A panel cointegrating rank test with structural breaks and cross-sectional dependence

Arsova, Antonia; Karaman Örsal, Deniz Dilan

Published in:
Econometrics and Statistics

DOI:
10.1016/j.ecosta.2020.05.002

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Econometrics and Statistics 17 (2021) 107–129

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

A panel cointegrating rank test with structural breaks and cross-sectional dependence

Antonia Arsova[a,b,*], Deniz Dilan Karaman Örsal[c,d]

[a] Faculty of Statistics, TU Dortmund University, Vogelpothsweg 78, Dortmund 44227, Germany
[b] RWI - Leibniz Institute for Economic Research, Hohenzollernstraße 1–3, Essen 45128, Germany
[c] Faculty of Business, Economics and Social Sciences, Universität Hamburg, Von-Melle-Park 5, Hamburg 20146, Germany
[d] Institute of Economics, Leuphana University Lüneburg, Universitätsallee 1, Lüneburg 21335, Germany

ARTICLE INFO

Article history:
Received 15 April 2019
Revised 11 May 2020
Accepted 11 May 2020
Available online 20 June 2020

JEL classification:
C12
C15
C33

Keywords:
Panel cointegrating rank test
Structural breaks
Cross-sectional dependence
Likelihood-ratio
Time trend

ABSTRACT

A new panel cointegrating rank test which allows for a linear time trend with breaks and cross-sectional dependence is proposed. The new correlation-augmented inverse normal (CAIN) test is based on a modification of the inverse normal method and combines the *p*-values of individual likelihood-ratio trace statistics by assuming that the number of breaks and break points are known. A Monte Carlo study demonstrates its robustness to cross-sectional dependence and its superior size and power properties compared to other meta-analytic tests used in practice. The test is applied to investigate the long-run relationship between regional house prices and personal income in the United States in view of the structural break introduced by the Global Financial Crisis.

1. Introduction

Panel unit root and cointegration tests have been developed since the early 2000s with the aim to increase the power of single-unit tests. The so-called “first generation” tests rely on the assumption of independence between the panel units. In macroeconomic panel data, however, cross-sectional dependence arises naturally due to common shocks or spillover effects. If not accounted for, it may bias the outcome of the tests by inflating the type-I error rate above the nominal significance level. Another issue often observed in longer time series is that of structural breaks; it may also invalidate the test results if left unattended.

Panel cointegration testing in the presence of structural breaks and cross-sectional dependence has only recently gained attention from researchers, leading to the development of the so called “third-generation tests”. Recently, two residual-based panel cointegration tests have been proposed in the literature, which allow for structural breaks and cross-sectional dependence (Westerlund and Edgerton, 2008; Banerjee and Carrion-i Silvestre, 2015). However, on the one hand, these tests are not appropriate to find out the number of cointegrating relations within a system of variables. On the other hand, the way the cross-sectional dependence is accounted for in these tests, namely by unobserved common factors, leads to

* Corresponding author at: Faculty of Statistics, TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany.
E-mail addresses: arsova@statistik.tu-dortmund.de, antonia.arsova@rwi-essen.de (A. Arsova), deniz.oersal@uni-hamburg.de (D.D. Karaman Örsal).

https://doi.org/10.1016/j.ecosta.2020.05.002

limitations of their application in practice. In Westerlund and Edgerton (2008) the common factors are assumed to be strictly stationary, which might be seen as restrictive. Banerjee and Carrion-i Silvestre (2015) relax this assumption at the cost of diminished interpretability of their test's outcome: cointegration between the observed variables occurs only in the case of stationary factors, otherwise "cointegration [...] up to the inclusion of $I(1)$ factors is possible", which essentially precludes cointegration in the classical sense.

To overcome these shortcomings, this study proposes the first cointegrating rank test for panel data with both structural breaks and cross-sectional dependence, so that the number of cointegrating relations in a system of variables can be established. This likelihood-based test not only allows for heterogeneous, albeit known, locations of the structural breaks in the data generating process (DGP) of each unit, but is also suitable to determine the cointegrating rank of the *observed* data, which is of primary interest in practice. For this new test we extend the cointegrating rank test of Trenkler et al. (2008) (henceforth the TSL test) to panel data with a general form of cross-sectional dependence. The source of cross-sectional dependence may be, but is not limited to, unobserved common factors, as commonly assumed in the literature. The TSL test admits structural breaks in the deterministic part of the DGP, i.e. in the level and/or in the trend slope, but not in the cointegrating vector. This is in line with the interpretation of the latter as a long-run equilibrium relationship between the variables in the system. A further advantage of the TSL test, and consequently of its new panel extension, is that the breaks are allowed for both under the null and under the alternative hypothesis, as they do not affect the stochastic properties of the DGP. Our preference for the TSL test over another likelihood-based alternative, the cointegrating rank test of Johansen et al. (2000) (henceforth JMN test), is motivated by its superior finite-sample properties, demonstrated by Trenkler et al. (2008) and our own simulations.

We extend the TSL test to the panel setting resorting to a new p-value combination method which allows the p-values to be correlated. In comparison to traditional pooling of individual test statistics, p-value combination approaches offer much more flexibility, as they allow the specification of the deterministic terms, the lag order, the number and the location of the breaks, and even the time span of the data to vary over cross-sections. In addition to accommodating more heterogeneous data generating processes and being applicable to unbalanced panels, tests based on p-value combinations can also exhibit better finite sample properties in comparison to tests based on pooling of individual test statistics. For example, Örsal and Arsova (2017) demonstrate superior size and power properties of their meta-analytic panel cointegration tests in comparison to the panel test based on a standardized average of the individual test statistics.

Recent research on panel unit root and cointegration testing has also benefited significantly from the "reinvention" of already existing methods for combining possibly dependent p-values. One example is the modification of the Bonferroni procedure proposed by Simes (1986) and employed by Hanck (2013) in his panel unit root test. In a Monte Carlo study he demonstrates its robustness to cross-sectional dependence induced by common factors. This avenue is further explored in the direction of cointegrating rank testing in dependent panels by Arsova and Örsal (2020), who also show empirically that a sufficient condition for the validity of Simes' procedure (namely that the joint distribution of the test statistics is *multivariate totally positive* (MTP_2), see Sarkar (1998) for a definition) is not violated in a common-factor-driven panel framework. In the present work we adopt a new, augmented version of a p-value combination method initially proposed by Stouffer et al. (1949) — the inverse normal test. The novelty of our approach lies in that we explicitly model the degree of cross-sectional correlation between the probits of the individual statistics and use it as a variance-inflation factor in the panel test statistic. In this regard the proposed test is similar to the modified inverse normal method of Hartung (1999), which also uses an estimate of the cross-sectional correlation of the probits, however in combination with a somewhat arbitrary correction factor κ. Hartung's approach has been used to account for cross-sectional dependence in panel unit root tests by Demetrescu et al. (2006) and Costantini and Lupi (2013). Our estimator for the unobserved correlation of the probits is modelled as a function of a quantity easily measurable in practice, namely the average absolute cross-sectional correlation of the residuals of the individual VAR models. In a Monte Carlo study we demonstrate that our correlation-augmented inverse normal (CAIN) method for combining p-values of individual TSL tests has good size and power properties in finite samples. Its performance is preferable to that of Hartung's method over the whole range of possible values for the cross-sectional correlation. The test can be used to determine the cointegrating rank of the observed time series even when the cross-sectional dependence is driven by unobserved common factors, without decomposing the data into idiosyncratic and common components and testing these separately.

The remainder of the paper is organized as follows. The next section briefly reviews the test of Trenkler et al. (2008), while its extension to panel data by a modification of the normal inverse method is given in Section 3. Section 4 presents the response surface approach to modeling the cross-sectional correlation between the probits of the individual test statistics. Section 5 discusses the results of a Monte Carlo study comparing the CAIN-TSL test with other meta-analytic approaches and with the test of Westerlund (2006). Section 6 illustrates the use of the CAIN-TSL test by investigating the equilibrium relationship between real house prices and real personal income in the USA, and the last section concludes. Supplementary material is provided in the Appendix.

2. The TSL test for the cointegrating rank

Our aim is to develop a panel test for the cointegrating rank which allows for structural breaks in the deterministic parts of the DGP. A natural first step is to take an existing single-unit test and to extend it to the panel setting. Two well-known

alternatives are the likelihood-based test of Johansen et al. (2000) and its GLS detrended counterpart proposed by Trenkler et al. (2008). Both tests allow for up to two breaks in the level and/or trend slope, whereas the break dates are assumed to be known. Theoretically more than two breaks could be specified, however response surface regressions approximating the limiting distributions of the test statistics are available for a maximum of two breaks for both tests. Including more than two breaks may also be seen as problematic in time series of a limited time span.

Trenkler et al. (2008) find that the JMN test with a single trend break displays excessive size distortions in systems of large dimensions while their test is correctly sized. Its preferable finite-sample properties are therefore our main motivation for extending the TSL test to the panel setting. Next we briefly describe its assumptions and the computation of the test statistic.

Let the observed data $Y_{it} = \left(Y_{1,it}, \ldots, Y_{m,it}\right)'$ for cross-sectional unit i $(i = 1, \ldots, N)$ be generated by a stochastic VAR(s_i) process X_{it} added to a deterministic process. The latter consists of a constant, linear time trend and structural breaks in both the level and the trend slope at known individual-specific time(s) τ_i:

$$Y_{it} = \mu_{0i} + \mu_{1i} t + \delta_{0i} d_{it} + \delta_{1i} b_{it} + X_{it}, \quad t = 1, \ldots, T_i. \tag{1}$$

Here μ_{ji} and δ_{ji} $(j = 0, 1)$ are unknown $(m \times 1)$ parameter vectors, while d_{it} and b_{it} are dummy variables defined by $d_{it} = b_{it} = 0$ for $t < \tau_i$, and $d_{it} = 1$ and $b_{it} = t - \tau_i + 1$ for $t \geq \tau_i$. We note that both the break dates and the number of breaks are assumed to be known. The number of breaks (one or two) is allowed to vary across units. The break dates are assumed to occur at individual-specific fixed fractions of the sample size: $\tau_i = [T_i \lambda_i]$ with $0 < \underline{\lambda_i} < \lambda_i < \overline{\lambda_i} < 1$, where $\underline{\lambda_i}$ and $\overline{\lambda_i}$ are specified real numbers and $[\cdot]$ denotes the integer part of the argument. In other words, the breaks are assumed not to occur in the very beginning or in the very end of the sample, while $\underline{\lambda_i}$ and $\overline{\lambda_i}$ are allowed to be arbitrarily close to 0 and 1, respectively. The stochastic processes X_{it} are assumed to be at most $I(1)$ and cointegrated with cointegrating rank r_i, $0 \leq r_i \leq m$:

$$X_{it} = A_{1i} X_{i,t-1} + \ldots + A_{s_i i} X_{i,t-s_i} + \varepsilon_{it}, \quad t = 1, \ldots, T_i. \tag{2}$$

It is assumed that the $(m \times 1)$ vector $\varepsilon_{it} \sim iid(0, \Omega_i)$, where Ω_i is a positive definite matrix for each i. Further it is assumed that ε_{it} have finite moments of order $(4 + \nu)$ for some $\nu > 0, \forall i$.

The computation of the individual TSL test statistics proceeds by estimating the deterministic terms by reduced rank regression taking into account the structural breaks, and then computing a likelihood-ratio (LR) trace statistic from the trend-adjusted observations. For details on the procedure we refer to Trenkler et al. (2008).

3. The CAIN-TSL panel cointegrating rank test

The main constituent of the new panel cointegrating rank test is the inverse normal method for combining p-values, which is briefly discussed next alongside its variants employed in panel unit root testing. Our proposed modification is outlined in Section 3.2, while the novel estimator of the correlation between the probits, on which the modification is based, is presented in Section 4.

3.1. The inverse normal method

Let p_i denote the p-values of the individual TSL statistics for units $i = 1, \ldots, N$. Let t_i denote the corresponding probits, i.e. $t_i = \Phi^{-1}(p_i)$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Assuming independence of the individual test statistics (and hence of their p-values and the corresponding probits t_i), the inverse normal test has a standard $N(0, 1)$ limiting distribution:

$$t = \frac{\sum_{i=1}^{N} \Phi^{-1}(p_i)}{\sqrt{N}} = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N}} \Rightarrow N(0, 1). \tag{3}$$

The inverse normal test was first introduced to the panel unit root testing literature by Choi (2001). He points out two advantageous properties of the method, which we also wish to exploit for our panel test: first, superior performance compared to other p-value combination alternatives under independence, and second, applicability to panels with both finite and infinite cross-sectional dimension N. Panel unit root tests for dependent panels based on the inverse normal method have been proposed by Demetrescu et al. (2006) and Costantini and Lupi (2013). Both studies implement Hartung's (1999) modification of the weighted inverse normal method, which accommodates constant dependence between the original test statistics. Assuming multivariate normal distribution of the probits, Hartung (1999) captures the dependence by a single correlation coefficient ρ_{probit}, which can be interpreted as a "mean correlation approximating the case of possibly different correlations between the transformed statistics". The variance of the denominator in (3) is then augmented with an estimator of the correlation between the individual probits $\hat{\rho}^*_{\text{probit}}$. For simplicity, we present the test statistic assuming unit weights for all probits:

$$t\left(\hat{\rho}^*_{\text{probit}}, \kappa\right) = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N + \left(N^2 - N\right)\left(\hat{\rho}^*_{\text{probit}} + \kappa \cdot \sqrt{\frac{2}{(N+1)}}(1 - \hat{\rho}^*_{\text{probit}})\right)}}. \tag{4}$$

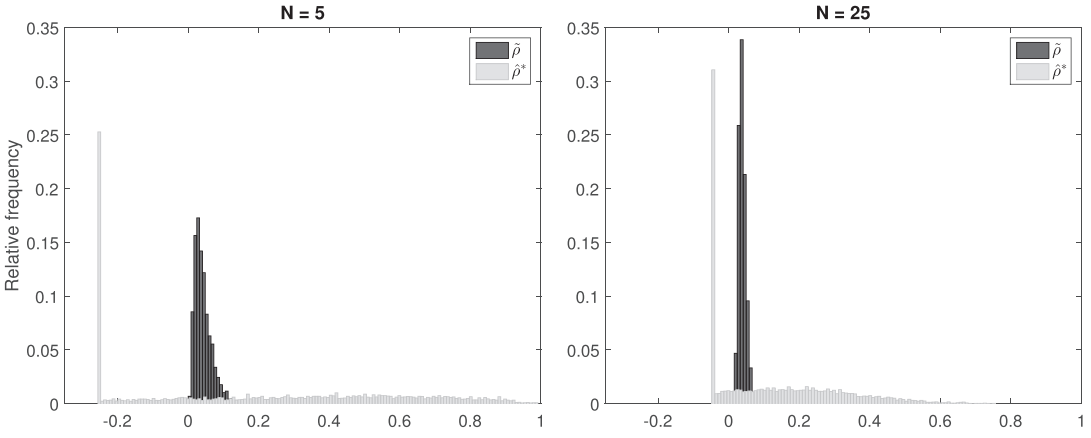

Fig. 1. Empirical distributions of Hartung's estimator $\hat{\rho}^*_{\text{probit}}$ and the CAIN estimator $\tilde{\rho}_{\text{probit}}$ for the correlation between the probits under $H_0: r_i = 0, \forall i$, 3-variate VAR(2) process with a multifactor-error structure and a diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$, $T = 100$, 5000 replications.

Hartung (1999) proposes an unbiased and consistent estimator $\hat{\rho}^*_{\text{probit}} = \max\{-\frac{1}{N-1}, \hat{\rho}_{\text{probit}}\}$, where $\hat{\rho}_{\text{probit}} = 1 - \frac{1}{N-1}\sum_{i=1}^{N}(t_i - \frac{1}{N}\sum_{i=1}^{N} t_i)^2$. The correction term $\kappa\sqrt{\frac{2}{(N+1)}}(1 - \hat{\rho}^*_{\text{probit}})$, which simply scales the standard deviation of $\hat{\rho}_{\text{probit}}$ by a factor κ, aims to avoid a systematic underestimation of the denominator in (4). For the parameter κ Hartung (1999) suggests two alternative values: $\kappa_1 = 0.2$ and $\kappa_2 = 0.1 \cdot (1 + \frac{1}{N-1} - \hat{\rho}^*_{\text{probit}})$, where κ_2 is suitable mainly for smaller values of $\hat{\rho}^*_{\text{probit}}$. However, he provides no guidance as to where the threshold between "small" and "large" $\hat{\rho}^*_{\text{probit}}$ should be, despite the value of κ having a significant effect on the test's finite-sample properties. Beside this ambiguity, our simulation results indicate two potentially serious issues with Hartung's (1999) estimator of the correlation between the probits. First, as it is computed from the single available observation on the N-dimensional vector of probits $(t_1, \ldots, t_N)$, its consistency rate is rather slow. Relaxing the assumption of constant correlation between the probits, Demetrescu et al. (2006) show that the consistency rate of $\hat{\rho}^*_{\text{probit}}$ is $o_{p(1)}$. Second, as our own simulations demonstrate, in a non-negligible proportion of all cases $\hat{\rho}^*_{\text{probit}}$ estimates the correlation between the probits at its natural lower bound, $-\frac{1}{N-1}$, i.e. it may be underestimated. This could lead to incorrect value of the variance of the test statistic and compromise the test's outcome. An illustration of both these issues is provided in Fig. 1.

3.2. The correlation-augmented inverse normal test

To overcome these shortcomings, we propose a new, improved version of the modified inverse normal test for combination of correlated individual TSL test statistics for cointegration with structural breaks. This new test, which we name CAIN-TSL test, is based on Hartung's (1999) test statistic (4) with unit weights. It differs from it in that it employs a novel empirical estimate $\tilde{\rho}_{\text{probit}}$ of the average correlation between the probits, which will be defined in Section 4.

The CAIN panel test statistic for the composite null hypothesis $H_0: r_i = r, \forall i = 1, \ldots, N$ against the alternative $H_1: r_i > r$ for at least one i is given by

$$t(\tilde{\rho}) = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N + (N^2 - N) \cdot \tilde{\rho}_{\text{probit}}}}. \tag{5}$$

We argue that the correlation between the probits can be inferred from its source, that is, the dependence between the shocks, or innovations, to the individual VAR processes (2). Unobserved common shocks or spillover effects will lead to dependence between the elements of ε_{it} and ε_{jt}, $i, j = 1, \ldots, N$, which will, in turn, induce dependence between the observed data and between the individual TSL test statistics for the cointegrating rank. As a measure of dependence we employ Pearson's correlation coefficient, as capturing linear dependence suffices for our purposes. Let $Y_{l,it}$ and $Y_{l,jt}$, for $l = 1, \ldots, m$ and $i, j = 1, \ldots, N$ denote the same variables for different cross-sectional units i and j, and $\rho_{li,kj} := \text{corr}(\varepsilon_{l,it}, \varepsilon_{k,jt})$ for $l, k = 1, \ldots, m$ and $i, j = 1, \ldots, N$ denote the pairwise cross-sectional correlations of the elements of ε_{it} and ε_{jt}. In order to be able to quantify the relationship between the correlation of the innovations and the correlation of the probits, we make the following assumptions.

Assumption 1. The average absolute pairwise cross-sectional correlation between the innovations to the same variable converges to some fixed value $\rho_\varepsilon > 0$ for all t as $N \to \infty$:

$$\lim_{N\to\infty} \frac{1}{mN(N-1)} \sum_{i \neq j}^{N} \sum_{l=1}^{m} |\rho_{li,lj}| = \rho_\varepsilon. \tag{6}$$

Assumption 2. The average absolute pairwise cross-sectional correlation between the innovations to different variables converges to zero for all t as $N \to \infty$:

$$\lim_{N\to\infty} \frac{1}{m(m-1)N(N-1)} \sum_{i\neq j}^{N} \sum_{l\neq k}^{m} |\rho_{li,kj}| = 0. \tag{7}$$

These assumptions are motivated by the fact that in practice strong correlations are less likely to occur between the innovations to different variables across units compared to those between the same variables. Differentiating between "strong" and "weak" correlations in this way and averaging only the former aims to avoid underestimation of the degree of dependence between the innovations. Regarding taking absolute values, we argue that even negative correlations between the innovations to the DGPs lead to positive correlation between the individual test statistics. With simple averaging large correlations of the opposite sign will cancel out; a better way to quantify the degree of cross-sectional dependence is therefore to look at the mean absolute correlation. Assumptions 1 and 2 thus facilitate the consistent estimation of an average absolute correlation coefficient $\hat{\rho}_{\varepsilon}$, which will subsequently be employed in the computation of an estimate $\tilde{\rho}_{\text{probit}}$ of the correlation between the probits. We note that the proposed panel test is robust to a certain degree of deviation from Assumption 2, as demonstrated by our Monte Carlo simulations. Therefore, albeit seemingly restrictive, it should not hinder the applicability of the test in practice. Furthermore, both assumptions hold when a spatial type of dependence is assumed, or when the dependence is driven by variable-specific common factors. As such we describe unobserved shocks, which affect the same variable over the cross-sections, but do not or only marginally affect other variables.

We next turn our attention to the correlation between the probits. In order to accommodate a certain degree of heterogeneity in the correlation coefficients $\rho_{\text{probit}_{i,j}} := \text{corr}(t_i, t_j)$, we make the same assumptions as Demetrescu et al. (2006), who prove the standard normal limiting distribution of Hartung's (1999) test statistic in this case.

Assumption 3.

$$\lim_{N\to\infty} \frac{1}{N(N-1)} \sum_{i\neq j}^{N} \rho_{\text{probit}_{i,j}} = \rho_{\text{probit}}, \quad \rho_{\text{probit}} \in (0,1), \quad \text{and} \tag{8}$$

$$\lim_{N\to\infty} \frac{1}{N(N-1)} \sum_{i\neq j} (\rho_{\text{probit}_{i,j}} - \rho_{\text{probit}})^2 = 0. \tag{9}$$

Assumption 4. The individual TSL test statistics have a Gaussian copula.

Assumption 3 allows the correlation matrix of the probits to gradually approach a constant correlation matrix as $N \to \infty$, which may be viewed as a consequence of the assumed convergence of the pairwise innovation correlations in Assumption 1. Assumption 4, as shown by Demetrescu et al. (2006), is a necessary and sufficient condition for the limiting distribution of the inverse normal test to be standard normal. Whether this assumption is met in practice for individual unit root or cointegration statistics, is not known; we leave such investigation for future research. However, as our Monte Carlo study demonstrates, despite the possibility that Assumption 4 might not hold, in the presence of cross-sectional dependence the proposed CAIN test performs better than the standard inverse normal method without correlation augmentation, especially when the null hypothesis of no cointegration is tested for. Hence in practice one would still be better off by correcting for existing cross-sectional dependence rather than ignoring it.

We postulate that ρ_{probit} can be inferred from its origin ρ_{ε}, which, in turn, can be consistently estimated in practice as N and T grow. It is, however, difficult to derive analytically how correlation between the innovations translates into correlation between the LR statistics, as the latter are complex non-linear functions of the observed data. We therefore resort to simulation methods to estimate this link. In a large-scale simulation study we estimate the average correlation of the probits, which we denote $\tilde{\bar{\rho}}_{\text{probit}}$, for different values of the absolute cross-sectional correlation between the innovations ρ_{ε}, controlling for the system dimension m, the hypothesized cointegrating rank r, and the time and cross-sectional dimensions T and N, respectively. We then estimate the relationship $\tilde{\bar{\rho}}_{\text{probit}} = g(\rho_{\varepsilon}, m, r)$ by response surface regression. As in practice ρ_{ε} is unobserved, we replace it by a consistent estimate $\hat{\rho}_{\varepsilon}$ to compute $\tilde{\rho}_{\text{probit}} = g(\hat{\rho}_{\varepsilon}, m, r)$ to use in the test statistic (5). This simulation study is described in detail in Section 4.

Conjecture 1. *Provided that Assumptions 1–4 hold, the limiting distribution of the test statistic $t(\tilde{\rho})$ under the null hypothesis is standard normal.*

As noted by Hartung (1999) and by Demetrescu et al. (2006), the statistic converges to the $N(0, 1)$ distribution when the estimator of the correlation between the probits is consistent. The proposed estimator $\tilde{\rho}_{\text{probit}}$ is based on an empirical estimate of the link between the cross-sectional correlation of the innovations to the individual VAR processes and the correlation between the probits of the individual TSL statistics. Since there is no analytic expression for it, no proof of the conjecture can be provided in the strict mathematical sense. Nevertheless, in the Appendix we outline the arguments by which $\tilde{\rho}_{\text{probit}}$ is expected to be consistent.

4. Response surface regressions for the correlation of the individual probits

Simulations design. We estimate the relationship between ρ_ε and ρ_{probit} in a large-scale simulation study. The data are generated according to (1) with the stochastic processes X_{it} following an m-variate VAR(1) Toda processes (see e.g. Toda, 1995) for $m = 2, 3, 4, 5$:

$$X_{it} = \begin{pmatrix} I_d & 0_{(d\times r)} \\ 0_{(r\times d)} & \Psi_r \end{pmatrix} X_{i,t-1} + \varepsilon_{it}, \tag{10}$$

where I_d denotes the identity matrix of dimension $d = m - r$, $0_{(d\times r)}$ is a zero matrix of the corresponding dimension and r is the cointegrating rank of the process. Ψ_r is a diagonal matrix of dimension $(r \times r)$ with diagonal elements $\psi_i \sim i.i.d.U(0, 1)$, $i = 1, \ldots, r$ which govern the dynamics of the stationary elements of the process for $r > 0$; the uniform (0,1) distribution of ψ_i provides for generality. As the individual LR trace statistics of the TSL test are invariant to the values of the deterministic terms μ_{0i}, μ_{1i}, δ_{0i} and δ_{1i} for given VAR order s_i and break date(s) τ_i (Trenkler et al., 2008, p. 340), we have set $\mu_{ji} = \delta_{ji} = 0,\ j = 0, 1$.

For the estimation of the individual LR trace statistics for each $i = 1, \ldots, N$ the number of breaks is randomly chosen between 1 and 2 with 50% chance each, and the break locations are chosen at relative sample length(s) $\lambda_i \sim i.i.d.U(0.15, 0.85)$. In the case of two breaks the minimal distance between them is set to $0.2T$. These values are in line with what is usually assumed in the literature on cointegration with structural breaks – the breaks are not allowed to be too close to the beginning or to the end of the sample, neither are they allowed to be too close to each other.

The $(T \times Nm)$ vector of innovations $\varepsilon_{NT} = (\varepsilon_1, \ldots, \varepsilon_T)'$, where $\varepsilon_t = (\varepsilon_{1,1t}, \ldots, \varepsilon_{m,1t}, \ldots, \varepsilon_{1,Nt}, \ldots, \varepsilon_{m,Nt})$, are drawn from a multivariate normal distribution with variance-covariance matrix Σ, where $\Sigma = R \otimes \Omega$ is generated as in Wagner and Hlouskova (2010) with

$$R = \begin{pmatrix} 1 & \rho_\varepsilon & \cdots & \rho_\varepsilon \\ \rho_\varepsilon & 1 & \cdots & \ddots \\ \vdots & \ddots & \ddots & \rho_\varepsilon \\ \rho_\varepsilon & \cdots & \rho_\varepsilon & 1 \end{pmatrix}_{(N\times N)}, \tag{11}$$

and $\Omega_{(m\times m)}$ being a random correlation matrix generated independently for each replication as described in Costantini and Lupi (2013, p. 283).

For each value of m, testing $H_0 : r_i = r_0$ for $r_0 = 0, \ldots, m-1$ is considered in a separate experiment. We let ρ_ε vary over a grid of 24 equally spaced values in the range [0.04, 0.96], in order to be able to adequately fit an interpolating curve to the estimated average values $\bar{\tilde{\rho}}_{\text{probit}}$ of ρ_{probit} over $\hat{\rho}_\varepsilon$. Each simulation experiment is repeated 100,000 times for three combinations of the panel dimensions T and N: $(T, N) \in \{(500, 5), (500, 10), (1000, 5)\}$.

Estimation of the average correlation of the probits $\bar{\tilde{\rho}}_{\text{probit}}$ *from simulated data*. For each combination (T, N) an average $\bar{\tilde{\rho}}_{\text{probit}}$ is computed by means of Fisher's Z-transformation from the $(N \times N)$ correlation matrix of the probits based on the $(100,\ 000 \times N)$ matrix of independent observations on $(t_1, \ldots, t_N)$.

For given ρ_ε and T, the estimated $\bar{\tilde{\rho}}_{\text{probit}}$ is practically invariant to the number of cross-sections. This fact is not surprising, as it is rational to expect that the correlation between the individual probits would depend on the degree of dependence between the processes of any two cross-sections, but not on the number of units. Preliminary simulations corroborate this conjecture – see the right panel of Fig. 2 in the Appendix. For robustness, the estimated $\bar{\tilde{\rho}}_{\text{probit}}$'s for all three combinations of (T, N) are subsequently modeled in the response surface regressions for $\bar{\tilde{\rho}}_{\text{probit}} = g(\rho_\varepsilon, m, r)$. In practice, by the Law of Large Numbers, increasing N would lead to more precise estimation of $\hat{\rho}_\varepsilon$ and subsequently also of $\tilde{\rho}_{\text{probit}}$. Because of the large number of replications in the current simulation exercise, considering a wider range of values for N is not necessary and we concentrate only on values which are typical for macro-econometric studies.

With regard to the choice of values for the parameter T, we motivate it by the fact that the estimated $\bar{\tilde{\rho}}_{\text{probit}}$ converges to ρ_{probit} from below as $T \to \infty$ (see the left panel of Fig. 2), with the differences between $T = 500$ and $T = 1000$ being virtually negligible for all practical purposes. Indeed, for small T the estimated $\tilde{\rho}_{\text{probit}}$ might be lower than its asymptotic large-T value and could thus potentially get overestimated. We argue, however, that the use of the large-T $\tilde{\rho}_{\text{probit}}$ even for small T's in the panel setting would have beneficial rather than detrimental effects. It is well known that individual likelihood-based cointegration tests tend to be oversized for $H_0 : r = 0$ when T is small, and these size distortions can get magnified in the panel setting as the cross-sectional dimension increases (see, e.g., Demetrescu and Hanck, 2012 and Arsova and Örsal, 2018). Using the asymptotic $\tilde{\rho}_{\text{probit}}$ might help mitigate this issue, as slight overestimation of the cross-sectional correlation of the probits might inflate the variance of the panel statistic thus offsetting the inherent size distortions of the individual tests.

Response surface regressions. We now turn our attention to modelling the relationship $\bar{\tilde{\rho}}_{\text{probit}} = g(\rho_\varepsilon, m, r)$. To obtain the response surface g, we regress $\bar{\tilde{\rho}}_{\text{probit}}$ on polynomials of the system dimension m, the cointegrating rank under the null hypothesis r, and the mean absolute correlation between the innovations ρ_ε. No constant is included in the regression and all regressors are multiples of ρ_ε, so that all estimates $\bar{\tilde{\rho}}_{\text{probit}}$ of ρ_{probit} are equal to 0 when $\rho_\varepsilon = 0$.

Table 1
Coefficients of the response surface $\tilde{\rho}_{\text{probit}} = g(\rho_\varepsilon, m, r)$ for the correlation between the probits of the individual TSL statistics.

Term	Estimated coefficient	Std. Err.	t-value	$P>\|t\|$	95% Confidence interval
ρ_ε^2	0.6319575	0.0357830	17.66	0	(0.561742, 0.702173)
		0.0309826	20.4	0	(0.571162, 0.692753)
$\sqrt{m}\cdot\rho_\varepsilon^2$	−0.5193669	0.0251431	−20.66	0	(−0.568704, −0.470030)
		0.0211871	−24.51	0	(−0.560941, −0.477792)
$\sqrt{m}\cdot\rho_\varepsilon^4$	0.2721753	0.0032355	84.12	0	(0.265826, 0.278524)
		0.0033330	81.66	0	(0.265635, 0.278716)
$\frac{r}{m}\cdot\rho_\varepsilon^2$	0.1821374	0.0307753	5.92	0	(0.121749, 0.242526)
		0.0260690	6.99	0	(0.130983, 0.233292)
$\frac{r}{m}\cdot\rho_\varepsilon^4$	−0.0856903	0.0322216	−2.66	0.008	(−0.148917, −0.022463)
		0.0293583	−2.92	0.004	(−0.143299, −0.028082)
$(r\cdot\rho_\varepsilon)^2$	0.0041125	0.0008434	4.88	0	(0.002458, 0.005768)
		0.0007581	5.43	0	(0.002625, 0.005600)
$r\cdot\rho_\varepsilon^2$	0.0766267	0.0076098	10.07	0	(0.061694, 0.091559)
		0.0070263	10.91	0	(0.062839, 0.090414)
$r\cdot\rho_\varepsilon^4$	−0.1008678	0.0057871	−17.43	0	(−0.112224, −0.089512)
		0.0060007	−16.81	0	(−0.112643, −0.089093)
$\sqrt{m-r}\cdot\rho_\varepsilon^2$	0.1874919	0.0348379	5.38	0	(0.119131, 0.255853)
		0.0276625	6.78	0	(0.133211, 0.241773)
$\frac{1}{m-r}\cdot\rho_\varepsilon^2$	0.1410229	0.0168932	8.35	0	(0.107874, 0.174172)
		0.0150575	9.37	0	(0.111476, 0.170570)
$\frac{1}{m-r}\cdot\rho_\varepsilon^4$	−0.2029126	0.0126466	−16.04	0	(−0.227728, −0.178097)
		0.0120002	−16.91	0	(−0.226460, −0.179365)
$(m-r)^2\cdot\rho_\varepsilon^2$	0.0052557	0.0009492	5.54	0	(0.003393, 0.007118)
		0.0008073	6.51	0	(0.003672, 0.006840)
$(m-r)^4\cdot\rho_\varepsilon^4$	−0.0000327	0.0000167	−1.96	0.050	(−0.000065, 0.000000)
		0.0000179	−1.83	0.068	(−0.000068, 0.000002)

The second row for each term presents robust standard errors and the statistics computed therewith.

The goodness of fit measure of the estimated regression is $R^2 = 0.9993$, which renders the approximation very good for practical purposes. All estimated coefficients, which are significant at the 5%-level, are listed in Table 1. They can be used to compute the estimate $\tilde{\rho}_{\text{probit}}$ of the unknown correlation of the probits given m, r and the estimated $\hat{\rho}_\varepsilon$ for any system dimension $m \leq 5$. The CAIN-TSL test statistic is then computed as in (5) with $\tilde{\rho}_{\text{probit}}$ in the place of the unknown ρ_{probit}.

Estimation of the average absolute cross-sectional correlation between the process innovations. The cornerstone of our proposed solution $\tilde{\rho}_{\text{probit}} = g(\rho_\varepsilon, m, r)$ is a consistent estimator $\hat{\rho}_\varepsilon$ of the average absolute cross-sectional correlation of the process innovations. In this regard we follow Pesaran (2015) and estimate $\hat{\rho}_\varepsilon$ from the residuals of the individual VAR(s_i) models under the null hypothesis $H_0 : r = 0$.

That is, for $\hat{\rho}_{li,lj}$ denoting the estimated sample correlation between the residuals for variable l in units i and j, $\hat{\rho}_\varepsilon$ is to be estimated as:

$$\hat{\rho}_\varepsilon = \frac{2}{mN(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{l=1}^{m} |\hat{\rho}_{li,lj}|. \tag{12}$$

Estimating $\hat{\rho}_\varepsilon$ only under $H_0 : r_i = 0, \forall i$ in the beginning of the sequential testing procedure for $r = 0, \ldots, m-1$ requires further justification. Ideally, the correlation between the probits would be inferred from the correlation between the residuals under each null hypothesis and then used in the correlation-augmented inverse normal panel test. In order to keep the procedure feasible for practical work, however, we decide to do this inference only once by determining $\hat{\rho}_\varepsilon$ only under the null of no cointegration, similarly to the lag orders s_i of the VAR processes. We note that the estimation error in $\hat{\rho}_\varepsilon$ is negligible under the true $H_0 : r_i = 0, \forall i, m$, while $\hat{\rho}_\varepsilon$ gets only slightly underestimated for higher true cointegrating ranks (see Table 6 in the Appendix). The resulting slight underestimation of ρ_{probit}, however, would be negligible in practice for moderate ρ_ε; for higher ρ_ε it could even be beneficial, given that the TSL test is severely undersized when testing for higher ranks under the null.

Step-by-step outline of the testing procedure. In order to illustrate the simplicity of the proposed testing approach, we summarize it with the following five steps.

1. Compute the TSL statistic and its corresponding p-value as outlined in Trenkler et al. (2008) under $H_0 : r = 0$ for each unit. As a by-product, save the residuals from the estimated individual VAR models.
2. Estimate the $(Nm \times Nm)$ sample correlation matrix of the residuals. If the time span of the series varies over the cross-sections, estimation of the correlation matrix would be based on the balanced panel restricted by the shortest time series. Compute $\hat{\rho}_\varepsilon$ as in (12).
3. Estimate $\tilde{\rho}_{\text{probit}} = g(\hat{\rho}_\varepsilon, m, r)$ using the response surface coefficients in Table 1.
4. Compute the panel test statistic by combining the probits of the individual p-values as in (5).

5. The CAIN-TSL test can be applied for testing $H_0 : r_i = r, \forall i$ at each step $r = 0, \ldots, m-1$ of the sequential rank testing procedure. If the composite null hypothesis $H_0 : r_i = 0, \forall i$ is rejected, repeat steps 1 and 4, without re-estimating $\hat{\rho}_\varepsilon$. As $\tilde{\rho}_{\text{probit}}$ depends on the cointegrating rank r, it needs to be re-estimated. The null hypothesis is rejected at significance level α if $t(\tilde{\rho}) < -z_{1-\alpha}$, where $z_{1-\alpha}$ is the $(1-\alpha)$-th quantile of the standard normal distribution.

5. Monte Carlo study

5.1. Simulation study design

The finite sample properties of the CAIN-TSL test are first examined by simulations in an empirically relevant case using a three-variate VAR(2) DGP as in the study of Wagner and Hlouskova (2010). The test is next compared to the panel cointegration test of Westerlund (2006), which as well allows for structural breaks in the deterministic terms.

The general form of the DGP is:

$$Y_{it} = \mu_{0i} + \mu_{1i} t + \delta_{0i} d_{it} + \delta_{1i} b_{it} + X_{it}, \tag{13}$$

$$X_{it} = \begin{pmatrix} a^i_{11} & 0 & 0 \\ 0 & a^i_{12} & 0 \\ 0 & 0 & a^i_{13} \end{pmatrix} X_{i,t-1} + \begin{pmatrix} a^i_{21} & 0 & 0 \\ 0 & a^i_{22} & 0 \\ 0 & 0 & a^i_{23} \end{pmatrix} X_{i,t-2} + u_{it}, \tag{14}$$

$$u_{it} = \gamma_i' f_t + \varepsilon_{it}, \tag{15}$$

$$\varepsilon_{it} \sim i.i.d.\, N(0, \Omega_i). \tag{16}$$

The cointegrating properties of the process are determined by the roots q^i_{1j}, q^i_{2j} of the autoregressive polynomial, which are linked to the coefficients of the autoregressive matrices by $a^i_{1j} = \frac{1}{q^i_{1j}} + \frac{1}{q^i_{2j}}$ and $a^i_{2j} = -\frac{1}{q^i_{1j} q^i_{2j}}$, $j = 1, 2, 3$.

Following Wagner and Hlouskova (2010), for a system with cointegrating rank zero we set $q^i_{1j} = 1$ and $q^i_{2j} \sim U(1.8, 3)$, $j = 1, 2, 3$. Power is investigated in a setting when all roots are sufficiently away from unity (case A) and also in a near-unit root setting (case B). For cointegrating rank one in case A we let $q^i_{11} \sim U(1.3, 1.7)$ or in case B, $q^i_{11} \sim U(1, 1.3)$, while $q^i_{21} \sim U(1.5, 2.5)$, $q^i_{1j} = 1$ and $q^i_{2j} \sim U(1.8, 3)$ for $j = 2, 3$ in both cases. Finally, for cointegrating rank two we again let $q^i_{11} \sim U(1.3, 1.7)$ for case A or $q^i_{11} \sim U(1, 1.3)$ for case B. The remaining roots for both cases are $q^i_{12}, q^i_{2j} \sim U(1.5, 2.5)$ for $j = 1, 2$, while $q^i_{13} = 1$ and $q^i_{23} \sim U(1.8, 3)$. All roots are drawn separately for each unit.

We let a 3-dimensional vector of variable-specific common factors $f_t \sim i.i.d.\, N(0, I_3)$ drive the cross-sectional dependence through heterogeneous loadings. The factor loadings γ_i are simulated as diagonal $(k \times m)$ matrices with (a) $i.i.d.\, U(-0.4, 0.4)$, (b) $i.i.d.\, U(0, 1)$ or (b) $i.i.d.\, U(-1, 3)$ entries, drawn separately for each unit. The robustness of the CAIN-TSL test against violations of Assumption 2 is investigated by letting γ_i be an unrestricted matrix with $i.i.d.\, U(0, 1)$ or $i.i.d.\, U(-1, 3)$ elements. This corresponds to all factors affecting all variables simultaneously.

Throughout the simulation study we use $\mu_{ji} = \delta_{ji} = 0$, $j = 0, 1$, and simulate the number of breaks, break locations, Ω_i and ε_{it} in the same way as explained in Section 4. The processes X_{it} are initialised with 0 and the first 50 observations are discarded to mitigate the effect of initial values. We consider all combinations of $T \in \{100, 200\}$ and $N \in \{5, 15, 25\}$. The lag order is assumed to be known and is set to its true value. The simulations are carried out in GAUSS and the number of replications is 5000. Nominal significance level $\alpha = 0.05$ applies in all cases.

The performance of the CAIN-TSL test is compared to that of other p-value combination methods commonly applied in the literature. Using the p-values from the individual TSL test statistics as building blocks, we combine them into different panel statistics by employing (a) the standard inverse normal test without correction for the cross-sectional dependence; (b) both variants of Hartung's modified inverse normal test with κ_1 and κ_2, respectively, and (c) the multiple testing procedure of Simes (1986). For the latter test the individual p-values of the test statistics are ordered in ascending way as $p_{(1)} \leq \cdots \leq p_{(N)}$, and the joint null hypothesis $H_0 : r_i = r, \forall i$, is rejected at significance level α if $p_{(i)} \leq \frac{i\alpha}{N}$ for any $i = 1, \ldots, N$.

5.2. Simulation results

The size and power results under $H_0 : r_i = 0, \forall i$ when the true rank is zero and one, respectively, are presented in Table 2. In the following discussion average values of the parameters, computed over all replications, are denoted by a long bar.

With diagonal factor loading matrix $\gamma_i \sim i.i.d. U(-0.4, 0.4)$, the cross-sectional dependence is very weak: $\overline{\hat{\rho}_\varepsilon} = 0.089$ and $\overline{\hat{\rho}_\varepsilon} = 0.068$ for $T = 100$ and $T = 200$, respectively, while the mean estimated correlation between the probits for the CAIN-TSL test is $\overline{\tilde{\rho}_{\text{probit}}} = 0.001$. In this case all variants of the inverse normal test, including the CAIN-TSL test, become undersized as N grows. This results from the individual TSL test being slightly undersized; the size distortions get magnified in the panel setting as N increases. Hartung's test with κ_1 has the most severe size distortions, inline with the findings of Demetrescu

Table 2
Monte Carlo study results, not near-unit root processes (case A) and near-unit root processes (case B). Empirical size and power under $H_0 : r_i = 0,\ \forall i = 1, \ldots, N$.

Test	True cointegrating rank 0: Size						True cointegrating rank 1, case A: Power						True cointegrating rank 1, case B: Power					
	T=100			T=200			T=100			T=200			T=100			T=200		
	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-0.4, 0.4)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-0.4, 0.4)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-0.4, 0.4)$					
Inverse normal	0.05	0.04	0.03	0.04	0.04	0.03	0.56	0.92	0.99	0.99	1.00	1.00	0.16	0.27	0.36	0.58	0.92	0.99
Hartung κ_1	0.04	0.02	0.01	0.04	0.01	0.01	0.43	0.68	0.79	0.96	1.00	1.00	0.13	0.15	0.17	0.51	0.80	0.93
Hartung κ_2	0.05	0.04	0.03	0.05	0.03	0.02	0.46	0.73	0.82	0.97	1.00	1.00	0.16	0.23	0.28	0.54	0.85	0.95
Simes	0.06	0.06	0.07	0.05	0.05	0.06	0.36	0.50	0.57	0.94	1.00	1.00	0.14	0.17	0.19	0.48	0.68	0.78
CAIN-TSL	0.05	0.04	0.03	0.04	0.04	0.03	0.56	0.92	0.99	0.99	1.00	1.00	0.16	0.27	0.35	0.58	0.92	0.99
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$					
Inverse normal	0.05	0.05	0.05	0.04	0.05	0.05	0.49	0.84	0.95	0.98	1.00	1.00	0.13	0.23	0.30	0.49	0.84	0.95
Hartung κ_1	0.04	0.02	0.01	0.04	0.02	0.01	0.35	0.56	0.66	0.93	1.00	1.00	0.10	0.12	0.12	0.42	0.65	0.79
Hartung κ_2	0.05	0.04	0.03	0.05	0.04	0.04	0.38	0.62	0.71	0.93	1.00	1.00	0.13	0.18	0.20	0.44	0.71	0.84
Simes	0.05	0.06	0.06	0.05	0.05	0.06	0.29	0.39	0.45	0.89	0.99	1.00	0.11	0.13	0.14	0.37	0.51	0.60
CAIN-TSL	0.04	0.04	0.04	0.04	0.04	0.04	0.48	0.83	0.93	0.98	1.00	1.00	0.13	0.21	0.26	0.49	0.82	0.94
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$					
Inverse normal	0.06	0.08	0.11	0.06	0.09	0.12	0.42	0.71	0.82	0.95	1.00	1.00	0.13	0.22	0.29	0.42	0.69	0.82
Hartung κ_1	0.04	0.03	0.04	0.05	0.04	0.04	0.28	0.42	0.47	0.85	0.97	0.99	0.10	0.11	0.12	0.32	0.47	0.57
Hartung κ_2	0.06	0.06	0.06	0.06	0.06	0.07	0.31	0.47	0.53	0.86	0.97	0.99	0.12	0.16	0.17	0.35	0.54	0.64
Simes	0.06	0.06	0.06	0.05	0.05	0.06	0.24	0.31	0.33	0.79	0.93	0.96	0.10	0.11	0.11	0.28	0.37	0.42
CAIN-TSL	0.04	0.05	0.05	0.05	0.05	0.06	0.38	0.61	0.70	0.94	1.00	1.00	0.12	0.15	0.17	0.38	0.59	0.70
	Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$						Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$						Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$					
Inverse normal	0.05	0.05	0.06	0.05	0.06	0.07	0.66	0.95	0.99	1.00	1.00	1.00	0.20	0.37	0.49	0.70	0.96	0.99
Hartung κ_1	0.04	0.02	0.02	0.04	0.02	0.02	0.52	0.77	0.87	0.98	1.00	1.00	0.17	0.23	0.27	0.62	0.90	0.97
Hartung κ_2	0.05	0.04	0.04	0.05	0.05	0.04	0.55	0.80	0.89	0.99	1.00	1.00	0.20	0.31	0.37	0.65	0.93	0.98
Simes	0.06	0.06	0.06	0.06	0.05	0.06	0.45	0.60	0.68	0.97	1.00	1.00	0.17	0.22	0.26	0.58	0.81	0.89
CAIN-TSL	0.04	0.03	0.03	0.04	0.04	0.04	0.63	0.92	0.98	1.00	1.00	1.00	0.18	0.29	0.35	0.67	0.93	0.99
	Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$						Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$						Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$					
Inverse normal	0.07	0.12	0.15	0.08	0.14	0.18	0.68	0.94	0.98	0.99	1.00	1.00	0.26	0.44	0.54	0.73	0.95	0.98
Hartung κ_1	0.04	0.05	0.05	0.05	0.05	0.06	0.54	0.79	0.87	0.98	1.00	1.00	0.21	0.29	0.33	0.68	0.91	0.96
Hartung κ_2	0.06	0.07	0.08	0.06	0.07	0.08	0.57	0.82	0.89	0.98	1.00	1.00	0.23	0.36	0.41	0.70	0.93	0.98
Simes	0.05	0.05	0.06	0.05	0.05	0.05	0.49	0.67	0.73	0.97	1.00	1.00	0.20	0.28	0.32	0.65	0.87	0.93
CAIN-TSL	0.06	0.07	0.08	0.07	0.09	0.10	0.64	0.89	0.94	0.99	1.00	1.00	0.22	0.32	0.37	0.69	0.91	0.96

Notes: Rejection frequencies at the 5% the significance level, 5000 replications. Power is not size-adjusted, as size-adjustment would not be available in practice. Some results with size-adjusted power are available in Table 9 in the Appendix.

et al. (2006) that it is not suitable for more than $N = 5$ units when the cross-sectional correlation is low. The inverse normal and the CAIN-TSL tests perform best in this case both in terms of size and power.

Diagonal factor loading matrix with $U(0, 1)$ entries generates moderate cross-sectional correlation between the innovations: $\overline{\hat{\rho}_\varepsilon} = 0.18$, while the estimated average correlations between the probits are already around $\overline{\tilde{\rho}_{\text{probit}}} = 0.007$ and $\overline{\tilde{\rho}_{\text{probit}}} = 0.005$ for $T = 100$ and $T = 200$, respectively. The inverse normal test has size close to the nominal one, along with the CAIN-TSL test, while both Hartung's tests again become undersized as N grows. Considering power, the inverse normal test performs best because the nominator of the panel test statistic is inflated by the unattended cross-sectional dependence. The CAIN-TSL test has slightly lower power compared to it for $T = 100$ in case B, but in case A and for $T = 200$ in case B it performs equally well.

Relatively strong cross-sectional correlation between the innovations is introduced by a diagonal factor loading matrix with $U(-1, 3)$ entries: $\overline{\hat{\rho}_\varepsilon} = 0.413$, with estimated average correlation between the probits $\overline{\tilde{\rho}_{\text{probit}}}$ varying between 0.020 and 0.046. In this case the inverse normal test and Hartung's test with κ_2 become oversized for large N, while the size of Hartung's test with κ_1 and that of CAIN-TSL fluctuate around the desired 5% level. In terms of power, in both cases A and B the CAIN-TSL test outperforms Hartung's tests and the test of Simes', being second only compared to the oversized standard inverse normal test. The explanation for this power gain of CAIN-TSL in comparison to Hartung's tests is simple: it is due to the $\tilde{\rho}_{\text{probit}}$ estimator being much more precise than the $\hat{\rho}^*_{\text{probit}}$ one (see Fig. 1). Therefore the variance of the test-statistic does not get overestimated, which would lead to lower power. We note that due to the high correlation between the cross-sections, the power increase over N for all tests is less significant compared to the case of low cross-sectional correlation.

This can be explained by the fact that the different cross-sections exhibit similar dynamics and in a sense carry very much the same information. Hence the marginal value (in terms of power) added by each individual unit is lower compared to when the cross-sectional dependence is weaker and the information set gets therefore richer as N grows.

Lastly, with unrestricted factor loadings we investigate the performance of the tests when all variables are assumed to be correlated to the same extent, on average, over the cross-sections. When $\gamma_i \sim U(0, 1)$, $\overline{\hat{\rho}_\varepsilon} = 0.36$ and the inverse normal test becomes oversized for high N, while both Hartung's κ_1 and the CAIN-TSL tests tend to become undersized. The CAIN-TSL test performs best in terms of power without becoming oversized for both cases A and B. When $\gamma_i \sim U(-1, 3)$, $\overline{\hat{\rho}_\varepsilon} = 0.44$ the CAIN-TSL test tends to become oversized as N grows with size reaching 10% for $T = 200$ and $N = 25$ at the 5% level, while Hartung's κ_1 test observes the nominal size. As these experimental settings violate Assumption 2 for the CAIN-TSL test, we conclude that it is robust to such violations when the estimated mean absolute cross-sectional correlation between the innovations $\hat{\rho}_\varepsilon$ is low to moderate (≤ 0.35), or when the number of cross-sections is small. For higher estimated $\hat{\rho}_\varepsilon$ or when $N > 10$ we'd rather recommend to use Hartung's modification with $\kappa_1 = 0.2$.

Size and power under $H_0 : r_i = 1, \forall i$ for cases A and B are presented in Tables 7 and 8 in the Appendix, respectively. The undersized individual TSL test leads to all five panel tests being severely undersized. Their power nevertheless increases with N for T sufficiently large. For case A the inverse normal test exhibits the highest power, closely followed by the CAIN-TSL test. The same holds for case B when the cross-sectional correlation is low to moderate. For larger correlation the CAIN-TSL test has already slightly lower power than Hartung's tests. This is because it precisely corrects for the high cross-sectional correlation between the probits, which turns out to be unnecessary given the severely undersized results of the individual TSL tests. In these cases the inverse normal test is again the most powerful test (without becoming oversized under the true null $H_0 : r_i = 1, \forall i$). In terms of size-adjusted power (see Table 9 in the Appendix) the CAIN-TSL test is exactly as powerful as the standard inverse normal.

These results can be summarized as follows. When there is only low mean absolute cross-sectional correlation $\hat{\rho}_\varepsilon$, in terms of size the CAIN-TSL performs comparably to the standard inverse normal and Hartung's κ_2 test, being as powerful as the inverse normal and more powerful than both Hartung's tests. This conclusion holds regardless of the null hypothesis under consideration. When the cross-sectional correlation is high, the CAIN-TSL test offers best size-power trade-off under the null hypothesis of no cointegration. This is particularly important in view of the sequential rank testing procedure, which begins with $H_0 : r_i = 0, \forall i$. Also, even when Assumption 2 is violated, the CAIN-TSL test may be used to test for no cointegration when the estimated $\hat{\rho}_\varepsilon$ is low to moderate, or when the number of cross-sections is not too large. Otherwise Hartung's test with κ_1 is recommended, as it controls the size better. When testing for higher cointegrating ranks under high cross-sectional correlation, then it might be preferable to employ the inverse normal test without any correction for the cross-sectional dependence, or Hartung's κ_2 test. They would yield high power without running the risk of an inflated Type I error probability, due to the lower empirical size of the individual TSL test under such null hypotheses. Simes' simple intersection test, although almost correctly sized in all settings, is in general, least powerful.

5.3. Comparison with Westerlund's (2006) test

To help position the CAIN-TSL test in the growing literature on panel cointegration, we evaluate its finite-sample properties in comparison with other similar tests. Our work is closest to that of Banerjee and Carrion-i Silvestre (2015), who propose a panel no-cointegration test accommodating level shifts, trend breaks and even breaks in the cointegrating relation. However, in the case of level and trend breaks their number and locations are restricted to be homogeneous over cross-sections, which may not always be the case in practice. Furthermore, their approach relies on first decomposing the observed time series into common and idiosyncratic components and determining their stochastic properties separately. It

Table 3
Comparison between the CAIN-TSL test and Westerlund's (2006) test under factor-driven cross-sectional dependence.

	Size			Power			Size-adjusted power		
$T\backslash N$	5	15	25	5	15	25	5	15	25
	CAIN-TSL								
100	0.007	0.000	0.000	0.636	0.867	0.938	0.764	0.967	0.994
200	0.020	0.007	0.009	0.813	0.979	0.998	0.858	0.990	0.999
	Westerlund's (2006) test								
100	0.491	0.974	0.999	0.997	1.000	1.000	0.937	1.000	1.000
200	0.393	0.935	0.993	1.000	1.000	1.000	1.000	1.000	1.000

Note: Rejection frequencies at the 5% nominal level based on 5000 replications. Size and power for the CAIN-TSL test are computed under $H_0 : r_i = 0, \forall i = 1, ..., N$, while the true cointegrating rank is zero and one, respectively. Size and power for the Westerlund's (2006) test are computed under $H_0 : r_i = 1, \forall i = 1, ...N$, while the true cointegrating rank is one and zero, respectively.

is, therefore, difficult to draw conclusions based on their test whether the observed variables actually exhibit cointegration or not, which is of primary interest in practice. Hence we rather compare the CAIN-TSL test with another test, which focuses on the cointegration properties of the observed variables.

Westerlund (2006) proposes an LM test for the null hypothesis of cointegration in panel data allowing for structural breaks in the deterministic terms of the data generating process. The heterogeneous number of breaks and the break dates are not necessarily known, but may be estimated from the data. The panel test statistic is computed as the normalized sum of the individual LM test statistics, where the first two moments of the statistic depend only on the number of breaks in each cross-section, but not on their locations or other nuisance parameters. Assuming independence of the cross-sectional units, Westerlund (2006) shows that the limiting distribution of the statistic is standard normal. For the general case of cross-sectionally dependent panels he proposes a bootstrapping algorithm. Nevertheless, he examines the performance of the test without bootstrap correction for a DGP featuring an unobserved common factor in the innovations, and he notes that it is "remarkably robust to moderate degrees of cross-sectional correlation". Hence it is this simpler version of the test which we employ next.

The comparative Monte Carlo study between the CAIN-TSL test and Westerlund's (2006) test is based on a cointegrated bi-variate VAR(2) process, employed by Dolado and Lütkepohl (1996), amended with multifactor error structure; details on this DGP are given in the Appendix. Table 3 summarizes the results. It is evident that although the CAIN-TSL test is undersized, its power increases significantly as N grows. Its size-adjusted power is slightly lower than that of Westerlund's (2006) test when N is as small as 5. However, Westerlund's (2006) test is severely oversized, despite the fact that the average correlation ρ_ε between the innovations for $i \neq j$ is only 0.3. Hence in practice the CAIN-TSL test offers a clear advantage in terms of better control of the Type-I error while maintaining good power properties.

5.4. Performance in higher-order VAR models

As described in Section 4, the link between ρ_ε and ρ_{probit} underlying the proposed CAIN-TSL test is estimated from a response surface regression; the data for which is simulated from a VAR(1) model. We note that, as long as the lag order is consistently estimated by common information criteria (e.g. Akaike's (AIC), Schwarz' (BIC) or the modified AIC of Qu and Perron (2007)), the CAIN-TSL testing procedure remains valid for a VAR process of any lag order. This is so, because the individual LR trace statistics are computed by reduced rank regression, which essentially concentrates out the (consistently estimated) short-term dynamics at its first step, leaving their asymptotic distribution invariant to the value of s. It is, nevertheless, of interest to examine the finite-sample performance of the test in DGPs with higher lag orders. We therefore employ the amended bivariate cointegrated VAR(p) process of Dolado and Lütkepohl (1996) with $s = \{2, 4, 6\}$ to evaluate and compare the properties of the CAIN-TSL test in these settings. Details on the DGP can be found in the Appendix; the lag order is assumed to be known. Table 4 presents the results.

The first thing to notice is that the individual TSL test ($N = 1$) is undersized, but as the lag order increases, it approaches the correct size from below. This observation extends to the panel test, where for fixed s and T the size distortion gets magnified as N grows. For a single cross-section with fixed T the power diminishes as the lag order increases, most likely due to the loss of degrees of freedom as the parameter space expands. This observation, too, carries over to the panel test. Nevertheless, increasing the number of cross-sections for fixed T and s still provides power gains, albeit more modest for higher values of s. Hence, a high cross-sectional dimension cannot fully offset the low power of the test caused by a low time dimension if the lag order is high. For optimal results, VAR models with high lag order are to be employed only if sufficient number of time observations are available.

Table 4
Performance of the CAIN-TSL test in higher-order VAR models.

DGP		Size				Power			
	$T\backslash N$	1	5	15	25	1	5	15	25
VAR(2)	100	0.028	0.007	0.000	0.000	0.389	0.636	0.867	0.938
	200	0.041	0.020	0.007	0.009	0.418	0.813	0.979	0.998
VAR(4)	100	0.048	0.020	0.003	0.002	0.214	0.251	0.303	0.342
	200	0.057	0.025	0.008	0.003	0.313	0.487	0.674	0.781
VAR(6)	100	0.050	0.024	0.006	0.003	0.167	0.198	0.230	0.243
	200	0.044	0.020	0.004	0.001	0.271	0.392	0.529	0.630

Note: Rejection frequencies at the 5% nominal level based on 5000 replications. Size and power are computed under $H_0 : r_i = 0, \forall i = 1, ...N$, while the true cointegrating rank is zero and one, respectively.

6. Testing for cointegration between US regional house prices and personal income

We illustrate the use of the CAIN-TSL test by applying it to investigate a presumed cointegrating relationship between regional house prices and personal income in the United States (US). The evidence on the existence of a long-run relationship between house prices and macroeconomic fundamentals in the US housing market is rather mixed. On the one hand, using 23 years of quarterly data until 2002 on 95 metropolitan areas, Gallin (2006) fails to find support for a stable link between house prices and income. On the other hand, analyzing yearly data from 1975 to 2003 for 49 states, Holly et al. (2010) do establish a long-run equilibrium between real house prices and real per capita income with coefficients $(-1, 1)$, "as predicted by the theory".

We aim to shed some light on this issue by employing the CAIN-TSL test and the newest available data at quarterly frequency, spanning from 1983Q3 to 2018Q1. Such a long window with 141 time observations predisposes single-unit unit root and cointegration tests to have good finite-sample properties. The longer time series, however, come at a price. The global financial crisis of 2007–2008 has caused a structural break in many economic and financial time series, with personal income and house prices being no exception. Therefore, when trying to establish a cointegrating relationship between these two variables, this structural break has to be taken into account in order to avoid spurious inference.

In formulating the econometric model we follow Holly et al. (2010) and look at a log-linear relationship between real house prices, p_{it}, and the macroeconomic fundamentals real per capita disposable income, y_{it}, and real cost of borrowing net of house prices appreciation/depreciation, c_{it}. We model it in a VAR(s_i) framework as $Y_{it} = (p_{it}, y_{it}, c_{it})'$ for each state $i = 1, \ldots, N$, allowing for a level shift and trend break in the deterministic terms:

$$Y_{it} = \mu_{i0} + \mu_{i1} t + \delta_{i0} d_{it} + \delta_{i1} b_{it} + X_{it}, \quad t = 1, \ldots, T, \tag{17}$$

$$X_{it} = A_{1i} X_{i,t-1} + \cdots + A_{p_i,i} X_{i,t-s_i} + u_{it}. \tag{18}$$

The structural break is assumed to take place at a known date, and namely in the third quarter of 2007, which marks the beginning of the US mortgage and credit crisis. Therefore, $\tau = \lambda T = 99$ with $\lambda = 0.7$. Hence the dummy variables d_{it} and b_{it} take values $d_{it} = b_{it} = 0$ for $t < \tau$ and $d_{it} = 1$ and $b_{it} = t - \tau + 1$ for $t \geq \tau$. To account for potential lag effects, however, we check for robustness of this choice by shifting the break date by one or two quarters ahead.

The variables p_{it}, y_{it} and c_{it} are computed as in Holly et al. (2010). We use state-level data for all 50 states and the District of Columbia ($N = 51$). Information on the data sources is listed in Table 10 in the Appendix. Although available, data for eight years prior to 1983 has not been included into the analysis in order to avoid issues arising from heteroscedasticity due to the Great Moderation. The time span of the remaining data is nevertheless sufficiently large.

Each variable is first tested for a unit root by a univariate ADF test. Perron (1989) has shown that not accounting for an existing structural break reduces the test's ability to reject a false null hypothesis of a unit root, so any non-rejections would require further investigation. Rejections, however, would point to stationarity. To reflect their trending behaviour we allow for an intercept and a linear time trend in the models for the income and house prices variables, while for net cost of borrowing only an intercept is included. The results are presented in the first panel of Table 11 in the Appendix. For y_{it} and p_{it} we observe rejections of the unit root null at the 5% significance level for only two and four states, respectively. In contrast, c_{it} is largely classified as stationary with rejections at the 5% level for 37 states and the largest p-value not exceeding 0.14. These results are in line with those of Holly et al. (2010). We therefore next focus only on the house prices and personal income variables.

To investigate whether the income and house price variables are nonstationary also when a structural break is allowed for in the level and the slope of the DGP, we apply the test of Popp (2008). His innovational-outlier (IO) type test is suitable for the application at hand as it allows the break to gradually take effect and permits a break to occur under both the null and the alternative hypotheses. Furthermore, the limiting distribution of the test statistic with an endogenous selection of the break date is the same as in the case of a known break, as is the case here. We again choose to employ a single-unit unit root test, as deciding upon the stochastic properties of the variables on a unit-by-unit basis and not at the panel level allows us to exclude stationary time series from the subsequent cointegration analysis. Table 12 in the Appendix summarizes the

Table 5
Results of the CAIN-TSL test with structural breaks for US house prices dataset.

	Break in 2007Q3			Break in 2007Q4			Break in 2008Q1		
	$\hat{s}_i$	$LR^{\text{trace}}_{\text{TSL}}$	p-value	$\hat{s}_i$	$LR^{\text{trace}}_{\text{TSL}}$	p-value	$\hat{s}_i$	$LR^{\text{trace}}_{\text{TSL}}$	p-value
Alabama	5	5.81	0.930	4	5.61	0.940	–	–	–
Arizona	7	8.57	0.719	4	7.75	0.794	4	8.11	0.761
Arkansas	–	–	–	–	–	–	–	–	–
California	7	13.02	0.301	7	13.54	0.262	7	15.47	0.151
Colorado	7	9.40	0.637	7	8.50	0.725	7	8.04	0.767
Connecticut	3	8.94	0.682	2	6.74	0.874	2	6.31	0.902
Delaware	5	6.19	0.910	5	6.55	0.887	–	–	–
District of Columbia	–	–	–	–	–	–	–	–	–
Florida	–	–	–	–	–	–	–	–	–
Georgia	5	3.93	0.989	8	15.36	0.157	7	17.60	0.077*
Idaho	–	–	–	8	11.67	0.412	–	–	–
Illinois	–	–	–	–	–	–	–	–	–
Indiana	8	5.40	0.949	8	6.65	0.880	7	9.34	0.640
Iowa	8	9.73	0.603	8	8.26	0.747	7	7.40	0.822
Kansas	6	6.45	0.894	6	7.56	0.811	6	11.16	0.458
Kentucky	8	9.57	0.619	8	10.39	0.535	–	–	–
Louisiana	8	4.26	0.983	8	4.71	0.973	8	6.51	0.889
Maine	–	–	–	–	–	–	–	–	–
Maryland	5	8.21	0.753	4	7.59	0.808	4	7.97	0.773
Massachusetts	4	10.06	0.569	5	9.00	0.675	4	9.17	0.657
Michigan	8	5.36	0.951	8	6.25	0.906	7	6.38	0.897
Minnesota	8	5.17	0.958	8	6.91	0.862	7	8.60	0.714
Mississippi	8	8.01	0.772	8	11.12	0.464	8	12.27	0.358
Missouri	7	6.45	0.894	7	7.64	0.804	7	11.17	0.457
Montana	6	7.39	0.825	6	8.26	0.747	7	3.97	0.988
Nebraska	7	9.68	0.608	8	11.88	0.393	7	12.74	0.320
Nevada	–	–	–	–	–	–	–	–	–
New Hampshire	6	9.75	0.600	5	11.73	0.406	4	8.78	0.696
New Jersey	2	6.15	0.912	4	5.17	0.958	6	5.90	0.925
New Mexico	5	8.06	0.767	5	11.66	0.413	4	15.65	0.143
New York	5	6.16	0.912	4	5.28	0.954	4	5.04	0.962
North Carolina	8	9.04	0.672	8	20.55	0.027**	–	–	–
North Dakota	7	9.92	0.583	5	15.59	0.147	7	19.29	0.043**
Ohio	8	6.43	0.895	7	8.89	0.687	7	11.04	0.470
Oklahoma	5	7.29	0.833	6	3.96	0.988	5	7.49	0.816
Oregon	–	–	–	–	–	–	–	–	–
Pennsylvania	5	14.23	0.219	6	18.49	0.057*	5	16.76	0.101
Rhode Island	5	10.05	0.570	4	11.16	0.460	4	10.47	0.526
South Carolina	6	8.55	0.721	8	2.99	0.997	–	–	–
South Dakota	8	13.40	0.273	8	14.47	0.203	8	13.75	0.246
Tennessee	5	5.10	0.961	4	4.52	0.978	8	9.15	0.659
Texas	7	6.86	0.866	7	7.27	0.834	6	9.95	0.578
Utah	–	–	–	–	–	–	–	–	–
Vermont	5	9.83	0.593	4	9.29	0.647	4	9.63	0.611
Virginia	8	8.35	0.740	8	11.30	0.446	8	11.91	0.389
Washington	8	13.52	0.265	8	12.35	0.353	–	–	–
West Virginia	–	–	–	–	–	–	–	–	–
Wisconsin	8	7.70	0.799	8	8.41	0.733	7	8.63	0.711
Wyoming	7	7.63	0.806	6	6.53	0.888	6	11.83	0.396
Alaska	8	15.65	0.145	8	12.63	0.330	8	15.82	0.136
Hawaii	8	6.75	0.874	8	7.93	0.777	6	8.91	0.684
	CAIN-TSL								
$t(\tilde{\rho})$	2.603			1.818			0.723		
$\hat{\rho}_\varepsilon$	0.426			0.421			0.416		
$\hat{\rho}^o_\varepsilon$	0.039			0.038			0.038		
$\tilde{\rho}_{\text{probit}}$	0.055			0.054			0.052		

Notes: $LR^{\text{trace}}_{\text{TSL}}$ denotes the LR trace statistic of Trenkler et al. (2008). The lag order $\hat{s}_i$ is selected by the MAIC criterion of Qu and Perron (2007), augmented to account for the structural break. $\hat{\rho}_\varepsilon$ denotes the average absolute pairwise cross-sectional correlation between the innovations to the same variables in the panel. $\hat{\rho}^o_\varepsilon$ denotes the average absolute pairwise cross-sectional correlation between the innovations to the different variables in the panel. $\tilde{\rho}_{\text{probit}}$ denotes the estimated correlation between the individual probits.

results. Allowing for a structural break around the time of the beginning of the financial crisis does indeed result in rejection of the unit root null at the 5% in some cases.

Excluding these few states where either variable is deemed stationary, we proceed to testing for cointegration with the CAIN-TSL test. The lag order of each individual VAR process has been selected by the modified AIC (MAIC) criterion of Qu and Perron (2007), whose computation has been augmented by dummy variables to account for the structural break. Table 5 presents the results. We note that for a break located in 2007Q3 not a single rejection of the no-cointegration null is observed. If the break is set to one quarter later, a single rejection at the 5% significance level emerges, and namely for North Carolina. For a break in 2008Q1 the null of zero cointegrating rank is rejected at 5% only for North Dakota. Turning to the CAIN-TSL test, we first note that the estimated average absolute correlation between the innovations to the same variables in the panel, $\hat{\rho}_{\varepsilon}$, is about 0.42. It is comparable to the magnitude of the correlations of the variables in first differences between regions reported by Holly et al. (2010) and points to the existence of significant cross-sectional dependence, as expected. This value of $\hat{\rho}_{\varepsilon}$ leads to an estimated correlation between the probits $\tilde{\rho}_{\text{probit}}$ of about 0.05. The estimate of the average absolute pairwise correlation between the innovations to the different variables is 0.039 and below, pointing that Assumption 2 is not violated. The highly positive CAIN-TSL statistics, far from the rejection region in the left tail of the standard normal distribution, only further corroborate the lack of cointegration between real house prices and real per capita disposable income in the selected period.

These findings are in stark contrast with those of Holly et al. (2010), who do establish that cointegration exists at the panel level. An important difference to their study is our extended sample. Holly et al. (2010) use data until 2003 and do not cover the subsequent housing bubble in the early 2000s, which affected more than twenty US states and began to collapse in the mid-2000s, leading to the subprime mortgage crisis. Hu and Oxley (2018) find evidence that following the dot-com crisis many regional house price bubbles have formed on a state level within different time frames instead of a single synchronous divergence from the intrinsic value of the house prices. Our results suggest that the house price dynamics have diverged from the evolution of the real per capita disposable income for most, if not all US states in this period.

Indeed, testing for (no) cointegration as a method to identify existent price bubbles is not new; see, e.g. McDonald and Stokes (2013), and Hu and Oxley (2018). While the panel cointegration test demonstrates the presence of a housing bubble at the aggregate level, the results at the state level are mixed. This outcome could be due to differences in "residential land use regulations and geographic land scarcity" among the states (Huang and Tang, 2012). According to the latter authors, these constraints on the housing supply side together with the subprime mortgage crisis drive the booms and busts in the US house prices. At the state level the data signalizes that a housing bubble is not present in every US state, but rather that localized housing bubbles are observed. This makes it possible for investors to diversify risk (see Huang, 2019) by investing in housing assets in different states, e.g. where geographical land scarcity is less pronounced. Furthermore, the results also indicate that the main driver of the housing bubble hasn't been the cost of borrowing, which was thought to be main cause for the collapse of the financial institutions (see Glaeser et al., 2010).

One question, however, poses further interest. Is there evidence in the data for another, newer departure of the house prices from the income fundamental in the period following the burst of the housing bubble? To answer it, we apply the intersection-type panel cointegration test by Arsova and Örsal (2020) to the data sample from 2008Q1 to 2018Q1 ($T = 41$). The p_{it} and y_{it} variables are found to be $I(1)$ for most states by the ADF test (see the second panel of Table 11 in the Appendix). The real net cost of borrowing c_{it} is, as in the full sample, mostly $I(0)$; the results are not reported to save space. No major economic event has led to a structural break in this period, hence we employ the single-unit cointegration test of Saikkonen and Lütkepohl (2000) without breaks for each state with nonstationary house prices and income variables. The individual p-values are then combined in a panel test by the Simes' procedure. The results are presented in Table 13 in the Appendix. Allowing for a trend in the variables but not in the error-correction (EC) term leads to a rejection of the no-cointegration null hypothesis in 13 states at the 5% level. Two of these values are smaller than the corresponding critical values of Simes, leading to a rejection at the 5% level of the composite null hypothesis of no-cointegration for all states. Hence we find evidence that the US real house prices have not (yet) departed from the equilibrium relationship with real disposable income in the period following the burst of the last house prices bubble.

7. Conclusion

In this paper we propose a new meta-analytic approach (CAIN) to test for the cointegrating rank in panels where structural breaks and cross-sectional dependence are allowed for. It is an extension of the likelihood-based rank test of Trenkler et al. (2008) (TSL), and requires only the p-values of the individual LR trace statistics of the TSL test. The CAIN-TSL test is based on a modification of the popular inverse normal method for p-values combination, employing a novel estimator for the unknown correlation between the probits. We propose a way to estimate this correlation as a function of the system dimension, the cointegrating rank under the null hypothesis and the average absolute cross-sectional correlation between the residuals of the individual VAR models in first differences. The latter is easily estimable in practice and provides an easy-to-interpret measure of the degree of cross-sectional dependence. In a Monte Carlo study we demonstrate the superior properties of the CAIN-TSL test in comparison with other meta-analytic approaches recently proposed in the panel unit-root literature. An application of the test to investigate a presumed long-run equilibrium relationship between house prices and personal income in a panel of 51 US states provides an illustration of its usefulness in practice.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue. The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Acknowledgments

Financial support by the German Research Foundation (DFG) through the project KA-3145/1-2 is gratefully acknowledged. The authors also thank two anonymous referees and the associate editor for many helpful comments and suggestions.

Appendix A

Consistency of the proposed estimator. For the approximate $N(0, 1)$ limiting distribution of the CAIN-TSL test statistic (5) to hold, the estimator $\tilde{\rho}_{\text{probit}}$ must be asymptotically unbiased and asymptotically consistent (Hartung, 1999; Demetrescu et al., 2006). That is, we need to show that

$$\lim_{N\to\infty} \mathbb{E}\left(\frac{1}{N^2}(N + (N^2 - N)\tilde{\rho}_{\text{probit}})\right) = \rho_{\text{probit}}, \quad \text{and} \tag{19}$$

$$\lim_{N\to\infty} \text{Var}\left(\frac{1}{N^2}(N + (N^2 - N)\tilde{\rho}_{\text{probit}})\right) = 0. \tag{20}$$

As discussed, we propose to estimate $\tilde{\rho}_{\text{probit}}$ as a continuous function g of the dimension of the system m, the hypothesized cointegrating rank r and the mean absolute correlation ρ_ε between the innovations of the individual DGPs. The function g results from an ordinary least squares (OLS) regression of mean values $\tilde{\bar{\rho}}_{\text{probit}}$, estimated from simulated data, onto r, m and ρ_ε. The simulations design reflects the assumptions made and allows for as much heterogeneity in the data-generating processes over the different cross-sections as possible. The estimated values of ρ_{probit} are then averaged by means of Fisher's Z-transformation to mitigate any potential bias. Therefore, assuming that the estimated response surface $\tilde{\rho}_{\text{probit}} = g(\rho_\varepsilon, m, r)$ correctly represents the relationship between $\tilde{\rho}_{\text{probit}}$ and ρ_ε, (19) follows from the unbiasedness of the OLS estimator and the unbiasedness of $\hat{\rho}_\varepsilon$ as a sample mean.

Establishing (20), i.e. $\lim_{N\to\infty} \text{Var}(\tilde{\rho}_{\text{probit}}) = 0$, however, is more difficult, if not impossible. Treating m and r as fixed and using the coefficients of the response surface g from Table 1, $\tilde{\rho}_{\text{probit}}$ can be represented as $\tilde{\rho}_{\text{probit}} = a\hat{\rho}_\varepsilon^2 + b\hat{\rho}_\varepsilon^4$. Here the parameters a and b depend solely on m and r and can thus be considered as fixed as well. (20) will then hold iff $\lim_{N\to\infty} \text{Var}(\hat{\rho}_\varepsilon^2) = 0$ and $\lim_{N\to\infty} \text{Var}(\hat{\rho}_\varepsilon^4) = 0$.

We note that $\hat{\rho}_\varepsilon$ is computed as the average of the absolute values of the sample pairwise residual correlations $\hat{\rho}_{li,lj}$, $i \neq j$:

$$\hat{\rho}_\varepsilon = \frac{2}{mN(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{l=1}^{m} |\hat{\rho}_{li,lj}|. \tag{21}$$

Denoting the sample correlation matrix as $\hat{P}$, the population correlation matrix as P and the sample size as n, Neudecker and Wesselman (1990) show that under fairly general conditions $\text{vec}(\hat{P})$ follows a multivariate normal distribution with mean $\text{vec}(P)$ and variance-covariance matrix $\Sigma_{\rho_\varepsilon}$, such that $\left|\left|\Sigma_{\rho_\varepsilon}\right|\right| = O_p\left(n^{-1}\right)$. Therefore, $\text{vec}(\hat{P})$ follows a folded multivariate normal distribution. However, although the folded normal distribution is closed under scale transformations, it is not a stable distribution (Tsagris et al., 2014). That is, the sum of random variables from a folded normal distribution does not follow a folded normal distribution. The distribution of $\hat{\rho}_\varepsilon$ is, to the best of the authors' knowledge, of unknown form, and no analytic results regarding the orders of magnitude of its higher moments can be established at this point. We leave this issue for further research. We note that in practice, provided that the residuals from the individual VAR models are homoscedastic and serially uncorrelated, $\hat{\rho}_\varepsilon$ can be consistently inferred from them as in (21); consistency follows by the

Table 6
True and estimated ρ_ε under $H_0 : r = 0$.

True ρ_ε	Estimated $\hat{\rho}_\varepsilon$		
	True rank 0	True rank 1	True rank 2
0.2	0.200	0.196	0.192
0.4	0.400	0.392	0.384
0.6	0.600	0.589	0.577
0.8	0.800	0.786	0.771

Notes: Results from large-scale simulation study with $m = 3$, $T = 500$ and $N = 5$.

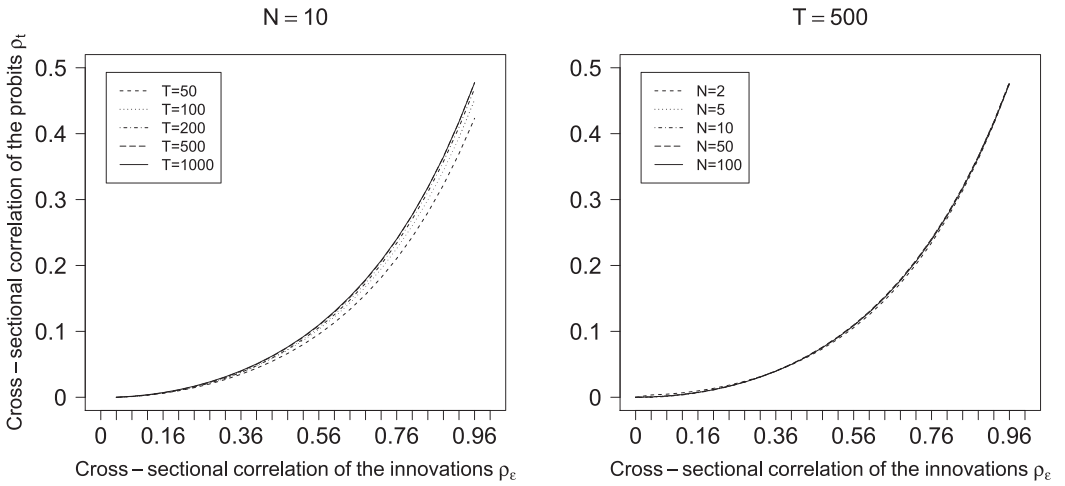

Fig. 2. Cross-sectional correlation of the probits against cross-sectional correlation of the innovations, based on a bivariate system with true cointegrating rank zero, testing for $H_0 : r_i = 0, \forall i$.

Table 7
Monte Carlo study results, case A: not near-unit root processes. Empirical size and power under $H_0 : r_i = 1, \ \forall i = 1, \ldots, N$.

Test	True cointegrating rank 1, hypothesized rank 1: Size						True cointegrating rank 2, hypothesized rank 1: Power					
	T=100			T=200			T=100			T=200		
	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-0.4, 0.4)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-0.4, 0.4)$					
Inverse normal	0.001	0.000	0.000	0.003	0.000	0.000	0.41	0.78	0.92	0.99	1.00	1.00
Hartung κ_1	0.001	0.000	0.000	0.005	0.000	0.000	0.28	0.46	0.56	0.97	1.00	1.00
Hartung κ_2	0.002	0.000	0.000	0.007	0.001	0.000	0.32	0.54	0.63	0.98	1.00	1.00
Simes	0.010	0.006	0.009	0.016	0.017	0.016	0.23	0.27	0.27	0.95	1.00	1.00
CAIN-TSL	0.001	0.000	0.000	0.003	0.000	0.000	0.41	0.77	0.91	0.99	1.00	1.00
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$					
Inverse normal	0.002	0.000	0.000	0.006	0.001	0.000	0.40	0.75	0.88	0.99	1.00	1.00
Hartung κ_1	0.004	0.000	0.000	0.006	0.001	0.000	0.26	0.42	0.51	0.97	1.00	1.00
Hartung κ_2	0.005	0.001	0.000	0.008	0.001	0.000	0.28	0.49	0.58	0.97	1.00	1.00
Simes	0.009	0.007	0.004	0.020	0.019	0.024	0.20	0.24	0.25	0.95	1.00	1.00
CAIN-TSL	0.002	0.000	0.000	0.005	0.001	0.000	0.39	0.72	0.85	0.99	1.00	1.00
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$					
Inverse normal	0.004	0.003	0.002	0.015	0.017	0.017	0.40	0.67	0.79	0.99	1.00	1.00
Hartung κ_1	0.003	0.001	0.001	0.014	0.007	0.005	0.24	0.34	0.41	0.95	1.00	1.00
Hartung κ_2	0.004	0.002	0.001	0.018	0.012	0.011	0.27	0.40	0.47	0.96	1.00	1.00
Simes	0.008	0.007	0.005	0.024	0.018	0.019	0.18	0.21	0.21	0.92	0.99	1.00
CAIN-TSL	0.002	0.001	0.000	0.010	0.008	0.005	0.34	0.55	0.63	0.98	1.00	1.00
	Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$						Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$					
Inverse normal	0.003	0.002	0.002	0.009	0.007	0.006	0.43	0.78	0.90	0.99	1.00	1.00
Hartung κ_1	0.003	0.001	0.001	0.008	0.004	0.003	0.30	0.49	0.60	0.98	1.00	1.00
Hartung κ_2	0.005	0.002	0.002	0.011	0.007	0.005	0.33	0.56	0.67	0.98	1.00	1.00
Simes	0.011	0.010	0.007	0.019	0.014	0.018	0.24	0.29	0.31	0.96	1.00	1.00
CAIN-TSL	0.002	0.001	0.000	0.006	0.002	0.002	0.40	0.69	0.81	0.99	1.00	1.00
	Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$						Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$					
Inverse normal	0.003	0.002	0.002	0.009	0.007	0.006	0.45	0.73	0.82	0.99	1.00	1.00
Hartung κ_1	0.003	0.001	0.001	0.008	0.004	0.003	0.30	0.46	0.53	0.97	1.00	1.00
Hartung κ_2	0.005	0.002	0.002	0.011	0.007	0.005	0.33	0.52	0.58	0.97	1.00	1.00
Simes	0.011	0.010	0.007	0.019	0.014	0.018	0.24	0.29	0.32	0.94	0.99	1.00
CAIN-TSL	0.002	0.001	0.000	0.006	0.002	0.002	0.39	0.60	0.68	0.99	1.00	1.00

Notes: Rejection frequencies at the 5% significance level, 5000 replications. Power is not size-adjusted, as size-adjustment would not be available in practice. Some results with size-adjusted power are available in Table 9 in the Appendix.

Table 8
Monte Carlo study results, case B: near-unit root processes. Empirical size and power under $H_0 : r_i = 1,\ \forall i = 1, \ldots, N$.

Test	True cointegrating rank 1, hypothesized rank 1: Size						True cointegrating rank 2, hypothesized rank 1: Power					
	T=100			T=200			T=100			T=200		
	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$	$N=5$	$N=15$	$N=25$
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-0.4, 0.4)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-0.4, 0.4)$					
Inverse normal	0.000	0.000	0.000	0.001	0.000	0.000	0.02	0.01	0.01	0.37	0.66	0.83
Hartung κ_1	0.000	0.000	0.000	0.002	0.000	0.000	0.03	0.01	0.00	0.34	0.55	0.70
Hartung κ_2	0.000	0.000	0.000	0.003	0.000	0.000	0.03	0.02	0.01	0.38	0.65	0.81
Simes	0.005	0.003	0.002	0.008	0.005	0.006	0.04	0.03	0.04	0.33	0.46	0.53
CAIN-TSL	0.000	0.000	0.000	0.001	0.000	0.000	0.02	0.01	0.01	0.37	0.66	0.83
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$					
Inverse normal	0.000	0.000	0.000	0.002	0.000	0.000	0.03	0.02	0.01	0.38	0.67	0.82
Hartung κ_1	0.001	0.000	0.000	0.001	0.000	0.000	0.02	0.01	0.00	0.32	0.53	0.66
Hartung κ_2	0.001	0.000	0.000	0.002	0.000	0.000	0.03	0.02	0.01	0.36	0.63	0.76
Simes	0.004	0.002	0.001	0.008	0.005	0.006	0.04	0.03	0.03	0.30	0.41	0.47
CAIN-TSL	0.000	0.000	0.000	0.002	0.000	0.000	0.03	0.01	0.01	0.37	0.64	0.79
	Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$						Diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$					
Inverse normal	0.000	0.000	0.000	0.003	0.001	0.000	0.04	0.05	0.05	0.39	0.64	0.76
Hartung κ_1	0.000	0.000	0.000	0.002	0.000	0.000	0.02	0.02	0.02	0.31	0.47	0.58
Hartung κ_2	0.001	0.000	0.000	0.003	0.001	0.001	0.03	0.03	0.04	0.34	0.55	0.65
Simes	0.003	0.002	0.001	0.008	0.004	0.005	0.03	0.03	0.03	0.27	0.35	0.40
CAIN-TSL	0.000	0.000	0.000	0.001	0.000	0.000	0.03	0.02	0.02	0.34	0.52	0.61
	Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$						Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(0, 1)$					
Inverse normal	0.000	0.000	0.000	0.002	0.000	0.000	0.03	0.02	0.02	0.39	0.66	0.80
Hartung κ_1	0.000	0.000	0.000	0.001	0.000	0.000	0.03	0.01	0.01	0.37	0.57	0.71
Hartung κ_2	0.001	0.000	0.000	0.002	0.000	0.000	0.04	0.02	0.01	0.41	0.67	0.81
Simes	0.004	0.003	0.002	0.008	0.006	0.006	0.05	0.05	0.04	0.37	0.52	0.62
CAIN-TSL	0.000	0.000	0.000	0.001	0.000	0.000	0.02	0.01	0.01	0.35	0.57	0.68
	Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$						Unrestricted factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$					
Inverse normal	0.000	0.000	0.000	0.002	0.001	0.001	0.05	0.04	0.06	0.41	0.64	0.75
Hartung κ_1	0.000	0.000	0.000	0.003	0.001	0.001	0.04	0.02	0.02	0.38	0.56	0.67
Hartung κ_2	0.001	0.000	0.000	0.004	0.001	0.001	0.05	0.04	0.04	0.42	0.64	0.76
Simes	0.005	0.003	0.004	0.009	0.005	0.006	0.05	0.05	0.05	0.39	0.54	0.61
CAIN-TSL	0.000	0.000	0.000	0.001	0.000	0.000	0.04	0.02	0.02	0.37	0.51	0.60

Notes: Rejection frequencies at the 5% significance level, 5000 replications. Power is not size-adjusted, as size-adjustment would not be available in practice. Some results with size-adjusted power are available in Table 9 in the Appendix.

Law of Large Numbers as $N \to \infty$ treating m as fixed. Therefore, consistency of $\tilde{\rho}_{\text{probit}}$ should follow from the consistency of the estimator $\hat{\rho}_{\varepsilon}$ and the Continuous Mapping Theorem for the mapping g.

An illustration based on simulation evidence is provided in Fig. 3. It is clear that increasing the cross-sectional dimension of the panel leads to more concentrated distributions of both $\hat{\rho}_{\varepsilon}$ and $\tilde{\rho}_{\text{probit}}$.

DGP of Dolado and Lütkepohl (1996), employed for the Monte Carlo studies in Sections 5.3 and 5.4 The simulations are based on the following DGP for Y_{it}, where the stochastic component $X_{it} = (X_{1t}, X_{2t})'$ follows the cointegrated VAR(s_i) process employed by Dolado and Lütkepohl (1996):

$$Y_{it} = \tilde{d}_{it} + X_{it}, \tag{22}$$

$$\Delta X_{it} = \begin{pmatrix} -\beta & \beta \\ 0 & 0 \end{pmatrix} X_{i,t-1} + \begin{pmatrix} 0.5 & 0.3 \\ 0 & 0.5 \end{pmatrix} \Delta X_{i,t-p+1} + u_{it}, \tag{23}$$

$$u_{it} = \gamma_i f_t + e_{it}. \tag{24}$$

The innovations u_{it} are augmented by a (2×1)-vector of unobserved common factors f_t such that $f_t \sim N(0, I_2)$. The factor loading matrix γ_i is diagonal with uniformly distributed entries $\sim U(-1, 3)$, while the idiosyncratic errors are generated as $e_{it} \sim N(0, I_2)$. The linear trend term $\tilde{d}_{it} = (d_{it}, 0)'$ affects only the first variable in the system, where d_{it} features one or two breaks with 50% chance each at individual-specific break fractions $\lambda_{ij} \in (0.15T, 0.85T)$. In particular, $d_{it} = \delta_{ij} \cdot (1, t)'$, where the intercept and trend parameters are $\delta_{i1} \sim N(1, I_2)$ in the first sub-sample with $t < [\lambda_{i1} T]$, $\delta_{i2} \sim N(2, I_2)$ in the second sub-sample with $t < [\lambda_{i2} T]$, and $\delta_{i3} \sim N(0, I_2)$ in the third. To simulate cointegrated process with rank one, we set $\beta = 1$, while for no cointegration (rank is zero) $\beta = 0$. The lag order, the breaks number and the locations of the breaks are assumed to be known.

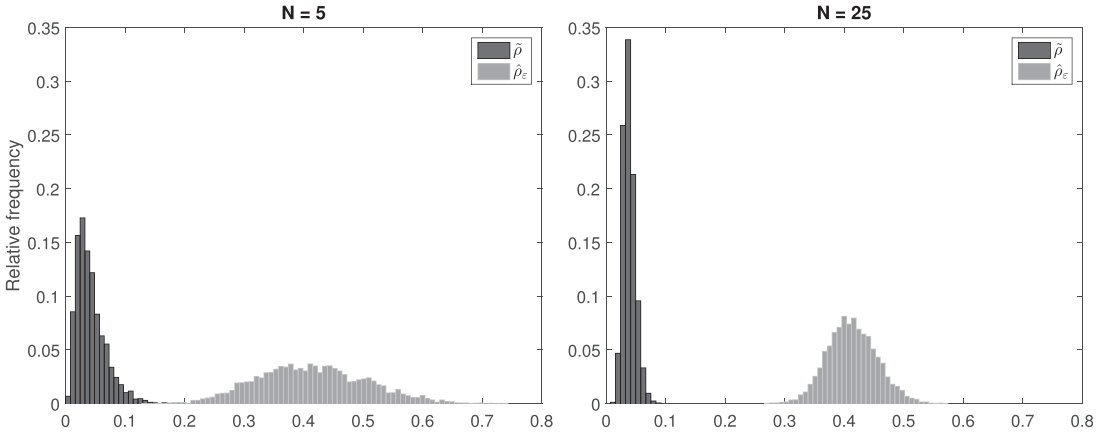

Fig. 3. Empirical distributions of $\hat{\rho}_\varepsilon$ and $\tilde{\rho}_{\text{probit}}$ under $H_0 : r_i = 0, \forall i$, based on a 3-variate VAR(2) process with a multifactor-error structure and a diagonal factor loading matrix $\gamma_i \sim i.i.d.U(-1, 3)$, $T = 100$, 5000 replications.

Table 9
Size-adjusted power for true rank one under $H_0 : r_i = 0, \forall i = 1, ...N$.

	$T = 100$			$T = 200$		
	$N = 5$	$N = 15$	$N = 25$	$N = 5$	$N = 15$	$N = 25$
Diagonal factor loadings $\gamma_i \sim i.i.d.U(-1, 3)$						
Inverse normal	0.40	0.62	0.70	0.95	1.00	1.00
Hartung κ_1	0.30	0.49	0.54	0.85	0.98	0.99
Hartung κ_2	0.29	0.43	0.47	0.82	0.96	0.98
CAIN-TSL	0.40	0.62	0.70	0.94	1.00	1.00
Unrestricted factor loadings $\gamma_i \sim i.i.d.U(-1, 3)$						
Inverse normal	0.62	0.84	0.90	0.97	1.00	1.00
Hartung κ_1	0.57	0.81	0.86	0.97	1.00	1.00
Hartung κ_2	0.54	0.77	0.80	0.97	1.00	1.00
CAIN-TSL	0.62	0.85	0.90	0.99	1.00	1.00
Unrestricted factor loadings $\gamma_i \sim i.i.d.U(-1, 3)$, near-unit root processes						
Inverse normal	0.20	0.25	0.28	0.67	0.85	0.91
Hartung κ_1	0.22	0.31	0.32	0.67	0.90	0.95
Hartung κ_2	0.21	0.31	0.32	0.66	0.90	0.95
CAIN-TSL	0.20	0.25	0.28	0.66	0.85	0.91

Rejection frequencies at 5% nominal level, 5000 replications.

Table 10
US house prices dataset: data description.

Variable	Description	ID	Source
$P_{it,g}$	CPI_{CA} for California; CPI_{FL} for Florida; CPI_{GA} for Georgia; CPI_{TX} for Texas; CPI_{MI} for Michigan; CPI_{IL} for Illinois, Indiana and Wisconsin; CPI_{MA} for Massatchusetts and New Hampshire; CPI_{PA} for Delaware and Maryland; CPI_{NY} for New York, New Jersey and Pennsylvania; CPI_{West} for Alaska, Arizona, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Utah, Washington and Wyoming; CPI_{US} for all remaining states.		
CPI_{US}	Consumer Price Index: Total All Items for the United States; Index 2010=100, Seasonally Adjusted	CPALTT01USQ661S	OECD
CPI_{CA}	Consumer Price Index for All Urban Consumers: All items in San Francisco-Oakland-Hayward, CA (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA422SA0	U.S. Bureau of Labor Statistics
CPI_{FL}	Consumer Price Index for All Urban Consumers: All items in Miami-Fort Lauderdale-West Palm Beach, FL (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA320SA0	U.S. Bureau of Labor Statistics
CPI_{GA}	Consumer Price Index for All Urban Consumers: All items in Atlanta-Sandy Springs-Roswell, GA (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA319SA0	U.S. Bureau of Labor Statistics
CPI_{TX}	Consumer Price Index for All Urban Consumers: All items in Houston-The Woodlands-Sugar Land, TX (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA318SA0	U.S. Bureau of Labor Statistics
CPI_{MI}	Consumer Price Index for All Urban Consumers: All items in Detroit-Warren-Dearborn, MI (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA208SA0	U.S. Bureau of Labor Statistics
CPI_{IL}	Consumer Price Index for All Urban Consumers: All items in Chicago-Naperville-Elgin, IL-IN-WI (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA207SA0	U.S. Bureau of Labor Statistics
CPI_{MA}	Consumer Price Index for All Urban Consumers: All items in Boston-Cambridge-Newton, MA-NH (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA103SA0	U.S. Bureau of Labor Statistics
CPI_{PA}	Consumer Price Index for All Urban Consumers: All items in Philadelphia-Camden-Wilmington, PA-NJ-DE-MD (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA102SA0	U.S. Bureau of Labor Statistics
CPI_{NY}	Consumer Price Index for All Urban Consumers: All items in New York-Newark-Jersey City, NY-NJ-PA (CBSA); Index 1982–1984=100; Not Seasonally Adjusted	CUURA101SA0	U.S. Bureau of Labor Statistics
CPI_{West}	Consumer Price Index for All Urban Consumers: All items in West; Index 1982–1984=100; Not Seasonally Adjusted	CUUR0400SA0	U.S. Bureau of Labor Statistics
$P_{it,h}$	All-Transactions House Price Index; Index 1980:Q1=100; Not Seasonally Adjusted	**STHPI	U.S. Federal Housing Finance Agency
PD_{it}	Approximated by the difference $PInc_{it}$ - G_{it}		
$PInc_{it}$	United States, BEA, Personal Income by Major Component (SQ4), Personal Income, USD		Macrobond
G_{it}	United States, BEA, Personal Income by Major Component (SQ4), Employee and Self-Employed Contributions for Government Social Insurance, USD		Macrobond
POP_{it}	Population in number of persons by state; United States, BEA, Personal Income Summary: Personal Income, Population, per Capita Personal Income (CA1), Population, States; Population (Midperiod) is used from 2017Q1 to 2018Q1; linear interpolation to yield quarterly data, carried out by Macrobond		Macrobond
RB_t	Long-Term Government Bond Yields: 10-year: Main (Including Benchmark) for the United States	IRLTLT01USQ156N	OECD
p_{it}	Natural logarithm of the US State real house price index, $p_{it} = \ln(P_{it,h}/P_{it,g})$		
y_{it}	Natural logarithm of the US State real per capita disposable income, $y_{it} = \ln(PD_{it}/(POP_{it} \times P_{it,g}))$		
r_{it}	US State real long-term interest rate, $r_{it} = RB_t/100 - ln(P_{it,g}/P_{i,t-1,g})$		
c_{it}	US State real cost of borrowing net of real house price appreciation/depreciation, $c_{it} = r_{it} - \Delta p_{it}$		

Notes: OECD stands for OECD, Main Economic Indicators - complete database. Where available, ID denotes the variable ID in FRED (Federal Reserve Economic Data). ** stands for the two-letter abbreviation of the corresponding state. Not seasonally adjusted CPI series have been seasonally adjusted by the Census X12 method. All indices have been rebased to 1980Q1=100. Quarterly CPI indices have been derived from monthly and bi-monthly series by taking average values of the three months in a quarter.

Table 11
ADF unit root test results for US house prices dataset.

	Sample 1983Q1 – 2018Q1 (T=141)									Sample 2008Q1 – 2018Q1 (T=41)					
	y_{it}			p_{it}			c_{it}			y_{it}			p_{it}		
	$\hat{s}_i$	t-stat	p-value	$\hat{s}_i$	t-stat	p-value	$\hat{s}_i$	t-stat	p-value	$\hat{s}_i$	t-stat	p-value	$\hat{s}_i$	t-stat	$p-$value
Alabama	3	−2.690	0.243	8	−1.232	0.899	3	−4.676	0.0002***	1	−2.860	0.186	3	−0.746	0.962
Arizona	4	−1.702	0.746	7	−1.494	0.827	3	−3.112	0.028**	4	−2.910	0.170	3	−1.341	0.863
Arkansas	4	−4.106	0.008***	4	−1.031	0.936	3	−3.276	0.018**	4	−2.744	0.225	4	−3.764	0.029**
California	4	−2.121	0.530	4	−0.823	0.960	4	−2.571	0.101	6	−2.994	0.146	5	−5.748	0.0001***
Colorado	4	−1.267	0.892	7	−1.742	0.727	8	−3.134	0.026**	4	−1.947	0.612	4	−1.280	0.879
Connecticut	1	−1.666	0.761	4	−1.694	0.749	3	−3.379	0.013**	1	−2.900	0.173	3	−2.190	0.482
Delaware	5	−1.489	0.829	8	−1.203	0.906	3	−3.552	0.008***	3	−2.394	0.377	1	−0.884	0.948
District of Columbia	3	−1.935	0.631	5	−0.764	0.966	4	−4.662	0.0002***	6	−1.941	0.615	1	−3.731	0.031**
Florida	4	−2.092	0.545	5	−0.831	0.960	3	−2.936	0.044**	5	−2.725	0.232	3	−4.132	0.012**
Georgia	3	−2.516	0.320	7	−0.884	0.954	3	−3.528	0.009***	3	−2.370	0.389	4	−1.002	0.933
Idaho	4	−1.626	0.778	7	−1.635	0.774	4	−3.970	0.002***	4	−2.733	0.230	3	−1.280	0.879
Illinois	3	−2.865	0.177	8	−0.928	0.949	3	−2.931	0.044**	4	−2.830	0.195	3	−1.680	0.742
Indiana	1	−3.414	0.054*	7	−1.026	0.936	4	−2.712	0.074*	6	−2.613	0.277	3	−1.007	0.932
Iowa	5	−2.887	0.170	7	−1.482	0.831	4	−2.645	0.087*	5	−1.846	0.664	3	−1.290	0.876
Kansas	4	−2.146	0.515	5	−1.625	0.778	4	−2.851	0.054*	5	−1.709	0.729	3	−0.829	0.954
Kentucky	3	−2.277	0.443	8	−0.993	0.941	4	−2.817	0.059*	5	−2.771	0.216	3	−0.669	0.969
Louisiana	4	−1.729	0.733	5	−3.667	0.028**	7	−2.895	0.048**	3	−2.753	0.222	1	−1.421	0.840
Maine	3	−2.168	0.503	6	−1.844	0.678	3	−4.106	0.001***	3	−2.239	0.457	3	−1.281	0.879
Maryland	3	−1.764	0.717	6	−0.701	0.971	3	−2.894	0.049**	3	−1.478	0.821	3	−2.799	0.206
Massachusetts	4	−2.588	0.287	7	−2.374	0.391	3	−3.033	0.034**	3	−2.729	0.231	3	−2.228	0.462
Michigan	4	−3.301	0.070*	8	−0.939	0.948	3	−3.670	0.006***	5	−2.928	0.165	4	−2.562	0.299
Minnesota	5	−2.468	0.344	8	−0.883	0.954	8	−3.457	0.011**	4	−2.741	0.226	4	−1.871	0.651
Mississippi	2	−1.398	0.858	6	−1.605	0.787	5	−3.804	0.004***	6	−4.305	0.008***	3	−1.055	0.924
Missouri	6	−2.597	0.282	8	−0.960	0.945	3	−2.505	0.117	4	−3.335	0.075*	3	−0.923	0.943
Montana	1	−2.918	0.160	8	−1.474	0.834	7	−6.917	0.000***	1	−1.658	0.752	3	−1.450	0.830
Nebraska	1	−3.222	0.084*	8	−1.474	0.834	6	−2.404	0.142	1	−1.783	0.695	3	−1.038	0.927
Nevada	5	−1.289	0.887	6	−0.811	0.961	3	−3.750	0.004***	1	−3.097	0.121	4	−3.998	0.017**
New Hampshire	5	−2.190	0.491	7	−1.370	0.866	3	−3.073	0.031**	6	−3.047	0.133	3	−1.833	0.670
New Jersey	3	−2.137	0.521	5	−1.574	0.798	3	−3.039	0.034**	1	−3.011	0.142	3	−2.257	0.447
New Mexico	4	−1.588	0.793	6	−1.971	0.612	3	−3.511	0.009***	4	−2.378	0.385	3	−1.410	0.843
New York	2	−3.249	0.079*	6	−2.151	0.513	3	−3.651	0.006***	1	−4.028	0.015*	3	−1.539	0.799
North Carolina	1	−2.815	0.194	8	−1.158	0.914	3	−2.766	0.066*	1	−2.450	0.350	3	−0.396	0.984
North Dakota	5	−2.244	0.461	7	−2.098	0.542	6	−8.977	0.000***	1	−0.404	0.984	5	−1.291	0.876
Ohio	3	−3.419	0.053*	8	−0.894	0.953	3	−3.159	0.025**	6	−2.749	0.223	4	−1.153	0.907
Oklahoma	5	−1.810	0.695	4	−4.464	0.002***	3	−2.804	0.060*	4	−1.011	0.931	1	−1.421	0.840
Oregon	3	−2.131	0.524	5	−1.182	0.910	5	−3.722	0.005***	4	−2.281	0.434	6	−2.119	0.520
Pennsylvania	2	−3.295	0.071*	6	−1.304	0.883	3	−2.577	0.100	5	−3.005	0.143	3	−1.312	0.871
Rhode Island	4	−2.213	0.478	5	−1.569	0.801	5	−3.659	0.006***	1	−3.066	0.128	3	−1.511	0.809
South Carolina	1	−3.021	0.130	7	−1.206	0.905	3	−3.649	0.006***	4	−2.533	0.312	3	−0.482	0.981
South Dakota	1	−2.186	0.494	8	−3.389	0.057*	7	−11.940	0.000***	6	−2.542	0.308	3	−0.927	0.943
Tennessee	1	−2.698	0.239	8	−1.188	0.909	4	−4.151	0.001***	6	−2.630	0.270	3	−0.595	0.974
Texas	4	−1.498	0.826	4	−1.161	0.914	3	−2.479	0.123	1	−1.776	0.698	4	−0.920	0.944
Utah	4	−1.885	0.657	5	−1.650	0.768	5	−3.351	0.015**	4	−2.614	0.276	4	−1.909	0.632
Vermont	3	−2.081	0.552	6	−0.859	0.957	3	−4.142	0.001***	6	−2.992	0.147	5	−1.487	0.818
Virginia	4	−1.841	0.680	5	−0.610	0.977	3	−2.921	0.046**	4	−2.686	0.247	3	−2.406	0.371
Washington	2	−3.069	0.118	4	−0.935	0.948	4	−2.521	0.113	4	−2.673	0.253	4	−1.472	0.823
West Virginia	1	−2.947	0.151	7	−3.306	0.070*	4	−7.782	0.000***	3	−3.348	0.073*	1	−1.815	0.679
Wisconsin	3	−3.177	0.093*	8	−0.603	0.977	4	−2.434	0.134	6	−3.293	0.082*	3	−0.745	0.963
Wyoming	3	−2.147	0.515	6	−4.539	0.002***	5	−5.230	0.000***	3	−1.684	0.740	1	−1.410	0.843
Alaska	1	−3.667	0.028**	8	−2.392	0.382	7	−7.267	0.000***	4	−1.690	0.738	1	−2.175	0.490
Hawaii	4	−2.317	0.422	7	−3.705	0.025**	3	−2.705	0.076*	3	−1.845	0.664	1	−1.838	0.668

Notes: Intercept and trend included for p_{it} and y_{it}, only intercept included for c_{it}. $\hat{s}_i$ denotes the lag order chosen by AIC with a maximum of 8 lags for the 1983Q1–2018Q1 sample and with a maximum of 6 lags for the 2008Q1–2018Q1 sample, respectively. We are grateful to Christoph Hanck for providing us with the GAUSS code for the computation of the $p-$values.

Table 12
Results of Popp's (2008) unit root test with structural breaks for US house prices dataset.

	Real per capita disposable income y_{it}						Real house price index p_{it}					
	Break in 2007Q3		Break in 2007Q4		Break in 2008Q1		Break in 2007Q3		Break in 2007Q4		Break in 2008Q1	
	$\hat{s}_i$	t-stat	$\hat{s}_i$	t-stat	$\hat{s}_i$	t-stat	$\hat{s}_i$	t-stat	$\hat{s}_i$	t-stat	$\hat{s}_i$	t-stat
Alabama	0	−4.112*	7	−3.433	2	−4.650**	6	−1.675	6	−1.769	6	−1.955
Arizona	3	−3.216	3	−3.206	3	−4.009*	0	−1.785	0	−1.933	0	−2.051
Arkansas	3	−6.382**	3	−5.962**	3	−7.184**	3	−3.519	3	−3.534	3	−3.386
California	3	−3.188	3	−3.187	3	−3.578	3	−3.700	3	−3.669	3	−3.317
Colorado	3	−4.055*	3	−3.663	3	−3.778	6	−3.465	6	−3.475	6	−3.331
Connecticut	4	−3.433	4	−3.343	4	−3.173	3	−3.167	3	−3.236	3	−3.128
Delaware	8	−2.586	8	−2.272	8	−2.203	7	−3.903	7	−4.126*	7	−4.321**
District of Columbia	3	−4.646**	3	−4.566**	3	−4.901**	4	−2.593	4	−2.789	4	−2.893
Florida	2	−4.327**	3	−4.389**	3	−4.797**	3	−3.126	3	−3.106	6	−3.742
Georgia	2	−3.205	2	−2.390	2	−2.426	6	−3.295	6	−3.337	6	−3.342
Idaho	8	−4.300**	8	−4.176*	8	−4.850**	6	−3.419	5	−3.204	6	−3.694
Illinois	2	−4.311**	2	−4.121*	2	−4.780**	7	−4.502**	7	−4.619**	7	−4.638**
Indiana	0	−2.600	2	−2.856	2	−3.248	6	−1.558	8	−1.795	8	−1.518
Iowa	4	−2.797	4	−2.829	4	−2.856	8	−2.994	8	−2.921	8	−2.636
Kansas	3	−3.247	3	−3.234	0	−1.836	4	−1.939	4	−2.042	6	−2.158
Kentucky	2	−3.223	2	−3.226	3	−4.276**	0	−1.347	0	−1.403	7	−2.176
Louisiana	7	−2.787	7	−3.164	3	−3.946*	0	−2.808	0	−3.027	0	−3.161
Maine	2	−2.323	2	−2.286	2	−2.395	8	−4.488**	8	−4.382**	8	−4.394**
Maryland	2	−2.012	2	−2.121	2	−2.163	4	−2.664	4	−2.874	6	−3.972*
Massachusetts	3	−2.964	3	−3.032	3	−3.083	5	−3.901	3	−3.243	5	−3.775
Michigan	8	−2.306	8	−1.989	3	−3.018	6	−3.261	6	−3.293	6	−2.551
Minnesota	4	−2.673	4	−2.897	4	−3.138	7	−3.807	7	−3.829	7	−3.622
Mississippi	1	−3.339	1	−3.285	1	−3.964*	0	−1.737	0	−1.833	0	−1.944
Missouri	3	−3.716	3	−3.711	3	−4.189*	7	−3.357	7	−3.458	7	−3.563
Montana	0	−2.255	0	−2.389	0	−2.302	7	−3.337	7	−3.540	7	−3.586
Nebraska	0	−3.375	0	−3.389	0	−3.439	6	−2.184	6	−2.151	6	−1.862
Nevada	4	−5.629**	4	−4.708**	4	−4.303**	3	−4.053*	3	−3.801	3	−3.440
New Hampshire	2	−2.147	2	−2.160	3	−2.554	6	−3.793	6	−3.864	6	−3.799
New Jersey	2	−2.632	2	−2.847	2	−2.953	3	−3.315	3	−3.342	3	−3.236
New Mexico	7	−2.162	7	−2.267	7	−2.358	3	−1.957	3	−2.359	3	−2.375
New York	0	−3.255	0	−3.367	0	−3.381	5	−3.798	5	−3.739	5	−3.649
North Carolina	0	−3.171	8	−3.368	2	−4.419**	6	−3.195	6	−3.288	6	−3.441
North Dakota	4	−2.005	4	−1.564	4	−1.306	6	−2.220	6	−2.438	6	−2.363
Ohio	2	−3.134	2	−3.205	2	−3.791	3	−0.581	3	−0.638	6	−1.065
Oklahoma	4	−1.663	4	−1.942	4	−1.569	3	−3.640	3	−3.718	3	−3.723
Oregon	2	−3.559	2	−3.592	2	−4.326**	6	−4.658**	4	−4.571**	4	−4.439**
Pennsylvania	0	−3.371	0	−3.440	2	−3.889	5	−3.515	5	−3.537	5	−3.457
Rhode Island	2	−2.575	2	−2.638	2	−2.734	4	−3.773	4	−3.775	3	−3.200
South Carolina	3	−3.840	0	−3.667	3	−4.815**	6	−1.717	6	−1.802	6	−2.027
South Dakota	6	−3.266	6	−2.882	6	−2.373	7	−2.486	7	−2.555	7	−2.535
Tennessee	0	−2.653	0	−2.620	0	−2.896	6	−2.981	6	−3.064	6	−3.134
Texas	2	−3.039	2	−3.184	3	−3.913	3	−2.124	3	−2.289	3	−2.325
Utah	3	−4.164*	3	−4.283**	3	−5.602**	6	−4.584**	4	−4.309**	6	−4.482**
Vermont	2	−2.595	2	−2.680	2	−2.749	5	−3.584	5	−3.729	5	−3.554
Virginia	3	−2.737	3	−2.872	3	−3.092	3	−2.433	3	−2.507	6	−3.925
Washington	5	−4.106*	5	−4.164*	5	−4.619**	3	−3.105	3	−3.350	3	−3.286
West Virginia	0	−5.115**	0	−5.064**	0	−4.803**	4	−1.724	4	−1.999	4	−2.094
Wisconsin	0	−3.362	0	−3.244	2	−3.157	2	−1.865	2	−1.989	2	−1.701
Wyoming	2	−3.156	2	−3.364	2	−2.838	3	−3.124	3	−3.320	3	−3.504
Alaska	8	−3.457	8	−3.410	8	−2.519	7	−2.675	7	−2.830	7	−2.949
Hawaii	2	−1.883	2	−2.004	2	−2.212	3	−3.512	3	−3.557	3	−3.565

Notes: A single break in both intercept and trend included. Optimal lag order denoted by $\hat{s}_i$. The 5% and 10% critical values are −4.2736 and −3.9417, respectively, for a break in 2007Q3; −4.2638 and −3.9312, respectively, for a break in 2007Q4; and −4.2536 and −3.9203, respectively, for a break in 2008Q1.

Table 13
Results of the Simes-SL test for US house prices dataset, sample 2008Q1–2018Q1.

	Trend allowed in the EC term			Trend orthogonal to the EC term			
	$\hat{s}_i$	LR_{SL}^{trace}	p-value	$\hat{s}_i$	LR_{SL}^{trace}	p-value	Simes' crit. value
Utah	7	14.39	0.083	7	25.85	0.000*	0.001
Washington	4	18.28	0.018	4	18.41	0.001*	0.002
Louisiana	8	6.63	0.704	8	14.42	0.006	0.003
North Dakota	8	19.62	0.010	8	14.65	0.006	0.005
Idaho	2	15.24	0.060	2	14.25	0.007	0.006
Oklahoma	3	10.94	0.259	3	13.82	0.009	0.007
Oregon	4	12.47	0.161	4	12.92	0.013	0.008
Wyoming	3	13.72	0.105	3	12.68	0.015	0.009
South Carolina	2	13.74	0.104	2	10.84	0.033	0.010
Rhode Island	2	8.26	0.517	2	10.79	0.034	0.011
Texas	4	10.58	0.288	4	10.34	0.041	0.013
Colorado	7	11.89	0.194	7	10.19	0.044	0.014
Missouri	2	8.81	0.456	2	10.00	0.048	0.015
Montana	2	11.74	0.203	2	9.50	0.059	0.016
North Carolina	6	8.32	0.510	6	8.66	0.083	0.017
Arizona	2	15.90	0.047	1	7.78	0.119	0.018
Kentucky	2	8.41	0.500	2	7.53	0.132	0.019
Illinois	2	10.89	0.263	2	7.31	0.143	0.020
Minnesota	2	9.18	0.417	2	7.26	0.146	0.022
New Hampshire	2	6.06	0.767	2	7.06	0.158	0.023
Tennessee	2	9.59	0.376	2	6.81	0.174	0.024
Virginia	2	6.00	0.774	1	6.79	0.175	0.025
Ohio	2	7.14	0.646	2	6.69	0.182	0.026
New Jersey	2	7.11	0.650	2	6.65	0.185	0.027
West Virginia	1	10.66	0.281	1	6.58	0.190	0.028
New Mexico	2	6.42	0.728	2	6.53	0.194	0.030
Alaska	1	6.27	0.745	1	6.39	0.204	0.031
Kansas	1	5.92	0.782	1	6.34	0.207	0.032
Massachusetts	2	6.27	0.745	2	6.16	0.222	0.033
Wisconsin	2	7.77	0.574	2	5.91	0.243	0.034
Pennsylvania	2	7.18	0.642	2	5.75	0.258	0.035
Connecticut	2	5.83	0.792	2	5.61	0.272	0.036
Vermont	2	6.71	0.696	1	5.43	0.290	0.038
Georgia	2	6.86	0.679	2	5.39	0.294	0.039
Alabama	2	7.98	0.548	2	5.25	0.308	0.040
Hawaii	5	12.96	0.137	2	5.13	0.322	0.041
Michigan	4	10.40	0.303	4	5.03	0.333	0.042
South Dakota	2	5.39	0.835	2	4.62	0.381	0.043
Nebraska	2	7.32	0.625	2	4.58	0.387	0.044
Delaware	2	6.11	0.762	2	4.31	0.422	0.045
Iowa	2	7.58	0.595	7	4.04	0.461	0.047
Maine	2	3.99	0.940	2	3.75	0.504	0.048
Indiana	2	5.66	0.809	2	3.01	0.624	0.049
Maryland	2	4.74	0.891	2	2.81	0.658	0.050

Notes: LR_{SL}^{trace} denotes the LR trace statistic of Saikkonen and Lütkepohl (2000). The lag order $\hat{s}_i$ is selected by the MAIC criterion of Qu and Perron (2007). The results are presented sorted by the $p-$value of the test with trend orthogonal to the EC term, as required for the Simes procedure. p-values which are smaller than the corresponding critical value of Simes are marked by an asterisk. Arkansas, California, District of Columbia, Florida, Mississippi, Nevada and New York are excluded as the unit root null hypothesis has been rejected by the ADF test at the 5% level for either y_{it} or p_{it} or both.

References

Arsova, A., Örsal, D.D.K., 2018. Likelihood-based panel cointegration test in the presence of a linear time trend and cross-sectional dependence. Econom. Rev. 37 (10), 1033–1050.

Arsova, A., Örsal, D.D.K., 2020. Intersection tests for the cointegrating rank in dependent panel data. Commun. Stat. - Simul. Comput. 4, 918–941.

Banerjee, A., Carrion-i Silvestre, J.L., 2015. Cointegration in panel data with structural breaks and cross-section dependence. J. Appl. Econom. 30 (1), 1–23.

Choi, I., 2001. Unit root tests for panel data. J. Int. Money Financ. 20, 249–272.

Costantini, M., Lupi, C., 2013. A simple panel-CADF test for unit roots. Oxf. Bull. Econ. Stat. 75 (2), 276–296.

Demetrescu, M., Hanck, C., 2012. Unit root testing in heteroscedastic panels using the cauchy estimator. J. Bus. Econ. Stat. 30 (2), 256–264.

Demetrescu, M., Hassler, U., Tarcolea, A.I., 2006. Combining significance of correlated statistics with application to panel data. Oxf. Bull. Econ. Stat. 68 (5), 647–663.

Dolado, J.J., Lütkepohl, H., 1996. Making wald tests work for cointegrated VAR systems. Econ. Rev. 15 (4), 369–386.

Gallin, J., 2006. The long run relationship between house prices and income: evidence from local housing markets. Real Estate Econ. 34, 417–438.

Glaeser, E.L., Gottlieb, J.D., Gyourko, J., 2010. Can Cheap Credit Explain the Housing Boom? Technical Report, 16230. The National Bureau of Economic Research Working Paper.

Hanck, C., 2013. An intersection test for panel unit roots. Econ. Rev. 32 (2), 183–203.
Hartung, J., 1999. A note on combining dependent tests of significance. Biometr. J. 41 (7), 849–855.
Holly, S., Pesaran, H.M., Yamagata, T., 2010. A spatio-temporal model of house prices in the USA. J. Econom. 158, 160–173.
Hu, Y., Oxley, L., 2018. Bubbles in US regional house prices: evidence from house price-income ratios at the State level. Appl. Econ. 50 (29), 3196–3229.
Huang, H., Tang, Y., 2012. Residential land use regulation and the us housing price cycle between 2000 and 2009. J. Urban Econ. 71 (1), 93–99.
Huang, M., 2019. A nationwide or localized housing crisis? evidence from structural instability in US housing price and volume cycles. Comput. Econ. 53, 1547–1563.
Johansen, S., Mosconi, R., Nielsen, B., 2000. Cointegration analysis in the presence of structural breaks in the deterministic trend. Econom. J. 3, 216–249.
McDonald, J.F., Stokes, H.H., 2013. Monetary policy, mortgage rates and the housing bubble. Econ. Financ. Res. 1 (1), 82–91.
Neudecker, H., Wesselman, A.M., 1990. The asymptotic variance matrix of the sample correlation matrix. Linear Algebra Appl. 127, 589–599.
Örsal, D.D.K., Arsova, A., 2017. Meta-analytic cointegrating rank tests for dependent panels. Econ. Stat. 2, 61–72.
Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57, 1361–1401.
Pesaran, M.H., 2015. Testing weak cross-sectional dependence in large panels. Econ. Rev. 34 (6–10), 1089–1117.
Popp, S., 2008. New innovational outlier unit root test with a break at an unknown time. J. Stat. Comput. Simul. 78 (12), 1145–1161.
Qu, Z., Perron, P., 2007. A modified information criterion for cointegration tests based on a VAR approximation. Econ. Theory 23, 638–685.
Saikkonen, P., Lütkepohl, H., 2000. Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process. J. Time Ser. Anal. 21 (4), 435–456.
Sarkar, S.K., 1998. Some probability inequalities for ordered MTP2 random variables: a proof of the simes conjecture. Ann. Stat. 26 (2), 494–504.
Simes, R.J., 1986. An improved Bonferroni procedure for multiple tests of significance. Biometrika 73 (3), 751–754.
Stouffer, S., DeVinney, L., Suchmen, E., 1949. The American Soldier: Adjustment During Army Life. Princeton University Press, Princeton, NJ.
Toda, H.Y., 1995. Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions. Econ. Theory 11 (5), 1015–1032.
Trenkler, C., Saikkonen, P., Lütkepohl, H., 2008. Testing for the cointegrating rank of a VAR process with level shift and trend break. J. Time Ser. Anal. 29 (2), 331–358.
Tsagris, M., Beneki, C., Hassani, H., 2014. On the folded normal distribution. Mathematics 2, 12–28.
Wagner, M., Hlouskova, J., 2010. The performance of panel cointegration methods: results from a large scale simulation study. Econom. Rev. 29 (2), 182–223.
Westerlund, J., 2006. Testing for panel cointegration with multiple structural breaks. Oxf. Bull. Econ. Stat. 68 (1), 101–132.
Westerlund, J., Edgerton, D.L., 2008. A simple test for cointegration in dependent panels with structural breaks. Oxf. Bull. Econ. Stat. 70 (5), 665–704.