Specification testing with grouped fixed effects

Pigini, Claudia and Pionati, Alessandro and Valentini, Francesco

4 July 2023

Specification testing with grouped fixed effects

Claudia Pigini* Alessandro Pionati† Francesco Valentini ‡

Preliminary draft

Abstract

We propose a bootstrap generalized Hausman test for the correct specification of unobserved heterogeneity in fixed-effects panel data models. We consider as null hypotheses two scenarios in which the unobserved heterogeneity is either time-invariant or specified as additive individual and time effects. We contrast the standard fixed-effects estimators with the recently developed two-step grouped fixed-effects estimator, that is consistent in the presence of time-varying heterogeneity under minimal specification and distributional assumptions for the unobserved effects. The Hausman test exploits the general formulation for the variance of the vector of contrasts and critical values are computed via parametric percentile bootstrap, so as to account for the non-centrality of the asymptotic χ^2 distribution arising from the incidental parameters and approximation biases. Monte Carlo evidence shows that the test has correct size and good power in both linear and non linear specification.

Keywords: Additive effects, Asymptotic bias, Hausman test, Parametric bootstrap, Time-varying heterogeneity

JEL Classification: C12, C23, C25

*Marche Polytechnic University (Italy). E-mail: c.pigini@univpm.it

†Marche Polytechnic University (Italy).Corresponding Author.Address: Department of Economics and Social Sciences, P.le Martelli 8, 60121 Ancona (Italy) E-mail: a.pionati@pm.univpm.it

‡Marche Polytechnic University (Italy). E-mail: f.valentini@staff.univpm.it

1 Introduction

Correct specification of unobserved heterogeneity is crucial in panel data modeling. For long, empirical applications have only considered individual time-constant fixed effects, but the assumption of time-invariant unobserved heterogeneity is often hardly tenable in practice, especially over a long time dimension. Therefore the current mainstream approach has become to include both subject and time fixed effects, in order to achieve credible identification of the effects of interest. The simplest and most widely employed setup is the specification of additive individual and time heterogeneity, namely the two-way fixed-effects model, that in the linear model is equivalent to the two-way correlated random effects approach (Wooldridge, 2021). Fernández-Val and Weidner (2016) provide analytical and jackknife bias corrections for the maximum likelihood (ML) estimator of nonlinear models with additive fixed effects, which is plagued by the incidental parameters problem.

While of simple implementation, the two-way fixed-effects specification fails to capture the specific impact common factors may have on each subject. There is now an important stream of literature on developing identification results and estimation strategies for models with interactive time and individual fixed effects. Contributions have been spurred by the seminal paper of Bai (2009), who provides identification results along with the asymptotics for the interactive fixed effect estimator in linear models. More recently, interactive fixed-effects have been introduced in nonlinear panel data and network models by Chen et al. (2021).

Testing the assumptions on the specification of time-varying unobserved heterogeneity has also received considerable attention in the recent econometric literature. Bartolucci et al. (2015) propose a Hausman-type test for the hypothesis of time-constant unobserved heterogeneity in generalized linear models where conditional ML estimators are compared with first-differences or pairwise conditional ML estimators. In the context of large stationary panel models, the factor specification could be tested by comparing additive with interactive fixed-effects models, on the basis of the Hausman test illustrated by Bai (2009) and its fixed-T version, derived by Westerlund (2019). However, as pointed out by Castagnetti et al. (2015b), the Hausman-type test fails to reject the null hypothesis when individual factor loadings are independent across equations. To overcome this issue, Castagnetti et al. (2015a) propose an alternative max-type test for the null hypothesis of time-invariant unobserved heterogeneity. While general, the interactive effects/factor specification imposes a parametric structure for the unobserved heterogeneity that may or may not be true under the alternative hypothesis.

In this paper we propose a generalized Hausman test for the fixed-effects specification, in both linear and nonlinear models and where the unobserved heterogeneity is either only individual or additive. The test contrasts fixed-effects ML estimators with the Two-Way Grouped Fixed Effects (TW-GFE henceforth) approach, recently put forward by Bonhomme et al. (2022a). Their proposal is based on a first-step data-driven approximation of the unobserved heterogeneity, which is clustered by the *kmeans* algorithm that uses individual and time-series moments to assign individual and time group memberships. Cluster dummies then enter the model specification as interactive effects and the associated parameters are estimated along with the regression coefficients in the second step. The resulting second-step estimator is consistent in the presence of unspecified forms of the time-varying unobserved heterogeneity with minimal assumptions on the unobserved components, which makes it a perfect candidate to contrast with the fixed-effects estimators that are consistent only with time-constant or time-varying additive heterogeneity.

Under specific choices for the number of clusters outlined by Bonhomme et al. (2022a) for the first step, it can be shown that the TW-GFE estimator is asymptotically normal, so that the Hausman statistic (Hausman, 1978) has asymptotic χ^2 distribution. However, as it might be difficult to verify which estimator is more efficient than the other under the null hypothesis, we rely on the generalized estimator for the variance of the vector of contrasts proposed by Bartolucci et al. (2015). In addition, the asymptotic χ^2 distribution is non-central because of two sources of asymptotic bias: the incidental parameters problem, that in nonlinear models plagues both estimators, and the approximation bias, that affects the TW-GFE. We therefore compute critical values of the test statistic distribution by means of parametric percentile bootstrap (MacKinnon, 2006). The main advantage of this procedure lies in the bootstrap distributions correctly capturing the non-centrality without the need for any bias correction of either estimator. The proposed strategy exploits the results on bootstrap inference for fixed-effects models provided by Kim and Sun (2016) and Higgins and Jochmans (2022).[1]

Literature review This paper relates to the stream of literature that has studied fixed-effects panel data models with grouped structures for the unobserved heterogeneity. Discrete heterogeneity has long been considered within the random-effects approach (Heckman and Singer, 1984), especially by a large body of statistical literature; see, for instance, MacLahlan and Peel (2000) on finite-mixture models and Bartolucci et al. (2012) on latent Markov models. On the contrary, the investigation of grouped patterns of heterogeneity in fixed-effects models is relatively recent in the econometric literature.

Hahn and Moon (2010) study the asymptotic bias arising form the incidental parameters problem in nonlinear panel data models where unobserved heterogeneity is assumed to be discrete with a finite number of support points. Bester and Hansen (2016) investigate the asymptotic behavior of the ML estimator for nonlinear models with grouped effects, under the assumption that subjects are clustered according to some external known classification. Models with unknown grouped membership are studied by Su et al. (2016), who propose penalized techniques for the estimation of models where regularization by classifier-Lasso shrinks individual effects to group coefficients, and by Ando and Bai (2016) who consider unobserved group factor structures in linear models with interactive fixed effects.

Discrete unobserved heterogeneity can serve as a regularization device that allows to identify the parameters of interest in panel data models with time-varying individual effects but not necessarily characterized by a factor structure. In this vein, Bonhomme and Manresa (2015) introduce a GFE estimator for linear models where the discrete heterogeneity is assumed to follow time-varying grouped patterns and cluster membership is left unrestricted. By contrast, the TW-GFE estimator by Bonhomme et al. (2022a) is consistent even with unspecified forms of time-varying unobserved heterogeneity. While using discretization as an approximation device introduces an asymptotic bias, the function of the unobserved heterogeneity they consider encompasses a variety of specifications, such as additive and interactive effects, under minimal distributional assumptions. This makes the TW-GFE estimator a simple and potentially very attractive tool for practitioners.

Outline The rest of the paper is organized as follows: Section 2 briefly outlines the fixed-effects models with individual and additive unobserved heterogeneity and recalls Bonhomme et al.'s

[1]As the test statistic is non-pivotal, pairs bootstrap cannot be used for asymptotic refinement in this case, unless a pre-pivoting strategy is applied (Cavaliere et al., 2022).

TW-GFE estimator. Section 3 illustrates the proposed approach. Section 4 presents the results of the simulation study. Finally, Section 5 concludes.

2 Background

Consider a panel data setup where subjects are indexed by $i = 1, \dots, N$ and time occasions are indexed by $t = 1, \dots, T$. Throughout the paper, we assume that observations are independent, conditional on the observed covariates and unobserved heterogeneity, and that the models are *static*. The traditional specification of fixed-effects models depicts unobserved heterogeneity as individual-specific intercepts, so that the conditional distribution of the response variable y_{it} given an r-vector exogenous covariates x_{it} is of the type

$$y_{it}|x_{it}, \theta_0, \alpha_{i0} \sim f(\cdot|x_{it}'\theta_0 + \alpha_{i0}), \tag{1}$$

where θ_0 is the vector of parameters of interest, α_{i0} denotes the permanent individual effect, and $f(\cdot)$ is a generic known density function, as in Chen et al. (2021). When (1) is a linear regression model, consistent OLS estimators of θ can be trivially obtained on the basis of standard demeaning or first-differences transformations, whereas ML estimators in non-linear models are consistent but exhibit a bias in their limiting distribution under rectangular array asymptotics (Li et al., 2003), unless probability formulations admit sufficient statistics for the individual intercepts (Andersen, 1970; Chamberlain, 1980). Therefore bias reduction techniques, such as analytical or jackknife corrections, are required (Hahn and Newey, 2004). These estimators are usually referred to as the one-way fixed-effects (OW-FE) estimators.

In order to account for time-varying heterogeneity, the widespread approach is to include common time effects, that enter the specification in an additive manner. The model is then of the type

$$y_{it}|x_{it}, \theta_0, \alpha_{i0}, \zeta_{t0} \sim f(\cdot|x_{it}'\theta_0 + \alpha_{i0} + \zeta_{t0}), \tag{2}$$

where ζ_{t0} represents such time-varying heterogeneity. Similarly to the case with only individual effects, a consistent estimator of θ can be obtained under suitable transformations when a linear regression model is specified, while bias corrections have to be implemented for ML estimators Fernández-Val and Weidner (2016). We denote them as two-way fixed-effects (TW-FE) estimators.

In this paper, we use the TW-GFE estimator to contrast with the OW-FE and TW-FE estimators so as to perform specification tests and possibly detect more sophisticated structures for the unobserved heterogeneity. Consider the following model formulation

$$y_{it}|x_{it}, \theta_0, \alpha_{it0} \sim f(\cdot|x_{it}'\theta_0 + \alpha_{it0}). \tag{3}$$

The time-varying unobserved heterogeneity α_{it0} is characterized by Assumption 1(b) in Bonhomme et al. (2022a), which states that there exist two vectors ξ_{i0} of dimension d_ξ and λ_{t0} of dimension d_λ, and a Lipschitz-continuous function in its first argument $\alpha(\cdot)$, such that $\alpha_{it0} = \alpha(\xi_{i0}, \lambda_{t0})$. In order to extend the assumptions made by Bonhomme et al. (2022a) to the structures for the unobserved heterogeneity in models (1) and (2), let us reconcile the three specifications under the following characterization for the latent traits:

$$\alpha_{it0} : \begin{cases} \alpha_{i0} \equiv \alpha(\xi_{i0}) & \text{in } (1) \\ \alpha_{i0} + \zeta_{t0} \equiv \alpha(\xi_{i0}, \lambda_{t0}) & \text{in } (2) \\ \alpha_{it0} \equiv \alpha(\xi_{i0}, \lambda_{t0}) & \text{in } (3) \end{cases} \tag{4}$$

$\alpha(\cdot)$ satisfies the requirements in Assumption 1(a) and 1(b) in Bonhomme et al. (2022a), that will be discussed later in more detail. It is also important to the GFE strategy that covariates are affected by the same source of heterogeneity, so that x_{it} depends on μ_{it0}, where $\mu_{it0} = \mu(\xi_{i0}, \lambda_{t0})$, with $\mu(\cdot)$ satisfying the same requirements as $\alpha(\cdot)$. The vectors ξ_{i0} and λ_{t0} are low-dimensional, with $d_\xi = d_\lambda = 1$ as a leading case.

The first-step estimation of the TW-GFE approach deals with the classification of subjects and time occasions into two different sets of groups. Classification relies on performing *kmeans* clustering twice, using the vectors of moments $h_i = \frac{1}{T}\sum_{t=1}^{T} h(y_{it}, x_{it})$ and $w_t = \frac{1}{N}\sum_{i=1}^{N} w(y_{it}, x_{it})$ of fixed dimensions, satisfying Assumption 2 in Bonhomme et al. (2022a) and the conditions for Lemma S1 in Bonhomme et al. (2022b). In particular, both vectors have to be *informative* about ξ_{i0} and λ_{t0}, respectively, meaning that ξ_{i0} can be uniquely recovered from h_i for large T and λ_{t0} can be uniquely recovered from w_t for large N.[2] The two *kmeans* clustering procedures return a number of K groups for the subjects and a different number of L groups for the time occasions, from which two sets of dummies identifying the related group memberships are created.

Cluster dummies for the cross-sectional and time dimensions are then interacted and enter the linear index of the model specified for the response variable as (grouped) fixed effects. Under the regularity conditions collected in Assumption S2 and the requirements in Lemma S1 of Bonhomme et al. (2022b), the two-step TW-GFE estimator is then proved (Corollary S2 *ibidem*) to have asymptotic expansion

$$\tilde{\theta} = \theta_0 + J(\theta_0)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}(\theta_0) + O_p\left(\frac{1}{T} + \frac{1}{N} + \frac{KL}{NT}\right) + O_p(K^{-\frac{2}{d_\xi}} + L^{-\frac{2}{d_\lambda}}) + o_p\left(\frac{1}{\sqrt{NT}}\right), \tag{5}$$

as N, T, K, L tend to infinity, such that $KL/(NT)$ tends to zero. In the above expression, $J(\cdot)$ and s_{it} are the expected Hessian and the score associated with the likelihood function. Three main different sources of bias can be identified: the $1/T$ and $1/N$ terms depend on the number of time occasions and subjects used for h_i and w_t in the classification step; the KL/NT term reflects the estimation of KL group-specific parameters using NT observations; the $K^{-\frac{2}{d_\xi}} + L^{-\frac{2}{d_\lambda}}$ terms refer to the approximation bias arising from the discretization of ξ_{i0} and λ_{t0} via *kmeans*.

The $O_p(\cdot)$ terms in the above expansion can be shown to become $O_p(1/T + 1/N)$ under suitable choices for the number of groups, K and L, and for $d_\xi = d_\lambda = 1$. In particular, Bonhomme et al. (2022a) suggest the following rules:

$$\hat{K} = \min_{K\geq 1}\{K : \hat{Q}(K) \leq \gamma \hat{V}_{h_i}\}, \qquad \hat{L} = \min_{L\geq 1}\{L : \hat{Q}(L) \leq \gamma \hat{V}_{w_t}\}, \tag{6}$$

where $Q(\cdot)$ is the objective function of the *kmeans* problem, V_h and V_w are the variability of

[2] In practice, $\underset{T\to\infty}{\text{plim}}\, h_i = \varphi(\xi_{i0})$ and $\underset{N\to\infty}{\text{plim}}\, w_t = \phi(\lambda_{t0})$, where $\varphi(\cdot)$ and $\phi(\cdot)$ are unknown Lipschitz-continuous functions.

the moments h_i and w_t, respectively, and $\gamma \in (0,1]$ is a user-specified parameter. Smaller values of γ yield a larger number of groups: lowering this value is suggested if moments are weakly informative about unobserved heterogeneity. In our simulation study we experiment with different values of γ.

3 Specification tests

We propose a generalized Hausman test for the specification of the unobserved heterogeneity considering, as null hypotheses, the model specifications portrayed by Equations (1) and (2). The OW-FE and TW-FE estimators are consistent, with an asymptotic bias in case of nonlinear models. The TW-GFE estimator is also consistent but always asymptotically biased. In the presence of more sophisticated forms of unobserved heterogeneity – different from those in (1) and (2) – such as a factor structure, only the TW-GFE estimator is consistent.

To state our main theorem we make the following assumptions.

Assumption 1. *Asymptotics: as* $N, T \to \infty$, $N/T \to \rho^2$, *with* $0 < \rho < \infty$.

Assumption 2. *Unobserved Heterogeneity:*
(i) There exist ξ_{i0} *of fixed dimension* d_ξ *and* λ_{t0} *of fixed dimension* d_λ *and two functions* $\alpha(\cdot)$ *and* $\mu(\cdot)$ *that are Lipschitz continuous in their first argument, such that* $\alpha_{it0} = \alpha(\xi_{i0}, \lambda_{t0})$ *and* $\mu_{it0} = \mu(\xi_{i0}, \lambda_{t0})$*; (ii) the supports of* ξ_{i0} *and* λ_{t0} *are compact.*

Assumption 3. *Sampling: (i)* $(y_{it}, x'_{it})'$, *for* $i = 1, \dots N$ *and* $t = 1, \dots T$, *are i.i.d given* ξ_{i0} *and* λ_{t0}*; (ii)* ξ_{i0} *and* λ_{t0} *are also i.i.d.*

Assumption 4. *Regularity: Let* $\ell_{it}(\alpha_{it}, \theta) = \ln f(Y_{it}|Y_{i,t-1}, X_{it}, \alpha_{it}, \theta)$ *and let* $\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \ell_{it}(\bar{\alpha}(\theta, \xi_{i0}, \lambda_{t0}), \theta)$ *be the target likelihood (Arellano and Hahn, 2007) and* $\bar{\alpha}(\theta, \xi, \lambda) = \underset{\alpha}{\operatorname{argmax}}\ \mathbb{E}_{\xi_{i0}=\xi, \lambda_{t0}=\lambda}(\ell_{it}(\alpha, \theta))$:
(i) $\ell_{it}(\theta, \alpha)$ *is three time differentiable in* (θ, α)*;* θ_0 *is an interior point of the parameter space* Θ*;* Θ *is compact;*
(ii) ℓ_{it} *is strictly concave as a function of* α, $\inf_{\xi,\lambda,\theta} \mathbb{E}_{\xi_{i0}=\xi,\lambda_{t0}=\lambda}\left(-\frac{\partial^2 \ell_{it}(\bar{\alpha}(\theta,\xi,\lambda),\theta)}{\partial\alpha\partial\alpha'}\right) > 0$;
$\mathbb{E}[\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \ell_{it}(\bar{\alpha}(\theta, \xi_{i0}, \lambda_{t0}), \theta)]$ *has an unique maximum at* θ_0 *on* Θ*, and it's second derivative is* $-J < 0$
(iii) Regularity conditions on boundedness of moments and asymptotic covariances in Bonhomme et al. (2022b) Assumption S2 (iv,v) apply.

Assumption 1 depicts rectangular array asymptotics, which is required for the characterization of the asymptotic normal distribution of the considered estimators. Assumption 2 gives the minimal properties of the unobserved heterogeneity in the Bonhomme et al. (2022a) setting. Assumption 3 is more restrictive than that usually required to characterize the asymptotic distribution of ML estimators under rectangular-array asymptotics for fixed-effects models with time heterogeneity. For example, Fernández-Val and Weidner (2016) assume independence over i while relaxing time independence by allowing for α-mixing.[3] Assumption 3 is instead required

[3] See Fernández-Val and Weidner (2016), Assumption 4.1 (ii).

for consistency of the TW-GFE, which effectively rules out the possibility of applying the proposed test to models with (i) feedback effects and (ii) unobserved heterogeneity that depends on dynamic factors . The conditions stated in Assumption 4 are standard requirements for a well-posed maximization problem.

Under Assumptions 1-4 the OW-FE and TW-FE estimators of θ, $\hat{\theta}$, for models (1) and (2), respectively, have the following asymptotic distribution

$$\sqrt{NT}(\hat{\theta}-\theta_0) \xrightarrow{d} N(B;\ I(\theta_0)^{-1}),$$

where B is constant and equal to ρC for the OW-FE estimator, while it is equal to $\rho C_1 + \rho^{-1}C_2$ for the TW-FE estimator. Notice that, in the case of informational orthogonality between the structural and nuisance parameters, $B = 0$, such as in the linear model. Finally, $I(\theta_0)$ is the Information matrix.

Consistency of the TW-GFE estimator relies on Lemma S1 in Bonhomme et al. (2022b) and Assumptions 3 - 4. The resulting asymptotic expansion in Corollary S2 is the one reported in Equation (5). In order to derive the asymptotic distribution of the TW-GFE estimator $\tilde{\theta}$, we need to provide minimal characterization of the $O_p(1/T + 1/N)$ term in said asymptotic expansion, under the appropriate choices for the number of groups suggested by Bonhomme et al. (2022a).

Assumption 5. *The $O_p(1/T + 1/N)$ term takes the form*

$$\frac{D_1}{T} + \frac{D_2}{N} + o_p\left(\frac{1}{T} \vee \frac{1}{N}\right),$$

where D_1 and D_2 are constant.

This assumption is in the spirit of Corollary 2 in Bonhomme et al. (2022a), according to which the $O_p(1/T)$ term in the asymptotic expansion for the one-way GFE estimator with only time-constant unobserved heterogeneity is $E/T + o_p(1/T)$, where E is constant. The asymptotic distribution of $\tilde{\theta}$ can now be characterized by the following theorem, the proof of which follow from standard arguments of ML estimation.

Theorem 1. *Suppose that Assumptions 1-5 and Lemma S1 and Corollary S2 of Bonhomme et al. (2022b) hold, and let $d_\xi = d_\lambda = 1$ then*

$$\sqrt{NT}(\tilde{\theta}-\theta_0) \xrightarrow{d} N\left(D;\ J(\theta_0)^{-1}\right),$$

where $D = D_1\rho + D_2\rho^{-1}$.

Asymptotic normality of both estimators allows us to derive a Hausman test to contrast $\hat{\theta}$ and $\tilde{\theta}$. Under the null hypothesis of correct specification for the unobserved heterogeneity we have

$$H = NT\hat{\delta}'\widehat{W}^{-1}\hat{\delta} \xrightarrow{d} \chi^2_{r,\omega}, \tag{7}$$

where $\hat{\delta} = \hat{\theta} - \tilde{\theta}$ and $\widehat{W}$ is a consistent estimator of its variance W. Under H_0, $\underset{N,T\to\infty}{\text{plim}}\ \hat{\delta} = \delta$, where $\delta = B - D$. Therefore the limiting distribution is a χ^2 with r degrees of freedom and non-centrality parameter $\omega = \delta'\delta$.

A test for H_0 based on H poses two problems. First, there is no guarantee that using the traditional formulation of the Hausman test (Hausman, 1978) will provide a positive definite $\widehat{W}$, as it is difficult to establish which estimator is more efficient under the null hypothesis. Secondly, quantiles of the limiting distribution of H are unknown because of the non-centrality parameter ω. In order to tackle the first issue, we rely on a generalized formulation for the variance of $\hat{\delta}$, put forward by Bartolucci et al. (2015). In particular, let us denote the estimator of the variance of $\hat{\delta}$ as

$$\widehat{W} = NT\left[M\widehat{V}(\hat{\theta},\tilde{\theta})M'\right],$$

where $M = [\mathrm{I}_r, -\mathrm{I}_r]$, I_r is an identity matrix, and

$$\hat{V}(\hat{\theta},\tilde{\theta}) = \begin{pmatrix} I(\hat{\theta}) & 0 \\ 0 & J(\tilde{\theta}) \end{pmatrix}^{-1} S\left(\hat{\theta},\tilde{\theta}\right) \begin{pmatrix} I(\hat{\theta}) & 0 \\ 0 & J(\tilde{\theta}) \end{pmatrix}^{-1},$$

with

$$S\left(\hat{\theta},\tilde{\theta}\right) = \sum_{i=1}^{N}\sum_{t=1}^{T} \begin{pmatrix} g_{it}(\hat{\theta}) \\ s_{it}(\tilde{\theta}) \end{pmatrix} \left(g_{it}(\hat{\theta})', \quad s_{it}(\tilde{\theta})'\right),$$

where $g_{it}(\hat{\theta})$ and $s_{it}(\tilde{\theta})$ are the scores of the observational log-likelihoods associated to the OW-FE (or TW-FE) and TW-GFE approaches, respectively.

We address the second issue concerning the non-centrality of the limiting chi-square distribution by performing parametric percentile bootstrap (MacKinnon, 2006). This approach is in the spirit of Kim and Sun (2016) and Higgins and Jochmans (2022). In particular, they both outline general conditions that guarantee the asymptotic validity of the bootstrap test. We assume these conditions hold in the present context as well, most of which overlap with those used for the above results.

Bootstrap observations, y_{it}^*, are generated under the null hypotheses of correct specification, depicted by Equations (1) or (2). Accordingly, samples are obtained using the OW-FE or TW-FE estimates, with the exogenous covariates held constant. Under the assumptions 1-5 and Lemma 1 and Corollary 1 of Bonhomme et al. (2022b) hold, the bootstrap statistic H^* becomes

$$H^* = NT\hat{\delta}^{*\prime}\left(\widehat{W}^*\right)^{-1}\hat{\delta}^* \xrightarrow{d^*} \chi^2_{r,\omega},$$

where $\xrightarrow{d^*}$ denotes convergence in distribution of the bootstrap measure. Therefore, percentiles can be obtained as

$$q^*_{1-\alpha} = \inf\left\{q^* : \Pr{}^*\left(H^* \leq q^*\right) \geq (1-\alpha)\right\},$$

where $\Pr^*$ denotes the probability conditional on the bootstrap sample.

4 Simulation study

In the following, we describe the design and report the results of an extensive simulation study where we investigate the test empirical size and power properties for the linear and probit model.

4.1 Linear model

We design a Monte Carlo experiment where observations are generated by a linear regression model with two exogenous covariates. We consider two scenarios for the null hypotheses, where the unobserved heterogeneity is specified as in models (1) and (2).

In particular, in the case of only individual time-constant effects, for $i = 1, \dots, N$ and $t = 1, \dots, T$, we generate samples according to the following equation, which we denote as DGP1:

$$\begin{aligned} y_{it} &= x_{it1}\theta_1 + x_{it2}\theta_2 + \alpha_i + \varepsilon_{it}, \\ x_{itj} &= \Gamma_i + N(0,1), \quad \text{for} \quad j = 1, 2, \end{aligned} \tag{8}$$

where $\alpha_i = \varrho\Gamma_i + \sqrt{(1-\varrho^2)}A_i$, with $A_i, \Gamma_i \sim N(0,3)$, and $\varrho = 0.5$. Finally, ε_{it} is an idiosyncratic standard normal error term. We let the coefficients $\theta = (\theta_1, \theta_2)' = (1,2)'$. With this design we explore the size properties of the proposed test comparing the OW-FE with the TW-GFE estimators. Similarly, when we allow for additive individual and time effects, samples are generated according to:

$$\begin{aligned} y_{it} &= x_{it1}\theta_1 + x_{it2}\theta_2 + \alpha_i + \zeta_t + \varepsilon_{it}, \\ x_{itj} &= \Gamma_i + \zeta_t + N(0,1), \quad \text{for} \quad j = 1, 2, \end{aligned} \tag{9}$$

where $\zeta_t \sim N(0,1)$. This design is denoted by DGP2 and it will be used to contrast the TW-FE with the TW-GFE estimators under the null hypothesis of additive effects. In order to investigate the power of the proposed test, the scenario generated under the alternative hypothesis is a linear panel data model with interactive fixed effects. Specifically, samples are generated according to a simplified version of the design outlined by Bai (2009), with one latent factor:

$$\begin{aligned} y_{it} &= x_{it1}\theta_1 + x_{it2}\theta_2 + \alpha_i\zeta_t + \varepsilon_{it}, \\ x_{it} &= \Gamma_i\zeta_t + N(0,1). \end{aligned} \tag{10}$$

We denote the above design as DGP3.

For each scenario, we consider $N = 50, 100$, $T = 10, 20$, and 399 bootstrap draws for each of the 1000 Monte Carlo replications. Only for DGP1 we add $T = 5$ to better highlight our results. It is worth recalling that the performance of the TW-GFE estimator is closely linked to the number of groups chosen for the first-step *kmeans* clusterings. Even under rule (6), this number depends on the variability in the data, which affects how informative h_i, w_t are about the unobserved heterogeneity, and the user-defined parameter γ. We account for the former by allowing for large variances in the stochastic components of the data generating processes, and for the latter by running scenarios where $\gamma = 1, 0.5, 0.05$, resulting in the selection of an increasing number of clusters.

Tables 1 and 2 report the results of the Monte Carlo experiments in four cases. First we compare the OW-FE and the TW-GFE estimators using DGP1 to study the empirical size, while the power of the proposed test is investigated under the alternative process described by DGP3. We then turn to the comparison between the TW-FE and the TW-GFE estimators in the setting where heterogeneity is specified as additive effects under the null hypothesis in

DGP2 and that of DGP3 under the alternative. The tables report the average of the Hausman test H in (7) across simulations, along with the empirical size based on the quantile of a central χ^2_2 (A p.05) and the bootstrap rejection rate (B p.05). We also present the average bias (Bias), standard deviation (SD) and ratio between standard error and SD for both elements of $\hat{\theta}$ and $\tilde{\theta}$. For the TW-GFE estimator, we also report the average selected number of groups according to (6) (Avg $\hat{K}$ and Avg $\hat{L}$).

As expected, the empirical rejection rate based on the asymptotic percentile of the centered chi-square distribution does not attain the nominal size (5%), as it fails to account for the noncentrality parameter arising from the approximation bias in the TW-GFE. It is in fact worth to notice that, even with the large number of groups obtained with $\gamma = 0.05$, the bias of the TW-GFE is sensibly larger than that of the OW-FE, and TW-FE estimators. Indeed the bias of the TW-GFE worsens with fewer clusters, as testified by the results with $\gamma = 0.5, 1$ and the resulting values of the Hausman test. We highlight that γ parameter seem to have no effect on the test value. Nonetheless, the bootstrap distribution is able to mimic the non-centrality of the chi-square, giving rise to a rejection rate close to the nominal one, with a improving performance as T increases. The bootstrap test also presents good rejection rates when the true data generating process has a factor structure for the unobserved heterogeneity. Increasing the number of individuals or enriching information embodied in factors may give power to our test.

The performance of the proposed test can be compared with that of the max-type test put forward by Castagnetti et al. (2015a) to detect factor structures in a linear framework. In the context here considered, the CRT test can be implemented by considering the null hypothesis of no factor structure as a model nested within DGP3 and defined by $H_0 : \zeta_t = \zeta$. [4] The max-type test statistics is

$$S = \max_{1\leq t\leq T} \left[N(\hat{\zeta}_t - \hat{\bar{\zeta}})' \hat{\Sigma}_t^{-1} (\hat{\zeta}_t - \hat{\bar{\zeta}}) \right].,$$

where factors are estimated using the common correlated effects approach by Pesaran (2006),[5] $\hat{\bar{\zeta}}$ is the sample mean of $\hat{\zeta}_t$ and the $\hat{\Sigma}_t$ is and estimate of the asymptotic factor covariance matrix (crf Equation 10 in Castagnetti et al., 2015a). The test statistic S has an asymptotic Gumbel distribution. Table 3 reports the empirical size and power of the CRT test under the hypothesis of no factor structure (DGP1) and the alternative hypothesis of a factor model with 1 latent factor (DGP3). As also confirmed by the Authors, the test attains the correct size only for larger values of T (approximately $T = 30$), which rules out the CRT test as a viable alternative to our approach in short panels.

[4] Castagnetti et al. (2015a) also proposed a max-type test for $H_0 : \Gamma_i = \Gamma$, that is expression 10 collapsing to a model with only time effects only.

[5] It is worth recalling that the approach by Castagnetti et al. (2015a) can in general be implemented in models with heterogeneous slopes.

Table 1: Simulation results: Linear model, OW-FE vs TW-GFE

		DGP1																		DGP3			
		$\gamma = 0.05$						$\gamma = 0.5$						$\gamma = 1$						$\gamma = 0.05$			
		T=5		T=10		T=20		T=5		T=10		T=20		T=5		T=10		T=20		T=10		T=20	
	N	50	100	50	100	50	100	50	100	50	100	50	100	50	100	50	100	50	100	50	100	50	100
Hausman	H	1.276	1.498	1.515	1.724	1.987	2.200	2.779	4.371	3.865	5.952	5.624	7.953	3.160	4.732	4.794	7.704	7.048	11.081	13.471	25.311	25.654	45.201
	A p.05	0.018	0.018	0.024	0.034	0.064	0.075	0.119	0.258	0.208	0.420	0.338	0.525	0.152	0.299	0.298	0.501	0.421	0.641	0.503	0.659	0.644	0.729
	B p.05	0.042	0.042	0.048	0.042	0.045	0.059	0.045	0.036	0.047	0.040	0.044	0.048	0.033	0.033	0.047	0.046	0.052	0.050	0.542	0.662	0.676	0.748
$\hat{\theta}_1$	Bias1	0.005	0.001	0.001	0.000	-0.002	0.000	0.005	0.001	0.001	0.000	-0.002	0.000	0.005	0.001	0.001	0.000	-0.002	0.000	0.403	0.403	0.406	0.406
	SD1	0.072	0.052	0.047	0.034	0.033	0.023	0.072	0.052	0.047	0.034	0.033	0.023	0.072	0.052	0.047	0.034	0.033	0.023	0.075	0.054	0.057	0.040
	SE/SD1	0.982	0.966	1.009	0.988	0.979	1.017	0.982	0.966	1.009	0.988	0.979	1.017	0.982	0.966	1.009	0.988	0.979	1.017	0.861	0.846	0.771	0.794
$\hat{\theta}_2$	Bias2	-0.001	-0.001	-0.001	-0.001	0.002	0.000	-0.001	-0.001	-0.001	-0.001	0.002	0.000	-0.001	-0.001	-0.001	-0.001	0.002	0.000	0.403	0.406	0.405	0.409
	SD2	0.071	0.051	0.047	0.032	0.033	0.023	0.071	0.051	0.047	0.032	0.033	0.023	0.071	0.051	0.047	0.032	0.033	0.023	0.075	0.052	0.055	0.040
	SE/SD2	1.001	0.980	0.999	1.029	0.973	1.010	1.001	0.980	0.999	1.029	0.973	1.010	1.001	0.980	0.999	1.029	0.973	1.010	0.855	0.874	0.810	0.780
$\tilde{\theta}_1$	Bias1	-0.001	-0.005	-0.002	-0.004	-0.005	-0.003	-0.017	-0.032	-0.018	-0.025	-0.016	-0.017	-0.018	-0.034	-0.023	-0.036	-0.023	-0.028	0.332	0.328	0.334	0.336
	SD1	0.078	0.056	0.050	0.035	0.035	0.024	0.092	0.066	0.062	0.045	0.046	0.030	0.109	0.080	0.077	0.057	0.055	0.037	0.114	0.097	0.094	0.088
	SE/SD1	0.993	0.970	0.998	0.998	0.946	1.004	0.951	0.958	0.910	0.906	0.810	0.870	0.868	0.859	0.792	0.778	0.735	0.779	0.564	0.465	0.474	0.358
$\tilde{\theta}_2$	Bias2	-0.013	-0.014	-0.008	-0.009	-0.003	-0.005	-0.049	-0.065	-0.033	-0.044	-0.017	-0.025	-0.049	-0.077	-0.038	-0.059	-0.026	-0.040	0.327	0.324	0.331	0.335
	SD2	0.077	0.055	0.051	0.034	0.035	0.024	0.091	0.064	0.064	0.043	0.047	0.030	0.105	0.080	0.075	0.055	0.055	0.038	0.115	0.103	0.101	0.093
	SE/SD2	1.005	0.983	0.981	1.030	0.949	0.990	0.970	0.993	0.887	0.943	0.790	0.883	0.907	0.864	0.822	0.807	0.733	0.753	0.561	0.440	0.442	0.340
	Avg$\hat{K}$	45.347	85.877	46.200	88.350	47.304	91.572	26.755	40.726	29.483	46.521	33.799	56.047	18.355	24.706	21.429	30.020	26.408	39.381	9.487	12.929	8.789	10.227
	Avg$\hat{L}$	1.269	1.222	1.364	1.324	1.436	1.395	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	4.041	3.930	5.480	5.374

1000 Monte Carlo (MC) replications. "H" is the average of the Hausman test statistic, across MC replications. "p.05" denotes the rejection rate for a nominal size of 5%. "A p.05" is based on the 95^{th} percentile of central a χ^2_2 distribution. "B p.05" is based on the 95^{th} percentile of the empirical distribution determined via 399 bootstrap replications. "Bias" is the mean bias, "SD" and "SE" denote the standard deviation over the MC replications and the average estimated standard error, respectively, those are reported for two regressors. "Avg $\hat{K}$" and "Avg $\hat{L}$" report the average number of groups for individuals and time occasions obtained in the first step.

4.2 Probit model

We also investigate the small sample properties of our test in a nonlinear setting, specifically considering the probit model. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$, we generate samples according to the following equation, which we denote as DGP1-NL:

$$\begin{aligned} y_{it} &= \mathbf{I}\left(x_{it}\theta + \alpha_i + \varepsilon_{it} \geq 0\right), \\ x_{it} &= \kappa\left[\Gamma_i + N(0,1)\right], \end{aligned} \tag{11}$$

where $\mathbf{I}(\cdot)$ is an indicator function, $\alpha_i = \varrho\Gamma_i + \sqrt{(1-\varrho^2)}A_i$, $A_i, \Gamma_i \sim N(1,1)$, $\varrho = 0.5$, and ε_{it} is an idiosyncratic standard normal error term. The slope parameter θ is set equal to 1. Following Hahn and Newey (2004), the variance contributions of the three elements in the linear index are roughly 0.2, 1, and 1, respectively, so that we rescale the variance of x_{it} by letting $\kappa = \sqrt{(1/10)}$. When we allow for additive individual and time effects, samples are generated according to

$$\begin{aligned} y_{it} &= \mathbf{I}\left(x_{it}\theta + \alpha_i + \zeta_t + \varepsilon_{it} \geq 0\right), \\ x_{it} &= \kappa\left[\Gamma_i + \zeta_t + N(0,1)\right], \end{aligned} \tag{12}$$

where $\zeta_t \sim N(1,1)$ and the sum $\alpha_i + \zeta_t$ is rescaled to have unit variance. This design is denoted by DGP2-NL. We evaluate the power of the proposed test in scenarios generated under the alternative hypothesis of interactive fixed effects with one latent factor, that is

$$\begin{aligned} y_{it} &= \mathbf{I}\{x_{it}\theta + \alpha_i\zeta_t + \varepsilon_{it} \geq 0\}, \\ x_{it} &= \kappa[\Gamma_i\zeta_t + N(0,1)], \end{aligned} \tag{13}$$

where again the product $\alpha_i\zeta_t$ is rescaled to have unit variance. We refer to this design as DGP3-NL. For each experiment, we consider $N = 50, 100$, $T = 10, 20$, and 399 bootstrap draws for each of the 1000 Monte Carlo replications. When dealing with DGP1-NL, we try different values of $\gamma = 0.25, 0.5, 1$ in order to evaluate sensitivity of the test to number of groups.

Tables 4 and 5 show the simulation results for experiments where we compare the OW-FE with TW-GFE estimators under DGP1-NL and DGP3-NL, and the TW-FE and TW-GFE estimators under DGP2-NL and DGP3-NL, respectively. As expected, the bootstrap test approaches the correct size under smaller values of γ, i.e. when we impose greater number of groups, while the one based on the asymptotic critical value fails to account for the noncentrality parameter of chi-square distribution. The power analysis shows that our test behaves well when OW-FE is contrasted with TW-GFE under DGP3-NL, while the power decreases drastically when TW-FE is contrasted with TW-GFE under the same scenario. This is due to the ability of two-way fixed effects specification to approximate a model with interacted effects, as the relatively small bias of TW-FE estimator with respect to TW-GFE makes clear. TW-GFE is less biased with respect to OW-FE under both DGP1-NL and DGP3-NL and under all values of γ, while, under DGP2-NL TW-GFE exhibits a larger bias than its counterpart due to the approximation issues mentioned above.

Table 2: Simulation results: Linear model, TW-FE vs TW-GFE

		DGP2 $\gamma = 0.05$				**DGP3** $\gamma = 0.05$			
		T=10		T=20		T=10		T=20	
	N	50	100	50	100	50	100	50	100
Hausman	H	3.195	3.397	3.269	3.082	13.445	24.429	22.660	44.798
	A p.05	0.145	0.156	0.147	0.126	0.533	0.631	0.618	0.720
	B p.05	0.042	0.041	0.054	0.039	0.550	0.642	0.657	0.737
$\hat{\theta}_1$	Bias1	0.001	-0.001	-0.000	-0.000	0.408	0.403	0.409	0.408
	SD1	0.048	0.034	0.032	0.024	0.076	0.054	0.057	0.041
	SE/SD1	1.003	0.989	1.024	0.955	0.855	0.854	0.788	0.778
$\hat{\theta}_2$	Bias2	-0.000	-0.001	0.001	-0.000	0.401	0.404	0.407	0.408
	SD2	0.048	0.033	0.034	0.023	0.072	0.054	0.056	0.042
	SE/SD2	0.993	1.003	0.978	0.992	0.906	0.856	0.791	0.755
$\tilde{\theta}_1$	Bias1	0.011	-0.003	0.010	0.002	0.334	0.333	0.344	0.337
	SD1	0.094	0.068	0.057	0.041	0.118	0.099	0.095	0.092
	SE/SD1	0.916	0.893	0.869	0.880	0.549	0.457	0.468	0.345
$\tilde{\theta}_2$	Bias2	-0.006	-0.022	0.004	-0.006	0.322	0.326	0.339	0.334
	SD2	0.096	0.068	0.056	0.043	0.115	0.104	0.096	0.095
	SE/SD2	0.894	0.896	0.888	0.848	0.565	0.438	0.462	0.334
	Avg$\hat{K}$	42.512	78.067	44.353	83.367	9.753	12.726	8.724	10.724
	Avg$\hat{L}$	7.037	7.777	11.232	12.698	4.084	3.918	5.342	5.350

1000 Monte Carlo (MC) replications. "H" is the average of the Hausman test statistic, across MC replications. "p.05" denotes the rejection rate for a nominal size of 5%. "A p.05" is based on the 95^{th} percentile of a central χ^2_2 distribution. "B p.05" is based on the 95^{th} percentile of the empirical distribution determined via 399 bootstrap replications. "Bias" is the mean bias, "SD" and "SE" denote the standard deviation over the MC replications and the average estimated standard error, respectively (reported for two regressors) . "Avg $\hat{K}$" and "Avg $\hat{L}$" report the average number of groups for individuals and time occasions obtained in the first step.

Table 3: CRT test for no factor structure

	DGP1			**DGP3**	
N	$T=5$	$T=10$	$T=20$	$T=10$	$T=20$
50	1.000	0.623	0.043	1.000	1.000
100	1.000	0.823	0.028	1.000	1.000

Table 4: Simulation results: Probit model, OW-FE vs TW-GFE

		DGP1-NL												DGP3-NL			
		$\gamma = 0.25$				$\gamma = 0.5$				$\gamma = 1$				$\gamma = 0.25$			
		T=10		T=20		T=10		T=20		T=10		T=20		T=10		T=20	
	N	50	100	50	100	50	100	50	100	50	100	50	100	50	100	50	100
Hausman	H	0.926	1.406	0.855	1.039	2.327	4.615	2.436	4.949	4.049	8.140	4.673	9.276	13.107	31.054	19.591	51.124
	A p.05	0.028	0.086	0.032	0.047	0.210	0.552	0.226	0.606	0.448	0.827	0.539	0.886	0.783	0.964	0.925	0.994
	B p.05	0.040	0.027	0.050	0.039	0.021	0.017	0.033	0.027	0.019	0.012	0.019	0.020	0.731	0.950	0.835	0.988
$\hat{\theta}$	Bias	0.141	0.135	0.067	0.066	0.141	0.135	0.067	0.066	0.141	0.135	0.067	0.066	0.747	0.731	0.639	0.632
	SD	0.282	0.193	0.174	0.122	0.282	0.193	0.174	0.122	0.282	0.193	0.174	0.122	0.281	0.224	0.175	0.140
	SE/SD	0.907	0.934	0.971	0.975	0.907	0.934	0.971	0.975	0.907	0.934	0.971	0.975	0.707	0.627	0.741	0.652
$\tilde{\theta}$	Bias	0.110	0.090	0.066	0.051	0.017	0.005	0.003	-0.003	-0.066	-0.080	-0.049	-0.053	0.176	0.126	0.142	0.086
	SD	0.284	0.193	0.184	0.125	0.253	0.174	0.166	0.117	0.239	0.165	0.158	0.113	0.307	0.202	0.203	0.138
	SE/SD	0.912	0.924	0.940	0.961	0.946	0.962	0.984	0.977	0.941	0.958	0.999	0.980	0.832	0.847	0.874	0.869
	Avg $\hat{K}$	20.445	27.279	25.870	36.731	13.713	16.601	18.482	23.625	8.555	9.640	12.172	14.160	9.619	11.456	12.635	15.373
	Avg $\hat{L}$	1.992	1.996	2.302	2.283	1.281	1.249	1.362	1.338	1.010	1.011	1.001	1.000	6.724	7.369	10.186	11.654

1000 Monte Carlo (MC) replications. "H" is the average of the Hausman test statistic, across MC replications. "p.05" denotes the rejection rate for a nominal size of 5%. "A p.05" is based on the 95^{th} percentile of a central χ^2_1 distribution. "B p.05" is based on the 95^{th} percentile of the empirical distribution determined via 399 bootstrap replications. "Bias" is the mean bias, "SD" and "SE" denote the standard deviation over the MC replications and the average estimated standard error, respectively. "Avg $\hat{K}$" and "Avg $\hat{L}$" report the average number of groups for individuals and time occasions obtained in the first step.

5 Final remarks

We propose a specification test for the form of the unobserved heterogeneity in panel data models. The test is based on the recently proposed grouped fixed-effects approach and served to detect departures from the commonly assumed time-invariant or additive fixed-effects specifications.

The proposed approach is a generalized Hausman test whose asymptotic distribution is a non-central chi square because of the bias arising from incidental parameters, for both the ML estimators that are being contrasted (at least in the non-linear case), and from the approximation error induced by discretization of the unobserved heterogeneity that is performed with the grouped fixed-effects approach. The use of bootstrap critical values successfully corrects the empirical size of the test and yields satisfactory power properties. The proposed test also emerges as a viable alternative to existing procedures with short panel datasets.

It is worth to mention that the work is still preliminary and it will be soon further developed. In particular, we aim at extending the simulation study so as to consider:

- A wider set of scenarios under both H_0 and H_1, considering in particular $N = 500$. In addition, the behavior of the test statistic will be studied under different setting for the TW-GFE estimator, that is, using more values for the user-defined γ parameter and a data generating process that allows us to govern how informative population moments are about the unobserved heterogeneity.

- Investigation of the finite-sample performance of the proposed test under departures from the required assumptions (for instance, violations of the sampling assumptions for the TW-GFE).

Table 5: Simulation results: Probit model, TW-FE vs TW-GFE

		DGP2-NL $\gamma = 0.25$				**DGP3-NL** $\gamma = 0.25$			
		T=10		T=20		T=10		T=20	
	N	50	100	50	100	50	100	50	100
Hausman	H	0.924	0.847	1.574	1.868	1.882	4.352	2.370	4.433
	A p.05	0.031	0.030	0.102	0.144	0.152	0.367	0.196	0.381
	B p.05	0.037	0.043	0.039	0.031	0.093	0.221	0.158	0.337
$\hat{\theta}$	Bias	0.210	0.166	0.108	0.092	0.297	0.275	0.210	0.194
	SD	0.316	0.192	0.182	0.130	0.268	0.183	0.169	0.126
	SE/SD	0.869	0.982	0.971	0.948	0.851	0.867	0.883	0.828
$\tilde{\theta}$	Bias	0.263	0.177	0.215	0.170	0.185	0.120	0.137	0.090
	SD	0.364	0.206	0.223	0.149	0.299	0.189	0.207	0.135
	SE/SD	0.829	0.963	0.903	0.916	0.859	0.905	0.851	0.888
	Avg $\hat{K}$	13.188	16.010	17.141	21.847	9.784	11.490	12.528	15.121
	Avg $\hat{L}$	6.824	7.466	10.412	11.897	6.736	7.289	10.114	11.616

1000 Monte Carlo (MC) replications. "H" is the average of the Hausman test statistic, across MC replications. "p.05" denotes the rejection rate for a nominal size of 5%. "A p.05" is based on the 95^{th} percentile of a central χ^2_1 distribution. "B p.05" is based on the 95^{th} percentile of the empirical distribution determined via 399 bootstrap replications. "Bias" is the mean bias, "SD" and "SE" denote the standard deviation over the MC replications and the average estimated standard error, respectively. "Avg $\hat{K}$" and "Avg $\hat{L}$" report the average number of groups for individuals and time occasions obtained in the first step.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society, Series B*, 32:283–301.

Ando, T. and Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31(1):163–191.

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, 43:381.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bartolucci, F., Belotti, F., and Peracchi, F. (2015). Testing for time-invariant unobserved heterogeneity in generalized linear models for panel data. *Journal of Econometrics*, 184(1):111–123.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2012). *Latent Markov models for longitudinal data*. CRC Press.

Bester, C. A. and Hansen, C. B. (2016). Grouped effects estimators in fixed effects models. *Journal of Econometrics*, 190(1):197–208.

Bonhomme, S., Lamadon, T., and Manresa, E. (2022a). Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643.

Bonhomme, S., Lamadon, T., and Manresa, E. (2022b). Supplement to "discretizing unobserved heterogeneity". *Econometrica supplementary material*, 90(2):1–21.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Castagnetti, C., Rossi, E., and Trapani, L. (2015a). Inference on factor structures in heterogeneous panels. *Journal of econometrics*, 184(1):145–157.

Castagnetti, C., Rossi, E., and Trapani, L. (2015b). Testing for no factor structures: On the use of hausman-type statistics. *Economics Letters*, 130:66–68.

Cavaliere, G., Gonçalves, S., and Nielsen, M. Ø. (2022). Bootstrap inference in the presence of bias. *arXiv preprint arXiv:2208.02028*.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47:225–238.

Chen, M., Fernández-Val, I., and Weidner, M. (2021). Nonlinear factor models for network and panel data. *Journal of Econometrics*, 220(2):296–324.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312.

Hahn, J. and Moon, H. R. (2010). Panel data models with finite number of multiple equilibria. *Econometric Theory*, 26(3):863–881.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, pages 1251–1271.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, pages 271–320.

Higgins, A. and Jochmans, K. (2022). Bootstrap inference for fixed-effect models. *arXiv preprint arXiv:2201.11156*.

Kim, M. S. and Sun, Y. (2016). Bootstrap and k-step bootstrap bias corrections for the fixed effects estimator in nonlinear panel data models. *Econometric Theory*, 32(6):1523–1568.

Li, H., Lindsay, B. G., and Waterman, R. P. (2003). Efficiency of projected score methods in rectangular array asymptotics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):191–208.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *Economic Record*, 82:S2–S18.

MacLahlan, G. and Peel, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Su, L., Shi, Z., and Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264.

Westerlund, J. (2019). Testing additive versus interactive effects in fixed-t panels. *Economics Letters*, 174:5–8.

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*.