Unit Root Testing in Heteroskedastic Panels using the Cauchy Estimator*

Matei Demetrescu† Christoph Hanck‡

Preliminary version: June 25, 2010

Abstract

The so-called Cauchy estimator uses the sign of the first lag as instrument variable in autoregressions, and the resulting IV t-type statistic has a standard normal limiting distribution even in the unit root case. Thus, nonstandard asymptotics of the usual unit root tests such as the augmented Dickey-Fuller [ADF] test can be avoided. Moreover, the ADF test is affected by unconditional heteroskedasticity; but the paper shows that, by using as instrument some nonlinear transformation behaving asymptotically like the sign, limiting normality of the t-type statistic is maintained under unconditional heteroskedasticity when the series to be tested has no deterministic trends. Neither estimation of the so-called variance profile nor bootstrap procedures are required to this end, unlike for the ADF test. When adjusting the differences for deterministic components, however, the null distribution of the Cauchy test for a unit root becomes non-standard, reminiscent of the ADF test. In fact, the Cauchy test has power in the same $1/T$ neighborhoods as the ADF test, irrespective of whether a deterministic trend is present in the data or not. The standard normality of the examined Cauchy test can be exploited to build a panel unit root test under cross-sectional dependence with an orthogonalization procedure. The panel test does not require any N asymptotics to establish the limiting distribution, but the paper's analysis of the joint N, T asymptotics for the panel statistic suggests that N should be smaller than T. To render the test applicable when the number of cross-sectional units is larger than the number of time observations, shrinkage estimators of the involved covariance matrix are used. The performance of the discussed procedures is found to be satisfactory in finite samples. An empirical application to (a panel of) GDP prices illustrates the inferential impact of dealing with nonstationary volatility.

Keywords: Integrated process, Time-varying variance, Nonstationary volatility, Asymptotic normality, Cross-dependent panel, Joint asymptotics
JEL classification: C12 (Hypothesis Testing), C22 (Time-Series Models), C23 (Models with Panel Data)

*The authors would like to thank Jörg Breitung for helpful comments.

†Goethe University Frankfurt, Applied Econometrics, Grüneburgplatz 1 (Postfach RuW 49), D-60629 Frankfurt, Germany, e-mail: deme@wiwi.uni-frankfurt.de.

‡Rijksuniversiteit Groningen, Department of Economics, Econometrics and Finance, Nettelbosje 2, 9747AE Groningen, Netherlands, email: c.h.hanck@rug.nl.

1 Motivation

Instrumental variable [IV] estimation is typically used to deal with regressor endogeneity. But, provided that the instrument is suitably chosen, it has turned out to be a valuable tool in unit root econometrics as well. Focusing on nonlinear transformations of variables used as instruments for the very same variables, So and Shin (1999) or Phillips et al. (2004) have established interesting properties of the nonlinear IV estimation procedure. So and Shin deal with the so-called Cauchy estimator, where the sign of the first lag is used as an instrument; Phillips et al. examine several other types of transformations of the first lag: (regularly) integrable and asymptotically homogenous (with the sign being a function homogenous of order 0).

The t-type statistic based on the Cauchy estimator has a standard normal limiting distribution irrespective of the type of autoregressive root (stationary, unit, or explosive) in the series examined; this leads to a simple way of obtaining confidence intervals, as the pivotal standard normal distribution holds for stationary as well as nonstationary cases. See So and Shin (1999) for a discussion. Chang (2002) examines IV estimation where the instrument is a regularly integrable transformation. For unit, and stationary, roots, a standard normal distribution arises as well. For both tests, asymptotic normality is established under the null of a unit root by resorting to variants of martingale difference central limit theorems [CLT]; hence, the deterministic component in levels has to be adaptively (recursively) removed (So and Shin, 1999).[1] Using recursive adjustment is an advantage rather than a drawback: recursive removal of deterministic components has been proven to be power-effective when testing for unit roots (Leybourne et al., 2005).

Moreover, in panels exhibiting correlation across the N units with T observations each, a panel test can be obtained by an orthogonalization procedure (Shin and Kang, 2006). No N-asymptotics is required in establishing the limiting distribution of the panel unit root test. And the Cauchy test can be easily used in a nonlinear time series framework, following Shin and Lee (2001); in contrast, this is difficult to do for OLS estimation. In cross-correlated panels, Chang's (2002) individual test statistics are asymptotically independent.

This paper takes a closer look at the IV test based on the Cauchy estimator (or asymptotically equivalent choices of the instrument) in the unit root case. We are able to contribute to the literature on unit root and panel unit root testing in several important respects.

Our analysis in Section 2 will show the properties of the Cauchy unit root test to hinge on the type of deterministic component present in the data. As long as the deterministic component is not "too trending" (e.g. a constant non-zero mean), additional attractive properties hold: we establish here the Cauchy test's robustness to unconditional heteroskedasticity of unknown shape. In such cases, the augmented Dickey-Fuller [ADF] tests are affected even asymptotically, having

[1] See Demetrescu (2010) for a unified treatment of recursive adjustment for deterministic components of general form, as well as Taylor (2002) and Kuzin (2005) for dealing with seasonal deterministics.

asymptotic distributions expressed in terms of time-transformed, rather than standard, Wiener processes, and require either resampling (using the i.i.d./wild bootstrap), or the estimation of the so-called variance profile, as proposed by Cavaliere and Taylor (2007a,b). Moreover, the Cauchy test has local power in the same type of $1/T$ neighborhoods of the unity like the ADF test in spite of exhibiting standard normal asymptotics under the null. If the differenced series require, however, adjustment for deterministic components (as would be the case when the data exhibit a linear trend), the asymptotic null distribution is not standard normal anymore, but can be written as a functional of (time-transformed) Wiener or Ornstein-Uhlenbeck [OU] processes, similar to the ADF case.

In Section 3, we examine panel unit root testing based on the Cauchy test and establish standard normality of the orthogonalization procedure proposed by Shin and Kang (2006) under joint N, T-asymptotics. Accounting for unconditional heteroskedasticity is relevant for panel unit root tests just like it is for univariate tests: Hanck (2009) demonstrates that several popular second-generation panel unit root tests cease to work reliably under unconditional heteroskedasticity, with some of the tests exhibiting empirical size as high as 60%. Here, the robustness to unconditional heteroskedasticity is shown to carry over from the univariate to the panel test. The admissible rates for N, however, turn out to be slower than $T^{0.25}$, also because Shin and Kang's procedure requires orthogonalization with an estimated $N \times N$ covariance matrix. In any case the orthogonalization procedure induces the need of having larger T than N. The slight drawback of requiring N to be small compared to T can be easily overcome: we use shrinkage estimators of the covariance matrix such that the test works reliably for larger N.

Section 4 provides an application of the procedures for testing the stationarity of GDP prices, highlighting the differences in inferences provided by tests that are, or are not, robust to unconditional heteroskedasticity.

Before proceeding to the derivations, we establish some notation. Let boldface symbols denote column vectors and boldface capital symbols matrices, and denote by $\mathrm{diag}(a_1, \ldots, a_N) = \mathrm{diag}(a_i)$ a diagonal matrix with a_i on its main diagonal. Let $\|\cdot\|$ denote the Euclidean vector norm and the induced matrix norm, and $\|\cdot\|_r$ both the L_r vector norm, $\sqrt[r]{\sum |\cdot|^r}$, and the L_r norm of a random variable, $\sqrt[r]{\mathrm{E}(|\cdot|^r)}$, or vector. The probabilistic Landau symbols $O_p(\cdot)$ and $o_p(\cdot)$ have their usual meaning, and exact orders of magnitude are denoted by $\Theta_p(\cdot)$. Let $\mathbb{I}(\cdot)$ denote the indicator function, and C be a generic constant whose value may change from line to line.

2 The univariate unit root test

2.1 Model and assumptions

Let us begin by examining the univariate case. The data generating process [DGP] is given by

$$y_t = d_t + x_t, \quad t = 1, \dots, T,$$

i.e. the usual additive component representation. The deterministic component d_t is assumed to be known up to multiplicative constants (covering e.g. an intercept, a linear trend, or a break at known time). To ease the exposition, we consider only the two standard situations where d_t is a constant, or d_t is a linear trend plus constant. On the one hand, they are the most common in econometric practice, and, on the other hand, they point out the limitations of the Cauchy test in relation to the deterministic component: the test's behavior turns out to be quite different for the two types of deterministics.

The purely stochastic part of the model is given by the following assumption.

Assumption 1. *Let*

$$\Delta x_t = \phi x_{t-1} + \sum_{j=1}^{p} a_j \Delta x_{t-j} + \varepsilon_t,$$

with $x_0 = o_p(T^{0.5})$, stable roots of the lag polynomial $A(L) = 1 - \sum_{j=1}^{p} a_j L^j$, and white noise innovations ε_t.

The null hypothesis of a unit root is parameterized in the above error-correction representation by $\phi = 0$, and we test against stationary alternatives, $-2 < \phi < 0$.

The Cauchy test is based on IV estimation of the model using the sign of the lagged level as instrument for the lagged level and the untransformed Δy_{t-j}, $j = 1, \dots, p$, as instruments for themselves. If x_t were observed, a standard normal t-type statistic would result asymptotically under the null given finite-variance i.i.d. innovations, see So and Shin (1999). Actually, they use a slightly different version involving the IV regression of the prewhitened differences on the lagged level. It is equivalent to our version under the null and the local alternative (cf. the proof of Proposition 1), although not under a fixed alternative. But ours is the "textbook" IV estimation procedure, and has been extensively used in this form before; see e.g. Chang (2002) and Demetrescu (2009).

In the following subsection, we show that adjusting for deterministic components is the most important issue with IV tests of the examined type. Using the sign as an instrument, however, is not necessary; transformations of the lagged level behaving for integrated processes like the sign in the limit are allowed for; see Shin and Kang (2006) and Assumption 2 below. Still, for lack of a better denomination, we shall call the resulting IV tests "Cauchy tests" as well, even if they are only asymptotically equivalent to the test based on the sign instrument under the null, see the discussion following Assumption 2.

Denoting with tildes the adjusted levels and differences (see Subsection 2.2), the test regression becomes in the general form

$$\widetilde{\Delta y}_t = \widehat{\phi}\widetilde{y}_{t-1} + \sum_{j=1}^{p} \widehat{a}_j \widetilde{\Delta y}_{t-j} + \widehat{\varepsilon}_t,$$

estimated by IV with instrument $h\left(\widetilde{y}_{t-1}\right)$ for $\widetilde{y}_{t-1}$, where the so-called instrument generating function $h(\cdot)$ is specified in Assumption 2. The IV t-type statistic for testing the null of a unit root is given in the above error correction representation by

$$t_{IV} = \frac{\widehat{\phi} - 0}{s.e.(\widehat{\phi})},$$

with $\widehat{\phi}$ the IV estimator of ϕ,

$$\widehat{\phi} = \left(\sum_{t=p+2}^{T} \mathbf{v}_{t-1}\mathbf{x}'_{t-1} \right)^{-1} \sum_{t=p+2}^{T} \mathbf{v}_{t-1}\widetilde{\Delta y}_t,$$

and

$$s.e.(\widehat{\phi}) = \sqrt{\widehat{\sigma}^2 \left[\left(\sum_{t=p+2}^{T} \mathbf{x}_{t-1}\mathbf{v}'_{t-1} \right) \left(\sum_{t=p+2}^{T} \mathbf{v}_{t-1}\mathbf{v}'_{t-1} \right)^{-1} \left(\sum_{t=p+2}^{T} \mathbf{v}_{t-1}\mathbf{x}'_{t-1} \right) \right]^{-1}_{1,1}},$$

where $\widehat{\sigma}^2$ is the residual variance estimator, $\mathbf{x}_{t-1} = \left(\widetilde{y}_{t-1}, \widetilde{\Delta y}_{t-1}, \ldots, \widetilde{\Delta y}_{t-p}\right)'$ is the vector of regressors, and $\mathbf{v}_{t-1} = \left(h\left(\widetilde{y}_{t-1}\right), \widetilde{\Delta y}_{t-1}, \ldots, \widetilde{\Delta y}_{t-p}\right)'$ the vector of instruments. The index $1,1$ denotes the first diagonal element of a square matrix. In what concerns the instruments for the lagged level, we require the instrument generating function $h(\cdot)$ to belong to a particular subclass of functions asymptotically homogenous of order 0 (cf. Park and Phillips, 1999).

Assumption 2. *Let* $h(x) = g(x)\, I(|x| \leq m) + \operatorname{sgn}(x)\, I(|x| > m)$ *where* $g(x)$ *is odd and continuous,* $m \geq 0$ *is fixed and* $I(\cdot)$ *is the indicator function.*

For $g(x) = x$ and $m = 0$, $h(x) = \operatorname{sgn}(x)$. Phillips et al. (2004, Lemma 4.1) show that the sign function, and functions behaving asymptotically like it, enjoy certain asymptotic optimality properties in a class of bounded instrument generating functions. We can therefore expect tests based on t_{IV} to perform well. More generally for any $h(x)$, when x_t is integrated of order one, it becomes increasingly improbable as t grows that x_t takes values within $\pm m$ from the origin. Thus, the sign part dominates asymptotically; see also Lemma $4B$ in the Appendix. Letting $m \to \infty$ at small rates would likely not affect the argument, but there is little value added in pursuing this topic. With x_t integrated of order zero, however, the asymptotic equivalence is not given anymore. In the stationary case, $\text{sgn}(\widetilde{y}_{t-1})$ and $h(\widetilde{y}_{t-1})$ are different (although dependent) processes, and the small-sample performance of the Cauchy test under the alternative can be improved by judicious choice of g; see the Huber-type instruments approach of Shin and Kang (2006) and the section on small-sample behavior.

Extending the work of So and Shin (1999), we shall examine the behavior of the Cauchy test both under the null $\phi = 0$ and under a sequence of local alternatives in $1/T$ neighborhoods of the unit root. Local power is an important attribute of unit root tests, as it gives some indication regarding the behavior of the test when the alternative is true, but close to the null. In particular we shall prove that the Cauchy test has power in the same type of $1/T$-neighborhoods of the unity as the ADF test.

Assumption 3. *Let $\phi = -c/T$ with $c \geq 0$.*

Under standard assumptions for the innovations, e.g. i.i.d. sampling and a suitable moment condition, the normalized process x_t converges weakly to an OU process (see e.g. Phillips, 1987). Let J_c denote the standard OU process, the solution of the stochastic differential equation $dJ_c(s) = -cJ_c(s)\,ds + dW(s)$ where W is a standard Wiener process. Note that $J_0(s) = W(s)$.

Here, however, the innovations ε_t are allowed to be unconditionally heteroskedastic. Many potentially integrated time series exhibit such behavior. Following e.g. Cavaliere and Taylor (2007b), we require a multiplicative component structure for the innovations; but we relax their i.i.d. assumption to martingale differences [md] with weak moment conditions. We are actually closer to the setup of Cavaliere and Taylor (2009), but keep the deterministically varying variance for comparability with the earlier literature. Our results arguably hold under Cavaliere and Taylor's (2009) assumptions as well.

Assumption 4. *The innovations ε_t are variance-modulated, $\varepsilon_t = \sigma_t \epsilon_t$, such that*

1. *ϵ_t is an md sequence with $\mathrm{E}(\epsilon_t^2) = 1$ having uniformly bounded conditional (on $\{\epsilon_{t-1}, \epsilon_{t-2}, \ldots\}$) density functions such that $\exists r > 4$ with $\sup_t \|\epsilon_t\|_r < C < \infty$.*
2. *$\sigma_t = \omega(t/T)$ where $\omega(\cdot)$ is a bounded positive function on $[-\infty; 1]$, piecewise Lipschitz.*

Remark 1. Because σ_t depends on T, Assumption 4 implies that the series of innovations ε_t is actually a triangular array that should strictly speaking carry the additional index $\varepsilon_{t,T}$. For simplicity, we drop the T subscript henceforth.

Requiring the innovations ε_t to have no atoms (or poles of the density function) is not uncommon, see the literature on convergence to local time (used e.g. for Chang's 2002 IV panel unit root test); see the more recent work of Wang and Phillips (2009). It could most likely be relaxed here, as it is employed for examining some characteristics of cumulated sums in the neighborhood of the origin and we do not require a "full" weak convergence result. But the assumption simplifies the proofs so we stick to it. The moment condition is also standard in the unit root literature; see e.g. Chang and Park (2002).

Variance modulation has become a topic in the unit root and panel unit root testing literature; see the recent surge of contributions discussing unconditional heteroskedasticity (besides the authors mentioned above, Kim et al. (2002) and Hamori and Tokihisa (1997) contribute to this literature, among others). Under the DGP implied by Assumption 4, weak convergence still holds, but to a time-transformed OU process (the stochastic differential equation describing the limit becomes $dJ_c^\eta(s) = -cJ_c^\eta(s)\,ds + \omega(s)\,dW(s)$). In order to describe the solution, define the so-called variance profile,

$$\eta(s) = \left(\int_0^1 \omega^2(r)\,dr\right)^{-1} \int_0^s \omega^2(r)\,dr,$$

and let $\overline{\omega}^2 = \int_0^1 \omega^2(r)\,dr$. If $\omega(s)$ is constant, $\eta(s) = s$ and the standard case is recovered. The limiting behavior of the partial sums is described by the following lemma.

Lemma 1. *It holds under Assumptions 1, 3, and 4 that*

$$\frac{1}{\sqrt{T}} x_{[sT]} \Rightarrow \frac{\overline{\omega}}{A(1)} J_c(\eta(s))$$

as $T \to \infty$.

Proof: Along the lines of Cavaliere (2004, Lemma 3).

2.2 The detrending scheme

Before proceeding to the asymptotic analysis, the schemes to recursively adjust for deterministic components are examined in more detail, as they are essential to the asymptotics of the Cauchy test.

Under the null, the lagged levels require an adjustment scheme that does not affect the martingale difference property of the cross-product of instrument and innovation (see So and Shin, 1999, and the proofs in the Appendix); hence the use of recursive adjustment for the lagged level. Recursive adjustment implies OLS fitting of the deterministic component at time $t-1$ using the sample up to $t-1$. Shin and So (2001) analyze the ADF test with recursive demeaning; for recursive detrending see Taylor (2002) (in a seasonal framework) or Rodrigues (2006).

Adjusting the differences Δy_t for deterministic components poses quite some problems because of the way the adjustment influences the asymptotic properties of the Cauchy test. We shall argue that, if the trend component of the differences is weak enough (such as a non-zero mean in the levels, which is differenced away), the asymptotics are not affected when not accounting for the weak trend component. If not, one has to adjust (for which one can resort to usual OLS adjustment), and there is an asymptotic effect.

If only a constant is to be removed, $d_t = \mu$, the recursive scheme for the lagged levels becomes

$$\widetilde{y}^{\mu}_{t-1} = y_{t-1} - \frac{1}{t-1}\sum_{j=1}^{t-1} y_j = \widetilde{x}^{\mu}_{t-1}$$

The differences are not affected and do not require deterministic adjustment. The implications of recursive demeaning on the DGP are analyzed in the following lemma.

Lemma 2. *Define $\widetilde{J}^{\mu}_c(s)$ to be the recursively demeaned OU-Process, $\widetilde{J}^{\mu}_c(s) = J_c(s) - \frac{1}{s}\int_0^s J_c(s)dr$ with $\widetilde{J}^{\mu}_c(0) = 0$ a.s.; then*

$$\frac{1}{\sqrt{T}}\widetilde{x}^{\mu}_{[sT]} \Rightarrow \frac{\overline{\omega}}{A(1)}\widetilde{J}^{\mu}_c(\eta(s))$$

under the assumptions of Lemma 1.

Proof: Since the OU process has integrable paths, Proposition 2 in Demetrescu (2010) applies, leading to the desired result.

With no need to adjust the differences, the test regression becomes

$$\Delta y_t = \widehat{\phi}\widetilde{y}^{\mu}_{t-1} + \sum_{j=1}^{p}\widehat{a}_j \Delta y_{t-j} + \widehat{\varepsilon}_t, \tag{1}$$

estimated by IV with $h\left(\widetilde{y}^{\mu}_{t-1}\right)$ as an instrument for $\widetilde{y}^{\mu}_{t-1}$. Denote t^{μ}_{IV} the t-type statistic resulting from IV estimation of (1).

If, on the other hand, a linear trend is present in the data, the recursive scheme delivers for the lagged levels

$$\widetilde{y}^{\tau}_{t-1} = y_{t-1} + \frac{2}{t-1}\sum_{j=1}^{t-1} y_j - \frac{6}{t(t-1)}\sum_{j=1}^{t-1} j y_j = \widetilde{x}^{\tau}_{t-1};$$

The implications of recursive adjustment on the DGP are analyzed in the following lemma.

Lemma 3. *Define $\widetilde{J}^{\tau}_c(s)$ to be the recursively detrended OU-Process, $\widetilde{J}^{\tau}_c(s) = J_c(s) + \frac{2}{s}\int_0^s J_c(r)\,dr - \frac{6}{s^2}\int_0^s rJ_c(r)\,dr$ with $\widetilde{J}^{\tau}_c(0) = 0$ a.s.; then*

$$\frac{1}{\sqrt{T}}\widetilde{x}^{\tau}_{[sT]} \Rightarrow \frac{\overline{\omega}}{A(1)}\widetilde{J}^{\tau}_c(\eta(s))$$

under the assumptions of Lemma 1.

Proof: Analogous to the proof of Lemma 2.

The differences have a non-zero mean and do require adjustment (i.e. demeaning), see Proposition 3 below. So one has to work with

$$\widetilde{\Delta} y_{t-j} = \widehat{\phi}\widetilde{y}^{\tau}_{t-1} + \sum_{j=1}^{p}\widehat{a}_j \widetilde{\Delta} y_{t-j} + \widehat{\varepsilon}_t, \tag{2}$$

with $h\left(\widetilde{y}_{t-1}^{\tau}\right)$ as instrument for $\widetilde{y}_{t-1}^{\tau}$ and $\widetilde{\Delta y}_{t-j}$ as instruments for themselves, where $\widetilde{\Delta y}_{t-j}$ are suitably adjusted differences. Usual demeaning is good enough to this end, leading to

$$\widetilde{\Delta y}_t = \Delta y_t - \frac{1}{T}\sum_{t=2}^{T}\Delta y_t;$$

denote by t_{IV}^{τ} the resulting t-type statistic.

Another possibility would be to include the deterministic component in the test regression,

$$\Delta y_t = \widehat{m} + \widehat{\phi}\widetilde{y}_{t-1}^{\tau} + \sum_{j=1}^{p}\widehat{a}_j \Delta y_{t-j} + \widehat{\varepsilon}_t, \tag{3}$$

and use for testing the t-type statistic from instrumental variable estimation of the above equation, say $\bar{t}_{IV}^{\tau}$. This does make a difference, though not an essential one; see Proposition 4.

Recursive adjustment of the differenced series is not an option, since it leads to inconsistent filtering of the stochastic component at the beginning of the sample. See e.g. Demetrescu (2010) for details.

2.3 Asymptotic results

We examine the case with a constant mean first. Here, the t-type statistic behaves nicely, as summarized in the following proposition.

Proposition 1. *With $y_t = \mu + x_t$, it holds under Assumptions 1 through 4 that*

$$t_{IV}^{\mu} \xrightarrow{d} \int_0^1 \operatorname{sgn}\left(\widetilde{J}_c^{\mu}\left(\eta\left(s\right)\right)\right) dW\left(\eta\left(s\right)\right) - \frac{c}{A(1)}\int_0^1 \operatorname{sgn}\left(\widetilde{J}_c^{\mu}\left(\eta\left(s\right)\right)\right) J_c\left(\eta\left(s\right)\right) ds$$

as $T \to \infty$.

Proof: See the Appendix.

Corollary 1. *Under the null hypothesis $c = 0$, it holds that*

$$t_{IV}^{\mu} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof: Under the null $c = 0$, we have that

$$t_{IV}^{\mu} = \frac{\overline{\omega}}{\widehat{\sigma}T^{0.5}}\sum_{t=p+2}^{T}\frac{\operatorname{sgn}\left(\widetilde{x}_{t-1}^{\mu}\right)\varepsilon_t}{\overline{\omega}} + o_p(1),$$

to which we apply a suitable CLT for martingale difference arrays. Condition (a) of Theorem 24.3 of Davidson (1994) is fulfilled, see the discussion preceding Equation (6) in the proof of Proposition 1, and Condition (b) is easily established given finiteness of 4th order moments of ε_t. Since $\widehat{\sigma}$ converges to $\overline{\omega}$, see Equation (6) again, the result follows.

Remark 2. The intuition behind the corollary is that robustness to heteroskedasticity is obtained because the sign transformation discounts the large variability of the lagged level to either 1 or -1 irrespective of how the volatility changes in time. Remarkably, heteroskedasticity-consistent standard errors are not required.

Remark 3. From Proposition 1 and Assumption 3 we see that, in spite of standard asymptotics under H_0, t_{IV}^{μ} has local power in the same $1/T$ neighborhoods as the ADF test; a distribution that does depend on the variance profile emerges under the local alternative, as is the case of the ADF test too.

Remark 4. Furthermore, the limiting distribution highlights the nonstandard nature of the situation even for $c = 0$. In particular, it is shown in the proof that $\widehat{\phi}$ is superconsistent. See also Theorem 1(ii) in So and Shin (1999).

Examining the proof one finds that lag augmentation with $p \to \infty$ such that $p^{-1} + p/T^{\kappa} \to 0$ for some $\kappa \in \left(0, \min\{\frac{1}{2} - \frac{2}{r}; \frac{1}{4}\}\right)$ (including logarithmic rates) does not affect the asymptotic normality under the null (nor the distribution under the local alternative). Data dependent lag choice should work like in the ADF case; see the proof of the proposition, where the asymptotic covariance matrix of the estimators is shown to be lower triangular, as well as the subsection containing the Monte Carlo examination of the test's small-sample behavior. The finding can immediately be extended to the case where the short run component is a finite-order invertible $ARMA$ process and the $AR(p)$ process is only an approximation. It is not clear, however, whether the rate for p to which information criteria based on the IV residuals lead is still logarithmic. Generalizations for s-summable $AR(\infty)$ processes (see Chang and Park, 2002, for the ADF case) require e.g. a different proof of Lemma 4 E in the Appendix and is not pursued here.

The analogy to the locally best invariant [LBI] test for a unit root (see Tanaka, 1996), based on the squared difference between the last and the first observation, is striking. Up to the normalizing factor (which includes a suitable long-run variance estimator), the LBI test basically consists of the square of the cumulated innovations, while the IV test can be reduced under the null to the sum of the same innovations, but weighted with different signs (see the proof of Corollary 1). When squared, the IV test has the same asymptotic null distribution as the LBI test, and both are robust to unconditional heteroskedasticity under the null; see Cavaliere (2004) for the discussion of the LBI test.

However, the locally best invariant test and the IV test are not asymptotically equivalent under neither the null nor the local alternative, and the analysis is also different under a fixed alternative $\phi < 0$. For the IV test for instance, standard instrumental regression asymptotics apply under the alternative, leading to $\sqrt{T}$-consistent estimation of the parameter ϕ and thus to consistency of the IV unit root test. See the following proposition and also So and Shin (1999, Theorem 1(ii)).

Proposition 2. *With* $y_t = \mu + x_t$ *and* $-2 < \phi < 0$*, it holds under Assumptions 1, 2 and 4 that*

$$t_{IV}^{\mu} \xrightarrow{p} -\infty$$

as $T \to \infty$.

Proof: Obvious and omitted.

If a linear trend is indeed present in the data, it has to be dealt with, as pointed out by the following proposition.

Proposition 3. *With* $y_t = \mu + \tau\, t + x_t$ *and* $\tau \neq 0$*, it holds under Assumptions 1 through 4 that*

$$\left| t_{IV}^{\mu} \right| \xrightarrow{p} \infty$$

as $T \to \infty$.

Proof: See the Appendix.

Remark 5. The ADF test behaves nicer in this respect. Namely, the t-type statistic is standard normal if there is a neglected linear trend and the ADF regression includes a constant (West, 1988); this can be exploited to build union of rejections when one is not sure about the presence of a linear trend in the data (Harvey et al., 2009). The above Proposition gives incentive to rather detrend when one is not sure there about the nature of d_t.

Remark 6. The result of Proposition 3 holds as well if y_{t-1} is recursively detrended in (1): the critical issue is ignoring the non-zero mean of the differences, $\mathrm{E}(\Delta y_t) = \tau$. If, however, $T^{-0.5} \sum \mathrm{E}(\Delta y_t) \to 0$, not adjusting the differences and correctly adjusting the levels does lead to asymptotic normality.

Moving on to the analysis of the detrended test, note that the two choices for demeaning the differences mentioned in the previous subsection are slightly different in the resulting distributions (see Proposition 4 below), but not in their implications for the asymptotic behavior: when $c = 0$, limiting standard normality is not given for either of the two.

Proposition 4. *With* $y_t = \mu + \tau\, t + x_t$*, it holds under Assumptions 1 through 4 that*

$$\begin{aligned} t_{IV}^{\tau} \xrightarrow{d} & \int_0^1 \operatorname{sgn}\left(\widetilde{J}_c^{\tau}(\eta(s))\right) dW(\eta(s)) - \frac{c}{A(1)} \int_0^1 \operatorname{sgn}\left(\widetilde{J}_c^{\tau}(\eta(s))\right) J_c(\eta(s))\, ds \\ & - \left(W(1) \int_0^1 \operatorname{sgn}\left(\widetilde{J}_c^{\tau}(\eta(s))\right) ds - \frac{c}{A(1)} \int_0^1 \operatorname{sgn}\left(\widetilde{J}_c^{\tau}(\eta(s))\right) ds \int_0^1 J_c(\eta(s))\, ds \right) \end{aligned}$$

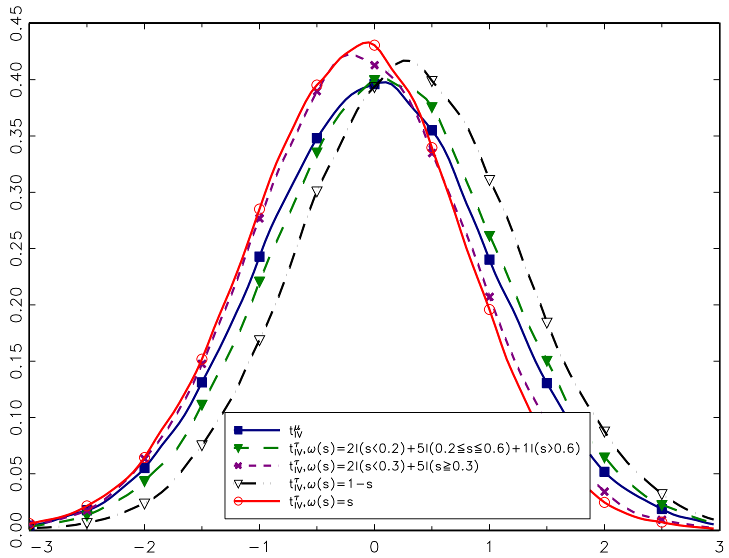


Figure 1: Null densities of t^{μ}_{IV} and t^{τ}_{IV} for different $\omega(s)$

and

$$\bar{t}^{\tau}_{IV} \xrightarrow{d} \frac{t^{\tau}_{IV}}{\sqrt{1 - \left(\int_0^1 \operatorname{sgn}\left(\widetilde{J}^{\tau}_c\left(\eta\left(s\right)\right)\right) ds \right)^2}}$$

as $T \to \infty$, *irrespective of whether* $\tau \neq 0$ *or not.*

Proof: See the Appendix.

Remark 7. The result analogous to Proposition 2 holds as well, guaranteeing consistency under a fixed alternative.

In the detrending case, the Cauchy estimator looses its good properties under the null of a unit root even when there is actually no linear trend in the data. The problem is that demeaning differences induces a component of order $O_p\left(T^{-0.5}\right)$ which affects the asymptotic distribution. Figure 1 illustrates the effect through plotting the (kernel-density estimated) simulated densities (obtained from 100,000 draws of the corresponding functionals) of t^{μ}_{IV} and t^{τ}_{IV} for different $\omega(s)$. As predicted from Proposition 1, the density of t^{μ}_{IV} is that of the standard normal. The densities of t^{τ}_{IV} are however clearly not standard normal. The 5%-quantiles of the distributions of t^{τ}_{IV} are in a range $[-1.3, -1.73]$, such that using standard normal critical values would lead to somewhat size-distorted tests. In particular, one cannot argue that standard normal critical values would produce liberal or conservative tests uniformly for all $\omega(s)$. For instance, using standard normal critical values yields undersized tests for the case of negatively trending variances ($\omega(s) = 1 - s$) and oversized tests for positively trending variances ($\omega(s) = s$).

Other schemes for demeaning the differences can be used, of course (e.g. GLS demeaning). The above proposition suggests, however, that the asymptotic distributions will change accordingly. We are not aware of a demeaning scheme for the differences that would change the finding in the sense that the scheme would lead to asymptotic normality of t^{τ}_{IV} for $c = 0$: any $\sqrt{T}$-consistent demeaning scheme will lead to a nonstandard distribution, which is slightly disappointing considering the nice behavior under demeaning.

Remark 8. Chang (2002) does not face this problem. From her work it can be seen that square-root consistent estimation of the mean of the differences does not affect the asymptotics: essentially, her choice of an instrument leads to standard normality through a martingale difference CLT requiring normalization lower than $\sqrt{T}$.

2.4 Monte Carlo examinations

This section investigates the size and power of the Cauchy unit root test with demeaning only, with a special emphasis on the robustness to unconditional heteroskedasticity.

Following Cavaliere and Taylor (2008), we use the following simple DGP:

$$\begin{aligned} y_t &= \mu + x_t \\ x_t &= \rho x_{t-1} + u_t \qquad t = 1, \ldots, T \end{aligned}$$

To gauge the effect of serial correlation, we let u_t be an MA(1) process, obtained by applying the filter $\psi(L) = 1 + \psi L$ to the possibly heteroskedastic innovations ε_t, where $\psi \in \{-0.5, 0, 0.5\}$. We alternatively also consider the AR(1) case $u_t = \eta u_{t-1} + \varepsilon_t$.

To introduce nonstationary volatility into the DGP, we generate a permanent break in the innovation variance of standard normal variates ε_t at time $\lfloor \tau T \rfloor$, where $\text{Var}(\varepsilon_t) = 1$ for $t = 1, \ldots, \lfloor \tau T \rfloor$ and $\text{Var}(\varepsilon_t) = 1/\delta^2$ for $t = \lfloor \tau T \rfloor + 1, \ldots, T$. We consider $\tau \in \{0.1, 0.5, 0.9\}$, corresponding to early, middle and late breaks (such a design ensures that neither regime dominates asymptotically), and $\delta \in \{1/5, 1, 5\}$ to generate positive ($\delta = 1/5$) and negative ($\delta = 5$) breaks, respectively. The case $\delta = 1$ covers the benchmark homoskedastic case.

In all simulations a constant is removed as described above. Since all tests considered are then invariant to the value of μ we set $\mu = 0$. The instrument generating function (cf. Assumption 2) is specified as $m = 1$ and $g(x) = x$; following Shin and Kang (2006), $\widetilde{y}^{\mu}_{t-1}$ is standardized by $\widehat{\sigma}$ to make the choice of m less dependent on the volatility of the series.[2] When $\psi \neq 0$ we choose the number of lagged differences p using Akaike's criterion. To study size, we let $\rho = 1$ corresponding to $\phi = 0$. In the power experiments we take $\rho = 0.8$.

To gauge the effectiveness of the Cauchy test under nonstationary volatility, we compare it to Cavaliere and Taylor's (2008) recent wild bootstrap version of the $\mathcal{M}$ tests of (Ng and Perron, 2001). We refer to Cavaliere and Taylor (2008) for a detailed description of their approach.

[2]Experimentation with other choices for m yielded results slightly inferior to those to be reported below.

Table 1: Size of the Cauchy and bootstrap $\mathcal{M}$ tests

		$\tau = 0.1$					$\tau = 0.5$					$\tau = 0.9$				
	T	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
$\psi = -0.5$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.144	.081	.087	.071	.063	.130	.081	.074	.069	.058	.139	.095	.070	.067	.050
	$\mathcal{MSB}^b$	.133	.076	.082	.071	.059	.110	.076	.068	.066	.057	.122	.093	.079	.067	.050
	$\mathcal{MZ}_t^b$	.147	.083	.087	.070	.063	.133	.082	.076	.069	.058	.135	.088	.068	.068	.049
	t_{IV}^μ	.025	.027	.044	.042	.051	.022	.027	.041	.053	.056	.020	.028	.023	.016	.023
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.135	.072	.072	.077	.059	.126	.084	.076	.077	.063	.135	.080	.075	.075	.064
	$\mathcal{MSB}^b$	.122	.068	.072	.075	.055	.118	.076	.072	.074	.062	.122	.077	.073	.073	.064
	$\mathcal{MZ}_t^b$	.138	.075	.073	.077	.059	.129	.083	.074	.078	.063	.138	.079	.076	.075	.063
	t_{IV}^μ	.022	.024	.040	.043	.044	.028	.028	.034	.050	.045	.026	.028	.038	.049	.048
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.082	.028	.042	.047	.037	.109	.058	.055	.061	.052	.128	.079	.075	.079	.064
	$\mathcal{MSB}^b$	.077	.027	.039	.045	.035	.104	.057	.053	.061	.052	.125	.077	.072	.076	.063
	$\mathcal{MZ}_t^b$	.084	.029	.043	.048	.037	.110	.060	.055	.062	.054	.131	.082	.076	.081	.064
	t_{IV}^μ	.028	.046	.054	.042	.017	.013	.010	.016	.019	.031	.027	.029	.034	.045	.048
$\psi = 0$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.060	.052	.052	.053	.050	.056	.057	.054	.051	.053	.062	.057	.058	.060	.055
	$\mathcal{MSB}^b$	.058	.046	.051	.051	.049	.052	.057	.053	.055	.053	.037	.042	.050	.058	.053
	$\mathcal{MZ}_t^b$	.061	.054	.052	.051	.052	.056	.061	.055	.051	.054	.073	.061	.060	.062	.058
	t_{IV}^μ	.072	.067	.067	.065	.056	.111	.105	.092	.080	.086	.092	.085	.073	.076	.073
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.056	.053	.050	.049	.049	.051	.056	.051	.051	.054	.059	.058	.053	.046	.050
	$\mathcal{MSB}^b$	.052	.051	.050	.047	.047	.052	.056	.050	.056	.052	.056	.058	.053	.045	.049
	$\mathcal{MZ}_t^b$	.056	.053	.051	.048	.048	.052	.056	.050	.049	.052	.061	.059	.052	.047	.050
	t_{IV}^μ	.059	.056	.057	.052	.050	.057	.062	.058	.061	.053	.065	.064	.060	.053	.059
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.072	.066	.058	.053	.054	.062	.054	.055	.052	.049	.057	.050	.052	.049	.050
	$\mathcal{MSB}^b$	.069	.066	.058	.055	.054	.061	.055	.056	.052	.048	.057	.049	.050	.048	.047
	$\mathcal{MZ}_t^b$	.072	.067	.059	.054	.055	.062	.056	.056	.051	.048	.058	.049	.053	.047	.051
	t_{IV}^μ	.044	.054	.047	.047	.057	.062	.061	.060	.057	.053	.060	.053	.055	.056	.055
$\psi = 0.5$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.050	.060	.059	.055	.055	.061	.065	.071	.067	.056	.190	.173	.110	.097	.071
	$\mathcal{MSB}^b$	.049	.062	.057	.056	.054	.062	.065	.070	.066	.054	.190	.182	.118	.099	.074
	$\mathcal{MZ}_t^b$	.046	.060	.058	.057	.054	.056	.064	.068	.065	.057	.182	.163	.105	.092	.069
	t_{IV}^μ	.006	.014	.032	.035	.042	.008	.018	.026	.033	.038	.011	.014	.017	.018	.016
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.048	.063	.056	.052	.052	.040	.062	.061	.052	.048	.045	.056	.056	.056	.051
	$\mathcal{MSB}^b$	.050	.065	.054	.050	.053	.043	.060	.060	.054	.047	.043	.056	.053	.055	.051
	$\mathcal{MZ}_t^b$	.046	.061	.057	.050	.052	.039	.062	.060	.052	.048	.043	.060	.056	.055	.051
	t_{IV}^μ	.007	.017	.025	.026	.037	.006	.015	.024	.031	.036	.007	.019	.029	.029	.034
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.019	.018	.028	.036	.030	.016	.032	.054	.057	.041	.030	.049	.053	.055	.050
	$\mathcal{MSB}^b$	.021	.019	.027	.035	.030	.015	.030	.054	.057	.043	.033	.050	.054	.054	.051
	$\mathcal{MZ}_t^b$	.019	.017	.027	.035	.031	.017	.033	.054	.057	.043	.030	.049	.052	.056	.049
	t_{IV}^μ	.011	.014	.017	.012	.015	.009	.004	.011	.015	.018	.004	.014	.024	.030	.034

Nominal 5% level. 5000 replications, 500 bootstrap replications for the $\mathcal{M}$ tests. ψ defines an MA(1) error term process for the errors ε_t.

Table 2: Size of the Cauchy and bootstrap $\mathcal{M}$ tests

		$\tau = 0.1$					$\tau = 0.5$					$\tau = 0.9$				
	T	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
$\eta = -0.5$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.044	.032	.033	.033	.035	.048	.032	.033	.030	.034	.101	.092	.051	.051	.051
	$\mathcal{MSB}^b$	.038	.029	.031	.033	.035	.043	.030	.029	.030	.035	.098	.099	.058	.051	.049
	$\mathcal{MZ}_t^b$	.044	.032	.035	.035	.035	.049	.034	.034	.031	.035	.095	.084	.050	.049	.047
	t_{IV}^μ	.051	.040	.046	.057	.055	.054	.067	.069	.077	.076	.054	.046	.036	.035	.039
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.035	.027	.036	.037	.045	.031	.029	.037	.040	.040	.036	.028	.033	.038	.042
	$\mathcal{MSB}^b$	.032	.025	.033	.038	.044	.026	.026	.034	.040	.040	.033	.025	.031	.034	.041
	$\mathcal{MZ}_t^b$	.035	.030	.037	.037	.042	.035	.030	.039	.042	.041	.037	.029	.033	.041	.043
	t_{IV}^μ	.048	.041	.036	.047	.047	.043	.045	.043	.047	.051	.042	.038	.038	.044	.050
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.010	.011	.022	.020	.029	.032	.024	.034	.030	.034	.036	.025	.032	.040	.038
	$\mathcal{MSB}^b$	.010	.010	.020	.022	.027	.029	.023	.032	.030	.034	.032	.021	.032	.040	.036
	$\mathcal{MZ}_t^b$	.011	.013	.022	.021	.029	.034	.024	.034	.030	.035	.037	.027	.033	.040	.038
	t_{IV}^μ	.061	.043	.033	.037	.023	.034	.038	.040	.039	.042	.045	.038	.043	.050	.046
$\eta = 0.5$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.093	.071	.055	.053	.046	.117	.086	.061	.056	.057	.252	.211	.123	.092	.078
	$\mathcal{MSB}^b$	.094	.073	.055	.055	.050	.120	.086	.062	.058	.055	.242	.213	.131	.102	.082
	$\mathcal{MZ}_t^b$	.088	.068	.055	.052	.045	.114	.081	.060	.054	.055	.246	.202	.118	.086	.076
	t_{IV}^μ	.017	.032	.047	.050	.052	.023	.037	.055	.059	.054	.022	.028	.030	.027	.028
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.079	.072	.058	.051	.060	.083	.071	.054	.054	.051	.084	.070	.050	.056	.054
	$\mathcal{MSB}^b$	.082	.073	.056	.051	.059	.085	.077	.054	.052	.053	.084	.068	.049	.056	.054
	$\mathcal{MZ}_t^b$	.073	.070	.060	.049	.058	.078	.068	.056	.056	.051	.081	.069	.049	.054	.053
	t_{IV}^μ	.016	.033	.045	.050	.052	.016	.036	.040	.047	.055	.015	.032	.042	.047	.042
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.028	.014	.019	.019	.025	.023	.034	.039	.041	.042	.060	.050	.050	.049	.049
	$\mathcal{MSB}^b$	.029	.015	.019	.019	.026	.022	.034	.038	.041	.042	.066	.052	.051	.050	.049
	$\mathcal{MZ}_t^b$	.027	.013	.019	.021	.026	.023	.035	.039	.041	.043	.056	.050	.048	.048	.049
	t_{IV}^μ	.013	.015	.020	.023	.019	.007	.011	.028	.034	.042	.012	.028	.040	.044	.043

Nominal 5% level. 5000 replications, 500 bootstrap replications for the $\mathcal{M}$ tests. η defines an AR(1) error term process for the errors ε_t.

Table 1 reports the size of the tests for the case of MA error terms. We see that both the Cauchy test t_{IV}^μ and the bootstrap $\mathcal{M}$ tests control size very well for T sufficiently large and any pattern of variance break and serial dependence. Further, t_{IV}^μ is level α throughout. It can however be rather undersized for small T, e.g. for $\psi = 0.5$, $\delta = 1/5$ and $\tau = 0.1$. On the other hand, the bootstrap $\mathcal{M}$ tests can be severely oversized for small T, with the empirical size sometimes almost four times the nominal one for $\psi = 0.5$, $\delta = 1/5$ and $\tau = 0.9$. In either case, the distortions vanish as $T \to \infty$. All tests are also capable of handling the baseline case of no heteroskedasticity, $\delta = 1$, although a larger T is required for the bootstrap tests. Table 2 reports analogous results for an AR(1) error process. Again, the bootstrap tests are mostly oversized for small T. For $\eta = 0.5$, $\delta = 1/5$ and $\tau = 0.9$, the bootstrap tests have empirical size five times the nominal one. The Cauchy test can again be somewhat undersized. All in all, we believe it is consistent with most analysts' loss functions to argue that t_{IV}^μ offers an improvement in small

Table 3: Power of the Cauchy and bootstrap $\mathcal{M}$ tests

		$\tau = 0.1$					$\tau = 0.5$					$\tau = 0.9$				
	T	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
$\psi = -0.5$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.498	.598	.947	.998	.998	.354	.402	.756	.936	.956	.276	.242	.421	.619	.632
	$\mathcal{MSB}^b$	.455	.562	.939	.999	.998	.301	.362	.727	.927	.952	.214	.220	.415	.624	.639
	$\mathcal{MZ}_t^b$	.510	.606	.947	.997	.998	.363	.409	.758	.933	.954	.250	.238	.403	.601	.619
	t_{IV}^μ	.046	.087	.207	.322	.411	.038	.065	.137	.199	.279	.030	.041	.073	.098	.145
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.362	.399	.680	.825	.813	.363	.415	.695	.828	.812	.361	.393	.689	.822	.819
	$\mathcal{MSB}^b$	.333	.372	.665	.812	.799	.332	.382	.676	.812	.802	.327	.359	.663	.806	.806
	$\mathcal{MZ}_t^b$	.370	.409	.688	.829	.820	.379	.421	.695	.831	.814	.372	.404	.696	.828	.821
	t_{IV}^μ	.036	.071	.195	.296	.397	.037	.079	.197	.306	.395	.036	.074	.207	.301	.386
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.111	.061	.168	.257	.253	.206	.212	.446	.597	.588	.373	.416	.690	.811	.803
	$\mathcal{MSB}^b$	.103	.055	.160	.250	.247	.199	.204	.440	.592	.584	.343	.391	.673	.802	.798
	$\mathcal{MZ}_t^b$	.116	.065	.177	.262	.260	.215	.220	.451	.599	.592	.390	.431	.704	.817	.807
	t_{IV}^μ	.025	.061	.173	.236	.198	.023	.033	.077	.132	.204	.040	.082	.205	.319	.390
$\psi = 0$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.338	.669	.990	1.00	1.00	.238	.437	.879	.992	1.00	.233	.408	.714	.874	.954
	$\mathcal{MSB}^b$	.289	.615	.984	1.00	1.00	.198	.391	.853	.988	.999	.132	.344	.715	.887	.964
	$\mathcal{MZ}_t^b$	.347	.674	.988	1.00	1.00	.241	.440	.873	.989	.999	.239	.386	.662	.829	.927
	t_{IV}^μ	.363	.611	.938	.994	1.00	.385	.572	.886	.981	.998	.284	.421	.696	.889	.972
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.273	.550	.905	.964	.988	.257	.529	.891	.957	.979	.273	.533	.889	.956	.978
	$\mathcal{MSB}^b$	.231	.505	.893	.960	.984	.226	.490	.871	.951	.974	.235	.480	.872	.952	.977
	$\mathcal{MZ}_t^b$	.275	.555	.906	.964	.986	.263	.538	.893	.957	.978	.280	.537	.889	.956	.978
	t_{IV}^μ	.270	.510	.912	.995	1.00	.261	.494	.905	.992	1.00	.276	.503	.915	.994	1.00
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.115	.203	.366	.471	.554	.156	.313	.664	.793	.869	.268	.530	.881	.948	.970
	$\mathcal{MSB}^b$	.100	.186	.356	.460	.547	.153	.305	.657	.793	.869	.239	.492	.871	.946	.970
	$\mathcal{MZ}_t^b$	.122	.210	.372	.477	.557	.160	.320	.661	.793	.866	.279	.550	.886	.949	.971
	t_{IV}^μ	.069	.151	.397	.630	.792	.163	.318	.691	.905	.982	.269	.507	.908	.990	1.00
$\psi = 0.5$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.076	.290	.889	.995	.999	.071	.150	.593	.893	.953	.172	.165	.365	.611	.683
	$\mathcal{MSB}^b$	.066	.245	.869	.991	.999	.058	.130	.549	.873	.950	.173	.171	.356	.615	.697
	$\mathcal{MZ}_t^b$	.082	.301	.890	.994	.999	.074	.159	.596	.892	.951	.167	.155	.355	.581	.658
	t_{IV}^μ	.024	.085	.292	.492	.599	.023	.067	.161	.268	.376	.016	.032	.042	.056	.098
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.072	.214	.707	.868	.902	.070	.210	.693	.848	.869	.065	.218	.684	.842	.874
	$\mathcal{MSB}^b$	.059	.192	.681	.857	.895	.057	.187	.670	.834	.861	.051	.190	.659	.830	.864
	$\mathcal{MZ}_t^b$	.077	.225	.709	.873	.903	.076	.227	.697	.849	.870	.067	.222	.694	.842	.873
	t_{IV}^μ	.021	.072	.251	.454	.583	.014	.067	.242	.432	.582	.016	.060	.253	.450	.574
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.013	.047	.182	.300	.311	.044	.124	.420	.618	.658	.055	.209	.681	.837	.864
	$\mathcal{MSB}^b$	.012	.041	.170	.289	.304	.040	.119	.409	.615	.655	.048	.185	.656	.827	.857
	$\mathcal{MZ}_t^b$	.013	.049	.189	.307	.318	.046	.131	.426	.620	.657	.060	.220	.695	.840	.867
	t_{IV}^μ	.011	.029	.108	.177	.157	.011	.017	.070	.149	.223	.020	.065	.267	.439	.573

Nominal 5% level. 5000 replications, 500 bootstrap replications for the $\mathcal{M}$ tests. ψ defines an MA(1) error term process for the errors ε_t.

Table 4: Power of the Cauchy and bootstrap $\mathcal{M}$ tests

		$\tau = 0.1$					$\tau = 0.5$					$\tau = 0.9$				
	T	30	50	100	150	200	30	50	100	150	200	30	50	100	150	200
$\eta = -0.5$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.224	.367	.869	.983	.999	.169	.211	.555	.797	.937	.168	.152	.308	.506	.593
	$\mathcal{MSB}^b$	.194	.327	.850	.978	.999	.140	.181	.516	.773	.929	.138	.153	.296	.516	.600
	$\mathcal{MZ}_t^b$	.234	.387	.872	.980	.998	.174	.216	.562	.796	.933	.159	.148	.296	.483	.569
	t_{IV}^μ	.132	.309	.780	.946	.978	.133	.267	.613	.809	.892	.096	.128	.215	.321	.391
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.145	.230	.647	.773	.847	.140	.228	.639	.767	.838	.138	.220	.627	.800	.841
	$\mathcal{MSB}^b$	.121	.199	.621	.757	.835	.124	.197	.612	.746	.824	.118	.194	.598	.784	.831
	$\mathcal{MZ}_t^b$	.155	.239	.654	.776	.851	.146	.239	.646	.769	.841	.146	.232	.632	.803	.842
	t_{IV}^μ	.108	.231	.736	.924	.980	.103	.228	.726	.937	.976	.095	.219	.717	.927	.976
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.030	.043	.136	.192	.273	.082	.102	.335	.494	.610	.145	.234	.636	.794	.833
	$\mathcal{MSB}^b$	.025	.038	.127	.183	.264	.078	.094	.325	.488	.607	.126	.208	.609	.782	.827
	$\mathcal{MZ}_t^b$	.034	.045	.142	.198	.280	.086	.109	.343	.498	.614	.157	.252	.652	.801	.835
	t_{IV}^μ	.058	.095	.318	.433	.399	.065	.121	.381	.647	.799	.105	.234	.722	.923	.973
$\eta = 0.5$																
$\delta = 1/5$	$\mathcal{MZ}_\alpha^b$	.110	.264	.829	.976	.999	.101	.154	.523	.793	.939	.207	.184	.332	.572	.663
	$\mathcal{MSB}^b$	.095	.234	.799	.973	.998	.094	.138	.494	.778	.931	.200	.201	.334	.575	.671
	$\mathcal{MZ}_t^b$	.113	.273	.827	.973	.998	.102	.157	.527	.792	.936	.199	.173	.327	.545	.633
	t_{IV}^μ	.073	.239	.652	.888	.976	.060	.169	.490	.713	.856	.035	.059	.139	.239	.328
$\delta = 1$	$\mathcal{MZ}_\alpha^b$	.101	.241	.701	.848	.923	.097	.212	.666	.814	.875	.089	.221	.660	.845	.887
	$\mathcal{MSB}^b$	.092	.222	.675	.829	.912	.089	.191	.639	.796	.868	.079	.204	.633	.826	.877
	$\mathcal{MZ}_t^b$	.104	.247	.705	.851	.925	.098	.218	.666	.816	.877	.094	.228	.669	.846	.887
	t_{IV}^μ	.051	.181	.586	.848	.957	.053	.168	.580	.854	.958	.048	.165	.568	.839	.954
$\delta = 5$	$\mathcal{MZ}_\alpha^b$	.024	.054	.167	.250	.347	.059	.119	.388	.549	.669	.078	.219	.655	.832	.867
	$\mathcal{MSB}^b$	.021	.050	.157	.239	.338	.057	.115	.379	.545	.666	.071	.201	.628	.824	.862
	$\mathcal{MZ}_t^b$	.027	.060	.176	.254	.350	.059	.123	.388	.552	.668	.083	.230	.671	.836	.870
	t_{IV}^μ	.015	.044	.203	.301	.295	.013	.061	.252	.489	.702	.047	.164	.571	.827	.949

Nominal 5% level. 5000 replications, 500 bootstrap replications for the $\mathcal{M}$ tests. η defines an AR(1) error term process for the errors ε_t.

sample size over the bootstrap $\mathcal{M}$ tests.

Tables 3 and 4 report power of the tests for the MA and AR error cases. Given that size-adjusted critical values are not available in practice we do not report size-adjusted power (Horowitz and Savin, 2000). All tests considered are consistent in that power tends to one as $T \to \infty$. Prima facie, the bootstrap $\mathcal{M}$ tests mostly appear to be much more powerful. For instance, for $T = 200$, $\psi = -0.5$, $\delta = 1/5$ and $\tau = 0.1$, the bootstrap $\mathcal{M}$ tests have power almost equal to one whereas t_{IV}^μ only achieves a power of 0.41. However, the examination of cases where all tests control the nominal size reveals that these power gains are entirely driven by the size distortion. For instance, for $\psi = 0$, $\delta = 1$ and $\tau = 0.1$ all tests have empirical size close to 5% and nearly identical power. The choice of one of the tests in a given application could therefore depend on a specific loss function attaching weights to false rejections and acceptances. Other considerations, such as the computational ease of t_{IV}^μ relative to the bootstrap tests may also play a role.

3 Panel unit root tests

For the panel analysis, we shall focus on the unit root case with no deterministic trends, as it is the one with normality and heteroskedasticity-robustness. Otherwise, all problems due to nonstandard distribution appear, and in particular lack of invariance to the variance profile. Hence, in the detrending case, one rather ought to use the ADF test with some fix along the lines of Cavaliere and Taylor (2007a,b), a topic left for further research.

3.1 Model and assumptions

Let $y_{i,t}$ be the observed series, generated as

$$y_{i,t} = \mu_i + x_{i,t}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T,$$

The stochastic component $x_{i,t}$ is generated unit-wise by an autoregressive process of order $p_i + 1$ with a possible unit root,

$$(1 - \rho_i L)\, A_i\,(L)\, x_{i,t} = \varepsilon_{i,t}$$

where the p_i characteristic roots of the polynomials A_i all belong to the stability region. The model written in error correction form is

$$\Delta x_{i,t} = \phi_i x_{i,t-1} + \sum_{j=1}^{p} a_{ij} \Delta x_{i,t-j} + \varepsilon_{i,t}, \tag{4}$$

with $\phi_i = \rho_i - 1$. Under the null of a unit root it holds $\rho_i = 1$ or $\phi_i = 0$.

Throughout the section, we shall assume the DGP and the instruments to satisfy the univariate assumptions individually. The lag orders, for instance, may be heterogenous, but we assume the maximal lag order to be finite (and set the "missing" autoregressive parameters in units with lower actual order to zero).

Assumption 5. *The unit-specific dynamics each satisfy Assumption 1 with* $\sup_i p_i < p$, $i = 1, \ldots, N$ *and the unit-specific instrument generating functions each satisfy Assumption 2 with* $\sup_i m_i < M$ *and* $\sup_{|x| \leq m_i} g_i(x) < G$, $i = 1, \ldots, N$, *for some* p, G *and* M *not depending on* T *or* N.

Hanck (2009) demonstrates that many popular second-generation panel unit root tests (e.g., Pesaran, 2007; Breitung and Das, 2005; Demetrescu et al., 2006; Moon and Perron, 2004) fail under unconditional heteroskedasticity. Specifically, some tests attain an empirical size of over 60% at a nominal 5% level. As such, they cannot be recommended for empirical application whenever e.g. variance breaks are a relevant concern. On the contrary, the Cauchy test's univariate robustness to unconditional heteroskedasticity prevents such failure in the panel case as well. Now, the test suggested in Demetrescu et al. (2006) is based on combining unit-specific evidence against unit roots from individual ADF tests, and fails because of the ADF test's lack of robustness to unconditional heteroskedasticity; when replacing individual ADF tests with individual Cauchy tests, the test works more reliably, see below.

Under cross-sectional independence, a panel test statistic can be constructed immediately based on the single-unit Cauchy tests due to their standard asymptotics. The simplest statistic is obtained by summing the individual ones and dividing by $\sqrt{N}$, leading to a standard normal panel test statistic. This holds true when allowing for $N \to \infty$; but $N \to \infty$ is not a necessary condition for normality. As a peculiarity of the IV estimation procedure, it makes no difference asymptotically if one assumes homogenous autoregressive roots across the panel and uses its IV t-type statistic, or if one uses the averaged t-type statistics from N individual Cauchy tests (i.e. allowing explicitly for heterogeneity of the autoregressive roots). For panel unit root tests based on ADF statistics, see Levin et al. (2002) and Im et al. (2003), assuming homogeneity when building the test statistic does make a difference in terms of local power; see Westerlund and Breitung (2009). In the ADF case, the denominators of the two panel test statistics are different, whereas in the IV case they are asymptotically the same, as can be easily checked.[3]

Under cross-correlation, the Cauchy panel unit root test requires orthogonalization, since the individual test statistics are correlated. See Shin and Kang (2006), who conduct their analysis under a fixed-N assumption. They propose several test statistics, all ultimately based on the joint distribution of the statistics $\hat{\tau}_{i,IV}$ resulting from an IV regression of the prewhitened and orthogonalized differences on the lagged levels. Concretely, let

$$\overline{\varepsilon}_{i,t} = \Delta y_{i,t} - \sum_{j=1}^{p} \overline{a}_{ij} \Delta y_{i,t-j}$$

and $\overline{\boldsymbol{\varepsilon}}_t = (\overline{\varepsilon}_{1,t}, \ldots, \overline{\varepsilon}_{N,t})'$ be the prewhitened differences; as estimates $\overline{a}_{ij}$, Shin and Kang (2006) suggest the use of the OLS estimates under the null $\rho_i = 1$. (Just as well, one could use residuals from the N individual Cauchy or ADF unit root regressions.) Then, compute the sample covariance matrix

$$\widehat{\boldsymbol{\Sigma}}_{\varepsilon} = \frac{1}{T-p} \sum_{t=p+2}^{T} \overline{\boldsymbol{\varepsilon}}_t \overline{\boldsymbol{\varepsilon}}_t'$$

and let $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{-1} = \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\Gamma}}'$ be a suitable LU decomposition. Denote the orthogonalized, prewhitened differences by

$$\boldsymbol{\varepsilon}_t^* = \widehat{\boldsymbol{\Gamma}}' \overline{\boldsymbol{\varepsilon}}_t.$$

Finally, the orthogonalized statistics $\hat{\tau}_{i,IV}$ are given by

$$\hat{\tau}_{i,IV} = \frac{\sum_{t=p+2}^{T} h_i\big(\widetilde{y}_{i,t-1}^{\mu}\big) \varepsilon_{i,t}^*}{\sqrt{\sum_{t=p+2}^{T} h_i^2\big(\widetilde{y}_{i,t-1}^{\mu}\big)}},$$

[3] This, however, is only true for the t-type statistics and not for the estimators of the autoregressive roots.

where $\varepsilon^*_{i,t}$ are the N elements of $\boldsymbol{\varepsilon}^*_t$. According to Shin and Kang (2006), these are equivalent to using as instruments transformations of the lagged levels standardized using the residual variance estimators.

Define $\boldsymbol{\tau}_{IV} = (\hat{\tau}_{1,IV}, \dots, \hat{\tau}_{N,IV})'$ the vector stacking the individual orthogonalized statistics of Shin and Kang (2006). Assuming a fixed N (after which sequential asymptotics, first $T \to \infty$ followed by $N \to \infty$, applies trivially), the resulting joint distribution of $\boldsymbol{\tau}_{IV}$ is multivariate normal with zero mean and unity covariance matrix under their conditions. The result holds under our assumptions as well, in particular under unconditional heteroskedasticity.

Proposition 5. *Under Assumption 5 with* $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})'$ *such that* $\boldsymbol{\varepsilon}_t = \boldsymbol{\Omega}^{0.5}(t/T)\,\boldsymbol{\epsilon}_t$, *where* $\boldsymbol{\epsilon}_t$ *is an* N*-dimensional md sequence with uniformly bounded* N*-dimensional conditional density functions and unity covariance matrix such that* $\exists\, r > 4$ *with* $\sup_t \|\boldsymbol{\epsilon}_t\|_r < C < \infty$, *and* $\boldsymbol{\Omega}(\cdot)$ *is an* $N \times N$ *matrix of piecewise continuous functions on* $[-\infty; 1]$, $\boldsymbol{\Omega}(s)$ *positive definite* $\forall s$, *it holds under the null* $\rho_i = 1$, $i = 1, \dots, N$, *that*

$$\boldsymbol{\tau}_{IV} \xrightarrow{d} \mathcal{N}_N(\mathbf{0}_N, \boldsymbol{I}_N),$$

with $\boldsymbol{I}_N$ *the* $N \times N$ *identity matrix.*

Proof: By multivariate extension of Proposition 1 when $\phi_i = 0$.

Remark 9. Alternatively, one can also examine the joint distribution of the N individual Cauchy tests as studied in the previous section. With macropanels, there is information to be gained from single-unit tests as well and thus it may be of interest to check these first, followed by an overall panel analysis based on their joint distribution. (This would be the standard procedure in multiple testing situations.) Define $\mathbf{t}_{IV} = \left(t^{\mu}_{1,IV}, \dots, t^{\mu}_{N,IV}\right)'$ and let $\widehat{\boldsymbol{\Xi}}$ be the sample correlation matrix of $\boldsymbol{h}(\widetilde{\boldsymbol{y}}^{\mu}_{t-1}) \odot \widehat{\boldsymbol{\varepsilon}}_t$ (the elementwise product), where the vector $\widehat{\boldsymbol{\varepsilon}}_t$ contains the N stacked residuals at time t from IV estimation of the N unit-specific error-correction models in (4), $\boldsymbol{h} = (h_1, \dots, h_N)'$ and $\widetilde{\boldsymbol{y}}^{\mu}_{t-1} = (\widetilde{y}^{\mu}_{1,t-1}, \dots, \widetilde{y}^{\mu}_{N,t-1})'$. The matrix $\widehat{\boldsymbol{\Xi}}$ does converge in distribution to a random correlation matrix,

$$\widehat{\boldsymbol{\Xi}}_{i,j} \xrightarrow{d} \int_0^1 \operatorname{sgn}\left(\widetilde{J}^{\mu}_{i,c}\left(\eta_i\left(s\right)\right)\right) \operatorname{sgn}\left(\widetilde{J}^{\mu}_{j,c}\left(\eta_j\left(s\right)\right)\right) ds;$$

we conjecture that the limiting distribution of $\mathbf{t}_{IV}$ is mixed Gaussian, so orthogonalizing $\mathbf{t}_{IV}$ by $\widehat{\boldsymbol{\Xi}}$ leads to a vector of N independent standard normal random variables. The assumed DGP allows for time-varying correlation as well, so the limit of $\widehat{\boldsymbol{\Xi}}$ is only the "average" correlation; see also the discussion following Assumption 6.

Remark 10. The panel tests are consistent against the alternative of at least one stationary unit, as the straightforward multivariate extension of Proposition 2 indicates.

We now turn to the panel tests available under the above assumptions. Shin and Kang consider the IPS-type statistic $\overline{\tau}_{IV} = N^{-1/2} \sum_{i=1}^{N} \hat{\tau}_{i,IV}$. We do not analyze their Wald-type statistic W_{IV}, for which they show $W_{IV} \xrightarrow{d} \chi^2_N$ for fixed N. This is a two-sided test, and one-sided versions that focus on the relevant alternative $\phi_i < 0$ for some i are likely to be more powerful.[4] Other tests discussed by Shin and Kang (2006) are as follows. Defining $p_i = \Phi(\hat{\tau}_{i,IV})$ for Φ the cdf of the standard normal distribution, Proposition 5 guarantees that $p_i \xrightarrow{d} \mathcal{U}[0,1]$ (with $\mathcal{U}$ the uniform distribution) under H_0, where p_i and p_j are independent for $i \neq j$. Hence, the Fisher-type meta statistics $P_{IV} = -2 \sum_{i=1}^{N} \ln(p_i)$ and $Z_{IV} = N^{-1/2} \sum_{i=1}^{N} \Phi^{-1}(p_i)$ are available. The asymptotic (as $T \to \infty$) null distributions are well-known to be $P_{IV} \xrightarrow{d} \chi^2_{2N}$ and $\overline{\tau}_{IV} \xrightarrow{d} \mathcal{N}(0,1)$. Note that, due to standard asymptotics of the Cauchy test (at least for a fixed N), $\overline{\tau}_{IV} = Z_{IV}$.

3.2 Joint N, T asymptotic results

Under the simplifying assumption of a fixed N, the discussed asymptotics could be seen as rather a time series problem. While we do not share the view that such assumptions—destined to make the asymptotics more tractable—render the tests unusable, it is obvious that they do not cover the entire spectrum of possible N, T combinations, and we provide a joint asymptotic analysis. But we now require panel-specific assumptions regarding the innovations; in particular, we assume a factor structure of the panel innovations.

Assumption 6. *Let $\boldsymbol{\varepsilon}_t := \boldsymbol{\Lambda}'\boldsymbol{\nu}_t + \tilde{\boldsymbol{\varepsilon}}_t$, where the common factors $\boldsymbol{\nu}_t$ and the idiosyncratic factors $\tilde{\boldsymbol{\varepsilon}}_t$ are such that*

(a) $\boldsymbol{\Lambda} = \{\boldsymbol{\lambda}_i'\}_{i=1,\dots,N}$ is an $N \times L$ matrix, $1 \leq L$ fixed, such that $\boldsymbol{\lambda}_i \neq \mathbf{0}_L$ $\forall i$ and $\sup_{i,j} |\lambda_{ij}| < C < \infty$;

(b) $\boldsymbol{\nu}_t$ and $\tilde{\varepsilon}_{i,t}$, $i = 1, \dots, N$, are independent sequences;

(c) $\tilde{\varepsilon}_{i,t}$ satisfy each Assumption 4 with $\sup_i T^{-1.5} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{E}\left(\left|\epsilon^2_{i,s} - 1\right|\left|\epsilon^2_{i,t} - 1\right|\right) \to 0$ as $T \to \infty$ and some uniformly (in N) bounded variance function ω_i;

(d) $\boldsymbol{\nu}_t$ satisfies the heteroskedastic md assumption in Proposition 5, $\boldsymbol{\nu}_t = \boldsymbol{\Omega}^{0.5}\boldsymbol{\upsilon}_t$, such that, for each pair $1 \leq k, l \leq L$, $T^{-1.5} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{E}\left(|\upsilon_{k,s}\upsilon_{l,s} - \mathbb{I}(k=l)|\,|\upsilon_{k,t}\upsilon_{l,t} - \mathbb{I}(k=l)|\right) \to 0$ as $T \to \infty$.

Assumptions similar to Assumption 6(a) have been used by Bai and Ng (2004). Under the assumed factor structure, the innovations $\boldsymbol{\varepsilon}_t$ have at time t a covariance matrix

$$\mathrm{E}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = \boldsymbol{\Lambda}\boldsymbol{\Omega}\left(t/T\right)\boldsymbol{\Lambda}' + \mathrm{diag}\left(\omega_i^2\left(t/T\right)\right);$$

Moreover, their "average" covariance matrix is, following the univariate case,

$$\overline{\boldsymbol{\Omega}} = \boldsymbol{\Lambda} \int_0^1 \boldsymbol{\Omega}\left(s\right) ds \boldsymbol{\Lambda}' + \mathrm{diag}\left(\int_0^1 \omega_i^2\left(s\right) ds\right)$$

[4]Unreported simulations that are available upon request confirm this claim.

The covariance matrix depends on t, hence orthogonalization is never exact; but the sample covariance approaches in a certain sense $\overline{\boldsymbol{\Omega}}$ as $T \to \infty$, so the orthogonalization procedure will work asymptotically; see the proof of Proposition 6 for details. The panel exhibits strong cross-correlation in the sense that $\left\|\overline{\boldsymbol{\Omega}}\right\| = \Theta(N)$ under Assumption 6.

The uniform higher-order cross-product moment conditions implied by independence of the idiosyncratic factors and the assumed summability conditions ensure the minimal degree of homogeneity across the panel that is required for joint asymptotics (implying e.g. convergence at the same rate of the sample cross-covariances). The summability assumptions in (c) and (d) even allow for some degree of long-memory in the variance of the factors and ensures $\sqrt[4]{T}$-consistency of their sample variances. Sequential, or fixed-N, asymptotics do not resort to such assumptions since $T \to \infty$ leads to joint normality, and correlation (taken care for by orthogonalization) is the only form of cross-sectional dependence.

The main result of the subsection is given in the following Proposition about the behavior of $\overline{\tau}_{IV}$. Analogous results can be derived for the other panel tests introduced above.

Proposition 6. *Under Assumptions 5 and 6, it holds as $N, T \to \infty$ such that $N^2/\sqrt{T} \to 0$ that*

$$\overline{\tau}_{IV} \xrightarrow{d} \mathcal{N}(0,1)$$

Proof: See the Appendix.

Remark 11. If letting $p \to \infty$, there will be some trade-off between the rate of p and the rate of N: the less approximation error (cf. the proof of Proposition 1) is present in each single-unit statistic, the smaller their cumulated effect across the panel, and the more units (i.e. higher N-rates) can be considered without affecting $\overline{\tau}_{IV}$'s asymptotic standard normality under the null.

Remark 12. With $N \to \infty$, a fixed fraction of the units should exhibit stationarity to ensure consistency of the panel tests.

The upper bound $N = o\left(T^{0.25}\right)$ suggests that T should be much larger than N; this is the consequence of having to estimate $N(N-1)/2$ covariances. Anyway, N must be smaller than T to ensure positive definiteness of the sample covariance matrix. Should N be indeed larger than T, it suggests itself to make simplifying assumptions about the covariance matrix to ensure a positive definite estimate. Hartung (1999) assumes equicorrelation, and gives an algorithm on how to estimate it based on as little as one time observation. Hartung's method is easily applied to combine standard normal individual t-type statistics, and it is only natural to do so with the dependent single unit Cauchy tests $t^{\mu}_{i,IV}$.

The simplification is extreme, although the method is quite robust to deviations from the equicorrelation, here $Cov(t^{\mu}_{i,IV}, t^{\mu}_{j,IV}) = c$ for $i \neq j, \quad i, j = 1, \ldots, N$; see Hartung (1999) and

Demetrescu et al. (2006). See the following subsection for details. Alternatively, we can use shrinkage covariance estimators; again see below.

3.3 Small-sample behavior

We now augment the DGPs from Section 2.4 to investigate the panel case. The interesting issue is the behavior of the orthogonalization procedure, and we hence simulate without short-run dynamics; their effect has been discussed extensively for the univariate case. Following Chang (2002), we shall include one lagged difference to capture the effect of not knowing the true lag order in practice.

Assuming, like for the univariate Monte Carlo experiments, the expectation of the observed process to be zero, we have directly

$$y_{i,t} = \rho_i y_{i,t-1} + \varepsilon_{i,t} \qquad i = 1, \dots, N,\ t = 1, \dots, T$$

The variance-breaking error processes are now standard normal variates $\tilde{\varepsilon}_{i,t}$, where $\mathrm{Var}(\tilde{\varepsilon}_{i,t}) = 1$ for $t = 1, \dots, \lfloor \tau_i T \rfloor$ and $\mathrm{Var}(\tilde{\varepsilon}_{i,t}) = 1/\delta^2$ for $t = \lfloor \tau_i T \rfloor + 1, \dots, T$. We again consider $\tau_i = \tau \in \{0.1, 0.5, 0.9\}$ and $\delta \in \{1/5, 1, 5\}$ but now also allow for heterogeneous break dates. Specifically, we introduce the variance break δ for $i = 1 \dots, N/2$ at $\tau = 1/4$ and at τ as specified above for $i = N/2 + 1, \dots, N$.[5] Finally, we consider two patterns of cross-sectional correlation among the error terms $\epsilon_{i,t}$.

A. *Independence*: Let $\epsilon_{i,t} = \tilde{\epsilon}_{i,t}$ and $\tilde{\boldsymbol{\varepsilon}}_t = (\tilde{\epsilon}_{1,t}, \dots, \tilde{\epsilon}_{N,t})' \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_N)$.

B. *Factor Structure*: Let $\epsilon_{i,t} := \lambda_i \cdot \nu_t + \tilde{\epsilon}_{i,t}$, where ν_t are i.i.d. $\mathcal{N}(0,1)$ and $\lambda_i \sim \mathcal{U}(-1, 3)$.

When $\boldsymbol{\phi} := (\phi_1, \dots, \phi_N)' = \mathbf{0}_N$, the panel null is true, such that we study the size of the tests. To analyze power of the tests, we draw the ϕ_i from the uniform distribution on $[-0.1, 0]$.

In this section we additionally experiment with the Hartung (1999) approach to capture cross-sectional dependence between the panel units. He assumes constant correlation and proposes to consistently estimate the off-diagonal element ξ of the correlation matrix by $\hat{\xi}^* = \max(-1/(N-1), \hat{\xi})$, where $\hat{\xi} = 1 - 1/(N-1)\sum_{i=1}^N \big(t^\mu_{i,IV} - N^{-1}\sum_{i=1}^N t^\mu_{i,IV}\big)^2$ to form the following panel test statistic:

$$t_{\hat{\xi}^*,\kappa} = \frac{\sum_{i=1}^N t^\mu_{i,IV}}{\sqrt{N + (N^2 - N)\left(\hat{\xi}^* + \kappa\sqrt{\frac{2}{N+1}}(1 - \hat{\xi}^*)\right)}}$$

Here, $\kappa = 0.1 \cdot (1 + 1/(N+1) - \hat{\xi}^*)$ is a parameter designed to improve the small sample behaviour of the test statistic. Should $t^\mu_{i,IV}$ be asymptotically jointly normal, $t_{\hat{\xi}^*,\kappa} \xrightarrow{d} \mathcal{N}(0,1)$ under H_0. The test rejects for large negative values. See Demetrescu et al. (2006) for the use of this method with ADF tests to conduct panel unit root tests.

[5] We waive to analyze whether consistent estimation of the break date could lead to better panel unit root tests and instead only view the variance breaks as a nuisance parameter against which robustness is to be achieved.

Table 5: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T N	6	16	26	46	6	16	26	46	6	16	26	46	6	16	26	46
$\delta = 1/5$																	
	50	.048	.047	.039	.033	.363	.648	.738	.701	.056	.046	.040	.030	.342	.647	.748	.695
$\overline{\tau}_{IV}$	100	.053	.060	.053	.047	.775	.990	.999	1.00	.045	.054	.049	.048	.769	.985	.998	1.00
	200	.056	.062	.057	.057	.988	1.00	1.00	1.00	.059	.069	.058	.054	.979	1.00	1.00	1.00
	50	.033	.033	.032	.039	.255	.456	.546	.576	.044	.035	.034	.034	.231	.456	.542	.580
P_{IV}	100	.046	.044	.047	.044	.681	.966	.996	1.00	.040	.048	.039	.050	.679	.955	.993	1.00
	200	.059	.054	.059	.068	.982	1.00	1.00	1.00	.055	.059	.059	.065	.969	1.00	1.00	1.00
	50	.035	.030	.022	.015	.130	.130	.104	.063	.044	.033	.026	.017	.116	.136	.096	.069
$t_{\hat{\xi}^*,\kappa}$	100	.044	.043	.034	.028	.396	.496	.575	.632	.044	.046	.030	.032	.399	.524	.565	.625
	200	.062	.051	.045	.038	.885	.984	.998	1.00	.067	.053	.043	.039	.869	.983	.997	1.00
$\delta = 1$																	
	50	.043	.044	.041	.033	.333	.587	.688	.639	.049	.055	.043	.045	.282	.452	.501	.461
$\overline{\tau}_{IV}$	100	.056	.050	.051	.041	.727	.975	.998	1.00	.051	.053	.056	.052	.596	.869	.945	.990
	200	.053	.052	.054	.052	.964	1.00	1.00	1.00	.052	.068	.049	.052	.919	.997	1.00	1.00
	50	.032	.034	.036	.039	.218	.418	.492	.533	.044	.042	.037	.040	.200	.344	.384	.410
P_{IV}	100	.051	.045	.042	.044	.620	.929	.985	.999	.043	.050	.045	.046	.526	.829	.924	.978
	200	.051	.047	.050	.052	.948	1.00	1.00	1.00	.051	.063	.050	.049	.911	.997	1.00	1.00
	50	.037	.031	.027	.015	.116	.121	.104	.059	.048	.057	.053	.054	.126	.104	.090	.077
$t_{\hat{\xi}^*,\kappa}$	100	.052	.044	.033	.025	.334	.432	.461	.483	.059	.062	.062	.068	.332	.390	.394	.430
	200	.052	.044	.035	.031	.813	.957	.988	.999	.058	.078	.067	.086	.776	.919	.952	.978
$\delta = 5$																	
	50	.044	.041	.038	.031	.058	.071	.068	.072	.054	.048	.046	.053	.129	.114	.112	.083
$\overline{\tau}_{IV}$	100	.052	.050	.051	.048	.085	.096	.130	.187	.049	.052	.048	.050	.233	.217	.188	.199
	200	.036	.031	.038	.040	.365	.689	.859	.971	.039	.039	.047	.037	.438	.491	.548	.603
	50	.037	.035	.028	.035	.054	.060	.065	.083	.048	.045	.036	.043	.205	.223	.194	.106
P_{IV}	100	.045	.041	.048	.047	.097	.121	.175	.237	.049	.051	.046	.041	.437	.545	.553	.517
	200	.032	.029	.032	.036	.338	.674	.844	.965	.029	.032	.035	.031	.659	.856	.920	.956
	50	.039	.028	.025	.012	.056	.053	.049	.051	.052	.058	.053	.052	.110	.125	.132	.126
$t_{\hat{\xi}^*,\kappa}$	100	.041	.039	.033	.027	.083	.094	.103	.080	.059	.059	.059	.057	.252	.284	.318	.306
	200	.032	.021	.011	.008	.277	.442	.545	.645	.065	.058	.064	.064	.565	.643	.666	.674

Nominal 5% level. 5000 replications. $\tau = 0.1$.

Table 5 reports rejection rates for Shin and Kang's (2006) $\overline{\tau}_{IV}$ and P_{IV} as well as Hartung's (1999) $t_{\hat{\xi}^*,\kappa}$ test based on univariate Cauchy tests. Here, we report the case $\tau = 0.1$. Tables B.1 and B.2 in Appendix B provide results for $\tau = 0.5$ and $\tau = 0.9$. Size is well-controlled throughout under both independence and cross-sectional dependence, with a few exceptions for the P_{IV} test, e.g. for $\delta = 1/5$ and $\tau = 1/2$. More generally, the size of $\overline{\tau}_{IV}$ is somewhat more accurate than that of P_{IV} or $t_{\hat{\xi}^*,\kappa}$, which prompts us to recommend its use in practice. As regards power, all tests are consistent as $T \to \infty$ for any configuration of τ and δ. Power increases in N provided T is sufficiently large. Once more, $\overline{\tau}_{IV}$ emerges as the most attractive choice in that its power

Table 6: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests, heterogenous variance breaks

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T N	6	16	26	46	6	16	26	46	6	16	26	46	6	16	26	46
$\delta = 1/5$																	
	50	.062	.056	.053	.039	.454	.741	.826	.737	.063	.054	.050	.040	.464	.765	.832	.723
$\overline{\tau}_{IV}$	100	.067	.067	.063	.058	.811	.992	.999	1.00	.064	.075	.057	.055	.808	.989	1.00	1.00
	200	.072	.078	.072	.062	.980	1.00	1.00	1.00	.063	.067	.059	.058	.979	1.00	1.00	1.00
	50	.061	.062	.072	.063	.365	.628	.739	.690	.058	.064	.070	.061	.367	.650	.735	.681
P_{IV}	100	.071	.084	.086	.091	.747	.980	.998	1.00	.065	.082	.082	.082	.734	.972	.998	1.00
	200	.070	.093	.095	.104	.973	1.00	1.00	1.00	.073	.080	.082	.086	.976	1.00	1.00	1.00
	50	.063	.069	.064	.061	.238	.304	.332	.294	.069	.059	.060	.058	.238	.296	.292	.272
$t_{\hat{\xi}^*,\kappa}$	100	.071	.066	.062	.060	.518	.691	.770	.866	.070	.074	.059	.049	.487	.677	.760	.850
	200	.066	.076	.066	.055	.897	.988	.998	1.00	.072	.069	.065	.048	.876	.986	.998	1.00
$\delta = 5$	50	.030	.038	.036	.040	.094	.132	.146	.133	.040	.054	.051	.038	.152	.158	.149	.120
$\overline{\tau}_{IV}$	100	.036	.043	.030	.032	.254	.452	.594	.719	.047	.037	.034	.043	.318	.375	.404	.441
	200	.047	.042	.040	.044	.611	.932	.987	1.00	.047	.053	.044	.041	.610	.742	.822	.907
	50	.024	.034	.032	.037	.059	.100	.116	.117	.028	.038	.036	.026	.164	.210	.206	.133
P_{IV}	100	.031	.037	.031	.042	.202	.375	.492	.610	.036	.027	.023	.033	.369	.592	.666	.716
	200	.042	.033	.036	.048	.552	.902	.983	.999	.046	.034	.028	.026	.740	.922	.974	.996
	50	.023	.011	.005	.001	.051	.039	.016	.008	.048	.052	.046	.055	.101	.109	.105	.107
$t_{\hat{\xi}^*,\kappa}$	100	.029	.022	.016	.008	.150	.188	.183	.153	.062	.064	.064	.068	.267	.306	.290	.314
	200	.036	.031	.024	.016	.402	.614	.686	.774	.067	.075	.081	.079	.622	.738	.758	.788

Nominal 5% level. 5000 replications. $\tau = 0.1$.

tends to be higher than that of the other tests, although there are some cases where P_{IV} is more powerful. The $t_{\hat{\xi}^*,\kappa}$ test appears to have slightly lower power.

Additionally, Table 6 reports results for the case of a heterogenous break in variances. We again report results for $\tau = 0.1$ here and provide the remaining cases in the Appendix. (It would be redundant and it is therefore omitted to report $\delta = 1$ again.) We notice that the tests also provide reasonable size control under this scenario, with once more $\overline{\tau}_{IV}$ as the best-performing variant. Power is good and comparable to the homogenous case.

As pointed out above, *the* drawback of Shin and Kang's (2006) panel unit root test is the requirement that $T > N$ in order to obtain an invertible variance-covariance matrix $\widehat{\boldsymbol{\Sigma}}_\varepsilon$. This may not be the case in practice. Moreover, if T is only moderately larger than N, the finite-sample performance of the test will suffer. We therefore employ a recent proposal by Ledoit and Wolf (2004) to obtain an estimate $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ that allows for a panel statistic for any configuration of T and N. They propose to construct a weighted version of $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ and the identity matrix $\boldsymbol{I}$, written as $\boldsymbol{S}_T = \kappa_{1T}\boldsymbol{I} + \kappa_{2T}\widehat{\boldsymbol{\Sigma}}_\varepsilon$. Specifically, κ_{1T} and κ_{2T} are constructed as follows. Define

$$\bar{b}_T^2 = \frac{1}{N}\left[\sum_{t=1}^{T}\left(\frac{\overline{\boldsymbol{\varepsilon}}_t'\overline{\boldsymbol{\varepsilon}}_t}{T}\right)^2 - \frac{1}{T}\,\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^2)\right].$$

Further, $m_T = \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_\varepsilon)/N$, $d_T^2 = \mathrm{tr}[(\widehat{\boldsymbol{\Sigma}}_\varepsilon - m_T\boldsymbol{I})(\widehat{\boldsymbol{\Sigma}}_\varepsilon - m_T\boldsymbol{I})']/N$, $b_T^2 = \min(\bar{b}_T^2, d_T^2)$ and $a_T^2 = d_T^2 - b_T^2$. Then, $\kappa_{1T} = m_T \cdot b_T^2/d_T^2$ and $\kappa_{2T} = a_T^2/d_T^2$. The full-rank matrix $\boldsymbol{I}$ ensures that $\boldsymbol{S}_T$ is invertible even if $T < N$. The (generally misspecified, but invertible) structure imposed by adding $\kappa_{1T}\boldsymbol{I}$ to the unbiased estimator $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ introduces a finite-sample bias in $\boldsymbol{S}_T$. Yet, the weights κ_{1T} and κ_{2T} are shown to be optimal in the sense that $\boldsymbol{S}_T$ asymptotically (for $N, T \to \infty$ jointly) has minimum expected loss in the class of (possibly random) linear combinations of $\boldsymbol{I}$ and $\widehat{\boldsymbol{\Sigma}}_\varepsilon$, including infeasible ones that use hindsight knowledge of the true covariance matrix in the construction of κ_{1T} and κ_{2T}. Ledoit and Wolf (2004) show the joint asymptotics to be a good guide to finite samples where N and T are of the same magnitude, including the case $T < N$.

Table 7: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests with shrinkage

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T N	16	26	56	106	16	26	56	106	16	26	56	106	16	26	56	106
$\delta = 1/5$																	
	50	.066	.079	.093	.106	.879	.976	1.00	1.00	.075	.065	.093	.083	.869	.978	1.00	1.00
$\overline{\tau}_{IV}$	100	.069	.067	.088	.107	.997	1.00	1.00	1.00	.070	.073	.076	.086	.995	1.00	1.00	1.00
	200	.069	.069	.081	.094	1.00	1.00	1.00	1.00	.075	.072	.070	.080	1.00	1.00	1.00	1.00
	50	.047	.054	.052	.054	.733	.886	.997	1.00	.045	.036	.035	.017	.704	.877	.989	.999
P_{IV}	100	.061	.066	.073	.094	.990	1.00	1.00	1.00	.064	.060	.056	.041	.984	.999	1.00	1.00
	200	.076	.079	.098	.112	1.00	1.00	1.00	1.00	.067	.071	.068	.048	1.00	1.00	1.00	1.00
	50	.045	.047	.033	.019	.229	.212	.155	.111	.051	.042	.032	.020	.220	.206	.150	.097
$t_{\hat{\xi}^*,\kappa}$	100	.051	.052	.040	.032	.633	.726	.840	.934	.061	.047	.040	.036	.637	.710	.833	.920
	200	.057	.056	.046	.030	.990	.999	1.00	1.00	.062	.055	.048	.036	.988	.999	1.00	1.00
$\delta = 1$																	
	50	.057	.072	.088	.106	.838	.965	1.00	1.00	.043	.034	.020	.004	.528	.633	.713	.654
$\overline{\tau}_{IV}$	100	.055	.055	.068	.087	.995	1.00	1.00	1.00	.051	.047	.042	.022	.902	.973	.997	.998
	200	.058	.065	.079	.088	1.00	1.00	1.00	1.00	.057	.056	.058	.042	.996	1.00	1.00	1.00
	50	.039	.036	.034	.027	.638	.835	.989	1.00	.022	.011	.001	.000	.324	.367	.257	.007
P_{IV}	100	.045	.045	.045	.053	.975	.999	1.00	1.00	.034	.029	.020	.002	.854	.954	.992	.990
	200	.050	.058	.063	.066	1.00	1.00	1.00	1.00	.053	.043	.046	.026	.997	1.00	1.00	1.00
	50	.037	.027	.016	.010	.162	.131	.066	.025	.049	.047	.047	.045	.136	.112	.098	.073
$t_{\hat{\xi}^*,\kappa}$	100	.042	.034	.021	.014	.507	.569	.632	.726	.070	.063	.073	.074	.471	.481	.511	.533
	200	.044	.037	.028	.017	.969	.996	1.00	1.00	.076	.071	.088	.093	.933	.971	.992	.997
$\delta = 5$	50	.025	.033	.023	.028	.103	.168	.270	.383	.013	.010	.003	.001	.085	.076	.044	.025
$\overline{\tau}_{IV}$	100	.036	.037	.051	.050	.428	.604	.898	.990	.024	.022	.012	.007	.302	.312	.336	.354
	200	.042	.043	.049	.057	.895	.982	1.00	1.00	.032	.031	.026	.017	.611	.688	.801	.918
	50	.006	.004	.000	.000	.029	.037	.019	.002	.001	.000	.000	.000	.047	.021	.001	.000
P_{IV}	100	.013	.014	.011	.004	.289	.389	.638	.818	.005	.003	.000	.000	.470	.488	.403	.131
	200	.026	.022	.018	.016	.850	.965	1.00	1.00	.017	.010	.003	.000	.881	.939	.982	.985
	50	.006	.001	.000	.000	.012	.007	.000	.000	.040	.043	.044	.045	.099	.104	.109	.103
$t_{\hat{\xi}^*,\kappa}$	100	.016	.008	.003	.000	.136	.129	.103	.060	.062	.053	.058	.061	.282	.291	.299	.312
	200	.024	.012	.006	.004	.590	.678	.797	.898	.072	.066	.063	.063	.708	.741	.753	.757

Nominal 5% level. 5000 replications. $\tau = 0.1$.

We therefore present some additional simulations gauging the effectiveness of Shin and Kang's (2006) tests using shrinkage, allowing us to also consider the case $T < N$. In particular we now take $N \in \{16, 26, 56, 106\}$. Table 7 reports rejection rates for $\tau = 0.1$; Tables B.5 and B.6 in the Appendix provide the cases $\tau = 0.5$ and $\tau = 0.9$. The P_{IV} test is now sometimes drastically undersized especially for cases where N is much larger than T. On the other hand, $\overline{\tau}_{IV}$ mostly performs quite well even with shrinkage and in cases where $N > T$, although predictably somewhat less accurately than when one can use an estimator $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$ that unbiasedly estimates the true covariance matrix. In terms of size, $t_{\hat{\xi}^*,\kappa}$ that does not require shrinkage emerges as a serious competitor when $N > T$. However, $\overline{\tau}_{IV}$ is substantially more powerful than $t_{\hat{\xi}^*,\kappa}$ for small and intermediate T whenever size is comparable. These results further strengthen the above recommendation to employ $\overline{\tau}_{IV}$ in practice. Reassuringly, we observe that the undersizedness of the shrinkage versions of P_{IV} does not destroy its consistency in that P_{IV} remains powerful at least for sufficiently large T.

In a final set of experiments, we study the size properties of the tests under cross-unit cointegration, a dependence structure which has been shown to be challenging for many existing second-generation panel unit root tests (Banerjee et al., 2005). Consider

DGP C: Stack $Y_t = (y_{1,t}, \ldots, y_{N/2,t})'$ and $X_t = (y_{N/2+1,t}, \ldots, y_{N,t})'$. Let

$$\begin{pmatrix} \Delta Y_t \\ \Delta X_t \end{pmatrix} = \boldsymbol{\alpha}\boldsymbol{\beta}' \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \tilde{\boldsymbol{\varepsilon}}_t,$$

$$\boldsymbol{\alpha} = -\tilde{\alpha} \cdot \begin{pmatrix} \boldsymbol{I}_{N/2} & \mathbf{0}_{N/2} \\ \mathbf{0}_{N/2} & \boldsymbol{I}_{N/2} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta}' = \begin{pmatrix} \boldsymbol{I}_{N/2} & -\boldsymbol{I}_{N/2} \\ \mathbf{0}_{N/2} & \boldsymbol{B}_{N/2} \end{pmatrix}.$$

This yields $N/2$ 'within-country' cointegration relationships $y_{i,t} - y_{N/2+i,t}$, $i = 1, \ldots, N/2$. 'Cross-cointegration' obtains when $\text{rk}(\boldsymbol{B}_{N/2}) > 0$. If, say, the first row of $\boldsymbol{B}_{N/2}$ is $(1,\ -1,\ \mathbf{0}'_{N/2-2})$, then $y_{N/2+1,t}$ and $y_{N/2+2,t}$ cointegrate with $(1, -1)'$. We put $\tilde{\alpha} = 0.05$ and $\text{rk}(\boldsymbol{B}_{N/2}) = 2$. Due to the moving-average component induced by the cointegrating relationships we construct the test regressions with $[4(T/100)^{2/9}]$ lagged differences. Table 8 shows that the tests also seem to be able to handle this scenario, although size is less well-controlled here. In particular, $t_{\hat{\xi}^*,\kappa}$ tends to be undersized. These distortions occur mainly for short and wide panels. This is not surprising in view of the results of e.g. Hanck (2008), who shows that time series size distortions will accumulate in N. P_{IV} and $\overline{\tau}_{IV}$ sometimes exhibit some upward size distortions. In general, the versions of the tests with shrinkage perform as good as those without in comparably wide panels. As before, shrinkage allows us to consider panels where $N > T$. (Since DGP C ensures $I(1)$-ness for all $y_{i,t}$, no power study applies here.)

All in all, we conclude that the applicability of Shin and Kang's (2006) tests in practical applications is much wider than was previously recognized.

Table 8: Size of the Shin and Kang and Demetrescu et al. Panel Tests under Cross-Cointegration

		$\tau = 0.1$								$\tau = 0.9$							
		Shrinkage				No shrinkage				Shrinkage				No shrinkage			
	T N	16	26	56	106	6	16	26	46	16	26	56	106	6	16	26	46
$\delta = 1/5$																	
	50	.132	.135	.205	.269	.116	.094	.083	.049	.029	.028	.017	.010	.049	.031	.024	.023
$\overline{\tau}_{IV}$	100	.188	.169	.199	.251	.227	.148	.139	.116	.035	.034	.023	.009	.065	.035	.032	.011
	200	.230	.214	.247	.280	.290	.226	.211	.189	.104	.101	.101	.096	.135	.093	.074	.041
	50	.039	.030	.025	.017	.052	.038	.032	.030	.010	.009	.005	.000	.028	.029	.029	.035
P_{IV}	100	.095	.075	.068	.072	.144	.081	.073	.063	.015	.012	.006	.003	.039	.028	.032	.025
	200	.173	.154	.171	.183	.204	.172	.156	.141	.072	.063	.054	.041	.101	.078	.072	.056
	50	.007	.004	.002	.000	.016	.010	.004	.001	.007	.002	.000	.000	.017	.007	.006	.002
$t_{\hat{\xi}^*,\kappa}$	100	.035	.025	.012	.003	.049	.031	.024	.011	.010	.008	.003	.000	.023	.011	.006	.003
	200	.088	.073	.066	.049	.069	.089	.074	.066	.049	.037	.024	.009	.067	.047	.038	.025
$\delta = 1$																	
	50	.090	.102	.137	.153	.101	.082	.058	.041	.062	.068	.060	.085	.063	.059	.041	.027
$\overline{\tau}_{IV}$	100	.123	.143	.128	.155	.169	.119	.099	.072	.105	.086	.088	.091	.149	.095	.077	.052
	200	.178	.169	.180	.201	.242	.175	.156	.141	.158	.153	.134	.143	.221	.158	.123	.111
	50	.023	.016	.010	.006	.043	.030	.023	.026	.016	.010	.006	.002	.027	.020	.014	.020
P_{IV}	100	.053	.051	.034	.021	.095	.052	.045	.030	.045	.030	.015	.013	.080	.041	.028	.025
	200	.118	.097	.094	.085	.162	.123	.105	.080	.100	.089	.070	.063	.149	.103	.072	.064
	50	.006	.005	.000	.000	.010	.006	.002	.001	.005	.003	.000	.000	.010	.005	.002	.001
$t_{\hat{\xi}^*,\kappa}$	100	.015	.013	.002	.000	.022	.016	.009	.004	.015	.009	.001	.000	.021	.014	.006	.003
	200	.054	.040	.025	.014	.055	.056	.050	.034	.039	.039	.019	.010	.043	.057	.033	.021
$\delta = 5$	50	.001	.001	.000	.000	.012	.005	.003	.005	.070	.067	.072	.074	.073	.047	.040	.025
$\overline{\tau}_{IV}$	100	.001	.000	.000	.000	.008	.001	.001	.000	.105	.097	.086	.087	.144	.083	.067	.053
	200	.011	.005	.002	.000	.036	.011	.005	.000	.169	.140	.131	.127	.227	.133	.117	.098
	50	.000	.000	.000	.000	.008	.003	.004	.006	.015	.009	.004	.002	.031	.016	.020	.015
P_{IV}	100	.000	.000	.000	.000	.002	.001	.001	.000	.038	.031	.019	.008	.079	.037	.029	.023
	200	.002	.001	.000	.000	.017	.006	.004	.002	.102	.078	.060	.048	.153	.093	.073	.055
	50	.001	.000	.000	.000	.007	.000	.001	.000	.003	.001	.000	.000	.011	.003	.001	.000
$t_{\hat{\xi}^*,\kappa}$	100	.000	.000	.000	.000	.002	.000	.000	.000	.011	.010	.003	.000	.020	.011	.007	.002
	200	.001	.001	.000	.000	.004	.002	.000	.000	.038	.034	.016	.006	.043	.045	.034	.019

Nominal 5% level. 5000 replications.

4 A robust analysis of OECD GDP prices

We now apply the tests to investigate whether panel time series of price levels of GDP can be treated as stationary. Price Level of GDP is the Purchasing Power Parity over GDP divided by the exchange rate times 100. The PPP of GDP is the national currency value divided by the real value in international dollars. The PPP and the exchange rate are both expressed as national currency units per US dollar (Heston et al., 2009). The data is taken from Heston et al. (2009), item [8]. The reference country is the United States.

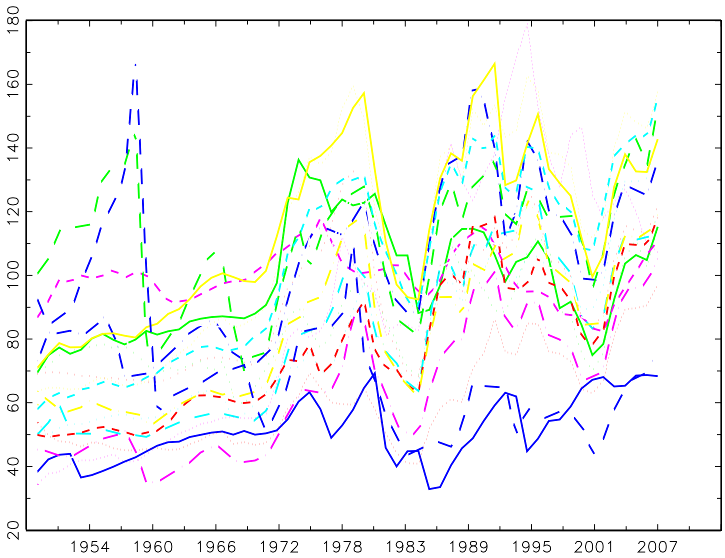
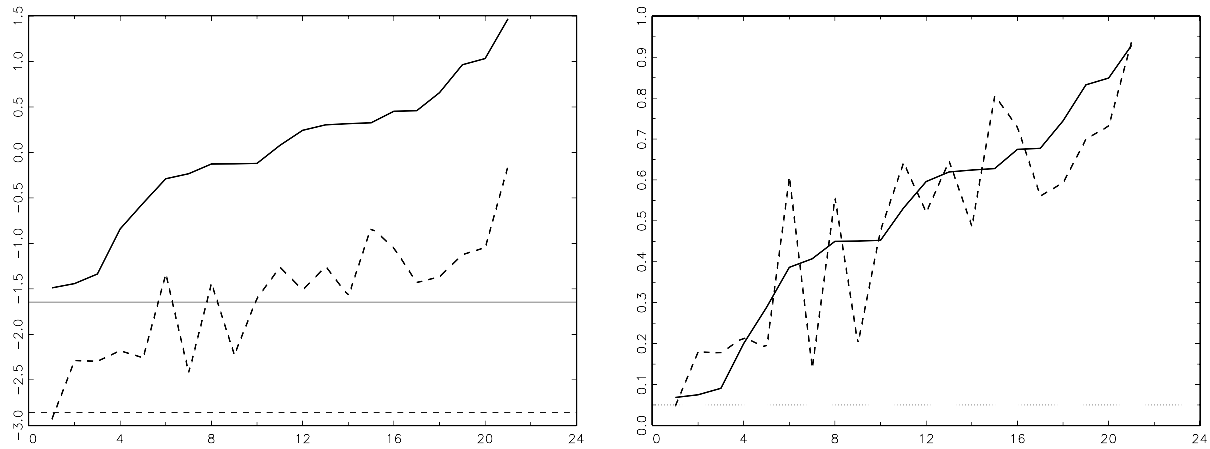


Figure 2: OECD GDP prices

Figure 2 plots the price of GDP for all OECD countries for which a complete record of observations from 1950 to 2007 is available.[6] Overall, the series do not appear to mean-revert, and as such one would not expect well-designed unit root tests to reject the null hypothesis. It is readily apparent that most series were characterized by relatively tranquil behavior until the early 1970s. Afterwards, the volatility of the series is markedly larger. Hence, the panel has undergone a variance break after around a third of the available observations. The break can easily be dated as the breakdown of the Bretton Woods system, after which exchange rates were no longer fixed, which translated into higher volatility of GDP prices. (Conveniently, no dating of breaks is necessary for our approach, though.) It is therefore useful to model the panel as driven by unconditionally heteroskedastic innovations. As discussed in previous sections, standard (panel) unit root tests are likely to produce misleading inference in the presence of such variance breaks. Moreover, the time series clearly comove. This pronounced effect is due to common global macroeconomic shocks, but also the common reference country. Hence, panel tests that can handle cross-sectional dependence are required in this application. Finally, the series do not exhibit trending behavior, such that the assumptions underlying the Cauchy test will be met in this application.

Figure 3 reports sorted Cauchy (solid) and standard augmented Dickey-Fuller (dashed) unit root test statistics and p-values for the OECD countries, sorted according to the Cauchy statistics. The horizontal lines are the 5% critical values. We observe that none of the Cauchy tests rejects

[6] These countries are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Turkey and the United Kingdom.



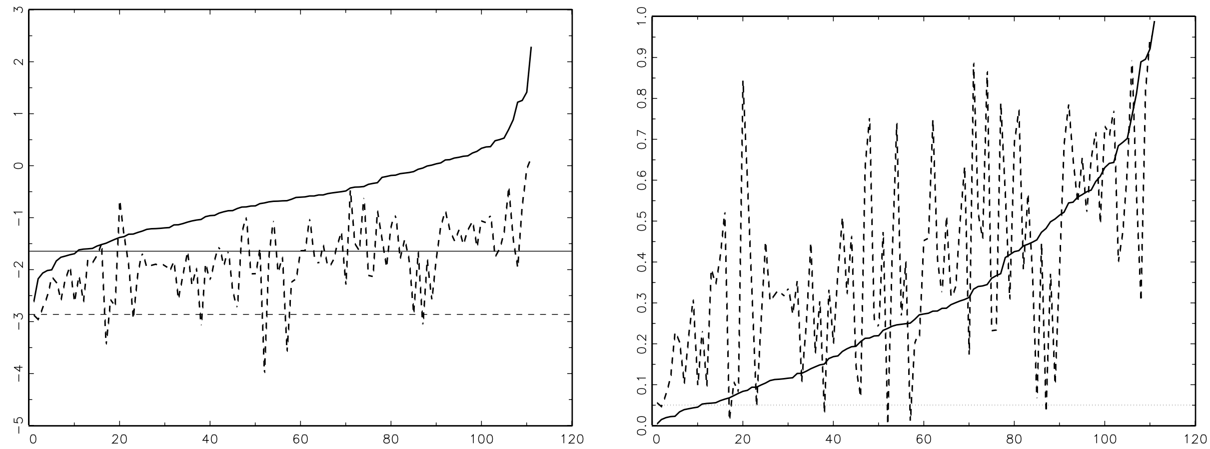
Left Panel: Test Statistics. Right Panel: p-values.
Solid: Cauchy. Dashed: ADF. The horizontal lines are 5% critical values.

Figure 3: Unit Root Tests for OECD GDP Prices

the unit root null. On the other hand, there is one country (Canada) for which a standard ADF test would have provided evidence against nonstationarity of the price level of GDP. That the ADF tests are more rejective under variance breaks is consistent with the Monte Carlo evidence of among others Hanck (2009) (see above) who demonstrates that standard ADF tests reject too frequently under unconditional heteroskedasticity. The p values demonstrate that the correlation between the standard and robust tests is important, but clearly less than unity. (Their correlation is 86.3%, while that of the test statistics is 86.5%.)

We now turn to the panel analysis. We calculate the heteroskedasticity-robust versions derived above and contrast the findings with those provided by popular second-generation panel unit root tests. Consistent with the eyeball analysis of Figure 2, none of the robust tests rejects. Hartung's $t_{\hat{\xi}^*,\kappa}$ yields a p-value of 0.308, the $\overline{\tau}_{IV}$ statistic equals -1.11 and $P_{IV} = 47.53$, where the appropriate critical value is 58.12. The panel tests t_{rob} from Breitung and Das (2005), t_a^* by Moon and Perron (2004), $t_{\hat{\xi}^*,\kappa}$ based on standard ADF statistics and $CIPS^*$ by Pesaran (2007) have test statistics -0.044, -1.20, 0.51 (p-value) and -2.36, respectively. Hence, consistent with the undersizedness reported in Hanck (2009) we do not observe a rejection for t_a^*. t_{rob} and $t_{\hat{\xi}^*,\kappa}$ do not reject, either. On the other hand, the oversizedness of $CIPS^*$ translates into a rejection. We interpret this rejection to be an artefact of the Bretton Woods upward variance break rather than as evidence for the stationarity of GDP prices. The more suitable, in our view, heteroskedasticity-robust panel tests support the notion of GDP price nonstationarity.

We next consider all countries in the Penn World Tables for which a complete record of observations from 1960 to 2007 is available. This yields $N = 111$ (of a possible total of 190) countries and $T = 48$ time series observations. It would therefore not be possible to calculate the



Left Panel: Test Statistics. Right Panel: p-values.
Solid: Cauchy. Dashed: ADF. The horizontal lines are 5% critical values.

Figure 4: Unit Root Tests for Penn World Table GDP Prices, Full Panel

standard Shin and Kang (2006) panel statistic. We therefore now employ the shrinkage method. This wider data set is not obviously characterized by a variance break, as the many developing countries now contained in the panel were less affected by the breakdown of Bretton Woods (plots are available upon request). Hence, it is difficult to a priori advocate adoption of any particular test here. For this panel, a mixed result arises, with $\overline{\tau}_{IV}$ rejecting with a statistic of -2.46 but P_{IV} and $t_{\hat{\xi}^*,\kappa}$ accepting with a test statistic and p-value of 217.50 (the critical value is 257.76) and 0.14, respectively. Only one of the second generation tests rejects, viz. t_{rob} with a test statistic of -3.96. The other tests accept with test statistics of -1.43 (t_a^*), -0.58 ($t_{\hat{\xi}^*,\kappa}$ based on standard ADF statistics) and -2.05 ($CIPS^*$).

The Hartung $t_{\hat{\xi}^*,\kappa}$ test also lends itself well to a sample sensitivity analysis in that it, through pooling test statistics, directly accommodates unbalanced panels. We therefore recalculate the $t_{\hat{\xi}^*,\kappa}$ statistic based on single-unit Cauchy tests using all available time series observations for a country provided T_i is at least $T_{\min} = 20$ to mitigate extreme small sample distortions. This increases the average time series sample size to roughly $\bar{T} = 49$ and the panel width to $N = 164$. The p-value then is 0.283, yielding the same conclusion as for the 1960-to-2007 panel. However, the value is much closer to the one observed for the OECD panel (0.308). This is intuitive because the unbalanced panel adds observations from the 1950s, mostly from OECD countries. Finally, we vary $T_{\min}$ from 20 to 10, 30 and 40, finding panel widths of 187, 163 and 113 and average lengths of 45, 49 and 54. The p-values then are 0.267, 0.284 and 0.329, suggesting that the conclusions are not sensitive to the choice of $T_{\min}$.

The mixed signals among the different panel tests call for combination procedures for panel unit root tests along the lines of Harvey et al. (2009) or Bayer and Hanck (2009), a topic left for further research. Inspecting the single country test statistics (cf. Figure 4), we observe, however, that both the robust and the standard approach only declare few (10 for the Cauchy test and 8 for ADF) single time series to be stationary. These numbers are roughly consistent with the rejections one would expect to see if all single null hypotheses are true in a multiple testing situation such a present one, viz. $0.05 \cdot 111 \approx 6$. Hence, evidence for stationarity of GDP prices appears to be weak.

5 Concluding remarks

We analyzed nonlinear instrumental unit root and panel unit root tests. The focus was on the so-called Cauchy estimator, where the sign of the lagged level is taken as an instrument for the lagged level itself.

In spite of the Cauchy test having a standard normal distribution under the null, our analysis showed the test to have power in the same $1/T$ neighborhood of the unit root as the ADF test. Moreover, we established the result under unconditional heteroskedasticity, with the byproduct that the asymptotic null distribution is invariant to the variance profile.

Our findings apply to the case where the series to be tested exhibits at most a non-zero mean. If the differences require adjustment for deterministic components, the standard asymptotics is lost. Still, the Cauchy test with detrending has power in $1/T$ neighborhoods of the unit root as well.

The standard asymptotics and robustness to unconditional heteroskedasticity were exploited to establish a robust panel unit root test robust to unconditional heteroskedasticity. The panel test is based on an orthogonalization procedure with an estimated covariance matrix. The assumptions under which joint N, T asymptotics hold suggested that N should be smaller than T. To extend the applicability of the panel test to situations where T is comparable to, or smaller than, N, we suggested the use of shrinkage covariance matrix estimators. The test performed well in small samples.

An empirical application to price levels of GDP in OECD and other countries shows the differential inference heteroskedasticity robust (panel) unit root tests may provide in practice.

Appendix A Proofs

Note: C stands for a generic constant not depending on either T or N. Sums run from $p+2$ to T unless specified otherwise.

Lemma 4. *Let either $\widetilde{x}_t = \widetilde{x}_t^{\mu}$ or $\widetilde{x}_t = \widetilde{x}_t^{\tau}$. It holds under Assumptions 1 through 4 as $T \to \infty$ that*

A. $\mathrm{E}\left(\left|\sum_{k\geq 0} b_k \varepsilon_{t-k}\right|^r\right) < C \ \forall t \leq T$ *if* $\sum_{k\geq 0} |b_k| < C < \infty$;

B. $\mathrm{E}\left(\left|h\left(\widetilde{x}_{t-1}\right) - \mathrm{sgn}\left(\widetilde{x}_{t-1}\right)\right|^{\beta}\right) \leq Ct^{-0.5} \ \forall \beta > 0$;

C. $\sum \left(h\left(\widetilde{x}_{t-1}\right) - \mathrm{sgn}\left(\widetilde{x}_{t-1}\right)\right) \varepsilon_t = O_p(T^{0.25})$;

D. $\sum h\left(\widetilde{x}_{t-1}\right) \widetilde{x}_{t-1} = O_p\left(T^{1.5}\right)$;

E. $\sum h\left(\widetilde{x}_{t-1}\right) \Delta x_{t-j} = o_p\left(T^{0.75}\right)$ *uniformly for* $0 \leq j \leq K_T$ *where* $K_T = CT^{\delta}$ *for some* $0 < \delta < 1/2 - 2/r$;

F. $\sum \widetilde{x}_{t-1} \Delta x_{t-j} = O_p\left(T\right)$ *uniformly for* $0 \leq j \leq K_T$.

Proof of Lemma 4

A. Note that uniform boundedness of the expectation is equivalent to uniform boundedness of the L_r norm of $\sum_{k\geq 0} b_k \varepsilon_{t-k}$. Then, by using the Minkowski inequality and the properties of the L_r norm, it follows that

$$\left\|\sum_{k\geq 0} b_k \varepsilon_{t-k}\right\|_r \leq \sum_{k\geq 0} |b_k| \left\|\varepsilon_{t-k}\right\|_r,$$

whose r.h.s. is uniformly bounded due to absolute summability of $\{b_k\}$ and uniform boundedness of $\mathrm{E}(|\varepsilon_t|^r)$.

B. The joint density of the innovations $\varepsilon_1, \ldots, \varepsilon_T$ does not have any poles or atoms, otherwise the conditional densities would themselves exhibit some; as a consequence, $\widetilde{x}_{t-1}/\sqrt{T}$ – which, thanks to the Beveridge-Nelson decomposition, can be written as a linear combination of $\varepsilon_t/\sqrt{T}$'s and a vanishing term – has itself continuous probability density function g_t bounded uniformly in T. This implies that $P\left(|\widetilde{x}_{t-1}| < m\right) \leq Ct^{-0.5}$. The result follows with

$$\mathrm{E}\left(\left|h\left(\widetilde{x}_{t-1}\right) - \mathrm{sgn}\left(\widetilde{x}_{t-1}\right)\right|^{\beta}\right) \leq \max_{|\widetilde{x}|<m} g_t^{\beta}\, \mathrm{E}\left(\left|I\left(|\widetilde{x}_{t-1}| < m\right)\right|\right) \leq Ct^{-0.5}$$

because $\mathrm{E}(|I\left(|\widetilde{x}_{t-1}| < m\right)|) = P\left(|\widetilde{x}_{t-1}| < m\right)$ and g_t, being continuous, is bounded on the compact interval $[-m, m]$. Note that weak convergence of $\widetilde{x}_{t-1}/\sqrt{T}$ is not strong enough to establish the result as it cannot guarantee the rate at which $P\left(|\widetilde{x}_{t-1}| < m\right)$ vanishes, cf. the proof of Lemma A.1(i) in Shin and So (2000).

C. We have thanks to the md property that $\mathrm{Var}\left(\left|\sum\left(h\left(\widetilde{x}_{t-1}\right) - \mathrm{sgn}\left(\widetilde{x}_{t-1}\right)\right)\varepsilon_t\right|\right) = \sum \mathrm{Var}\left(\left|\left(h\left(\widetilde{x}_{t-1}\right) - \mathrm{sgn}\left(\widetilde{x}_{t-1}\right)\right)\varepsilon_t\right|\right)$. Use now the law of iterated expectations to write $\mathrm{Var}\left(\left|\left(h\left(\widetilde{x}_{t-1}\right) - \mathrm{sgn}\left(\widetilde{x}_{t-1}\right)\right)\varepsilon_t\right|\right)$ as a sum,

$$\mathrm{E}\left(\left|\left(h\left(\widetilde{x}_{t-1}\right) - \mathrm{sgn}\left(\widetilde{x}_{t-1}\right)\right)^2 \varepsilon_t^2\right| \,\middle|\, |\widetilde{x}_{t-1}| < m\right) P\left(|\widetilde{x}_{t-1}| < m\right) + \mathrm{E}\left(\left|\left(h\left(\widetilde{x}_{t-1}\right) - \mathrm{sgn}\left(\widetilde{x}_{t-1}\right)\right)^2 \varepsilon_t^2\right| \,\middle|\, |\widetilde{x}_{t-1}| \geq m\right) P\left(|\widetilde{x}_{t-1}| \geq m\right)$$

The second conditional expectation on the r.h.s. is zero, and the first can be bounded by $C\,\mathrm{E}(\varepsilon_t^2)$. Since the probability $P\left(|\widetilde{x}_{t-1}| < m\right)$ is of order $t^{-0.5}$, the variance of the sum is of order $O(\sqrt{T})$ leading to the desired result.

D. Write with $\mathrm{E}\left(\left|T^{-1}\widetilde{x}_{t-1}^2\right|\right) < C$

$$\begin{aligned}
\mathrm{E}\left|\sum_{t=p+2}^{T} h\left(\widetilde{x}_{t-1}\right) \widetilde{x}_{t-1}\right| &\leq \sqrt{\mathrm{E}\sum_{t=p+2}^{T} h^2\left(\widetilde{x}_{t-1}\right) \mathrm{E}\sum_{t=p+2}^{T} \widetilde{x}_{t-1}^2} \\
&\leq CT^{1.5}
\end{aligned}$$

since $\mathrm{E}\big(\sum_{t=p+2}^{T} h^2(\widetilde{x}_{t-1})\big) \leq C \sum_{t=p+2}^{T} \mathrm{E}\big(\mathrm{sgn}^2(\widetilde{x}_{t-1})\big) = O(T)$.

E. Note first that $A(L)\Delta x_t = \varepsilon_t - c/T x_{t-1}$. With $\sum_{k\geq 0} b_k L^k = A^{-1}(L)$, where $b_0 = 1$ and the coefficients b_k decay exponentially, it follows that, $\forall j$, $\Delta x_{t-j} = \sum_{k\geq 0} b_k \varepsilon_{t-j-k} - c/T \sum_{k\geq 0} b_k x_{t-j-k-1}$. But $x_{t-j-k-1} = O_p(T^{0.5})$, see Lemmas 1 and 2, and the coefficients b_k are absolutely summable; as a consequence, it holds that

$$\sum_{t=p+2}^{T} \mathrm{sgn}(\widetilde{x}_{t-j}) \Delta x_{t-1} = \sum_{t=p+2}^{T} \mathrm{sgn}(\widetilde{x}_{t-1}) \sum_{k\geq 0} b_k \varepsilon_{t-j-k} + O_p\left(T^{0.5}\right).$$

Using arguments similar to those in the proof of item C (in particular making use of the absolute summability of the coefficients b_k), the result follows from examining the quantity $\sum \mathrm{sgn}(\widetilde{x}_{t-1}) \sum_{k\geq 0} b_k \varepsilon_{t-j-k}$ instead of $\sum h(\widetilde{x}_{t-1}) \sum_{k\geq 0} b_k \varepsilon_{t-j-k}$.

Write now for some $s > \max\{j, p+2\}$

$$\begin{aligned}
&\sum_{t=p+2}^{T} \mathrm{sgn}(\widetilde{x}_{t-1}) \sum_{k\geq 0} b_k \varepsilon_{t-j-k} \\
&\quad = \sum_{t=p+2}^{s} \mathrm{sgn}(\widetilde{x}_{t-1}) \sum_{k\geq 0} b_k \varepsilon_{t-j-k} + \sum_{t=s+1}^{T} \mathrm{sgn}(\widetilde{x}_{t-1}) \sum_{k\geq 0} b_k \varepsilon_{t-j-k}
\end{aligned}$$

which further equals

$$\begin{aligned}
&\sum_{t=p+2}^{s} \mathrm{sgn}(\widetilde{x}_{t-1}) \sum_{k\geq 0} b_k \varepsilon_{t-j-k} \\
&\quad + \sum_{t=s+1}^{T} \mathrm{sgn}(\widetilde{x}_{t-s}) \left(\sum_{k=0}^{s-j-1} b_k \varepsilon_{t-j-k} + \sum_{k\geq s-j} b_k \varepsilon_{t-j-k} \right) \\
&\quad + \sum_{t=s+1}^{T} \left(\mathrm{sgn}(\widetilde{x}_{t-1}) - \mathrm{sgn}(\widetilde{x}_{t-s})\right) \sum_{k\geq 0} b_k \varepsilon_{t-j-k}.
\end{aligned}$$

Now, we obviously have that $\mathrm{E}\big(\big|\sum_{t=p+2}^{s} \mathrm{sgn}(\widetilde{x}_{t-1}) \sum_{k\geq 0} b_k \varepsilon_{t-j-k}\big|\big) \leq Cs$. Furthermore, note that

$$\sum_{t=s+1}^{T} \mathrm{sgn}(\widetilde{x}_{t-s}) \sum_{k=0}^{s-j-1} b_k \varepsilon_{t-j-k} = O_p\left(T^{0.5}\right)$$

since it can be written as a sum of md terms having uniformly bounded variance. Moreover,

$$\mathrm{E}\left(\sum \mathrm{sgn}(\widetilde{x}_{t-s}) \sum_{k\geq s-j} b_k \varepsilon_{t-j-k} \right) = O\left(T \sum_{k\geq s-j} b_k \right)$$

or $O\left(Te^{-(s-j)}\right)$ due to the fact that the coefficients b_k have exponential decay. So the result follows by letting $s \to \infty$ as $T \to \infty$ at an appropriate rate, higher than K_T, if

$$\sum_{t=s+1}^{T} \left(\mathrm{sgn}(\widetilde{x}_{t-1}) - \mathrm{sgn}(\widetilde{x}_{t-s})\right) \sum_{k\geq 0} b_k \varepsilon_{t-j-k} = o_p\left(T^{0.75}\right)$$

uniformly in $0 \leq j \leq K_T$. To this end, apply Hölder's inequality with $1 = (r-1)/r + 1/r$ to obtain

$$\begin{aligned}
&\left| \sum_{t=s+1}^{T} \left(\mathrm{sgn}(\widetilde{x}_{t-1}) - \mathrm{sgn}(\widetilde{x}_{t-s})\right) \sum_{k\geq 0} b_k \varepsilon_{t-j-k} \right| \\
&\quad \leq \sqrt[\frac{r}{r-1}]{\sum_{t=s+1}^{T} \left(\mathrm{sgn}(\widetilde{x}_{t-1}) - \mathrm{sgn}(\widetilde{x}_{t-s})\right)^{\frac{r}{r-1}}} \sqrt[r]{\sum_{t=s+1}^{T} \left(\sum_{k\geq 0} b_k \varepsilon_{t-j-k} \right)^r}.
\end{aligned}$$

Given finiteness of rth order moments of ε_t and item A of this lemma, the second term of the r.h.s. is of order $O_p\left(T^{1/r}\right)$, and we only have to examine the rate of $\sum |\text{sgn}\left(\widetilde{x}_{t-1}\right) - \text{sgn}\left(\widetilde{x}_{t-s}\right)|^{r/(r-1)}$. Note that, for suitable C,

$$\begin{aligned}
&\mathrm{E}\left(\sum_{t=s+1}^{T} |\text{sgn}\left(\widetilde{x}_{t-1}\right) - \text{sgn}\left(\widetilde{x}_{t-s}\right)|^{r/(r-1)}\right) \\
&\quad = C\sum_{t=s+1}^{T} \mathrm{E}\big(|\text{sgn}\left(\widetilde{x}_{t-1}\right) - \text{sgn}\left(\widetilde{x}_{t-s}\right)|\big) \\
&\quad = C\sum_{t=s+1}^{T} P\left(|\text{sgn}\left(\widetilde{x}_{t-1}\right) - \text{sgn}\left(\widetilde{x}_{t-s}\right)| = 2\right).
\end{aligned}$$

Each probability on the r.h.s. is given by the (unconditional) probability of a change of sign from $t-s$ to t. To assess this probability, write

$$\begin{aligned}
&P\left(|\text{sgn}\left(\widetilde{x}_{t-1}\right) - \text{sgn}\left(\widetilde{x}_{t-s}\right)| = 2\right) \\
&\quad \leq 2P\left(|\widetilde{x}_{t-1} - \widetilde{x}_{t-s}| > |\widetilde{x}_{t-s}|\right) \\
&\quad = 2\,\mathrm{E}\big\{P\big[\,|\widetilde{x}_{t-1} - \widetilde{x}_{t-s}| > \left|\widetilde{x}_{t-s}\right| \bigm| \widetilde{x}_{t-s}\big]\big\} \\
&\quad \leq 2\,\mathrm{E}\left(\frac{1}{|\widetilde{x}_{t-s}|^{\alpha}}\,\mathrm{E}\left(\,|\widetilde{x}_{t-1} - \widetilde{x}_{t-s}|^{\alpha}|\,\widetilde{x}_{t-s}\right)\right)
\end{aligned}$$

where the generalized Markov inequality with some $0 < \alpha < 1$ was used in the last step (conditional on $\widetilde{x}_{t-s}$). Moreover,

$$\widetilde{x}_{t-1} - \widetilde{x}_{t-s} = x_{t-1} - x_{t-s} + \frac{s}{(t-s)(t-1)}\sum_{j=1}^{t-s} x_j - \frac{1}{t-1}\sum_{j=t-s+1}^{t-1} x_j,$$

so $\mathrm{E}\big(|\widetilde{x}_{t-1} - \widetilde{x}_{t-s}|^{\alpha}\big) = O\big(s^{\alpha/2}\big)$. If $T^{-0.5}\widetilde{x}_{t-s}$ has bounded pdf in the neighborhood of the origin (uniformly in t), as argued in the proof of item B, it can be shown that $\mathrm{E}\left(\frac{T^{\alpha/2}}{|\widetilde{x}_{t-s}|^{\alpha}}\right)$ is bounded uniformly in s (and t) as follows:

Let $g_t(x)$ be the relevant density functions; since $\exists\, C > 0$ such that $\sup_{|x|\leq C, \forall t} g_t(x) < C^* < \infty$, write

$$\begin{aligned}
\mathrm{E}\big(|x|^{-\alpha}\big) &= \int_{-C}^{C} \frac{1}{|x|^{\alpha}} f(x)\mathrm{d}x + \int_{C}^{\infty} \frac{1}{|x|^{\alpha}}(f(x) + f(-x))\mathrm{d}x \\
&\leq C^* f(x)\int_{-C}^{C} \frac{1}{|x|^{\alpha}}\mathrm{d}x + \frac{2}{C^{\alpha}}.
\end{aligned}$$

The uniform bound of the expectation follows due to finiteness of the improper integral for $0 < \alpha < 1$.

Hence, there exists a constant C such that

$$P\left(|\text{sgn}\left(\widetilde{x}_{t-1}\right) - \text{sgn}\left(\widetilde{x}_{t-s}\right)| = 2\right) \leq C\left(\frac{s}{T}\right)^{\alpha/2};$$

note also that $\exists \kappa > 0$ such that $\left(\frac{1}{4} + \frac{1}{r}\right)\frac{2}{1-\kappa} < 1$, so choose α in between. Take now e.g. $s = [T^{\kappa}]$ to obtain after some algebra

$$\sum_{t=p+2}^{T} |\text{sgn}\left(\widetilde{x}_{t-1}\right) - \text{sgn}\left(\widetilde{x}_{t-s}\right)|^{r/(r-1)} = o_p\left(T^{0.75-1/r}\right)$$

leading to the desired bound uniformly: one can always choose $\kappa \in (\delta; 1/2 - 2/r)$.

F. We shall only prove this for the demeaning case; the proof for detrending is established by a straightforward modification of the arguments. Recall, $\widetilde{x}_{t-1} = x_{t-1} - \frac{1}{t-1}\sum_{j=1}^{t-1} x_j$. Then,

$$\begin{aligned}\Delta\widetilde{x}_{t-1} &= \Delta x_{t-1} - \frac{1}{t-1}\sum_{j=1}^{t-1} x_j + \frac{1}{t-2}\sum_{j=1}^{t-2} x_j \\ &= \Delta x_{t-1} - \frac{x_{t-1}}{t-1} + \left(\frac{1}{t-2} - \frac{1}{t-1}\right)\sum_{j=1}^{t-2} x_j\end{aligned}$$

Hence,

$$\mathrm{E}\big(|\Delta\widetilde{x}_{t-1} - \Delta x_{t-1}|\big) \leq Ct^{-0.5} \tag{5}$$

since $\mathrm{E}\big(|x_{t-1}|\big) \leq Ct^{0.5}$. Then, write

$$\widetilde{x}_t^2 = \widetilde{x}_{t-1}^2 + 2\widetilde{x}_{t-1}\Delta\widetilde{x}_t + \Delta\widetilde{x}_t^2;$$

by summing over $t = 1, \ldots, T$ and rearranging the terms, it follows that

$$\sum_{t=2}^{T} \widetilde{x}_{t-1}\Delta\widetilde{x}_t = \frac{1}{2}\left(\widetilde{x}_T^2 - \widetilde{x}_1^2 - \sum_{t=2}^{T}\Delta\widetilde{x}_t^2\right) = O_p(T).$$

At this point, note that $\sum \widetilde{x}_{t-1}\Delta x_{t-j} = \sum\left(\widetilde{x}_{t-j-1} + \sum_{i=t-j}^{t-1}\Delta\widetilde{x}_i\right)\Delta x_{t-j}$. The result follows if

$$\sum\left(\sum_{i=t-j}^{t-1}\Delta\widetilde{x}_i\Delta x_{t-j}\right) = O_p(T)$$

We know from Chang and Park (2002, Lemma 3.2(b)) that $\sum\big(\sum_{i=t-j}^{t-1}\Delta x_i\Delta x_{t-j}\big)$ is $O_p(T)$, magnitude holding uniformly in j for $j < K_T$ even when $K_T = o\left(T^{0.5}\right)$; and it is straightforward to check that it holds under the local alternative as well. Moreover, (5) shows the difference between $\Delta\widetilde{x}_{t-1}$ and Δx_{t-1} to be negligible given our moment conditions, thus completing the proof.

Proof of Proposition 1

After recursively demeaning the levels and building differences, it holds that $\widetilde{y}_t^\mu = \widetilde{x}_t^\mu$ and $\Delta y_t = \Delta x_t$. Then, denoting $\mathbf{z}_{t-1} = (\Delta x_{t-1}, \ldots, \Delta x_{t-p})'$, we have for the IV estimator and IV t-type statistic

$$\begin{pmatrix}\widehat{\phi}\\ \widehat{\mathbf{a}}\end{pmatrix} = \begin{pmatrix}\sum h\left(\widetilde{x}_{t-1}^\mu\right)\widetilde{x}_{t-1}^\mu & \sum h\left(\widetilde{x}_{t-1}^\mu\right)\mathbf{z}_{t-1}'\\ \sum \mathbf{z}_{t-1}\widetilde{x}_{t-1}^\mu & \sum\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\end{pmatrix}^{-1}\begin{pmatrix}\sum h\left(\widetilde{x}_{t-1}^\mu\right)\Delta x_t\\ \sum\mathbf{z}_{t-1}\Delta x_t\end{pmatrix}.$$

Usual IV algebra leads to

$$\begin{aligned}\begin{pmatrix}\widehat{\phi}\\ \widehat{\mathbf{a}}-\mathbf{a}\end{pmatrix} &= \begin{pmatrix}\sum h\left(\widetilde{x}_{t-1}^\mu\right)\widetilde{x}_{t-1}^\mu & \sum h\left(\widetilde{x}_{t-1}^\mu\right)\mathbf{z}_{t-1}'\\ \sum \mathbf{z}_{t-1}\widetilde{x}_{t-1}^\mu & \sum\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\end{pmatrix}^{-1}\begin{pmatrix}\sum h\left(\widetilde{x}_{t-1}^\mu\right)\varepsilon_t\\ \sum\mathbf{z}_{t-1}\varepsilon_t\end{pmatrix}\\ &\quad -\frac{c}{T}\begin{pmatrix}\sum h\left(\widetilde{x}_{t-1}^\mu\right)\widetilde{x}_{t-1}^\mu & \sum h\left(\widetilde{x}_{t-1}^\mu\right)\mathbf{z}_{t-1}'\\ \sum \mathbf{z}_{t-1}\widetilde{x}_{t-1}^\mu & \sum\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\end{pmatrix}^{-1}\begin{pmatrix}\sum h\left(\widetilde{x}_{t-1}^\mu\right)x_{t-1}\\ \sum\mathbf{z}_{t-1}x_{t-1}\end{pmatrix}.\end{aligned}$$

Premultiply now both sides of the above equation by the diagonal matrix $\boldsymbol{L}_T = \mathrm{diag}\big(T, \sqrt{T}, \ldots, \sqrt{T}\big)$, and "insert" $\boldsymbol{R}_T\boldsymbol{R}_T^{-1}$, $\boldsymbol{R}_T = \mathrm{diag}\big(\sqrt{T}, \sqrt{T}, \ldots, \sqrt{T}\big)$ in the two terms on the r.h.s.; given the magnitude orders of the involved sample cross-product moments, it is immediately seen that $\widehat{\phi}$ is superconsistent, and that $\widehat{\mathbf{a}}$ is $\sqrt{T}$-consistent, just like in the ADF case. In fact, the matrix

$$\boldsymbol{L}_T^{-1}\begin{pmatrix}\sum h\left(\widetilde{x}_{t-1}^\mu\right)\widetilde{x}_{t-1}^\mu & \sum h\left(\widetilde{x}_{t-1}^\mu\right)\mathbf{z}_{t-1}'\\ \sum \mathbf{z}_{t-1}\widetilde{x}_{t-1}^\mu & \sum\mathbf{z}_{t-1}\mathbf{z}_{t-1}'\end{pmatrix}\boldsymbol{R}_T^{-1}$$

converges in distribution to a lower triangular matrix.

This also implies "consistency" of the residual variance estimator

$$\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=p+2}^{T}\left(\Delta y_t - \widehat{\phi}\widetilde{y}^{\mu}_{t-1} - \widehat{\mathbf{a}}'\mathbf{z}_{t-1}\right)^2$$

as follows. With $\Delta y_t - \widehat{\phi}\widetilde{y}^{\mu}_{t-1} - \widehat{\mathbf{a}}'\mathbf{z}_{t-1} = \varepsilon_t + o_p(1)$ and thus $\widehat{\sigma}^2 = \frac{1}{T}\sum \varepsilon_t^2 + o_p(1)$, it follows from Cavaliere and Taylor (2009, Corollary A.1) that

$$\widehat{\sigma}^2 = \overline{\omega}^2 + o_p(1). \tag{6}$$

Then, the same standard algebra leads to

$$t^{\mu}_{IV} = \frac{\frac{1}{\sqrt{T}}A_T}{\widehat{\sigma}\sqrt{\frac{1}{T}B_T}},$$

with

$$\begin{aligned} A_T &= \sum_{t=p+2}^{T} h\left(\widetilde{x}^{\mu}_{t-1}\right)\left(\varepsilon_t - \frac{c}{T}x_{t-1}\right) \\ &\quad - \sum_{t=p+2}^{T} h\left(\widetilde{x}^{\mu}_{t-1}\right)\mathbf{z}'_{t-1}\left(\sum_{t=p+2}^{T}\mathbf{z}_{t-1}\mathbf{z}'_{t-1}\right)^{-1}\sum_{t=p+2}^{T}\mathbf{z}_{t-1}\left(\varepsilon_t - \frac{c}{T}x_{t-1}\right) \end{aligned}$$

and

$$\begin{aligned} B_T &= \sum_{t=p+2}^{T} h^2\left(\widetilde{x}^{\mu}_{t-1}\right) \\ &\quad - \sum_{t=p+2}^{T} h\left(\widetilde{x}^{\mu}_{t-1}\right)\mathbf{z}'_{t-1}\left(\sum_{t=p+2}^{T}\mathbf{z}_{t-1}\mathbf{z}'_{t-1}\right)^{-1}\sum_{t=p+2}^{T}\mathbf{z}_{t-1}h\left(\widetilde{x}^{\mu}_{t-1}\right). \end{aligned}$$

It is relatively straightforward to show that $B_T = T + o_p(T)$ as follows. First, the probability of $\widetilde{x}^{\mu}_{t-1}$ to belong to $[-m, m]$ is of order $t^{-0.5}$, and as such $h\left(\widetilde{x}^{\mu}_{t-1}\right) = \operatorname{sgn}\left(\widetilde{x}^{\mu}_{t-1}\right) + O_p\left(t^{-0.5}\right)$. See also Lemma 4 B. Second, $\sum \mathbf{z}_{t-1}\mathbf{z}'_{t-1}$ is $\Theta_p(T)$, and third, $\sum h\left(\widetilde{x}^{\mu}_{t-1}\right)\mathbf{z}'_{t-1} = o_p\left(T^{0.75}\right)$ which is proved in Lemma 4 item E above.

Further, from $\left[h\left(\widetilde{x}^{\mu}_{t-1}\right) - \operatorname{sgn}\left(\widetilde{x}^{\mu}_{t-1}\right)\right]^2 = O_p\left(t^{-1}\right)$ we have $\sum_{t=p+2}^{T}\left[h\left(\widetilde{x}^{\mu}_{t-1}\right) - \operatorname{sgn}\left(\widetilde{x}^{\mu}_{t-1}\right)\right]^2 = O_p(\ln T)$ and hence, by Cauchy-Schwarz, for any $\gamma > 0$,

$$\begin{aligned} -\frac{c}{T}\sum_{t=p+2}^{T}\left[h\left(\widetilde{x}^{\mu}_{t-1}\right) - \operatorname{sgn}\left(\widetilde{x}^{\mu}_{t-1}\right)\right]x_{t-1} &\leqslant -c\sqrt{\sum_{t=p+2}^{T}\left[h\left(\widetilde{x}^{\mu}_{t-1}\right) - \operatorname{sgn}\left(\widetilde{x}^{\mu}_{t-1}\right)\right]^2\frac{1}{T^2}\sum_{t=p+2}^{T}x^2_{t-1}} \\ &= o_p(T^{\gamma})O_p(1) = o_p(T^{\gamma}) = O_p(T^{0.25}) \end{aligned}$$

Using item C of Lemma 4, we then obtain that

$$\sum_{t=p+2}^{T} h\left(\widetilde{x}^{\mu}_{t-1}\right)\left(\varepsilon_t - \frac{c}{T}x_{t-1}\right) = \sum_{t=p+2}^{T}\operatorname{sgn}\left(\widetilde{x}^{\mu}_{t-1}\right)\left(\varepsilon_t - \frac{c}{T}x_{t-1}\right) + O_p(T^{0.25}).$$

Further, $\sum_{t=p+2}^{T}\mathbf{z}_{t-1}\left(\varepsilon_t - \frac{c}{T}x_{t-1}\right) = O_p(T^{1/2}) - \frac{c}{T}\sum_{t=p+2}^{T}\mathbf{z}_{t-1}x_{t-1}$. By Cauchy-Schwarz,

$$T^{-3/2}\sum_{t=p+2}^{T}\mathbf{z}_{t-1}x_{t-1} \leqslant \sqrt{\frac{1}{T}\sum_{t=p+2}^{T}\mathbf{z}'_{t-1}\mathbf{z}_{t-1}\frac{1}{T^2}\sum_{t=p+2}^{T}x^2_{t-1}} = O_p(1)$$

Thus, the second summand of A_T is $o_p(T^{0.75})O_p(T^{-1})O_p(T^{1/2}) = o_p(T^{0.25})$. Then, item F of the lemma can also be used to show that

$$\begin{aligned} A_T &= \sum_{t=p+2}^{T} \operatorname{sgn}\left(\widetilde{x}_{t-1}^{\mu}\right)\left(\varepsilon_t - \frac{c}{T}x_{t-1}\right) + O_p\left(T^{0.25}\right) \\ &= \sum_{t=p+2}^{T} \operatorname{sgn}\left(\frac{A(1)}{\overline{\omega}}\frac{\widetilde{x}_{t-1}^{\mu}}{\sqrt{T}}\right)\varepsilon_t \\ &\quad -\frac{\overline{\omega}}{A(1)}\frac{c}{\sqrt{T}}\sum_{t=p+2}^{T} \operatorname{sgn}\left(\frac{A(1)}{\overline{\omega}}\frac{\widetilde{x}_{t-1}^{\mu}}{\sqrt{T}}\right)\frac{A(1)}{\overline{\omega}}\frac{1}{\sqrt{T}}x_{t-1} + o_p\left(\sqrt{T}\right). \end{aligned}$$

The numerator of t_{IV}^{μ}, $A_T/\sqrt{T}$, converges to the desired limit as follows. With the arguments used by So and Shin (1999) in the proof of their Theorem 1(ii), the first summand has $\overline{\omega}\int_0^1 \operatorname{sgn}\big(\widetilde{J}_c^{\mu}\left(\eta\left(s\right)\right)\big)dW\left(\eta\left(s\right)\right)$ as weak limit, and convergence of the second summand follows immediately due to the CMT. Considering that $B_T/T \xrightarrow{p} 1$, the convergence of $\widehat{\sigma}$ to $\overline{\omega}$ (cf. Eq. 6) completes the result.

Proof of Proposition 3

If a linear trend is present in the data, i.e. $d_t = \mu + \tau\, t$ and $\tau \neq 0$, recursive demeaning does not remove it. In fact,

$$\widetilde{y}_t^{\mu} = \widetilde{x}_t^{\mu} + \tau\left(\frac{t}{2} - 1\right),$$

so $\widetilde{y}_t^{\mu}/t = O_p(1)$ uniformly in $1 \leq t \leq T$, with the linear trend dominating the stochastic one. Also, $\mathrm{E}(\Delta y_t) = \tau$. This affects of course the quantities A_T and B_T from the proof of Proposition 1, but not the magnitude order of the residual variance estimator.

More precisely, we have with $\mathbf{z}_{t-1}^{*} = (\tau + \Delta x_{t-1}, \ldots, \tau + \Delta x_{t-p})'$ that

$$\begin{pmatrix} \widehat{\phi} \\ \widehat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \sum h\left(\widetilde{y}_{t-1}^{\mu}\right)\widetilde{y}_{t-1}^{\mu} & \sum h\left(\widetilde{y}_{t-1}^{\mu}\right)\mathbf{z}_{t-1}^{*}{}' \\ \sum \mathbf{z}_{t-1}^{*}\widetilde{y}_{t-1}^{\mu} & \sum \mathbf{z}_{t-1}^{*}\mathbf{z}_{t-1}^{*}{}' \end{pmatrix}^{-1} \begin{pmatrix} \sum h\left(\widetilde{y}_{t-1}^{\mu}\right)\Delta y_t \\ \sum \mathbf{z}_{t-1}^{*}\Delta y_t \end{pmatrix}. \tag{7}$$

Note that the following exact magnitude orders for cross-products hold true (due to the non-zero mean of the differences and the and the linear trend dominating the behavior of $\widetilde{y}_t^{\mu}$):

$$\begin{aligned} \sum_{t=p+2}^{T} h\left(\widetilde{y}_{t-1}^{\mu}\right)\mathbf{z}_{t-1}^{*} &= \Theta_p\left(T\right) \\ \sum_{t=p+2}^{T} \mathbf{z}_{t-1}^{*}\widetilde{y}_{t-1}^{\mu} &= \Theta_p\left(T^2\right) \\ \sum_{t=p+2}^{T} \mathbf{z}_{t-1}^{*}\mathbf{z}_{t-1}^{*}{}' &= \Theta_p\left(T\right) \\ \sum_{t=p+2}^{T} \mathbf{z}_{t-1}^{*}\Delta y_t &= \Theta_p\left(T\right) \end{aligned}$$

and in particular

$$\begin{aligned} \sum_{t=p+2}^{T} h\left(\widetilde{y}_{t-1}^{\mu}\right)\widetilde{y}_{t-1}^{\mu} &= \Theta_p\left(T^2\right) \\ \sum_{t=p+2}^{T} h\left(\widetilde{y}_{t-1}^{\mu}\right)\Delta y_t &= \Theta_p\left(T\right). \end{aligned}$$

Similarly to the proof of Proposition 1, premultiply both sides of Equation (7) with the (suitably redefined) diagonal matrix $\boldsymbol{L}_T = \text{diag}\,(T, 1, \ldots, 1)$, and insert $\boldsymbol{R}_T \boldsymbol{R}_T^{-1}$, $\boldsymbol{R}_T = \text{diag}\big(T, T, \ldots, T\big)$, in its r.h.s., to obtain that $T\widehat{\phi} = O_p(1)$ and $\widehat{\mathbf{a}} = O_p(1)$, which ensures that $\widehat{\sigma}^2 = O_p(1)$. Furthermore, it is straightforward to show that

$$B_T = O_p(T);$$

but A_T now contains, due to the non-removal of the mean of Δy_t, a term of the type $C \sum h\left(\widetilde{y}_{t-1}^{\mu}\right)$ which is $\Theta_p(T)$ and the t-type statistic diverges even under the null hypothesis of a unit root.

Proof of Proposition 4

The result is obtained along the lines of the proof of Proposition 1. In fact, one has to show that the terms involving differences vanish asymptotically and then resort to the same weak convergence arguments, but now accounting for the effects of demeaning the differences Δx_t.

For that, we need to check the validity of results analogous to Lemma 4, items E and F when demeaning Δx_t. This is indeed the case when $\sqrt{T}$-consistent demeaning is applied. Item E reduces thanks to Item B to

$$\sum \text{sgn}\left(\widetilde{x}_{t-1}\right) \left(\Delta x_t - \overline{\Delta x}\right) = \sum \text{sgn}\left(\widetilde{x}_{t-1}\right) \Delta x_t - \overline{\Delta x} \sum \text{sgn}\left(\widetilde{x}_{t-1}\right);$$

the second term on the r.h.s. is $O_p(T^{0.5})$, and the first term dominates, as required. For item F, write

$$\sum \widetilde{x}_{t-1} \widetilde{\Delta x}_{t-j} = \sum \widetilde{x}_{t-1} \Delta x_{t-j} - \overline{\Delta x} \sum \widetilde{x}_{t-1} = O_p\left(T\right)$$

since $\widetilde{x}_{t-1} = O_p(T^{0.5})$. The first part follows.

The second part is itself immediately established with the CMT by noting that, when including an intercept in the error correction model, $B_T = \sum \left(\text{sgn}\big(\widetilde{x}_{t-1}^{\tau}\big) - \overline{\text{sgn}\left(\widetilde{x}^{\tau}\right)}\right)^2 + o_p(T)$ instead of $T + o_p(T)$.

Proof of Proposition 6

Let us first analyze the behavior of the sample covariance matrix of $\overline{\boldsymbol{\varepsilon}}_t$. Namely, we shall prove that

$$\left\| \frac{1}{T} \sum_{t=p+2}^{T} \overline{\boldsymbol{\varepsilon}}_t \overline{\boldsymbol{\varepsilon}}_t' - \overline{\boldsymbol{\Omega}} \right\| = O_p\left(NT^{-0.25}\right).$$

It actually suffices to show that $T^{-1} \sum \varepsilon_{i,t} \varepsilon_{j,t}$ is $\sqrt[4]{T}$-consistent at a uniform rate over i, j; the norm of an $N \times N$ matrix with bounded elements is known to be $O\left(N\right)$. In order to establish the consistency at a uniform rate of the sample covariances, we make use of the factor structure of the innovations. We namely have that

$$\begin{aligned}
\frac{1}{T} \sum_{t=p+2}^{T} \varepsilon_{i,t} \varepsilon_{j,t} &= \frac{1}{T} \sum_{t=p+2}^{T} \left(\boldsymbol{\lambda}_i' \boldsymbol{\nu}_t + \widetilde{\varepsilon}_{i,t}\right) \left(\boldsymbol{\lambda}_j' \boldsymbol{\nu}_t + \widetilde{\varepsilon}_{j,t}\right) \\
&= \frac{1}{T} \sum_{t=p+2}^{T} \boldsymbol{\lambda}_i' \boldsymbol{\nu}_t \boldsymbol{\nu}_t' \boldsymbol{\lambda}_j + \frac{1}{T} \sum_{t=p+2}^{T} \boldsymbol{\lambda}_i' \boldsymbol{\nu}_t \widetilde{\varepsilon}_{j,t} + \frac{1}{T} \sum_{t=p+2}^{T} \widetilde{\varepsilon}_{i,t} \boldsymbol{\lambda}_j' \boldsymbol{\nu}_t + \frac{1}{T} \sum_{t=p+2}^{T} \widetilde{\varepsilon}_{i,t} \widetilde{\varepsilon}_{j,t}.
\end{aligned}$$

The sample variance of the cross-products $\boldsymbol{\nu}_t \widetilde{\varepsilon}_{j,t}$ and $\widetilde{\varepsilon}_{i,t} \widetilde{\varepsilon}_{j,t}$ for $i \neq j$ vanishes at rate T^{-1} due to the independence of the factors; their expectation is 0. Furthermore,

$$\frac{1}{T} \sum_{t=p+2}^{T} \widetilde{\varepsilon}_{i,t}^2 \xrightarrow{p} \overline{\omega}_i^2$$

and

$$\frac{1}{T} \sum_{t=p+2}^{T} \boldsymbol{\nu}_t \boldsymbol{\nu}_t' \xrightarrow{p} \int_0^1 \boldsymbol{\Omega}\left(s\right) ds$$

along the lines of Corollary A.1 of Cavaliere and Taylor (2009). The uniform boundedness of ω_i across the panel ensures a uniform rate of convergence to $\overline{\omega}_i^2$, but it is the summability condition from Assumption 6 that guarantees the variance of the above sample variances to vanish at rate $T^{0.5}$ uniformly in i. Thus the sample covariances of $\boldsymbol{\varepsilon}_t$ are $T^{-0.25}$-consistent for the respective elements of $\overline{\boldsymbol{\Omega}}$.

Moving on to the main part of the proof, we have that

$$\overline{\tau}_{IV} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\sum_{t=p+2}^{T} h_i\left(\widetilde{y}_{i,t-1}^{\mu}\right) \varepsilon_{i,t}^{*}}{\sqrt{\sum_{t=p+2}^{T} h_i^2\left(\widetilde{y}_{i,t-1}^{\mu}\right)}}.$$

With Lemma 4 B we have that $h_i^2\left(\widetilde{y}_{i,t-1}^{\mu}\right) = 1 + O_p(t^{-0.5})$. Since $\sum_{t=p+2}^{T} h_i\left(\widetilde{y}_{i,t-1}^{\mu}\right) \varepsilon_{i,t}^{*} = O_p(\sqrt{T})$, it follows that

$$\frac{\sum_{t=p+2}^{T} h_i\left(\widetilde{y}_{i,t-1}^{\mu}\right) \varepsilon_{i,t}^{*}}{\sqrt{\sum_{t=p+2}^{T} h_i^2\left(\widetilde{y}_{i,t-1}^{\mu}\right)}} = \frac{1}{\sqrt{T}} \sum_{t=p+2}^{T} h_i\left(\widetilde{y}_{i,t-1}^{\mu}\right) \varepsilon_{i,t}^{*} + O_p(T^{-0.5});$$

furthermore,

$$\begin{aligned}
\overline{\tau}_{IV} &= \frac{1}{\sqrt{TN}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} h_i\left(\widetilde{y}_{i,t-1}^{\mu}\right) \varepsilon_{i,t}^{*} + O_p\left(T^{-0.5} N^{0.5}\right) \\
&= \frac{1}{\sqrt{TN}} \sum_{i=1}^{N} \sum_{t=p+2}^{T} h_i\left(\widetilde{y}_{i,t-1}^{\mu}\right) \varepsilon_{i,t}^{*} + o_p\left(1\right) \\
&= \frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \mathbf{h}_t' \widehat{\boldsymbol{\Gamma}}' \overline{\boldsymbol{\varepsilon}}_t + o_p\left(1\right)
\end{aligned}$$

with obvious notation $\mathbf{h}_t = \left(h_i\left(\widetilde{y}_{i,t-1}^{\mu}\right)\right)_{i=1,\ldots,N}'$. The effect of the difference between $\widehat{\boldsymbol{\Gamma}}'$ and $\boldsymbol{\Gamma}'$ (the LU decomposition of $\overline{\boldsymbol{\Omega}}^{-1}$) is quantified as follows:

$$\frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \mathbf{h}_t' \widehat{\boldsymbol{\Gamma}}' \overline{\boldsymbol{\varepsilon}}_t = \frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \mathbf{h}_t' \boldsymbol{\Gamma}' \overline{\boldsymbol{\varepsilon}}_t + \frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \mathbf{h}_t' \big(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}'\big) \overline{\boldsymbol{\varepsilon}}_t.$$

Since the instruments are uniformly bounded, we have, defining $\widetilde{\boldsymbol{\imath}}' = \boldsymbol{\imath}_N' \big(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}'\big)$ with $\boldsymbol{\imath}_N$ a vector of N ones, that

$$\left\| \frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \mathbf{h}_t' \left(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}'\right) \overline{\boldsymbol{\varepsilon}}_t \right\| \leq C \left\| \frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \widetilde{\boldsymbol{\imath}}' \overline{\boldsymbol{\varepsilon}}_t \right\| \leq C \frac{1}{\sqrt{N}} \left\| \widetilde{\boldsymbol{\imath}}' \right\| \left\| \frac{1}{\sqrt{T}} \sum_{t=p+2}^{T} \overline{\boldsymbol{\varepsilon}}_t \right\|.$$

The latter norm on the r.h.s. of the above inequality is $O_p\left(N^{0.5}\right)$, since $\overline{\varepsilon}_{i,t} = \varepsilon_{i,t} + O_p(T^{-0.5})$ (whether estimating by imposing the unit root or not) and the elements of the vector are uniformly bounded in probability. The norm $\|\boldsymbol{\imath}_N\|$ is $O\left(N^{0.5}\right)$; hence the norm $\|\widetilde{\boldsymbol{\imath}}'\|$ is $O_p\left(N^{0.5}T^{-0.25}\right)$ because the $T^{-0.25}$-consistency at a uniform rate of the sample covariance matrix translates into the same uniform convergence rate for the elements of $\widehat{\boldsymbol{\Gamma}}$, and $\widehat{\boldsymbol{\Gamma}}$ is the LU decomposition of a matrix whose norm is $O_p(N)$ so $\|\boldsymbol{\Gamma}'\| = O\big(\sqrt{N}\big)$ and $\|\widehat{\boldsymbol{\Gamma}}'\| = O_p\big(\sqrt{N}\big)$. Summing up,

$$\frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \mathbf{h}_t' \big(\widehat{\boldsymbol{\Gamma}}' - \boldsymbol{\Gamma}'\big) \overline{\boldsymbol{\varepsilon}}_t = O_p\left(\frac{N}{T^{0.25}}\right) = o_p\left(1\right).$$

For the term $\frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \mathbf{h}_t' \boldsymbol{\Gamma}' \overline{\boldsymbol{\varepsilon}}_t$ we use Lemma 4 E to conclude that

$$\overline{\tau}_{IV} = \frac{1}{\sqrt{TN}} \sum_{t=p+2}^{T} \boldsymbol{\imath}_N' \boldsymbol{\Gamma}' \boldsymbol{\varepsilon}_t + o_p\left(N^{0.5}T^{0.25}\right).$$

The quantity $N^{-0.5}\boldsymbol{\iota}_N'\boldsymbol{\Gamma}'\boldsymbol{\varepsilon}_t$ is a md array; given the finiteness of its 4th order moments (cf. Assumption 6), the second condition of the CLT for md arrays, Theorem 24.3 of Davidson (1994), is fulfilled; checking the first condition amounts to showing that $T^{-1}\sum_{t=p+2}^{T}\left(N^{-0.5}\boldsymbol{\iota}_N'\boldsymbol{\Gamma}'\boldsymbol{\varepsilon}_t\right)^2 \xrightarrow{p} 1$ as follows. Write the quantity as

$$\frac{1}{N}\boldsymbol{\iota}_N'\boldsymbol{\Gamma}'\frac{1}{T}\sum_{t=p+2}^{T}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{\Gamma}\boldsymbol{\iota}_N$$

and recall that $\left\|\frac{1}{T}\sum_{t=p+2}^{T}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' - \left(\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\right)^{-1}\right\| = O_p\left(T^{-0.25}N\right)$ from which the result follows.

Appendix B Additional Simulation Results

Table B.1: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T \ N	6	16	26	46	6	16	26	46	6	16	26	46	6	16	26	46
$\delta = 1/5$																	
	50	.086	.075	.056	.038	.430	.659	.668	.571	.073	.063	.064	.041	.420	.637	.663	.556
$\overline{\tau}_{IV}$	100	.081	.084	.068	.057	.761	.970	.996	1.00	.083	.078	.081	.061	.757	.970	.992	.999
	200	.070	.072	.078	.069	.967	1.00	1.00	1.00	.073	.080	.078	.066	.963	1.00	1.00	1.00
	50	.108	.124	.126	.056	.382	.639	.690	.537	.087	.110	.123	.052	.372	.596	.659	.511
P_{IV}	100	.096	.136	.142	.167	.728	.956	.992	1.00	.098	.112	.130	.147	.713	.948	.985	.998
	200	.091	.112	.144	.148	.958	1.00	1.00	1.00	.082	.116	.123	.136	.955	1.00	1.00	1.00
	50	.114	.116	.106	.095	.312	.464	.559	.638	.097	.107	.111	.111	.275	.424	.481	.537
$t_{\hat{\xi}^*,\kappa}$	100	.090	.109	.102	.097	.562	.773	.848	.929	.098	.098	.108	.106	.519	.720	.809	.888
	200	.089	.094	.099	.070	.869	.987	.998	1.00	.081	.090	.090	.084	.856	.978	.993	1.00
$\delta = 5$																	
	50	.044	.033	.035	.040	.104	.137	.103	.073	.041	.030	.032	.031	.135	.152	.133	.092
$\overline{\tau}_{IV}$	100	.044	.042	.041	.038	.412	.718	.834	.852	.049	.043	.039	.041	.406	.566	.629	.656
	200	.048	.054	.052	.042	.847	.997	.999	1.00	.049	.056	.050	.046	.778	.938	.979	.995
	50	.036	.033	.035	.032	.069	.105	.107	.057	.026	.026	.030	.030	.106	.163	.172	.092
P_{IV}	100	.041	.041	.042	.043	.299	.552	.684	.760	.041	.030	.028	.035	.366	.562	.674	.770
	200	.045	.047	.047	.048	.768	.988	.999	1.00	.046	.045	.043	.036	.797	.967	.988	.999
	50	.037	.022	.016	.010	.060	.037	.016	.005	.042	.054	.058	.050	.090	.070	.079	.067
$t_{\hat{\xi}^*,\kappa}$	100	.045	.033	.034	.019	.164	.167	.141	.097	.059	.071	.076	.077	.238	.265	.259	.253
	200	.048	.042	.030	.025	.507	.685	.764	.863	.058	.089	.103	.093	.647	.773	.825	.857

Nominal 5% level. 5000 replications. $\tau = 0.5$.

Table B.2: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T N	6	16	26	46	6	16	26	46	6	16	26	46	6	16	26	46
$\delta = 1/5$																	
	50	.071	.077	.070	.064	.236	.402	.514	.512	.071	.058	.059	.060	.233	.330	.415	.381
$\overline{\tau}_{IV}$	100	.064	.073	.068	.070	.465	.817	.949	.993	.063	.065	.067	.057	.465	.731	.849	.943
	200	.067	.059	.045	.075	.830	.993	1.00	1.00	.067	.062	.073	.057	.829	.986	.999	1.00
	50	.086	.102	.099	.095	.235	.372	.464	.492	.073	.079	.081	.079	.218	.315	.396	.382
P_{IV}	100	.075	.094	.093	.108	.440	.779	.912	.984	.068	.084	.083	.084	.426	.710	.835	.933
	200	.076	.068	.067	.107	.811	.992	1.00	1.00	.064	.077	.090	.087	.810	.983	1.00	1.00
	50	.087	.103	.104	.103	.203	.344	.427	.559	.077	.099	.102	.132	.193	.266	.324	.355
$t_{\hat{\xi}^*,\kappa}$	100	.076	.087	.088	.084	.372	.596	.679	.783	.076	.084	.100	.116	.355	.513	.580	.642
	200	.078	.066	.054	.056	.662	.892	.948	.991	.069	.089	.099	.110	.675	.877	.932	.977
$\delta = 5$																	
	50	.042	.045	.034	.031	.338	.623	.732	.668	.044	.046	.042	.043	.290	.452	.500	.436
$\overline{\tau}_{IV}$	100	.054	.049	.060	.046	.748	.983	.998	1.00	.061	.044	.050	.051	.616	.875	.942	.981
	200	.052	.048	.068	.048	.972	1.00	1.00	1.00	.052	.052	.062	.055	.913	.996	.999	1.00
	50	.033	.034	.028	.033	.227	.411	.521	.546	.037	.036	.037	.046	.213	.334	.376	.402
P_{IV}	100	.044	.040	.053	.047	.624	.940	.990	.998	.046	.046	.050	.047	.541	.826	.923	.973
	200	.051	.048	.059	.049	.954	1.00	1.00	1.00	.047	.047	.046	.054	.906	.997	.999	1.00
	50	.035	.037	.021	.016	.113	.094	.076	.033	.048	.044	.052	.052	.119	.107	.077	.067
$t_{\hat{\xi}^*,\kappa}$	100	.045	.040	.036	.026	.343	.418	.421	.442	.059	.068	.068	.066	.324	.378	.385	.399
	200	.056	.043	.043	.027	.798	.954	.987	1.00	.059	.075	.077	.087	.772	.913	.940	.974

Nominal 5% level. 5000 replications. $\tau = 0.9$.

Table B.3: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests, heterogenous variance breaks

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T N	6	16	26	46	6	16	26	46	6	16	26	46	6	16	26	46
$\delta = 1/5$																	
	50	.077	.069	.058	.036	.441	.685	.781	.711	.073	.076	.052	.045	.425	.682	.750	.695
$\overline{\tau}_{IV}$	100	.075	.068	.064	.056	.778	.983	.997	1.00	.075	.077	.062	.062	.770	.978	.998	1.00
	200	.068	.074	.071	.068	.970	1.00	1.00	1.00	.065	.069	.074	.068	.970	1.00	1.00	1.00
	50	.088	.098	.096	.050	.388	.634	.736	.630	.076	.101	.082	.059	.345	.606	.693	.641
P_{IV}	100	.090	.110	.109	.126	.722	.972	.997	1.00	.082	.102	.101	.116	.719	.960	.994	.999
	200	.078	.104	.116	.124	.968	.999	1.00	1.00	.068	.086	.114	.122	.960	1.00	1.00	1.00
	50	.088	.090	.094	.087	.302	.443	.531	.582	.078	.098	.090	.094	.253	.393	.429	.473
$t_{\hat{\xi}^*,\kappa}$	100	.094	.093	.091	.080	.537	.744	.826	.926	.084	.095	.080	.090	.510	.701	.787	.866
	200	.080	.084	.080	.070	.882	.986	.998	1.00	.073	.078	.082	.089	.863	.982	.998	1.00
$\delta = 5$																	
	50	.038	.042	.031	.036	.101	.174	.187	.139	.041	.038	.038	.043	.154	.193	.196	.158
$\overline{\tau}_{IV}$	100	.050	.042	.042	.041	.355	.647	.793	.880	.045	.037	.050	.041	.366	.515	.580	.631
	200	.054	.050	.045	.045	.740	.984	1.00	1.00	.056	.058	.042	.040	.723	.878	.944	.976
	50	.034	.039	.032	.040	.072	.120	.146	.127	.026	.020	.033	.039	.122	.200	.236	.175
P_{IV}	100	.041	.040	.042	.046	.267	.480	.642	.754	.032	.028	.034	.037	.347	.580	.687	.774
	200	.047	.045	.042	.047	.680	.960	.995	1.00	.045	.042	.037	.027	.762	.942	.982	.998
	50	.026	.022	.012	.005	.058	.036	.020	.011	.046	.050	.043	.048	.088	.098	.078	.074
$t_{\hat{\xi}^*,\kappa}$	100	.039	.031	.026	.015	.179	.188	.171	.141	.057	.073	.083	.078	.248	.270	.292	.282
	200	.046	.035	.030	.018	.484	.642	.724	.839	.069	.090	.086	.078	.645	.764	.791	.836

Nominal 5% level. 5000 replications. $\tau = 0.5$.

Table B.4: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests, heterogenous variance breaks

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T N	6	16	26	46	6	16	26	46	6	16	26	46	6	16	26	46
$\delta = 1/5$																	
	50	.068	.066	.056	.050	.348	.602	.699	.680	.056	.065	.055	.045	.345	.578	.658	.639
$\overline{\tau}_{IV}$	100	.067	.084	.066	.052	.654	.959	.988	.999	.075	.065	.058	.066	.684	.942	.984	.998
	200	.066	.082	.080	.058	.941	1.00	1.00	1.00	.063	.072	.068	.057	.935	.999	1.00	1.00
	50	.072	.069	.065	.065	.271	.516	.594	.628	.055	.065	.059	.051	.276	.474	.552	.588
P_{IV}	100	.074	.089	.093	.084	.618	.932	.980	.997	.069	.080	.080	.083	.635	.916	.973	.997
	200	.078	.095	.102	.101	.930	.999	1.00	1.00	.070	.080	.087	.082	.922	.999	1.00	1.00
	50	.073	.080	.064	.062	.196	.304	.356	.378	.062	.060	.058	.069	.197	.258	.251	.267
$t_{\hat{\xi}^*,\kappa}$	100	.080	.085	.069	.069	.464	.664	.756	.842	.078	.082	.075	.089	.450	.620	.678	.766
	200	.075	.080	.071	.055	.795	.962	.992	.999	.072	.085	.081	.082	.806	.957	.984	.999
$\delta = 5$																	
	50	.034	.032	.036	.035	.099	.155	.188	.154	.051	.048	.044	.045	.173	.210	.212	.200
$\overline{\tau}_{IV}$	100	.046	.045	.044	.038	.390	.755	.893	.970	.050	.046	.050	.045	.438	.589	.689	.800
	200	.050	.050	.054	.044	.853	.995	1.00	1.00	.058	.056	.054	.044	.782	.942	.977	.998
	50	.028	.025	.032	.036	.080	.130	.168	.176	.040	.036	.040	.042	.152	.234	.260	.235
P_{IV}	100	.042	.037	.034	.040	.324	.635	.798	.926	.042	.032	.037	.038	.408	.636	.774	.884
	200	.054	.047	.039	.042	.818	.990	1.00	1.00	.052	.048	.040	.031	.816	.972	.996	1.00
	50	.026	.010	.011	.006	.067	.069	.064	.054	.041	.047	.055	.046	.102	.098	.091	.090
$t_{\hat{\xi}^*,\kappa}$	100	.039	.030	.020	.017	.235	.326	.334	.329	.066	.078	.088	.094	.268	.320	.318	.316
	200	.055	.044	.027	.022	.641	.840	.909	.976	.064	.080	.098	.082	.705	.833	.870	.896

Nominal 5% level. 5000 replications. $\tau = 0.9$.

Table B.5: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests with shrinkage

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T N	16	26	56	106	16	26	56	106	16	26	56	106	16	26	56	106
$\delta = 1/5$																	
	50	.107	.110	.131	.156	.775	.930	.996	1.00	.091	.101	.109	.093	.753	.887	.984	.998
$\overline{\tau}_{IV}$	100	.096	.096	.104	.132	.989	.998	1.00	1.00	.097	.092	.096	.088	.981	.998	1.00	1.00
	200	.085	.097	.094	.112	1.00	1.00	1.00	1.00	.072	.083	.082	.080	1.00	1.00	1.00	1.00
	50	.157	.180	.247	.300	.729	.893	.991	1.00	.100	.121	.104	.045	.667	.804	.933	.947
P_{IV}	100	.143	.168	.228	.330	.977	.997	1.00	1.00	.112	.113	.101	.054	.966	.994	1.00	1.00
	200	.123	.143	.187	.266	1.00	1.00	1.00	1.00	.092	.107	.094	.054	1.00	1.00	1.00	1.00
	50	.131	.129	.121	.110	.523	.647	.791	.903	.109	.120	.126	.136	.447	.550	.671	.738
$t_{\hat{\xi}^*,\kappa}$	100	.118	.104	.088	.085	.772	.873	.964	.995	.108	.108	.115	.124	.713	.809	.916	.965
	200	.100	.094	.073	.063	.982	.996	1.00	1.00	.092	.098	.095	.101	.977	.995	1.00	1.00
$\delta = 5$																	
	50	.055	.064	.082	.103	.550	.765	.978	1.00	.027	.021	.003	.001	.235	.236	.169	.069
$\overline{\tau}_{IV}$	100	.053	.069	.076	.096	.927	.990	1.00	1.00	.041	.036	.016	.008	.658	.771	.837	.835
	200	.060	.065	.076	.088	.999	1.00	1.00	1.00	.049	.054	.038	.033	.964	.990	1.00	1.00
	50	.030	.022	.023	.013	.274	.404	.679	.873	.007	.003	.000	.000	.105	.079	.005	.000
P_{IV}	100	.042	.048	.044	.039	.784	.929	.999	1.00	.023	.018	.000	.000	.585	.714	.750	.439
	200	.056	.053	.050	.053	.995	1.00	1.00	1.00	.039	.036	.025	.006	.974	.995	1.00	1.00
	50	.027	.015	.006	.002	.046	.032	.006	.000	.061	.055	.054	.051	.074	.070	.077	.062
$t_{\hat{\xi}^*,\kappa}$	100	.039	.033	.019	.011	.215	.201	.144	.079	.072	.081	.089	.076	.301	.287	.302	.298
	200	.042	.036	.022	.014	.752	.821	.942	.985	.085	.096	.098	.097	.805	.854	.888	.903

Nominal 5% level. 5000 replications. $\tau = 0.5$.

Table B.6: Size and Power of the Shin and Kang and Demetrescu et al. Panel Tests with shrinkage

		Independence								Factor Structure							
		Size				Power				Size				Power			
	T N	16	26	56	106	16	26	56	106	16	26	56	106	16	26	56	106
$\delta = 1/5$																	
	50	.094	.106	.128	.141	.498	.655	.920	.993	.062	.059	.041	.023	.391	.485	.655	.780
$\overline{\tau}_{IV}$	100	.076	.097	.107	.140	.843	.960	1.00	1.00	.064	.048	.035	.025	.765	.894	.981	.997
	200	.072	.076	.093	.110	.994	1.00	1.00	1.00	.060	.060	.042	.030	.986	.998	1.00	1.00
	50	.096	.095	.104	.060	.421	.541	.766	.877	.032	.019	.001	.000	.218	.206	.104	.026
P_{IV}	100	.082	.102	.119	.143	.780	.930	.999	1.00	.040	.022	.002	.000	.635	.737	.818	.821
	200	.075	.083	.120	.138	.990	1.00	1.00	1.00	.047	.043	.011	.000	.980	.996	1.00	1.00
	50	.081	.088	.092	.088	.306	.384	.506	.582	.082	.095	.119	.136	.240	.263	.295	.300
$t_{\hat{\xi}^*,\kappa}$	100	.073	.086	.069	.065	.544	.662	.804	.895	.093	.097	.120	.144	.491	.555	.666	.730
	200	.068	.058	.058	.042	.883	.944	.990	1.00	.084	.098	.122	.147	.877	.932	.983	.994
$\delta = 5$																	
	50	.052	.060	.070	.095	.760	.935	.999	1.00	.039	.031	.013	.001	.459	.539	.584	.509
$\overline{\tau}_{IV}$	100	.058	.063	.069	.092	.994	1.00	1.00	1.00	.058	.045	.039	.018	.883	.951	.989	.998
	200	.054	.061	.063	.074	1.00	1.00	1.00	1.00	.056	.049	.048	.041	.996	.999	1.00	1.00
	50	.033	.032	.027	.026	.506	.734	.964	.999	.019	.008	.000	.000	.250	.271	.111	.001
P_{IV}	100	.047	.042	.042	.048	.963	.997	1.00	1.00	.047	.032	.013	.000	.821	.915	.972	.964
	200	.046	.051	.054	.056	1.00	1.00	1.00	1.00	.046	.037	.035	.023	.997	1.00	1.00	1.00
	50	.028	.023	.011	.006	.095	.083	.025	.005	.050	.050	.054	.039	.091	.092	.060	.043
$t_{\hat{\xi}^*,\kappa}$	100	.038	.029	.018	.015	.418	.426	.454	.450	.068	.065	.065	.066	.379	.403	.400	.407
	200	.042	.036	.025	.014	.960	.988	1.00	1.00	.065	.075	.078	.077	.913	.951	.977	.991

Nominal 5% level. 5000 replications. $\tau = 0.9$.

References

Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. *Econometrica 72*(4), 1127–1177.

Banerjee, A., M. Marcellino, and C. Osbat (2005). Testing for PPP: Should we use panel methods? *Empirical Economics 30*(1), 77–91.

Bayer, C. and C. Hanck (2009). Combining non-cointegration tests. *METEOR Research Memorandum 9*(12), 1–34.

Breitung, J. and S. Das (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*(3), 259–292.

Cavaliere, G. and A. M. R. Taylor (2007a). Bootstrap Unit Root Tests For Time Series With Nonstationary Volatility. *Econometric Theory 24*(1), 43–71.

Cavaliere, G. and A. M. R. Taylor (2007b). Testing for Unit Roots in Time Series Models with Non-Stationary Volatility. *Journal of Econometrics 140*(2), 919–947.

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

Cavaliere, G. and A. M. R. Taylor (2009). Heteroskedastic Time Series with a Unit Root. *Econometric Theory 25*(5), 1228–1276.

Chang, Y. (2002). Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency. *Journal of Econometrics 110*(2), 261–292.

Chang, Y. and J. Y. Park (2002). On The Asymptotics Of Adf Tests For Unit Roots. *Econometric Reviews 21*(4), 431–447.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Demetrescu, M. (2009). Panel Unit Root Testing with Nonlinear Instruments for Infinite-Order Autoregressive Processes. *Journal of Time Series Econometrics 1*(2), article 3.

Demetrescu, M. (2010). Recursive Adjustment for General Deterministic Components and Improved Tests for the Cointegration Rank. Working paper, Department of Economics, Goethe University Frankfurt.

Demetrescu, M., U. Hassler, and A.-I. Tarcolea (2006, October). Combining significance of correlated statistics with application to panel data. *Oxford Bulletin of Economics and Statistics 68*(5), 647–663.

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

Hanck, C. (2008). The error-in-rejection probability of meta analytic panel tests. *Economics Letters 101*, 27–30.

Hanck, C. (2009). Nonstationary-Volatility Robust Panel Unit Root Tests and the Great Moderation. Working paper, Maastricht University.

Hartung, J. (1999). A note on combining dependent tests of significance. *Biometrical Journal 41*, 849–855.

Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2009). Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. *Econometric Theory 25*, 587–636.

Heston, A., R. Summers, and B. Aten (2009). Penn world table version 6.3. Technical report, Center for International Comparisons of Production, Income and Prices, University of Pennsylvania.

Horowitz, J. L. and N. Savin (2000). Empirically relevant critical values for hypothesis tests: A bootstrap approach. *Journal of Econometrics 95*, 375–389.

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

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

Kuzin, V. (2005). Recursive Demeaning and Deterministic Seasonality. *Statistics & Probability Letters 72*(3), 195–204.

Ledoit, O. and M. Wolf (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88*, 365–411.

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

Leybourne, S., T. H. Kim, and P. Newbold (2005). Examination of Some More Powerful Modifications of the Dickey-Fuller Test. *Journal of Time Series Analysis 26*(3), 355–369.

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

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

Park, J. Y. and P. C. B. Phillips (1999). Asymptotics for Nonlinear Transformations of Integrated Time Series. *Econometric Theory 15*(3), 269–298.

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. (1987). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika 74*(3), 535–547.

Phillips, P. C. B., J. Y. Park, and Y. Chang (2004). Nonlinear Instrumental Variable Estimation of an Autoregression. *Journal of Econometrics 118*(1-2), 219–246.

Rodrigues, P. M. M. (2006). Properties of Recursive Trend-Adjusted Unit Root Tests. *Economics Letters 91*(3), 413–419.

Shin, D. and O. Lee (2001). Tests for Asymmetry in Possibly Nonstationary Time Series Data. *Journal of Business and Economic Statistics 19*(2), 233–244.

Shin, D. W. and S. Kang (2006). An Instrumental Variable Approach for Panel Unit Root Tests under Cross-Sectional Dependence. *Journal of Econometrics 134*(1), 215–234.

Shin, D. W. and B. S. So (2000). Gaussian Tests for Seasonal Unit Roots Based on Cauchy Estimation and Recursive Mean Adjustments. *Journal of Econometrics 99*(1), 107–137.

Shin, D. W. and B. S. So (2001). Recursive Mean Adjustment for Unit Root Tests. *Journal of Time Series Analysis 22*(5), 595–612.

So, B. S. and D. W. Shin (1999). Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals. *Econometric Theory 15*(2), 165–176.

Tanaka, K. (1996). *Time Series Analysis: Nonstationary and noninvertible distribution theory.* Wiley-Interscience.

Taylor, A. M. R. (2002). Regression-Based Unit Root Tests With Recursive Mean Adjustment for Seasonal and Nonseasonal Time Series. *Journal of Business & Economic Statistics 20*(2), 269–281.

Wang, Q. and P. Phillips (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory 25*(3), 710–738.

West, K. D. (1988). Asymptotic Normality, When Regressors Have a Unit Root. *Econometrica 56*(6), 1397–1417.

Westerlund, J. and J. Breitung (2009). Myths and Facts about Panel Unit Root Tests. Working papers in economics no 380, Gothenburg University.