Center for European, Governance and Economic Development Research

Discussion Papers

Number 314 – June 2017

HETEROSKEDASTICITY-ROBUST UNIT ROOT TESTING FOR TRENDING PANELS

Helmut Herwartz
Simone Maxand
Yabibal M. Walle

Georg-August-Universität Göttingen

ISSN: 1439-2305

Heteroskedasticity-robust unit root testing for trending panels

Helmut Herwartz*, Simone Maxand and Yabibal M. Walle

Georg-August-University Göttingen, Germany.

Abstract Standard panel unit root tests (PURTs) are not robust to breaks in innovation variances. Consequently, recent papers have proposed PURTs that are pivotal in the presence of volatility shifts. The applicability of these tests, however, has been restricted to cases where the data contains only an intercept, and not a linear trend. This paper proposes a new heteroskedasticity-robust PURT that works well for trending data. Under the null hypothesis, the test statistic has a limiting Gaussian distribution. Simulation results reveal that the test tends to be conservative but shows remarkable power in finite samples.

JEL Classification: C23, C12, Q40.

Keywords: Panel unit root tests, nonstationary volatility, cross-sectional dependence, near epoch dependence, energy use per capita.

*Address correspondence to: Helmut Herwartz, Chair of Econometrics, Georg-August-University Göttingen, Humboldtallee 3, D-37073 Göttingen, Germany, Telephone: +49 551 39 27350, fax: +49 551 397279, e-mail: *hherwartz@uni-goettingen.de*.

1 Introduction

Most macroeconomic time series experience occasional breaks or trending behavior in their unconditional variances. For instance, Sensier and van Dijk (2004) document that, during the period 1959–1999, about 80% of 214 U.S. macroeconomic time series they studied displayed breaks in their unconditional volatility. It is also well-known that volatilities of several macroeconomic series were significantly lower during the period 1984–2007 than in earlier decades, a phenomenon called the 'Great Moderation' (see, for instance, Stock and Watson, 2003). However, business cycle volatilities rose again during the recent global economic and financial crises. Whether the 'great recession' marks the end of the Great Moderation or was just a short interruption within an ongoing Great Moderation is still debated.[1] In any case, the debate—or even the very notion of Great Moderation for that matter—underscores the fact that time-varying volatility of macroeconomic series is more of a rule rather than an exception.

The potential consequences of variance shifts on univariate unit root tests have been investigated by, among others, Hamori and Tokihisa (1997), Kim et al. (2002), Cavaliere (2004), and Cavaliere and Taylor (2007). These studies find that the (augmented) Dickey-Fuller (Dickey and Fuller, 1979) tests have seriously distorted empirical sizes—and, hence, provide deceptive inference—if volatility varies over time. The same problem carries over to panel unit root tests (PURTs), as shown in Demetrescu and Hanck (2012a,b) and Herwartz et al. (2016). In particular, widely applied PURTs such as those suggested in Levin et al. (2002) and Breitung and Das (2005) are no longer pivotal if the homoskedasticity assumption is violated (Herwartz et al., 2016).

To deal with the above problem, a few heteroskedasticity-robust PURTs have been proposed recently. In consecutive papers, Demetrescu and Hanck (2012a,b) suggest PURTs that are built on the so-called Cauchy estimator. As the sign function of Cauchy instrumenting reduces the lagged level series to -1 and 1—irrespective of the underlying time varying volatility—these tests are argued to be robust to heteroskedasticity. Herwartz et al. (2016) show that the non-Cauchy version of the test in Demetrescu and Hanck (2012a), which was initially proposed in Herwartz and Siedenburg (2008), is robust to volatility shifts. Another heteroskedasticity-robust PURT has been suggested by Westerlund (2014). This test utilizes the information contained in group-specific

[1] See, for instance, Gadea-Rivas et al. (2014) for a concise survey on this debate.

variances.

While these heteroskedasticity-robust tests also remain pivotal under a fairly general form of cross-sectional and serial correlation, they, however, do not work for detrended data. Namely, available detrending schemes introduce nuisance parameters that affect the limiting distribution of the tests under variance breaks (Herwartz et al., 2016; Westerlund, 2014). This problem significantly limits the applicability of the tests as many macroeconomic time series exhibit trending behavior. In fact, Westerlund (2015, p. 454) states that

> "*...for many economic time series, a linear trend, rather than a constant, might be considered appropriate as the default specification, ... This is certainly true for series such as GDP, industrial production, money supply and consumer or commodity prices, where trending behavior is evident.*"

In this paper, we propose a new heteroskedasticity-robust PURT. Most importantly, the test can be applied to detrended data and its limiting distribution (under the null hypothesis) is free of nuisance parameters. The construction of the test is simple. We begin by detrending the data according to the method suggested in Demetrescu and Hanck (2014), and trace the effects of the detrending scheme on the (detrended) integrated level data. The drift term is estimated as the unconditional mean of first-differenced series. Taking account of volatility breaks, level detrending and drift estimation, we construct a test statistic that exhibits an asymptotic Gaussian distribution under the panel unit root null hypothesis. To prove asymptotic normality we rely on central limit theory for near-epoch dependent processes as discussed, e.g., in Davidson (1994). Simulation results show that the proposed test works well in finite samples, and has satisfactory power which is comparable with the power of the tests in Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2012a) under homoskedasticity.

As an empirical illustration, we examine whether energy use per capita is trend or difference stationary. Using data from 23 OECD economies over the period 1960–2014, we find that energy use per capita is generally integrated of order one. However, results from unit root testing for rolling fixed-length time spans show that the series could be characterized as trend stationary for forty-years windows that start between 1963 and 1968.

Section 2 sketches the panel unit root testing problem and describes two of the existing heteroskedasticity-robust PURTs. Section 3 discusses ways of handling serial correlation and deterministic terms. Section 4 introduces the proposed test statistic and states its asymptotic distribution. The finite sample performance of the new test is evaluated by means of a Monte Carlo study documented in Section 5. As an empirical illustration, the stationarity of energy use per capita is examined in Section 6. Section 7 concludes. Proofs of the asymptotic results are provided in the Appendix.

2 Homogeneous panel unit root testing

In this section we first describe the panel unit root testing problem and formalize cross-sectional dependence and heteroskedasticity. Next, we present the White-type heteroskedasticity-robust PURTs suggested in Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2012a).

2.1 The first order panel autoregression

A first order panel autoregression under nonstationary volatility and a linear trend can be specified as

$$\boldsymbol{y}_t = \boldsymbol{\mu} + (1-\rho)\boldsymbol{\delta} t + \rho \boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \quad t = 1, ..., T, \tag{1}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $\boldsymbol{y}_{t-1} = (y_{1,t-1}, \ldots, y_{N,t-1})'$, $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{Nt})'$ are $N \times 1$ vectors, and $\boldsymbol{e}_t$ is heterogeneously distributed with mean zero and covariance Ω_t. Furthermore, the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)'$ stacks panel-specific trend parameters, and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$ contains panel-specific intercepts. The specification in (1) formalizes an empirically relevant panel unit root testing problem of distinguishing between a random walk with drift on the one hand and a trend stationary process on the other hand (Pesaran, 2007). PURTs are used to test the hypothesis $H_0 : \rho = 1$ against $H_1 : \rho < 1$ in (1).

To formalize cross-sectional dependence and heteroskedasticity, we adopt the following assumptions about the vector of error terms $\boldsymbol{e}_t$ as in Herwartz et al. (2016) with strengthened moment conditions:

Assumptions $\mathcal{A}$.

(i) $\boldsymbol{e}_t$ *is serially uncorrelated with mean* 0 *and covariance* Ω_t.

(ii) Ω_t *is a positive definite matrix with eigenvalues* $\lambda_t^{(1)} \leq \lambda_t^{(2)} \leq \ldots \leq \lambda_t^{(N)}$ *and* $\lambda_t^{(N)} < \bar{c} < \infty$, $\lambda_t^{(1)} > \underline{c} > 0$ *for all* t.
(iii) $E[u_{it}^p u_{jt}^p u_{kt}^p u_{lt}^p] < \infty$ *for all* i, j, k, l *and* $p = 1, 2$, *where* $u_{\bullet t}$, $\bullet \in \{i, j, k, l\}$ *denote typical elements of* $\boldsymbol{u}_t = \Omega_t^{-1/2}\boldsymbol{e}_t$. *Here we set* $\Omega_t^{1/2} = \Gamma_t \Lambda_t^{1/2} \Gamma_t'$, *where* Λ_t *is a diagonal matrix of eigenvalues of* Ω_t *and the columns of* Γ_t *are the corresponding eigenvectors.*

$\mathcal{A}(i)$ restricts the error terms to be serially uncorrelated. Ways of handling higher order serial correlation will be described later. The assumption that the fourth order moments of e_{it} (or u_{it} by implication of $\mathcal{A}(ii)$) should be finite ($\mathcal{A}(iii)$ for $p = 1$) is standard in the (panel) unit root literature. The stronger assumption of finiteness of moments up to order eight ($p = 2$) will allow to apply asymptotic theory for near-epoch dependent processes. While $\mathcal{A}(ii)$ captures so-called weak forms of cross-sectional dependence such as spatial panel models (for more details on spatial panel models see, e.g., Anselin, 2013) and seemingly unrelated regressions, it rules out strong forms of cross-sectional dependence that might be traced back to the presence of common factors. Since $\mathrm{tr}(\Omega_t) = \sum_{i=1}^N \lambda_t^{(i)}$, $\mathcal{A}(ii)$ covers both discrete covariance breaks as well as smoothly trending variances.

2.2 Heteroskedasticity-robust tests

2.2.1 The White-type test

Herwartz and Siedenburg (2008) propose a PURT based on a White-type covariance estimator. Setting $\boldsymbol{\mu} = \boldsymbol{\delta} = 0$ in (1), the test statistic is given by

$$t_{HS} = \frac{\sum_{t=1}^T \boldsymbol{y}_{t-1}' \Delta \boldsymbol{y}_t}{\sqrt{\sum_{t=1}^T \boldsymbol{y}_{t-1}' \hat{\boldsymbol{e}}_t \hat{\boldsymbol{e}}_t' \boldsymbol{y}_{t-1}}} \xrightarrow{d} N(0,1), \quad \hat{\boldsymbol{e}}_t = \Delta \boldsymbol{y}_t = \boldsymbol{e}_t. \tag{2}$$

Originally, t_{HS} was proposed as an alternative to the test in Breitung and Das (2005) for finite samples where the cross-sectional dimension is relatively large in comparison with the time series dimension. Recently, Herwartz et al. (2016) show that time-varying volatility does not affect the pivotalness of t_{HS}.

2.2.2 The White-type Cauchy test

Demetrescu and Hanck (2012a) suggest a heteroskedasticity-robust PURT based on the

'Cauchy' estimator which instruments the lagged level by its sign. This statistic reads as

$$t_{DH} = \frac{\sum_{t=1}^{T} sgn(\boldsymbol{y}_{t-1})'\Delta\boldsymbol{y}_t}{\sqrt{\sum_{t=1}^{T} sgn(\boldsymbol{y}_{t-1})'\hat{\boldsymbol{e}}_t\hat{\boldsymbol{e}}_t' sgn(\boldsymbol{y}_{t-1})}} \xrightarrow{d} N(0,1), \tag{3}$$

where $sgn(\cdot)$ denotes the sign function.

Two further heteroskedasticity-robust PURTs that we are aware of are those proposed in Demetrescu and Hanck (2012b) and Westerlund (2014). A common limitation of all these PURTs, however, is that in the presence of linear trends (i.e., $\boldsymbol{\delta} \neq 0$ in (1)), applying standard detrending schemes does not retain the pivotalness of the tests if the data exhibit variance breaks.

3 Deterministic terms and serial correlation

In this section, we discuss how serial correlation and deterministic terms are handled in panel unit root testing under variance breaks.

3.1 Short-run dynamics

To eliminate short-run serial correlation from the data, prewhitening is an important procedure which leaves the limiting distribution of the tests unaffected (Breitung and Das, 2005). This procedure requires estimating individual-specific autoregressions of the first differences under H_0, i.e.,

$$\Delta y_{it} = \sum_{j=1}^{p_i} b_{ij}\Delta y_{i,t-j} + e_{it}. \tag{4}$$

Prewhitened data is then obtained as

$$\widehat{y}_{it} = y_{it} - \widehat{b}_{i1}y_{i,t-1} - \ldots - \widehat{b}_{ip_i}y_{i,t-p_i}, \tag{5}$$

and

$$\widehat{\Delta y}_{it} = \Delta y_{it} - \widehat{b}_{i1}\Delta y_{i,t-1} - \ldots - \widehat{b}_{ip_i}\Delta y_{i,t-p_i}. \tag{6}$$

Any consistent lag-length selection criterion can be applied to decide upon the lag orders p_i. In cases where both short-run dynamics and deterministic patterns are present in the data, prewhitening should precede detrending. The prewhitening regression should include an intercept term if the model features linear time trends under the alternative hypothesis.

3.2 Deterministic terms

Removing the trend in (1) by means of popular schemes such as OLS, GLS or recursive detrending renders the PURTs to depend on the drift terms in $\boldsymbol{\mu}$, and, hence, requires bias-correction terms. Moreover, the bias-correction becomes highly complicated with the presence of variance breaks. The detrending procedures in Breitung and Das (2005) and Demetrescu and Hanck (2014) do not require bias adjustment terms as long as the homoskedasticity assumption is maintained. With time-varying volatility, however, both detrending methods affect the pivotalness of PURTs, including t_{HS} and t_{DH}. As the test we are proposing utilizes the detrending scheme in Demetrescu and Hanck (2014), we briefly outline it here. This method involves recursively detrending the lagged level variable to obtain

$$\tilde{\boldsymbol{y}}_{t-1} = \boldsymbol{y}_{t-1} + \frac{2}{t-1}\sum_{j=1}^{t-1}\boldsymbol{y}_{\boldsymbol{j}} - \frac{6}{t(t-1)}\sum_{j=1}^{t-1} j\boldsymbol{y}_j. \tag{7}$$

Since $\Delta\boldsymbol{y}_t$ has non-zero mean, it has to be demeaned. One choice is to center $\Delta\boldsymbol{y}_t$ in the usual way as

$$\Delta\boldsymbol{y}_t^* = \Delta\boldsymbol{y}_t - \frac{1}{T}\sum_{t=2}^{T}\Delta\boldsymbol{y}_t, \tag{8}$$

where T in the denominator replaces $T-1$ for notational convenience. Demetrescu and Hanck (2014) show that, under homoskedasticity, $\Delta\boldsymbol{y}_t$ could also be centered by means of forward demeaning instead of (8). In the presence of heteroskedasticity, both full sample centering and forward demeaning affect the pivotalness of even the heteroskedasticity-robust tests t_{HS} and t_{DH} and, hence, invoke marked size distortions (see Demetrescu and Hanck (2014) for rigorous arguments on this issue). As forward demeaning additionally leads to relatively large power losses in comparison with full sample centering, the test proposed in this work relies on full sample demeaning.

4 Panel unit root test for trending series with time-varying volatility

The heteroskedasticity-robust test we propose builds upon the White-type test given in (2) and the detrending scheme described by (7) and (8). Instead of providing the test statistic in a compact form, we first consider a modified version of the numerator of t_{HS}

in (2). With (7) and (8) the summands of the numerator of this modification can be rewritten as

$$\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^* = \tilde{\boldsymbol{y}}_{t-1}'\hat{\boldsymbol{e}}_t = \sum_{i=1}^{t-1}\left(a_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \frac{1}{T}a_{i,t-1}\sum_{k=2}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\right), \tag{9}$$

where

$$\Delta\boldsymbol{y}_t^* = \hat{\boldsymbol{e}}_t = \boldsymbol{e}_t - \frac{1}{T}\sum_{t=2}^{T}\boldsymbol{e}_t, \tag{10}$$

and finite weighting coefficients $a_{i,t-1}$ read as

$$a_{i,t-1} = 1 + \frac{2}{t-1}(t-i) - 3\left(1 - \frac{(i-1)i}{(t-1)t}\right). \tag{11}$$

Derived from data detrended according to (7) and (8), the expression in (9) has a non-zero expectation in the absence of homoskedasticity under the null hypothesis of a panel unit root. The theoretical version of the new test statistic, henceforth denoted by τ, can be seen as a modification of t_{HS} with adjustments for the non-zero mean in the numerator, and corresponding changes for the variance (in the denominator). Specifically, the test statistic with theoretical moments is given by

$$\tau = \frac{\sum_{t=2}^{T}\frac{1}{\sqrt{NT}}\left(\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^* - \nu_t\right)}{\sqrt{\frac{1}{NT}\left(E\left[\sum_{t=2}^{T}\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*\right]^2 - \left(\sum_{t=2}^{T}\nu_t\right)^2\right)}}, \tag{12}$$

where $\nu_t = E[\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*]$.

Unlike in Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2014), where the White-type covariance estimator is applied, the more complicated form of $\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*$ invokes the following representation of the variance of the numerator in (12):

$$s_{NT}^2 := \frac{1}{NT}\left(E\left[\sum_{t=2}^{T}\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*\right]^2 - \left(\sum_{t=2}^{T}\nu_t\right)^2\right) = \zeta_1 - \zeta_2 + \zeta_3 + \zeta_4 + \zeta_5 - \frac{1}{NT}\left(\sum_{t=2}^{T}\nu_t\right)^2. \tag{13}$$

The expansion of the expectation in (13) yields components $\zeta_1, \ldots, \zeta_5$ which can be shown to correspond to the following quantities

$$\zeta_1 = \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}\bar{a}_{i,s-1}\bar{a}_{j,t-1}\left(\mathrm{tr}(\Omega_i\Omega_j)+\mathrm{tr}(\Omega_i)\mathrm{tr}(\Omega_j)\right)$$

$$\zeta_2 = \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=i+1}^{T}\tilde{a}_{i,s-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i\Omega_s)$$

$$\zeta_3 = \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\tilde{a}^2_{i,t-1}\mathrm{tr}(\Omega_i\Omega_t) \tag{14}$$

$$\zeta_4 = \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\bar{a}^2_{i,t-1}\left(E\left[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2\right]+\sum_{j=1,j\neq i,t}^{T}\mathrm{tr}(\Omega_i\Omega_j)\right)$$

$$\zeta_5 = \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\left(E\left[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2\right]+\sum_{j=1,j\neq i,t,s}^{T}\mathrm{tr}(\Omega_i\Omega_j)\right),$$

where $\tilde{a}_{i,t-1} = \left(1-\frac{1}{T}\right)a_{i,t-1}$ and $\bar{a}_{i,t-1} = \frac{1}{T}a_{i,t-1}$ with coefficients $a_{i,t-1}$ defined in (11). Similarly,

$$\nu_t = E[\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*] = -\sum_{i=1}^{t-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i). \tag{15}$$

The new test is then the empirical version of τ in (12), i.e.,

$$\hat{\tau} = \frac{\sum_{t=2}^{T}\frac{1}{\sqrt{NT}}\left(\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^* - \hat{\nu}_t\right)}{\hat{s}_{NT}}, \tag{16}$$

where estimators of ν_t and the variance components are based on the estimation of the traces of the covariance matrices Ω_i. More precisely, we replace $\mathrm{tr}(\Omega_i)$ by $\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i$, $\mathrm{tr}(\Omega_i\Omega_j)$ by $\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j$ and $E\left[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2\right]$ by $(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2$ where $\hat{\boldsymbol{e}}_i$ is a vector of centered residuals (first differences) as defined in (10). Detailed representations of $\hat{\nu}_t$ and $\hat{s}^2_{NT}$ are given in the Appendix. The following proposition states the asymptotic normality of the statistic in (16).

Proposition 1. *Under assumptions $\mathcal{A}$ the test statistic in (16) is asymptotic normally distributed, i.e., for $N,T\to\infty$ with $N/T^2\to 0$*

$$\hat{\tau}\xrightarrow{d}\mathcal{N}(0,1). \tag{17}$$

The proof of Proposition 1 is based on a central limit theorem for near-epoch dependent sequences and is given in the Appendix. As it will turn out, the additional requirement of $N/T^2\to 0$ is necessary for $\hat{\tau}$ to fulfill the conditions of the central limit theorem, as well as for applying the test to prewhitened data (Herwartz et al., 2016).

5 Monte Carlo study

5.1 The simulation design

To evaluate the finite sample properties of the proposed test $\hat{\tau}$, we consider the following DGPs taken from Pesaran (2007):

$$\text{DGP1:} \quad \boldsymbol{y}_t = \boldsymbol{\mu} + (\boldsymbol{j} - \boldsymbol{\rho}) \odot \boldsymbol{\beta} t + \boldsymbol{\rho} \odot \boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \quad t = -50, \ldots, T, \tag{18}$$

$$\text{DGP2:} \quad \boldsymbol{y}_t = \boldsymbol{\mu} + (\boldsymbol{j} - \boldsymbol{\rho}) \odot \boldsymbol{\beta} t + \boldsymbol{\rho} \odot \boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t = \boldsymbol{b} \odot \boldsymbol{\epsilon}_{t-1} + \boldsymbol{e}_t, \tag{19}$$

where bold entries indicate vectors of dimension $N \times 1$, $\boldsymbol{j}$ is a vector of ones and $\odot$ denotes the Hadamard product. The DGP1 formalizes AR(1) models with serially uncorrelated innovations while DGP2 introduces AR(1) disturbances. Both DGPs formalize a panel random walk with drift under the null hypothesis, and a panel of trend stationary processes with individual effects under the alternative. Empirical size is obtained by setting $\boldsymbol{\rho} = \boldsymbol{j}$ and power is simulated as $\boldsymbol{\rho} = 0.9\boldsymbol{j}$.[2] Individual effects, trend parameters as well as serial correlation of innovations are modeled as in Pesaran (2007): $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$, $\mu_i \sim \text{iid}\, U(0, 0.02)$ and $\boldsymbol{b} = (b_1, \ldots, b_N)'$, $b_i \sim \text{iid}\, U(0.2, 0.4)$.

To separate the issue of cross-sectional correlation from variance breaks, we employ the decomposition

$$\Omega_t = \Phi_t^{1/2} \Psi \Phi_t^{1/2},$$

where $\Phi_t = diag(\sigma_{1t}^2, \ldots, \sigma_{Nt}^2)$ and Ψ is a (time invariant) correlation matrix characterizing Ω_t. Cross-sectional independence is obtained by setting Ψ to an identity matrix of order N. We generate a weak form of cross-sectional correlation by means of the spatial autoregressive (SAR) error structure used in Herwartz and Siedenburg (2008). Specifically, we take Ψ_{SAR} that is implied by the SAR model

$$\boldsymbol{e}_t = (I_N - \Theta W)^{-1} \boldsymbol{\xi}_t, \quad \text{with} \quad \Theta = 0.8 \quad \text{and} \quad \boldsymbol{\xi}_t \sim \text{iid}\, N(\mathbf{0}, I_N),$$

where W is the so-called spatial weights matrix. In this particular case, W is a row normalized symmetric contiguity matrix of the '*g ahead and g behind*' structure, with

[2] Results for DGPs with heterogeneous autogregressive coefficients under the alternative hypothesis, i.e., $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_N), \rho_i \sim \text{iid}\, U(0.85, 0.95)$, are qualitatively identical and available upon request. Moreover, recent papers, e.g., Homm and Breitung (2012), also consider power against explosive alternatives ($\rho > 1$). Using a right-sided testing, the proposed test $\hat{\tau}$ is powerful against the alternative that $\boldsymbol{\rho} = 1.03\boldsymbol{j}$, even for $T = 25$. The corresponding simulation results are available upon request.

$g = 1$ (see, e.g., Kelejian and Prucha, 1999). The resulting covariance matrix of $\boldsymbol{e}_t$ is given by $\Omega_{SAR} = ((I_N - \Theta W)'(I_N - \Theta W))^{-1}$, and Ψ_{SAR} is the correlation matrix implied by Ω_{SAR}.

Cross-section specific volatility shifts are generated as

$$\sigma_{it}^2 = \begin{cases} \sigma_{i1}^2, & \text{if } t < \lfloor \gamma_i T \rfloor,\ (0 < \gamma_i < 1) \\ \sigma_{i2}^2, & \text{otherwise,} \end{cases}$$

where γ_i refers to the time a variance break occurs and $\lfloor \gamma_i T \rfloor$ denotes the integer part of $\gamma_i T$. In the homoskedastic case, $\sigma_{i1} = \sigma_{i2} = 1$. We introduce heteroskedasticity by changing the post-break variance to $\sigma_{i2} = 1/3$, for a negative variance break, and to $\sigma_{i2} = 3$, for a positive one. Regarding the timing of the variance breaks, we consider scenarios of homogeneously early ($\gamma_i = 0.2$) or late ($\gamma_i = 0.8$) variance breaks for all panel units.[3] Data are generated for all combinations of $N \in [50, 100, 250]$ and $T \in [25, 50, 100, 250]$. To mitigate the potential impacts of initial values on our analysis, we generate and discard 50 presample observations.

5.2 Simulation results

In the following we discuss simulation results on the finite sample performance of the proposed test statistic $\hat{\tau}$ in comparison with two of the existing heteroskedasticity-robust tests (t_{HS} and t_{DH}). For the new test, we also document results for its theoretical counterpart τ determined from the true covariance matrices Ω_t (see (12)). Presenting simulation results for both $\hat{\tau}$ and τ is meant to highlight finite sample performance of $\hat{\tau}$ that can be traced back to the use of moment estimators.

5.2.1 Cross-sectionally independent panels

Simulation results for data generated according to DGP1 for cross-sectionally independent panels are documented in Table 1. Results in the upper panel of this table show that, under homoskedasticity, the recursive detrending scheme in Demetrescu and Hanck (2014) leaves the pivotalness of heteroskedsticity-robust tests unaffected. With respect to rejection frequencies under the alternative hypothesis, it can be seen that using estimated

[3] Main findings of the simulation exercise remain qualitatively unaffected by consideration of randomly distinct break moments $\gamma_i \sim \text{iid}\, U(0.1, 0.9)$. These results are available upon request.

Table 1: Empirical rejection frequencies, cross-sectionally independent panels

		5%								10%							
		size				power				size				power			
N	T	τ	$\hat{\tau}$	HS	DH	τ	$\hat{\tau}$	HS	DH	τ	$\hat{\tau}$	HS	DH	τ	$\hat{\tau}$	HS	DH
Constant variance (HOM)																	
50	25	5.6	4.6	4.8	4.5	40.5	32.0	31.1	19.4	11.1	10.0	9.8	9.9	55.2	49.9	48.3	32.3
50	50	4.6	5.5	4.8	4.3	98.9	98.6	97.4	77.2	10.4	11.1	9.7	9.0	99.6	99.7	99.4	87.7
50	100	4.6	4.3	3.5	4.0	100.0	100.0	100.0	100.0	9.7	9.7	8.4	8.6	100.0	100.0	100.0	100.0
50	250	3.8	4.2	3.6	4.0	100.0	100.0	100.0	100.0	9.0	9.9	8.6	8.6	100.0	100.0	100.0	100.0
100	25	5.4	3.4	4.3	4.7	63.8	50.8	49.6	28.9	11.4	8.2	9.5	9.5	76.1	68.4	67.8	45.6
100	50	5.2	5.5	4.7	4.7	100.0	100.0	100.0	96.2	11.0	10.8	9.7	9.9	100.0	100.0	100.0	98.8
100	100	4.5	4.9	3.7	4.4	100.0	100.0	100.0	100.0	9.4	10.2	8.8	8.7	100.0	100.0	100.0	100.0
100	250	4.9	4.6	3.7	4.4	100.0	100.0	100.0	100.0	10.3	10.0	8.4	9.0	100.0	100.0	100.0	100.0
250	25	5.4	2.0	4.5	4.8	92.7	81.0	83.2	54.2	10.8	5.9	9.5	9.7	96.1	92.1	93.4	70.1
250	50	4.9	4.1	3.8	4.2	100.0	100.0	100.0	100.0	10.4	9.5	9.0	9.2	100.0	100.0	100.0	100.0
250	100	4.7	4.7	3.6	4.1	100.0	100.0	100.0	100.0	9.6	9.5	8.0	8.7	100.0	100.0	100.0	100.0
250	250	4.6	5.0	3.9	4.0	100.0	100.0	100.0	100.0	10.1	10.3	8.4	9.0	100.0	100.0	100.0	100.0
Early negative variance shift (NEG)																	
50	25	4.9	1.4	0.0	0.0	9.6	4.3	0.0	0.0	10.2	4.4	0.0	0.0	17.3	12.1	0.0	0.0
50	50	5.1	3.0	0.0	0.0	48.8	48.2	0.0	0.1	10.7	7.9	0.0	0.0	63.8	66.2	0.1	0.6
50	100	4.7	4.0	0.0	0.0	99.9	99.9	37.8	36.1	10.0	8.9	0.0	0.0	100.0	100.0	54.4	51.4
50	250	4.6	4.5	0.0	0.0	100.0	100.0	100.0	100.0	9.9	10.2	0.0	0.0	100.0	100.0	100.0	100.0
100	25	5.4	0.4	0.0	0.0	12.5	3.3	0.0	0.0	10.5	2.6	0.0	0.0	22.3	12.3	0.0	0.0
100	50	5.2	2.3	0.0	0.0	75.5	72.4	0.0	0.0	10.6	6.0	0.0	0.0	85.8	87.1	0.0	0.1
100	100	4.9	3.9	0.0	0.0	100.0	100.0	57.0	57.5	10.6	8.9	0.0	0.0	100.0	100.0	73.7	73.0
100	250	5.2	4.5	0.0	0.0	100.0	100.0	100.0	100.0	10.5	10.3	0.0	0.0	100.0	100.0	100.0	100.0
250	25	5.0	0.0	0.0	0.0	18.6	0.8	0.0	0.0	10.1	0.4	0.0	0.0	29.2	6.9	0.0	0.0
250	50	5.0	1.0	0.0	0.0	97.8	96.1	0.0	0.0	10.7	4.4	0.0	0.0	99.2	98.9	0.0	0.0
250	100	5.0	3.4	0.0	0.0	100.0	100.0	84.4	87.1	10.1	7.9	0.0	0.0	100.0	100.0	93.4	94.0
250	250	5.0	4.8	0.0	0.0	100.0	100.0	100.0	100.0	10.0	9.3	0.0	0.0	100.0	100.0	100.0	100.0
Late positive variance shift (POS)																	
50	25	4.6	1.6	17.7	11.0	57.7	35.6	77.8	50.8	9.9	7.0	30.7	21.3	73.6	64.3	92.3	68.4
50	50	3.3	3.2	26.0	13.1	99.0	98.8	99.8	95.4	8.2	8.2	39.6	23.4	99.9	99.9	100.0	98.7
50	100	3.1	3.3	21.0	12.2	100.0	100.0	100.0	100.0	7.6	8.4	35.0	21.9	100.0	100.0	100.0	100.0
50	250	2.6	2.7	18.4	12.3	100.0	100.0	100.0	100.0	7.3	8.0	32.1	22.4	100.0	100.0	100.0	100.0
100	25	4.7	0.5	29.1	15.6	85.5	55.6	92.7	72.9	10.0	3.7	44.7	27.0	92.6	83.4	99.0	86.9
100	50	4.6	2.9	34.4	17.4	100.0	100.0	100.0	99.9	9.4	7.8	50.8	29.6	100.0	100.0	100.0	100.0
100	100	3.5	2.9	30.0	16.3	100.0	100.0	100.0	100.0	8.0	7.2	46.7	28.1	100.0	100.0	100.0	100.0
100	250	3.4	3.1	28.8	15.4	100.0	100.0	100.0	100.0	8.3	8.1	45.3	26.8	100.0	100.0	100.0	100.0
250	25	4.4	0.1	57.3	29.4	99.7	85.0	99.7	94.6	10.4	1.1	72.3	44.1	99.9	98.3	100.0	98.8
250	50	3.4	1.9	56.0	27.7	100.0	100.0	100.0	100.0	9.0	6.3	73.7	42.0	100.0	100.0	100.0	100.0
250	100	4.0	3.7	53.7	26.8	100.0	100.0	100.0	100.0	9.0	8.7	71.8	41.7	100.0	100.0	100.0	100.0
250	250	3.1	3.3	57.3	29.1	100.0	100.0	100.0	100.0	7.4	8.1	72.3	44.0	100.0	100.0	100.0	100.0

Notes: τ, $\hat{\tau}$, HS and DH refer to the PURT statistics given in (12), (16), (2) and (3) respectively. Power is not size adjusted. All results are based on 5000 replications. Data is generated according to DGP1 in (18) and all tests are computed on detrended data.

covariance matrices induces considerable power loss under a small time dimension $T = 25$. However, this power loss vanishes with increasing T. Furthermore, the new test $\hat{\tau}$ is generally as powerful as t_{HS} and more powerful than t_{DH}. Hence, it is worthwhile noting that our adjustment for obtaining robustness to time-varying volatility does not come at a cost of reduced power. In view of the fact that the reported empirical powers are not size adjusted, the power estimates for $\hat{\tau}$ are rather remarkable.

When early negative variance breaks are introduced, t_{HS} and t_{DH} display zero rejection frequencies under the null hypothesis. On the contrary, $\hat{\tau}$ holds remarkable size control, except for small T ($T = 25$) where it is substantially undersized. These size distortions, however, improve markedly as the time dimension increases to $T = 50$. The new test also has significant power under early variance breaks although it is less than the power under homoskedasticity. In comparison with $\hat{\tau}$, the White-type tests t_{HS} and t_{DH} have substantially weaker power, with both tests showing almost zero probability of rejecting the alternative hypothesis until the time dimension increases to $T = 100$.

Size distortions of t_{HS} and t_{DH} are also observed when a late positive volatility shift is considered, but this time with huge oversizings. On the contrary, $\hat{\tau}$ displays a fairly good size precision. Consistent with results in Herwartz et al. (2016) for non-trending data, power seems to be unaffected by late positive variance breaks but reduced by early negative volatility shifts. In general, simulation results documented in Table 1 demonstrate not only the risk of using t_{HS} and t_{DH} for trending time series, but also the satisfactory finite sample performance of $\hat{\tau}$ for trending heteroskedastic data.

5.2.2 DGPs with cross-sectionally correlated panels

The left-hand side block of Table 2 documents simulation results for $\hat{\tau}$ applied on data generated according to DGP1 for weakly correlated panels. Results available upon request show that size distortions of t_{HS} and t_{DH} observed for cross-sectionally independent panels (Table 1) carry over to panels with weak forms of cross-sectional correlation. Hence, we focus on the implications of cross-sectional correlation for the new test $\hat{\tau}$. Confirming the asymptotic considerations, a relatively larger cross-sectional dimension N is required for the empirical size of $\hat{\tau}$ to come closer to the nominal significance levels. Moreover, the statistic $\hat{\tau}$ is less powerful under the SAR(1) model than under independent panels—a result consistent with those documented in Herwartz et al. (2016) for non-trending series.

Table 2: Empirical rejection frequencies of $\hat{\tau}$, diverse scenarios

		DGP1, SAR(1) model				DGP2, Independence				DGP2, SAR(1) model			
		5%		10%		5%		10%		5%		10%	
N	T	size	power	size	power	size	power	size	power	size	power	size	power
Constant variance (HOM)													
50	25	4.3	16.2	11.2	28.9	0.0	0.0	0.0	0.1	0.0	0.1	0.5	0.7
50	50	3.8	60.4	10.3	78.0	0.1	29.7	0.8	50.3	0.2	8.1	1.2	20.6
50	100	3.6	100.0	9.3	100.0	1.3	100.0	3.7	100.0	1.2	95.4	4.1	98.7
50	250	3.2	100.0	8.6	100.0	2.9	100.0	6.7	100.0	1.9	100.0	5.5	100.0
100	25	4.8	24.7	10.6	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2
100	50	4.4	89.7	10.5	95.6	0.1	58.7	0.3	77.6	0.1	17.7	0.9	36.8
100	100	4.2	100.0	10.0	100.0	1.0	100.0	3.0	100.0	1.0	100.0	4.0	100.0
100	250	3.8	100.0	9.2	100.0	2.5	100.0	6.0	100.0	2.2	100.0	6.4	100.0
250	25	3.7	42.6	8.9	61.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
250	50	4.5	99.9	9.9	100.0	0.0	94.0	0.1	98.3	0.1	48.1	0.6	68.4
250	100	4.1	100.0	9.1	100.0	0.4	100.0	1.3	100.0	0.8	100.0	2.9	100.0
250	250	4.5	100.0	10.2	100.0	2.6	100.0	5.9	100.0	2.5	100.0	6.4	100.0
Early negative variance shift (NEG)													
50	25	2.9	5.1	7.7	13.7	0.0	0.0	0.0	0.0	0.1	0.0	0.5	0.3
50	50	3.3	20.8	8.9	38.1	0.1	0.4	0.4	2.0	0.3	0.9	1.4	3.0
50	100	3.7	79.2	9.6	90.4	0.7	87.7	3.0	95.9	0.8	30.3	3.1	55.3
50	250	3.5	100	8.5	100.0	3.0	100.0	8.0	100.0	1.7	100.0	6.3	100.0
100	25	2.0	4.8	5.7	13.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
100	50	3.3	35.3	8.5	53.4	0.0	0.2	0.1	0.7	0.2	0.5	0.6	2.1
100	100	4.6	97.7	10.1	99.3	0.6	99.6	2.0	100.0	0.9	66.3	2.9	84.8
100	250	4.4	100.0	9.5	100.0	3.7	100.0	8.5	100.0	2.3	100.0	6.0	100.0
250	25	0.6	3.6	3.1	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
250	50	3.0	61.9	6.9	78.1	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.9
250	100	4.2	100.0	9.2	100.0	0.2	100.0	1.2	100.0	0.5	97.9	2.0	99.5
250	250	4.6	100.0	9.8	100.0	3.9	100.0	9.5	100.0	2.5	100.0	7.6	100.0
Late positive variance shift (POS)													
50	25	1.8	14.8	7.9	34.7	0.0	0.3	0.4	4.4	0.2	0.7	1.5	4.5
50	50	1.8	56.7	7.1	80.1	0.8	63.1	4.7	83.6	0.6	16.4	3.8	37.9
50	100	1.4	99.0	6.0	99.8	2.2	100.0	6.6	100.0	0.9	85.8	4.8	96.8
50	250	1.1	100.0	5.6	100.0	2.8	100.0	7.1	100.0	0.9	100.0	4.9	100.0
100	25	2.0	24.8	7.2	50.0	0.0	0.2	0.2	3.2	0.1	0.5	0.8	4.4
100	50	2.6	89.8	7.6	97.0	1.2	91.9	4.4	98.4	1.0	39.9	4.1	66.0
100	100	2.6	100.0	7.7	100.0	3.0	100.0	8.2	100.0	1.6	99.7	5.8	100.0
100	250	2.2	100.0	7.0	100.0	3.1	100.0	8.3	100.0	2.0	100.0	6.0	100.0
250	25	0.8	48.1	4.3	77.4	0.0	0.0	0.0	1.1	0.0	0.2	0.1	2.9
250	50	3.1	99.9	8.0	100.0	0.4	100.0	3.7	100.0	0.8	85.0	4.2	95.7
250	100	3.0	100.0	8.2	100.0	3.1	100.0	9.0	100.0	2.1	100.0	6.3	100.0
250	250	2.8	100.0	7.7	100.0	4.1	100.0	10.1	100.0	2.5	100.0	7.8	100.0

Notes: Data is generated according to DGP1 in (18) for results in the left-hand side block, while DGP2 in (19) is used to generate data for results documented in the middle and right-hand side blocks of the table. Testing is performed on detrended data. For DGP2, detrending is preceded by prewhitening. Power is not size adjusted and all results are based on 5000 replications.

5.2.3 DGPs with serially correlated innovations

To evaluate how the proposed test $\hat{\tau}$ performs for data with serially correlated disturbances, we generate data according to DGP2 in (19) and subject it to prewhitening before detrending. The corresponding simulation results are documented in the middle- and right-hand side blocks of Table 2. The results show that serial correlation and the ensuing prewhitening procedure entail marked size distortions for small time dimensions. This result could be explained by noting that estimation errors arising from the prewhitening procedure introduce finite sample correlations between the lagged level and first differenced series, thereby inducing a non-zero mean to the numerator of the test statistic in (16). However, size distortions vanish as T grows, and empirical power grows in T and N.

5.2.4 Summary of simulation results

The simulation results reported in Table 1 show that existing heteroskedasticity-robust PURTs exhibit huge size distortions (either undersizing or oversizing) when applied to detrended data with time-varying volatility. The proposed test, however, performs remarkably well in this scenario. Results documented in Table 2 show that the new test has fairly good finite sample properties even when the data are not only trending and heteroskedastic, but also cross-sectionally and serially correlated. Therefore, the new test should be helpful in (often complex) empirical applications. However, results not reported here for space considerations show that $\hat{\tau}$ does not remain pivotal under strong forms of cross-sectional dependence such as factor structures (Pesaran, 2007). An effective way of panel unit root testing under strong forms of cross-sectional correlation is to remove the common factor from the data (see for example Bai and Ng, 2004 and Moon and Perron, 2004). While the test in Westerlund (2014) uses this approach, it is, however, not pivotal in the presence of linear trends.

6 Is energy use per capita trend or difference stationary?

6.1 Background

Whether energy use per capita is trend or difference stationary has been intensively investigated in the past two decades. The growing interest in testing the stationarity of per capita energy consumption is attributed to three main reasons (e.g., Hsu et al., 2008; Narayan and Smyth, 2007). First, knowing the direction of causality between per capita energy use and economic growth has gained significant policy relevance as it has direct implications on governments' involvement in global efforts to reduce greenhouse gas emissions. On the one hand, if causality runs from energy consumption to growth, reductions in energy use will have adverse effects on economic growth and, hence, generates reluctance on the part of policy makers to commit to substantial energy use reductions. On the other hand, if causality runs from growth to energy use, and not vice versa, reductions in energy consumption will not be harmful for economic growth. The order of integration of energy use per capita has implications on testing and interpreting the (causal) relationship between energy use and GDP per capita. For instance, Granger causality tests employing level vector autoregressions could be misleading if the series are nonstationary and not cointegrated. Conversely, Granger causality testing by means of variables in levels will be appropriate if the series are either stationary or cointegrated. Consequently, unit root testing is routinely performed before testing for cointegration between energy use and GDP per capita.

Second, stationarity of energy use per capita has implications for the effectiveness of energy policies such as import tariffs on fuels and vehicles or carbon taxes on transportation fuels. In particular, if energy consumption is a stationary process, it will return to its trend after a policy shock. This implies that energy saving policies will have transitory effects only. On the other hand, if energy consumption contains a unit root, such policies will have a permanent impact. Furthermore, nonstationarity implies that (permanent) shocks to energy use are more likely to affect other sectors of the economy as well as macroeconomic aggregates (Narayan and Smyth, 2007).

Third, the order of integration of energy consumption has implications for forecasting energy demand. For instance, if energy consumption is trend stationary, its past behaviour offers valuable information to forecast future energy demand. However, if energy consumption is a unit root process, it does not follow a predictable path and, hence, forecasting energy demand will be more difficult than in the stationary case.

Efforts to test for a unit root in energy use per capita have initially relied on univariate tests.[4] Most of these studies, including Glasure and Lee (1998), Beenstock et al. (1999) and McAvinchey and Yannopoulos (2003) report that the null hypothesis of an I(1) energy consumption series can not be rejected at conventional levels of significance. As an exception to this general conclusion, Altinay and Karagol (2004) document evidence in favor of characterizing energy use in Turkey during 1950–2000 as a trend stationary process. However, given the low power of univariate tests in finite samples, it is not clear if the failure to reject the null of a unit root is an evidence of a truly I(1) series. To circumvent this problem, a few studies have recently applied PURTs to examine the stationarity of energy use per capita. Results have been generally mixed, however. For instance, Joyeux and Ripple (2007) employ the PURTs suggested in Levin et al. (2002) and Im et al. (2003) and find that energy consumption measures are I(1). Narayan and Smyth (2007), on the other hand, report that the unit root null hypothesis can be rejected at the 10% level of significance for 56 of the 182 countries they considered. However, they find strong evidence of a (trend) stationary energy consumption by employing the PURT of Im et al. (2003). Nevertheless, these results should be seen with caution as the studies employ standard PURTs, which are not pivotal if the series exhibit volatility shifts.

6.2 Panel unit root test results

In this section, we study the order of integration of energy use per capita using the heteroskedasticity-robust test suggested in this paper, $\hat{\tau}$, vis-a-vis heteroskedasticity-robust tests of Herwartz and Siedenburg (2008) and Demetrescu and Hanck (2012a). We analyse annual data of energy use per capita (kilogram of oil equivalent per capita) obtained from World Development Indicators.[5] In this data set, energy use refers to "*use of primary energy before transformation to other end-use fuels, which is equal to*

[4] See Hsu et al. (2008) for a review of the empirical literature on unit root testing of energy use per capita.

[5] `www.data.worldbank.org`. Accessed on September 23, 2016.

indigenous production plus imports and stock changes, minus exports and fuels supplied to ships and aircraft engaged in international transport." The study covers 23 OECD economies that are selected according to data availability, from 1960 to 2014.[6] As transforming the series into natural logarithms before undertaking unit root testing is a standard practice in the literature, we test for unit roots both on original series as well as their logarithmic values.

To get an impression if variances in the energy use per capita series exhibit significant changes over time, we plot variance profiles in Figure 1. Variance profiles $\hat{\vartheta}_i(w)$ are computed as

$$\hat{\vartheta}_i(w) = \frac{\sum_{t=1}^{\lfloor sT \rfloor} \hat{\eta}_{it}^2 + (wT - \lfloor wT \rfloor)\hat{\eta}_{i\lfloor wT \rfloor + 1}^2}{\sum_{t=1}^{T} \hat{\eta}_{it}^2}, 0 \leq w \leq 1, \tag{20}$$

where the $\hat{\eta}_{it}$'s are obtained as residuals from AR(1) regressions of the series. Plotting $\hat{\vartheta}_i(w)$ against w, it is straightforward to see that a homoskedastic series would fall on the 45° line and deviations from the diagonal indicate time varying variances. Figure 1 reveals that time-varying variances characterize energy per capita series in most cross section members.

Panel unit root test results are reported in Table 3. Results for all the tests overwhelmingly show that energy use per capita has a unit root. This evidence is consistent with the findings of most of the empirical studies on the area, except, e.g., Narayan and Smyth (2007). However, it is well-known that unit root test results often depend on the specific time period chosen for study. To address this caveat, we perform panel unit root testing on rolling windows of 40 years. Corresponding results depicted in Figure 2 show that while energy use per capita is difference stationary for most of the period, it could be considered trend stationary—at least at the 10 percent significance level—for the sample periods starting between 1965 and 1968. It is worthwhile noting that $\hat{\tau}$ has the lowest $p-$value of the three tests in almost all the considered periods and could suggest an inferential outcome which is distinct from that of the other two tests. In particular, for the period spanning 1966-2005 and based on the 5 percent significance level, $\hat{\tau}$ implies that log energy per capita series can be considered trend stationary while the other two tests suggest to treat the series as difference stationary. Moreover, our

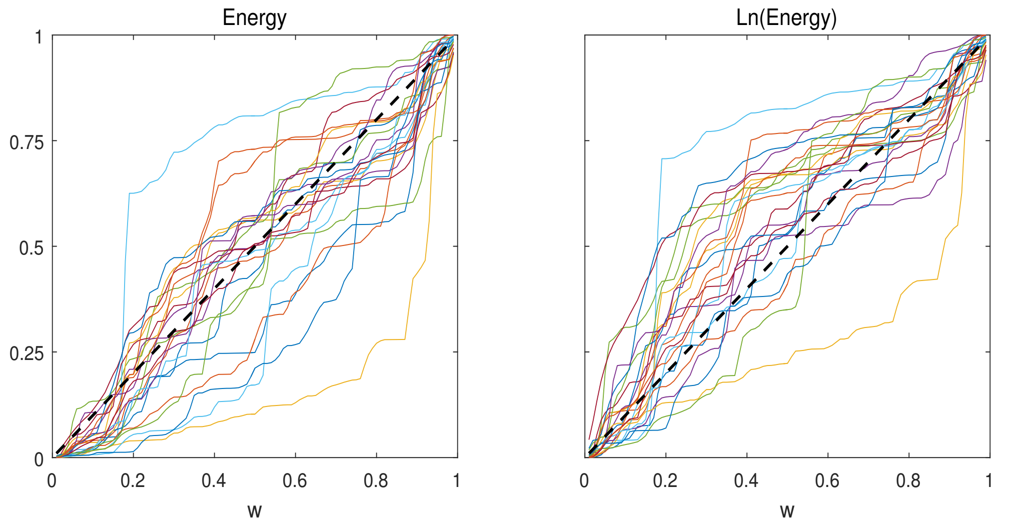
[6] The economies are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom and the United States.

Table 3: Is energy use per capita trend or difference stationary?

Period	Energy use p.c.						Ln (energy use p.c.)					
	y			Δy			y			Δy		
	$\hat{\tau}$	HS	DH	$\hat{\tau}$	HS	DH	$\hat{\tau}$	HS	DH	$\hat{\tau}$	HS	DH
Full period												
1960-2014	0.71	0.55	1.36	**-2.79**	**-3.18**	**-2.56**	1.56	1.24	1.37	**-2.86**	**-2.88**	**-2.64**
50 years window												
1960-2009	0.46	0.43	1.24	**-2.65**	**-2.82**	**-2.17**	0.86	0.91	1.39	**-2.64**	**-2.40**	**-2.13**
1961-2010	-0.18	-0.01	0.93	**-2.82**	**-2.81**	**-2.21**	0.77	0.84	1.34	**-2.81**	**-2.59**	**-2.02**
1962-2011	-0.18	0.00	0.40	**-2.71**	**-2.81**	**-2.32**	0.73	0.83	1.04	**-2.69**	**-2.56**	**-2.01**
1963-2012	-0.29	-0.10	0.18	**-2.69**	**-2.85**	**-2.57**	0.28	0.55	0.64	**-2.76**	**-2.66**	**-2.17**
1964-2013	-0.57	-0.36	-0.31	**-2.67**	**-3.25**	**-2.61**	0.04	0.39	-0.33	**-2.84**	**-2.91**	**-2.74**
1965-2014	-0.27	-0.14	-0.66	**-2.70**	**-3.11**	**-2.18**	0.19	0.48	-0.85	**-2.79**	**-2.83**	**-2.51**

Notes: Reported numbers are estimates of the panel unit root tests $\hat{\tau}$, t_{HS} and t_{DH}. Testing is performed on data that is first prewhitened and then recursively detrended. The lag order used for prewhitening is selected based on the AIC criterion, with the maximum lag lenth set to two. 'Ln' denotes the natural logarithmic transformation. Bold entries represent cases in which the panel unit root null hypothesis is rejected with 5% significance.

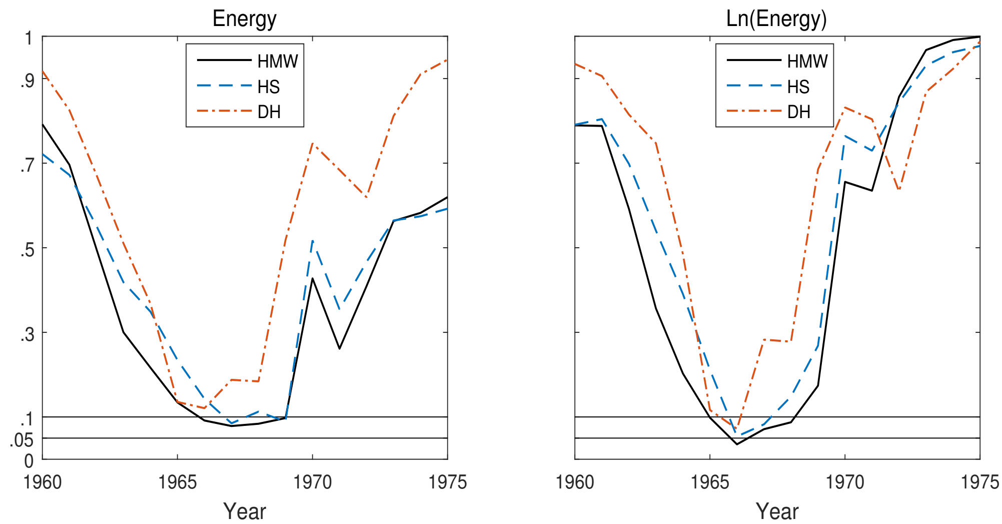
Figure 1: Estimated variance profiles



Notes: 'Ln' denotes the natural logarithmic transformation.

results also highlight the risk of deciding on stationarity of series using one specific time window.

Figure 2: Panel unit root testing over 40-years windows



Notes: The figures depict $p-$values from the panel unit root tests $\hat{\tau}$ (HMW), t_{HS} and t_{DH}. 'Year' represents the year at which the 40-years sample period begins. For further notes, see Table 3.

7 Conclusions

In this paper, we suggested a new panel unit root test (PURT) that works well when the series are trending and exhibit time-varying volatility. The test makes use of the recursive detrending scheme suggested in Demetrescu and Hanck (2014), and the construction of the test statistic fully accounts for non-zero expectation of the pooled panel regression estimator and the variance of its centered counterpart. Accordingly, the resulting test statistic has a Gaussian limiting distribution. Monte Carlo simulation results show that the test has satisfactory finite sample properties. In particular, the test tends to be conservative, while it shows remarkable power. Hence, this test should be useful in panel unit root testing of several trending macroeconomic and financial time series such as GDP per capita, industrial production, money supply and commodity prices.

The empirical illustration examined the order of integration of energy use per capita. Results using data from 23 OECD economies for the period 1960-2014 show that energy use per capita is often difference stationary. Yet, there are also a few sub-periods for which the series could be considered as trend stationary.

A particular limitation of the suggested test is that it does not perform well under a strong form of cross-sectional dependence. An effective way of panel unit root testing under strong forms of cross-sectional correlation is to remove the common factor from the data (Bai and Ng, 2004; Moon and Perron, 2004). Consequently, it appears worthwhile to see in a future research if such an approach would yield a panel unit root test that works for strongly correlated panels with trending and heteroskedastic time series.

Acknowledgements

We thank Jörg Breitung and Matei Demetrescu for helpful comments and suggestions.

Appendix

In order to prove Proposition 1 we proceed in three steps. First, stating Lemmas 1 and 2 below we are explicit on the order properties of the variance s_{NT}^2 in (13) and define a mixing array which is essential to prove the asymptotic result for our test statistic (Part A.1). Second, before we derive asymptotic normality for $\hat{\tau}$ defined in (16), we establish a corresponding result for τ assuming that time specific expectations and variances are known (A.2). Third, we discuss the stochastic properties of the estimated moments $\hat{\nu}_t$ and $\hat{s}_{NT}^2$ and build upon the result for τ to finally derive the Gaussian limit distribution for $\hat{\tau}$ and thus, to prove Proposition 1 (A.3). The following derivations proceed under the null hypothesis and assumptions $\mathcal{A}$. Furthermore, we assume $N/T^2 \to 0$ throughout.

A.1 - Variance order and mixing array

Recalling from Section 4, the detrending scheme in (7) and (8) obtains coefficients $a_{i,t-1}$, finite for all $i < t$ and $t \leq T$, i.e.,

$$a_{i,t-1} = 1 + \frac{2}{t-1}(t-i) - 3\left(1 - \frac{(i-1)i}{(t-1)t}\right).$$

Let $\tilde{a}_{i,t-1} = \left(1 - \frac{1}{T}\right) a_{i,t-1}$ and $\bar{a}_{i,t-1} = \frac{1}{T} a_{i,t-1}$. The mean of $\tilde{\boldsymbol{y}}_{t-1}' \Delta \boldsymbol{y}_t^*$ is

$$\begin{aligned}
\nu_t = E\left[\tilde{\boldsymbol{y}}_{t-1}' \Delta \boldsymbol{y}_t^*\right] &= E\left[\sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\right)\right] \\
&= -\sum_{i=1}^{t-1}\bar{a}_{i,t-1}E\left[\boldsymbol{e}_i'\boldsymbol{e}_i\right] \\
&= -\sum_{i=1}^{t-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i),
\end{aligned} \tag{21}$$

since $E[\boldsymbol{e}_i'\boldsymbol{e}_k] = 0$ for all $i \neq k$. For the variance, we have

$$s_{NT}^2 = \frac{1}{NT}\left(E\left[\sum_{t=2}^{T}\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*\right]^2 - \left(\sum_{t=2}^{T}\nu_t\right)^2\right) = \zeta_1 - \zeta_2 + \zeta_3 + \zeta_4 + \zeta_5 - \frac{1}{NT}\left(\sum_{t=2}^{T}\nu_t\right)^2, \tag{22}$$

where the sums ζ_i, $i = 1, \ldots, 5$, are defined as in (14). Since

$$\frac{1}{NT}\left(\sum_{t=2}^{T}\nu_t\right)^2 = \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\text{tr}(\Omega_i)^2 + \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T-1}\sum_{s=t+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\text{tr}(\Omega_i)^2$$
$$+ \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}\bar{a}_{i,s-1}\bar{a}_{j,t-1}\text{tr}(\Omega_i)\text{tr}(\Omega_j)$$

we can rewrite s_{NT}^2 in (22) as

$$s_{NT}^2 = \tilde{\zeta}_1 - \zeta_2 + \zeta_3 + \tilde{\zeta}_4 + \tilde{\zeta}_5, \tag{23}$$

where ζ_2 and ζ_3 are defined in (14) and

$$\tilde{\zeta}_1 = \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}\bar{a}_{i,s-1}\bar{a}_{j,t-1}\text{tr}(\Omega_i\Omega_j)$$
$$\tilde{\zeta}_4 = \frac{1}{NT}\sum_{i=1}^{T}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\left(E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \text{tr}(\Omega_i)^2 + \sum_{j=1\neq i,t}^{T}\text{tr}(\Omega_i\Omega_j)\right)$$
$$= \frac{1}{NT}\sum_{i=1}^{T}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\left(E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \text{tr}(\Omega_i)^2\right) + \frac{1}{NT}\sum_{i=1}^{T}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\sum_{j=1\neq i,t}^{T}\text{tr}(\Omega_i\Omega_j)$$
$$= \tilde{\zeta}_{41} + \tilde{\zeta}_{42}$$
$$\tilde{\zeta}_5 = \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\left(E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \text{tr}(\Omega_i)^2 + \sum_{j=1\neq i,t,s}^{T}\text{tr}(\Omega_i\Omega_j)\right)$$
$$= \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\left(E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \text{tr}(\Omega_i)^2\right)$$
$$+ \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\sum_{j=1\neq i,t,s}^{T}\text{tr}(\Omega_i\Omega_j)$$
$$= \tilde{\zeta}_{51} + \tilde{\zeta}_{52}.$$

The following lemma characterizes the variance in more detail.

Lemma 1. *Under assumptions $\mathcal{A}$ the variance s_{NT}^2 is of order $\mathcal{O}(T)$. Moreover, $s_{NT}^2/T > 0$ for all $N, T \geq 1$.*

Proof. First, we determine the order of s_{NT}^2 as provided in (23). Then, for instance,

$$\frac{\tilde{\zeta}_1}{T} = \frac{2}{NT^4}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}a_{i,s-1}a_{j,t-1}\text{tr}(\Omega_i\Omega_j)$$

is bounded in T and in N because $\text{tr}(\Omega_i\Omega_j) = \mathcal{O}(N)$, i.e. $\tilde{\zeta}_1 = \mathcal{O}(T)$. Analogously, it follows $\zeta_2 = \mathcal{O}(T)$, $\zeta_3 = \mathcal{O}(T)$, $\tilde{\zeta}_{42} = \mathcal{O}(1)$ and $\tilde{\zeta}_{52} = \mathcal{O}(T)$. Furthermore, since

$E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2] - \text{tr}(\Omega_i)^2 = \mathcal{O}(N)$, one has $\tilde{\zeta}_{41} = \mathcal{O}(T^{-1})$ and $\tilde{\zeta}_{51} = \mathcal{O}(1)$. Altogether, $s_{NT}^2 = \mathcal{O}(T)$.

Secondly, the variance is greater or equal to zero, $s_{NT}^2 \geq 0$, by definition (cf. equation (22)). To see that s_{NT}^2/T strictly exceeds zero for all $N, T \geq 1$, the variance of the numerator

$$\sum_{t=2}^{T} \frac{1}{\sqrt{NT}} \sum_{i=1}^{t-1} \left(a_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \frac{1}{T} a_{i,t-1} \sum_{k=2}^{T} \boldsymbol{e}_i'\boldsymbol{e}_k \right)$$
$$= \underbrace{\sum_{t=2}^{T} \frac{1}{\sqrt{NT}} \sum_{i=1}^{t-1} a_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t}_{=:\mathcal{X}_1} - \underbrace{\sum_{t=2}^{T} \frac{1}{\sqrt{NT}} \sum_{i=1}^{t-1} \frac{1}{T} a_{i,t-1} \sum_{k=2}^{T} \boldsymbol{e}_i'\boldsymbol{e}_k}_{=:\mathcal{X}_2}$$

can be rewritten as $s_{NT}^2 = \text{Var}[\mathcal{X}_1 - \mathcal{X}_2] = \text{Var}[\mathcal{X}_1] + \text{Var}[\mathcal{X}_2] - 2 \cdot \text{Cov}[\mathcal{X}_1, \mathcal{X}_2]$. The components of $\text{Var}[\mathcal{X}_1]$ consist of terms $\text{tr}(\Omega_i\Omega_t)/N$ which are strictly positive, $\text{tr}(\Omega_i\Omega_j)/N \geq \lambda_i^{(1)}(\sum_{l=1}^{N} \lambda_j^{(l)})/N > 0$ for all $i, j = 1, \ldots, T$, $N \geq 1$, and eigenvalues $\lambda_i^{(1)}, \lambda_j^{(l)} > 0$ (from assumption $\mathcal{A}(ii)$). Hence, it can be shown that $\text{Var}[\mathcal{X}_1]/T > 0$ for all $N, T \geq 1$. Furthermore, the variance terms $\text{Var}[\mathcal{X}_1] + \text{Var}[\mathcal{X}_2]$ can be shown to dominate the covariance term $2 \cdot \text{Cov}[\mathcal{X}_1, \mathcal{X}_2]$ so that $s_{NT}^2/T = \text{Var}[\mathcal{X}_1 - \mathcal{X}_2]/T > 0$ for all $N, T \geq 1$.

□

To show the asymptotic normality of $\hat{\tau}$ in (16), we employ a central limit theorem for near-epoch dependent sequences. For this, we define a mixing array[7]

$$\boldsymbol{V}_{T,t} = \left(\boldsymbol{e}_t, \sum_{k=t+1}^{T} \boldsymbol{e}_k \right). \tag{24}$$

The generated sigma algebra corresponds to

$$\mathcal{F}_{T,t-m}^{t+m} = \sigma(\boldsymbol{V}_{T,s}, t - m \leq s \leq t + m) = \sigma(\boldsymbol{e}_{t-m}, \ldots, \boldsymbol{e}_{t+m}, \mathcal{E}_{T,t-m}, \ldots, \mathcal{E}_{T,t+m}),$$

where $\mathcal{E}_{T,t+m} := \sum_{k=t+m+1}^{T} \boldsymbol{e}_k$. In particular, $\mathcal{F}_{T,-\infty}^{t} = \sigma(\ldots, \boldsymbol{e}_t, \ldots, \mathcal{E}_{T,t})$. This definition of the sigma algebra is similar to the one used in Lemma 3 of Demetrescu and Hanck (2014), but contains the vector $(\boldsymbol{e}_t, \mathcal{E}_{T,t})$ instead of the sum of the two entries. Using the notation of Davidson (1994) we state the following result:

Lemma 2. *$\boldsymbol{V}_{T,t}$ in (24) is α-mixing of size $-\beta$ for $0 \leq \beta < \infty$.*

[7] For simplicity the subscript N is omitted here, since the process is near-epoch dependent with respect to the time dimension.

Proof. To show the mixing property of $\boldsymbol{V}_{T,t}$ consider the sequence

$$\alpha_m = \sup_t \sup_{A\in\mathcal{F}^{\infty}_{T,t+m},B\in\mathcal{F}^{t}_{T,-\infty}} \left|P(A\cap B) - P(A)P(B)\right|$$

for all $T \geq 1$ and events A and B. The second supremum is taken with respect to the sigma algebras $\mathcal{F}^t_{T,-\infty} = \sigma(\ldots,\boldsymbol{e}_t,\ldots,\mathcal{E}_{T,t})$ and $\mathcal{F}^{\infty}_{T,t+m} = \sigma(\boldsymbol{e}_{t+m},\ldots,\mathcal{E}_{T,t+m},\ldots)$. Noticing that the $\boldsymbol{e}_i$'s are uncorrelated, dependence between A and B (i.e., $|P(A\cap B) - P(A)P(B)\big| > 0$) can only occur by involving terms of $\mathcal{E}_{T,t}$. More precisely, the sums $\mathcal{E}_{T,t} = \sum_{k=t}^{T}\boldsymbol{e}_k$ and $\mathcal{E}_{T,t+m} = \sum_{k=t+m+1}^{T}\boldsymbol{e}_k$ both include error terms $\{\boldsymbol{e}_{t+m+1},\ldots,\boldsymbol{e}_T\}$ such that $\alpha_m \neq 0$ for events

$$A,B\in\mathcal{F}^t_{T,-\infty}\cap\mathcal{F}^{\infty}_{T,t+m} = \sigma\left(\mathcal{E}_{T,t+m},\mathcal{E}_{T,t+m+1},\ldots\right)\subseteq\sigma\left(\boldsymbol{e}_{t+m+1},\boldsymbol{e}_{t+m+2},\ldots,\boldsymbol{e}_T\right). \quad (25)$$

For increasing m the number of random variables generating the sigma algebra decreases. For $m > T-t-1$ the generated sigma algebra in (25) is the empty set. Thus, $\alpha_m = 0$ for $m > T-t-1$ for all $T\geq 1$ and $-\infty \leq t \leq \infty$. It follows $\alpha_m = \mathcal{O}(m^{-\beta})$ for all $0\leq\beta<\infty$. □

A.2 - Asymptotic distribution with true moments

In the following, $\hat{\nu}_t$ and $\hat{s}_{NT}$ are substituted by their theoretical counterparts so that asymptotic normality of τ defined in (12) is shown first. To prove asymptotic normality of τ we rewrite the numerator from (12) as

$$\sum_{t=2}^{T}\frac{1}{\sqrt{NT}}(\tilde{\boldsymbol{y}}'_{t-1}\Delta\boldsymbol{y}^*_t - \nu_t) = \sum_{t=2}^{T}\frac{1}{\sqrt{NT}}\left(\sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}'_i\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}'_i\boldsymbol{e}_k\right) - \nu_t\right).$$

From the variance

$$s^2_{NT} = E\left[\left(\sum_{t=2}^{T}\frac{1}{\sqrt{NT}}\sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}'_i\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}'_i\boldsymbol{e}_k\right) - \nu_t\right)^2\right],$$

a standardized sequence is given by

$$\begin{aligned} X_{NT,t} :=& \frac{1}{\sqrt{NT}}(\tilde{\boldsymbol{y}}'_{t-1}\Delta\boldsymbol{y}^*_t - \nu_t)\Big/ s_{NT} \\ =& \frac{1}{\sqrt{NT}}\sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}'_i\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}'_i\boldsymbol{e}_k + \bar{a}_{i,t-1}\mathrm{tr}(\Omega_i)\right)\Big/ s_{NT}. \end{aligned} \quad (26)$$

A central limit theorem (CLT) for $\tau = \sum_{t=2}^{T}X_{NT,t}$ that controls for near-epoch dependence (NED) of $X_{NT,t}$ holds if the following conditions of Corollary 24.7 in Davidson (1994) are fulfilled:

(a) $X_{NT,t}$ is $\mathcal{F}^t_{T,-\infty}$ measurable with $E\left[X_{NT,t}\right] = 0$ and $E\left[\left(\sum_{t=2}^{T} X_{NT,t}\right)^2\right] = 1$.

(b) There exists a constant array $\{c_{NT,t}\}$ such that $\sup_{T,t} ||X_{NT,t}/c_{NT,t}||_r < \infty$ for $r > 2$.

(c) $X_{NT,t}$ is L_2-NED of size -1 on $\boldsymbol{V}_{T,t}$ which is α-mixing of size $-r/(r-2)$.

(d) $\sup_T\{T(\max_{1\leq t\leq T} c_{NT,t})^2\} < \infty$.

Lemma 3. *Under assumptions $\mathcal{A}$ on the error terms and $N/T^2 \to 0$ the conditions (a)-(d) are fulfilled for the sequence $X_{NT,t}$ in (26) and the mixing process $\boldsymbol{V}_{T,t}$ in (24).*

From Corollary 24.7 in Davidson (1994) and Lemma 3 asymptotic normality of τ in (12) follows directly and can be stated as

Corollary 1. *Under assumptions $\mathcal{A}$ and $N/T^2 \to 0$,*

$$\tau = \sum_{t=2}^{T} X_{NT,t} \xrightarrow{d} \mathcal{N}(0,1), \quad N,T \to \infty.$$

Remark. *The CLT in T holds for all $N \geq 1$, in particular for $N \to \infty$. The joint limit $N,T \to \infty$, furthermore, provides convergence of the sums of $\mathbf{e}_i'\mathbf{e}_t$ and thus, ensures that the assumptions of the CLT are fulfilled. Note that we show asymptotic normality in the joint limit $N,T \to \infty$ instead of the sequential limit applying the convergence properties following, for instance, from Theorem 4.4 of Billingsley (1999).*

Proof of Lemma 3. **Condition (a):** As it is a function of measurable random variables, $X_{NT,t} = f(\boldsymbol{e}_1, \dots, \boldsymbol{e}_t, \mathcal{E}_t)$ is measurable with respect to $\mathcal{F}^t_{T,-\infty}$. The sequence $X_{NT,t}$ is centered and standardized such that $E[X_{NT,t}] = 0$ and $E\left[\left(\sum_{t=2}^{T} X_{NT,t}\right)^2\right] = 1$ follow directly.

Condition (b): Let the array of constants be equal to $\{c_{NT,t}\} = \{1/s_{NT}\}$ and set $r = 4$.

Then,

$$\begin{aligned}
\Big|\Big|X_{NT,t}/c_{NT,t}\Big|\Big|_4 &= E\left[\Big|X_{NT,t}\Big|^4 \Big/ c^4_{NT,t}\right]^{\frac{1}{4}} = E\left[\Big|X_{NT,t}\Big|^4 \cdot s^4_{NT}\right]^{\frac{1}{4}} \\
&= \left[E\Big|\frac{1}{\sqrt{NT}}\sum_{i=1}^{t-1}\Big(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k + \bar{a}_{i,t-1}\mathrm{tr}(\Omega_i)\Big)\Big|^4\right]^{\frac{1}{4}} \\
&= \Big(E\Big|\frac{1}{\sqrt{NT}}\sum_{i=1}^{t-1}\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \sum_{i=1}^{t-1}\bar{a}_{i,t-1}\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k + \sum_{i=1}^{t-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i)\Big)\Big|^4\Big)^{\frac{1}{4}} \\
&\leq \Big(\frac{1}{N^2T^2}E\Big|\sum_{i=1}^{t-1}\tilde{a}_{i,t-1}\sum_{l=1}^{N}e_{li}e_{lt}\Big|^4\Big)^{\frac{1}{4}} + \Big(\frac{1}{N^2T^6}E\Big|\sum_{i=1}^{t-1}a_{i,t-1}\sum_{\substack{k=2\\k\neq t}}^{T}\sum_{l=1}^{N}e_{li}e_{lk}\Big|^4\Big)^{\frac{1}{4}} \\
&\quad + \Big(\frac{1}{N^2T^6}E\Big|\sum_{i=1}^{t-1}a_{i,t-1}\mathrm{tr}(\Omega_i)\Big)\Big|^4\Big)^{\frac{1}{4}} \\
&< \infty \qquad \text{for all } N, T.
\end{aligned} \tag{27}$$

The inequality holds by virtue of the Minkowski inequality. The first part in (27) is finite with similar reasoning as in equation (19) of Herwartz et al. (2016), i.e. nonzero expectations arise only from terms involving $(e_{li}e_{lt})^4$ or $e^2_{li}e^2_{lj}e^4_{mt}$, $i \neq j$. The second and the third term contain the product of error terms from the same time period ($e_{li}e_{lk}$ with $i = k$). Thus, for finiteness we need to assume finiteness up to order eight, $E|e_{lt}|^8 < \infty$, which was formulated in assumption $\mathcal{A}(iii)$. Furthermore, noticing that N and T can be related by means of $N/T^2 \to 0$, the denominator controls for increasing N and T adequately.

Condition (c): To verify this condition, $X_{NT,t}$ is shown to be near-epoch dependent on $\boldsymbol{V}_{T,t}$ meaning that

$$\Big(E\left[X_{NT,t} - E[X_{NT,t}|\ \mathcal{F}^{t+m}_{T,t-m}]\right]^2\Big)^{1/2} \leq c_{NT,t}\rho_m, \tag{28}$$

where ρ_m is a sequence of order $\mathcal{O}(m^{-1})$ and $c_{NT,t}$ is the positive constant defined in condition (b).

The expectation of $X_{NT,t}$ conditioned on the m neighboring sigma algebras is

$$\begin{aligned}
E\left[X_{NT,t}|\ \mathcal{F}^{t+m}_{T,t-m}\right] &= E\left[\frac{1}{\sqrt{NT}}\Big(\sum_{i=1}^{t-1}(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k) - \nu_t\Big)\Big/ s_{NT}\Big|\ \mathcal{F}^{t+m}_{T,t-m}\right] \\
&= \frac{1}{\sqrt{NT}}\Big(\sum_{i=1}^{t-1}E\left[\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\Big|\ \mathcal{F}^{t+m}_{T,t-m}\right] - \nu_t\Big)\Big/ s_{NT},
\end{aligned} \tag{29}$$

with

$$\sum_{i=1}^{t-1} E\left[\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k \middle| \mathcal{F}_{T,t-m}^{t+m}\right]$$

$$= \sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}E\left[\boldsymbol{e}_i'\boldsymbol{e}_t \middle| \mathcal{F}_{T,t-m}^{t+m}\right] - \bar{a}_{i,t-1}E\left[\boldsymbol{e}_i'\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_k \middle| \mathcal{F}_{T,t-m}^{t+m}\right]\right)$$

$$= \sum_{i=t-m}^{t-1}\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \sum_{i=1}^{t-m-1}\tilde{a}_{i,t-1}E(\boldsymbol{e}_i'\boldsymbol{e}_i) - \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=t-m,k\neq t}^{T}\boldsymbol{e}_k. \qquad (30)$$

Here, parts of the conditional expectations cancel out because of measurability or zero covariance of the corresponding random variables. Inserting (30) into (29) obtains

$$E\left[X_{NT,t}|\mathcal{F}_{T,t-m}^{t+m}\right] = \frac{1}{\sqrt{NT}}\left(\sum_{i=t-m}^{t-1}\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \sum_{i=1}^{t-m-1}\tilde{a}_{i,t-1}E(\boldsymbol{e}_i'\boldsymbol{e}_i)\right.$$

$$\left. - \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{\substack{k=t-m\\k\neq t}}^{T}\boldsymbol{e}_k + \sum_{i=1}^{t-1}\bar{a}_{i,t-1}\mathrm{tr}(\Omega_i)\right) \Big/ s_{NT}$$

$$= \frac{1}{\sqrt{NT}}\sum_{i=t-m}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\right.$$

$$\left. + \bar{a}_{i,t-1}\mathrm{tr}(\Omega_i) - \bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\right) \Big/ s_{NT}.$$

Hence, the condition for NED sequences in (28) is fulfilled by noticing

$$\left(E\left[X_{NT,t} - E[X_{NT,t}|\mathcal{F}_{T,t-m}^{t+m}]\right]^2\right)^{\frac{1}{2}}$$

$$= \left(\frac{1}{s_{NT}\sqrt{NT}}E\left[\sum_{i=1}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k\right) - \nu_t\right.\right.$$

$$\left.\left. - \sum_{i=t-m}^{t-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k + \bar{a}_{i,t-1}\mathrm{tr}(\Omega_i) - \bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\right)\right]^2\right)^{\frac{1}{2}}$$

$$= \frac{1}{s_{NT}\sqrt{NT}}\left(E\left[\sum_{i=1}^{t-m-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\left(\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k - \mathrm{tr}(\Omega_i)\right)\right) + \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\right]^2\right)^{\frac{1}{2}}$$

$$= c_{NT,t}\left(\frac{1}{NT}E\left[\sum_{i=1}^{t-m-1}\left(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\left(\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k - \mathrm{tr}(\Omega_i)\right)\right) + \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\right]^2\right)^{\frac{1}{2}}$$

$$=: c_{NT,t}\,\rho_m.$$

In order to show that $\rho_m = \mathcal{O}(m^{-1})$, we apply Minkowski's inequality:

$$\rho_m = \Bigg(\frac{1}{NT}E\Bigg[\sum_{i=1}^{t-m-1}\Big(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\Big(\sum_{\substack{k=2\\k\neq t}}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k - \text{tr}(\Omega_i)\Big)\Big) + \sum_{i=t-m}^{t-1}\bar{a}_{i,t-1}\boldsymbol{e}_i'\sum_{k=2}^{t-m}\boldsymbol{e}_k\Bigg]^2\Bigg)^{\frac{1}{2}}$$

$$\leq \frac{1}{\sqrt{T}}\underbrace{\left(\frac{1}{N}E\left[\sum_{i=1}^{t-m-1}\Big(\tilde{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_t - \bar{a}_{i,t-1}\Big(\sum_{k=2,k\neq t}^{T}\boldsymbol{e}_i'\boldsymbol{e}_k - \text{tr}(\Omega_i)\Big)\Big)\right]^2\right)^{\frac{1}{2}}}_{=\mathcal{O}(\sqrt{t-m-1})}$$

$$+\frac{1}{\sqrt{T}}\underbrace{\left(\frac{1}{N}E\left[\sum_{i=t-m}^{t-1}\sum_{k=2}^{t-m}\bar{a}_{i,t-1}\boldsymbol{e}_i'\boldsymbol{e}_k\right]^2\right)^{\frac{1}{2}}}_{=\mathcal{O}((m-1)(t-m)/T)}. \tag{31}$$

The error terms have finite fourth order moments and, hence, dividing by $\sqrt{T}$ the L_2-norms are bounded for all $m, t, T \geq 1$. Furthermore, for $m \geq t-1$ the sums are zero such that $\rho_m = 0$. Consequently, for $0 \leq \beta < \infty$ we have $m^\beta \rho_m = \mathcal{O}(1)$ if $m < t-1$, because both m and ρ_m are bounded, and if $m \geq t-1$ because $\rho_m = 0$. Thus, $\rho_m = \mathcal{O}(m^{-\beta})$ for every $0 \leq \beta < \infty$ and especially, for $\beta = 1$ such that $\rho_m = \mathcal{O}(m^{-1})$.

Furthermore, from Lemma 2 it follows that $\boldsymbol{V}_{T,t}$ is mixing of size $-\beta$ for $\beta \geq 0$. In particular, for $r = 4$ the order of convergence is $-\beta = -r/(r-2) = -2$ as considered in condition (b) of Theorem 24.6 in Davidson (1994).

Condition (d): To show that this condition holds for $c_{NT,t} = 1/s_{NT}$, notice that s_{NT}^2 is of order $\mathcal{O}(T)$ following Lemma 1. Together with $s_{NT}^2/T > 0$ this directly indicates the finiteness required by condition (d):

$$\sup_T \left\{T\left(\max_{1\leq t\leq T} c_{NT,t}\right)^2\right\} = \sup_T \frac{T}{s_{NT}^2} < \infty.$$

□

A.3 - Asymptotic distribution with estimated moments

Mean estimation

The representation in (21) reduces the estimation of ν_t to the estimation of terms such as $\text{tr}(\Omega_i)$ so that convergence is assured by the increasing panel and time dimensions N and T. For the model residuals evaluated under the null hypothesis $\hat{\boldsymbol{e}}_t = \Delta\boldsymbol{y}_t^*$ the estimator is explicitly given as

$$\hat{\nu}_t = -\sum_{i=1}^{t-1}\bar{a}_{i,t-1}\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i.$$

The following lemma states convergence of $\frac{1}{\sqrt{NT}}(\hat{\nu}_t - \nu_t)$ so that the theoretical counterpart ν_t can be used to prove asymptotic normality of $\hat{\tau}$.

Lemma 4. *Under assumptions $\mathcal{A}$,*

$$\frac{1}{\sqrt{NT}}(\hat{\nu}_t - \nu_t) \xrightarrow{p} 0, \quad \textit{for} \quad N, T \to \infty.$$

Proof. To show the convergence in probability, we rewrite

$$\frac{1}{\sqrt{NT}}(\hat{\nu}_t - \nu_t) = \frac{1}{\sqrt{NT}} \left(\sum_{i=1}^{t-1} \frac{1}{T} a_{i,t-1} \left(\hat{\boldsymbol{e}}_i' \hat{\boldsymbol{e}}_i - E[\boldsymbol{e}_i' \boldsymbol{e}_i] \right) \right).$$

From (10) we have $\hat{\boldsymbol{e}}_i = \boldsymbol{e}_i - \frac{1}{T} \sum_{t=2}^{T} \boldsymbol{e}_t$. For finite T the variance and covariance of the estimator $\hat{\boldsymbol{e}}_i$ differ from corresponding moments of $\boldsymbol{e}_i$. However, asymptotically they are equivalent. For instance, for any $i = 1, \ldots, T$,

$$\begin{aligned} E\left[\hat{\boldsymbol{e}}_i' \hat{\boldsymbol{e}}_i\right] &= E\left[\Big(\boldsymbol{e}_i - \frac{1}{T} \sum_{t=2}^{T} \boldsymbol{e}_t \Big)' \Big(\boldsymbol{e}_i - \frac{1}{T} \sum_{t=2}^{T} \boldsymbol{e}_t \Big) \right] \\ &= E\left[\boldsymbol{e}_i' \boldsymbol{e}_i\right] - 2E\left[\boldsymbol{e}_i' \Big(\frac{1}{T} \sum_{t=2}^{T} \boldsymbol{e}_t \Big) \right] + E\left[\Big(\frac{1}{T} \sum_{t=2}^{T} \boldsymbol{e}_t \Big)' \Big(\frac{1}{T} \sum_{t=2}^{T} \boldsymbol{e}_t \Big) \right] \\ &= \Big(1 - \frac{2}{T} \Big) E\left[\boldsymbol{e}_i' \boldsymbol{e}_i\right] + \frac{1}{T^2} \sum_{t=2}^{T} E[\boldsymbol{e}_t' \boldsymbol{e}_t] \to E\left[\boldsymbol{e}_i' \boldsymbol{e}_i\right], \quad T \to \infty. \end{aligned} \tag{32}$$

Similarly, the higher moments converge, i.e. $E[(\hat{\boldsymbol{e}}_i{}' \hat{\boldsymbol{e}}_i)^2] \to E[(\boldsymbol{e}_i' \boldsymbol{e}_i)^2]$ for $T \to \infty$.

Applying these results and the Markov inequality we have

$$
\begin{aligned}
P\left(\left|\frac{\hat{\nu}_t-\nu_t}{\sqrt{NT}}\right|>\varepsilon\right) &< \frac{E\left[\frac{1}{NT}\left(\hat{\nu}_t-\nu_t\right)^2\right]}{\varepsilon^2}\\
&= \frac{1}{\varepsilon^2}E\left[\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{t-1}\frac{1}{T}a_{i,t-1}\left(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i-E[\boldsymbol{e}_i'\boldsymbol{e}_i]\right)\right)^2\right]\\
&= \frac{1}{\varepsilon^2 NT^3}E\left[\left(\sum_{i=1}^{t-1}a_{i,t-1}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i-E[\boldsymbol{e}_i'\boldsymbol{e}_i])\right)^2\right]\\
&= \frac{1}{\varepsilon^2 NT^3}\left[\sum_{i=1}^{t-1}a_{i,t-1}^2E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i-E[\boldsymbol{e}_i'\boldsymbol{e}_i])^2\right]\right.\\
&\quad\left.+2\sum_{i=1}^{t-1}\sum_{j=i+1}^{t-1}a_{i,t-1}a_{j,t-1}E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i-E[\boldsymbol{e}_i'\boldsymbol{e}_i])\left(\hat{\boldsymbol{e}}_j'\hat{\boldsymbol{e}}_j-E[\boldsymbol{e}_j'\boldsymbol{e}_j]\right)\right]\right]\\
&= \frac{1}{\varepsilon^2T^3}\underbrace{\sum_{i=1}^{t-1}a_{i,t-1}^2\frac{1}{N}\left(E[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2]-(\mathrm{tr}(\Omega_i))^2\right)}_{\mathcal{O}(T)}\\
&\quad+\frac{2}{\varepsilon^2NT^3}\sum_{i=1}^{t-1}\sum_{j=i+1}^{t-1}a_{i,t-1}a_{j,t-1}\underbrace{E\left[\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i\hat{\boldsymbol{e}}_j'\hat{\boldsymbol{e}}_j-E[\boldsymbol{e}_i'\boldsymbol{e}_i]E[\boldsymbol{e}_j'\boldsymbol{e}_j]\right]}_{=0}\\
&\longrightarrow 0 \quad \text{for} \quad \varepsilon>0, \quad N,T\to\infty.
\end{aligned}
$$

□

Variance estimation

According to the representation of s_{NT}^2 in (23) the variance estimator is

$$\hat{s}_{NT}^2=\hat{\hat{\zeta}}_1-\hat{\zeta}_2+\hat{\zeta}_3+\hat{\hat{\zeta}}_4+\hat{\hat{\zeta}}_5, \tag{33}$$

where

$$\hat{\hat{\zeta}}_1 = \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{j=i+1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=j+1}^{T}\bar{a}_{i,s-1}\bar{a}_{j,t-1}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j)^2$$

$$\hat{\zeta}_2 = \frac{2}{NT}\sum_{i=1}^{T-1}\sum_{s=i+1}^{T}\sum_{t=i+1}^{T}\tilde{a}_{i,s-1}\bar{a}_{i,t-1}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_s)^2$$

$$\hat{\zeta}_3 = \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2$$

$$\hat{\hat{\zeta}}_4 = \frac{1}{NT}\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\bar{a}_{i,t-1}^2\left(\underbrace{(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2 - (\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2}_{=0} + \sum_{j=1,j\neq i,t}^{T}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j)^2\right)$$

$$\begin{aligned}\hat{\hat{\zeta}}_5 &= \hat{\hat{\zeta}}_{51} + \hat{\hat{\zeta}}_{52}\\ &= \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\underbrace{\left((\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2 - (\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_i)^2\right)}_{=0}\\ &\quad + \frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\sum_{j=1,j\neq i,t,s}^{T}(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_j)^2.\end{aligned}$$

However, unlike $\frac{1}{\sqrt{NT}}(\hat{\nu}_t - \nu_t) \xrightarrow{p} 0$, the difference $\hat{s}_{NT}^2 - s_{NT}^2$ does not converge in probability. To determine the order of this difference, we consider the components in (33) separately. For $N, T \to \infty$, the orders of the differences of $\hat{\hat{\zeta}}_1$, $\hat{\zeta}_2$, $\hat{\zeta}_3$ and $\hat{\hat{\zeta}}_{52}$ from their theoretical counterparts can be derived in the same form. As an example, we consider

$$\hat{\zeta}_3 - \zeta_3 = \sum_{i=1}^{T-1}\frac{1}{NT}\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2 - \mathrm{tr}(\Omega_i\Omega_t)\right]. \tag{34}$$

To define the order of $\hat{\zeta}_3 - \zeta_3$ we use $E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2\right] \to E\left[(\boldsymbol{e}_i'\boldsymbol{e}_t)^2\right]$ and $E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_s)^2\right] \to E\left[(\boldsymbol{e}_i'\boldsymbol{e}_t)^2(\boldsymbol{e}_i'\boldsymbol{e}_s)^2\right]$ which can be derived similarly to (32). Accordingly, the difference in

(34) has mean zero but its variance does not vanish. More precisely,

$$
\begin{aligned}
&E\left[\left(\sum_{i=1}^{T-1}\frac{1}{NT}\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\operatorname{tr}(\Omega_i\Omega_t)\right]\right)^2\right]\\
&=\frac{1}{N^2T^2}\left(\sum_{i=1}^{T-1}E\left[\left(\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\operatorname{tr}(\Omega_i\Omega_t)\right]\right)^2\right]\right.\\
&\quad\left.+2\sum_{i=1}^{T-2}\sum_{j=i+1}^{T-1}E\left[\left(\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\operatorname{tr}(\Omega_i\Omega_t)\right]\right)\left(\sum_{t=j+1}^{T}\tilde{a}_{j,t-1}^2\left[(\hat{\boldsymbol{e}}_j'\hat{\boldsymbol{e}}_t)^2-\operatorname{tr}(\Omega_j\Omega_t)\right]\right)\right]\right)\\
&=\frac{1}{N^2T^2}\left(\sum_{i=1}^{T-1}\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^4\left[E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^4\right]-(\operatorname{tr}(\Omega_i\Omega_t))^2\right]\right.\\
&\qquad+\sum_{i=1}^{T-1}2\sum_{t=i+1}^{T}\sum_{s=t+1}^{T}\tilde{a}_{i,t-1}^2\tilde{a}_{i,s-1}^2\left[E\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_s)^2\right]-\operatorname{tr}(\Omega_i\Omega_t)\operatorname{tr}(\Omega_i\Omega_s)\right]\\
&\quad\left.+2\sum_{i=1}^{T-2}\sum_{j=i+1}^{T-1}E\left[\left(\sum_{t=i+1}^{T}\tilde{a}_{i,t-1}^2\left[(\hat{\boldsymbol{e}}_i'\hat{\boldsymbol{e}}_t)^2-\operatorname{tr}(\Omega_i\Omega_t)\right]\right)\left(\sum_{t=j+1}^{T}\tilde{a}_{j,t-1}^2\left[(\hat{\boldsymbol{e}}_j'\hat{\boldsymbol{e}}_t)^2-\operatorname{tr}(\Omega_j\Omega_t)\right]\right)\right]\right)\\
&=\frac{1}{N^2T^2}\left(\mathcal{O}(NT^2)+\mathcal{O}(NT^3)+\mathcal{O}(NT^4)\right)\\
&=\mathcal{O}\left(T^2/N\right). \qquad (35)
\end{aligned}
$$

Assuming weak cross-sectional dependence the order in N follows similarly to the derivation of the order $\operatorname{tr}(\Omega_i\Omega_t)=\mathcal{O}(N)$. Consequently, $\hat{\zeta}_3-\zeta_3=\mathcal{O}_p\left(\sqrt{\operatorname{Var}[\hat{\zeta}_3]}\right)=\mathcal{O}_p(T/\sqrt{N})$. Similar arguments apply for $\tilde{\zeta}_1$, ζ_2 and $\tilde{\zeta}_{52}$. Moreover, we obtain

$$\hat{\tilde{\zeta}}_{51}-\tilde{\zeta}_{51}=\frac{2}{NT}\sum_{i=1}^{T-2}\sum_{s=i+1}^{T-1}\sum_{t=s+1}^{T}\bar{a}_{i,t-1}\bar{a}_{i,s-1}\left(\operatorname{tr}(\Omega_i)^2-E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2]\right)=\mathcal{O}_p(1),$$

because we have $\left(\operatorname{tr}(\Omega_i)^2-E[(\boldsymbol{e}_i'\boldsymbol{e}_i)^2]\right)/N=\mathcal{O}(1)$ from the proof of Lemma 1. Combining these arguments, convergence of the remaining term $\hat{\tilde{\zeta}}_4-\tilde{\zeta}_4\xrightarrow{p}0$ follows directly. By implication, $(\hat{s}_{NT}^2-s_{NT}^2)=\mathcal{O}_p(T/\sqrt{N})+\mathcal{O}_p(1)$.

Proof of Proposition 1. Asymptotic normality of $\hat{\tau}$ stated in Proposition 1 follows from the asymptotic behaviour of $\hat{\nu}$ and $\hat{s}_{NT}^2$, Corollary 1 and a Taylor approximation of $\hat{\tau}$ in the true variance s_{NT}^2. Noticing that $\frac{1}{\sqrt{NT}}(\hat{\nu}_t-\nu_t)=o_p(1)$, we define the empirical version of the test statistic $\hat{\tau}$ as a function of $\hat{s}_{NT}^2$ as

$$\hat{\tau}=\tau(\hat{s}_{NT}^2)=\sum_{t=2}^{T}\frac{1}{\sqrt{NT}}(\tilde{\boldsymbol{y}}_{t-1}'\Delta\boldsymbol{y}_t^*-\nu_t)\Big/\hat{s}_{NT}.$$

Applying the first-order Taylor expansion in s_{NT}^2, asymptotic normality of $\hat{\tau}$ follows as:

$$\begin{aligned}
\hat{\tau} &= \sum_{t=2}^{T} \frac{1}{\sqrt{NT}} (\tilde{\boldsymbol{y}}_{t-1}' \Delta \boldsymbol{y}_t^* - \nu_t) \Big/ \hat{s}_{NT} = \tau(s_{NT}^2 + (\hat{s}_{NT}^2 - s_{NT}^2)) \\
&= \tau(s_{NT}^2) + (\hat{s}_{NT}^2 - s_{NT}^2) \cdot \frac{\partial \tau}{\partial (s_{NT}^2)} + o_p(1) \\
&= \tau(s_{NT}^2) + (\hat{s}_{NT}^2 - s_{NT}^2) \left(-\frac{1}{2} \sum_{t=2}^{T} \frac{1}{\sqrt{NT}} (\tilde{\boldsymbol{y}}_{t-1}' \Delta \boldsymbol{y}_t^* - \nu_t) \right) \cdot (s_{NT}^2)^{-3/2} + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0,1) + \left(\mathcal{O}_p \left(T/\sqrt{N} \right) + \mathcal{O}_p(1) \right) \mathcal{O}_p \left(\sqrt{T} \right) \left(\mathcal{O}(T) \right)^{-3/2} + o_p(1), \\
&\quad = \mathcal{N}(0,1) + o_p(1), \quad \text{for} \quad N, T \to \infty.
\end{aligned}$$

Convergence of the first term to the standard normal distribution is stated in Corollary 1, and, hence Proposition 1 follows.[8]

□

[8]To see that the remainder term is $o_p(1)$, consider, for instance, the expansion of second order $\left(\hat{s}_{NT}^2 - s_{NT}^2\right)^2 \cdot \frac{\partial \tau}{\partial^2 (s_{NT}^2)} = \left(\mathcal{O}_p \left(T/\sqrt{N} \right) + \mathcal{O}_p(1) \right)^2 \mathcal{O}_p(\sqrt{T})(\mathcal{O}(T))^{-5/2} = o_p(1)$ for $N, T \to \infty$.

References

Altinay, G., Karagol, E., 2004. Structural break, unit root, and the causality between energy consumption and GDP in Turkey. Energy Economics 26 (6), 985–994.

Anselin, L., 2013. Spatial econometrics: methods and models. Vol. 4. Springer Science & Business Media.

Bai, J., Ng, S., 2004. A panic attack on unit roots and cointegration. Econometrica 72 (4), 1127–1178.

Beenstock, M., Goldin, E., Nabot, D., April 1999. The demand for electricity in Israel. Energy Economics 21 (2), 168–183.

Billingsley, P., 1999. Convergence of probability measures, 2nd Edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, a Wiley-Interscience Publication.

Breitung, J., Das, S., 2005. Panel unit root tests under cross sectional dependence. Statistica Neerlandica 59 (4), 414–433.

Cavaliere, G., 2004. Unit root tests under time varying variances. Econometric Reviews 23 (4), 259–292.

Cavaliere, G., Taylor, A. M. R., 2007. Testing for unit roots in time series with non-stationary volatility. Journal of Econometrics 140 (2), 919–947.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press.

Demetrescu, M., Hanck, C., 2012a. A simple nonstationary-volatility robust panel unit root test. Economics Letters 117 (2), 10–13.

Demetrescu, M., Hanck, C., 2012b. Unit root testing in heteroscedastic panels using the Cauchy estimator. Journal of Business and Economic Statistics 30 (2), 256–264.

Demetrescu, M., Hanck, C., 2014. Robust inference for near-unit root processes with time-varying error variances. Econometric Reviews (forthcoming).

Dickey, D. A., Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Gadea-Rivas, M. D., Gómez-Loscos, A., Pérez-Quirós, G., 2014. The two greatest. Great recession vs. great moderation. Banco de Espana Working Papers 1423, Banco de Espana.

Glasure, Y. U., Lee, A.-R., 1998. Cointegration, error-correction, and the relationship between GDP and energy: The case of South Korea and Singapore. Resource and Energy Economics 20 (1), 17–25.

Hamori, S., Tokihisa, A., 1997. Testing for a unit root in the presence of a variance shift. Economics Letters 57 (3), 245–253.

Herwartz, H., Siedenburg, F., 2008. Homogenous panel unit root tests under cross sectional dependence: Finite sample modifications and the wild bootstrap. Computational Statistics and Data Analysis 53 (1), 137–150.

Herwartz, H., Siedenburg, F., Walle, Y. M., 2016. Heteroskedasticity robust panel unit root testing under variance breaks in pooled regressions. Econometric Reviews 35 (5), 727–750.

Homm, U., Breitung, J., 2012. Testing for speculative bubbles in stock markets: A comparison of alternative methods. Journal of Financial Econometrics 10 (1), 198–231.

Hsu, Y.-C., Lee, C.-C., Lee, C.-C., 2008. Revisited: Are shocks to energy consumption permanent or temporary? New evidence from a panel SURADF approach. Energy Economics 30 (5), 2314–2330.

Im, K. S., Pesaran, H. M., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. Journal of Econometrics 115 (1), 53–74.

Joyeux, R., Ripple, R. D., 2007. Household energy consumption versus income and relative standard of living: A panel approach. Energy Policy 35 (1), 50–60.

Kelejian, H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International economic review 40 (2), 509–533.

Kim, T. H., Leybourne, S., Newbold, P., 2002. Unit root tests with a break in innovation variance. Journal of Econometrics 109 (2), 365–387.

Levin, A., Lin, C. F., Chu, C. J., 2002. Unit root tests in panel data: asymptotic and finite-sample properties. Journal of Econometrics 108 (1), 1–24.

McAvinchey, I. D., Yannopoulos, A., 2003. Stationarity, structural change and specification in a demand system: the case of energy. Energy Economics 25 (1), 65–92.

Moon, H. R., Perron, B., 2004. Testing for a unit root in panels with dynamic factors. Journal of Econometrics 122 (1), 81–126.

Narayan, P. K., Smyth, R., 2007. Are shocks to energy consumption permanent or temporary? Evidence from 182 countries. Energy Policy 35 (1), 333–341.

Pesaran, M. H., 2007. A simple panel unit root test in the presence of cross section dependence. Journal of Applied Econometrics 22 (2), 265–312.

Sensier, M., van Dijk, D., 2004. Testing for volatility changes in u.s. macroeconomic time series. Review of Economics and Statistics 86 (3), 833–839.

Stock, J. H., Watson, M. W., 2003. Has the business cycle changed and why? NBER Macroeconomic Annual 17, 159–230.

Westerlund, J., 2014. Heteroscedasticity robust panel unit root tests. Journal of Business & Economic Statistics 32 (1), 112–135.

Westerlund, J., 2015. The effect of recursive detrending on panel unit root tests. Journal of Econometrics 185 (2), 453–467.