CESifo WORKING PAPERS

10509
2023

Original Version: June 2023
This Version: April 2024

Heterogeneous Autoregressions in Short T Panel Data Models

M. Hashem Pesaran, Liying Yang

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10509*

Heterogeneous Autoregressions in Short T Panel Data Models

Abstract

This paper considers a first-order autoregressive panel data model with individual-specific effects and heterogeneous autoregressive coefficients defined on the interval (-1; 1], thus allowing for some of the individual processes to have unit roots. It proposes estimators for the moments of the cross-sectional distribution of the autoregressive (AR) coefficients, assuming a random coefficient model for the autoregressive coefficients without imposing any restrictions on the fixed effects. It is shown the standard generalized method of moments estimators obtained under homogeneous slopes are biased. Small sample properties of the proposed estimators are investigated by Monte Carlo experiments and compared with a number of alternatives, both under homogeneous and heterogeneous slopes. It is found that a simple moment estimator of the mean of heterogeneous AR coefficients performs very well even for moderate sample sizes, but to reliably estimate the variance of AR coefficients much larger samples are required. It is also required that the true value of this variance is not too close to zero. The utility of the heterogeneous approach is illustrated in the case of earnings dynamics.

JEL-Codes: C220, C230, C460.

Keywords: heterogeneous dynamic panels, neglected heterogeneity bias, short T panels, earnings dynamics.

M. Hashem Pesaran
Department of Economics
University of Southern California
Los Angeles / CA / USA
pesaran@usc.edu

Liying Yang
Sauder School of Business
University of British Columbia
Vancouver / BC / Canada
liying.yang@sauder.ubc.ca

April 3, 2024
We are grateful to two anonymous reviewers for most helpful and constructive comments. We have also benefited greatly from helpful comments and suggestions by Alexander Chudik, Ron Smith and Hayun Song. The working paper version of this paper was completed when Liying Yang was a Ph.D. student at Department of Economics, University of Southern California. She is now a postdoctoral research fellow at the Sauder School of Business at the University of British Columbia.

1 Introduction

The importance of cross-sectional heterogeneity in panel regressions is becoming increasingly recognized in the literature. When the time dimension of the panel, T, is short, significant advances have been made in the case of random coefficient models with strictly exogenous regressors, for example, Chamberlain (1992), Wooldridge (2005), and Graham and Powell (2012). A trimmed version of the mean group estimator proposed by Pesaran and Smith (1995) can also be applied to ultra short T panels when the regressors are strictly exogenous. See Pesaran and Yang (2023). In contrast, there are only a few papers that consider the estimation of heterogeneous dynamic panels when the time dimension is short.

There are some limitations to applying existing estimation methods to such heterogeneous short T dynamic panels. The generalized method of moments (GMM) estimators applied after first differencing by Anderson and Hsiao (1981, 1982), Arellano and Bond (1991), Blundell and Bond (1998), and Chudik and Pesaran (2021), allow for intercept heterogeneity but not for possible heterogeneity in the autoregressive (AR) coefficients, and as shown in this paper, can lead to biased estimates and distorted inference. Gu and Koenker (2017) and Liu (2023) consider the estimation of panel AR(1) models with exogenous regressors using Bayesian techniques. While they assume random coefficients on strictly exogenous regressors, they still impose homogeneity on the AR coefficients. The mean group estimator and the hierarchical Bayesian estimator proposed by Hsiao et al. (1999) allow for heterogeneity but require that T is reasonably large relative to the cross section dimension, n.

For moderate values of T, analytical, Bootstrap and Jackknife bias correction approaches have also been proposed to deal with the small sample bias of the mean group and other related estimators. See, for example, Pesaran and Zhao (1999), Okui and Yanagi (2019) and Okui and Yanagi (2020). Even with bias corrections, n cannot be too large compared with T, since a valid inference based on the asymptotic distribution often requires $nT^{-c} \to 0$, for some constant $c > 2$. In short, none of the above approaches are appropriate and can lead to seriously biased estimates and distorted inference when T is small and fixed as $n \to \infty$. Nonetheless, heterogeneity in dynamics can play an important role in many empirical studies using short T panel data models. Examples include, earnings dynamics studied by Meghir and Pistaferri (2004), unemployment dynamics by Browning and Carro (2014), and firm's growth by Liu (2023)

This paper considers a relatively simple panel AR(1) model, but allows for both individual fixed effects and heterogeneous AR coefficients, ϕ_i, where some of the individual processes, $\{y_{it}\}$, could have unit roots, $\phi_i = 1$. We eliminate the fixed effects by first differencing, $\Delta y_{it} = y_{it} - y_{i,t-1}$, and establish conditions under which the mean and variance of ϕ_i can be identified from the autocovariances of Δy_{it}, averaged over i. We show that existing GMM estimators of $E(\phi_i) = \mu_\phi$ are asymptotically biased, and derive analytical expressions for their bias in simple cases. We then propose estimators for the moments of ϕ_i, in particular, $E(\phi_i)$ and $E(\phi_i^2)$, using cross-sectional averages of the autocorrelation coefficients of the first differences. In terms of the estimation approach, the most relevant paper to ours is by Robinson (1978), who considered a random coefficient AR(1) model without fixed effects. Assuming the "usual" stationary conditions, he proposed identifying the moments of ϕ_i as functions of autocovariances of y_{it}.

In particular, we propose two new estimators for the moments $\theta_s = E(\phi_i^s)$ for $s = 1, 2, ..., T-3$. A relatively simple estimator based on autocorrelations of first differences, denoted by FDAC, and a generalized method of moments (GMM) estimator based on autocovariances of first differences, which we denote by HetroGMM. We also consider estimation of $Var(\phi_i) = \sigma_\phi^2 = \theta_2 - \theta_1^2$, when the true value of σ_ϕ^2 is not too close to zero. We do not make any assumptions about the fixed effects and allow them to have arbitrary correlations with ϕ_i, but require the underlying AR(1) processes to be stationary after first differencing and assume ϕ_i and the error variances are independently distributed. It is possible to extend our analysis to higher-order panel AR processes and dynamic panels with exogenous regressors. However, these important extensions are outside the scope of the present paper.

We compare FDAC and HetroGMM estimators to a kernel-weighted likelihood estimator proposed by Mavroeidis et al. (2015), MSW. Assuming independently distributed Gaussian errors with cross-sectional heteroskedasticity, MSW show that the unknown distribution of heterogeneous coefficients can be identified, provided the linear operator that maps the unknown distribution to the joint distribution of data is complete (or "invertible"). They provide an estimation algorithm for the parametric version of their estimator assuming the heterogeneous coefficients, including the intercepts and ϕ_i, follow a multivariate normal distribution. The estimation algorithm becomes computationally very demanding if the parametric assumption about the distribution of ϕ_i is relaxed.

We investigate small sample properties of FDAC and HetroGMM estimators using Monte Carlo (MC) experiments. The simulations show that the relatively simple FDAC estimator performs better than the HetroGMM estimator uniformly across different sample sizes, and is robust to non-Gaussian errors and conditional error heteroskedasticity.

We also compare the small sample properties of the FDAC estimator of μ_ϕ with several GMM estimators proposed in the literature for homogenous AR panels, including the popular Arellano and Bond (1991), AB, and Blundell and Bond (1998), BB, estimators. We refer to these as HomoGMM estimators, to be distinguished from the HetroGMM estimator proposed in this paper. The simulation results confirm the neglected heterogeneity bias of the HomoGMM estimators, and show that the FDAC estimator of μ_ϕ performs well for all values of $T = 4, 6, 10$ and $n = 100, 1,000$ and $5,000$, so long as the underlying processes are stationary after first differencing. This is true for bias, root mean square errors, and size. Both FDAC and HetroGMM estimators are robust to the presence of unit roots and non-Gaussian errors, but can be subject to bias and size distortions if the distribution of the initial values, y_{i0}, significantly depart from stationarity. Similar comparative outcomes are also obtained when estimating σ_ϕ^2, except that much larger sample sizes (n and/or T) are required for reliable estimation and inference. In addition, it is important that the true value of σ_ϕ^2 is not too close to the boundary value of 0. When n and T are not sufficiently large, estimates of σ_ϕ^2 obtained using the plugging estimator, $\hat{\sigma}_\phi^2 = \hat{\theta}_2 - \hat{\theta}_1^2$, can be negative. This occurs with a high frequency when $n = 100$, and $T = 5$. The occurrence of negative estimates declines rapidly when $T = 10$ and $n \geq 1,000$.

Using Monte Carlo experiments we also provide a limited comparison of the MSW and

FDAC estimators of μ_ϕ, and find that in general, the MSW estimator does not have satisfactory small-sample performance under the data generating process in the paper. As the MSW estimator depends on the assumed Gaussian distribution of ϕ_i, it can be severely biased with uniformly and categorically distributed ϕ_i that we consider in our MC experiments.

Finally, we provide an empirical application using five and ten yearly samples from the PSID dataset over the 1976–1995 period to estimate the persistence of real earnings. To this end, we extend the basic panel AR(1) model to allow for linear trends. Following the empirical literature we report estimates for three educational categories (high school dropouts, high school graduates and college graduates) and all three categories combined. We find comparable estimates for the linear trend coefficients across sub-periods and educational categories, around 2 per cent per annum. The FDAC estimates of mean persistence (μ_ϕ) for the sub-periods 1991–1995 and 1986–1995 fall in the range of $0.570 - 0.734$, and tend to rise with the level of educational attainment, with college graduates showing the highest degree of persistence. No such patterns are observed for other estimates, which are around $0.3, 0.9$ and 0.41 for the AB, BB and MSW estimators, respectively. The FDAC estimates of σ_ϕ^2 for all three categories combined are statistically significant and are given by 0.100 (0.042) and 0.129 (0.023) for the sub-periods 1991–1995 ($n = 1,366$) and 1986–1995 ($n = 1,139$), respectively, providing further evidence of heterogeneity in real earnings persistence.

The rest of the paper is set out as follows. Section 2 sets out the model and assumptions. Section 3 derives the autocovariances of the first differences, $\Delta y_{it} = y_{it} - y_{i,t-1}$, and establishes conditions under which they are stationary. Section 4 shows that the HomoGMM estimators are biased in the heterogeneous panel AR(1) models. Section 5 establishes conditions under which the moments of ϕ_i can be identified from the autocorrelation functions of first differences. Section 6 proposes FDAC and HetroGMM estimators of the moments of ϕ_i. The respective asymptotic distributions are also derived. Section 7 evaluates the performance of FDAC, HetroGMM, HomoGMM, and MSW estimators by Monte Carlo simulations. Section 8 presents the empirical application, and Section 9 concludes. Some of the mathematical derivations, Monte Carlo evidence and additional empirical results are provided in an online supplement.

2 Model and assumptions

We consider the following first-order autoregressive panel data model

$$y_{it} = \alpha_i + \phi_i y_{i,t-1} + u_{it}, \text{ for } i = 1, 2, ..., n, \tag{2.1}$$

where the fixed effects, α_i, are restricted, $\alpha_i = \mu_i (1 - \phi_i)$. This restriction is necessary for y_{it} to have a fixed mean irrespective of whether $\phi_i = 1$ or $|\phi_i| < 1$. If α_i is unrestricted, a linear trend is introduced in y_{it} when $\phi_i = 1$. The restriction on α_i is not binding when $|\phi_i| < 1$. We impose the restriction since we will be considering a mixture of processes with and without unit roots. With $\alpha_i = \mu_i(1 - \phi_i)$, (2.1) can be written equivalently as

$$y_{it} - \mu_i = \phi_i (y_{i,t-1} - \mu_i) + u_{it}, \text{ for } i = 1, 2, ..., n. \tag{2.2}$$

Suppose that y_{it} is generated starting at time $t = -M_i \leq 0$ with the initial value, $y_{i,-M_i}$. We assume observations on all the n units are available over the periods $t = 1, 2, 3, ..., T$, yielding a total of nT observations $\{y_{i1}, y_{i2}, ..., y_{iT}, \; i = 1, 2, ..., n\}$. The parameters of interest are first and higher order moments of ϕ_i, which we denote by $\theta_s = E(\phi_i^s)$, $s = 1, 2, ..., T-2$. The key feature of our analysis is to allow for a high degree of parameter heterogeneity when T is short as $n \to \infty$. We allow ϕ_i to take any values in the non-explosive interval $[-1+\epsilon, 1]$ for some $\epsilon > 0$, which includes the unit root case, $\phi_i = 1$ for some of the units, but rules out a negative unit root, namely it is required that $\inf_i(1+\phi_i) > 0$. We are able to accommodate distributions of ϕ_i with a non-zero mass on $\phi_i = 1$, by basing our estimation of θ_s on autocorrelations of first differences, $\Delta y_{it} = y_{it} - y_{i,t-1}$, rather than the autocovariances of y_{it} considered by Robinson (1978). As examples, we consider a uniform distribution of ϕ_i defined over the interval $(-1, 1-\epsilon]$ with $\epsilon > 0$, and a categorical distribution where ϕ_i takes two values, ϕ_H (high) and ϕ_L (low), with probabilities $(1-\pi)$ and π, respectively. The unit root case arises when $\epsilon = 0$ (for the uniform distribution), and $\phi_H = 1$ with $0 < \phi_L < 1$ (for the categorical distribution). Our analysis does not allow for a negative unit root, namely when $\phi_i = -1$.

The key identification assumption is the stationarity of the first differences. First differencing of (2.1) eliminates the fixed effects, $\alpha_i = \mu_i(1-\phi_i)$, but does not remove the effects of initial values, $y_{i,-M_i}$, on the first differences when T is small. Under slope heterogeneity, the effects of initial values on first differences do not vanish for processes whose ϕ_i falls in the stable region, $-1 < \phi_i < 1$, unless they are all initialized at a distant past, namely only if $M_i \to \infty$, otherwise the realized values y_{it} and/or their first differences $\{\Delta y_{it}$, for $t = 2, 3, ..., T\}$ will depend on $y_{i,-M_i} - \mu_i$. Including the observations y_{i0} amongst the realizations does not resolve the problem, since we move one period backward and the distribution of $y_{i,-1}$ must still be specified and so on. For processes with unit roots, $\phi_i = 1$, we have $\Delta y_{it} = u_{it}$, and initialization will not be an issue, at least not for the unit-root AR(1) process.

To accommodate the possible mixture of stationary and unit-root processes and achieve identification of the moments of ϕ_i, we make the following assumptions regarding unit-specific parameters, $\boldsymbol{\psi}_i = (\mu_i, \phi_i, \sigma_i^2)'$, the error terms, u_{it}, and the initial value deviations, $y_{i,-M_i} - \mu_i$ for $i = 1, 2, ..., n$, where $\sigma_i^2 = Var(u_{it})$.

Assumption 1 *(individual effects) The individual specific means, μ_i, are bounded, $sup_i |\mu_i| < C$.*

Assumption 2 *(errors) Conditional on $\boldsymbol{\psi}_i = (\mu_i, \phi_i, \sigma_i^2)'$, the errors, u_{it}, are cross-sectionally and serially independent over i and t, with zero means, $E(u_{it}^2) = \sigma_i^2$, and $\sup_{i,t} E|u_{it}|^4 < C < \infty$.*

Assumption 3 *(error variances) (a) The error variances, σ_i^2, are independent draws from a common probability distribution such that $E(\sigma_i^2) = \sigma^2$, where $0 < c < \sigma_i^2$, and $\sigma^2 < C < \infty$. (b) σ_i^2 are distributed independently of ϕ_i.*

Assumption 4 *(autoregressive coefficients) (a) The autoregressive coefficients, ϕ_i, for $i = 1, 2, ..., n$, are independent draws from a common probability distribution, defined on the closed interval $\phi_i \in [-1+\epsilon, 1]$, for some $\epsilon > 0$, with mean $E(\phi_i) = \mu_\phi$ and variance $Var(\phi_i) = \sigma_\phi^2 \geq 0$.*

Assumption 5 *(initialization) The process* $\{y_{it}\}$ *is initialized with* $y_{i,-M_i}$*, where* $M_i \in \mathbb{N} = \{0, 1, 2, ...\}$*, and* $y_{i,-M_i} - \mu_i$ *is given and bounded,* $sup_i |y_{i,-M_i} - \mu_i| < C$*.*

Assumption 1 imposes minimal restrictions on μ_i or on the fixed effects α_i for $|\phi_i| < c < 1$. But as noted earlier, to ensure that y_{it} is not subject to a drift, as it is standard in the unit root literature, α_i is set to 0 when $\phi_i = 1$. Assumptions 2 and 3 are standard in the literature on short T dynamic panels. They allow for cross-sectional as well as conditional time series heteroskedasticity, such as GARCH effects, but rule out unconditional time series heteroskedasticity. Denoting the available information at time $t-1$ by $\mathcal{I}_{i,t-1}$, $E(u_{it}^2 | \mathcal{I}_{i,t-1})$ could be time-varying, so long as $E(u_{it}^2) = \sigma_i^2$ as required by Assumption 2.

Assumptions 4 and 5 ensure that Δy_{it} is covariance stationary if $M_i \to \infty$, without requiring y_{it} to be stationary for all n units in the panel.

3 Autocovariances of first differences

Before setting out our approach to the identification of $\theta_s = E(\phi_i^s)$, we need to derive expressions for the autocovariances of Δy_{it}. Given the available data and after first differencing (2.1), we have

$$\Delta y_{it} = \phi_i \Delta y_{i,t-1} + \Delta u_{it}, \text{ for } t = 2, 3, ..., T. \tag{3.1}$$

Also setting $t = 1$ and using (2.2) we obtain

$$\Delta y_{i1} = -(1 - \phi_i)(y_{i0} - \mu_i) + u_{i1}. \tag{3.2}$$

Iterating (3.1) forward from $t = 2$ and using the above expression for Δy_{i1}, we obtain

$$\Delta y_{it} = u_{it} - (1 - \phi_i) \sum_{\ell=1}^{t-1} \phi_i^{\ell-1} u_{i,t-\ell} - \phi_i^{t-1}(1 - \phi_i)(y_{i0} - \mu_i). \tag{3.3}$$

It is clear that in general, Δy_{it} depends on $y_{i0} - \mu_i$, and Assumption 5 is required if we are to eliminate the impact of initial values on the autocovariances of Δy_{it}. Iterating equation (2.2) forward from $y_{i,-M_i}$ to $t = 0$ we have

$$y_{i0} - \mu_i = \phi_i^{M_i}(y_{i,-M_i} - \mu_i) + \sum_{\ell=0}^{M_i-1} \phi_i^{\ell} u_{i,-\ell}. \tag{3.4}$$

Substituting $y_{i0}-\mu_i$ from (3.4) in (3.3) now yields

$$\Delta y_{it} = u_{it} - (1-\phi_i)\sum_{\ell=1}^{M_i+t-1}\phi_i^{\ell-1}u_{i,t-\ell} + \phi_i^t R_i\left(y_{i,-M_i}\right), \text{ for } t=2,3,...,T, \tag{3.5}$$

where $R_i\left(y_{i,-M_i}\right) = -\phi_i^{M_i-1}(1-\phi_i)\left(y_{i,-M_i}-\mu_i\right)$. For a fixed T, the remainder term, R_i, does not vanish unless $M_i \to \infty$. Note that under Assumption 5 $\sup_i \left|y_{i,-M_i}-\mu_i\right| < C$, and $\left|R_i\left(y_{i,-M_i}\right)\right| \leq \left|\phi_i\right|^{M_i-1}\left|1-\phi_i\right|\left|y_{i,-M_i}-\mu_i\right| \leq C\left|\phi_i\right|^{M_i-1}\left|1-\phi_i\right|$, and $\left|R_i\left(y_{i,-M_i}\right)\right| \to 0$, for *all* i (irrespective of whether $\phi_i=1$ or $|\phi_i|<1$), *if and only if* $M_i \to \infty$. Under this condition

$$\Delta y_{it} = u_{it} - (1-\phi_i)\sum_{\ell=1}^{\infty}\phi_i^{\ell-1}u_{i,t-\ell}, \tag{3.6}$$

and the available first differences, Δy_{it} for $t=2,3,...,T$, do not depend on y_{i0}, and can be used to derive expressions for $\gamma_\Delta(h) = E\left(\Delta y_{it}\Delta y_{i,t-h}\right)$ for $h=0,1,...,T-2$. But first, we need to establish that these autocovariances do exist, particularly given that we are allowing for some y_{it} processes to have unit roots. This requirement is easily established when the distribution of ϕ_i is categorical. In this case we have

$$\gamma_\Delta(h) = \pi E\left(\Delta y_{it}\Delta y_{i,t-h}\left||\phi_i| < c < 1\right.\right) + (1-\pi)E\left(\Delta y_{it}\Delta y_{i,t-h}\left|\phi_i = 1\right.\right),$$

where $0<\pi\leq 1$. By application of Minkowski's inequality to (3.6) we have (for $p \geq 1$)

$$\left\|\Delta y_{it}\right\|_p \leq \left\|u_{it}\right\|_p + \sum_{\ell=1}^{\infty}\left[\left\|\phi_i^{\ell-1}\right\|_p + \left\|\phi_i^{\ell}\right\|_p\right]\left\|u_{i,t-\ell}\right\|_p,$$

where $\left\|\Delta y_{it}\right\|_p = E\left(\left|\Delta y_{it}\right|^p\right)^{1/p}$. By Assumption 2 $sup_{i,t}\left\|u_{it}\right\|_4 < C$, and for units with $|\phi_i| < c < 1$, we have $\left\|\phi_i^\ell\right\|_p = \left|\phi_i\right|^\ell < c^\ell$. Hence, conditional on $|\phi_i| < c < 1$, we have $\left\|\Delta y_{it}\right\|_4 \leq \frac{2C}{1-c} < \infty$. Also by Cauchy-Schwarz inequality $\left|E\left(\Delta y_{it}\Delta y_{i,t-h}\right)\right| \leq \left[E\left(\Delta y_{it}\right)^2 E\left(\Delta y_{i,t-h}\right)^2\right]^{1/2}$, and $\sup_i\left|E\left(\Delta y_{it}\Delta y_{i,t-h}\left||\phi_i|<c<1\right.\right)\right| < \infty$. In the unit root case

$$\begin{aligned} E\left(\Delta y_{it}\Delta y_{i,t-h}\left|\phi_i = 1\right.\right) &= E(\sigma_i^2) < C, \text{ for } h=0, \\ &= 0, \text{ for } h>0, \end{aligned}$$

and overall $\left|\gamma_\Delta(h)\right| < \infty$, for $h \leq T-2$. Existence of $\gamma_\Delta(h)$ when ϕ_i is distributed uniformly over the closed interval $[0,1]$ involves some algebra and is established in Section S.3 of the online supplement.

General expressions for the mean, variance and autocovariances of the first differences (covering unit root processes) are given in the following lemma and will be used in our subsequent analysis.

Lemma 1 *Consider the panel AR(1) model given by (2.2), and suppose that Assumptions 1-5 hold, and $M_i \to \infty$ for all units with $|\phi_i| < 1$. Then for all $i = 1, 2, ..., n$*

$$E\left(\Delta y_{it}\right) = 0, \quad E\left(\Delta y_{it}^2\right) = \sigma^2 E\left(\frac{2}{1+\phi_i}\right), \tag{3.7}$$

$$E\left(\Delta y_{it}\Delta y_{i,t-h}\right) = -\sigma^2 E\left[\left(\frac{1-\phi_i}{1+\phi_i}\right)\phi_i^{h-1}\right], \text{ for } h = 1, 2, ..., T-2, \tag{3.8}$$

and

$$E\left(\phi_i \Delta y_{it}\Delta y_{i,t-h}\right) = -\sigma^2 E\left[\left(\frac{1-\phi_i}{1+\phi_i}\right)\phi_i^{h}\right], \text{ for } h = 1, 2, ..., T-2. \tag{3.9}$$

A proof is provided in Section S.2 of the online supplement.

Remark 1 *Assumption 5 which in effect requires all processes $\{y_{it}, i = 1, 2, ..., n\}$ with $|\phi_i| <$ 1 are initialized from a distant past, could be restrictive. Although this assumption is required for our theoretical derivations, we do investigate the implications of relaxing it using Monte Carlo experiments. See sub-sections 7.5 and S.8.4 in the online supplement.*

Our identification and estimation strategy is based on matching sample estimates of autocorrelations of first differences (denoted as ρ_h) with first and higher order moments of ϕ_i. But before providing the details of our proposed estimators, we first show that the HomoGMM estimators of $E(\phi_i)$ that neglect heterogeneity of ϕ_i over i are biased even as $n \to \infty$, for any fixed T, and inferences based on them could be misleading. It is recognized that neglecting heterogeneity in dynamic panels can lead to biased estimates, but to the best of our knowledge, there is no formal analysis of the extent of the bias for short T panels. In the case of heterogeneous dynamic panels when both n and T are large, Pesaran and Smith (1995) provide expressions for asymptotic bias of fixed effects estimators.

4 Neglected heterogeneity bias

Under homogeneity where $\phi_i = \phi$ for all i, ϕ can be consistently estimated by the method of moments after eliminating α_i, for example by first differencing. We begin our analysis by showing the HomoGMM estimators are biased when ϕ_i are heterogeneous. The extent of the bias depends on the degree of heterogeneity. To simplify the exposition, without loss of generality, we consider the case where $T = 4$, the minimum value required for identification of $\mu_\phi = E(\phi_i)$ under heterogeneity established in Section 5. For the Anderson-Hsiao (AH) estimator, $\hat{\phi}_{AH} = \left(\sum_{i=1}^n \Delta y_{i4}\Delta y_{i2}\right) / \left(\sum_{i=1}^n \Delta y_{i3}\Delta y_{i2}\right)$, and using (3.1) for $t = 4$ we have

$$\hat{\phi}_{AH} = \frac{n^{-1}\sum_{i=1}^n \phi_i \Delta y_{i3}\Delta y_{i2}}{n^{-1}\sum_{i=1}^n \Delta y_{i3}\Delta y_{i2}} + \frac{n^{-1}\sum_{i=1}^n \Delta u_{i4}\Delta y_{i2}}{n^{-1}\sum_{i=1}^n \Delta y_{i3}\Delta y_{i2}}. \tag{4.1}$$

Since $E\left(\Delta u_{i4}\Delta y_{i2}\right) = 0$, then under Assumptions 1 to 5 and assuming $M_i \to \infty$ for units with $|\phi_i| < 1$, we have (as $n \to \infty$)

$$\hat{\phi}_{AH} \to_p \frac{\sum_{i=1}^n E\left(\phi_i \Delta y_{i3}\Delta y_{i2}\right)}{\sum_{i=1}^n E\left(\Delta y_{i3}\Delta y_{i2}\right)},$$

where $E\left(\Delta y_{i3}\Delta y_{i2}\right)$ and $E\left(\phi_i \Delta y_{i3}\Delta y_{i2}\right)$ are given by (3.8) and (3.9), respectively. Using these results

$$\hat{\phi}_{AH} \to_p \frac{E\left[\left(\frac{1-\phi_i}{1+\phi_i}\right)\phi_i\right]}{E\left[\left(\frac{1-\phi_i}{1+\phi_i}\right)\right]}. \tag{4.2}$$

In the homogeneous case ($\phi_i = \phi$), we have $\hat{\phi}_{AH} \to_p \mu_\phi = \phi$, as expected. Under heterogeneity, $\hat{\phi}_{AH}$ is clearly not a consistent estimator of $E(\phi_i)$. The extent of the asymptotic bias of the AH estimator depends on the distribution of ϕ_i. Exact expressions for the neglected heterogeneity bias of the AH estimator under uniform and categorical distributions are summarized in the following proposition.

Proposition 1 *Consider the Anderson-Hsiao estimator of μ_ϕ, $\hat{\phi}_{AH}$, given by (4.1), and suppose $\mu_\phi = E(\phi_i)$ in the heterogeneous panel AR(1) model given by (2.2). Suppose Assumptions 1–5 hold, and $M_i \to \infty$ for all i with $|\phi_i| < 1$. Then $\hat{\phi}_{AH}$ is asymptotically biased as an estimator of μ_ϕ. For $T = 4$, the asymptotic bias of the AH estimator is given by*

$$plim_{n\to\infty}\left(\hat{\phi}_{AH} - \mu_\phi\right) = \frac{2\left(1+\mu_\phi\right)\left[\frac{1}{1+\mu_\phi} - E\left(\frac{1}{1+\phi_i}\right)\right]}{E\left(\frac{1-\phi_i}{1+\phi_i}\right)}, \tag{4.3}$$

and $plim_{n\to\infty}\hat{\phi}_{AH} \leq \mu_\phi$*. The equality holds if and only if* $\phi_i = \phi = \mu_\phi$*, for all* i*.*

A proof is provided in Section S.4 of the online supplement.

The asymptotic biases of the AB and BB estimators under heterogeneous slopes are derived in Section S.5 of the online supplement. The magnitude of the asymptotic bias of AH, AB and BB estimators depends on the distribution of ϕ_i. For example, suppose that ϕ_i are random draws from a uniform distribution centered at $E(\phi_i) = \mu_\phi > 0$, with $\phi_i = \mu_\phi + v_i$, where $v_i \sim IIDU[-a, a]$, $a > 0$.[1] Then

$$plim_{n\to\infty}\left(\hat{\phi}_{AH} - \mu_\phi\right) = \frac{2(1+\mu_\phi)\left[\delta - \frac{1}{2}ln\left(\frac{1+\delta}{1-\delta}\right)\right]}{ln\left(\frac{1+\delta}{1-\delta}\right) - a}, \tag{4.4}$$

where $\delta = a/(1+\mu_\phi) \leq \left(1-\mu_\phi\right)/(1+\mu_\phi) < 1$. It is easily seen that $\hat{\phi}_{AH} - \mu_\phi \to 0$ with $a \to 0$. The magnitudes of the asymptotic bias of the AH estimator for $\mu_\phi \in \{0.4, 0.5\}$ and $a = 0.5$ are around -0.186 and -0.204, respectively, which are very close to the corresponding simulated bias in Tables S.8 and S.9 in the online supplement.

In the case where ϕ_i follows a categorical distribution, $\phi_i = \phi_L$ $(0 < \phi_L < 1)$ with probability π and $\phi_i = \phi_H > \phi_L$ with probability $1-\pi$, we have

$$plim_{n\to\infty}\left(\hat{\phi}_{AH} - \mu_\phi\right) = \frac{-2\pi(1-\pi)(\phi_H - \phi_L)^2}{\pi(1-\phi_L)(1+\phi_H) + (1-\pi)(1+\phi_L)(1-\phi_H)}.$$

As to be expected the asymptotic bias is negative, and its magnitude depends on the degree of dispersion of ϕ_i which is given by $Var(\phi_i) = \sigma_\phi^2\ \pi(1-\pi)(\phi_H - \phi_L)^2$. The unit root case arises for the units with $\phi_H = 1$.

Asymptotic bias, even if small, can lead to substantial size distortions when n is sufficiently large. See sub-section 7.3 for Monte Carlo evidence on the bias and size distortions of AH and other HomoGMM estimators.

5 Identification of moments of the AR coefficients

In this section, we formally establish conditions necessary for identification of $E(\phi_i^s)$ without making any specific distributional assumptions on ϕ_i. Suppose Assumptions 1 to 5 hold.

[1] To ensure that $|\phi_i| \leq 1$ we also require that $a \leq 1 - \mu_\phi$.

We consider the minimum number of periods needed to consistently estimate $E(\phi_i^s)$, for $s = 1, 2, ..., S$. Denote the h^{th}-order autocorrelation coefficients of Δy_{it} as ρ_h given by

$$\rho_h = \frac{E\left(\Delta y_{it} \Delta y_{i,t-h}\right)}{E\left[\left(\Delta y_{it}\right)^2\right]}, \tag{5.1}$$

for $h = 1, 2, ...,$ with $|\rho_h| \leq 1$. Since by assumption ϕ_i and σ_i^2 are independently distributed (see part (b) of Assumption 3), then using the results in Lemma 1 we have

$$\rho_h = \frac{E\left(\Delta y_{it} \Delta y_{i,t-h}\right)}{E\left[\left(\Delta y_{it}\right)^2\right]} = -\frac{E\left[\left(\frac{1-\phi_i}{1+\phi_i}\right) \phi_i^{h-1}\right]}{2E\left(\frac{1}{1+\phi_i}\right)}, \tag{5.2}$$

for $h = 1, 2, ...,$ with $|\rho_h| \leq 1$.

Suppose that ρ_h can be consistently estimated by the moment estimators of $E\left(\Delta y_{it} \Delta y_{i,t-h}\right)$ and $E\left[\left(\Delta y_{it}\right)^2\right]$. Then the identification condition of $E(\phi_i^s)$ can be derived by the system of equations in (5.2). For $h = 1$, $2E\left(\frac{1}{1+\phi_i}\right)\rho_1 = -E\left(\frac{1-\phi_i}{1+\phi_i}\right) = 1 - 2E\left(\frac{1}{1+\phi_i}\right)$, which can be equivalently written as $2E\left(\frac{1}{1+\phi_i}\right) = \frac{1}{1+\rho_1}$. Using this result and noting that for $h = 2$, $2E\left(\frac{1}{1+\phi_i}\right)\rho_2 = -2 + E\left(\phi_i\right) + 2E\left(\frac{1}{1+\phi_i}\right)$, we have

$$E\left(\phi_i\right) = \frac{1 + 2\rho_1 + \rho_2}{1 + \rho_1}. \tag{5.3}$$

Similarly, for $h = 3$ we have $2E\left(\frac{1}{1+\phi_i}\right)\rho_3 = -E\left(2\phi_i - 2 - \phi_i^2 + \frac{2}{1+\phi_i}\right)$, which yields

$$E\left(\phi_i^2\right) = \frac{1 + 2\rho_1 + 2\rho_2 + \rho_3}{1 + \rho_1}. \tag{5.4}$$

For $h = 4$, $2E\left(\frac{1}{1+\phi_i}\right)\rho_4 = -E\left(2\phi_i^2 - \phi_i^3 - 2\phi_i + 2 - \frac{2}{1+\phi_i}\right)$, and upon using the results of the lower-order moments we obtain

$$E\left(\phi_i^3\right) = \frac{1 + 2\rho_1 + 2\rho_2 + 2\rho_3 + \rho_4}{1 + \rho_1}. \tag{5.5}$$

Higher-order moments of ϕ_i can be obtained similarly. To identify the s^{th} order moment of ϕ_i requires consistent estimation of ρ_h for $h = 1, 2, ..., s + 1$. In general, we must have $T \geq s + 3$, as $n \to \infty$ to identify $E\left(\phi_i^s\right)$.

Remark 2 *Note that under homogeneity where $\phi_i = \phi$ for all i, using (5.2) we have*

$$\rho_h = \frac{E\left(\Delta y_{it} \Delta y_{i,t-h}\right)}{E\left[\left(\Delta y_{it}\right)^2\right]} = -\frac{1}{2}\phi^{h-1}\left(1 - \phi\right), \text{ for } h = 1, 2, ..., T - 2. \tag{5.6}$$

For $h = 1$ under homogeneity, $\rho_1 = -(1-\phi)/2$ and ϕ can be estimated by $\hat{\phi}_{Homo} = 1+2\hat{\rho}_{1,nT}$. In this case for identification of ϕ, we need $T \geq 2$. This result also follows if we let $\rho_h = \phi\rho_{h-1}$ in (5.3) $E(\phi_i) = \phi = 1 + \frac{\rho_1+\phi\rho_1}{1+\rho_1}$, which is satisfied when $\rho_1 = -(1-\phi)/2$.

6 Estimation of the moments of the AR coefficients

We now turn our attention to consistent estimation of the moments of ϕ_i, namely $\theta_s = E(\phi_i^s)$, for $s = 1, 2$, and 3. We consider a simple moment estimator which we refer to as the first differenced autocorrelation (FDAC) estimator, and a GMM-type estimator that we refer to as HetroGMM to be distinguished from the GMM estimators proposed in the literature for estimation of the homogeneous AR coefficient assuming $\phi_i = \phi$ for all i.

6.1 First differenced autocorrelation (FDAC) estimator

The FDAC estimator uses the sample analogs of autocorrelations of the first differences, ρ_h given by (5.1), in equations (5.3), (5.4) and (5.5) to obtain consistent estimators of $\theta_s = E(\phi_i^s)$ for $s = 1, 2$ and 3, respectively. Specifically, using $\{\Delta y_{it},\ t = 2, 3, ..., T; i = 1, 2, ..., n\}$, ρ_h can be consistently estimated by

$$\hat{\rho}_{h,nT} = \frac{(T-h-1)^{-1}\sum_{t=h+2}^{T}\left[n^{-1}\sum_{i=1}^{n}\Delta y_{it}\Delta y_{i,t-h}\right]}{(T-1)^{-1}\sum_{t=2}^{T}\left[n^{-1}\sum_{i=1}^{n}\left(\Delta y_{it}\right)^2\right]}, \text{ for } h = 1, 2, ..., T-2. \tag{6.1}$$

Then plugging these estimators in (5.3)–(5.5) we have the following FDAC estimators

$$\hat{\theta}_{1,FDAC} = \widehat{E(\phi_i)} = \frac{1 + 2\hat{\rho}_{1,nT} + \hat{\rho}_{2,nT}}{1+\hat{\rho}_{1,nT}}, \text{ for } T \geq 4, \tag{6.2}$$

$$\hat{\theta}_{2,FDAC} = \widehat{E\left(\phi_i^2\right)} = \frac{1 + 2\hat{\rho}_{1,nT} + 2\hat{\rho}_{2,nT} + \hat{\rho}_{3,nT}}{1+\hat{\rho}_{1,nT}}, \text{ for } T \geq 5, \tag{6.3}$$

and

$$\hat{\theta}_{3,FDAC} = \widehat{E\left(\phi_i^3\right)} = \frac{1 + 2\hat{\rho}_{1,nT} + 2\hat{\rho}_{2,nT} + 2\hat{\rho}_{3,nT} + \hat{\rho}_{4,nT}}{1+\hat{\rho}_{1,nT}}, \text{ for } T \geq 6. \tag{6.4}$$

These estimators can also be viewed as moment estimators that place equal weights on the cross-section averages, $n^{-1}\sum_{i=1}^{n}\Delta y_{it}\Delta y_{i,t-h}$, for different t. This makes sense since under our assumptions for each t, $\Delta y_{it}\Delta y_{i,t-h}$ are cross-sectionally independent with finite second-order moments, and by the law of large numbers $n^{-1}\sum_{i=1}^{n}\Delta y_{it}\Delta y_{i,t-h} \to_p E(\Delta y_{it}\Delta y_{i,t-h})$, and hence $\hat{\rho}_{h,nT} \to_p E\left(\Delta y_{it}\Delta y_{i,t-h}\right)/E\left(\Delta y_{it}\right)^2 = \rho_h$ as $n \to \infty$. Using this result and noting that $1 + \hat{\rho}_{1,nT} \to_p 1 + \rho_1 > 0$, it then readily follows that $\hat{\theta}_{1,FDAC} \to \frac{1+2\rho_1+\rho_2}{1+\rho_1} = \theta_1 = E(\phi_i)$. Similarly, $\hat{\theta}_{s,FDAC} \to_p E(\phi_i^s)$, for $s = 2$ and 3. Since $\Delta y_{it}\Delta y_{i,t-h}$ for $h = 1, 2, ..., T-2$ have second order moments, it also follows that the convergence of $\hat{\theta}_{s,FDAC}$ to θ_s is in the mean squared error sense which is stronger than convergence in probability.

6.2 Generalized method of moments estimator based on autocovariances

The FDAC estimator is a plug-in type estimator and needs not be efficient. An alternative and arguably more efficient approach would be to base the estimation of θ_s directly on the sample moments of $E\left(\Delta y_{it}\Delta y_{i,t-h}\right)$ and then use standard results from the GMM literature to obtain asymptotically optimum weighted moment conditions rather than equally weighted moments which might not be efficient. In practice, the differences between the two approaches could depend on the degree of heterogeneity and the sampling uncertainty associated with the GMM weights. The relative performance of FDAC and heterogeneous GMM estimators of θ_s will be investigated by Monte Carlo simulations.

6.2.1 Heterogeneous generalized method of moments (HetroGMM) estimator of $E(\phi_i)$

Given (5.1), the moment condition (5.3) can be written equivalently as

$$\theta_1[E\left(\Delta y_{it}\right)^2 + E\left(\Delta y_{it}\Delta y_{i,t-1}\right)] = E\left(\Delta y_{it}\right)^2 + 2E\left(\Delta y_{it}\Delta y_{i,t-1}\right) + E\left(\Delta y_{it}\Delta y_{i,t-2}\right), \quad (6.5)$$

which yields $T-3$ moment conditions for $t = 4, 5, ..., T$, requiring that $T \geq 4$. These moment conditions can be written more compactly as

$$E\left[M_{nt}(\theta_{1,0})\right] = 0, \text{ for } t = 4, 5, ..., T, \quad (6.6)$$

where $M_{nt}(\theta_{1,0}) = n^{-1}\sum_{i=1}^{n} m_{it}(\theta_{1,0})$, $m_{it}(\theta_{1,0}) = \theta_1 h_{it} - g_{it}$,

$$h_{it} = \left(\Delta y_{it}\right)^2 + \Delta y_{it}\Delta y_{i,t-1}, \text{ and } g_{it} = \left(\Delta y_{it}\right)^2 + 2\Delta y_{it}\Delta y_{i,t-1} + \Delta y_{it}\Delta y_{i,t-2}. \quad (6.7)$$

To optimally combine the moment conditions in (6.6) set

$$\mathbf{h}_{iT} = (h_{i4}, h_{i5}, ..., h_{iT})', \tag{6.8}$$

$$\text{and } \mathbf{g}_{iT} = (g_{i4}, g_{i5}, ..., g_{iT}). \tag{6.9}$$

Then $\mathbf{M}_{nT}(\theta_1) = (m_{n,4}(\theta_1), m_{n,5}(\theta_1), ..., m_{n,T}(\theta_1))' = \mathbf{G}_{nT} - \mathbf{H}_{nT}\theta_1$, where $\mathbf{G}_{nT} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_{iT}$ and $\mathbf{H}_{nT} = n^{-1}\sum_{i=1}^{n}\mathbf{h}_{iT}$. Using (6.5), it readily follows that $E\left[\mathbf{M}_{nT}(\theta_{1,0})\right] = \mathbf{0}$. The HetroGMM estimator of θ_1 is given by

$$\hat{\theta}_{1,HetroGMM} = \text{argmin}_{\theta_1}\left(\mathbf{G}_{nT} - \mathbf{H}_{nT}\theta_1\right)' \mathbf{A}_{nT}\left(\mathbf{G}_{nT} - \theta_1\mathbf{H}_{nT}\right),$$

where $\mathbf{A}_{nT}$ is a $(T-3)\times(T-3)$ positive definite stochastic weight matrix, and for any $T \geq 4$, it tends to a non-stochastic positive definite matrix $\mathbf{A}_T$ as $n \to \infty$. The most efficient HetroGMM estimator is given by

$$\hat{\theta}_{1,HetroGMM}\left(\mathbf{A}_T^*\right) = \left(\mathbf{H}_{nT}'\mathbf{A}_T^*\mathbf{H}_{nT}\right)^{-1}\mathbf{H}_{nT}'\mathbf{A}_T^*\mathbf{G}_{nT}, \tag{6.10}$$

where $\mathbf{A}_T^* = \mathbf{S}_T^{-1}(\theta_1)$ is the optimal weight matrix with

$$\mathbf{S}_T(\theta_1) = Var\left(\sqrt{n}\mathbf{M}_{nT}(\theta_1)\right) = nVar\left(\mathbf{G}_{nT} - \theta_1\mathbf{H}_{nT}\right) = nVar\left[n^{-1}\sum_{i=1}^{n}\left(\mathbf{g}_{iT} - \theta_1\mathbf{h}_{iT}\right)\right].$$

Given (6.5), $E\left(\mathbf{g}_{iT} - \theta_{1,0}\mathbf{h}_{iT}\right) = \mathbf{0}$, and $\mathbf{g}_{iT} - \theta_{1,0}\mathbf{h}_{iT}$ are cross-sectionally independent, then

$$\mathbf{S}_T(\theta_{1,0}) = \frac{1}{n}\sum_{i=1}^{n}E\left[\left(\mathbf{g}_{iT} - \theta_{1,0}\mathbf{h}_{iT}\right)\left(\mathbf{g}_{iT} - \theta_{1,0}\mathbf{h}_{iT}\right)'\right]. \tag{6.11}$$

It is difficult to derive an analytical expression for $\mathbf{S}_T(\theta_{1,0})$, but for a given value of θ_1, $\mathbf{S}_T(\theta_1)$ can be consistently estimated by its sample mean given by

$$\hat{\mathbf{S}}_T\left(\theta_1\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{g}_{iT} - \theta_1\mathbf{h}_{iT}\right)\left(\mathbf{g}_{iT} - \theta_1\mathbf{h}_{iT}\right)', \text{ for } n > T-3. \tag{6.12}$$

A standard two-step GMM estimator of θ_1 can now be obtained using $\hat{\theta}_{1,FDAC}$ given by (6.2) as an initial estimate to consistently estimate the optimal weight matrix, $\mathbf{S}_T^{-1}(\theta_{1,0})$, in the first step. Substituting $\hat{\theta}_{1,FDAC}$ into (6.12) yields the following two-step HetroGMM estimator

$$\hat{\theta}_{1,HetroGMM} = \left[\mathbf{H}_{nT}'\hat{\mathbf{S}}_T^{-1}\left(\hat{\theta}_{1,FDAC}\right)\mathbf{H}_{nT}\right]^{-1}\left[\mathbf{H}_{nT}'\hat{\mathbf{S}}_T^{-1}\left(\hat{\theta}_{1,FDAC}\right)\mathbf{G}_{nT}\right], \tag{6.13}$$

where
$$\hat{\mathbf{S}}_T\left(\hat{\theta}_{1,FDAC}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{g}_{iT} - \hat{\theta}_{1,FDAC}\mathbf{h}_{iT}\right)\left(\mathbf{g}_{iT} - \hat{\theta}_{1,FDAC}\mathbf{h}_{iT}\right)'. \tag{6.14}$$

It is also possible to obtain an iterated version of the above, where $\hat{\theta}_{1,HetroGMM}$ is used to obtain a new estimate of $\hat{\mathbf{S}}_T(\theta_1)$, namely $\hat{\mathbf{S}}_T(\hat{\theta}_{1,HetroGMM})$, and so on. But there seems little gain in doing so since $\hat{\theta}_{1,HetroGMM}$ is asymptotically efficient.

The above results are summarized in the following theorem.

Theorem 1 *Consider the panel AR(1) model given by (2.2) and suppose that Assumptions 1–5 hold, $T \geq 4$, and $M_i \to \infty$ for all i with $|\phi_i| < 1$. Then the HetroGMM estimator of $\theta_1 = E(\phi_i)$ given by (6.13) is asymptotically efficient. The asymptotic distribution of $\hat{\theta}_{1,HetroGMM}$ is given by*
$$\sqrt{n}\left(\hat{\theta}_{1,HetroGMM} - \theta_{1,0}\right) \to_d N\left(0, V_{\theta_1}\right), \tag{6.15}$$
where $\theta_{1,0}$ is the true value of θ_1, $V_{\theta_1}^{-1} = plim_{n\to\infty}\left(\mathbf{H}'_{nT}\mathbf{S}_T^{-1}(\theta_{1,0})\mathbf{H}_{nT}\right)$, $\mathbf{H}_{nT} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{h}_{iT}$, and $\mathbf{S}_T^{-1}(\theta_{1,0})$ and $\mathbf{h}_{iT}$ are defined by (6.11) and (6.8), respectively. The asymptotic variance of $\hat{\theta}_{1,HetroGMM}$ can be estimated consistently by $n^{-1}\left[\mathbf{H}'_{nT}\hat{\mathbf{S}}_T^{-1}\left(\hat{\theta}_{1,FDAC}\right)\mathbf{H}_{nT}\right]^{-1}$, where $\hat{\mathbf{S}}_T\left(\hat{\theta}_{1,FDAC}\right)$ is given by (6.14).

Our use of the FDAC estimator as an initial estimator for the two-step GMM estimator is based on the observations that FDAC exploits the stationarity properties of moments in the first differences and is based on more information as compared to the first step GMM estimator. As an example, consider the exact identified case when $T = 4$. Then (see (6.7))
$$\hat{\theta}_{1,HetroGMM} = \frac{n^{-1}\sum_{i=1}^{n}(\Delta y_{i4})^2 + 2n^{-1}\sum_{i=1}^{n}\Delta y_{i4}\Delta y_{i,3} + n^{-1}\sum_{i=1}^{n}\Delta y_{i4}\Delta y_{i,2}}{n^{-1}\sum_{i=1}^{n}(\Delta y_{i4})^2 + n^{-1}\sum_{i=1}^{n}\Delta y_{i4}\Delta y_{i,3}},$$
as compared to $\hat{\theta}_{1,FDAC}$ given by (6.2) which can be written equivalently
$$\hat{\theta}_{1,FDAC} = \frac{\left(\frac{1}{3}\right)\sum_{t=2}^{4}\left[\frac{1}{n}\sum_{i=1}^{n}(\Delta y_{it})^2\right] + \sum_{t=3}^{4}\left[\frac{1}{n}\sum_{i=1}^{n}\Delta y_{it}\Delta y_{i,t-1}\right] + \left[\frac{1}{n}\sum_{i=1}^{n}\Delta y_{i4}\Delta y_{i,4-2}\right]}{\left(\frac{1}{3}\right)\sum_{t=2}^{4}\left[\frac{1}{n}\sum_{i=1}^{n}(\Delta y_{it})^2\right] + \left(\frac{1}{2}\right)\sum_{t=3}^{4}\left[\frac{1}{n}\sum_{i=1}^{n}\Delta y_{it}\Delta y_{i,t-1}\right]}.$$

Both estimators converge to $\theta_{1,0}$ at the rate of $\sqrt{n}$, but $\hat{\theta}_{1,FDAC}$ exploits the stationary properties of the $(\Delta y_{it})^2$ and $\Delta y_{it}\Delta y_{i,t-1}$ more effectively. Specifically, $n^{-1}\sum_{i=1}^{n}(\Delta y_{i4})^2$ and $(1/3)\sum_{t=2}^{4}\left[n^{-1}\sum_{i=1}^{n}(\Delta y_{it})^2\right]$ converge to the same limit, but the latter makes use of $(\Delta y_{i2})^2$ and $(\Delta y_{i3})^2$ obervations as well as $(\Delta y_{i4})^2$. Similarly, $2n^{-1}\sum_{i=1}^{n}\Delta y_{i4}\Delta y_{i3}$ and $\sum_{t=3}^{4}\left[n^{-1}\sum_{i=1}^{n}\Delta y_{it}\Delta y_{i,t-1}\right]$ converge to the same limit, but the latter makes use of $\Delta y_{i3}\Delta y_{i2}$ in addition to $\Delta y_{i4}\Delta y_{i3}$.

Remark 3 *Both FDAC and HetroGMM estimators should work fine asymptotically under $E(u_{it}^2) = \sigma_{it}^2$, so long as the time variations of σ_{it}^2 is stationary, in a sense that $E(\sigma_{it}^2) = \sigma_i^2$. One important example is when u_{it} has a stationary GARCH specification. This property is illustrated in the Monte Carlo simulations where we consider the properties of the proposed estimators with and without GARCH effects.*

6.2.2 Generalized method of moments estimator of $E(\phi_i^2)$

Similarly, the HetroGMM estimator of $\theta_2 = E\left(\phi_i^2\right)$ can be obtained based on the equation below for $t = 5, 6, ..., T$,

$$\begin{aligned} &\theta_2\left[E\left[(\Delta y_{it})^2\right] + E\left(\Delta y_{it}\Delta y_{i,t-1}\right)\right] \\ =&E\left[(\Delta y_{it})^2\right] + 2E\left(\Delta y_{it}\Delta y_{i,t-1}\right) + 2E\left(\Delta y_{it}\Delta y_{i,t-2}\right) + E\left(\Delta y_{it}\Delta y_{i,t-3}\right). \end{aligned} \tag{6.16}$$

Let

$$\mathbf{h}_{2,iT} = (h_{2,i5}, h_{2,i6}, ..., h_{2,iT})' \tag{6.17}$$

$$\text{and } \mathbf{g}_{2,iT} = (g_{2,i5}, g_{2,i6}, ..., g_{2,iT})' \tag{6.18}$$

with $h_{2,it} = (\Delta y_{it})^2 + \Delta y_{it}\Delta y_{i,t-1}$ and $g_{2,it} = (\Delta y_{it})^2 + 2\Delta y_{it}\Delta y_{i,t-1} + 2\Delta y_{it}\Delta y_{i,t-2} + \Delta y_{it}\Delta y_{i,t-3}$. Denote $\mathbf{G}_{2,nT} = n^{-1}\sum_{i=1}^{n}\mathbf{g}_{2,iT}$, and $\mathbf{H}_{2,nT} = n^{-1}\sum_{i=1}^{n}\mathbf{h}_{2,iT}$, where $\mathbf{G}_{2,nT}$ and $\mathbf{H}_{2,nT}$ are $(T-4)\times 1$ vectors (with $T > 4$). Then, the two-step HetroGMM estimator of the second moment can be derived as

$$\hat{\theta}_{2,HetroGMM} = \left[\mathbf{H}_{2,nT}'\hat{\mathbf{S}}_{2,T}^{-1}\left(\hat{\theta}_{2,FDAC}\right)\mathbf{H}_{2,nT}\right]^{-1}\left[\mathbf{H}_{2,nT}'\hat{\mathbf{S}}_{2T}^{-1}\left(\hat{\theta}_{2,FDAC}\right)\mathbf{G}_{2,nT}\right], \tag{6.19}$$

where the initial estimator can be the FDAC estimator of θ_2 given by equation (6.3), and $\hat{\mathbf{S}}_{2,T}(\theta_2) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{g}_{2,iT} - \theta_2\mathbf{h}_{2,iT}\right)\left(\mathbf{g}_{2,iT} - \theta_2\mathbf{h}_{2,iT}\right)'$. Finally, the asymptotic distribution of $\hat{\theta}_{2,HetroGMM}$ is given by

$$\sqrt{n}\left(\hat{\theta}_{2,HetroGMM} - \theta_{2,0}\right) \rightarrow_d N\left(0, V_{\theta_2}\right), \tag{6.20}$$

where $\theta_{2,0}$ is the true value of θ_2, and V_{θ_2} can be consistently estimated by

$$\hat{V}_{\theta_2} = \left[\mathbf{H}'_{2,nT}\hat{\mathbf{S}}^{-1}_{2,T}\left(\hat{\theta}_{2,HetroGMM}\right)\mathbf{H}_{2,nT}\right]^{-1}. \tag{6.21}$$

6.3 Plug-in estimator of σ^2_ϕ

Consider now the estimation of $\sigma^2_\phi = Var(\phi_i)$, and recall that in terms of $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ we have $\sigma^2_\phi = \theta_2 - \theta_1^2$. Therefore, a plug-in estimator of σ^2_ϕ is given by

$$\hat{\sigma}^2_\phi = \hat{\theta}_2 - \left(\hat{\theta}_1\right)^2, \tag{6.22}$$

which is an asymptotically valid estimator of σ^2_ϕ if $\hat{\theta}_2 - \left(\hat{\theta}_1\right)^2 > 0$. This condition will be met for n sufficiently large, noting that $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2\right)'$ is a consistent estimator of $\boldsymbol{\theta}_0 = (\theta_{1,0}, \theta_{2,0})'$. The asymptotic distribution of $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)'$, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to_d N(0, \mathbf{V}_{\boldsymbol{\theta}})$ is derived in Section S.6 of the online supplement. Then using the Delta method it follows that $\sqrt{n}\left(\hat{\sigma}^2_\phi - \sigma^2_{\phi,0}\right) \to_d N(0, V_{\sigma^2})$, where $\sigma^2_{\phi,0} = \theta_{2,0} - \theta^2_{10}$ denotes the true value of σ^2_ϕ, and $V_{\sigma^2} = (-2\theta_{1,0}, 1)\mathbf{V}_{\boldsymbol{\theta}}(-2\theta_{1,0}, 1)'$. V_{σ^2} can be consistently estimated by $\hat{V}_\sigma = \left(-2\hat{\theta}_1, 1\right)\hat{\mathbf{V}}_{\boldsymbol{\theta}}\left(-2\hat{\theta}_1, 1\right)'$, where $\hat{\mathbf{V}}_{\boldsymbol{\theta}}$ is a consistent estimator of $\mathbf{V}_{\boldsymbol{\theta}}$ given by (S.13) in the online supplement. However, it is important to bear in mind that the asymptotic distribution of $\hat{\sigma}^2_\phi$ is valid only in the locality of the true value of σ^2_ϕ, and only if this true value is sufficiently away from the boundary value of 0. In practice, we recommend using the plug-in estimator of σ^2_ϕ only when n is large, in excess of $1,000$, judging by the Monte Carlo evidence to be discussed below.

7 Monte Carlo experiments

7.1 Data generating process of Monte Carlo experiments

For each $i = 1, 2, ..., n$, the process $\{y_{it}\}$ is generated starting at time $t = -M_i + 1$, with the initial value $y_{i,-M_i}$ using

$$y_{it} = \mu_i(1 - \phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}, \text{ for } t = -M_i + 1, -M_i + 2, ..., 0, 1, 2, ..., T. \tag{7.1}$$

We experiment with two distributions to generate $\phi_i \in (-1, 1]$: (a) uniform and (b) categorial. Under the former we set $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$. To distinguish between cases when $|\phi_i| < 1$ for all i and when $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ with $\phi_i = 1$ for some i, we fix $a = 0.5$ and consider the values of $\mu_\phi = 0.4$ and 0.5, with $E(\phi_i) = \mu_\phi$ and $\sigma_\phi^2 = a^2/3 = 0.083$. Under case (b), we generate $\phi_i = \phi_H$ (high) and $\phi_i = \phi_L$ (low) with probabilities $1-\pi$ and π, respectively. Two sets of parameter values for (ϕ_H, ϕ_L, π) are considered: $(0.8, 0.5, 0.85)$ with $|\phi_i| < 1$ for all i, and $(1, 0.5, 0.95)$ with $\phi_i \in (-1, 1]$ for all i. Then $\mu_\phi = E(\phi_i) = \phi_L\pi + \phi_H(1-\pi) = 0.545$ and 0.525, and $\sigma_\phi^2 = \left[\phi_L^2\pi + \phi_H^2(1-\pi)\right] - \mu_\phi^2 = 0.011$ and 0.012, respectively. The individual-specific means of $\{y_{it}\}$ are generated as $\mu_i = \phi_i + \eta_i$ with $\eta_i \sim IIDN(0, 1)$, allowing for a non-zero correlation between μ_i and ϕ_i.

We consider two choices when generating ε_{it}: Gaussian $\varepsilon_{it} \sim IIDN(0, 1)$, and non-Gaussian $\varepsilon_{it} = (e_{it} - 2)/2$, with $e_{it} \sim IID\chi_2^2$, where χ_2^2 is a chi-squared variate with two degrees of freedom, for all i and t. $\{h_{it}\}$ is generated as a $GARCH(1,1)$ process, namely $h_{it}^2 = \sigma_i^2(1 - \psi_0 - \psi_1) + \psi_0 h_{i,t-1}^2 + \psi_1(h_{i,t-1}\varepsilon_{i,t-1})^2$, with $\sigma_i^2 \sim IID\left(0.5 + 0.5z_i^2\right)$ and $z_i \sim IIDN(0, 1)$. We set $\psi_0 = 0.6$ and $\psi_1 = 0.2$, with the initial values $h_{i,-M_i} = \sigma_i$.[2] The case where errors are conditionally homoskedastic over time is obtained as a special case by setting $\psi_0 = \psi_1 = 0$.

We generate the initial values of $\{y_{it}\}$ as $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$ for all i. Again the choice of M_i is set depending on whether $|\phi_i| < 1$ or $\phi_i = 1$. For the former case, we set $M_i = 100$, which applies to all the units when ϕ_i is uniformly distributed as it is not known which $\phi_i = 1$, and units with $\phi_i < 1$ in the case of categorical-distributed ϕ_i. For draws with $\phi_i = 1$ in the categorical distribution, we set $M_i = 1$ such that y_{it} for $t = 1, 2, ..., T$ has finite moments as T is fixed in our design.

To check the robustness of the results to non-stationary initialization for $|\phi_i| < 1$, when the processes start from a finite date in the past, we conduct two sets of experiments, one set with $M_i = 1$, and another set with $M_i = 3$ for all i.

The estimation of the moments of ϕ_i, $\mu_\phi = E(\phi_i)$ and $\sigma_\phi^2 = Var(\phi_i)$, are based on

[2]Our approach also allows the coefficients of the $GARCH(1,1)$ model to be heterogeneous across i, so long as they are drawn from the same common distribution. But to keep the MC design simple, we are only reporting for the case where ψ_0 and ψ_1 are homogeneous.

$\{y_{it}^{(r)}$, for $i = 1, 2, ..., n; t = 1, 2, ..., T\}$, where r denotes the r^{th} replication of DGP in (7.1). We carry out 2,000 replications for the experiments that compare the small sample performances of FDAC, HetroGMM, and a number of estimators proposed in the literature for the homogeneous slope case (denoted by HomoGMM), specifically, the estimators proposed by Anderson and Hsiao (1981, 1982) (AH), Arellano and Bond (1991) (AB), Blundell and Bond (1998) (BB), and the augmented Anderson-Hsiao (AAH) estimator proposed by Chudik and Pesaran (2021), as well as the FDLS estimator due to Han and Phillips (2010).[3] For experiments that compare our proposed estimator with the MSW estimator in Mavroeidis et al. (2015), we use 1,000 replications as it takes a substantial amount of time to compute the MSW estimator.[4] To save space, the tables summarize the results of the MC experiments are all included in the online supplement.

7.2 Comparison of FDAC and HetroGMM estimators

7.2.1 MC results for estimation of μ_ϕ

Bias, root mean square errors (RMSE), and size of tests of FDAC and HetroGMM estimators of $\mu_\phi = E(\phi_i)$ with uniformly distributed ϕ_i are summarized in Table S.1 in the online supplement. The results with categorically distributed ϕ_i are shown in Table S.2 in the online supplement. These tables provide results for the sample size combinations $T = 4, 5, 6, 10$ and $n = 100, 1000, 5000$, in the case of Gaussian errors without GARCH effects. The parameters of distributions are chosen to distinguish between cases where $|\phi_i| < 1$ and $\phi_i \in (-1, 1]$, with the related results displayed in the left and right panels of the tables, respectively.

In line with our theoretical results, both FDAC and HetroGMM estimators offer reliable estimates for μ_ϕ in the case of heterogeneous short T panels under both uniform and categorical distributions. The categorical distribution yields marginally lower RMSEs, which

[3] We have downloaded the codes of the AH, AB, BB, and AAH estimators from the supplementary materials of Chudik and Pesaran (2021) using the link: `https://www.econ.cam.ac.uk/people-files/emeritus/mhp1/fp21/CP_AAH_paper_July_2021_codes_and_data.zip`. We are grateful to Alexander Chudik for making the codes publicly available.

[4] We have downloaded the codes of the MSW estimator used in empirical applications from the supplementary materials of Mavroeidis et al. (2015) using the link: `https://drive.google.com/file/d/1hdRFpcWo3r88YV_5Kc4Our-siCYGSBDN/view?usp=sharing`. We are grateful to Yuya Sasaki for also sharing the codes of the MSW estimator used in their Monte Carlo experiments by private correspondence.

is largely due to the fact that σ_ϕ^2 is much smaller under the categorical distribution around 0.012, as compared to 0.083 under the uniform distribution. More importantly, the magnitudes of bias, RMSE, and size are very similar irrespective of whether $|\phi_i| < 1$ or $\phi_i \in (-1, 1]$. This result holds even if a fixed proportion of units have unit roots, as is the case with the categorical distribution where $\phi_i = 1$ in the case of 5 per cent of all units in the sample. The empirical power functions for FDAC and HetroGMM estimators of μ_ϕ are displayed in Figure S.1 of the online supplement for the uniformly distributed AR coefficients with $\phi_i \in (-1, 1]$ in the baseline case (Gaussian errors and no GARCH effects). The power functions for the other experiments are very similar and can be obtained from the authors upon request.

Compared with the HetroGMM estimator, the FDAC estimator has uniformly smaller biases across all sample size combinations, lower RMSE, and greater power for $T = 4, 5, 6$, and $n = 100, 1000$, and 5000. The differences between the two estimators of μ_ϕ become negligible only when $T = 10$. In the light of our discussion in sub-section 6.2, this could be because the FDAC estimator uses averages of the individual sample moments both over time and across all units given the stationary properties of the autocovariances of the first differences, and thus it is not subject to the many moment problem that could adversely impact the HetroGMM estimator. Consequently, tests based on the FDAC estimator are not adversely affected as T is increased with n small, and its size is mostly around the nominal size of five per cent. However, tests based on the HetroGMM estimator tend to over-reject slightly as T is increased when n is relatively small ($n = 100$). For example, for the uniform distribution with $\phi_i \in (-1, 1]$ and $n = 100$, the size of the tests of $\mu_\phi = 0.5$ based on the HetroGMM estimator rises from 5.7 to 10.5 per cent when T is increased from 4 to 10. These findings are in line with the results obtained in the literature when GMM is applied to homogeneous dynamic panels.[5]

As can be seen from the empirical power functions in Figure S.1, the tests based on FDAC and HetroGMM estimators can not reject $\mu_\phi = 1$ with 100 per cent certainty due to the small

[5]For GMM estimators with many moment conditions, some of the moment conditions can be weak. The small-sample bias associated with the weak moments will result in substantial size distortions, which become more severe with greater weights on the weak moments. See also Section 6 of Chudik and Pesaran (2021).

sample sizes with $n = 100$. But as n and T increase, the empirical power functions become steeper, illustrating an enhanced ability to discern deviations from the null hypothesis.

7.2.2 MC results for estimation of σ_ϕ^2

As discussed in sub-section 6.3, the FDAC and HetroGMM estimators of σ_ϕ^2 are consistent so long as the true value of σ_ϕ^2, namely $\theta_{2,0} - \theta_{0,1}^2$ is not too close to the boundary value of zero. Also to avoid negative estimates of the plug-in estimator of σ_ϕ^2 given by (6.22) we need n to be sufficiently large. Table S.3 in the online supplement summarizes the number of replications, out of $2,000$, with negative or close to zero estimates (defined as estimates below 0.0001) for the baseline experiments and sample size combinations $n = 100, 1000, 2500, 5000$ and $T = 5, 6, 10$. The frequencies of the HetroGMM estimator are noticeably higher than those of the FDAC estimator for small T and n. When $n = 100$, a sizeable proportion of the estimates of σ_ϕ^2 are negative, suggesting that $n = 100$ is not sufficiently large for the asymptotic properties to hold. However, as to be expected, the number of negative estimates declines rapidly as n and T are increased. Accordingly, we only focus on samples with $n \geq 1000$, and report the bias and RMSE of the estimates of σ_ϕ^2 for sample size combinations $n = 1000, 2500, 5000$ and $T = 5, 6, 10$. The results for the positive estimates are summarized in Table S.4 of the online supplement for uniformly distributed ϕ_i. For these sample size combinations, we only encounter very few negative estimates and none when $n = 5000$ and $T \geq 6$.[6]

Overall, both FDAC and HetroGMM estimators of σ_ϕ^2 perform well when $T = 10$ or n is large, with comparable performances whether $|\phi| < 1$ or $\phi_i \in (-1, 1]$. However, the FDAC estimator performs much better for smaller values of T and n, as can be seen from the larger bias and RMSE of the HetroGMM estimator.

The empirical power functions for FDAC and HetroGMM estimators of σ_ϕ^2 are shown in Figure S.2 of the online supplement for the uniformly distributed AR coefficients with $\phi_i \in (-1, 1]$ in the baseline case (Gaussian errors and no GARCH effects). The empirical

[6] We did not consider estimating σ_ϕ^2 under the categorical distributions of ϕ_i since the associated true values of σ_ϕ^2 are too close to zero.

power functions are flat around the true value of σ_ϕ^2 for $T = 5$ and n small. When $T = 5$, large values of n are required to achieve reasonable power in the locality of the null hypothesis. The power improves rapidly as T and n are increased and, in line with the earlier results, the FDAC estimator performs better than the HetroGMM estimator.

7.2.3 Robustness

The FDAC estimators seem to be reasonably robust to departures from Gaussian errors and the presence of GARCH effects. Table S.5 of the online supplement provides results for the four combinations of error distributions, Gaussian and non-Gaussian, without and with GARCH effects for estimation of μ_ϕ. This table reports the results for the uniformly distributed AR coefficients with $\phi_i \in (-1, 1]$ and $\mu_\phi = 0.5$. We obtain similar results when we generate ϕ_i following a categorical distribution. The RMSE and size distortions of the FDAC estimator increase only slightly as we move from Gaussian to non-Gaussian errors and as we allow for GARCH effects. In contrast, the HetroGMM estimator is much more adversely affected by departures from Gaussian errors. Its bias and RMSE are much higher, with large size distortions, particularly with small n ($n = 100$). The performances of both estimators are adversely affected when non-Gaussian errors are combined with GARCH effects. Estimation of σ_ϕ^2 is similarly adversely affected when we allow for non-Gaussian errors as well as GARCH effects. The related simulation results are summarized in Table S.6 of the online supplement.

Overall, the FDAC estimator outperforms the HetroGMM estimator and seems to be reasonably robust to non-Gaussian errors and GARCH effects. It is also simple to compute. In what follows we focus on the estimation of μ_ϕ and compare the FDAC estimator with the HomoGMM estimators as well as the MSW estimator that allows for slope heterogeneity.

7.3 Comparison of FDAC and HomoGMM estimators

Tables S.8 and S.9 in the online supplement summarize the results comparing the FDAC estimator with FDLS, AH, AAH, AB and BB estimators, where ϕ_i is uniformly distributed, $\phi_i = \mu_\phi + v_i$ and $v_i \sim IIDU[-a, a]$ with $a = 0.5$ and $\mu_\phi = 0.4$ ($|\phi_i| < 1$) and $\mu_\phi = 0.5$

($\phi_i \in (-1, 1]$). We use the sample size combinations, $T = 4, 6, 10$, and $n = 100, 1000, 5000$, in the baseline case where the errors are Gaussian without GARCH effects. The simulation results with the other error processes are available upon request.

In line with our theoretical derivations, the HomoGMM estimators that neglect heterogeneity are severely biased and show large size distortions, whilst the bias of the FDAC estimator is close to zero and its size is around the five per cent nominal level, irrespective of whether $|\phi_i| < 1$ or $\phi_i \in (-1, 1]$. Also, with increases in n and/or T, the biases of the HomoGMM estimators do not shrink to zero and, as a result, the size distortions of the HomoGMM estimators become even more pronounced. The simulation results also confirm the magnitude of the asymptotic bias of the AH estimator given by (4.4) in Section 4, and those of AB and BB estimators provided in Section S.5 of the online supplement.

Since it is not known if the heterogeneity bias is serious, it is natural to ask if the FDAC estimator continues to perform equally well under homogeneity ($\phi_i = \mu_\phi = 0.5$ for all i), and if its performance under homogeneity is comparable to those of HomoGMM estimators of ϕ. Accordingly, we also computed bias, RMSE, and size of the FDAC and HomoGMM estimators under slope homogeneity ($a = 0$) with $\mu_\phi = 0.5$. The results for Gaussian errors without GARCH effects are summarized in Table S.10 of the online supplement. As can be seen, the FDAC estimator continues to perform well even under slope homogeneity. Its bias is close to zero and shows only a small degree of size distortions when $n = 100$. In terms of assumptions, the FDAC estimator is closest to the FDLS estimator under homogeneity. Figure S.3 in the online supplement compares the empirical power functions of FDAC and FDLS estimators. Compared to the FDLS estimator, the FDAC estimator makes use of higher order autocorrelation of first differences that are not needed for identification of μ_ϕ under homogeneity. As a result, the FDLS estimator is marginally more powerful than the FDAC for small $T = 4$, while the opposite is the case for $T = 10$.

When comparing the FDAC and the other HomoGMM estimators (such as AAH, BB, or AB) one needs to be cautious however, since these estimators do allow for the distribution of y_{i0} to depart from the steady state distribution of $\{y_{it}\}$. With this in mind, we note that the

FDAC estimator performs well when compared to AH, AAH and AB estimators, although it is marginally less efficient when compared to the BB estimator. Also, the FDAC estimator has less size distortion and better power performance compared to all HomoGMM estimators as T is increased. In short, these results demonstrate the FDAC estimator is reliable and has desirable small-sample performance even in homogeneous panels with stationary outcome processes.

Figure S.4 in the online supplement shows the empirical power functions for the FDAC estimator under homogeneity with $\phi_i = \mu_\phi = 0.5$ for all i and heterogeneity with uniformly distributed $\phi_i \in (-1, 1]$ ($\mu_\phi = 0.5$), in the cases of Gaussian and non-Gaussian errors without GARCH effects. The empirical power functions for the FDAC estimator in the cases of Gaussian errors without and with GARCH effects are displayed in Figure S.5 of the online supplement. The power functions become steeper as n and T increase. In general, the power of the FDAC estimator is similar under heterogeneous and homogeneous ϕ_i. Consistent with the previous findings, with non-Gaussian errors and/or GARCH effects, particularly for small $n = 100$, the power functions become noticeably flatter, and the size distortions become more pronounced.

7.4 Comparison of FDAC and MSW estimators

This section compares the small-sample performance of the FDAC estimator with the MSW estimator by Mavroeidis et al. (2015). Table S.11 in the online supplement reports bias, RMSE, and size of the FDAC and MSW estimators for μ_ϕ for $T = 4, 6, 10$, and $n = 100, 1000$, with uniformly distributed ϕ_i and Gaussian errors without GARCH effects. The left and right panels of the table report results for $\mu_\phi = 0.4$ ($|\phi_i| < 1$) and $\mu_\phi = 0.5$ ($\phi_i \in (-1, 1]$), respectively. The performance of the FDAC estimator is in line with the ones already discussed and as noted earlier is not affected by whether some $\phi_i = 1$ or not. In contrast, the MSW estimator performs rather poorly in the presence of a high degree of heterogeneity in ϕ_i and shows large biases and substantial size distortions across the examined sample sizes. In the case of $\phi_i \in (-1, 1]$, the MSW estimator shows greater bias, RMSE, and size distortions.

7.5 Non-stationary initializations

Since the first differences of y_{it} do not depend on the initial values when $\phi_i = 1$, non-stationary initialization matters only if $|\phi_i| < 1$. In this case, using (3.4) it is clear initial values matter only when M_i is small. Therefore, to investigate the robustness of the FDAC estimator to different initializations we consider relatively small values of $M_i = 1$ and 3 for all i, compared with the baseline case where we set $M_i = 100$ for all units with $|\phi_i| < 1$. The initial values are generated as $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$, compared to their steady state values of $b = 0$ and $\kappa_i = 1/(1 - \phi_i^2)$, respectively. When ϕ_i are generated from a categorical distribution we set $M_i = 1$ for all units with $\phi_H = 1$.

We consider both uniformly and categorically distributed ϕ_i. The results for the uniformly distributed ϕ_i under the three initializations $M_i \in \{100, 3, 1\}$ are summarized in Table S.12 of the online supplement. Similar results when ϕ_i follow the categorical distribution is given in Table S.13 of the online supplement. It is clear that the FDAC estimator is adversely affected when $M_i = 1$ and displays bias and substantial size distortions. As to be expected, the magnitude of the bias is not affected by n but falls sharply with T. As a result when $M_i = 1$ we observe substantial size distortions when n is large. Comparing the upper and lower panels, as the first differences of a unit root process are not affected by the initial values, having some ϕ_i being close to one mitigates the negative impact of non-stationary initializations on the FDAC estimator. These impacts are more pronounced for categorically distributed ϕ_i where the variances of ϕ_i are smaller, as shown in Table S.13 versus Table S.12. More importantly, as to be expected, the bias and size distortion of FDAC disappear as M_i is increased. When moving from $M_i = 1$ to $M_i = 3$, the bias and size distortion shrink fast, with only a slight size distortion observed when $M_i = 3$.

We also consider the relative performance of the FDAC and HomoGMM estimators under different initialization scenarios, for both cases of homogeneous and heterogeneous panels. Results for the homogenous case when $\phi_i = \mu_\phi = 0.5$ are summarized in Table S.14, and results for the heterogenous case are provided in Tables S.15 and S.16 for cases where $\mu_\phi = 0.4$ ($|\phi_i| < 1$) and $\mu_\phi = 0.5$ ($\phi_i \in (-1.1]$), respectively. In the homogeneous case, when $M_i = 1$, the FDAC, FDLS, and BB estimators all show sizeable bias and size distortions that do not vanish as n increases. Also, as to be expected, under homogeneity, the AH, AAH, and AB estimators are robust to non-stationary initialization and have similar performances across different values of M_i. In the case of heterogeneous panels, the performance of the FDAC estimator is as discussed above. For the HomoGMM estimators, the magnitude of neglected heterogeneity bias is smaller with less serious size distortions when $M_i = 1$ or 3, as compared to $M_i = 100$ (which approximately corresponds to the stationary case). The AH estimator seems to be an exception. Nonetheless, the HomoGMM estimators exhibit substantial size distortions across most of the considered sample sizes, leading to incorrect inference.

In short, for moderate values of M_i (in the case of our experiments when $M_i > 3$), the performance of the FDAC estimator is satisfactory even when $y_{i0} - \mu_i$ are not drawn from the steady distribution of the underlying processes, $\{y_{it} - \mu_i\}$. Comparisons of the FDAC and HomoGMM estimators also highlight the trade-off that exists between the "non-stationary initialization" bias of the FDAC estimator and the neglected heterogenous bias of the HomoGMM estimator. It remains a challenge to simultaneously deal with heterogeneity of ϕ_i and the non-stationarity of the initial values.

8 Empirical application: heterogeneity in earnings dynamics

8.1 Literature review of estimation of earnings dynamics

Estimating earnings equations is crucial for answering some of the most important economic questions.[7] Variance of earnings has been modeled and decomposed to measure income uncertainties in Lillard and Weiss (1979), MaCurdy (1982), Carroll and Samwick (1997), Meghir and Pistaferri (2004), Altonji et al. (2013) and to quantify earnings mobility in

[7] See p. 58 in Guvenen (2009) for a brief summary of several economic inquiries hinging on the estimation of earnings functions.

Lillard and Willis (1978) and Geweke and Keane (2000). The covariance structures between earnings and other households' characteristics, for example, work hours, consumptions and savings, have been studied by Abowd and Card (1989), Hubbard et al. (1995), Guvenen (2007), and Alan et al. (2018).

Among these studies, a homogeneous AR or ARMA process is often used as a component when modeling *innovations* in earnings processes. Based on the Restricted Income Profiles model that assumes homogenous linear trends proposed in MaCurdy (1982), MaCurdy (1982) and Hubbard et al. (1995), obtained close to unit root estimates for the AR(1) coefficient, ranging from 0.946 to 0.998.[8] Following this literature, a unit root assumption was imposed in Carroll and Samwick (1997) and Meghir and Pistaferri (2004). On the other hand, using the Heterogeneous Income Profiles, by assuming unit-specific linear trends, Lillard and Weiss (1979) obtained estimates of the AR(1) coefficient (assumed to be homogeneous) ranging from 0.153 to 0.860 for a sample with PhD degrees. Guvenen (2009) using PSID data obtained estimates ranging from 0.809 to 0.899.[9]

There are also a number of studies that allow for heterogeneity in the AR(1) coefficients. Prominent examples are Browning et al. (2010), Alan et al. (2018), Browning et al. (2010), and Gu and Koenker (2017). These studies are typically based on panels with a moderate time dimension and make parametric assumptions regarding the distribution of the AR(1) coefficients; often using a Bayesian framework.[10] The application of the FDAC estimator to earnings equation allows for heterogeneity in the AR(1) coefficients without making any strong parametric assumptions, even when T is as small as 5. Also because of first differencing prior to estimation, the FDAC estimator is robust to unobserved individual-specific characteristics and is not subject to misspecification bias that could arise when log real wages are filtered for individual-specific characteristics before investigating the dynamics of the earnings process.

[8]See Table 5 on p. 111 in MaCurdy (1982) using an ARMA(1,1) process. See Table 2 on p. 380 in Hubbard et al. (1995) based on an AR(1) process.

[9]See Tables 2, 4, 6 and 7 in Lillard and Weiss (1979), Table 1 on p. 64 in Guvenen (2009), and the abstract of Gu and Koenker (2017).

[10]See pp. 227–232 in Browning and Ejrnæs (2013) for a comprehensive survey of heterogeneity in parameters of earnings functions.

8.2 A heterogeneous panel AR(1) model of earnings dynamics with linear trends

We consider estimating the earnings equation with fixed effects, heterogeneous autoregressive coefficients, without imposing any restrictions on the joint distributions of α_i, ϕ_i, and y_{i0}. However, to accommodate growth in real earnings we extend our baseline model in (2.1) to allow for linear trends:

$$y_{it} = \alpha_i + g_i(1 - \phi_i)t + \phi_i y_{i,t-1} + u_{it}, \tag{8.1}$$

where $y_{it} = log(earnings_{it}/p_t)$, $earnings_{it}$ is the reported earnings of individual i in year t, p_t is a general price, and g_i is the growth rate of real earnings for individual i. (8.1) can be written equivalently as

$$\tilde{y}_{it}(g_i) = b_i + \phi_i \tilde{y}_{i,t-1}(g_i) + u_{it},$$

with $\tilde{y}_{it}(g_i) = y_{it} - g_i t$ and $b_i = \alpha_i - g_i\phi_i$. For $|\phi_i| < 1$, the steady state distribution of y_{it} can now be derived using

$$y_{it} = b_i + g_i t + \sum_{s=0}^{\infty} \phi_i^s u_{i,t-s}. \tag{8.2}$$

When T is sufficiently large, individual-specific growth rates, g_i, can be estimated $\sqrt{T}$-consistently by running individual least squares regressions of y_{it} on an intercept and a linear trend, and then using the residuals from these regressions to estimate the moments of ϕ_i. This approach requires n and T to be both large. In the case of the present empirical application where T is short (5 or 10), we provide estimates of the moments of ϕ_i assuming that $g_i = g$ for individuals within a given group, but allow g to differ across groups, classified by the educational attainment levels. $\sqrt{n}$-consistent estimators of g can be obtained either from the pooled regression of y_{it} on fixed effects and a common linear trend, namely

$$\hat{g}_{FE} = \left[\sum_{t=1}^{T}\left(t - \frac{(T+1)}{2}\right)^2\right]^{-1}\left[\sum_{t=1}^{T}(\bar{y}_{\circ t} - \bar{y}_{\circ\circ})t\right], \tag{8.3}$$

with $\bar{y}_{\circ t} = n^{-1}\sum_{i=1}^{n} y_{it}$ and $\bar{y}_{\circ\circ} = T^{-1}\sum_{t=1}^{T}\bar{y}_{\circ t}$, or after first differencing of (8.2) by

$$\hat{g}_{FD} = \frac{\sum_{t=2}^{T}\sum_{i=1}^{n}\Delta y_{it}}{n(T-1)}. \tag{8.4}$$

For small T there is little to choose between these two estimators, and they are identical when $T = 2$. Given either of the above estimators, generically denoted by $\hat{g}$, $\tilde{y}_{it}(\hat{g}) = y_{it} - \hat{g}t$ can now be used to estimate the moments of ϕ_i using the FDAC or MSW procedures.[11]

In addition to the FDAC estimates, we also present estimates based on four estimation methods assuming homogeneous slope coefficients, namely AAH, AB, and BB estimators proposed by Chudik and Pesaran (2021), Arellano and Bond (1991), and Blundell and Bond (1998), and the MSW estimator of Mavroeidis et al. (2015). Following Meghir and Pistaferri (2004), individuals in each time series sample are divided into three education categories, where "HSD" refers to high school dropouts with less than 12 years of education, "HSG" refers to high school graduates with at least 12 but less than 16 years of education, and "CLG" refers to college graduates with at least 16 years of education.[12] To allow for possible time variations in the estimates of mean earnings persistence we provide estimates for $five$ and *ten* yearly non-overlapping sub-periods. The five yearly samples are 1976–1980, 1981–1985, 1986–1990, and 1991–1995. The ten yearly samples are 1976–1985, 1981–1990 and 1991–1995. For each sub-period, we provide estimates for all categories combined, as well as separate estimates for the three educational sub-categories.[13] To save space, the results for the last five and ten yearly samples are given in the paper. The estimates for the earlier sub-periods are provided in the online supplement.

Table 1 gives the estimates of mean earnings persistence, $\mu_\phi = E(\phi_i)$, and the common linear trend coefficient, g, for the sub-periods 1991–1995 ($T = 5$) and 1986–1995 ($T = 10$). The estimates of g are on average around 2 per cent per annum with some modest variations across the sub-samples and educational categories. The HomoGMM estimates (AAH, AB and BB) differ a great deal, both over sub-periods and across educational categories. The AAH estimates are all around 0.50 and show little variations across the two sub-periods and

[11] Consistent estimation of $E(\phi_i)$ in the presence of heterogeneity in both ϕ_i and g_i requires moderate to large values of T. The approach used in the empirical literature whereby y_{it} are first de-meaned and de-trended for each i prior to the estimation of $E(\phi_i)$ is subject to Nickell (1981) bias in the case of short T panels, even if $E(\phi_i) = \phi$.

[12] The sample for all individuals in both 5 and 10 yearly samples covered $3,113$ individuals with consecutive observations of nine years or more, and 36,325 individual-year observations.

[13] From 1997 PSID data are updated every two years. We confine our analysis to the years 1976 to 1995 to construct panels with 5 and 10 consecutive years.

the educational categories. The AB estimates tend to be quite low and are not statistically significant for two of the educational categories in the shorter sub-period ($T = 5$). In contrast, the BB estimates are much larger and in many instances are close to unity. For example, for the sub-period 1986–1995 ($T = 10$), the BB estimates of earnings persistence for the three educational categories HSD, HSG and CLG are 0.923 (0.003), 0.914 (0.003) and 0.992 (0.004), respectively, with standard errors in brackets.

Table 1: Estimates of mean persistence ($\mu_\phi = E(\phi_i)$) of log real earnings in a panel AR(1) model with a common linear trend using PSID data over 1991–1995 and 1986–1995

	1991–1995, $T = 5$				1986–1995, $T = 10$			
	All categories	Category by education			All categories	Category by education		
		HSD	HSG	CLG		HSD	HSG	CLG
Homogeneous slopes								
AAH	0.526	0.490	0.547	0.447	0.546	0.569	0.535	0.522
	(0.046)	(0.072)	(0.061)	(0.072)	(0.028)	(0.024)	(0.033)	(0.038)
AB	0.278	0.105	0.320	-0.013	0.311	0.310	0.335	0.232
	(0.081)	(0.147)	(0.097)	(0.133)	(0.039)	(0.045)	(0.044)	(0.070)
BB	0.488	0.872	0.602	0.964	0.880	0.923	0.914	0.992
	(0.059)	(0.031)	(0.042)	(0.074)	(0.004)	(0.003)	(0.003)	(0.004)
Heterogeneous slopes								
FDAC	0.586	0.582	0.567	0.635	0.636	0.580	0.611	0.734
	(0.042)	(0.132)	(0.056)	(0.065)	(0.023)	(0.071)	(0.028)	(0.040)
MSW	0.437	0.431	0.436	0.452	0.458	0.459	0.452	0.460
	(0.040)	(0.044)	(0.043)	(0.045)	(0.054)	(0.038)	(0.046)	(0.063)
Common linear trend	0.023	0.008	0.027	0.020	0.019	0.024	0.020	0.013
n	1,366	127	832	407	1,139	109	689	341

Notes: The estimates are based on $y_{it} = \alpha_i + g(1-\phi_i)t + \phi_i y_{i,t-1} + u_{it}$, where $y_{it} = log(earnings_{it}/p_t)$ using the PSID data over the sub-periods 1991–1995 and 1986–1995. "HSD" refers to high school dropouts with less than 12 years of education, "HSG" refers to high school graduates with at least 12 but less than 16 years of education, and "CLG" refers to college graduates with at least 16 years of education. $\hat{g}_{FD}$ is computed by (8.4), then μ_ϕ is estimated based on $\tilde{y}_{it} = y_{it} - \hat{g}_{FD}t$. "AAH", "AB", and "BB" denote the 2-step GMM estimators by Chudik and Pesaran (2021), Arellano and Bond (1991), and Blundell and Bond (1998). The FDAC estimator is calculated by (6.2). "MSW" denotes the estimator by Mavroeidis et al. (2015).

We also find sizeable differences in the estimates of mean earnings persistence when we consider the FDAC and MSW estimators. The MSW estimates are all around 0.45 and do not vary with the level of educational attainment. In contrast, the FDAC estimates are somewhat larger (lie in the range of 0.570–0.734) and rise with the level of educational attainment. This pattern can be seen in both sub-periods. For example, for the longer sub-period (1986-1995), the mean persistence for HSD, HSG and CLG categories are estimated to be 0.580 (0.071), 0.611 (0.028) and 0.735 (0.040), respectively. Similar results are obtained for the other sub-periods. See Tables S.19 and S.20 of the supplement. Interestingly, the higher earnings persistence of the college graduate category is a prominent feature of the FDAC estimates for all sub-periods. This result is also in line with a number of theoretical arguments in the literature in terms of higher mobility of college graduates and their relative job stability, for example, Carroll and Samwick (1997) and Carneiro et al. (2023).

Although we have not developed a formal statistical test of the heterogeneity ϕ_i, the estimates of σ_ϕ^2 provide a good indication of the degree of within-group heterogeneity. Estimates of σ_ϕ^2 based on MSW and FDAC procedures for the various sub-periods are given in Tables S.21–S.23 of the online supplement. The FDAC estimates are much larger than the MSW estimates. For example, for the sub-period 1986–1995 the MSW estimates of σ_ϕ^2 are all around 0.011 with standard errors in the range of 0.005–0.011, whilst the FDAC estimates of σ_ϕ^2 for the same sub-period are 0.122 (0.06), 0.12 (0.031) and 0.141 (0.036) for the three educational categories of HSD, HSG and CLG, respectively. The degree of within-group heterogeneity also seems to vary over time. For example, for the shorter sub-period (1991-1995), the FDAC estimates of σ_ϕ^2 are generally smaller with larger standard errors for the two categories of HSG and CLG.

9 Conclusion

This paper considers the estimation of heterogeneous panel AR(1) models with short T, as $n \to \infty$. It allows for individual fixed effects and proposes estimating the moments of the AR(1) coefficients, $E(\phi_i^s)$, for $s = 1, 2, ..., S$, using the autocorrelation functions of first differences. It is shown that the standard GMM estimators proposed in the literature for short T homogeneous panels are inconsistent in the presence of slope heterogeneity. Analytical expressions for the bias are derived and shown to be very close to estimates obtained from stochastic simulations.

We propose two moment based estimators. A simple estimator based on autocorrelations of first differences, denoted by FDAC, and a GMM estimator based on autocovariances of first differences denoted by HetroGMM. Both estimators allow for some of the cross section units to have unit roots.

The small sample properties of the proposed estimators are investigated using Monte Carlo experiments. It is shown that the FDAC estimators of μ_ϕ and σ_ϕ^2 perform much better than the corresponding HetroGMM estimator. We also find that quite large samples might be required for reliable estimation of σ_ϕ^2, assuming that the true value of σ_ϕ^2 is not too close to zero.

The simulation results also show that the FDAC estimator of μ_ϕ is robust to different distributions of autoregressive coefficients and error processes. Further, we find that the FDAC estimator performs well even under homogeneous AR(1) coefficients. The magnitudes of bias and RMSE of the FDAC estimator are comparable to the HomoGMM estimators, and the size of the tests based on the FDAC estimator is mostly around the 5 per cent nominal level. But when initializations of the outcome processes deviate from their associated steady state distributions, the FDAC estimator could suffer from bias and size distortions. There is a trade-off between heterogeneity bias and the bias due to the non-stationary initializations.

The utility of the FDAC estimators of μ_ϕ and σ_ϕ^2 is illustrated by an empirical application using 1976–1995 PSID data to estimate heterogeneous AR(1) panels in log real earnings with a common linear trend. We provide estimates of μ_ϕ and σ_ϕ^2 over a number of 5 and 10 yearly sub-periods, with 3 educational groupings. The estimates of μ_ϕ differ systematically across the education groups, with the mean persistence of real earnings rising with the level of educational attainments (high school dropouts, high school graduates, and college graduates). The estimates of σ_ϕ^2 differ across periods and levels of educational attainment but do not display any particular patterns.

It is important to acknowledge that the scope of the present paper is limited, with a number of remaining challenges: (a) allowing for individual-specific time-varying covariates, and (b) simultaneously dealing with heterogeneity and non-stationary initializations. It is not clear that such extensions will be possible without relaxing the assumption that T is short and fixed, as $n \to \infty$. But these are clearly important topics for future research.

References

Abowd, J. M. and D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica 57*, 411–445.

Alan, S., M. Browning, and M. Ejrnæs (2018). Income and consumption: a micro semistructural analysis with pervasive heterogeneity. *Journal of Political Economy 126*, 1827–1864.

Altonji, J. G., A. A. Smith Jr, and I. Vidangos (2013). Modeling earnings dynamics. *Econometrica 81*, 1395–1454.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76*, 598–606.

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*, 47–82.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies 58*, 277–297.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 115–143.

Browning, M. and J. M. Carro (2014). Dynamic binary outcome models with maximal heterogeneity. *Journal of Econometrics 178*, 805–823.

Browning, M. and M. Ejrnæs (2013). Heterogeneity in the dynamics of labor earnings. *Annual Review of Econonomics 5*, 219–245.

Browning, M., M. Ejrnæs, and J. Alvarez (2010). Modelling income processes with lots of heterogeneity. *The Review of Economic Studies 77*, 1353–1381.

Carneiro, A., P. Portugal, P. Raposo, and P. M. Rodrigues (2023). The persistence of wages. *Journal of Econometrics 233*, 596–611.

Carroll, C. D. and A. A. Samwick (1997). The nature of precautionary wealth. *Journal of Monetary Economics 40*, 41–71.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica: Journal of the Econometric Society 60*, 567–596.

Chudik, A. and M. H. Pesaran (2021). An augmented Anderson–Hsiao estimator for dynamic short-T panels. *Econometric Reviews 41*, 416–447.

Geweke, J. and M. Keane (2000). An empirical analysis of earnings dynamics among men in the PSID: 1968–1989. *Journal of Econometrics 96*, 293–356.

Graham, B. S. and J. L. Powell (2012). Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica 80*, 2105–2152.

Gu, J. and R. Koenker (2017). Unobserved heterogeneity in income dynamics: an empirical Bayes perspective. *Journal of Business & Economic Statistics 35*, 1–16.

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review 97*, 687–712.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics 12*, 58–79.

Han, C. and P. C. Phillips (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory 26*, 119–151.

Hsiao, C., M. H. Pesaran, and A. Tahmiscioglu (1999). Bayes estimation of short-run coefficients in dynamic panel data models. In *Analysis of Panel Data and Limited Dependent Variable Models*. Cambridge University Press.

Hubbard, R. G., J. Skinner, and S. P. Zeldes (1995). Precautionary saving and social insurance. *Journal of Political Economy 103*, 360–399.

Lillard, L. A. and Y. Weiss (1979). Components of variation in panel earnings data: American scientists 1960-70. *Econometrica: Journal of the Econometric Society 47*, 437–454.

Lillard, L. A. and R. J. Willis (1978). Dynamic aspects of earning mobility. *Econometrica 46*, 985–1012.

Liu, L. (2023). Density forecasts in panel data models: a semiparametric Bayesian perspective. *Journal of Business & Economic Statistics 41*, 1–15.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics 18*, 83–114.

Mavroeidis, S., Y. Sasaki, and I. Welch (2015). Estimation of heterogeneous autoregressive parameters with short panel data. *Journal of Econometrics 188*, 219–235.

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica 72*, 1–32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society 49*, 1417–1426.

Okui, R. and T. Yanagi (2019). Panel data analysis with heterogeneous dynamics. *Journal of Econometrics 212*, 451–475.

Okui, R. and T. Yanagi (2020). Kernel estimation for panel data with heterogeneous dynamics. *The Econometrics Journal 23*, 156–175.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*, 79–113.

Pesaran, M. H. and L. Yang (2023). Trimmed mean group estimation of average treatment effects in ultra short-T panels with correlated heterogeneous coefficients. Working paper, arXiv preprint arXiv:2310.11680.

Pesaran, M. H. and Z. Zhao (1999). Bias reduction in estimating long-run relationships from dynamic heterogeneous panels. In *Analysis of Panel Data and Limited Dependent Variable Models*. Cambridge University Press.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics 5*, 163–168.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics 87*, 385–390.

Online Supplement to

"Heterogeneous Autoregressions in Short T Panel Data Models "

M. Hashem Pesaran

University of Southern California, and Trinity College, Cambridge

Liying Yang

Postdoctoral research fellow, University of British Columbia

April 3, 2024

S.1 Introduction

This online supplement is organized as follows. Section S.2 provides a proof of Lemma 1 under the stationarity of the first differences, $\Delta y_{it} = y_{it} - y_{i,t-1}$. Section S.3 further illustrates the convergence property with uniformly distributed autoregressive coefficients, ϕ_i. Section S.5 derives expressions for the analytical bias of the AB and BB estimators under heterogeneity of ϕ_i when $T = 4$. Section S.6 derives the asymptotic covariance matrix for the HetroGMM estimator of the first two moments and its consistent estimator. Section S.7 provides formulae for empirical power functions of the tests based on our proposed estimators in the Monte Carlo simulations. Section S.8 provides additional Monte Carlo evidence. Section S.9 describes the sample (1976–1995) of the Panel Study of Income Dynamics (PSID) data used in the empirical application, and provides estimation results for a number of sub-periods in addition to the ones reported in the main paper.

S.2 Proof of Lemma 1: Existence of autocovariances of first differences

We first establish conditions under which first differences, Δy_{it}, are covariance stationary for any given t and i. Consider the result (3.5) in the main paper which we reproduce here for convenience:

$$\Delta y_{it} = u_{it} - (1 - \phi_i) \sum_{\ell=1}^{M_i+t-1} \phi_i^{\ell-1} u_{i,t-\ell} - \phi_i^{M_i+t-1}(1 - \phi_i)\left(y_{i,-M_i} - \mu_i\right), \tag{S.1}$$

for $t = 2, 3, ..., T$, where $R_i\left(y_{i,-M_i}\right) = -\phi_i^{M_i-1}(1 - \phi_i)\left(y_{i,-M_i} - \mu_i\right)$. Assuming ϕ_i and u_{it} are independently distributed and since the initial values, $y_{i,-M_i} - \mu_i$, are given, we have

$$E\left|y_{it}\right| \leq E\left|u_{it}\right| + \sum_{\ell=1}^{M_i+t-1} E\left|\phi_i^{\ell-1}(1 - \phi_i)\right| E\left|u_{i,t-\ell}\right| + E\left[\left|\phi_i^{M_i+t-1}(1 - \phi_i)\right|\right]\left|y_{i,-M_i} - \mu_i\right|.$$

Also since $\phi_i \in (-1, 1]$, then $E\left|\phi_i^s(1 - \phi_i)\right| \leq c^s$, for some $c < 1$, and we have

$$\sup_{i,t} E\left[\left|y_{it}\right| \left|(y_{i,-M_i} - \mu_i)\right.\right] \leq \sup_{i,t}\left|u_{it}\right| \left[1 + \frac{1 - c^{M_i+t-1}}{1 - c}\right] + c^{M_i+t-1}\left|y_{i,-M_i} - \mu_i\right| < C < \infty.$$

Hence, $E\left|y_{it}\right|$ exists for all values of $\phi_i \in (-1, 1]$ and is given by

$$E\left(y_{it}\left|y_{i,-M_i} - \mu_i\right.\right) = -E\left[\phi_i^{M_i+t-1}(1-\phi_i)\right]\left(y_{i,-M_i} - \mu_i\right).$$

It is clear that, since $t = 1, 2, ..., T$ and T is finite, then $E\left(y_{it}\right)$ varies with t and in general depends on the initial values, $y_{i,-M_i}$. $E\left(y_{it}\left|y_{i,-M_i} - \mu_i\right.\right)$ is time-invariant if and only if $M_i \to \infty$, and hence unconditionally we have $E(y_{it}) = 0$, for all i and t, if $M_i \to \infty$.

By Cauchy-Schwarz inequality $\left|\gamma_\Delta(h)\right| = \left|E\left(\Delta y_{it}\Delta y_{i,t-h}\right)\right| \leq \left[E\left(\Delta y_{it}\right)^2 E\left(\Delta y_{i,t-h}\right)^2\right]^{\frac{1}{2}}$, thus for the existence of autocovariances of Δy_{it}, it is sufficient to show that $E\left(\Delta y_{it}\right)^2 < \infty$. Using (3.6) in the main paper, it readily follows that

$$E\left(\Delta y_{it}\right)^2 = E\left[E\left(\left(\Delta y_{it}\right)^2\left|\phi_i, \sigma_i^2\right.\right)\right] = E\left(\sigma_i^2\right) + E\left[(1-\phi_i)^2\sum_{\ell=1}^{\infty}\phi_i^{2(\ell-1)}\sigma_i^2\right], \tag{S.2}$$

and given the independence of σ_i^2 and ϕ_i (see Assumption 3 in the main paper) we have

$$E\left(\Delta y_{it}\right)^2 = \sigma^2 + \sigma^2\sum_{\ell=1}^{\infty}E\left[(1-\phi_i)^2\phi_i^{2(\ell-1)}\right] \leq \sigma^2 + \sigma^2\sum_{s=0}^{\infty}E\left[(1-\phi_i)^2\phi_i^{2s}\right].$$

We now show that $\sum_{s=0}^{\infty}E\left[(1-\phi_i)^2\phi_i^{2s}\right]$ is convergent for any probability distribution of ϕ_i defined over the interval $(-1, +1]$. Note that for any finite M

$$\begin{aligned}\sum_{s=0}^{M}E\left[(1-\phi_i)^2\phi_i^{2s}\right] &= E\left[\sum_{s=0}^{M}(1-\phi_i)^2\phi_i^{2s}\right] = E\left[\frac{(1-\phi_i)^2\left(1-\phi_i^{2M+2}\right)}{1-\phi_i^2}\right] \\ &= E\left[\frac{(1-\phi_i)\left(1-\phi_i^{2M+2}\right)}{1+\phi_i}\right],\end{aligned}$$

where $1 + \phi_i > \epsilon > 0$, and $-1 < \phi_i \leq 1$.

$$\frac{(1-\phi_i)\left(1-\phi_i^{2M+2}\right)}{1+\phi_i} \leq (1/\epsilon)(1 + |\phi_i| + \left|\phi_i^{2M+2}\right| + \left|\phi_i^{2M+3}\right|).$$

Hence,

$$\sum_{s=0}^{M}E\left[(1-\phi_i)^2\phi_i^{2s}\right] \leq (1/\epsilon)(1 + |\phi_i| + \left|\phi_i^{2M+2}\right| + \left|\phi_i^{2M+3}\right|).$$

But since $\phi_i \in (-1, 1]$, $E\left|\phi_i^\ell\right| \leq 1$ for any $\ell = 0, 1, ...$, and it follows that $\sum_{s=0}^{M}E\left[(1-\phi_i)^2\phi_i^{2s}\right]$ $\leq 4/\epsilon$ for any finite M and as $M \to \infty$. Therefore, it follows that $\left|\gamma_\Delta(h)\right| < C$, as required.

Having established the existence of $\gamma_\Delta(h)$, using (S.2) and recalling that under Assumption 3) in the main paper ϕ_i and σ_i^2 are independently distributed we have

$$\begin{aligned} Var(\Delta y_{it}) &= \gamma_\Delta(0) = E\left[\sigma_i^2 + (1-\phi_i)^2 \sum_{\ell=1}^{\infty} \phi_i^{2(\ell-1)} \sigma_i^2\right] \\ &= E\left[\sigma_i^2 + \frac{(1-\phi_i)^2}{1-\phi_i^2}\sigma_i^2\right] = 2\sigma^2 E\left(\frac{1}{1+\phi_i}\right). \end{aligned}$$

Similarly, to derive $\gamma_\Delta(h) = E\left(\Delta y_{it} \Delta y_{i,t-h}\right)$ we first note that $M_i \to \infty$, then using (S.2) we have

$$\Delta y_{it} = u_{it} - (1-\phi_i) \sum_{\ell=1}^{\infty} \phi_i^{\ell-1} u_{i,t-\ell},$$

$$\Delta y_{i,t-h} = u_{i,t-h} - (1-\phi_i) \sum_{\ell=1}^{\infty} \phi_i^{\ell-1} u_{i,t-\ell-h},$$

and for $h = 1, 2, ...,$

$$\begin{aligned} E\left(\Delta y_{it} \Delta y_{i,t-h}\right) &= E\left[(1-\phi_i)^2 \left(\sum_{\ell=1}^{\infty} \phi_i^{\ell-1} u_{i,t-\ell}\right)\left(\sum_{\ell=1}^{\infty} \phi_i^{\ell-1} u_{i,t-\ell-h}\right)\right] \\ &\quad -E\left[(1-\phi_i)\left(\sum_{\ell=1}^{\infty} \phi_i^{\ell-1} u_{i,t-\ell} u_{i,t-h}\right)\right]. \end{aligned}$$

First, we consider the second term, and note that

$$E\left[(1-\phi_i)\left(\sum_{\ell=1}^{\infty} \phi_i^{\ell-1} u_{i,t-\ell} u_{i,t-h}\right)\right] = E\left[\sigma_i^2 (1-\phi_i) \phi_i^{h-1}\right].$$

Also

$$E\left[(1-\phi_i)^2 \left(\sum_{\ell=1}^{\infty} \phi_i^{\ell-1} u_{i,t-\ell}\right)\left(\sum_{\ell=1}^{\infty} \phi_i^{\ell-1} u_{i,t-\ell-h}\right)\right] = E\left[\sigma_i^2 (1-\phi_i)^2 \left(\phi_i^h + \phi_i^{h+2} + \phi_i^{h+4} + ...\right)\right].$$

Hence

$$E(\Delta y_{it} \Delta y_{i,t-h}) = -E\left[\sigma_i^2 (1-\phi_i) \phi_i^{h-1}\right] + E\left[\sigma_i^2 (1-\phi_i)^2 \left(\phi_i^h + \phi_i^{h+2} + \phi_i^{h+4} + ...\right)\right],$$

and since ϕ_i and σ_i^2 are independently distributed we have

$$E(\Delta y_{it} \Delta y_{i,t-h}) = -E\left(\sigma_i^2\right) E\left[(1-\phi_i)\phi_i^{h-1} - (1-\phi_i)^2 \left(\phi_i^h + \phi_i^{h+2} + \phi_i^{h+4} + ...\right)\right],$$

As before, for all $\phi_i \in (-1, 1]$, we have $E\left|(1-\phi_i)^2 \phi_i^{h+s}\right| \leq E\left|(1-\phi_i)^2 \left|\phi_i\right|^{h+s}\right| \leq c^{h+s}$, where

$c < 1$, and the series is convergent, and we have

$$E(\Delta y_{it}\Delta y_{i,t-h}) = -E\left(\sigma_i^2\right) E\left[(1-\phi_i)\phi_i^{h-1} - \frac{(1-\phi_i)\phi_i^h}{1+\phi_i}\right],$$

or

$$E(\Delta y_{it}\Delta y_{i,t-h}) = -E\left(\sigma_i^2\right) E\left[\frac{(1-\phi_i)\phi_i^{h-1}}{1+\phi_i}\right], \text{ for } h = 1, 2, ..., \tag{S.3}$$

as required. Similarly

$$\begin{aligned} E\left(\phi_i\Delta y_{it}\Delta y_{i,t-h}\right) &= -E\left(\sigma_i^2\right) E\left[(1-\phi_i)\phi_i^h - \frac{(1-\phi_i)\phi_i^{h+1}}{1+\phi_i}\right] \\ &= -E\left(\sigma_i^2\right) E\left[\frac{(1-\phi_i)\phi_i^h}{1+\phi_i}\right], \text{ for } h = 1, 2, ... \end{aligned}$$

The results of Lemma 1 are now established noting that $E\left(\sigma_i^2\right) = \sigma^2$.

S.3 Examples: uniform distributions

It is also instructive to consider the important case where ϕ_i is uniformly distributed. First suppose that $\phi_i \sim Uniform(0, a]$ for $0 < a \leq 1$, then $E\left(\phi_i^\ell\right) = \frac{a^\ell}{\ell+1}$, and

$$E\left[(1-\phi_i)^2\phi_i^{2s}\right] = \frac{a^{2s}}{2s+1} - \frac{2a^{2s+1}}{2s+2} + \frac{a^{2s+2}}{2s+3}.$$

Hence

$$\sum_{s=0}^{\infty} E\left[(1-\phi_i)^2\phi_i^{2s}\right] = \sum_{s=0}^{\infty}\left(\frac{a^{2s}}{2s+1} - \frac{2a^{2s+1}}{2s+2} + \frac{a^{2s+2}}{2s+3}\right).$$

When $a < 1$, all the three individual sums in the above expression are bounded by $C/(1-a^2)$. However, this does not follow when $a = 1$, and the series $\sum_{s=0}^{\infty}\frac{1}{2s+1}$, $\sum_{s=0}^{\infty}\frac{2}{2s+2}$, and $\sum_{s=0}^{\infty}\frac{1}{2s+3}$, diverge individually. Hence, to investigate the convergence property of $\sum_{s=0}^{\infty} E\left[(1-\phi_i)^2\phi_i^{2s}\right]$ when $a = 1$, we need to consider all the terms together. For $a = 1$,

$$\sum_{s=0}^{\infty} E\left[(1-\phi_i)^2\phi_i^{2s}\right] = \sum_{s=0}^{\infty}\left(\frac{1}{2s+1} - \frac{2}{2s+2} + \frac{1}{2s+3}\right),$$

and after some algebra we have

$$\frac{1}{2s+1} - \frac{2}{2s+2} + \frac{1}{2s+3} = \frac{2}{(2s+1)(2s+2)(2s+3)},$$

$$\sum_{s=0}^{\infty} E\left[(1-\phi_i)^2\phi_i^{2s}\right] = \sum_{s=0}^{\infty} \frac{2}{(2s+1)(2s+2)(2s+3)} < C < \infty.$$

Similarly, for $\phi_i \sim Unifrom(-1+\epsilon, 0]$ we have $E\left(\phi_i^{\ell}\right) = -\frac{(-1)^{\ell}(1-\epsilon)^{\ell}}{\ell+1}$, and we have

$$E\left[(1-\phi_i)^2\phi_i^{2s}\right] = \frac{(1-\epsilon)^{2s}}{2s+1} + \frac{2(1-\epsilon)^{2s+1}}{2s+2} + \frac{(1-\epsilon)^{2s+2}}{2s+3},$$

and $\sum_{s=0}^{\infty} E\left[(1-\phi_i)^2\phi_i^{2s}\right]$ is convergent for $\epsilon > 0$, and diverges if $\epsilon = 0$. The latter case is ruled out under Assumption 4 in the main paper, which establishes the necessity of ruling out the boundary value of $\phi_i = -1$.

S.4 Proof of Proposition 1: Neglected heterogeneity bias of the AH estimator

Result (4.3) follows directly from (4.2), after subtracting $E(\phi_i)$ from both sides. Also, since $\phi_i \in [-1+\epsilon, 1]$, for some $\epsilon > 0$, then $1 + E(\phi_i) > 0$, and $E\left(\frac{1-\phi_i}{1+\phi_i}\right) > 0$. Since $1/(1+\phi_i)$ is a convex function of ϕ_i on $[-1+\epsilon, 1]$, then by Jensen inequality $E\left(\frac{1}{1+\phi_i}\right) \geq \frac{1}{1+E(\phi_i)}$, and it follows that $plim_{n\to\infty}\hat{\phi}_{AH} \leq E(\phi_i) = \mu_{\phi}$. Since $1 + \mu_{\phi} = 1 + E(\phi_i) > 0$, the asymptotic bias is zero if and only if $\frac{1}{1+\mu_{\phi}} = E\left(\frac{1}{1+\phi_i}\right)$, and due to the convexity of $1/(1+\phi_i)$, this condition is met only if $\phi_i = \mu_{\phi}$ for all i.

S.5 Neglected heterogeneity bias in AB and BB estimators

The AB estimator proposed by Arellano and Bond (1991) is based on the following moment conditions:[S1]

$$E(y_{is}\Delta u_{it}) = 0, \text{ for } i = 1, 2, ..., n, s = 1, 2, ..., t-2, \text{ and } t = 3, 4, ..., T, \tag{S.4}$$

which can also be written as $E[y_{is}(\Delta y_{it} - \phi_i \Delta y_{i,t-1})] = 0$, with $(T-1)(T-2)/2$ moment conditions in total. When $T = 4$, under homogeneity of ϕ_i, the AB moment conditions are given by $E[y_{i1}(\Delta y_{i3} - \phi\Delta y_{i2})] = 0$, $E[y_{i1}(\Delta y_{i4} - \phi\Delta y_{i3})] = 0$, and $E[y_{i2}(\Delta y_{i4} - \phi\Delta y_{i3})] = 0$.

[S1] See equation (8) on p. 5 in Chudik and Pesaran (2021).

With a fixed weight matrix $\mathbf{W}_{AB}$, the AB estimator can be written as

$$\hat{\phi}_{AB} = \left(\bar{\boldsymbol{z}}_{na}'\boldsymbol{W}_{AB}\bar{\boldsymbol{z}}_{na}\right)^{-1}\left(\bar{\boldsymbol{z}}_{na}'\boldsymbol{W}_{AB}\bar{\boldsymbol{z}}_{nb}\right), \tag{S.5}$$

where $\bar{\boldsymbol{z}}_{na} = n^{-1}\left(\sum_{i=1}^{n} y_{i1}\Delta y_{i2}, \sum_{i=1}^{n} y_{i1}\Delta y_{i3}, \sum_{i=1}^{n} y_{i2}\Delta y_{i3}\right)'$, and
$\bar{\boldsymbol{z}}_{nb} = n^{-1}\left(\sum_{i=1}^{n} y_{i1}\Delta y_{i3}, \sum_{i=1}^{n} y_{i1}\Delta y_{i4}, \sum_{i=1}^{n} y_{i2}\Delta y_{i4}\right)'$.

Using (2.2) in the main paper

$$y_{it} = \mu_i + \phi_i^{M_i+t}(y_{i,-M_i} - \mu_i) + \sum_{\ell=0}^{M_i+t-1} \phi_i^{\ell} u_{i,t-\ell}, \tag{S.6}$$

and assuming that μ_i is distributed independently of $\{u_{it}\}$ (as assumed under AB) then using (3.5) in the main paper and (S.6) we have

$$\begin{aligned}
&E\left(y_{i,t-h}\Delta y_{it}|\alpha_i,\phi_i,\sigma_i^2\right)\\
=&E\left[\left(\mu_i + \phi_i^{M_i+t-h}(y_{i,-M_i} - \mu_i) + \sum_{\ell=0}^{M_i+t-h-1} \phi_i^{\ell} u_{i,t-h-\ell}\right)\right.\\
&\left.\times\left(u_{it} - (1-\phi_i)\sum_{\ell=1}^{M_i+t-1}\phi_i^{\ell-1}u_{i,t-\ell} - \phi_i^{M_i+t-1}(1-\phi_i)\left(y_{i,-M_i} - \mu_i\right)\right)\middle|\,\alpha_i,\phi_i,\sigma_i^2\right]\\
=&E\left[-(1-\phi_i)\left(\sum_{\ell=0}^{M_i+t-h-1}\phi_i^{h-1+2\ell}u_{i,t-h-\ell}^2\right) - (1-\phi_i)\phi_i^{2M_i+2t-h-1}\left(y_{i,-M_i} - \mu_i\right)\middle|\,\phi_i,\sigma_i^2\right]
\end{aligned}$$

As $M_i \to \infty$ for $|\phi_i| < 1$ (with finite M_i for $\phi_i = 1$),

$$E\left(y_{i,t-h}\Delta y_{it}|\alpha_i,\phi_i,\sigma_i^2\right) = \begin{cases} 0, & \text{for } \phi_i = 1 \text{ and } h = 2,3,..., \\ -\frac{\sigma_i^2\phi_i^{h-1}}{1+\phi_i}, & \text{for } |\phi_i| < 1 \text{ and } h = 1,2,.... \end{cases} \tag{S.7}$$

Given (S.7), if $Pr(\phi_i = 1) = 0$ and $\phi_i \in (-1,1]$, we have

$$\begin{aligned}
\boldsymbol{z}_a = \underset{n\to\infty}{plim}\,\bar{\boldsymbol{z}}_{na} &= -\left(E\left(\frac{\sigma_i^2}{1+\phi_i}\right), E\left(\frac{\sigma_i^2\phi_i}{1+\phi_i}\right), E\left(\frac{\sigma_i^2}{1+\phi_i}\right)\right)',\\
\text{and}\quad \boldsymbol{z}_b = \underset{n\to\infty}{plim}\,\bar{\boldsymbol{z}}_{nb} &= -\left(E\left(\frac{\sigma_i^2\phi_i}{1+\phi_i}\right), E\left(\frac{\sigma_i^2\phi_i^2}{1+\phi_i}\right), E\left(\frac{\sigma_i^2\phi_i}{1+\phi_i}\right)\right)'.
\end{aligned}$$

Since ϕ_i is distributed independently of σ_i^2, for uniformly distributed $\phi_i = \mu_\phi + v_i$ with v_i

$\sim IIDU(-a,a)$, $a>0$ and $\phi_i \in (-1,1]$,

$$\underset{n\to\infty}{plim}\left(\hat{\phi}_{AB}-E(\phi_i)\right)=\left(\boldsymbol{z}_a'\boldsymbol{W}_{AB}\boldsymbol{z}_a\right)^{-1}\left(\boldsymbol{z}_a'\boldsymbol{W}_{AB}\boldsymbol{z}_b\right)-\mu_\phi, \tag{S.8}$$

where $\boldsymbol{z}_a=-\sigma^2(c_\phi,1-c_\phi,c_\phi)'$ and $\boldsymbol{z}_a=-\sigma^2(1-c_\phi,\mu_\phi-1+c_\phi,1-c_\phi)'$ with $\sigma^2=E(\sigma_i^2)$ and $c_\phi=E\left(\frac{1}{1+\phi_i}\right)=\frac{1}{2a}\ln\left(\frac{1+\mu_\phi+a}{1+\mu_\phi-a}\right)$.

In addition to (S.4), consider the following moment conditions, used in the system GMM estimator proposed by Blundell and Bond (1998), and note that under homogeneity we have[S2]

$$E\left[\Delta y_{i,t-1}(\mu_i(1-\phi_i)+u_{it})\right]=E[\Delta y_{i,t-1}(y_{it}-\phi y_{i,t-1})=0,\text{ for } i=1,2,...,n,\text{ and } t=3,4,...,T. \tag{S.9}$$

For $T=4$, with a given weight matrix $\boldsymbol{W}_{BB}$, the BB estimator based on the moment conditions in (S.4) and (S.9) is given by

$$\hat{\phi}_{BB}=\left(\bar{\boldsymbol{z}}_{nc}'\boldsymbol{W}_{BB}\bar{\boldsymbol{z}}_{nc}\right)^{-1}\left(\bar{\boldsymbol{z}}_{nc}'\boldsymbol{W}_{BB}\bar{\boldsymbol{z}}_{nd}\right), \tag{S.10}$$

where

$$\bar{\boldsymbol{z}}_{nc}=n^{-1}\left(\sum_{i=1}^n y_{i1}\Delta y_{i2},\sum_{i=1}^n y_{i1}\Delta y_{i3},\sum_{i=1}^n y_{i2}\Delta y_{i3},\sum_{i=1}^n y_{i2}\Delta y_{i2},\sum_{i=1}^n y_{i3}\Delta y_{i3}\right)',$$

$$\text{and}\quad \bar{\boldsymbol{z}}_{nd}=n^{-1}\left(\sum_{i=1}^n y_{i1}\Delta y_{i3},\sum_{i=1}^n y_{i1}\Delta y_{i4},\sum_{i=1}^n y_{i2}\Delta y_{i4},\sum_{i=1}^n y_{i3}\Delta y_{i2},\sum_{i=1}^n y_{i4}\Delta y_{i3}\right)'.$$

Using (3.5) in the main paper and (S.6), similarly, we can derive the following equations as $M_i\to\infty$ for $|\phi_i|<1$,

$$E\left(y_{it}\Delta y_{i,t-h}\right)=E\left(\frac{\sigma_i^2\phi_i^h}{1+\phi_i}\right),\text{ for } h=0,1,2,..., \tag{S.11}$$

and for $\phi_i=1$ and finite M_i, $E(\Delta y_{i,t-1}y_{it})=E(\Delta y_{i,t-1}y_{i,t-1})=\sigma_i^2$. In the case of $Pr(\phi_i=$

[S2]See equation (9) on p. 5 in Chudik and Pesaran (2021).

$1) = 0$ with $\phi_i \in (-1,1]$, it follows that

$$\boldsymbol{z}_c = \underset{n\to\infty}{plim}\bar{\boldsymbol{z}}_{nc} = \left(-E\left(\frac{\sigma_i^2}{1+\phi_i}\right), -E\left(\frac{\sigma_i^2\phi_i}{1+\phi_i}\right), -E\left(\frac{\sigma_i^2}{1+\phi_i}\right), E\left(\frac{\sigma_i^2}{1+\phi_i}\right), E\left(\frac{\sigma_i^2}{1+\phi_i}\right)\right)',$$

$$\boldsymbol{z}_d = \underset{n\to\infty}{plim}\bar{\boldsymbol{z}}_{nd}$$
$$= \left(-E\left(\frac{\sigma_i^2\phi_i}{1+\phi_i}\right), -E\left(\frac{\sigma_i^2\phi_i^2}{1+\phi_i}\right), -E\left(\frac{\sigma_i^2\phi_i}{1+\phi_i}\right), E\left(\sigma_i^2 - \frac{\sigma_i^2}{1+\phi_i}\right), E\left(\sigma_i^2 - \frac{\sigma_i^2}{1+\phi_i}\right)\right)'.$$

Since ϕ_i is distributed independently of σ_i^2, for uniformly distributed $\phi_i = \mu_\phi + v_i$ with $v_i \sim IIDU(-a,a)$, $a > 0$ and $\phi_i \in (-1,1]$,

$$plim_{n\to\infty}\left[\hat{\phi}_{BB} - E(\phi_i)\right] = \left(\boldsymbol{z}_c'\boldsymbol{W}_{BB}\boldsymbol{z}_c\right)^{-1}\left(\boldsymbol{z}_c'\boldsymbol{W}_{BB}\boldsymbol{z}_d\right), \tag{S.12}$$

where $\boldsymbol{z}_c = \sigma^2(-c_\phi, -1+c_\phi, -c_\phi, c_\phi, c_\phi)'$ and $\boldsymbol{z}_d = \sigma^2(c_\phi - 1, 1-\mu_\phi - c_\phi, -1+c_\phi, 1-c_\phi, 1-c_\phi)'$, with $E(\sigma_i^2) = \sigma^2$, and $c_\phi = E\left(\frac{1}{1+\phi_i}\right) = \frac{1}{2a}\ln\left(\frac{1+\mu_\phi+a}{1+\mu_\phi-a}\right)$.

To approximate the values of the asymptotic bias of AB and BB estimators corresponding to our Monte Carlo experiments, we replace $\boldsymbol{W}_{AB}$ and $\boldsymbol{W}_{BB}$ by the simulated weight matrices[S3] with $a = 0.5$, $\mu_\phi \in \{0.4, 0.5\}$, and Gaussian errors without GARCH effects for $T = 4$, and $n = 5,000$. In this case, the biases of AB and BB estimators are around -0.055 and -0.045 for $\mu_\phi = 0.4$, and -0.062 and -0.044 for $\mu_\phi = 0.5$, respectively. These results are close to the simulated bias of these estimators reported in Tables S.8 ($\mu_\phi = 0.4$) and S.9 ($\mu_\phi = 0.5$) for $T = 4$ and $n = 5,000$.

S.6 Asymptotic variances of the first two moments

Suppose that Assumptions 1–5 in the main paper hold, $T \geq 5$, and $M_i \to \infty$. The asymptotic distribution of $\hat{\boldsymbol{\theta}}_{HetroGMM} = (\hat{\theta}_{1,HetroGMM}, \hat{\theta}_{2,HetroGMM})'$ is given by

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{HetroGMM} - \boldsymbol{\theta}_0) \to_d N(0, \mathbf{V}_{\boldsymbol{\theta}}),$$

[S3]The simulated weight matrices are calculated as the average of the weight matrices used in calculating the two-step AB and BB estimators across 2,000 replications.

with $\boldsymbol{\theta}_0 = (\theta_{1,0}, \theta_{2,0})'$, and

$$\mathbf{V}_{\boldsymbol{\theta}}^{-1} = plim_{n\to\infty}(\mathbf{H}'_{\boldsymbol{\theta},nT}\mathbf{S}^{-1}_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0)\mathbf{H}_{\boldsymbol{\theta},nT}),$$

where $\mathbf{H}_{\boldsymbol{\theta},nT} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{H}_{\boldsymbol{\theta},iT}$,

$$\mathbf{H}_{\boldsymbol{\theta},iT} = \begin{pmatrix} \mathbf{h}_{iT} & \mathbf{0}_{(T-3)\times 1} \\ \mathbf{0}_{(T-4)\times 1} & \mathbf{h}_{2,iT} \end{pmatrix},$$

$\mathbf{S}_{\boldsymbol{\theta},T}(\boldsymbol{\theta}_0) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{g}_{\boldsymbol{\theta},iT} - \mathbf{H}_{\boldsymbol{\theta},iT}\boldsymbol{\theta}_0)(\mathbf{g}_{\boldsymbol{\theta},iT} - \mathbf{H}_{\theta,iT}\boldsymbol{\theta}_0)'$, $\mathbf{g}_{\boldsymbol{\theta},iT} = (\mathbf{g}'_{iT}, \mathbf{g}'_{2,iT})'$, and $\mathbf{h}_{iT}$, $\mathbf{h}_{2,iT}$, $\mathbf{g}_{iT}$ and $\mathbf{g}_{2,iT}$ are given by (6.8), (6.17), (6.9) and (6.18) in the main paper, respectively. $\mathbf{V}_{\boldsymbol{\theta}}$ can be consistently estimated by

$$\hat{\mathbf{V}}_{\boldsymbol{\theta}} = \left(\mathbf{H}'_{\boldsymbol{\theta},nT}\hat{\mathbf{S}}^{-1}_{\boldsymbol{\theta},T}(\hat{\boldsymbol{\theta}}_{HetroGMM})\mathbf{H}_{\boldsymbol{\theta},nT}\right)^{-1}, \tag{S.13}$$

with

$$\hat{\mathbf{S}}_{\boldsymbol{\theta},T}(\hat{\boldsymbol{\theta}}_{HetroGMM}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{g}_{\boldsymbol{\theta},iT} - \mathbf{H}_{\boldsymbol{\theta},iT}\hat{\boldsymbol{\theta}}_{HetroGMM})(\mathbf{g}_{\boldsymbol{\theta},iT} - \mathbf{H}_{\theta,iT}\hat{\boldsymbol{\theta}}_{HetroGMM})'.$$

S.7 Empirical power functions

The test statistics for $\mu_\phi = E(\phi_i)$ and $\sigma^2_\phi = Var(\phi_i)$ are given by

$$S_{N,\mu}\left(\mu_\phi\right) = \frac{\hat{\mu}_\phi - \mu_\phi}{\left[\widehat{Var}(\hat{\mu}_\phi)\right]^{1/2}} \quad \text{and} \quad S_{N,\sigma}\left(\sigma^2_\phi\right) = \frac{\hat{\sigma}^2_\phi - \sigma^2_\phi}{\left[\widehat{Var}\left(\hat{\sigma}^2_\phi\right)\right]^{1/2}},$$

respectively, where FDAC and HetroGMM estimators of $\hat{\mu}_\phi = \hat{\theta}_1$ are given by (6.2) and (6.13) in the main paper, respectively. $\hat{\sigma}^2_\phi$ is computed as the plug-in estimator given by (6.22) in the main paper. In the Monte Carlo experiments, the empirical power functions (EPF) are computed as the simulated rejection frequencies for replications $r = 1, 2, ..., R$:

$$EPF_R(\mu_\phi) = R^{-1}\sum_{r=1}^{R} I\left[\left|\frac{\hat{\mu}^{(r)}_\phi - \mu_\phi}{\left[\widehat{Var}\left(\hat{\mu}_\phi\right)^{(r)}\right]^{1/2}}\right| > 1.96\right],$$

and

$$EPF_R(\sigma_\phi^2) = R^{-1} \sum_{r=1}^{R} I \left[\left| \frac{\left(\hat{\sigma}_\phi^2\right)^{(r)} - \sigma_\phi^2}{\left[\widehat{Var\left(\hat{\sigma}_\phi^2\right)}^{(r)} \right]^{1/2}} \right| > 1.96 \right].$$

S.8 Monte Carlo evidence

S.8.1 Comparison of FDAC and HetroGMM estimators

Tables S.1 and S.2 summarize bias, RMSE, and size of FDAC and HetroGMM estimators of $\mu_\phi = E(\phi_i)$ with uniformly and categorically distributed ϕ_i, respectively, in the case of Gaussian errors without GARCH effects for the sample size combinations $n = 100, 1000, 5000$ and $T = 4, 5, 6, 10$. The empirical power functions of FDAC and HetroGMM estimators of μ_ϕ with uniformly distributed $\phi_i \in [-1 + \epsilon, 1]$ for some $\epsilon > 0$ are shown in Figure S.1.

Table S.3 reports the frequency where FDAC and HetroGMM estimates of σ_ϕ^2 are either negative or very close to zero, using the threshold $\left(\hat{\sigma}_\phi^2\right)^{(r)} < 0.0001$, for replication $r = 1, 2, ..., 2000$, respectively, with uniformly distributed ϕ_i and Gaussian errors without GARCH effects for $n = 100, 1000, 2500, 5000$ and $T = 5, 6, 10$. Table S.4 summarizes simulated outcomes with positive estimates of $\sigma_\phi^2 = Var(\phi_i)$ with uniformly distributed ϕ_i in the case of Gaussian errors without GARCH effects for the sample size combinations $n = 100, 1000, 5000$ and $T = 5, 6, 10$. The empirical power functions of FDAC and HetroGMM estimators of σ_ϕ^2 (for simulated outcomes of positive estimates) are shown in Figure S.2 with $n = 1000, 2500, 5000$ and $T = 5, 6, 10$.

For the four combinations of error distributions, Gaussian and non-Gaussian, without and with GARCH effects, Tables S.5 and S.6 summarize simulation results of the estimation of μ_ϕ and σ_ϕ^2 (for simulated outcomes of positive estimates), respectively, for uniformly distributed $\phi_i \in [-1 + \epsilon, 1]$ for some $\epsilon > 0$ with $\mu_\phi = 0.5$. Table S.7 reports the frequency where estimates of σ_ϕ^2 are not positive.

Table S.1: Bias, RMSE, and size of FDAC and HetroGMM estimators of $\mu_\phi = E(\phi_i)$ in a heterogeneous panel AR(1) model with uniformly distributed ϕ_i and Gaussian errors without GARCH effects

		Bias		RMSE		Size (×100)		Bias		RMSE		Size (×100)	
			Hetro		Hetro		Hetro		Hetro		Hetro		Hetro
T	n	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM
		$\mu_\phi = 0.4$ with $\lvert\phi_i\rvert < 1$						$\mu_\phi = 0.5$ with $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$					
4	100	-0.008	-0.043	0.174	0.297	7.0	5.2	0.003	-0.013	0.176	0.264	7.6	5.7
4	1,000	0.000	-0.004	0.057	0.086	5.1	5.8	0.000	0.003	0.057	0.080	5.1	5.2
4	5,000	0.000	-0.002	0.025	0.038	3.8	4.5	0.001	0.002	0.026	0.036	5.4	5.4
5	100	-0.003	0.006	0.134	0.163	6.3	7.3	-0.003	0.008	0.134	0.157	7.1	8.5
5	1,000	0.000	0.001	0.043	0.052	5.1	5.1	-0.001	0.002	0.042	0.051	5.1	5.8
5	5,000	0.000	0.000	0.019	0.023	4.4	4.9	0.001	0.002	0.019	0.022	4.3	4.3
6	100	-0.004	0.010	0.111	0.120	6.3	7.4	-0.004	0.009	0.112	0.119	7.1	8.6
6	1,000	-0.001	0.001	0.037	0.040	5.8	5.2	-0.002	0.001	0.035	0.039	4.5	6.2
6	5,000	0.000	0.000	0.016	0.018	4.4	5.2	0.000	0.001	0.016	0.017	4.7	4.9
10	100	0.000	0.013	0.078	0.077	6.5	10.8	-0.003	0.009	0.079	0.077	6.3	10.5
10	1,000	0.000	0.001	0.026	0.026	5.8	6.5	-0.001	0.001	0.025	0.026	4.8	5.7
10	5,000	0.000	0.000	0.011	0.012	5.3	5.4	0.000	0.001	0.011	0.012	5.3	5.8

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$ for $i = 1, 2, ..., n$, and $t = -99, -98, ..., T$, with $\varepsilon_{it} \sim IIDN(0,1)$ and cross-sectional heteroskedasticity, $h_{it} = \sigma_i$, where $\sigma_i^2 \sim IID(0.5 + 0.5z_i^2)$ and $z_i \sim IIDN(0,1)$. The heterogeneous AR(1) coefficients are generated by uniform distributions: $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi \in \{0.4, 0.5\}$. The initial values are generated as $(y_{i,-100} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$ for all i. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. FDAC and HetroGMM estimators of μ_ϕ are computed based on (6.2) and (6.13) in the main paper, respectively. The asymptotic variances are estimated by the Delta method. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$.

Table S.2: Bias, RMSE, and size of FDAC and HetroGMM estimators of $\mu_\phi = E(\phi_i)$ in a heterogeneous panel AR(1) model with categorically distributed ϕ_i and Gaussian errors without GARCH effects

		Bias		RMSE		Size (×100)		Bias		RMSE		Size (×100)	
			Hetro		Hetro		Hetro		Hetro		Hetro		Hetro
T	n	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM
		$\mu_\phi = 0.545$ with $\lvert\phi_i\rvert < 1$						$\mu_\phi = 0.525$ with $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$					
4	100	0.000	-0.034	0.164	0.264	7.5	6.8	-0.002	-0.028	0.169	0.267	7.4	6.7
4	1,000	-0.001	-0.005	0.053	0.077	5.2	5.0	0.001	-0.002	0.054	0.074	6.2	4.5
4	5,000	0.000	0.000	0.025	0.035	5.1	5.4	0.001	0.002	0.025	0.033	5.8	4.9
5	100	0.001	0.004	0.118	0.144	7.0	7.0	-0.003	0.003	0.118	0.142	5.9	7.0
5	1,000	-0.001	-0.002	0.037	0.045	4.4	4.7	0.001	0.001	0.039	0.046	6.4	5.1
5	5,000	-0.001	0.000	0.017	0.021	4.9	5.2	0.001	0.002	0.017	0.021	5.8	5.2
6	100	-0.001	0.006	0.099	0.108	6.9	8.6	-0.001	0.007	0.097	0.107	6.4	7.8
6	1,000	0.000	-0.001	0.031	0.034	5.1	5.1	0.001	0.002	0.032	0.035	5.3	5.9
6	5,000	0.000	0.000	0.014	0.016	4.8	5.7	0.001	0.001	0.014	0.016	6.0	5.5
10	100	-0.001	0.004	0.064	0.064	5.9	8.8	0.000	0.006	0.066	0.067	5.4	10.8
10	1,000	0.000	0.000	0.021	0.021	4.2	5.3	0.001	0.002	0.021	0.021	4.6	5.4
10	5,000	0.000	0.000	0.009	0.010	5.0	5.7	0.000	0.001	0.010	0.010	5.4	4.9

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -M_i + 1, -M_i + 2, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity. The heterogeneous AR(1) coefficients are generated by categorical distributions: $\Pr(\phi_i = \phi_L) = \pi$ and $\Pr(\phi_i = \phi_H) = 1 - \pi$, where $(\phi_H, \phi_L, \pi) = (0.8, 0.5, 0.85)$ with $|\phi_i| < 1$ for all i and $(1, 0.5, 0.95)$ with $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ and all i. The initial values are given by $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$, where $M_i = 100$ for units with $|\phi_i| < 1$, and $M_i = 1$ for units with $\phi_i = 1$. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. FDAC and HetroGMM estimators of μ_ϕ are computed based on (6.2) and (6.13) in the main paper, respectively. The asymptotic variances are estimated by the Delta method. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$.

Figure S.1: Empirical power functions for FDAC and HetroGMM estimators of $\mu_\phi = E(\phi_i)$ ($\mu_{\phi,0} = 0.5$) in a heterogeneous AR(1) panel with uniformly distributed $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ and Gaussian errors without GARCH effects

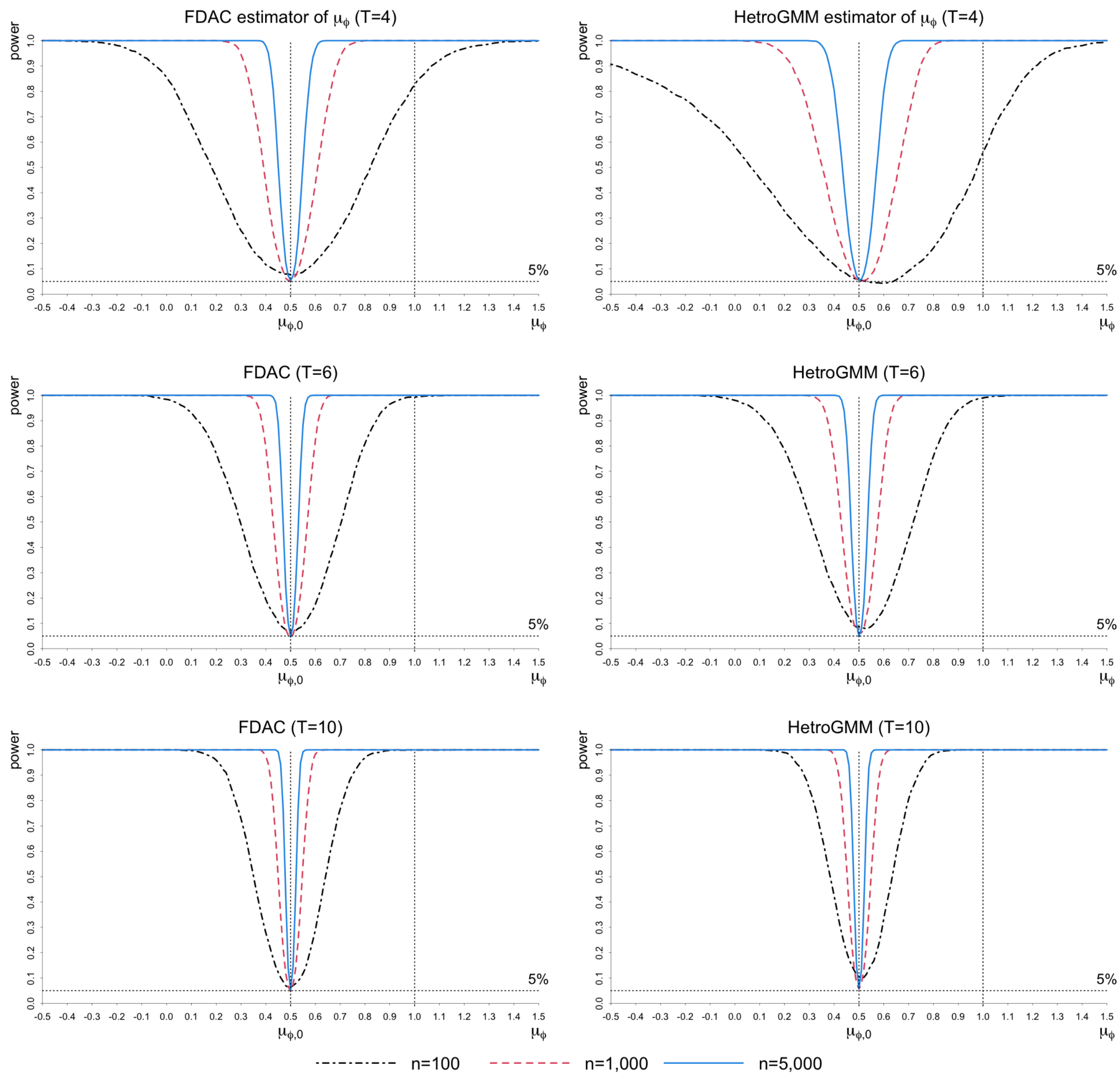


Table S.3: Frequency of FDAC and HetroGMM estimators of $\sigma_\phi^2 = Var(\phi_i)$ being negative with uniformly distributed ϕ_i and Gaussian errors without GARCH effects

		$\sigma_\phi^2 = 0.083$ with $\lvert\phi_i\rvert < 1$		$\sigma_\phi^2 = 0.083$ with $\phi_i \in [-1+\epsilon, 1]$ $(\epsilon > 0)$	
T	n	FDAC	Hetro GMM	FDAC	Hetro GMM
5	100	34.4	42.1	33.4	45.2
5	1,000	5.9	21.3	6.4	20.6
5	2,500	0.8	9.9	0.7	9.8
5	5,000	0.0	2.8	0.0	2.0
6	100	28.2	34.3	28.1	36.3
6	1,000	1.9	8.4	1.7	8.2
6	2,500	0.1	1.8	0.2	1.4
6	5,000	0.0	0.2	0.0	0.1
10	100	15.6	20.6	15.4	19.9
10	1,000	0.1	0.0	0.1	0.2
10	2,500	0.0	0.0	0.0	0.0
10	5,000	0.0	0.0	0.0	0.0

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -99, -M_i+2, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity. The heterogeneous AR(1) coefficients are generated by uniform distributions: $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi \in \{0.4, 0.5\}$. The initial values are given by $(y_{i,-100} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$ for all i. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator of σ_ϕ^2 is computed by plugging (6.2) and (6.3) into (6.22) in the main paper, and the HetroGMM estimator of σ_ϕ^2 is computed by plugging (6.13) and (6.19) into (6.22) in the main paper. The asymptotic variances are estimated by the Delta method. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The figure in the cell denotes the frequency (multiplied by 100) of occurrences where the estimate of σ_ϕ^2 is negative or close to zero, $\left(\hat{\sigma}_\phi^2\right)^{(r)} < 0.0001$, for replication r over 2,000 replications.

Table S.4: Bias, RMSE, and size of FDAC and HetroGMM estimators of $\sigma_\phi^2 = Var(\phi_i)$ in a heterogeneous panel AR(1) model with uniformly distributed ϕ_i and Gaussian errors without GARCH effects

		Bias		RMSE		Size (×100)		Bias		RMSE		Size (×100)	
			Hetro		Hetro		Hetro		Hetro		Hetro		Hetro
T	n	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM
		$\sigma_\phi^2 = 0.083$ with $\lvert\phi_i\rvert < 1$ and $\mu_\phi = 0.4$						$\sigma_\phi^2 = 0.083$ with $\phi_i \in [-1+\epsilon, 1]$ for $\epsilon > 0$ and $\mu_\phi = 0.5$					
5	1,000	0.005	0.032	0.047	0.079	3.0	3.2	0.006	0.029	0.047	0.076	2.0	2.6
5	2,500	0.000	0.012	0.033	0.053	4.6	2.9	0.000	0.010	0.031	0.053	4.0	2.8
5	5,000	-0.001	0.002	0.024	0.041	4.9	2.8	0.000	0.003	0.023	0.041	4.9	3.7
6	1,000	0.001	0.007	0.038	0.052	3.8	3.0	0.002	0.008	0.038	0.050	4.5	2.4
6	2,500	0.000	0.001	0.026	0.036	5.5	3.7	0.000	0.001	0.025	0.035	5.1	3.3
6	5,000	0.000	0.000	0.018	0.027	5.8	5.0	0.001	0.001	0.018	0.026	5.1	5.1
10	1,000	-0.001	-0.002	0.023	0.027	4.0	5.0	0.000	-0.002	0.024	0.027	5.3	5.4
10	2,500	0.000	-0.001	0.015	0.017	5.2	6.0	0.000	-0.001	0.015	0.017	4.6	5.2
10	5,000	0.000	-0.001	0.011	0.012	5.4	5.8	0.000	0.000	0.010	0.012	4.4	4.5

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$ for $i = 1, 2, ..., n$, and $t = -99, -98, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects, where the heterogeneous AR(1) coefficients are generated by uniform distributions. The FDAC estimator of σ_ϕ^2 is computed by plugging (6.2) and (6.3) into (6.22), and the HetroGMM estimator of σ_ϕ^2 is computed by plugging (6.13) and (6.19) into (6.22) in the main paper. The asymptotic variances are estimated by the Delta method. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The total number of replications is $2,000$. But the reported results are based on simulated outcomes with $\left(\hat{\sigma}_\phi^2\right)^{(r)} \geq 0.0001$. The frequencies with negative outcomes, by sample sizes and estimation method, are reported in Table S.3 of the online supplement. See also the footnotes to Table S.1 for further details of the DGP used.

Figure S.2: Empirical power functions for FDAC and HetroGMM estimators of $\sigma_\phi^2 = Var(\phi_i)$ ($\sigma_{\phi,0}^2 = 0.083$) in a heterogeneous AR(1) panel with uniformly distributed $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ and Gaussian errors without GARCH effects

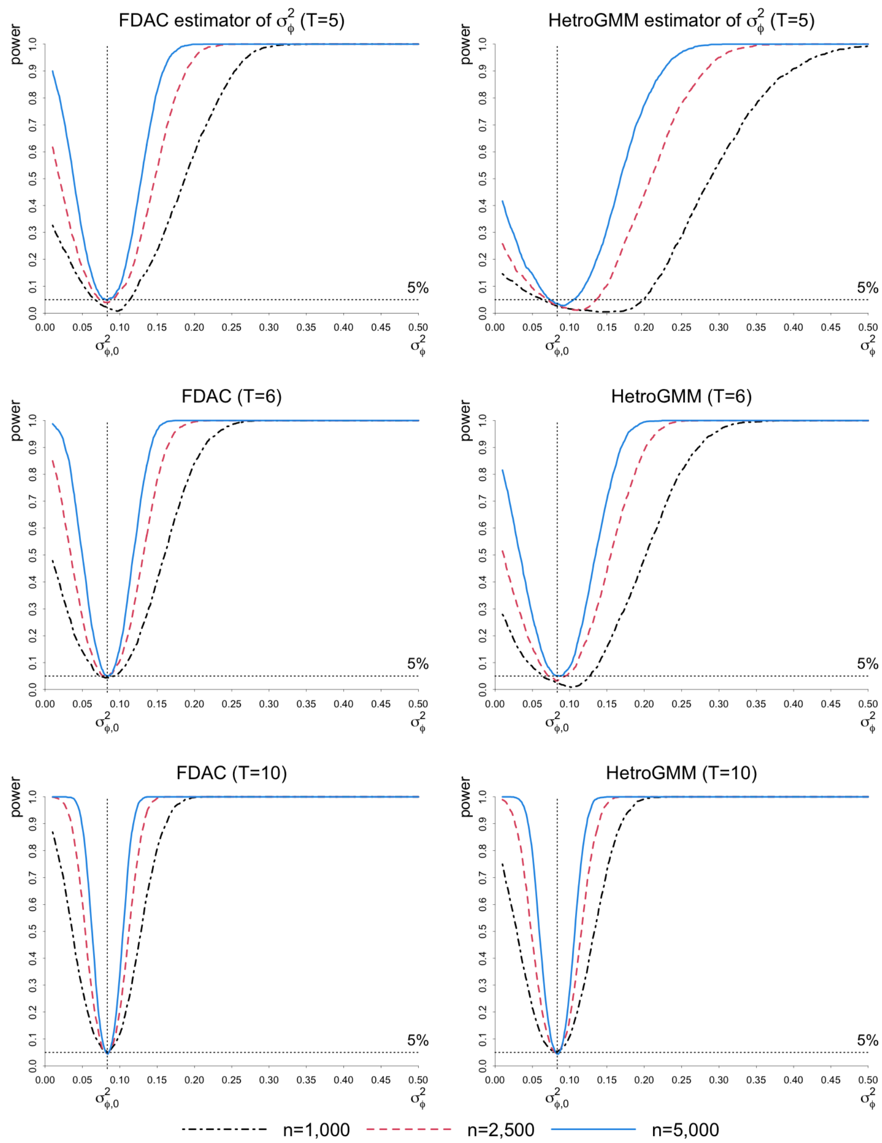


Table S.5: Bias, RMSE, and size of FDAC and HetroGMM estimators of $\mu_\phi = E(\phi_i) = 0.5$ in a heterogeneous panel AR(1) model with uniformly distributed $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ under different error processes

		Bias		RMSE		Size (×100)		Bias		RMSE		Size (×100)	
			Hetro		Hetro		Hetro		Hetro		Hetro		Hetro
T	n	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM	FDAC	GMM
		Gaussian errors without GARCH effects						Non-Gaussian errors without GARCH effects					
4	100	0.003	-0.013	0.176	0.264	7.6	5.7	-0.008	-0.113	0.214	0.607	9.5	8.3
4	1,000	0.000	0.003	0.057	0.080	5.1	5.2	0.002	-0.008	0.071	0.116	5.8	5.6
4	5,000	0.001	0.002	0.026	0.036	5.4	5.4	0.000	-0.002	0.031	0.050	5.2	5.2
5	100	-0.003	0.008	0.134	0.157	7.1	8.5	-0.005	0.019	0.151	0.189	7.4	11.1
5	1,000	-0.001	0.002	0.042	0.051	5.1	5.8	0.001	0.004	0.051	0.064	5.4	5.8
5	5,000	0.001	0.002	0.019	0.022	4.3	4.3	0.000	0.002	0.023	0.029	5.0	5.1
6	100	-0.004	0.009	0.112	0.119	7.1	8.6	-0.002	0.027	0.125	0.134	6.5	11.5
6	1,000	-0.002	0.001	0.035	0.039	4.5	6.2	0.001	0.006	0.042	0.047	5.5	6.4
6	5,000	0.000	0.001	0.016	0.017	4.7	4.9	0.000	0.001	0.019	0.021	5.6	5.1
10	100	-0.003	0.009	0.079	0.077	6.3	10.5	-0.001	0.023	0.085	0.081	6.4	14.4
10	1,000	-0.001	0.001	0.025	0.026	4.8	5.7	0.002	0.006	0.028	0.028	5.9	7.1
10	5,000	0.000	0.001	0.011	0.012	5.3	5.8	0.000	0.001	0.013	0.013	4.9	5.0
		Gaussian errors with GARCH effects						Non-Gaussian errors with GARCH effects					
4	100	0.000	-0.017	0.205	0.306	9.0	6.4	-0.028	-0.082	0.302	1.029	13.5	8.3
4	1,000	0.000	0.002	0.069	0.095	5.7	5.2	-0.002	-0.016	0.117	0.181	6.1	5.7
4	5,000	0.000	0.001	0.031	0.043	6.2	5.4	-0.001	-0.003	0.059	0.087	5.1	4.3
5	100	-0.004	0.010	0.159	0.178	9.1	8.9	-0.015	0.008	0.215	0.251	11.1	12.3
5	1,000	-0.001	0.003	0.051	0.061	5.3	6.2	-0.001	0.004	0.085	0.092	6.7	5.9
5	5,000	0.000	0.002	0.023	0.027	4.7	4.9	0.000	0.001	0.045	0.045	5.3	4.9
6	100	-0.006	0.009	0.133	0.137	8.1	10.5	-0.007	0.017	0.181	0.166	10.5	14.0
6	1,000	-0.002	0.002	0.042	0.046	4.6	6.0	0.000	0.002	0.072	0.065	5.7	6.9
6	5,000	0.000	0.002	0.019	0.021	5.0	4.9	-0.001	-0.001	0.038	0.034	5.4	6.0
10	100	-0.005	0.008	0.095	0.087	6.9	13.2	-0.005	0.008	0.130	0.104	9.4	19.4
10	1,000	-0.001	0.001	0.031	0.030	4.9	6.3	0.001	-0.001	0.053	0.040	6.0	9.7
10	5,000	0.000	0.001	0.014	0.014	5.7	6.1	0.000	-0.002	0.027	0.020	5.4	5.9

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -99, -98, ..., T$, where the heterogeneous AR(1) coefficients are generated by the uniform distribution: $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi = 0.5$. The standardized errors, ε_{it}, are generated as Gaussian, $\varepsilon_{it} \sim IIDN(0,1)$, or non-Gaussian, $\varepsilon_{it} = (e_{it} - 2)/2$ with $e_{it} \sim IID\chi^2_2$. The GARCH effect is generated as $h^2_{it} = \sigma^2_i(1-\psi_0-\psi_1) + \psi_0 h^2_{i,t-1} + \psi_1(h_{i,t-1}\varepsilon_{i,t-1})^2$, with $\sigma^2_i \sim IID\left(0.5 + 0.5z^2_i\right)$ and $z_i \sim IIDN(0,1)$, where $\psi_0 = 0.6$ and $\psi_1 = 0.2$, with $h_{i,-M_i} = \sigma_i$. In the case of no GARCH effects, $\psi_0 = \psi_1 = 0$. The initial values are given by $(y_{i,-100} - \mu_i) \sim IIDN(b, \kappa\sigma^2_i)$ with $b = 1$ and $\kappa = 2$ for all i. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator is calculated based on (6.2) in the main paper, and its asymptotic variance is estimated by the Delta method. The HetroGMM estimator and its asymptotic variance are calculated by (6.13) and (6.14) in the main paper. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$.

Table S.6: Bias, RMSE, and size of FDAC and HetroGMM estimators of $\sigma^2_\phi = Var(\phi_i) = 0.083$ in a heterogeneous panel AR(1) model with uniformly distributed $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ under different error processes

		Bias		RMSE		Size (×100)		Bias		RMSE		Size (×100)	
T	n	FDAC	Hetro GMM	FDAC	Hetro GMM	FDAC	Hetro GMM	FDAC	Hetro GMM	FDAC	Hetro GMM	FDAC	Hetro GMM
		Gaussian errors without GARCH effects						Non-Gaussian errors without GARCH effects					
5	1,000	0.006	0.029	0.047	0.076	2.0	2.6	0.005	0.039	0.047	0.088	2.3	4.4
5	2,500	0.000	0.010	0.031	0.053	4.0	2.8	0.000	0.013	0.033	0.059	5.0	4.1
5	5,000	0.000	0.003	0.023	0.041	4.9	3.7	0.001	0.004	0.024	0.046	4.4	3.0
6	1,000	0.002	0.008	0.038	0.050	4.5	2.4	-0.001	0.011	0.038	0.055	3.3	3.4
6	2,500	0.000	0.001	0.025	0.035	5.1	3.3	-0.002	-0.001	0.026	0.039	5.5	2.6
6	5,000	0.001	0.001	0.018	0.026	5.1	5.1	0.000	-0.002	0.018	0.030	5.1	4.2
10	1,000	0.000	-0.002	0.024	0.027	5.3	5.4	-0.002	-0.003	0.024	0.029	5.2	6.3
10	2,500	0.000	-0.001	0.015	0.017	4.6	5.2	-0.001	-0.002	0.015	0.018	5.1	5.8
10	5,000	0.000	0.000	0.010	0.012	4.4	4.5	0.000	-0.001	0.011	0.013	5.6	5.9
		Gaussian errors with GARCH effects						Non-Gaussian errors with GARCH effects					
5	1,000	0.012	0.043	0.054	0.093	2.2	3.0	0.024	0.080	0.076	0.151	3.4	5.1
5	2,500	0.002	0.016	0.037	0.061	2.9	3.3	0.010	0.046	0.056	0.101	3.4	4.8
5	5,000	0.001	0.008	0.028	0.047	4.3	3.5	0.004	0.025	0.044	0.076	4.3	3.5
6	1,000	0.004	0.016	0.044	0.059	3.3	2.4	0.011	0.032	0.058	0.087	2.8	3.8
6	2,500	0.000	0.004	0.029	0.041	4.6	2.8	0.003	0.014	0.044	0.059	4.1	3.9
6	5,000	0.001	0.003	0.022	0.031	4.8	4.6	0.002	0.006	0.034	0.046	4.9	3.8
10	1,000	0.000	-0.002	0.029	0.031	6.1	5.1	0.001	-0.002	0.039	0.039	4.4	6.0
10	2,500	0.000	-0.001	0.018	0.021	4.8	5.3	0.000	-0.002	0.028	0.028	6.0	5.9
10	5,000	0.000	0.000	0.013	0.014	5.0	4.6	0.001	-0.001	0.021	0.021	5.5	5.2

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -99, -98, ..., T$, where the heterogeneous AR(1) coefficients are generated by the uniform distribution: $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi = 0.5$. The standardized errors, ε_{it}, are generated as Gaussian, $\varepsilon_{it} \sim IIDN(0,1)$, or non-Gaussian, $\varepsilon_{it} = (e_{it} - 2)/2$ with $e_{it} \sim IID\chi^2_2$. The GARCH effect is generated as $h^2_{it} = \sigma^2_i(1-\psi_0-\psi_1) + \psi_0 h^2_{i,t-1} + \psi_1(h_{i,t-1}\varepsilon_{i,t-1})^2$, with $\sigma^2_i \sim IID\left(0.5 + 0.5z^2_i\right)$ and $z_i \sim IIDN(0,1)$, where $\psi_0 = 0.6$ and $\psi_1 = 0.2$, with $h_{i,-M_i} = \sigma_i$. In the case of no GARCH effects, $\psi_0 = \psi_1 = 0$. The initial values are given by $(y_{i,-100} - \mu_i) \sim IIDN(b, \kappa\sigma^2_i)$ with $b = 1$ and $\kappa = 2$ for all i. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator of σ^2_ϕ is calculated by plugging (6.2) and (6.3) into (6.22) in the main paper. The HetroGMM estimator of σ^2_ϕ is calculated by plugging (6.13) and (6.19) into (6.22) in the main paper. The asymptotic variances are estimated by the Delta method. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$. But the reported results are based on simulated outcomes with $\left(\hat{\sigma}^2_\phi\right)^{(r)} \geq 0.0001$. The frequencies with negative outcomes, by sample sizes and estimation method, are reported in Table S.7.

Table S.7: Frequency of FDAC and HetroGMM estimators of $\sigma_\phi^2 = Var(\phi_i) = 0.083$ being negative with uniformly distributed $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ under different error processes

		Without GARCH effects				With GARCH effects			
		Gaussian		Non-Gaussian		Gaussian		Non-Gaussian	
T	n	FDAC	Hetro GMM	FDAC	Hetro GMM	FDAC	Hetro GMM	FDAC	Hetro GMM
5	1,000	6.3	20.6	6.7	27.2	10.4	24.6	18.6	32.9
5	2,500	0.7	9.8	1.4	15.6	2.2	12.9	8.8	26.0
5	5,000	0.0	1.9	0.0	7.2	0.1	4.7	3.4	17.8
6	1,000	1.7	8.2	2.4	13.4	3.6	11.1	11.5	19.4
6	2,500	0.1	1.3	0.1	3.6	0.4	2.8	3.9	10.6
6	5,000	0.0	0.0	0.0	0.4	0.0	0.3	1.7	4.8
10	1,000	0.1	0.2	0.0	0.4	0.3	0.6	2.8	2.7
10	2,500	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.4
10	5,000	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -99, -M_i + 2, ..., T$ where the heterogeneous AR(1) coefficients are generated by uniform distributions with $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ and all i. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The figure denotes the frequency (multiplied by 100) of occurrences where the estimate of σ_ϕ^2 is negative or close to zero, $\left(\hat{\sigma}_\phi^2\right)^{(r)} < 0.0001$, for replication r over 2,000 replications. See also the footnotes to Table S.6.

S.8.2 Comparison of the FDAC estimator with FDLS, AH, AAH, AB, and BB estimators

Tables S.8–S.9 report bias, RMSE, and size of the FDAC, FDLS, AH, AAH, AB, and BB estimators with $\phi_i = \mu_\phi + v_i$, $v_i \sim IIDU(-a, a)$, $\mu_\phi \in \{0.4, 0.5\}$, $a = 0.5$, and Gaussian errors without GARCH effects. Table S.10 summarizes simulation results of FDAC and the above HomoGMM estimators with homogeneous $\phi_i = \mu_\phi = 0.5$ and Gassuain errors without GARCH effects.

Figure S.3 compares the empirical power functions of FDAC and FDLS estimators under homogeneity of ϕ_i for $T = 4, 10$, and $n = 5,000$. Figures S.4 and S.5 plot the empirical power functions of the FDAC estimator in homogeneous ($\phi_i = \mu_\phi = 0.5$ for all i) and heterogeneous panel AR(1) panels, where the heterogeneous AR(1) coefficients are generated by the above uniform distribution with $\phi_i \in (-1, 1]$ and $\mu_\phi = 0.5$, under different error processes for $T = 4, 10$ and $n = 100, 1000, 5000$.

Table S.8: Bias, RMSE, and size of FDAC, FDLS, AH, AAH, AB, and BB estimators of $\mu_\phi = E(\phi_i) = 0.4$ in a heterogeneous panel AR(1) model with uniformly distributed $|\phi_i| < 1$ and Gaussian errors without GARCH effects

		Bias						RMSE						Size (×100)					
T	n	FDAC	FDLS	AH	AAH	AB	BB	FDAC	FDLS	AH	AAH	AB	BB	FDAC	FDLS	AH	AAH	AB	BB
4	100	0.005	-0.057	-0.124	-0.007	-0.081	-0.041	0.182	0.168	1.930	0.261	0.236	0.157	8.2	8.2	15.5	18.2	14.9	16.2
4	1,000	-0.001	-0.061	-0.176	-0.066	-0.061	-0.046	0.058	0.079	0.223	0.109	0.092	0.066	6.7	23.2	33.6	30.8	19.4	19.8
4	5,000	0.000	-0.062	-0.185	-0.073	-0.055	-0.044	0.026	0.065	0.195	0.077	0.063	0.049	5.4	78.1	85.8	81.4	46.8	56.4
6	100	0.001	-0.060	-0.178	-0.025	-0.086	-0.026	0.114	0.127	0.258	0.123	0.163	0.107	7.4	10.3	32.8	18.4	27.7	24.4
6	1,000	0.000	-0.061	-0.143	-0.033	-0.052	-0.022	0.036	0.072	0.155	0.046	0.068	0.039	4.1	40.8	71.5	18.9	30.0	15.5
6	5,000	0.000	-0.062	-0.140	-0.033	-0.046	-0.020	0.016	0.064	0.142	0.036	0.050	0.024	5.0	96.8	100.0	60.1	70.6	31.8
10	100	0.000	-0.061	-0.121	-0.031	-0.060	-0.018	0.078	0.102	0.159	0.084	0.110	0.080	5.2	13.1	57.5	41.5	48.1	48.4
10	1,000	0.000	-0.061	-0.092	-0.022	-0.033	-0.002	0.025	0.067	0.098	0.033	0.044	0.024	4.8	63.6	85.2	17.7	32.9	11.5
10	5,000	0.000	-0.061	-0.090	-0.021	-0.029	0.000	0.011	0.063	0.091	0.024	0.032	0.011	5.2	100.0	100.0	44.8	70.9	9.6

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -99, -98, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The heterogeneous AR(1) coefficients are generated by the uniform distribution: $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi = 0.4$. The initial values are given by $(y_{i,-100} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$ for all i. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator is calculated by (6.2) in the main paper, and its asymptotic variance is estimated by the Delta method. "FDLS" denotes the first difference least square estimator proposed by Han and Phillips (2010). "AH", "AAH", "AB", and "BB" denote the 2-step GMM estimators proposed by Anderson and Hsiao (1981, 1982), Chudik and Pesaran (2021), Arellano and Bond (1991), and Blundell and Bond (1998). The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$.

Table S.9: Bias, RMSE, and size of FDAC, FDLS, AH, AAH, AB, and BB estimators of $\mu_\phi = E(\phi_i) = 0.5$ in a heterogeneous panel AR(1) model with uniformly distributed $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ and Gaussian errors without GARCH effects

		Bias						RMSE						Size (×100)					
T	n	FDAC	FDLS	AH	AAH	AB	BB	FDAC	FDLS	AH	AAH	AB	BB	FDAC	FDLS	AH	AAH	AB	BB
4	100	0.003	-0.053	-0.070	-0.012	-0.109	-0.034	0.181	0.171	2.255	0.263	0.305	0.166	8.3	8.2	15.2	16.1	16.3	17.2
4	1,000	-0.001	-0.057	-0.190	-0.062	-0.072	-0.043	0.057	0.076	0.256	0.125	0.113	0.068	6.2	19.8	30.0	29.5	18.5	16.4
4	5,000	0.000	-0.057	-0.204	-0.075	-0.065	-0.043	0.025	0.062	0.216	0.080	0.076	0.050	5.1	70.1	78.6	74.1	40.6	45.2
6	100	0.001	-0.056	-0.224	-0.021	-0.119	-0.022	0.111	0.127	0.314	0.133	0.202	0.114	7.2	9.6	35.8	21.0	30.8	27.8
6	1,000	0.000	-0.057	-0.171	-0.034	-0.074	-0.021	0.035	0.069	0.184	0.047	0.092	0.040	4.4	35.0	73.7	20.5	38.0	15.1
6	5,000	0.000	-0.057	-0.166	-0.034	-0.068	-0.020	0.016	0.060	0.169	0.037	0.073	0.025	5.1	93.3	100.0	61.2	84.3	29.1
10	100	0.000	-0.057	-0.164	-0.024	-0.086	-0.016	0.077	0.101	0.201	0.087	0.136	0.085	5.2	12.0	66.4	41.0	54.5	51.5
10	1,000	0.000	-0.057	-0.123	-0.018	-0.059	-0.003	0.025	0.063	0.129	0.031	0.069	0.025	4.6	55.3	94.6	13.0	58.1	11.9
10	5,000	0.000	-0.057	-0.120	-0.017	-0.058	0.000	0.011	0.058	0.121	0.021	0.061	0.012	4.9	99.8	100.0	29.5	97.3	10.2

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -99, -98, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The heterogeneous AR(1) coefficients are generated by the uniform distribution: $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi = 0.5$. The initial values are given by $(y_{i,-100} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$ for all i. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator is calculated by (6.2) in the main paper, and its asymptotic variance is estimated by the Delta method. "FDLS" denotes the first difference least square estimator proposed by Han and Phillips (2010). "AH", "AAH", "AB", and "BB" denote the 2-step GMM estimators proposed by Anderson and Hsiao (1981, 1982), Chudik and Pesaran (2021), Arellano and Bond (1991), and Blundell and Bond (1998). The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$.

Table S.10: Bias, RMSE, and size of FDAC, FDLS, AH, AAH, AB, and BB estimators of ϕ ($\phi_0 = 0.5$) in a homogeneous panel AR(1) model with Gaussian errors without GARCH effects

		Bias						RMSE						Size (×100)					
T	n	FDAC	FDLS	AH	AAH	AB	BB	FDAC	FDLS	AH	AAH	AB	BB	FDAC	FDLS	AH	AAH	AB	BB
4	100	0.004	0.002	0.564	0.084	-0.023	-0.006	0.169	0.150	9.211	0.298	0.226	0.141	7.7	5.9	8.3	12.7	10.0	13.2
4	1,000	0.000	0.001	0.021	0.072	-0.004	-0.002	0.054	0.049	0.206	0.209	0.070	0.044	6.2	5.6	4.2	18.4	5.9	5.2
4	5,000	0.000	0.000	0.002	0.023	0.000	0.000	0.024	0.022	0.086	0.116	0.031	0.020	5.4	4.9	5.2	9.2	5.0	5.0
6	100	0.001	-0.001	-0.072	0.023	-0.041	-0.001	0.098	0.104	0.220	0.152	0.130	0.088	8.1	6.6	15.9	22.6	16.9	21.1
6	1,000	0.000	0.000	-0.007	0.001	-0.006	0.000	0.031	0.034	0.066	0.040	0.038	0.026	4.5	5.9	6.4	6.7	6.4	6.3
6	5,000	0.000	0.000	-0.001	0.000	-0.001	0.000	0.014	0.015	0.030	0.014	0.017	0.012	4.9	5.3	5.4	5.2	6.5	5.8
10	100	0.000	-0.001	-0.049	0.006	-0.035	-0.004	0.063	0.074	0.103	0.073	0.080	0.057	5.1	6.0	34.0	44.3	39.6	43.6
10	1,000	0.000	0.000	-0.004	0.000	-0.004	0.000	0.020	0.024	0.028	0.016	0.021	0.016	5.0	5.1	8.5	8.8	8.7	8.6
10	5,000	0.000	0.000	-0.001	0.000	0.000	0.000	0.009	0.011	0.013	0.007	0.009	0.007	5.6	4.7	5.2	5.9	6.2	5.3

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -99, -98, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The AR(1) coefficients are generated to be homogeneous: $\phi_i = \mu_\phi = 0.5$ for all i. The initial values are given by $(y_{i,-100} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$ for all i. For each experiment, $(\alpha_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator is calculated by (6.2) in the main paper, and its asymptotic variance is estimated by the Delta method. "FDLS" denotes the first difference least square estimator proposed by Han and Phillips (2010). "AH", "AAH", "AB", and "BB" denote the 2-step GMM estimators proposed by Anderson and Hsiao (1981, 1982), Chudik and Pesaran (2021), Arellano and Bond (1991), and Blundell and Bond (1998). The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$.

Figure S.3: Empirical power functions for FDAC and FDLS estimators of $\phi_0 = 0.5$ in a homogeneous AR(1) panel with Gaussian errors without GARCH effects

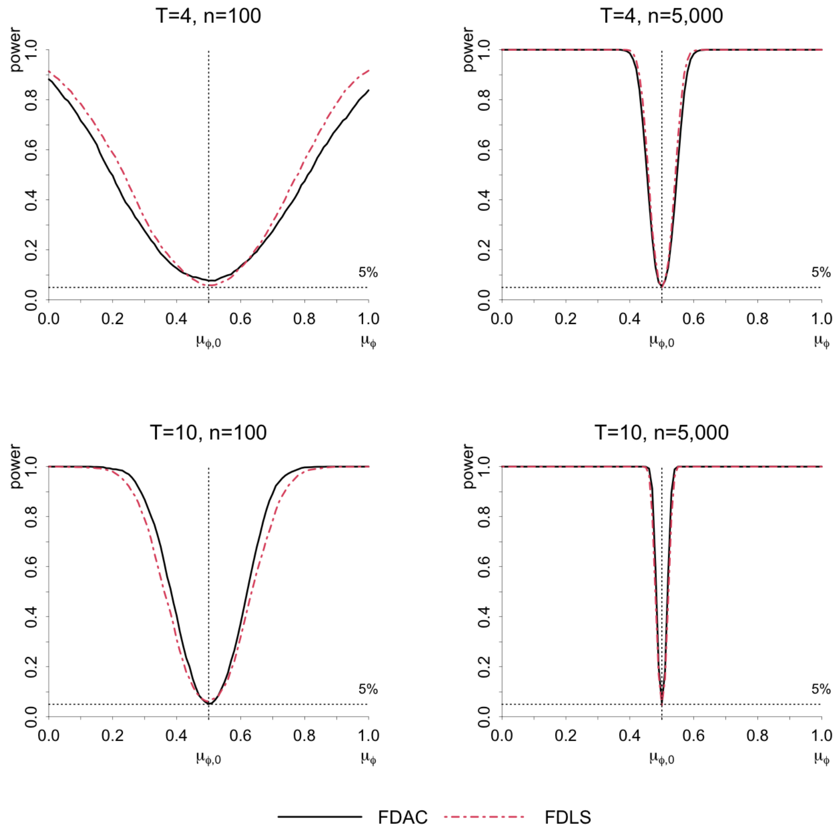


Figure S.4: Empirical power functions for the FDAC estimator of $\mu_\phi = E(\phi_i)$ ($\mu_{\phi,0} = 0.5$) in homogeneous and heterogeneous AR(1) panels with Gaussian and non-Gaussian error processes without GARCH effects

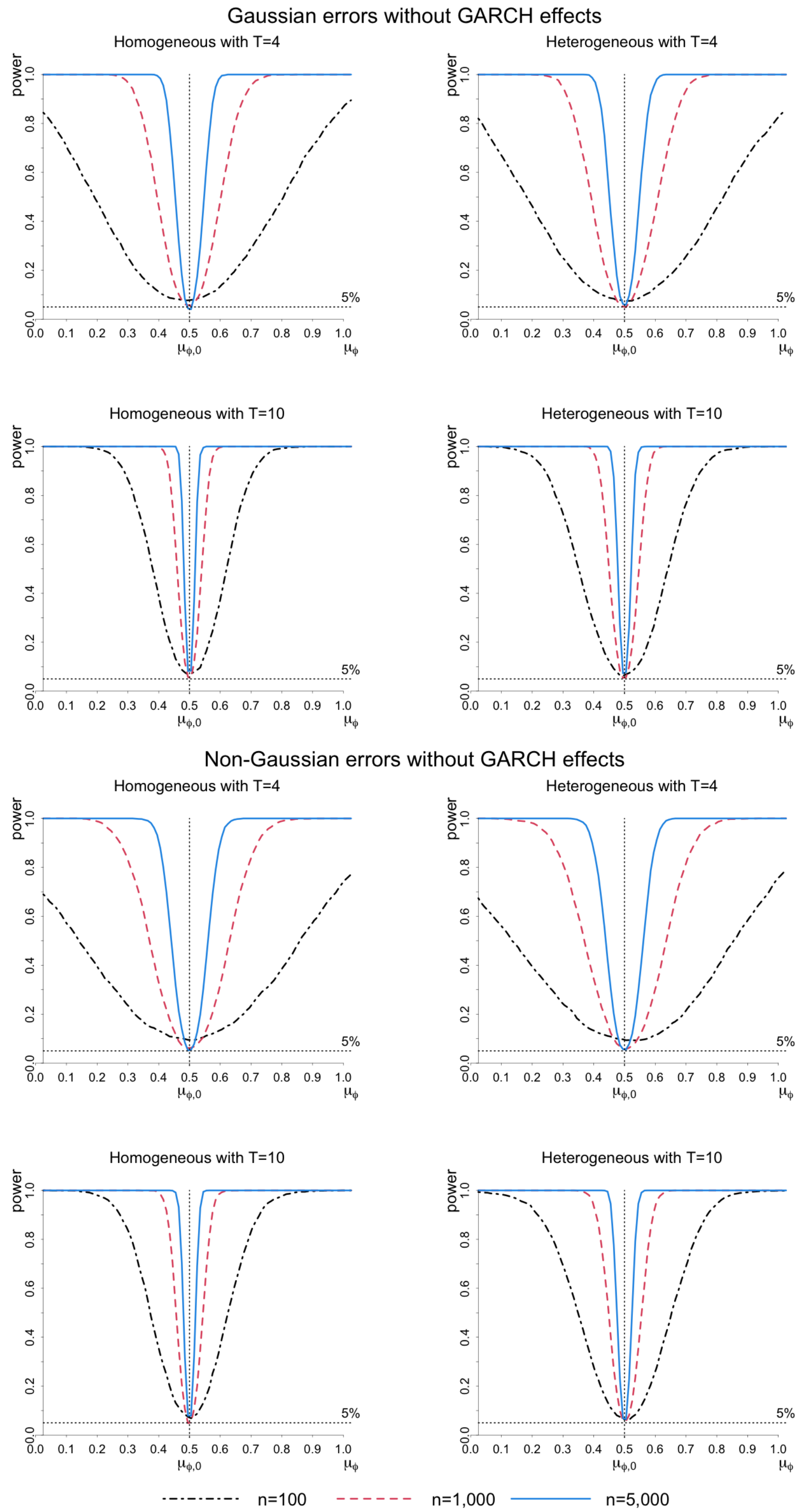
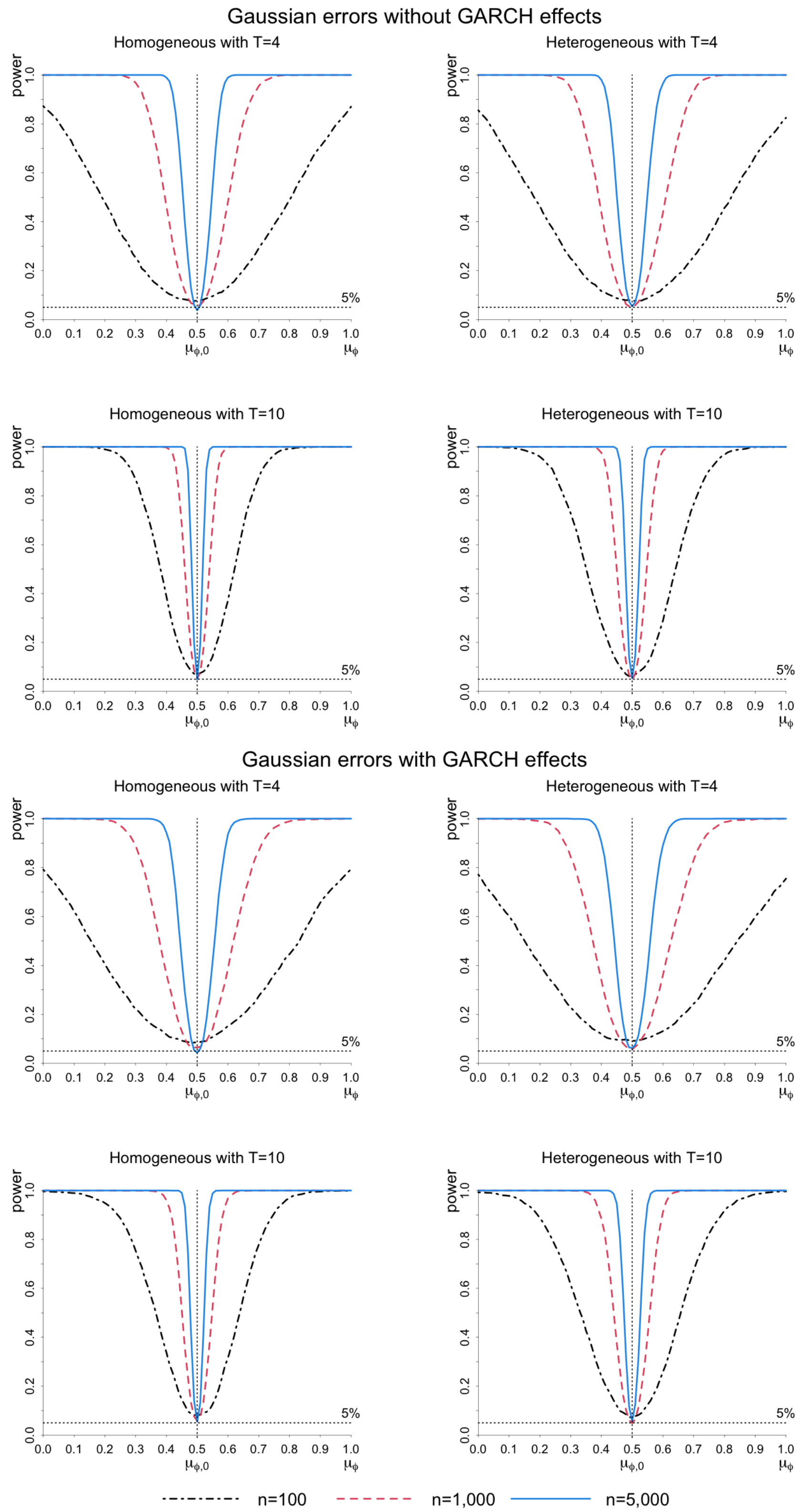


Figure S.5: Empirical power functions for the FDAC estimator of $\mu_\phi = E(\phi_i)$ ($\mu_{\phi,0} = 0.5$) in homogeneous and heterogeneous AR(1) panels with Gaussian errors without and with GARCH effects



S.8.3 Comparison of FDAC and MSW estimator

Table S.11 summarizes bias, RMSE, and size of FDAC and MSW estimators of $\mu_\phi = E(\phi_i)$ with uniformly distributed ϕ_i in the case of Gaussian errors without GARCH effects for the sample size combinations $n = 100, 1000$ and $T = 4, 6, 10$.

Table S.11: Bias, RMSE, and size of FDAC and MSW estimators of $\mu_\phi = E(\phi_i)$ in heterogeneous panel AR(1) models with uniformly distributed ϕ_i and Gaussian errors without GARCH effects

		$\mu_\phi = 0.4$ with $\lvert\phi_i\rvert < 1$						$\mu_\phi = 0.5$ with $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$					
		Bias		RMSE		Size ($\times$100)		Bias		RMSE		Size ($\times$100)	
T	n	FDAC	MSW	FDAC	MSW	FDAC	MSW	FDAC	MSW	FDAC	MSW	FDAC	MSW
4	100	-0.005	-0.145	0.177	0.157	8.0	82.0	-0.005	-0.207	0.175	0.221	8.7	84.3
4	1,000	0.000	-0.128	0.056	0.130	4.7	100.0	0.000	-0.194	0.056	0.196	5.0	100.0
6	100	-0.004	-0.144	0.113	0.155	5.7	79.3	-0.004	-0.202	0.111	0.215	5.5	81.2
6	1,000	-0.001	-0.129	0.037	0.130	6.3	100.0	-0.001	-0.187	0.036	0.189	5.2	100.0
10	100	-0.001	-0.146	0.079	0.158	6.4	71.2	-0.001	-0.198	0.079	0.213	6.7	74.3
10	1,000	0.000	-0.141	0.026	0.143	5.7	100.0	-0.001	-0.192	0.025	0.194	5.9	100.0

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -99, -98, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects, where the heterogeneous AR(1) coefficients are generated by uniform distributions. The FDAC estimator is calculated by (6.2) in the main paper. The asymptotic variance is estimated by the Delta method. "MSW" denotes the estimator proposed by Mavroeidis et al. (2015). The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. Due to the extensive computations required for the implementation of the MSW estimator, the number of replications is $1,000$. See also footnotes to Table S.1.

S.8.4 Simulation results with different initializations

Tables S.12 and S.13 summarize the bias, RMSE, and size of the FDAC estimator of $E(\phi_i)$ with uniformly and categorically distributed ϕ_i, respectively, under different initializations $M_i = 100, 3, 1$ for all i, (except a case of categorically distributed ϕ_i where $M_i = 100$ for units with $\phi_i = \phi_L = 0.5$ and $M_i = 1$ for units with $\phi_i = \phi_H = 1$). Table S.14 reports the bias, RMSE, and sizes of the FDAC, FDLS, AH, AAH, AB, and BB estimators in homogeneous panels for $M_i = 100, 3, 1$ for all i. The simulation results for heterogeneous panels with uniformly distributed ϕ_i are shown in Table S.15 for $\mu_\phi = 0.4$, and Table S.16 for $\mu_\phi = 0.5$. Table S.17 summarizes results of FDAC and MSW estimators in both homogeneous and heterogeneous panels for different initializations with $M_i = 100, 1$ for all i. .

Table S.12: Bias, RMSE, and size of the FDAC estimator of $\mu_\phi = E(\phi_i)$ in a heterogeneous panel AR(1) model with uniformly distributed ϕ_i, Gaussian errors without GARCH effects, and different initializations

		Bias			RMSE			Size (×100)		
T	n/M_i	100	3	1	100	3	1	100	3	1
$\mu_\phi = 0.4$ with $\lvert\phi_i\rvert < 1$										
4	100	-0.008	0.004	0.052	0.174	0.180	0.194	7.0	7.9	8.5
4	1,000	0.000	0.016	0.063	0.057	0.060	0.087	5.1	5.6	18.2
4	5,000	0.000	0.014	0.062	0.025	0.029	0.068	3.8	8.5	64.8
5	100	-0.003	0.008	0.041	0.134	0.136	0.147	6.3	7.2	8.8
5	1,000	0.000	0.014	0.048	0.043	0.046	0.066	5.1	6.5	19.0
5	5,000	0.000	0.012	0.048	0.019	0.023	0.052	4.4	9.2	65.2
6	100	-0.004	0.007	0.034	0.111	0.113	0.124	6.3	6.2	8.8
6	1,000	-0.001	0.011	0.038	0.037	0.038	0.054	5.8	5.1	16.4
6	5,000	0.000	0.010	0.038	0.016	0.019	0.042	4.4	9.0	60.9
10	100	0.000	0.005	0.015	0.078	0.082	0.084	6.5	7.2	7.0
10	1,000	0.000	0.007	0.020	0.026	0.026	0.033	5.8	6.6	12.8
10	5,000	0.000	0.006	0.021	0.011	0.013	0.024	5.3	7.5	41.9
$\mu_\phi = 0.5$ with $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$										
4	100	0.003	0.008	0.050	0.176	0.174	0.191	7.6	7.8	9.0
4	1,000	0.000	0.012	0.058	0.057	0.057	0.084	5.1	5.2	17.0
4	5,000	0.001	0.013	0.057	0.026	0.029	0.063	5.4	7.8	57.9
5	100	-0.003	0.007	0.039	0.134	0.136	0.145	7.1	6.7	8.5
5	1,000	-0.001	0.010	0.044	0.042	0.044	0.063	5.1	6.2	17.1
5	5,000	0.001	0.010	0.044	0.019	0.022	0.048	4.3	9.8	59.2
6	100	-0.004	0.003	0.030	0.112	0.114	0.121	7.1	7.3	9.2
6	1,000	-0.002	0.007	0.035	0.035	0.037	0.051	4.5	6.2	15.9
6	5,000	0.000	0.008	0.035	0.016	0.018	0.039	4.7	8.8	56.6
10	100	-0.003	0.001	0.016	0.079	0.078	0.083	6.3	5.9	8.1
10	1,000	-0.001	0.004	0.019	0.025	0.026	0.032	4.8	5.9	11.9
10	5,000	0.000	0.005	0.019	0.011	0.012	0.022	5.3	7.3	36.6

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -M_i + 1, -M_i + 2, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The heterogeneous AR(1) coefficients are generated by the uniform distribution: $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi \in \{0.4, 0.5\}$ The initial values are given by $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$, where $M_i \in \{100, 3, 1\}$ for all i. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator is calculated based on (6.2) in the main paper, and its asymptotic variance is estimated by the Delta method. The HetroGMM estimator and its asymptotic variance are calculated by (6.13) and (6.14) in the main paper. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$.

Table S.13: Bias, RMSE, and size of the FDAC estimator of $\mu_\phi = E(\phi_i)$ in a heterogeneous panel AR(1) model with categorically distributed ϕ_i, Gaussian errors without GARCH effects, and different initializations

		Bias			RMSE			Size (×100)		
T	n/M_i	100	3	1	100	3	1	100	3	1
$\mu_\phi = 0.545$ with $\lvert\phi_i\rvert < 1$										
4	100	0.000	0.010	0.079	0.164	0.168	0.193	7.5	8.0	11.1
4	1,000	-0.001	0.012	0.081	0.053	0.055	0.099	5.2	5.9	30.1
4	5,000	0.000	0.012	0.082	0.025	0.027	0.085	5.1	7.8	91.1
5	100	0.001	0.008	0.058	0.118	0.121	0.138	7.0	6.6	10.5
5	1,000	-0.001	0.009	0.061	0.037	0.040	0.073	4.4	5.8	33.0
5	5,000	-0.001	0.009	0.060	0.017	0.020	0.063	4.9	9.0	91.7
6	100	-0.001	0.008	0.046	0.099	0.101	0.112	6.9	6.8	10.2
6	1,000	0.000	0.007	0.048	0.031	0.033	0.058	5.1	6.3	32.4
6	5,000	0.000	0.008	0.047	0.014	0.016	0.049	4.8	8.1	89.2
10	100	-0.001	0.004	0.024	0.064	0.065	0.072	5.9	5.2	7.4
10	1,000	0.000	0.004	0.025	0.021	0.021	0.034	4.2	5.6	22.4
10	5,000	0.000	0.004	0.025	0.009	0.010	0.027	5.0	7.0	75.0
$\mu_\phi = 0.525$ with $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$										
4	100	-0.002	0.003	0.068	0.169	0.165	0.182	7.4	7.6	8.7
4	1,000	0.001	0.003	0.072	0.054	0.055	0.091	6.2	6.2	25.2
4	5,000	0.001	0.005	0.073	0.025	0.024	0.077	5.8	5.5	83.9
5	100	-0.003	0.001	0.051	0.118	0.117	0.132	5.9	6.6	9.6
5	1,000	0.001	0.002	0.050	0.039	0.039	0.064	6.4	5.4	25.3
5	5,000	0.001	0.003	0.051	0.017	0.017	0.054	5.8	5.7	82.6
6	100	-0.001	0.001	0.037	0.097	0.100	0.106	6.4	7.2	8.2
6	1,000	0.001	0.001	0.038	0.032	0.032	0.050	5.3	6.0	23.2
6	5,000	0.001	0.002	0.039	0.014	0.014	0.041	6.0	5.5	76.4
10	100	0.000	0.002	0.016	0.066	0.066	0.066	5.4	5.9	5.7
10	1,000	0.001	0.001	0.020	0.021	0.021	0.029	4.6	5.7	14.9
10	5,000	0.000	0.001	0.020	0.010	0.010	0.022	5.4	5.8	55.8

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -M_i + 1, -M_i + 2, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The heterogeneous AR(1) coefficients are generated by the categorical distribution: $\Pr(\phi_i = \phi_L) = \pi$ and $\Pr(\phi_i = \phi_H) = 1-\pi$, where $(\phi_H, \phi_L, \pi)' = (0.8, 0.5, 0.85)'$ with $|\phi_i| < 1$ for all i and $(1, 0.5, 0.95)'$ with $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ and all i. The initial values are given by $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$, where $M_i \in \{100, 3, 1\}$ for all i, except a case with $M_i = 100$ for units with $\phi_i = \phi_L = 0.5$ and $M_i = 1$ for units with $\phi_i = \phi_H = 1$. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator is calculated based on (6.2) in the main paper, and its asymptotic variance is estimated by the Delta method. The HetroGMM estimator and its asymptotic variance are calculated by (6.13) and (6.14) in the main paper. The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. The number of replications is $2,000$.

Table S.14: FDAC, FDLS, AH, AAH, AB, and BB estimators of ϕ ($\phi_0 = 0.5$) in a homogeneous panel AR(1) model with Gaussian errors without GARCH effects, and different initializations

			FDAC			FDLS			AH			AAH			AB			BB		
	T	n/M_i	100	3	1	100	3	1	100	3	1	100	3	1	100	3	1	100	3	1
Bias	4	100	0.004	0.007	0.086	0.002	0.009	0.072	0.564	0.412	0.255	0.084	0.090	0.093	-0.023	-0.026	-0.017	-0.006	-0.016	-0.016
	4	1,000	0.000	0.006	0.090	0.001	0.005	0.069	0.021	0.016	0.027	0.072	0.065	0.073	-0.004	-0.001	-0.003	-0.002	-0.014	-0.023
	4	5,000	0.000	0.006	0.090	0.000	0.005	0.069	0.002	0.004	0.006	0.023	0.023	0.017	0.000	0.001	0.000	0.000	-0.014	-0.023
	6	100	0.001	0.003	0.046	-0.001	0.005	0.038	-0.072	-0.070	-0.086	0.023	0.026	0.028	-0.041	-0.039	-0.032	-0.001	-0.008	-0.011
	6	1,000	0.000	0.003	0.049	0.000	0.002	0.037	-0.007	-0.008	-0.013	0.001	0.001	-0.001	-0.006	-0.003	-0.003	0.000	-0.008	-0.011
	6	5,000	0.000	0.003	0.049	0.000	0.003	0.037	-0.001	-0.001	-0.004	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.008	-0.011
	10	100	0.000	0.002	0.024	-0.001	0.003	0.019	-0.049	-0.044	-0.053	0.006	0.008	0.005	-0.035	-0.031	-0.030	-0.004	-0.005	-0.009
	10	1,000	0.000	0.002	0.025	0.000	0.001	0.018	-0.004	-0.005	-0.006	0.000	0.000	0.000	-0.004	-0.003	-0.003	0.000	-0.003	-0.003
	10	5,000	0.000	0.002	0.025	0.000	0.001	0.019	-0.001	-0.001	-0.002	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	-0.003	-0.003
RMSE	4	100	0.169	0.164	0.185	0.150	0.148	0.171	9.211	9.900	12.229	0.298	0.301	0.292	0.226	0.216	0.185	0.141	0.136	0.135
	4	1,000	0.054	0.053	0.105	0.049	0.048	0.085	0.206	0.201	0.262	0.209	0.198	0.216	0.070	0.066	0.055	0.044	0.047	0.047
	4	5,000	0.024	0.025	0.093	0.022	0.022	0.072	0.086	0.087	0.109	0.116	0.116	0.106	0.031	0.030	0.025	0.020	0.024	0.029
	6	100	0.098	0.095	0.107	0.104	0.103	0.115	0.220	0.210	0.242	0.152	0.156	0.155	0.130	0.124	0.112	0.088	0.085	0.084
	6	1,000	0.031	0.031	0.058	0.034	0.034	0.051	0.066	0.068	0.077	0.040	0.038	0.036	0.038	0.037	0.033	0.026	0.027	0.027
	6	5,000	0.014	0.014	0.051	0.015	0.015	0.040	0.030	0.029	0.034	0.014	0.014	0.013	0.017	0.017	0.015	0.012	0.014	0.016
	10	100	0.063	0.062	0.069	0.074	0.073	0.080	0.103	0.098	0.108	0.073	0.074	0.072	0.080	0.077	0.073	0.057	0.057	0.057
	10	1,000	0.020	0.020	0.032	0.024	0.024	0.030	0.028	0.028	0.030	0.016	0.017	0.016	0.021	0.021	0.019	0.016	0.016	0.016
	10	5,000	0.009	0.009	0.026	0.011	0.011	0.021	0.013	0.013	0.013	0.007	0.007	0.007	0.009	0.009	0.009	0.007	0.008	0.007
Size (×100)	4	100	7.7	7.2	11.1	5.9	5.9	11.9	8.3	7.3	7.3	12.7	12.8	13.1	10.0	8.5	9.1	13.2	12.3	11.5
	4	1,000	6.2	4.8	41.5	5.6	5.4	31.2	4.2	5.0	5.1	18.4	16.7	17.8	5.9	5.3	5.7	5.2	7.0	7.8
	4	5,000	5.4	6.2	96.8	4.9	5.8	87.7	5.2	5.6	6.0	9.2	8.9	7.0	5.0	5.2	5.8	5.0	9.2	17.8
	6	100	8.1	7.0	10.4	6.6	6.0	10.2	15.9	13.9	16.0	22.6	21.1	22.4	16.9	17.0	16.6	21.1	20.0	20.4
	6	1,000	4.5	5.9	37.3	5.9	5.2	20.3	6.4	6.0	7.6	6.7	7.2	5.9	6.4	7.1	6.8	6.3	7.8	8.4
	6	5,000	4.9	5.1	94.0	5.3	4.6	70.3	5.4	5.2	6.4	5.2	4.9	4.2	6.5	5.1	5.1	5.8	10.2	16.0
	10	100	5.1	5.6	8.6	6.0	5.1	8.2	34.0	32.7	35.6	44.3	44.8	43.7	39.6	37.3	38.0	43.6	43.4	47.1
	10	1,000	5.0	5.2	23.4	5.1	5.1	12.0	8.5	8.1	8.2	8.8	9.3	8.8	8.7	8.0	7.9	8.6	8.3	9.8
	10	5,000	5.6	5.1	77.1	4.7	5.1	42.2	5.2	5.8	5.0	5.9	5.4	6.1	6.2	6.1	5.3	5.3	7.8	8.8

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -M_i + 1, -M_i + 2, ..., T$, where $\phi_i = 0.5$ for all i, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The initial values are given by $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$, where $M_i \in \{100, 3, 1\}$ for all i. See also the notes to Table S.10.

Table S.15: FDAC, FDLS, AH, AAH, AB, and BB estimators of $\mu_\phi = E(\phi_i) = 0.4$ in a heterogeneous panel AR(1) model with uniformly distributed $|\phi_i| < 1$, Gaussian errors without GARCH effects, and different initializations

			FDAC			FDLS			AH			AAH			AB			BB		
	T	n/M_i	100	3	1	100	3	1	100	3	1	100	3	1	100	3	1	100	3	1
Bias	4	100	0.005	0.015	0.072	-0.057	-0.044	-0.007	-0.124	-0.052	0.013	-0.007	0.014	0.024	-0.081	-0.050	-0.002	-0.041	-0.046	-0.031
	4	1,000	-0.001	0.017	0.074	-0.061	-0.050	-0.014	-0.176	-0.183	-0.212	-0.066	-0.060	-0.049	-0.061	-0.021	0.018	-0.046	-0.053	-0.036
	4	5,000	0.000	0.017	0.073	-0.062	-0.051	-0.014	-0.185	-0.192	-0.216	-0.073	-0.066	-0.050	-0.055	-0.018	0.023	-0.044	-0.054	-0.035
	6	100	0.001	0.011	0.044	-0.060	-0.050	-0.029	-0.178	-0.186	-0.217	-0.025	-0.018	-0.004	-0.086	-0.059	-0.024	-0.026	-0.027	-0.018
	6	1,000	0.000	0.012	0.046	-0.061	-0.054	-0.033	-0.143	-0.152	-0.177	-0.033	-0.026	-0.013	-0.052	-0.022	0.011	-0.022	-0.023	-0.011
	6	5,000	0.000	0.012	0.046	-0.062	-0.054	-0.032	-0.140	-0.148	-0.171	-0.033	-0.027	-0.013	-0.046	-0.018	0.015	-0.020	-0.022	-0.009
	10	100	0.000	0.007	0.025	-0.061	-0.055	-0.044	-0.121	-0.124	-0.147	-0.031	-0.026	-0.019	-0.060	-0.041	-0.022	-0.018	-0.014	-0.010
	10	1,000	0.000	0.007	0.025	-0.061	-0.057	-0.047	-0.092	-0.101	-0.118	-0.022	-0.019	-0.012	-0.033	-0.016	0.006	-0.002	-0.002	0.008
	10	5,000	0.000	0.007	0.025	-0.061	-0.058	-0.046	-0.090	-0.099	-0.115	-0.021	-0.018	-0.011	-0.029	-0.012	0.010	0.000	0.001	0.010
RMSE	4	100	0.182	0.175	0.191	0.168	0.164	0.164	1.930	2.949	3.145	0.261	0.274	0.261	0.236	0.218	0.203	0.157	0.151	0.150
	4	1,000	0.058	0.059	0.094	0.079	0.071	0.055	0.223	0.228	0.258	0.109	0.102	0.080	0.092	0.069	0.064	0.066	0.071	0.059
	4	5,000	0.026	0.031	0.078	0.065	0.056	0.027	0.195	0.201	0.225	0.077	0.071	0.056	0.063	0.035	0.035	0.049	0.057	0.040
	6	100	0.114	0.111	0.123	0.127	0.123	0.120	0.258	0.260	0.295	0.123	0.121	0.116	0.163	0.145	0.131	0.107	0.103	0.103
	6	1,000	0.036	0.038	0.059	0.072	0.065	0.050	0.155	0.164	0.188	0.046	0.041	0.033	0.068	0.049	0.044	0.039	0.039	0.034
	6	5,000	0.016	0.020	0.049	0.064	0.056	0.036	0.142	0.150	0.173	0.036	0.030	0.019	0.050	0.027	0.024	0.024	0.026	0.017
	10	100	0.078	0.077	0.085	0.102	0.100	0.099	0.159	0.161	0.184	0.084	0.084	0.086	0.110	0.099	0.093	0.080	0.079	0.080
	10	1,000	0.025	0.026	0.036	0.067	0.063	0.054	0.098	0.107	0.123	0.033	0.031	0.027	0.044	0.033	0.030	0.024	0.024	0.025
	10	5,000	0.011	0.013	0.028	0.063	0.059	0.048	0.091	0.100	0.116	0.024	0.021	0.016	0.032	0.018	0.017	0.011	0.011	0.015
Size (×100)	4	100	8.2	7.6	8.9	8.2	8.8	8.1	15.5	15.2	16.8	18.2	18.8	14.9	14.9	11.5	10.2	16.2	14.7	13.5
	4	1,000	6.7	5.3	26.5	23.2	16.8	6.9	33.6	34.8	39.9	30.8	26.0	18.6	19.4	7.0	7.2	19.8	21.6	13.9
	4	5,000	5.4	10.0	82.8	78.1	60.1	9.2	85.8	86.8	89.2	81.4	74.9	54.0	46.8	11.8	13.9	56.4	70.9	37.8
	6	100	7.4	6.8	9.7	10.3	8.8	8.3	32.8	34.1	37.4	18.4	17.8	15.4	27.7	20.7	19.4	24.4	23.4	24.8
	6	1,000	4.1	6.8	26.8	40.8	31.4	14.4	71.5	75.7	81.2	18.9	13.9	6.8	30.0	12.4	12.1	15.5	16.1	10.6
	6	5,000	5.0	11.8	81.6	96.8	91.5	49.2	100.0	100.0	100.0	60.1	42.4	12.4	70.6	20.2	19.0	31.8	36.5	11.9
	10	100	5.2	6.0	7.7	13.1	11.9	11.7	57.5	58.0	64.4	41.5	42.3	42.0	48.1	43.7	39.8	48.4	49.1	51.3
	10	1,000	4.8	5.9	17.6	63.6	58.7	39.6	85.2	90.2	94.6	17.7	16.0	9.8	32.9	17.5	14.1	11.5	12.6	14.9
	10	5,000	5.2	8.6	59.7	100.0	99.8	96.0	100.0	100.0	100.0	44.8	35.0	15.6	70.9	23.2	20.2	9.6	8.3	23.6

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i)+\phi_i y_{i,t-1}+h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -M_i+1, -M_i+2, ..., T$, where $\phi_i = \mu_\phi + v_i$ with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi = 0.4$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The initial values are given by $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$, where $M_i \in \{100, 3, 1\}$ for all i. See also the notes to Table S.8.

Table S.16: FDAC, FDLS, AH, AAH, AB, and BB estimators of $\mu_\phi = E(\phi_i) = 0.5$ in a heterogeneous panel AR(1) model with uniformly distributed $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$, Gaussian errors without GARCH effects, and different initializations

			FDAC			FDLS			AH			AAH			AB			BB		
	T	n/M_i	100	3	1	100	3	1	100	3	1	100	3	1	100	3	1	100	3	1
Bias	4	100	0.003	0.014	0.066	-0.053	-0.040	-0.003	-0.070	-0.926	-0.064	-0.012	0.003	0.013	-0.109	-0.075	-0.032	-0.034	-0.043	-0.034
	4	1,000	-0.001	0.015	0.069	-0.057	-0.045	-0.009	-0.190	-0.198	-0.229	-0.062	-0.059	-0.050	-0.072	-0.033	-0.005	-0.043	-0.053	-0.044
	4	5,000	0.000	0.016	0.068	-0.057	-0.046	-0.009	-0.204	-0.210	-0.236	-0.075	-0.070	-0.059	-0.065	-0.030	0.001	-0.043	-0.054	-0.044
	6	100	0.001	0.010	0.040	-0.056	-0.046	-0.025	-0.224	-0.232	-0.267	-0.021	-0.017	-0.007	-0.119	-0.089	-0.057	-0.022	-0.026	-0.022
	6	1,000	0.000	0.011	0.043	-0.057	-0.050	-0.028	-0.171	-0.180	-0.206	-0.034	-0.029	-0.019	-0.074	-0.039	-0.010	-0.021	-0.023	-0.015
	6	5,000	0.000	0.011	0.043	-0.057	-0.050	-0.027	-0.166	-0.174	-0.197	-0.034	-0.029	-0.019	-0.068	-0.033	-0.004	-0.020	-0.021	-0.012
	10	100	0.000	0.006	0.022	-0.057	-0.051	-0.040	-0.164	-0.166	-0.189	-0.024	-0.018	-0.012	-0.086	-0.068	-0.050	-0.016	-0.015	-0.012
	10	1,000	0.000	0.006	0.023	-0.057	-0.053	-0.042	-0.123	-0.132	-0.148	-0.018	-0.015	-0.008	-0.059	-0.036	-0.015	-0.003	-0.001	0.008
	10	5,000	0.000	0.006	0.023	-0.057	-0.053	-0.041	-0.120	-0.128	-0.144	-0.017	-0.014	-0.007	-0.058	-0.032	-0.010	0.000	0.003	0.012
RMSE	4	100	0.181	0.171	0.186	0.171	0.165	0.167	2.255	42.237	8.912	0.263	0.267	0.260	0.305	0.265	0.236	0.166	0.159	0.159
	4	1,000	0.057	0.057	0.089	0.076	0.069	0.055	0.256	0.260	0.295	0.125	0.115	0.107	0.113	0.084	0.070	0.068	0.074	0.066
	4	5,000	0.025	0.030	0.072	0.062	0.052	0.025	0.216	0.223	0.249	0.080	0.076	0.064	0.076	0.046	0.031	0.050	0.058	0.049
	6	100	0.111	0.108	0.119	0.127	0.123	0.122	0.314	0.314	0.355	0.133	0.128	0.124	0.202	0.176	0.155	0.114	0.107	0.107
	6	1,000	0.035	0.037	0.056	0.069	0.062	0.048	0.184	0.194	0.219	0.047	0.043	0.036	0.092	0.064	0.049	0.040	0.040	0.035
	6	5,000	0.016	0.019	0.045	0.060	0.052	0.032	0.169	0.176	0.200	0.037	0.032	0.023	0.073	0.040	0.021	0.025	0.025	0.019
	10	100	0.077	0.076	0.083	0.101	0.099	0.098	0.201	0.202	0.225	0.087	0.087	0.089	0.136	0.121	0.111	0.085	0.081	0.081
	10	1,000	0.025	0.025	0.034	0.063	0.059	0.050	0.129	0.138	0.153	0.031	0.029	0.026	0.069	0.049	0.037	0.025	0.024	0.025
	10	5,000	0.011	0.013	0.026	0.058	0.055	0.043	0.121	0.129	0.145	0.021	0.018	0.013	0.061	0.035	0.018	0.012	0.012	0.016
Size (×100)	4	100	8.3	6.9	9.0	8.2	8.6	8.0	15.2	14.9	15.8	16.1	16.0	14.1	16.3	11.5	10.7	17.2	15.0	13.9
	4	1,000	6.2	6.0	24.4	19.8	14.5	6.3	30.0	30.9	34.2	29.5	25.4	21.3	18.5	8.8	7.1	16.4	19.4	16.1
	4	5,000	5.1	9.8	78.8	70.1	51.0	6.7	78.6	79.8	81.7	74.1	67.7	56.1	40.6	16.9	5.6	45.2	61.9	47.6
	6	100	7.2	6.7	9.7	9.6	8.2	8.2	35.8	36.2	40.6	21.0	18.1	16.4	30.8	24.5	21.4	27.8	24.6	25.6
	6	1,000	4.4	5.8	24.8	35.0	27.2	11.7	73.7	76.3	80.6	20.5	16.0	10.5	38.0	19.4	10.9	15.1	15.5	11.1
	6	5,000	5.1	10.2	77.5	93.3	84.6	37.3	100.0	100.0	100.0	61.2	47.3	23.3	84.3	39.1	8.6	29.1	30.2	14.9
	10	100	5.2	6.3	7.8	12.0	11.1	11.3	66.4	67.8	71.5	41.0	42.4	41.8	54.5	49.4	46.4	51.5	50.4	52.0
	10	1,000	4.6	5.2	16.2	55.3	50.1	32.6	94.6	96.2	97.4	13.0	11.2	7.1	58.1	34.0	19.2	11.9	12.0	15.7
	10	5,000	4.9	8.0	52.9	99.8	99.6	91.1	100.0	100.0	100.0	29.5	21.7	8.6	97.3	67.1	20.2	10.2	9.8	26.2

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i)+\phi_i y_{i,t-1}+h_{it}\varepsilon_{it}$, for $i = 1, 2, ..., n$, and $t = -M_i+1, -M_i+2, ..., T$, where $\phi_i = \mu_\phi + v_i$ with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi = 0.5$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The initial values are given by $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$, where $M_i \in \{100, 3, 1\}$ for all i. See also the notes to Table S.9.

Table S.17: Bias, RMSE, and size of FDAC and MSW estimators of $\mu_\phi = E(\phi_i)$ in heterogeneous and homogeneous panel AR(1) models with Gaussian errors without GARCH effects and different initializations

		Bias				RMSE				Size (×100)			
		FDAC		MSW		FDAC		MSW		FDAC		MSW	
T	n/M_i	100	1	100	1	100	1	100	1	100	1	100	1
$\mu_\phi = 0.4$ with uniformly distributed $\lvert\phi_i\rvert < 1$ for all i													
4	100	-0.005	0.054	-0.145	-0.092	0.177	0.190	0.157	0.113	8.0	7.8	82.0	42.4
4	1,000	0.000	0.062	-0.128	-0.082	0.056	0.086	0.130	0.086	4.7	18.1	100.0	98.3
6	100	-0.004	0.038	-0.144	-0.073	0.113	0.121	0.155	0.097	5.7	8.7	79.3	27.6
6	1,000	-0.001	0.038	-0.129	-0.064	0.037	0.053	0.130	0.068	6.3	16.7	100.0	91.5
10	100	-0.001	0.021	-0.146	-0.059	0.079	0.084	0.158	0.090	6.4	7.3	71.2	17.5
10	1,000	0.000	0.020	-0.141	-0.055	0.026	0.033	0.143	0.060	5.7	12.4	100.0	76.1
$\mu_\phi = 0.5$ with uniformly distributed $\phi_i \in [-1+\epsilon, 1]$ for some $\epsilon > 0$ and all i													
4	100	-0.005	0.049	-0.207	-0.130	0.175	0.187	0.221	0.148	8.7	7.7	84.3	54.4
4	1,000	0.000	0.057	-0.194	-0.119	0.056	0.081	0.196	0.122	5.0	15.3	100.0	99.9
6	100	-0.004	0.031	-0.202	-0.102	0.111	0.118	0.215	0.124	5.5	7.4	81.2	36.2
6	1,000	-0.001	0.034	-0.187	-0.091	0.036	0.050	0.189	0.094	5.2	16.2	100.0	97.9
10	100	-0.001	0.015	-0.198	-0.077	0.079	0.083	0.213	0.107	6.7	6.4	74.3	21.2
10	1,000	-0.001	0.017	-0.192	-0.072	0.025	0.031	0.194	0.076	5.9	10.4	100.0	87.1
$\phi_i = \mu_\phi = 0.5$ for all i													
4	100	0.002	0.067	-0.201	-0.155	0.166	0.184	0.208	0.164	7.4	10.4	98.1	86.7
4	1,000	-0.001	0.080	-0.182	-0.135	0.054	0.097	0.183	0.137	5.7	29.7	100.0	100.0
6	100	0.002	0.038	-0.199	-0.132	0.097	0.105	0.205	0.143	7.4	9.0	98.5	75.0
6	1,000	-0.001	0.041	-0.185	-0.121	0.031	0.052	0.186	0.123	4.4	25.8	100.0	100.0
10	100	0.002	0.017	-0.198	-0.111	0.065	0.067	0.205	0.126	7.7	7.4	95.6	50.5
10	1,000	-0.001	0.020	-0.194	-0.106	0.020	0.029	0.195	0.108	4.8	16.4	100.0	99.9

Notes: The DGP is given by $y_{it} = \mu_i(1-\phi_i) + \phi_i y_{i,t-1} + h_{it}\varepsilon_{it}$, for $i = 1,2,...,n$, and $t = -M_i + 1, -M_i + 2, ..., T$, featuring Gaussian standardized errors with cross-sectional heteroskedasticity without GARCH effects. The heterogeneous AR(1) coefficients are generated by uniform distributions: $\phi_i = \mu_\phi + v_i$, with $v_i \sim IIDU[-a, a]$, $a = 0.5$ and $\mu_\phi \in \{0.4, 0.5\}$. In the homogeneous case, $\phi_i = \mu_\phi = 0.5$ for all i. The initial values are given by $(y_{i,-M_i} - \mu_i) \sim IIDN(b, \kappa\sigma_i^2)$ with $b = 1$ and $\kappa = 2$, where $M_i \in \{100, 1\}$ for all i. For each experiment, $(\alpha_i, \phi_i, \sigma_i)'$ are generated differently across replications. The FDAC estimator is calculated by (6.2) in the main paper, and its asymptotic variance is estimated by the Delta method. "MSW" denotes the estimator proposed by Mavroeidis et al. (2015). The estimation is based on $\{y_{i1}, y_{i2}, ..., y_{iT}\}$ for $i = 1, 2, ..., n$. The nominal size of the tests is set to 5 per cent. Due to the extensive computations required for the implementation of the MSW estimator, the number of replications is $1,000$.

S.9 Empirical results for other sub-periods of the PSID

Table S.18 shows the distribution of cross-sectional observation numbers by year based on the sample selection criterion in Meghir and Pistaferri (2004). For different sub-periods, Tables S.19 and S.20 report the estimates of mean persistence of log real earnings in a panel AR(1) model with a common linear trend, and Tables S.21–S.23 report the estimates of σ_ϕ^2 of the heterogeneous persistence parameters, ϕ_i.

Table S.18: Distribution of individual observation numbers by year

Year	Number of observations
1976	1,600
1977	1,663
1978	1,706
1979	1,773
1980	1,800
1981	1,868
1982	1,884
1983	1,933
1984	1,972
1985	2,012
1986	2,053
1987	2,083
1988	2,091
1989	2,008
1990	1,907
1991	1,831
1992	1,711
1993	1,576
1994	1,471
1995	1,384
Total	36,325

Notes: The sample selection criteria of Meghir and Pistaferri (2004) are summarized as the following. (i) Individuals are from the "core" sample, i.e., the 1968 SRC cross-section sample and the 1968 Census sample. (ii) Individuals are continuously heads of their families. (iii) Over the respective observed period, the range of individuals' ages is 25 to 55. (iv) Individuals are males. (v) Individuals have nine years or more observations of usable (non-zero and not top-coded) money income of labor $earnings_{it}$. (vi) Individuals have no missing records of education or race over their sample periods. (vii) Observations with only self-employed status are dropped. (viii) Observations of outcome variables $y_{it} = log(earnings_{it}/p_t)$ with outlying deviations $\Delta y_{it} > 5$ or $\Delta y_{it} < -1$ are dropped.

Table S.19: Estimates of mean persistence ($\mu_\phi = E(\phi_i)$) of log real earnings in a panel AR(1) model with a common linear trend using the PSID data over the sub-periods 1976–1980, 1981–1985 and 1986–1990

	1976–1980, $T=5$				1981–1985, $T=5$				1986–1990, $T=5$			
	All	Category by education			All	Category by education			All	Category by education		
	categories	HSD	HSG	CLG	categories	HSD	HSG	CLG	categories	HSD	HSG	CLG
Homogeneous slopes												
AAH	0.527	0.545	0.489	0.560	0.481	0.426	0.465	0.598	0.499	0.725	0.426	0.491
	(0.051)	(0.079)	(0.084)	(0.070)	(0.038)	(0.083)	(0.046)	(0.072)	(0.035)	(0.093)	(0.041)	(0.065)
AB	0.326	0.346	0.076	0.623	0.219	0.286	0.178	-0.066	0.281	0.239	0.305	0.131
	(0.109)	(0.151)	(0.148)	(0.207)	(0.071)	(0.092)	(0.092)	(0.214)	(0.089)	(0.303)	(0.100)	(0.171)
BB	0.905	0.916	0.898	0.916	0.957	0.939	0.962	1.041	0.939	0.897	0.929	0.978
	(0.012)	(0.015)	(0.015)	(0.028)	(0.005)	(0.009)	(0.006)	(0.014)	(0.011)	(0.027)	(0.012)	(0.014)
Heterogeneous slopes												
FDAC	0.589	0.567	0.595	0.607	0.602	0.428	0.596	0.844	0.675	0.760	0.604	0.805
	(0.037)	(0.062)	(0.056)	(0.079)	(0.039)	(0.076)	(0.053)	(0.056)	(0.032)	(0.083)	(0.042)	(0.056)
MSW	0.419	0.388	0.434	0.452	0.420	0.378	0.439	0.452	0.429	0.427	0.427	0.450
	(0.060)	(0.058)	(0.045)	(0.030)	(0.058)	(0.055)	(0.031)	(0.031)	(0.056)	(0.048)	(0.056)	(0.046)
Common linear trend	0.023	0.029	0.021	0.021	0.025	0.036	0.019	0.032	0.018	0.009	0.021	0.014
n	1,312	363	641	308	1,489	283	855	351	1,654	201	994	459

Notes: The estimates are based on the heterogeneous panel AR(1) model with a common linear trend, $y_{it} = \alpha_i + g(1-\phi_i)t + \phi_i y_{i,t-1} + u_{it}$, where $y_{it} = log(earnings_{it}/p_t)$ using the PSID data over the sub-periods 1976–1980, 1981–1985, and 1986–1990. "HSD" refers to high school dropouts with less than 12 years of education, "HSG" refers to high school graduates with at least 12 but less than 16 years of education, and "CLG" refers to college graduates with at least 16 years of education. The common trend, g, is estimated by $\hat{g}_{FD} = n^{-1}(T-1)^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta y_{it}$. Then the estimation for μ_ϕ is based on $\tilde{y}_{it} = y_{it} - \hat{g}_{FD}t$ for $t = 1, 2, ..., T$. "AAH", "AB", and "BB" denote different 2-step GMM estimators proposed by Chudik and Pesaran (2021), Arellano and Bond (1991), and Blundell and Bond (1998). The FDAC estimator is calculated by (6.2), and its asymptotic variance is estimated by the Delta method. "MSW" denotes the kernel-weighted estimator in Mavroeidis et al. (2015) and is calculated based on a parametric assumption that $(\alpha_i, \phi_i)|y_{i1}$ follows a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{V})$ with initial values given by $\boldsymbol{\mu} = (5, 0.5)$, $\sigma_\alpha = 2$, $\sigma_\phi = 0.4$, $corr(\alpha_i, \phi_i) = 0.5$ with $\sigma_u = 0.5$.

Table S.20: Estimates of mean persistence ($\mu_\phi = E(\phi_i)$) of log real earnings in a panel AR(1) model with a common linear trend using the PSID data over the sub-periods 1976–1985 and 1981–1990

	1976–1985, $T = 10$				1981–1990, $T = 10$			
	All	Category by education			All	Category by education		
	categories	HSD	HSG	CLG	categories	HSD	HSG	CLG
Homogeneous slopes								
AAH	0.615	0.532	0.587	0.632	0.579	0.545	0.529	0.654
	(0.044)	(0.040)	(0.045)	(0.027)	(0.030)	(0.038)	(0.027)	(0.043)
AB	0.471	0.402	0.391	0.348	0.265	0.261	0.273	0.388
	(0.048)	(0.054)	(0.061)	(0.051)	(0.041)	(0.053)	(0.038)	(0.059)
BB	0.960	0.922	0.962	1.001	0.958	0.956	0.961	0.978
	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Heterogeneous slopes								
FDAC	0.643	0.554	0.637	0.766	0.628	0.614	0.600	0.734
	(0.028)	(0.052)	(0.041)	(0.054)	(0.025)	(0.057)	(0.033)	(0.042)
MSW	0.443	0.397	0.443	0.474	0.458	0.453	0.446	0.541
	(0.060)	(0.047)	(0.067)	(0.062)	(0.030)	(0.041)	(0.025)	(0.064)
Common linear trend	0.024	0.026	0.021	0.029	0.023	0.031	0.019	0.025
n	885	201	458	226	1,046	170	620	256

Notes: The estimates are based on the heterogeneous panel AR(1) model with a common linear trend, $y_{it} = \alpha_i + g(1-\phi_i)t + \phi_i y_{i,t-1} + u_{it}$, where $y_{it} = log(earnings_{it}/p_t)$ using the PSID data over the sub-periods 1976–1985 and 1981–1990. "HSD" refers to high school dropouts with less than 12 years of education, "HSG" refers to high school graduates with at least 12 but less than 16 years of education, and "CLG" refers to college graduates with at least 16 years of education. The common trend, g, is estimated by $\hat{g}_{FD} = n^{-1}(T-1)^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta y_{it}$. Then the estimation for μ_ϕ is based on $\tilde{y}_{it} = y_{it} - \hat{g}_{FD}t$ for $t = 1, 2, ..., T$. "AAH", "AB", and "BB" denote different 2-step GMM estimators proposed by Chudik and Pesaran (2021), Arellano and Bond (1991), and Blundell and Bond (1998). The FDAC estimator is calculated by (6.2), and its asymptotic variance is estimated by the Delta method. "MSW" denotes the kernel-weighted estimator in Mavroeidis et al. (2015) and is calculated based on a parametric assumption that $(\alpha_i, \phi_i)|y_{i1}$ follows a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{V})$ with initial values given by $\boldsymbol{\mu} = (5, 0.5)$, $\sigma_\alpha = 2$, $\sigma_\phi = 0.4$, $corr(\alpha_i, \phi_i) = 0.5$ with $\sigma_u = 0.5$. .

Table S.21: Estimates of variance of heterogeneous persistence (σ_ϕ^2) of log real earnings in a panel AR(1) model with a common linear trend using the PSID data over the sub-periods 1991–1995 and 1986–1995

	1991–1995, $T = 5$				1986–1995, $T = 10$			
	All	Category by education			All	Category by education		
	categories	HSD	HSG	CLG	categories	HSD	HSG	CLG
FDAC	0.100	0.204	0.081	0.091	0.129	0.122	0.120	0.141
	(0.042)	(0.100)	(0.054)	(0.090)	(0.023)	(0.060)	(0.031)	(0.036)
MSW	0.012	0.011	0.011	0.010	0.015	0.010	0.011	0.014
	(0.003)	(0.009)	(0.004)	(0.007)	(0.005)	(0.011)	(0.005)	(0.011)
n	1,366	127	832	407	1,139	109	689	341

Notes: The estimates are based on the heterogeneous panel AR(1) model with a common linear trend, $y_{it} = \alpha_i + g(1-\phi_i)t + \phi_i y_{i,t-1} + u_{it}$, where $y_{it} = log(earnings_{it}/p_t)$ using the PSID data over the sub-periods 1991–1995 and 1986–1995. "HSD" refers to high school dropouts with less than 12 years of education, "HSG" refers to high school graduates with at least 12 but less than 16 years of education, and "CLG" refers to college graduates with at least 16 years of education. The common trend, g, is estimated by $\hat{g}_{FD} = n^{-1}(T-1)^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta y_{it}$. Then the estimation for σ_ϕ^2 is based on $\tilde{y}_{it} = y_{it} - \hat{g}_{FD}t$ for $t = 1, 2, ..., T$. The The FDAC estimator of σ_ϕ^2 is calculated by (6.22), and its asymptotic variance is estimated by the Delta method. "MSW" denotes the kernel-weighted maximum likelihood estimator in Mavroeidis et al. (2015).

Table S.22: Estimates of variance of heterogeneous persistence (σ_ϕ^2) of log real earnings in a panel AR(1) model with a common linear trend using the PSID data over the sub-periods 1976–1985 and 1981–1990

	1976–1985, $T = 10$				1981–1990, $T = 10$			
	All	Category by education			All	Category by education		
	categories	HSD	HSG	CLG	categories	HSD	HSG	CLG
FDAC	0.095	0.139	0.100	0.001	0.150	0.104	0.171	0.113
	(0.028)	(0.049)	(0.043)	(0.046)	(0.022)	(0.058)	(0.026)	(0.046)
MSW	0.016	0.013	0.013	0.013	0.011	0.011	0.010	0.014
	(0.007)	(0.010)	(0.010)	(0.013)	(0.003)	(0.008)	(0.003)	(0.012)
n	885	201	458	226	1,046	170	620	256

Notes: The estimates are based on the heterogeneous panel AR(1) model with a common linear trend, $y_{it} = \alpha_i + g(1-\phi_i)t + \phi_i y_{i,t-1} + u_{it}$, where $y_{it} = log(earnings_{it}/p_t)$ using the PSID data over the sub-periods 1976–1985 and 1981–1990. "HSD" refers to high school dropouts with less than 12 years of education, "HSG" refers to high school graduates with at least 12 but less than 16 years of education, and "CLG" refers to college graduates with at least 16 years of education. The common trend, g, is estimated by $\hat{g}_{FD} = n^{-1}(T-1)^{-1}\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta y_{it}$. Then the estimation for σ_ϕ^2 is based on $\tilde{y}_{it} = y_{it} - \hat{g}_{FD}t$ for $t = 1, 2, ..., T$. The FDAC estimator is calculated by (6.22), and its asymptotic variance is estimated by the Delta method. "MSW" denotes the estimator proposed by Mavroeidis et al. (2015). See also the notes to Table S.20.

Table S.23: Estimates of variance of heterogeneous persistence (σ_{ϕ}^2) of log real earnings in a panel AR(1) model with a common linear trend using the PSID data over the sub-periods 1976–1980, 1981–1985, and 1986–1990

	1976–1980, $T = 5$				1981–1985, $T = 5$				1986–1990, $T = 5$			
	All	Category by education			All	Category by education			All	Category by education		
	categories	HSD	HSG	CLG	categories	HSD	HSG	CLG	categories	HSD	HSG	CLG
FDAC	0.038	0.072	0.025	0.013	0.089	0.040	0.093	0.032	0.126	0.095	0.111	0.151
	(0.056)	(0.078)	(0.099)	(0.098)	(0.037)	(0.068)	(0.052)	(0.072)	(0.040)	(0.105)	(0.056)	(0.054)
MSW	0.015	0.014	0.013	0.009	0.015	0.014	0.010	0.009	0.015	0.011	0.013	0.011
	(0.004)	(0.008)	(0.005)	(0.006)	(0.004)	(0.008)	(0.002)	(0.005)	(0.004)	(0.008)	(0.004)	(0.007)
n	1,312	363	641	308	1,489	283	855	351	1,654	201	994	459

Notes: The estimates are based on the heterogeneous panel AR(1) model with a common linear trend, $y_{it} = \alpha_i + g(1 - \phi_i)t + \phi_i y_{i,t-1} + u_{it}$, where $y_{it} = log(earnings_{it}/p_t)$ using the PSID data over the sub-periods 1976–1980, 1981–1985, and 1986–1990. "HSD" refers to high school dropouts with less than 12 years of education, "HSG" refers to high school graduates with at least 12 but less than 16 years of education, and "CLG" refers to college graduates with at least 16 years of education. The common trend, g, is estimated by $\hat{g}_{FD} = n^{-1}(T-1)^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} \Delta y_{it}$. Then the estimation for σ_{ϕ}^2 is based on $\tilde{y}_{it} = y_{it} - \hat{g}_{FD} t$ for $t = 1, 2, ..., T$. The FDAC estimator is calculated by (6.22), and its asymptotic variance is estimated by the Delta method. "MSW" denotes the estimator proposed by Mavroeidis et al. (2015). See also the notes to Table S.19.

References

Anderson, T. W., and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76, 598-606.

Anderson, T. W., and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18, 47-82.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* 58, 277–297.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87, 115–143.

Chudik, A. and M. H. Pesaran (2021). An augmented Anderson-Hsiao estimator for dynamic short-T panels. *Econometric Reviews* 1–32.

Han, C. and P. C. Phillips (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory* 26, 119–151.

Mavroeidis, S., Y. Sasaki, and I. Welch (2015). Estimation of heterogeneous autoregressive parameters with short panel data. *Journal of Econometrics* 188, 219–235.

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica* 72, 1–32.