UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

A homogeneous approach to testing for Granger non-causality in heterogeneous panels

Juodis, A.; Karavias, Y.; Sarafidis, V.

DOI
10.1007/s00181-020-01970-9

Publication date
2021

Document Version
Final published version

Published in
Empirical Economics

License
CC BY

Link to publication

Citation for published version (APA):
Juodis, A., Karavias, Y., & Sarafidis, V. (2021). A homogeneous approach to testing for Granger non-causality in heterogeneous panels. *Empirical Economics*, *60*(1), 93–112. https://doi.org/10.1007/s00181-020-01970-9

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:11 Oct 2026

Empirical Economics (2021) 60:93–112
https://doi.org/10.1007/s00181-020-01970-9

A homogeneous approach to testing for Granger non-causality in heterogeneous panels

Artūras Juodis[1] · **Yiannis Karavias**[2] · **Vasilis Sarafidis**[3,4]

Received: 30 September 2020 / Accepted: 22 October 2020 / Published online: 23 November 2020
© The Author(s) 2020

Abstract
This paper develops a new method for testing for Granger non-causality in panel data models with large cross-sectional (N) and time series (T) dimensions. The method is valid in models with homogeneous or heterogeneous coefficients. The novelty of the proposed approach lies in the fact that under the null hypothesis, the Granger-causation parameters are all equal to zero, and thus they are homogeneous. Therefore, we put forward a pooled least-squares (fixed effects type) estimator for these parameters only. Pooling over cross sections guarantees that the estimator has a $\sqrt{NT}$ convergence rate. In order to account for the well-known "Nickell bias", the approach makes use of the well-known Split Panel Jackknife method. Subsequently, a Wald test is proposed, which is based on the bias-corrected estimator. Finite-sample evidence shows that the resulting approach performs well in a variety of settings and outperforms existing procedures. Using a panel data set of 350 U.S. banks observed during 56 quarters, we test for Granger non-causality between banks' profitability and cost efficiency.

Keywords Panel data · Granger causality · VAR · "Nickell bias" · Bias correction · Fixed effects

JEL Classification C12 · C13 · C23 · C33

1 Introduction

Predictive causality and feedback between variables is one of the main subjects of applied time series analysis. Granger (1969) provided a definition that allows formal statistical testing of the hypothesis that one variable is not temporally related to (or does not "Granger-cause") another one. Besides time series models, this hypothe-

✉ Artūras Juodis
A.Juodis@uva.nl

Extended author information available on the last page of the article

sis is also important in panel data analysis when examining relationships between macroeconomic or microeconomic variables.

The seminal paper of Holtz-Eakin et al. (1988) provided one of the early contributions to the panel data literature on Granger non-causality testing. Using Anderson and Hsiao (1982) type moment conditions, the authors put forward a Generalised Method of Moments (GMM) testing framework for short T panels with homogeneous coefficients. Unfortunately, this approach is less appealing when T is sizeable. This is due to the well-known problem of using too many moment conditions, which often renders the usual GMM-based inference highly inaccurate. While there exist alternative fixed T procedures that can be applicable to cases where T is large (e.g. those of Binder et al. 2005; Karavias and Tzavalis 2017; Juodis 2013; Arellano 2016; Juodis 2018), these methods are designed to estimate panels with homogeneous slope parameters only. Thus, when feedback based on past own values is heterogeneous (i.e. the autoregressive parameters vary across individuals), inferences may not be valid even asymptotically.

For the reasons above, one of the most popular approaches among practitioners has been the one proposed by Dumitrescu and Hurlin (2012), which can accommodate heterogeneous slopes under both null and alternative hypotheses. Their approach is reminiscent of the so-called IPS panel unit root test for heterogeneous panels proposed by Im et al. (2003) and involves averaging of individual Wald statistics. The resulting standardized Wald test statistic has asymptotic normal limit as $T \to \infty$ followed by $N \to \infty$. However, this approach does not account for "Nickell" bias, and therefore, it is theoretically justified only for sequences with $N/T^2 \to 0$, as it is the case with standard Mean-Group type approaches.[1]

The aim of this paper is to propose a new test for Granger non-causality that explicitly accounts for "Nickell" bias and is valid in both homogeneous and heterogeneous panels. The novelty of our approach comes from exploiting the fact that under the null hypothesis, while the individual effects and the autoregressive parameters may be heterogeneous across individuals, the Granger-causation parameters are all equal to zero and thus they are homogeneous. We therefore propose the use of a pooled estimator for these parameters only. Pooling over cross sections guarantees that the estimator has the faster $\sqrt{NT}$ convergence rate.

The pooled estimator suffers from the incidental parameters problem of Neyman and Scott (1948) due to the presence of the predetermined regressors, see, e.g. Nickell (1981) and Karavias and Tzavalis (2016). This result implies that standard tests for pooled estimators do not control size asymptotically, unless $N << T$. To overcome this problem, we use the idea of Split Panel Jackknife (SPJ) of Dhaene and Jochmans (2015) and construct an estimator that is free from the "Nickell bias". This type of bias correction works very well under circumstances that are empirically relevant: moderate time dimension, heterogeneous nuisance parameters, and high persistence, as argued by Dhaene and Jochmans (2015), Fernández-Val and Lee (2013) and Chambers (2013), respectively. Furthermore, Chudik et al. (2018) argue that SPJ procedures are suitable

[1] For panels with a fixed-T dimension, and under normality of the innovations, Dumitrescu and Hurlin (2012) propose centering their test statistic using moments of an appropriate F distribution rather than χ^2. However, the modified statistic is not standard normal for fixed-T (even under normality of the innovations) because the suggested approximation assumes that regressors are strictly exogenous.

so long as $N/T^3 \to 0$. Thus, we test the null hypothesis of Granger non-causality by using a Wald test based on our bias-corrected estimator.

A Monte Carlo study shows that the proposed method has good finite sample properties even in panels with a moderate time dimension. In contrast, the Wald statistic of Dumitrescu and Hurlin (2012) can suffer from substantial size distortions, especially when $T << N$. In terms of power, the proposed method appears to dominate the method of Dumitrescu and Hurlin (2012), especially so in panels with N and T both large.

Using a panel data set of 350 U.S. banks observed during the period 2006:Q1-2019:Q4, we test for Granger non-causality between banks' profitability and cost efficiency. The null hypothesis is rejected in all cases, except for large banks during a period spanning the financial crisis (2007–2009) and prior to the introduction of the Dodd–Frank Act in 2011. This outcome may be conducive of past moral hazard-type behaviour of large financial institutions.

The remainder of the present paper is organized as follows: Sect. 2 sets up the model and the hypothesis of interest. Section 3 outlines the SPJ estimator and the proposed test statistic. Section 4 studies the finite sample performance of the approach using Monte Carlo experiments. Section 5 presents the empirical illustration, and Sect. 6 concludes.

2 Testing framework

We consider a simple linear dynamic panel data model with a single covariate $x_{i,t}$:

$$y_{i,t} = \phi_{0,i} + \sum_{p=1}^{P} \phi_{p,i} y_{i,t-p} + \sum_{q=1}^{Q} \beta_{q,i} x_{i,t-q} + \varepsilon_{i,t}; \quad t = 1, \ldots, T, \tag{2.1}$$

for $i = 1, \ldots, N$, where $\phi_{0,i}$ captures the individual-specific fixed effects, $\varepsilon_{i,t}$ denotes the innovation for individual i at time t, $\phi_{p,i}$ denotes the heterogeneous autoregressive coefficients and $\beta_{q,i}$ denotes the heterogeneous feedback coefficients or Granger causation parameters.[2] Thus, we assume that $y_{i,t}$ follows an ARDL(P,Q) process; more generally, $y_{i,t}$ can be considered as one of the equations of a joint VAR model for $(y_{i,t}, x_{i,t})'$. Such bivariate system is studied for simplicity of presentation, as our results are straightforwardly extendable to multivariate systems.[3]

The null hypothesis that the time series $x_{i,t}$ does not Granger-cause (linearly) the time series $y_{i,t}$ can be formulated as a set of linear restrictions on the β's in Eq. (2.1):

$$H_0: \quad \beta_{q,i} = 0, \quad \text{for all } i \text{ and } q, \tag{2.2}$$

[2] Since the model above is observed over T time periods, it is implicitly assumed that $y_{i,-P+1}, y_{i,-P+2}, \ldots, y_{i,0}$ are observed, and so are $x_{i,-Q+1}, x_{i,-Q+2}, \ldots, x_{i,0}$.

[3] Also, to save space, we do not provide an exposition for how to test bi-directional causality, which can take place in a similar manner by expressing x as a function of own lags and lagged values of y.

against the alternative

$$H_1: \quad \beta_{q,i} \neq 0 \quad \text{for some } i \text{ and } q. \tag{2.3}$$

The model, null and alternative hypotheses presented here are as in Dumitrescu and Hurlin (2012). Similarly to the case of panel unit root testing, rejection of the null hypothesis should be interpreted as evidence of the existence of a large enough number of cross-sectional units i in which the null hypothesis is violated (see, e.g. Pesaran 2012).

3 Approach

Equation (2.1) can be re-written as follows:

$$y_{i,t} = z'_{i,t}\boldsymbol{\phi}_i + \boldsymbol{x}'_{i,t}\boldsymbol{\beta}_i + \varepsilon_{i,t}, \tag{3.1}$$

where $\boldsymbol{z}_{i,t} = (1, y_{i,t-1}, \ldots, y_{i,t-P})'$ and $\boldsymbol{x}_{i,t} = (x_{i,t-1}, \ldots, x_{i,t-Q})'$ are column vectors of order $1 + P$ and Q, respectively, while $\boldsymbol{\phi}_i = (\phi_{0,i}, \ldots, \phi_{P,i})'$ and $\boldsymbol{\beta}_i = (\beta_{1,i}, \ldots, \beta_{Q,i})'$ denote the corresponding parameter vectors.

Define $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,T})'$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})'$, both of which are column vectors of order T, and let $\boldsymbol{Z}_i = (\boldsymbol{z}_{i,1}, \ldots, \boldsymbol{z}_{i,T})'$ be a matrix of dimension $[T \times (1 + P)]$, and $\boldsymbol{X}_i = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,T})'$, a matrix of dimension $[T \times Q]$. Equation (3.1) can be expressed in vector form as

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\phi}_i + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i. \tag{3.2}$$

Observe that under the null hypothesis of Granger non-causality, the true coefficient vector of $\boldsymbol{X}_i$ equals zero. Thus, assuming homogeneity in $\boldsymbol{\beta}_i$, Eq. (3.2) becomes

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\phi}_i + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i. \tag{3.3}$$

In what follows, we shall use the above model specification to estimate the common parameters $\boldsymbol{\beta}$. In particular, we propose the following least-squares (fixed effects type) estimator of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{M}_{\boldsymbol{Z}_i}\boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{M}_{\boldsymbol{Z}_i}\boldsymbol{y}_i\right), \tag{3.4}$$

where $\boldsymbol{M}_{\boldsymbol{Z}_i}$ denotes a $[T \times T]$ matrix that projects on the orthogonal complement of $\boldsymbol{Z}_i$, i.e. $\boldsymbol{M}_{\boldsymbol{Z}_i} = \boldsymbol{I}_T - \boldsymbol{Z}_i\left(\boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1}\boldsymbol{Z}_i'$. The estimator in Eq. (3.4) generalizes the standard FE estimator, as the latter imposes that *all* slope coefficients are homogeneous, including the autoregressive parameters (see, e.g. Hahn and Kuersteiner 2002). Note that for this estimator to be well defined, a sufficient number of $\boldsymbol{M}_{\boldsymbol{Z}_i}$ matrices should be nonzero. As in that paper, we limit our attention to balanced panels, and so the necessary condition is $T > 1 + P$, which ensures that the coefficients $\boldsymbol{\phi}_i$ are estimable.

The model in (2.1) belongs to a class of panel data models with nonadditive unobserved heterogeneity studied in Fernández-Val and Lee (2013). In particular, under

Conditions 1–2 of that paper, which restrict $\boldsymbol{q}_{i,t} = (y_{i,t}, x_{i,t})'$ to be a strong mixing sequence, conditional on all time-invariant effects, with at least $4 + \delta$ moments (for some $\delta > 0$), the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is readily available. Note that the aforementioned restriction rules out non-stationary and local-to-unity dynamics in $\boldsymbol{y}_i$ and $\boldsymbol{X}_i$.

In order to facilitate further discussion, we shall adapt the conclusions of Theorem 1 in Fernández-Val and Lee (2013) to the present setup:

Theorem 3.1 *Under Conditions 1–2 (Fernández-Val and Lee 2013) and given* $N/T \to a^2 \in [0; \infty)$ *as* $N, T \to \infty$ *jointly:*

$$\sqrt{NT}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} \boldsymbol{J}^{-1} N\left(-a\boldsymbol{b}, \boldsymbol{V}\right). \tag{3.5}$$

The Hessian matrix $\boldsymbol{J}$ in our case is given by:

$$\boldsymbol{J} = \text{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_{\boldsymbol{Z}_i} \boldsymbol{X}_i, \tag{3.6}$$

while the exact form of $\boldsymbol{V}$ and $\boldsymbol{b}$ depends on the underlying assumptions of $\varepsilon_{i,t}$. For example, if $\varepsilon_{i,t}$ are independent and identically distributed (i.i.d.) over i and t, i.e. $\varepsilon_{i,t} \sim i.i.d.(0, \sigma^2)$, then

$$\boldsymbol{V} = \sigma^2 \boldsymbol{J}. \tag{3.7}$$

The vector $\boldsymbol{b}$ captures the incidental parameter bias of the common parameter estimator, which is induced by estimation of $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N$. We will not elaborate on the exact form of this matrix, as it is not needed for the purposes of this paper.[4]

Although $\hat{\boldsymbol{\beta}}$ is consistent, the asymptotic distribution of the estimator is not centered around zero under sequences where N and T grow at a similar rate. The presence of bias invalidates any asymptotic inference because the bias is of the same order as the variance (that is, unless $a = 0$). In particular, the use of $\hat{\boldsymbol{\beta}}$ for Granger non-causality testing of $H_0 : \boldsymbol{\beta}_0 = \boldsymbol{0}_Q$ will not lead to a test with correct asymptotic size. As a result, the Wald test statistic:

$$W = NT\hat{\boldsymbol{\beta}}'\left(\boldsymbol{J}^{-1}\boldsymbol{V}\boldsymbol{J}^{-1}\right)^{-1}\hat{\boldsymbol{\beta}}, \tag{3.8}$$

converges to a non-central $\chi^2(Q)$ distribution under the null hypothesis even if $\boldsymbol{J}$ and $\boldsymbol{V}$ are assumed to be known.

The above discussion implies that $\hat{\boldsymbol{\beta}}$ should not be used in the construction of the Wald test statistic (3.8). Instead, we suggest the use of the same test statistic, but based on an alternative estimator that is free from the asymptotic bias term $-a\boldsymbol{b}$. Below, we shall focus on a bias-corrected estimator constructed based on the Jackknife principle, using the Half Panel Jackknife (HPJ) procedure of Dhaene and Jochmans (2015).

[4] For more details on the exact form of all matrices in Theorem 3.1, the interested reader is referred to Fernández-Val and Lee (2013).

Given a balanced panel with an even number of time series observations, the HPJ estimator is defined as

$$\tilde{\boldsymbol{\beta}} \equiv 2\hat{\boldsymbol{\beta}} - \frac{1}{2}\left(\hat{\boldsymbol{\beta}}_{1/2} + \hat{\boldsymbol{\beta}}_{2/1}\right), \tag{3.9}$$

where $\hat{\boldsymbol{\beta}}_{1/2}$ and $\hat{\boldsymbol{\beta}}_{2/1}$ denote the FE estimators of $\boldsymbol{\beta}$ based on the first $T_1 = T/2$ observations, and the last $T_2 = T - T_1$ observations, respectively. The HPJ estimator can be decomposed into a sum of two terms:

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \left(\hat{\boldsymbol{\beta}} - \frac{1}{2}\left(\hat{\boldsymbol{\beta}}_{1/2} + \hat{\boldsymbol{\beta}}_{2/1}\right)\right) = \hat{\boldsymbol{\beta}} + T^{-1}\hat{\boldsymbol{b}}, \tag{3.10}$$

where the second component implicitly estimates the bias term in (3.5). The use of this estimator can be justified in our setting given that the bias of $\hat{\boldsymbol{\beta}}$ is of order (T^{-1}) and thus satisfies the expansion requirement of Dhaene and Jochmans (2015). Although there do exist alternative ways of splitting the panel to construct a bias-corrected estimator, as shown in Dhaene and Jochmans (2015), the HPJ estimator minimizes the higher order bias in the class of Split Panel Jackknife (SPJ), provided that the data are stationary. For this reason, we limit our attention to Eq. (3.9).

Corollary 3.1 *Under Conditions 1–2 of* Fernández-Val and Lee (2013) *and given* $N/T \to a^2 \in [0; \infty)$ *as* $N, T \to \infty$ *jointly:*

$$\hat{W}_{HPJ} = NT\tilde{\boldsymbol{\beta}}'\left(\hat{\boldsymbol{J}}^{-1}\hat{\boldsymbol{V}}\hat{\boldsymbol{J}}^{-1}\right)^{-1}\tilde{\boldsymbol{\beta}} \xrightarrow{d} \chi^2(Q), \tag{3.11}$$

where, assuming $\varepsilon_{i,t} \sim i.i.d.(0, \sigma^2)$,

$$\hat{\boldsymbol{J}} = \frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{M}_{\boldsymbol{Z}_i}\boldsymbol{X}_i$$

$$\hat{\boldsymbol{V}} = \hat{\sigma}^2\hat{\boldsymbol{J}}$$

$$\hat{\sigma}^2 = \frac{1}{N(T-1-P)-Q}\sum_{i=1}^{N}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}\right)'\boldsymbol{M}_{\boldsymbol{Z}_i}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}\right).$$

The proof of this corollary follows from the corresponding results in Fernández-Val and Lee (2013) and Dhaene and Jochmans (2015). The formula for $\hat{\boldsymbol{V}}$ can be easily modified to allow for heteroskedasticity in both cross-sectional and time-series dimensions, based, e.g. on the clustered-covariance matrix estimator of Arellano (1987). For instance, cross-sectional heteroskedasticity can be accommodated by setting

$$\hat{\boldsymbol{V}} = \frac{1}{N(T-1-P)-Q}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_{\boldsymbol{Z}_i}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}_i'\boldsymbol{M}_{\boldsymbol{Z}_i}\boldsymbol{X}_i, \tag{3.12}$$

where $\hat{\boldsymbol{\varepsilon}}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}$. Given the recent results in Chudik et al. (2018), we conjecture that for the HPJ approach to work it is only necessary to assume $N/T^3 \to 0$.

Remark 3.1 An alternative homogeneous estimator is available by taking into account the fact that under the null hypothesis, not only $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i but also $\beta_1 = \beta_2 = \ldots = \beta_Q = 0$. Therefore, letting $\boldsymbol{x}_{i,-1} = (x_{i,0}, \ldots, x_{i,T-1})'$, one can also consider the following restricted fixed effects type estimator:

$$\hat{\beta}_1 = \left(\sum_{i=1}^{N} \boldsymbol{x}'_{i,-1} \boldsymbol{M}_{\boldsymbol{Z}_i} \boldsymbol{x}_{i,-1} \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{x}'_{i,-1} \boldsymbol{M}_{\boldsymbol{Z}_i} \boldsymbol{y}_i \right). \tag{3.13}$$

This estimator is attractive because, under the null hypothesis, it does not require specifying a value for Q. However, the resulting Wald test statistic is expected to have lower power compared to that in Eq. (3.11).

Remark 3.2 Jackknife is by no means the only approach that corrects the incidental parameters bias of the FE estimator. Alternatively, one can consider an analytical bias-correction, as in Hahn and Kuersteiner (2002) and Fernández-Val and Lee (2013). However, the analytical approach has several practical limitations such as the need to specify a kernel function and the corresponding bandwidth. In this respect, the HPJ approach of Dhaene and Jochmans (2015) has some clear advantages.

4 Monte Carlo simulation

4.1 Design

To illustrate the performance of the new testing procedure, we adapt the Monte Carlo setup of Binder et al. (2005) and Juodis (2018). In particular, we assume that the bivariate vector $\boldsymbol{y}_{i,t} = (y_{i,t}, x_{i,t})'$ is subject to the following VAR(1) process:

$$\boldsymbol{y}_{i,t} = \boldsymbol{\Phi}_i \boldsymbol{y}_{i,t-1} + \boldsymbol{\varepsilon}_{i,t}; \quad \boldsymbol{\varepsilon}_{i,t} \sim N(\boldsymbol{0}_2, \boldsymbol{\Sigma}), \tag{4.1}$$

for all $i = 1, \ldots, N$, and $t = 1, \ldots, T$. The vector $\boldsymbol{y}_{i,t}$ is assumed to be initialized in a distant past, in particular we set $\boldsymbol{y}_{i,-50} = \boldsymbol{0}_2$ and discard the first 50 observations in estimation.

In order to simplify parametrization, our baseline setup specifies that some of the design matrices are common for all i. In particular, we adopt Design 2 of Juodis (2018) for the error variance matrix, setting

$$\boldsymbol{\Sigma} \equiv \begin{pmatrix} \sigma^2_{\varepsilon_y} & \sigma_{\varepsilon_{y,x}} \\ \sigma_{\varepsilon_{y,x}} & \sigma^2_{\varepsilon_x} \end{pmatrix} = \begin{pmatrix} 0.07 & 0.05 \\ 0.05 & 0.07 \end{pmatrix}. \tag{4.2}$$

Matrix $\boldsymbol{\Phi}_i$ is set equal to

$$\boldsymbol{\Phi}_i = \begin{pmatrix} \alpha_i & \beta_i \\ -0.5 & \rho \end{pmatrix}, \tag{4.3}$$

where in the homogeneous case we impose $\alpha_i = \alpha = 0.4$ while in the heterogeneous case $\alpha_i = \alpha + \xi_i^{(y)} = 0.4 + \xi_i^{(y)}$, $\xi_i^{(y)} \sim i.i.d.U\,[-.15, .15]$. ρ alternates such that

$\rho = \{0.4; 0.8\}$. This parameter controls the degree of persistence in $x_{i,t}$, which can be either moderate ($\rho = 0.4$) or high ($\rho = 0.8$).

The main parameter of interest is β_i. For $\beta_i = 0$, the $\boldsymbol{\Phi}_i$ matrix is lower triangular so that $x_{i,t}$ does not Granger-cause $y_{i,t}$. In this case, the empirical rejection rate corresponds to the size of the test. On the other hand, for $\beta_i \neq 0$, the empirical rejection rate reflects power. In order to cover a broad range of possible alternative hypotheses, we consider the following schemes:

1. (Homogeneous). $\beta_i = \beta$ for all i. $\beta = \{0.00; 0.02; 0.03; 0.05\}$.
2. (Heterogeneous). $\beta_i = \beta + \xi_i^{(x)}$, $\xi_i^{(x)} \sim i.i.d.U\,[-0.1; 0.1]$, where β is as in the homogeneous case.

The homogeneous design covers the classical pooled setup of Holtz-Eakin et al. (1988). On the other hand, heterogeneity introduced in the second design is qualitatively closer to Dumitrescu and Hurlin (2012). Note that in the heterogeneous case $\mathrm{E}[\beta_i] = \beta$.

Given that the procedure of Dumitrescu and Hurlin (2012) is primarily used in medium-size macro-panels, we focus on combinations of (N, T) that better reflect these applications. In particular, we limit our attention to the following 9 combinations:

$$N = \{50; 100; 200\}; \quad T = \{20; 50; 100\}. \tag{4.4}$$

We consider the following test statistics:

- "DHT"—the Dumitrescu and Hurlin (2012) Wald test statistic given by[5]

$$\widetilde{W}_{DH} = \sqrt{\frac{N}{2P}\frac{T-2P-5}{T-P-3}}\left(\left(\frac{T-2P-3}{T-2P-1}\right)\frac{1}{N}\sum_{i=1}^{N} W_i - P\right). \tag{4.5}$$

- "HPJ"—the proposed pooled Wald test statistic in Eq. (3.11), which is based on the HPJ bias-corrected estimator.

Inference is conducted at the 5% level of significance. The total number of Monte Carlo replications is set to 5, 000. Size-adjusted power is reported.

In an alternative setup, we also consider heteroskedastic innovations, where the top-diagonal entry of the variance–covariance matrix $\boldsymbol{\Sigma}$ in Eq. (4.2), $\sigma^2_{\varepsilon_y}$, is scaled by $\xi_i^{(\varepsilon)} \sim i.i.d.U\,[0, 2]$, such that $\mathrm{E}\left[\sigma^2_{\varepsilon_y,i}\right] = \sigma^2_{\varepsilon_y}\mathrm{E}\left[\xi_i^{(\varepsilon)}\right] = 0.07$.

4.2 Results

This section provides a brief summary of the simulation results, which are reported in Tables 3, 4, 5 and 6 in Appendix A. In specific,

- (*size*) when the degree of persistence in $x_{i,t}$ is moderate, such that $\rho = 0.4$, both HPJ and DHT tests perform similarly. In particular, empirical size is fairly close to

[5] The authors also propose an alternative Wald test statistic that is not centered. However, in the present setup we prefer using DHT because it provides better size control.

its nominal value in most circumstances, with some size distortions observed when $T << N$, especially for DHT. On the other hand, for $\rho = 0.8$, the performance of both tests deteriorates. This is particularly so for DHT, where in 8 out of 18 cases size exceeds 20%. In fact, for the case where $N = 200$ & $T = 20$ size is over 50%. On the other hand, HPJ appears to be more reliable and size remains below 15% under all circumstances.

- (*power*) for $\rho = 0.4$ HPJ dominates DHT almost uniformly in terms of power. Similar conclusions can be drawn for $\rho = 0.8$. Note that on average, for any fixed value of N, power increases with T at a higher rate for HPJ than DHT, which reflects the $\sqrt{NT}$ convergence rate of the bias-corrected least-squares estimator employed by the HPJ test.
- (*homogeneous vs heterogeneous models*) The performance of the tests in the heterogeneous model is similar to the homogeneous one in terms of both size and power.
- (*homoskedasticity vs heteroskedasticity*) The results are similar in terms of both size and power under homoskedasticity and heteroskedasticity. This implies that heteroskedasticity does not distort the performance of the tests, once appropriately accounted for.

In summary, the above results suggest that HPJ has good finite sample properties even in panels with a moderate time dimension. In contrast, DHT can suffer from substantial size distortions, especially when $T << N$. Moreover, in terms of power, HPJ dominates DHT, especially so in panels where N and T are both large.[6]

5 Illustration: Granger causality evidence on bank profitability and efficiency

We perform Granger non-causality tests in order to examine the sign and the type of temporal relation between banks profitability and cost efficiency. We employ panel data from a random sample of 350 U.S. banking institutions, each one observed over 56 time periods, namely 2006:Q1-2019:Q4. This data set has also been used by Cui et al. (2020), albeit in a different context related to the estimation of a spatial dynamic panel model with common factors. The data are publicly available, and they have been downloaded from the Federal Deposit Insurance Corporation (FDIC) website.[7]

5.1 Data and model specification

We consider the following specification:

$$y_{i,t} = \phi_{0,i} + \sum_{p=1}^{P} \phi_{p,i} y_{i,t-p} + \sum_{q=1}^{Q} \beta_{q,i} x_{i,t-q} + \varepsilon_{i,t}, \tag{5.1}$$

[6] In further simulations, we have studied cases where both y and x are drawn based on a VAR(2) process with either homogeneous or heterogeneous coefficients. The results are similar to those already reported here, and so we refrain from discussing these further.

[7] See https://www.fdic.gov/.

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where y denotes *profitability*, which is proxied by the return on assets (ROA), defined as annualized net income after taxes expressed as a percentage of average total assets, and x denotes the time-varying operational *cost efficiency* of bank i at period t, to be defined shortly. The parameters of the model above are described in Sect. 2. For the purposes of the present illustration, we shall focus on the unidirectional link (one-way causation) from cost efficiency to profitability. In addition, we shall impose $P = Q$.

A measure of cost efficiency has been constructed based on a cost frontier model using a translog functional form, two outputs and three inputs. In particular, following Altunbas et al. (2007), we specify

$$\begin{aligned}\ln TC_{i,t} = &\sum_{h=1}^{3} \gamma_h \ln P_{h,i,t} + \sum_{h=1}^{2} \delta_h \ln Y_{h,i,t} + 0.5 \sum_{m=1}^{2}\sum_{n=1}^{2} \mu_{mn} \ln Y_{m,i,t} \ln Y_{n,i,t} \\ &+ \sum_{m=1}^{3}\sum_{n=1}^{3} \pi_{mn} \ln P_{m,i,t} \ln P_{n,i,t} + \sum_{m=1}^{2}\sum_{n=1}^{3} \xi_{mn} \ln Y_{m,i,t} \ln P_{n,i,t} + \eta_i + \tau_t + \upsilon_{it},\end{aligned} \tag{5.2}$$

where TC represents total cost, while Y_1 and Y_2 denote two outputs, net loans and securities, respectively; Y_1 is defined as gross loans minus reserves for loan loss provision. Y_2 is the sum of securities held to maturity and securities held for sale. P_1, P_2 and P_3 denote three input prices, namely the price of capital, price of labour and price of loanable funds. The model above is estimated using two-way fixed effects regression. The bank-specific, time-varying operational inefficiency component is captured by the sum of the two fixed effects, i.e. $\eta_i + \tau_t$. Subsequently, cost efficiency, $x_{i,t}$ is computed as follows:

$$x_{i,t} = e^{\min\{\hat{\eta}_i + \hat{\tau}_t\}_{i,t} - (\hat{\eta}_i + \hat{\tau}_t)}, \tag{5.3}$$

which ensures that larger scores imply higher cost efficiency such that the most efficient bank scores one.

We initially test for Granger non-causality using Eq. (5.1) based on the entire sample, i.e. all 350 banks during 2006:Q1-2019:Q4. Subsequently, we split banks into two groups based on their average size, which is proxied by the natural logarithm of banks total assets. The grouping of banks is performed using a k-means algorithm, as advocated, e.g. in Lin and Ng (2012) and Sarafidis and Weber (2015). In addition, we distinguish between two subperiods, namely "Basel II" (2006:Q1-2010:Q4) and a period under the Dodd-Frank Act "DFA" (2011:Q1-2019:Q4). Basel II represents the second of the Basel Accords and constitutes recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision (BCBS).[8] The DFA is a federal law enacted towards the end of 2010, aiming "to promote the financial stability of the United States by improving accountability and transparency in the financial system, to end "too big to fail", to protect the American taxpayer by ending bailouts, to protect consumers from abusive financial services practices, and

[8] Basel II was eventually superseded by the Basel III framework internationally.

Table 1 Summary statistics for bank size

	Mean	SD	Min	Max
Small banks	11.31	.599	9.71	12.29
Large banks	13.28	1.07	12.31	18.89

for other purposes".[9] In a nutshell, the DFA has instituted a new failure-resolution regime, which seeks to ensure that losses resulting from bad decisions by managers are absorbed by equity and debt holders, thus potentially reducing moral hazard.

5.2 Results

Table 1 reports summary statistics for the two groups of banks in terms of their size, proxied by the natural logarithm of the average value (over time) of total assets.

Table 2 reports results for the Wald test statistic and its p value for the null hypothesis $H_0 : \beta_{q,i} = 0 \quad \text{for all } i \text{ and } q$. We also report the estimated number of lags employed, $\hat{P}$, which is obtained using BIC,[10] as well as estimates for the pooled estimator (standard errors in parentheses) of the Granger-causation parameters, defined in Eq. (3.9) and denoted as $\hat{\beta}$. When $\hat{P} = 1$, $\hat{\beta} = \hat{\beta}_1$ in Eq. (5.1), whereas for $\hat{P} > 1$ we report the sum of the estimates of $\beta_q, q = 1, \ldots, \hat{P}$, i.e. $\hat{\beta} = \sum_{q=1}^{\hat{P}} \hat{\beta}_q$. The variance–covariance matrix of the pooled estimator, $\widehat{V}$, is computed as in Eq. (3.12), i.e. it accommodates cross-sectional heteroskedasticity. For the purposes of comparison, we also report the mean-group estimator of the Granger-causation parameters, $\hat{\beta}_{MG}$, computed using the sample mean (across i) of the corresponding individual-specific regression estimates.

The top panel corresponds to the entire sample of 350 banks. Column "Full" reports results for the entire period of the sample, i.e. 2006:Q1-2019:Q4. Columns "Basel II" and "DFA" present results for two different subperiods, namely 2006:Q1-2010:Q4 and 2011:Q1-2019:Q4, respectively. The middle panel contains results for "small-sized" banks, followed by "large-sized" banks at the bottom panel.

As we can see, in almost all cases the null hypothesis is rejected at the 1% level of significance, which implies that cost efficiency Granger-causes profitability, i.e. past values of x contain information that helps to predict y over and above the information contained in past values of y. The only exception occurs when it comes to large banks during Basel II, where the null hypothesis is not rejected, with a p value approximately equal to 0.509. This result is important because it signifies potential moral hazard-type behaviour prior to the introduction of the DFA; such outcome is consistent with findings in the existing literature, such as those of Cui et al. (2020) and Zhu et al. (2020). However, following the introduction of DFA, the null of Granger non-causality is rejected for large banks as well.

In regards to the remaining quantities, in most cases $\widehat{P} = 1$, i.e. the optimal lagged value of x and y equals unity except for large banks during DFA, where $\widehat{P} = 2$. As

[9] See https://www.cftc.gov/sites/default/files/idc/groups/public/@swaps/documents/file/hr4173_enrolledbill.pdf.

[10] To ensure BIC is consistent under both under null and alternative hypotheses, we estimate P under the alternative, thus allowing for heterogeneity of the Granger causation parameters.

Table 2 Results for the HPJ-based Wald test approach

	Full	Basel II	DFA
All banks			
Wald-stat.	19.67	7.69	10.22
p value	[.000]	[.006]	[.001]
$\widehat{P}$	1	1	1
$\hat{\beta}$	.266 (.038)	.476 (.047)	.186 (.031)
$\hat{\beta}_{MG}$	.349 (.082)	.284 (.161)	.371 (.097)
N	350	350	350
T	56	20	36
Small banks			
Wald-stat.	12.2	7.32	10.74
p value	[.000]	[.007]	[.001]
$\widehat{P}$	1	1	1
$\hat{\beta}$	.244 (.045)	.575 (.059)	.189 (.031)
$\hat{\beta}_{MG}$	.338 (.099)	.525 (.208)	.302 (.116)
N	211	211	211
T	56	20	36
Large banks			
Wald-stat.	9.13	.436	12.65
p value	[.003]	[.509]	[.000]
$\widehat{P}$	1	1	2
$\hat{\beta}$	.346 (.025)	.132 (.019)	.423 (.036)
$\hat{\beta}_{MG}$	.366 (.142)	− .082 (.252)	.477 (.168)
N	139	139	139
T	56	20	36

For $\hat{P} = 1$, $\hat{\beta} = \hat{\beta}_1$ in Eq. (5.1). For $\hat{P} > 1$ $\hat{\beta} = \sum_{q=1}^{\hat{P}} \hat{\beta}_q$. Standard errors in parentheses

expected, the Granger-causation parameters are statistically significant at the 5% level, except for $\hat{\beta}_{MG}$ when the null hypothesis of Granger non-causality is not rejected.

We have also run Granger non-causality tests based on the method of Dumitrescu and Hurlin (2012) (the "DHT" test statistic) using the Stata algorithm developed by Lopez and Weber (2017).[11] The results are identical when it comes to lag model selection using BIC. However, as it turns out, this time the null hypothesis of Granger non-causality is rejected in *all* cases, including for the sample of large banks during the subperiod under Basel II. In particular, in this case the DHT statistic equals 2.58 with a p value of 0.0099. Given that the result is marginal at the 1% level of significance, and taking into account the potentially substantial size distortions observed in the simulations for the DHT test when $T = 20$, one is inclined to trust the outcome of the HPJ-based Wald test reported in Table 2.

[11] We do not report the results to save space. They are available from the authors upon request.

6 Conclusions

This paper considers the problem of Granger non-causality testing in panels with large cross-sectional and time series dimensions. First, we put forward a pooled fixed effects type estimator for the Granger-causation parameters, which makes use of the fact that, under the null hypothesis, these parameters are all equal to zero and, thus, they are homogeneous. Pooling over cross sections guarantees that the estimator has a $\sqrt{NT}$ convergence rate. In order to account for the well-known "Nickell bias", we make use of the Split Panel Jackknife procedure of Dhaene and Jochmans (2015). Subsequently, a Wald test is proposed, which is based on the bias-corrected fixed effects type estimator. The resulting approach is valid irrespective of whether the alternative hypothesis is homogeneous or heterogeneous, or whether the autoregressive parameters vary across individuals or not, so long as T is (at least moderately) large.

The statistical model considered in this paper rules out any forms of the cross-sectional dependence in $\varepsilon_{i,t}$. This restriction can be easily relaxed if one is willing to assume that cross-sectional dependence is strong, generated by an unobserved factor component, $\boldsymbol{\lambda}_i' \boldsymbol{f}_t$. In particular, in this case one can use either the Common Correlated Effects (CCE) approach of Pesaran (2006)/Chudik and Pesaran (2015) combined with HPJ as in Juodis et al. (2020), or the PC estimator of Bai (2009)/Ando and Bai (2017). In these setups, the HPJ-based statistic provides a natural starting point, as the finite T corrections proposed by Dumitrescu and Hurlin (2012) are not feasible. In panels with homogeneous autoregressive parameters and T fixed, one can employ the GMM framework of Robertson and Sarafidis (2015) and the linear GMM estimator of Juodis and Sarafidis (2020).[12] We leave these avenues for future research.

Acknowledgements We are delighted to contribute this paper to a special issue in honour of Professor Badi Baltagi, who has made an enormous contribution to the field of econometrics. We are grateful to the Guest Editors and all referees involved. We would like to thank participants at the IPDC2019 conference (Vilnius) for useful comments and suggestions. Part of this research project was conducted while the first author was visiting the Applied Macroeconomic Research Division (TMTS) at the Bank of Lithuania (BoL). The views expressed in this paper are those of authors and do not necessarily represent the official views of the Bank of Lithuania or the Eurosystem. Financial support from the Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged by Juodis. Sarafidis gratefully acknowledges financial support from the Australian Research Council, under research Grant No. DP-170103135.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted

[12] It is possible to use alternative estimators for this class of models in fixed-T panels, such as those reviewed by Juodis and Sarafidis (2018).

by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix A: Monte Carlo Results

See Tables 3, 4, 5 and 6.

Table 3 Empirical rejection rates

N	T	$\rho = 0.4$								$\rho = 0.8$							
		$\beta = 0$		$\beta = .02$		$\beta = .03$		$\beta = .05$		$\beta = 0$		$\beta = .02$		$\beta = .03$		$\beta = .05$	
		HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT
50	20	9.1	8.8	9.8	8.7	14.0	10.4	30.3	17.4	14.4	21.5	7.6	9.6	10.7	12.1	22.6	18.0
	50	7.1	6.7	12.9	9.7	26.8	14.7	62.7	30.4	9.5	10.5	13.5	10.3	25.5	17.9	61.7	37.2
	100	5.7	4.7	27.4	13.5	53.9	20.9	91.1	52.1	7.9	8.3	29.4	14.6	58.8	25.3	94.3	63.2
100	20	11.3	12.6	13.3	8.1	19.8	9.4	46.3	19.4	14.1	35.9	11.7	8.2	17.3	11.5	40.6	22.1
	50	7.1	6.6	25.6	13.2	52.2	22.2	91.0	48.4	9.5	15.3	22.0	13.0	48.3	24.3	91.1	54.9
	100	5.9	4.8	48.3	17.3	81.1	33.0	99.5	75.9	6.7	7.9	55.6	21.2	88.6	44.1	99.8	89.7
200	20	10.9	15.1	21.5	11.8	40.2	18.2	77.7	37.0	14.3	55.5	19.2	10.6	35.0	16.4	69.2	35.7
	50	5.8	8.3	46.1	16.9	79.8	31.2	99.8	72.5	9.6	22.1	38.4	20.6	77.0	38.9	99.8	82.1
	100	5.1	6.3	77.8	22.9	98.0	47.2	100	95.4	7.0	11.4	84.4	32.9	99.3	63.5	100	99.3

Homogeneous model, with homoskedastic innovations

In the homogeneous model, $\alpha_i = \alpha = 0.4$ for all i. Moreover, $\beta_i = \beta$ for all i, where $\beta = 0$ under the null (size) and $\beta = \{0.02; 0.03; 0.05\}$ under alternative hypotheses (power). The innovations of the process for $y_{i,t}$ are homoskedastic, with mean zero and variance equal to 0.07, i.e. $\sigma^2_{\varepsilon_y} = 0.07$. Size-adjusted power is reported

Table 4 Empirical rejection rates

N	T	$\rho = 0.4$								$\rho = 0.8$							
		$\beta = 0$		$E[\beta_i] = .02$		$E[\beta_i] = .03$		$E[\beta_i] = .05$		$\beta = 0$		$E[\beta_i] = .02$		$E[\beta_i] = .03$		$E[\beta_i] = .05$	
		HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT
50	20	8.7	8.9	10.3	8.0	16.6	9.8	31.3	16.7	13.7	22.5	8.7	8.7	14.0	10.2	23.9	18.3
	50	6.6	6.4	14.4	10.3	30.0	15.3	63.3	29.0	8.5	11.9	13.3	10.0	30.0	17.8	62.5	35.2
	100	5.3	5.7	27.9	11.5	54.9	19.1	93.2	52.3	6.9	7.4	29.5	15.8	62.5	28.8	95.8	68.4
100	20	10.6	11.1	11.9	9.5	22.2	13.7	47.0	23.8	14.7	36.8	9.2	9.0	18.3	13.8	37.7	22.0
	50	6.8	7.5	25.7	11.3	48.1	18.7	91.2	44.7	9.1	15.5	24.1	14.9	48.0	25.3	90.8	58.4
	100	5.5	5.7	49.6	18.6	81.7	30.8	100	72.9	6.6	8.3	58.5	26.7	89.1	46.8	100	90.5
200	20	10.3	13.8	24.4	15.1	41.4	21.2	78.9	41.1	14.9	56.4	19.5	15.9	32.3	23.4	69.5	44.2
	50	5.8	9.4	49.2	18.1	80.1	30.8	99.6	71.4	11.1	22.5	41.6	23.1	75.4	41.3	99.8	84.2
	100	6.5	6.8	75.1	24.2	98.1	48.9	100	94.7	8.4	11.0	83.9	35.3	99.5	66.0	100	99.1

Heterogeneous model, with homoskedastic innovations

In the heterogeneous model, $\alpha_i = \alpha + \xi_i^{(y)}$ with $\alpha = 0.4$ and $\xi_i^{(y)} \sim i.i.d.U[-0.15, 0.15]$, such that $E[\alpha_i] = 0.4$. Under the null, $\beta_i = \beta = 0$. Under the alternative hypothesis, $\beta_i = \beta + \xi_i^{(x)}$ with $\beta = \{0.02; 0.03; 0.05\}$ and $\xi_i^{(x)} \sim i.i.d.U[-0.10, 0.10]$. The innovations of the process for $y_{i,t}$ are homoskedastic, with mean zero and variance equal to 0.07, i.e. $\sigma^2_{\varepsilon_y} = 0.07$. Size-adjusted power is reported

Table 5 Empirical rejection rates

N	T	$\rho = 0.4$								$\rho = 0.8$							
		$\beta = 0$		$\beta = .02$		$\beta = .03$		$\beta = .05$		$\beta = 0$		$\beta = .02$		$\beta = .03$		$\beta = .05$	
		HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT
50	20	8.7	9.4	9.6	8.4	14.3	13.4	29.2	25.0	11.2	24.4	8.6	6.8	12.8	10.3	22.8	21.3
	50	8.0	7.5	12.0	12.9	25.6	23.1	54.7	52.2	9.7	12.3	11.3	11.5	23.1	22.8	55.1	55.6
	100	6.2	5.8	23.7	20.2	45.7	39.2	88.3	78.1	7.9	9.1	30.2	24.4	57.5	47.1	93.3	88.4
100	20	9.3	10.9	13.6	11.7	21.1	18.8	42.0	41.0	11.6	36.9	9.7	10.0	17.1	13.5	35.8	32.8
	50	6.2	6.4	23.4	20.3	46.5	40.3	87.8	80.1	8.8	15.1	21.1	17.9	42.4	36.6	86.4	82.2
	100	5.5	5.6	45.2	28.3	77.0	59.7	99.3	97.0	5.8	8.2	55.6	40.6	87.0	73.6	99.9	99.6
200	20	9.8	15.2	19.6	15.2	37.5	28.4	71.9	63.9	11.6	57.6	17.8	12.2	31.4	23.9	65.3	53.9
	50	6.2	9.2	42.1	28.6	73.9	54.3	99.1	95.2	7.5	24.1	39.5	29.1	73.7	59.2	99.2	96.8
	100	5.2	4.6	73.6	51.4	96.8	86.0	100	100	7.4	11.3	80.4	58.1	98.7	92.8	100	100

Homogeneous model, with heteroskedastic innovations

In the homogeneous model, $\alpha_i = \alpha = 0.4$, and $\beta_i = \beta$ for all i, where $\beta = 0$ under the null (size) and $\beta = \{0.02; 0.03; 0.05\}$ under alternative hypotheses (power). The innovations of the process for $y_{i,t}$ are heteroskedastic, i.e. $\sigma^2_{\varepsilon_y}$ is scaled by $\xi_i^{(\varepsilon)} \sim i.i.d.U\,[0, 2]$, such that $E\left[\sigma^2_{\varepsilon_y,i}\right] = \sigma^2_{\varepsilon_y} E\left[\xi_i^{(\varepsilon)}\right] = 0.07$. Size-adjusted power is reported

Table 6 Empirical rejection rates

N	T	$\rho = 0.4$								$\rho = 0.8$							
		$\beta = 0$		$E[\beta_i] = .02$		$E[\beta_i] = .03$		$E[\beta_i] = .05$		$\beta = 0$		$E[\beta_i] = .02$		$E[\beta_i] = .03$		$E[\beta_i] = .05$	
		HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT	HPJ	DHT
50	20	9.6	8.8	6.4	10.4	12.1	14.1	23.8	25.0	13.4	23.3	8.7	10.8	12.6	15.2	25.3	23.5
	50	7.7	7.5	13.3	11.8	23.2	21.0	55.7	51.5	9.1	10.8	12.9	15.7	24.0	26.7	59.8	61.9
	100	7.4	4.5	21.3	19.1	44.1	39.8	85.2	83.0	7.5	7.0	28.2	25.4	56.4	49.3	93.3	91.2
100	20	9.7	11.7	12.2	10.6	21.9	16.8	42.6	35.6	11.5	36.4	11.0	8.8	19.9	13.3	39.5	30.0
	50	6.7	7.2	22.7	19.2	46.1	33.6	85.7	77.0	8.5	15.4	22.7	20.0	46.7	39.1	86.9	82.8
	100	6.6	6.1	43.6	27.0	75.4	58.1	99.4	97.3	7.7	10.0	54.7	37.5	86.0	72.7	100.0	99.5
200	20	8.5	15.7	22.2	16.8	37.4	25.4	75.2	58.4	10.8	58.3	18.9	11.9	32.2	19.6	67.5	50.6
	50	6.3	7.2	41.8	30.2	74.9	59.3	99.1	96.7	8.4	24.3	39.8	31.4	75.3	60.1	99.0	98.2
	100	6.9	5.8	67.8	45.8	96.1	82.7	100.0	100.0	7.7	12.2	79.6	61.3	99.0	94.5	100.0	100.0

Heterogeneous model, with heteroskedastic innovations

In the heterogeneous model, $\alpha_i = \alpha + \xi_i^{(y)}$ with $\alpha = 0.4$, and $\xi_i^{(y)} \sim i.i.d.U[-0.15, 0.15]$, such that $E[\alpha_i] = 0.4$. Under the null, $\beta_i = \beta = 0$. Under the alternative hypothesis, $\beta_i = \beta + \xi_i^{(x)}$ with $\beta = \{0.02; 0.03; 0.05\}$ and $\xi_i^{(x)} \sim i.i.d.U[-0.10, 0.10]$. The innovations of the process for $y_{i,t}$ are heteroskedastic, i.e. $\sigma^2_{\varepsilon_y}$ is scaled by $\xi_i^{(\varepsilon)} \sim i.i.d.U[0, 2]$, such that $E\left[\sigma^2_{\varepsilon_y,i}\right] = \sigma^2_{\varepsilon_y} E\left[\xi_i^{(\varepsilon)}\right] = 0.07$. Size-adjusted power is reported

References

Altunbas Y, Carbo S, Gardener EP, Molyneux P (2007) Examining the relationships between capital, risk and efficiency in European banking. Eur Financ Manag 13:49–70

Anderson TW, Hsiao C (1982) Formulation and estimation of dynamic models using panel data. J Econom 18:47–82

Ando T, Bai J (2017) Clustering huge number of financial time series: a panel data approach with high-dimensional predictors and factor structures. J Am Stat Assoc 112:1182–1198

Arellano M (1987) Computing robust standard errors for within-groups estimators. Oxf Bull Econ Stat 49:431–434

Arellano M (2016) Modeling optimal instrumental variables for dynamic panel data models. Res Econ 70:238–261

Bai J (2009) Panel data models with interactive fixed effects. Econometrica 77:1229–1279

Binder M, Hsiao C, Pesaran MH (2005) Estimation and inference in short panel vector autoregressions with unit root and cointegration. Econom Theory 21:795–837

Chambers MJ (2013) Jackknife estimation of stationary autoregressive models. J Econom 172:142–157

Chudik A, Pesaran MH (2015) Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. J Econom 188:393–420

Chudik A, Pesaran MH, Yang J-C (2018) Half-panel Jackknife fixed effects estimation of panels with weakly exogenous regressor. J Appl Econom 33:816–836

Cui G, Sarafidis V, Yamagata T (2020) Large IV estimation of spatial dynamic panels with interactive effects: large sample theory and an application on bank attitude toward risk. Working Paper MPRA Paper No. 102488, Munich Personal RePEc Archive

Dhaene G, Jochmans K (2015) Split-panel Jackknife estimation of fixed-effect models. Rev Econ Stud 82:991–1030

Dumitrescu E-I, Hurlin C (2012) Testing for Granger non-causality in heterogeneous panels. Econ Model 29:1450–1460

Fernández-Val I, Lee J (2013) Panel data models with nonadditive unobserved heterogeneity: estimation and inference. Quant Econ 4:453–481

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424–438

Hahn J, Kuersteiner G (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. Econometrica 70(4):1639–1657

Holtz-Eakin D, Newey WK, Rosen HS (1988) Estimating vector autoregressions with panel data. Econometrica 56:1371–1395

Im KS, Pesaran M, Shin Y (2003) Testing or unit roots in heterogeneous panels. J Econom 115:53–74

Juodis A (2013) A note on bias-corrected estimation in dynamic panel data models. Econ Lett 118:435–438

Juodis A (2018) First difference transformation in panel VAR models: robustness, estimation and inference. Econom Rev 37:650–693

Juodis A, Sarafidis V (2018) Fixed T dynamic panel data estimators with multifactor errors. Econom Rev 37:893–929

Juodis A, Sarafidis V (2020) A linear estimator for factor-augmented fixed-T panels with endogenous regressors. J Bus Econ Stat:1–15

Juodis A, Karabıyık H, Westerlund J (2020) On the robustness of the pooled CCE estimator. J Econom **(forthcoming)**

Karavias Y, Tzavalis E (2016) Local power of fixed-T panel unit root tests with serially correlated errors and incidental trends. J Time Ser Anal 37:222–239

Karavias Y, Tzavalis E (2017) Local power of panel unit root tests allowing for structural breaks. Econom Rev 36:1123–1156

Lin C, Ng S (2012) Estimation of panel data models with parameter heterogeneity when group membership is unknown. J Econom Methods 1:42–55

Lopez L, Weber S (2017) Testing for Granger causality in panel data. Stata J 4:972–984

Neyman J, Scott EL (1948) Consistent estimation from partially consistent observations. Econometrica 16:1–32

Nickell S (1981) Biases in dynamic models with fixed effects. Econometrica 49:1417–1426

Pesaran MH (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74:967–1012

Pesaran MH (2012) On the interpretation of panel unit root tests. Econ Lett 116:545–546
Robertson D, Sarafidis V (2015) IV estimation of panels with factor residuals. J Econom 185:526–541
Sarafidis V, Weber N (2015) A partially heterogeneous framework for analyzing panel data. Oxf Bull Econ Stat 77:274–296
Zhu H, Sarafidis V, Silvapulle MJ (2020) A new structural break test for panels with common factors. Econom J 23:137–155

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Artūras Juodis[1] · Yiannis Karavias[2] · Vasilis Sarafidis[3,4]

Yiannis Karavias
i.karavias@bham.ac.uk

Vasilis Sarafidis
vasilis.sarafidis@monash.edu

1 Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

2 Department of Economics, University of Birmingham, Birmingham, UK

3 Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia

4 BI Norwegian Business School, Oslo, Norway