ECONOMETRIC REVIEWS
2023, VOL. 42, NO. 4, 358–392
https://doi.org/10.1080/07474938.2023.2178141

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

Martin Wagner[a,b,c] and Karsten Reichold[d]

[a]Department of Economics, University of Klagenfurt, Klagenfurt, Austria; [b]Bank of Slovenia, Ljubljana, Slovenia; [c]Institute for Advanced Studies, Vienna, Austria; [d]Department of Statistics, TU Dortmund University, Dortmund, Germany

ABSTRACT

We develop group-mean fully modified OLS (FM-OLS) estimation and inference for panels 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 integrated regressors – allowed to contain drifts – to be endogenous and, as usual in the panel literature, we include individual-specific fixed effects and also allow for individual-specific time trends. We consider a fixed cross-section dimension and asymptotics in the time dimension only. Within this setting, we develop cross-section dependence robust inference for the group-mean estimator. In both the simulations and an illustrative application estimating environmental Kuznets curves (EKCs) for carbon dioxide emissions we compare our group-mean FM-OLS approach with a recently proposed pooled FM-OLS approach of de Jong and Wagner.

KEYWORDS

Cointegrating polynomial regression; cross-section dependence; drift; fully modified OLS; group-mean estimation; panel data

JEL CLASSIFICATION

C13;C23;Q20

1. Introduction

We develop *group-mean* fully modified OLS (FM-OLS) estimation and inference for panels of cointegrating polynomial regressions (CPRs) in a large time ($T \to \infty$) and finite cross-section (N fixed) setting. CPRs, a term coined by Wagner and Hong (2016), include deterministic variables as well as integrated processes, potentially with drifts, and their powers as regressors. The stochastic regressors are allowed to be endogenous and the stationary errors are allowed to be serially correlated. For notational brevity, we only discuss a simple specification, the cubic CPR with only one integrated regressor, see (1) and (2) below. The cubic and – probably even more so – the quadratic single regressor CPR are the most widely-used specifications for the analysis of, e.g., environmental Kuznets curves (EKCs). Thus, considering the cubic case simplifies notation considerably whilst containing all elements required for a typical EKC analysis. All results extend, at the price of increased notational rather than mathematical complexity, straightforwardly to higher-order powers and multiple integrated regressors, compare for a pure time series setting Wagner and Hong (2016). With respect to deterministic regressors we consider individual-specific intercepts only or individual-specific intercepts and individual-specific linear trends; this can also be generalized without additional mathematical complexities to more general deterministic regressors.

The article is closely related to de Jong and Wagner (2022), who consider *pooled* FM-OLS-type estimators for CPRs in a setting with both a large cross-section and time dimension and

CONTACT Martin Wagner Martin.Wagner@aau.at Department of Economics, University of Klagenfurt, Klagenfurt, Austria.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/07474938.2023.2178141.

with a cross-sectional i.i.d. structure. Considering a finite cross-section dimension and asymptotic analysis only for a large time series dimension renders it, of course, impossible to develop a joint or sequential asymptotic normality result for the group-mean FM-OLS estimator.[1] However, the finite cross-section dimension offers some room to consider a more general setting than de Jong and Wagner (2022) in two important ways: First, we allow for the presence of *drifts*, i.e., linear time trends, in the integrated regressors, which are a prominent feature in many macroeconomic and financial time series. The presence of drifts, as is known from standard unit root and cointegration analysis, see, e.g., West (1988), can lead to asymptotic normality of estimated coefficients in the time series unit root case. We show that similar results hold also in the CPR case, in which higher-order polynomial trends are the dominant features of the powers of the integrated regressors with drifts. It turns out that whether and if so for which slope coefficients asymptotic normality prevails depends, in addition to the presence of drifts, also upon the presence or absence of individual-specific linear trends in the regression model. In this respect, it is important to note that for applying the developed estimators and tests no knowledge concerning the presence or absence of drifts is required. Second, we allow for very general forms of cross-section dependence by providing *robust* test statistics that lead to asymptotically valid inference despite cross-section dependence. As is well-known, for macro-panels, which is an important difference to classical micro-panels, the assumption of cross-sectional independence is very likely unrealistic. Consequently, being able to perform cross-section dependence robust inference in conjunction with our group-mean estimator increases applicability substantially, *nota bene* without the need to posit a specific model for cross-section dependence like, e.g., a factor structure.

In a simulation study, we compare the group-mean estimators, both OLS and FM-OLS, with the pooled FM-OLS estimator of de Jong and Wagner (2022). In addition to assessing estimator performance, we also compare the performance, i.e., null rejection probabilities and "size-corrected" power, of a variety of tests based upon these estimators. Many of the results are as expected and in line with asymptotic theory, e.g., the strong negative effects of error serial correlation, endogeneity, and cross-section dependence on the performance of the estimators, where – as expected – the group-mean OLS estimator is most strongly affected. By construction, the pooled FM-OLS estimator leads in most cases to the smallest bias and RMSE. The overall conclusion for hypothesis testing is to use the cross-section dependence robust version of tests based on the group-mean FM-OLS estimator. These tests are, by construction, least affected by cross-section dependence and are much less affected than, e.g., the test based on the pooled FM-OLS estimator by large values of error serial correlation and regressor endogeneity (and are the only ones asymptotically valid in the presence of cross-section dependence). Furthermore, even in the absence of cross-section dependence, the cross-section dependence robust test statistics perform at least at par with the non-robust counterparts. Altogether, this makes the cross-section dependence robust tests based on the group-mean FM-OLS estimator the preferred choice.

We briefly illustrate the developed methodology by estimating EKCs for carbon dioxide emissions using the same data sets as de Jong and Wagner (2022), i.e., two long data sets with $N = 6$ and $N = 19$ countries and about $T = 130$ observations over time and one wide data set with $N = 89$ countries and $T = 54$ observations over time. The EKC hypothesis postulates an inverted 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

[1]Given that many macro-panel data sets have a small cross-section dimension, e.g., also two of the data sets used in our illustration with six and 19 countries, it is not ex ante clear that it is always necessary or beneficial to consider large cross-section dimensions. Of course, in situations with N large compared to T, asymptotics in N in addition to T is important and useful. One main value added that large N asymptotics provides – at the standard $\sqrt{N}$-rate – in addition to large T asymptotics, is unconditional asymptotic normality of estimators (under appropriate assumptions). Of course, in case of large N, especially large with respect to T, asymptotics in N is an important aspect. However, unconditional asymptotic normality is not necessary for asymptotic standard inference, which can be based on a conditional asymptotic normality result when $T \to \infty$ only.

postulated by Simon Kuznets (1955) in his 1954 presidential address to the American Economic Association.[2] A key finding in our illustrative application is that cross-section robust inference makes a difference. The coefficient to the third-order power of the logarithm of GDP per capita is significantly different from zero for the wide data set only, both with and without individual-specific linear trends included. Relying upon standard inference only indicates the necessity for a cubic specification also for the two long data sets for one of the specifications. The group-mean FM-OLS-based turning points (TPs) for the long data sets are larger than those found in de Jong and Wagner (2022) for the $N=6$ data set and very similar for the $N=19$ data set. For the wide data set group-mean FM-OLS estimation leads to very small or no TPs. For this data set, pooled estimation leads to more *plausible* results.

The article is organized as follows: Section 2 presents the setting, the assumptions, and the theoretical results, separated – for didactic reasons – in three subsections according to different settings concerning the absence or presence of drifts. Section 3 contains some illustrative simulation results. Section 4 briefly illustrates the method by estimating EKCs for carbon dioxide emissions using, as mentioned above, the same data sets as de Jong and Wagner (2022) and Section 5 briefly summarizes and concludes. The proofs are relegated to Appendix A and Appendix B provides the country list for the wide data set. Supplementary Material available upon request contains additional simulation results.

We use the following notation: $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$ and $\text{diag}(\cdot)$ denotes a diagonal matrix. With $\Rightarrow$, $\xrightarrow{p}$ and $\xrightarrow{d}$ we denote weak convergence, convergence in probability and convergence in distribution, respectively, 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. $\text{Var}(z)$ denotes the covariance matrix of a vector z and $\text{Cov}(z_1, z_2)$ denotes the cross-covariance matrix of two vectors z_1 and z_2.

2. Theory

As mentioned in the introduction, in this article, we discuss the cubic specification with a single unit root regressor only. With respect to deterministic regressors, we allow for either individual-specific intercepts (i.e., fixed effects) only or individual-specific intercepts and individual-specific linear time trends. The integrated regressors x_{it} *potentially* contain individual-specific drifts μ_i, i.e.,:

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

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

Mainly to relate the article to de Jong and Wagner (2022), see the discussion below Assumption 3, we use the same assumptions as in that (companion) article. Thus, we assume that the cross-sectionally independent error processes $\{\eta_{it}\}_{t\in\mathbb{Z}} := \{(u_{it}, v_{it})'\}_{t\in\mathbb{Z}}$ are random linear processes fulfilling a functional central limit theorem similar to (Phillips and Moon, 1999, Lemma 3), i.e.,:

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

[2]The empirical EKC literature started about 30 years ago, 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 modeling 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, 2007; 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.

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:

$$\Omega_i := \begin{pmatrix} \Omega_{u_iu_i} & \Omega_{u_iv_i} \\ \Omega_{v_iu_i} & \Omega_{v_iv_i} \end{pmatrix}. \tag{4}$$

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

$$\Delta_i := \begin{pmatrix} \Delta_{u_iu_i} & \Delta_{u_iv_i} \\ \Delta_{v_iu_i} & \Delta_{v_iv_i} \end{pmatrix}, \tag{5}$$

with consequently $\Omega_i = \Delta_i + \Delta_i' - \Sigma_i$, where Σ_i is the random contemporaneous covariance matrix. More specifically, this leads to the assumption:

Assumption 1. *The random processes $\{\eta_{it}\}_{t\in\mathbb{Z}}$ are independent across $i = 1, ..., N$, the random matrices (Δ_i, Σ_i) are independent of the Wiener processes $W_i(r)$ for $i = 1, ..., N$ and Ω_i is positive definite almost surely for $i = 1, ..., N$.*

Given the primary focus on the slope parameter vector $\beta := (\beta_1, \beta_2, \beta_3)'$, it is convenient to use *uniform notation*, $\tilde{y}_{it}$ and $\tilde{X}_{it}$, for both demeaned and demeaned and linearly detrended variables. In the demeaning only case, we thus have:

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

with analogously defined quantities for x_{it} (and its powers), u_{it} and v_{it}. Stacking defines:

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

In case of demeaning and linear detrending, we have, using generic notation z_{it}, for y_{it}, x_{it} and its powers:

$$\tilde{z}_{it} := z_{it} - \frac{4T - 6t + 2}{T - 1}\overline{z_{i.}} - \frac{-6T + 12t - 6}{(T-1)(T+1)}\sum_{t=1}^{T}\left(\frac{t}{T}\right) z_{it}, \tag{8}$$

leading to a correspondingly demeaned and detrended stacked vector $\tilde{X}_{it}$, with $\tilde{u}_{it}$ and $\tilde{v}_{it}$ again defined analogously.[3]

The exact form of the results depends, in addition to the specification of the deterministic components in the regression equation, also on whether the regressors x_{it} include a (non-zero) drift or not. It is therefore convenient to structure the discussion according to the following cases: zero drifts $\mu_i = 0$, $i = 1, ..., N$, non-zero drifts $\mu_i \neq 0$, $i = 1, ..., N$ and the general case $\mu_i \in \mathbb{R}$, $i = 1, ..., N$.

2.1. Zero drifts

To complete the formulation of the assumptions required in this case, define $G_T := \text{diag}(T^{-1}, T^{-3/2}, T^{-2})$ and $A_i := (1, 2\int_0^1 B_{v_i}(r)dr, 3\int_0^1 B_{v_i}^2(r)dr)'$. To capture the effects of demeaning and demeaning and linear detrending – or of the "removal" of more general trend functions

[3]Clearly, more general (asymptotically) regular trend functions can be considered, e.g., higher-order polynomial time trends. A trend function $D(r), 0 \leq r \leq 1$ is called asymptotically regular, if $\int_0^1 D(r)D(r)'dr$ is positive definite.

– define for an (integrable stochastic process) $P(r)$ and an asymptotically regular trend function $D(r)$ for $0 \leq r \leq 1$:

$$\tilde{P}(r) := P(r) - D(r)\left(\int_0^1 D(s)D(s)'ds\right)^{-1}\int_0^1 D(s)P(s)ds, \tag{9}$$

which for the case of demeaning, of course, simplifies to $\tilde{P}(r) = P(r) - \int_0^1 P(s)ds$.[4] The notation allows to (generically) define $\tilde{\mathbf{B}}_{v_i}(r) := (\tilde{B}_{v_i}(r), \widetilde{B^2_{v_i}}(r), \widetilde{B^3_{v_i}}(r))'$, corresponding to the deterministic specification considered. Using this notation, we assume:

Assumption 2. *For $i = 1, ..., N$ and $0 \leq r \leq 1$ it holds that:*

(a) $T^{1/2}G_T\tilde{X}_{i\lfloor rT\rfloor} \Rightarrow \tilde{\mathbf{B}}_{v_i}(r)$,

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

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

with all quantities converging jointly.

As usual in FM-OLS type estimation, consistent non-parametric kernel estimators of long-run covariances and half long-run covariances – based on the OLS residuals $\hat{u}_{it}$ from (1) and $v_{it} = \Delta x_{it}$ – are required. This in turn requires appropriate kernel and bandwidth choices, compare, e.g., Jansson (2002).[5]

Assumption 3. *The cross-sectionally independent 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 $i = 1, ..., N$. By definition, this implies cross-sectional independence and consistency of $\hat{\Omega}_i := \hat{\Delta}_i + \hat{\Delta}_i' - \hat{\Sigma}_i$ for $i = 1, ..., N$.*

For brevity, we abstain from formulating primitive assumptions that generate our Assumptions 2 and 3 that are, in fact, convergence results. 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 (2022) who in turn build upon Phillips and Moon (1999). However, in a finite N setting, as considered in this article, one can replace the random linear process framework without any (substantial) loss with more classical assumptions as posited, e.g., in Wagner and Hong (2016) in a time series setting. As discussed below in Remark 4, the random linear process framework provides *fundamental* value added only in case $N \to \infty$, see also the discussion in de Jong and Wagner (2022).

We are now ready to define the group-mean FM-OLS estimator as the cross-sectional average of the individual-specific FM-OLS estimators (as developed in Wagner and Hong, 2016) of the coefficient vector β. More precisely, we define for $i = 1, ..., N$ the FM-OLS estimator of β from the ith cross-section member – computed from individual-specifically demeaned or individual-specifically demeaned and linearly detrended data – as:

$$\hat{\beta}^+(i) := \left(\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\sum_{t=1}^T \tilde{X}_{it}\tilde{y}_{it}^+ - C_i\right), \tag{10}$$

[4] As is well-known, in case of demeaning and linear detrending, $\tilde{P}(r) = P(r) - (4-6r)\int_0^1 P(s)ds - (-6+12r)\int_0^1 sP(s)ds$.

[5] To maintain cross-sectional independence of the individual-specific estimators, the long-run covariance matrix estimators need to be cross-sectionally independent as well. The asymptotic analysis considered in de Jong and Wagner (2022), with also $N \to \infty$ after $T \to \infty$, allows for more flexibility in this respect.

where $\tilde{y}_{it}^{+} := \tilde{y}_{it} - \Delta x_{it}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i}$ and $C_i := \hat{\Delta}_{v_i u_i}^{+}(T, 2\sum_{t=1}^{T} x_{it}, 3\sum_{t=1}^{T} x_{it}^{2})'$, with $\hat{\Delta}_{v_i u_i}^{+} := \hat{\Delta}_{v_i u_i} - \hat{\Delta}_{v_i v_i}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i}$.[6] The cross-sectional average of $\hat{\beta}^{+}(i)$ defines the *group-mean* FM-OLS estimator:

$$\hat{\beta}^{+} := \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{+}(i). \tag{11}$$

The following proposition derives its asymptotic distribution as the time series dimension $T \to \infty$, for fixed cross-section dimension N.

Proposition 1. *Let the data be generated by (1) and (2) with $\mu_i = 0$, $i = 1, ..., N$ and let Assumptions 1–3 be in place. Then it holds for $T \to \infty$, conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, ..., N$, that:*

$$G_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}(0, V^{+}), \tag{12}$$

where $\mathcal{N}(0, V^{+})$ denotes a normal distribution with expectation zero and conditional covariance matrix:

$$V^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i \cdot v_i}\left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1} = \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i \cdot v_i}\tilde{M}_{ii}^{-1}, \tag{13}$$

with $\Omega_{u_i \cdot v_i} := \Omega_{u_i u_i} - \Omega_{u_i v_i}\Omega_{v_i v_i}^{-1}\Omega_{v_i u_i} > 0$ equal to the conditional variance of $B_{u_i \cdot v_i}(r) := B_{u_i}(r) - \Omega_{u_i v_i}\Omega_{v_i v_i}^{-1}B_{v_i}(r)$ and $\tilde{M}_{ii}$ defined by the last equality.

Under our assumptions, the natural consistent estimator of V^{+} is:

$$\hat{V}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i \cdot v_i}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T\right)^{-1} = G_T^{-1}\hat{S}^{+}G_T^{-1}, \tag{14}$$

with $\hat{\Omega}_{u_i \cdot v_i} := \hat{\Omega}_{u_i u_i} - \hat{\Omega}_{u_i v_i}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i}$ and $\hat{S}^{+}$ 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 (14) 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 $\hat{\beta}^{+}$ converge at different rates, an issue discussed in detail in, e.g., Sims et al. (1990, Section 4) or Wagner and Hong (2016, Section 2.2, p. 1297). It suffices to assume that the constraint matrix fulfills the (asymptotic) restriction posited in the following corollary.

Corollary 1. *Let the data be generated by (1) and (2) with $\mu_i = 0$, $i = 1, ..., N$, and let Assumptions 1–3 be in place. Consider s linearly independent restrictions collected in:*

$$H_0 : R\beta = r, \tag{15}$$

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

$$W^{+} := \left(R\hat{\beta}^{+} - r\right)'\left(R\hat{S}^{+}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right) \tag{16}$$

is chi-squared distributed with s degrees of freedom as $T \to \infty$. In case s = 1, of course, a t-type test can also be considered:

[6]Note that performing FM-OLS calculations for a time series dimension ranging from $t = 1, ..., T$ *implicitly* assumes that observations are available for $t = 0, ..., T$ as the construction of $\tilde{y}_{it}^{+}$ implies that one loses the first observation.

$$t^+ := \frac{R\hat{\beta}^+ - r}{\sqrt{R\hat{S}^+ R'}}, \tag{17}$$

which is under the null hypothesis asymptotically standard normally distributed as $T \to \infty$.

Inference on α_i and δ_i is also possible. Similarly, as an observation for later when drifts are considered, inference on μ_i using, e.g., augmented Dickey-Fuller type regressions can also be performed.

Remark 1. The group-mean estimator is robust to many forms of cross-section dependence, i.e., it remains consistent with a zero mean Gaussian mixture limiting distribution despite cross-section dependence. Of course, the covariance matrix of the asymptotic distribution changes – depending upon the form and extent of cross-section dependence. Given that we consider a fixed N setting, it suffices to simply consider a multivariate version of our assumptions ensuring joint convergence of all quantities for $i = 1, \ldots, N$.

The key quantity required for robust inference is (a consistent estimator of) the asymptotic covariance matrix of the group-mean FM-OLS estimator in case of cross-section dependence. To this end, denote with $\tilde{M}_{ij} := \int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_j}(r)'dr$ and with $\mathbf{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}$ the "constant" in the quadratic covariation of the processes $B_{u_i\cdot v_i}(r)$ and $B_{u_j\cdot v_j}(r)$.[7] The asymptotic covariance matrix of the group-mean estimator given in (11) changes from the expression given in (13) to the "sandwich" form:

$$V_{\text{rob}}^+ := \frac{1}{N^2}\sum_{i,j=1}^{N} \mathbf{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}\tilde{M}_{ii}^{-1}\tilde{M}_{ij}\tilde{M}_{jj}^{-1}. \tag{18}$$

It is important to note that the above result allows for very general forms of cross-section dependencies, as long as V_{rob}^+ is invertible. As an (extreme) example, consider the case $x_{it} = x_t$ for $i = 1, \ldots, N,$ i.e., the integrated regressor is the same for all cross-section members, which is an extreme form of cross-unit cointegration, compare Wagner and Hlouskova (2009). In this case, $\tilde{M}_{ii} = \tilde{M}_{jj} = \tilde{M}_{ij} = \tilde{M}$ for $i,j = 1, \ldots, N$ and $V_{\text{rob}}^+ = \left(\frac{1}{N^2}\sum_{i,j=1}^{N} \mathbf{\Omega}_{u_i\cdot v;u_j\cdot v}\right)\tilde{M}^{-1}$, using simplified notation $\Delta x_t = v_t$ in Ω. The term in brackets simplifies in this case to $\frac{1}{N^2}1_N'\mathbf{\Omega}_{u\cdot v}1_N$, with $1_N := [1, \ldots, 1]' \in \mathbb{R}^N$ and $\mathbf{\Omega}_{u\cdot v} = \mathbf{\Omega}_{uu} - \mathbf{\Omega}_{uv}\Omega_{vv}^{-1}\Omega_{vu}$. Thus, positive definiteness of $\mathbf{\Omega}_{u\cdot v}$ is in this example sufficient for robust inference. This example highlights the wide applicability of robust inference based on the group-mean estimator, without having to posit a model for cross-section dependence, e.g., common stochastic trends or a factor structure.[8]

A consistent estimator of the asymptotic covariance matrix V_{rob}^+ is given by:

$$\begin{aligned}\hat{V}_{\text{rob}}^+ &:= \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\mathbf{\Omega}}_{u_i\cdot v_i;u_j\cdot v_j}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T\right)^{-1}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{jt}'G_T\right)\left(G_T\sum_{t=1}^{T}\tilde{X}_{jt}\tilde{X}_{jt}'G_T\right)^{-1} \\ &= G_T^{-1}\frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\mathbf{\Omega}}_{u_i\cdot v_i;u_j\cdot v_j}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{jt}'\right)\left(\sum_{t=1}^{T}\tilde{X}_{jt}\tilde{X}_{jt}'\right)^{-1}G_T^{-1} \\ &=: G_T^{-1}\hat{S}_{\text{rob}}^+G_T^{-1},\end{aligned} \tag{19}$$

[7]Given that we consider the quadratic covariation between Brownian motions, this constant is, of course, simply the covariance between $B_{u_i\cdot v_i}(1)$ and $B_{u_j\cdot v_j}(1)$.

[8]We abstain from positing an explicit set of assumptions for brevity as the discussion in the remark makes clear that any set of sufficient assumptions has to extend the *marginal* assumptions posited so far to hold jointly with cross-section dependence allowed for. Clearly, in the presence of cross-section dependence the estimators of the joint long-run covariance matrix will not feature cross-sectional independence by construction, compare Footnote 5.

with $\hat{S}^{+}_{\text{rob}}$ defined by the last equality. Since $\Omega_{u_i\cdot v_i;u_j\cdot v_j} = \Omega_{u_iu_j} - \Omega_{u_iv_i}\Omega^{-1}_{v_iv_i}\Omega_{v_iu_j} - \Omega_{u_jv_j}\Omega^{-1}_{v_jv_j}\Omega_{v_ju_i} + \Omega_{u_iv_i}\Omega^{-1}_{v_iv_i}\Omega_{v_iv_j}\Omega^{-1}_{v_jv_j}\Omega_{v_ju_j}$, we obtain the estimator $\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}$ by replacing the unknown long-run variances and covariances in the expression just given for $\Omega_{u_i\cdot v_i;u_j\cdot v_j}$ by consistent estimators. Robust Wald-type and t-type test statistics can now be defined similarly to the Wald-type and t-type test statistics defined in (16) and (17), with $\hat{S}^{+}_{\text{rob}}$ as defined in (19) in place of $\hat{S}^{+}$, i.e.:

$$W^{+}_{\text{rob}} := \left(R\hat{\beta}^{+} - r\right)'\left(R\hat{S}^{+}_{\text{rob}}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right), \tag{20}$$

$$t^{+}_{\text{rob}} := \frac{R\hat{\beta}^{+} - r}{\sqrt{R\hat{S}^{+}_{\text{rob}}R'}}, \tag{21}$$

which are under the null hypothesis asymptotically chi-squared distributed with s degrees of freedom and standard normally distributed, respectively, as $T \to \infty$.

Remark 2. In panel data settings, often time effects rather than individual-specific time trends are considered – most commonly in conjunction with individual effects – in a two-way effects specification. 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-sectional averages of time series limits. In the two-way case, with individual-specific intercepts and time effects, the transformed regressor vector, e.g., is given by $\check{X}_{it} := \tilde{X}_{it} - \frac{1}{N}\sum_{j=1}^{N}\tilde{X}_{jt}$, with $\tilde{X}_{it}$, $i = 1, ..., N$ as defined in (7). This leads to a partial sum limit (compare Assumption 2) 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 $\check{u}^{+}_{it} := \check{u}_{it} - \Delta x_{it}\hat{\Omega}^{-1}_{v_iv_i}\hat{\Omega}_{v_iu_i}$ and $\check{u}_{it} := \tilde{u}_{it} - \frac{1}{N}\sum_{j=1}^{N}\tilde{u}_{jt}$. Under appropriate assumptions $\frac{1}{N}\sum_{j=1}^{N}\tilde{\mathbf{B}}_{v_j}(r)$ fulfills a law of large numbers for $N \to \infty$. A corresponding result is the basis for the derivation of the large N and large T asymptotic distribution of the pooled estimator in de Jong and Wagner (2022) in the two-way effects case.

Remark 3. Considering time effects in a cross-sectionally homogenous case, with $\Delta_i = \Delta$ a.s. and $\Sigma_i = \Sigma$ a.s. for $i = 1, ..., N$, allows to alternatively adjust the group-mean estimator to achieve asymptotically valid inference by using $\check{y}^{+}_{it} := \check{y}_{it} - \Delta\check{x}_{it}\hat{\Omega}^{-1}_{vv}\hat{\Omega}_{vu}$, where $\check{y}_{it} := \tilde{y}_{it} - \frac{1}{N}\sum_{j=1}^{N}\tilde{y}_{jt}$, with $\tilde{y}_{it}$ as defined in (6) for $i = 1, ..., N$, as transformed dependent variable and:[9]

$$\check{C}_i := \hat{\Delta}^{+}_{vu}\left(\left(\frac{N-1}{N}\right)^2\left(T, 2\sum_{t=1}^{T}x_{it}, 3\sum_{t=1}^{T}x^2_{it}\right)' + \frac{1}{N^2}\sum_{j\neq i}\left(T, 2\sum_{t=1}^{T}x_{jt}, 3\sum_{t=1}^{T}x^2_{jt}\right)'\right), \tag{22}$$

as additive correction term when estimating the parameters of the ith equation with FM-OLS. This leads to the following homogeneous group-mean estimator:

$$\check{\beta}^{+}_{\text{HOM}} := \frac{1}{N}\sum_{i=1}^{N}\check{\beta}^{+}(i), \tag{23}$$

where:

[9]In this case, e.g., the homogenous long-run covariance matrix Ω can be estimated by the cross-sectional 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.

$$\check{\beta}^{+}(i) := \left(\sum_{t=1}^{T} \check{X}_{it}\check{X}_{it}'\right)^{-1}\left(\sum_{t=1}^{T} \check{X}_{it}\check{y}_{it}^{+} - \check{C}_i\right), \quad i = 1, ..., N. \tag{24}$$

The asymptotic (conditional) covariance matrix of the homogenous group-mean FM-OLS estimator is given by $V_{\text{HOM}}^{+} := \Omega_{u\cdot v}\frac{1}{N^2}\sum_{i=1}^{N}\tilde{M}_{ii}^{-1}$, compare (13). The homogenous versions of the Wald- and t-type statistics follow straightforwardly.

Remark 4. 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 in case of cross-sectional independence that:

$$\sqrt{N}G_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}(\Omega_{u_i\cdot v_i}\tilde{M}_{ii}^{-1})\right), \tag{25}$$

as $N \to \infty$ after $T \to \infty$. An estimator of the covariance matrix of this limiting distribution is given by $N\hat{V}^{+}$, with $\hat{V}^{+}$ the "finite N" covariance matrix estimator given below Proposition 1 in (14).

2.2. Non-zero drifts

Let us now consider the case with non-zero drifts, i.e., $\mu_i \neq 0$, $i = 1, ..., N$. In this case, the integrated regressor:

$$x_{it} = \mu_i + x_{i,t-1} + v_{it} = \mu_i t + \sum_{s=1}^{t} v_{is} + x_{i0} = \mu_i t + x_{it}^{o} + x_{i0}, \tag{26}$$

is asymptotically *dominated* by the deterministic linear trend $\mu_i t$ rather than the stochastic trend $x_{it}^{o} := \sum_{s=1}^{t} v_{is}$. For later usage define $\tilde{X}_{it}^{o}$ similarly to $\tilde{X}_{it}$ in (7), with x_{it}^{o} and its powers in place of x_{it} and its powers.

The implications of the dominance of a deterministic trend component on unit root and cointegration analysis have been already investigated in the linear time series case by West (1988), and, in the context of FM-OLS estimation, in Phillips and Hansen (1990, Remark (e), p. 105). For the second and third powers of x_{it}, the higher-order deterministic (monomial) quadratic or cubic time trends are the dominant elements. This, of course, leads to asymptotic normality results similar to those of West (1988) in a *linear cointegration setting*. However, in our context, the deterministic trend will not be dominant in $\tilde{x}_{it}$, when *both* demeaning and linear detrending take place. In this case, the deterministic component is exactly annihilated in the demeaned and detrended variable $\tilde{x}_{it}$. Consequently, in this case, the coefficient to the first power of the integrated regressor will have a unit root type asymptotic distribution rather than a normal asymptotic distribution.[10]

It is maybe worth mentioning that the presence of non-zero drifts $\mu_i \neq 0$ does not imply changes in the construction of the transformed dependent variable $\tilde{y}_{it}^{+}$. Commencing from $\Delta x_{it} = \mu_i + v_{it}$ immediately leads to:

$$\tilde{y}_{it}^{+} = \tilde{y}_{it} - \Delta x_{it}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i} = \tilde{y}_{it} - v_{it}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i} - \mu_i\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}. \tag{27}$$

This in turn implies that $\tilde{u}_{it}^{+} = \tilde{u}_{it} - v_{it}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i} - \mu_i\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}$. Consequently, the scaled partial sum process $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor}\tilde{u}_{it}^{+}$ diverges, being non-centered. Nevertheless, $\sum_{t=1}^{T}\tilde{X}_{it} = 0$ implies that –

[10]For a full analysis of the impacts of the presence of deterministic trends in the regression equation and/or the regressors for a *more general* CPR specification – in the time series case – see Reichold and Wagner (2022).

after appropriate scaling – the cross product term $\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^{+}$ converges (conditionally) to a Gaussian mixture limit (with integrator and integrand independent of each other) plus an additive bias term to be subtracted. This is the key result allowing for asymptotically valid inference based upon the group-mean FM-OLS estimator. Thus, the definition and computation of the group-mean FM-OLS estimator is unaffected by or invariant to the presence of non-zero drifts.

Remark 5. In relation to the above, a word of caution may be in order concerning long-run covariance estimation, typically based on the OLS residuals $\hat{u}_{it}$ of (1) in conjunction with the first difference Δx_{it} of x_{it}. If one uses, as is sometimes done, an estimator that does not center the variables prior to autocovariance estimation, the resultant estimator will diverge due to non-zero expectation μ_i of Δx_{it}. By construction, $\hat{u}_{it}$ does not have to be centered in any of our specifications as they all include at least an intercept as deterministic variable. If it is known that $\mu_i = 0$, also Δx_{it} need not be centered.

Extending Proposition 1 to the case of non-zero drifts requires the definition of a few additional quantities, including the scaling matrices $H_T := \text{diag}(T^{-3/2}, T^{-5/2}, T^{-7/2})$ and $K_T := \text{diag}(T^{-1}, T^{-5/2}, T^{-7/2})$. Furthermore, for $i = 1, ..., N$ define:

$$J_i(r) := \begin{pmatrix} \mu_i & & \\ & \mu_i^2 & \\ & & \mu_i^3 \end{pmatrix} \begin{pmatrix} r - 1/2 \\ r^2 - 1/3 \\ r^3 - 1/4 \end{pmatrix} =: \mathcal{D}(\mu_i) \begin{pmatrix} r - 1/2 \\ r^2 - 1/3 \\ r^3 - 1/4 \end{pmatrix}, \tag{28}$$

$$L_i(r) := \begin{pmatrix} 1 & & \\ & \mu_i^2 & \\ & & \mu_i^3 \end{pmatrix} \begin{pmatrix} \tilde{B}_{v_i}(r) \\ r^2 - r + 1/6 \\ r^3 - 9/10r + 1/5 \end{pmatrix} =: \mathcal{E}(\mu_i) \begin{pmatrix} \tilde{B}_{v_i}(r) \\ r^2 - r + 1/6 \\ r^3 - 9/10r + 1/5 \end{pmatrix}. \tag{29}$$

Proposition 2. *Let the data be generated by (1) and (2) with $\mu_i \neq 0$, $i = 1, ..., N$ and let Assumptions 1, 2 for $\tilde{X}_{it}^{o}$ and 3 be in place.*

(i) *In case individual-specific intercepts but no individual-specific linear trends are included in (1), it holds for $T \to \infty$, conditional upon Δ_i and Σ_i for $i = 1, ..., N$ that:*

$$H_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\alpha}^{+}\right), \tag{30}$$

with $V_{\alpha}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i \cdot v_i}(\int_0^1 J_i(r)J_i(r)'dr)^{-1}$ for $i = 1, ..., N$.

(ii) *In case individual-specific intercepts and linear trends are included in (1), it holds for $T \to \infty$, conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, ..., N$ that:*

$$K_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\alpha,\delta}^{+}\right), \tag{31}$$

with $V_{\alpha,\delta}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i \cdot v_i}(\int_0^1 L_i(r)L_i(r)'dr)^{-1}$ for $i = 1, ..., N$.

Proposition 2 shows that the two cases – with or without individual-specific trends – lead to different asymptotic distributions of the group-mean FM-OLS estimator. Case (i), without individual-specific trends, leads to a *West*-type asymptotic normality result for all elements of β, more clearly (unconditionally) visible in case Δ_i and Σ_i are considered nonrandom. It is convenient to rewrite V_{α}^{+} as:

$$V_{\alpha}^{+} = \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i \cdot v_i}\left(\mathcal{D}(\mu_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\mu_i)\right)^{-1}. \tag{32}$$

This leads immediately to *two* estimators of V_{α}^{+}, one similar to the estimator $\hat{V}^{+}$ given in (14) and the second commencing from the closed form expression for the limit result, i.e.:

$$\hat{V}_{\alpha}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i\cdot v_i}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'H_T\right)^{-1} = H_T^{-1}\hat{S}^{+}H_T^{-1}, \tag{33}$$

and:

$$\tilde{V}_{\alpha}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i\cdot v_i}\left(\mathcal{D}(\hat{\mu}_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\hat{\mu}_i)\right)^{-1}, \tag{34}$$

with, e.g., $\hat{\mu}_i := \frac{1}{T}\sum_{t=1}^{T}\Delta x_{it}$.

In Case (ii), with individual-specific intercepts and linear trends included, the coefficient to $\tilde{x}_{it}$ has, as mentioned above, a *unit-root*-type limiting distribution and only the coefficients to the higher-order powers have a *West*-type asymptotic normal distribution. This implies that a "direct" estimator of $V_{\alpha,\delta}^{+}$, similar in spirit to $\tilde{V}_{\alpha}^{+}$, can only be constructed for the lower 2×2 block, i.e.:

$$\hat{V}_{\alpha,\delta}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i\cdot v_i}\left(K_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'K_T\right)^{-1} = K_T^{-1}\hat{S}^{+}K_T^{-1}, \tag{35}$$

and:

$$\tilde{V}_{\alpha,\delta}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i\cdot v_i}\begin{pmatrix} \frac{1}{T^2}\sum_{t=1}^{T}(\tilde{x}_{it})^2 & \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x_{it}^2} & \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x_{it}^3} \\ \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x_{it}^2} & \hat{\mu}_i^4/180 & \hat{\mu}_i^5/120 \\ \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x_{it}^3} & \hat{\mu}_i^5/120 & 9\hat{\mu}_i^6/700 \end{pmatrix}^{-1}. \tag{36}$$

The above considerations lead to the following corollary summarizing the test options in case of non-zero drifts.

Corollary 2. *Let the data be generated by (1) and (2) with $\mu_i \neq 0$, $i = 1,\ldots,N$ and let Assumptions 1, 2 for $\tilde{X}_{it}^{o}$ and 3 be in place. Consider s linearly independent restrictions collected in $H_0 : R\beta = r$, with $R \in \mathbb{R}^{s\times 3}$, $r \in \mathbb{R}^{s}$ and assume that there exists a nonsingular matrix $G_R \in \mathbb{R}^{s\times s}$ and a matrix $R^* \in \mathbb{R}^{s\times 3}$ of rank s such that $\lim_{T\to\infty} G_R R H_T = R^*$ (in the individual-specific intercepts only case) or $\lim_{T\to\infty} G_R R K_T = R^*$ (in the individual-specific intercepts and linear trends case).*

In both, the individual-specific intercepts only and the individual-specific intercepts and linear trends cases, the Wald- and (in case $s = 1$) t-type statistics:

$$W^{+} = \left(R\hat{\beta}^{+} - r\right)'\left(R\hat{S}^{+}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right), \tag{37}$$

$$t^{+} = \frac{R\hat{\beta}^{+} - r}{\sqrt{R\hat{S}^{+}R'}}, \tag{38}$$

already defined in (16) and (17), are under the null hypothesis chi-squared distributed with s degrees of freedom and standard normally distributed, respectively, as $T \to \infty$.

Furthermore, in the individual-specific intercepts only case, the test statistics can alternatively (asymptotically equivalently) be defined as:

$$W_{\alpha}^{+} := \left(R\hat{\beta}^{+} - r\right)'\left(R\tilde{S}_{\alpha}^{+}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right), \tag{39}$$

$$t_{\alpha}^{+} := \frac{R\hat{\beta}^{+} - r}{\sqrt{R\tilde{S}_{\alpha}^{+}R'}}, \tag{40}$$

with $\tilde{S}_{\alpha}^{+} := H_T\tilde{V}_{\alpha}^{+}H_T$.

In the individual-specific intercepts and linear trends case, the test statistics can alternatively (asymptotically equivalently) be defined as:

$$W_{\alpha,\delta}^{+} := \left(R\hat{\beta}^{+} - r\right)'\left(R\tilde{S}_{\alpha,\delta}^{+}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right), \tag{41}$$

$$t_{\alpha,\delta}^{+} := \frac{R\hat{\beta}^{+} - r}{\sqrt{R\tilde{S}_{\alpha,\delta}^{+}R'}}, \tag{42}$$

with $\tilde{S}_{\alpha,\delta}^{+} := K_T\tilde{V}_{\alpha,\delta}^{+}K_T$. *Under the null hypothesis, the four additionally considered test statistics are asymptotically chi-squared or standard normally distributed, respectively, as* $T \to \infty$.

Remark 6. Similar to Remark 1 in Subsection 2.1, the results can be extended to allow for cross-section dependence; based again upon any suitable modification of the assumptions to ensure the necessary joint convergence results. The precise form of the asymptotic results will depend upon the deterministic components in (1). With individual-specific intercepts only, the covariance matrix of the asymptotic distribution is, in case of cross-section dependence, given by:

$$\begin{aligned} V_{\alpha,\mathrm{rob}}^{+} &:= \frac{1}{N^2}\sum_{i,j=1}^{N}\Omega_{u_i\cdot v_i;u_j\cdot v_j}\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)J_j(r)'dr\left(\int_0^1 J_j(r)J_j(r)'dr\right)^{-1} \\ &= \frac{1}{N^2}\sum_{i,j=1}^{N}\Omega_{u_i\cdot v_i;u_j\cdot v_j}\left(\mathcal{D}(\mu_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\mu_j)\right)^{-1}. \end{aligned} \tag{43}$$

In case that both individual-specific intercepts and linear trends are included in (1), the covariance matrix of the asymptotic distribution is given by:

$$\begin{aligned} V_{\alpha,\delta,\mathrm{rob}}^{+} &:= \frac{1}{N^2}\sum_{i,j=1}^{N}\Omega_{u_i\cdot v_i;u_j\cdot v_j}\left(\int_0^1 L_i(r)L_i(r)'dr\right)^{-1}\int_0^1 L_i(r)L_j(r)'dr\left(\int_0^1 L_j(r)L_j(r)'dr\right)^{-1} \\ &=: \frac{1}{N^2}\sum_{i,j=1}^{N}\Omega_{u_i\cdot v_i;u_j\cdot v_j}\mathcal{C}(i,j), \end{aligned} \tag{44}$$

with $\mathcal{C}(i,j)$ defined by the last equality. Considering again:

$$\hat{S}_{\mathrm{rob}}^{+} = \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{jt}'\right)\left(\sum_{t=1}^{T}\tilde{X}_{jt}\tilde{X}_{jt}'\right)^{-1}, \tag{45}$$

as defined already in (19), immediately leads to consistent estimators in both cases, given by $\hat{V}_{\alpha,\mathrm{rob}}^{+} := H_T^{-1}\hat{S}_{\mathrm{rob}}^{+}H_T^{-1}$ or $\hat{V}_{\alpha,\delta,\mathrm{rob}}^{+} := K_T^{-1}\hat{S}_{\mathrm{rob}}^{+}K_T^{-1}$, respectively. Entirely analogously to Remark 1, using $\hat{S}^{+}_{\text{rob}}$ in the definition of the robust test statistics W^{+}_{rob} and t^{+}_{rob} (in case $s=1$) given in (20) and (21), leads to chi-squared and standard normal inference, respectively, as $T\to\infty$.

Note for completeness that the test statistics W^{+}_{α} and t^{+}_{α} defined in (39) and (40) in Corollary 2 can also be "robustified" straightforwardly. Considering:

$$\tilde{V}^{+}_{\alpha,\text{rob}} := \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i v_i;u_j\cdot v_j}\left(\mathcal{D}(\hat{\mu}_i)\begin{pmatrix}1/12 & 1/12 & 3/40\\ 1/12 & 4/45 & 1/12\\ 3/40 & 1/12 & 9/112\end{pmatrix}\mathcal{D}(\hat{\mu}_j)\right)^{-1} \tag{46}$$

and $\tilde{S}^{+}_{\alpha,\text{rob}} := H_T\tilde{V}^{+}_{\alpha,\text{rob}}H_T$ allows to define:

$$W^{+}_{\alpha,\text{rob}} := \left(R\hat{\beta}^{+}-r\right)'\left(R\tilde{S}^{+}_{\alpha,\text{rob}}R'\right)^{-1}\left(R\hat{\beta}^{+}-r\right), \tag{47}$$

$$t^{+}_{\alpha,\text{rob}} := \frac{R\hat{\beta}^{+}-r}{\sqrt{R\tilde{S}^{+}_{\alpha,\text{rob}}R'}}. \tag{48}$$

Analogously, $W^{+}_{\alpha,\delta}$ and $t^{+}_{\alpha,\delta}$ defined in (41) and (42) can be "robustified" by constructing a "direct" estimator of $V^{+}_{\alpha,\delta,\text{rob}}$. To be precise, $\tilde{V}^{+}_{\alpha,\delta}$ defined in (36) has to be replaced by:

$$\tilde{V}^{+}_{\alpha,\delta,\ \text{rob}} := \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}\tilde{\mathcal{C}}(i,j) := \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}\tilde{\mathcal{A}}(i)^{-1}\tilde{\mathcal{B}}(i,j)\tilde{\mathcal{A}}(j)^{-1}, \tag{49}$$

$$\tilde{\mathcal{A}}(i) := \begin{pmatrix}\frac{1}{T^2}\sum_{t=1}^{T}(\tilde{x}_{it})^2 & \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{it}} & \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{it}}\\ \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{it}} & \hat{\mu}_i^4/180 & \hat{\mu}_i^5/120\\ \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{it}} & \hat{\mu}_i^5/120 & 9\hat{\mu}_i^6/700\end{pmatrix}, \qquad i=1,\dots,N, \tag{50}$$

$$\tilde{\mathcal{B}}(i,j) := \begin{pmatrix}\frac{1}{T^2}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{jt} & \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{jt}} & \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{jt}}\\ \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{jt}} & \hat{\mu}_i^2\hat{\mu}_j^2/180 & \hat{\mu}_i^2\hat{\mu}_j^3/120\\ \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{jt}} & \hat{\mu}_i^3\hat{\mu}_j^2/120 & 9\hat{\mu}_i^3\hat{\mu}_j^3/700\end{pmatrix}, \qquad i,j=1,\dots,N. \tag{51}$$

Based upon this, defining $\tilde{S}^{+}_{\alpha,\delta,\text{rob}} := K_T\tilde{V}^{+}_{\alpha,\text{rob}}K_T$ leads to the robust versions of the "direct" test statistics, i.e.:

$$W^{+}_{\alpha,\delta,\text{rob}} := \left(R\hat{\beta}^{+}-r\right)'\left(R\tilde{S}^{+}_{\alpha,\delta,\text{rob}}R'\right)^{-1}\left(R\hat{\beta}^{+}-r\right), \tag{52}$$

$$t^{+}_{\alpha,\delta,\text{rob}} := \frac{R\hat{\beta}^{+}-r}{\sqrt{R\tilde{S}^{+}_{\alpha,\delta,\text{rob}}R'}}. \tag{53}$$

Under the null hypothesis, the test statistics are asymptotically chi-squared or standard normally distributed, respectively, as $T \to \infty$.

Remark 7. In case of individual-specific intercepts in (1) only, the OLS estimator also allows for asymptotically valid standard inference, as noted by West (1988) in the context of linear time series cointegrating regressions. Proper scaling by a consistent estimator of the long-run variance of the errors u_{it} suffices. Therefore, in this case, one can consider a group-mean OLS estimator:

$$\hat{\beta} := \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}(i), \tag{54}$$

with:

$$\hat{\beta}(i) := \left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{y}_{it}, \quad i = 1, ..., N. \tag{55}$$

Under the assumptions of Proposition 2, it holds for $T \to \infty$, conditional upon Δ_i and Σ_i for $i = 1, ..., N$, that:

$$H_T^{-1}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_iu_i}\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\right). \tag{56}$$

Therefore, exactly as discussed in Corollary 2, group-mean OLS-based Wald- and t-type test statistics can be defined using two different estimators of the covariance matrix, analogous to using either $\hat{S}^+$ or $\tilde{S}_\alpha^+$, where in both matrices $\hat{\Omega}_{u_i \cdot v_i}$ is replaced by $\hat{\Omega}_{u_iu_i}$ for $i = 1, ..., N$. More precisely, constructing:

$$\hat{S} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_iu_i}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}, \tag{57}$$

$$\tilde{S}_\alpha := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_iu_i}H_T\left(\mathcal{D}(\hat{\mu}_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\hat{\mu}_i)\right)^{-1}H_T, \tag{58}$$

with, as before, $H_T = \text{diag}(T^{-3/2}, T^{-5/2}, T^{-7/2})$, $\mathcal{D}(\hat{\mu}_i) = \text{diag}(\hat{\mu}_i, \hat{\mu}_i^2, \hat{\mu}_i^3)$ and $\hat{\Omega}_{u_iu_i}$ an estimator of the long-run variance of u_{it}, allows to define corresponding Wald- and (in case $s = 1$) t-type statistics:

$$W := (R\hat{\beta} - r)'(R\hat{S}R')^{-1}(R\hat{\beta} - r), \tag{59}$$

$$t := \frac{R\hat{\beta} - r}{\sqrt{R\hat{S}R'}} \tag{60}$$

and

$$W_\alpha := (R\hat{\beta} - r)'(R\tilde{S}_\alpha R')^{-1}(R\hat{\beta} - r), \tag{61}$$

$$t_\alpha := \frac{R\hat{\beta} - r}{\sqrt{R\tilde{S}_\alpha R'}}. \tag{62}$$

Furthermore, similar to Remarks 1 and 6, cross-section dependence can be accommodated, i.e., the group-mean OLS estimator can also be used to perform robust inference, again in two ways. One variant is given by:

$$W_{\text{rob}} := \left(R\hat{\beta} - r\right)'\left(R\hat{S}_{\text{rob}}R'\right)^{-1}\left(R\hat{\beta} - r\right), \tag{63}$$

$$t_{\text{rob}} := \frac{R\hat{\beta} - r}{\sqrt{R\hat{S}_{\text{rob}}R'}}, \tag{64}$$

with $\hat{S}_{\text{rob}}$ similar to $\hat{S}^+_{\text{rob}}$ as defined in (19), but with $\hat{\Omega}_{u_iu_j}$ in place of $\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}$. The second possibility resembles the result discussed in Remark 6. The corresponding test statistics are given by:

$$W_{\alpha,\text{rob}} := \left(R\hat{\beta} - r\right)'\left(R\tilde{S}_{\alpha,\text{rob}}R'\right)^{-1}\left(R\hat{\beta} - r\right), \tag{65}$$

$$t_{\alpha,\text{rob}} := \frac{R\hat{\beta} - r}{\sqrt{R\tilde{S}_{\alpha,\text{rob}}R'}}, \tag{66}$$

with $\tilde{S}_{\alpha,\text{rob}}$ similar to $\tilde{S}^+_{\alpha,\text{rob}}$, but with $\hat{\Omega}_{u_iu_j}$ in place of $\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}$ in $\tilde{V}^+_{\alpha,\text{rob}}$ as defined in (46). Under the null hypothesis, all considered test statistics are asymptotically chi-squared or standard normally distributed, respectively, as $T \to \infty$.

2.3. Zero or non-zero drifts

We are now ready to discuss the "general" case concerning drifts, with drifts present or absent in any cross-section member. It is important to stress again that for using the developed estimators and tests based upon them no knowledge concerning the presence or absence of drifts is required. As in the previous subsection, it is convenient to first discuss the case with individual-specific intercepts only on the one hand and the case with individual-specific intercepts and linear trends on the other hand separately.

In the individual-specific intercepts only case, it follows from a combination of the results of Propositions 1 and 2 that the asymptotic behavior of the group-mean estimator only depends on the individual-specific estimators $\hat{\beta}^+(i)$ calculated from cross-section members with zero drifts, since these converge at a slower rate than the estimators corresponding to cross-section members with non-zero drifts in the integrated regressor. It is clear that this "sorts out itself" in the limiting distributions and there are no implications for either the definition or the usage of the considered test statistics.

In case of individual-specific intercepts and linear trends, Proposition 2 shows that the coefficient to the first power of the integrated regressor, β_1, is estimated with (the standard unit root) rate T, irrespective of whether a non-zero drift is present or not. Therefore, the limiting distribution of the first component of $\hat{\beta}^+$ will depend upon all cross-section member-specific estimates of β_1. For β_2 and β_3, the situation is exactly as in the individual-specific intercepts only case, with the limiting distribution only depending upon the individual-specific estimators corresponding to cross-section members with zero drifts in the integrated regressor.

For notational convenience only, consider the cross-section members ordered in $i = 1, \ldots, N_0$ cross-section members with zero drifts and $i = N_0 + 1, \ldots, N$ cross-section members with non-zero drifts; noting that N_0 can range from zero (non-zero drifts in all cross-section members) to N (all cross-section members with zero drifts in x_{it}). Furthermore, define the following scaling matrices:

$$Q_T := \begin{cases} G_T & \text{if } N_0 > 0 \\ H_T & \text{if } N_0 = 0 \end{cases} \quad \text{and} \quad R_T := \begin{cases} G_T & \text{if } N_0 > 0, \\ K_T & \text{if } N_0 = 0. \end{cases} \tag{67}$$

Proposition 3. *Let the data be generated by (1) and (2) with $\mu_i \in \mathbb{R}$, $i = 1, \ldots, N$ and let Assumptions 1, 2 for $\tilde{X}^o_{it}$ and 3 be in place.*

(i) *In case individual-specific intercepts but no individual-specific linear trends are included in (1), it holds for $T \to \infty$, conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, \dots, N$ that:*

$$Q_T^{-1}\left(\hat{\beta}^+ - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{N_0}^+\right), \tag{68}$$

with:

$$V_{N_0}^+ := \begin{cases} \dfrac{1}{N^2}\displaystyle\sum_{i=1}^{N_0} \boldsymbol{\Omega}_{u_i \cdot v_i} \left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1} & \text{if } N_0 > 0, \\ V_\alpha^+ & \text{if } N_0 = 0. \end{cases} \tag{69}$$

(ii) *In case individual-specific intercepts and linear trends are included in (1), it holds for $T \to \infty$, conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, \dots, N$ that:*

$$R_T^{-1}\left(\hat{\beta}^+ - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{N_0}^+\right), \tag{70}$$

with:

$$V_{N_0}^+ := \begin{cases} \dfrac{1}{N^2}\displaystyle\sum_{i=1}^{N_0} \boldsymbol{\Omega}_{u_i \cdot v_i} \left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1} & \\ + \dfrac{1}{N^2}\displaystyle\sum_{i=N_0+1}^{N} \boldsymbol{\Omega}_{u_i \cdot v_i} \begin{pmatrix} \left(\int_0^1 L_i(r)L_i(r)'dr\right)^{-1}_{[1,1]} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \text{if } N_0 > 0, \\ V_{\alpha,\delta}^+ & \text{if } N_0 = 0, \end{cases} \tag{71}$$

with $_{[1,1]}$ indicating the (1, 1) element of the (3×3 inverted) matrix.

The second term in the covariance matrix $V_{N_0}^+$ in item (ii) in case $N_0 > 0$ reflects the above-mentioned fact that the coefficient to the first power of the integrated regressor is estimated at rate T irrespective of whether the drift is zero or non-zero – as in either case linear detrending removes a potential deterministic linear trend from the corresponding regressor. The asymptotic distribution immediately leads to Wald- and t-type test statistics.

Corollary 3. *Let the data be generated by (1) and (2) with $\mu_i \in \mathbb{R}$, $i = 1, \dots, N$ and let Assumptions 1, 2 for $\tilde{X}_{it}^o$ and 3 be in place. Consider s linearly independent restrictions collected in $H_0 : R\beta = r$ with $R \in \mathbb{R}^{s \times 3}$, $r \in \mathbb{R}^s$ and assume that there exists a nonsingular matrix $G_R \in \mathbb{R}^{s \times s}$ and a matrix $R^* \in \mathbb{R}^{s \times 3}$ of rank s such that $\lim_{T\to\infty} G_R R Q_T = R^*$ (in the individual-specific intercepts only case) or $\lim_{T\to\infty} G_R R R_T = R^*$ (in the individual-specific intercepts and linear trends case). In both, the individual-specific intercepts only and the individual-specific intercepts and linear trends case, the Wald- and (in case $s = 1$) t-type statistics:*

$$W = \left(R\hat{\beta}^+ - r\right)'\left(R\hat{S}^+R'\right)^{-1}\left(R\hat{\beta}^+ - r\right), \tag{72}$$

$$t = \frac{R\hat{\beta}^+ - r}{\sqrt{R\hat{S}^+R'}}, \tag{73}$$

already defined in (16) and (17), are under the null hypothesis chi-squared distributed with s degrees of freedom and standard normally distributed, respectively, as $T \to \infty$.

Remark 8. As in the previous subsections, cf. Remarks 1 and 6, the group-mean FM-OLS estimator remains consistent with a zero mean (conditional) normal limiting distribution in case of cross-section dependencies; with the assumptions adjusted correspondingly. The key input for performing robust inference is again a consistent estimator of the covariance matrix of the asymptotic distribution.

In case individual-specific intercepts only are included (1), the asymptotic covariance matrix is, in case of cross-section dependence, given by:

$$V^{+}_{N_0,\text{rob}} := \begin{cases} \frac{1}{N^2}\sum_{i,j=1}^{N_0} \Omega_{u_i\cdot v_i;u_j\cdot v_j}\tilde{M}_{ii}^{-1}\tilde{M}_{ij}\tilde{M}_{jj}^{-1} & \text{if } N_0 > 0,\\ V^{+}_{\alpha,\text{rob}} & \text{if } N_0 = 0. \end{cases} \tag{74}$$

In case both individual-specific intercepts and linear trends are included in (1), the asymptotic covariance matrix is given by:

$$V^{+}_{N_0,\text{rob}} := \begin{cases} \frac{1}{N^2}\sum_{i,j=1}^{N_0} \Omega_{u_i\cdot v_i;u_j\cdot v_j}\tilde{M}_{ii}^{-1}\tilde{M}_{ij}\tilde{M}_{jj}^{-1} & \\ +\frac{1}{N^2}\sum_{i,j=N_0+1}^{N} \Omega_{u_i\cdot v_i;u_j\cdot v_j}\begin{pmatrix} \mathcal{C}(i,j)_{[1,1]} & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix} & \\ +\frac{1}{N^2}\sum_{i=1}^{N_0}\sum_{j=N_0+1}^{N} \Omega_{u_i\cdot v_i;u_j\cdot v_j}\begin{pmatrix} \mathcal{F}(i,j)_{[1,1]} & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix} & \text{if } N_0 > 0,\\ +\frac{1}{N^2}\sum_{i=N_0+1}^{N}\sum_{j=1}^{N_0} \Omega_{u_i\cdot v_i;u_j\cdot v_j}\begin{pmatrix} \mathcal{K}(i,j)_{[1,1]} & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix} & \\ V^{+}_{\alpha,\delta,\text{rob}} & \text{if } N_0 = 0, \end{cases} \tag{75}$$

with:

$$\mathcal{F}(i,j) := \left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)L_j(r)'dr\left(\int_0^1 L_j(r)L_j(r)'dr\right)^{-1}, \tag{76}$$

$$\mathcal{K}(i,j) := \left(\int_0^1 L_i(r)L_i(r)'dr\right)^{-1}\int_0^1 L_i(r)\tilde{\mathbf{B}}_{v_j}(r)'dr\left(\int_0^1 \tilde{\mathbf{B}}_{v_j}(r)\tilde{\mathbf{B}}_{v_j}(r)'dr\right), \tag{77}$$

for $i,j = 1,\ldots,N$.

For performing robust inference, however, the fact that the asymptotic covariance matrices are case-dependent with respect to both N_0 and whether or not individual-specific linear trends are included in (1), has no consequences. The robust test statistics W^{+}_{rob} and t^{+}_{rob} defined in (20) and (21), using $\hat{S}^{+}_{\text{rob}}$ defined in (19), lead to chi-squared and standard normal inference, respectively, under the null hypothesis as $T\to\infty$. This follows using similar arguments as in Proposition 3 and Corollary 3.

We abstain from a detailed discussion of constructing test statistics based on "direct" estimators of the covariance matrix. Doing so would, in practice, necessitate knowledge concerning the presence or absence of non-zero drifts in the integrated regressors in the individual cross-section members. Whilst this knowledge, as unlikely as this may be, could in some applications indeed be available and one could construct individual-specific "direct" estimators, we do not provide – notationally more cumbersome rather than mathematically more complicated – details here. For the same reason, we also abstain from considering OLS rather than FM-OLS estimation in the cross-section members with non-zero drifts and do not define a *mixed* OLS-FM-OLS group-mean estimator. The corresponding analysis is, again, notationally more cumbersome rather than mathematically more complex.

3. Finite sample performance

We generate, commencing from de Jong and Wagner (2022), data according to (1) and (2), i.e.:

$$y_{it} = \alpha_i + \delta_i t + x_{it}\beta_1 + x_{it}^2\beta_2 + x_{it}^3\beta_3 + u_{it}, \tag{78}$$

$$x_{it} = \mu_i + x_{i,t-1} + v_{it}, \quad x_{i0} = 0, \tag{79}$$

with slope parameters $\beta_1 = 5$, $\beta_2 = -3$ and $\beta_3 = 0.3$. The regression errors u_{it} and v_{it} are generated as:

$$u_{it} = \rho_{1i}u_{i,t-1} + \varepsilon_{it} + \rho_{2i}\nu_{it}, \tag{80}$$

$$v_{it} = 0.1(\nu_{it} + 0.5\nu_{i,t-1}), \tag{81}$$

with $(\varepsilon_{1t}, ..., \varepsilon_{Nt})' \sim \mathcal{N}(0, \Sigma)$ and $(\nu_{1t}, ..., \nu_{Nt})' \sim \mathcal{N}(0, \Sigma)$, i.i.d. across $t = 0, 1, ..., T$, where:

$$\Sigma = \begin{pmatrix} 1 & \rho_3 & \cdots & \rho_3 \\ \rho_3 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_3 \\ \rho_3 & \cdots & \rho_3 & 1 \end{pmatrix}. \tag{82}$$

The parameters ρ_{1i} and ρ_{2i} control the level of serial correlation in the error terms u_{it} and the extent of regressor endogeneity, respectively, whereas the parameter ρ_3 controls the extent of cross-section dependence. The parameters ρ_{i1}, ρ_{i2} are cross-sectionally i.i.d. and independent of $(\varepsilon_{it}, \nu_{it})'$, $t = 1, ..., T$. 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 $\rho_1, \rho_2 \in \{0, 0.3, 0.6, 0.9\}$.[11] Furthermore, we also consider $\rho_3 \in \{0, 0.3, 0.6, 0.9\}$. The individual effects α_i are i.i.d. $\mathcal{N}(-45, 5)$ and independent of all other random quantities. For the individual-specific time trends we consider two cases: (i) $\delta_i = 0$ for $i = 1, ..., N$ and (ii) δ_i i.i.d. $\mathcal{N}(-0.01, 0.01)$, independent of all other random quantities. In the former case, the variables are demeaned and in the second case, the variables are demeaned and linearly detrended for the construction of the estimators, compare (10) and (11).

With respect to drifts, μ_i, we consider three cases: Two boundary cases, one with all drift parameters equal to zero, i.e., $\mu_i = \mu = 0$, and one with all drift parameters equal to $\mu_i = \mu = 0.02$.[12] Furthermore, we consider an "intermediate case," with half of the individual-specific drifts equal to zero and the other half equal to 0.02. The simulation setting covers all combinations of $N \in \{10, 20, 100\}$ and $T \in \{100, 250, 500\}$. For every setting considered, the

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

Our way of introducing cross-section dependence is inspired by Wagner and Hlouskova (2009) who consider three specifications for modeling cross-section dependence. We consider their constant correlation setting.

[12]Setting all non-zero drift parameters equal to 0.02 is for simplicity only. The results are very similar when the non-zero drifts are independently drawn from the interval $[0.01, 0.03]$. The point value for $\mu = 0.02$ and the interval $[0.01, 0.03]$ are inspired by the arithmetic means of the annual GDP per capita growth rates for 19 countries in the long data set analyzed in Section 4. The country-specific arithmetic means range from 0.013 to 0.024, and the arithmetic mean over all countries of the country-specific mean growth rates is equal to 0.018.

number of replications is 5,000 and all test decisions are performed at the 5% nominal level. The reported results rely upon long-run covariance estimation using the Bartlett kernel in conjunction with the data-dependent bandwidth rule of Andrews (1991). As indicated in the introduction, the Supplementary Material contains a number of additional tables and figures.

We start by considering bias and root mean squared error (RMSE) of three estimators: The group-mean OLS estimator, labeled $\hat{\beta}$, the group-mean FM-OLS estimator $\hat{\beta}^+$ and the pooled FM-OLS estimator of de Jong and Wagner (2022), labeled $\hat{\beta}_P^+$.[13] In general, see as an illustration the results for β_1, with $\mu_i \neq 0$ for $i = 1, \ldots, N$, in Tables 1 and 2, the presence of individual-specific trends adversely affects estimator performance, both in terms of bias and RMSE. This almost necessarily implies, as will be seen below, a corresponding detrimental impact also on test performance.[14] As expected, increasing the sample size, either the cross-section dimension N or (with a stronger positive effect) the time series dimension T, leads to improved performance. As also expected, increasing any of the ρ-parameters that govern error serial correlation, regressor endogeneity, or cross-section dependence, respectively, leads to performance deterioration. In this respect, it turns out that RMSE is more strongly affected by cross-section dependence than bias, which does not react strongly to cross-section dependence. By construction, as the pooled FM-OLS estimator estimates only one set of slope coefficients, the pooled FM-OLS estimator mostly outperforms the group-mean FM-OLS estimator both in terms of bias and RMSE. Only for β_1 in the individual-specific intercepts only case, see Tables 1 and 2, the group-mean FM-OLS estimator leads in several cases to smaller bias than the pooled FM-OLS estimator (more pronounced for large ρ-values and smaller sample sizes), albeit in conjunction with higher RMSE. However, this is not the case for β_2 and β_3, see Tables 9–16 in the Supplementary Material, and should thus not be over-interpreted. Increasing values of ρ_1, ρ_2 lead to performance advantages of group-mean FM-OLS over group-mean OLS with – as expected – basically no differences between these two estimators for $\rho_1, \rho_2 = 0$.

The (asymptotic) implications of the absence or presence of drifts manifest themselves also in the finite sample results. In the individual-specific intercepts only case, bias and RMSE of all components of the OLS and FM-OLS group-mean estimators of β are smaller in the presence than in the absence of drifts; compare, e.g., for β_1 Table 1 with Table 7 in the Supplementary Material. Exactly in line with asymptotic theory (Proposition 3), bias and RMSE of the OLS and FM-OLS group-mean estimators of β_1 are not affected by the absence or presence of drifts in the individual-specific intercepts and linear trends case, compare Table 2 with Table 8 in the Supplementary Material.

To assess test performance, we consider in total five different test statistics evaluated under the null hypothesis by means of empirical null rejection probabilities and under a sequence of 20 alternatives by means of "size-corrected" power. We consider two test statistics based on the group-mean OLS estimator: The first is a *textbook* version of a group-mean OLS estimator-based test, labeled W_{TB}, using in $\hat{S}$, as defined in (57), instead of $\hat{\Omega}_{u_i u_i}$ a textbook variance estimator given by $\hat{\sigma}^2_{u_i} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}^2_{it}$. This test serves as a "textbook" OLS test benchmark and leads to asymptotically valid inference only when $\rho_{i1} = \rho_{i2} = \rho_3 = 0$ for $i = 1, \ldots, N$. The second group-mean OLS-based test statistic is W_{rob} as defined in (63). As discussed in Remark 7, asymptotic validity of this test for all values of the ρ-parameters hinges critically upon drifts being present in all cross-section members, which in practice is almost certainly unknown. We, of course, consider both standard and robust inference based on the group-mean OLS estimator, i.e., W^+ as defined in (16) and W^+_{rob} as defined in (20). Finally, for comparison, we also include the Wald-type test

[13] de Jong and Wagner (2022) do not consider the case of individual-specific linear trends but consider time effects, compare Remark 2. It is straightforward to adjust – and implement – the pooled FM-OLS estimator to include individual-specific (linear) time trends, using demeaned and linearly detrended observations.

[14] Tables 7 and 8 in the Supplementary Material provide the corresponding results for $\mu_i = 0$ for $i = 1, \ldots, N$.

Table 1. Bias and RMSE of the estimators of β_1 in the individual-specific intercepts only case with non-zero drifts.

		$N=10$			$N=20$			$N=100$		
T	ρ_1, ρ_2	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$
					Bias, $\rho_3=0$					
100	0	−0.01	−0.01	−0.00	−0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.23	0.04	0.01	0.24	0.05	0.01	0.24	0.05	0.01
	0.6	0.88	0.21	0.10	0.90	0.22	0.09	0.90	0.22	0.08
	0.9	3.23	0.54	0.94	3.24	0.52	0.95	3.23	0.57	0.89
250	0	0.00	0.00	0.00	0.00	0.00	−0.00	−0.00	−0.00	−0.00
	0.3	0.08	0.01	0.00	0.08	0.01	0.00	0.08	0.01	0.00
	0.6	0.36	0.06	0.03	0.36	0.06	0.03	0.36	0.06	0.02
	0.9	1.79	0.24	0.33	1.78	0.24	0.31	1.79	0.25	0.30
500	0	0.00	0.00	−0.00	−0.00	−0.00	−0.00	0.00	0.00	−0.00
	0.3	0.03	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00
	0.6	0.14	0.02	0.01	0.14	0.02	0.01	0.14	0.02	0.01
	0.9	0.86	0.13	0.11	0.86	0.12	0.11	0.87	0.14	0.10
					Bias, $\rho_3=0.9$					
100	0	−0.02	−0.02	−0.01	−0.03	−0.02	−0.02	0.00	0.00	0.01
	0.3	0.21	0.02	0.00	0.20	0.01	−0.01	0.24	0.05	0.03
	0.6	0.86	0.18	0.11	0.86	0.17	0.10	0.90	0.23	0.15
	0.9	3.23	0.39	0.68	3.22	0.35	0.72	3.28	0.47	0.90
250	0	−0.00	−0.00	−0.00	−0.01	−0.01	−0.00	−0.00	−0.00	0.00
	0.3	0.08	0.01	0.00	0.08	−0.00	0.00	0.08	0.01	0.01
	0.6	0.36	0.06	0.03	0.35	0.05	0.03	0.36	0.07	0.04
	0.9	1.78	0.22	0.23	1.77	0.17	0.22	1.81	0.22	0.27
500	0	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	0.00	0.00	−0.00
	0.3	0.03	0.00	0.00	0.03	−0.00	−0.00	0.03	0.00	−0.00
	0.6	0.14	0.02	0.01	0.13	0.01	0.00	0.14	0.02	0.01
	0.9	0.84	0.10	0.09	0.85	0.10	0.07	0.84	0.12	0.09
					RMSE, $\rho_3=0$					
100	0	0.47	0.48	0.11	0.33	0.33	0.07	0.15	0.15	0.03
	0.3	0.66	0.63	0.16	0.50	0.44	0.10	0.31	0.20	0.04
	0.6	1.32	0.97	0.29	1.14	0.71	0.19	0.95	0.37	0.10
	0.9	4.03	2.69	1.37	3.66	1.95	1.14	3.32	1.04	0.93
250	0	0.14	0.14	0.05	0.10	0.10	0.03	0.04	0.04	0.01
	0.3	0.21	0.19	0.07	0.16	0.14	0.05	0.10	0.06	0.02
	0.6	0.50	0.33	0.13	0.44	0.24	0.09	0.38	0.12	0.04
	0.9	2.15	1.16	0.65	1.97	0.85	0.49	1.83	0.46	0.34
500	0	0.05	0.05	0.03	0.04	0.04	0.02	0.02	0.02	0.01
	0.3	0.08	0.08	0.04	0.06	0.05	0.02	0.04	0.02	0.01
	0.6	0.20	0.13	0.07	0.17	0.09	0.04	0.15	0.05	0.02
	0.9	1.06	0.54	0.32	0.95	0.39	0.23	0.89	0.22	0.13
					RMSE, $\rho_3=0.9$					
100	0	1.10	1.11	0.63	1.05	1.06	0.57	1.03	1.04	0.53
	0.3	1.50	1.46	0.84	1.41	1.40	0.77	1.38	1.36	0.71
	0.6	2.54	2.25	1.34	2.40	2.13	1.23	2.32	2.05	1.14
	0.9	6.46	6.37	4.13	6.14	5.82	3.73	5.95	5.58	3.42
250	0	0.35	0.35	0.24	0.35	0.35	0.22	0.35	0.35	0.22
	0.3	0.50	0.49	0.34	0.50	0.49	0.31	0.50	0.48	0.30
	0.6	0.95	0.82	0.57	0.96	0.82	0.53	0.95	0.81	0.51
	0.9	3.48	2.87	2.04	3.50	2.85	1.93	3.44	2.79	1.84
500	0	0.14	0.14	0.11	0.14	0.14	0.10	0.14	0.14	0.10
	0.3	0.21	0.20	0.15	0.21	0.20	0.15	0.21	0.20	0.15
	0.6	0.42	0.35	0.27	0.41	0.35	0.25	0.43	0.35	0.25
	0.9	1.84	1.41	1.05	1.80	1.38	1.01	1.85	1.40	1.00

Note: The column labels $\hat{\beta}_1$, $\hat{\beta}_1^+$, and $\hat{\beta}_{P,1}^+$ denote the group-mean OLS estimator, the group-mean FM-OLS estimator, and the pooled FM-OLS estimator, respectively, of β_1.

Table 2. Bias and RMSE of the estimators of β_1 in the individual-specific intercepts and linear trends case with non-zero drifts.

		$N = 10$			$N = 20$			$N = 100$		
T	ρ_1, ρ_2	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$
					Bias, $\rho_3 = 0$					
100	0	−0.01	−0.01	−0.00	−0.00	0.00	−0.00	0.00	0.00	0.00
	0.3	0.31	0.10	0.03	0.32	0.11	0.03	0.32	0.11	0.02
	0.6	1.23	0.67	0.30	1.25	0.69	0.27	1.24	0.69	0.25
	0.9	4.69	3.96	2.91	4.73	3.97	2.82	4.73	3.99	2.74
250	0	0.00	0.00	0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00
	0.3	0.13	0.03	0.01	0.13	0.03	0.01	0.13	0.03	0.01
	0.6	0.58	0.24	0.10	0.57	0.23	0.09	0.57	0.23	0.08
	0.9	3.06	2.22	1.34	3.02	2.19	1.25	3.04	2.21	1.18
500	0	0.00	0.00	0.00	−0.00	−0.00	−0.00	0.00	0.00	0.00
	0.3	0.06	0.01	0.00	0.06	0.01	0.00	0.07	0.01	0.00
	0.6	0.30	0.10	0.04	0.30	0.09	0.03	0.30	0.10	0.03
	0.9	1.92	1.23	0.60	1.92	1.23	0.56	1.94	1.25	0.53
					Bias, $\rho_3 = 0.9$					
100	0	−0.03	−0.03	−0.02	−0.03	−0.03	−0.02	0.01	0.01	0.01
	0.3	0.29	0.08	0.04	0.28	0.07	0.03	0.32	0.12	0.07
	0.6	1.21	0.65	0.46	1.20	0.63	0.44	1.25	0.71	0.49
	0.9	4.75	4.01	3.53	4.73	3.94	3.46	4.75	4.09	3.56
250	0	−0.00	0.00	−0.00	−0.01	−0.01	−0.00	−0.00	0.00	0.00
	0.3	0.13	0.03	0.02	0.12	0.02	0.01	0.13	0.03	0.02
	0.6	0.58	0.24	0.16	0.56	0.22	0.15	0.58	0.24	0.17
	0.9	3.07	2.25	1.84	3.00	2.15	1.76	3.11	2.27	1.82
500	0	−0.00	−0.00	−0.00	−0.01	−0.01	−0.01	0.00	0.00	0.00
	0.3	0.06	0.01	0.01	0.06	0.00	−0.00	0.06	0.01	0.01
	0.6	0.29	0.09	0.07	0.30	0.09	0.06	0.29	0.10	0.07
	0.9	1.89	1.21	0.94	1.94	1.23	0.93	1.91	1.25	0.93
					RMSE, $\rho_3 = 0$					
100	0	0.49	0.50	0.14	0.34	0.35	0.09	0.15	0.15	0.04
	0.3	0.72	0.66	0.20	0.55	0.47	0.13	0.38	0.23	0.06
	0.6	1.59	1.20	0.45	1.44	0.99	0.35	1.28	0.75	0.27
	0.9	5.19	4.54	3.11	4.99	4.29	2.91	4.78	4.06	2.75
250	0	0.15	0.15	0.07	0.10	0.10	0.04	0.05	0.05	0.02
	0.3	0.25	0.21	0.09	0.19	0.15	0.06	0.14	0.07	0.03
	0.6	0.69	0.42	0.19	0.63	0.34	0.14	0.59	0.26	0.09
	0.9	3.31	2.51	1.52	3.14	2.33	1.34	3.06	2.24	1.20
500	0	0.06	0.06	0.03	0.04	0.04	0.02	0.02	0.02	0.01
	0.3	0.11	0.08	0.05	0.09	0.06	0.03	0.07	0.03	0.01
	0.6	0.34	0.17	0.09	0.32	0.14	0.07	0.30	0.11	0.04
	0.9	2.05	1.39	0.72	1.98	1.31	0.63	1.95	1.27	0.55
					RMSE, $\rho_3 = 0.9$					
100	0	1.15	1.16	0.70	1.09	1.11	0.64	1.06	1.07	0.60
	0.3	1.57	1.52	0.95	1.49	1.46	0.88	1.44	1.41	0.82
	0.6	2.77	2.44	1.60	2.64	2.32	1.49	2.54	2.23	1.39
	0.9	7.19	6.73	5.27	6.95	6.39	4.97	6.75	6.27	4.83
250	0	0.37	0.37	0.27	0.36	0.37	0.26	0.37	0.37	0.25
	0.3	0.54	0.52	0.38	0.54	0.51	0.36	0.53	0.51	0.35
	0.6	1.11	0.90	0.67	1.11	0.89	0.64	1.11	0.90	0.62
	0.9	4.39	3.67	2.90	4.34	3.59	2.79	4.37	3.62	2.75
500	0	0.16	0.16	0.13	0.15	0.15	0.12	0.16	0.16	0.12
	0.3	0.24	0.22	0.18	0.23	0.22	0.17	0.24	0.22	0.17
	0.6	0.54	0.40	0.32	0.53	0.40	0.31	0.54	0.41	0.31
	0.9	2.68	2.04	1.59	2.67	2.01	1.54	2.69	2.03	1.54

Note: See note of Table 1.

based on the pooled estimator of de Jong and Wagner (2022), labeled as W_{P}^{+}.[15] Specifically, we consider the null hypothesis $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$. To assess power, we generate data for a sequence of 20 alternative values for the vector β. Reflecting the different convergence rates of the components of β, we choose (including also the null values) 21 equidistant values for β_1 in the interval $[5, 7]$, for β_2 in the interval $[-3, -2]$ and for β_3 in the interval $[0.3, 0.7]$. The selection of tests does not include the "direct" tests as they do not provide any extra value added. The simulations have shown that for small values of T they are very conservative, with empirical null rejection probabilities often very close to zero, and for large values of T their performance is (as expected) very similar to the performance of their "non-direct" counterparts.

As indicated already above, the tests – as an immediate consequence of estimator performance – generally also perform better in the individual-specific intercepts only case than when linear trends are also included. This effect becomes more pronounced for increasing ρ-parameters, see and compare, e.g., Tables 3 and 4 for the results in case $\mu_i \neq 0$, $i = 1, \ldots, N$.[16] Many of the observed features are in line with expectations: First, size distortions increase with increasing ρ-parameters. This effect occurs most visibly for W_{TB}, which, as mentioned, only leads to asymptotically valid inference in case all ρ-parameters are equal to zero. If N is *large* compared to T, we observe the phenomenon of "size-divergence" (see, e.g., Wagner and Hlouskova, 2009), i.e., increasing size distortions for increasing N and fixed (small) T.[17] The (relative) behavior of W^{+} and W_{rob}^{+} is also as expected: Both tests are, by construction, less adversely affected than, e.g., W when ρ_{i1}, ρ_{i2} increase, at least for small values of ρ_3. Increasing ρ_3 leads to smaller size distortions – partly substantially smaller size distortions – of W_{rob}^{+} than of W^{+}. This indicates that robust inference indeed works. The group-mean OLS-based robust test W_{rob} is much less affected by increasing ρ_3 than one would expect, with this being driven by our DGP that generates strong contemporaneous cross-section dependence for large values of ρ_3. The test based on the pooled FM-OLS estimator of de Jong and Wagner (2022) is very strongly adversely affected by cross-section dependence, visible already for $\rho_3 = 0.3$. W_{P}^{+} is strongly outperformed by W_{rob}^{+} and even by W_{rob} in case of cross-section dependence. Altogether, in case of unknown forms of error serial correlation, regressor endogeneity, and cross-section dependence, W_{rob}^{+} is the overall best performing test with the smallest size distortions under the null hypothesis. W_{rob}^{+} performs similarly to W^{+} even when all ρ-parameters are equal to zero and is thus, from the null rejection probabilities perspective, the best choice.

We close the simulation section by looking at "size-corrected" power. Figures 1 and 2 display results for $T = 100$, $\rho_1, \rho_2 = 0.6$ and $\mu_i \neq 0$, $i = 1, \ldots, N$ for the individual-specific intercepts only and the individual-specific intercepts and linear trends cases, respectively.[18] Some observations emerge: First, whilst the empirical null rejection probabilities are hardly affected by the absence or presence of drifts, size-corrected power is higher when all drifts are non-zero. Second, larger values of ρ_{i1}, ρ_{i2} lead to smaller size-corrected power. Third, size-corrected power increases unequivocally with an increasing time dimension T, whereas increasing N has only minor impact on size-corrected power in case of cross-section dependence. Fourth, effectively by construction, the test based on the pooled estimator of de Jong and Wagner (2022) exhibits the highest size-corrected power (which, however, has to be seen in conjunction with the very large size distortions in case of cross-section dependence). Fifth, size-corrected power is often the second highest for W_{rob}^{+} and is for large values of ρ_3 closely followed by size-corrected power of W_{rob}. These findings, in conjunction with the behavior under the null hypothesis, lead to the conclusion that

[15]For the pooled estimator the so-called standard covariance estimator is used, see de Jong and Wagner (2022) for details.

[16]Tables 17 and 18 in the Supplementary Material provide the corresponding results for $\mu_i = 0$ for $i = 1, \ldots, N$. The absence or presence of drifts exhibits very limited impact on the null rejection probabilities.

[17]The test based on the pooled estimator of de Jong and Wagner (2022) is particularly strongly affected by size-divergence.

[18]Figures 4 and 5 in the Supplementary Material provide the corresponding results for $\mu_i = 0$ for $i = 1, \ldots, N$.

Table 3. Empirical null rejection probabilities of Wald-type tests for $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$ in the individual-specific intercepts only case with non-zero drifts.

		$N=10$					$N=20$					$N=100$				
T	ρ_1, ρ_2	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P
								$\rho_3 = 0$								
100	0	0.06	0.07	0.08	0.08	0.06	0.07	0.08	0.09	0.09	0.07	0.06	0.07	0.07	0.07	0.06
	0.3	0.22	0.11	0.12	0.12	0.11	0.25	0.13	0.13	0.13	0.11	0.40	0.25	0.11	0.11	0.11
	0.6	0.58	0.20	0.15	0.15	0.18	0.67	0.27	0.15	0.14	0.20	0.95	0.68	0.16	0.16	0.35
	0.9	0.89	0.29	0.26	0.25	0.55	0.94	0.32	0.24	0.23	0.71	1.00	0.54	0.34	0.23	0.99
250	0	0.05	0.06	0.07	0.06	0.06	0.05	0.06	0.06	0.06	0.05	0.05	0.05	0.06	0.06	0.05
	0.3	0.26	0.11	0.10	0.11	0.09	0.27	0.12	0.10	0.10	0.08	0.47	0.25	0.09	0.09	0.08
	0.6	0.69	0.22	0.13	0.13	0.12	0.77	0.32	0.12	0.12	0.12	0.99	0.81	0.14	0.12	0.16
	0.9	0.97	0.39	0.23	0.22	0.30	0.99	0.47	0.21	0.18	0.36	1.00	0.86	0.29	0.17	0.76
500	0	0.05	0.05	0.05	0.06	0.06	0.05	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05
	0.3	0.25	0.09	0.09	0.09	0.08	0.28	0.10	0.09	0.09	0.08	0.45	0.22	0.08	0.08	0.07
	0.6	0.69	0.19	0.11	0.11	0.10	0.77	0.27	0.10	0.10	0.10	0.99	0.77	0.12	0.10	0.10
	0.9	0.98	0.40	0.20	0.20	0.19	0.99	0.51	0.20	0.18	0.22	1.00	0.92	0.28	0.13	0.41
								$\rho_3 = 0.3$								
100	0	0.13	0.07	0.15	0.08	0.21	0.18	0.08	0.21	0.08	0.33	0.41	0.07	0.43	0.08	0.70
	0.3	0.33	0.12	0.22	0.13	0.28	0.40	0.14	0.27	0.13	0.41	0.67	0.17	0.49	0.13	0.75
	0.6	0.66	0.21	0.26	0.17	0.36	0.74	0.24	0.31	0.17	0.49	0.94	0.41	0.52	0.18	0.81
	0.9	0.91	0.33	0.37	0.31	0.63	0.94	0.35	0.41	0.29	0.75	0.99	0.44	0.62	0.32	0.94
250	0	0.18	0.06	0.20	0.06	0.26	0.28	0.06	0.29	0.06	0.41	0.61	0.06	0.62	0.06	0.79
	0.3	0.44	0.11	0.27	0.11	0.32	0.55	0.12	0.36	0.11	0.47	0.81	0.13	0.67	0.10	0.82
	0.6	0.78	0.21	0.31	0.15	0.36	0.86	0.25	0.40	0.15	0.50	0.97	0.36	0.69	0.14	0.84
	0.9	0.97	0.38	0.42	0.27	0.51	0.98	0.40	0.47	0.24	0.64	1.00	0.50	0.72	0.25	0.91
500	0	0.25	0.06	0.25	0.06	0.32	0.40	0.06	0.41	0.06	0.51	0.76	0.06	0.76	0.06	0.87
	0.3	0.52	0.09	0.31	0.10	0.37	0.65	0.10	0.46	0.10	0.55	0.88	0.11	0.79	0.10	0.89
	0.6	0.82	0.17	0.35	0.13	0.39	0.89	0.19	0.50	0.12	0.57	0.97	0.25	0.81	0.13	0.89
	0.9	0.98	0.35	0.46	0.23	0.46	0.99	0.39	0.58	0.24	0.63	1.00	0.45	0.83	0.22	0.91
								$\rho_3 = 0.6$								
100	0	0.30	0.08	0.33	0.09	0.40	0.43	0.07	0.45	0.08	0.58	0.76	0.07	0.78	0.08	0.89
	0.3	0.52	0.14	0.41	0.15	0.48	0.62	0.13	0.52	0.13	0.64	0.88	0.14	0.80	0.14	0.91
	0.6	0.78	0.22	0.45	0.21	0.54	0.85	0.23	0.56	0.19	0.69	0.97	0.28	0.82	0.20	0.93
	0.9	0.95	0.37	0.56	0.38	0.72	0.96	0.38	0.63	0.38	0.84	0.99	0.43	0.87	0.42	0.97
250	0	0.41	0.06	0.43	0.07	0.48	0.56	0.06	0.58	0.06	0.66	0.86	0.06	0.86	0.07	0.93
	0.3	0.64	0.11	0.49	0.12	0.53	0.77	0.10	0.64	0.11	0.71	0.94	0.12	0.88	0.11	0.94
	0.6	0.87	0.20	0.54	0.17	0.57	0.93	0.21	0.67	0.16	0.74	0.99	0.25	0.89	0.17	0.95
	0.9	0.98	0.38	0.63	0.32	0.68	0.99	0.38	0.73	0.32	0.81	1.00	0.42	0.91	0.33	0.96
500	0	0.50	0.06	0.51	0.06	0.55	0.68	0.06	0.68	0.06	0.73	0.92	0.06	0.93	0.06	0.96
	0.3	0.73	0.10	0.56	0.10	0.60	0.84	0.10	0.72	0.10	0.77	0.97	0.10	0.94	0.10	0.96
	0.6	0.91	0.16	0.59	0.14	0.62	0.95	0.16	0.75	0.14	0.78	0.99	0.17	0.95	0.14	0.97
	0.9	0.99	0.34	0.69	0.26	0.67	1.00	0.35	0.80	0.28	0.82	1.00	0.35	0.96	0.26	0.98
								$\rho_3 = 0.9$								
100	0	0.66	0.09	0.68	0.10	0.70	0.81	0.08	0.83	0.09	0.84	0.97	0.09	0.98	0.10	0.98
	0.3	0.80	0.15	0.74	0.17	0.75	0.90	0.13	0.86	0.16	0.87	0.99	0.15	0.98	0.16	0.99
	0.6	0.92	0.22	0.78	0.25	0.79	0.96	0.21	0.88	0.25	0.89	1.00	0.23	0.99	0.25	0.99
	0.9	0.98	0.44	0.84	0.50	0.87	0.99	0.44	0.91	0.49	0.94	1.00	0.45	0.98	0.50	0.99
250	0	0.71	0.06	0.73	0.06	0.73	0.86	0.06	0.87	0.06	0.88	0.98	0.06	0.98	0.07	0.99
	0.3	0.86	0.10	0.77	0.12	0.78	0.94	0.10	0.89	0.11	0.91	0.99	0.11	0.99	0.12	0.99
	0.6	0.95	0.18	0.81	0.17	0.80	0.98	0.17	0.91	0.17	0.91	1.00	0.18	0.99	0.18	0.99
	0.9	1.00	0.38	0.85	0.38	0.85	1.00	0.39	0.93	0.38	0.93	1.00	0.40	0.99	0.38	0.99
500	0	0.76	0.05	0.76	0.06	0.77	0.89	0.06	0.89	0.06	0.90	0.99	0.06	0.99	0.06	0.99
	0.3	0.89	0.09	0.80	0.09	0.80	0.95	0.10	0.91	0.10	0.91	0.99	0.09	0.99	0.10	0.99
	0.6	0.96	0.14	0.82	0.13	0.82	0.99	0.15	0.92	0.14	0.92	1.00	0.14	0.99	0.13	0.99
	0.9	1.00	0.31	0.86	0.28	0.85	1.00	0.32	0.94	0.29	0.94	1.00	0.32	0.99	0.28	0.99

Note: The column labels are as defined in the main text of Section 3.

Table 4. Empirical null rejection probabilities of Wald-type tests for $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$ in the individual-specific intercepts and linear trends case with non-zero drifts.

		$N = 10$					$N = 20$					$N = 100$				
T	ρ_1, ρ_2	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P
									$\rho_3 = 0$							
100	0	0.06	0.08	0.09	0.09	0.07	0.07	0.08	0.09	0.09	0.07	0.06	0.07	0.08	0.08	0.07
	0.3	0.26	0.16	0.14	0.15	0.13	0.30	0.19	0.15	0.15	0.14	0.55	0.39	0.15	0.15	0.15
	0.6	0.74	0.40	0.30	0.30	0.35	0.83	0.53	0.36	0.35	0.47	0.99	0.94	0.71	0.65	0.94
	0.9	0.98	0.81	0.83	0.82	0.99	1.00	0.90	0.92	0.91	1.00	1.00	1.00	1.00	1.00	1.00
250	0	0.06	0.06	0.06	0.07	0.06	0.05	0.06	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06
	0.3	0.35	0.18	0.12	0.12	0.10	0.42	0.23	0.11	0.11	0.09	0.77	0.58	0.12	0.12	0.10
	0.6	0.91	0.62	0.29	0.29	0.21	0.96	0.78	0.36	0.35	0.26	1.00	1.00	0.76	0.70	0.65
	0.9	1.00	0.94	0.92	0.91	0.89	1.00	0.98	0.97	0.97	0.98	1.00	1.00	1.00	1.00	1.00
500	0	0.05	0.05	0.06	0.06	0.06	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.06	0.06	0.05
	0.3	0.42	0.20	0.09	0.09	0.09	0.55	0.32	0.10	0.10	0.08	0.95	0.85	0.12	0.11	0.08
	0.6	0.98	0.78	0.24	0.24	0.14	1.00	0.95	0.35	0.33	0.16	1.00	1.00	0.82	0.76	0.40
	0.9	1.00	0.99	0.94	0.93	0.70	1.00	1.00	1.00	0.99	0.89	1.00	1.00	1.00	1.00	1.00
									$\rho_3 = 0.3$							
100	0	0.11	0.08	0.15	0.09	0.17	0.17	0.08	0.20	0.09	0.28	0.39	0.08	0.42	0.09	0.63
	0.3	0.35	0.16	0.22	0.15	0.24	0.44	0.18	0.28	0.15	0.37	0.72	0.25	0.50	0.15	0.71
	0.6	0.79	0.42	0.40	0.32	0.45	0.86	0.49	0.49	0.37	0.59	0.98	0.71	0.77	0.47	0.89
	0.9	0.99	0.81	0.86	0.83	0.97	0.99	0.87	0.92	0.89	0.99	1.00	0.97	0.99	0.98	1.00
250	0	0.16	0.06	0.17	0.07	0.20	0.24	0.06	0.26	0.07	0.33	0.58	0.06	0.59	0.06	0.75
	0.3	0.50	0.17	0.25	0.12	0.26	0.61	0.19	0.34	0.12	0.39	0.87	0.26	0.66	0.12	0.78
	0.6	0.92	0.56	0.42	0.28	0.37	0.96	0.65	0.56	0.32	0.52	1.00	0.81	0.83	0.39	0.87
	0.9	1.00	0.90	0.91	0.87	0.89	1.00	0.93	0.95	0.91	0.97	1.00	0.98	1.00	0.97	1.00
500	0	0.19	0.06	0.20	0.06	0.25	0.33	0.06	0.34	0.06	0.41	0.72	0.06	0.73	0.06	0.82
	0.3	0.59	0.17	0.27	0.10	0.29	0.72	0.21	0.41	0.11	0.47	0.93	0.29	0.77	0.10	0.85
	0.6	0.97	0.65	0.41	0.22	0.36	0.99	0.76	0.57	0.25	0.53	1.00	0.88	0.88	0.32	0.89
	0.9	1.00	0.96	0.93	0.86	0.78	1.00	0.98	0.97	0.93	0.92	1.00	0.99	1.00	0.97	1.00
									$\rho_3 = 0.6$							
100	0	0.30	0.08	0.34	0.10	0.39	0.44	0.08	0.48	0.09	0.59	0.78	0.08	0.80	0.09	0.89
	0.3	0.54	0.15	0.43	0.17	0.48	0.66	0.16	0.55	0.16	0.66	0.90	0.19	0.83	0.16	0.91
	0.6	0.85	0.38	0.57	0.33	0.61	0.90	0.41	0.67	0.33	0.75	0.99	0.51	0.89	0.38	0.94
	0.9	0.99	0.78	0.88	0.79	0.95	1.00	0.81	0.93	0.83	0.98	1.00	0.89	0.99	0.90	1.00
250	0	0.40	0.07	0.42	0.07	0.46	0.56	0.06	0.58	0.07	0.66	0.88	0.06	0.88	0.06	0.93
	0.3	0.68	0.15	0.50	0.13	0.53	0.80	0.14	0.64	0.12	0.71	0.95	0.16	0.90	0.12	0.94
	0.6	0.94	0.42	0.61	0.25	0.61	0.97	0.45	0.73	0.25	0.76	1.00	0.51	0.93	0.27	0.95
	0.9	1.00	0.83	0.91	0.79	0.90	1.00	0.85	0.95	0.81	0.95	1.00	0.90	0.99	0.84	1.00
500	0	0.47	0.06	0.48	0.06	0.52	0.67	0.06	0.67	0.06	0.73	0.93	0.06	0.93	0.06	0.96
	0.3	0.74	0.13	0.55	0.10	0.57	0.85	0.14	0.72	0.10	0.76	0.98	0.15	0.95	0.12	0.97
	0.6	0.97	0.42	0.64	0.19	0.62	0.98	0.45	0.78	0.19	0.79	1.00	0.50	0.96	0.21	0.97
	0.9	1.00	0.85	0.91	0.71	0.84	1.00	0.87	0.95	0.73	0.94	1.00	0.91	0.99	0.78	0.99
									$\rho_3 = 0.9$							
100	0	0.67	0.10	0.70	0.11	0.71	0.83	0.09	0.86	0.10	0.87	0.98	0.10	0.98	0.10	0.98
	0.3	0.83	0.16	0.77	0.18	0.77	0.92	0.16	0.89	0.17	0.89	0.99	0.17	0.98	0.18	0.99
	0.6	0.94	0.31	0.83	0.32	0.83	0.97	0.32	0.91	0.32	0.91	1.00	0.34	0.99	0.33	0.99
	0.9	1.00	0.72	0.94	0.75	0.95	1.00	0.73	0.97	0.76	0.98	1.00	0.75	0.99	0.77	1.00
250	0	0.72	0.06	0.73	0.07	0.74	0.87	0.07	0.88	0.07	0.89	0.98	0.07	0.98	0.07	0.99
	0.3	0.87	0.12	0.78	0.12	0.79	0.94	0.12	0.91	0.13	0.91	0.99	0.13	0.99	0.13	0.99
	0.6	0.97	0.27	0.83	0.21	0.83	0.99	0.27	0.92	0.22	0.93	1.00	0.28	0.99	0.23	0.99
	0.9	1.00	0.69	0.94	0.66	0.94	1.00	0.69	0.97	0.66	0.97	1.00	0.70	1.00	0.68	1.00
500	0	0.77	0.06	0.77	0.06	0.78	0.89	0.06	0.89	0.06	0.90	0.99	0.06	0.99	0.06	0.99
	0.3	0.90	0.10	0.81	0.10	0.81	0.96	0.11	0.91	0.10	0.92	1.00	0.10	0.99	0.11	0.99
	0.6	0.98	0.23	0.84	0.16	0.84	0.99	0.24	0.93	0.16	0.93	1.00	0.26	0.99	0.16	0.99
	0.9	1.00	0.64	0.94	0.54	0.93	1.00	0.65	0.97	0.54	0.97	1.00	0.67	1.00	0.55	1.00

Note: See note of Table 3.

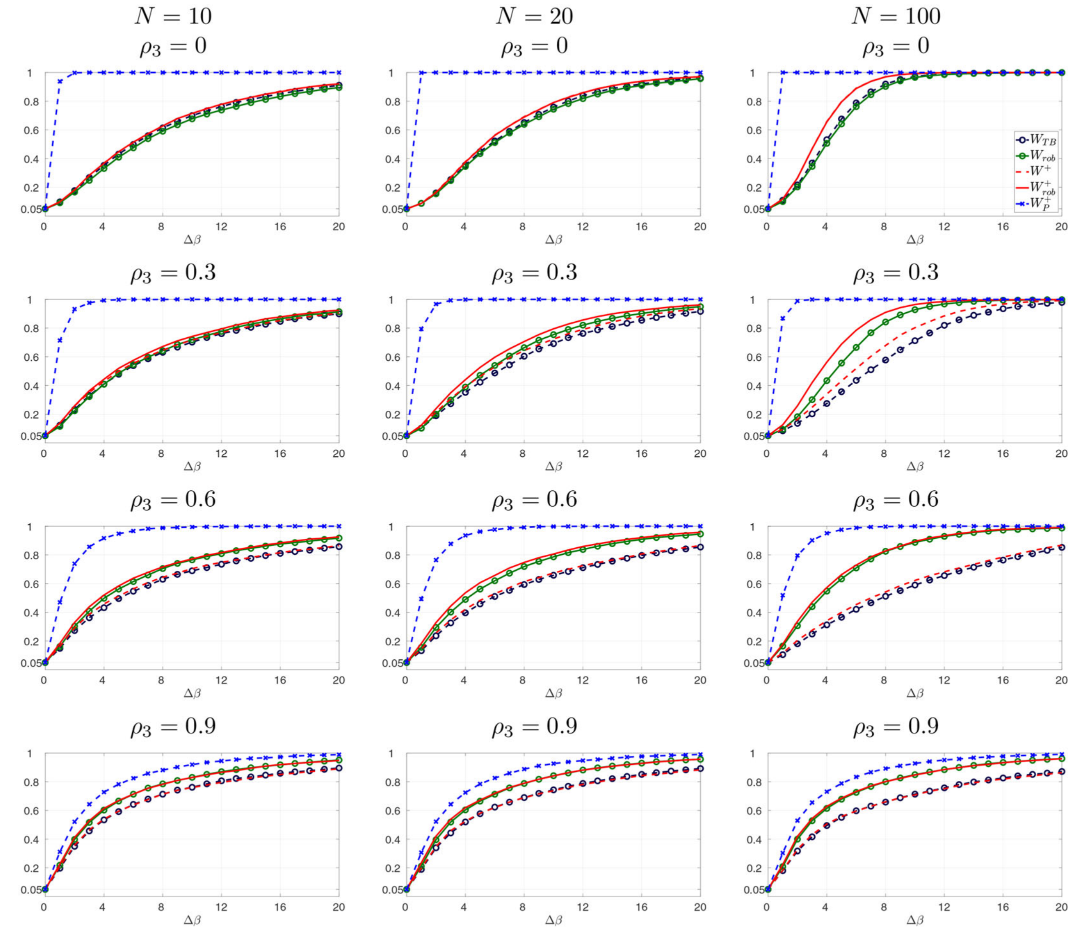

Figure 1. Size-corrected power of the tests for $T = 100$ and $\rho_1, \rho_2 = 0.6$ in the individual-specific intercepts only case with non-zero drifts. Note: The axis label $\Delta\beta$ indicates, see also the description in the main text, the difference between the parameter vector under the null hypothesis and for the considered alternatives, i.e., $\beta_{H_1} = \beta + j \times \Delta\beta$, with $\beta = (5, -3, 0.3)'$, $\Delta\beta = (0.1, 0.05, 0.02)'$ and $j = 0, 1, \ldots, 20$ (displayed on the horizontal axis).

in applications, where one typically does not know the dependence structure, it is the best choice to use W^+_{rob}, i.e., the robust version of the group-mean FM-OLS-based test statistic.[19]

4. An illustration: the environmental Kuznets curve for carbon dioxide emissions

In this section, we briefly illustrate the group-mean FM-OLS estimator as well as inference based upon it by estimating EKCs for carbon dioxide (CO_2) emissions. The dependent variable is the logarithm of per capita CO_2 emissions and the explanatory variables are log per capita GDP and its powers. We consider both the quadratic and the cubic specification as well as the inclusion of individual-specific intercepts only and of both individual-specific intercepts and linear trends. Long-run covariance estimation uses the Bartlett kernel and the Andrews (1991) bandwidth selection rule.

We use exactly the same data as de Jong and Wagner (2022). These are the *long data set* with $N = 19$ countries for $T = 136$ years and the *wide data set* with $N = 89$ countries and $T = 54$ years.

[19]Note that even in the absence of cross-section dependence, i. e., $\rho_3 = 0$, the non-robust version of the Wald-type test, W^+, does not have larger size-corrected power than W^+_{rob}.

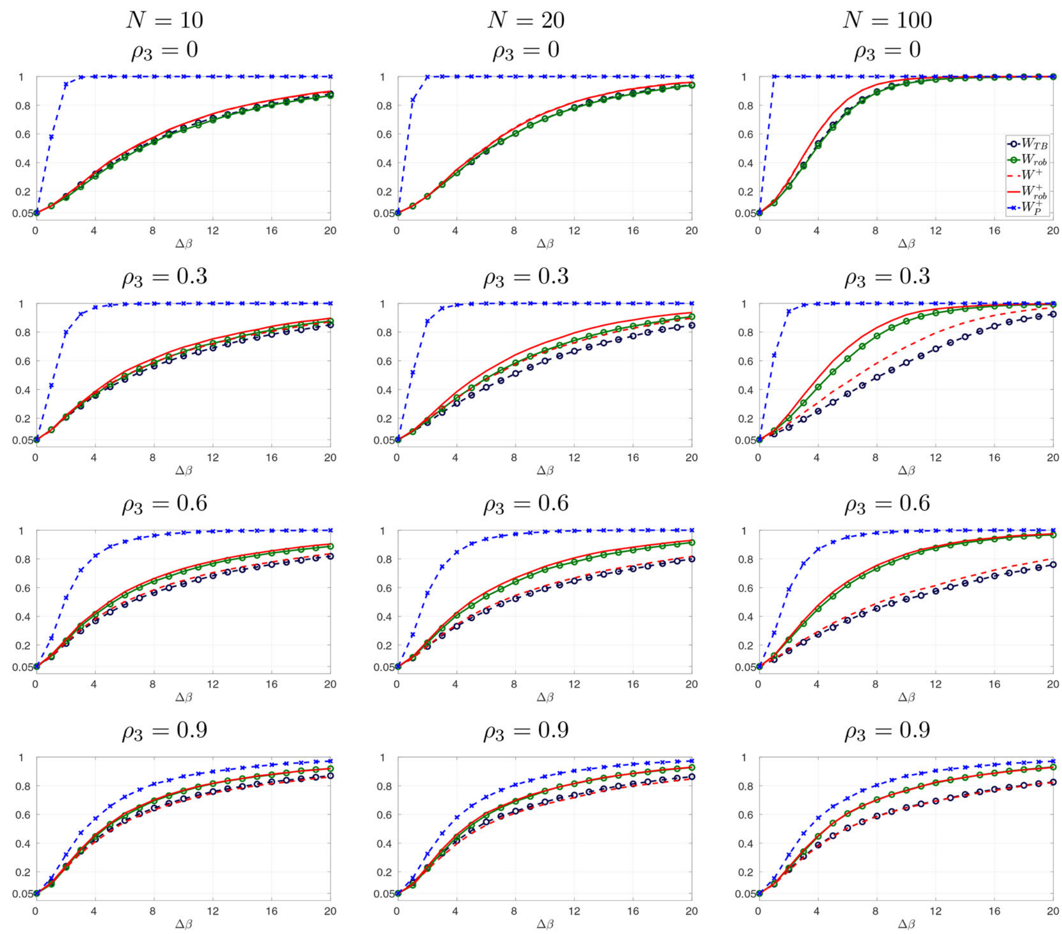

Figure 2. Size-corrected power of the tests for $T = 100$ and $\rho_1, \rho_2 = 0.6$ in the individual-specific intercepts and linear trends only case with non-zero drifts. Note: See note of Figure 1.

The long data set has originally been used in Wagner et al. (2020) and ranges from 1878 to 2013 for 19 early industrialized countries.[20] We also consider a subset comprising six of these 19 countries analyzed in more detail in a seemingly unrelated regression setting in Wagner et al. (2020). These six countries are Austria (AT), Belgium (BE), Finland (FI), the Netherlands (NL), Switzerland (CH) and the United Kingdom (UK), with data for these countries available from 1870 to 2013, leading to a sample size of $T = 144$. The country list for the wide data set, with time span 1960–2013, is available in Table B.1 in Appendix B.

Table 5 shows all estimation results – including standard and robust t-statistics – as well as the implied turning points (TPs). To facilitate comparison with de Jong and Wagner (2022) the TPs obtained in that paper are also included in the rows labeled "TP de J&W." The upper panel considers individual-specific intercepts only and the lower panel considers individual-specific intercepts and linear trends. The left block-column shows the results for the quadratic

[20]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 (2022). 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. (2020).

Table 5. Group-mean fully modified OLS EKC estimation results.

	Quadratic specification			Cubic specification		
	$N=6$	$N=19$	$N=89$	$N=6$	$N=19$	$N=89$
Individual-specific intercepts only						
β_1	7.63	8.24	9.16	−26.22	0.46	1061.85
	(**9.66**)	(**15.38**)	(**3.26**)	(*−1.65*)	(0.04)	(**2.53**)
	[**6.06**]	[**6.46**]	[**3.06**]	[−1.03]	[0.02]	[**2.46**]
β_2	−0.38	−0.42	−0.44	3.43	0.43	−148.79
	(**−8.65**)	(**−13.98**)	(**−2.44**)	(*1.92*)	(0.33)	(**−2.72**)
	[**−5.43**]	[**−5.83**]	[**−2.30**]	[1.20]	[0.14]	[**−2.65**]
β_3				−0.14	−0.03	6.80
				(**−2.13**)	(−0.64)	(**2.84**)
				[−1.33]	[−0.27]	[**2.78**]
TP GM	20,951	19,470	35,596	16,854	19,587	4,211
				548	1	510
TP de J&W	14,051	20,054	531,260	–	–	43,231
				–	–	443
Individual-specific intercepts and linear trends						
β_1	9.92	8.74	11.54	15.83	26.69	−952.79
	(**15.58**)	(**18.58**)	(**4.72**)	(1.57)	(**2.87**)	(**−2.75**)
	[**12.22**]	[**8.16**]	[**4.11**]	[1.18]	[1.45]	[**−2.77**]
β_2	−0.48	−0.43	−0.59	−1.18	−2.51	114.93
	(**−14.06**)	(**−17.38**)	(**−3.86**)	(−1.04)	(**−2.43**)	(**2.55**)
	[**−10.95**]	[**−7.41**]	[**−3.40**]	[−0.78]	[−1.21]	[**2.56**]
β_3				0.03	0.08	−4.71
				(0.67)	(**2.11**)	(**−2.40**)
				[0.50]	[1.03]	[**−2.42**]
TP GM	33,743	25,889	17,027	1.2×10^7	–	–
				94,276	–	–
TP de J&W	23,967	26,284	72,329	–	–	29,519
				–	–	578

Notes: Standard *t*-statistics, defined in (17), in parentheses and robust *t*-statistics, defined in (21), in square brackets. *Italic* numbers indicate significance at the 10% nominal level and **bold** numbers indicate significance at the 5% significance level. The turning points based on the group-mean estimator (TP GM) are computed as $\exp\left(-\frac{\hat{\beta}_1}{2\hat{\beta}_2}\right)$ in the quadratic case and as $\exp\left(-\frac{\hat{\beta}_2}{3\hat{\beta}_3}(\pm1)\left(-\frac{\hat{\beta}_1}{3\hat{\beta}_3}+(\frac{\hat{\beta}_2}{3\hat{\beta}_3})^2\right)^{1/2}\right)$ in the cubic case. The symbol "–" indicates the absence of turning points for the estimated polynomial. The row labeled "TP de J&W" contains the turning points given in de Jong and Wagner (2022, Tables 7 and 8) using the pooled FM-OLS estimator in a slightly different specification with, in the lower panel (common) time effects instead of individual-specific linear time trends.

specification and the right block-column shows the results for the cubic specification. The first question to be addressed concerns the polynomial degree of the EKC, i.e., whether a cubic specification has to be considered or the quadratic specification suffices. With respect to this question, it turns out that robust inference leads to different conclusions than standard inference. For both $N=6$ and $N=19$ the use of robust inference leads to insignificant coefficients to the third power of the logarithm of per capita GDP; for both the intercept only and the intercept and trend case. For the wide data set with $N=89$ the cubic specification is required, in the sense that both standard and – more importantly – robust *t*-statistics indicate significance of the third-order coefficient for both specifications of the deterministic component.

Based on the above, we focus on the findings with the quadratic specification for the $N=6$ and $N=19$ data sets. For both specifications of the deterministic components, the coefficient to the squared logarithm of per capita GDP is (significantly) negative, with both standard and robust *t*-statistics. For $N=6$, the TPs differ substantially between the group-mean estimator and the pooled estimator of de Jong and Wagner (2022) and are substantially larger for the

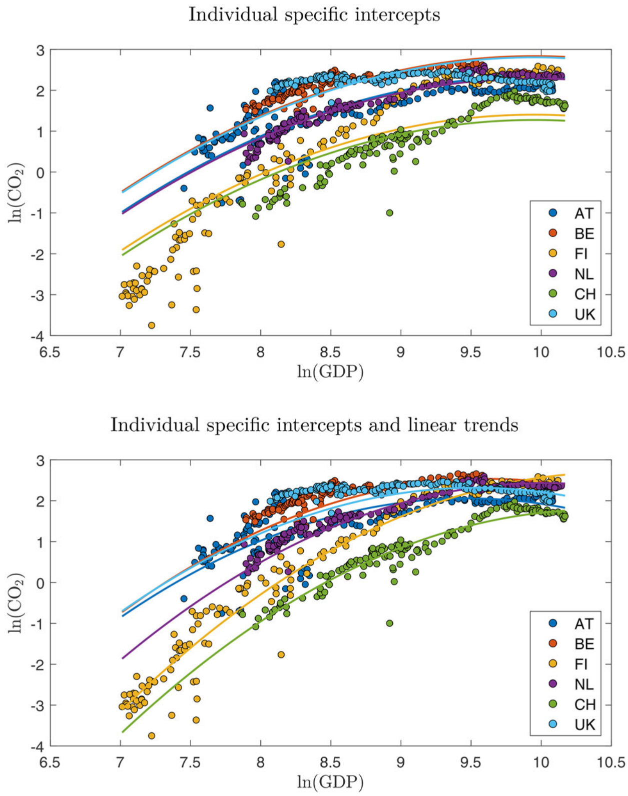

Figure 3. Scatter plot and estimated EKC relationship for CO_2 emissions over the period 1870–2013 for the $N = 6$ data set. Notes: The curves display the results of inserting 144 equidistant points from the sample range of $\ln(\text{GDP})$ in the quadratic relationship estimated with group-mean fully modified OLS estimator and adding the individual-specific intercepts (top panel) or the individual-specific intercepts and linear time trends (bottom panel), with corresponding values of the time trend given by $t = 1, \ldots, 144$.

group-mean estimator. For $N = 19$ the differences in the TPs between the group-mean and pooled estimators are negligible.[21]

Figure 3 shows the impact of including individual-specific linear trends (in the lower graph) in addition to individual-specific intercepts only (in the upper graph) on the estimated EKCs for $N = 6$. Including individual-specific linear time trends (obviously) leads to a better fit, in particular for Finland and Switzerland, both for the low GDP values, i.e., for the beginning of the sample period, and the high GDP values, i.e., for the end of the sample period. Thus, the different "average levels" of log per capita emissions are well captured by the individual-specific intercepts, the individual-specific trends allow in addition to account to some extent for "curvature differences" across countries. On the question of poolability of the EKC across these countries see also Wagner et al. (2020), who in fact only find evidence for – in the words of that paper – *partial poolability* of the slope coefficients for Belgium, the Netherlands, and the UK. Against this background, this empirical section is to be interpreted merely as an illustration. For larger values of N, of course, the seemingly unrelated regressions-based analysis of Wagner et al. (2020) is not feasible and one needs to resort to panel-type methods of one kind or another with the corresponding cross-sectional pooling imposed.

[21]The sample range for the $N = 6$ data set is from 1,725 to 26,102 and for the $N = 19$ data set the sample range is from 794 to 31,933 (measured in 1990 Geary-Khamis dollars). Therefore, the group-mean TP when including individual-specific intercepts and linear trends is out of sample for the $N = 6$ data set.

Let us close this illustration section with a brief look at the cubic specification results for the $N = 89$ wide data set. One striking feature for this data set is that the signs of the coefficient to the third power differ between the two specifications of the deterministic component, with $\beta_3 > 0$ in the intercept only case and $\beta_3 < 0$ in the intercept and linear trend case.[22] The group-mean estimator leads to two TPs at small values in the intercept only case, with the larger TP corresponding to U-type behavior, and to a monotonic relationship in the intercept and linear trend specification. For the wide data set, it thus appears that the pooled estimator leads to – notwithstanding all issues concerning poolability – more "useful" TPs.

5. Summary and conclusions

This article extends the toolkit for parameter estimation and inference in panels of CPRs with a group-mean FM-OLS approach, which complements the pooled FM-OLS approach of de Jong and Wagner (2022). The consideration of a group-mean rather than a pooled estimation approach is not the only difference between the two papers. This article gains a lot of mileage from considering a fixed cross-section setting, which allows to include two features not considered in de Jong and Wagner (2022). First, we allow for the (potential) presence of drifts in the integrated regressors, which increases applicability substantially. Second, we provide cross-section robust inference for the group-mean OLS and FM-OLS estimators. Asymptotically valid inference is, as discussed, possible under minimal restrictions on the form and extent of cross-section dependence. No specific model of cross-section dependence, e.g., a factor structure, has to be posited. It is important to stress again that computation of the developed estimators and tests does not require any knowledge concerning the presence or absence of drifts and/or cross-section dependence.

The simulation results are, by and large, as expected, with one important exception regarding hypothesis testing: Using the cross-section robust version of the group-mean FM-OLS estimator-based tests is unequivocally the best choice, as the robust version of the tests performs at least as good as the non-robust version of the tests even in the absence of cross-section dependence. The test based on the pooled estimator of de Jong and Wagner (2022) is very strongly adversely affected by cross-section dependence.

The illustrative application conveys two messages: First, cross-section robust inference makes a difference. In our illustration, it indicates, unlike standard inference, that the quadratic specification is sufficient for the long data sets and that a cubic formulation is only required for the wide data set. The wide data set with $N = 89$ (larger than $T = 54$) indicates potential advantages of the pooled estimator in case of large cross-section dimension compared to the time series dimension, i.e., benefits of resorting to an asymptotic approximation also in the cross-section dimension.

Two (related) issues remain open for future research: First, an analysis of the asymptotic behavior of the group-mean estimator in the two-way fixed effects case, i.e., with both individual- and time-specific fixed effects. This will require, second, asymptotic analysis in a large time and large cross-section setting, which is in any case important for panels with N large compared to T. Letting $N \to \infty$ requires that potential cross-section dependence will have to be considered more restrictively than in our fixed N setting; not only with respect to robust inference but also for obtaining, e.g., a sequential (unconditional) asymptotic normality result for the estimated coefficients.

[22] $\beta_3 > 0$ implies that the fitted polynomial diverges to plus infinity for log per capita GDP tending to infinity. Consequently, in case of TPs being present, the larger TP corresponds to U-type rather than an inverted U-type behavior. Note for completeness, see de Jong and Wagner (2022, Table 8), that the pooled estimator leads to negative third-order coefficients for both specifications for this data set.

Appendices

A. Proofs

Proof of Proposition 1. The starting point is:

$$G_T^{-1}\left(\hat{\beta}^+ - \beta\right) = \frac{1}{N}\sum_{i=1}^{N} G_T^{-1}\left(\hat{\beta}^+(i) - \beta\right)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T\right)^{-1}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^+ - G_TC_i\right), \tag{83}$$

with:

$$\tilde{u}_{it}^+ := \tilde{u}_{it} - \Delta x_{it}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i} = \tilde{u}_{it} - (\mu_i + v_{it})\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}. \tag{84}$$

Since $\mu_i = 0$ for all $i = 1, ..., N$, it follows directly from Assumptions 2 and 3 that:

$$G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T \xrightarrow{d} \int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr, \tag{85}$$

$$G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^+ \xrightarrow{d} \int_0^1 \tilde{\mathbf{B}}_{v_i}(r)dB_{u_i\cdot v_i}(r) + \Delta_{v_iu_i}^+A_i, \tag{86}$$

and:

$$G_TC_i \xrightarrow{d} \Delta_{v_iu_i}^+A_i, \tag{87}$$

where $\Delta_{v_iu_i}^+ := \Delta_{v_iu_i} - \Delta_{v_iv_i}\Omega_{v_iv_i}^{-1}\Omega_{v_iu_i}$ and A_i as given in the main text, with all quantities converging jointly. This immediately implies – for the parameter estimator corresponding to the ith equation – that:

$$G_T^{-1}\left(\hat{\beta}^+(i) - \beta\right) \xrightarrow{d} \left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)dB_{u_i\cdot v_i}(r). \tag{88}$$

Conditional upon Δ_i, Σ_i and $W_{v_i}(r)$, the limiting distribution given in (88) is normal with expectation zero and covariance matrix $\Omega_{u_i\cdot v_i}(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr)^{-1}$. Cross-sectional independence (Assumption 1) thus implies the – conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, ..., N$ – asymptotic normality result for the group-mean FM-OLS estimator given in the main text in (12) and (13). □

Proof of Corollary 1. Under the null hypothesis, the Wald-type statistic given in (16) is equal to:

$$W^+ = \left((G_RRG_T)G_T^{-1}(\hat{\beta}^+ - \beta)\right)'\left((G_RRG_T)\hat{V}^+(G_RRG_T)'\right)^{-1}$$
$$\times\left((G_RRG_T)G_T^{-1}(\hat{\beta}^+ - \beta)\right). \tag{89}$$

With the (asymptotic) restriction on the constraint matrix R posited in the main text in place and with $\hat{V}^+ = G_T^{-1}\hat{S}^+G_T^{-1}$ converging in distribution to V^+, it follows from Proposition 1 that:

$$W^+ \xrightarrow{d} (R^*\mathcal{Z})'(R^*V^+R^{*'})^{-1}(R^*\mathcal{Z}), \tag{90}$$

with $\mathcal{Z}$ conditionally $\mathcal{N}(0, V^+)$ distributed. This shows the conditional – and hence unconditional – asymptotic chi-squared null distribution of the Wald-type statistic. In case $s = 1$ analogous arguments lead to the result for the t-type test. □

Proof of Remark 1. Similar arguments as used in the proof of Proposition 1 show that:

$$G_T^{-1}\left(\hat{\beta}^+ - \beta\right) \xrightarrow{d} \frac{1}{N}\sum_{i=1}^{N}\left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)dB_{u_i\cdot v_i}(r). \tag{91}$$

Conditional upon Δ, Σ and $W_v(r)$, the limiting distribution given in (91) is normal with expectation zero and covariance matrix:

$$\begin{aligned}
&\mathrm{Var}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)dB_{u_i\cdot v_i}(r)\right\} \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N}\mathrm{Cov}\left\{\left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)dB_{u_i\cdot v_i}(r),\right. \\
&\qquad \left.\left(\int_0^1 \tilde{\mathbf{B}}_{v_j}(r)\tilde{\mathbf{B}}_{v_j}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{v_j}(r)dB_{u_j\cdot v_j}(r)\right\} \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N}\mathbf{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}\tilde{M}_{ii}^{-1}\tilde{M}_{ij}\tilde{M}_{jj}^{-1} = V_{\mathrm{rob}}^{+},
\end{aligned} \tag{92}$$

where $\tilde{M}_{ij} = \int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_j}(r)'dr$ and $\mathbf{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}$ is the constant in the quadratic covariation of the processes $B_{u_i\cdot v_i}(r)$ and $B_{u_j\cdot v_j}(r)$ and is defined in the main text.

It is straightforward to verify that $\hat{V}_{\mathrm{rob}}^{+} = G_T^{-1}\hat{S}_{\mathrm{rob}}^{+}G_T^{-1}$ converges in distribution to V_{rob}^{+}. Therefore, the null limiting distributions of W_{rob}^{+} and t_{rob}^{+} can be derived with exactly the same arguments as used in the proof of Corollary 1. □

Proof of Proposition 2. We first consider the case with individual-specific intercepts but no individual-specific linear trends included in (1). The proof for the case with both individual-specific intercepts and individual-specific linear trends included in (1) is considered afterwards and is based upon similar arguments.

(i) Similar to the proof of Proposition 1 the starting point is given by:

$$\begin{aligned}
H_T^{-1}\left(\hat{\beta}^{+} - \beta\right) &= \frac{1}{N}\sum_{i=1}^{N}H_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'H_T\right)^{-1}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^{+} - H_TC_i\right).
\end{aligned} \tag{93}$$

By definition of $\tilde{u}_{it}^{+}$ (see, e. g., the proof of Proposition 1) it follows that:

$$\begin{aligned}
\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^{+} &= \sum_{t=1}^{T}\tilde{X}_{it}(\tilde{u}_{it} - v_{it}\hat{\mathbf{\Omega}}_{v_iv_i}^{-1}\hat{\mathbf{\Omega}}_{v_iu_i}) - \mu_i\hat{\mathbf{\Omega}}_{v_iv_i}^{-1}\hat{\mathbf{\Omega}}_{v_iu_i}\sum_{t=1}^{T}\tilde{X}_{it} \\
&= \sum_{t=1}^{T}\tilde{X}_{it}(\tilde{u}_{it} - v_{it}\hat{\mathbf{\Omega}}_{v_iv_i}^{-1}\hat{\mathbf{\Omega}}_{v_iu_i}),
\end{aligned} \tag{94}$$

where the last equality follows from the fact that by construction $\sum_{t=1}^{T}\tilde{X}_{it} = 0$. This implies:

$$\begin{aligned}
&H_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}H_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'H_T\right)^{-1}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}(\tilde{u}_{it} - v_{it}\hat{\mathbf{\Omega}}_{v_iv_i}^{-1}\hat{\mathbf{\Omega}}_{v_iu_i}) - H_TC_i\right).
\end{aligned} \tag{95}$$

As the deterministic trends (asymptotically) dominate the elements of $\tilde{X}_{it}$, it follows that $T^{1/2}H_T\tilde{X}_{i\lfloor rT\rfloor} \Rightarrow J_i(r)$, $H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it} \xrightarrow{d} \int_0^1 J_i(r)dB_{u_i}(r)$, $H_T\sum_{t=1}^{T}\tilde{X}_{it}v_{it} \xrightarrow{d} \int_0^1 J_i(r)dB_{v_i}(r)$ and $H_TC_i = o_{\mathbb{P}}(1)$, with all quantities converging jointly, with $J_i(r)$ as defined in the main text in (28).[23] This immediately implies – for the parameter estimator from the ith equation – that:

$$H_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \xrightarrow{d} \left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)dB_{u_i\cdot v_i}(r). \tag{96}$$

[23]For more details we refer to Reichold and Wagner (2022, Lemma 2).

Conditional upon Δ_i, Σ_i and $W_{v_i}(r)$, the limiting distribution given in (96) is normal with expectation zero and covariance matrix $\Omega_{u_i\cdot v_i}(\int_0^1 J_i(r)J_i(r)'dr)^{-1}$. Cross-sectional independence (Assumption 1) thus implies the – conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, ..., N$ – asymptotic normality result for the group-mean estimator given in the main text in (30).

(ii) Analogously, the starting point for showing (31) is given by:

$$\begin{aligned} K_T^{-1}\left(\hat{\beta}^+ - \beta\right) &= \frac{1}{N}\sum_{i=1}^N K_T^{-1}\left(\hat{\beta}^+(i) - \beta\right) \\ &= \frac{1}{N}\sum_{i=1}^N \left(K_T\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}'K_T\right)^{-1}\left(K_T\sum_{t=1}^T \tilde{X}_{it}\tilde{u}_{it}^+ - K_T C_i\right). \end{aligned} \tag{97}$$

As described in the main text, as a result of demeaning and linear detrending, the linear trend that asymptotically dominates x_{it} is exactly annihilated in $\tilde{x}_{it}$. This is reflected in the following joint convergence results that can be derived with similar calculations as in Reichold and Wagner (2022, Proof of Lemma 2). First, $T^{1/2}K_T\tilde{X}_{i\lfloor rT\rfloor} \Rightarrow L_i(r)$, with $L_i(r)$ as defined in the main text in (29). Moreover:

$$K_T\sum_{t=1}^T \tilde{X}_{it}\tilde{u}_{it} \xrightarrow{d} \int_0^1 L_i(r)dB_{u_i}(r) + (\Delta_{v_iu_i}, 0, 0)', \tag{98}$$

$$\begin{aligned} K_T\sum_{t=1}^T \tilde{X}_{it}\tilde{u}_{it} &\xrightarrow{d} \int_0^1 L_i(r)dB_{u_i}(r) + (\Delta_{v_iu_i}, 0, 0)', \\ K_T\sum_{t=1}^T \tilde{X}_{it}v_{it} &\xrightarrow{d} \int_0^1 L_i(r)dB_{v_i}(r) + (\Delta_{v_iv_i}, 0, 0)' \end{aligned} \tag{99}$$

and $K_TC_i \xrightarrow{d} (\Delta_{v_iu_i}^+, 0, 0)'$. The remaining parts of the proof are similar to the corresponding parts of the proof of (i) and are therefore omitted. □

Proof of Corollary 2. The proof is based on similar arguments as the proof of Corollary 1 and therefore omitted. □

Proof of Remark 6. For sake of brevity, we only consider the individual-specific intercepts only case here in detail. The proof is entirely analogous for the individual-specific intercepts and linear trends case.

It follows from the proof of Proposition 2 that:

$$H_T^{-1}\left(\hat{\beta}^+ - \beta\right) \xrightarrow{d} \frac{1}{N}\sum_{i=1}^N \left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)dB_{u_i\cdot v_i}(r). \tag{100}$$

Conditional upon Δ, Σ and $W_v(r)$, the limiting distribution given in (100) is normal with expectation zero and covariance matrix:

$$\begin{aligned} &\mathrm{Var}\left\{\frac{1}{N}\sum_{i=1}^N \left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)dB_{u_i\cdot v_i}(r)\right\} \\ &= \frac{1}{N^2}\sum_{i,j=1}^N \mathrm{Cov}\left\{\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)dB_{u_i\cdot v_i}(r),\right. \\ &\qquad \left.\left(\int_0^1 J_j(r)J_j(r)'dr\right)^{-1}\int_0^1 J_j(r)dB_{u_j\cdot v_j}(r)\right\} \\ &= \frac{1}{N^2}\sum_{i,j=1}^N \Omega_{u_i\cdot v_i;u_j\cdot v_j}\left(\mathcal{D}(\mu_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\mu_j)\right)^{-1} = V_{\alpha,\mathrm{rob}}^+, \end{aligned}$$

where $\Omega_{u_i\cdot v_i;u_j\cdot v_j}$ is the constant in the quadratic covariation of the processes $B_{u_i\cdot v_i}(r)$ and $B_{u_j\cdot v_j}(r)$ and is defined in the main text.

It is straightforward to verify that both $\hat{V}^{+}_{\alpha,\text{rob}} = H_T^{-1}\hat{S}^{+}_{\text{rob}}H_T^{-1}$ and:

$$\tilde{V}^{+}_{\alpha,\text{rob}} = \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i\dot{v}_i;u_j\cdot v_j}\left(\mathcal{D}(\hat{\mu}_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\hat{\mu}_j)\right)^{-1} \tag{101}$$

converge in distribution to $V^{+}_{\alpha,\text{rob}}$. Therefore, the limiting distributions of W^{+}_{rob} and $W^{+}_{\alpha,\text{rob}}$ can be shown to be chi-squared with s degrees of freedom under the null hypothesis using exactly the same arguments as in the proof of Corollary 1. Similarly, in case $s = 1$, t^{+}_{rob} and $t^{+}_{\alpha,\text{rob}}$ can be shown to be asymptotically standard normally distributed under the null hypothesis. □

Proof of Proposition 3. The case $N_0 = 0$ is contained in the (proof of) Proposition 2. The results for $N_0 > 0$ follow from combining the results of Propositions 1 and 2. As in the proof of Proposition 2, we commence with the individual-specific intercepts only case before turning to the individual-specific intercepts and linear trends case.

(i) First note that the appropriate scaling matrix for the individual-specific estimators $\hat{\beta}^{+}(i)$ calculated from cross-section members with zero drifts in the integrated regressor is G_T, whereas the appropriate scaling matrix for the individual-specific estimators $\hat{\beta}^{+}(i)$ calculated from cross-section members with non-zero drifts in the integrated regressor is H_T. This implies:

$$\begin{aligned} G_T^{-1}\left(\hat{\beta}^{+} - \beta\right) &= \frac{1}{N}\sum_{i=1}^{N} G_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N_0} G_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) + \frac{1}{N}\sum_{i=N_0+1}^{N} G_T^{-1}H_T H_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N_0} G_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) + o_{\mathbb{P}}(1), \end{aligned} \tag{102}$$

where the last equality follows from $G_T^{-1}H_T = \text{diag}(T^{-1/2}, T^{-1}, T^{-3/2})$ and $H_T^{-1}(\hat{\beta}^{+}(i) - \beta) = O_{\mathbb{P}}(1)$ for $i = N_0 + 1, \ldots, N$. Hence, the asymptotic behavior of the group-mean estimator only depends on the individual-specific estimators $\hat{\beta}^{+}(i)$ calculated from cross-section members with zero drifts in the integrated regressor, since these converge at a slower rate than the estimators corresponding to cross-section members with non-zero drifts. The rest of the proof is analogous to the proof of Proposition 1 and therefore omitted.

(ii) As in (i), the appropriate scaling matrix depends upon the absence or presence of a non-zero drift in the integrated regressor. In the former case, the appropriate scaling matrix is again given by G_T, whereas it is given by K_T in the presence of a non-zero drift. Therefore:

$$\begin{aligned} & G_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} G_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N_0} G_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) + \frac{1}{N}\sum_{i=N_0+1}^{N} G_T^{-1}K_T K_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N_0} G_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) + \frac{1}{N}\sum_{i=N_0+1}^{N}\begin{bmatrix} T\left(\hat{\beta}_1^{+}(i) - \beta_1\right) & & \\ & o_{\mathbb{P}}(1) & \\ & & o_{\mathbb{P}}(1) \end{bmatrix}, \end{aligned} \tag{103}$$

where $\hat{\beta}_1^{+}(i)$ denotes the first element of $\hat{\beta}^{+}(i)$. The last equality follows from $G_T^{-1}K_T = \text{diag}(1, T^{-1}, T^{-3/2})$ and $K_T^{-1}(\hat{\beta}^{+}(i) - \beta) = O_{\mathbb{P}}(1)$, for $i = N_0 + 1, \ldots, N$. In contrast to (i), the limiting distribution of the first component of $\hat{\beta}^{+}$ depends upon all cross-section member-specific estimates of β_1, reflecting the fact that the coefficient to the first power of the integrated regressor is estimated at rate T irrespective of whether the drift is zero or non-zero – as in any case linear detrending removes a potentially present linear trend from the corresponding regressor. For the coefficients β_2 and β_3, the situation is exactly as in (i), with the limiting distribution only depending upon the individual-specific estimators corresponding to cross-section members with zero drifts in the integrated regressor, since these are converging at a slower rate than the estimators corresponding to cross-section members with non-zero drifts. The rest of the proof is similar to the proof of Proposition 1 and therefore omitted. □

Proof of Corollary 3. The proof is based on similar arguments as the proofs of Corollary 1 and 2 and therefore omitted. □

Proof of Remark 8. The case $N_0 = 0$ has already been considered in the proof of Remark 6. The results for $N_0 > 0$ follow from combining the results of Proposition 3 and Remark 1, compare also the proofs of Remarks 1 and 6. The proof is therefore omitted. □

B. Country List for the Wide Data Set

Table B.1. Country list for the wide data set.

Albania	Algeria	Angola	Argentina	Australia
Austria	Bahrain	Barbados	Belgium	Bolivia
Brazil	Bulgaria	Cambodia	Cameroon	Canada
Chile	China	Colombia	Costa Rica	Côte d'Ivoire
Cyprus	Denmark	Dominican Republic	DR Congo	Ecuador
Egypt	Ethiopia	Finland	France	Germany
Ghana	Greece	Guatemala	Hong Kong	Hungary
Iceland	India	Indonesia	Iran	Iraq
Ireland	Israel	Italy	Jamaica	Japan
Jordan	Kenya	Luxembourg	Madagascar	Mali
Malta	Mexico	Morocco	Mozambique	Myanmar
Netherlands	New Zealand	Niger	Nigeria	Norway
Pakistan	Peru	Philippines	Poland	Portugal
Romania	Saudi Arabia	Senegal	Singapore	South Africa
South Korea	Spain	Sri Lanka	Saint Lucia	Sudan
Sweden	Switzerland	Syria	Tanzania	Trinidad and Tobago
Tunisia	Turkey	Uganda	United Kingdom	United States
Uruguay	Venezuela	Vietnam	Yemen	

Acknowledgments

The authors gratefully acknowledge partial financial support from Deutsche Forschungsgemeinschaft via the Collaborative Research Center SFB823 *Statistical Modelling of Nonlinear Dynamic Processes* (Projects A3 and A4). Furthermore, we thank the editor Esfandiar Maasoumi and two anonymous reviewers for a number of insightful comments that have led to significant changes and improvements of the paper. In addition, we thank Fabian Knorre for his suggestion to present the empirical results as in Figure 3 and conference participants at the 2021 Annual Conference of the Verein für Socialpolitik in Regensburg. The views expressed in this paper are 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.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Martin Wagner http://orcid.org/0000-0002-6123-4797

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59(3):817–858. doi:10.2307/2938229

Darvas, Z. (2007). Estimation bias and inference in overlapping autoregressions: implications for the Target-Zone literature. *Oxford Bulletin of Economics and Statistics* 0(0):070921170652008–070921170652008. doi:10.1111/j.1468-0084.2007.00488.x

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

de Jong, R. M., Wagner, M. (2022). Panel cointegrating polynomial regression analysis and the environmental Kuznets curve. Forthcoming in *Econometrics and Statistics* doi:10.1016/j.ecosta.2022.03.005
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*, Cambridge: MIT Press, pp. 13–56.
Holtz-Eakin, D., Selden, T. M. (1995). Stoking the fires? CO_2 emissions and economic growth. *Journal of Public Economics* 57(1):85–101. doi:10.1016/0047-2727(94)01449-X
Ibragimov, R., Phillips, P. C. B. (2008). Regression asymptotics using martingale convergence methods. *Econometric Theory* 24(4):888–947. doi:10.1017/S0266466608080365
Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18(6):1449–1459. doi:10.1017/S0266466602186087
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(2):147–161. doi:10.1006/jeem.1993.1039
Malenbaum, W. (1978). *World Demand for Raw Materials in 1985 and 2000*. New York, NY: McGraw-Hill, EMJ Mining Information Services.
Park, J. Y., Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* 69(1):117–161. doi:10.1111/1468-0262.00180
Phillips, P. C. B., Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies* 57(1):99–125. doi:10.2307/2297545
Phillips, P. C. B., Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5):1057–1111. doi:10.1111/1468-0262.00070
Reichold, K., Wagner, M. (2022). *Cointegrating Polynomial Regressions with Integrated Regressors with Drift: Fully Modified OLS Estimation and Inference*. New York, NY: Mimeo.
Sims, C. A., Stock, J. H., Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica* 58(1):113–144. doi:10.2307/2938337
Stern, D. I. (2004). The rise and fall of the environmental Kuznets curve. *World Development* 32(8):1419–1439. doi:10.1016/j.worlddev.2004.03.004
Svensson, L. E. O. (1992). An interpretation of recent research on exchange rate target zones. *Journal of Economic Perspectives* 6(4):119–144. doi:10.1257/jep.6.4.119
Wagner, M. (2015). The environmental Kuznets curve, cointegration and nonlinearity. *Journal of Applied Econometrics* 30(6):948–967. doi:10.1002/jae.2421
Wagner, M., Grabarczyk, P., Hong, S. H. (2020). Fully modified OLS estimation and inference for seemingly unrelated cointegrating polynomial regressions and the environmental Kuznets curve for carbon dioxide emissions. *Journal of Econometrics* 214(1):216–255. doi:10.1016/j.jeconom.2019.05.012
Wagner, M., Hlouskova, J. (2009). The performance of panel cointegration methods: results from a large scale simulation study. *Econometric Reviews* 29(2):182–223. doi:10.1080/07474930903382182
Wagner, M., Hong, S. H. (2016). Cointegrating polynomial regressions: fully modified OLS estimation and inference. *Econometric Theory* 32(5):1289–1315. doi:10.1017/S0266466615000213
West, K. D. (1988). Asymptotic normality, when regressors have a unit root. *Econometrica* 56(6):1397–1417. doi:10.2307/1913104
Yandle, B., Bjattarai, M., Vijayaraghavan, M. (2004). Environmental Kuznets curves: a review of findings, methods, and policy implications. *Research Study 02.1 update, PERC.*