Latent grouped structures in panel data: a review

Pionati, Alessandro

Marche Polytechnic University

12 March 2025

Latent grouped structures in panel data: a review*

Alessandro Pionati[a]

[a]Marche Polytechnic University (Italy)

E-mail: a.pionati@pm.univpm.it

Department of Economics and Social Sciences, P.le Martelli 8, 60121 Ancona (Italy)

Preliminary draft

Abstract

Latent group structures in panel data models are a new and powerful approach to deal with unobserved heterogeneity in a parsimonious way. This review, with a special focus on grouped structure in unobservable traits, first analyzes the limits and opportunities of Bonhomme and Manresa (2015a)'s Grouped Fixed Effects (GFE) estimator, also discussing the literature it contributed to create. A rich selection of models enhancing clustered heterogeneity at a slope level, starting from Su et al. (2016a), is then presented. A short section investigates how the applied literature has employed in practice the GFE. Finally, the GFE of Bonhomme et al. (2022) is presented in detail together with its limits and advantages.

Keywords: Grouped Fixed Effects, Fixed Effects, Discrete Heterogeneity

JEL Classification: C13,C23

*I am grateful to Claudia Pigini, Francesco Valentini, Pavel Čížek, Laura Serlenga and Silvia Sarpietro for their helpful comments and suggestions.

1 Introduction

Correct specification of unobserved heterogeneity (UH) is crucial in panel data models. The fixed effects (FE) approach, the most widely employed technique in empirical applications (Hsiao, 2022), directly estimates the parameters concerning unobservable individual-specific traits along with other quantities that are usually of interest for the analyst. Another popular approach, known as the random effects (RE) approach models the UH parameters as random variables. Despite being popular, the RE approach assumes the absence of correlation between the unobservable traits and the regressors included in the model specification, which is not usually the case in economics. The FE procedure instead could be employed even if such a correlation is present in the data and that is why it has become the most employed tool in empirical economic studies. In this review, we focus on fixed effects models only.

In the fixed effects panel data literature different specifications of UH - from simple to more involved ones - are studied, according to the (assumed) form of the latent characteristics.

Let us start by defining a simple linear panel data model:

$$y_{it} = x_{it}\theta + u_{it} \tag{1}$$

For $i = 1, \ldots, N$ and $t = 1, \ldots, T$ we observe N individuals for T time spans and x_{it} gathers the p regressors used in the analysis associated with a conformable vector of parameters of interest θ, which is supposed to be the main object of analysis. We define u_{it} as the idiosyncratic error: however, let us suppose for the moment that this component of the model has a slightly more complex shape such as:

$$u_{it} = \alpha_i + \epsilon_{it} \tag{2}$$

If this is the case, the error term involves an individual-specific (time-invariant) component α_i. Controlling for this unobserved individual heterogeneity in models such as (1) is straightforward using the so-called one-way fixed effects approach (OW-FE henceforth). At the same time, in contexts in which the heterogeneity is supposed to be time-varying we have that a suitable specification for the heterogeneity takes form:

$$u_{it} = \alpha_i + \lambda_t + \epsilon_{it} \tag{3}$$

The additive representation for the unobservable traits is the current mainstream approach in empirical panel data modeling (Wooldridge, 2021). Models like (1) with an additive specification for heterogeneity such as expression 3 are usually estimated with the so-called two-way fixed effects approach (TW-FE).

Both OW-FE and TW-FE estimations are carried out using the standard least squares approach, based upon de-meaning procedures, which under mild assumptions gives consistent and asymptotically normal distributed estimators of the parameters of interest (Hsiao, 2022).

Generalizations of model (1) with both time invariant and time-varying specification of heterogeneity to well-studied nonlinear models, such as the logit or probit panel data models, have been object of interest in econometric literature: Maximum likelihood (ML) estimation of parameters of interest in most common nonlinear models could be carried out, but the resulting estimator is known to suffer from the incidental parameters problem (IPP): when T is small, the noise in the estimation of the fixed effects influences the estimation of the common parameters, unlike what happens in linear models. Under fixed-T asymptotics (Neyman and Scott, 1948; Lancaster, 2000) the IPP leads to inconsistency of the MLE. Under large N, T asymptotics - with N,T growing at the same rate (Li et al., 2003) - ML estimators are generally consistent, but they exhibit an asymptotic bias which can be removed. A number of solutions is actually offered to the analyst: analytically removing the bias (Hahn and Newey, 2004; Carro, 2007; Bester and Hansen, 2009; Hahn and Kuersteiner, 2011; Bartolucci et al., 2016; Fernández-Val and Weidner, 2016), adopting a jackknife estimator (Hahn and Newey, 2004; Dhaene and Jochmans, 2015; Fernández-Val and Weidner, 2016) or relying on conditional inference (Andersen, 1970; Chamberlain, 1980) - which indeed works also when T is small as individual nuisance parameters are not actually estimated - for models with OW-FE (Bartolucci and Nigro, 2010, 2012) and generalizations to TW-FE (Charbonneau, 2017; Bartolucci et al., 2024). For details see Arellano and Hahn (2007) and Valentini et al. (2023).

Another specification for UH that has acquired growing importance in recent studies is the one portrayed by a factor structure. The two seminal contributions are the common correlated effects by Pesaran (2006) and the interacted fixed effects (IFE) by Bai (2009). The latter, originally focused on linear models only, has been extended to popular nonlinear models by Chen et al. (2021). Models with a factor structure generally assume that unobservables in (1) have latent characterization, such as

$$u_{it} = \alpha_i' \lambda_t + \epsilon_{it} \tag{4}$$

where individual specific loadings ($R \times 1$) and time-varying unobserved factors ($R \times 1$) enter in the model in a multiplicative way (the number of factors is R).

While the Pesaran (2006)'s approach is based on proxying the common factors via cross-sectional averages of the regressors and the dependent variable, the Bai (2009)'s approach directly estimates the common factors along with the parameters of interest. Bai (2009)'s procedure, based on an iterative least squares minimization which makes use of principal component analysis for factors extraction, gives consistent and asymptotically normal distributed estimators. However, despite its increasing popularity in panel data econometrics the IFE estimation is hindered by three issues: i) the objective function is not globally convex, meaning that it is likely to run into local minima (Moon and Weidner, 2023) ii) the reliability of the iterative procedure crucially depends on the consistency of the parameter estimates chosen as the starting point for the algorithm (Hsiao, 2018) iii) the number of factors should be known *a priori* (Bai, 2009). The latter issue has been addressed by Moon and Weidner (2015), while the generalization of the Bai's model to predetermined covariates is discussed in Moon and Weidner (2017). In a nonlinear framework, estimation issues are amplified (Chen et al., 2021). Although recent contributions - such as the introduction of the nuclear norm regularization (Moon and Weidner, 2023) for a well-behaved optimization problem - have tried to deal with the aforementioned challenges, it is straightforward that a clear trade-off between the simplicity of the FE and the flexibility of factor models could be found[1].

The following literature review wants to explore a different topic. The idea of exploiting latent group structures in panel data is not new, especially in statistical literature where the presence of models assuming discrete heterogeneity is consolidated. In econometrics, Hahn and

[1]The Hausman test developed in Kapetanios et al. (2023) allows one to test for the correct specification of UH between a multiplicative factor or a simpler structure .

Moon (2010) argued that the group structure may have sound foundations in economic models where multiplicity of Nash equilibria could be found: they consider a nonlinear panel data model where the parameter of interest is common to individuals whereas the fixed effects have finite support.

However, in fixed effects literature, two recent contributions (Bonhomme and Manresa (2015a) and Bonhomme et al. (2022), BM15 and BLM22 henceforth) have attracted attention on the possibility of cleverly using clustering procedures in addition to fixed effects modelling for a more precise estimation of parameters of interest: we refer to these methods as *grouped fixed effects* (GFE) estimators. Since the two contributions are affine in the spirit but rather different in the asymptotic framework, we discuss them and the literature flows that have spurred from them in two different Sections. The choice of BM15 as a starting point of the analysis of GFE methods and as one of the main object of the focus in this review when dealing with discrete UH is twofold: i) the contribution has gained undoubted popularity, the estimation procedure is convenient for applied economists and intuitive with respect to the other contributions explored above; moreover there are a number of studies in applied microeconometrics literature that explore the possibility of employing BM15 approach, especially in health and labor economics literature ii) BM15 could be seen as the most relevant forerunner of the BLM22 approach, that is the actual object of the thesis. The main differences between the BM15 and the BLM22 approach are that the first contribution supposes that the UH is discrete, that a latent group structure is present in the unobservable components, and that, after a proper (*a priori* unknown) number of cluster is chosen and a suitable clustering procedure is applied, it is possible to consistently estimates grouped fixed effects in a linear framework, via group membership dummies interacted with time dimension, in a standard maximum likelihood procedure. In this way, the grouped fixed effects approach allows the analyst to deal with the unobservable latent group structure.

BLM22 postulates instead a different goal for clustering: the main idea is that unspecified and possibly time-varying UH could be approximated throughout grouped fixed effects. Clustering serves here as an approximation tool. BLM22 works for both linear and nonlinear models, but, unlike BM15, does not allow for dynamic models when time-varying heterogeneity is involved.

At the same time the econometrics literature has also focused on the pursuit of heterogeneity in slope coefficients: the idea of clustering individuals with the goal of estimating different slope coefficients has been exploited by the seminal contribution of Su et al. (2016a), in which the novel C-Lasso approach has been firstly introduced. Part of the paper is also devoted to main recent contributions to this different way of conceiving clustered latent structures in panel data: main challenges and non-trivial issues are analyzed as well.

The paper is organized as follows: Section 2 deals with latent clustered patterns of unobservable heterogeneity, introduces the BM15 approach, explains the possible issues and discusses the possible solutions appeared in the following literature, together with some of the most interesting extensions of this method; Section 3 discusses the literature on heterogeneous grouped slopes; Section 4 introduces relevant empirical studies that have employed grouped fixed effects or procedures for estimating heterogeneous slopes; Section 5 describes the BLM22 method in details, discusses the major limits and briefly examinates the subsequent literature. Finally, Section 6 concludes.

2 Latent clustered patterns of UH

In this Section the BM15 model is briefly discussed together with relevant contributions in the literature that have tried to improve the main drawbacks.

2.1 BM15 and its literature: methodological and practical challenges

BM15 deal with the latent grouped structure by estimating grouped fixed effects and accounting for cluster heterogeneity in the unobservables. In BM15 the object of interest is a linear panel data model, possibly dynamic, in which the UH is supposed to be discrete and described by a number of support points G_0, that is the number of unobservable clusters in data. The number of clusters is finite and needs to be specified by the analyst; moreover this number is not estimated along with the other parameters but selected using a suitable Information Criterion. The model of interest is then (in the static specification):

$$y_{it} = x_{it}'\theta + \alpha_{g_{(i)}t} + \epsilon_{it}, \tag{5}$$

for $i = 1, \ldots, N, t = 1, \ldots, T, g_i = 1, \ldots, G_0$ and where the $\alpha_{g_{(i)}t}$ are the grouped fixed effects to be estimated and each individual is classified in one of the $g_{(i)} \ldots \hat{G}$ groups. To account for time heterogeneity, the model also includes time trajectories assumed to be cluster-specific.

The BM15 approach introduces a number of advantages. From an economic standpoint it is completely reasonable to suppose that some economic phenomena have grouped structure in observables and unobservables components: consider for instance the "convergence clubs" theory for economic growth revisited by Lin and Ng (2012)[2] or the relationship between democracy and income, put forward by Acemoglu et al. (2008), and revisited - with a rich discussion in the following literature - by BM15. Another advantage is that GFE lead to a more parsimonious specification by shrinking the number of parameters while accounting for time heterogeneity: as it is well-known, the estimation of standard one-way or two-way fixed effects models in large datasets comes with the estimation of a non-negligible number of parameters. Moreover, as discussed in BM15, the GFE estimator could represent a flexible yet valid alternative to Bai's IFE estimator, as it is proven to perform better in terms of bias when the underlying structure of heterogeneity is clustered, at least in simulations.

A two-step iterative algorithm assigns cluster memberships in the first step and estimates grouped fixed effects in the second step. In the clustering step they propose to solve the optimal group assignment for model 5:

$$\hat{g}_i(\theta, \alpha) = \text{argmin}_{g \in \{1,\ldots,G\}} \sum_{t=1}^{T} (y_{it} - x_{it}'\theta - \alpha_{g(i)t})^2,$$

In the second step, a standard least squares problem is solved after individuals are classified in groups

$$(\hat{\theta}, \hat{\alpha}) = \text{argmin}_{(\theta,\alpha) \in \Theta \times A^{GT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}'\theta - \alpha_{\hat{g}_{(i)}t})^2,$$

[2] In Lin and Ng (2012) application the grouped heterogeneity is present at a common parameters level.

where $\Theta \times \mathcal{A}^{GT}$ is the parameter space. Iteration until numerical convergence follows. The Authors suggest two different algorithms for clustering, the first one is based on the *classification likelihood* principle (Bryant and Williamson, 1978) and explores all possible classification of N individuals in G groups: the task could be easily overwhelming for a large dataset, since the computational burden increases with N. The second algorithm is based on the well known *k-means* method (Pollard, 1981) and is suited for larger scale problems. However, this second method crucially requires to try different starting guesses for the *k-means* problem. The solution of the least squares problem does not generally pose any issue, since it relies on a standard maximum likelihood approach with groups specific dummies interacted with time dummies. One of the most relevant aspects on which the literature has focused after the BM15 paper - discussed in more details below - is the research of faster algorithms or different techniques for retrieving group memberships.

The proper number of groups is specified by the econometrician and it is not known *a priori*. Following the literature on determining the number of factors in factors models (Bai and Ng, 2002; Bai, 2009) BM15 propose an information criterion based procedure for selecting the number of groups. Their Bayesian Information Criterion (BIC) writes:

$$BIC(G) = \frac{1}{NT}\sum_{i}^{N}\sum_{t}^{T}(y_{it} - x_{it}'\hat{\theta}^{(G)} - \hat{\alpha}_{\hat{g}_{(i)}^{(G)}t})^2 + \hat{\sigma}^2\frac{GT+N+C}{NT}\ln(NT) \tag{6}$$

where $\hat{\sigma}^2$ is a consistent estimator of the variance of u_{it} and C is the number of common parameters. This Information Criterion has proven to perform reasonably well in simulations in terms of suggesting the right number of groups, and that is why all the following literature tend to suggest different information criterion more than different rules for the number of groups. The BM15 GFE is a consistent and asymptotically normal distributed estimator. Moreover, the GFE estimator and the infeasible least squares estimator with known population groups are asymptotically equivalent. In order to get these results some relevant assumptions on group structure are needed. The G population groups must be well-separated and in the limit groups must be populated by a large number of individuals. The separation assumption fails, for instance, when the number of groups is underidentified. Further assumptions are stated on the difference between two α^0_{gt} different (in g) processes, the resulting process supposed to be a strong-mixing one. This assumption is needed to bound misclassification probabilities. As predetermined covariates are allowed, GFE estimation could be carried out on dynamic models

In BM15 framework there is a distinction between the under- and the overidentification of the real number of clusters, the latter being less "dangerous" than the former: the case for $\hat{G} > G_0$ could be seen as the inclusion in the model specification of irrelevant regressors, and, under BM15 asymptotic framework does not cause inconsistency of the common parameters θ but results in biased estimates of grouped effects. It is mainly an efficiency problem. Under the $\hat{G} < G_0$ scenario, the bias is present in both common parameters and grouped fixed effects and could be seen as an omitted variable bias problem.

Another concern for the econometrician is the misclassification problem: even under the hypothesis that $G_0 = \hat{G}$ there is the possibility that, in small samples, individuals are assigned to wrong groups. Although misclassification mainly represents a short panel problem, it could lead to dispersion on the finite sample performances of the GFE. However, it is worth noticing that in BM15's apparatus of simulations the misclassification rate is not really an issue when G_0 is small and, even in presence of high misclassification rates when G_0 is increasing ($G_0 = 10$ groups yield to 55% misclassification rate in their experiment), wrong classification is not problematic if G_0 is small with respect to the sample size N. As a final note, under fixed-T asymptotics the GFE estimator provides valid inference on pseudo true value of the parameters only and a suitable fixed-T estimator of its variance is needed. In Bonhomme and Manresa (2015b) a formula for an analytical variance is provided together with an alternative bootstrap-based estimator, obtained resampling unit-specific blocks of information (x_i, y_i). Under large N, T asymptotics a simple estimator for the variance matrix, which hinges on deviations from the group-specific means, is discussed in Bonhomme and Manresa (2015b).

2.2 Extensions to BM15

BM15 contribution has generated a rich literature that attempts to address some of its limitations. In the following Sections we investigate how the literature has specifically studied grouped fixed effects-related topics: how to select the number of groups, how to ease the computational burden of the clustering procedure, how to extend the original BM15 framework to nonlinear models and other miscellaneous topics.

2.2.1 Number of groups, cluster memberships and heteroskedasticity

As described above, the identification of the right number of groups is crucial for a correctly specified model. The BIC criterion introduced in BM15 (see Equation (6)) presents good to perfect performances in simulation, and it is known not to overidentify the number of groups when N, T grow to infinity, since the real number of latent groups is fixed and does not vary in N, T. Models featuring grouped heterogeneity in slopes make extensive use of Bayesian Information Criterion for choosing the number of clusters (see below in Section 3). At the same time, in empirical economic applications it is not always easy to reconcile an economic meaningful number of clusters with the outcome of the application of the BIC. As an example, consider the empirical application on the relationship between income and democracy originally studied by Acemoglu et al. (2008) and revisited by BM15[3]: following the BIC guidance would lead to an unclear choice of the number of clusters and the Authors suggest the criterion of parameter stability - evaluate the stability of the parameters of interest by varying the number of groups - in order to find 4 groups, the choice also driven by economic intuition.

Another key issue in BM15 is the problem of misclassification. Under their asymptotic framework the GFE estimator attains the perfect classification rate in the limit and, from a small sample perspective, it is proven to be reliable in simulations. However, the literature has further analyzed the cluster membership topic.

The theoretical contribution of Dzemski and Okui (2021) investigates the convergence rate of grouped fixed effects estimator in a linear model without covariates in presence of individual variances and potential misclassified units in the limit. The Authors managed to prove uniform consistency of their simplified GFE in presence of a misclassification rate, assuming that it vanishes at a sufficiently fast rate[4]. The possibility of misclassification in their framework is due

[3] See Section 4 for details.

[4] See Condition 4 in the paper.

to an individual error variance large with respect to the sample size. The main limit of this result is that - up to now - applies to a regressors-free model only.

Loyo and Boot (2024) introduces a novel approach for estimation of BM15 GFE when group-specific variances entail information: necessary condition in the original GFE is that cluster specific means are well-separated in order to consistently identify non-overlapping groups. Loyo and Boot (2024) investigates the case for well-separated group variances when cluster means are (possibly) not: hence their framework features a cluster-level heteroskedasticity. By virtue of the novel estimation of cluster-specific variances, the Authors introduce a slightly different objective function - called Grouped Square Root (GSR) - robust to singularities that may arise in presence of ill-posed clustering outcomes, such as one individual in one group, leading to meaningless and problematic cluster-variance estimation. Of course, with respect to BM15, further distributional assumptions are required, since the GSR is basically a finite mixture model approach with cluster heteroskedasticity. The number of groups is identified using a BIC-style information criterion.

Furthermore, under suitable assumptions, their GSR-GFE estimator is consistent and asymptotically normal: in this vein, the Loyo and Boot (2024)'s contribution widens the range of application of the GFE with minimal further assumptions on group specific variances.

2.2.2 Computational issues

Classification Likelihood (CL) is the clustering algorithm initially proposed in BM15, but it is hard to apply it in real world datasets due to the dimensions of the grouping problem and the number of iterations required for a solution. In this spirit, *k-means* stands as an optimal alternative to CL and has become the standard tool in GFE applications, even though it may fall into local minima and requires the econometrician to evaluate different starting values.

Some contributions in literature have developed different clustering procedures w.r.t *k-means* in order to ease the GFE estimation. Chetverikov and Manresa (2022) introduce a grouped fixed effects estimator for linear models which has asymptotic properties comparable to BM15 while being computationally straightforward to compute. The spectral GFE estimation is a three-step process: the first step gives a preliminary estimate of the parameter(s) of interest, the second step concerns the classification of individuals in groups, and the third step is a simple OLS estimation with cluster-specific dummies. The clustering step is *k-means*-free and closely resembles a splitting algorithm; since the preliminary estimation step basically contemplates the computation of the eigenvalues of a potentially large matrix and the third step is embodied anyway in the standard BM15 procedure, this novel spectral approach is computationally simple. The spectral GFE is $\sqrt{NT}$-consistent, asymptotically normal and unbiased under the same set of BM15 assumptions plus a relevant new one concerning the structure imposed on covariates, which are supposed to exhibit a latent factor structure. Finally, Chetverikov and Manresa (2022) suggest the use of information criteria in order to choose the right number of groups.

Mugnier (2022) introduces a three-step procedure to estimate BM15 baseline linear model (5) which is computationally trivial and results in an estimator with desirable asymptotic properties. It relies on nuclear norm regularization (Moon and Weidner, 2015) as a first step for a preliminary estimate of parameters of interest. Then a simple iterative clustering algorithm (*k-means*-free) based upon the distance between residuals obtained in the first step carries out the estimated number of groups. The third step is an OLS estimation with interaction between cluster specific and time dummies. Though fast, the procedure requires the specification of two hyperparameters, one for the nuclear norm problem and one for the clustering process: they can be both selected by cross validation.

Lewis et al. (2023) suggest employing a novel fuzzy clustering algorithm to estimate parameters in model (5). Their approach manages to exactly replicate results in BM15, while being computationally way less intensive. The main idea behind the procedure is that the BM15 problem could be rewritten as a GMM problem, for which maximization procedure are easily available. It can be viewed as a procedure for approximating the more computationally involved BM15 estimator. The fuzzy clustering objective function depends on a parameter m which strictly controls how well we are approximating the GFE, the approximation improving as $m \to 1^+$. Lewis et al. (2023) show in simulation that their approach leads to an exact replication of BM15 for values of m close to the limit. Moreover, the fuzzy approach exhibits a substantial computational speed improvement w.r.t. standard BM15 algorithm when big datasets are involved, at least in simulations.

2.2.3 Nonlinearities

BM15 contemplate a linear model. In order to overcome this limit, econometricians have proposed extensions of the GFE idea to nonlinear models. Under the assumptions that groups are known, Bester and Hansen (2016) study a nonlinear model with clustered structure for discrete UH; the group memberships are supposed to be known due to some external classification, such as information entailed in quantiles. In this vein, the Bester and Hansen (2016) framework is different from BM15's one due to the uncertainty on group memberships.

Mugnier (2023) studies the model:

$$P(Y_{it} = y_{it} | X_{i1}, \dots, X_{it}, g_i^0, \alpha^0_{g_i^0 t}) = h(y_{it}, X_{it}\beta + \alpha^0_{g_i^0 t}) \tag{7}$$

where the link function $h(\cdot)$ is unknown. The model (7) is the nonlinear counterpart of the original BM15 model. As in BM15 the number of groups is unknown and the true underlying UH is discrete. In order to carry out asymptotic analysis Mugnier (2023) further assumes that both the number of groups and the functional form of $h(\cdot)$ are known. While the former assumption seems reasonable and is usually driven by the economic nature of the data (e.g., Poisson distribution for count data), it is not always easy to understand how the number of groups should be chosen according to Mugnier (2023), and, beyond heuristic arguments such as performing a grid search on G, the issue is left open.

In order to find the nonlinear GFE estimator (NGFE), Mugnier (2023) generalizes the BM15 algorithm to nonlinear models. However, finding global minima for the NGFE problem may be hard since different starting values for the initialization of parameters are required. The resultant NGFE estimator is consistent and, under large N, T asymptotics, $N/T \to \infty$, $N/T^b \to 0$ with $b > 1$, exhibits an asymptotic normal distribution. Interestingly, NGFE is free of the incidental parameter problem due to the noiseless estimation of the group specific fixed effects.

Ando and Bai (2023) - the nonlinear extension of Ando and Bai (2016)[5], study a slightly different family of nonlinear models with latent structure for heterogeneity. In their framework,

[5] Ando and Bai (2016) - studies a linear model where UH is supposed to follow a latent grouped factor structure. Alike Ando and Bai (2023), the estimation is carried out optimizing a SCAD penalized likelihood function.

which contemplates a generic exponential family of distributions nesting most used nonlinear models, the UH is supposed to hold a latent factor structure (close to that in Bai (2009)) where factors own in turn a clustered pattern themselves. As an example, Ando and Bai (2023)'s approach allows the econometrician to estimate a standard binary choice model like:

$$f(y_{it}|x_{it}, \beta_{g_i}, f_{t,g_i}, \lambda_{i,g_i}) = \Phi(x_{it}\beta_{g_i} + f'_{t,g_i}\lambda_{i,g_i})^{y_{it}}(1 - \Phi(x_{it}\beta_{g_i} + f'_{t,g_i}\lambda_{i,g_i})^{1-y_{it}} \tag{8}$$

where f_{t,g_i} is an $r_{g_i} \times 1$ vector of group-specific unobservable factors and λ_{i,g_i} represents the factor loadings. β_{g_i} refer to group specific heterogeneous slopes[6]

Estimation of model (8) may be challenging because the issues usually involved in estimation of factor models for nonlinear models add to the GFE's peculiar ones. The proposed algorithm is *k-means*-based for the clustering step and relies on usual principal component analysis for extraction of factors. Both the number of groups and factors are estimated using a tailored BIC.

2.2.4 Other extensions

This Section considers two relevant departures from BM15 original framework: the possibility to exploit latent clustered structure in a non-parametric context (Vogt and Linton, 2017) and GFE applied to quantile regression models (Gu and Volgushev, 2019).

Vogt and Linton (2017) study a simple model like:

$$Y_{it} = m_i(X_{it}) + u_{it} \tag{9}$$

where $m_i(\cdot)$ are individual unknown functions to be estimated nonparametrically. m_i functions posses an underlying grouped structure and individuals belonging to the same group share the same regression curve. The number of groups nor the memberships are known *a priori* and need to be estimated. The error term in model (9) is supposed to have an additive two-way characterization.

Estimation of regression curves in model (9) employs Nadaraya-Watson smoother, while the number of groups is chosen according to a threshold procedure, which could be further refined applying *k-means* grouping on preliminary estimates of group memberships given by the threshold-based procedure. As it is well known in nonparametric literature, a proper choice for the bandwidth when employing kernel estimators is required, even though, in a following contribution Vogt and Linton (2020) develop an estimation method for model (9) which is bandwidth-free.

Gu and Volgushev (2019) discuss the introduction of grouped fixed effects in quantile regression models. They study the conditional τ-quantile function:

$$q_{i,\tau} = x_{it}\beta_0^\tau + \alpha_{0i}(\tau) \tag{10}$$

where $\alpha_{0i}(\tau)$ hides a cluster structure and can take $i = 1, \ldots, G$ values. G is unknown. Number of groups and group specific parameters are estimated jointly by means of an algorithm which,

[6]A complete discussion of models with clustered heterogeneity in slopes is carried out in Section 3. The Ando and Bai (2023) contribution is reviewed here for the main innovation in fixed effects structure more than in heterogeneity slopes.

after taking individual fixed effects as preliminary estimates, compute GFE minimizing a suitable objective function and select the lowest BIC as a function of different G's.

The resultant GFE quantile estimator, under usual assumptions made in the quantile regression literature, holds the oracle property[7] and is consistent and asymptotically normal. Moreover, the proposed BIC selects the right number of groups with probability approaching one.

3 The pursuit for heterogeneity in slope coefficients

The review has hitherto considered clustered heterogeneity in fixed effects only. Another popular option is to consider heterogeneity at the level of the regression parameters, allowing for the presence of different slope parameters, one for each latent group in the dataset. This approach is not at odds with the standard FE way of modelling unobservable traits, as long as individual dummies can enter the model specification with no particular concerning. The seminal contribution in this stream of literature is that of Su et al. (2016a), in which the novel classification Lasso approach allows one to jointly estimate heterogeneous slope parameters and cluster memberships, supposed to be unknown. However, the idea of a latent grouped structure in slopes parameter has been studied in relevant papers before Su et al. (2016a).

Sun (2005) studies a method to detect grouped structure in the slope coefficients of a linear panel data model. Here the (unknown) group memberships are retrieved through a multinomial logistic regression. This approach could be seen as if the data were generated by a mixture of distributions, each possessing its own distribution $f_g(x_{it}, y_{it}, \beta_g)$ and is very close to a finite mixture model one (MacLahlan and Peel, 2000): this is why Sun (2005) estimates the model parameters with an Expectation Maximization algorithm.

Lin and Ng (2012) consider a simple model like:

$$y_{it} = x_{it}\beta_g + u_{it} \tag{11}$$

in which an unknown number of groups G is supposed to exist, and a consequent number of G slope coefficients β_g needs to be estimated. Controlling for time-invariant heterogeneity in u_{it} with standard OW-FE is also possible in their framework. The Authors proposed two different clustering procedures, a pseudo-threshold method and a *k-means* based method. The *k-means* method is quite peculiar in the homogeneous slopes framework because since the Su et al. (2016a) contribution, the C-Lasso technique and its improvement have become the main estimation tool. The pseudo-threshold method supposes the existence of unobservable threshold(s) completely determining the group membership(s). The conditional *k-means* approach is based on an iterative *k-means* procedure: after an initial random assignment of all individuals to G groups, the residual sum of squares (RSS) is computed for each group and the algorithm stops when it is not possible to minimize the SSR by moving individuals between groups. When N is large, this operation could be computationally expensive. In addition, the number of groups requires to be specified by the analyst. Lin and Ng (2012) propose two methods, a Bayesian Information Criterion and a sequential application of the Pesaran and Yamagata (2008) test for

[7]The obtained estimates by GFE are the same that one would have by knowing the classification of each individual *a priori*.

parameter homogeneity: the latter works by sequentially splitting the dataset and consequently doubling the number of groups until the null hypothesis of parameter homogeneity cannot be rejected for the subsamples. Both are proven to be effective in simulations.

Sarafidis and Weber (2015) study a linear model like

$$y_{itg} = x_{itg}\beta_g + u_{itg} \tag{12}$$

with heterogeneous slopes in a fixed T, large N framework. Unlike model (11), the heterogeneity involves the whole structure of the model and the focal point is on correctly estimating the number of groups and membership and then running G different regressions. Cluster memberships and number of clusters are unknown. The Authors propose a clustering algorithm similar to the one in Lin and Ng (2012), basically a *k-means* based procedure, but here the focus is on minimizing the following information criterion:

$$N\log\left(\frac{RSS}{NT}\right) + f(G)\theta_N$$

where RSS is the residual sum of squares in each sub-G regression, $f(G)$ is an unspecified function increasing in G and θ_N is a sequence of constants chosen to satisfy $lim_{N\to\infty}\frac{\theta_N}{N} = 0$. In practice $f(G) = G$ and $\theta_N = \sqrt{N}$. The procedure should be repeated for a different number of clusters in order to choose the cardinality of G: this is possible because the information criterion does not overfit or underfit, unlike standard BIC or AIC.

Su et al. (2016a) consider the estimation of panel data models, linear and nonlinear, in presence of heterogeneous slopes and individual fixed effects (α_i) via a classifier-Lasso (C-Lasso) approach. The reason underlying the employ of a penalized approach is the following: the Lasso can deal with sparsity of parameters and usually in empirical applications it does make sense to assume that β_i slope parameters, for $i = 1, \ldots, N$, is a form of overparameterization whenever G with $G < N$ coefficients are enough: from this standpoint the Lasso could introduce a useful shrinkage. A relevant aspect here is that, unlike BM15, the number of groups G does not increase with the sample size. They consider the following optimization problem for model 11 (Penalized Profile Likelihood - PPL - approach) :

$$Q(\beta_i, \mu_g) = Q_P(\beta_i) + \frac{\lambda}{N}\sum_i^N \prod_g^G ||\beta_i - \mu_g|| \tag{13}$$

where

$$Q_P(\beta_i) = 1/NT\sum_i \sum_t \psi(y_{it}, x_{it}; \beta_i, \alpha(\beta_i))$$

is the profile log-likelihood (Hahn and Newey, 2004), $\psi(\cdot)$ is the density function of $y_{it}|x_{it}, \alpha_i, \beta_i$, λ is a tuning parameter and μ_g is the vector of shrunk (clustered) slope parameters, meaning that $\beta_i = \sum_g \mu_g \mathbf{I}\{i \in G_g\}$. Alike to a standard Lasso estimator, that shrinks some parameters to zero, the novel C-Lasso estimator shrinks some of the β_i parameters to the μ_g set. Under large N, T asymptotics, the obtained estimator has desirable properties under mild assumptions[8].

[8] In particular the Penalized Profile Likelihood C-Lasso estimator is consistent and asymptotically normal under the standard assumptions for dynamic nonlinear models (Hahn and Kuersteiner, 2011). It also has the property of Classification Consistency, meaning that each individual is correctly classified in

Equation (13) depicts a non-convex optimization problem that requires different starting points and an iterative algorithm for the solution. However, Su et al. (2016b) discuss how the proposed algorithm is robust to perturbation of initial values. The choice of the proper number of groups is made by minimizing a suitable Information Criterion.

The Penalized Profile Likelihood approach could be extended to the slightly different Penalized GMM approach - also introduced in the paper - when the object of interest is a dynamic model or a model with endogeneity.[9] Another interesting feature of C-Lasso approach in non-linear models is that the Dhaene and Jochmans (2015) split panel jackknife bias correction is feasible and proven effective in simulations.

The framework depicted in Su et al. (2016a) has attracted a wide popularity, and it has been extended in a number of contributions. One of the main problems in the latent grouped structure is the choice of the number of clusters. Lu and Su (2017) propose a LM type test for the choice of the right number of groups in the Su et al. (2016a) C-Lasso procedure. It crucially hinges on the estimation of the residuals using the post-C-Lasso estimator. The testing procedure works by sequentially increasing the number of clusters until the null hypothesis $H_0(G) : G_0 = \hat{G}$ is not rejected. So $H_0(1)$ is a test for homogeneity in slopes coefficients, $H_0(2)$ is a test for the presence of two groups and so on. The test outperforms the BIC introduced by Su et al. (2016a) in simulations.

Su and Ju (2018) extend the C-Lasso allowing for interacted fixed effects in a linear framework: this challenge is not trivial because in addition to the number of latent clusters also the number of factors is unknown. They consider the model:

$$y_{it} = x_{it}\beta_g + \lambda_i' f_t + \epsilon_{it} \tag{14}$$

where the factor component does not have a clustered structure. Estimation of parameters in Equation (14) is carried out in two steps, the first one is to obtain estimates of β_g using the C-Lasso methods explained above and then estimates of the factors and factors loading are computed using the Bai and Ng (2002)'s Principal Components method. The procedure is called Penalized Principal Component (PPC) estimation, and it is affine - from a conceptual and algorithmic standpoint - to the Penalized Profile Likelihood. The number of factors is chosen to minimize an Information Criterion as well as the number of groups, the former being strictly computed before the latter. The estimators obtained with the Penalized Principal Component procedure are consistent, asymptotically normal and enjoy the oracle classification property under mild assumptions. However, it is worth commenting on the effects of an over/under estimation of either the number of groups or the number of factors $\hat{R}$. If $\hat{R}$ is underspecified but $\hat{G}$ is not, the slopes coefficients are poorly estimated; if $\hat{R}$ is right but $\hat{G}$ is overspecified, coefficients are not generally biased, but the root mean square error tends to increase; if $\hat{R}$ is overspecified but $\hat{G}$ is right there are no concerns. To sum up, the same conclusions on the effects of the wrong choice on the number of groups formulated by BM15 apply: the overidentification of $\hat{G}$ leads to inefficiency while the underidentification of $\hat{G}$ leads to inconsistency.

the limit for N, T and the Oracle Property, meaning that the obtained estimates are the same the one would have by knowing the classification of each individual *a priori*.

[9]The main difference between the two approaches is that the Penalized Profile Likelihood approach has the oracle property in all cases while the PGMM procedure does not generally hold this feature.

Wang et al. (2018) study a model like (11) with additive individual fixed effects α_i and time effects λ_t. Their contribution is chiefly algorithmic since they introduce a novel approach called Panel - CARDS (clustering algorithm in regression via data-driven segmentation), basically a penalized likelihood approach in the spirit of Su et al. (2016a) with a different penalization based on the within and between group difference of slope coefficients. The number of groups is indirectly[10] chosen via an Information Criterion.

One interesting feature of the Wang et al. (2018) method is that works, in a more general large N, T framework, under the assumption that G_0 is not fixed and can diverge, unlike both BM15 and Su et al. (2016a), which strictly assume fixed G_0. Furthermore, under their assumptions the slopes parameters are consistent, asymptotically normal and hold the oracle property.

Liu et al. (2020) address the estimation of both model (11) and its non-linear counterpart in a context of penalty-free estimation, hinging instead on M-estimator models theory. Consequently, the Authors do not introduce a C-Lasso procedure but employ an iterative *k-means* method as in BM15.

Their approach has two main advantages: i) since it is based on standard M-estimation procedure, it is a tuning parameter-free approach and this avoids possible ambiguity in the choice, for instance, of the λ parameter in Su et al. (2016a) C-lasso ii) the consistency of the proposed estimator is guaranteed when the number of estimated clusters is either over- or exactly identified.

The proposed clustering algorithm is a modified version of the *k-means* based one in BM15 where clustered slopes instead of grouped fixed effects are estimated. After an initial random assignment of individuals to G groups and an initial estimation of $\beta_i, i = 1 \dots G$ is performed, the algorithm keeps moving individuals between groups and reassigning memberships until the likelihood is maximized. Due to the problem of initial starting values, the procedure should be repeated for different starting guesses. In order to find the right number of groups, the following information criterion is proposed:

$$PC(G) = \hat{\Psi}(\hat{\phi}) - \eta G \tag{15}$$

$$\hat{G} = \text{argmax}_{G \in G_{max}} PC(G) \tag{16}$$

where $\hat{\Psi}$ is the likelihood function, $\hat{\phi}$ is the vector of all nuisance and slopes parameters, $\eta > 0$ is a tuning parameter and G_{max} is an upper bound. The upper bound is required because, although the method is robust to overidentification of the number of groups, a more parsimonious estimation is usually more desirable, even though from a practical point of view the choice of a relatively large number of groups is implicitly suggested.

Under large N, T, rectangular array asymptotics and fixed G, the Liu et al. (2020) estimator is consistent, asymptotically normal and achieves the perfect classification in the limit.

One limit of the Su et al. (2016a) contribution is that it does not allow for time varying group heterogeneity in slopes, although Su and Ju (2018) discuss the introduction of interactive fixed effects - a flexible way to control for time variation in unobservables - in the C-Lasso approach.

[10]The Information criterion described in Wang et al. (2018) directly suggests the value of a tuning parameter that influences the number of groups. In practice the CARDS method estimates a number of β_i depending on an hyperparameter, and individuals with the same β_i are then classified in the same cluster. In this vein, it works the opposite way with respect to BM15 where the number of groups directly minimizes the BIC.

Okui and Wang (2021) fill the gap in the literature by putting forward an estimation method for time varying clustered slopes in presence of structural breaks. They study a model like

$$y_{it} = x_{it}\beta_{g,t} + u_{it} \tag{17}$$

where slopes are group-specific and change over time for groups: the time variability is modeled with structural breaks, meaning that the slopes change after each break for each group. Each cluster has a number m_g of breaks, but group's membership does not change over time nor the number of groups; the number of breaks increases with T.

Along with group memberships and heterogeneous slopes, the number and date of structural breaks need to be estimated. The Authors propose an algorithm which is a blend of BM15's one adapted for heterogeneous slopes in one step and a penalized likelihood approach in another step. In particular, in the first step, after a random assignment of individuals in groups, maximization of a penalized likelihood is performed in order to get preliminary estimates of $\beta_{g,t}$ and time breaks. Then the group membership is updated by minimizing a least squares problem. The procedure is iterated up to numerical convergence.

Together with the selection of two hyperparameters for the penalized problem, a proper number of groups is required. The choice is performed with a Bayesian Information Criterion that closely resembles the one in BM15, extended to the presence of structural breaks

$$BIC(G) = \frac{1}{NT}\sum_{j}^{m-1}\sum_{T_j-1}^{T_j}\sum_{i}^{N}(y_{it} - x_{it}\hat{\beta}_{g_i,j})^2 + \hat{\sigma}^2\frac{n_p(G)+N}{NT}\ln NT$$

where $n_p(G)$ is the total number of parameters estimated for a given G and $\hat{\sigma}^2$ is the estimated variance of the idiosyncratic error term. One caveat in the use of this tool lies in the fact that the number of parameters does not directly increase with G since larger G's could lead to fewer structural breaks detected: despite this theoretical pitfall, the Bayesian Information Criterion is proven to identify the right number of groups in simulations.

Under a mild set of assumptions, which closely resembles the BM15 asymptotic framework, Okui and Wang (2021) show that their estimator is consistent, asymptotically normal distributed and that both dates and number of the structural breaks are consistently estimated in presence of an unknown group structure.

A powerful extension to Okui and Wang (2021) is the Lumsdaine et al. (2023)'s proposal in which the authors study a linear model with heterogeneous slopes in presence of a single[11] structural break, common for all groups, that changes the group memberships and possibly the number of groups. In this vein, the structural break changes the behavior of both slopes coefficients and latent patterns. In Lumsdaine et al. (2023) framework the data of the structural break, the number of groups and the membership before and after the break are unknown and are jointly estimated.

The parameters' estimation is carried out by minimizing the following least squares problem for model 17:

$$(\hat{s}, \hat{\beta}_g, \hat{g}_i) = \text{argmin}_{s,\beta,g_i}\left[\sum_{t}^{s-1}\sum_{i}^{N}(y_{it} - x_{it}\beta_{gi,A})^2 + \sum_{s}^{T}\sum_{i}^{N}(y_{it} - x_{it}\beta_{g_{i,B}})^2\right]$$

[11]Even though the case for multiple structural breaks is discussed as a possible extension.

where $g_{i,A}$ refers to group membership before the break and $g_{i,B}$ refers to memberships after the break and s is the time when the break occurs. Since the complete research in the parameter space is unfeasible, Lumsdaine et al. (2023) offer a multistep algorithm for the purpose. The procedure is essentially a *k-means* based algorithm, so no penalized likelihood is involved.

The choice of the number of them is performed with an Information Criterion: the main challenge here is that the number of groups may change before and after the break. The Information Criterion writes:

$$IC = \log \hat{Q} + n_p 3 \frac{\ln(NT)}{NT}$$

where $\hat{Q} = 1/NT \left[\sum_t^{s-1} \sum_i^N (y_{it} - x_{it}\beta_{gi,A})^2 + \sum_s^T \sum_i^N (y_{it} - x_{it}\beta_{g_{i,B}})^2 \right]$ and n_p is the total number of parameters.

The estimator is consistent and achieves perfect classification of group membership and the date of the structural break for large N, T asymptotics and NT^{-d} with $d > 0$.

In the context of slope heterogeneity, the contribution of Miao et al. (2020) studies a panel threshold model with unknown cluster structure in both slopes and thresholds. More precisely, the model of interest is:

$$y_{it} = x_{it}\beta_g + x_{it}\delta_g d_{it}(\gamma_g) + \alpha_i + \epsilon_{it}$$

where $d_{it}(\gamma_g) = \mathbf{I}(q_{it} \leq \gamma)$, and q_{it} is a scalar threshold variable. Cluster heterogeneity is involved here for both slopes and threshold coefficients and individuals classified in the same group share the same coefficients profile. The number of groups and relative membership is unknown and is estimated along with the slope and the threshold parameters. The BM15-like three-step iterative algorithm proposed by the Authors jointly estimates the parameters of interest. Finally, Miao et al. (2020) developed a sequential testing procedure articulated in three Likelihood Ratio tests in order to carry inference on threshold effects: the analyst is allowed to test for specific values of γ_g, for the presence of common effects and if the hypothesis is not rejected, to test for specific values of the common effect. Testing for the absence of threshold effects at all could be achieved with a sup-Wald test statistic. The number of groups is selected by means of a Bayesian Information Criterion.

4 Empirical applications

This Section discusses relevant applied contributions that use models for latent structures in unobservables or in slopes: in particular, it focuses on studies that examine health and environmental issues, the labour market and the presence of sorting effects in it. These fields are analyzed since they are prominent objects of studies in microeconometrics. Finally, the BM15 application to the relationship between income and democracy (originally in Acemoglu et al. (2008)) is analyzed together with all the subsequent new interpretations that have appeared in the literature. Applications are selected among the most impactful in the literature.

4.1 Health and environment

Health Oberlander et al. (2017) examines the relationship between quality of diet - measured by food production data at the country level - and social globalisation together with trade openness, using a GFE approach (BM15 version). The idea behind this specification is that country-specific unobservable characteristics, such as dietary preferences, may exhibit grouped patterns. A BIC criterion is used to select the correct number of groups: this results in $G = 6$ clusters (for 70 countries). It is worth noticing that dividing countries into $G = 4$ groups according to the World Bank's income classification and estimating the model with these *a priori* cluster fixed effects does not minimize the objective function with respect to the GFE approach. Western and richer countries tend to belong to the same clusters, while South American, Asian and African countries are clustered in specific groups.

Moreover, a graphical inspection of the grouped specific time trends and the supply of animal protein (a crucial measure of dietary health in the literature) highlights the relevance of estimating grouped fixed effects, as the clusters show different patterns over time. The Oberlander et al. (2017)'s results point in the direction that globalisation is associated with a positive and significant effect on the supply of animal protein and sugar, as well as on the average body mass index.

Janys and Siflinger (2024) investigate the effects of abortion on mental health using Swedish data in a grouped fixed effects framework. They consider a linear model where the dependent variable equals one if the woman has experienced mental health issues at time t: the object of interest of the analysis is the parameter associated to the abortion event at time t. The GFE approach allows Janys and Siflinger (2024) to estimate group-specific trajectories of affected women over time. The number of groups is chosen according to the BIC criterion of BM15, resulting in $G = 2$ groups, which in turn represent high- and low- risk women, the latter being the largest part of the sample. Interestingly, the GFE approach does not find a significant effect of abortion on mental health, a result that is absent from literature, which makes extensive use of standard FE methods. The main conclusion is that, once proper controlling for latent risky behavior with GFE, the positive association between mental status and abortion found in the literature disappears.

Environment Grunewald et al. (2017) contribute to the literature on the ambiguous effect of income inequality on per capita emissions: economic growth is strongly associated with increase in polluting emissions, even though the phenomenon is not homogeneous across countries. The Authors suggest the use of the BM15's GFE to take into account historical divergences across countries in terms of polluting behaviours. Moreover, inequality measures may exhibit small time variation, causing poor performance of FEs estimators: GFE is indeed less affected by this drawback. Grunewald et al. (2017) find that per capita income and inequality, proxied by Gini index, have a strong positive effect on CO2 emissions in both TW-FE standard and GFE specifications, even though the latter approach leads to more precise estimates for country specific emission-income elasticities for relevant countries such as the U.S. and India. Five clusters are estimated using an information criterion, the main driver of the classification, according to the authors, being the level of energy intensity across countries. Furthermore, the Environmental Kuznets Curve hypothesis holds for both FE and GFE models.

Johar et al. (2022) studies the application of GFE to model resilience, defined in their framework as ability to withstand financial hardship caused by natural disasters. They estimate a linear dynamic model using financial hardship[12] as dependent variable, which accounts for grouped heterogeneity: individuals are classified into $G = 3$ groups according to the BM15 information criterion. Johar et al. contemplate both time invariant GFE (meaning that, unlike

[12]For all details on how financial hardship is measured see Johar et al. (2022).

BM15, no interaction between time trajectories and clusters enters the model) and interaction between cluster dummies and lagged financial hardship in the specification: this choice allows them to interpret the latter as the measure of resilience and to study the different responses of different individuals to disastrous events. The three clusters find a natural interpretation as differently vulnerable parts of society.

4.2 Labour market and sorting effects

The seminal contribution of Abowd et al. (1999) investigates a novel approach for estimating sorting effects in the labour market: the Authors are interested in whether better workers land in high-paying firms. In order to evaluate this hypothesis, Abowd et al. (1999) study a TW-FE based approach for estimating the correlation between firm- and worker-specific fixed effects in a standard wage equation. The Abowd et al. (1999)'s idea could actually be seen as a way to decompose the variability of earnings into firm and worker specific variability plus the sorting effect, which is real when the firm- and worker-specific fixed effects' correlation is positive and indeed, using a large French dataset, Abowd et al. (1999) found the presence of assortative matching. However, this estimation procedure has been recognized as biased and bias corrections have been proposed by Andrews et al. (2008) and Kline et al. (2020).

Bonhomme et al. (2019)'s contribution (BLM19 henceforth) is tightly related to Abowd et al. (1999)'s one, since it addresses the estimation of sorting effects in the labour market. However, BLM19 assume discrete firm- and worker-specific latent types (so-called two-sided heterogeneity) and propose to apply *k-means* clustering to group firms according to the distribution of wages within each firm: this feature introduces an unrestricted relationship between workers and firms latent traits and wages, instead of the Abowd et al. (1999)'s linear one. The consequent dimensionality reduction could be deciding since the datasets usually employed in this literature include thousands of different firms (Bonhomme et al., 2023) and standard FE's estimation may be cumbersome. In a second step, parameters related to worker's mobility and wage are estimated with a finite mixture model[13], assuming that workers belong to a finite number of classes. Moreover, BLM19 approach allows one to estimate a model that accounts for worker-firm complementarities, that is, a model with interaction between firm and worker heterogeneity. The BLM19 framework is also suitable for dynamic models.

The model is then evaluated using labour market data from Sweden. BLM19 identify 6 latent types for workers and 10 clusters for firms. A strong sorting effect, measured as the correlation between worker and firm heterogeneity, can be identified while BLM19 find little evidence of complementarities in the Swedish labour market. However, as it is well known in the literature (Bonhomme et al., 2023), studies on matched employer-employee datasets have not reached a single conclusion.

Finally, Lentz et al. (2018) extend the BLM19 model - keeping latent discrete types and a finite mixture model approach - by introducing a more precise estimation technique for the job transition probability parameters. The Authors employ a modified Classification Likelihood algorithm for clustering firms instead of the *k-means* routine. The empirical analysis is performed on Danish labour market data and reveals strong sorting effects between workers and firms latent

[13]Under additional distributional assumptions, the BM15 GFE could be seen as a finite mixture model approach (Bonhomme and Manresa, 2015b).

types. With respect to BLM19, the number of types has increased, as the Authors estimate a mixture model with 14 worker types and 24 firm types.

4.3 Income and democracy

The original paper by Acemoglu et al. (2008) investigates the relationship between democracy and income with a simple linear model. Considering 150 countries every 5 years over the period 1960-2000. Acemoglu et al. find that the strong and positive relationship between per capita GDP and the democracy indicator disappears when controlling for additively specified country and year fixed effects. This evidence suggests that keeping in account country-specific unobservable historical paths over time may lead to a more precise estimation.

The BM15's GFE, allowing one to control for hidden grouped trajectories over time, is a suitable tool to refine estimation in the Acemoglu et al. (2008) model. They consider a subset composed by $N = 90$ countries observed from 1970-2000 with a 5-years frequency ($T = 7$). Countries are classified into $G = 4$ groups as parameters of interest are shown to be stable to variation in the number of groups. The four clusters have full economic interpretability and are identified as: i) high-democracy countries ii) low-democracy countries iii) early transition to democracy countries iv) late transition to democracy countries. It is worth noting that in a model specification with both time-varying GFE and country specific FE's the income effect on democracy is not statistically different from zero, consistent with Acemoglu et al. (2008)'s findings.

The application on income and democracy has been reinvestigated with models assuming latent clustered heterogeneity in unobservables (Loyo and Boot, 2024) or in slopes (Lu and Su, 2017; Wang et al., 2018; Okui and Wang, 2021).

Once allowing for heterogeneous slopes, Lu and Su (2017)'s testing procedure strongly rejects the hypothesis of homogeneous income-related parameters and finds $G = 3$ groups. The estimated effects differ among clusters, with one group exhibiting a significant and positive relationship between income and democracy, while the residual ones show the opposite effect. However, since the clustering procedure is entirely data driven, Lu and Su (2017) perform further analyses in order to understand the drivers of the memberships' allocation. The Authors estimate a multinomial logit model of cluster membership, finding that date of independence and long-run economic growth are the key drivers of the classification.

Both Wang et al. (2018) and Okui and Wang (2021) confirm the conjecture that heterogeneous slopes lead to a better identification of income effects. More specifically, the Panel-CARDS procedure of Wang et al. (2018) finds 4 groups with different effects and magnitudes among them, while the Okui and Wang (2021) approach, allowing for structural breaks, again suggests 4 groups. Three of these have one structural break (at group specific points in time), while the last one does not exhibit any break and includes the countries classified by BM15 in the "high democracy" group. Okui and Wang (2021) estimate different income effects within clusters, too.

Finally, Loyo and Boot (2024) apply the GFE-GSR method and find different results with respect to BM15: although the Authors gather countries in 4 groups, they suggest that, by exploiting information in cluster specific variances, the relationship between income and democracy is positive and significant.

5 Discretizing unobserved heterogeneity

Bonhomme et al. (2022) study grouped fixed effects in a framework where unobservable traits have a latent grouped structure, whilst heterogeneity may have a nondiscrete, potentially very complex and unspecified, yet subject to mild assumption, functional form.

The model in exam is:

$$y_{it}|x_{it}, \beta_0, \alpha_{it0} \sim f(y_{it}|x'_{it}\beta_0 + \alpha_{it0}). \tag{18}$$

where α_{it0} parameterizes an unspecified shape for unobservable traits - subject to assumptions described below - and $f(\cdot)$ is a generic functional form. The BLM22 approach works for both linear and nonlinear models, so $f(\cdot)$ may be the identity function or the CDFs which gives raise to popular nonlinear models such as standard normal/logistic distribution.

According to the structure of heterogeneity, BLM22 has derived two different estimators, one slightly more involved than the other: the OW-GFE estimator, which can not handle time-varying heterogeneity, hence allowing one to estimate models with unspecified time-invariant heterogeneity only - and the TW-GFE, which instead can take into account dynamic behaviour of unobservable traits.

BLM22 propose to use clustering of fixed effects as a way to *approximate* the heterogeneity, whose functional form is in practice unknown. The procedure works in two steps: in a first clustering step sample averages at individual and time-specific level are computed from data, the *k-means* algorithm is applied to these sample moments and a number K and L of groups for individual and time occasions are found. The idea underlying the first step is that sample moments contain information on unobservable traits that can be exploited by grouping individuals and time occasions. In the second estimation step the time and individual specific dummies are interacted and KL grouped fixed effects are estimated in a standard ML procedure. The aforementioned procedure is peculiar of the TW-GFE, while in the OW-GFE this routine is identical though simplified, since it does not feature the time dimension[14]. Moreover, the BLM22 GFE approach could be seen as a *regularization* procedure, since it intrinsically leads to a more parsimonious estimation as the number of UH parameters is strictly smaller with respect to standard ML-FE specification.

BLM22 differ with respect to BM15 since in the latter contribution the UH is supposed to be discrete and completely described by an unknown number G_0 of support points: here instead the clustering approach is used as a way to approximate the heterogeneity, the goodness of the approximation strictly related to the number of groups and to the information entailed in the sample moments.

Assumptions In order to employ the BLM22 GFE, among the others standard ones[15], two assumptions must be met. The first one concerns the structure of the UH, the second one the information entailed in the tool used for clustering, the sample moments.

The assumption on UH parameterizes α_{it0} as a smooth, Lipschitz-continuous function of

[14] In the OW-GFE procedure only individual specific moments are used for clustering and K fixed effects are estimated, where $K \ll N$.

[15] All assumptions are listed and discussed in details in Bonhomme et al. (2022) and are standard assumptions for a well-behaved likelihood problem. It is worth noticing that, in order to employ the TW-GFE only, an additional assumption on i.i.d data is required, meaning that TW-GFE can not be used when dynamic linear/nonlinear models are involved.

latent individual and time specific traits, $\xi_{i0} and \lambda_{t0}$. The dimension of the latent types (d_{ξ_i} and d_{λ_t}) is required to be small, usually equal to one. This assumption is of paramount importance, since the link between the unknown heterogeneity α_{it0} and the latent types ξ_{i0}, λ_{t0} is left unrestricted and may be potentially very complex.

The second assumption imposes that moments used in the clustering procedure should be informative on latent traits of individuals and time occasions. In the limit, the difference between moments and the unspecified UH vanishes at a parametric rate. Theoretically speaking, the BLM22 approach also works employing moments of different order than first: this choice, however, should be justified by the richer amount of information entailed in moments other that the first one[16].

Number of groups Since the *k-means* algorithm is an unsupervised clustering method, the number of groups to be found in the data should be provided by the econometrician. Of course, the number of groups is not known in practice and BLM22 introduced a rule in order to find the proper number of groups. The rule writes:

Proposition 1. *Number of groups:*
The number of groups K and L are chosen according to

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

where $Q(\cdot)$ is the objective function of the kmeans *problem, V_h and V_w are the variability of the individual moments h_i and time-specific moments w_j, respectively, and $\gamma \in (0, 1]$ is a user-specified hyperparameter;*

The rule states that the number of groups K and L is increased until the order of the objective function of the *k-means* problem and the variability of moments share the same order. Of course, when considering OW-GFE, only the $\hat{K} = \min_{K \geq 1}\{K : \hat{Q}(K) \leq \gamma \hat{V}_{h_i}\}$ part is taken into account.

The choice of the hyperparameter γ is deciding because it indirectly controls the number of groups, and it is discussed in Subsection 5.1 below. How many groups are considered is crucial in BLM22 framework for reasons different from those explained above for BM15's GFE: while in the latter setting the true number of groups is finite but unknown, and hence could be consistently estimated, in the BLM22 framework the groups serve as an approximation device and their number actually controls the goodness of approximation and consequently the severity of the incidental parameters bias. The role of K (and L) is clear when analyzing the asymptotic behavior of the GFE.

Asymptotics OW-GFE and TW-GFE present a slightly different asymptotic behavior and are analyzed separately. Consider the OW-GFE first. It is a generally biased estimator for both linear and nonlinear models and we can distinguish between three distinct sources of bias: an $O_p(\frac{1}{T})$ component refers to an incidental parameter bias that originates from the limited number of observations used in the first step of the GFE when computing individual moments,

[16]For instance, Loyo and Boot (2024) suggest the use of second moments if variances are more informative than sample averages in the clustering procedure, an idea that is closely related in the spirit to their extension of BM15 to variance-based GFE.

an incidental parameters bias of order $O_p(\frac{1}{T})$ is due to the noise in the estimation of grouped fixed effects[17] and an $O_p(K^{-\frac{2}{d_{\xi_i}}})$ term, with d_{ξ_i} usually assumed to be 1, refers to the approximation bias induced by the *k-means* discretization of the UH. In this sense, the number of groups K controls the trade-off between the two: the larger is K, the better is the approximation, but the number of group fixed effects estimated in the model grows consequently, worsening the incidental parameter bias. The employment of Rule 1 asymptotically leads to a number of groups in the order of $\sqrt{T}$, meaning that the leading order of the bias is $O_p\left(\frac{1}{T}\right)$.

BLM22 manage to prove that the $O_p(1/T)$ term is equal to $C/T + o_p(1/T)$ for some constant C. Consequently, the OW-GFE could be bias corrected employing Dhaene and Jochmans (2015)'s half-panel jackknife estimator. The distribution of the resulting bias-corrected OW-GFE estimator is asymptotically normal and centered at the truth.

The asymptotics for the TW-GFE closely resembles the OW-GFE's one, with two main differences: i) the asymptotic behaviour accounts for interactions between individual and time-specific cluster dummies in the model, ii) TW-GFE is characterised by a bias which is not proven to be constant, making standard bias correction techniques unfeasible. The TW-GFE is a generally biased estimator for both linear and nonlinear models and three different sources of bias could be found: an $O_p(K^{-\frac{2}{d_{\xi_i}}} + L^{-\frac{2}{d_{\lambda_t}}})$ and $O_p\left(\frac{1}{N} + \frac{1}{T}\right)$ terms represent the extension to time dimension of the bias-related $O_p(\cdot)$ elements discussed above for OW-GFE and they refer respectively to the *k-means* approximation bias and to the incidental parameter bias due to limited time and individual occasions for the moments. A new $O_p(\frac{KL}{NT})$ term comes instead from the estimation of the KL group specific parameters using NT observations. Symmetrically to OW-GFE, TW-GFE asymptotic behaviour could be simplified under a suitable choice of the number of groups K and L following the rule 1 and the hypothesis that $d_{\xi_i} = d_{\lambda_t} = 1$: if the rule is applied the leading term becomes $O_p\left(\frac{1}{N} + \frac{1}{T}\right)$. Unlike OW-GFE, the bias around which the distribution of the TW-GFE is centered is not proven to be constant, making the jackknife procedure impossible to apply.

5.1 Challenges

Despite the clear asymptotic framework and building-up procedure for the BLM22 GFE, its use in real world datasets is not clear-cut.

The most relevant aspect is how to determine the proper number of groups in order to have an effective GFE estimation. The issue is tightly related to the choice of the hyperparameter γ, since it directly controls the amount of clusters. One possible choice in order to find a proper value of γ is to perform a grid search, evaluating many possible values and selecting K, L driven by the parameters' stability together with economic intuition. Another option is to select the hyperparameter using cross-validation, which aims to minimizing forecasting error in linear case or to optimize metrics derived from the confusion matrix in binary choice models.

Another key point in the GFE framework is how to account for time-varying heterogeneity in a complete manner. The TW-GFE allows one to deal with unspecified time-varying UH, but it lacks methods for bias correction. Under a proper choice of the number of groups, the TW-GFE presents an asymptotic bias of the same order of the TW-FE, and this bias could be severe,

[17] This is the standard incidental parameters bias described in the Introduction and it hampers estimation of grouped fixed effects in nonlinear models only.

especially when T is short. BLM22 actually developed a version of the OW-GFE which can deal with time-varying UH, which works by interacting time dummies with the cluster-specific dummies, closely resembling what happens in the BM15 GFE. The time-varying OW-GFE is consistent but also biased, with a bias of the order $1/T + K/N$.

After applying the rule for K, the order of the bias of the time-varying OW-GFE becomes $1/T + \sqrt{T}/N$, although the rate of convergence is too slow to apply conventional bias reduction methods, meaning that unbiased GFE estimators that can manage time-varying UH are not available in the BLM22 framework.

5.2 Linear panel regressions with two-way unobserved heterogeneity

The Freeman and Weidner (2023)'s contribution, object of the paragraph, is closely related to that of BLM22, as it addresses the approximation of UH by means of a grouped structure. Their analysis pertains to a linear model like:

$$y_{it} = x_{it}\beta + h(\alpha_i, \lambda_t) + \epsilon_{it} \tag{19}$$

where $h(\cdot)$ is an unknown real-valued function depending on unobserved fixed effects. The singular value decomposition of the function $h(\alpha_i, \lambda_t)$ decomposes it as an infinite sum of multiplied factors and factor loadings.

In this vein, Freeman and Weidner developed a grouped fixed effects estimator to consistently estimate parameters of interest in Equation (19). Freeman and Weidner adopt a two-way additive specification for GFE, meaning that the set of individual- and time-cluster- specific dummies enter the model in an additive way. The GFE estimation follows a three-step algorithm: first, preliminary estimates of factors and factor loadings are computed using standard principal components techniques. The preliminary estimates of factors serve as inputs for a hierarchical clustering algorithm. In the final step, newly found group-specific individual and time dummies are used to estimate additive GFE. The clustering algorithm intrinsically leads to the formation of clusters of small size, such as 2-3 individuals per cluster: this approach conveniently relieves the analyst of the burden of choice the proper number of groups.

Under conditions outlined in Freeman and Weidner (2023), the additive two-way GFE estimator is consistent but asymptotically biased. Since the grouped structure actually introduces an approximation, the additive GFE has an approximation error that is of the order of $1/min(N,T)$. If rectangular array asymptotics is assumed, this error reduces to an order of $1/\sqrt{NT}$. To mitigate this bias, Freeman and Weidner (2023) propose a jackknife estimator in the spirit of Dhaene and Jochmans (2015), even though it lacks any formal discussion on its theoretical properties.

6 Conclusions

Latent grouped structure estimation in panel data is a new and flexible approach to either deal with UH or to pursue a better model specification. Starting from the relevant contribution of Bonhomme and Manresa (2015a), which addresses the estimation of discrete grouped fixed effects in linear models, the review focuses on the perks and limitations of this approach, following the relevant literature spurred by it. Then the affine in spirit yet different in construction contribution of Su et al. (2016a), which contemplates heterogeneity at a slopes level is discussed together with other papers that have extended and improved the C-Lasso approach. Finally, the Bonhomme et al. (2022)'s GFE is presented and analyzed, while its main limits are discussed.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and democracy. *American economic review*, 98(3):808–842.

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.

Ando, T. and Bai, J. (2023). Large-scale generalized linear models for longitudinal data with grouped patterns of unobserved heterogeneity. *Journal of Business & Economic Statistics*, 41(3):983–994.

Andrews, M. J., Gill, L., Schank, T., and Upward, R. (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society Series A: Statistics in Society*, 171(3):673–697.

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.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bartolucci, F., Bellio, R., Salvan, A., and Sartori, N. (2016). Modified profile likelihood for fixed-effects panel data models. *Econometric Reviews*, 35(7):1271–1289.

Bartolucci, F. and Nigro, V. (2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a root-n-consistent conditional estimator. *Econometrica*, 78(2):719–733.

Bartolucci, F. and Nigro, V. (2012). Pseudo conditional maximum likelihood estimation of the dynamic logit model for binary panel data. *Journal of Econometrics*, 170(1):102–116.

Bartolucci, F., Pigini, C., and Valentini, F. (2024). Mcmc conditional maximum likelihood for the two-way fixed-effects logit. *Econometric Reviews*, pages 1–26.

Bester, C. A. and Hansen, C. (2009). A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics*, 27(2):131–148.

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

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., and Setzler, B. (2023). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*, 41(2):291–322.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739.

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

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

Bonhomme, S. and Manresa, E. (2015b). Supplement to grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3).

Bryant, P. and Williamson, J. A. (1978). Asymptotic behaviour of classification maximum likelihood estimates. *Biometrika*, 65(2):273–281.

Carro, J. M. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, 140(2):503–528.

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

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal*, 20(3):S1–S13.

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

Chetverikov, D. and Manresa, E. (2022). Spectral and post-spectral estimators for grouped panel data models. *arXiv preprint arXiv:2212.13324*.

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030.

Dzemski, A. and Okui, R. (2021). Convergence rate of estimators of clustered panel models with misclassification. *Economics Letters*, 203:109844.

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.

Freeman, H. and Weidner, M. (2023). Linear panel regressions with two-way unobserved heterogeneity. *Journal of Econometrics*, 237(1):105498.

Grunewald, N., Klasen, S., Martínez-Zarzoso, f. I., and Muris, C. (2017). The trade-off between income inequality and carbon dioxide emissions. *Ecological Economics*, 142:249–256.

Gu, J. and Volgushev, S. (2019). Panel data quantile regression with grouped fixed effects. *Journal of Econometrics*, 213(1):68–91.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27(6):1152–1191.

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.

Hsiao, C. (2018). Panel models with interactive effects. *Journal of Econometrics*, 206(2):645–673.

Hsiao, C. (2022). *Analysis of panel data*. Number 64. Cambridge university press.

Janys, L. and Siflinger, B. (2024). Mental health and abortions among young women: Time-varying unobserved heterogeneity, health behaviors, and risky decisions. *Journal of Econometrics*, 238(1):105580.

Johar, M., Johnston, D. W., Shields, M. A., Siminski, P., and Stavrunova, O. (2022). The economic impacts of direct natural disaster exposure. *Journal of Economic Behavior & Organization*, 196:26–39.

Kapetanios, G., Serlenga, L., and Shin, Y. (2023). Testing for correlation between the regressors and factor loadings in heterogeneous panels with interactive effects. *Empirical Economics*, pages 1–49.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of econometrics*, 95(2):391–413.

Lentz, R., Piyapromdee, S., Robin, J.-M., et al. (2018). On worker and firm heterogeneity in wages and employment mobility: Evidence from danish register data. *PIER Discussion Papers*, 91.

Lewis, D. J., Melcangi, D., Pilossoph, L., and Toner-Rodgers, A. (2023). Approximating grouped fixed effects estimation via fuzzy clustering regression. *Journal of Applied Econometrics*, 38(7):1077–1084.

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.

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

Liu, R., Shang, Z., Zhang, Y., and Zhou, Q. (2020). Identification and estimation in panel models with overspecified number of groups. *Journal of Econometrics*, 215(2):574–590.

Loyo, J. A. and Boot, T. (2024). Grouped heterogeneity in linear panel data models with heterogeneous error variances. *Journal of Business & Economic Statistics*, pages 1–13.

Lu, X. and Su, L. (2017). Determining the number of groups in latent panel structures with an application to income and democracy. *Quantitative Economics*, 8(3):729–760.

Lumsdaine, R. L., Okui, R., and Wang, W. (2023). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, 233(1):45–65.

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

Miao, K., Su, L., and Wang, W. (2020). Panel threshold regressions with latent group structures. *Journal of Econometrics*, 214(2):451–481.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579.

Moon, H. R. and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195.

Moon, H. R. and Weidner, M. (2023). Nuclear norm regularized estimation of panel regression models.

Mugnier, M. (2022). A simple and computationally trivial estimator for grouped fixed effects models. *arXiv preprint arXiv:2203.08879*.

Mugnier, M. (2023). A simple and computationally trivial estimator for grouped fixed effects models. *arXiv preprint arXiv: 2203.08879*.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, pages 1–32.

Oberlander, L., Disdier, A.-C., and Etilé, F. (2017). Globalisation and national trends in nutrition and health: A grouped fixed-effects approach to intercountry heterogeneity. *Health economics*, 26(9):1146–1161.

Okui, R. and Wang, W. (2021). Heterogeneous structural breaks in panel data models. *Journal of Econometrics*, 220(2):447–473.

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

Pesaran, M. H. and Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of econometrics*, 142(1):50–93.

Pollard, D. (1981). Strong consistency of k-means clustering. *The annals of statistics*, pages 135–140.

Sarafidis, V. and Weber, N. (2015). A partially heterogeneous framework for analyzing panel data. *Oxford Bulletin of Economics and Statistics*, 77(2):274–296.

Su, L. and Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics*, 206(2):554–573.

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

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

Sun, Y. X. (2005). Estimation and inference in panel structure models.

Valentini, F., Pigini, C., and Bartolucci, F. (2023). Advances in maximum likelihood estimation of fixed-effects binary panel data models. In *Trends and Challenges in Categorical Data Analysis: Statistical Modelling and Interpretation*, pages 275–315. Springer.

Vogt, M. and Linton, O. (2017). Classification of non-parametric regression functions in longitudinal data models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(1):5–27.

Vogt, M. and Linton, O. (2020). Multiscale clustering of nonparametric regression curves. *Journal of Econometrics*, 216(1):305–325.

Wang, W., Phillips, P. C., and Su, L. (2018). Homogeneity pursuit in panel data models: Theory and application. *Journal of Applied Econometrics*, 33(6):797–815.

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