zones. *Journal of Economic Perspectives* **6**, 119–144.

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

Wagner, M., Hlouskova, J. (2009). The Performance of Panel Cointegration Methods: Results From a Large Scale Simulation Study. *Econometric Reviews* **29**, 182–223.

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

Wagner, M., Grabarczyk, P., Hong, S.H. (2018). Fully Modified OLS Esimation and Inference for Seemingly Unrelated Cointegrating Polynomial Regressions and the Enivironmental Kuznets Curve for Carbon Dioxide Emissions. *Mimeo.*

Yandle, B., Bjattarai, M., Vijayaraghavan, M. (2004). Environmental Kuznets Curves: A Review of Findings, Methods, and Policy Implications. Research Study 02.1 update, PERC.

Appendix: Proofs

Proof of Proposition 1.

Assumptions 1 and 2 directly imply that

$$G_T \sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}' G_T \xrightarrow{d} \int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)' dr \tag{18}$$

$$G_T \sum_{t=1}^{T} \tilde{X}_{it}\tilde{u}_{it}^{+} \xrightarrow{d} \int_0^1 \tilde{\mathbf{B}}_{v_i}(r) dB_{u_i \cdot v_i}(r) + \Delta_{v_i u_i}^{+} A_i, \tag{19}$$

converge jointly, with $\tilde{u}_{it}^{+} := \tilde{u}_{it} - v_{it}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i}$, $\Delta_{v_i u_i}^{+} := \Delta_{v_i u_i} - \Delta_{v_i v_i}\Omega_{v_i v_i}^{-1}\Omega_{v_i u_i}$ and A_i as given in the main text.

This immediately implies – for the parameter estimator from the i-th equation – that

$$\begin{aligned} G_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) &= \left(G_T \sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}' G_T\right)^{-1} \left(G_T \sum_{t=1}^{T} \tilde{X}_{it}\tilde{u}_{it}^{+} - G_T C_i\right) \\ &\xrightarrow{d} \left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)' dr\right)^{-1} \int_0^1 \tilde{\mathbf{B}}_{v_i}(r) dB_{u_i \cdot v_i}(r). \end{aligned} \tag{20}$$

Conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ the limiting distribution given in (20) is normal with expectation zero and covariance matrix $\Omega_{u_i \cdot v_i}\left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)' dr\right)^{-1}$. This in turn implies the – conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, \ldots, N$ – asymptotic normality result for the group-mean estimator given in the main text in (10) and (11). □

Proof of Lemma 1.

Under the null hypothesis, with the assumptions on the restriction matrix R in place, the Wald-type statistic given in (14) can be written as

$$\begin{aligned} W &= \left((G_R R G_T) G_T^{-1}(\hat{\beta}_{\text{GM}}^{+} - \beta)\right)' \left((G_R R G_T) G_T^{-1} \hat{S}_{\text{GM}} G_T^{-1} (G_R R G_T)'\right)^{-1} \left((G_R R G_T) G_T^{-1}(\hat{\beta}_{\text{GM}}^{+} - \beta)\right) \\ &\xrightarrow{d} (R^* \times \mathcal{Z})' \left(R^* V_{\text{GM}} R^{*\prime}\right)^{-1} (R^* \times \mathcal{Z}), \end{aligned} \tag{21}$$

with $\mathcal{Z}$, as shown in Proposition 1, conditionally $\mathcal{N}(0, V_{\text{GM}})$ distributed. This shows the conditional – and hence unconditional – asymptotic chi-squared null distribution of the Wald-type statistic. □