Lin, Yicong; Reuvers, Hanno

Working Paper

Cointegrating Polynomial Regressions With Power Law Trends: Environmental Kuznets Curve or Omitted Time Effects?

Tinbergen Institute Discussion Paper, No. TI 2022-092/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Lin, Yicong; Reuvers, Hanno (2022) : Cointegrating Polynomial Regressions With Power Law Trends: Environmental Kuznets Curve or Omitted Time Effects?, Tinbergen Institute Discussion Paper, No. TI 2022-092/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/273805

TI 2022-092/III
Tinbergen Institute Discussion Paper

Cointegrating Polynomial Regressions With Power Law Trends: Environmental Kuznets Curve or Omitted Time Effects?

Yicong Lin[1]
Hanno Reuvers[2]

1 Vrije Universiteit Amsterdam and Tinbergen Institute

2 Erasmus University Rotterdam

Cointegrating Polynomial Regressions With Power Law Trends: Environmental Kuznets Curve or Omitted Time Effects?

Yicong Lin*[1,2] and Hanno Reuvers[3]

[1]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam

[2]Tinbergen Institute

[3]Department of Econometrics, Erasmus University Rotterdam

December, 2022

Abstract

The environmental Kuznets curve predicts an inverted U-shaped relationship between air pollution and economic growth. Current analyses frequently employ models that restrict nonlinearities in the data to be explained by economic growth only. We propose a Global Trend Augmented Cointegrating Polynomial Regression (GTACPR) to allow for nonlinearities in time and economic growth. The theoretical properties of the GTACPR are established. Empirically, a single global trend accurately captures all nonlinearities for all the countries studied, leading to a linear relationship between GDP and CO_2. This suggests that the environmental improvement of the last years is due to factors different from GDP.

Keywords: Cointegration Testing, Environmental Kuznets Curve, Cointegrating Polynomial Regression, Power Law Trends.

JEL Classification: C12, C13, C32, Q56

*Corresponding author: Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, the Netherlands. E-mail address: yc.lin@vu.nl.

1 Introduction

On p. 370 of their seminal paper, Grossman and Krueger (1995) conclude:

> "*Contrary to the alarmist cries of some environmental groups, we find no evidence that economic growth does unavoidable harm to the natural habitat. Instead, we find that while increases in GDP may be associated with worsening environmental conditions in very poor countries, air and water quality appear to benefit from economic growth once some critical level of income has been reached.*"

The quote above suggests an inverted U-shaped relationship between environmental degradation and economic growth. This relationship is currently known as the Environmental Kuznets Curve (EKC) and it forms an active research area. Its relevance becomes clear if we look at some forecasts of long-run economic growth. The projected GDP per capita growth of the world is about 2.1% per year for the next decades (chapter 3 in Nordhaus 2013; Gillingham and Nordhaus 2018) and this growth is partially powered by carbon-based energy resources, water usage, and material consumption. In absence of an EKC, economic growth will place more and more stress on the environment. Alternatively, if the EKC exists, then the inverted U-shape eventually implies a turning point after which economic growth and environmental improvement go hand in hand. Due to such considerations, there is now, some 25 years after its first conception, a rich literature that (1) reports on the experimental evidence on the existence/nonexistence of the EKC, (2) provides the economic theory to explain the EKC, and/or (3) refines the econometric tools that are used to analyse the EKC.[1]

Driven by contradictory empirical results as well as the variability in estimated turning points, the EKC has been criticised on two main points. First, the income variable was initially treated as a stationary variable whereas later research shows that the unit root hypothesis often cannot be rejected (see Galeotti et al. 2009, p. 553; Stern 2017, p. 14–15). Nonstationarity has further implications because EKC regressions include higher integer powers of GDP as well. This combination of nonstationarity and nonlinearity places the EKC in the nonlinear cointegration literature and appropriate econometric techniques should be employed. Such techniques have been developed in Wagner (2015) and Wagner and Hong (2016) under the name of *Cointegrating Polynomial Regressions (CPRs)*. CPRs contain deterministic variables, integrated processes, and their integer powers. Multivariate extensions of CPRs, Seemingly Unrelated Cointegrating Polynomial Regressions, are discussed in Wagner et al. (2020) and Lin and Reuvers (2022).

As a second point of critique, there is an ongoing debate on the model specification. Various functional forms can describe the relationship between national income and the pollution variable.

[1] Further references to these specific areas of research can be found in the review articles by Dasgupta et al. (2002), Stern (2004), and Carson (2009) among others.

The quadratic specification is widespread but cubic relationships (Harbaugh et al., 2002; Wagner, 2015) and double-nonlinear transformation (Lin et al., 2020) are also in use. Various specification tests are helpful while deciding on the right parametric specification (Hong and Phillips, 2010; Wang and Phillips, 2012; Wang et al., 2018). Alternatively, one could resort to nonparametric estimation procedures altogether (Wang and Phillips, 2009; Linton and Wang, 2016). Whereas such modelling approaches do allow for a more flexible relationship, they also assume that nonlinear environmental effects are solely attributable to economic growth. Relevant variables are thus potentially missing from the model specification. Such omitted variables are a valid concern because advances in green technology, pollution policy, and environmental awareness, may all influence pollution levels. However, such data is typically available for short time spans only (and for that reason often excluded from the model). Time effects can control for time-variation in unobserved effects (Vollebergh et al., 2009).

Current developments on nonlinear cointegration emphasize the role of the nonstationary regressor yet pay less attention to time effects. Time effects are important. The small simulation exercise in Table 1 illustrates the point. Foreshadowing our proposed model, we consider a multivariate setting with a global nonlinear, smooth time trend. The global trend is omitted by the researcher and a quadratic EKC specification is estimated: $y_{i,t} = \tau_{1,i} + \tau_{2,i}t + \phi_{1,i}x_{i,t} + \phi_{2,i}x_{i,t}^2 + u_{i,t}$ $(i = 1, \ldots, 3)$, where $x_{i,t}$ and $y_{i,t}$ are unit-specific variables measuring income and environmental pollution, respectively. We test $H_0 : \phi_{2,1} = \phi_{2,2} = \phi_{2,3} = 0$ because a significantly negative coefficient in front of $x_{i,t}^2$ is typically interpreted as evidence of an EKC.[2] Panel (A) reveals exacerbated rejection frequency as curvature caused by the global deterministic trend is mistakenly interpreted as curvature caused by the income variable. In other words, negative and significant coefficients in front of squared GDP are possibly caused by omitted nonlinear deterministic trends rather than being indicative of an EKC. To be on the safe side, we recommend researchers include a nonlinear trend component in their model specification. If unnecessary, then this is rather innocuous. Indeed, Panel (B) of Table 1 shows that significant results for nonlinear economic growth effects continue to be found with modest losses in statistical power.

Our contributions are fourfold. First, we propose the *Global Trend Augmented Cointegrating Polynomial Regression (GTACPR)*. This multivariate model features a global power law trend to capture time effects. Power law trends have been employed to model non-constant growth rates in technology indices (Duggal et al., 2007) and production functions (Klein et al., 2004). Within the EKC context, this flexible trend can capture common time effects that are implicit in omitted variables. Alternatively, as in Li and Linton (2020), the reader can view the global flexible

[2]For the moment, we focus on the curvature parameter. Clearly, for an inverted U-shaped relationship the coefficient in front of the linear term should be positive.

Table 1: The rejection rate (in %) when testing $H_0 : \phi_{2,1} = \phi_{2,2} = \phi_{2,3} = 0$. (A) Falsely inflated rejections of $H_0 : \phi_{2,1} = \phi_{2,2} = \phi_{2,3} = 0$ when time effects are omitted. (B) Adding an additional global deterministic trend to the model specification hardly influences the power of the test $H_0 : \phi_{2,1} = \phi_{2,2} = \phi_{2,3} = 0$. That is, significant coefficients in front of $x_{i,t}^2$ remain significant after adding a redundant flexible global trend.

	Panel (A): Omitted Global Trend			Panel (B): Redundant Global Trend	
DGP	$y_{i,t} = \tau_g t^\theta + \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + u_{i,t}$			$y_{i,t} = \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + \phi_2 x_{i,t}^2 + u_{i,t}$	
Model	$y_{i,t} = \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + \phi_{2,i} x_{i,t}^2 + u_{i,t}$			Correct Specification	$y_{i,t} = \tau_g t^\theta + \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + \phi_{2,i} x_{i,t}^2 + u_{i,t}$
$\tau_g\,(\times 10^{-5})$	FM-SOLS	FM-SUR	ϕ_2	SimNLS	SimNLS
0	6.30	6.63	0	6.93	5.70
-0.5	13.27	12.77	-0.5	9.20	6.97
-1	30.07	27.50	-1	14.97	9.97
-1.5	46.23	41.57	-1.5	30.47	20.70
-2	56.60	50.30	-2	55.33	40.53
-2.5	64.50	56.00	-2.5	81.73	69.20
-3	68.60	57.57	-3	93.97	89.50

Note 1: For illustrative purpose, we consider a stylised example in this introduction. The exact parametrisation is available in Section of the Supplementary Material. More elaborate simulation results based on the empirical application are reported as simulation DGP2 in Section 4.
Note 2: FM-SOLS and FM-SUR are documented in Wagner et al. (2020). The results in Panel (B) are based on simulation-based inference, see Section 3.2.

trend as an outside option (next to the income variable) to describe nonlinearities in the data. Limiting distributions for estimators in models with purely deterministic power law trends have been reported in Phillips (2007), Robinson (2012), and Gao et al. (2020). The presence of integrated variables requires an alternative asymptotic framework. Moreover, due to endogeneity, approaches assuming pre-determined integrated regressors (Park and Phillips, 1999, 2001; Chang et al., 2001) are inappropriate and we instead opt for a proof along the lines of Chan and Wang (2015). Our resulting limiting distribution is non-standard because (1) the scaling matrix with convergence rates is non-diagonal and parameter-dependent, and (2) second-order bias terms are present. Second, we propose a simulation-based approach to conduct inference. Monte Carlo simulations show clear benefits of this simulation-based approach in terms of size control compared to existing methods. Third, in the spirit of Choi and Saikkonen (2010), we report a multivariate KPSS-type test to verify the stationarity of the error process thus enabling researchers to avoid spurious results or misspecified cointegrating relations. Fourth and finally, in the empirical application, we investigate the EKC for Austria, Belgium, Finland, the Netherlands, Switzerland, and the UK over the period 1870–2014. Nonparametric estimates and tests confirm that the global trend captures all nonlinearities in the data. Nonlinear effects in log per capita GDP (and thus also evidence for an EKC) are absent. We recommend researchers check whether their EKC conclusions are robust to the inclusion of power law trends.

Finally, some words on notation. The integer part of the number $a \in \mathbb{R}^+$ is denoted by $[a]$. For a vector $\boldsymbol{x} \in \mathbb{R}^n$, its p-norm is denoted by $\|\boldsymbol{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. For a matrix $\boldsymbol{A}$, say of dimension $(n \times m)$, the induced p-norm is defined as $\|\boldsymbol{A}\|_p = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}\|_p / \|\boldsymbol{x}\|_p$. We will omit the subscripts whenever $p = 2$. The $(n \times n)$ identity matrix is denoted $\boldsymbol{I}_n$ and $\boldsymbol{\imath}_n$ signifies an n-dimensional

column vector with all entries equal to 1. The block-diagonal matrix $\text{diag}[\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n]$ stacks the matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ along its diagonal. We omit the integration bounds whenever the integration interval is $[0, 1]$. The symbol "$\overset{d}{=}$" stands for equality in distribution, and "$\longrightarrow_p$" and "$\longrightarrow_d$" denote convergence in probability and in distribution. If convergence occurs conditionally on the sample, then we add a superscript "*" to the standard notation. Finally, the generic constant C can change from line to line.

2 The model and NLS estimation

Our model enriches the Seemingly Unrelated Cointegrating Polynomial Regressions (SUCPRs) from Wagner et al. (2020) with a flexible deterministic trend. That is, each individual series in the system is affected by specific deterministic variables and integrated regressors (and their integer powers) while a *global* flexible trend describes nonlinear behaviour that is prevalent across all series. The resulting *Global Trend Augmented Cointegrating Polynomial Regression (GTACPR)* is given by:

$$y_{i,t} = \tau_g t^\theta + \tau_{1,i} + \tau_{2,i} t + \sum_{j=1}^{p_i} \phi_{j,i} x_{i,t}^j + u_{i,t}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2.1}$$

where $\theta \in \Theta(\varepsilon)$ with $\Theta(\varepsilon) = \{\theta \in [\theta_L, \theta_U] : \ |\theta| > \varepsilon, |\theta - 1| > \varepsilon\}$ and $-1 < \theta_L \leq \theta_U < \infty$.[3] Model (2.1) has two important features: (1) the heterogeneity in the cross-sectional dimension reflects the observed differences in EKC curves (see, e.g., Mazzanti and Musolesi 2013), and (2) a global trend to separate the nonlinearities in economic growth from the nonlinearities in time. Regarding the latter, it is the common parameter τ_g that drives this identification.[4] Our choice for a power law global trend is motivated by being flexible enough to have nonlinear behavior in the deterministic component yet avoiding a too flexible specification in which the deterministic component can proxy stochastic trends (Phillips, 1998). Alternatively, we write $y_{i,t} = \tau_g t^\theta + \boldsymbol{z}_{i,t}' \boldsymbol{\beta}_i + u_{i,t}$, where $\boldsymbol{z}_{i,t} = \left[1, t, x_{i,t}, \ldots, x_{i,t}^{p_i}\right]'$ and $\boldsymbol{\beta}_i = \left[\tau_{1,i}, \tau_{2,i}, \phi_{1,i}, \ldots, \phi_{p_i,i}\right]'$. Finally, we stack all N equations in (2.1) in matrix form to retrieve

$$\boldsymbol{y}_t = \tau_g t^\theta \boldsymbol{\iota}_N + \boldsymbol{Z}_t' \boldsymbol{\beta} + \boldsymbol{u}_t, \qquad t = 1, \ldots, T, \tag{2.2}$$

with $\boldsymbol{y}_t = \left[y_{1,t}, \ldots, y_{N,t}\right]'$, $\boldsymbol{Z}_t = \text{diag}\left[\boldsymbol{z}_{1,t}, \ldots, \boldsymbol{z}_{N,t}\right]$, and the vector $\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_N'\right]'$ of length $p = 2N + \sum_{i=1}^N p_i$ containing all local parameters.

We consider nonlinear least squares (NLS) estimators of the unknown parameters in (2.1). Define

[3] The parameter $\varepsilon > 0$ ensures that the global trend remains distinguishable from the unit-specific intercepts and linear trends. We use $\varepsilon = 0.05$. Changing this value has little influence on the outcomes because the estimated θ is generally different from both 0 and 1.

[4] The importance of distinguishing income- and time-effects is discussed in detail in Vollebergh et al. (2009).

the objective function $Q_T(\theta, \tau_g, \boldsymbol{\beta}) = \frac{1}{2}\sum_{t=1}^{T} \left\| \boldsymbol{y}_t - \tau_g t^{\theta} \boldsymbol{\imath}_N - \boldsymbol{Z}_t' \boldsymbol{\beta} \right\|^2$ and compute

$$\left(\widehat{\theta}_T, \widehat{\tau}_{g,T}, \widehat{\boldsymbol{\beta}}_T \right) = \underset{(\theta, \tau_g, \boldsymbol{\beta}) \in \Theta(\varepsilon) \times \mathbb{R} \times \mathbb{R}^p}{\arg\min} Q_T(\theta, \tau, \boldsymbol{\beta}). \tag{2.3}$$

The optimization problem in (2.3) is easy to solve. Given θ, model (2.2) is linear-in-parameters. The minimizer for $(\tau_g, \boldsymbol{\beta})$ has a closed-form expression

$$\begin{bmatrix} \tau_g(\theta) \\ \boldsymbol{\beta}(\theta) \end{bmatrix} = \left(\sum_{t=1}^{T} \boldsymbol{Z}_t(\theta) \boldsymbol{Z}_t'(\theta) \right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{Z}_t(\theta) \boldsymbol{y}_t \right),$$

where $\boldsymbol{Z}_t'(\theta) = \left[t^{\theta} \boldsymbol{\imath}_N \ \boldsymbol{Z}_t' \right]$. We subsequently minimize the concentrated criterion function $\widetilde{Q}_T(\theta) = Q_T\big(\theta, \tau_g(\theta), \boldsymbol{\beta}(\theta)\big)$ to obtain $\widehat{\theta}_T$. At last, we plug in $\widehat{\theta}_T$ and recover $\widehat{\tau}_{g,T}$ and $\widehat{\boldsymbol{\beta}}_T$ through a final OLS estimation.

Remark 1

Keeping the powers of $x_{i,t}$ fixed allows us to test for their significance and thereby distinguish between nonlinearities caused by deterministic and stochastic trends. This is important for our empirical application on the EKC, see Section 5. Hu et al. (2021) study a model with a flexible power of the integrated regressor. That is, these authors derive the limiting distribution of the NLS estimators for β and γ when $y_t = \beta |x_t|^{\gamma} + u_t$ with $\beta \neq 0$.

Remark 2

The GTACPR of (2.1) *can be extended in several directions. First, integer powers of deterministic trends can be added as long as $\Theta(\varepsilon)$ is adjusted accordingly (to avoid collinearity). Second, multiple explanatory variables can be included. Examples within the EKC literature are: trade openness (Jalil and Feridun, 2011), energy prices (Al-Mulali and Ozturk, 2016), and educational level (Maranzano et al., 2021). For nonstationary variables, conditions similar to those on $\{x_{i,t}\}$ should be fulfilled (Assumption 2 below). Stationary variables should be strictly exogenous. To avoid elaborate notation, we focus on the baseline specification in* (2.1).

3 Asymptotic theory

We subsequently study the asymptotic properties of the NLS estimators. We first collect all the unknown parameters in $\boldsymbol{\gamma} = \left[\theta, \tau_g, \boldsymbol{\beta}' \right]'$. This vector is assumed to be an element of the parameter space $\boldsymbol{\Gamma} = \Theta(\varepsilon) \times \mathbb{R}^{1+p}$. The true parameter vector is $\boldsymbol{\gamma}_0 = \left[\theta_0, \tau_{g,0}, \boldsymbol{\beta}_0' \right]'$.

Assumption 1

The global trend is relevant, i.e., $\tau_{g,0} \neq 0$.

Assumption 2

Let $\boldsymbol{\zeta}_t = [\eta_t', \boldsymbol{\varepsilon}_t']'$ *be a sequence of i.i.d. random vectors with* $\mathbb{E}(\boldsymbol{\zeta}_t) = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbb{E}\big(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t'\big) \succ 0$, *and* $\mathbb{E}\left\|\boldsymbol{\zeta}_t\right\|^{2q} < \infty$ *for some* $q > 2$.

(a) $u_t = \sum_{k=0}^{\infty} \psi_k \eta_{t-k}$ *with* $\sum_{k=1}^{\infty} k|\psi_k| < \infty$.

(b) $\boldsymbol{x}_t = \sum_{s=1}^{t} \boldsymbol{v}_s$, *where* $\boldsymbol{v}_t = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\varepsilon}_{t-k}$, $\sum_{k=0}^{\infty} \|\boldsymbol{\Psi}_k\| < \infty$, *and* $\det\left(\sum_{k=0}^{\infty} \boldsymbol{\Psi}_k\right) \neq 0$.

The first assumption is needed to avoid identification issues. That is, if $\tau_{g,0} = 0$, then θ is not identified and the Davies problem arrises when testing $H_0 : \tau_i = 0$ (see Davies (1977, 1987)). Such complications are not investigated here and this is further reflected in our model specification (2.1). That is, we consider *flexible* powers of the deterministic trends but *fixed* powers of the stochastic trends, hence allowing us to test zero restrictions on (elements of) $\boldsymbol{\beta}$. This is of crucial importance in the EKC application while determining whether nonlinear effects in the economic growth variables $(x_{i,t})$ remain significant after nonlinear time trends have been added to the model. Assumption 1 has been relaxed in the literature albeit for different models. Baek et al. (2015) and Cho and Phillips (2018) study the asymptotic behaviour of a quasi-likelihood ratio test when Assumption 1 is violated and the conditional mean of the data contains strictly stationary regressors and a flexible time trend. Whereas we do not study the consequences of violations of Assumption 1 theoretically, we do provide some empirical robustness checks. First, Monte Carlo simulations indicate that adding a redundant global trend has little impact. Second, we propose a heuristic verification of Assumption 1 in our empirical application (see p. 21, Section 5).

Assumption 2 excludes cointegration among elements of $\boldsymbol{x}_t$ and defines this vector as the partial sum of a short memory process. This implies $T^{-1/2}\sum_{t=1}^{[rT]}\left[\begin{smallmatrix}\boldsymbol{u}_t\\ \boldsymbol{v}_t\end{smallmatrix}\right] \longrightarrow_d \boldsymbol{B}(r) = \left[\begin{smallmatrix}\boldsymbol{B}_u(r)\\ \boldsymbol{B}_v(r)\end{smallmatrix}\right]$ where $\boldsymbol{B}(r)$ is an $2N$-dimensional vector Brownian motion with covariance $\boldsymbol{\Omega} = \left[\begin{smallmatrix}\boldsymbol{\Omega}_{uu} & \boldsymbol{\Omega}_{uv}\\ \boldsymbol{\Omega}_{vu} & \boldsymbol{\Omega}_{vv}\end{smallmatrix}\right]$. The one-sided covariance matrix $\boldsymbol{\Delta} = \left[\begin{smallmatrix}\boldsymbol{\Delta}_{uu} & \boldsymbol{\Delta}_{uv}\\ \boldsymbol{\Delta}_{vu} & \boldsymbol{\Delta}_{vv}\end{smallmatrix}\right] = \sum_{h=0}^{\infty}\mathbb{E}\left(\left[\begin{smallmatrix}\boldsymbol{u}_t\boldsymbol{u}_{t+h} & \boldsymbol{u}_t\boldsymbol{v}_{t+h}'\\ \boldsymbol{v}_t\boldsymbol{u}_{t+h} & \boldsymbol{v}_t\boldsymbol{v}_{t+h}'\end{smallmatrix}\right]\right)$ is partitioned similarly. Subscripts refer to specific elements. For example, $\boldsymbol{B}_{v_i}$ and $\boldsymbol{\Delta}_{v_iu_j}$ denote the i^{th} and $(i,j)^{\text{th}}$ elements of $\boldsymbol{B}_v$ and $\boldsymbol{\Delta}_{vu}$, respectively.

Additional notation is:

(1) $\boldsymbol{D}_{(i),T} = \operatorname{diag}\left[1, T, T^{1/2}, T, \ldots, T^{p_i/2}\right]$ to scale the deterministic and stochastic trends within each equation. Define $\boldsymbol{D}_{Z,T} = \operatorname{diag}\left[\boldsymbol{D}_{(1),T}, \ldots, \boldsymbol{D}_{(N),T}\right]$, $\boldsymbol{D}_{\theta_0,T} = \sqrt{T}\left[\begin{smallmatrix}T^{\theta_0} & & \\ & T^{\theta_0} & \\ & & \boldsymbol{D}_{Z,T}\end{smallmatrix}\right]$ and $\boldsymbol{L}_{\tau_{g,0},T} = \left[\begin{smallmatrix}1 & -\tau_{g,0}\ln T & \\ 0 & 1 & \\ & & \boldsymbol{I}_p\end{smallmatrix}\right]$ for the full system of equation. Finally, set $\boldsymbol{G}_{\boldsymbol{\gamma}_0,T} = \boldsymbol{D}_{\theta_0,T}\boldsymbol{L}_{\tau_{g,0},T}^{\prime -1}$.

(2) Define $\boldsymbol{j}_i(r) = \left[1, r, B_{v_i}(r), B^2_{v_i}(r), \ldots, B^{p_i}_{v_i}(r)\right]'$, $\boldsymbol{J}_Z(r) = \text{diag}\left[\boldsymbol{j}_1(r), \ldots, \boldsymbol{j}_N(r)\right]$. Moreover, let $\boldsymbol{J}(r;\boldsymbol{\gamma}_0) = \left[\tau_{g,0} r^{\theta_0} \ln r\, \boldsymbol{\imath}_N, r^{\theta_0}\boldsymbol{\imath}_N, \boldsymbol{J}'_Z(r)\right]'$.

(3) Let $\boldsymbol{b}_i = \left[\boldsymbol{0}_{1\times 2}, 1, 2\int \boldsymbol{B}_{v_i}(r)dr, \ldots, p_i \int \boldsymbol{B}^{p_i-1}_{v_i}(r)dr\right]'$ and $\boldsymbol{\mathcal{B}}_{vu} = \left[\boldsymbol{0}_{1\times 2}, \boldsymbol{b}'_1\boldsymbol{\Delta}_{v_1u_1}, \ldots, \boldsymbol{b}'_N\boldsymbol{\Delta}_{v_Nu_N}\right]'$ for the second-order bias terms.

Theorem 1

Under Assumptions 1-2, as $T \to \infty$ (for a fixed N), we have

$$\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}\left(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0\right) \longrightarrow_d \left(\int \boldsymbol{J}(r;\boldsymbol{\gamma}_0)\boldsymbol{J}'(r;\boldsymbol{\gamma}_0)\, dr\right)^{-1} \left(\int \boldsymbol{J}(r;\boldsymbol{\gamma}_0)\, d\boldsymbol{B}_u(r) + \boldsymbol{\mathcal{B}}_{vu}\right) =: \boldsymbol{\mathcal{J}}(\boldsymbol{\gamma}_0).$$

The proof of Theorem 1 is closely related to the work by Chan and Wang (2015). These authors provide the asymptotic distribution of NLS estimators under a set of general conditions in univariate, nonstationary time series models (see their theorem 3.1). The results in Chan and Wang (2015) and Wang et al. (2018) suggest that Assumption 2 can be replaced by a long memory specification for $\Delta\boldsymbol{x}_t$. However, long memory parameters enter the limiting distribution and inference will be complicated further. We illustrate Theorem 1 with two examples.

Example 1

We consider $y_t = \tau t^\theta + u_t$ with innovations satisfying Assumption 2. The limiting distribution of the parameter estimators depends solely on the mean square Riemann-Stieltjes integrals $\int \tau_0 r^{\theta_0}\ln(r)dB_u$ and $\int r^{\theta_0}dB_u$, and is therefore normally distributed (e.g., section 2.3 in Tanaka 2017). We have

$$\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ T^{\theta_0+\frac{1}{2}}\tau_0\ln(T) & T^{\theta_0+\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \widehat{\theta}_T - \theta_0 \\ \widehat{\tau}_T - \tau_0 \end{bmatrix} \longrightarrow_d \text{N}\left(\boldsymbol{0}, \Omega_{uu}(2\theta_0+1)^3 \begin{bmatrix} 2\tau_0^2 & -\tau_0(2\theta_0+1) \\ -\tau_0(2\theta_0+1) & (2\theta_0+1)^2 \end{bmatrix}^{-1}\right).$$

The scaling matrix in the LHS depends on θ_0 and is non-diagonal. The dependence on θ_0 is unavoidable but asymptotic results for the case of a diagonal scaling matrix are obtainable. Noting that $\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ 0 & T^{\theta_0+\frac{1}{2}}/\ln(T) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\tau_0 & 1/\ln(T) \end{bmatrix} \begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ T^{\theta_0+\frac{1}{2}}\tau_0\ln(T) & T^{\theta_0+\frac{1}{2}} \end{bmatrix}$ *and* $\begin{bmatrix} 1 & 0 \\ -\tau_0 & 1/\ln(T) \end{bmatrix} \overset{T\to\infty}{\to} \begin{bmatrix} 1 & 0 \\ -\tau_0 & 0 \end{bmatrix}$*, we have*

$$\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & 0 \\ 0 & T^{\theta_0+\frac{1}{2}}/\ln(T) \end{bmatrix} \begin{bmatrix} \widehat{\theta}_T - \theta_0 \\ \widehat{\tau}_T - \tau_0 \end{bmatrix} \longrightarrow_d \begin{bmatrix} 1/\tau_0 \\ -1 \end{bmatrix} \text{N}\left(\boldsymbol{0}, \Omega_{uu}(2\theta_0+1)^3\right).$$

This limiting distribution coincides with the result in theorem 6.3 of Phillips (2007).

Example 2

If $y_t = \tau t^\theta + \phi x_t + u_t$, *then* $\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & \\ T^{\theta_0+\frac{1}{2}}\tau_0 \ln(T) & T^{\theta_0+\frac{1}{2}} \end{bmatrix}_T \begin{bmatrix} \widehat{\theta}_T-\theta_0 \\ \widehat{\tau}_T-\tau_0 \\ \widehat{\phi}_T-\phi_0 \end{bmatrix}$ *converges to*

$$\begin{bmatrix} \int \big(\tau_0 r^{\theta_0}\ln(r)\big)^2 dr & \int \tau_0 r^{2\theta_0}\ln(r)dr & \int \tau_0 r^{\theta_0}\ln(r)B_v dr \\ \int \tau_0 r^{2\theta_0}\ln(r)dr & \int r^{2\theta_0}dr & \int r^{\theta_0}B_v dr \\ \int \tau_0 r^{\theta_0}\ln(r)B_v dr & \int r^{\theta_0}B_v dr & \int B_v^2 dr \end{bmatrix}^{-1} \left(\begin{bmatrix} \int \tau_0 r^{\theta_0}\ln(r)dB_u \\ \int r^{\theta_0}dB_u \\ \int B_v dB_u \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \Delta_{vu} \end{bmatrix} \right).$$

This distribution exhibits second-order bias when $\Delta_{vu} \neq 0$, *or when* B_u *and* B_v *are correlated.*

Two features of the limiting distribution of $\boldsymbol{G}_{\gamma_0,T}\big(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0\big)$ deserve further comments. First, as emphasised in Examples 1–2, the scaling matrix $\boldsymbol{G}_{\gamma_0,T}$ features two uncommon properties: (1) this matrix depends on the true parameters $\boldsymbol{\tau}_{g,0}$ and θ_0, and (2) $\boldsymbol{G}_{\gamma_0,T}$ is not diagonal. These peculiarities are caused by the nonlinearity and nonstationarity of the model. More specifically, these features can be traced back to the presence of the global trend. Limiting distributions with a similar mathematical structure can be found in the structural breaks literature, cf. model setting II.b of Perron and Zhu (2005) and its detailed analysis in Beutner et al. (2022).

Second, the nonstationary regressor $x_{i,t}$ enters the model (2.1) through a polynomial transformation of the form $g(x_{i,t}, \boldsymbol{\phi}_i) = \phi_{i,1}x_{i,t} + \ldots + \phi_{i,p_i}x_{i,t}^{p_i}$ $(i = 1, 2, \ldots, N)$. In the terminology of Park and Phillips (2001), this part of the regression function is a linear combination of H_0-regular functions. It is well-documented in the literature, e.g., Chang et al. (2001) and Chan and Wang (2015), that this leads to second-order bias terms and hence nonstandard inference (except for the special case of strictly exogenous nonstationary regressors).

3.1 Consistent long-run covariance matrix estimation

Correcting for second-order bias terms typically involves estimating long-run variance (LRV) matrices. This subsection establishes that the NLS residuals can be used to construct consistent kernel estimators for the LRV matrices $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}$. Defining $\boldsymbol{V}_t(\boldsymbol{\gamma}) = \big[\boldsymbol{u}_t(\boldsymbol{\gamma})', \Delta\boldsymbol{x}_t'\big]'$ with $\boldsymbol{u}_t(\boldsymbol{\gamma}) = \boldsymbol{y}_t - \tau_g t^\theta \boldsymbol{\iota}_N - \boldsymbol{Z}_t'\boldsymbol{\beta}$, these LRV estimators are defined as

$$\widehat{\boldsymbol{\Delta}}_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t} k\left(\frac{|t-s|}{b_T}\right)\boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)\boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)', \qquad \widehat{\boldsymbol{\Omega}}_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} k\left(\frac{|t-s|}{b_T}\right)\boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)\boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)', \tag{3.1}$$

for some kernel $k(\cdot)$ and bandwidth b_T. The first N elements of $\boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)$ are the elements of the residual vector $\widehat{\boldsymbol{u}}_t = \boldsymbol{y}_t - \widehat{\tau}_{g,T}\, t^{\widehat{\theta}_T}\boldsymbol{\iota}_N - \boldsymbol{Z}_t'\widehat{\boldsymbol{\beta}}_T$. The remaining elements are $\Delta\boldsymbol{x}_t = \boldsymbol{v}_t$.

Assumption 3

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

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

(c) The bandwidth satisfies $\{b_T\} \subseteq (0, \infty)$ *and* $\lim_{T\to\infty} \left(b_T^{-1} + T^{-1/2} b_T \ln T\right) = 0$.

The conditions on the kernel function $k(\cdot)$, Assumptions 3(a)–(b), are identical to those in Jansson (2002). Jansson (2002) remarks that these assumptions "*would appear to be satisfied by any kernel in actual use*". Commonly used kernels such as the Bartlett, Parzen, and Quadratic Spectral kernels indeed satisfy all these assumptions. Assumption 3(c) differs from the usual requirement, $\lim_{T\to\infty} \left(b_T^{-1} + T^{-1/2} b_T\right) = 0$, by a factor $\ln T$. The difference is caused by the estimation error in $\widehat{\theta}_T$. This error causes the residuals $\{\widehat{\boldsymbol{u}}_t\}$ to be less close to the innovations $\{\boldsymbol{u}_t\}$ and we balance this by including autocovariance matrices of higher lags at a slower pace.

Theorem 2

Under Assumptions 1-3, we have $\widehat{\boldsymbol{\Delta}}_T \longrightarrow_p \boldsymbol{\Delta}$ *and* $\widehat{\boldsymbol{\Omega}}_T \longrightarrow_p \boldsymbol{\Omega}$ *as* $T \to \infty$.

3.2 Simulation-based inference

The limiting distribution in Theorem 1 is nonpivotal and thus not directly suited for inference. Some popular solutions for linear-in-parameters cointegration models are: Saikkonen's (1992) dynamic least squares, the integrated modified OLS and fixed-b approaches by Vogelsang and Wagner (2014), and the fully modified approach, e.g., Phillips and Hansen (1990). For a nonlinear-in-parameter model as in (2.1), a preliminary Monte Carlo (MC) exercise[5] shows poor performance for fully modified inference but promising results for a simulation-based approach. We pursue the latter method for the remainder of this paper.

The main idea behind the simulation-based approach is to replace nuisance parameters with consistent estimates and to rely on MC simulations to approximate the limiting distribution. The empirical quantiles of these MC draws allow us to conduct inference. Wang et al. (2018) shows that the simulation approach is asymptotically justified in several model specifications. We adapt their algorithm to the current setting and prove its asymptotic validity.

Algorithm 1 (simulation-based inference)

STEP*1: Estimate* $\widehat{\boldsymbol{\gamma}}_T$ *and use the residuals* $\{\widehat{\boldsymbol{u}}_t\}$ *to compute the estimators* $\widehat{\boldsymbol{\Delta}}_T$ *and* $\widehat{\boldsymbol{\Omega}}_T$ *from* (3.1).

STEP*2: Repeat for* $j = 1, \ldots, J$,

[5]The details are available in Section S8 of the Supplementary Material. The analytical results in that appendix also suggest that the convergence speed of $\widehat{\theta}_T$ to θ_0 is too slow to recover the standard zero-mean Gaussian limiting distribution.

(a) Draw random variables $\{\boldsymbol{e}_t\}_{t=1}^{M_T}$ *i.i.d. from* $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_{2N})$.

(b) Compute $\begin{bmatrix} \widehat{\boldsymbol{\mu}}_t \\ \widehat{\boldsymbol{v}}_t \end{bmatrix} = \widehat{\boldsymbol{\Omega}}_T^{1/2} \boldsymbol{e}_t$ *and the partial sum* $\widehat{\boldsymbol{\chi}}_t = \left[\widehat{\chi}_{1,t}, \ldots, \widehat{\chi}_{N,t}\right]' = \sum_{s=1}^{t} \widehat{\boldsymbol{v}}_s$.

(c) Let $\widehat{\boldsymbol{J}}\big(t; \widehat{\boldsymbol{\gamma}}_T\big) = \left[\widehat{\tau}_{g,T}\, t^{\widehat{\theta}_T} \ln t\, \boldsymbol{\imath}_N, t^{\widehat{\theta}_T} \boldsymbol{\imath}_N, \widehat{\mathcal{Z}}_t'\right]'$, *where* $\widehat{\mathcal{Z}}_t = \operatorname{diag}\left[\widehat{\boldsymbol{z}}_{1,t}, \ldots, \widehat{\boldsymbol{z}}_{N,t}\right]$, $\widehat{\boldsymbol{z}}_{i,t} = \left[1, t, \widehat{\chi}_{i,t}, \ldots, \widehat{\chi}_{i,t}^{p_i}\right]'$, $i = 1, \ldots, N$. *For given* M_T, *construct the* j^{th} *simulated draw as*

$$\widehat{\boldsymbol{\mathcal{J}}}^{(j)}\left(\widehat{\boldsymbol{\gamma}}_T, \widehat{\boldsymbol{\Omega}}_T, \widehat{\boldsymbol{\Delta}}_{vu}^{-}\right) = \left\{\boldsymbol{G}_{\widehat{\gamma}_T, M_T}^{\prime -1}\left[\sum_{t=1}^{M_T} \widehat{\boldsymbol{J}}\big(t; \widehat{\boldsymbol{\gamma}}_T\big) \widehat{\boldsymbol{J}}\big(t; \widehat{\boldsymbol{\gamma}}_T\big)'\right] \boldsymbol{G}_{\widehat{\gamma}_T, M_T}^{-1}\right\}^{-1} \times \left\{\boldsymbol{G}_{\widehat{\gamma}_T, M_T}^{\prime -1}\left[\sum_{t=1}^{M_T} \widehat{\boldsymbol{J}}\big(t; \widehat{\boldsymbol{\gamma}}_T\big)\, \widehat{\boldsymbol{\mu}}_t\right] + \widehat{\boldsymbol{\mathcal{B}}}_{vu}^{-}\right\},$$

where $\widehat{\boldsymbol{\Delta}}_{vu}^{-}$ *is a consistent estimator of the lower-left subblock of* $\boldsymbol{\Delta}^{-} = \begin{bmatrix} \boldsymbol{\Delta}_{uu}^{-} & \boldsymbol{\Delta}_{uv}^{-} \\ \boldsymbol{\Delta}_{vu}^{-} & \boldsymbol{\Delta}_{vv}^{-} \end{bmatrix} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}'$, $\widehat{\boldsymbol{\mathcal{B}}}_{vu}^{-} = \left[\mathbf{0}_{1\times 2}, \widehat{\boldsymbol{b}}_1' \widehat{\boldsymbol{\Delta}}_{v_1 u_1}^{-}, \ldots, \widehat{\boldsymbol{b}}_N' \widehat{\boldsymbol{\Delta}}_{v_N u_N}^{-}\right]'$, $\widehat{\boldsymbol{b}}_i = \left[\mathbf{0}_{1\times 2}, 1, 2\frac{1}{M_T}\sum_{t=1}^{M_T}\left(\frac{\widehat{\chi}_{i,t}}{\sqrt{M_T}}\right), \ldots, p_i \frac{1}{M_T}\sum_{t=1}^{M_T}\left(\frac{\widehat{\chi}_{i,t}}{\sqrt{M_T}}\right)^{p_i - 1}\right]'$ *is an approximation of bias terms.*

STEP*3: Use the empirical quantiles of elements of* $\left\{\widehat{\boldsymbol{\mathcal{J}}}^{(1)}, \ldots, \widehat{\boldsymbol{\mathcal{J}}}^{(J)}\right\}$ *for inference.*

Algorithm 1 uses a discretisation in M_T steps to approximate the limiting distribution of the parameters. In practice, and in accordance with Theorem 3, we can take $M_T = T$. Remark 3 details how simulation-based inference can be used to conduct hypothesis testing. Discussions on size and power are also presented there.

Theorem 3

Suppose Assumptions 1-3 hold. Let $\{M_T\} \subseteq (0, \infty)$ *with* $\lim_{T\to\infty} M_T/T \leq \kappa$, $\kappa < \infty$, *then*

$$\left\{\boldsymbol{G}_{\widehat{\gamma}_T, T}^{\prime -1}\left[\sum_{t=1}^{M_T} \widehat{\boldsymbol{J}}\big(t; \widehat{\boldsymbol{\gamma}}_T\big) \widehat{\boldsymbol{J}}\big(t; \widehat{\boldsymbol{\gamma}}_T\big)'\right] \boldsymbol{G}_{\widehat{\gamma}_T, T}^{-1}\right\}^{-1} \left\{\boldsymbol{G}_{\widehat{\gamma}_T, T}^{\prime -1}\left[\sum_{t=1}^{M_T} \widehat{\boldsymbol{J}}\big(t; \widehat{\boldsymbol{\gamma}}_T\big)\, \widehat{\boldsymbol{\mu}}_t\right] + \widehat{\boldsymbol{\mathcal{B}}}_{vu}^{-}\right\} \longrightarrow_{d^*} \left(\int \boldsymbol{J}(r; \boldsymbol{\gamma}_0) \boldsymbol{J}(r; \boldsymbol{\gamma}_0)'\, dr\right)^{-1} \left(\int \boldsymbol{J}(r; \boldsymbol{\gamma}_0)\, d\boldsymbol{B}_u(r) + \boldsymbol{\mathcal{B}}_{vu}\right), \quad (3.2)$$

in probability, as $T \to \infty$ *(for a fixed* N*).*

Theorem 3 establishes the asymptotic validity of the simulation approach. That is, for a large enough J, the empirical quantiles of the simulated distribution will coincide with the asymptotic distribution. Two remarks are important. First, even though the simulation algorithm is adapted from Wang et al. (2018), the proof of Theorem 3 is not. In particular, the method of proof is similar to Theorem 1 and continues to allow for the endogeneity of the regressors. Second, the simulation approach mimics the stochastic integrals in the limiting distribution directly. It, therefore, suffices

to draw normally distributed random variables in Step 2(a) and use consistent long-run covariance estimates to replicate the covariance structure of the underlying Brownian motions. Compared to a bootstrap procedure, this simulation approach has the advantage of avoiding tedious NLS re-estimation on bootstrap samples but forsakes possible asymptotic refinements.

Remark 3

Step 3 in Algorithm 1 has been kept general for notational convenience. An illustrative example is as follows. Assume we are interested in $H_0 : \phi_{2,1} = 0$ *(irrelevance of the regressor* $x_{1,t}^2$*) when*

$$y_{i,t} = \tau_g t^\theta + \tau_{1,i} + \tau_{2t,i} t + \phi_{1,i} x_{i,t} + \phi_{2,i} x_{i,t}^2 + u_{i,t}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

Under H_0*, we have* $T^{3/2}\widehat{\phi}_{2,1} = \boldsymbol{e}_6' \boldsymbol{G}_{\boldsymbol{\gamma}_0,T}\big(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0\big) \longrightarrow_d \boldsymbol{e}_6' \boldsymbol{\mathcal{J}}(\boldsymbol{\gamma}_0)$ *with* $\boldsymbol{e}_k$ *being the* k^{th} *basis vector in* $\mathbb{R}^{2+p}$*. Denoting the empirical* ζ*-quantiles of* $\big\{\boldsymbol{e}_6'\widehat{\boldsymbol{\mathcal{J}}}^{(1)}, \ldots, \boldsymbol{e}_6'\widehat{\boldsymbol{\mathcal{J}}}^{(J)}\big\}$ *by* c_ζ*, a test of size* α *will reject for* $T^{3/2}\widehat{\phi}_{2,1} < c_{\alpha/2}$ *or* $T^{3/2}\widehat{\phi}_{2,1} > c_{1-\alpha/2}$*. Under the alternative* $\phi_{2,1} \neq 0$*, we rewrite the test statistic as* $T^{3/2}\widehat{\phi}_{2,1} = T^{3/2}\big(\widehat{\phi}_{2,1} - \phi_{2,1}\big) + T^{3/2}\phi_{2,1}$*. Statistical power is guaranteed because the simulation approach mimics the asymptotic distribution and is thus bounded, whereas the second term diverges.*

3.3 Test for the null of cointegration

The correct specification of the nonlinear cointegrating relation will result in a stationary error process $\{u_t\}_{t\in\mathbb{Z}}$. We consider a KPSS-type test statistic for the null of stationarity. The candidate statistic is $\widetilde{K}_T^+ = T^{-2}\sum_{t=1}^T \left\| \widehat{\boldsymbol{\Omega}}_{u.v}^{-1/2} \sum_{i=\ell}^t \widehat{\boldsymbol{u}}_i^+ \right\|^2$, where $\widehat{\boldsymbol{u}}_t^+ = \boldsymbol{y}_t - \widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\Delta\boldsymbol{x}_t - \widehat{\tau}_{g,T}\, t^{\widehat{\theta}_T}\boldsymbol{\imath}_N - \boldsymbol{Z}_t'\widehat{\boldsymbol{\beta}}_T$ and $\widehat{\boldsymbol{\Omega}}_{u.v}$ is a consistent estimator of $\boldsymbol{\Omega}_{u.v} = \boldsymbol{\Omega}_{uu} - \boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{\Omega}_{vu}$. This statistic is stochastically bounded under the null hypothesis but diverges under the alternative. Rejections of the null hypothesis are an indication of a spurious relationship and/or an incorrect functional form of the nonlinear cointegrating relationship. For some settings the asymptotic null distribution of K_T^+ is known, e.g., Kwiatkowski et al. (1992) and Wagner and Hong (2016).

The estimation of $\boldsymbol{\theta}$ contaminates the limiting distribution of $\widetilde{K}_T^+$ with nuisance parameters.[6] Choi and Saikkonen (2010), Wagner and Hong (2016), and Lin and Reuvers (2022), have shown that subsampling can resolve this issue. We will follow their approach and use subsamples of size q_T to compute the test statistics.

Theorem 4

Under Assumptions 1-3 and if $\lim_{T\to\infty}\left(q_T^{-1} + (\ln T)\left(\frac{q_T}{T}\right)^{\theta_L+\frac{1}{2}}\right) = 0$*, then for any* $\ell \in \{1, \ldots, T -$

[6]Proposition 5 in Wagner and Hong (2016) shows that the limiting distribution of $\widetilde{K}_T^+$ is free of nuisance parameters if $\boldsymbol{\theta}_0$ is known and only a single integrated regressor occurs with integer powers greater than one. This result does not carry over to the current setting because of the estimation error in $\widehat{\boldsymbol{\theta}}_T$.

$q_T + 1\}$, *we have*

$$K^+_{q_T,\ell} = \frac{1}{q_T} \sum_{t=\ell}^{\ell+q_T-1} \left\| \frac{1}{\sqrt{q_T}} \widehat{\boldsymbol{\Omega}}^{-1/2}_{u.v} \sum_{i=\ell}^{t} \widehat{\boldsymbol{u}}^+_i \right\|^2 \longrightarrow_d \int \|\boldsymbol{W}(r)\|^2 \, dr, \tag{3.3}$$

as $T \to \infty$ *(for a fixed* N*), where* $\boldsymbol{W}(\cdot)$ *denotes an* N*-dimensional standard Brownian motion.*

Theorem 4 does not provide any guidance on the choices for the starting value ℓ and the subsample size q_T. First, for a given q_T, Choi and Saikkonen (2010) argue that the use of a single subsample (instead of all T observations) implies a significant loss of power. We follow their example and combine all $M = [T/q_T]$ subresidual series of length q_T using a Bonferroni procedure. That is, we create subresiduals series by selecting adjacent blocks of q_T residuals while alternating between the start and end of the sample. We calculate the KPSS-type test statistic for each subseries, say $K_1, \ldots, K_M$, and reject the null of stationarity at significance α whenever $\max\{K_1, \ldots, K_M\}$ exceeds $c_{\alpha/M}$ which is defined by $\mathbb{P}\left(\int \|\boldsymbol{W}(r)\|^2 dr \geq c_{\alpha/M}\right) = \alpha/M$. Finally, we select the block size q_T using Romano and Wolf's (2001) minimum volatility rule. The approach is now completely data-driven.

4 Simulations

This section lists various Monte Carlo simulations showing that the asymptotic approximations from Section 3 provide useful guidance in finite samples. Further details on the implementation are as follows. The long-run covariance matrices in (3.1) are computed using the Barlett kernel, $k(x) = 1 - |x|$ for $|x| \leq 1$ (and zero otherwise), and the bandwidth selection method described in Andrews (1991). Simulated limiting distributions are based on $J = 299$ replicates and we set $M_T = T$ (Algorithm 1). We test at 5% significance and report results based on $3,000$ Monte Carlo replications.

DGP1: Empirical size and power of the coefficient tests

This DGP follows Wagner et al. (2020). It augments their quadratic seemingly unrelated cointegrating polynomial regression model with a global flexible trend. That is, we consider

$$y_{i,t} = \tau_g t^\theta + \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + \phi_{2,i} x^2_{i,t} + u_{i,t}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{4.1}$$

and compute $u_{i,t}$ and $\Delta x_{i,t} = v_{i,t}$ recursively as

$$u_{i,t} = \rho_1 u_{i,t-1} + \varepsilon_{i,t} + \rho_2 e_{i,t}, \qquad v_{i,t} = e_{i,t} + 0.5 e_{i,t-1}.$$

All recursions are initialized from zero, i.e., $x_{i,0} = u_{i,0} = e_{i,0} = 0$, $i = 1, \ldots, N$. The innovations $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \ldots, \varepsilon_{N,t}]'$ and $\boldsymbol{e}_t = [e_{1,t}, \ldots, e_{N,t}]'$ are drawn independently as $\boldsymbol{\varepsilon}_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon})$ and $\boldsymbol{e}_t \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{ee})$, where

$$\boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \begin{bmatrix} 1 & \rho_3 & \cdots & \rho_3 \\ \rho_3 & 1 & \cdots & \rho_3 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_3 & \rho_3 & \cdots & 1 \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma}_{ee} = \begin{bmatrix} 1 & \rho_4 & \cdots & \rho_4 \\ \rho_4 & 1 & \cdots & \rho_4 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_4 & \rho_4 & \cdots & 1 \end{bmatrix}.$$

Regarding the global trend in (4.1), we set $\tau_g = -0.2$ and consider $\theta \in \{0.8, 1.3, 1.8\}$. All other coefficient values are inspired by Wagner et al. (2020). That is, $\tau_{1,i} = 1$, $\tau_{2,i} = 1$ and $\phi_{1,i} = 5$ are identical across equations.[7] Also, we let $\rho_1 = \rho_2 = \rho_3 = \rho_4$ and redefine these four parameters as ρ. We vary $\rho \in \{0, 0.3, 0.6, 0.8\}$, $N \in \{3, 5, 10\}$, and $T \in \{150, 300, 600\}$. In line with the typical EKC application, we test for the significance of $x_{i,t}^2$. We set $\phi_{2,i} = 0$ for $i = 1, \ldots, N$ and report the empirical size of the single equation test for $H_0 : \phi_{2,1} = 0$ and the joint test for $H_0 : \phi_{2,1} = \ldots = \phi_{2,N} = 0$.

For $\theta_0 = 1.3$, the empirical size of various tests are displayed in Table 2.[8] These tests are based on four estimators: (1) the NLS estimator with simulated critical values as in Section 3.2 (SimNLS); (2) the NLS estimator with simulated critical values and the true value for $\theta_0 = 1.3$ (SimNLS(θ_0)); (3) the FM-SOLS estimator based on $\theta_0 = 1.3$ (FM-SOLS(θ_0)); and (4) the FM-SUR estimator based on $\theta_0 = 1.3$ (FM-SUR(θ_0)). The main findings are as follows:

(a) The simulation-based approaches SimNLS and SimNLS(θ_0) offer better size control. The size improvements are particularly pronounced when $T = 150$ and $\rho = 0.8$. The differences in the empirical size of SimNLS and SimNLS(θ_0) are small.

(b) Size distortions are more severe when N increases and/or a joint test is performed. The same observation was made in Wagner et al. (2020). The behaviour of the simulation-based and fully modified tests is the opposite in these cases. SimNLS and SimNLS(θ_0) tend to become conservative whereas FM-SOLS and FM-SUR are oversized.

We subsequently simulate power curves.[9] The specification of the single equation test and the joint test is as before but we now vary $\phi_{2,1} = \ldots = \phi_{2,N}$ over the set $[-0.008, -0.007, \ldots, 0]$. We

[7] This homogenous parametrisation is particularly convenient to study the impact of the cross-sectional dimension. That is, we can vary N without having to provide additional parameter values. DGP2 is directly inspired by the empirical application and thus more realistic.

[8] The results for $\theta = 0.8$ and $\theta = 1.8$ are qualitatively the same. For brevity, we do not include these results in the main paper. The interested reader can find such simulation results in Section S6 of the Supplementary Material.

[9] Power curves are computationally more intensive. We economize computational time by (1) reducing the number of MC replicates to 1,000 and (2) investigating a subset of all possible parameter configurations.

Table 2: The empirical size (in %) of the tests $H_0 : \phi_{2,1} = 0$ and the joint test for $H_0 : \phi_{2,1} = \ldots = \phi_{2,N} = 0$ with $\phi_{2,i}$ denoting the coefficient in front of $x_{i,t}^2$. Results are based on: simulated inference with θ estimated by NLS (SimNLS), simulated inference with known $\theta = 1.3$ (SimNLS(θ_0)), and two FM estimators for systems as in Wagner et al. (2020) with known $\theta = 1.3$ (FM-SOLS(θ_0) and FM-SUR(θ_0)).

$\theta_0 = 1.3$	$N = 3$				$N = 5$				$N = 10$			
ρ	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)
Panel A: Single-equation test												
$T = 150$												
0	4.77	4.63	9.53	10.77	4.43	4.90	10.03	12.80	4.37	4.47	9.70	16.63
0.3	4.80	4.77	10.47	11.97	4.53	4.63	10.00	13.00	4.23	4.27	11.60	19.00
0.6	4.83	4.53	11.90	12.93	3.93	4.17	11.87	16.50	5.13	4.90	14.23	31.70
0.8	4.50	4.67	14.00	19.10	4.90	4.60	15.23	26.53	5.27	4.43	16.83	56.47
$T = 300$												
0	4.43	4.10	8.03	8.63	4.43	4.20	7.33	8.33	4.50	4.67	8.60	11.73
0.3	4.20	4.60	8.13	9.20	4.37	4.77	8.97	9.80	4.43	4.23	8.67	12.83
0.6	5.80	5.70	10.23	11.80	4.97	4.97	10.17	12.73	4.40	4.30	10.83	18.77
0.8	4.97	4.43	11.17	13.40	4.60	4.53	12.07	19.37	4.03	3.53	13.67	36.53
$T = 600$												
0	4.67	4.67	7.00	7.20	4.77	4.73	6.57	7.53	4.27	4.23	7.37	9.20
0.3	4.77	4.37	7.67	7.77	4.43	4.73	7.37	7.40	4.80	4.67	7.90	9.57
0.6	5.60	5.07	9.43	9.50	5.43	5.43	8.77	10.30	5.30	4.83	9.57	14.10
0.8	4.70	4.50	8.73	9.20	5.10	5.23	9.27	13.87	5.80	5.20	11.47	24.63
Panel B: Joint test												
$T = 150$												
0	4.23	4.03	12.70	15.10	4.10	4.07	14.50	21.57	3.63	3.47	25.67	51.23
0.3	4.80	4.57	14.37	17.33	3.80	3.70	19.50	26.90	3.70	3.77	31.63	59.73
0.6	4.27	4.03	18.03	22.07	3.80	3.57	23.67	37.77	3.13	3.03	40.27	82.17
0.8	3.13	2.93	23.37	30.47	3.33	2.73	31.60	57.20	2.00	1.50	49.73	82.60
$T = 300$												
0	4.80	4.73	10.07	10.87	4.30	4.37	10.63	14.83	3.77	3.80	18.17	31.80
0.3	4.77	4.97	11.60	12.93	4.67	4.67	14.03	17.70	3.40	3.47	19.13	36.53
0.6	4.93	4.40	14.33	14.87	3.77	3.90	17.50	25.63	3.10	3.10	29.20	59.97
0.8	3.87	3.40	17.40	20.97	3.33	2.77	22.97	38.80	2.60	2.13	37.87	86.33
$T = 600$												
0	4.27	4.53	7.37	8.03	4.33	4.13	8.50	10.83	3.97	4.07	12.90	19.23
0.3	4.93	5.17	9.23	10.00	4.80	4.53	10.57	12.30	4.60	4.50	14.87	24.03
0.6	4.07	3.80	11.63	12.43	4.57	4.73	12.30	16.97	4.30	4.23	21.73	38.80
0.8	5.00	4.53	12.77	14.37	3.57	3.80	15.67	25.33	3.60	3.57	26.43	66.63

take $\rho = 0.3$, $\theta = 1.3$, and $N = 3$ as the baseline scenario and subsequently vary these quantities one by one. Figures 1–2 in the supplement show the results. As expected, power increases with increasing sample size, and as $\phi_{2,i}$ moves away from zero.

DGP2: Illustrative simulations in line with the empirical application

Our second set of simulations is tailored toward the EKC application. That is, we employ parametrizations that mimic the data. Generally speaking, we first estimate the baseline model specification on the data and subsequently fit a VAR(1) specification on the stacked vector of residuals and first-differenced explanatory variables.[10] In line with the empirical application, these simulations use $N = 6$ and $T = 145$. All results are displayed in Figure 3. Below, we motivate the simulation settings in view of the EKC application and draw conclusions.

(a) *Correctly specified model*: The specification $y_{i,t} = \tau_g t^\theta + \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + \phi_{2,i} x_{i,t}^2 + u_{i,t}$ with $\phi_{2,i} = 0$ is estimated on the data. We subsequently move $\phi_{2,1} = \ldots = \phi_{2,6}$ away from zero in the DGP and check whether we can detect the resulting curvature caused by the integrated variable. Power curves for the individual and joint test for the coefficients in front of $x_{i,t}^2$ are found in Figures 3(a) and 3(b) in the supplement, respectively. Clearly, nonlinear effects due to $x_{i,t}^2$ are detectable. The statistical power varies across units because (contrary to DGP1) time series properties are now heterogeneous across equations.

(b) *Redundant global trend*: Assumption 1 requires the global trend to be relevant. This simulation DGP investigates how violations of this assumption affect the typical EKC coefficient test. We obtain parameter values by fitting the model $y_{i,t} = \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + \phi_{2,i} x_{i,t}^2 + u_{i,t}$ with $\phi_{2,i} = 0$. As in (a), we vary $\phi_{2,1} = \ldots = \phi_{2,6}$ and test for the significance of these parameters. The solid lines in Figures 3(c) and 3(d) in the supplement are power curves obtained using the correctly specified DGP whereas markers indicate the power when a redundant global trend is estimated as well. The redundant trend has virtually no influence on the statistical power of the coefficient tests of the first five series. There is a power loss for $i = 6$. An inspection of the coefficients offers an explanation. The estimated coefficients in front of the global trend are mostly small (10^{-10} to 10^{-9}) and thus irrelevant. However, in a fraction of cases the flexible trend mimics the curvature in the sixth series causing the quadratic stochastic trend to become insignificant. As reported in the introduction, the power of the joint test does not suffer from the inclusion of a redundant trend.

(c) *KPSS test*: Nonstationary residuals are an indication of model misspecification. That is, either the regression is spurious or the functional form of the cointegrating relation is misspecified.

[10] All details on the simulation designs for DGP2(a)–(c) are available in Section S5 of the Supplementary Material.

We look at the latter situation. The simulation DGP is the quadratic GTACPR as in DGP2(a) but the quadratic component is missing in the fitted model. The empirical rejection frequency of the KPSS test (Figure 3(e) in the supplement) is signalling that there are specification issues. However, a comparison with Figures 3(a) - 3(d) also reveals that if the source of misspecification is known, then a dedicated coefficient test leads to a higher power.

5 Empirical application

We examine the evidence for an EKC for a collection of 18 countries over the period 1870–2014 ($T = 145$). Economic growth is measured by GDP and we use carbon dioxide (CO_2) emissions as a proxy for air pollution. The origin of these data is as follows. We use population and GDP data from the Maddison Project (see https://www.rug.nl/ggdc/historicaldevelopment/maddison/). Our carbon dioxide observations are fossil-fuel CO_2 emissions as made available by the Carbon Dioxide Information Analysis Center (CDIAC, see https://cdiac.ess-dive.lbl.gov). The CDIAC database ceased operation in 2017 causing these time series to be available until 2014. Both GDP and CO_2 emissions are expressed per capita and subsequently log-transformed. In accordance with the notation of this paper, we will denote them by $x_{i,t}$ and $y_{i,t}$, respectively. The same data (or subsets thereof) have also been studied by Wagner (2015), Chan and Wang (2015), Wang et al. (2018), Wagner et al. (2020), and Lin and Reuvers (2022).[11] This conveniently allows us to compare results. All user choices (kernel specification, bandwidth selection, etc.) are kept the same as during the simulation study (see page 13).

5.1 An illustration using Belgian data

Prior to the analysis of a multivariate specification, we will first discuss several features of the individual time series (hence omitting subscripts "i"). The example throughout this narrative is Belgium (Figure 1).[12] An inverted U-shaped relationship between GDP and CO_2 (both in log per capita) is clearly visible in Figure 1(a) and behavior like this has triggered research on the Environmental Kuznets Curve. However, the time heat map also shows that time is almost monotonically increasing along the curve. Time effects – e.g., increasing global environmental awareness, worldwide advances in sustainable technologies – can be valid alternative explanations for these nonlinearities and their omission can falsely exaggerate the influence of GDP. It is for this reason that we develop

[11] The stationarity properties of the series have been discussed in these papers already. We will not repeat this analysis but refer the interested reader to Section S7 of the Supplement.

[12] The data for Austria, Belgium, and Finland are mentioned in both Wagner (2015) and Wagner et al. (2020) to behave in line with the EKC. We discuss Belgium in the main text but the interested reader can find the same figures for Austria and Finland in Section S7.3. Qualitatively, the findings for these other two countries are the same.

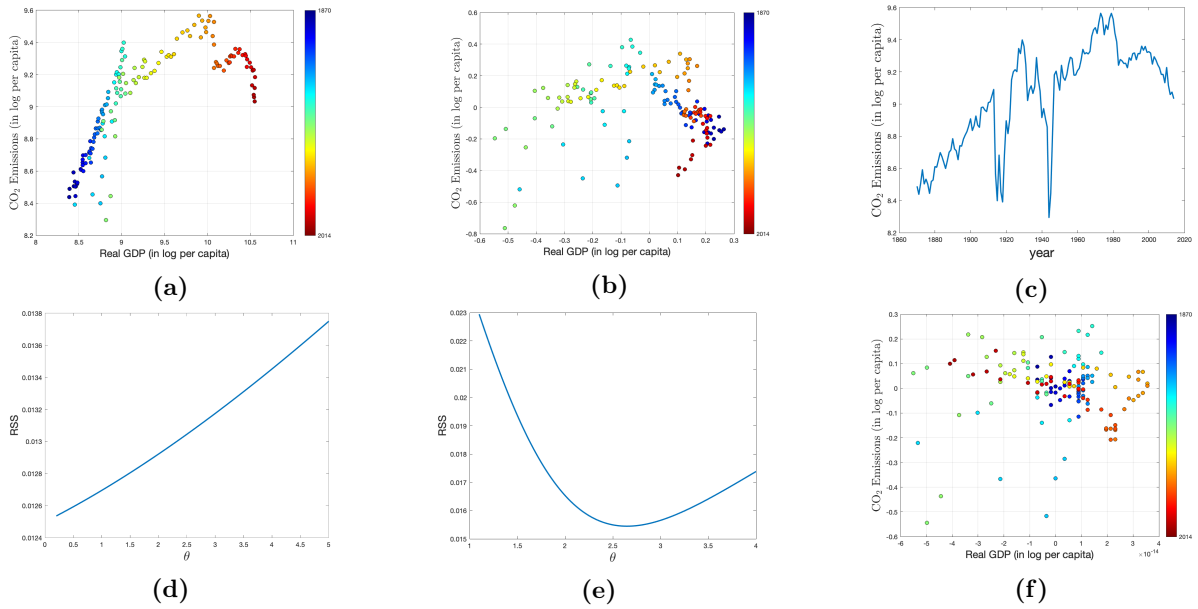


Figure 1: Overview graphs for Belgium over 1870-2014. **(a)** log(GDP) versus log(CO_2) (both per capita). **(b)** The same series as in subfigure (a), but now using detrended variables. **(c)** The log per capita CO_2 emissions time series for Belgium over time. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ. **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

and analyse the GTACPR. More evidence for the importance of time effects is available in Figure 1(b). This figure depicts the same per capita series after detrending.[13] The inverted U-shape is now (visually) less pronounced or even absent.

Finally, let us depart from a traditional linear cointegration specification: $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + u_t$. This model cannot incorporate any nonlinear behaviour over time and is, therefore, ill-suited to fit the data displayed in Figure 1(c). Cointegrating polynomial regressions use integer powers of x_t to describe the curvature over time. More generally, as in Hu et al. (2021), we can allow for an integrated regressor with a flexible power and estimate $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$. The residual sum of squares (RSS) of the NLS estimator for this specification is shown in Figure 1(d). The absence of a minimum at $\theta = 2$ casts doubt on the commonly used quadratic specification in x_t. Additionally, the lack of any minimum might be interpreted as a sign that log per capita GDP is not the source of nonlinearity. This finding is not specific for Belgium. There are no minima in the RSS for 15 out of 18 countries (see Section S7.4). For the remaining three countries – Denmark, France and the Netherlands – minima are found at $\widehat{\theta}_{DK} = 1.46$, $\widehat{\theta}_{FR} = 3.61$ and $\widehat{\theta}_{NL} = 1.28$, respectively. Alternatively, we can describe the nonlinearity in the data using a flexible deterministic trend as in

[13] The Perron and Yabu (2009) test allows us to test for the presence of a deterministic trend irrespectively of the series being trend-stationary or having a unit root. The results of this test (see supplement) indicate that log per capita GDP is likely to have a deterministic trend component. It is thus recommended to have a deterministic trend in the model for log per capita CO_2 emissions and the visual inspection of the relationship between GDP and CO_2 emissions (in log per capita) should take place after partialling out this deterministic trend.

$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t$. The RSS in Figure 1(e) now exhibits a clear minimum. Further empirical analysis on individual countries (see Section S7.5 of the Supplement) suggests that: (1) the inclusion of a flexible time trend renders all quadratic effects in squared log per capita GDP insignificant, and (2) models remain well-specified after removing quadratic income effects from the model. These results suggest – albeit in a univariate setting – that flexible time trends give a more satisfactory (or at least competing) description of the nonlinearities in the data.

5.2 Seemingly unrelated regression

The interpretation of a country-specific flexible deterministic trend is complicated because of its high collinearity with GDP per capita. The multivariate analysis of this section allows us to separate country-specific environmental improvements caused by national income growth from global environmental improvements. We study the following six countries ($N = 6$): Austria, Belgium, Finland, the Netherlands, Switzerland, and the UK. The motivation behind this choice is as follows. First, based on data series to ours, Piaggio and Padilla (2012), Mazzanti and Musolesi (2013), and Wagner et al. (2020) report considerable evidence of parameter heterogeneity across countries.[14] The evidence in Mazzanti and Musolesi (2013) is anecdotal in the sense that these authors consider groups of similar countries and find different results for different groups. The lack of overlap among confidence intervals of country-specific parameters has also been interpreted as a sign of heterogeneity (section 4.2 in Piaggio and Padilla (2012)). Wagner et al. (2020) explicitly test for various forms of poolability and conclude that pooling is (at most) appropriate for small subgroups of countries. This lack of parameter homogeneity justifies a multivariate approach with a small N rather than a panel setting. Admittedly, in the current time-series setting, studying "large N" is also infeasible since consistent estimators for $(2N \times 2N)$ long-run covariance matrices are required. Second, prior studies already refute the existence of a carbon dioxide EKC for several countries, and little seems lost by excluding these countries from the outset.[15] That is, we consider the same countries as in Wagner et al. (2020), who decide on these countries because their prior cointegration analysis "*leads to evidence for a quadratic cointegrating EKC including a constant and linear trend*".

Having decided on the set of countries, we subsequently study the effect of the global flexible

[14]Parameter heterogeneity is also reported for other data, e.g., List and Gallet (1999), Dijkgraaf and Vollebergh (2005).

[15]Most parameters in the GCPR are country-specific. The estimation accuracy of these parameters should deteriorate little when focusing on a subset of countries. Losses will occur in the precision of the estimators for τ_g and θ. There is thus a trade-off between accurate global trend estimation (improving with large N) and accurate LRV estimation (deteriorating with large N). To strike a balance and to connect to the recent literature, we continue the analysis of Wagner et al. (2020) and take $N = 6$.

Table 3: Parameter estimates and test results for Models (M1)–(M3). The joint p-value refers to the test with null hypothesis $H_0 : \phi_{2,1} = \ldots = \phi_{2,6} = 0$ and is thus inapplicable for Model (M3).

Model	Omitted Global Trend						Global Trend		
	(M1)						(M2)		(M3)
	FM-SOLS		FM-SUR		SimNLS		SimNLS		SimNLS
	$\phi_{1,i}$	$\phi_{2,i}$	$\phi_{1,i}$	$\phi_{2,i}$	$\phi_{1,i}$	$\phi_{2,i}$	$\phi_{1,i}$	$\phi_{2,i}$	$\phi_{1,i}$
Austria	9.37^{***}	-0.43^{***}	3.96^{*}	-0.16	6.42^{***}	-0.28	3.08^{***}	-0.09	1.73^{***}
Belgium	11.78^{***}	-0.59^{***}	9.92^{***}	-0.50^{***}	12.36^{***}	-0.62^{**}	7.68^{***}	-0.36	1.01^{***}
Finland	16.00^{***}	-0.72^{***}	15.07^{***}	-0.68^{***}	17.18^{***}	-0.78^{*}	15.19^{***}	-0.65	2.22^{***}
Netherlands	10.68^{***}	-0.51^{***}	9.58^{***}	-0.46^{***}	9.27^{***}	-0.44^{*}	4.97^{***}	-0.20	1.33^{***}
Switzerland	8.17^{***}	-0.27^{***}	7.29^{***}	-0.23^{***}	8.11^{***}	-0.28	0.58^{*}	0.10	2.55^{***}
UK	9.28^{***}	-0.47^{***}	7.93^{***}	-0.40^{***}	9.16^{***}	-0.46^{*}	4.93^{***}	-0.21	1.33^{***}
Joint p-value	0.00		0.00		0.16		0.39		—
KPSS-statistic	3.45		5.10		3.46		3.48		3.78
$\widehat{\tau}\, t^{\widehat{\theta}}$							$-0.012\, t^{1.263}$		$-1.374 \cdot 10^{-5} t^{2.450}$

Note: Asterisks denote rejection of the null hypothesis at the ***1%, **5%, and *10% significance level. Depending on the specific table entry, the null hypothesis refers to coefficient(s) being zero or a well-specified cointegrating relation.

trend on EKC evidence. Table 3 shows the estimation results of the quadratic EKC specification

$$y_{i,t} = \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + \phi_{2,i} x_{i,t}^2 + u_{i,t}. \tag{M1}$$

This setting (possibly with the additional constraint $\tau_{2,i} = 0$) has been explored in numerous papers, for example, Selden and Song (1994), Piaggio and Padilla (2012), Chan and Wang (2015), Wagner (2015), Wang et al. (2018), and Wagner et al. (2020). For Model (M1), an inverted-U relationship results when $\phi_{1,i} > 0$ and $\phi_{2,i} < 0$ and empirical evidence hereof is traditionally interpreted as the existence of an EKC. If these coefficients have the correct signs, then the country's turning point – the level of economic growth at which environmental improvement starts – can be computed as $\exp(-\phi_{1,i}/2\phi_{2,i})$. We assess the parameter values and their significance using FM-SOLS and FM-SUR (repeating the analysis of Wagner et al. (2020) for ease of comparison) and the simulated approach of Section 3.2. Regardless of the estimation method and country, all coefficient signs are in agreement with the EKC hypothesis. The parameters $\phi_{1,i}$ are generally significantly different from zero but the significance of $\phi_{2,i}$ does vary across estimation methods. FM-SOLS and FM-SUR typically (strongly) reject $H_0 : \phi_{2,i} = 0$ $(i = 1, \ldots, 6)$ whereas evidence against these null hypotheses is less pronounced for the simulation-based approach. The same behaviour emerges when testing $\phi_{2,1} = \ldots = \phi_{2,6} = 0$ jointly. This pattern reminds us of the simulation results in Table 2 where the cross-sectional dimensions $N = 5$ and $N = 10$ cause over-sized tests for FM-SOLS and FM-SUR and conservative tests for simulation-based inference. The KPSS test does not indicate any signs of misspecification. Overall, Model (M1) leads to considerable evidence in favour of a quadratic cointegrating EKC.

The reported evidence in favour of the EKC should not come as surprise. First, the set of

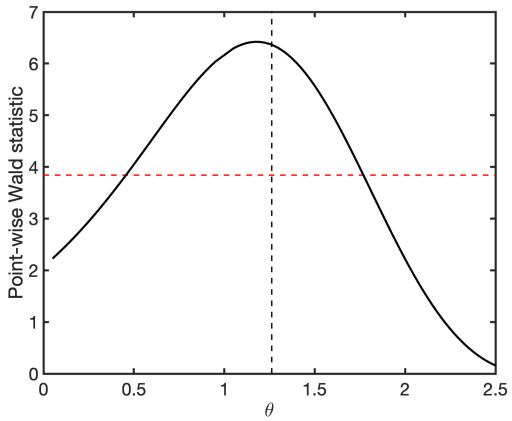


Figure 2: The magnitude of the Wald test for fixed values of θ when testing $H_0 : \tau_g = 0$ under Model (M2). Dash lines display the 95% quantile of a chi-squared distributed random variable with 1 degree of freedom (red) and the NLS estimate $\widehat{\theta} = 1.263$ for specification (M2).

countries was selected based on these criteria. Second, the visualisations of the data clearly suggest nonlinear effects (recall Figures 1(a) and 1(c) for the case of Belgium). With Model (M1) being restrictive in the sense that nonlinearities over time are solely incorporable through $x_{i,t}^2$, we expect this variable to be important. In line with our proposed GTACPR framework, we subsequently add a global flexible trend and estimate

$$y_{i,t} = \tau_g t^\theta + \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + \phi_{2,i} x_{i,t}^2 + u_{i,t}. \tag{M2}$$

From a statistical perspective, the term $\tau_g t^\theta$ opens a different channel through which nonlinearities can be described. We refer back to the introduction for further elaboration on this point. From an economic perspective, $\tau_g t^\theta$ captures changes in CO_2 emissions that are common across series and thus unrelated to changes in national GDPs. Parameter inference for Model (M2) is also reported in Table 3. The contributions of $x_{i,t}^2$ are insignificant for both individual countries and all countries jointly. How about the significance of the global trend? The standard Wald test for $\tau_g = 0$ is invalid because θ is unidentified under the null hypothesis (see Assumption 1 and the related discussion). As a heuristic alternative, we vary θ over the interval $[0, 2.5]$ and compute Wald statistics while assuming θ to be fixed. Comparing these Wald statistics to the 95% quantile of a $\chi^2(1)$-distributed random variable (critical value: 3.842), the range of θ-values from about 0.5 to 1.75 implies a significant global trend (Figure 2). Having estimated $\widehat{\theta} = 1.263$, our analysis suggests that the global trend and not GDP per capita is the source of nonlinearity. Before interpreting this result, we first verify whether the model with $\phi_{2,1} = \ldots = \phi_{2,6} = 0$ shows signs of misspecification.

Omitting insignificant parameters from the previous model specification, we arrive at

$$y_{i,t} = \tau_g t^\theta + \tau_{1,i} + \tau_{2,i} t + \phi_{1,i} x_{i,t} + u_{i,t}. \tag{M3}$$

Model (M3) is linear in log per capita GDP. The positive parameter estimates for $\phi_{1,i}$ imply that *at a given point in time* increases in economic growth imply increases in CO_2 emissions. However, as $\widehat{\tau}_g = -1.374 \times 10^{-5}$ and $\widehat{\theta} = 2.45$, there will be common emission reductions over time. Also, the omission of the quadratic terms in log per capita GDP does not seem to result in a misspecified model. First, the KPSS test does not reject the null of cointegration. Second, there is no (visual) evidence that the linear functional form of (M3) is inappropriate. To arrive at this last conclusion, we compute $\widetilde{y}_{i,t} = y_{i,t} - \widehat{\tau}_g t^{\widehat{\theta}} - \widehat{\tau}_{1,i} - \widehat{\tau}_{2,i} t$ and employ the nonparametric kernel estimator from Wang and Phillips (2009) to estimate $\widetilde{y}_{i,t} = f(x_{i,t}) + \widetilde{u}_{i,t}$ for each individual country. Figure 3 shows the nonparametric estimate in blue and the fit of Model (M3) in red. After the removal of the global trend, there are some temporary departures from linearity but there is little curvature overall and certainly no visual turning point. We further conduct a formal statistical test for the null of linearity using the model specification test as outlined in section 3 of Wang and Phillips (2016). The results are reported in Table S7 in the supplement. Based on the full sample, linearity is rejected for Austria only. A comparison with the 95% confidence intervals of the kernel estimate (Figure 3) suggests that this rejection is caused by the sharp decline in CO_2 emissions during World War II. We subsequently repeat the analysis using the $T = 69$ observations after 1945. Linearity is never rejected.[16] All this aligns well with our earlier findings of a relevant global trend and irrelevant quadratic effects in log GDP per capita.

The preceding analysis suggests that the global flexible trend captures omitted determinants of CO_2 emission levels that have been decreasing over time. In their analysis, Grossman and Krueger (1995) already included a global deterministic trend in their model because they "*did not want to attribute to national income growth any improvements in local environmental quality that might actually be due to global advances in the technology for environmental preservation or to an increased global awareness of the severity of environmental problems*". Indeed, since reliable data on green technology adaptation[17] and global awareness is scarcely available (certainly for time horizons allowing for a cointegration analysis), these variables are likely missing and thus requiring a proxy. Similar remarks are applicable to variables such as pollution control policies[18]. In reduced-form models, an EKC finding is typically explained by national income being the proxy for these omitted

[16]The properties of nonparametric kernel estimators in nonlinear cointegration models have been studied by Wang and Phillips (2009) and Wang and Phillips (2016), among others. The latter reference is particularly relevant because it establishes that kernel estimators remain consistent and asymptotically (mixed) normal under serially correlated errors and endogeneity. These papers do not include deterministic trends in the DGP. However, we conjecture that detrending does not affect the asymptotic properties of the kernel estimator due to the high convergence rates of the trend parameters in comparison to the slow convergence rates of the nonparametric estimator. We take a bandwidth $h = T^{-1/3}$ in implementation. Matlab functions for nonparametric kernel regression and specification tests are available at `https://github.com/HannoReuvers`.

[17]Nordhaus (2014) discusses the link between climate change and technological changes.

[18]A policy variable, 'Repudiation of Contracts by Government', was included by Panayotou (1997) to proxy the quality of environmental policies and institutions.

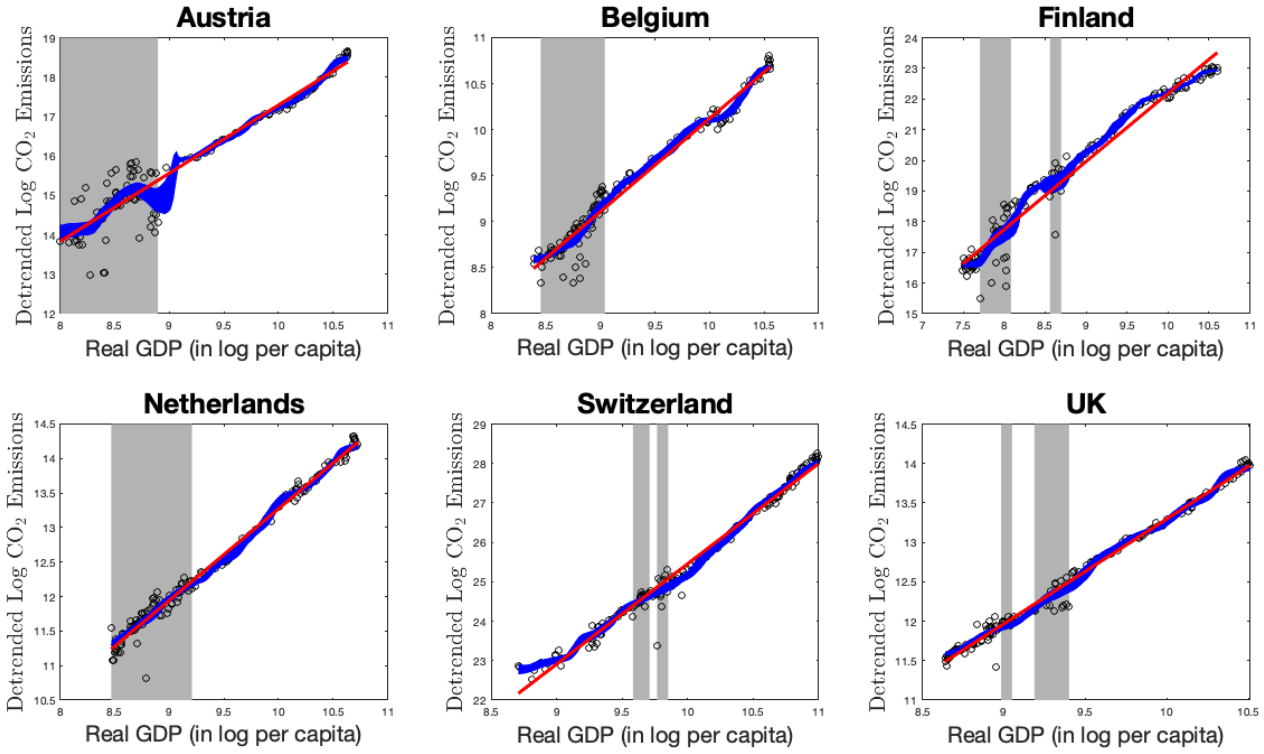


Figure 3: The 95% confidence intervals of the non-parametric kernel estimate for the relationship between GDP and CO_2 emissions (blue) after removal of the country-specific and joint flexible deterministic trends. The red line is the linear fit from Model (M3). As the sample covers the years 1870–2014 there are several observations during World War I and World War II. The affected ranges of GDP are indicated in grey.

variables. That is, at higher levels of national income, countries have access to cleaner technologies and their citizens show greater appreciation for the environment and pollution legislation. The current analysis contradicts these income effects and points towards improvements being captured by a global trend.

Our final model specification (M3) is linear in log GDP per capita. Moreover, for a given year, the coefficient estimates suggest that increasing national income by 1% implies an *increase* in CO_2 emissions of about 1%–2.5% (depending on the country). This result seems plausible for non-carbon-neutral economies. However, CO_2 emissions in Austria, Belgium, Finland, the Netherlands, Switzerland, and the UK are jointly reducing at the end of the sample. What causes these global emission reductions? Mazzanti and Musolesi (2013) suggest that conglomerates of countries anticipate and respond to international climate agreements such as the Rio Convention (1992) and Kyoto Protocol (adopted in 1997; operational since 2005). Interestingly, the latter agreement contains emission reduction targets to be reached in 2020, and such "working-towards-a-common-reduction-deadline" does point towards a time effect.[19] Alternatively, given our sample of European countries,

[19]According to the Doha amendment of the Kyoto Protocol, the reduction commitments were 92% (over the period 2008–2012) and 80% (over the period 2013–2020) of 1990 emission levels for Austria, Belgium, Finland, the Netherlands, and the UK. For Switzerland, the reduction target was also 92% (over the period 2008–2012) but 84.2% (over the period 2013–2020). (source: https://unfccc.int/files/kyoto_protocol/application/pdf/kp_doha_amendment_english.pdf).

EU-coordinated emission reduction efforts like the EU Emissions Trading System can be a driving force behind these common emission decreases.

Acknowledgements

Earlier versions of this paper have been presented at the 2019 CFE London meeting, the Econometrics Internal Seminar at Erasmus University Rotterdam, and the Brownbag Seminar at Vrije Universiteit Amsterdam. We gratefully acknowledge the comments of the participants. We extend our thanks to Eric Beutner, Dick van Dijk, Stephan Smeekes, and Xiaohu Wang for their valuable feedback. All remaining errors are our own.

References

Al-Mulali, U. and I. Ozturk (2016). The investigation of Environmental Kuznets Curve hypothesis in the advanced economies: The role of energy prices. *Renewable and Sustainable Energy Reviews 54*, 1622–1631.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Andrews, D. W. K. and Y. Sun (2004). Adaptive local polynomial whittle estimation of long-range dependence. *Econometrica 72*, 569–614.

Baek, Y. I., S. J. Cho, and P. C. B. Phillips (2015). Testing linearity using power transforms of regressors. *Journal of Econometrics 187*, 376–384.

Beutner, E., Y. Lin, and S. Smeekes (2022). GLS estimation and confidence sets for the date of a single break in models with trends. *Econometric Reviews*. Forthcoming.

Carson, R. T. (2009). The Environmental Kuznets Curve: seeking empirical regularity and theoretical structure. *Review of Environmental Economics and Policy 4*, 3–23.

Chan, N. and Q. Wang (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics 185*, 182–195.

Chang, Y., J. Y. Park, and P. C. B. Phillips (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal 4*, 1–36.

Cho, J. S. and P. C. B. Phillips (2018). Sequentially testing polynomial model hypotheses using power transforms of regressors. *Journal of Applied Econometrics 33*, 141–159.

Choi, I. and P. Saikkonen (2010). Tests for nonlinear cointegration. *Econometric Theory 26*, 682–709.

Dasgupta, S., B. K. H. Wang, and D. Wheeler (2002). Confronting the Environmental Kuznets Curve. *Journal of Economic Perspectives 16*, 147–168.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*, 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*, 33–43.

Dijkgraaf, E. and H. R. J. Vollebergh (2005). A test for parameter homogeneity in CO_2 panel EKC estimations. *Environmental and Resource Economics 32*, 229–239.

Duggal, V. G., C. Saltzman, and L. R. Klein (2007). Infrastructure and productivity: An extension to private infrastructure and it productivity. *Journal of Econometrics 140*, 485–502.

Galeotti, M., M. Manera, and A. Lanza (2009). On the robustness of robustness checks of the environmental Kuznets curve hypothesis. *Environmental and Resource Economics 42*, 551.

Gao, J., O. Linton, and B. Peng (2020). Inference on a semiparametric model with global power law and local nonparametric trends. *Econometric Theory 36*, 223–249.

Gillingham, P. C. K. and W. Nordhaus (2018). Uncertainty in forecasts of long-run economic growth. *Proceedings of the National Academy of Sciences 115*, 5409–5414.

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

Harbaugh, W. T., A. Levinson, and D. M. Wilson (2002). Reexamining the empirical evidence for an environmental Kuznets curve. *Review of Economics and Statistics 84*, 541–551.

Hong, S. H. and P. C. B. Phillips (2010). Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business & Economic Statistics 28*, 96–114.

Hu, Z., P. C. B. Phillips, and Q. Wang (2021). Nonlinear cointegrating power function regression with endogeneity. *Econometric Theory 37*, 1173–1213.

Jalil, A. and M. Feridun (2011). The impact of growth, energy and financial development on the environment in China: A cointegration analysis. *Energy Economics 33*, 284–291.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory 18*, 1449–1459.

Klein, L. R., V. G. Duggal, and C. Saltzman (2004). Contributions of input-output analysis to the understanding of technological change: the information sector in the United States. In E. Dietzenbacher and M. L. Lahr (Eds.), *Wassily Leontief and Input-output Economics*, Chapter 17, pp. 311–336. Cambridge University Press.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics 54*, 159–178.

Li, S. and O. Linton (2020). When will the covid-19 pandemic peak? *Journal of Econometrics 220*, 130–157.

Lin, Y. and H. Reuvers (2022). Fully modified estimation in cointegrating polynomial regressions: Extensions and Monte Carlo comparison. Working paper.

Lin, Y., Y. Tu, and Q. Yao (2020). Estimation for double-nonlinear cointegration. *Journal of Econometrics 216*, 175–191.

Linton, O. and Q. Wang (2016). Nonparametric transformation regression with nonstationary data. *Econometric Theory 32*, 1–29.

List, J. A. and C. A. Gallet (1999). The environmental Kuznets curve: Does one size fit all? *Ecological Economics 31*, 409–423.

Maranzano, P., J. P. C. Bento, and M. Manera (2021). The role of education and income inequality on environmental quality. A panel data analysis of the EKC hypothesis on OECD. Nota di Lavoro 8.2021, Milano, Italy: Fondazione Eni Enrico Mattei.

Mazzanti, M. and A. Musolesi (2013). The heterogeneity of carbon kuznets curves for advanced countries: Comparing homogeneous, heterogeneous and shrinkage/bayesian estimators. *Applied Economics 45*, 3827–3842.

Nordhaus, W. D. (2013). *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World*. Yale University Press.

Nordhaus, W. D. (2014). The perils of the learning model for modeling endogenous technological change. *The Energy Journal 35*, 1–13.

Panayotou, T. (1997). Demystifying the environmental kuznets curve: Turning a black box into a policy tool. *Environment and Development Economics 2*, 465–484.

Park, J. Y. (2002). An invariance principle for sieve bootstrap in time series. *Econometric Theory 18*, 469–490.

Park, J. Y. and P. C. B. Phillips (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory 15*, 269–298.

Park, J. Y. and P. C. B. Phillips (2001). Nonlinear regressions with integrated time series. *Econometrica 69*, 117–161.

Perron, P. and T. Yabu (2009). Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics 151*, 56–69.

Perron, P. and X. Zhu (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics 129*, 65–119.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica 66*, 1299–1325.

Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory 23*, 557–614.

Phillips, P. C. B. and B. E. Hansen (1990). Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies 57*, 99–125.

Piaggio, M. and E. Padilla (2012). CO_2 emissions and economic activity: Heterogeneity across countries and non-stationary series. *Energy Policy 46*, 370–381.

Robinson, P. M. (2012). Inference on power law spatial trends. *Bernoulli 18*, 644–677.

Romano, J. P. and M. Wolf (2001). Subsampling intervals in autoregressive models with linear time trend. *Econometrica 69*, 1283–1314.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory 8*, 1–27.

Selden, T. M. and D. Song (1994). Environmental quality and development: Is there a Kuznets curve for air pollution emissions? *Journal of Environmental Economics and Management 27*, 147–162.

Stern, D. I. (2004). The rise and fall of the environmental kuznets curve. *World Development 32*, 1419–1439.

Stern, D. I. (2017). The environmental Kuznets curve after 25 years. *Journal of Bioeconomics 19*, 7–28.

Tanaka, K. (2017). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. John Wiley & Sons.

Vogelsang, T. J. and M. Wagner (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics 178*, 741–760.

Vollebergh, H. R. J., B. Melenberg, and E. Dijkgraaf (2009). Identifying reduced-form relations with panel data: The case of pollution and income. *Journal of Environmental Economics and Management 58*, 27–42.

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

Wagner, M., P. Grabarczyk, and S. H. Hong (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*, 216–255.

Wagner, M. and S. H. Hong (2016). Cointegrating polynomial regressions: Fully modified OLS estimation and inference. *Econometric Theory 32*, 1289–1315.

Wang, Q. and P. C. B. Phillips (2009). Structural nonparametric cointegrating regression. *Econometrica 77*, 1901–1948.

Wang, Q. and P. C. B. Phillips (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics 40*, 727–758.

Wang, Q. and P. C. B. Phillips (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory 32*, 359–401.

Wang, Q., D. Wu, and K. Zhu (2018). Model checks for nonlinear cointegrating regression. *Journal of Econometrics 207*, 261–284.

Appendix A: Proofs for main theorems

We collect the proofs of the main theorems in this section. Further information is available in the Supplement.

Proof of Theorem 1: In view of the identity $\|\boldsymbol{a}+\boldsymbol{b}\|^2 = \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 + 2\boldsymbol{a}'\boldsymbol{b}$, we have

$$Q_T(\boldsymbol{\gamma}) = \frac{1}{2}\sum_{t=1}^{T}\|\boldsymbol{y}_t - \boldsymbol{Z}_t'\boldsymbol{\beta}\|^2 - \tau_g\sum_{t=1}^{T}t^{\theta}\left(\boldsymbol{y}_t - \boldsymbol{Z}_t'\boldsymbol{\beta}\right)'\boldsymbol{\imath}_N + \frac{1}{2}N\tau_g^2\sum_{t=1}^{T}t^{2\theta}.$$

The proof proceeds along the lines of Lemma 1 of Andrews and Sun (2004) and Theorem 3.1 of Chan and Wang (2015). The proofs separate into two parts. The first part uses a Taylor expansion of $Q_T(\boldsymbol{\gamma})$ around $Q_T(\boldsymbol{\gamma}_0)$ to recover a quadratic approximation for $Q_T(\boldsymbol{\gamma})$ on the set $\boldsymbol{\Gamma}_{\delta,k_T} \subseteq \boldsymbol{\Gamma}$. In the second part, we obtain the limiting distribution from this quadratic approximation.

Part 1: Let $\{k_T, T \geq 1\}$ denote a deterministic sequence such that $k_T \to \infty$ as $T \to \infty$. Define $\boldsymbol{\Gamma}_{\delta,k_T} = \{\boldsymbol{\gamma} \in \boldsymbol{\Gamma} :\ \|\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\| \leq k_T,\ \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\| \leq \delta\}$ and select a $\delta > 0$ such that $Q_T(\cdot)$ is twice differentiable on $\{\boldsymbol{\gamma} \in \mathbb{R}^{p+2} :\ \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\| \leq \delta\} \subset \boldsymbol{\Gamma}$. For any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_{\delta,k_T}$, the Taylor expansion of $Q_T(\boldsymbol{\gamma})$ around $\boldsymbol{\gamma}_0$ reads

$$\begin{aligned} Q_T(\boldsymbol{\gamma}) - Q_T(\boldsymbol{\gamma}_0) &= \dot{Q}_T'(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\ddot{Q}_T(\bar{\boldsymbol{\gamma}})(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) \\ &= \dot{Q}_T'(\boldsymbol{\gamma}_0)(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\left[\ddot{Q}_T(\bar{\boldsymbol{\gamma}}) - \ddot{Q}_T(\boldsymbol{\gamma}_0) - \ddot{Q}_{T,2}(\boldsymbol{\gamma}_0) + \ddot{Q}_{T,1}(\boldsymbol{\gamma}_0)\right](\boldsymbol{\gamma}-\boldsymbol{\gamma}_0), \end{aligned} \tag{A.1}$$

where $\bar{\boldsymbol{\gamma}}$ is a point on the line segment connecting $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_0$, and the various derivatives of Q_T are

$$\dot{Q}_T(\boldsymbol{\gamma}_0) = -\sum_{t=1}^{T}\begin{bmatrix}\boldsymbol{\imath}_N'\tau_{g0}\,t^{\theta_0}\ln t \\ \boldsymbol{\imath}_N'\,t^{\theta_0} \\ \boldsymbol{Z}_t\end{bmatrix}\boldsymbol{u}_t,$$

$$\ddot{Q}_T(\boldsymbol{\gamma}) = \sum_{t=1}^{T}\begin{bmatrix} -\tau_g t^{\theta}(\ln t)^2\left(\boldsymbol{y}_t - \boldsymbol{Z}_t'\boldsymbol{\beta}\right)'\boldsymbol{\imath}_N + 2N\tau_g^2 t^{2\theta}(\ln t)^2 & -t^{\theta}\ln t\left(\boldsymbol{y}_t - \boldsymbol{Z}_t'\boldsymbol{\beta}\right)'\boldsymbol{\imath}_N + 2N\tau_g t^{2\theta}\ln t & \tau_g t^{\theta}\ln t\,\boldsymbol{\imath}_N'\boldsymbol{Z}_t' \\ -t^{\theta}\ln t\left(\boldsymbol{y}_t - \boldsymbol{Z}_t'\boldsymbol{\beta}\right)'\boldsymbol{\imath}_N + 2N\tau_g t^{2\theta}\ln t & Nt^{2\theta} & t^{\theta}\boldsymbol{\imath}_N'\boldsymbol{Z}_t' \\ \tau_g t^{\theta}\ln t\,\boldsymbol{Z}_t\boldsymbol{\imath}_N & t^{\theta}\boldsymbol{Z}_t\boldsymbol{\imath}_N & \boldsymbol{Z}_t\boldsymbol{Z}_t' \end{bmatrix},$$

$$\begin{aligned} \ddot{Q}_T(\boldsymbol{\gamma}_0) &= \sum_{t=1}^{T}\begin{bmatrix} N\tau_{g0}^2 t^{2\theta_0}(\ln t)^2 - \tau_{g0}t^{\theta_0}(\ln t)^2\boldsymbol{u}_t'\boldsymbol{\imath}_N & N\tau_{g0}t^{2\theta_0}\ln t - t^{\theta_0}\ln t\,\boldsymbol{u}_t'\boldsymbol{\imath}_N & \tau_{g0}t^{\theta_0}\ln t\,\boldsymbol{\imath}_N'\boldsymbol{Z}_t' \\ N\tau_{g0}t^{2\theta_0}\ln t - t^{\theta_0}\ln t\,\boldsymbol{u}_t'\boldsymbol{\imath}_N & Nt^{2\theta_0} & t^{\theta_0}\boldsymbol{\imath}_N'\boldsymbol{Z}_t' \\ \tau_{g0}t^{\theta_0}\ln t\,\boldsymbol{Z}_t\boldsymbol{\imath}_N & t^{\theta_0}\boldsymbol{Z}_t\boldsymbol{\imath}_N & \boldsymbol{Z}_t\boldsymbol{Z}_t' \end{bmatrix} \\ &= \sum_{t=1}^{T}\begin{bmatrix} N\tau_{g0}^2 t^{2\theta_0}(\ln t)^2 & N\tau_{g0}t^{2\theta_0}\ln t & \tau_{g0}t^{\theta_0}\ln t\,\boldsymbol{\imath}_N'\boldsymbol{Z}_t' \\ N\tau_{g0}t^{2\theta_0}\ln t & Nt^{2\theta_0} & t^{\theta_0}\boldsymbol{\imath}_N'\boldsymbol{Z}_t' \\ \tau_{g0}t^{\theta_0}\ln t\,\boldsymbol{Z}_t\boldsymbol{\imath}_N & t^{\theta_0}\boldsymbol{Z}_t\boldsymbol{\imath}_N & \boldsymbol{Z}_t\boldsymbol{Z}_t' \end{bmatrix} - \sum_{t=1}^{T}t^{\theta_0}\ln t\begin{bmatrix}\tau_{g0}\ln t & 1 & \\ 1 & 0 & \\ & & \mathbf{O}\end{bmatrix}\boldsymbol{u}_t'\boldsymbol{\imath}_N \\ &=: \ddot{Q}_{T,1}(\boldsymbol{\gamma}_0) - \ddot{Q}_{T,2}(\boldsymbol{\gamma}_0). \end{aligned}$$

For simplicity, let $R_T(\bar{\boldsymbol{\gamma}},\boldsymbol{\gamma}_0) = \frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\left[\ddot{Q}_T(\bar{\boldsymbol{\gamma}}) - \ddot{Q}_T(\boldsymbol{\gamma}_0) - \ddot{Q}_{T,2}(\boldsymbol{\gamma}_0)\right](\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)$, $\boldsymbol{A}_T := \boldsymbol{G}_{\boldsymbol{\gamma}_0,T}'^{-1}\ddot{Q}_{T,1}(\boldsymbol{\gamma}_0)\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}^{-1}$,

and $\boldsymbol{b}_T := -\boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,T}\dot{Q}_T(\boldsymbol{\gamma}_0)$. We finally arrive at

$$\begin{aligned} Q_T(\boldsymbol{\gamma}) - Q_T(\boldsymbol{\gamma}_0) &= -\boldsymbol{b}_T'\big[\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\big] + \frac{1}{2}\big[\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\big]'\boldsymbol{A}_T\big[\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\big] + R_T(\bar{\boldsymbol{\gamma}},\boldsymbol{\gamma}_0) \\ &= \frac{1}{2}\left[\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) - \boldsymbol{A}_T^{-1}\boldsymbol{b}_T\right]'\boldsymbol{A}_T\left[\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) - \boldsymbol{A}_T^{-1}\boldsymbol{b}_T\right] - \frac{1}{2}\boldsymbol{b}_T'\boldsymbol{A}_T^{-1}\boldsymbol{b}_T + R_T(\bar{\boldsymbol{\gamma}},\boldsymbol{\gamma}_0). \end{aligned} \tag{A.2}$$

Part 2: For any $\varepsilon > 0$, let $\boldsymbol{\Gamma}_T(\varepsilon) = \big\{\boldsymbol{\gamma} \in \boldsymbol{\Gamma} : \ \big\|\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) - \boldsymbol{A}_T^{-1}\boldsymbol{b}_T\big\| \le \varepsilon\big\}$. We shall show that the minimum of $Q_T(\cdot)$ over $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_T(\varepsilon)$ is attained in the interior of $\boldsymbol{\Gamma}_T(\varepsilon)$. The next two statements are proven later:

(a) $\sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}_{\delta,k_T}}\left\|\boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,T}\big[\ddot{Q}_T(\boldsymbol{\gamma}) - \ddot{Q}_T(\boldsymbol{\gamma}_0)\big]\boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,T}\right\| = o_p(1)$;

(b) $\boldsymbol{A}_T^{-1}\boldsymbol{b}_T = O_p(1)$, where $\boldsymbol{A}_T \xrightarrow{T\to\infty}_d \boldsymbol{A}_\infty$ with $\mathbb{P}\left(\boldsymbol{A}_\infty > 0\right) = 1$.

Given claim (b), for any $\varepsilon > 0$, we have $\mathbb{P}\left(\boldsymbol{\Gamma}_T(\varepsilon) \subset \boldsymbol{\Gamma}_{\delta,k_T}\right) \to 1$ as $T \to \infty$ because $\|\boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,T}\| \to 0$. Define $\boldsymbol{\gamma}_T^* = \boldsymbol{\gamma}_0 + \boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,T}\boldsymbol{A}_T^{-1}\boldsymbol{b}_T$. Clearly, $\boldsymbol{\gamma}_T^*$ is an interior point of $\boldsymbol{\Gamma}_T(\varepsilon)$ as long as $\varepsilon > 0$. Subsequently select a $\boldsymbol{\gamma}_\varepsilon \in \partial\boldsymbol{\Gamma}_T(\varepsilon)$, i.e., $\boldsymbol{\gamma}_\varepsilon$ is a boundary point of $\boldsymbol{\Gamma}_T(\varepsilon)$. From (A.2), we have

$$\begin{aligned} Q_T(\boldsymbol{\gamma}_\varepsilon) - Q_T(\boldsymbol{\gamma}_T^*) &= \left[Q_T(\boldsymbol{\gamma}_\varepsilon) - Q_T(\boldsymbol{\gamma}_0)\right] - \left[Q_T(\boldsymbol{\gamma}_T^*) - Q_T(\boldsymbol{\gamma}_0)\right] \\ &= \left[\frac{1}{2}\boldsymbol{\mu}_T'\boldsymbol{A}_T\boldsymbol{\mu}_T - \frac{1}{2}\boldsymbol{b}_T'\boldsymbol{A}_T^{-1}\boldsymbol{b}_T + R_T(\bar{\boldsymbol{\gamma}}_\varepsilon,\boldsymbol{\gamma}_0)\right] - \left[-\frac{1}{2}\boldsymbol{b}_T'\boldsymbol{A}_T^{-1}\boldsymbol{b}_T + R_T(\bar{\boldsymbol{\gamma}}_T^*,\boldsymbol{\gamma}_0)\right] \\ &= \frac{1}{2}\boldsymbol{\mu}_T'\boldsymbol{A}_T\boldsymbol{\mu}_T + R_T(\bar{\boldsymbol{\gamma}}_\varepsilon,\boldsymbol{\gamma}_0) - R_T(\bar{\boldsymbol{\gamma}}_T^*,\boldsymbol{\gamma}_0) = \frac{1}{2}\boldsymbol{\mu}_T'\boldsymbol{A}_T\boldsymbol{\mu}_T + o_p(1), \end{aligned}$$

where $\boldsymbol{\mu}_T$ a random vector with $\|\boldsymbol{\mu}_T\| = \varepsilon$, and $\bar{\boldsymbol{\gamma}}_\varepsilon$ is a point on the line segment connecting $\boldsymbol{\gamma}_\varepsilon$ and $\boldsymbol{\gamma}_0$. The point $\bar{\boldsymbol{\gamma}}_T^*$ is defined similarly. Moreover, the final equality follows from

$$R_T(\bar{\boldsymbol{\gamma}},\boldsymbol{\gamma}_0) \le \frac{1}{2}\left\|\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\right)\right\|^2\left\{\sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}_{\delta,k_T}}\left\|\boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,T}\big[\ddot{Q}_T(\boldsymbol{\gamma}) - \ddot{Q}_T(\boldsymbol{\gamma}_0)\big]\boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,T}\right\| + \left\|\boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,T}\ddot{Q}_{T,2}(\boldsymbol{\gamma}_0)\boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,T}\right\|\right\},$$

claim (a), and $\left\|\boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,T}\ddot{Q}_{T,2}(\boldsymbol{\gamma}_0)\boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,T}\right\| \le o_p(1)\left|T^{-1}\sum_{t=1}^T\left(\frac{t}{T}\right)^{\theta_0+1/2}\right| = o_p(1)$. The second part of claim (b), $\mathbb{P}\left(\boldsymbol{A}_\infty > 0\right) = 1$, implies that $\mathbb{P}\big(\frac{1}{2}\boldsymbol{\mu}_T'\boldsymbol{A}_T\boldsymbol{\mu}_T > 0\big) \to 1$ as $T \to \infty$. Therefore, $\mathbb{P}\big(Q_T(\boldsymbol{\gamma}_\varepsilon) > Q_T(\boldsymbol{\gamma}_T^*)\big) \to 1$ for any boundary point $\boldsymbol{\gamma}_\varepsilon$ and the minimum of Q_T must be attained at an interior point of $\boldsymbol{\Gamma}_T(\varepsilon)$, say $\widehat{\boldsymbol{\gamma}}_T(\varepsilon)$. As in Andrews and Sun (2004), we can now construct a sequence $\{\widehat{\boldsymbol{\gamma}}_T\}$ such that $\widehat{\boldsymbol{\gamma}}_T = \widehat{\boldsymbol{\gamma}}_T(J_T^{-1}) \in \boldsymbol{\Gamma}_T(J_T^{-1})$, where $J_T \to \infty$, satisfying the first-order conditions $\mathbb{P}\Big(\dot{Q}_T\big(\widehat{\boldsymbol{\gamma}}_T\big) = 0\Big) \to 1$ as $T \to \infty$. As a result, we obtain

$$\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0) = \boldsymbol{A}_T^{-1}\boldsymbol{b}_T + o_p(1). \tag{A.3}$$

It remains to verify claims (a) and (b). We consider the sequence $\boldsymbol{\Gamma}_{\delta,k_T}$ for $k_T = \tilde{\kappa}\ln T$ and

$\tilde{\kappa} > 0$. There exists $T^* > 0$ such that whenever $T > T^*$,

$$\boldsymbol{\Gamma}_{\delta,k_T} \subset \{\boldsymbol{\gamma} \in \boldsymbol{\Gamma} : \ \|\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)\| \leq \tilde{\kappa} \ln T\} \subset \mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0),$$

where $\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)$ is given in (S.22), and $\kappa = C\tilde{\kappa}$ with some constant $C > 0$. Claim (a) thus holds if $\sup_{\boldsymbol{\gamma} \in \mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)} \left\| \boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,T} \big[\ddot{Q}_T(\boldsymbol{\gamma}) - \ddot{Q}_T(\boldsymbol{\gamma}_0)\big] \boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,T} \right\| = o_p(1)$. Since N is fixed, we can bound $\boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,T} \big[\ddot{Q}_T(\boldsymbol{\gamma}) - \ddot{Q}_T(\boldsymbol{\gamma}_0)\big] \boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,T}$ element-wise. Using the identity $(\boldsymbol{y}_t - \boldsymbol{Z}_t'\boldsymbol{\beta})' \boldsymbol{\imath}_N = N\tau_{g0} t^{\theta_0} - \big(\boldsymbol{\beta} - \boldsymbol{\beta}_0\big)' \boldsymbol{Z}_t \boldsymbol{\imath}_N + \boldsymbol{u}_t' \boldsymbol{\imath}_N$ and Lemma S.3.3, it is easily shown that the supremum of each element is indeed $o_p(1)$. Claim (b) follows directly from the weak convergence results in Lemma S.3.2. That is, $\boldsymbol{A}_T \longrightarrow_d \int \boldsymbol{J}(r;\boldsymbol{\gamma}_0)\boldsymbol{J}(r;\boldsymbol{\gamma}_0)'\, dr$ and $\boldsymbol{b}_T \longrightarrow_d \int \boldsymbol{J}(r;\boldsymbol{\gamma}_0)\, d\boldsymbol{B}_u(r) + \boldsymbol{\mathcal{B}}_{vu}$ as $T \to \infty$. Theorem 1 now follows from (A.3). ■

Proof of Theorem 2 The proof is to a large extent an application of theorem 2 in Jansson (2002). We provide the details in the Supplement. ■

Proof of Theorem 3 We abbreviate $M = M_T$. By simple rearrangements, we obtain

$$\begin{aligned}
&\left\{ \boldsymbol{G}'^{-1}_{\widehat{\boldsymbol{\gamma}}_T,M} \left[\sum_{m=1}^{M} \widehat{\boldsymbol{J}}\big(m;\widehat{\boldsymbol{\gamma}}_T\big) \widehat{\boldsymbol{J}}\big(m;\widehat{\boldsymbol{\gamma}}_T\big)' \right] \boldsymbol{G}^{-1}_{\widehat{\boldsymbol{\gamma}}_T,M} \right\}^{-1} \left\{ \boldsymbol{G}'^{-1}_{\widehat{\boldsymbol{\gamma}}_T,M} \left[\sum_{m=1}^{M} \widehat{\boldsymbol{J}}\big(m;\widehat{\boldsymbol{\gamma}}_T\big)\, \widehat{\boldsymbol{\mu}}_m \right] + \widehat{\boldsymbol{\mathcal{B}}}^{-}_{vu} \right\} \\
&= \boldsymbol{S}_M^{-1} \left\{ \boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,M} \left[\sum_{m=1}^{M} \widehat{\boldsymbol{J}}\big(m;\boldsymbol{\gamma}_0\big) \widehat{\boldsymbol{J}}\big(m;\boldsymbol{\gamma}_0\big)' \right] \boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,M} + \boldsymbol{R}_{1,M} \right\}^{-1} \\
&\quad \times \left\{ \boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,M} \left[\sum_{m=1}^{M} \widehat{\boldsymbol{J}}\big(m;\boldsymbol{\gamma}_0\big)\, \widehat{\boldsymbol{\mu}}_m \right] + \boldsymbol{S}'^{-1}_M \widehat{\boldsymbol{\mathcal{B}}}^{-}_{vu} + \boldsymbol{R}_{2,M} \right\},
\end{aligned} \tag{A.4}$$

while having defined $\widehat{\boldsymbol{J}}\big(m;\boldsymbol{\gamma}_0\big) = \Big[\tau_{g0}\, m^{\theta_0} \ln m\, \boldsymbol{\imath}_N, m^{\theta_0} \boldsymbol{\imath}_N, \widehat{\mathcal{Z}}_m'\Big]'$ and the quantities

(a) $\boldsymbol{S}_M := \boldsymbol{G}_{\boldsymbol{\gamma}_0,M} \boldsymbol{G}^{-1}_{\widehat{\boldsymbol{\gamma}}_T,M} = \begin{bmatrix} M^{\theta_0 - \widehat{\theta}_T} & 0 & 0 \\ (\tau_{g0} - \widehat{\tau}_{g,T}) M^{\theta_0 - \widehat{\theta}_T} \ln M & M^{\theta_0 - \widehat{\theta}_T} & \boldsymbol{0}_{1\times p} \\ \boldsymbol{0}_{p\times 1} & \boldsymbol{0}_{p\times 1} & \boldsymbol{I}_p \end{bmatrix}$,

(b) $\boldsymbol{R}_{1,M} = \boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,M} \sum_{m=1}^{M} \big[\widehat{\boldsymbol{J}}\big(m;\widehat{\boldsymbol{\gamma}}_T\big) \widehat{\boldsymbol{J}}\big(m;\widehat{\boldsymbol{\gamma}}_T\big)' - \widehat{\boldsymbol{J}}\big(m;\boldsymbol{\gamma}_0\big) \widehat{\boldsymbol{J}}\big(m;\boldsymbol{\gamma}_0\big)'\big] \boldsymbol{G}^{-1}_{\boldsymbol{\gamma}_0,M}$,

(c) $\boldsymbol{R}_{2,M} = \boldsymbol{G}'^{-1}_{\boldsymbol{\gamma}_0,M} \sum_{m=1}^{M} \big[\widehat{\boldsymbol{J}}\big(m;\widehat{\boldsymbol{\gamma}}_T\big) - \widehat{\boldsymbol{J}}\big(m;\boldsymbol{\gamma}_0\big)\big]\, \widehat{\boldsymbol{\mu}}_m$.

(a) We have $\boldsymbol{S}_M \longrightarrow_p \boldsymbol{I}_{p+2}$. To see this, note that $M^{|\widehat{\theta}_T - \theta_0|} = \exp\big(\big(\frac{\ln M}{T^{\theta_0+1/2}} O_p(1)\big) \longrightarrow_p 1$ and $\big|\widehat{\tau}_{g,T} - \tau_{g0}\big| \ln M = \big|\frac{T^{\theta_0+1/2}}{\ln T}\big(\widehat{\tau}_{g,T} - \tau_{g0}\big)\big| \frac{\ln T \ln M}{T^{\theta_0+1/2}} = o_p(1)$.

(b) Looking at the elements of $\boldsymbol{R}_{1,M}$, we conclude that $\boldsymbol{R}_{1,M} = o_p^*(1)$ if results similar to those in Lemma S.3.3(i)-(iii) continue to hold. Two conditions need to be verified:

(b1) $\mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_T \in \mathcal{N}_{\kappa,M}(\boldsymbol{\gamma}_0)\right) \to 1$ with $\mathcal{N}_{\kappa,M}(\boldsymbol{\gamma}_0)$ similarly defined to (S.22),

(b2) the stochastic order of terms remains the same when replacing $\boldsymbol{Z}_m$ by $\widehat{\mathcal{Z}}_m$, conditional on the sample $(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_T, \boldsymbol{y}_T)$.

For condition (b1), using set inclusions similar to those below (A.3), it suffices to show $\widehat{\boldsymbol{\gamma}}_T \in \left\{\boldsymbol{\gamma} \in \boldsymbol{\Gamma} : \left\|\boldsymbol{G}_{\boldsymbol{\gamma}_0,M}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\right\| \leq \tilde{\kappa}\ln M\right\}$ with large probability for some $\tilde{\kappa} > 0$. This is trivial by $\left\|\boldsymbol{G}_{\boldsymbol{\gamma}_0,M}(\widehat{\boldsymbol{\gamma}}_T-\boldsymbol{\gamma}_0)\right\| \leq \left\|\boldsymbol{G}_{\boldsymbol{\gamma}_0,M}\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}^{-1}\right\| \left\|\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}(\widehat{\boldsymbol{\gamma}}_T-\boldsymbol{\gamma}_0)\right\| = O_p(1)$, where $\boldsymbol{G}_{\boldsymbol{\gamma}_0,M}\boldsymbol{G}_{\boldsymbol{\gamma}_0,T}^{-1} = O(1)$. Continuing with (b2), by independence between $\{\boldsymbol{e}_m\}$ and $\left\{\widehat{\boldsymbol{\Omega}}_T, \widehat{\boldsymbol{\Delta}}_{vu}^{-}\right\}$, the consistency of $\widehat{\boldsymbol{\Omega}}_T$, and a FCLT for an i.i.d. sequence, we may have

$$\frac{1}{\sqrt{M}}\sum_{m=1}^{[rM]}\begin{bmatrix}\widehat{\boldsymbol{\mu}}_m\\ \widehat{\boldsymbol{v}}_m\end{bmatrix} = \widehat{\boldsymbol{\Omega}}_T^{1/2}\frac{1}{\sqrt{M}}\sum_{m=1}^{[rM]}\boldsymbol{e}_n \longrightarrow_{d^*} \boldsymbol{B}(r), \tag{A.5}$$

in probability, c.f. Section 2 of Park (2002). Since $\left\{\widehat{\mathcal{Z}}_m\right\}$ contains partial sum processes of $\{\widehat{\boldsymbol{v}}_m\}$, its integer powers and deterministic terms, (b2) is satisfied.

(c) We have $\left\|\boldsymbol{R}_{2,M}\right\| \leq C\sum_{j=1}^{4}\left|\boldsymbol{R}_{2,M,j}\right|$ where

$$\begin{aligned}\left|\boldsymbol{R}_{2,M,1}\right| &= M^{-1/2}\left|\sum_{m=1}^{M} M^{-\theta_0}\left(m^{\widehat{\theta}_T}-m^{\theta_0}\right)\widehat{\boldsymbol{\mu}}_m'\boldsymbol{\imath}_N\right| = O_p^*\left(\frac{\ln M}{T^{\theta_0+1/2}}\right) = o_p^*(1),\\ \left|\boldsymbol{R}_{2,M,2}\right| &= M^{-1/2}\left|\sum_{m=1}^{M} M^{-\theta_0}\left(m^{\widehat{\theta}_T}-m^{\theta_0}\right)\ln\frac{m}{M}\widehat{\boldsymbol{\mu}}_m'\boldsymbol{\imath}_N\right| = T^{-(\theta_0+1/2)}O_p^*(1) = o_p^*(1),\end{aligned} \tag{A.6}$$

$\left|\boldsymbol{R}_{2,M,3}\right| = M^{-1/2}\left|\widehat{\tau}_{g,T}-\tau_{g0}\right|\left|\sum_{m=1}^{M} M^{-\theta_0}\left(m^{\widehat{\theta}_T}-m^{\theta_0}\right)\ln m\,\widehat{\boldsymbol{\mu}}_m'\boldsymbol{\imath}_N\right| = O_p^*\left(\frac{\ln T(\ln M)^2}{T^{2\theta_0+1}}\right) = o_p^*(1)$, and $\left|\boldsymbol{R}_{2,M,4}\right| = M^{-1/2}\left|\widehat{\tau}_{g,T}-\tau_{g0}\right|\left|\sum_{m=1}^{M}\left(\frac{m}{M}\right)^{\theta_0}\ln m\,\widehat{\boldsymbol{\mu}}_m'\boldsymbol{\imath}_N\right| = O_p^*\left(\frac{\ln T\ln M}{T^{\theta_0+1/2}}\right) = o_p^*(1)$. All these stochastic orders are a consequence of (A.5) and a straightforward modification of Lemma S.3.2. Overall, we have $\boldsymbol{R}_{2,M} = o_p^*(1)$.

It remains to look at the leading terms in (A.4). The elements of $\widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\Delta}}$ are always multiplicative in the construction. From $\boldsymbol{S}_M \longrightarrow_p \boldsymbol{I}_{p+2}$, (A.5), and Lemma S.3.2, we have

$$\boldsymbol{G}_{\boldsymbol{\gamma}_0,M}'^{-1}\left[\sum_{m=1}^{M}\widehat{\boldsymbol{J}}(m;\boldsymbol{\gamma}_0)\,\widehat{\boldsymbol{\mu}}_m\right]+\boldsymbol{S}_M'^{-1}\widehat{\boldsymbol{\mathcal{B}}}_{vu}^{-} \longrightarrow_{d^*} \int_0^1 \boldsymbol{J}(r;\boldsymbol{\gamma}_0)\,d\boldsymbol{B}_u(r)+\begin{bmatrix}\boldsymbol{0}_{2\times 1}\\ \boldsymbol{\Omega}_{v_1u_1}\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{\Omega}_{v_Nu_N}\boldsymbol{b}_N\end{bmatrix}+\begin{bmatrix}\boldsymbol{0}_{2\times 1}\\ \boldsymbol{\Delta}_{v_1u_1}^{-}\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{\Delta}_{v_Nu_N}^{-}\boldsymbol{b}_N\end{bmatrix}, \tag{A.7}$$

in probability. The last two terms in (A.7) equal $\boldsymbol{\mathcal{B}}_{vu}$, because $\boldsymbol{\Omega}+\boldsymbol{\Delta}^{-} = (\boldsymbol{\Delta}+\boldsymbol{\Delta}'-\boldsymbol{\Sigma})+(\boldsymbol{\Sigma}-\boldsymbol{\Delta}') = \boldsymbol{\Delta}$. Similarly,

$$\boldsymbol{G}_{\boldsymbol{\gamma}_0,M}'^{-1}\left[\sum_{m=1}^{M}\widehat{\boldsymbol{J}}(m;\boldsymbol{\gamma}_0)\widehat{\boldsymbol{J}}(m;\boldsymbol{\gamma}_0)'\right]\boldsymbol{G}_{\boldsymbol{\gamma}_0,M}^{-1} \longrightarrow_{d^*} \int_0^1 \boldsymbol{J}(r;\boldsymbol{\gamma}_0)\boldsymbol{J}(r;\boldsymbol{\gamma}_0)'\,dr, \tag{A.8}$$

in probability. The theorem follows after combining the limiting distribution of these leading terms

through (A.4). ■

Proof of Theorem 4 The proof follows from a functional central limit theorem for linear processes and the continuous mapping theorem. See the Supplement for further details. ■

Supplemental Appendix to:

Cointegrating Polynomial Regressions With Power Law Trends: Environmental Kuznets Curve or Omitted Time Effects?

Yicong Lin[1,2] and Hanno Reuvers[3]

[1]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam

[2]Tinbergen Institute

[3]Department of Econometrics, Erasmus University Rotterdam

December, 2022

Abstract

This document contains further details on our simulation study, mathematical proofs, and empirical application.

(1) Section S1: the simulation DGP used in the section Introduction.

(2) Section S2: the derivation of limiting distribution in Example 3.1.

(3) Section S3: proofs for some auxiliary lemmas that are used to show the main theorems.

(4) Section S4: further details on Theorems 3.2 and 3.4.

(5) Section S5: details on the simulation DGPs 2(a)–2(c).

(6) Section S6: additional simulation results.

(7) Section S7: further empirical results including a detailed study on univariate models.

(8) Section S8: some discussion on the invalidity of fully modified OLS estimators.

S1 Simulation DGP Used for Introduction

The simulation DGPs of the introduction are based on the data for Austria, Belgium and Finland. Parameter values are (nonlinear) least squares estimates and innovations are mean-zero normally distributed random variables with a covariance matrix estimated from the residuals and $\Delta \boldsymbol{x}_t$. The specific parametrization for the *model with global trend* is

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \tau_g t^{2.21} \boldsymbol{\iota}_3 - \begin{bmatrix} 8.89 \\ 4.16 \\ 16.39 \end{bmatrix} + \begin{bmatrix} 0.0017 \\ 0.0122 \\ 0.0108 \end{bmatrix} t + \begin{bmatrix} 2.015 x_{1,t} \\ 1.477 x_{2,t} \\ 2.703 x_{3,t} \end{bmatrix} + \boldsymbol{u}_t, \tag{S.1}$$

where $[\boldsymbol{u}_t' \ \Delta \boldsymbol{x}_t']' \sim \mathrm{N}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}})$ with

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 18.86 & * & * & * & * & * \\ 1.35 & 2.02 & * & * & * & * \\ 3.68 & 2.65 & 1.88 & * & * & * \\ 0.10 & 0.38 & 1.46 & 0.83 & * & * \\ 0.45 & 0.09 & 0.22 & 0.07 & 0.18 & * \\ 0.26 & 0.14 & 0.56 & 0.15 & 0.14 & 0.24 \end{bmatrix} \times 10^{-2}.$$

The simulations investigating the influence of the redundant trend follow

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} -1.01 \\ 8.64 \\ -5.15 \end{bmatrix} + \begin{bmatrix} -0.0111 \\ 0.0058 \\ 0.0163 \end{bmatrix} t + \begin{bmatrix} 1.103 x_{1,t} \\ -0.001 x_{2,t} \\ 1.232 x_{3,t} \end{bmatrix} + \phi_2 \begin{bmatrix} x_{1,t}^2 \\ x_{2,t}^2 \\ x_{3,t}^2 \end{bmatrix} + \boldsymbol{u}_t,$$

where $[\boldsymbol{u}_t' \ \Delta \boldsymbol{x}_t']' \sim \mathrm{N}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}})$ with

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 18.57 & * & * & * & * & * \\ 4.03 & 4.02 & * & * & * & * \\ 11.82 & 9.90 & 35.88 & * & * & * \\ 0.56 & 0.64 & 1.79 & 0.83 & * & * \\ 0.47 & 0.20 & 0.45 & 0.07 & 0.18 & * \\ 0.43 & 0.35 & 0.91 & 0.15 & 0.14 & 0.24 \end{bmatrix} \times 10^{-2}.$$

Compared to the empirical application (and thus also DGP2 in Section 4), the main differences are the smaller N and the omission of serial correlation in both the innovations and the increments of the integrated variables. These modifications allow us to showcase the influence of the omitted global trend while not having to worry about the effects of long-run covariance estimation on statistical size.

S2 Limiting Distribution for Example 3.1

Invoking Theorem 3.1, we have

$$\begin{bmatrix} T^{\theta_0+\frac{1}{2}} & \\ T^{\theta_0+\frac{1}{2}}\tau_0\ln(T) & T^{\theta_0+\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \widehat{\theta}_T-\theta_0 \\ \widehat{\tau}_T-\tau_0 \end{bmatrix} \longrightarrow_d \begin{bmatrix} \int\left(\tau_0 r^{\theta_0}\ln(r)\right)^2 dr & \int \tau_0 r^{2\theta_0}\ln(r)dr \\ \int \tau_0 r^{2\theta_0}\ln(r)dr & \int r^{2\theta_0}dr \end{bmatrix}^{-1}\begin{bmatrix} \int \tau_0 r^{\theta_0}\ln(r)dB_u \\ \int r^{\theta_0}dB_u \end{bmatrix}.$$

It remains to show that the quantity in the RHS is normally distributed with a mean and variance as in the example. Consider an arbitrary vector $\boldsymbol{c}=[c_1,c_2]'$ and define

$$A_{\boldsymbol{c}}=\boldsymbol{c}'\begin{bmatrix} \int \tau_0 r^{\theta_0}\ln(r)dB_u \\ \int r^{\theta_0}dB_u \end{bmatrix}=\int\left[c_1\tau_0 r^{\theta_0}\ln(r)+c_2 r^{\theta_0}\right]dB_u \overset{d}{=}\Omega_{uu}^{1/2}\int\left[c_1\tau_0 r^{\theta_0}\ln(r)+c_2 r^{\theta_0}\right]dW_u.$$

Gaussianity is preserved under mean square integration (see, e.g., section 4.6 in Soong (1973)) and we proceed to the mean and variance of $A_{\boldsymbol{c}}$. From (4.190) in the same reference, it is not hard to obtain $\mathbb{E}(A_{\boldsymbol{c}})=\Omega_{uu}^{1/2}\int\left[c_1\tau_0 r^{\theta_0}\ln(r)+c_2 r^{\theta_0}\right]d\mathbb{E}\big(W_u\big)=0$. Moreover, (2.16) in Tanaka (2017) yields

$$\mathbb{V}\mathrm{ar}\left(A_{\boldsymbol{c}}\right)=\Omega_{uu}\int\left[c_1\tau_0 r^{\theta_0}\ln(r)+c_2 r^{\theta_0}\right]^2 dr=\Omega_{uu}\boldsymbol{c}'\begin{bmatrix} \int\left(\tau_0 r^{\theta_0}\ln(r)\right)^2 dr & \int \tau_0 r^{2\theta_0}\ln(r)dr \\ \int \tau_0 r^{2\theta_0}\ln(r)dr & \int r^{2\theta_0}dr \end{bmatrix}\boldsymbol{c}.$$

Our choice of $\boldsymbol{c}$ was arbitrary and thus $\left[\begin{smallmatrix} \int \tau_0 r^{\theta_0}\ln(r)dB_u \\ \int r^{\theta_0}dB_u \end{smallmatrix}\right]\sim \mathrm{N}\left(\mathbf{0},\Omega_{uu}\left[\begin{smallmatrix} \int\left(\tau_0 r^{\theta_0}\ln(r)\right)^2 dr & \int \tau_0 r^{2\theta_0}\ln(r)dr \\ \int \tau_0 r^{2\theta_0}\ln(r)dr & \int r^{2\theta_0}dr \end{smallmatrix}\right]\right)$. Finally, use $\int\left(r^{\theta_0}\ln(r)\right)^2 dr=\frac{2}{(2\theta_0+1)^3}$, $\int r^{2\theta_0}\ln(r)dr=-\frac{1}{(2\theta_0+1)^2}$, and basic linear algebra to recover the result.

S3 Auxiliary Lemmas

Lemma S.3.1. *(i) For $a_L>-1$, we have $\sup_{a\geq a_L}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a\right|\leq C$.*

(ii) Under Assumption 3.2, for any $a\geq a_L>-\frac{1}{2}$, and any $k\geq 0$, $\mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a(\ln t)^k u_{i,t}\right)^2\leq C(\ln T)^{2k}$, $i=1,\ldots,N$.

(iii) Under Assumption 3.2, for some a_L and a_U such that $-\frac{1}{2}<a_L<a_U<\infty$, and any $k\geq 0$, $\mathbb{E}\left(\sup_{a\in[a_L,a_U]}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a(\ln t)^k u_{i,t}\right|\right)\leq C(\ln T)^k$, $i=1,\ldots,N$.

(iv) If a_L and a_U satisfy $-1<a_L<a_U<\infty$, and if $k=0,1,2,\ldots$, then

$$\sup_{a\in[a_L,a_U]}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k-\int_0^1 r^a(\ln r)^k dr\right|\leq C\frac{(\ln T)^{k+1}}{T^{1+\min(a_L,0)}}.$$

Proof ***(i)*** This is shown in lemma 4 of Robinson (2012). ***(ii)*** Note that

$$\begin{aligned}\mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a}(\ln t)^{k}u_{i,t}\right)^{2} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{t}{T}\right)^{a}\left(\frac{s}{T}\right)^{a}(\ln t)^{k}(\ln s)^{k}\,\mathbb{E}(u_{i,t}u_{i,s}) \\ &\leq \frac{(\ln T)^{2k}}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\frac{t}{T}\right)^{a}\left(\frac{s}{T}\right)^{a}\left|\mathbb{E}(u_{i,t}u_{i,s})\right| \leq 2\frac{(\ln T)^{2k}}{T}\sum_{t=1}^{T}\sum_{s=0}^{t-1}\left(\frac{t}{T}\right)^{a}\left(\frac{t-s}{T}\right)^{a}|\gamma_{i,s}|,\end{aligned} \tag{S.1}$$

where we define $\gamma_{i,s}=\mathbb{E}(u_{i,t}u_{i,t-s})$. For the given index ranges, we also have $|t-s|\leq t$ such that

$$\mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a}(\ln t)^{k}u_{i,t}\right)^{2} \leq 2(\ln T)^{2k}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2a}\sum_{s=0}^{\infty}|\gamma_{i,s}|. \tag{S.2}$$

The first summation in the RHS of (S.2) is bounded in view of Lemma S.3.1 *(i)* and $\sum_{s=0}^{\infty}|\gamma_{i,s}|<\infty$ due to Assumption 3.2(a) (cf. Appendix 3.A. in Hamilton (1994)). ***(iii)*** Using the equality $\frac{t}{T}=\sum_{s=0}^{t-1}\left[\left(\frac{s+1}{T}\right)^{a}-\left(\frac{s}{T}\right)^{a}\right]$ and a change in the order of summation, we find

$$\begin{aligned}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a}(\ln t)^{k}u_{i,t} &= \sum_{t=1}^{T}\sum_{s=0}^{t-1}\left[\left(\frac{s+1}{T}\right)^{a}-\left(\frac{s}{T}\right)^{a}\right](\ln t)^{k}u_{i,t} = \sum_{s=0}^{T-1}\left[\left(\frac{s+1}{T}\right)^{a}-\left(\frac{s}{T}\right)^{a}\right]\sum_{t=s+1}^{T}(\ln t)^{k}u_{i,t} \\ &= \left(\frac{1}{T}\right)^{a}\sum_{t=1}^{T}(\ln t)^{k}u_{i,t}+\sum_{s=1}^{T-1}\left[\left(\frac{s+1}{T}\right)^{a}-\left(\frac{s}{T}\right)^{a}\right]\left(\sum_{t=1}^{T}(\ln t)^{k}u_{i,t}-\sum_{t=1}^{s}(\ln t)^{k}u_{i,t}\right) \\ &= \left(\frac{1}{T}\right)^{a}\sum_{t=1}^{T}(\ln t)^{k}u_{i,t}+\sum_{t=1}^{T}(\ln t)^{k}u_{i,t}-\left(\frac{1}{T}\right)^{a}\sum_{t=1}^{T}(\ln t)^{k}u_{i,t}-\sum_{s=1}^{T-1}\left[\left(\frac{s+1}{T}\right)^{a}-\left(\frac{s}{T}\right)^{a}\right]\sum_{t=1}^{s}(\ln t)^{k}u_{i,t},\end{aligned}$$

and hence

$$\begin{aligned}\mathbb{E}\left(\sup_{a\in[a_L,a_U]}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a}(\ln t)^{k}u_{i,t}\right|\right) &\leq \mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ln t)^{k}u_{i,t}\right| \\ &\quad+\mathbb{E}\left(\sup_{a\in[a_L,a_U]}\left|\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left[\left(\frac{s+1}{T}\right)^{a}-\left(\frac{s}{T}\right)^{a}\right]\sum_{t=1}^{s}(\ln t)^{k}u_{i,t}\right|\right).\end{aligned} \tag{S.3}$$

For the first term in the RHS of (S.3), we have $\mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ln t)^{k}u_{i,t}\right| \leq \left(\mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ln t)^{k}u_{i,t}\right)^{2}\right)^{1/2} \leq C(\ln T)^{k}$ by Lemma S.3.1(ii) with $a=0$. For the second term, note that

$$\left|\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left[\left(\frac{s+1}{T}\right)^{a}-\left(\frac{s}{T}\right)^{a}\right]\sum_{t=1}^{s}(\ln t)^{k}u_{i,t}\right| \leq \frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left(\frac{s}{T}\right)^{a}\left|\left(1+\frac{1}{s}\right)^{a}-1\right|\left|\sum_{t=1}^{s}(\ln t)^{k}u_{i,t}\right|. \tag{S.4}$$

To deal with the supremum of $\left|\left(1+\frac{1}{s}\right)^{a}-1\right|$ over $[a_L,a_U]$, we define $g_a(x)=(1+x)^{a}-1$ for $0\leq x\leq 1$. If $-\frac{1}{2}<a\leq 1$, then $|g_a(x)|\leq|a|x$ by Bernoulli's inequality. If $a\geq 1$, then convexity of $g_a(x)$ implies

$$g_a(x)\leq(1-x)g_a(0)+xg_a(1)\leq(2^{a}-1)\,x.$$

We conclude that $|g_a(x)| \leq Cx$ for all $a_L \leq a \leq a_U$ and $x \in [0,1]$. Combining this result with (S.4), we have

$$\begin{aligned}
&\mathbb{E}\left(\sup_{a\in[a_L,a_U]}\left|\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left[\left(\frac{s+1}{T}\right)^a-\left(\frac{s}{T}\right)^a\right]\sum_{t=1}^{s}(\ln t)^k u_{i,t}\right|\right)\\
&\leq \mathbb{E}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left(\frac{s}{T}\right)^{a_L}\sup_{a\in[a_L,a_U]}\left|\left(1+\frac{1}{s}\right)^a-1\right|\left|\sum_{t=1}^{s}(\ln t)^k u_{i,t}\right|\right) \leq C\frac{1}{\sqrt{T}}\sum_{s=1}^{T-1}\left(\frac{s}{T}\right)^{a_L}\frac{1}{s}\,\mathbb{E}\left|\sum_{t=1}^{s}(\ln t)^k u_{i,t}\right|\\
&\leq CT^{-(a_L+1/2)}\sum_{s=1}^{T-1}s^{a_L-1/2}(\ln s)^k \leq C(\ln T)^k\left[\frac{1}{T}\sum_{s=1}^{T}\left(\frac{s}{T}\right)^{a_L-1/2}\right] \leq C(\ln T)^k,
\end{aligned}$$

where we used $\mathbb{E}\left|\sum_{t=1}^{s}(\ln t)^k u_{i,t}\right| \leq \left(\mathbb{E}\left(\sum_{t=1}^{s}(\ln t)^k u_{i,t}\right)^2\right)^{1/2} \leq Cs^{1/2}(\ln s)^k$ (the steps in the proof of (ii) require a small modification to establish this) to go to the last line, and (i) to obtain the final inequality. The proof is complete since we have bounded both terms in the RHS of (S.3). ***(iv)*** If we divide the integral into integration intervals of width $\frac{1}{T}$, then we find

$$\begin{aligned}
&\sup_{a\in[a_L,a_U]}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k-\int_0^1 r^a(\ln r)^k dr\right|\\
&=\sup_{a\in[a_L,a_U]}\left|\sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k dr-\sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}r^a(\ln r)^k dr\right|\\
&=\sup_{a\in[a_L,a_U]}\left|\frac{1}{T}\left(\frac{1}{T}\right)^a\left(\ln\frac{1}{T}\right)^k-\int_0^{1/T}r^a(\ln r)^k dr+\sum_{t=2}^{T}\int_{(t-1)/T}^{t/T}\left[\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k-r^a(\ln r)^k\right]dr\right|\\
&\leq \sup_{a\in[a_L,a_U]}\left|\left(\frac{1}{T}\right)^{a+1}\left(\ln\frac{1}{T}\right)^k\right|+\sup_{a\in[a_L,a_U]}\left|\int_0^{1/T}r^a(\ln r)^k dr\right|+\sup_{a\in[a_L,a_U]}\sum_{t=2}^{T}\int_{(t-1)/T}^{t/T}\left|\left(\frac{t}{T}\right)^a\left(\ln\frac{t}{T}\right)^k-r^a(\ln r)^k\right|\\
&=: Ia+Ib+Ic, \qquad\qquad (S.5)
\end{aligned}$$

using the triangle inequality. Clearly, Ia is bounded by $T^{-(a_L+1)}(\ln T)^k$. For Ib we can use the standard integral (cf. Adams and Essex (2016)), namely $\int_0^{1/T} r^a(\ln r)^k dr = \frac{(-1)^k}{a+1}\left(\frac{1}{T}\right)^{a+1}(\ln T)^k - \frac{k}{a+1}\int_0^{1/T} r^a(\ln r)^{k-1}dr$ for $k \neq -1$, to obtain

$$\begin{aligned}
\int_0^{1/T} r^a(\ln r)^k dr &= (-1)^k\left(\frac{1}{T}\right)^{a+1}\sum_{j=0}^{k-1}\frac{k!}{(k-j)!}\frac{1}{(a+1)^{1+j}}(\ln T)^{k-j}+(-1)^k\frac{k!}{(a+1)^k}\int_0^{1/T}r^a dr\\
&= (-1)^k\left(\frac{1}{T}\right)^{a+1}\sum_{j=0}^{k}\frac{k!}{(k-j)!}\frac{1}{(a+1)^{1+j}}(\ln T)^{k-j}.
\end{aligned}$$

We therefore conclude that

$$\begin{aligned}
Ib &\leq \sum_{j=1}^{k}\frac{k!}{(k-j)!}\sup_{a\in[a_L,a_U]}\frac{1}{(a+1)^{1+j}}\left(\frac{1}{T}\right)^{a+1}(\ln T)^{k-j} \leq \sum_{j=1}^{k}\frac{k!}{(k-j)!}\frac{1}{(a_L+1)^{1+j}}\left(\frac{1}{T}\right)^{a_L+1}(\ln T)^{k-j}\\
&\leq CT^{-(a_L+1)}(\ln T)^k.
\end{aligned}$$

It remains to bound the term Ic. Changing the integration variable to $r = \frac{t}{T} - s$ yields

$$Ic = \sup_{a\in[a_L,a_U]} \sum_{t=2}^{T} \int_0^{1/T} \left| \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k - \left(\frac{t}{T} - s\right)^a \left[\ln\left(\frac{t}{T} - s\right)\right]^k \right| ds. \tag{S.6}$$

We subsequently derive an upper bound for the integrand using an approach which mimics the derivations in (D.14) and (D.15) in Robinson (2012). For any $\frac{2}{T} \leq \ell \leq 1$ (such that $0 < s/\ell \leq \frac{1}{2}$), we have

$$\begin{aligned} \left|\ell^a(\ln \ell)^k - (\ell - s)^a\big(\ln(\ell - s)\big)^k\right| &= \left|\big[\ell^a - (\ell - s)^a\big](\ln \ell)^k + (\ell - s)^a\big[(\ln \ell)^k - (\ln(\ell - s))^k\big]\right| \\ &\leq \left|\big[\ell^a - (\ell - s)^a\big](\ln \ell)^k\right| + \left|(\ell - s)^a\Big[(\ln \ell)^k - (\ln(\ell - s))^k\Big]\right| \\ &= \ell^a\big|\ln \ell\big|^k\big|1 - (1 - s/\ell)^a\,\big| + \ell^a(1 - s/\ell)^a\left|(\ln \ell)^k - (\ln(\ell - s))^k\right| =: IIa + IIb, \end{aligned} \tag{S.7}$$

by the triangle inequality and the fact that $\big|(\ell - s)^a\big| = (\ell - s)^a$. For IIa similar arguments as those found below (S.4) give $\big|1 - (1 - x)^a\big| \leq Cx$, and hence

$$IIa \leq C\ell^{a_L}\big|\ln \ell\big|^k \frac{s}{\ell} \leq C\ell^{a_L - 1}\big|\ln \ell\big|^k s \leq C\ell^{a_L - 1}\big|\ln \ell\big|^k \frac{1}{T} \leq C\ell^{a_L - 1}\big(\ln T\big)^k \frac{1}{T}, \tag{S.8}$$

since $|\ln \ell| \leq |\ln T|$ for all $\frac{2}{T} \leq \ell \leq 1$. For IIb we first note that $\frac{1}{2} \leq 1 - s/\ell < 1$ and therefore $(1 - s/\ell)^a < (1 - s/\ell)^{-1} \leq 2$. Moreover, we use the factorization $p^n - q^n = (p - q)\sum_{j=0}^{n-1} p^{n-1-j}q^j$ to obtain[1]

$$\begin{aligned} \left|(\ln \ell)^k - (\ln(\ell - s))^k\right| &= \big|\ln \ell - \ln(\ell - s)\big| \left|\sum_{j=0}^{k-1} (\ln \ell)^{k-1-j}\big(\ln(\ell - s)\big)^j\right| \\ &= \big|\ln(1 - s/\ell)\big| \left|\sum_{j=0}^{k-1} (\ln \ell)^{k-1-j}\big(\ln(\ell - s)\big)^j\right| \leq \big|\ln(1 - s/\ell)\big| \sum_{j=0}^{k-1} |\ln \ell|^{k-1-j}\,|\ln(\ell - s)|^j \\ &\leq k\,|\ln(1 - s/\ell)|\,(\ln T)^{k-1} \leq 2k\frac{s}{\ell}\big(\ln T\big)^{k-1}, \end{aligned} \tag{S.9}$$

because $1/T \leq \ell - s < 1$ and thus $|\ln(\ell - s)| \leq \ln T$. Combining all previous results for IIb gives

$$IIb \leq C\ell^a \frac{s}{\ell}\big(\ln T\big)^{k-1} \leq C\ell^{a_L - 1}\big(\ln T\big)^{k-1}\frac{1}{T}.$$

Since $\frac{2}{T} \leq \ell \leq 1$, we use the bounds on IIa and IIb to bound the integrand of (S.6) as follows:

$$Ic \leq C \sup_{a\in[a_L,a_U]} \sum_{t=2}^{T} \int_0^{1/T} \left(\frac{t}{T}\right)^{a_L - 1} \frac{1}{T}(\ln T)^k ds \leq C\frac{(\ln T)^k}{T^2} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^{a_L - 1}.$$

The asymptotic order of $\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a_L - 1}$ relies on the values of a_L. We distinguish three cases: (1) if $a_L < 0$, then $\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a_L - 1} = T^{1-a_L}\sum_{t=1}^{T}\frac{1}{t^{1-a_L}} = T^{1-a_L}O(1)$, (2) if $a_L = 0$, then $\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a_L - 1} = T\sum_{t=1}^{T} t^{-1} =$

[1] For any $x > -1$, we have the inequality $\frac{x}{1+x} \leq \ln(1 + x) \leq x$. This implies that $|\ln(1 - s/\ell)| = -\ln\big(1 - s/\ell\big) \leq \frac{s/\ell}{1 - s/\ell} \leq 2\frac{s}{\ell}$.

$TO(\ln T)$, and (3) if $a_L > 0$, $\sum_{t=1}^{T} \left(\frac{t}{T}\right)^{a_L - 1} = O(T)$ by Lemma S.3.1(i). Overall, we have

$$Ic \leq C \frac{(\ln T)^k}{T^2} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^{a_L - 1} = O\left(\frac{(\ln T)^k}{T^{a_L+1}} \mathbb{1}_{\{a_L<0\}} + \frac{(\ln T)^{k+1}}{T} \mathbb{1}_{\{a_L=0\}} + \frac{(\ln T)^k}{T} \mathbb{1}_{\{a_L>0\}} \right). \tag{S.10}$$

It is seen that Ia, Ib, and Ic converge to zero as $T \to \infty$. The proof follows from (S.5). ■

Lemma S.3.2. *Let Assumption 3.2 hold. For any a such that $-\frac{1}{2} < a_L \leq a \leq a_U < \infty$, any $j \in \{1, 2, \ldots, p_i\}$, $i \in \{1, 2, \ldots, N\}$, and $k \in \{0, 1, 2, \ldots\}$, as $T \to \infty$, we have:*

(i) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{x_{i,t}}{\sqrt{T}}\right)^j u_{i,t} \longrightarrow_d \int_0^1 B_{v_i}^j(r) dB_{u_i}(r) + j \boldsymbol{\Delta}_{v_i u_i} \int_0^1 B_{v_i}^{j-1}(r) dr$,

(ii) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k u_{i,t} \longrightarrow_d \int_0^1 r^a (\ln r)^k dB_{u_i}(r)$,

(iii) $\frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k \left(\frac{x_{i,t}}{\sqrt{T}}\right)^j \longrightarrow_d \int_0^1 r^a (\ln r)^k B_{v_i}^j(r) dr$.

Proof For $r \in (0, 1]$, we define $f(r) = r^a (\ln r)^k$. Two partial sum processes are defined as $S_{i,T}(r) = \frac{1}{\sqrt{T}} \sum_{s=1}^{[rT]} u_{i,s}$, and $X_{i,T}(r) = \frac{1}{\sqrt{T}} x_{i,[rT]} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} v_{i,t}$. Finally, set $f_T(r) = \left(\frac{[rT]}{T}\right)^a \left(\ln \frac{[rT]}{T}\right)^k$ for $r \in \left[\frac{1}{T}, 1\right]$. ***(i)*** This result follows from lemma 1 of Hong and Phillips (2010). ***(ii)*** We have

$$\begin{aligned} \frac{1}{\sqrt{T}} \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k u_{i,t} &= f_T\left(\frac{t}{T}\right) \frac{u_{i,t}}{\sqrt{T}} = f_T\left(\frac{t}{T}\right) \left[S_{i,T}\left(\frac{t}{T}\right) - S_{i,T}\left(\frac{t-1}{T}\right) \right] \\ &= \left[f_T\left(\frac{t}{T}\right) S_{i,T}\left(\frac{t}{T}\right) - f_T\left(\frac{t-1}{T}\right) S_{i,T}\left(\frac{t-1}{T}\right) \right] - \left[f_T\left(\frac{t}{T}\right) - f_T\left(\frac{t-1}{T}\right) \right] S_{i,T}\left(\frac{t-1}{T}\right) \end{aligned} \tag{S.11}$$

and hence

$$\begin{aligned} &\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k u_{i,t} = \left(\frac{1}{T}\right)^a \left(\ln \frac{1}{T}\right)^k \frac{u_{i,1}}{\sqrt{T}} + \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k u_{i,t} \\ &\overset{\text{(S.11)}}{=} f_T\left(\frac{1}{T}\right) S_{i,T}\left(\frac{1}{T}\right) + \left[f_T(1) S_{i,T}(1) - f_T\left(\frac{1}{T}\right) S_{i,T}\left(\frac{1}{T}\right) \right] - \sum_{t=2}^{T} \left[f_T\left(\frac{t}{T}\right) - f_T\left(\frac{t-1}{T}\right) \right] S_{i,T}\left(\frac{t-1}{T}\right) \\ &\overset{f_T(1)=0}{=} -\sum_{t=2}^{T} \int_{(t-1)/T}^{t/T} S_{i,T}(r) df_T(r) \end{aligned} \tag{S.12}$$

where we used the fact that $S_{i,T}(\cdot)$ is piecewise constant. In view of Assumption 3.2, we can extend suitably extend the probability space and have the following uniformly strong approximation of the partial sum process $S_{i,T}$ (see for example page 562 of Phillips (2007)):

$$\sup_{1 \leq t \leq T} \left| S_{i,T}\left(\frac{t-1}{T}\right) - B_{u_i}\left(\frac{t-1}{T}\right) \right| = o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/q)}} \right), \tag{S.13}$$

for $q > 2$. Continuing from (S.12), this uniformly strong approximation gives

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{a}\left(\ln\frac{t}{T}\right)^{k}u_{i,t} &= -\sum_{t=2}^{T}\int_{(t-1)/T}^{t/T}B_{u_i}(r)df_T(r) + o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/q)}}\right)\\
&= -\int_{1/T}^{1}B_{u_i}(r)df_T(r) + o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/q)}}\right)\\
&= B_{u_i}\left(\frac{1}{T}\right)f_T\left(\frac{1}{T}\right) + \int_{1/T}^{1}f_T(r)dB_{u_i}(r) + o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/p)}}\right) \qquad \text{(S.14)}\\
&= \int_{0}^{1}f(r)dB_{u_i}(r) - \int_{0}^{1/T}f(r)dB_{u_i}(r) + B_{u_i}\left(\frac{1}{T}\right)f_T\left(\frac{1}{T}\right)\\
&\quad + \int_{1/T}^{1}\Big[f_T(r)-f(r)\Big]dB_{u_i}(r) + o_{a.s.}\left(\frac{1}{T^{(1/2)-(1/p)}}\right),
\end{aligned}$$

where the third line is obtained using integration by parts of the mean square Riemann-Stieltjes integral, c.f. theorem 2.7 in Tanaka (2017). It remains to show that $\int_0^{1/T}f(r)dB_{u_i}(r)$, $B_{u_i}\left(\frac{1}{T}\right)f_T\left(\frac{1}{T}\right)$, and $\int_{1/T}^{1}\Big[f_T(r)-f(r)\Big]dB_{u_i}(r)$ are asymptotically negligible. These quantities are zero mean so it suffices to show that their variances vanish as $T\to\infty$. By the isometry property and steps similar to those above (S.6), we have

$$\mathbb{V}\mathrm{ar}\left(\int_0^{1/T}f(r)dB_{u_i}(r)\right) = \Omega_{u_iu_i}\int_0^{1/T}\big[f(r)\big]^2dr \le CT^{-(2a_L+1)}(\ln T)^{2k}\to 0, \qquad \text{(S.15)}$$

as $T\to\infty$. Also, $\mathbb{V}\mathrm{ar}\left(B_{u_i}\left(\frac{1}{T}\right)f_T\left(\frac{1}{T}\right)\right) = \frac{1}{T}\Omega_{u_iu_i}\left[f_T\left(\frac{1}{T}\right)\right]^2 = \Omega_{u_iu_i}\left(\frac{1}{T}\right)^{2a_L+1}\left(\ln\frac{1}{T}\right)^{2k}\to 0$. To control the variance of $\int_{1/T}^{1}\Big[f_T(r)-f(r)\Big]dB_{u_i}(r)$, we look at

$$\begin{aligned}
\int_{1/T}^{1}\left|f(r)-f_T(r)\right|^2dr &= \sum_{t=2}^{T}\int_{(t-1)/T}^{t/T}\left|f(r)-\left(\frac{t-1}{T}\right)^{a}\left(\ln\frac{t-1}{T}\right)^{k}\right|^2dr\\
&= \sum_{t=1}^{T-1}\int_{t/T}^{(t+1)/T}\left|r^a\big(\ln r\big)^k-\left(\frac{t}{T}\right)^{a}\left(\ln\frac{t}{T}\right)^{k}\right|^2dr \qquad \text{(S.16)}\\
&= \sum_{t=1}^{T-1}\int_{0}^{1/T}\left|\left(\frac{t}{T}+s\right)^{a}\left[\ln\left(\frac{t}{T}+s\right)\right]^{k}-\left(\frac{t}{T}\right)^{a}\left(\ln\frac{t}{T}\right)^{k}\right|^2ds.
\end{aligned}$$

Now let $\ell\in\left\{\frac{1}{T},\frac{2}{T},\ldots,1\right\}$ and recall that $0\le s\le\frac{1}{T}$ (hence also $0\le\frac{s}{\ell}\le 1$). Using the triangle inequality, the expression in absolute values can be bounded as

$$\begin{aligned}
\Big|(\ell+s)^a(\ln(\ell+s))^k-\ell^a(\ln\ell)^k\Big| &= \Big|\left[(\ell+s)^a-\ell^a\right](\ln(\ell+s))^k+\ell^a\Big[(\ln(\ell+s))^k-(\ln\ell)^k\Big]\Big|\\
&\le \Big|\left[(\ell+s)^a-\ell^a\right](\ln(\ell+s))^k\Big|+\Big|\ell^a\Big[(\ln(\ell+s))^k-(\ln\ell)^k\Big]\Big| \qquad \text{(S.17)}\\
&= \ell^a\Big|\,(1+s/\ell)^a-1\Big|\,|\ln(\ell+s)|^k+\ell^a\Big|(\ln(\ell+s))^k-(\ln\ell)^k\Big| = IIc+IId.
\end{aligned}$$

By the inequality $|g_a(x)|\le Cx$ below (S.4) and the fact that $|\ln(\ell+s)|\le|\ln\ell|+|\ln(1+s/\ell)|\le\ln T+s/\ell$, we obtain $IIc\le C\ell^{a_L}\frac{s}{\ell}\left|\ln T+\frac{s}{\ell}\right|^k\le C\ell^{a_L-1}(\ln T)^k\frac{1}{T}$. Moreover, the factorisation $p^n-q^n=(p-$

$q) \sum_{j=0}^{n-1} p^{n-1-j} q^j$ yields

$$IId = \ell^a \left|\ln\left(1+s/\ell\right)\right| \left|\sum_{j=0}^{k-1} \left(\ln(\ell+s)\right)^{k-1-j} \left(\ln \ell\right)^j\right| \leq k\ell^{a_L} \frac{s}{\ell} \left|(\ln T)+1\right|^{k-1} \leq C\ell^{a_L-1}(\ln T)^{k-1}\frac{1}{T}. \quad \text{(S.18)}$$

By combination of the bounds on IIc and IId, we conclude that $\left|(\ell+s)^a(\ln(\ell+s))^k - \ell^a(\ln \ell)^k\right| \leq C\ell^{a_L-1}(\ln T)^k \frac{1}{T}$ and arrive at the following upper bound on the RHS of (S.16):

$$\begin{aligned}\int_{1/T}^1 |f(r) - f_T(r)|^2 \, dr &\leq C(\ln T)^{2k} \frac{1}{T^3} \sum_{t=1}^T \left(\frac{t}{T}\right)^{2(a_L-1)} \\ &= O\left(\frac{(\ln T)^{2k}}{T^{2(a_L+\frac{1}{2})}} \mathbb{1}_{\{a_L<\frac{1}{2}\}} + \frac{(\ln T)^{2k+1}}{T^2}\mathbb{1}_{\{a_L=\frac{1}{2}\}} + \frac{(\ln T)^{2k}}{T^2}\mathbb{1}_{\{a_L>\frac{1}{2}\}}\right).\end{aligned} \quad \text{(S.19)}$$

The RHS of (S.19) will go to zero as $T \to \infty$, thereby establishing that $\int_{1/T}^1 \Big[f_T(r) - f(r)\Big] dB_{u_i}(r)$ is also asymptotically negligible. The proof of part (ii) is now complete. ***(iii)*** We have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T \left(\frac{t}{T}\right)^a \left(\ln \frac{t}{T}\right)^k \left(\frac{x_{i,t}}{\sqrt{T}}\right)^j &= \sum_{t=2}^T \int_{(t-1)/T}^{t/T} f_T(r) X_{i,T}^j(r) dt \\ &= \int_0^1 f(r) X_{i,T}^j(r) dr + \int_{1/T}^1 \left[f_T(r) - f(r)\right] X_{i,T}^j(r) dr =: IIIa + IIIb.\end{aligned} \quad \text{(S.20)}$$

Given the CMT and $X_{i,T} \longrightarrow_d B_{v_i}$, term $IIIa$ will converge weakly to $\int_0^1 f(r) B_{v_i}^j(r) dr$ if we can show that $x \mapsto \int_0^1 f(r) x^j(r) dr$ is a continuous functional. Let $x, y \in D[0,1]$. Hölder's inequality implies

$$\begin{aligned}\left|\int_0^1 f(r)x^j(r)dr - \int_0^1 f(r)y^j(r)dr\right| &= \left|\int_0^1 f(r)\left(x^j(r) - y^j(r)\right)dr\right| \\ &\leq \int_0^1 |f(r)|dr \sup_{r\in[0,1]} |x^j(r) - y^j(r)| \leq C \sup_{r\in[0,1]} |x(r) - y(r)| \to 0,\end{aligned} \quad \text{(S.21)}$$

because $\int_0^1 |f(r)|dr = \frac{k!}{(1+a)^{k+1}}$ is bounded. Continuity of the functional now follows from (S.21). If we apply the Cauchy-Schwartz inequality to $IIIb$, then we find

$$IIIb \leq \left[\int_{1/T}^1 |f(r) - f_T(r)|^2 \, dr\right]^{1/2} \left[\int_{1/T}^1 X_{i,T}^{2j}(r) dr\right]^{1/2}.$$

Since $\int_{1/T}^1 |f(r) - f_T(r)|^2 = o(1)$ by (S.19) and $\int_{1/T}^1 X_{i,T}^{2j}(r)dr = \int X_{i,T}^{2j}(r)dr \longrightarrow_d \int B_{v_i}^{2j}(r)dr$. We conclude that $IIIb = o_p(1)$. Now combine the limiting results for $IIIa$ and $IIIb$ to complete the argument. ■

Lemma S.3.3. *For any $\kappa > 0$, define*

$$\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0) = \left\{\boldsymbol{\gamma} \in \boldsymbol{\Gamma} : T^{\theta_0+1/2}|\theta - \theta_0| \leq \kappa \ln T, T^{\theta_0+1/2}|\tau_g - \tau_{g0}| \leq \kappa(\ln T)^2, T^{1/2}\left\|\boldsymbol{D}_{Z,T}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\| \leq \kappa \ln T\right\}. \quad \text{(S.22)}$$

Assume N is fixed and $T \to \infty$. Let k_1, k_2 be any nonnegative integers. Let Assumption 3.2 hold.

(i) $\sup_{\boldsymbol{\gamma}\in\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)} (\ln T)^{k_2} T^{-1} \sum_{t=1}^T \left|T^{-\theta_0}\left(\tau_g t^\theta - \tau_{g0} t^{\theta_0}\right)(\ln t)^{k_1}\right| = o(1)$.

(ii) $\sup_{\boldsymbol{\gamma}\in\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)} N(\ln T)^{k_2}T^{-1}\sum_{t=1}^{T}\left|T^{-2\theta_0}\left(\tau_g t^{2\theta}-\tau_{g0}t^{2\theta_0}\right)(\ln t)^{k_1}\right| = o(1)$.

(iii) $\sup_{\boldsymbol{\gamma}\in\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)}(\ln T)^{k_2}T^{-1}\left|\sum_{t=1}^{T}T^{-\theta_0}\left(t^{\theta}-t^{\theta_0}\right)\boldsymbol{D}_{Z,T}^{-1}\boldsymbol{Z}_t\boldsymbol{\iota}_N\right| = o_p(1)$.

(iv) $\sup_{\boldsymbol{\gamma}\in\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)} T^{-1}\left|\sum_{t=1}^{T}T^{-\theta_0}\left[\left(\tau_g t^{\theta}-\tau_{g0}t^{\theta_0}\right)\ln t-\tau_{g0}\left(t^{\theta}-t^{\theta_0}\right)\ln T\right]\boldsymbol{D}_{Z,T}^{-1}\boldsymbol{Z}_t\boldsymbol{\iota}_N\right| = o_p(1)$.

(v) $\sup_{\boldsymbol{\gamma}\in\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)}(\ln T)^{k_2}T^{-1}\left|\sum_{t=1}^{T}T^{-2\theta_0}t^{\theta}(\ln t)^{k_1}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)'\boldsymbol{Z}_t\boldsymbol{\iota}_N\right| = o_p(1)$.

(vi) $\sup_{\boldsymbol{\gamma}\in\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)}(\ln T)^{k_2}T^{-1}\left|\sum_{t=1}^{T}T^{-2\theta_0}\left(\tau_g t^{\theta}-\tau_{g0}t^{\theta_0}\right)(\ln t)^{k_1}\boldsymbol{u}_t'\boldsymbol{\iota}_N\right| = o_p(1)$.

Proof We only show ***(i)***, ***(iii)***, ***(v)*** and ***(vi)***. The proof of the remaining results is similar and thus omitted.

(i) For any $\boldsymbol{\gamma}\in\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)$, by the triangular inequality and the mean-value theorem (MVT),

$$\begin{aligned}\left|T^{-\theta_0}\left(\tau_g t^{\theta}-\tau_{g0}t^{\theta_0}\right)(\ln t)^{k_1}(\ln T)^{k_2}\right| &= \left(\frac{t}{T}\right)^{\theta_0}\left|\tau_g\left(t^{\theta-\theta_0}-1\right)+\left(\tau_g-\tau_{g0}\right)\right|(\ln t)^{k_1}(\ln T)^{k_2}\\ &\le \left(\frac{t}{T}\right)^{\theta_0}\left[\left|\tau_g\right|\left|t^{\tilde{\theta}}(\theta-\theta_0)\ln t\right|+\left|\tau_g-\tau_{g0}\right|\right](\ln T)^{k_1+k_2}\\ &\le C\left(\frac{t}{T}\right)^{\theta_0}\frac{(\ln T)^{k_1+k_2+2}}{T^{\theta_0+1/2}},\end{aligned}$$

where $t^{|\tilde{\theta}|}\le T^{|\theta-\theta_0|}=\exp\left(|\theta-\theta_0|\ln T\right)\le C$ whenever T is sufficiently large. We obtain the first result due to Lemma S.3.1(i) and $\frac{(\ln T)^k}{T^{\theta_0+1/2}}=o(1)$ for any $k\ge 0$.

(iii) By Lemma S.3.1(i), Lemma S.3.2(iii), and the MVT,

$$(\ln T)^{k_2}T^{-1}\left|\sum_{t=1}^{T}T^{-\theta_0}\left(t^{\theta}-t^{\theta_0}\right)\boldsymbol{D}_{Z,T}^{-1}\boldsymbol{Z}_t\boldsymbol{\iota}_N\right| \le \sqrt{N}\,O\left(\frac{(\ln T)^{k_2+2}}{T^{\theta_0+1/2}}\right)\left\|T^{-1}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_0}\boldsymbol{D}_{Z,T}^{-1}\boldsymbol{Z}_t\right\| = o_p(1),$$

where the term $o_p(1)$ is uniform over $\boldsymbol{\gamma}\in\mathcal{N}_{\kappa,T}(\boldsymbol{\gamma}_0)$.

(v) By Part ***(i)*** and Lemma S.3.2(iii),

$$\begin{aligned}(\ln T)^{k_2}T^{-1}\left|\sum_{t=1}^{T}T^{-2\theta_0}t^{\theta}(\ln t)^{k_1}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)'\boldsymbol{Z}_t\boldsymbol{\iota}_N\right| &\le \sqrt{N}(\ln T)^{k_2}T^{1/2}\left\|\boldsymbol{D}_{Z,T}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)\right\|\\ &\quad\times\left\|T^{-(\theta_0+1/2)}T^{-1}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_0}(\ln t)^{k_1}\boldsymbol{D}_{Z,T}^{-1}\boldsymbol{Z}_t+T^{-(\theta_0+1/2)}o_p(1)\right\| = o_p(1)\end{aligned}$$

(vi) Using the MVT and Lemma S.3.2(ii), we obtain

$$(\ln T)^{k_2}T^{-1}\left|\sum_{t=1}^{T}T^{-2\theta_0}\left(\tau_g t^{\theta}-\tau_{g0}t^{\theta_0}\right)(\ln t)^{k_1}\boldsymbol{u}_t'\boldsymbol{\iota}_N\right| \le \sqrt{N}\,o\left(\frac{(\ln T)^{k_2}}{T^{\theta_0+1/2}}\right)\left\|T^{-1/2}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_0}\boldsymbol{u}_t\right\| = o_p(1).$$

The proof is completed. ■

S4 More Details on the Main Results

Proof of Theorem 3.2 We write $\widehat{\boldsymbol{\Delta}}_T \equiv \widehat{\boldsymbol{\Delta}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T)$ and $\widehat{\boldsymbol{\Omega}}_T \equiv \widehat{\boldsymbol{\Omega}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T)$ to make their dependence on the parameter estimator $\widehat{\boldsymbol{\gamma}}_T$ and bandwidth b_T explicit. Changing the summation indices, we can express the one-sided long-run covariance estimator as

$$\widehat{\boldsymbol{\Delta}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T) = \sum_{i=0}^{T-1} k\left(\frac{i}{b_T}\right)\left[\frac{1}{T}\sum_{t=1}^{T-i} \boldsymbol{V}_{t+i}(\widehat{\boldsymbol{\gamma}}_T)\boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)'\right] =: \widehat{\boldsymbol{\Sigma}}_T(\widehat{\boldsymbol{\gamma}}_T) + \widehat{\boldsymbol{\Gamma}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T),$$

where $\widehat{\boldsymbol{\Sigma}}_T(\widehat{\boldsymbol{\gamma}}_T) = T^{-1}\sum_{t=1}^{T} \boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)\boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)'$ and $\widehat{\boldsymbol{\Gamma}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T) = \sum_{i=1}^{T-1} k\left(\frac{i}{b_T}\right)\left[T^{-1}\sum_{t=1}^{T-i} \boldsymbol{V}_{t+i}(\widehat{\boldsymbol{\gamma}}_T)\boldsymbol{V}_t(\widehat{\boldsymbol{\gamma}}_T)'\right]$. Similarly, we have

$$\widehat{\boldsymbol{\Omega}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T) = \widehat{\boldsymbol{\Sigma}}_T(\widehat{\boldsymbol{\gamma}}_T) + \widehat{\boldsymbol{\Gamma}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T) + \widehat{\boldsymbol{\Gamma}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T)'.$$

Clearly, it suffices to study the asymptotic behavior of $\widehat{\boldsymbol{\Sigma}}_T(\widehat{\boldsymbol{\gamma}}_T)$ and $\widehat{\boldsymbol{\Gamma}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T)$. As the lower right subblock of $\boldsymbol{V}_{t+i}(\boldsymbol{\gamma})\boldsymbol{V}_t(\boldsymbol{\gamma})'$ equals $\boldsymbol{v}_{t+i}\boldsymbol{v}_t'$ (no parameter estimation uncertainty here), the consistency result for this subblock follows from the properties of $\{\boldsymbol{v}_t\}$ in Assumption 3.2, the kernel requirements in Assumption 3.3, and an application in Theorem 2 of Jansson (2002).

We proceed to the upper left subblocks of $\widehat{\boldsymbol{\Sigma}}_T(\widehat{\boldsymbol{\gamma}}_T)$ and $\widehat{\boldsymbol{\Gamma}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T)$. If the residuals are close enough to the true innovations, then the results from Theorem 2 of Jansson (2002) again applies. It suffices to show

$$T^{-1}\sum_{t=1}^{T}\left[\widehat{\boldsymbol{u}}_t\widehat{\boldsymbol{u}}_t' - \boldsymbol{u}_t\boldsymbol{u}_t'\right] \longrightarrow_p 0 \quad \text{and} \quad \sum_{i=1}^{T-1} k\left(i/b_T\right)\left(T^{-1}\sum_{t=1}^{T-i}\left[\widehat{\boldsymbol{u}}_{t+i}\widehat{\boldsymbol{u}}_t' - \boldsymbol{u}_{t+i}\boldsymbol{u}_t'\right]\right) \longrightarrow_p 0.$$

Using Lemmas S.3.1–S.3.2 and Theorem 3.1, the following key result follows immediately

$$\widehat{\boldsymbol{u}}_t = \boldsymbol{u}_t - \left(\widehat{\tau}_{g,T}\, t^{\widehat{\theta}_T} - \tau_{g0}\, t^{\theta_0}\right)\boldsymbol{\imath}_N - \boldsymbol{Z}_t'\left(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0\right) = \boldsymbol{u}_t + O_p\big(T^{-1/2}\ln T\big)\left(\frac{t}{T}\right)^{\theta_0} + O_p\left(T^{-1/2}\right)\left(\boldsymbol{D}_{Z,T}^{-1}\boldsymbol{Z}_t\right)'. \tag{S.1}$$

This implies $T^{-1}\sum_{t=1}^{T}\left[\widehat{\boldsymbol{u}}_t\widehat{\boldsymbol{u}}_t' - \boldsymbol{u}_t\boldsymbol{u}_t'\right] = O_p\big(T^{-1/2}\ln T\big)$ and

$$\left\|\sum_{i=1}^{T-1} k\left(i/b_T\right) T^{-1}\sum_{t=1}^{T-i}\left(\widehat{\boldsymbol{u}}_{t+i}\widehat{\boldsymbol{u}}_t' - \boldsymbol{u}_{t+i}\boldsymbol{u}_t'\right)\right\| \leq O_p\big(T^{-1/2}\ln T\big)\sum_{i=1}^{T-1}\left|k\left(i/b_T\right)\right| = O_p\big(T^{-1/2}b_T\ln T\big), \tag{S.2}$$

where the final step is due to lemma 1 of Jansson (2002). Clearly, both terms are asymptotically negligible under the assumption $T^{-1/2}b_T\ln T \to 0$ as $T\to\infty$. The limits of the two remaining subblocks of $\widehat{\boldsymbol{\Sigma}}_T(\widehat{\boldsymbol{\gamma}}_T)$ and $\widehat{\boldsymbol{\Gamma}}_T(\widehat{\boldsymbol{\gamma}}_T, b_T)$ are derived similarly. ■

Proof of Theorem 3.4 Without loss of generality, we set $\ell = 1$. Note that

$$q_T^{-1/2} \sum_{t=1}^{[rq_T]} \widehat{\boldsymbol{u}}_t^+ = q_T^{-1/2} \sum_{t=1}^{[rq_T]} \left(\boldsymbol{u}_t - \boldsymbol{\Omega}_{uv} \boldsymbol{\Omega}_{vv}^{-1} \boldsymbol{v}_t \right) - \sum_{j=1}^{3} \widetilde{\boldsymbol{R}}_{q_T,j}, \qquad r \in [0,1], \tag{S.3}$$

where the stochastic order of the remainder terms $\widetilde{\boldsymbol{R}}_{q_T,1}$–$\widetilde{\boldsymbol{R}}_{q_T,3}$ follows from Lemma S.3.3 and Theorem 3.1:

(a) $\widetilde{\boldsymbol{R}}_{q_T,1} = \left(\widehat{\boldsymbol{\Omega}}_{uv} \widehat{\boldsymbol{\Omega}}_{vv}^{-1} - \boldsymbol{\Omega}_{uv} \boldsymbol{\Omega}_{vv}^{-1} \right) q_T^{-1/2} \sum_{t=1}^{[rq_T]} \boldsymbol{v}_t = o_p(1)$,

(b) $\widetilde{\boldsymbol{R}}_{q_T,2} = q_T^{-1/2} \sum_{t=1}^{[rq_T]} \left(\widehat{\tau}_{g,T}\, t^{\widehat{\theta}_T} - \tau_{g0}\, t^{\theta_0} \right) \boldsymbol{\imath}_N = O_p\left(\ln T \left(\frac{q_T}{T} \right)^{\theta_0 + 1/2} \right)$,

(c) $\widetilde{\boldsymbol{R}}_{q_T,3} = q_T^{-1/2} \sum_{t=1}^{[rq_T]} \boldsymbol{Z}_t' \left(\widehat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0 \right) = O_p\left(\left(\frac{q_T}{T} \right)^{1/2} \right) q_T^{-1} \sum_{t=1}^{[rq_T]} \left(\boldsymbol{D}_{Z,q_T}^{-1} \boldsymbol{Z}_t \right)' \boldsymbol{D}_{Z,q_T} \boldsymbol{D}_{Z,T}^{-1} = O_p\left(\left(\frac{q_T}{T} \right)^{1/2} \right)$.

The theorem follows from (S.3), a functional central limit theorem for linear processes, the continuous mapping theorem, and the rate requirements. ■

S5 Details on Simulation DGPs 2(a)–2(c)

The parameters of simulation DGPs have been selected according to the following general procedure.

STEP 1: Load the data and estimate the model corresponding to the specification under H_0. The resulting coefficients $\widehat{\boldsymbol{\gamma}}_T$ and residual series $\{\widehat{\boldsymbol{u}}_t\}$ are stored.

STEP 2: Estimate a VAR(1) on the residuals, $\widehat{\boldsymbol{u}}_t = \boldsymbol{A}^{(1)} \widehat{\boldsymbol{u}}_{t-1} + \boldsymbol{\xi}_t^{(1)}$, and compute $\widehat{\boldsymbol{\xi}}_t^{(1)} = \widehat{\boldsymbol{u}}_t - \widehat{\boldsymbol{A}}^{(1)} \widehat{\boldsymbol{u}}_{t-1}$ for $t = 2, \ldots, T$.

STEP 3: Repeat STEP 2 for $\Delta \boldsymbol{x}_t$. Using obvious notation, the resulting filtered residuals are $\widehat{\boldsymbol{\xi}}_t^{(2)} = \Delta \boldsymbol{x}_t - \widehat{\boldsymbol{A}}^{(2)} \Delta \boldsymbol{x}_{t-1}$ for $t = 3, \ldots, T$.

STEP 4: Set $\widehat{\boldsymbol{\xi}}_t = [\widehat{\boldsymbol{\xi}}_t^{(1)\prime}, \widehat{\boldsymbol{\xi}}_t^{(2)\prime}]'$ and compute the $(2N \times 2N)$ covariance matrix estimate $\widehat{\boldsymbol{\Sigma}} = \frac{1}{T-2} \sum_{t=3}^{T} \widehat{\boldsymbol{\xi}}_t \widehat{\boldsymbol{\xi}}_t'$.

STEP 5: The simulated data is based on the parameters from STEP 1 – STEP 4. First, generate $\boldsymbol{\xi}_t = [\boldsymbol{\xi}_t^{(1)\prime}, \boldsymbol{\xi}_t^{(2)\prime}]' \overset{i.i.d.}{\sim} \mathrm{N}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}})$. Subsequently, we use the results from the VAR(1) models:

(a) Set $\boldsymbol{u}_0 = \boldsymbol{0}$ and construct innovations according to $\boldsymbol{u}_t = \widehat{\boldsymbol{A}}^{(1)} \boldsymbol{u}_{t-1} + \boldsymbol{\xi}_t^{(1)}$.

(b) Set $\Delta \boldsymbol{x}_0 = \boldsymbol{0}$, construct the increments of the integrated explanatory variables through the recursion $\Delta \boldsymbol{x}_t = \widehat{\boldsymbol{A}}^{(2)} \Delta \boldsymbol{x}_{t-1} + \boldsymbol{\xi}_t^{(2)}$, and compute the partial sums $\boldsymbol{x}_t = \sum_{s=1}^{t} \Delta \boldsymbol{x}_s$.

Given the simulated innovations and simulated integrated regressors, it remains to use $\widehat{\boldsymbol{\gamma}}_T$ to obtain the simulated dependent variables.

Three remarks follow. First, we explicitly choose individual VAR(1) models for $\{\widehat{\boldsymbol{u}}_t\}$ and $\{\widehat{\Delta \boldsymbol{x}}_t\}$ rather than a single $2N$-dimensional VAR(1) for the joint vector. Otherwise, with $N = 6$ and $T = 145$ in the data, the number of parameters in the autoregressive matrix would be $12^2 = 144$ which is rather close to the length of the data series. Similarly, the selection of the VAR(1) specification results from a trade-off between model parsimony and a simulation DGP with serial correlation. Second, we follow the literature and rebuild the

integrated explanatory variables as random walks *without* drift. Third, the specific values of all parameters (rounded to 2 decimals) are reported in the next subsections.

S5.1 Parameter values for DGP2(a)

The estimated parameters of the model are

$$\boldsymbol{y}_t = -1.37 \times 10^{-5} t^{2.45} \boldsymbol{\iota}_6 + \begin{bmatrix} -6.39 \\ -0.12 \\ -12.66 \\ -3.99 \\ -1.67{\times}10^{1} \\ -2.55 \end{bmatrix} + \begin{bmatrix} -4.9{\times}10^{-3} \\ 7.6{\times}10^{-3} \\ 1.05{\times}10^{-3} \\ 1.09{\times}10^{-3} \\ -2.7{\times}10^{-3} \\ 2.14{\times}10^{-4} \end{bmatrix} t + \begin{bmatrix} 1.73x_{1,t} \\ 1.01x_{2,t} \\ 2.22x_{3,t} \\ 1.33x_{4,t} \\ 2.55x_{5,t} \\ 1.33x_{6,t} \end{bmatrix} + \widehat{\boldsymbol{u}}_t.$$

The results for the VAR(1) specifications follow

$$\widehat{\boldsymbol{A}}^{(1)} = \begin{bmatrix} 0.74 & -0.04 & 0.02 & 0.32 & -0.34 & 0.41 \\ 0.05 & 0.68 & 0.01 & -0.30 & -0.05 & -0.01 \\ 0.16 & 1.05 & 0.62 & -0.54 & -0.29 & -0.55 \\ 0.04 & 0.12 & 0.03 & 0.41 & 0.00 & -0.21 \\ 0.10 & 0.31 & -0.08 & -0.29 & 0.67 & 0.01 \\ 0.05 & 0.08 & -0.03 & -0.12 & 0.02 & 0.44 \end{bmatrix}, \qquad \widehat{\boldsymbol{A}}^{(2)} = \begin{bmatrix} -0.09 & -0.32 & 0.21 & 0.96 & -0.22 & 0.55 \\ -0.04 & 0.02 & 0.39 & 0.10 & -0.15 & -0.00 \\ -0.08 & -0.10 & 0.38 & 0.15 & -0.07 & 0.35 \\ -0.37 & 0.38 & -0.13 & 0.25 & -0.00 & 0.26 \\ -0.18 & 0.19 & 0.05 & 0.18 & -0.07 & 0.41 \\ 0.03 & 0.18 & -0.05 & 0.02 & -0.06 & 0.47 \end{bmatrix},$$

and

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 6.13 & * & * & * & * & * & * & * & * & * & * & * \\ 0.39 & 0.70 & * & * & * & * & * & * & * & * & * & * \\ 0.22 & 1.00 & 7.19 & * & * & * & * & * & * & * & * & * \\ 0.59 & 0.06 & 0.29 & 0.91 & * & * & * & * & * & * & * & * \\ 0.57 & 0.15 & 0.68 & 0.38 & 0.01 & * & * & * & * & * & * & * \\ 0.36 & 0.13 & 0.37 & 0.01 & 0.13 & 0.46 & * & * & * & * & * & * \\ -0.29 & -0.02 & 0.32 & 0.03 & 0.18 & -0.02 & 0.43 & * & * & * & * & * \\ 0.13 & 0.07 & 0.22 & 0.04 & -0.06 & 0.06 & -0.00 & 0.14 & * & * & * & * \\ 0.21 & 0.11 & 0.21 & 0.09 & -0.03 & 0.05 & 0.04 & 0.10 & 0.19 & * & * & * \\ 0.23 & 0.18 & 0.28 & -0.07 & 0.01 & 0.04 & 0.03 & 0.11 & 0.10 & 0.33 & * & * \\ 0.10 & 0.07 & 0.25 & -0.01 & -0.34 & 0.07 & 0.01 & 0.06 & 0.05 & 0.05 & 0.20 & * \\ 0.02 & 0.02 & -0.03 & -0.00 & 0.03 & 0.02 & 0.05 & -0.01 & 0.02 & 0.01 & 0.02 & 0.08 \end{bmatrix} \times 10^{-2}.$$

S5.2 Parameter values for DGP2(b)

The estimated model specification is

$$\boldsymbol{y}_t = \begin{bmatrix} -1.01 \\ 8.64 \\ -5.15 \\ 3.78 \\ -17.84 \\ 11.84 \end{bmatrix} + \begin{bmatrix} -1.11{\times}10^{-2} \\ 5.78{\times}10^{-3} \\ 1.63{\times}10^{-2} \\ 7.69{\times}10^{-3} \\ -2.36{\times}10^{-2} \\ 4.69{\times}10^{-3} \end{bmatrix} t + + \begin{bmatrix} 1.10x_{1,t} \\ -1.44{\times}10^{-3}x_{2,t} \\ 1.23x_{3,t} \\ 4.42{\times}10^{-1}x_{4,t} \\ 2.73x_{5,t} \\ -3.21{\times}10^{-1}x_{6,t} \end{bmatrix} + \widehat{\boldsymbol{u}}_t.$$

The VAR(1) dynamics in the innovations and increments are governed by

$$\widehat{\boldsymbol{A}}^{(1)} = \begin{bmatrix} 0.78 & 0.01 & 0.15 & -0.07 & -0.39 & -0.07 \\ 0.03 & 0.87 & -0.01 & 0.02 & -0.02 & -0.07 \\ 0.06 & 1.29 & 0.60 & -0.37 & -0.01 & -0.35 \\ 0.04 & 0.36 & 0.01 & 0.50 & -0.05 & 0.06 \\ 0.03 & 0.38 & -0.03 & -0.07 & 0.60 & -0.06 \\ 0.03 & -0.05 & 0.01 & 0.08 & 0.02 & 0.59 \end{bmatrix}, \qquad \widehat{\boldsymbol{A}}^{(2)} = \begin{bmatrix} -0.09 & -0.32 & 0.21 & 0.96 & -0.22 & 0.55 \\ -0.04 & 0.02 & 0.39 & 0.10 & -0.15 & -0.00 \\ -0.08 & -0.10 & 0.38 & 0.15 & -0.07 & 0.35 \\ -0.37 & 0.38 & -0.13 & 0.25 & -0.00 & 0.26 \\ -0.18 & 0.19 & 0.05 & 0.18 & -0.07 & 0.41 \\ 0.03 & 0.18 & -0.05 & 0.02 & -0.06 & 0.47 \end{bmatrix},$$

and

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 5.65 & * & * & * & * & * & * & * & * & * & * & * \\ 0.68 & 1.13 & * & * & * & * & * & * & * & * & * & * \\ 1.29 & 1.79 & 8.56 & * & * & * & * & * & * & * & * & * \\ 0.84 & 0.58 & 1.33 & 1.23 & * & * & * & * & * & * & * & * \\ 0.67 & 0.30 & 1.16 & 0.71 & 2.32 & * & * & * & * & * & * & * \\ 0.09 & 0.25 & 0.60 & 0.17 & 0.16 & 0.72 & * & * & * & * & * & * \\ -0.09 & -0.01 & 0.24 & 0.06 & 0.19 & 0.04 & 0.43 & * & * & * & * & * \\ 0.15 & 0.19 & 0.39 & 0.13 & -0.04 & 0.01 & -0.00 & 0.14 & * & * & * & * \\ 0.24 & 0.20 & 0.40 & 0.20 & 0.01 & 0.04 & 0.04 & 0.10 & 0.19 & * & * & * \\ 0.24 & 0.32 & 0.43 & 0.19 & 0.09 & 0.01 & 0.03 & 0.11 & 0.10 & 0.33 & * & * \\ 0.09 & 0.12 & 0.28 & 0.04 & -0.34 & 0.08 & 0.01 & 0.06 & 0.05 & 0.05 & 0.20 & * \\ 0.00 & 0.01 & -0.07 & 0.02 & 0.02 & 0.09 & 0.05 & -0.01 & 0.02 & 0.01 & 0.02 & 0.08 \end{bmatrix} \times 10^{-2}.$$

S5.3 Parameter values for DGP2(c)

The simulation experiments regarding the performance of the KPSS test are based on the same specification as DGP2(a).

S6 Additional Simulation Results

S6.1 Empirical size for DGP1 when $\theta = 0.8$

Table S1: The empirical size (in %) of the single-equation t-tests $H_0 : \beta_{2,1} = 0$ and the joint Wald tests for $H_0 : \beta_{2,1} = \ldots = \beta_{2,N} = 0$ with $\beta_{2,i}$ denoting the coefficient in front of $x_{i,t}^2$. The Monte Carlo results are based on: simulated inference with θ estimated by NLS (SimNLS), simulated inference with known $\theta = 0.8$ (SimNLS(θ_0)), and two Fully Modified estimators for systems developed by Wagner et al. (2020) with known $\theta = 0.8$ (FM-SOLS(θ_0) and FM-SUR(θ_0)).

$\theta_0 = 0.8$	$N = 3$				$N = 5$				$N = 10$			
ρ	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)
$T = 150$												
0	4.03	3.93	9.10	10.47	4.67	4.67	10.03	12.90	4.40	4.50	10.80	16.67
0.3	4.60	4.50	9.77	11.07	4.53	4.50	10.37	13.07	5.07	4.90	11.90	19.50
0.6	4.53	4.47	10.57	12.60	4.30	4.23	11.87	16.27	4.60	4.40	13.57	29.83
0.8	4.33	4.50	13.80	18.47	4.70	4.63	15.60	27.07	4.33	4.30	16.30	56.73
$T = 300$												
0	4.20	4.23	7.87	8.67	4.87	4.87	7.87	9.40	4.37	4.40	8.27	10.97
0.3	5.27	5.23	8.47	9.50	4.47	4.50	9.23	10.83	4.47	4.50	8.87	12.93
0.6	4.50	4.63	9.47	11.00	5.10	4.83	10.10	12.90	4.27	4.47	10.63	18.57
0.8	4.60	4.40	12.07	14.47	4.43	4.30	12.17	18.13	5.47	5.23	14.00	35.00
$T = 600$												
0	4.43	4.47	6.83	7.47	4.23	4.37	6.53	6.93	4.23	4.17	7.27	9.10
0.3	4.97	5.03	7.57	8.60	5.13	4.93	7.37	8.10	4.70	4.93	8.10	9.60
0.6	5.27	5.40	8.27	9.50	5.17	4.93	8.43	9.33	5.03	4.97	9.60	14.57
0.8	4.13	4.30	8.83	10.10	4.93	4.77	9.73	14.10	5.20	4.93	10.90	23.70
Panel B: Joint test												
$T = 150$												
0	3.57	3.63	12.03	15.23	4.00	4.23	14.30	21.57	4.03	3.93	26.03	50.13
0.3	4.07	3.90	13.83	16.43	3.77	3.47	19.47	26.93	3.23	3.40	29.67	60.20
0.6	3.73	3.70	17.03	21.43	3.60	3.67	23.60	38.07	2.33	2.03	39.73	83.70
0.8	3.20	2.80	23.43	31.03	2.87	2.87	32.13	58.27	1.53	1.37	50.30	82.73
$T = 300$												
0	5.13	5.13	10.47	11.43	3.43	3.60	12.00	15.17	3.67	3.57	17.93	30.33
0.3	4.40	4.30	9.90	11.63	4.07	3.87	13.43	17.83	3.83	4.00	19.30	36.63
0.6	4.20	4.37	13.40	15.57	3.97	4.00	17.47	24.33	3.30	3.20	28.60	59.80
0.8	4.07	3.63	16.27	20.87	3.47	3.17	22.20	38.97	2.40	2.40	37.77	86.00
$T = 600$												
0	3.50	3.53	7.03	7.97	4.43	4.63	8.83	11.07	3.90	4.10	12.63	18.93
0.3	4.57	4.53	8.90	9.53	4.17	4.13	10.70	12.67	4.53	4.53	15.07	23.83
0.6	5.37	4.87	10.40	12.23	4.73	4.33	13.30	16.97	4.03	4.07	21.47	39.70
0.8	3.70	3.83	11.63	14.30	3.50	3.70	15.20	24.83	3.60	3.60	26.37	66.77

S6.2 Empirical size for DGP1 when $\theta = 1.8$

Table S2: The empirical size (in %) of the single-equation t-tests $H_0 : \beta_{2,1} = 0$ and the joint Wald tests for $H_0 : \beta_{2,1} = \ldots = \beta_{2,N} = 0$ with $\beta_{2,i}$ denoting the coefficient in front of $x_{i,t}^2$. The Monte Carlo results are based on: simulated inference with θ estimated by NLS (SimNLS), simulated inference with known $\theta = 1.8$ (SimNLS(θ_0)), and two Fully Modified estimators for systems developed by Wagner et al. (2020) with known $\theta = 1.8$ (FM-SOLS(θ_0) and FM-SUR(θ_0)).

$\theta_0 = 1.8$	$N = 3$				$N = 5$				$N = 10$			
ρ	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)	SimNLS	SimNLS(θ_0)	FM-SOLS(θ_0)	FM-SUR(θ_0)
$T = 150$												
0	4.70	4.80	9.87	11.37	4.13	4.23	10.27	12.30	4.43	4.37	9.80	15.60
0.3	3.97	3.80	9.67	10.97	4.43	4.50	9.83	13.13	4.77	4.50	11.40	18.37
0.6	5.03	4.63	12.53	14.77	4.70	4.83	12.27	17.07	3.63	3.37	12.27	29.63
0.8	5.43	5.47	14.93	18.93	4.97	4.60	14.80	27.27	5.23	4.90	16.30	56.23
$T = 300$												
0	4.53	4.80	7.23	8.10	4.40	4.43	7.80	9.63	4.47	4.73	8.53	11.77
0.3	4.27	4.50	8.10	9.37	5.07	5.00	9.23	9.97	4.37	4.20	8.90	12.80
0.6	6.23	5.87	10.17	12.23	4.73	4.77	10.03	13.63	4.77	4.30	10.50	18.83
0.8	4.50	4.43	11.00	13.80	4.57	4.43	11.97	18.70	4.10	3.73	13.70	36.73
$T = 600$												
0	4.13	4.37	7.33	7.83	4.07	4.13	6.90	7.90	4.97	4.97	6.87	8.17
0.3	4.70	4.90	8.33	8.90	4.77	4.83	6.93	8.13	4.80	4.67	8.10	10.57
0.6	5.77	6.10	8.70	9.03	5.17	5.23	8.23	9.37	5.53	5.53	9.27	13.80
0.8	4.73	4.80	9.07	10.33	4.57	4.77	9.63	12.80	4.93	4.73	12.03	25.87
Panel B: Joint test												
$T = 150$												
0	4.23	4.03	12.13	15.07	3.53	3.50	16.50	23.07	3.40	3.33	25.57	50.30
0.3	4.13	3.80	14.50	16.00	3.63	3.77	19.90	27.00	3.77	3.63	30.80	58.17
0.6	3.23	2.83	17.13	22.07	3.93	3.73	24.67	38.53	2.13	2.43	40.87	82.83
0.8	3.57	3.03	23.77	30.83	2.93	2.30	32.43	57.57	2.37	1.77	49.77	83.23
$T = 300$												
0	4.63	4.83	9.40	10.97	4.10	4.20	11.47	15.63	3.80	3.83	18.50	31.87
0.3	4.60	4.80	11.70	13.00	4.67	4.67	14.90	18.43	3.67	3.57	19.17	36.73
0.6	5.00	4.53	14.00	15.17	4.03	3.33	17.00	24.80	3.17	2.83	29.00	59.57
0.8	3.43	3.10	18.00	21.27	3.33	2.80	22.77	39.80	2.50	1.93	37.93	86.07
$T = 600$												
0	4.23	4.47	7.97	8.63	3.20	3.43	9.10	11.43	3.87	3.83	10.47	17.40
0.3	4.20	4.37	9.37	10.80	4.53	4.60	10.37	12.17	3.87	3.77	15.47	23.80
0.6	5.43	5.33	11.90	12.30	4.80	4.70	12.27	17.03	5.00	4.57	19.63	39.03
0.8	3.73	3.63	11.67	14.40	3.97	3.77	16.37	24.77	3.30	3.07	27.87	68.27

S6.3 Empirical power for DGP1

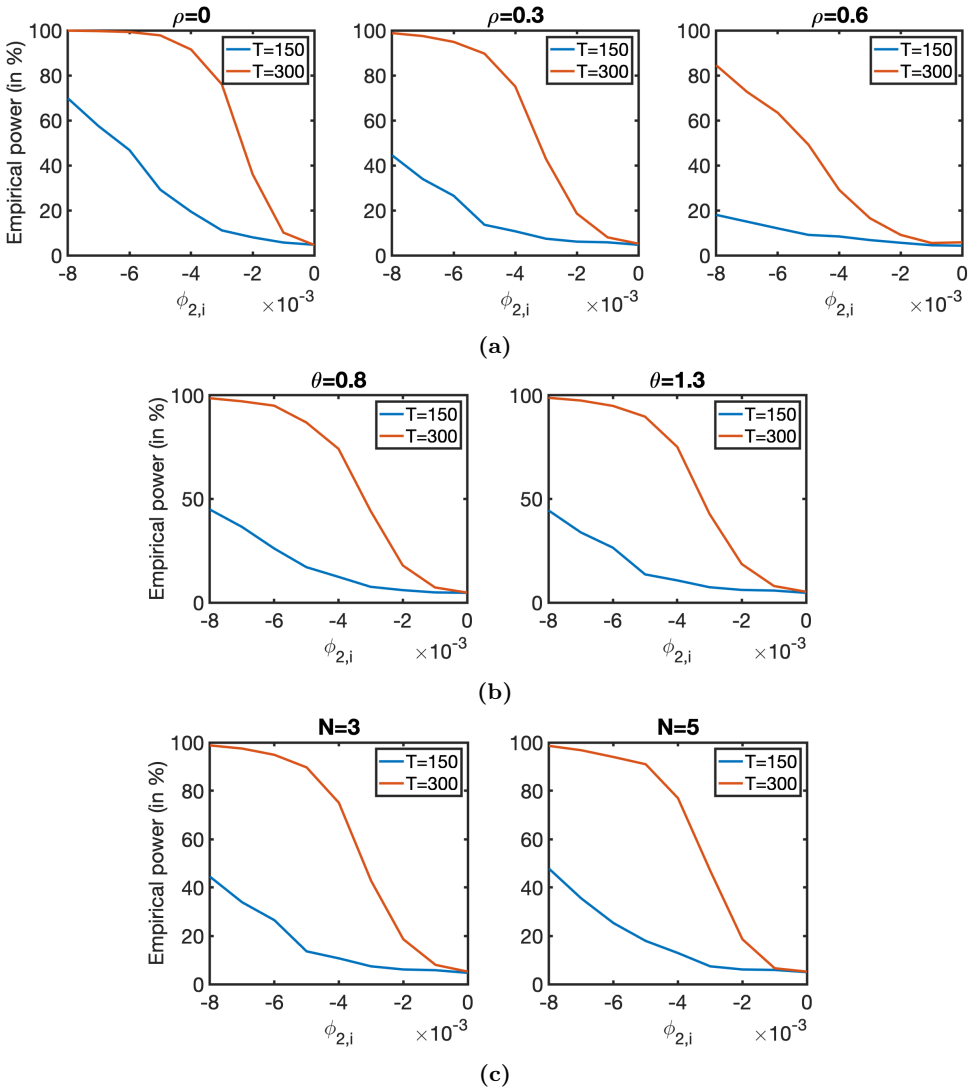


Figure 1: The power curves for the *single equation test* $H_0 : \phi_{2,1} = 0$. The reference model is DGP1 with $\rho = 0.3$, $\theta = 1.3$ and $N = 3$. We vary the parameters of this reference specification one-by-one while keeping the remaining two parameters fixed at their baseline values. Specifically, we study changes in: **(a)** the serial correlation and endogeneity parameter ρ, **(b)** the nonlinear deterministic time trend power, and **(c)** the cross-sectional dimension.

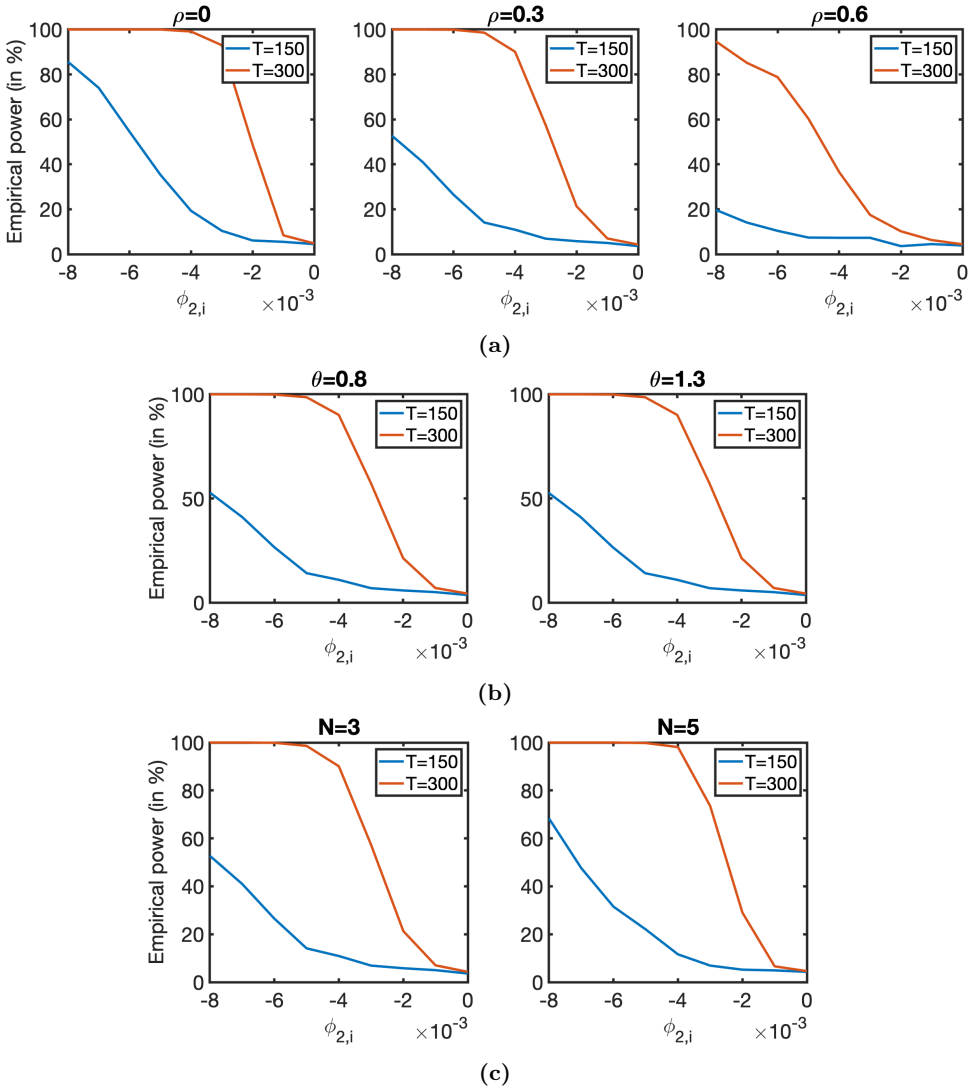


Figure 2: The power curves for the *joint test* for $H_0 : \phi_{2,1} = \ldots = \phi_{2,N} = 0$. The reference model is DGP1 with $\rho = 0.3$, $\theta = 1.3$ and $N = 3$. We vary the parameters of this reference specification one-by-one while keeping the remaining two parameters fixed at their baseline values. Specifically, we study changes in: **(a)** the serial correlation and endogeneity parameter ρ, **(b)** the nonlinear deterministic time trend power, and **(c)** the cross-sectional dimension.

S6.4 Empirical power and KPSS test for DGP2

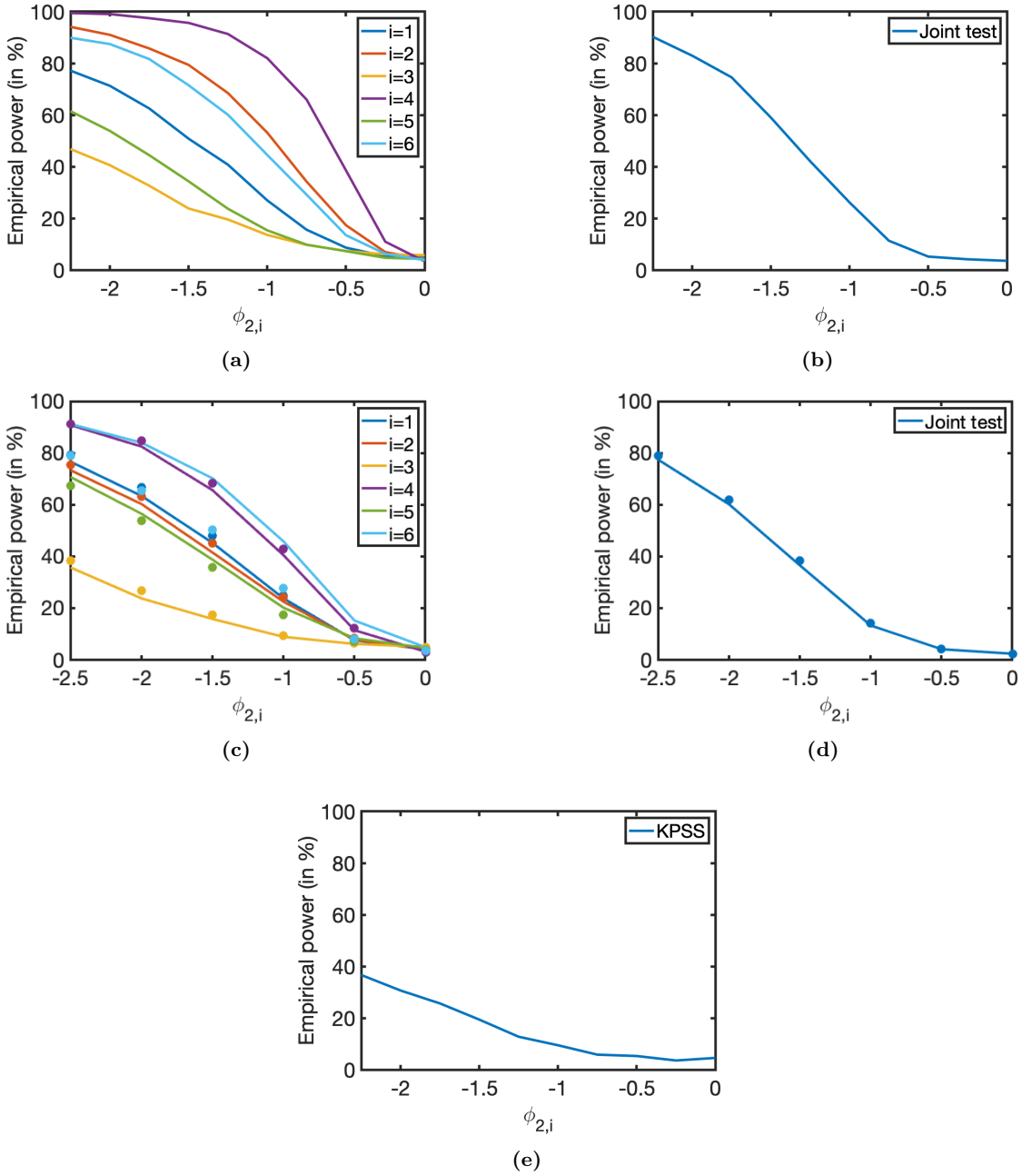


Figure 3: An overview of various power curves for DGP2. **(a)** Unit-specific power curves when testing $H_0 : \phi_{2,i} = 0$ $(i = 1, \ldots, 6)$ for a correctly specified model. **(b)** The power curve when testing $H_0 : \phi_{2,1} = \ldots = \phi_{2,6} = 0$ for a correctly specified model. **(c)** The empirical rejection frequencies for a correctly specified model (lines) and an estimation with a redundant global time trend (dots). Individual coefficients are tested. **(d)** As in (c), but now for the joint test $H_0 : \phi_{2,1} = \ldots = \phi_{2,6} = 0$. **(e)** The empirical power of the KPSS for a misspecified linear cointegrating relation.

S7 Further Empirical Results

The exact numbers may show (minor) differences from previously reported results due to differences in: (1) the time span of the data, (2) the implemented long-run covariance estimator, and (3) the scaling of the data. Related to scaling, we follow the official guidelines and multiply by 3.667 and 10^3 to convert thousand of metric tons of carbon into units of carbon dioxide. Since the data will be expressed in logarithms, this rescaling effectively amounts to a change of intercept.

S7.1 Unit root tests

Table S3: The t-statistics for the ADF and DF-GLS unit root tests. The columns with header 'const' and 'const & trend' refer to the inclusion of only an intercept or both intercept and linear trend. Rejection of the unit root hypothesis at a 10% and 5% level are indicated with one and two stars, respectively.

	ADF				DF-GLS			
	const		const & trend		const		const & trend	
	GDP	CO_2	GDP	CO_2	GDP	CO_2	GDP	CO_2
Australia	0.287	-2.549	-2.050	-1.986	2.046	1.379	-1.577	-0.732
Austria	-0.055	-2.118	-1.943	-2.738	1.478	-1.143	-1.655	-2.718^{*}
Belgium	0.153	-2.336	-1.705	-2.818	2.041	-0.794	-1.287	-2.644
Canada	-0.575	-1.133	-2.020	-1.120	1.117	0.874	-1.894	-0.387
Denmark	-0.235	-2.446	-2.326	-0.136	1.393	0.410	-1.505	0.084
Finland	-0.362	-1.327	-2.315	-3.248^{*}	0.420	-0.076	-1.155	-3.217^{**}
France	-0.557	-2.438	-1.823	-1.858	1.087	-0.267	-1.470	-1.212
Germany	-0.374	-3.099^{**}	-2.767	-3.971^{**}	1.195	-0.726	-2.474	-2.080
Italy	-0.252	-1.546	-1.759	-1.987	1.213	0.354	-1.240	-1.860
Japan	0.010	-0.862	-1.733	-0.941	1.382	0.504	-1.272	-0.878
Netherlands	-0.106	-1.629	-2.247	-3.106	1.378	0.213	-1.679	-2.818^{*}
Norway	-0.680	-2.044	-2.064	-2.318	0.749	0.331	-1.017	-1.292
Portugal	-1.432	-0.455	-1.697	-1.676	-0.708	0.593	-0.741	-1.923
Spain	0.402	-1.243	-1.354	-1.994	1.487	0.959	-1.077	-2.014
Sweden	-0.789	-2.075	-2.289	-1.625	0.258	0.180	-1.513	-0.968
Switzerland	-1.093	-1.963	-2.785	-1.989	2.272	0.368	-2.447	-1.237
UK	-0.179	-0.721	-1.262	-0.402	2.446	-0.622	-0.608	-0.013
USA	-0.349	-2.055	-2.871	-1.322	2.409	-0.101	-2.708^{*}	-0.812

Note: Asterisks denote rejection of the null hypothesis at the $^{***}1\%$, $^{**}5\%$, and $^{*}10\%$ significance level.

S7.2 Perron and Yabu (2009) test for deterministic trend coefficient

The Perron and Yabu (2009) test is used to test for the presence of a deterministic trend function in the log per capita GDP series, see Table S4. The test allows for integrated or stationary errors. The details of the procedure can be found on page 61 of Perron and Yabu (2009). The asymptotic distribution of this test statistic is standard normal (quantiles are $z_{0.95} = 1.645$, $z_{0.975} = 1.96$, and $z_{0.995} = 2.58$).

Table S4: Perron and Yabu (2009) test statistic for each of the 18 countries.

	$\widehat{PY}$
Australia	3.17
Austria	2.19
Belgium	3.52
Canada	3.33
Denmark	5.58
Finland	4.27

	$\widehat{PY}$
France	2.41
Germany	1.91
Italy	2.11
Japan	2.93
Netherlands	2.27
Norway	5.85

	$\widehat{PY}$
Portugal	2.16
Spain	2.31
Sweden	7.12
Switzerland	3.91
UK	3.60
USA	4.12

S7.3 Overviews for Austria and Finland

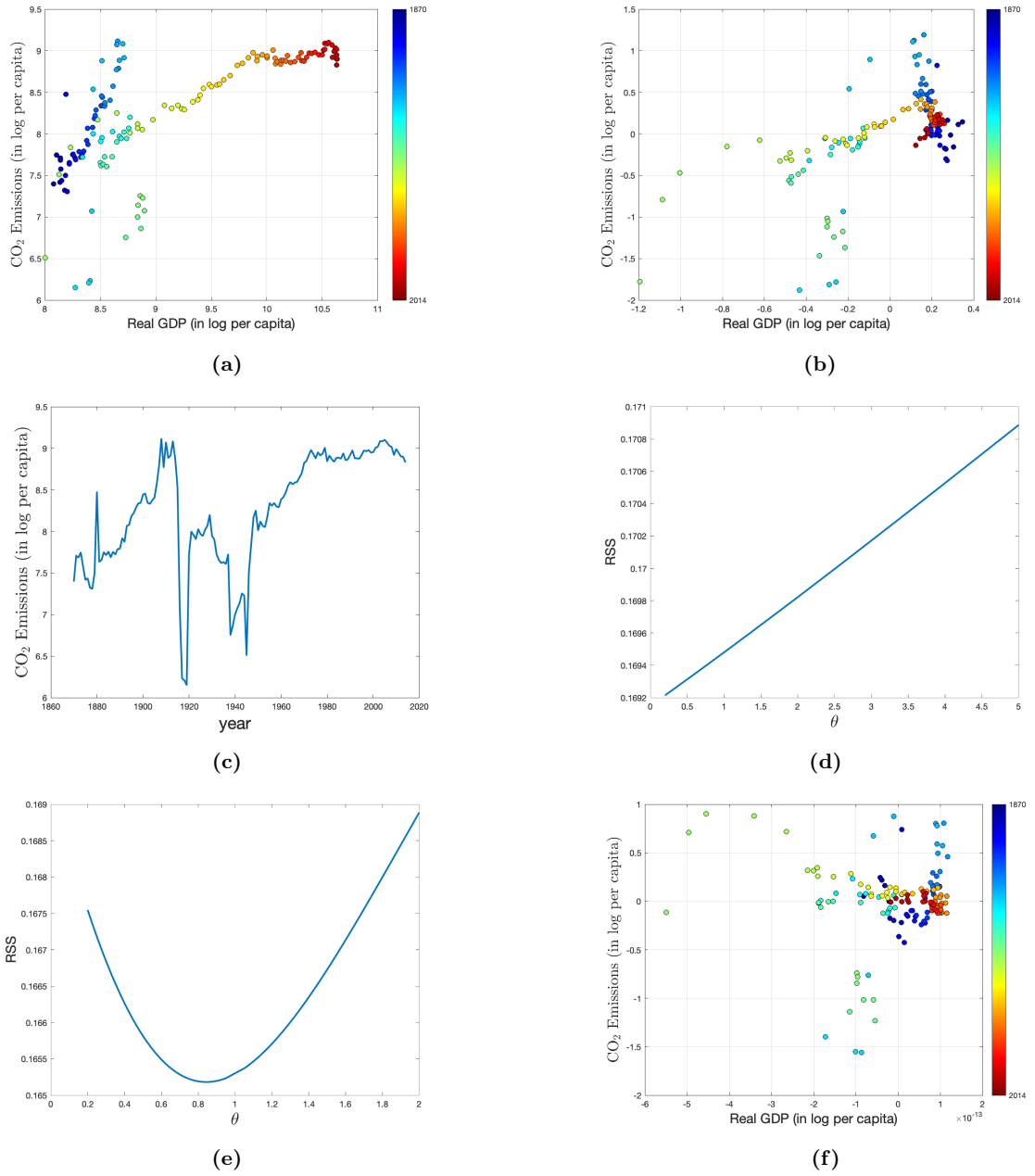


Figure 4: Overview graphs for Austria over 1870-2014. **(a)** log(GDP) versus log(CO_2) (both per capita). **(b)** As subfigure (a) but using detrended variables. **(c)** The log per capita CO_2 emissions time series for Austria. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ. **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

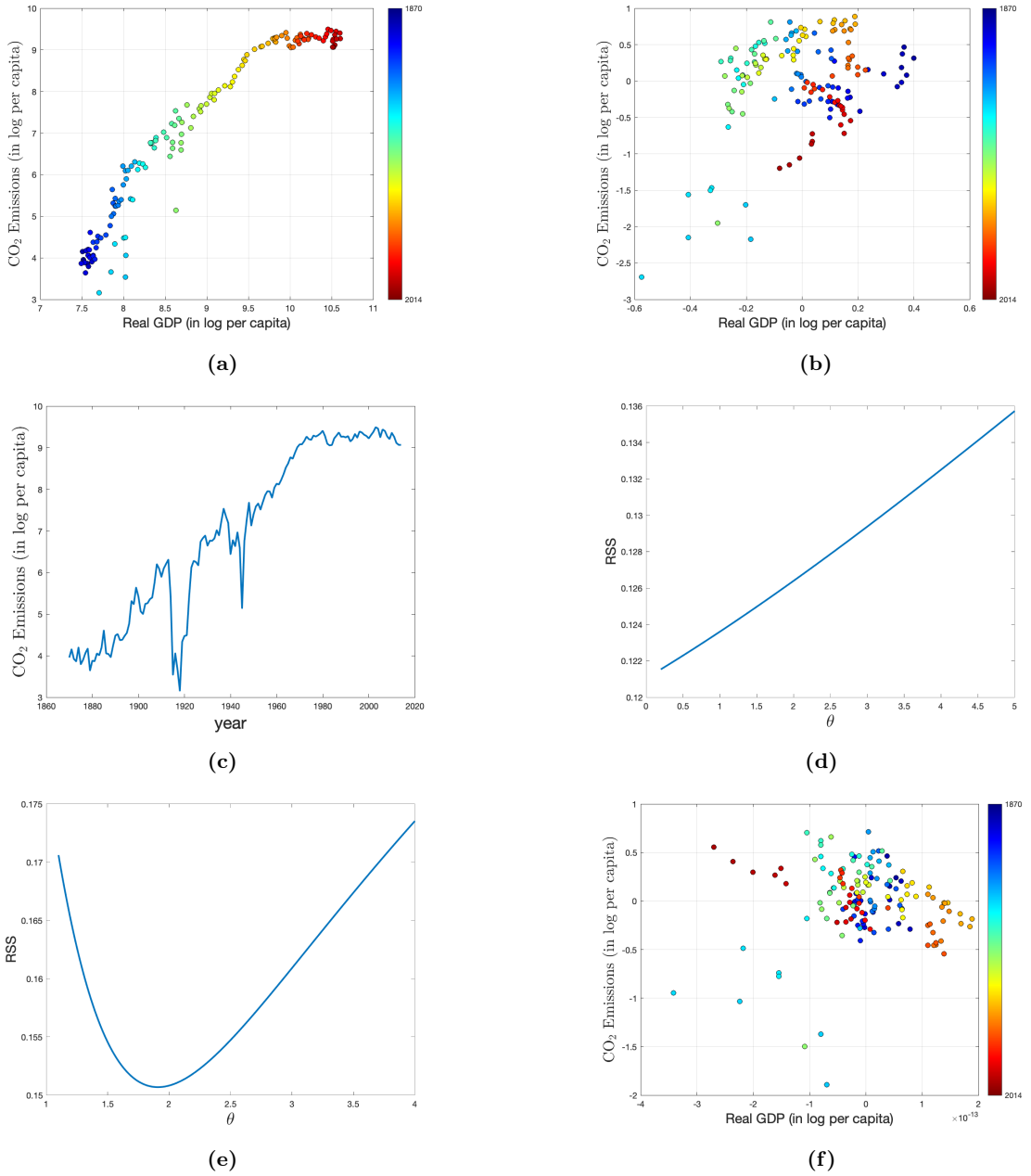


Figure 5: Overview graphs for Finland over 1870-2014. **(a)** log(GDP) versus log(CO_2) (both per capita). **(b)** As subfigure (a) but using detrended variables. **(c)** The log per capita CO_2 emissions time series for Finland. **(d)** The residual sum of squares (RSS) for the nonlinear model specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ. **(e)** The RSS as a function of θ for the flexible nonlinear trend specification $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi x_t + u_t$. **(f)** The relation between x_t and y_t after partialling out the constant, linear trend, and flexible deterministic trend.

S7.4 RSS(θ) for $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$

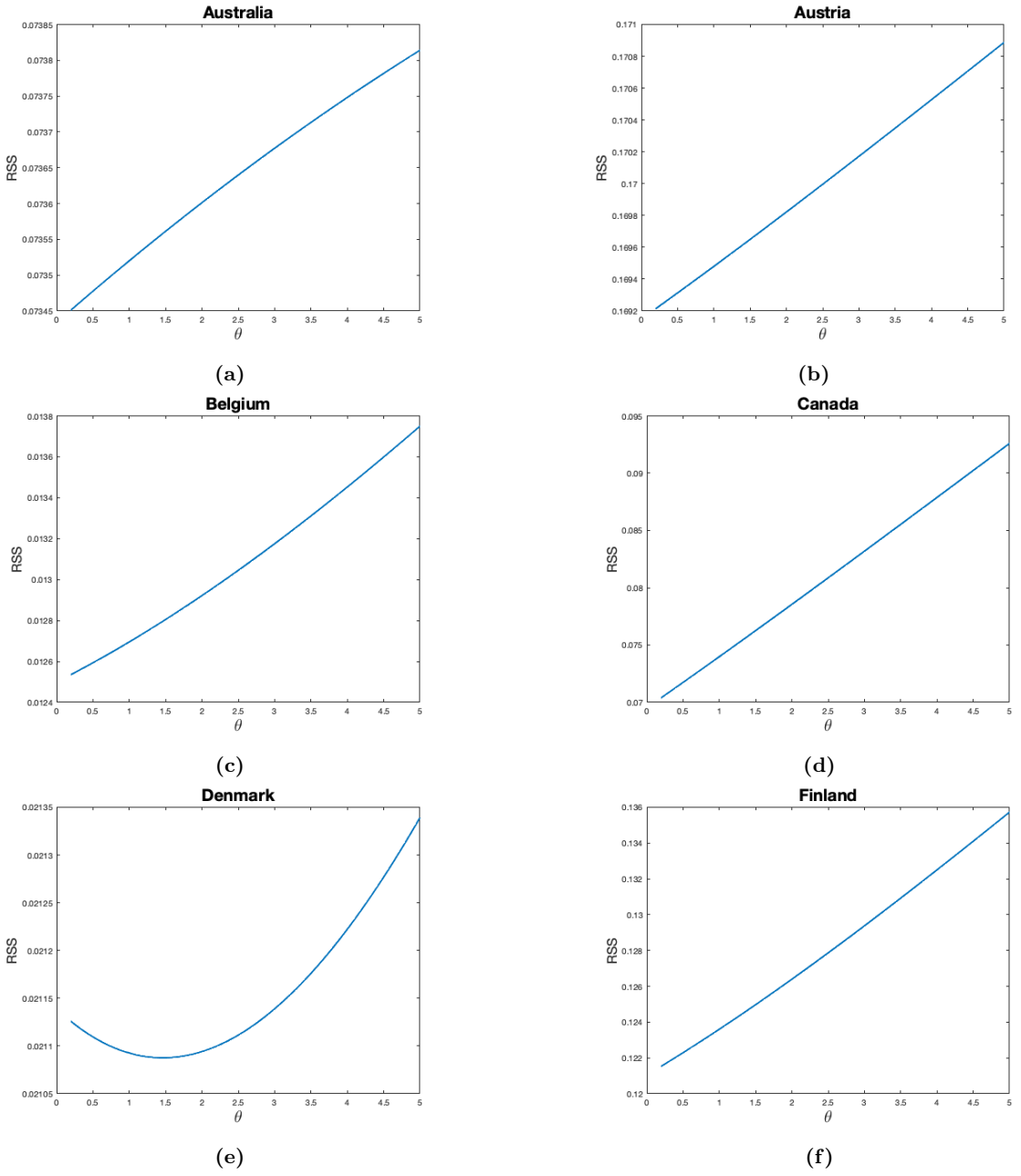
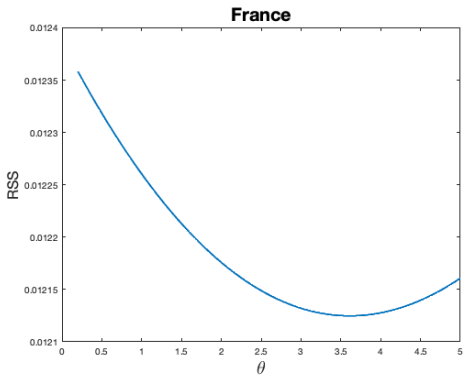
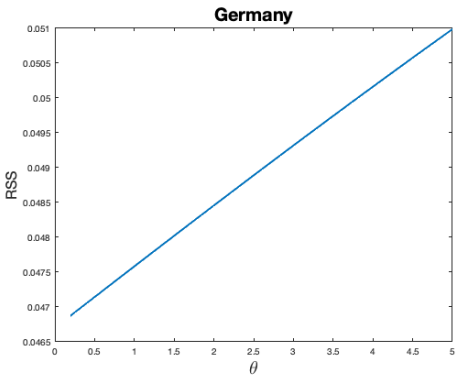


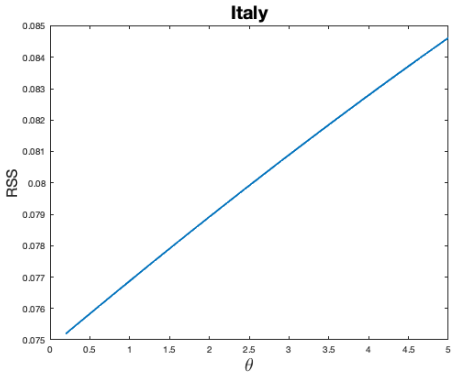
Figure 6: The residual sum of squares (RSS) for the nonlinear specification $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^\theta + u_t$ for various values of θ. This replicates Figure 1(d) of the main paper for all countries in the data set.



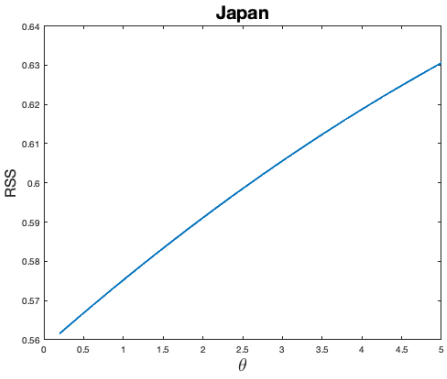
(g)



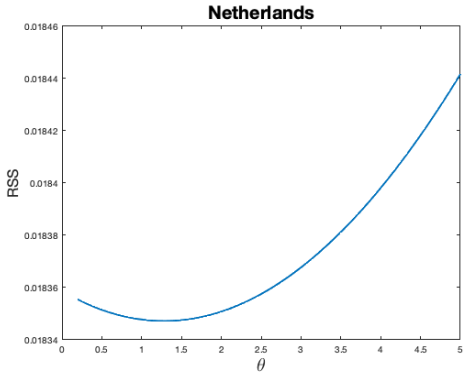
(h)



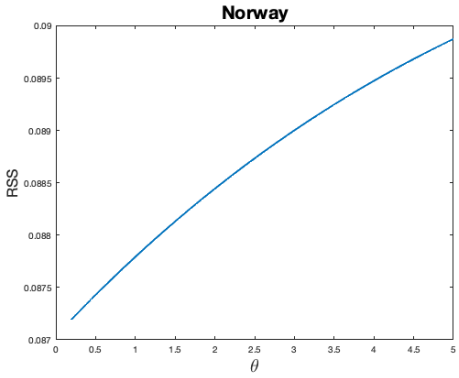
(i)



(j)

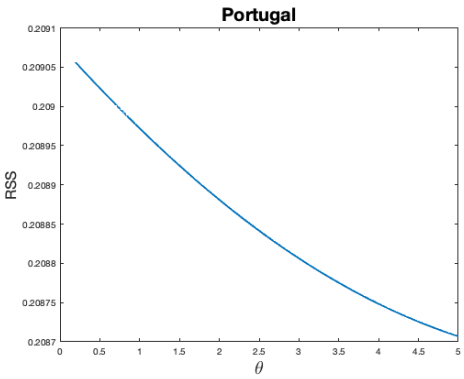


(k)

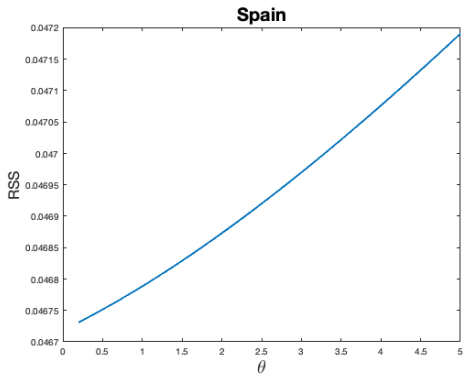


(l)

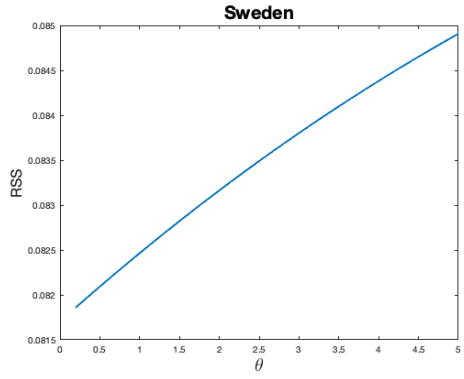
Continuation of Figure 6.



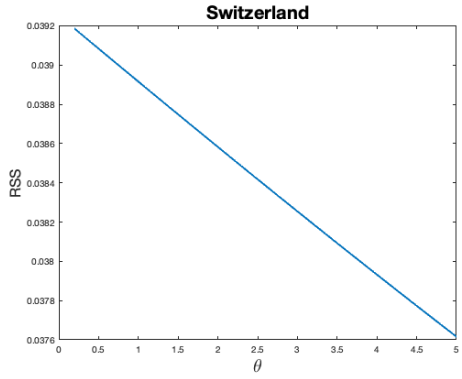
(m)



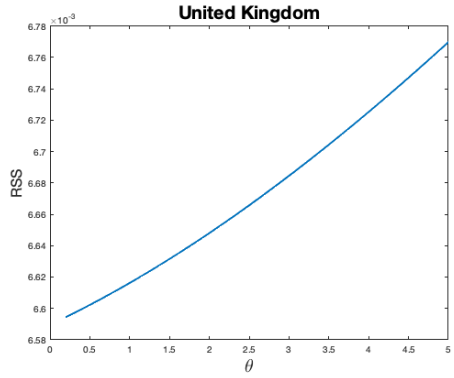
(n)



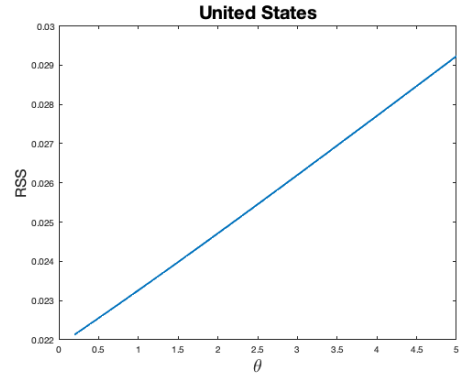
(o)



(p)



(q)



(r)

Continuation of Figure 6.

S7.5 Additional results for univariate models

Results of a more in-depth univariate analysis are collected in this section. We look at models (M1*)–(M4*) as listed in Table S5.

Table S5: An overview of the univariate models.

Model	Specification
(M1*)	$y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$
(M2*)	$y_t = \tau_1 + \tau_2 t + \tau_3 t^2 + \phi_1 x_t + \phi_2 x_t^2 + u_t$
(M3*)	$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t$
(M4*)	$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t$

All three models are of the form:

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t. \tag{S.1}$$

Model (M1*) is the specification above with $\tau_3 = 0$ and forces all nonlinearities to be captured through x_t^2. Specifications (M2*) and (M3*) include deterministic nonlinear time trends. For model (M2*), we allow for $\tau_3 \neq 0$ but fix $\theta = 2$. Model (S.1) without further restrictions is referred to as (M3*). In the latter model, the NLS estimator for θ is computed by a grid search over the values $\Theta = [0.05, 0.95] \cup [1.05, 10]$ and simulated inference is used (see Section 3.2 of the main paper). Table S6 illustrates how increasingly flexible nonlinear deterministic trends affect the parameter estimates for ϕ_1 and ϕ_2. Judging exclusively by the signs of $\widehat{\phi}_1$ and $\widehat{\phi}_2$, the EKC exists for 17 out of 18, 9 out of 18, and 8 out of 18 countries for (M1*), (M2*), and (M3*), respectively. Moreover, the significance of squared log per capita GDP (read: ϕ_2) reduces when nonlinear deterministic time trends are included. For model (M3*), ϕ_2 is never significantly different from zero at a 10% level and evidence in favour of EKC becomes rather meagre. The results of the univariate KPSS tests for these models can be found in Table S6 under "Stationarity tests". In general, the cointegrating relations seem well-specified except maybe for Belgium, Denmark, and UK.

The insignificance of ϕ_2 in model (M3*) suggests a final model specification, namely

$$y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t. \tag{M4*}$$

Model (M4*) specifies a linear cointegrating relation around a flexible time trend and does not incorporate nonlinear effects in log per capita GDP.[2] That is, the model specification does not allow for an EKC. As before, we check parameter estimates and test for stationarity of the error terms (the columns labeled "(M4*)" in Table S6). Some remarks concerning this final model specification are:

1. For Belgium, the fitted model reads

$$y_t = -0.049 + 0.0063t - 6.131 \times 10^{-6}\, t^{2.603} + 1.006 x_t + \hat{u}_t. \tag{S.2}$$

[2]Model specification (M4*) has the additional advantage of being invariant to the possible presence of a drift component in log per capita GDP, also see footnote 12 of the main text.

Table S6: Parameter estimates and output of the KPSS-type of test for stationarity as computed for model specifications (M1*)-(M4*). The column $\widehat{KPSS}$ and M_{opt} provide the numerical values of the KPSS tests and the number of chosen residual subblocks, respectively.

	Parameter estimates								Stationarity tests							
	(M1*)		(M2*)		(M3*)			(M4*)	(M1*)		(M2*)		(M3*)		(M4*)	
Country	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\theta}$	$\widehat{\phi}_1$	$\widehat{KPSS}$	M_{opt}	$\widehat{KPSS}$	M_{opt}	$\widehat{KPSS}$	M_{opt}	$\widehat{KPSS}$	M_{opt}
Australia	2.75	−0.17	−23.92***	1.40***	−12.19***	0.74	0.88	1.25***	1.49	9	1.79	9	1.80	9	1.35	9
Austria	7.13***	−0.30**	1.25	0.03	3.75***	−0.12	0.88	1.55***	1.03	7	1.07	7	1.65	7	1.63	7
Belgium	11.45***	−0.57***	10.03***	−0.49***	10.29***	−0.50	2.60	1.01***	1.91	9	2.92*	8	2.74*	8	2.28	9
Canada	12.72***	−0.64***	14.80	−0.77	−3.46***	0.25	0.56	1.14***	2.83*	7	2.60*	7	1.26	9	1.28	9
Denmark	14.52***	−0.65***	−2.80	0.25	−5.75***	0.39	2.03	1.68***	3.14*	9	3.30**	9	2.98*	9	1.58	8
Finland	16.86***	−0.76***	16.97***	−0.77***	22.58***	−1.06	1.87	3.95***	2.05	8	2.06	8	0.71	9	0.80	9
France	10.87***	−0.55***	3.14*	−0.12	3.31***	−0.13	2.09	1.00***	1.72	9	0.69	8	0.56	8	2.49	9
Germany	6.24***	−0.31***	−1.82	0.13	−4.42***	0.29	0.59	0.89***	2.63*	7	2.1	8	1.23	9	2.81*	7
Italy	11.76***	−0.55***	7.31**	−0.30	7.72***	−0.29	0.82	2.41***	4.18**	7	3.79**	8	0.79	7	0.78	7
Japan	9.86***	−0.52***	−4.27	0.29	1.16***	−0.00	0.05	1.15***	5.17***	8	3.93**	7	1.83	9	1.84	9
Netherlands	8.70***	−0.41***	1.49	−0.01	0.48**	0.05	1.86	1.32***	2.16	7	0.94	7	1.2	7	1.15	7
Norway	3.87	−0.16	−9.14**	0.51**	−1.10**	0.16	0.46	2.05***	2.53*	7	1.06	7	0.74	9	1.44	8
Portugal	0.09	0.04	−5.86***	0.42	−1.11**	0.15	0.05	1.69***	6.95***	8	5.28**	7	0.64	7	1.64	7
Spain	7.72***	−0.37***	1.98	−0.01	4.31***	−0.16	1.55	1.52***	2.78*	7	2.03	8	2.4	8	2.42	8
Sweden	10.91***	−0.44***	−9.08*	0.61**	0.43	0.17	0.46	3.48***	3.59**	7	1.27	7	0.73	7	0.80	7
Switzerland	8.57***	−0.29***	−7.86**	0.54***	−13.86***	0.83	2.98	2.63***	0.80	7	0.96	7	0.77	7	0.75	7
UK	9.32***	−0.47***	5.91***	−0.27***	4.13***	−0.18	3.04	0.80***	2.76*	9	4.25**	9	4.30**	9	3.98**	9
USA	8.67***	−0.44***	0.93	−0.03	−5.62***	0.35	0.92	0.95***	1.64	8	1.85	8	2.25	8	1.97	8

Note: Asterisks denote rejection of the null hypothesis at the ***1%, **5%, and *10% significance level. Depending on the specific table entry, the null hypothesis refers to either a coefficient being zero or (nonlinear) cointegration.

The flexible power on the linear trend is estimated to be $\widehat{\theta} = 2.603$ resulting in nonlinear behaviour over time. Moreover, the negative coefficient in front of $t^{2.603}$ provides a contribution that is sloping down over time. If time effects are ignored, then a 1% increase in GDP will lead to an estimated 1.006% increase in fossil-fuel CO_2 emissions.

2. The outcomes of the KPSS test do not point towards a misspecified cointegrating relation (Table S6). The flexible deterministic trend is generally sufficient to describe the nonlinear behaviour of the (univariate) log per capita CO_2 emissions over time, that is, *squared log per capita GDP is not needed in the univerariate models*. Visual proof is found in Figures 1(a), 1(b) and 1(f) where the incorporation of increasingly flexible time effects is seen to remove any apparent nonlinear relationship between log per capita GDP and CO_2 emissions.

Visualisations of the model fits are available in Figures 9–13.

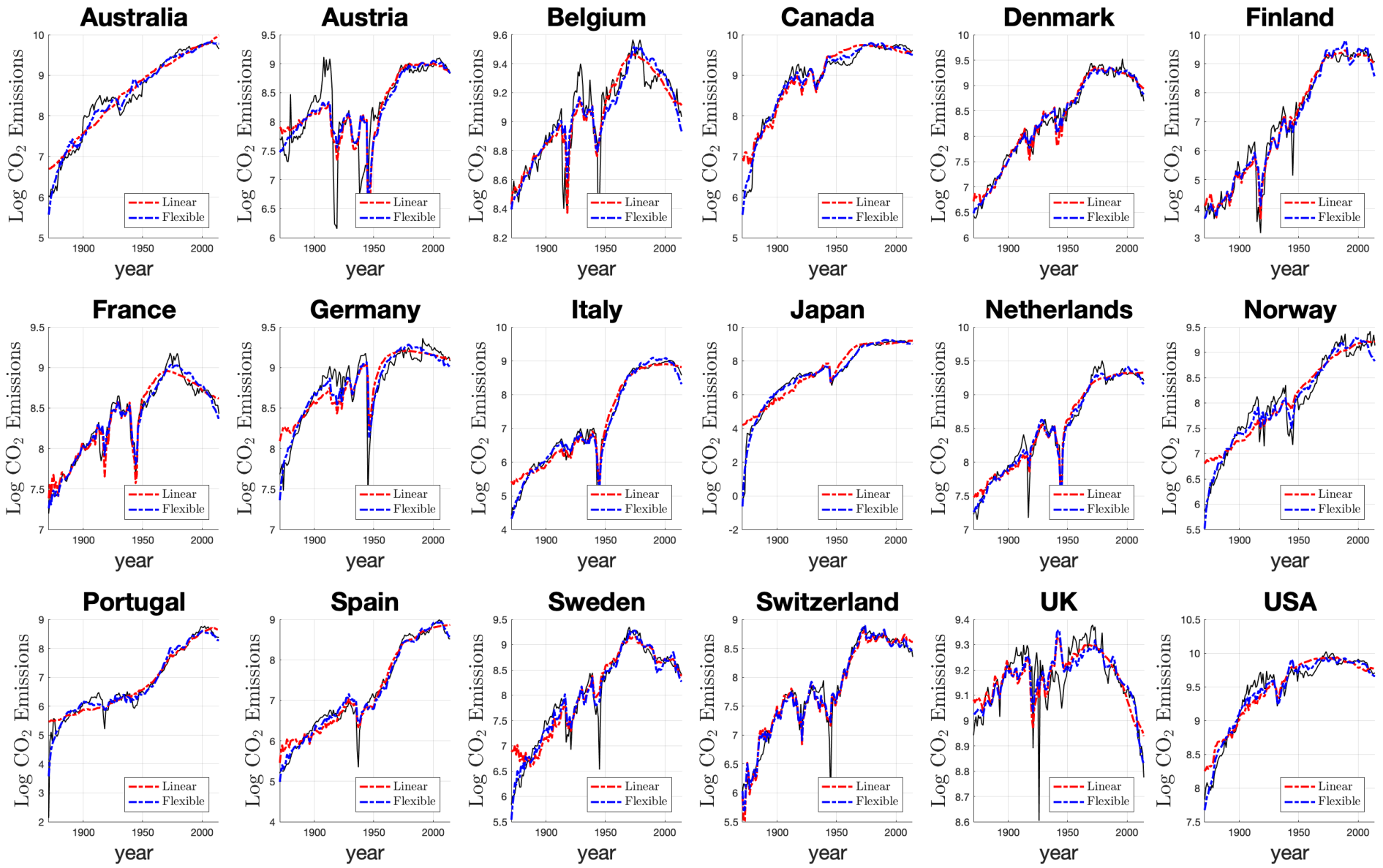


Figure 9: Estimation results for CO_2 emissions: actual values (black), fitted values under the CPR model $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$ (red), and fitted values under the GTACPR model $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + u_t$ (blue).

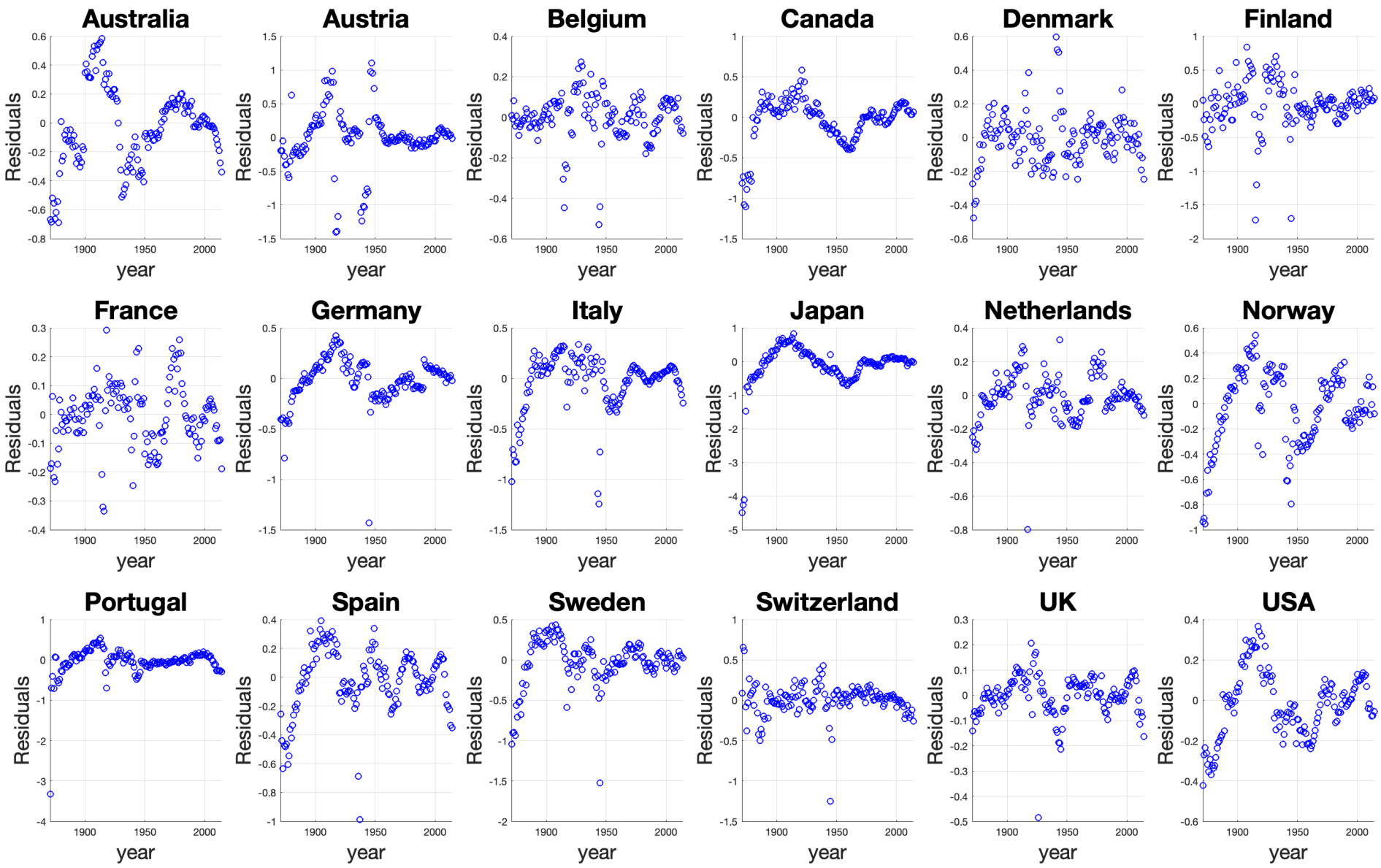


Figure 10: The residual series for each country under model specification (M1): $y_t = \tau_1 + \tau_2 t + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

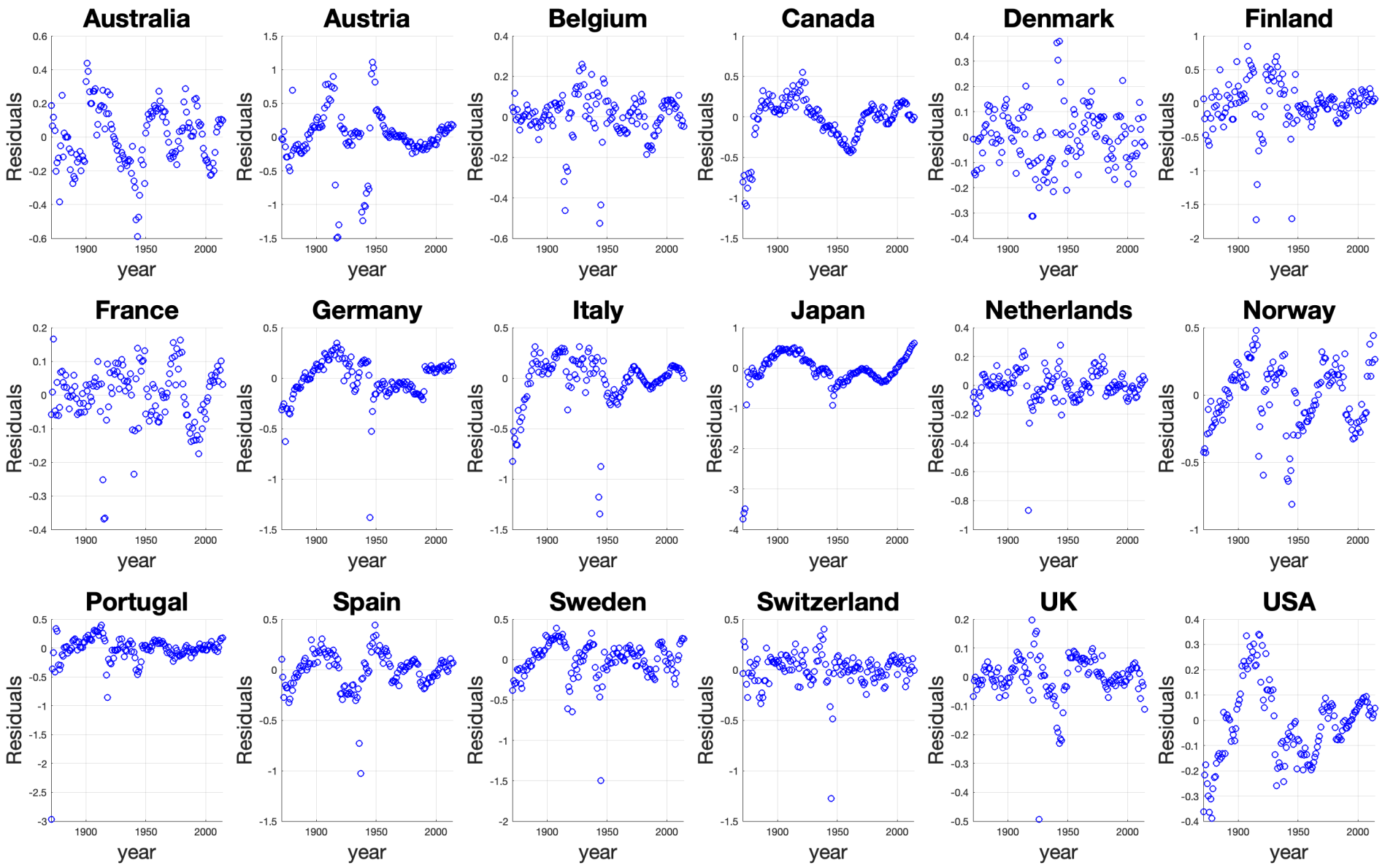


Figure 11: The residual series for each country under model specification (M2): $y_t = \tau_1 + \tau_2 t + \tau_3 t^2 + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

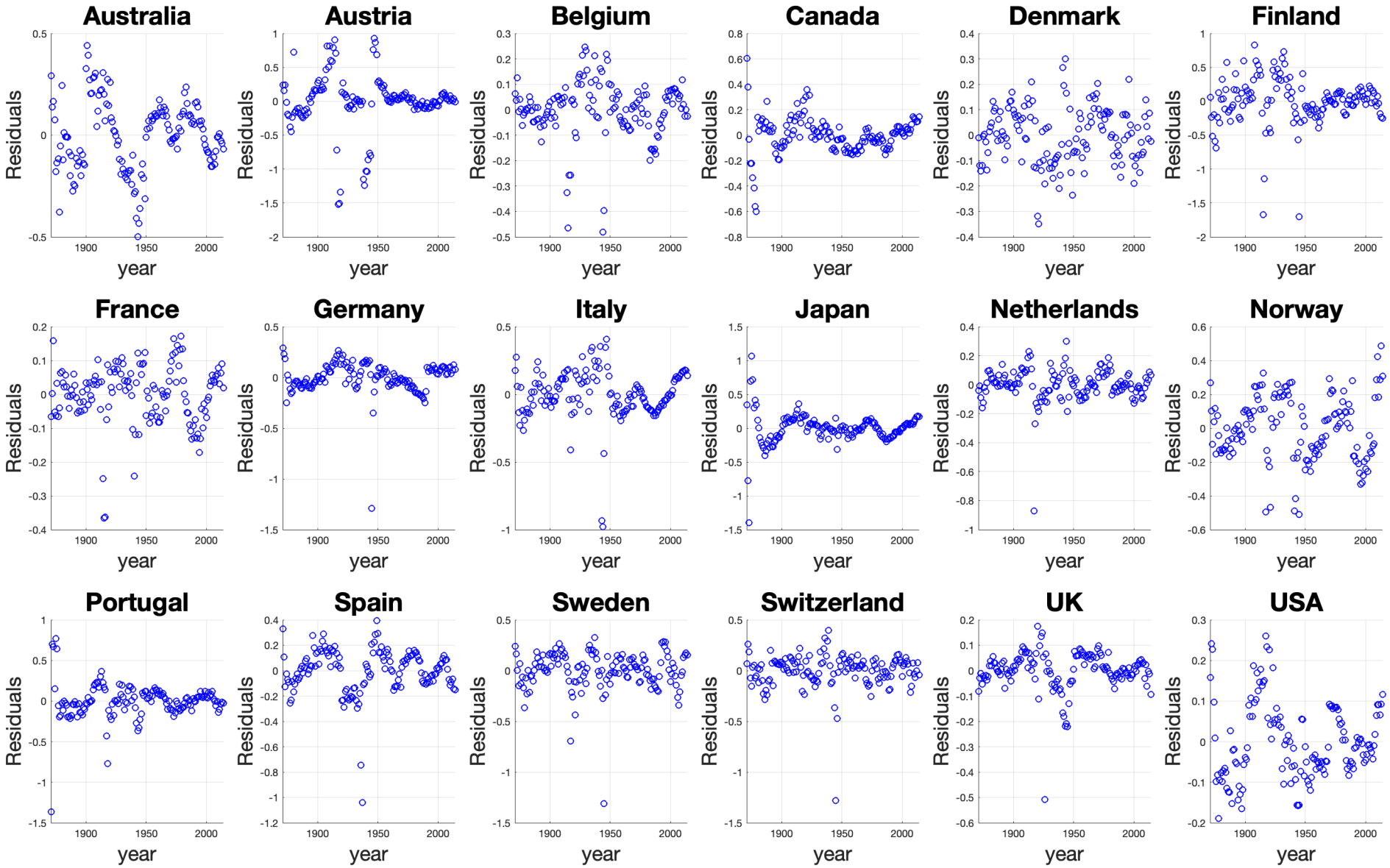


Figure 12: The residual series for each country under model specification (M3): $y_t = \tau_1 + \tau_2 t + \tau_3 t^\theta + \phi_1 x_t + \phi_2 x_t^2 + u_t$.

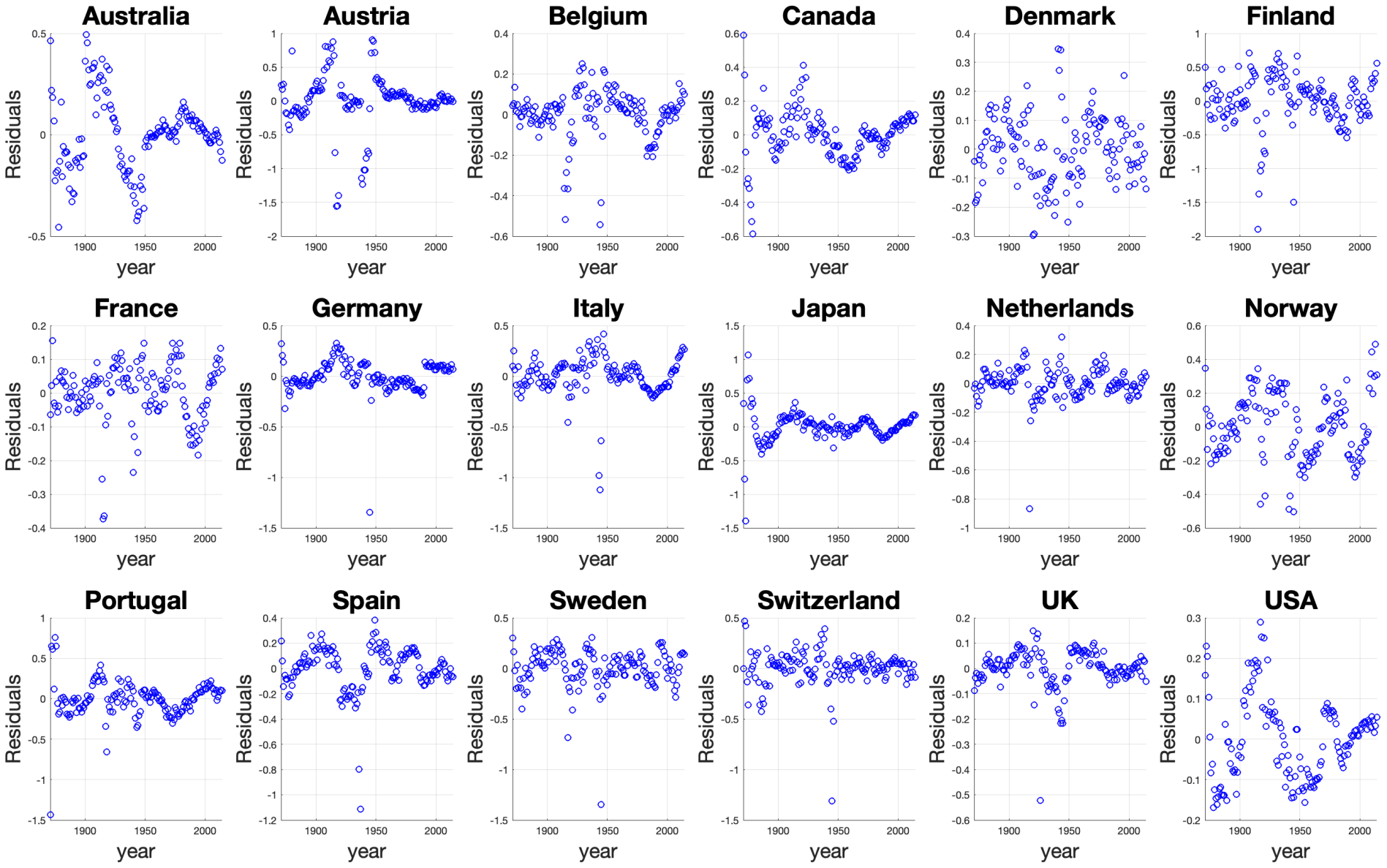


Figure 13: The residual series for each country under model specification (M4): $y_t = \tau_1 + \tau_2 t + \tau_3 t^{\theta} + \phi_1 x_t + u_t$.

S7.6 Nonparametric kernel estimator and linear fit

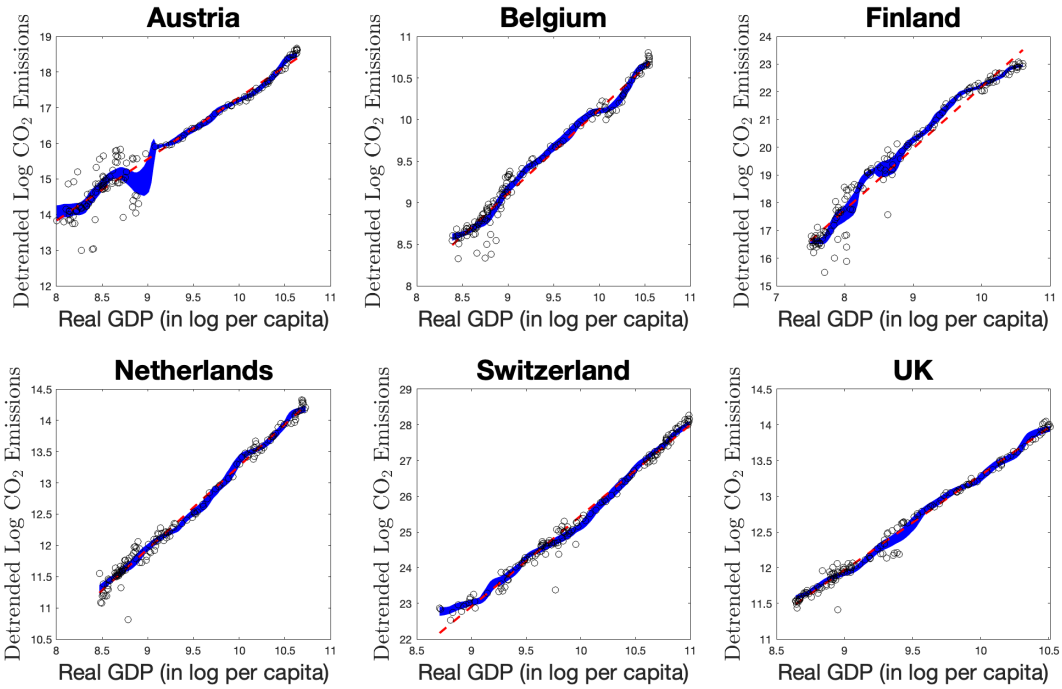


Figure 14: The 95% (point-wise) confidence intervals of the non-parametric kernel estimate for the relationship between GDP and CO_2 emissions (blue) after removal of the country-specific and joint flexible deterministic trends. The red dotted line is the linear fit. Results are based on the *full sample*.

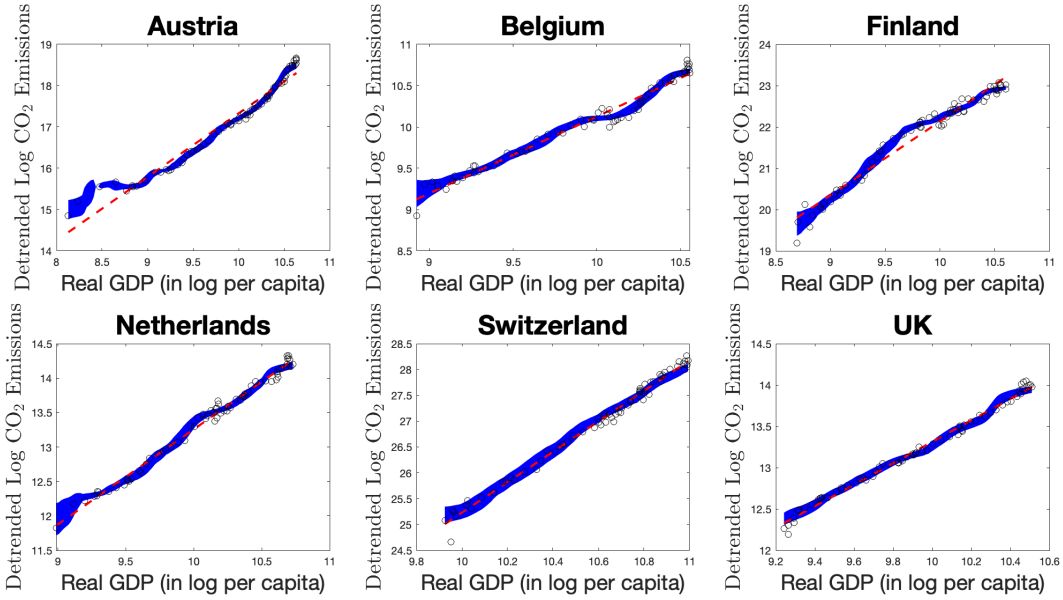


Figure 15: The 95% (point-wise) confidence intervals of the non-parametric kernel estimate for the relationship between GDP and CO_2 emissions (blue) after removal of the country-specific and joint flexible deterministic trends. The red dotted line is the linear fit. Results are based on *observations after World War II*.

The high p-values in Table S7 are caused by visually small deviations from the linear trend, see graphs above.

Also, the relatively small sample size (for nonparametric settings) might adversely affect power.

Table S7: Linearity test results. The linearity test is based on the model specification test documented in section 3 of Wang and Phillips (2016). The test is based on the integrated weighted squared deviations between the data and the linear model fit. We report the integration range, the (standardized) test statistic, and the p-value. Under H_0, the relationship between (detrended) log CO_2 emissions and log GDP per capita is linear.

	Full Sample			After World War II		
	range	$\frac{\phi}{\tau_0\sqrt{nh}}T_n$	p-value	range	$\frac{\phi}{\tau_0\sqrt{nh}}T_n$	p-value
Austria	[8.003,10.635]	8.987	0.000	[8.129,10.635]	0.162	0.871
Belgium	[8.389,10.553]	0.299	0.765	[8.923,10.553]	0.076	0.939
Finland	[7.494,10.602]	0.105	0.916	[8.693,10.602]	0.020	0.984
Netherlands	[8.469,10.728]	0.046	0.964	[8.990,10.728]	0.022	0.982
Switzerland	[8.708,10.993]	0.028	0.977	[9.925,10.993]	0.009	0.993
UK	[8.641,10.510]	0.022	0.982	[9.242,10.510]	0.021	0.984

Note: The asymptotic properties of $\phi T_n/\tau_0\sqrt{nh}$ are established in Wang and Phillips (2016). Under suitable conditions, $\phi T_n/\tau_0\sqrt{nh} \to L_W(1,0)$ as $n \to \infty$ with $L_W(1,0)$ denoting the sojourning time of a standard Brownian motion around zero during the time interval $[0,1]$. The p-values are computed using the cumulative distribution function of $L_W(1,0)$, see (2.11) in Dong et al. (2017).

S8 Simulation and Calculations Related to FMOLS

S8.1 Preliminary simulations

We consider $N = 1$ and test $H_0 : \phi_2 = 0$ versus $H_a : \phi_2 \neq 0$ with FMOLS. Specifically, we generate the data according to

$$y_t = \tau_1 + \tau_2 t + \tau_g t^{\,\theta} + \phi_1 x_t + \phi_2 x_t^2 + u_t, \tag{S.1}$$

where $x_t = \sum_{s=1}^{t} v_s$. The chosen parameter values are $\theta = 2$, $\boldsymbol{\tau} = [\tau_1, \tau_2, \tau_g]' = [7, 0.05, -5 \times 10^{-4}]'$, and $\boldsymbol{\phi} = [\phi_1, \phi_2]' = [5, 0]'$. These parameter values are representative. The disturbance vector $[u_t, v_t]'$ is generated from the VAR(1) specification[3]

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}, \qquad \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix} \overset{i.i.d.}{\sim} \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right). \tag{S.2}$$

We construct the autoregressive matrix $\boldsymbol{A}$ along the following two steps: (1) generate a (2×2) random matrix $\boldsymbol{U}$ from U$[0,1]$ to construct the orthogonal matrix $\boldsymbol{H} = \boldsymbol{U}\left(\boldsymbol{U}'\boldsymbol{U}\right)^{-1/2}$, and (2) compute $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{L}\boldsymbol{H}'$ with $\boldsymbol{L} = \mathrm{diag}[0.9, 0.7]$.

As shown in Figure 16, for sample sizes as large as $15,000$, the empirical size of the feasible FMOLS estimator seems to stabilisze at 11% whereas the infeasible estimator FMOLS(θ_0) yields an empirical size close to 5%. These results indicate poor finite sample performance of FMOLS or possible even a lack of asymptotic validity.

[3]We start the VAR recursions from $\left[\begin{smallmatrix} u_0 \\ v_0 \end{smallmatrix}\right] = \mathbf{0}$ and use a presample of 50 observations to reduce the influence of these initial values.

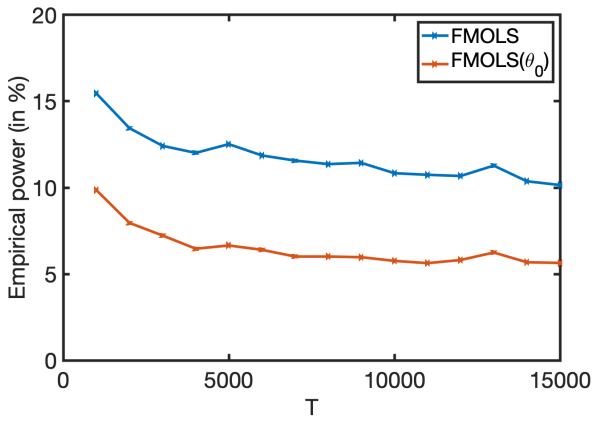


Figure 16: The empirical size of feasible and infeasible FMOLS estimators for a large range of sample sizes.

S8.2 Asymptotic properties of FMOLS

We comment on the asymptotic properties of the FMOLS estimator when $N = 1$. To shorten notation, the subscript 'i' in $x_{i,t}$, $y_{i,t}$ and p_i will be omitted. We analyse the asymptotic properties of $\widetilde{\boldsymbol{D}}_{\theta_0,T}\begin{bmatrix}\widehat{\tau}^{+}_{g,T}-\boldsymbol{\tau}_{g,0}\\ \widehat{\boldsymbol{\beta}}^{+}_{T}-\boldsymbol{\beta}_0\end{bmatrix}$, with $\widetilde{\boldsymbol{D}}_{\theta_0,T}=\sqrt{T}\begin{bmatrix}T^{\theta_0} & \mathbf{0}_{1\times(p+2)}\\ \mathbf{0}_{(p+2)\times 1} & \boldsymbol{D}_{(1),T}\end{bmatrix}$ and

$$\begin{bmatrix}\widehat{\tau}^{+}_{g,T}\\ \widehat{\boldsymbol{\beta}}^{+}_{T}\end{bmatrix}=\left(\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\boldsymbol{z}_t(\widehat{\theta}_T)'\right)^{-1}\left(\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)y_t^{+}-\boldsymbol{A}^{*}\right),$$

where $\boldsymbol{z}_t(\theta)=[t^{\theta},1,t,x_{1,t},\ldots,x_{1,t}^{p_1}]'$, and y_t^{+} and $\boldsymbol{A}^{*}$ are second-order bias corrections. That is, $y_t^{+}=y_t-\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\Delta\boldsymbol{x}_t$ and $\boldsymbol{A}^{*}=[\mathbf{0}'_{3\times1},\boldsymbol{A}_1^{*\prime}]'$ with $\boldsymbol{A}_1^{*}=\widehat{\boldsymbol{\Delta}}_{vu}^{+}\left[T,2\sum_{t=1}^{T}x_t,\ldots,p\sum_{t=1}^{T}x_t^{p-1}\right]'$ and $\widehat{\boldsymbol{\Delta}}_{vu}^{+}$ equals $\widehat{\boldsymbol{\Delta}}_{vu}^{+}=\widehat{\boldsymbol{\Delta}}_{vu}-\widehat{\boldsymbol{\Delta}}_{vv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\widehat{\boldsymbol{\Omega}}_{vu}$.

We now investigate how the estimation of θ affects the limiting distribution of the FMOLS estimator. By straightforward linear algebra manipulations, we find

$$\widetilde{\boldsymbol{D}}_{\theta_0,T}\begin{bmatrix}\widehat{\boldsymbol{\tau}}^{+}_{g,T}-\boldsymbol{\tau}_{g,0}\\ \widehat{\boldsymbol{\beta}}^{+}_{T}-\boldsymbol{\beta}_0\end{bmatrix}=\left(\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\boldsymbol{z}_t(\widehat{\theta}_T)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\right)^{-1}\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\left[\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\tilde{u}_t^{+}-\boldsymbol{A}^{*}\right],\tag{S.3}$$

where $\tilde{u}_t^{+}=\left(\boldsymbol{z}_t(\theta_0)-\boldsymbol{z}_t(\widehat{\theta}_T)\right)'\begin{bmatrix}\boldsymbol{\tau}_{g,0}\\ \boldsymbol{\beta}_0\end{bmatrix}+u_t-\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}\Delta x_t$. We will discuss $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\boldsymbol{z}_t(\widehat{\theta}_T)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}$ and $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\left[\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\tilde{u}_t^{+}-\boldsymbol{A}^{*}\right]$ separately after having enumerate several intermediate results.

Lemma S.8.4. *Define* $\widetilde{\boldsymbol{j}}(r;\theta_0)=\left[r^{\theta_0},1,r,B_v(r),\ldots,B_v^{p}(r)\right]'$ *and* $B_{u.v}=B_u-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}B_v$. *Then, under Assumptions 3.1-3.3, we have*

(i) $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\boldsymbol{z}_t(\widehat{\theta}_T)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\longrightarrow_d\int\widetilde{\boldsymbol{j}}(r;\theta_0)\widetilde{\boldsymbol{j}}(r;\theta_0)'dr$,

(ii) $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\left[\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\left(u_t-\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}v_t\right)-\boldsymbol{A}^{*}\right]\longrightarrow_d\int\widetilde{\boldsymbol{j}}(r;\theta_0)dB_{u.v}(r)$,

(iii) $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)'\begin{bmatrix}\boldsymbol{\tau}_{g,0}\\ \boldsymbol{\phi}_0\end{bmatrix}=O_p\Big(\ln T\Big)$,

(iv) $\sum_{t=1}^{T}\widetilde{\boldsymbol{D}}_{\theta_0,b_T}^{-1}\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)'\begin{bmatrix}\boldsymbol{\tau}_{g,0}\\ \boldsymbol{\phi}_0\end{bmatrix}=O_p\left((\ln T)^2T^{-(\theta_L+\frac{1}{2})}\right)$,

(v) $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)\left(u_t-\widehat{\boldsymbol{\Omega}}_{uv}\widehat{\boldsymbol{\Omega}}_{vv}^{-1}v_t\right)=o_p(1)$.

Proof ***(i)*** We can always add and subtract such that the LHS of (i) reads

$$\begin{aligned}\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\boldsymbol{z}_t(\widehat{\theta}_T)'\widetilde{\boldsymbol{D}}_{\theta_0,T}&=\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\boldsymbol{z}_t(\theta_0)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\\&+\left(\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\boldsymbol{z}_t(\widehat{\theta}_T)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}-\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\boldsymbol{z}_t(\theta_0)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\right).\end{aligned}\tag{S.4}$$

Lemma S.3.2(iii) implies that the first term in the RHS of (S.4) converges to $\int\widetilde{\boldsymbol{j}}(r;\theta_0)\widetilde{\boldsymbol{j}}(r;\theta_0)'dr$. It remains to show that the term in parenthesis vanishes. By $\sum_t\boldsymbol{a}_t\boldsymbol{a}_t'-\sum_t\boldsymbol{b}_t\boldsymbol{b}_t'=\sum_t(\boldsymbol{a}_t-\boldsymbol{b}_t)(\boldsymbol{a}_t-\boldsymbol{b}_t)'+\sum_t(\boldsymbol{a}_t-\boldsymbol{b}_t)\boldsymbol{b}_t'+\sum_t\boldsymbol{b}_t(\boldsymbol{a}_t-\boldsymbol{b}_t)'$ and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}&\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\widehat{\theta}_T)\boldsymbol{z}_t(\widehat{\theta}_T)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}-\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\boldsymbol{z}_t(\theta_0)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\right\|\\&\le\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)\right\|^2+2\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\boldsymbol{z}_t(\theta_0)\right\|\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)\right\|\\&\le\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)\right\|^2+2\sqrt{\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\boldsymbol{z}_t(\theta_0)\right\|^2}\sqrt{\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)\right\|^2}.\end{aligned}$$

We have $\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\boldsymbol{z}_t(\theta_0)\right\|^2=\operatorname{tr}\left(\sum_{t=1}^{T}\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\boldsymbol{z}_t(\theta_0)\boldsymbol{z}_t(\theta_0)'\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\right)\longrightarrow_d\operatorname{tr}\left(\int\widetilde{\boldsymbol{j}}(r;\theta_0)\widetilde{\boldsymbol{j}}(r;\theta_0)'dr\right)$. Next note that $\sum_{t=1}^{T}\left\|\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0))\right\|^2=\frac{1}{T}\sum_{t=1}^{T}[T^{-\theta_0}(t^{\widehat{\theta}_T}-t^{\theta_0})]^2$. We have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left[T^{-\theta_0}\left(t^{\widehat{\theta}_T}-t^{\theta_0}\right)\right]^2&\le C\left(\widehat{\theta}_T-\theta_0\right)\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_0}(\ln t)^2\\&\le CT^{-2(\theta_0+\frac{1}{2})}(\ln T)^2\left[T^{\theta_0+\frac{1}{2}}\left(\widehat{\theta}_T-\theta_0\right)\right]^2\sup_{\theta_L\le\theta\le\theta_U}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta}\right|=o_p(1),\end{aligned}\tag{S.5}$$

where we used the mean-value theorem and Lemma S.3.1(i). The claim follows. ***(ii)*** $\widehat{\boldsymbol{\Omega}}_{uv}$ and $\widehat{\boldsymbol{\Omega}}_{vv}$ consistently estimate $\boldsymbol{\Omega}_{uv}$ and $\boldsymbol{\Omega}_{vv}$, respectively (Theorem 3.2). It therefore suffices to look at the quantities $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\left(u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}\boldsymbol{v}_t\right)$ and $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\boldsymbol{A}^*$. Lemma S.3.2(ii) with $u_t^+=u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}v_t$ instead of u_t gives the limiting result $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(x_t/\sqrt{T}\right)^j u_t^+\longrightarrow_d\int_0^1\boldsymbol{B}_v^j(r)dB_{u.v}(r)+j\boldsymbol{\Delta}_{vu}^+\int_0^1\boldsymbol{B}_v^{j-1}(r)dr$, which implies

$$\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\left(u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}v_t\right)\longrightarrow_d\int\widetilde{\boldsymbol{j}}(r;\theta_0)dB_{u.v}(r)+\widetilde{\boldsymbol{\mathcal{B}}}_{vu}^{+},\tag{S.6}$$

where $\widetilde{\boldsymbol{\mathcal{B}}}_{vu}^{+}=\left[\boldsymbol{0}_{3\times1}',\boldsymbol{b}'\boldsymbol{\Delta}_{vu}^{+}\right]'$. The term $-\widetilde{\boldsymbol{D}}_{\boldsymbol{\theta}_0,T}^{-1}\boldsymbol{A}^*$ is constructed to asymptotically cancel out the term $\widetilde{\boldsymbol{\mathcal{B}}}_{vu}^{+}$ in the RHS of (S.6). ***(iii)*** Using $\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)=\left[t^{\widehat{\theta}_T}-t^{\boldsymbol{\theta}_0}\quad\boldsymbol{0}'\right]'$, we have

$$\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)'\left[\begin{smallmatrix}\tau_{g,0}\\ \beta_0\end{smallmatrix}\right]=\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\left(t^{\widehat{\theta}_T}-t^{\boldsymbol{\theta}_0}\right)\tau_{g,0}.$$

We note that the typical elements in the vector on the RHS are of the form $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_0}\tau_{g,0}\big(t^{\widehat{\theta}_T}-t^{\theta_0}\big)$ or $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{j}\tau_{g,0}\big(t^{\widehat{\theta}_T}-t^{\theta_0}\big)$. We show that both contributions are $O_p\left(\ln T\right)$. By the mean-value theorem and Lemma S.3.1(i),

$$\begin{aligned}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_0}\tau_{g,0}\big(t^{\widehat{\theta}_T}-t^{\theta_0}\big)\right| &\leq \left|\frac{1}{\sqrt{T}}\tau_{g,0}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_0}t^{\theta_0}\left(t^{\widehat{\theta}_T-\theta_0}-1\right)\right| \\ &\leq C|\tau_{g,0}|\left|T^{\theta_0+\frac{1}{2}}(\widehat{\theta}_T-\theta_0)\right|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_0}\ln t \\ &\leq C(\ln T)|\tau_{g,0}|\left|T^{\theta_0+\frac{1}{2}}(\widehat{\theta}_T-\theta_0)\right|\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_0}\right]=O_p(\ln T).\end{aligned} \tag{S.7}$$

Similarly, from the mean-value theorem and Cauchy-Schwartz inequality, we see that

$$\begin{aligned}&\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{j}\tau_{g,0}\big(t^{\widehat{\theta}_T}-t^{\theta_0}\big)\right| \leq \left|\frac{1}{\sqrt{T}}\tau_{g,0}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{j}t^{\theta_0}\left(t^{\widehat{\theta}_T-\theta_0}-1\right)\right| \\ &\leq C|\tau_{0k}|\left|T^{\theta_0+\frac{1}{2}}(\widehat{\theta}_T-\theta_0)\right|\frac{1}{T}\sum_{t=1}^{T}\left|\frac{x_{it}}{\sqrt{T}}\right|^{j}\left(\frac{t}{T}\right)^{\theta_0}\ln t \\ &\leq C(\ln T)|\tau_{g,0}|\left|T^{\theta_0+\frac{1}{2}}(\widehat{\theta}_T-\theta_0)\right|\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_{it}}{\sqrt{T}}\right)^{2j}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_L}}.\end{aligned} \tag{S.8}$$

From (S.7) and (S.8) we conclude that $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t(\theta_0)\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)'\left[\begin{smallmatrix}\tau_{g,0}\\ \boldsymbol{\beta}_0\end{smallmatrix}\right]=O_p(\ln T)$. ***(iv)*** Use $\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)=\left[t^{\widehat{\boldsymbol{\theta}}_T}-t^{\boldsymbol{\theta}_0}\quad \mathbf{0}'\right]'$ to obtain $\widetilde{\boldsymbol{D}}_{\theta_0,T}^{-1}\sum_{t=1}^{T}\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)\left(\boldsymbol{z}_t(\widehat{\theta}_T)-\boldsymbol{z}_t(\theta_0)\right)'\left[\begin{smallmatrix}\tau_{g,0}\\ \boldsymbol{\beta}_0\end{smallmatrix}\right]$ $=\begin{bmatrix}T^{-\theta_0}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(t^{\widehat{\theta}_T-t^{\boldsymbol{\theta}_0}}\right)^2\tau_{g,0}\\ \mathbf{0}\end{bmatrix}$. The absolute value of the nonzero element can be bounded as follows

$$\begin{aligned}&\left|\tau_{g,0}\frac{1}{T^{\theta_0+1/2}}\sum_{t=1}^{T}\left(t^{\widehat{\theta}_T}-t^{\theta_0}\right)^2\right| \leq |\tau_{g,0}|\frac{1}{T^{\theta_0+1/2}}\sum_{t=1}^{T}t^{2\theta_0}\left|t^{\widehat{\theta}_T-\theta_0}-1\right|\left|t^{\widehat{\theta}_T-\theta_0}-1\right| \\ &\leq C|\tau_{g,0}|\left|\widehat{\theta}_T-\theta_0^2\right|\frac{1}{T^{\theta_0+1/2}}\sum_{t=1}^{T}t^{2\theta_{0i}}(\ln t)^2 \\ &\leq C(\ln T)^2T^{-(\theta_L+\frac{1}{2})}|\tau_{g,0}|\left|T^{\theta_0+\frac{1}{2}}(\widehat{\theta}_T-\theta_0)\right|^2\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{2\theta_0}\right]=O_p\left(\frac{(\ln T)^2}{T^{\theta_L+\frac{1}{2}}}\right).\end{aligned}$$

(v) By similar steps as before, and invoking Theorem 3.2, it is easy to show that it suffices to bound $T^{-(\theta_0+\frac{1}{2})}\sum_{t=1}^{T}\left(t^{\widehat{\theta}_T}-t^{\theta_0}\right)\left(u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}v_t\right)$. Writing $u_t^+=u_t-\boldsymbol{\Omega}_{uv}\boldsymbol{\Omega}_{vv}^{-1}v_t$, we have

$$\begin{aligned}&T^{-(\theta_0+\frac{1}{2})}\sum_{t=1}^{T}\left(t^{\widehat{\theta}_T}-t^{\theta_0}\right)u_t^+=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_0}\left(t^{\widehat{\theta}_T-\theta_0}-1\right)u_t^+=\left(\widehat{\theta}_T-\theta_0\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ln t)\left(\frac{t}{T}\right)^{\theta_0}u_t^++o_p(1) \\ &=T^{-(\theta_0+\frac{1}{2})}\left[T^{\theta_0+\frac{1}{2}}\left(\widehat{\theta}_T-\theta_0\right)\right]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\ln\frac{t}{T}\right)\left(\frac{t}{T}\right)^{\theta_0}u_t^+ \\ &\quad+T^{-(\theta_0+\frac{1}{2})}\left[T^{\theta_0+\frac{1}{2}}\left(\widehat{\theta}_T-\theta_0\right)\right](\ln T)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{t}{T}\right)^{\theta_0}u_t^++o_p(1)=\frac{1}{T^{\theta_0+\frac{1}{2}}}O_p(1)+\frac{\ln T}{T^{\theta_0+\frac{1}{2}}}O_p(1).\end{aligned}$$

This establishes (v). ■

The currents upper bounds in the lemma above suggest that the RHS of (S.3) does not converge to a Gaussian mixture limiting distribution. The problematic expression is Lemma S.8.4(iii). That is, if θ is estimated, then $\boldsymbol{z}_t\big(\widehat{\theta}_T\big)-\boldsymbol{z}_t\big(\theta_0\big)$ does not convergence sufficiently fast to zero to obtain the standard stochastic integral.

References

Adams, R. A. and C. Essex (2016). *Calculus: A Complete Course*. Pearson Canada.

Dong, C., J. Gao, D. Tjøstheim, and J. Yin (2017). Specification testing for nonlinear multivariate cointegrating regressions. *Journal of Econometrics 200*, 104–117.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hong, S. H. and P. C. B. Phillips (2010). Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business & Economic Statistics 28*, 96–114.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory 18*, 1449–1459.

Perron, P. and T. Yabu (2009). Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics 151*, 56–69.

Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory 23*, 557–614.

Robinson, P. M. (2012). Inference on power law spatial trends. *Bernoulli 18*, 644–677.

Soong, T. T. (1973). *Random Differential Equations in Science and Engineering*. Academic Press, Inc.

Tanaka, K. (2017). *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. John Wiley & Sons.

Wagner, M., P. Grabarczyk, and S. H. Hong (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*, 216–255.

Wang, Q. and P. C. B. Phillips (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory 32*, 359–401.