City Research Online

City St George's, University of London

Citation: Fuertes, A-M. (2008). Sieve bootstrap t-tests on long-run average parameters. Computational Statistics & Data Analysis, 52(7), pp. 3354-3370. doi: 10.1016/j.csda.2007.11.014

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/6420/

Link to published version: https://doi.org/10.1016/j.csda.2007.11.014

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Sieve bootstrap t-tests on long run average parameters

Ana-Maria Fuertes*

Faculty of Finance, Cass Business School, City University,
106 Bunhill Row, London EC1Y 8TZ

First version: June 2005
This version: November 2007

Abstract

Panel estimators can provide consistent measures of a long-run average parameter even if the individual regressions are spurious. However, the t-test on this parameter is fraught with problems because the limit distribution of the test statistic is nonstandard and rather complicated, particularly in panels with mixed (non)stationary errors. A sieve bootstrap framework is suggested to approximate the distribution of the t-statistic. An extensive Monte Carlo study demonstrates that the bootstrap is quite useful in this context.

Keywords: Resampling; Threshold cointegration; Persistent errors; Size distorsion; Power

*Tel.: +44 (0)20 7040 8647; fax: +44 (0)20 7040 8881. *E-mail address:* a.fuertes@city.ac.uk.

1 Introduction

Most of the macroeconomic or financial variables researchers encounter are stochastic trend nonstationary (integrated of order one or I(1) for short) and the theoretical long run relationships that arise among them from arbitrage or market efficiency conditions have often proven rather elusive. The error term in the empirical regressions used to characterize such relationships, albeit truly stationary, can be observationally I(1) in finite samples. This may stem, first, from threshold or Markov cointegration due to transaction costs, lumpy costs of adjustment or major events such as changes in technology, government policy or the presence of bubbles in prices which can interrupt temporarily the adjustment towards an underlying long run equilibrium (Balke and Fomby, 1997; Chortareas et al., 2003; Psaradakis et al., 2004). Second, data aggregation over time or across individuals can induce highly persistent disequilibria as Taylor (2001) demonstrates in the context of spot exchange rate and relative prices. Third, long lags in the response of, say, energy demand to prices or wages to inflation can also result in seemingly nonstationary disequilibria. Against this background, the challenge is to extract the signal, that is, consistently estimate the long run (average) association between the variables and to test whether it satisfies specific theoretical restrictions. Making reliable inferences about these theoretical relationships is important both for forecasting and policy-making purposes.

The econometrics literature has recently established that one advantage of panels versus single time series is that the danger of nonsense regression through lack of cointegration is mitigated. In this context, Pesaran and Smith (1995) show that a cross-section regression for time averaged data produces consistent long run measures. Kao (1999) and Phillips and Moon (1999) develop multi-index asymptotic theory to demonstrate that the Least Squares Dummy Variable (LSDV) and Pooled OLS (POLS) estimators are gaussian and $\sqrt{N}$-consistent for a long run average effect. One important message of this literature is that long run relations are not exclusively associated with cointegrating regressions. For instance, purchasing power parity (PPP) or the long run relation between spot exchange rates and relative prices has been traditionally associated with mean-reverting real exchange rates. However, recent studies argue that it is possible to reconcile PPP and nonstationary real exchange rates because the equilibrium real exchange rate may be a moving function of unobserved I(1) factors (Coakley et al., 2004a).

The problem we seek to tackle is that the usual (or HAC robust) POLS and LSDV standard errors lead to severely distorted t-tests in the context of I(1)-error regressions. This problem will also occur in *iid* or autocorrelated I(0)-error panels when the true (cointegrating) coefficients are heterogeneous. More importantly, in heterogeneous panels with a mix of I(0)- and I(1)-error equations the asymptotic distribution of the LSDV and POLS estimators will depend in a complicated manner upon a nuisance parameter, the fraction of I(1)-error equations.

Our study has two aims. First, it seeks to extend the existing work on nonstationary panels by proposing a bootstrap 'solution' to the aforementioned inference problems. We make use of the sieve bootstrap which can be regarded as a nonparametric procedure. The basic idea of this bootstrap method is to approximate the error process by an AR model of order increasing with the sample size. The sieve bootstrap has been successfully employed to test for an autoregressive unit root (Psaradakis, 2001, 2003; Chang, 2004), to resample from cointegrating regressions (Chang et al., 2006), to conduct inference with VAR models (Inoue and Kilian, 2002a) and to construct prediction intervals for nonlinear time series using neural networks (Giordano et al., 2007). We develop two resampling algorithms that build both on the *fixed regressor bootstrap* of Hansen (2000) and on the *restricted residuals* approach of Nankervis and Savin (1996). One scheme is based on residual pretesting and constructs bootstrap samples that have the I(1) property by construction whereas the other scheme applies the sieve bootstrap method directly to the residuals — we call them, respectively, the pretesting sieve bootstrap (PSB) and the direct sieve bootstrap (DSB).

Second, the paper investigates via Monte Carlo simulation the effectiveness of the bootstrap in controlling the null rejection probability of the usual t-test and its power properties. We consider regression models driven by both AR and MA innovations and allow for heterogeneity across units. Unit root innovations with a negative MA component are also included — this data generating process (DGP) has attracted considerable attention in the literature because it produces observationally *iid* sequences (Psaradakis, 2003; Chang, 2004). The Monte Carlo design covers also: (i) probability distributions typical of economic data, (ii) near unit root or mixed I(0), I(1) innovations, (iii) threshold unit root behavior, and (iv) cross-section dependence.

Several interesting findings emerge. First, asymptotic t-tests on the average slope coefficient yield rejection rates of about 75% (at a nominal level of 5%) in simple homogeneous-slope panels when just one individual error term is unit root persistent and the remaining N-1 errors are white

noise. These large size distortions apply to more general nonstationary panel settings and are shown to worsen as the time dimension (T) increases. Second, the sieve bootstrap-t method proposed is shown to facilitate robust inference in a variety of settings which include *iid* or autocorrelated I(0)-error panels, I(1)-error panels and panels comprising a mix of I(0) and I(1) errors. The sieve bootstrap t-tests remain correctly sized in panel regressions with asymmetric, highly leptokurtic innovations, I(1) errors with a negative MA component, and cross-section dependence. In the problematic case of near-I(1) errors that may result from a threshold cointegrating mechanism, the DSB scheme is shown to work better than the PSB algorithm. Moreover, the unit root pretesting (PSB) effects some power loss in the t-tests. Hence, the practical recommendation that emerges from this study is to employ the DSB testing approach.

The paper is structured as follows. Section 2 outlines the model and assumptions. Section 3 discusses the asymptotic properties of the panel estimators under study. Section 4 describes the bootstrap techniques and Section 5 analyses the simulation findings. A final section concludes.

2 Panel model and assumptions

Let data be generated for N cross-section individuals (or units) and T time periods according to

$$y_{it} = \mu_i + \beta_i x_{it} + u_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T, \tag{1}$$

$$x_{it} = x_{i,t-1} + e_{it}, \quad u_{it} = \rho_i u_{i,t-1} + \nu_{it}, \tag{2}$$

$$\nu_{it} = \psi_i(L)\varepsilon_{it}, \ \varepsilon_{it} \sim iid(0, \sigma^2_{\varepsilon,i}), \tag{3}$$

where L is the lag operator and $\psi_i(z) = \sum_{j=0}^{\infty} \psi_{i,j} z^j$ with $\psi_{i,0} := 1$. The assumptions made are:

(A1) *The individual error process e_{it} is strictly stationary for all i.*

(A2) *Statistical independence of the processes e_{it} and ν_{is} for all t and s.*

(A3) *The coefficients μ_i and β_i are constant over time but may differ randomly over units, that is, $\mu_i \sim iid(\mu, \sigma^2_\mu)$, $\beta_i \sim iid(\beta, \sigma^2_\beta)$, and $(\mu_i, \beta_i)'$ are distributed independently of x_{it} and u_{it}.*

(A4) *The innovation sequence $\{\varepsilon_{it}\}$ satisfies $\mathbf{E}[\varepsilon_{it}] = 0$, $\mathbf{E}[\varepsilon^2_{it}] = \sigma^2_{\varepsilon,i} > 0$ and $\mathbf{E}[\varepsilon^4_{it}] < \infty$.*

(A5) *The sequence $\{\psi_{i,j}\}$ satisfies $|\sum_{j=0}^{\infty} \psi_{i,j}| > 0$, $\sum_{j=1}^{\infty} j|\psi_{i,j}| < \infty$ and $\psi_i(z) \neq 0$ for $|z| \leq 1$.*

If $|\rho_i| < 1$ for all i, we have a cointegrating panel. If $|\rho_i| = 1$ for all i, there is one unit root in u_{it} and we have a non-cointegrating, I(1)-error panel which is just a univariate (single-regressor) version of the setup in Phillips and Moon (1999; Sections 4 and 6). Assumption (A1) ensures that

x_{it} is an I(1) process. The exogeneity assumption (A2) allows us to build on the limit theory for pooled estimators of I(0)- or I(1)-error panel regressions developed by Phillips and Moon (1999) and Kao (1999). We allow for quite general temporal dependence in ν_{it} through (3) which builds on Wold's decomposition theorem — every weakly stationary, purely non-deterministic stochastic process can be written as a linear filter of uncorrelated random variables. Assumptions (A4) and (A5) are sufficiently general to accommodate weakly dependent processes of practical relevance such as the invertible ARMA with *iid* innovations where $|\psi_{i,j}|$ decays at rate $O(\lambda^j)$ as $j \to \infty$ for $\lambda \in (0,1)$. Thus the process ν_{it} admits an AR(∞) representation which can be approximated by

$$\nu_{it} = \theta_{i,1}\nu_{i,t-1} + ... + \theta_{i,p_i}\nu_{i,t-p_i} + \varepsilon_{it}^{p_i}, \tag{4}$$

where $\varepsilon_{it}^{p_i} = \varepsilon_{it} + \sum_{j=p_i+1}^{\infty} \theta_{i,j}\nu_{i,t-j}$. It follows that $\sum_{j=p_i+1}^{\infty} \theta_{i,j} = o(p_i^{-1})$ from assumption (A5) and so if p_i increases with T, the error in finite approximation of ν_{it} can be made arbitrarily small. In sum, the regression errors u_{it} are allowed to be heterogeneous across units in the I(0) or I(1) sense. Further stationary AR (or MA) dependence is also possible. Cross-section heteroskedasticity is allowed and, since $Var(u_{it}) = \nu_{i0} + tVar(\nu_{it})$ when $\rho_i = 1$, the presence of a unit root implies time-series heteroskedasticity also. The regression disturbances may be contemporaneously correlated across units, $cov(u_{it}, u_{jt}) \neq 0$. To simplify the exposition, the initial discussion abstracts from cross-section dependence but this issue is revisited in the simulations below.

The static equation (1) typifies the empirical framework of many cross-country studies of PPP, the Feldstein-Horioka puzzle or economic growth, inter alios. In the context of PPP, nominal exchange rates (y_{it}) are regressed against price differentials (x_{it}). One goal is to test for a unit long-run average price elasticity irrespective of the stationarity properties of the individual residual sequences. The main idea is to accommodate innovations that are observationally I(1) due to Balassa-Samuelson productivity effects and other real shocks, or stemming from measurement error, transaction costs and other market imperfections such as limits to arbitrage (Taylor, 2001; Coakley et al., 2004a). Several recent Feldstein-Horioka studies acknowledge protracted current account imbalances due to productivity and demographic shocks, so it seems important to account for observationally I(1) disturbances in country saving-investment regressions so as to measure global average effects (Taylor, 1998; Herbertsson and Zoega, 2000; Coakley et al., 2004b). Likewise, in growth studies output is regressed against the stock of physical capital and/or education, and

the error term captures technical progress. The individual equations are not long run equilibrium relations if technology is I(1), but the average capital (or education) elasticity $E(\beta_i)$ is still of interest for economists. As Temple (1999, p.126) puts it '[ ] given that the purpose of cross-country empirical work is often to arrive at generalizations about growth, the averages are important'.

Pesaran and Smith (1995) and Phillips and Moon (1999) were the first to note that an average or mean effect can be consistently estimated in panel equations with individual I(1) errors. By drawing an analogy with classical regression, Phillips and Moon (1999) define the *long run average* regression coefficient $\beta \equiv E(\Omega_{y_i x_i})/E(\Omega_{x_i x_i})$ where $\Omega_{y_i x_i}$ and $\Omega_{x_i x_i}$ are, respectively, the long run covariance of $Z_{it} = (y_{it}, x_{it})'$ and long run variance of x_{it}. Phillips and Moon (1999) and Kao (1999) prove theoretically that this mean effect β is consistently measured through the POLS and LSDV estimators ($\hat{\beta}$) when the individual errors u_{it} are I(0) or I(1). They further demonstrate that $\hat{\beta}$ converges to the normal distribution both in I(0)- and I(1)-error panels.

But there remain some inference problems. Suppose the goal is to test a hypothesis on the long run average, $H_0 : \beta = \beta_0$ and $H_1 : \beta \neq \beta_0$, using the t-statistic, $\hat{\Gamma} = (\hat{\beta} - \beta_0)/s_{\hat{\beta}}$ where $\hat{\beta}$ is

$$\hat{\beta}^{POLS} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \overline{y})(x_{it} - \bar{x})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x})^2}, \tag{5a}$$

or

$$\hat{\beta}^{LSDV} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \overline{y}_i)(x_{it} - \overline{x}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)^2}, \tag{5b}$$

with $\overline{y} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$ and likewise for $\overline{x}$. The standard error $s_{\hat{\beta}}$ is obtained from

$$s^2_{\hat{\beta}^{POLS}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2/(NT-2)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x})^2}, \tag{6a}$$

and

$$s^2_{\hat{\beta}^{LSDV}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2/(NT-N-1)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)^2}, \tag{6b}$$

if the regression disturbances are spherical, $u_{it} \sim iid(0, \sigma^2)$, or from an appropriate robust covariance matrix estimator in more general settings, e.g. heteroskedasticity and autocorrelation (HAC) Newey-West style corrections. However, this testing exercise is fraught with difficulties if (at least one of) the individual regression errors is I(1). The key issue is that the conventional formulae ($s^2_{\hat{\beta}}$) or HAC robust corrections dramatically underestimate the true dispersion of $\hat{\beta}$ in the I(1)-error case. Kao's (1999) simulations in a I(1)-error setup like ours, equations (1)-(2), but simplified to abstract from heteroskedasticity and further stationary autocorrelations reveals that:

a) the true dispersion of the LSDV estimator $\hat{\beta}$ is unaffected by T, in line with the $\sqrt{N}$-consistency property of $\hat{\beta}$, b) the theoretical standard error $s_{\hat{\beta}}$ falls rapidly with T. As a result, the true dispersion of the t-statistic grows significantly with T and so the t-statistic does not converge to any meaningful distribution. Moreover, when the true DGP is a heterogeneous-slope panel, even if $u_{it} \sim iid$ there is an additional error component reflected in the residuals which is I(1) and heteroskedastic, $(\beta_i - \beta)x_{it}$, because the LSDV and POLS estimators impose common slopes. For a practical analysis of the role of heterogeneity, see Fuertes and Kalotychou (2006).

Asymptotic covariance matrices for I(1)-error panels are derived in Phillips and Moon (1999) using multi-index asymptotics but they are quite cumbersome to estimate and this may partially explain the absence of applications in the literature as yet. They require kernel estimates of the long run covariance matrices for each i, and so their small sample properties are sensitive to bandwidth choice — kernel estimators can be substantially biased in small and moderately sized samples, yielding tests with finite sample properties that are very different from those predicted in large-sample theory. Moreover, inference based on these asymptotic covariance matrices will be problematic for mixed panels: if all error terms are I(0), the convergence rate of the POLS (LSDV) estimator is $T\sqrt{N}$, if all error terms are I(1) the convergence rate is $\sqrt{N}$. Hence, the appropriate normalization constant needed to derive the asymptotic covariance matrix is model dependent (it depends on a nuisance parameter, the fraction of I(1) errors) and difficult to obtain.

3 Sieve bootstrap tests

This section presents bootstrap procedures for inference on the long run average coefficient. Section 3.1 discusses the construction of pseudodata while Section 3.2 deals with the bootstrap p-values.

3.1 Generating bootstrap panel samples

Our approach is in the spirit of the sieve bootstrap proposed by Bühlmann (1997) which relies on the approximation of an infinite-dimensional, nonparametric model by a sequence of finite-dimensional parametric models such that the dimension increases with the sample size. Accordingly, the temporal dependence in the data is removed by an AR($p_i(T)$) approximation, a so-called sieve, where $p_i(T) \to \infty$ and $p_i(T) = o(T)$ as $T \to \infty$. Bühlmann demonstrates the asymptotic validity of the sieve bootstrap for a general class of nonlinear estimators. This is also proven by Psaradakis

(2001, 2004) and Chang (2004) for unit root tests, Chang et al. (2006) for cointegrating coefficients, and Inoue and Kilian (2002a) for smooth functions of VAR slope parameters and innovation variances. Due to observational equivalence, apart from the linear class of stochastic processes, the sieve bootstrap procedure may also be successful in cases where the series at hand is nonlinear but satisfies an α-mixing condition, an issue which is explored in Section 4 below.

Two different algorithms are suggested, called the *pretesting* sieve bootstrap (PSB) and the *direct* sieve bootstrap (DSB), which share three aspects. First, they build on the *fixed regressor bootstrap* approach of Hansen (2000) which amounts to treating the regressor as fixed in resampling, that is, $x^*_{it} \equiv x_{it}$. Second, they build on the idea of *restricted regression* in resampling (Nankervis and Savin, 1996; Li and Maddala, 1997) and so the scheme $y^*_{it} = \hat{\mu} + \beta_0 x^*_{it} + u^*_{it}$ or $y^*_{it} = \hat{\mu}_i + \beta_0 x^*_{it} + u^*_{it}$ is employed. Here $\hat{\mu}$ ($\hat{\mu}_i$) is the constrained POLS (LSDV) estimator of the intercept term, and u^*_{it} are the bootstrap innovations obtained by resampling (as detailed below) the restricted regression residuals. The latter are given by $\hat{u}_{it} = y_{it} - (\hat{\mu} + \beta_0 x_{it})$ for POLS and by $\hat{u}_{it} = y_{it} - (\hat{\mu}_i + \beta_0 x_{it})$ for LSDV. Third, the sieve order p_i is chosen through the Akaike Information Criterion (AIC) or Schwarz Bayesian Criterion (SBC) alongside a sequential 0.05-level testing down approach. Thus the sieve order p_i is the integer that minimizes $\log \hat{\sigma}^2_{\tilde{T}} + 2p_i/\tilde{T}$ with AIC and $\log \hat{\sigma}^2_{\tilde{T}} + 2p_i \log \tilde{T}/\tilde{T}$ with SBC, where $\hat{\sigma}^2_{\tilde{T}} = \tilde{T}^{-1} \sum_{t=p_i+1}^{T} \hat{\varepsilon}^2_{it}$ is the residual variance and $\tilde{T} = T - p_i$ is the number of observations used in estimating the sieve.

Pretesting sieve bootstrap (PSB) algorithm

This approach to resampling the regression errors, u_{it}, deals separately with their unit root nonstationary properties and with the remaining stationary dependence. To preserve the former, the order of integration of the error term for each panel member, $u_{it} \sim I(d_i), d_i \in \{0, 1\}$, is built into the bootstrap errors by construction — the presence of one unit root may be known a priori (from an existing theory or consensus empirical evidence) or otherwise pretested. To establish results of practical relevance, in the simulations the order of integration is identified by subjecting each individual residual sequence $\{\hat{u}_{it}\}$ to the ADF test using MacKinnon (1996) one-sided critical values. The augmentation order is selected by a 0.05-level testing down procedure from $k_{max} = 10$. Accordingly, $\hat{\nu}_{it} \equiv \hat{u}_{it}$ in the I(0) case and $\hat{\nu}_{it} \equiv \Delta \hat{u}_{it}$ in the I(1) case.

Next a finite AR(p_i) approximation or sieve, as given in (4), is consistently estimated by

single-equation OLS for each residual sequence $\{\hat{\nu}_{it}\}$ thereby allowing heterogeneity in individual autocorrelation structures and variances (cross-section heteroskedasticity) such that $\sigma^2_{\varepsilon,i} \neq \sigma^2_{\varepsilon,j}$. Since the individual-specific means of the POLS (or LSDV) residuals, $N^{-1}\sum_i \hat{\nu}_{it}$, are not necessarily zero, an intercept is included in the sieve. Alternatively, if T is moderate and the panel members are believed to be (near) homogeneous, efficiency gains can be obtained by employing a single sieve, i.e. by fitting an AR model to the pooled $NT \times 1$ residual vector $\hat{\nu}$. The bootstrap residuals u^*_{it} are constructed using any of two resampling tools, called b_1 and b_2, respectively.

In the b_1 version, B stationary sequences $\{\nu^*_{it}\}_{j=1}^B$ are generated recursively for each i using

$$\nu^*_{it} = \hat{\theta}_{0i} + \hat{\theta}_{1i}\nu^*_{i,t-1} + ... + \hat{\theta}_{pi}\nu^*_{i,t-p_i} + \varepsilon^*_{it},\ \varepsilon^*_{it} \sim iidN(0, \hat{\sigma}^2_{\varepsilon,i}), \tag{7}$$

starting from some random initial values $V_0^* = (\nu^*_{i0}, \nu^*_{i1}, ..., \nu^*_{i,p_i-1})'$, where $\hat{\sigma}^2_{\varepsilon,i}$ is the squared standard error of the sieve. Since using the same V_0^* in every bootstrap loop may alter the stationarity properties of $\hat{\nu}_{it}$, we follow the block initialization approach of Stine (1987) and divide the sequence $\{\hat{\nu}_{it}\}$ into $T - p_i + 1$ overlapping blocks of length p_i. A block is randomly selected (with replacement) for V_0^* in each bootstrap. Bootstrap residuals u^*_{it} are constructed with $\hat{d}_i$ unit roots imposed, $\Delta^{\hat{d}_i} u^*_{it} = \nu^*_{it}$. Rewriting the latter using partial sums or stochastic trends, for instance, for $\hat{d}_i = 1$ we generate $u^*_{it} = u^*_{i0} + \sum_{k=1}^t \nu^*_{ik}$ with $u^*_{i0} = 0$.

In b_2, the residual sequence $\{u^*_{it}\}$ is constructed from $\{\nu^*_{it}\}$ in the same manner. However, one important difference is that the pseudo-innovations ε^*_{it} are drawn with replacement from the sieve residuals for each unit i, after adjusting for location and scale, that is, from

$$\tilde{\varepsilon}_{it} = \sqrt{\frac{T}{T - \hat{p}_i}}(\hat{\varepsilon}_{it} - T^{-1}\sum_{t=1}^T \hat{\varepsilon}_{it}),$$

rather than assuming a gaussian distribution. The PSB test is denoted $\Gamma^*_{b_1}$ or $\Gamma^*_{b_2}$, respectively.

Direct sieve bootstrap (DSB) algorithm

This approach applies the sieve approximation directly to $\hat{u}_{it}$ rather than to the transformed residuals $\hat{\nu}_{it} \equiv \Delta^{d_i}\hat{u}_{it}$, $d_i \in \{0, 1\}$. Accordingly, B sequences $\{u^*_{it}\}_{j=1}^B$ are obtained for each i using

$$u^*_{it} = \hat{\gamma}_{0i} + \hat{\gamma}_{1i}u^*_{i,t-1} + ... + \hat{\gamma}_{pi}u^*_{i,t-p_i} + \varepsilon^*_{it},$$

assuming $\varepsilon^*_{it} \sim iidN(0, \hat{\sigma}^2_{\varepsilon,i})$ as in the b_1 scheme above, or drawing ε^*_{it} with replacement from the sieve residuals as in b_2. The DSB test is denoted $\tilde{\Gamma}^*_{b_1}$ or $\tilde{\Gamma}^*_{b_2}$, respectively, in each scheme.

This algorithm is motivated by the findings in Inoue and Kilian (2002b) regarding how to bootstrap persistent processes of unknown order of integration. They demonstrate that the standard bootstrap algorithm for unrestricted autoregressions is asymptotically valid for many I(1) processes. These include the important cases for applied work of a (near) random walk process with drift and high-order autoregressive (near) unit root processes with or without drift. In our context, the pooled LSDV (or POLS) residuals have zero mean by construction which is not the case for the individual residual sequences, $\{\hat{u}_{it}\}_{t=1}^{T}$. Thus in effect, the appropriate sieve for $\{\hat{u}_{it}\}_{t=1}^{T}$ has a non-zero intercept. For the validity of the DSB scheme, it is required that $(\hat{\gamma}_{0i}, \hat{\gamma}_{1i}, ..., \hat{\gamma}_{pi})'$ is consistent. It is well known that OLS satisfies the latter when u_{it} is stationary (Brockwell and Davies, 1991) and in the unit root case (West, 1988).

3.2 Bootstrap inference

In the spirit of the bootstrap-t method (see the monograph by Efron and Tibshirani, 1993; Ch. 12), for each of B bootstrap panel samples $\{(y_{it}^*, x_{it}^*)\}_{j=1}^{B}$ we calculate the bootstrap t-statistic $\hat{\Gamma}^* = (\hat{\beta}^* - \hat{\beta})/s_{\hat{\beta}}^*$ where $\hat{\beta}^*$ and $s_{\hat{\beta}}^*$ are, respectively, the bootstrap POLS (LSDV) slope estimate and its HAC robust standard error. The theoretical p-value of a two-tailed test is defined as $p_{\hat{\Gamma}} \equiv \mathrm{P}_{\beta_0}(|\Gamma| > |\hat{\Gamma}|)$ where $\hat{\Gamma}$ is the usual t-statistic computed from the observed sample and $\mathrm{P}_{\beta_0}(\cdot)$ indicates probability under the null. This p-value is estimated as $\hat{p}_{\hat{\Gamma}}^* \equiv \frac{1}{B}\sum_{j=1}^{B} I(|\hat{\Gamma}_j^*| > |\hat{\Gamma}|)$ where $\{\hat{\Gamma}_1^*, ..., \hat{\Gamma}_B^*\}$ is the sequence of bootstrap t-statistics and $I(\cdot)$ is an indicator function. The null hypothesis is rejected if the bootstrap p-value, $\hat{p}_{\hat{\Gamma}}^*$, falls below the nominal level α.

We consider a second method (an earlier use of the bootstrap) which makes use of the fact that, in the present context, the distribution of $\hat{\beta}$ is known to be asymptotically normal. The bootstrap estimator $\tilde{s}_{\hat{\beta}}^* = \sqrt{B^{-1}\sum_{j=1}^{B}(\hat{\beta}_j^* - \bar{\beta}^*)^2}$ is used to approximate the standard error of $\hat{\beta}$, where $\{\hat{\beta}_1^*, \hat{\beta}_2^*, ..., \hat{\beta}_B^*\}$ is the sequence of long run average coefficients (with mean $\bar{\beta}^*$) estimated for each of the bootstrap samples. The bootstrap-studentized t statistic (denoted $\tilde{\Gamma}^*$) obtained by substituting $\tilde{s}_{\hat{\beta}}^*$ for $s_{\hat{\beta}}$ in $\hat{\Gamma}$ is used to make inferences on the basis of the N(0,1) quantiles.

4 Finite sample properties

4.1 Simulation design

The finite sample behaviour of the sieve bootstrap t-tests on the long run average coefficient is now analyzed by means of Monte Carlo experiments. The following DGP is used

$$\begin{aligned} y_{it} &= \mu_i + \beta_i x_{it} + u_{it}, \quad i = 1, 2, ..., N, \quad t = 1, 2, ..., T, \\ \Delta x_{it} &= e_{it}, \; e_{it} = \pi_i e_{i,t-1} + \xi_{it}, \\ u_{it} &= \rho_i u_{i,t-1} + \nu_{it}, \end{aligned} \tag{8}$$

where ν_{it} obeys any of the stationary processes

$$\text{(AR) } \nu_{it} = \theta_i \nu_{i,t-1} + \varepsilon_{it}, \quad \text{(MA) } \nu_{it} = \varepsilon_{it} + \psi_i \varepsilon_{i,t-1}, \tag{9}$$

for $|\theta_i| < 1$. The innovations are generated according to

$$\left(\varepsilon_{it}, \xi_{it}\right)' \sim iidN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varphi\sigma^2 \\ \varphi\sigma^2 & \sigma^2 \end{pmatrix}\right],$$

where the signal-noise ratio is given by σ^2 and the endogeneity by φ. We set $\varphi = 0$ so that u_{it} and x_{is} are independent for all t, s. The initial set of simulations rules out contemporaneous cross-section dependence and is calibrated to match the sample dimensions and signal-noise ratio of the post Bretton-Woods monthly OECD spot exchange rate (y) and price differential (x) panel in Coakley and Fuertes (2001) — $\sigma^2 = 0.2, N = 15$, $T = 300$. The first $T_0 = 50$ observations are dropped for each i. The intercept is $\mu_i \sim iidU(-0.5, 0.5)$ and $\mu_i = 0$ for the LSDV and POLS simulations, respectively. All computations are programmed in GAUSS.

We set $\rho_i = 0$ for all i to simulate a cointegrating panel. For the I(1)-error panel, we set $\rho_i = 1$ for all i. Mixed I(0), I(1) panels are obtained by setting $\rho_i = 0$ for a fixed fraction of individuals $i = 1, 2, ..., [\lambda N]$ and $\rho_i = 1$ elsewhere. A wide spectrum of cases is considered, $\lambda = \{0.05, 0.2, 0.5, 0.8, 0.95\}$. We set $\pi_i = 0.5$ to introduce temporal dependence in Δx_{it} also.

The baseline $\nu_{it} \sim iid$ case is considered by setting $\theta_i = 0$ ($\psi_i = 0$). Next we allow for AR processes, $\theta_i \in \{0.5, 0.9\}$, and MA processes, $\psi_i \in \{0.5, 0.9\}$. Cross-section heteroskedasticity in ν_{it} is introduced through random coefficients $\theta_i \sim iidU(0.3, 0.5)$, $\psi_i \sim iidU(0.3, 0.5)$, first, and $\theta_i \sim iidU(0.2, 0.9)$, $\psi_i \sim iidU(0.2, 0.9)$, second, to consider different degrees of heterogeneity. Negatively correlated MA errors are introduced through $\psi_i \in \{-0.5, -0.8\}$ and $\psi_i \sim iidU(-0.9, -0.2)$. The

case $\psi_i = -0.8$ has attracted considerable attention in the unit root testing literature (for instance, see Psaradakis, 2001).

The assumption that the regression disturbances u_{it} are gaussian is relaxed by drawing ε_{it} from pdfs that have been shown to be empirically relevant in business and economics. These include a Student's t with five degrees of freedom ($D1 = t_5$), a shifted chi-squared ($D2 = \chi^2_5 - 5$) and, following Nankervis and Savin (1996), the highly leptokurtic mixture of normals $D3 = 0.8N(0,1)\&0.2N(0,16)$. $D1$ captures fat tails whereas $D2$ characterizes asymmetry and leptokurtosis. All distributions are rescaled to have unit variance.

Finally, the panel setup is generalised to accommodate threshold cointegration which is a plausible rationale for observationally I(1) errors. We generalize (5) as follows

$$u_{it} = \begin{cases} +c + \rho_i(u_{i,t-1} - c) + \nu_{it} & \text{if } u_{i,t-1} > c, \\ u_{i,t-1} + \nu_{it} & \text{if } c \geq u_{i,t-1} \geq -c, \\ -c + \rho_i(u_{i,t-1} - c) + \nu_{it} & \text{if } u_{i,t-1} < -c, \end{cases} \tag{10}$$

with $\nu_{it} \sim iidN(0,1)$. This model allows for discontinuous adjustment to equilibrium, namely, a 'band of inaction' around zero. For large positive (negative) disequilibria u_{it}, there is mean reversion towards c (-c). The idea is that only when u_{it} exceeds a critical threshold, do the benefits of adjustment exceed its costs and hence, economic agents act to move the system back towards equilibrium. Balke and Fomby (1997) simulate (10) using $\rho_i = 0.4$ and $c = \{3, 5, 10\}$ to represent geometrically ergodic processes satisfying $\alpha-$mixing conditions. The resulting sequence $\{u_{it}\}_{t=1}^T$ is observationally equivalent to an AR1 series with unconditional first-order autocorrelations of 0.90, 0.96 and 0.99, respectively. We use the same ρ_i and the wider range $c \in \{1, 2, 3, ..., 12\}$.

The hypothesis of interest is $H_0 : \beta = \beta_0$. All tests are based on the t statistic using the HAC Newey-West covariance matrix with truncation lag $L = \lfloor 4\left(\frac{T}{100}\right)^{2/9} \rfloor$. The experiments deal with two methods of inference. One uses the standard normal distribution (Γ). The other is a sieve bootstrap in its gaussian ($\Gamma^*_{b_1}$) or semi-parametric ($\Gamma^*_{b_2}$) form. Unless otherwise noted, the bootstrap tests are based on the AIC in selecting the sieve lag order $p_i \in \{0, 1, ..., \tilde{p}_T\}$ where 0 signifies the iid case. The maximum sieve order considered is $\tilde{p}_T = 10$ which corresponds to Buhlmann's (1997) criteria, $\tilde{p}_T = \lfloor a \log_{10} T \rfloor$, with $a = 4$.

Each of the Monte Carlo replications follows the steps: (i) Generate $Z_{it} = (y_{it}, x_{it})'$ data using (5)-(9); (ii) Test for $H_0 : \beta = \beta_0$ at the 5% significance level, and record $R = 1$ if rejection and 0 otherwise; (iii) Repeat the above two steps M times; (iv) Compute the rejection frequency of the

test, R/M. $M = 1,000$ Monte Carlo replications and $B = 500$ bootstrap repetitions are used.

4.2 Simulation results

Finite sample performance in terms of size (or level) is studied first. We start with homogeneous-slope panels, $\beta_i = 1$. Next, we allow for heterogeneity, $\beta_i \sim iidU(0.7, 1.3)$. The hypothesized value is $\beta_0 = 1$. To allow for some random variation, a confidence interval for the Type-I error probability estimate $\hat{\alpha}$ is formed using the (binomial) standard error estimator $\sigma_{\hat{\alpha}} = \sqrt{\alpha(1-\alpha)/M}$. For $\alpha = 0.05$ and $M = 1,000$ this gives the two-standard error confidence interval (0.036, 0.064).

Figure 1 displays the empirical distribution function (EDF) of the LSDV $t-$statistic over 1,000 Monte Carlo replications, alongside that of a standard normal. The plots correspond to panel DGPs where $\mu_i \sim iidU(-0.5, 0.5)$, $\beta_i = 1$, and $\pi_i = 0.5$. Plots A and B pertain to cointegrating ($\rho_i = 0$) regressions with *iid* errors ($\theta_i = 0$) and AR1 errors ($\theta_i = 0.9$), respectively. Plots C and D correspond, respectively, to a non-cointegrating panel ($\rho_i = 1$) and a mixed I(0)-, I(1)-error panel — $\rho_i = 0$ in 80% of the units and $\rho_i = 1$ elsewhere — with $\theta_i = 0.9$ in both.

[Figure 1 around here]

Plot A corroborates that standard asymptotic inference is valid in cointegrating panel regressions with *iid* errors. But when the error sequences are all I(0) but strongly autocorrelated, all I(1) or a mix the sampling variability of the t-statistic is well above that of the standard normal. However, its distribution remains symmetric. The EDF of the LSDV estimator $\hat{\beta}$ for the same four DGPs alongside a normal EDF with the same variance (Appendix Figure A1) lends support to the extant panel theory. The distribution of $\hat{\beta}$ is centered on the true value of unity and approximately normal except for the mixed I(0)- and I(1)-error panel where it shows significant leptokurtosis. The Appendix material for the paper is available at *www.cass.city.ac.uk/faculty/a.fuertes*.

Homogeneous slope coefficients

Table 1 reports the empirical size of the t-tests based on the LSDV estimator in homogeneous-slope panels. The results for the POLS estimator are quite similar (see Appendix Table A1).

[Table 1 around here]

In the *iid*-error case, inference based in the N(0,1) quantiles is reliable as one would expect and so Γ is correctly sized. Reassuringly, the same is true of the bootstrap-t approach ($\Gamma^*_{b_1}$ and $\Gamma^*_{b_2}$) with empirical sizes that clearly lie within the two standard error confidence limits. The rejection rates from the PSB and DSB methods are similar so, to preserve space, the table reports the former only. But the main issue is whether correct rejection probabilities are attained with I(0)-errors that are autocorrelated and possibly cross-sectionally heteroskedastic, and in I(1)-error panels.

As Table 1 shows, the test Γ is still correctly sized in cointegrating panels with either homogeneous or heterogeneous AR (MA) dependence in u_{it}. But a large degree of autocorrelation effects significantly oversized tests — for $\theta_i = 0.9$ and $\theta_i \sim U(0.2, 0.9)$ the rejection rate jumps to 34.9% and 16.4%, respectively. Reassuringly, the empirical significance level of the bootstrap tests is reasonably close to the nominal level, despite two potential pitfalls of our resampling approach. One is that we resample u_{it} and, hence, do not explicitly incorporate the information that x_{it} is I(1). Another is that we do not correct for sieve order and parameter uncertainty. For small T samples, the AR estimates from OLS are downward biased, particularly for strongly autocorrelated (persistent) series, so the bootstrap may not be as effective. However, the accuracy of the sieve can be substantially improved using finite-sample bias corrections such as the median-unbiased AR estimation approach of Andrews and Chen (1994). This issue is addressed below.

The right-hand side of Table 1 reports the empirical level of t-tests for I(1)-error regressions. The simulations confirm the theoretical result that conventional asymptotic inference (Γ) leads to unacceptably large size distortions of about 68% for the panel dimensions under study. By contrast, the sieve bootstrap tests attain essentially the correct level. Other intercept or slope parameter specifications give also qualitatively similar results. For instance, using $\beta_i = 0$, $\pi_i = \theta_i = 0$ in (5) so that y_{it} and x_{it} are now two independent random walks, and testing for $H_0 : \beta = 0$ gives rejection probabilities of 65.8% (Γ) and 4.5% ($\Gamma^*_{b_2}$). Hence, sieve bootstrap-t tests will not suggest a significant long run relationship when it actually does not exist.

Regarding the issue of pooled or individual resampling, Table 1 illustrates that both approaches (reported in normal and italic font, respectively) give the correct level for $T = 300$. One exception is the AR case for $\theta_i = 0.9$ where the pooled resampling generally works better. This may be because, by pooling the residuals $\hat{\nu}_{it}$, a larger sample is effectively used in the sieve approximation and so the downward bias problem (for θ_i near 1) is mitigated.

The alternative bootstrap-studentized t approach yields correctly sized bootstrap tests in I(0)-error panels as long as the autocorrelation in the errors is not strong — for $\theta_i > 0.9$ the bootstrap tests are somewhat distorted both for pooled and individual resampling. Likewise, there are some distortions in the context of I(1) errors. Detailed results are given in Appendix Tables A2(a,b). This suggests that computing the bootstrap version of the (improperly studentized) t-statistic $\hat{\Gamma}$ for each bootstrap sample is more effective than calculating the bootstrap standard error of $\hat{\beta}$ in order to studentize $\hat{\Gamma}$. Hence, we focus on the bootstrap-t approach hereafter.

Heterogeneous slope coefficients

In the context of I(1)-error panels with heterogeneous slopes β_i, the first-order differenced residuals measure $\nu_{it} + (\beta_i - \beta)\Delta x_{it}$ where ν_{it} is an AR1 or MA1 process with innovations $\varepsilon_{it} \sim iidN(0,1)$. The AR1 process $(\beta_i - \beta)\Delta x_{it}$ is orthogonal to ε_{it} and has variance $(\beta_i - \beta)^2 \sigma^2$. The residual sequence $\{\Delta \hat{u}_{it}\}$ for each $i = 1, ..., N$ is therefore a realization from AR1 or ARMA(1,1) processes with heterogeneous coefficients and cross-section heteroskedasticity. The DGPs considered are like those in Table 1 but allowing $\beta_i \sim U(0.7, 1.3)$. The empirical size of the asymptotic and bootstrap tests is, respectively, 72% and 5.6% on average for the different DGPs considered — results are reported in Table A3 in the Appendix. Hence, the sieve bootstrap tests generally attain the correct nominal size in heterogeneous-slope I(1) panels also.

Non-gaussian disturbances

The empirical level of the tests when the regression errors are non-gaussian I(1) is reported in Table 2. For the panel dimensions under consideration ($N = 15$, $T = 300$), the bootstrap tests still perform quite well for skewed and/or leptokurtic errors. Additional simulations for $T = 150$ ceteris paribus produce qualitatively similar results.

[Table 2 around here]

For $T = 60$, the bootstrap tests deteriorate slightly because the sieve approximation is less accurate, but they still clearly outperform the asymptotic test Γ. For instance, for $\theta_i = 0.9$ using D_1 (Student's t with 5 d.f.) errors the rejection rate of the LSDV t-test is 36.3% (Γ) and 3.6% ($\Gamma^*_{b_2}$) when the errors are I(0), and 76.3% (Γ) and 6.7% ($\Gamma^*_{b_2}$) when the errors are I(1). For $D3$ (normal

mixture) errors, the corresponding figures are 34.5% (Γ), 3.5% ($\Gamma^*_{b_2}$) and 78.1% (Γ), 6.6% ($\Gamma^*_{b_2}$), respectively. The results for I(0)-error panels are quite similar (see Appendix Table A4).

Mixed I(0), I(1) errors

Table 3 pertains to panels where a fraction λ of equations is cointegrating ($\rho_i = 0$) and the remainder are non-cointegrating ($\rho_i = 1$). As noted earlier, heterogeneous slopes induce I(1) errors in all equations irrespective of ρ_i, so $\beta_i = 1$ is adopted for these simulations. We consider $\lambda = \{0.05, 0.20, 0.50, 0.80, 0.95\}$ alongside the AR and MA parameters $\theta_i = \{0, 0.5, 0.9\}$ and $\psi_i = 0.5$.

[Table 3 around here]

One remarkable result in the *iid*-error panel case ($\theta_i = 0$) is the size distortion of the conventional t-test (Γ) at around 70% for all λ. Of particular interest is its large size distortion that appears for $\lambda = 0.95$, when virtually all the equations (14 out of 15) are cointegrating. This contrasts sharply with the correct size of Γ at 5.7% (POLS) and 4.6% (LSDV) in the counterpart case (Table 1) where all equations are cointegrating, $\lambda = 1$. As Table 3 illustrates, correct inferences can still be made through a sieve-bootstrap when the panel errors are a mix of I(1) and I(0) processes.

Negatively correlated MA errors

We now consider the problematic case where the noise in (5) has a negatively correlated MA1 component $\psi_i = \{-0.5, -0.8\}$. To conserve space we focus on the LSDV estimator and compare the Γ and $\Gamma^*_{b_2}$ tests. We consider as fraction of I(1) errors $\lambda = \{0.00, 0.20, 0.50, 0.80, 0.95\}$ and explore the effects that the choice of the sieve order $\hat{p}_i$ has on the $\Gamma^*_{b_2}$ test. On one hand, we fix the latter at $\hat{p}_i = \{1, 3, 5, 8, \tilde{p}_T\}$ with $\tilde{p}_T = \lfloor 4 \log_{10} T \rfloor$. On the other, we select $\hat{p}_i$ among $\{0, 1, 2, ..., \tilde{p}_T\}$ using either either the AIC, SBC or a sequential 0.05-level testing down approach.

The true sampling variability of $\hat{\beta}$ is underestimated by the HAC formulae when the noise (of at least one equation) is I(1), irrespective of the MA dependence and so the asymptotic test Γ is severely oversized at around 68%. For $\lambda = 0.5$ and $\psi_i = -0.5$, the empirical size of $\Gamma^*_{b_2}$ with sieve order choice as indicated in parenthesis is 4.2%(1), 5.4%(8), 6.1%($\tilde{p}_T$), 6.3%(AIC), 5.5%(SBC), and 5.7%(t-test); for $\psi_i = -0.8$, the results are 2.0%(1), 4.6%(8), 5.4%($\tilde{p}_T$), 5.3%(AIC), 3.7%(SBC) and 4.7%(t-test). Unsurprisingly, for $\psi_i = -0.8$ the PSB test appears rather conservative for small $\hat{p}_i$. However, it improves as the sieve order increases in the allowed range $\{1, 2, ..., \tilde{p}_T\}$. Sieve

orders $8 \leq p_i \leq \hat{p}_T = 10$ suffice to ensure correctly sized tests. The other specifications give similar results as detailed in Appendix Table A5. By contrast, the DSB test tends to overreject for small p_i. But again the distortions vanish as p_i increases, although larger sieve orders ($10 < p_i < 15$) are generally needed for the DSB scheme to attain the correct level.

Choice of sieve order and sample size effects

The foregoing analysis does not reveal systematic differences between the AIC and SBC for the choice of sieve order with one exception, the former criterion seems to be preferred over the latter in the context of large negative MA roots. This is in agreement with Kilian (1998) who examines these criteria in the context of bootstrapping (V)AR models and finds that the adverse consequences of over-parameterizing an AR model for bootstrapping purposes may be less severe than those of under-parameterizing it. The testing-down approach performs similarly to the AIC.

We analyse panels with N fixed at 15 and $T = \{60, 80, 100, ..., 350\}$, and T fixed at 150 with $N = \{10, 12, 15, ..., 45\}$. The focus is on the challenging mixed I(1)-, I(0)-error panels ($\lambda = 0.5$) with heterogeneous dependence of AR1 type, $\theta_i \sim iidU(0.2, 0.9)$, or negative MA1 type, $\psi_i \sim iidU(-0.9, -0.2)$; all other specifications are as in Table 3. The results in Appendix Figure A2 suggest that the error in null rejection probability of the bootstrap test is essentially insignificant for all these (N, T) combinations. For the test Γ, the size distortions worsen with T as expected.

Near unit roots and threshold unit root effects

So far we have used $\rho_i = 0$ or $\rho_i = 1$. The null rejection probability of the tests is now examined for various degrees of autocorrelation ($0 \leq \rho_i \leq 1$) in the regression errors. For near unit roots, the error u_{it} is observationally equivalent to an I(1) process in finite samples but, since $\rho_i < 1$, the first-order differencing of u_{it} will result in a non-invertible MA term that will make it hard for the sieve bootstrap to work. The plots in Figure 2 are for $N = 15$, $T = 300$, $\theta_i = 0$, $\psi_i = 0$ and $\beta_i = 1$.

[Figure 2 around here]

The performance of the test Γ worsens dramatically as ρ_i increases. In stark contrast, the bootstrap (PSB) test $\Gamma^*_{b_2}$ does a reasonably good job, especially for the exact I(1) case and when $\rho_i \leq 0.9$. However, the plots show that for $0.9 < \rho_i < 1$ the test $\Gamma^*_{b_2}$ is too conservative. This is because the

order of integration d_i is estimated using the ADF test whose power falls dramatically as the AR root approaches unity. The over-differencing of u_{it} makes it hard for the sieve bootstrap to work.

This pitfall of the PSB scheme can be mitigated by using, instead of the ADF test, a more powerful unit root test for which there are a number of good candidates (see Maddala and Kim, 1998). It turns out that applying the sieve approximation directly (DSB scheme) to the individual residual sequences $\{\hat{u}_{it}\}$ without differencing works quite well. Figure 2 shows that the DSB approach ($\tilde{\Gamma}^*_{b_2}$) does a reasonably good job in correcting the size distortions of conventional t-tests for near I(1)-error panels. Interestingly, the bootstrap test based on the LSDV estimator (that exploits the within variation in the data) appears superior to that based on the POLS estimator (that gives equal weight to the within and between variation) for $\rho_i \geq 0.97$. The foregoing analysis thus suggests that the DSB approach is more reliable than the PSB approach.

We now revisit the notion that (behavioral) threshold effects can make regression errors appear nonstationary in finite samples. Figure 3 reports results for homogeneous cointegrating regressions ($\beta_i = 1$) with threshold dynamics in the error term according to the equation (10) so that ρ_i switches between 0.4 and 1 over time. The reported results pertain to the DSB approach. For the POLS estimator, we also deploy a slightly modified bootstrap (denoted $\tilde{\Gamma}^{*MU}_{b_2}$) where the sieve approximation is based on the median-unbiased correction of Andrews and Chen (1994).

[Figure 3 around here]

As the width of the band-of-inaction increases with c, the asymptotic test Γ becomes dramatically oversized. The bootstrap test $\tilde{\Gamma}^*_{b_2}$ eliminates the size distortions. For large $c \geq 7$, when the threshold AR1 series u_{it} becomes observationally equivalent to a linear AR1 series with $\rho_i \geq 0.98$, finite-sample bias corrections notably improve the performance of the POLS-based bootstrap test.

Cross-section dependence

The Monte Carlo design includes now panels with cross-correlated errors. Two distinct setups are considered. One is an unobserved-factor residual model (DGP1) which has been widely used in recent years (see inter alios Pesaran, 2005; Coakley et al., 2004a, 2005). The regression disturbances contain an I(1) common effect, $f_t = f_{t-1} + \varepsilon_{f_t}$, $\varepsilon_{f_t} \sim iidN(0,1)$, and an individual-specific

(idiosyncratic) error ν_{it} with stationary AR (or MA) dependence as in (9). In particular, we adopt

$$u_{it} = \gamma_i f_t + \nu_{it}, \tag{11}$$

$$\nu_{it} = \theta_i \nu_{it} + \varepsilon_{it}, \tag{12}$$

and the innovations ε_{it} follow a multivariate normal distribution with mean zero and covariance matrix $E(\varepsilon_t \varepsilon_t') = I_N$ where I_N is the identity matrix. We set $\gamma_i = 1$ and $\gamma_i \sim iidU(0.5, 1.5)$ to allow, respectively, for homogeneous and heterogeneous factor loadings. The latter amounts to heterogeneous cross-section correlations. In the second setup (DGP2), the errors are generated according to $u_{it} = u_{i,t-1} + \nu_{it}$ with $\nu_{it} = \theta_i \nu_{it} + \varepsilon_{it}$ and a non-diagonal covariance matrix

$$E(\varepsilon_t \varepsilon_t') = \Omega_\varepsilon = \begin{bmatrix} 1 & \omega & ... & \omega \\ \omega & 1 & ... & \omega \\ . & . & . & . \\ \omega & \omega & ... & 1 \end{bmatrix},$$

as in O'Connell (1998), Coakley et al. (2004a) and Chang (2004). Two levels of homogeneous cross-correlation are used, $\omega = \{0.6, 0.9\}$. For ν_{it} we consider $\theta_i = 0, \theta_i \sim iidU(0.2, 0.5)$, and $\theta_i \sim iidU(0.7, 0.9)$. All other specifications are as in the experiments of Table 1. The simulations are for $N, T = \{(15, 300), (20, 40)\}$ which typify monthly and annual macroeconomic panels, respectively.

In order to control for the cross-section dependence, one can apply LSDV (or POLS) to an augmented regression where the cross-sectional averages $\bar{y}_t$ and $\bar{x}_t$ are included as additional regressors to proxy the unobserved common effect — this is the factor-augmented approach, also known as the common correlated effects (CCE) estimator, proposed by Pesaran (2005) and further analysed by Kapetanios et al.(2006). The rationale is as follows. Let $y_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + u_{it}$, $u_{it} = \boldsymbol{\phi}_i' \mathbf{f}_t + \nu_{it}$ where the idiosyncratic innovations ν_{it} are I(0) but possibly autocorrelated, and the unobserved common factors $\mathbf{f}_t$ are I(d), $d \in \{0, 1\}$. Averaging over units gives $\bar{y}_t - \boldsymbol{\beta}' \bar{\mathbf{x}}_t = \bar{\boldsymbol{\phi}}_i' \mathbf{f}_t + \bar{\nu}_t$, and the observables $\bar{y}_t$ and $\bar{\mathbf{x}}_t$ together are shown to form a sufficient basis for the consistent (large N) estimation of $\mathbf{f}_t$. Accordingly, we construct two t-statistics, one is based on the baseline regression (1) and the other on the factor-augmented counterpart. For each of these two statistics, we conduct inferences based on the standard normal quantiles (Γ and Γ_F, respectively) and the DSB distribution ($\tilde{\Gamma}^*_{b_2,C}$ and $\tilde{\Gamma}^*_{b_2,F}$). The bootstrap approach $\tilde{\Gamma}^*_{b_2,C}$ is as explained in Section 3.1, with a slight modification so that the cross-correlation structure is preserved. We now resample rows from the centered and scaled $T \times N$ residual matrix $\tilde{\varepsilon}_{it}$, namely, each draw is a $1 \times N$ vector $\varepsilon_t^* = (\varepsilon_{1t}^*, ..., \varepsilon_{Nt}^*)'$

as in Chang (2004) and Cerrato and Sarantis (2007). The factor-augmented bootstrap approach ($\Gamma^*_{b_2,F}$) does not require the latter because it deals with the cross-section dependence by estimating β through the factor-augmented regression. So it only departs from the approach described in Section 3.1 in that the bootstrap samples are now obtained through the 'augmented' resampling scheme $y^*_{it} = \hat{\mu}_i + \beta_0 x^*_{it} + \hat{\delta}_1 \bar{y}^*_t + \hat{\delta}_2 \bar{x}^*_t + v^*_{it}$ (instead of $y^*_{it} = \hat{\mu}_i + \beta_0 x_{it} + u^*_{it}$), where $\hat{\mu}_i, \hat{\delta}_1, \hat{\delta}_2$ are the constrained LSDV parameter estimates obtained from the observed data and $x^*_{it} \equiv x_{it}$, $\bar{y}^*_t \equiv y_{it}$ and $\bar{x}^*_t \equiv x_{it}$ are fixed across replications. In both cases, $\tilde{\Gamma}^*_{b_2,C}$ and $\tilde{\Gamma}^*_{b_2,F}$, the direct sieve bootstrap approximation (DSB) is based on the median-unbiased approach of Andrews and Chen (1994) which is particularly important for the small-T panel case.

Table 4 reports the rejection frequencies of the different t-tests, all of them based on the LSDV estimator $\hat{\beta}$ and HAC Newey-West covariance matrix.

[Table 4 around here]

The test Γ is seriously oversized both in DGP1 and DGP2, as expected. The test Γ_F works quite well in DGP1 (albeit with some small-sample size distortions) with two exceptions. One is when the idiosyncratic component ν_{it} is strongly autocorrelated. The other is when the factor loadings are heterogeneous in which case the LSDV residuals measure $(\gamma - \gamma_i) f_t + \nu_{it}$ which is I(1) and so the robust HAC standard errors underestimate the true residual autocorrelation. The latter problem can be easily avoided by utilizing the modified factor-augmented LSDV estimator proposed by Pesaran (2005) which allows for heterogeneity in the slope coefficients of $\bar{y}_t$ and $\bar{x}_t$ so that the residuals measure the idiosyncratic error ν_{it} as in the homogeneous loadings case. But the test Γ_F does not work well in DGP2 where the cross-section dependence does not stem from a common factor. Reassuringly, the two bootstrap tests essentially attain the correct nominal size.

Power analysis

So far the hypothesized β_0 has been chosen to match the true long run average effect $\beta \equiv E(\beta_i)$ in the Monte Carlo DGP. To construct power curves, the discrepancy $\beta_0 - \beta$ is allowed to vary in a range which is set relative to the noise-to-signal ratio in the DGP, namely, $\beta_0 - \beta = \pm\delta\sigma^2_\varepsilon/\sigma^2_\xi$, with δ varying between -0.3 and 0.3 at intervals of 0.02. For the test Γ, which suffers from large size distortions, we construct an unadjusted power curve (rejection frequency at the nominal 0.05 level) and an adjusted power curve (rejection rate at the 'true' 0.05 level). The 'true' 0.05 levels used

are the empirical critical values taken from the corresponding size experiment. The power of the bootstrap tests has not been size-adjusted because the previous analysis suggests that there is no need to do so and it would have doubled the already high computational costs of these experiments.

We focus on panels with a mixture of I(1) and I(0) errors ($\lambda = 0.5$), and further AR1 dependence $\theta_i = 0.5$; all other specifications are as in the size experiments reported in Table 3. Cross-section dependence is introduced as in DGP2 above with $\omega = 0.6$. The power curves are virtually symmetric around $\beta_0 - \beta = 0$ so Figure 4 plots the rejection frequencies for $\beta_0 - \beta > 0$ only.

[Figure 4 around here]

The DSB test ($\tilde{\Gamma}^*_{b_2}$) has essentially the same power as the size-adjusted asymptotic Γ test which is in line with theoretical results on the power of the bootstrap (see Horowitz, 2000). The curve labelled $\tilde{\Gamma}^*_{b_2,200}$ corresponds to a computationally cheaper test based on $B = 200$ (instead of 500) replications. The results indicate that there is a price to pay in terms of power by reducing the number of bootstrap samples. Unsurprisingly also, the unit root pretesting required in the PSB scheme ($\Gamma^*_{b_2}$) effects some power loss.

5 Concluding remarks

Recent theoretical studies have shown that panel estimators can provide consistent measures of a long run average effect in the presence of unit root disturbances. This is quite relevant in empirical applications because strongly autocorrelated disequilibrium errors can be observationally unit root persistent in finite samples. Regression residuals with nonstationary properties can stem from periodically collapsing price bubbles or other behavioural effects, transaction costs, lumpy costs of adjustment or changes in government policy inter alios

This paper aims to fill two gaps in the literature. First, a bootstrap framework is provided to facilitate inference in nonstationary panel regressions under weak assumptions about the disturbances. This is important since, for instance, macroeconomic panel regressions with a mix of observationally I(1) and I(0) errors are very common in practice and the non-standard statistical theory depends on a nuisance parameter, the fraction of I(1) error processes. It is shown that asymptotic t-tests for the long run average parameter in homogeneous I(0)-error panels yield rejection rates of about 75% at a nominal level of 5% when just one individual error term is I(1) and

the remaining errors are white noise. To circumvent these problems, we propose a sieve bootstrap method and consider two residual resampling schemes. One is a unit-root pretesting approach that constructs pseudo-innovations with the I(1) property by construction. The other approach applies the sieve approximation directly to the residuals.

Secondly, an extensive Monte Carlo analysis is provided. Panel data generating processes with I(0) errors, I(1) errors or a mix of both are used. In the context of gaussian errors, it is shown that the finite sample distribution of the LSDV estimator is essentially normal in the former two cases but not in the latter. To accommodate realistic settings in our experiments, we also include asymmetric and highly leptokurtic error distributions, near-I(1) innovations generated from both a linear AR mechanism and a threshold AR process, I(1) innovations with a negative MA component, and cross-section dependence. Our findings suggest that the sieve bootstrap method yields correctly sized t-tests under a wide range of scenarios. It turns out that bootstrapping an improperly studentized t-statistic gives more robust results than bootstrapping the long run coefficient of interest. The pretesting-based bootstrap shows some size distortions in the near-I(1) error case due to overdifferencing of the residuals and generally is inferior in terms of power.

The direct sieve bootstrap (together with median-unbiased corrections for the sieve approximation in small T panels) provides correctly sized t-tests for long run average effects in a variety of settings where the asymptotic tests are oversized. The power of the bootstrap test is reasonably good and comparable to that of the level-adjusted asymptotic test. Alternative bootstrap methods and refinements of the proposed approach warrant further research.

Acknowledgments

This is a revised version of Cass Business School FF-WP 26/05. I acknowledge the helpful comments of the editor, Erricos Kontoghiorghes, an anonymous associate editor and three anonymous referees for helpful suggestions. I am also grateful to Kit Baum, Jerry Coakley, Ron Smith, Zacharias Psaradakis and participants at the 9th Annual Meeting of the *Society for Computational Economics* in the University of Washington, Seattle, for their comments.

References

Andrews, D.W.K., Chen, H.-Y., 1994. Approximately median-unbiased estimation of autoregressive models, J Bus. Econom. Statist.12, 187-204.

Balke, N.S., Fomby, T.B., 1997. Threshold cointegration. Internat. Econ. Rev. 38, 627-643.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer-Verlag.

Bühlmann, P., 1997. Sieve bootstrap for time series. Bernouilli 3, 123-148.

Cerrato, M., Sarantis, N., 2007. A bootstrap panel unit root test under cross-sectional dependence with an application to PPP. Comput. Statist. Data Anal. 51, 4028-4037.

Chang, Y., 2004. Bootstrap unit root tests in panels with cross-sectional dependency. J. Econometrics 120, 263-293.

Chang, Y., Park, Y., Song, K., 2006. Bootstrapping cointegrating regressions. J. Econometrics 133, 703-739.

Chortareas, G., Kapetanios, G., Uctum, M., 2003. An investigation of current account solvency in Latin America using nonlinear stationarity tests. Studies Nonl. Dyn. Econometrics 8.

Coakley, J., Fuertes, A.M., 2001. Nonparametric cointegration analysis of real exchange rates. Applied Financial Econom. 11, 1-8.

Coakley, J., Flood, R., Fuertes, A.M., Taylor, M.P., 2004a. Long run purchasing power parity and general relativity, J. Internat. Money Finance, 24, 293-316.

Coakley, J., Fuertes, A.M., Spagnolo, F., 2004b. Is the Feldstein-Horioka puzzle history? Manchester School 72, 569-590.

Coakley, J., Fuertes, A.M., Smith, R.P., 2005. Unobserved heterogeneity in panel time series. Comput. Statist. Data Anal. 50, 2361-2380.

Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman and Hall/CRC, New York.

Fuertes, A.-M, Kalotychou, E., 2006. Early warning systems for sovereign debt crises: the role of heterogeneity. Comput. Statist. Data Anal. 51, 1420-1441.

Giordano, F., La Rocca, M., Perna, C., 2007. Forecasting Nonlinear Time Series with Neural Network Sieve Bootstrap. Comput. Statist. Data Anal. 51, 3871-3884.

Hansen, B., 2000. Testing for structural change in conditional models. J. Econometrics 97, 93-115.

Hertbertsson, T., Zoega, G., 2000. Trade surpluses and life-cycle saving behaviour. Econ. Letters 65, 227-237.

Horowitz, J., 2000. The bootstrap, In: Heckman, J.J. and Leamer, E.E. (Eds.), Handbook of Econometrics, Vol. V. Amsterdam: North Holland.

Inoue, A., Kilian, L., 2002a. Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models. Internat. Economic Rev. 43, 309-332.

Inoue, A., Kilian, L., 2002b. Bootstrapping autoregressive processes with possible unit roots. Econometrica 70, 377-391.

Kao, C., 1999. Spurious regression and residual-based tests for cointegration in panel data. J. Econometrics 90, 1-44.

Kilian, L., 1998. Accounting for lag-order uncertainty in autoregressions: the endogenous lag-order bootstrap algorithm. J. Time Series Anal. 19, 531-548.

Li, H., Maddala, G.S., 1997. Bootstrapping cointegrating regressions, J. Econometrics 80, 297-318.

Maddala, G.S., Kim, I.M., 1998. Unit roots, cointegration and structural change. Cambridge University Press.

MacKinnon, J. G., 1996. Numerical distribution functions for unit root and cointegration tests. J. Applied Econometrics 11, 601-618.

Nankervis, J. C., Savin, N. E., 1996. The level and power of the bootstrap t test in the AR(1) model with trend. J. Business Econ. Stat. 14, 161-168.

Pesaran, H., 2005. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967-1012.

Pesaran, M.H., Smith, R.P., 1995. Estimating long run relationships from dynamic heterogeneous panels. J. Econometrics 68, 79-113.

Phillips, P.C.B., Moon, H.R., 1999. Linear regression theory for nonstationary panel data. Econometrica 67, 1057-1111.

Psaradakis, H., 2001. Bootstrap Tests for an Autoregressive Unit root in the presence of weakly dependent errors. J. Time Series Anal. 22, 577-594.

Psaradakis, H., 2003. A sieve bootstrap test for stationarity. Stat. Prob. Letters 62, 263-274.

Psaradakis, H., Sola, M., Spagnolo, F., 2004. On Markov Error-Correction Models. J. Applied Econometrics 19, 69-88.

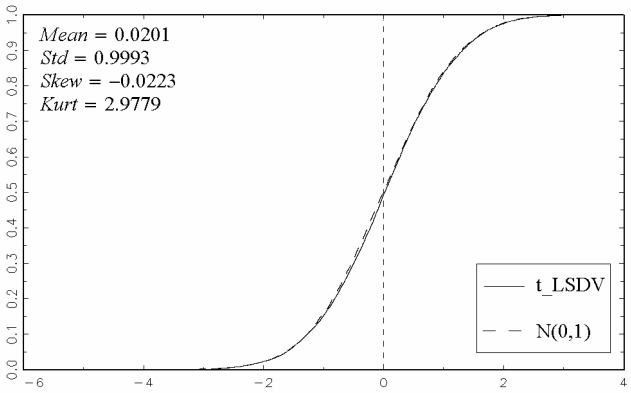
Stine, R. A., 1987. Estimating properties of autoregressive forecasts. J. Amer. Stat. Assoc. 82, 1072-1078.

Taylor, A.M., 1998. Saving, investment and international capital mobility in the twentieth century. J. Develop. Economics 57, 147-184.

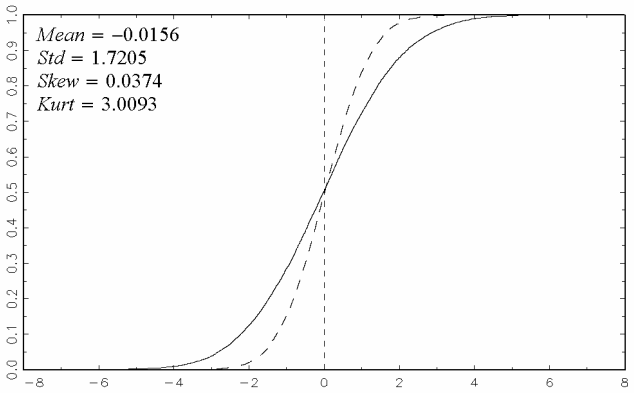
Taylor, A.M., 2001. Potential pitfalls for the purchasing power parity puzzle? sampling and specification biases in mean-reversion tests of the law of one price. Econometrica 69, 473-498.

Temple, J., 1999. The new growth evidence. J. Economic Literature 37, 112-156.

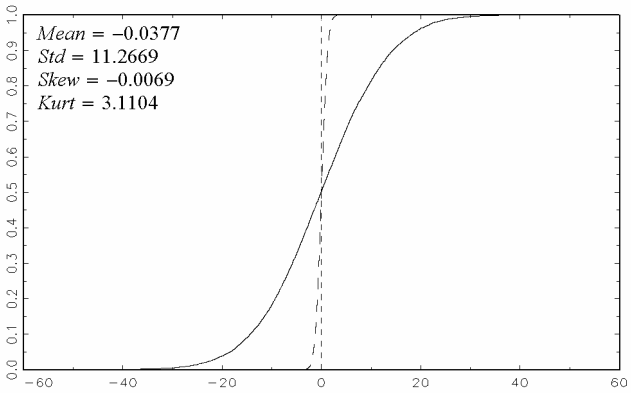
West, K. D., 1988. Asymptotic normality when regressors have a unit root. Econometrica 56, 1397-1418.



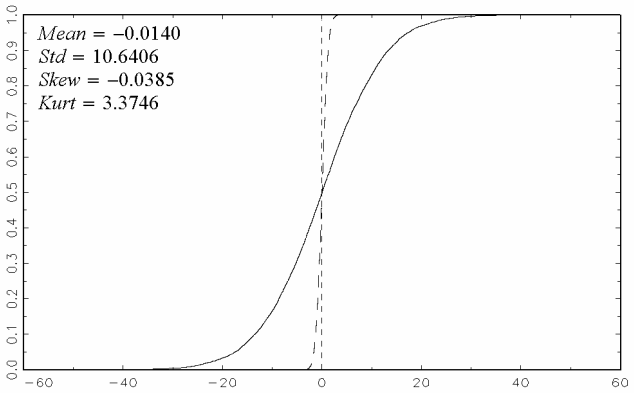
A) DGP with I(0)-errors, $\theta_i = 0$



B) DGP with I(0)-errors, $\theta_i = 0.9$



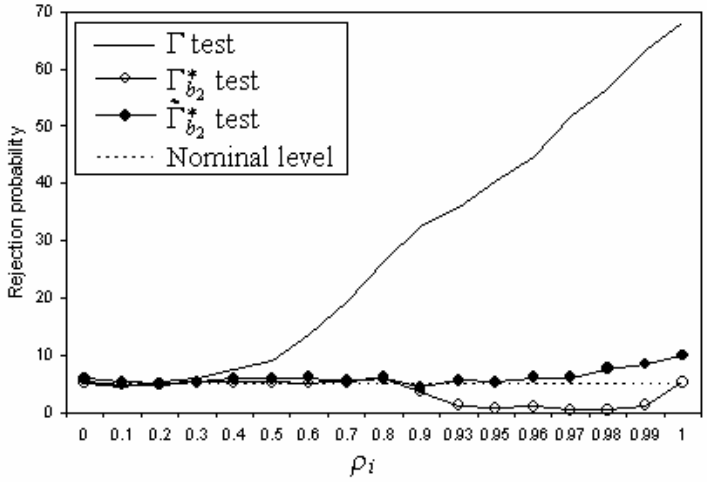
C) DGP with I(1)-errors, $\theta_i = 0.9$



D) DGP with mixed I(0)-, I(1)-errors, $\theta_i = 0.9$

Fig. 1. Empirical distribution function of *t*-statistics (LSDV estimator)

A) POLS estimator



B) LSDV estimator

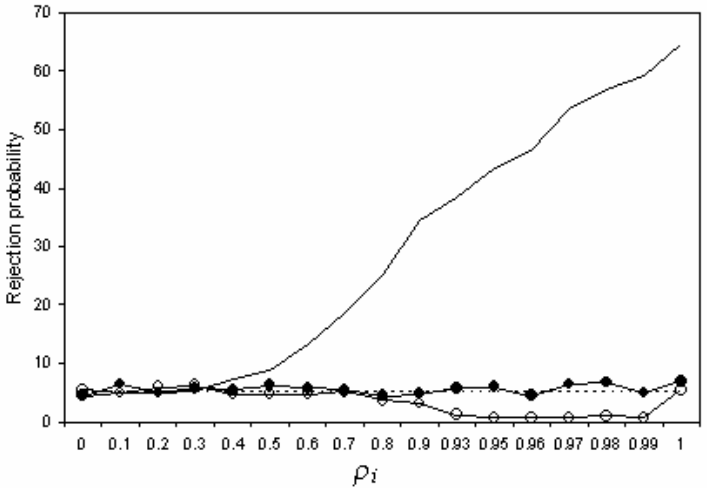
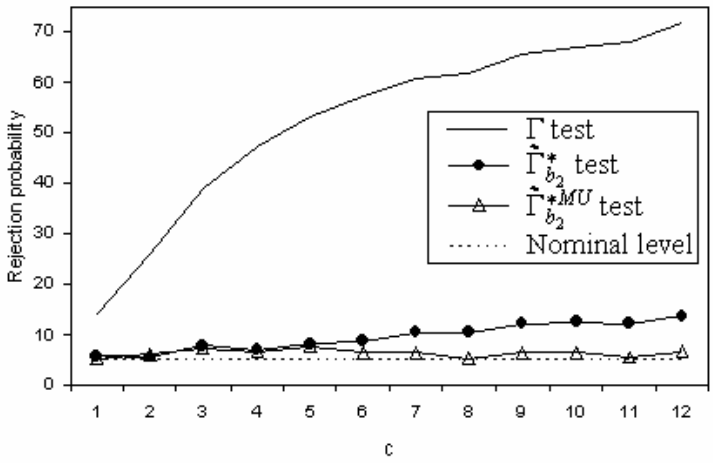


Fig. 2. Empirical size for various degrees of error persistence

A) POLS estimator



B) LSDV estimator

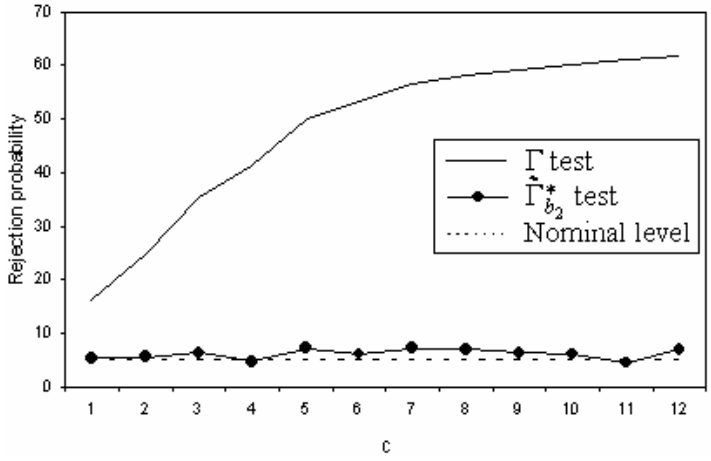


Fig. 3. Empirical size and threshold cointegration

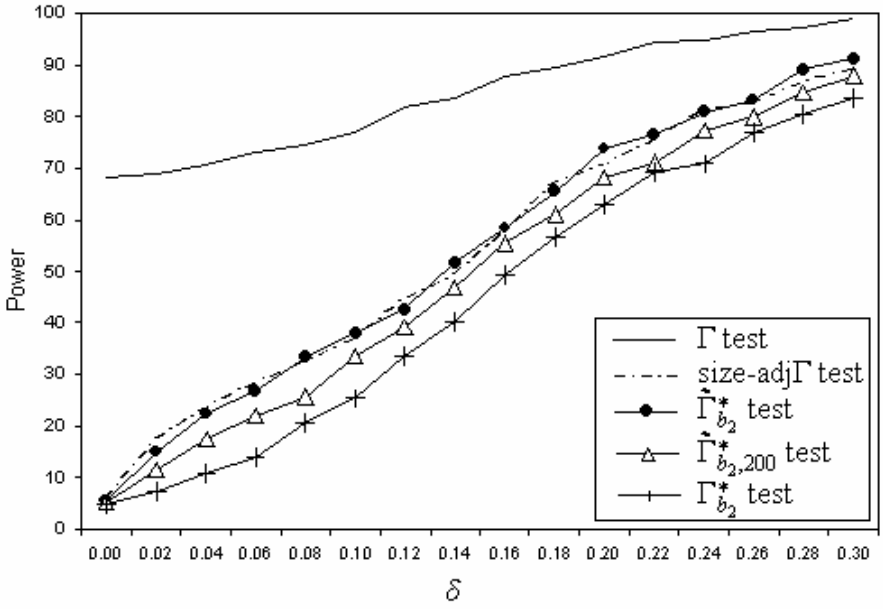


Fig. 4. Power curves. 5% level tests. Γ and size-adj Γ denote, respectively, the unadjusted and level adjusted asymptotic *t*-tests. $\tilde{\Gamma}^*_{b_2}$ and $\tilde{\Gamma}^*_{b_2,200}$ are DSB *t*-tests based on B=500 and B=200 bootstrap repetitions, respectively. $\Gamma^*_{b_2}$ is the PSB test based on B=500 bootstrap repetitions.

Table 1
Empirical size: homogeneous-slope panel, LSDV estimator

(AR)	(MA)	I(0) errors: $\rho_i = 0$					I(1) errors: $\rho_i = 1$				
		Γ	$\Gamma^*_{b_1}$		$\Gamma^*_{b_2}$		Γ	$\Gamma^*_{b_1}$		$\Gamma^*_{b_2}$	
θ_i	ψ_i		AIC	SBC	AIC	SBC		AIC	SBC	AIC	SBC
0.0	—	4.6	4.6	4.9	4.9	4.6	66.3	6.1	5.1	5.6	5.4
			5.6	*5.5*	*5.4*	*5.9*		*5.0*	*7.0*	*5.5*	*6.9*
0.5	—	8.9	5.6	6.4	5.7	5.9	68.6	5.8	5.0	6.2	5.0
			5.0	*5.1*	*4.8*	*5.1*		*5.4*	*5.2*	*5.4*	*4.9*
0.9	—	34.9	4.2	4.7	4.1	5.2	70.5	6.2	5.1	6.4	5.9
			3.2	*3.2*	*3.5*	*3.0*		*5.0*	*5.3*	*4.4*	*5.6*
—	0.5	5.8	5.1	6.4	5.2	6.9	69.1	4.7	5.3	5.4	5.4
			5.1	*5.9*	*4.6*	*6.2*		*4.8*	*4.3*	*5.2*	*4.4*
—	0.9	5.7	4.9	7.0	5.4	7.0	69.9	4.8	5.4	4.9	5.4
			5.9	*4.1*	*5.5*	*4.2*		*4.9*	*5.8*	*5.2*	*5.3*
U(0.3,0.5)	—	7.5	4.6	6.3	4.1	5.8	65.6	5.5	5.0	5.1	4.8
			4.7	*5.1*	*3.8*	*5.7*		*4.7*	5.6	5.6	6.1
U(0.2,0.9)	—	16.4	6.6	9.0	6.5	8.7	67.9	4.6	4.3	4.6	5.0
			4.7	*5.1*	*4.9*	*5.0*		*5.8*	*4.5*	*5.7*	*4.8*
—	U(0.3,0.5)	5.9	5.3	6.0	5.6	5.7	69.9	4.1	6.0	4.5	5.7
			6.1	*6.6*	*5.9*	*6.3*		*6.1*	*5.8*	*6.4*	*6.2*
—	U(0.2,0.9)	6.2	4.9	6.4	5.3	6.6	70.1	3.8	4.8	3.8	4.5
			5.6	*6.2*	*5.7*	*6.2*		*5.9*	*4.2*	*6.0*	*4.9*

5% level tests. Γ is based on N(0,1) quantiles. $\Gamma^*_{b_1}$ ($\Gamma^*_{b_2}$) is the gaussian (nonparametric) bootstrap test. d_i is estimated from $\hat{u}_{it}$ using the ADF test. The first and second row entries pertain to the sampling schemes where eq. (4) is estimated for $\hat{\nu}_{it} \equiv \Delta^{\hat{d}_i}\hat{u}_{it}$ by pooled OLS and individual OLS, resp. The sieve order is selected using AIC or SBC with $\tilde{p}_T = \lfloor 4\log_{10} T \rfloor$. N =15, T =300.

Table 2
Empirical size: non Gaussian I(1)-error panel

(AR) θ_i	(MA) ψ_i	pdf	POLS Γ	POLS $\Gamma^*_{b_1}$ AIC	POLS $\Gamma^*_{b_1}$ SBC	POLS $\Gamma^*_{b_2}$ AIC	POLS $\Gamma^*_{b_2}$ SBC	LSDV Γ	LSDV $\Gamma^*_{b_1}$ AIC	LSDV $\Gamma^*_{b_1}$ SBC	LSDV $\Gamma^*_{b_2}$ AIC	LSDV $\Gamma^*_{b_2}$ SBC
0.5	—	*D1*	69.8	6.4	5.3	6.6	5.3	68.5	5.6	5.5	5.5	5.1
		D2	70.1	5.2	6.1	5.3	6.2	68.2	6.1	4.6	5.8	4.6
		D3	68.9	6.0	5.1	5.6	5.6	68.6	6.1	5.2	6.1	5.2
0.9	—	*D1*	70.7	6.8	5.9	7.4	5.6	69.1	6.3	5.9	6.3	6.0
		D2	67.1	5.4	6.8	5.8	6.5	67.3	6.1	4.3	5.9	4.0
		D3	71.3	5.3	4.2	5.5	4.6	70.8	6.1	6.1	6.4	6.3
—	0.5	*D1*	69.9	5.3	6.1	6.1	6.7	66.8	5.1	5.6	5.3	6.1
		D2	67.9	7.5	4.3	7.3	4.4	64.9	5.3	4.8	5.8	4.7
		D3	69.6	5.9	7.3	6.1	7.3	66.1	6.8	6.3	6.8	6.2
—	0.9	*D1*	72.8	5.9	5.4	6.2	5.4	70.7	5.3	4.6	4.9	4.7
		D2	69.4	5.8	6.6	5.2	6.7	66.3	5.6	5.1	4.8	5.1
		D3	69.6	6.1	5.5	6.4	5.7	68.6	5.4	4.7	5.4	4.6
U(0.3,0.5)	—	*D1*	71.6	5.7	5.0	5.9	5.7	70.0	6.1	5.4	6.1	5.0
		D2	69.9	5.6	6.2	5.7	6.0	68.3	4.2	4.9	4.7	4.4
		D3	71.8	5.8	4.8	6.6	4.5	71.2	4.9	6.1	4.7	6.0
U(0.2,0.9)	—	*D1*	74.8	5.5	6.7	5.4	6.5	69.7	5.0	5.4	5.3	5.1
		D2	75.1	4.6	5.5	4.8	5.1	71.5	5.3	4.6	5.1	5.4
		D3	74.3	5.4	6.1	5.8	6.3	71.1	4.6	4.6	5.1	4.4
—	U(0.3,0.5)	*D1*	70.8	4.3	5.2	4.8	5.2	68.1	5.2	5.8	5.4	6.1
		D2	70.1	5.5	4.7	6.0	5.4	69.7	3.0	5.7	3.2	5.9
		D3	69.8	4.7	5.2	4.1	5.6	68.2	6.4	6.3	5.9	5.8
—	U(0.2,0.9)	*D1*	69.6	4.7	4.7	4.3	4.8	67.9	5.2	6.1	6.0	6.0
		D2	70.2	5.1	4.2	5.7	4.3	68.8	6.5	5.3	6.9	5.0
		D3	71.7	5.5	6.0	5.5	5.9	67.3	5.3	4.6	5.8	5.0

For $\Gamma^*_{b_1}$ and $\Gamma^*_{b_2}$ eq. (4) is fitted by single-equation OLS. *D1*, *D2* and *D3* are a t_5, a shifted χ^2_5 and a normal mixture 0.8N(0,1)&N(0,16), respectively. $\beta_i = 1$ for all i. N =15 and T =300.

Table 3
Empirical size: mixed I(1)-, I(0)-error panel

			POLS					LSDV				
	(AR)	(MA)	Γ	$\Gamma^*_{b_1}$		$\Gamma^*_{b_2}$		Γ	$\Gamma^*_{b_1}$		$\Gamma^*_{b_2}$	
$\lambda_{I(0)}$	θ_i	ψ_i		AIC	SBC	AIC	SBC		AIC	SBC	AIC	SBC
0.05	0.0	—	68.4	7.2	5.8	6.7	5.9	66.7	4.8	4.2	4.5	4.7
	0.5	—	72.6	5.4	5.8	5.2	6.0	70.0	4.7	5.5	5.0	5.8
	0.9	—	70.7	5.2	6.7	6.0	6.5	68.6	5.4	5.1	5.2	5.9
	—	0.5	69.9	5.0	6.4	4.7	6.3	68.2	6.1	5.5	6.4	5.2
0.20	0.0	—	72.9	7.1	6.2	6.9	6.1	69.1	4.7	5.4	5.8	5.8
	0.5	—	71.8	6.0	5.0	6.1	4.8	69.0	6.1	5.2	6.3	5.0
	0.9	—	75.2	6.1	6.2	6.1	6.5	71.2	6.2	5.4	5.6	6.1
	—	0.5	73.5	4.9	5.0	4.3	5.2	70.6	5.7	6.2	5.9	6.3
0.50	0.0	—	72.1	5.7	6.7	5.8	7.0	69.3	5.3	4.2	5.6	4.5
	0.5	—	76.3	6.2	4.7	5.9	4.7	72.7	4.8	5.7	4.8	5.8
	0.9	—	73.7	6.3	6.3	6.5	5.9	71.0	6.8	7.0	6.7	7.0
	—	0.5	69.1	6.2	6.2	6.5	6.5	66.2	4.9	6.1	4.6	6.2
0.80	0.0	—	78.6	7.3	7.5	6.8	8.2	71.6	4.8	6.5	5.2	5.8
	0.5	—	73.5	7.4	6.1	7.0	6.4	69.9	5.4	5.5	5.3	5.8
	0.9	—	73.4	8.5	7.1	8.2	7.4	71.2	6.4	6.0	5.9	6.0
	—	0.5	76.7	6.3	6.6	6.6	6.9	72.1	5.6	4.9	5.6	5.0
0.95	0.0	—	75.4	7.7	8.0	8.5	8.6	69.0	6.1	5.3	6.2	6.2
	0.5	—	73.7	9.3	7.6	9.8	7.0	67.2	6.9	6.5	6.8	7.1
	0.9	—	80.1	5.5	6.2	5.5	6.1	75.2	5.1	5.5	5.1	5.3
	—	0.5	78.8	4.0	7.1	4.1	7.4	71.5	7.6	5.0	7.5	5.2

$\lambda_{I(0)}$ is the fraction of I(0)-error equations, $\rho_i = 0$ for i=1,...,$\lfloor N\lambda_{I(0)} \rfloor$ and $\rho_i = 1$ elsewhere. $\lfloor \cdot \rfloor$ denotes the closest integer. For $\Gamma^*_{b_1}$ and $\Gamma^*_{b_2}$ eq. (4) is estimated individually by OLS for each i=1,2,...,N. The sieve order is chosen using AIC or SBC with $\tilde{p}_T = \lfloor 4 \log_{10} T \rfloor$.

Table 4
Empirical size: non-stationary error panel with cross-section dependence

A. Unobserved factor structure (DGP1)

factor loadings	(AR)	$N=15, T=300$				$N=20, T=40$			
γ_i	θ_i	Γ	Γ_F	$\Gamma^*_{b_2,C}$	$\Gamma^*_{b_2,F}$	Γ	Γ_F	$\Gamma^*_{b_2,C}$	$\Gamma^*_{b_2,F}$
1	0	57.6	6.0	4.6	5.2	22.7	8.6	6.1	5.4
	U(0.3,0.5)	58.2	9.4	4.8	4.5	23.4	10.6	4.6	5.0
	U(0.7,0.9)	54.2	23.6	4.5	5.0	23.8	21.2	4.0	4.9
U(0.5,1.5)	0	54.9	56.9	4.0	4.2	23.3	21.7	6.4	4.4
	U(0.3,0.5)	54.3	53.9	5.0	6.1	21.7	18.7	5.1	5.3
	U(0.7,0.9)	55.2	50.2	6.5	4.8	24.7	25.1	4.7	6.2

B. Non-spherical idiosyncratic disturbances (DGP2)

pairwise correlation	(AR)	$N=15, T=300$				$N=20, T=40$			
ω	θ_i	Γ	Γ_F	$\Gamma^*_{b_2,C}$	$\Gamma^*_{b_2,F}$	Γ	Γ_F	$\Gamma^*_{b_2,C}$	$\Gamma^*_{b_2,F}$
0.6	0	59.2	66.4	5.4	6.1	26.7	32.7	6.7	5.4
	U(0.3,0.5)	58.2	67.3	6.4	5.0	28.6	34.4	5.3	6.9
	U(0.7,0.9)	60.2	66.2	4.9	4.8	32.9	36.2	6.2	5.0
0.9	0	64.8	64.9	6.3	5.7	31.8	32.5	4.2	4.9
	U(0.3,0.5)	64.1	66.8	5.1	6.2	31.4	33.8	3.9	5.9
	U(0.7,0.9)	65.6	65.7	6.2	4.5	33.0	37.2	4.8	5.6

Results based on LSDV estimation of (1) or its factor-augmented version, denoted F. The test $\tilde{\Gamma}^*_{b_2,C}$ controls for cross-section dependence by resampling $N \times 1$ vectors $(\tilde{\varepsilon}^*_{1t}, \tilde{\varepsilon}^*_{2t}, ..., \tilde{\varepsilon}^*_{Nt})'$ from the centered and scaled residual matrix $\tilde{\varepsilon}$. $\Gamma^*_{b_2,F}$ is the factor-augmented bootstrap test.