Herwartz, Helmut; Siedenburg, Florian

Working Paper

The effects of variance breaks on homogenous panel unit root tests

Economics working paper / Christian-Albrechts-Universität Kiel, Department of Economics, No. 2009,07

Provided in cooperation with:

Christian-Albrechts-Universität Kiel (CAU)

C|A|U
Christian-Albrechts-Universität zu Kiel
Department of Economics
Economics Working Paper
No 2009-07

the effects of variance breaks on homogenous panel unit root tests

by Helmut Herwartz and Florian Siedenburg

The effects of variance breaks on homogenous panel unit root tests

Helmut Herwartz and Florian Siedenburg

Christian-Albrechts-University of Kiel

This draft: Juli 2009

Abstract Noting that many economic variables display occasional shifts in their second order moments, we investigate the performance of homogenous panel unit root tests in the presence of permanent volatility shifts. It is shown that in this case, panel unit root tests derived under time invariant innovation variances lose control over actual significance levels while the test proposed by Herwartz and Siedenburg (2008) retains size control. A simulation study of the finite sample properties confirms the theoretical results in finite samples. As an empirical illustration, we reassess evidence on the Fisher hypothesis.

JEL Classification: C23, C12, E40

Keywords: Panel unit root tests, variance breaks, cross sectional dependence, Fisher hypothesis.

Corresponding author: F. Siedenburg, Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D–24118 Kiel, Germany.

Email: fsiedenb@economics.uni-kiel.de. Phone: +49-431-8802225.

1 Introduction

Panel unit root tests (PURTs) have become a standard tool in macroeconometric applications. Making use of the cross sectional dimension allows to overcome power deficiencies of univariate unit root tests and helps to avoid the multiple testing problem. Moreover, a number of macroeconomic models postulate stationarity of some key variables. For instance, the purchasing power parity hypothesis implies stationarity of real exchange rates (see Taylor and Taylor, 2004 for a survey) or the Fisher hypothesis, which predicts real interest rates to be stationary (e.g. Herwartz and Reimers, 2006, 2009). First generation PURTs (e.g. Levin et al., 2002 or Im et al., 2003) rely on the assumption of cross sectionally independent error terms. Since the work of O'Connell (1998), however, it is widely recognized that a violation of this assumption leads to severe size distortions of first generation tests and, therefore, second generation tests relying on less restrictive assumptions have been suggested (see Hurlin and Mignon, 2007 and Breitung and Pesaran, 2008 for recent surveys). Two general directions of coping with the nuisance parameters invoked by the cross sectional dependence can be identified. On the one hand, approaches presuming a common factor structure for the error terms and, on the other hand, tests building on robust covariance estimators.

Second order invariance of model disturbances is an additional implicit assumption of PURTs. However, this assumption is quite restrictive, as many macroeconomic and financial variables are characterized by structural shifts in their unconditional volatility. In fact, what has become known as the 'Great Moderation' is a substantial decline in numerous macroeconomic key variables' volatility across the G7 economies since the mid 1980s (see, for instance, Kim and Nelson, 1999, McConnell and Perez-Quiroz, 2000 and Stock and Watson, 2003). The adverse effects of variance shifts on unit root tests for single time series have been studied by, among others, Hamori and Tokihisa (1997), Kim et al. (2002), Cavaliere (2004), and Cavaliere and Taylor (2007a,b, 2008). The main findings are that the (augmented) Dickey-Fuller (Dickey and Fuller, 1979, (A)DF henceforth) and other unit root tests asymptotically depend on nuisance parameters in the presence of permanent variance shifts. Hence, seriously distorted empirical type one errors and deceptive inference are the consequences of a violation of the implicit assumption of time invariant volatility. The magnitude of size distortions is shown to depend on specific break patterns. Generally, largest (positive) size distortions are observed for early negative and late positive shifts in the level of the process' unconditional variance. So far, only Hanck (2009b) attempts to generalize these results to the field of panel unit root testing. However, while he considers intersection tests for heterogenous panels which are constructed by combining the p-values obtained from volatility break robust univariate tests, this paper concentrates on the class of class of homogenous PURTs based on a pooled DF regression. We show that the second generation 'White-type' corrected PURT proposed in Herwartz and Siedenburg (2008) retains a Gaussian limiting distribution under discrete shifts of the innovation variance. In contrast, the first generation test of Levin et al. (2002) and the second generation test of Breitung and Das (2005) do not converge to a nuisance free limiting distribution in this case. Moreover, the local asymptotic power function of the test statistic is derived. It turns out that in absence of volatility breaks, its local asymptotic power equals those of the statistic proposed by Breitung and Das (2005), while in the presence of a volatility break, the power of depends on the timing and direction of the break. Deterministic terms and residual serial correlation are accounted for by detrending and prewhitening schemes proposed in Breitung (2000) and Breitung and Das (2005), respectively. While the prewhitening scheme works well even under second order moment instability, the detrending scheme invokes serious deviations of empirical type one errors from the nominal significance level if there is a break in the innovation variance.

As an illustrative example, we reconsider PURT based evidence on the Fisher hypothesis in Crowder (2003). Postulating a one-to-one comovement of nominal interest rates and expected rates of inflation, the Fisher hypothesis implies stationary real interest rates. The considered cross section of 9 developed economies over the period 1961Q2-2007Q2 mirrors core issues discussed in this paper, such as shifts in unconditional volatility and cross sectional dependence.

The remainder of the paper is organized as follows. A brief review of the effects of nonstationary volatility on univariate unit root tests is given in the next section. The panel model is introduced and asymptotic results for the considered PURTs are derived in Section 3. Section 4 provides the results of a Monte Carlo simulation study. The empirical illustration is presented in Section 5 and Section 6 concludes. Formal proofs are contained in the Appendix.

2 Effects of nonstationary volatility on univariate unit root tests

The effects of nonstationary volatility on unit root tests in the univariate case have been investigated by Hamori and Tokihisa (1997), Kim et al. (2002), Cavaliere (2004) and Cavaliere and Taylor (2007a,b, 2008). For illustrative purposes, we review the results of Hamori and Tokihisa (1997) who consider the most basic example of a single upward shift in the innovation variance of an autoregressive process of order one (AR(1)) without any deterministic terms. In particular, consider the following data generating process (DGP)

$$y_t = \rho y_{t-1} + e_t, \quad t = 1, ..., T. \tag{1}$$

In (1), the variance shift is modeled by means of the composite error term e_t, i.e.

$$e_t = \epsilon_t + \eta_t DU_t, \quad \epsilon_t \sim iid(0, \sigma_1^2), \quad \eta_t \sim iid(0, \sigma_2^2)$$
$$\text{and } DU_t = \begin{cases} 1, & \text{if } t > T_B, \ (1 < T_B < T) \\ 0, & \text{otherwise.} \end{cases}$$

Let $\lambda = T_B/T$ denote the ratio of pre-break to total sample period and $W(r)$ is a standard Brownian motion defined on $r \in [0, 1]$, then, as $T \to \infty$, the asymptotic distribution of the DF t-ratio of $\hat{\rho}-1$ is given by

$$t_{DF} \quad \xrightarrow{d} \quad \frac{\frac{1}{2}[\{W(1)\}^2-1]}{\sqrt{\int_0^1 W(r)dr^2 - \Xi\left[(1-\lambda)\int_0^\lambda W(r)dr + \lambda\int_\lambda^1\left(\frac{1-r}{r}\right)W(r)dr\right]}}, \qquad (2)$$

$$\text{where} \quad \Xi = \frac{(\sigma_2/\sigma_1)^2}{1+(\sigma_2/\sigma_1)^2(1-\lambda)}.$$

It is easy to verify that the nuisance parameters in the limiting distribution depend on the strength and the timing of the variance break. The standard DF case is covered by $\sigma_2 = 0$ and $\lambda = 0$ or $\lambda = 1$. Hamori and Tokihisa (1997) provide simulation evidence suggesting that a late positive variance shift leads to the largest (upward) bias of empirical type one errors. Kim et al. (2002) generalize the previous result to models with deterministic terms and propose a pivotal test for the unit root null hypothesis based on prior break date estimation. In a series of papers, Cavaliere (2004) and Cavaliere and Taylor (2007a,b, 2008) extend these results in three directions. First, they allow for a wider class of volatility processes, including multiple breaks and trending volatility. Second, they extend the analysis to the class of M-type of unit root tests proposed by Perron and Ng (1996), Stock (1999) and Ng and Perron (2001). Finally, they propose alternative volatility-break robust test procedures, such as a test based on the estimated variance profile, as well as tests based on simulation or resampling methods.

3 PURTs under nonstationary volatility

3.1 The autoregressive, heteroskedastic panel model

In the following, we study the effects of nonstationary volatility on homogenous PURTs. More specifically, the limiting distributions of alternative t-statistics obtained from pooled DF regressions are derived for a panel AR(1) model allowing for multiple and possibly heterogeneous breaks in the innovation variance as well as for weak cross sectional dependence. Weak cross sectional dependence as defined by Breitung and Pesaran (2008) is characterized by bounded eigenvalues of the covariance matrix as $N \to \infty$. This type of dependence includes, for instance, covariance matrices implied by all types of spatial panel models (Elhorst, 2003) but excludes dependence invoked through common factor models. The empirically relevant treatment of deterministic terms and serially correlated disturbances is discussed later.

The heteroskedastic panel model is given by

$$\boldsymbol{y}_t = \rho \boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \quad t = 1, ..., T, \tag{3}$$

where $\boldsymbol{y}_t = (y_{1t}, ..., y_{Nt})'$, $\boldsymbol{y}_{t-1} = (y_{1,t-1}, ..., y_{N,t-1})'$ and $\boldsymbol{e}_t = (e_{1t}, ..., e_{Nt})'$ are $N \times 1$ vectors and the index $i = 1, ..., N$ indicates the cross sectional units. The autoregressive coefficient ρ satisfies either $\rho = 1$ under the unit root null hypothesis or $|\rho| < 1$ under the stationary alternative hypothesis. The assumption of a homogenous AR coefficient under the alternative hypothesis could be relaxed to the case of different stationary coefficients for all cross sectional units i without loss of generality. In fact, Breitung and Pesaran (2008) point out that the power of both pooled and averaged (as for example the test proposed by Im et al., 2003) PURTs only depends on the average of the individual specific autoregressive coefficients. Hence, pooled PURTs are also powerful against the heterogeneous alternative where mean reverting behavior holds only for some nonzero fraction of the cross sectional units.

We make the following set of assumptions regarding the vector of errors $\boldsymbol{e}_t$:

Assumption 1 ($\mathcal{A}_1$)
(i) The error vector $\boldsymbol{e}_t \sim iid\,(\mathbf{0}, \Omega_t)$.
(ii) Ω_t *is a positive definite matrix with eigenvalues* $\lambda_1 \geq ... \geq \lambda_N$ *and* $\lambda_1 < \bar{c} < \infty$ *for all* t.
(iii) Finally, it is assumed that $E[e_{it}e_{jt}e_{kt}e_{lt}] < \infty$ *for all* i, j, k, l.

The assumptions $\mathcal{A}_1(i)$-$\mathcal{A}_1(iii)$ are basically the same as in Breitung and Das (2005) except that we allow for a time varying covariance matrix, Ω_t. $\mathcal{A}_1(i)$ rules out higher order serial correlation which will be considered later. $\mathcal{A}_1(ii)$ restricts the pattern of cross sectional dependence to the weak type dependence while the assumed existence of finite fourth order moments of e_{it} in $\mathcal{A}_1(iii)$ is a standard assumption in the (panel) unit root literature. Additionally, we make the following assumptions on Ω_t which further define the types of volatility breaks considered in this paper.

Assumption 2 ($\mathcal{A}_2$)
(i) $\Omega_t = \Omega_1$ *for* $t = 1, ...T_1$ *and* $\Omega_t = \Omega_2$ *for* $t = T + 1, ...T$.
(ii) Moreover, $T_1, T_2 \to \infty$ *as* $T \to \infty$ *with* $T_2 = T - T_1$ *and* $\frac{T_1}{T} \to \delta > 0$, $\frac{T_2}{T} \to 1 - \delta > 0$.
(iii) Finally, $\Omega_t = \Phi_t^{1/2} \Psi \Phi_t^{1/2}$ *for all* t*, with* $\Phi_t = diag(\sigma_{1t}^2, ..., \sigma_{Nt}^2)$.

Assumption $\mathcal{A}_2(i)$ restricts the number of variance break points to one. This assumption is made to simplify the analytical derivations. From the proofs in the Appendix, however, it will become clear that it is straightforward to modify the analytical derivations to account for multiple breaks. $\mathcal{A}_2(ii)$ requires that the pre- and post-break sample increase as $T \to \infty$, with the subsamples being some constant fractions δ and $(1 - \delta)$ of the total sample, respectively. This assumption is important in the derivation of the limiting distribution as it ensures convergence of partial sum processes to functionals of Brownian motions in each subsample. Assumption $\mathcal{A}_2(iii)$ defines the type of variance break considered. The decomposition $\Omega_t = \Phi_t^{1/2} \Psi \Phi_t^{1/2}$, where Ψ is the (time invariant) correlation matrix implied by Ω_t and $\Phi_t^{1/2}$ is a diagonal matrix of the idiosyncratic standard deviations, allows to separate the issues of cross sectional dependence and variance breaks. It further incorporates heterogeneity along the cross sectional dimension as idiosyncratic variances σ_{it}^2 and the strength of the variance breaks may differ. Obviously, $\mathcal{A}_2(iii)$ also covers the case where only a fraction of the cross sectional units feature a shift in the innovation variance.

3.2 Asymptotic size distortions of homogenous PURTs

Consider the AR(1) panel model defined in (3). The unit root null hypothesis, $H_0 : \rho = 1$, can be tested by means of the OLS t-ratio of $\hat{\phi}$ from the pooled DF regression

$$\Delta \boldsymbol{y}_t = \phi \boldsymbol{y}_{t-1} + \boldsymbol{e}_t,$$

with $\Delta \boldsymbol{y}_t = (y_{1t} - y_{1,t-1}, ..., y_{Nt} - y_{N,t-1})'$. The test statistic is

$$t_{OLS} = \frac{\sum_{t=1}^{T} \boldsymbol{y}'_{t-1} \Delta \boldsymbol{y}_t}{\sqrt{\sigma_e^2 \sum_{t=1}^{T} \boldsymbol{y}'_{t-1} \boldsymbol{y}_{t-1}}}, \tag{4}$$

where σ_e^2 is replaced by $\hat{\sigma}_e^2 = (NT)^{-1} \sum_{t=1}^{T} (\Delta \boldsymbol{y}_t - \widehat{\phi} \boldsymbol{y}_{t-1})' (\Delta \boldsymbol{y}_t - \widehat{\phi} \boldsymbol{y}_{t-1})$. The results in Levin et al. (2002) imply that under H_0 in (3) with cross sectionally independent and homoskedastic error terms with constant variance ($\Omega_t = \Omega = I_N \sigma_e^2$), t_{OLS} is asymptotically Gaussian as T, $N \to \infty$. Violations of the assumption of cross sectional independence can be overcome along the lines of Breitung and Das (2005) or Herwartz and Siedenburg (2008), either by means of robust covariance estimation or resampling methods. The effects of a break in the innovation variance on homogenous PURTs have not yet been studied. In the following, it is shown that in analogy to the univariate case, t_{OLS} does not converge to a nuisance free limiting distribution and, hence, loses control over the asymptotic size of the test.

Proposition 1 *Assume the panel DGP is given by (3) and assumptions* $(\mathcal{A}_1)$ *and* $(\mathcal{A}_2)$ *with* $\Psi = I_N$ *and* $\sigma_{i\bullet}^2 = \sigma_{e\bullet}^2 \, \forall \, i$ *and* $\bullet = 1, 2$. *Then, under* $H_0 : \rho = 1$ *and for* $T \to \infty$ *followed by* $N \to \infty$, $t_{OLS} \xrightarrow{d} N(0, \overline{\nu}_{OLS})$, $\overline{\nu}_{OLS} \neq 1$ *if* $\sigma_{e1}^2 \neq \sigma_{e2}^2$.

The proof of Proposition 1 is deferred to Section A.1 in the Appendix. The result directly shows that discrete shifts in the innovation variance induce nuisance parameters in the asymptotic distribution of the t_{OLS} PURT statistic. Moreover, given the specific form of $\overline{\nu}_{OLS}$ derived in the Appendix, it is clear that the direction and strength of the implied size distortion depend on the specification of the break. In particular, we have

$$\overline{\nu}_{OLS} = \frac{0.5\delta^2 \overline{\lambda_1^2} + \delta(1-\delta)\overline{\lambda_1 \lambda_2} + 0.5(1-\delta)^2 \overline{\lambda_2^2}}{\left(\delta \overline{\lambda_1} + (1-\delta)\overline{\lambda_2}\right) \left[0.5\delta^2 \overline{\lambda_1} + \delta(1-\delta)\overline{\lambda_1} + 0.5(1-\delta)^2 \overline{\lambda_2}\right]},$$

with $\overline{\lambda_\bullet} = N^{-1} \sum_{i=1}^{N} \lambda_{i\bullet}$, $\overline{\lambda_\bullet^2} = N^{-1} \sum_{i=1}^{N} \lambda_{i\bullet}^2$, where $\bullet = 1, 2$ refers to the pre- and post-break period, respectively and $\overline{\lambda_1 \lambda_2} = N^{-1} \sum_{i=1}^{N} \lambda_{i1} \lambda_{i2}$. To illustrate the size distortion invoked by variance breaks, Figure 1 depicts the asymptotic variance $\overline{\nu}_{OLS}$ for a continuity of breakpoints $\delta \in [0, 1]$ with

$$\overline{\lambda_1} = 1 \text{ and } \overline{\lambda_2} \in \{0.2, 0.33, 0.5, 0.66, 0.9, 1.1, 1.33, 1.5, 1.66, 1.8\}.$$

Figure 1 reveals that largest deviations of $\overline{\nu}_{OlS}$ from unity are characteristic for late positive and early negative variance breaks. This result is in line with findings for the time series case, where early negative and late positive variance shifts have been found to induce largest size distortions. However, in the time series case, both scenarios induce an upward size distortion, while the simulated values of $\overline{\nu}_{OLS}$ imply a downward size distortion in the case of a negative variance break. This is easily seen by noting that $\overline{\nu}_{OLS} < 1$ corresponds to less probability mass in the tails of the asymptotic distribution of t_{OLS} compared with the Gaussian distribution.

So far, results are derived under cross sectional independence and homoskedasticity. However, asymptotic size distortions carry over to the cross sectional dependence robust statistic t_{Rob} suggested by Breitung and Das (2005). Under weak form cross sectional dependence with a covariance structure characterized by bounded eigenvalues as $N \to \infty$ and time invariant innovation variance, the statistic retains a Gaussian limiting distribution by applying *panel corrected standard errors* (Beck and Katz, 1995). It is given as

$$t_{Rob} = \frac{\sum_{t=1}^{T} \boldsymbol{y}'_{t-1} \Delta \boldsymbol{y}_t}{\sqrt{\sum_{t=1}^{T} \boldsymbol{y}'_{t-1} \hat{\Omega} \boldsymbol{y}_{t-1}}}, \quad \text{with} \quad \hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{e}}_t \hat{\boldsymbol{e}}'_t. \tag{5}$$

Proposition 2 *Assume the panel DGP is given by (3) and assumptions* $(\mathcal{A}_1)$ *and* $(\mathcal{A}_2)$*. Then, under* $H_0 : \rho = 1$ *and for* $T \to \infty$ *followed by* $N \to \infty$*,* $t_{Rob} \xrightarrow{d} N(0, \overline{\nu}_{Rob})$*,* $\overline{\nu}_{Rob} \neq 1$ *if* $\sigma_{e1}^2 \neq \sigma_{e2}^2$*.*

The proof of Proposition 2 is given in Section A.2 of the Appendix. It turns out that

$$\overline{\nu}_{Rob} = \frac{0.5\delta^2 \overline{\lambda_1^2} + \delta(1-\delta)\overline{\lambda_1 \lambda_2} + 0.5(1-\delta)^2 \overline{\lambda_2^2}}{\left(\delta^2 - 0.5\delta^3\right) \overline{\lambda_1^2} + \left(1.5\delta + 0.5\delta^2 - 2\delta^3\right) \overline{\lambda_1 \lambda_2} + 0.5(1-\delta)^3 \overline{\lambda_2^2}}.$$

It is easy to verify that in absence of volatility breaks the results in Breitung and Das (2005) obtain as a special case with $\delta = 0$ or $\delta = 1$.

3.3 A volatility-break robust test

Herwartz and Siedenburg (2008) propose a test statistic, which is based on a 'White-type' covariance estimator, making use of residuals obtained under H_0. The test statistic and its asymptotic distribution are

$$t_{HS} = \frac{\sum_{t=1}^{T} \boldsymbol{y}'_{t-1}\Delta \boldsymbol{y}_t}{\sqrt{\sum_{t=1}^{T} \boldsymbol{y}'_{t-1}\check{\boldsymbol{e}}_t\check{\boldsymbol{e}}'_t\boldsymbol{y}_{t-1}}} \xrightarrow{d} N(0,1), \quad \check{\boldsymbol{e}}_t = \Delta \boldsymbol{y}_t = \boldsymbol{e}_t. \tag{6}$$

The statistic was originally proposed as an alternative to t_{Rob} in finite samples where the number of cross sectional units is relatively large compared to the time dimension. However, given the construction of the employed covariance estimator, Herwartz and Siedenburg (2008) conjecture that t_{HS} might be robust with respect to unknown patterns of (nonstationary) heteroskedasticity. Similarly, Hamori and Tokihisa (1997) suggest the White correction (with unrestricted residuals, however) as a potential means to appropriately cope with the nuisance invoked by a variance shift. The following Proposition states asymptotic Gaussianity of the statistic t_{HS} under a volatility break as defined by ($\mathcal{A}_2$).

Proposition 3 *Assume the DGP is given by (3) and Assumptions $\mathcal{A}_1$ and $\mathcal{A}_2$ hold and $\sigma^2_{e1} \neq \sigma^2_{e2}$. Then under $H_0 : \rho = 1$ and for $T \to \infty$ followed by $N \to \infty$, $t_{HS} \xrightarrow{d} N(0,1)$.*

The proof of Proposition 3 is derived in Section A.3 in the Appendix.

Even though the proof is laid out for a single break date, it is straightforward to extend it to scenarios of multiple break dates. A caveat of the asymptotic results is that they are obtained under sequential asymptotics. As it is shown in Phillips and Moon (1999), sequential asymptotics do not necessarily imply convergence if N and T approach infinity jointly. However, results in Breitung and Westerlund (2009) conjecture that the previous results might also apply if $\sqrt{N}/T \to 0$ as $T, N \to \infty$ jointly.

3.4 Local asymptotic power of t_{HS}

To verify that the test based on t_{HS} has asymptotic power in local-to-unity neighborhoods, the following Proposition states its asymptotic distribution under a sequence of local alternatives given by

$$H_l : \rho = 1 - \frac{c}{T\sqrt{N}}. \tag{7}$$

Proposition 4 *Under the sequence of local alternatives defined in (7), for $T \to \infty$ followed by $N \to \infty$, t_{HS} is asymptotically distributed as $N(-c\mu_l, 1)$, where*

$$\mu_l = \frac{0.5\delta^2\overline{\lambda_1} + \delta(1-\delta)\overline{\lambda_1} + 0.5(1-\delta)^2\overline{\lambda_2}}{\sqrt{0.5\delta^2\overline{\lambda_1^2} + \delta(1-\delta)\overline{\lambda_1\lambda_2} + 0.5(1-\delta)^2\overline{\lambda_2^2}}}.$$

The proof of Proposition 4 is deferred to Section A.4 in the Appendix. The result directly implies asymptotic power of the test in local-to-unity neighborhoods of order $O\left(T^{-1}N^{-1/2}\right)$ for models without individual time trends. Moreover, it is easy to see that in the case of time invariant volatility with $\delta = 1$, $\mu_l = \sqrt{0.5}\,\overline{\lambda_1}/\sqrt{\overline{\lambda_1^2}}$, implying the same local asymptotic power as obtained by Breitung and Das (2005) for the t_{Rob} statistic. Finally, a more detailed investigation of μ_l reveals that a downward (upward) shift of the innovation variance leads to asymptotically higher (lower) local power compared with the benchmark case of constant volatility.

3.5 Deterministic terms and serial correlation

In the following, we discuss data transformations suggested in Breitung and Meyer (1994), Breitung (2000) and Breitung and Das (2005) to cope with deterministic terms and residual serial correlation. In contrast to OLS-detrending and lag augmentation, these approaches allow to construct asymptotically pivotal test statistics without the necessity of applying bias correction terms.

3.5.1 Deterministic terms

If the DGP contains (cross section specific) deterministic intercepts or trends, the pooled regression in (3) is inappropriate. However, in contrast to the time series case, inclusion of deterministic components in the test regression invokes the so-called Nickell-bias (Nickell, 1981), present in dynamic panels with individual specific intercepts or trends. While the Nickell-bias can be accounted for by bias adjustment terms (as e.g. in Levin et al., 2002), it can be shown that elimination of individual intercepts by means of least square projections substantially reduces the power of the tests.

Consider the case of distinguishing a driftless random walk from a stationary process with individual specific intercept terms. The DGP can then be written as

$$\boldsymbol{y}_t = (1-\rho)\boldsymbol{\mu} + \rho\boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \tag{8}$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)'$ collects individual specific intercepts. Breitung and Meyer (1994) were the first to point out that the intercept terms can be efficiently removed by subtracting the first observation from the data as $\boldsymbol{y}_0$ is the best estimator of $\boldsymbol{\mu}$ under H_0. Hence, the pooled test regression is based on the transformed data

$$\Delta\boldsymbol{y}_t = \phi\boldsymbol{y}^*_{t-1} + \boldsymbol{e}_t, \quad \text{with} \quad \boldsymbol{y}^*_{t-1} = \boldsymbol{y}_{t-1} - \boldsymbol{y}_0.$$

Breitung and Meyer (1994) illustrate that the power of tests based on a regression with the transformed data does not depend on the individual effects and is hence superior to the power of tests based on least square demeaned data.

If the test is performed to discriminate a random walk with drift from a trend stationary process, the underlying DGP may be written as

$$\boldsymbol{y}_t = \boldsymbol{\mu} + (1-\rho)\boldsymbol{\beta}t + \rho\boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \tag{9}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_N)'$ is the vector of individual trend parameters. Moon et al. (2007) show that in this case of incidental trends, pooled t-ratio type tests only have asymptotic power in local neighborhoods shrinking at the rate of $O\left(\frac{1}{N^{1/4}T}\right)$. To obtain a test statistic which does not require bias correction terms, Breitung (2000) suggests the Helmert transformation to center the first differences of the data in a forward looking manner, i.e.

$$\begin{aligned} \Delta\boldsymbol{y}^*_t &= s_t\left[\Delta\boldsymbol{y}_t - \frac{1}{T-t}(\Delta\boldsymbol{y}_{t+1} + ... + \Delta\boldsymbol{y}_T)\right], \text{ and} \\ & s_t^2 = (T-t)/(T-t+1). \end{aligned} \tag{10}$$

Detrending of the test regression's right hand side variable proceeds as

$$\boldsymbol{y}^*_t = \boldsymbol{y}_t - \boldsymbol{y}_0 - \widehat{\boldsymbol{\beta}}_t = \boldsymbol{y}_t - \boldsymbol{y}_0 - \frac{\boldsymbol{y}_T - \boldsymbol{y}_0}{T}t. \tag{11}$$

Breitung (2000) demonstrates that the detrending in (10) and (11) is sufficient to remove the Nickell-bias and, hence, asymptotically pivotal test statistics can be obtained by running a pooled regression on the transformed data. This result, however, relies on $\Delta \boldsymbol{y}_t^*$ being white noise with constant variance. As this assumption is violated in our setting, it is unclear if the proposed detrending scheme still yields pivotal PURT statistics. We address this issue by means of the Monte Carlo study in Section 4.

3.5.2 Short run dynamics

For the case of serially correlated error terms, Breitung and Das (2005) prove that the pooled statistic in (5) remains asymptotically Gaussian if it is computed from prewhitened data. In contrast, ADF-type lag augmentation is shown to be insufficient to remove the effects of the short run dynamics if deterministic terms are present in the model. Prewhitening proceeds by running individual specific, ADF regressions under H_0, i.e.

$$\Delta y_{it} = \sum_{j=1}^{p_i} c_{ij} \Delta y_{i,t-j} + e_{it}. \qquad (12)$$

The estimates $\widehat{\boldsymbol{c}}_i = (\widehat{c}_{i1}, ..., \widehat{c}_{ip})$ are then used to obtain prewhitened data as

$$y_{it}^* = y_{it} - \widehat{c}_{i1} y_{i,t-1} - ... - \widehat{c}_{ip_i} y_{i,t-p_i} \qquad (13)$$

and

$$\Delta y_{it}^* = \Delta y_{it} - \widehat{c}_{i1} \Delta y_{i,t-1} - ... - \widehat{c}_{ip_i} \Delta y_{i,t-p_i}. \qquad (14)$$

The choice of lag lengths p_i can be based on any consistent lag-length selection criterion. Since the estimates $\widehat{c}_{ip}$ are $\sqrt{T}$ consistent, one would expect a somewhat larger time dimension for obtaining correctly sized tests than in the case of serially uncorrelated increments. If the DGP features both, short run dynamics and deterministic patterns, the data is first prewhitened and subsequently detrended as discussed in Section 3.5.1. Since the prewhitening regression is performed under H_0, an intercept term has to be included only if the model includes linear time trends under the alternative hypothesis.

4 Monte Carlo study

4.1 The simulation design

To illustrate the finite sample effects of volatility breaks on the considered homogenous PURTs, we consider three stylized scenarios:

$$
\begin{aligned}
\text{DGP1:}\quad \boldsymbol{y}_t =&\quad (1-\rho)\boldsymbol{\mu} + \rho\boldsymbol{y}_{t-1} + \boldsymbol{e}_t, \quad t = -50, ..., T,\\
\text{DGP2:}\quad \boldsymbol{y}_t =&\quad \boldsymbol{\mu} + (1-\rho)\boldsymbol{\beta}t + \rho\boldsymbol{y}_{t-1} + \boldsymbol{e}_t,\\
\text{DGP3:}\quad \boldsymbol{y}_t =&\quad (1-\rho)\boldsymbol{\mu} + \rho\boldsymbol{y}_{t-1} + \boldsymbol{u}_t, \quad \boldsymbol{u}_t = \boldsymbol{c} \circ \boldsymbol{u}_{t-1} + \boldsymbol{e}_t,
\end{aligned}
$$

where bold entries indicate vectors of dimension $N \times 1$ and $\circ$ denotes the Hadamard product. The first two DGPs formalize AR(1) models with serially uncorrelated errors, whereas the last one introduces AR(1) disturbances. DGPs 1 and 3 formalize the panel unit root against a panel stationary process with individual effects, while DGP 2 models a panel random walk with drift under H_0 or a panel of trend stationary processes with individual effects under the alternative. Rejection frequencies under H_0 are computed with $\rho = 1$ whereas empirical (size adjusted) power is calculated against the homogeneous alternatives $\rho = 1 - \frac{5}{T\sqrt{N}}$ or $\rho = 1 - \frac{5}{TN^{1/4}}$ for the cases featuring individual intercepts or trends, respectively. As mentioned in Section 3.1 that homogenous PURTs have power against heterogenous alternatives, it is important to note that the choice of a homogenous alternative is without loss of generality. Following Pesaran (2007), the deterministic terms are parameterized such that the processes display the same average trend properties under H_0 and the alternative hypothesis. In particular, $\boldsymbol{\mu} \sim iidU(0, 0.02)$, and $\boldsymbol{\beta} \sim iidU(0, 0.02)$. The parametrization of the short run dynamics in DGP 3 is also taken from Pesaran (2007), i.e. $\boldsymbol{c} \sim iidU(0.2, 0.4)$.

Six distinct scenarios for the covariance matrix Ω_t are simulated for each DGP. With regard to contemporaneous correlation, cases of cross sectionally independent, as well as of (weakly) contemporaneously correlated panels are considered. Three different scenarios are simulated with respect to volatility breaks: constant volatility as well as a late positive and an early negative variance shift. Cross sectionally uncorrelated data is generated by setting $\Psi = I_N$ and $\Phi_t = diag(\sigma_{et}^2)$. As demonstrated in Sections 3.2 and 3.3, the choice of cross sectionally homogenous variances is without loss of generality for the t_{Rob} and t_{HS} statistics but necessary to obtain asymptotic Gaussianity of t_{OLS} in the benchmark case of constant volatility. For the case of a contemporaneously correlated panel, a spatial autoregressive (SAR) error structure is presumed. The latter is specified as

$$\boldsymbol{e}_t = (I_N - \Theta W)^{-1}\boldsymbol{\varepsilon}_t, \quad \text{with} \quad \Theta = 0.8 \quad \text{and} \quad \boldsymbol{\varepsilon}_t \sim iidN(\mathbf{0}, diag(\sigma_{et}^2)),$$

where the so-called spatial weights matrix W is a row normalized symmetric contiguity matrix of the one-behind-one-ahead type (for more details on spatial panel models see e.g. Elhorst, 2003). In the following, we refer to this specification as an SAR(1) model. The resulting covariance matrix of $\boldsymbol{e}_t$ is given by $\Omega_t = \sigma_{et}^2(B'B)^{-1}$ with $B = (I_N - \Theta W)$. As mentioned above, three distinct variance patterns are simulated. Let $\sigma_{e\lfloor sT \rfloor} = \sigma_{e1}\mathbb{I}(s \leq s_B) + \sigma_{e2}\mathbb{I}(s > s_B)$, where $s_B \in [0, 1]$ indicates the timing of the variance break, $\lfloor sT \rfloor$ denotes the integer part of sT and $\mathbb{I}$ is the indicator function. In the homoskedastic case, we set $\sigma_{et} = \sigma_{e1}$, with $\sigma_{e1} = 1$. The break scenarios are taken from Cavaliere and Taylor (2007b) and are parameterized as $s_B = 0.2$ and $\sigma_{e2} = 1/3$ for the early negative break, while the late positive break is given by $s_B = 0.8$ and $\sigma_{e2} = 3$.

Data is generated for all combinations of $N \in [10, 50]$ and $T \in [10, 50, 100, 250]$. To ensure convergence of the process to its unconditional mean under the alternative hypothesis, 50 presample values are generated and discarded throughout. To compute empirical rejection probabilities under H_0, we calculate each PURT statistic for the appropriately transformed data and compare the resulting statistics with the 5% critical value of the Gaussian distribution. Reported estimates for local power are adjusted such that empirical type one errors equal 5%. Throughout, we use 5000 replications.

4.2 Results

Table 1 documents empirical rejection frequencies obtained for DGP1. The left hand side of Table 1 documents results obtained under cross sectional independence while entries on the right hand side refer to results obtained under a SAR(1) error model. Rejection frequencies under H_0 are reported to the left of size adjusted local power estimates in both cases.

The first block in the upper left panel corresponds to the benchmark case of cross sectional independence and time invariant innovation variances. In this setting, all employed statistics have a Gaussian limiting distribution and, hence, should display empirical rejection frequencies close to 5% as T and N become large. However, the documented results reflect some evidence of small sample size distortions. Empirical rejection frequencies obtained by t_{OLS} range around 7% for panels with $N = 10$, whereas application of t_{Rob} leads to undersizing for small values of T. Results obtained for the 'White-type' statistic t_{HS} display comparatively small deviations from the nominal level, especially if $N = 50$. Size adjusted local power estimates indicate that under full homogeneity, all three statistics are asymptotically equally powerful and that the chosen sample sizes are too small for local power estimates to fully converge. The right hand side of the first block presents results for the SAR(1) error model with constant volatility. While the OLS test is severely oversized in this instance, both robust tests remain asymptotically Gaussian. However, finite sample distortions observed for t_{HS} are slightly larger while the undersizing of t_{Rob} is less pronounced than in the case of cross sectional independence. Local power results show that all considered tests are less powerful if the data is cross sectionally correlated. This finding might be explained by noting that cross sectional correlation reduces the amount of independent information contained in the data (Hanck, 2009a).

In line with the theoretical results in Section 3.2, results obtained under an early negative variance break and cross sectional independence indicate a tendency of undersizing for t_{OLS} and t_{Rob}, where the downward bias of empirical rejection frequencies positively depends on the size of N. As mentioned before, this is in contrast to results for univariate unit root tests, where positive size distortions are reported (e.g. Kim et al., 2002, Cavaliere and Taylor, 2007b). Rejection frequencies obtained by the 'White-type' statistic t_{HS} display only minor deviations from the nominal significance level. Documented results under spatially correlated errors indicate that size distortions reported for t_{OLS} are less pronounced than under constant volatility since the upward distortion invoked by cross sectional dependence is somewhat dampened by the negative shift in the innovation variance. Empirical rejection frequencies of t_{Rob} reflect moderate oversizing for panels with $N = 10$ and $T \geq 50$ and tend to be undersized if $N = 50$. Empirical results for t_{HS} are only indicative of a moderate finite sample size distortions but are otherwise very similar to those results obtained under constant volatility. With regard to local power, the scenario of an early downward shift in the innovation variance is characterized by a steeper gradient of rejection frequencies with respect to the sample size. While local power estimates are significantly smaller than in the constant variance case for small panel dimensions, up to six percentage points (respectively four percentage points in the SAR(1) case) higher rejection frequencies are documented for the largest simulated panel. The finding of superior power in large samples is supported by the analytically derived location parameter μ_l. Increased asymptotic local power is implied by the absolute value of the location parameter, which becomes larger compared with the benchmark scenario under a downward break in the innovation variance.

If the innovation variance features an upward shift towards the end of the sample, empirical rejection frequencies for t_{OLS} are in the range of 11.4-14.5% for all combinations of N and T and cross sectional independence. Rejection frequencies for t_{Rob} depend on the relative magnitude of the time dimension: for T large relative to N, the unit root null hypothesis is rejected significantly too often while for N larger than T, the undersizing observed in the previous experiments persists. Observed upward distortions are in accordance with the theoretical results in Proposition 2 and quantitatively in line with results obtained in a similar setting for the univariate DF test (Hamori and Tokihisa, 1997). In contrast, most accurate size control is obtained by t_{HS}, with empirical errors in rejection frequencies ranging between 0.2 and 2.1 percentage points. If the data is cross sectionally correlated, positive size distortions observed for t_{OLS} and, to a lesser extent, t_{Rob}, are even more pronounced whilst t_{HS} retains comparatively accurate size control. Results obtained under the alternative hypothesis show that local power estimates are less sensitive to the sample size compared with the case of an early downward shift of innovation variances. However, in line with the asymptotic results in Proposition 4, an upward break in the innovation variance induces decreased local power estimates for the largest considered panel dimension.

Table 2 reports results for DGP2, with all test statistics computed on detrended data. For the benchmark scenario of constant variances and either cross sectional independence or a SAR(1) error structure, results under H_0 are similar to those obtained for DGP1. As before, a large T relative to N is required in order to obtain rejection probabilities close to 5% for t_{Rob} and t_{OLS} yields substantial size distortions under spatial correlation while t_{HS} provides reliable size control in most instances. Noting that local power is computed in a neighborhood of order $O(T^{-1}N^{-1/4})$, the results imply that local power of all three tests is substantially reduced compared with the intercepts only case of DGP1. For both scenarios of variance shifts, all tests based on detrended data lose size control. If there is a reduction in the innovation variance, the tests are characterized by empirical rejection frequencies which increase with the sample size. In contrast, empirical rejection frequencies of all tests tend to zero in the case of a late positive variance shift. As mentioned in Section 3.5.1, the employed detrending scheme is based on the assumption of constant innovation variances. Obviously, the violation of this assumption invokes substantial adverse effects on the performance of the considered PURTs. We do not comment local power results for the latter two scenarios featuring variance shifts, as corresponding size estimates of the tests appear prohibitive for applied research.

Table 3 document results for data featuring serially correlated disturbances. These results indicate a general tendency of the tests to overreject H_0 if T is small, with most severe size distortions observed in the case of $N = 50$ and $T = 10$. The latter observation, however, does not apply to t_{Rob}, which remains undersized for this panel dimension. Imprecise size estimates for panels with small T are also not surprising from a theoretical point of view. As mentioned in Section 3.5.2, the estimates $\widehat{c}_i$ in the prewhitening regression (12) are $\sqrt{T}$ consistent and, hence, a relatively large time dimension is required in order to fully remove the effects of serial correlation from the data. Conditional on this finding, results obtained under H_0 are qualitatively similar to those obtained for DGP1. In particular, an early negative variance shift diminishes rejection probabilities under H_0, while a late positive shift leads to increased rejections of H_0. Moreover, t_{HS} remains robust against time varying volatility and, as before, application of t_{OLS} leads to markedly oversized rejection rates if the data is cross sectionally correlated. Local power estimates are similar to those obtained for serially uncorrelated error terms (DGP1) with some loss of local power for small values of T.

4.3 Summary of simulation results

The main result obtained by the simulation study is that an early negative (late positive) variance shift invokes a downward (upward) distortion of rejection frequencies for PURTs derived under the assumption of invariant second order moments. If the DGP formalizes a random walk without drift under H_0, rejection rates obtained by the 'White-type' statistic t_{HS} are not affected by variance breaks. Results under the local alternative H_l and the largest considered sample size confirm the theoretical finding that local power is asymptotically higher (lower) under a downward (upward) shift in the innovation variance. However, local power estimates in smaller samples are not necessarily in line with this asymptotic result. For the scenario of a random walk with drift under H_0, the applied detrending scheme (Breitung, 2000) leads to deceptive inference if there is a break in the innovation variance. Prewhitening the data to remove the effect of serially correlated error terms leaves the main findings unaffected, however, a larger time dimension is required for the empirical type one errors of the tests to come reasonably close to the nominal level.

5 Testing the Fisher hypothesis by means of PURTs

5.1 Economic background

The Fisher hypothesis (Fisher, 1930) postulates a stable one-to-one relationship between nominal interest rates and the expected rate of inflation. This hypothesis has been investigated in numerous empirical studies (see e.g. Rose, 1988, Crowder, 2003, Cooray, 2003 or Herwartz and Reimers, 2006, 2009). In its simplest form, the Fisher hypothesis states that the nominal interest rate in country i at time t, R_{it}, comprises the ex-ante real interest rate, $E_{t-1}[r_{it}]$, and the ex-ante expected inflation rate, $E_{t-1}[\pi_{it}]$, i.e.

$$R_{it} = E_{t-1}[r_{it}] + E_{t-1}[\pi_{it}] + \upsilon_{it},$$

where υ_{it} denotes an uninformative forecast error. Under rational expectations, actual and expected values differ only by a white-noise error term, i.e. $\pi_{it} = E_{t-1}[\pi_{it}] + \nu_{it}^{(1)}$ and $r_{it} = E_{t-1}[r_{it}] + \nu_{it}^{(2)}$. Accordingly, the ex-post real interest rate can be expressed as

$$r_{it} = R_{it} - \pi_{it} + \nu_{it}, \tag{15}$$

with $\nu_{it} = \upsilon_{it} - \nu_{it}^{(1)} - \nu_{it}^{(2)}$. The representation in (15) is a starting point for empirical investigations of the Fisher hypothesis by means of unit root tests. If, for instance, inflation and nominal interest rates are found to be I(1) variables, the Fisher hypothesis would imply (1, -1) cointegration establishing a stationary real interest rate. In contrast, a finding of nominal interest rates being I(1) and inflation being I(0) would contradict the Fisher hypothesis.

Prevalence of the Fisher hypothesis is still a question open to empirical research. Using univariate unit root tests on data for 18 economies, Rose (1988) concludes that nominal interest rates follow a unit root process while inflation rates are stationary. On the other hand, Rapach and Weber (2004) report evidence in favor of both variables being integrated of order one, albeit not forming a cointegration relationship. Evidence favorable for a stable long run relationship between inflation and nominal interest rates is reported in Crowder (2003) and Herwartz and Reimers

(2006, 2009). However, assessments of the Fisher hypothesis based on first generation PURTs yield conflicting results. For instance, Crowder (2003) finds some evidence of stationary nominal interest rates based on the PURT of Levin et al. (2002) for a panel of 9 industrialized economies. In the latter case, it is argued that these results must be interpreted carefully, as first generation PURTs are generally prone to distorted rejection frequencies through (neglected) cross sectional correlation. However, as highlighted by Kaliva (2008), analyses of the Fisher hypothesis must explicitly account for time-varying volatility as interest and inflation data display marked discrete volatility shifts. In the following assessment of the Fisher hypothesis, we document the presence of volatility breaks and cross sectional dependence in inflation and interest rate panel data sets. Subsequently, the PURTs discussed above are applied to the data to compare the marginal impacts of accounting for both departures from the assumptions underlying first- (and second-) generation PURTs.

5.2 Data and preliminary analyses

The empirical illustration is based upon the same sample of 9 developed economies considered in Crowder (2003).[1] Data is drawn from the International Financial Statistics of the IMF at the quarterly frequency, ranging from 1961Q2 to 2007Q2.[2] Inflation rates π_i are annual changes of the CPIs. Nominal interest rates, R_{it}, are selected depending on data availability and real interest rates, r_{it}, are obtained as $r_{it} = R_{it} - \pi_{it}$. Table 4 contains country specific definitions of interest rate data. The sample data is depicted in Figure 2 and eyeball inspection reveals close accordance with the figures provided in Crowder (2003). Figure 3 illustrates the prevalence of cross sectional dependence and time varying volatility. The left hand side graph documents a high degree of comovement of US and UK real interest rates over

[1]These countries are: Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, the United Kingdom and the United States.

[2]CPI data for the Netherlands is drawn from Dutch national statistics office as IFS data displays discretionary jumps, leading to inflation rates ranging between +30% and -17%.

the sample period. This is not surprising, given that both economies are highly integrated in the world economy and face similar external shocks, as for instance, abrupt oil price swings. The right hand side graph displays the first differences of the two time series, confirming a substantial reduction of volatility around 1985, ending roughly a decade of rather high fluctuations of real interest rates.

The estimated variance profiles $\hat{\vartheta}_i(s)$ of the three variables under investigation are displayed in Figure 4 in order to get an impression of the volatility processes governing the sample data (see Cavaliere and Taylor (2007b) for details and alternative estimators of variance profiles). Variance profiles $\vartheta_i(s)$ are calculated as

$$\hat{\vartheta}_i(s) = \frac{\sum_{t=1}^{\lfloor sT \rfloor} \hat{e}_{it}^2 + (sT - \lfloor sT \rfloor)\hat{e}_{i\lfloor sT \rfloor+1}^2}{\sum_{t=1}^{T} \hat{e}_{it}^2}, \tag{16}$$

where the $\hat{e}_{it}$'s are residuals from the first order autoregression of the considered process. While a (perfectly) homoskedastic variance profile would be represented by the 45° line, time varying volatilities are characterized by marked deviations from the diagonal.

Inspection of Figure 4 reveals that time-varying variances are rather the rule more than an exception for most cross section members. Moreover, it is obvious that estimated variance profiles differ across countries. However, focussing on the overall picture, there is some evidence of an upward followed by a downward shift in the first half of the sample period for all three variables and most of the economies.

In the following, we analyze to what extend previous evidence on the Fisher hypothesis obtained via first generation PURTs might have been distorted by cross sectional correlation or (unconditional) volatility shifts.

5.3 Panel unit root test results

The first step of the empirical analysis is to prewhiten the raw data. We use the SIC to determine individual specific lag lengths and subsequently apply the prewhitening procedure discussed in Section 3.5.2. In order to obtain a balanced panel, the maximum of the individual lag lengths is applied to all cross sectional units, hence prewhitening regressions for most cross sectional units are likely moderately overfitted. We use 12, 5 and 8 lags of the first differenced series for prewhitening inflation, nominal interest, and real interest rates, respectively. Assuming that inflation as well as interest rates contain a non-zero mean under the stationary alternative, prewhitened data is centered by subtracting the first observations. All PURTs are then computed for the resulting balanced panels of prewhitened and centered data. Table 5 lists the results of the empirical application. Test statistics for the pooled PURTs are documented in columns 3-5. The numbers in parentheses are $p-$values obtained from the Gaussian CDF. Results for the three variables are listed by rows.

Using the statistic t_{OLS} to test the order of integration of the inflation rate yields a t-ratio of -3.52 and, hence, a rejection of the unit root null hypothesis at any conventional significance level. This result is in line with Crowder (2003), reporting a t-ratio -5.32 obtained via the Levin et al. (2002) procedure. Given that based on univariate tests, the unit root hypothesis is maintained for all sample economies, Crowder (2003) argues that the rejection of H_0 obtained by the PURT might be due to size bias, invoked by cross sectional dependence. Accordingly, we apply the robust t_{Rob} statistic proposed by Breitung and Das (2005). The resulting t-ratio of -2.45 is substantially smaller in absolute value, however, it still leads to a rejection of the null hypothesis. The relative impact of time varying volatility of the sample data on pooled PURTs might be assessed by application of the volatility break robust statistic t_{HS}. The resulting t-ratio of -1.85 is larger than the $t-$ratios obtained by t_{OLS} and t_{Rob} and the corresponding marginal significance level is 3.2%.

Qualitatively similar results are obtained for the nominal interest rate. By means of the first generation test statistic t_{OLS}, a t-ratio of -4.22 is calculated, which is substantially smaller in absolute value than -7.57 reported in Crowder (2003), but nevertheless leads to a clear rejection of H_0. Again, application of the robust tests increase marginal significance levels. The cross sectional dependence robust test statistic t_{Rob} still implies a rejection of H_0 at the 1% level. In this instance, application of t_{HS} might lead to a different test decision as the respective $p-$values of 4.8% is just below the widespread 5% threshold for a rejection of the PUR hypothesis.

Finally, we test for the panel unit root in the real interest rate. All tests yield results in support of panel stationarity of the real interest rate, and thus, of the Fisher hypothesis. Note however, that at the 5% significance level, even the volatility break robust test does not rule out the possibility of inflation and nominal interest rates being likewise panel stationary variables. Accordingly, one should be careful in interpreting stationarity of real interest rates as a cointegration relationship, linking two nonstationary variables.

6 Conclusions

In this paper we investigate the effects of discrete breaks in the innovation variance on homogenous panel unit root tests. It is shown that size distortions documented in the literature on univariate unit root tests under time varying variances carry over to the panel case.

The limiting distribution of first and second generation pooled PURTs under a discrete variance shift are derived and it is shown that only the 'White-type' PURT statistic proposed in Herwartz and Siedenburg (2008) remains asymptotically Gaussian under the unit root null hypothesis. Under local-to-unity alternatives, it turns out that local power depends on the particular pattern of breaks in the innovation variance. By means of a Monte Carlo study we analyze a variety of possible model settings, including deterministic trends, autocorrelated disturbances and cross sectional correlation. The simulation study reveals that the 'White-type' statistic offers most reliable size control in finite samples and is asymptotically as powerful as the statistic proposed by Breitung and Das (2005). Moreover, it turns out that the employed detrending scheme to account for linear time trends leads to deceptive inference for all analyzed statistics if there is a break in the innovation variance. As an empirical illustration, recent evidence on the Fisher hypothesis in Crowder (2003) is reconsidered. Based on data for a cross section of 9 developed economies, sampled over the period 1961Q2 - 2007Q2, the order of integration of inflation rates as well as of nominal and real interest rates is tested. The results illustrate the importance of robust panel unit root tests, accounting for nonstationary innovation variances and cross sectional dependence.

The results in this paper raise a number of issues for future research. Firstly, noting that the detrending scheme proposed in Breitung (2000) is apparently not applicable under time varying innovation variances, it appears promising to study alternative detrending schemes. Secondly, the assumed constancy of cross sectional correlation might not generally hold in empirical applications. It seems sensible to investigate how time varying patterns of cross sectional correlation affect the performance of PURTs and if the proposed robust statistic is also able to cope with this kind of nuisance appropriately. Finally, the focus of this paper was on PURTs which are pivotal only under weak cross sectional dependence. Extending the analysis to the case of strong form cross sectional dependence is a topic of immediate interest for future research.

A Appendix

A.1 Proof of Proposition 1

Basically, all subsequent proofs are extensions of the proofs in Breitung and Das (2005) to the case of discrete variance breaks. To derive the limiting distribution of t_{OLS} define

$$t_{OLS} = \frac{N^{-0.5}T^{-1}\sum_{t=1}^{T}\boldsymbol{y}_{t-1}\Delta\boldsymbol{y}_t}{\sqrt{N^{-1}T^{-2}\sum_{t=1}^{T}\hat{\sigma}_e^2\boldsymbol{y}_{t-1}'\boldsymbol{y}_{t-1}}} = \frac{a_{NT}}{\sqrt{b_{OLS}}}.$$

Consider the numerator first. Under H_0, it follows that

$$a_{NT} = N^{-0.5}T^{-1}\sum_{t=1}^{T}\boldsymbol{y}_{t-1}\Delta\boldsymbol{y}_t = N^{-0.5}T^{-1}\sum_{t=1}^{T}\boldsymbol{y}_{t-1}\boldsymbol{e}_t.$$

Noting that we may decompose Ω_t as

$$\Omega_t = \begin{cases} \Omega_1 = \Gamma\Lambda_1\Gamma', & \text{if } \quad 0 < t \leq T_1 \\ \Omega_2 = \Gamma\Lambda_2\Gamma', & \text{if } \quad T_1 < t \leq T. \end{cases},$$

where $\Lambda_\bullet = diag(\lambda_1, ..., \lambda_N)'$, $\bullet = 1, 2$, is a diagonal matrix of eigenvalues and Γ is the corresponding matrix of normalized eigenvectors, which remains unaffected by the shift in idiosyncratic variance components due to the assumed time invariant pattern of cross sectional correlation . Now that $\boldsymbol{u}_t = \Lambda_\bullet^{1/2}\Gamma'\boldsymbol{e}_t$ is an $N \times 1$ vector of cross sectionally independent error terms with unit variance and $\boldsymbol{z}_t = \Lambda_\bullet^{1/2}\Gamma'\boldsymbol{y}_t$, is an $N \times 1$ vector of mutually uncorrelated random walks, the numerator can be expressed as

$$\begin{aligned} a_{NT} &= N^{-0.5}T^{-1}\left[\sum_{t=1}^{T_1}\left(\sum_{s=1}^{t-1}\boldsymbol{u}_s\right)'\Gamma\Lambda_1\Gamma'\boldsymbol{u}_t\right. \\ &+ \left.\sum_{t=T_1+1}^{T}\left(\sum_{s=1}^{T_1}\boldsymbol{u}_s\right)'\Gamma\Lambda_1^{1/2}\Lambda_2^{1/2}\Gamma'\boldsymbol{u}_t + \sum_{T_1+1}^{T}\left(\sum_{s=T_1+1}^{t-1}\boldsymbol{u}_s\right)'\Gamma\Lambda_2\Gamma'\boldsymbol{u}_t\right], \quad (17) \end{aligned}$$

The terms in (17) are constructed such that summation always only comprises error terms with homogenous variances as, for instance, $T_1^{-1/2}\sum_{s=1}^{T_1}\boldsymbol{u}_s = \boldsymbol{z}_{T_1} \xrightarrow{d} \boldsymbol{W}(1)$ is a multivariate Gaussian random vector. Defining $\overline{\boldsymbol{z}}_{t-1} = \boldsymbol{z}_{t-1} - \boldsymbol{z}_{T_1}$, we have

$$a_{NT} = N^{-0.5}T^{-1}\left[\sum_{i=1}^{N}\lambda_{1i}\sum_{t=1}^{T_1}z_{i,t-1}u_{it} + \sum_{i=1}^{N}\lambda_{1i}^{1/2}\lambda_{2i}^{1/2}z_{iT_1}\sum_{t=T_1+1}^{T}u_{it} + \sum_{i=1}^{N}\lambda_{2i}\sum_{t=T_1+1}^{T}\overline{\boldsymbol{z}}_{t-1}'u_{it}\right].$$

To economize on space, in the following we throughout use the shorthand notations $\int W_i$ and $\int W_i dW_i$ instead of $\int W_i(r)dr$ and $\int W_i(r)dW_i(r)$. As $T, T_1 \to \infty$, common invariance principles for partial sum processes imply that

$$\begin{aligned} a_{NT} &\xrightarrow{d} N^{-0.5}\left[\delta \sum_{i=1}^{N} \lambda_{1i} \int_0^1 W_i dW_i + \sqrt{\delta(1-\delta)} \sum_{i=1}^{N} \sqrt{\lambda_{1i}\lambda_{2i}} W_{i,T_1}(1) W_{i,T_2}(1) \right. \\ &+ \left. (1-\delta) \sum_{i=1}^{N} \lambda_{2i} \int_0^1 W_i dW_i \right]. \end{aligned} \tag{18}$$

The subscripts in $W_{i,T_1}(1)$ and $W_{i,T_2}(1)$ in the medium term of the right hand side of (18) are chosen in order to highlight that both terms are the values of two uncorrelated Brownian motions at $r = 1$ with $T_2 = T - T_1$. Since $W_{i,T_1}(1)$ and $W_{i,T_2}(1)$ are independent Gaussian random variables and $E\left[\int_0^1 W_i dW_i\right] = 0$ while $Var\left[\int_0^1 W_i dW_i\right] = 0.5$, one obtains for from the central limit theorem for mean zero *iid* random variables that the numerator of the three test statistics t_{OLS}, t_{Rob}, and t_{HS} is given by

$$a_{NT} \xrightarrow{d} N(0, \overline{\sigma}^2), \quad \overline{\sigma}^2 = 0.5\delta^2 \overline{\lambda_1^2} + \delta(1-\delta)\overline{\lambda_1 \lambda_2} + 0.5(1-\delta)^2 \overline{\lambda_2^2}, \tag{19}$$

where $\overline{\lambda_\bullet^2} = N^{-1}\sum_{i=1}^N \lambda_\bullet^2$, with $\bullet = 1, 2$, and $\overline{\lambda_1\lambda_2} = N^{-1}\sum_{i=1}^N \lambda_1\lambda_2$ as $N \to \infty$.

Now consider the denominator of t_{OLS}. We have

$$\begin{aligned} b_{OLS} &= N^{-1}T^{-2}\widehat{\sigma}^2 \sum_{t=1}^{T} \boldsymbol{y}_{t-1}' \boldsymbol{y}_{t-1} \\ &= N^{-1}T^{-2}\widehat{\sigma}^2 \left[\sum_{t=1}^{T_1} \boldsymbol{z}_{t-1}' \Lambda_1 \boldsymbol{z}_{t-1} + T_1 T_2 \frac{\boldsymbol{z}_{t-1}}{\sqrt{T_1}}' \Lambda_1 \frac{\boldsymbol{z}_{t-1}}{\sqrt{T_1}} + \sum_{t=T_1+1}^{T} \overline{\boldsymbol{z}}_{t-1}' \Lambda_2 \overline{\boldsymbol{z}}_{t-1}\right]. \end{aligned}$$

As $T \to \infty$,

$$\begin{aligned} b_{OLS} &\xrightarrow{d} N^{-1}\left(N^{-1}\delta \sum_{i=1}^{N} \lambda_{1i} + N^{-1}(1-\delta)\sum_{i=1}^{N} \lambda_{2i}\right) \\ &\times \left[\delta^2 \sum_{i=1}^{N} \lambda_{1i} \int_0^1 W_i^2 + \delta(1-\delta)\sum_{i=1}^{N} \lambda_{1i} W_i(1)^2 + (1-\delta)^2 \sum_{i=1}^{N} \lambda_{2i} \int_0^1 W_i^2\right]. \end{aligned}$$

Letting $N \to \infty$, we obtain convergence in probability

$$b_{OLS} \xrightarrow{p} \left(\delta\overline{\lambda_1} + (1-\delta)\overline{\lambda_2}\right)\left[0.5\delta^2\overline{\lambda_1} + \delta(1-\delta)\overline{\lambda_1} + 0.5(1-\delta)^2\overline{\lambda_2}\right], \tag{20}$$

since $E[\int_0^1 W_i^2] = 0.5$ and $E[W_i(1)^2] = 1$. It is immediate from (20) that $b_{OLS} \neq \overline{\sigma}^2$, implying that t_{OLS} does not converge to a Gaussian limiting distribution if there is a break in the innovation variance, even under cross sectional independence and cross sectionally homogeneous variances.

□

A.2 Proof of Proposition 2

Since the numerator is the same for t_{OLS}, t_{Rob}, and t_{HS}, it suffices to consider the denominator to derive the asymptotic distribution of t_{Rob}. Specifically,

$$b_{Rob} = \sum_{t=1}^{T} \boldsymbol{y}'_{t-1}\widehat{\Omega}\boldsymbol{y}_{t-1}, \text{ with } \widehat{\Omega} = T^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{e}}_t\widehat{\boldsymbol{e}}_t{}' = T^{-1}\sum_{t=1}^{T}\boldsymbol{e}_t\boldsymbol{e}'_t + o_p(1).$$

Making use of the same decomposition as in (17) and dropping lower order terms yields

$$\begin{aligned}
b_{Rob} \;=\; & N^{-1}T^{-2}\left[\sum_{t=1}^{T_1}\boldsymbol{z}'_{t-1}\Lambda_1^{1/2}\left\{\frac{T_1}{T}\Lambda_1\left(T_1^{-1}\sum_{t=1}^{T_1}\boldsymbol{u}_t\boldsymbol{u}'_t\right)+\frac{T_2}{T}\Lambda_2\left(T_2^{-1}\sum_{t=T_1+1}^{T}\boldsymbol{u}_t\boldsymbol{u}'_t\right)\right\}\Lambda_1^{1/2}\boldsymbol{z}_{t-1}\right.\\
+ \; & T_1T_2\frac{\boldsymbol{z}_{T_1}}{\sqrt{T_1}}{}'\Lambda_1^{1/2}\left\{\frac{T_1}{T}\Lambda_1\left(T_1^{-1}\sum_{t=1}^{T_1}\boldsymbol{u}_t\boldsymbol{u}'_t\right)+\frac{T_2}{T}\Lambda_2\left(T_2^{-1}\sum_{t=T_1+1}^{T}\boldsymbol{u}_t\boldsymbol{u}'_t\right)\right\}\Lambda_1^{1/2}\frac{\boldsymbol{z}_{T_1}}{\sqrt{T_1}}\\
+ \; & \left.\sum_{t=T_1+1}^{T}\overline{\boldsymbol{z}}'_{t-1}\Lambda_2^{1/2}\left\{\frac{T_1}{T}\Lambda_1\left(T_1^{-1}\sum_{t=1}^{T_1}\boldsymbol{u}_t\boldsymbol{u}'_t\right)+\frac{T_2}{T}\Lambda_2\left(T_2^{-1}\sum_{t=T_1+1}^{T}\boldsymbol{u}_t\boldsymbol{u}'_t\right)\right\}\Lambda_2^{1/2}\overline{\boldsymbol{z}}_{t-1}\right].
\end{aligned}$$

As $T \to \infty$ and by noting that $T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}'_t$, $T_1^{-1}\sum_{t=1}^{T_1}\boldsymbol{u}_t\boldsymbol{u}'_t$ and $T_2^{-1}\sum_{t=T_1+1}^{T}\boldsymbol{u}_t\boldsymbol{u}'_t \xrightarrow{p} E[\boldsymbol{u}_t\boldsymbol{u}'_t] = I_N$, one obtains

$$\begin{aligned}
b_{Rob} \;\xrightarrow{d}\; & N^{-1}\left[\delta^3\sum_{i=1}^{N}\lambda_{1i}^2\int_0^1 W_i^2 + \delta^2(1-\delta)\sum_{i=1}^{N}\lambda_{1i}\lambda_{2i}\int_0^1 W_i^2\right.\\
+ \; & \delta^2(1-\delta)\sum_{i=1}^{N}\lambda_{1i}^2 W_i(1)^2 + \delta(1-\delta)^2\sum_{i=1}^{N}\lambda_1\lambda_{2i}W_i(1)^2\\
+ \; & \left.(1-\delta)^2\delta\sum_{i=1}^{N}\lambda_{1i}\lambda_{2i}\int_0^1 W_i^2 + (1-\delta)^3\sum_{i=1}^{N}\lambda_{2i}^2 W_i^2\right].
\end{aligned}$$

For $N \to \infty$ this yields

$$
\begin{aligned}
b_{Rob} &\xrightarrow{p} 0.5\left\{\delta^3\overline{\lambda_1^2} + \left(\delta^2(1-\delta)+(1-\delta)^2\delta\right)\overline{\lambda_1\lambda_2} + (1-\delta)^3\overline{\lambda_2^2}\right\} \\
&+ \left\{\delta^2(1-\delta)\overline{\lambda_1^2} + \delta(1-\delta)^2\overline{\lambda_1\lambda_2}\right\} \\
&= \left(\delta^2 - 0.5\delta^3\right)\overline{\lambda_1^2} + \left(1.5\delta + 0.5\delta^2 - 2\delta^3\right)\overline{\lambda_1\lambda_2} + 0.5(1-\delta)^3\overline{\lambda_2^2},
\end{aligned}
$$

establishing that

$$
\nu_{Rob} = \frac{0.5\delta^2\overline{\lambda_1^2} + \delta(1-\delta)\overline{\lambda_1\lambda_2} + 0.5(1-\delta)^2\overline{\lambda_2^2}}{(\delta^2 - 0.5\delta^3)\,\overline{\lambda_1^2} + (1.5\delta + 0.5\delta^2 - 2\delta^3)\,\overline{\lambda_1\lambda_2} + 0.5(1-\delta)^3\overline{\lambda_2^2}} \neq 1.
$$

The result in Breitung and Das (2005) with $\nu_{Rob} = 1$ holds as a special case if $\delta = 1$ or $\delta = 0$.

$\square$

A.3 Proof of Proposition 3

Finally, we show that $b_{HS} \to \overline{\sigma}^2$ for $T \to \infty$ followed by $N \to \infty$. With $\check{\boldsymbol{e}}_t = \boldsymbol{e}_t + o_p(1)$ and dropping the lower order term in the expression we obtain

$$
\begin{aligned}
b_{HS} &= N^{-1}T^{-2}\sum_{t=1}^{T}\boldsymbol{y}_{t-1}'\boldsymbol{e}_t\boldsymbol{e}_t'\boldsymbol{y}_{t-1} = N^{-1}T^{-2}\sum_{t=1}^{T}\boldsymbol{z}_{t-1}'\Lambda\boldsymbol{u}_t\boldsymbol{u}_t'\Lambda\boldsymbol{z}_{t-1} \\
&= N^{-1}T^{-2}\left[\sum_{t=1}^{T_1}\boldsymbol{z}_{t-1}'\Lambda_1\boldsymbol{u}_t\boldsymbol{u}_t'\Lambda_1\boldsymbol{z}_{t-1} + T_1T_2\frac{\boldsymbol{z}_{T_1}}{\sqrt{T_1}}'\Lambda_1^{1/2}\Lambda_2^{1/2}\boldsymbol{u}_t\boldsymbol{u}_t'\Lambda_1^{1/2}\Lambda_2^{1/2}\frac{\boldsymbol{z}_{T_1}}{\sqrt{T_1}}\right. \\
&+ \left.\sum_{t=T_1+1}^{T}\overline{\boldsymbol{z}}_{t-1}'\Lambda_2\boldsymbol{u}_t\boldsymbol{u}_t'\Lambda_2\overline{\boldsymbol{z}}_{t-1}\right].
\end{aligned}
$$

By Assumption ($\mathcal{A}_1$), restricting $E[u_{it}^4] < \infty$, we can define $\xi_i = u_{it}^2$, which is an *iid* random variable with $E[\xi_i] = 1$ and $Var[\xi_i] = \sigma_{\xi_i}^2 < \infty$. Hence, as $T \to \infty$,

$$
\begin{aligned}
b_{HS} &\xrightarrow{d} N^{-1}\left[\delta^2\sum_{i=1}^{N}\lambda_{1i}^2\int_0^1 W_i^2\xi_i + \delta(1-\delta)\sum_{i=1}^{N}\lambda_{1i}\lambda_{2i}W_i(1)^2\xi_i\right. \\
&+ \left.(1-\delta)^2\sum_{i=1}^{N}\lambda_{2i}^2\int_0^1 W_i^2\xi_i\right].
\end{aligned}
$$

Because of the independence of the ξ_i and the partial sum processes, as $N \to \infty$ this expression converges in probability

$$
b_{HS} \xrightarrow{p} 0.5\delta^2\overline{\lambda_1^2} + \delta(1-\delta)\overline{\lambda_1\lambda_2} + 0.5(1-\delta)^2\overline{\lambda_2^2} = \overline{\sigma}^2,
$$

verifying that $t_{HS} \xrightarrow{d} N(0,1)$.

$\square$

A.4 Proof of Proposition 4

The derivation of the limiting distribution of t_{HS} under the sequence of local alternatives $H_l : \rho = 1 - \frac{c}{T\sqrt{N}}$ is based on the respective proof for the statistic t_{Rob} in Breitung and Das (2005). First note that in local-to-unity neighborhoods as defined above, $\boldsymbol{z}_{[rT]} \xrightarrow{d} \boldsymbol{W}_i(r)$ for all $0 \leq c < \infty$. It follows that the numerator of t_{HS} is given by

$$a_{HS} = T^{-1}N^{-1/2}\sum_{t=1}^{T}\boldsymbol{y}_{t-1}\Delta\boldsymbol{y}_t = T^{-1}N^{-1/2}\sum_{t=1}^{T}\boldsymbol{u}_t'\Lambda\Delta\boldsymbol{z}_t - cT^{-2}N^{-1}\sum_{t=1}^{T}\boldsymbol{z}_{t-1}'\Lambda\boldsymbol{z}_{t-1}.$$

From the proof of Proposition A.1 it follows directly that the first term on the right hand side equals the numerator under the null hypothesis while the second term converges in probability

$$cT^{-2}N^{-1}\sum_{t=1}^{T}\boldsymbol{z}_{t-1}'\Lambda\boldsymbol{z}_{t-1} \xrightarrow{p} c\left[0.5\delta^2\overline{\lambda_1} + \delta(1-\delta)\overline{\lambda_1} + 0.5(1-\delta)^2\overline{\lambda_2}\right],$$

as $T \to \infty$, followed by $N \to \infty$. From the proof of Proposition 3 it follows that $t_{HS} \xrightarrow{d} N(-c\mu_l, 1)$, with

$$\mu_l = \frac{0.5\delta^2\overline{\lambda_1} + \delta(1-\delta)\overline{\lambda_1} + 0.5(1-\delta)^2\overline{\lambda_2}}{\sqrt{0.5\delta^2\overline{\lambda_1^2} + \delta(1-\delta)\overline{\lambda_1\lambda_2} + 0.5(1-\delta)^2\overline{\lambda_2^2}}}.$$

Again, the result in Breitung and Das (2005) with $\mu_{l,Rob} = \sqrt{0.5}\,\overline{\lambda_1}/\sqrt{\overline{\lambda_1^2}}$ obtains as a special case with $\delta = 0$ or $\delta = 1$.

$\square$

Acknowledgments: Financial support of Deutsche Forschungsgemeinschaft (DFG) (HE 2188/1-1) is gratefully acknowledged. The authors are also grateful for helpful comments provided by Jörg Breitung, Christoph Hanck, Matei Demetrescu and other discussants at 15th Conference on Panel Data 2009 in Bonn, Germany.

References

Beck, N., Katz, J. N., 1995. What to do (and not to do) with time-series cross-section data. American Political Science Review 89 (3), 634–647.

Breitung, J., 2000. The local power of some unit root tests for panel data. In: Baltagi, B. (Ed.), Advances in Econometrics, Vol. 15. JAI, Amsterdam, pp. 161–178.

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

Breitung, J., Meyer, W., 1994. Testing for unit roots in panel data: are wages on different bargaining levels cointegrated? Applied Economics 26 (4), 353–361.

Breitung, J., Pesaran, H. M., 2008. Unit roots and cointegration in panels. In: Matyas, L., Sevestre, P. (Eds.), The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice. Kluwer Academic Publishers, Dordrecht.

Breitung, J., Westerlund, J., 2009. Myths and facts about panel unit root tests. Mimeo.

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

Cavaliere, G., Taylor, A. M. R., 2007a. Testing for unit roots in autoregressions with multiple level shifts. Econometric Theory 23 (6), 1162–1215.

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

Cavaliere, G., Taylor, A. M. R., 2008. Bootstrap unit root tests for time series with nonstationary volatility 24 (1), 43–71, econometric Theory.

Cooray, A., 2003. The Fisher effect: a survey. The Singapore Economic Review 48 (2), 135–150.

Crowder, W. J., 2003. Panel estimates of the Fisher effect. University of Texas at Arlington, Discussion Paper.

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.

Elhorst, J. P., 2003. Specification and estimation of spatial panel data models. International Regional Science Review 26 (3), 244–268.

Fisher, I., 1930. The Theory of Interest. Macmillan, New York.

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

Hanck, C., 2009a. Cross sectional correlation robust tests for panel cointegration. Journal of Applied Statistics 36 (7), 817–833.

Hanck, C., 2009b. Nonstationary-volatility robust panel unit root tests and the great moderation. Mimeo.

Herwartz, H., Reimers, H. E., 2006. Panel nonstationarity tests of the Fisher hypothesis: an analysis of 114 economies during the period 1960-2004. Applied Econometrics and International Development 6 (3), 39–54.

Herwartz, H., Reimers, H. E., 2009. A functional coefficient approach to modeling the fisher hypothesis: World wide evidence. *forthcoming in:* Macroeconomic Dynamics.

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.

Hurlin, C., Mignon, V., 2007. Second generation panel unit root tests. Manuscript, THEMA-CNRS, University of Paris X.

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

Kaliva, K., 2008. The fisher effect, survey data and time-varying volatility. Empirical Economics 35 (1), 1–10.

Kim, C. J., Nelson, C. R., 1999. Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. Review of Economics and Statistics 81 (4), 608–616.

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.

McConnell, M. M., Perez-Quiroz, G., 2000. Output fluctuations in the United States: what has changed since the early 1980s? American Economic Review 90 (5), 1464–1476.

Moon, H. R., Perron, B., Phillips, P. C. B., 2007. Incidental trends and the power of panel unit root tests. Journal of Econometrics 141 (2), 416–459.

Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69 (6), 1519–1554.

Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49 (6), 1399–1416.

O'Connell, P. G. J., 1998. The overvaluation of purchasing power parity. Journal of International Economics 44 (1), 1–19.

Perron, P., Ng, S., 1996. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. Review of Economic Studies 63 (3), 435–463.

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.

Phillips, P. C. B., Moon, H. R., 1999. Linear regression limit theory for nonstationary panel data. Journal of Econometrics 67 (5), 1057–1111.

Rapach, D. E., Weber, C. E., 2004. Are real interest rates really nonstationary? new evidence from tests with good size and power. Journal of Macroeconomics 26 (3), 409–430.

Rose, A. K., 1988. Is the real interest rate stable? The Journal of Finance 43 (5), 1095–1112.

Stock, J. H., 1999. A class of tests for integration and cointegration. In: Engle, R. F., White, H. (Eds.), Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger. Oxford University Press, Oxford.

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

Taylor, A. M., Taylor, M. P., 2004. The purchasing power parity debate. Journal of Economic Perspectives 18 (4), 135–158.

Figure 1: Asymptotic variance of t_{OLS}, $\overline{\nu}_{OLS}$

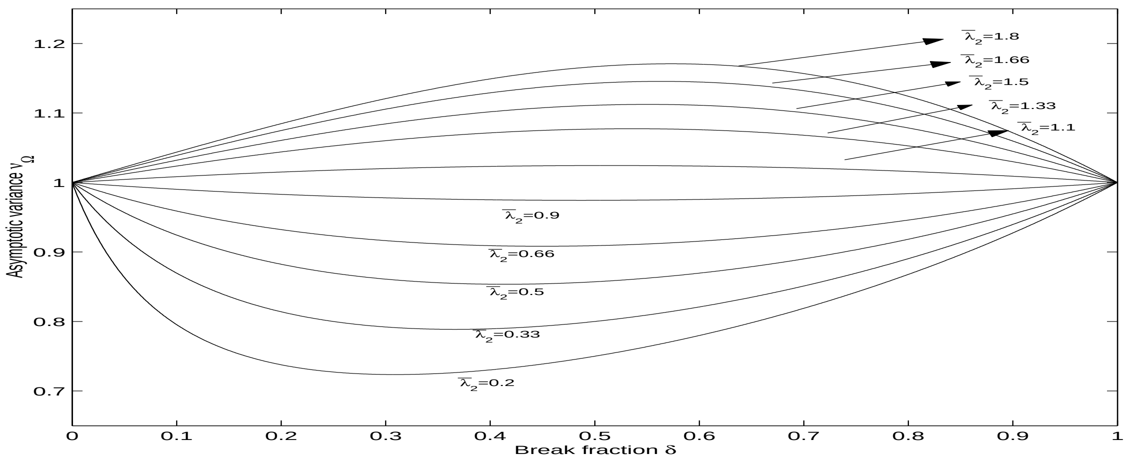

Figure 2: Nominal Interest Rates and Inflation rates, 1961Q2 - 2007Q2

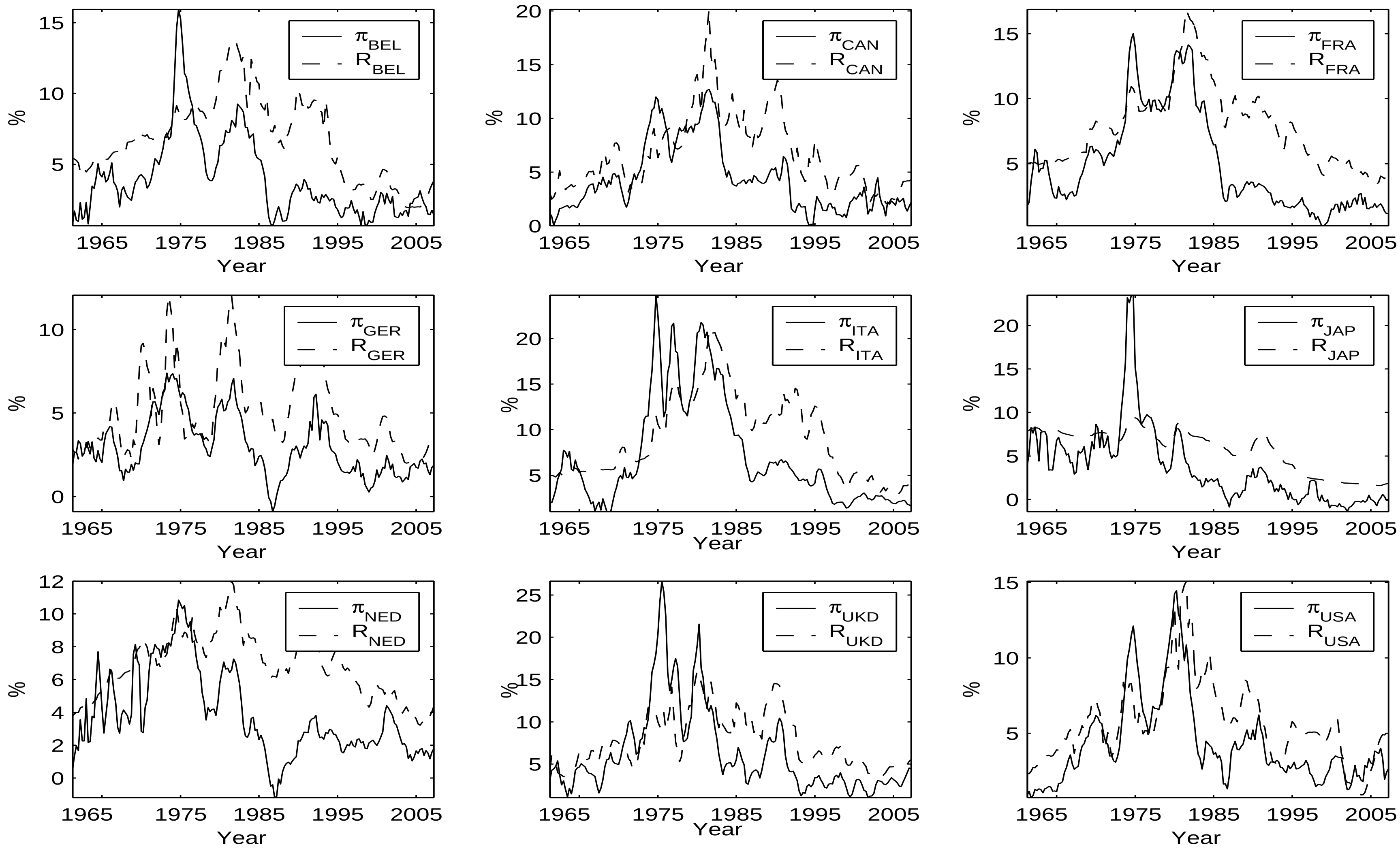

Figure 3: Real interest rates, levels and 1st differences, US vs. UK

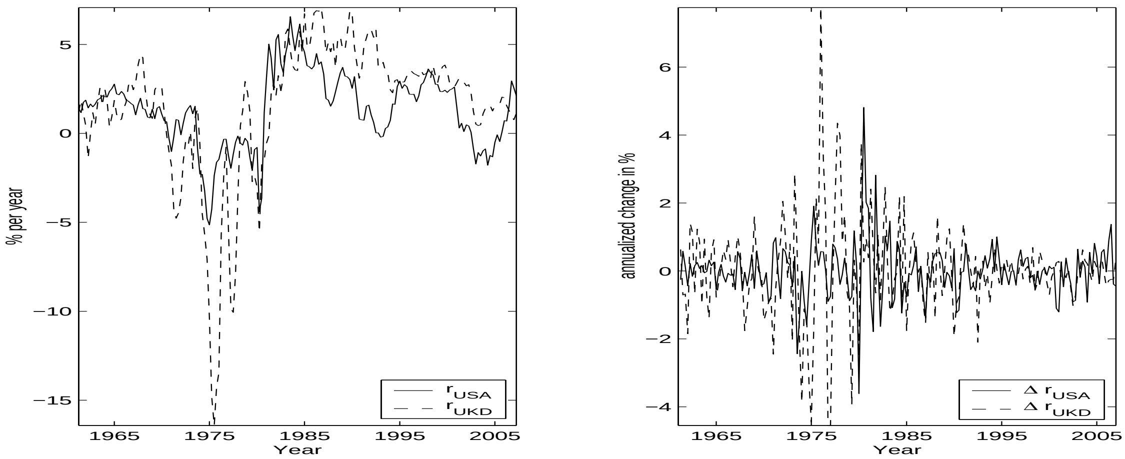

Figure 4: Estimated variance profiles

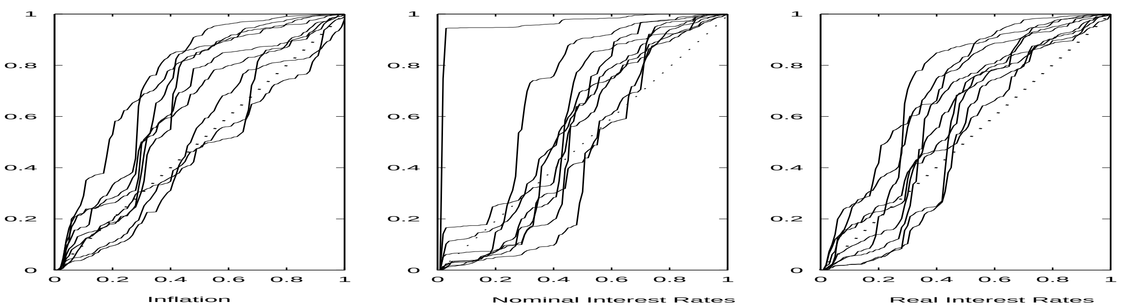

Table 1: DGP1

		CS independence						SAR(1) model					
		ρ=1			$\rho = 1 - \frac{5}{T\sqrt{N}}$			ρ=1			$\rho = 1 - \frac{5}{T\sqrt{N}}$		
N	*T*	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*
Constant variance													
10	10	6.9	2.0	5.8	36.1	35.1	30.0	23.8	4.7	5.8	19.9	18.1	14.9
10	50	6.8	5.0	6.5	39.1	38.6	37.4	25.2	8.2	8.5	20.3	17.0	16.7
10	100	6.9	6.1	6.7	43.8	43.7	44.3	24.4	7.1	7.1	25.0	20.2	20.5
10	250	7.2	6.8	7.2	57.0	56.5	55.8	25.1	8.3	8.5	28.3	21.1	20.9
50	10	5.2	0.0	5.3	47.0	46.9	38.1	20.1	0.5	5.3	21.4	20.5	18.2
50	50	5.8	1.3	5.7	59.0	57.9	57.2	20.8	4.3	7.0	26.8	24.8	23.0
50	100	5.2	2.3	5.3	74.8	74.5	73.3	20.1	4.5	5.9	36.9	33.9	33.5
50	250	5.5	3.9	5.6	84.8	84.4	84.2	21.1	5.7	6.4	44.3	39.4	38.5
Early negative variance shift													
10	10	3.3	0.2	5.8	12.5	13.3	10.2	15.2	1.9	4.6	8.3	7.5	9.0
10	50	4.5	2.7	6.7	9.9	9.7	10.0	19.9	5.8	6.9	8.3	7.4	6.9
10	100	3.8	2.9	6.0	17.3	17.3	14.9	18.7	6.7	7.7	11.0	9.1	8.4
10	250	4.2	3.6	6.1	38.2	38.6	35.6	21.1	7.5	7.9	19.5	15.3	13.5
50	10	1.5	0.0	4.4	8.3	10.1	6.7	13.0	0.0	5.4	7.4	7.8	5.7
50	50	2.8	0.1	5.5	19.6	20.6	17.4	15.4	1.4	5.8	11.9	11.5	11.6
50	100	2.7	0.2	5.5	50.5	49.7	49.4	16.1	2.0	6.6	22.4	20.3	19.2
50	250	3.0	1.5	5.4	92.1	90.8	90.9	15.8	3.1	6.2	49.3	43.1	42.1
Late positive variance shift													
10	10	12.7	4.6	3.5	20.3	20.5	22.3	29.2	8.5	3.9	14.4	13.7	13.5
10	50	13.5	10.8	6.0	24.9	24.1	23.6	31.9	12.0	6.2	19.1	15.3	14.6
10	100	12.7	11.5	6.3	28.6	28.0	26.2	32.7	13.1	6.7	19.8	16.8	14.6
10	250	14.5	13.8	7.1	30.7	30.5	27.0	34.8	14.4	8.2	19.3	14.9	13.5
50	10	11.4	0.1	3.4	22.5	22.7	25.5	26.8	1.9	3.5	12.7	13.3	14.5
50	50	13.0	4.0	5.2	31.6	30.0	31.2	30.2	9.1	5.8	17.8	17.0	16.6
50	100	12.8	8.2	6.3	34.4	32.5	33.0	29.3	11.3	6.3	20.6	19.3	17.8
50	250	12.7	10.1	6.0	40.8	39.6	39.4	30.2	12.5	6.6	21.6	20.4	18.8

Notes: *OLS*, *Rob* and *HS* refer to the PURT statistics defined in (4), (5),(6). Results are based on 5000 replications and the nominal size equals 5%. Local power results are size adjusted. Data is generated according to DGP1 and all tests are computed on demeaned data.

Table 2: Empirical rejection frequencies, DGP2

		CS independence						SAR(1) model					
		ρ=1			$\rho = 1 - \frac{5}{TN^{1/4}}$			ρ=1			$\rho = 1 - \frac{5}{TN^{1/4}}$		
N	*T*	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*
Constant variance													
10	10	7.4	1.8	5.4	16.2	16.3	14.5	23.7	5.1	5.4	10.6	10.4	8.9
10	50	6.9	4.9	6.3	17.5	17.3	15.7	22.7	7.0	7.0	11.3	10.1	8.9
10	100	6.1	5.2	5.7	18.7	18.6	19.1	21.0	6.6	6.7	12.0	11.3	11.1
10	250	5.8	5.5	5.7	19.7	19.2	19.3	22.1	6.9	6.9	12.8	11.7	11.8
50	10	6.6	0.0	5.2	22.6	23.3	19.3	21.6	0.7	5.2	12.8	13.2	11.2
50	50	5.6	1.1	5.2	23.6	24.1	22.9	20.9	3.9	6.1	13.8	13.0	12.7
50	100	5.8	2.6	5.6	24.7	24.8	24.3	20.8	4.9	6.5	12.7	12.6	11.6
50	250	5.2	3.5	5.2	28.1	28.3	27.6	21.4	5.4	6.1	15.0	13.7	13.7
Early negative variance shift													
10	10	8.4	0.9	3.9	20.7	19.5	23.7	23.9	4.6	3.3	11.5	11.4	13.5
10	50	14.6	9.7	9.3	13.8	13.8	15.0	32.2	12.6	8.5	8.7	8.0	8.4
10	100	15.4	12.6	11.2	13.2	12.6	13.9	34.0	15.1	10.0	10.2	9.0	8.1
10	250	15.2	14.2	11.4	16.1	15.8	14.8	33.1	15.3	9.9	10.5	9.5	10.0
50	10	11.9	0.0	5.4	32.8	30.9	34.0	26.7	0.1	5.0	15.6	14.8	17.3
50	50	22.1	2.2	16.0	21.9	19.3	25.3	35.8	7.2	10.3	11.5	10.9	12.6
50	100	23.4	7.6	18.3	23.1	21.6	24.6	37.9	11.9	12.6	11.6	10.5	10.6
50	250	23.5	14.3	19.0	30.4	28.5	30.5	38.7	14.8	13.1	14.9	13.8	13.8
Late positive variance shift													
10	10	0.1	0.0	0.0	17.4	15.9	14.5	1.8	0.2	0.1	13.6	13.6	13.2
10	50	0.0	0.0	0.0	18.4	18.7	17.9	2.1	0.4	0.2	12.3	12.0	12.0
10	100	0.0	0.0	0.0	20.1	19.9	20.0	1.9	0.5	0.2	13.2	12.9	13.4
10	250	0.1	0.1	0.1	17.4	17.4	17.2	2.2	0.8	0.5	11.9	11.6	11.4
50	10	0.0	0.0	0.0	25.1	15.2	10.4	0.0	0.0	0.0	14.4	12.7	10.4
50	50	0.0	0.0	0.0	31.2	27.5	18.8	0.0	0.0	0.0	19.2	19.3	18.2
50	100	0.0	0.0	0.0	35.6	31.8	27.6	0.0	0.0	0.0	18.9	19.1	19.2
50	250	0.0	0.0	0.0	35.2	33.4	32.6	0.0	0.0	0.0	18.8	19.3	18.9

Notes: Data is generated according to DGP2 and all tests are computed on detrended data. For further notes see Table 1.

Table 3: Empirical rejection frequencies, DGP3

		CS independence						SAR(1) model					
		ρ=1			$\rho = 1 - \frac{5}{T\sqrt{N}}$			ρ=1			$\rho = 1 - \frac{5}{T\sqrt{N}}$		
N	T	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*	*OLS*	*Rob*	*HS*
Constant variance													
10	10	7.5	1.8	6.0	24.9	24.6	20.1	22.7	3.3	4.4	16.1	14.7	13.0
10	50	7.0	5.0	6.6	36.9	36.9	36.8	24.6	7.5	7.6	19.4	16.6	15.9
10	100	7.1	6.2	6.9	42.0	41.3	41.4	24.4	7.6	7.8	22.2	18.4	18.7
10	250	7.2	6.7	7.0	55.9	55.7	55.4	24.3	8.4	8.6	27.3	21.5	20.9
50	10	14.5	0.0	12.2	27.3	27.6	19.6	28.2	0.2	6.5	16.6	16.1	14.9
50	50	6.8	1.6	6.6	54.9	54.1	53.1	22.6	4.3	6.9	25.1	24.4	23.9
50	100	6.5	2.9	6.5	68.3	68.6	67.5	21.2	5.5	7.0	33.4	30.1	30.0
50	250	5.9	4.4	6.1	83.9	84.3	83.5	21.3	5.8	6.2	43.6	40.3	40.7
Early negative variance shift													
10	10	5.5	0.5	7.9	8.0	8.3	6.2	17.5	2.0	6.4	7.6	7.1	6.2
10	50	4.3	2.7	6.2	10.5	10.6	9.1	18.6	5.5	6.4	8.2	7.2	6.5
10	100	4.4	3.3	7.0	15.6	16.0	14.1	19.2	6.5	7.1	9.5	8.5	9.0
10	250	4.9	4.4	6.9	35.0	34.4	31.7	19.8	7.5	8.1	17.4	12.6	12.5
50	10	9.4	0.0	12.4	4.1	5.6	1.9	22.6	0.0	8.3	5.4	6.0	3.8
50	50	5.0	0.1	9.4	13.9	15.1	12.4	19.0	1.5	7.0	9.6	9.2	9.4
50	100	4.0	0.4	6.9	45.0	45.4	43.1	17.5	2.8	6.7	22.4	19.5	19.0
50	250	3.1	1.3	5.6	91.1	89.6	90.0	16.9	3.4	7.2	45.8	39.7	38.3
Late positive variance shift													
10	10	8.1	1.2	4.9	13.6	14.2	13.4	20.8	2.6	3.9	12.0	12.8	10.5
10	50	14.7	11.7	6.8	22.7	23.2	20.7	33.0	12.3	6.7	15.4	13.4	12.8
10	100	14.1	12.4	6.5	27.3	26.3	24.6	33.2	13.8	7.6	17.2	14.5	13.6
10	250	14.1	13.7	7.4	30.6	30.2	26.0	34.6	14.7	8.1	18.3	14.9	13.8
50	10	10.7	0.0	6.7	13.5	13.6	13.1	23.6	0.2	5.4	11.3	11.1	9.3
50	50	18.8	6.8	8.0	28.6	28.2	28.4	33.9	10.8	7.8	16.0	14.7	13.6
50	100	16.0	9.2	7.1	35.2	33.5	33.3	31.1	11.5	7.0	19.4	17.5	16.6
50	250	14.2	11.5	6.8	39.1	37.4	37.6	31.3	12.9	6.8	21.0	19.3	17.9

Notes: Data is generated according to DGP3 and all tests are computed on prewhitened and centered data. For further notes see Table 1.

Table 4: Interest rates, definitions

Country	Label	Interest rate
Belgium	BEL	Treasury paper
Canada	CAN	Treasury Bill rate
France	FRA	Government Bond yield
Germany	GER	Call money rate
Italy	ITA	Government Bond yield medium-term
Japan	JAP	Lending rate
Netherlands	NED	Government Bond yield
United Kingdom	UKD	Treasury Bill rate
United States	USA	Treasury Bill Rate

Table 5: Empirical results

Variable	T	OLS	*Rob*	HS
π	172	-3.52	-2.45	-1.85
		(.000)	(.007)	(.032)
R	179	-4.22	-2.60	-1.67
		(.000)	(.005)	(.048)
r	176	-4.69	-3.49	-2.83
		(.000)	(.000)	(.002)

Notes: T denotes the number of included time series observation in the balanced panels. OLS, *Rob*, and HS refer to the PURT statistics defined in (4), (5),(6). Numbers in parentheses are $p-$values.