Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Bsc$^2 Econometrics and Economics$

Panel data models with Latent Heterogeneity: Econometric Methods, Algorithms and Empirical Application

Aditeya Gupta (558999)

Supervisor:	Dr. Wendun Wang
Second assessor:	A Venes Schmidt
Date final version:	1st July 2024

Abstract

This paper addresses latent heterogeneity in panel data models, crucial for accurate empirical economic analysis. Building on Bonhomme and Manresa's (2015) Group-Fixed Effect (GFE) model, it introduces a Three-Step Iterative GFE Algorithm to enhance computational efficiency and accuracy. Additionally, a novel Group Interactive Fixed Effects (GIFE) model integrates group-specific factor loadings into interactive fixed effects, allowing nuanced analysis of heterogeneous responses to common shocks (in cases where grouping structures are hypothesised). Monte Carlo simulations validate the robustness of the proposed algorithms in handling the different models, contributing to advanced methodologies in this research. Empirical applications include investigating the relationship between cumulative income and democracy. Multinomial logistic regressions are also used to explain the group memberships of the GFE and GIFE frameworks.

1 Introduction

Extensive evidence indicates the importance of considering unobservable heterogeneity while using panel data models in empirical economic studies. This heterogeneity can be found in either dimension of the panel data, across periods or individuals. As discussed by Hsiao (2003), disregarding latent heterogeneity in panel data studies can lead to incorrect inference and biased parameter estimation. Models that consider underlying heterogeneity have been demonstrably valuable in many settings. For example, Hsiao and Tahmiscioglu (1997) shows that heterogeneous slope coefficients may exist in US manufacturing firms' production functions. Conventional methods in econometrics account for heterogeneity by applying unit-specific, time-invariant effects. Furthermore, Hahn and Moon (2010) also claims that group structures are important in game theoretic models. Their research shows that group structures can be especially useful in games where many Nash equilibria are expected. However, such applications tend to model heterogeneity at the cost of model parsimony due to the so-called incidental parameter bias. Furthermore, they are unable to take other kinds of heterogeneity into account. This has led to a vast increase in research in this area over the last decade in this developing field. Therefore, this paper will specifically focus on replicating some parts of Bonhomme and Manresa (2015) and extending their model to a more general setting.

Bonhomme and Manresa (2015) proposes the Group-Fixed Effect (GFE) model framework. Their main model attempts to find latent grouping structures of individuals in a panel and allows for group-specific time-varying fixed effects (Equation 1). They further extend their model by allowing for unit-specific heterogeneity (equation 6) and group-specific coefficients (equation 7). Finally, they extend their model by accounting for both effects simultaneously (equation 8). This paper aims to replicate their empirical application using the simplest specification (equation 1) and the case with heterogeneous coefficients (equation 7).

As a first extension, this paper will propose an additional algorithm for the GFE model proposed by Bonhomme and Manresa (2015), named the Three-Step Iterative GFE Algorithm. This paper develops this algorithm for both the homogeneous (Section 3.1.2) and the heterogeneous cases (Section A.2) coefficients. The motivation for this procedure is to propose a more accurate algorithm than the basic one while being quicker and equally precise as the variable neighbourhood (VNS) search algorithm. Furthermore, this method combines the estimation methodologies of Bonhomme and Manresa (2015) and Bonhomme, Lamadon and Manresa (2022). It utilises algorithm 1 of Bonhomme and Manresa (2015) to find local minima. Then it uses those estimates to inform the algorithm of Bonhomme et al. (2022) to conduct a neighbourhood search. The algorithm continues iteratively until no further improvements are found. Finally, a Monte Carlo simulation is used to test the relative performance of the different algorithms in terms of the misclassification rates (the rate at which countries are classified into the wrong groups) and parameter estimates.

The second contribution of this paper is to propose a novel model and estimation process accounting for a type of interactive fixed effects model having group-dependent factor loadings, referred to as the Group Interactive Fixed Effects (GIFE) model. As discussed, the Group Fixed Effects model attempts to find time-varying fixed effects of latent groups in the data. However, another kind of heterogeneity in the data may be modelled using Interactive Fixed Effects models (Bai, 2009). Interactive fixed effects models assume that there are latent shocks during different periods. Subsequently, individuals in the data react to these shocks heterogeneously. Su and Ju (2018) have explored combining group-based coefficient heterogeneity and interactive fixed effects which allows for both these effects to be taken into account simultaneously. Furthermore, a Ando and Bai (2016) also proposes a model similar to the one considered in Section 3.2.2. Their model enforces group structures in the common shocks and also the factor loadings.

To this end, this paper proposes the Group Interactive Fixed Effects (GIFE) model. This model assumes that all individuals in the panel experience the same homogeneous common shocks. Within groups, individuals respond homogeneously, while responses differ across groups. There are many applications where using the GIFE can be useful. One such area of application is in environmental economics studies, where there are many common global climate initiatives (common shocks) but varied responses (factor loadings) by blocks of countries (Xu, Gao, Oka & Whang, 2022). Asset Pricing (and Financial economics) is another area where the GIFE can be applied. Financial systems are significantly affected by many latent and observable global effects, namely monetary policy changes, geopolitical events, or economic crises. Extensive evidence also shows that different asset classes/groups react heterogeneously to underlying shocks (Cieslak & Pang, 2021). This has led to the development of many types of asset pricing factor models like the Statistical Factor Model (Lai & Xing, 2008) (accounting for underlying factors) and the Fama French Models (Fama & French, 1993) (accounting for observable factors). Therefore, the GIFE model can be used in Asset Pricing to consider both observable and unobservable covariates, while accounting for group-based heterogeneous responses.

An important point to note is that the GIFE model is a nested model of the IFE model. This implies that both should be equally valid when the underlying data-generating process (DGP) follows the GIFE model. However, as noted in the Monte Carlo experiments in Bonhomme and Manresa (2015), even though the GFE model is a nested model of the IFE, the IFE tends to perform worse since it overfits noise. A similar case also holds for the relationship between the IFE and GIFE estimates. This is illustrated by the Monte Carlo simulation in this paper (Section 3.2.4), where the GIFE outperforms the IFE when the underlying DGP follows the GIFE model. One should also note that the IFE itself does not identify or estimate groups which may be important in empirical applications. To this end, this paper first details the regular IFE model and its estimation algorithm to provide a brief background on the method (Bai, 2009). Afterwards, the paper proposes an algorithm to estimate the GIFE model, by integrating the group fixed effects (GFE) estimation process with the interactive fixed effects methodology. To this end, the interactive fixed effects methodology has also been modified to enforce group structures in the factor loadings.

As a final step, this paper will consider an empirical study using the models considered in this paper. To this end, this paper applies the IFE, GFE, and GIFE models to accurately estimate the effect of cumulative income on democracy. The different models are used here to consider the possibility of various kinds of heterogeneity in this application. To gain further insight into the underlying drivers of group membership in the GFE and GIFE models, the effect of external covariates on group membership is also studied through multinomial logistic regressions. The detailed literature review in Section 4 will further explain the appropriateness of these models for this study.

The rest of the paper will be structured as follows. In Section 2, a literature review is performed, looking at past research in the area of Group-based Panel Data Models. In Section 3, the methodology is specified. This section consists of models this paper considers (GFE, IFE and the GIFE) and various algorithms to estimate them. The novel GIFE is proposed as well. Furthermore, Monte Carlo Simulations are also conducted to observe the finite sample performance of the different algorithms for the different specifications. In Section 4 the empirical application is discussed. Here, this paper also considers the main drivers of the underlying groups using covariates in a multinomial logistic regression. Overall, this paper replicates the results of Bonhomme and Manresa (2015) only in the empirical study in the GFE cases. All other results, including the Monte Carlo simulations are considered as extensions in this bachelor thesis.

2 Literature Review

A plethora of literature exists, detailing the estimation and existence of underlying grouping structures in panel data structures. The main complexity in such models lies in identifying the unknown underlying groupings in the data. As discussed by Shen and Huang (2010), the number of combinations in data partitioning is a Bell number and therefore extremely large. As detailed by W. Wang, Phillips and Su (2018), previous literature attempts to apply various methods to find these underlying grouping structures. The first approach uses prior data or knowledge to inform the grouping structures, as discussed in Bester and Hansen (2016). Such information can include the geographic locations of countries or firm industries. These estimation procedures are however prohibitive. Many panel data models may not have external variables that can be used to inform the grouping structures. Estimation also becomes unreliable when the number of groups is incorrectly specified or individuals in the panel are classified into the incorrect groups. To this end, better approaches have been devised to integrate group identification into the estimation process. The main approaches include K-means type algorithms, finite mixture models and machine learning methods. Finite mixture model literature can be traced back to papers like Kasahara and Shimotsu (2009) and Browning and Carro (2014). Both aimed to research statistical inference in discrete choice models, where the number of groups is known. Machine learning techniques have also been used to make estimations in this area. Particularly, Su, Shi and Phillips (2016) proposed a classifier-Lasso (C-Lasso) which uses a penalty term to enforce grouping structures in the parameter estimates. Subsequently, Lu and Su (2017) introduced a method to find the number of groups and Su, Wang and Jin (2018) expanded on this research by further developing estimation procedures for nonparametric and interactive fixed effects panels. To enhance the machine learning literature in this area, W. Wang et al. (2018) suggests the panel-Cards model. This algorithm is shown to asymptotically identify the true group structure and consistently estimate model parameters. Furthermore, Huang, Jin and Su (2020) apply and develop the C-Lasso method of Su et al. (2016)in cointegrated panels, allowing for endogeneity and non-stationarity.

The main studies, that apply the k-means type algorithm for group-based panel models, include Lin and Ng (2012), Sarafidis and Weber (2015), Bonhomme and Manresa (2015), Ando and Bai (2016) and Bonhomme et al. (2022). Of these papers, Lin and Ng (2012) and Sarafidis and Weber (2015) consider cases where the panel data models include group-based heterogeneous coefficients. Lin and Ng (2012) performs conditional clustering by applying an augmented k-means algorithm. The main idea of their algorithms is to minimize the sum of squared deviations within clusters. To this end, groups with individuals having similar coefficient (parameter) values are created. To further the research in the area of the K-means type algorithms, Bonhomme and Manresa (2015) proposes the Group Fixed Effects estimator. Their estimator and estimation procedure attempts to use a k-means type algorithm to simultaneously estimate the grouping structure in the panel while performing parameter inference. This method has strong asymptotic consistency properties for coefficient estimates in panels with a fixed number of individuals and infinite periods. The same holds when N (number of individuals) and T (time dimension) grow together. They apply a few different extensions to their models that take further nuances of panel models/data into account (further discussed in Section 3.1). Bonhomme et al. (2022) also proposes a Two Step Group fixed effects estimator. Their algorithm first groups individuals together and then applies a regression model (based on the estimated group structure) to make its estimates. Their procedure differs from other papers since it does not integrate the parameter and group estimation processes. Ando and Bai (2016) adds to this literature by considering, a factor component along with underlying grouping structures. This allows for complex relationships in the data to be captured.

Recent advancements in group-based panel models have led to significant developments and applications across various contexts. A notable extension involves addressing structural breaks within these models. Okui and Wang (2018) introduced a new model and estimation method specified to identify heterogeneous structural breaks in panel data. This method takes differences in the timing and magnitude of breaks into account across individual units, by utilizing a hybrid approach that combines grouped fixed effects with adaptive group fused Lasso. The approach has proven effective in detecting latent group structures and structural breaks, showing strong performance in both simulations and empirical applications. Building on this foundation, Lumsdaine, Okui and Wang (2023) Lumsdaine (2023) proposed models that accommodate changes in group membership or slope coefficients following structural breaks, offering a comprehensive framework for analyzing the dynamics of such changes in panel data. More recently, Y. Wang, Phillips and Su (2023) incorporated structural breaks into panel models, allowing for variations in group sizes and memberships before and after the breakpoint.

3 Methodology

This section outlines the methodology used throughout the remainder of the paper. First, the Group Fixed Effects Model and its variants (Section 3.1) are discussed (Bonhomme & Manresa, 2015). Here, a new algorithm for the model estimation is also proposed. It combines Algorithm 1 of Bonhomme and Manresa (2015) and applies an iterative form of the two-step GFE estimation procedure of Bonhomme et al. (2022). A comparison is also made to the preexisting algorithms of Bonhomme and Manresa (2015) by applying Monte Carlo simulations.

In the subsequent section, a brief introduction of the IFE model is detailed to provide a foundation (Section 3.2.1). Finally, the Grouped Interactive Fixed Effects Model, its adjoining estimations process and GIFE model extensions are discussed (Section 3.2.2 and 3.2.3). Monte Carlo simulations are also considered here to test the accuracy of the estimation process and misclassification rates (Section 3.2.4).

3.1 Group-Fixed Effects with Extensions: Model Specifications and Estimation Processes

The group-based panel models have been extensively studied in literature as discussed in Section 2. In this section, firstly, specific attention is given to the basic model of heterogeneity discussed by Bonhomme and Manresa (2015) as shown in equation 1.

$$y_{it} = x'_{it}\theta + \alpha_{g_i t} + \epsilon_{it}, \quad i = 1, \dots, N \quad \text{and} \quad t = 1, \dots, T \tag{1}$$

In this model specification, it is assumed that x_{it} is uncorrelated with ϵ_{it} but may have some correlation structure with the $\alpha_{g_i t}$ term. Furthermore, the groupings $g \in 1, ..., G$ and $\alpha_{g_i t} \forall i, t$ are left unrestricted to be estimated through statistical inference-based methods. Making parameter estimates implies solving the following optimisation problem in equation 2. Therefore, given values of θ and α, the group assignment of each individual takes on the optimisations in equation 3. Furthermore, with known groupings, (θ, α) can be estimated as a regular regression problem by solving the problem in equation 4.

$$(\hat{\theta}, \hat{\alpha}, \hat{\gamma}) = argmin_{(\theta,\alpha,\gamma)\in\Theta\times A^{GT}\times\Gamma_G} \{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x'_{it}\theta - \alpha_{g_i,t})^2\} \tag{2}$$

$$\hat{g}_i(\theta, \alpha) = argmin_{g\in\{1,\dots,G\}} \{\sum_{t=1}^{T}(y_{it} - x'_{it}\theta - \alpha_{gt})^2\} \tag{3}$$

$$(\hat{\theta}, \hat{\alpha}) = argmin_{(\theta,\alpha,\gamma)\in\Theta\times A^{GT}} \sum_{t=1}^{T}(y_{it} - x'_{it}\theta - \alpha_{\hat{g}_i(\theta,\alpha)t})^2 \tag{4}$$

Considering no simple solution to solve the optimisation in equation 2 exists, Bonhomme and Manresa (2015) propose an iterative k-means-like algorithm to reach a solution. To this end, Bonhomme and Manresa (2015) propose Algorithm 1. As one may notice, this algorithm uses an iterative process which alternates between equation 3 and 4 til numerical convergence.

One of the main issues in such algorithms is the dependence of the final estimation on the initial conditions. Therefore, it is usually necessary to attempt different initial conditions to find the starting point that yields the lowest sum squared residuals (SSR). By applying this algorithm with many different starting points, the solution is the iteration that obtains the lowest possible sum squared residuals value.

Algorithm 1 Iterative Estimation Algorithm

1: **Initialization:**
- Let $(\theta^{(0)}, \alpha^{(0)}) \in \Theta \times A_{GT}$ be some starting value.
- Set $s = 0$.

2: **while** not converged **do**
3:     **for** all $i \in \{1, \dots, N\}$ **do**
4:         Compute:

$$g_i^{(s+1)} = \arg \min_{g \in \{1,\dots,G\}} \sum_{t=1}^{T} (y_{it} - x'_{it}\theta^{(s)} - \alpha_{gt}^{(s)})^2$$

5:     **end for**
6:     Update parameters:

$$(\theta^{(s+1)}, \alpha^{(s+1)}) = \arg \min_{(\theta,\alpha) \in \Theta \times A_{GT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x'_{it}\theta - \alpha_{g_i^{(s+1)}t})^2$$

7:     Increment s:

$$s = s + 1$$

8: **end while**

The main issue in algorithm 1 is that it may necessitate a prohibitive number of starting points to reach the lowest optimum, especially in cases where many latent groups may exist. To this end, Bonhomme and Manresa (2015) also proposes a Variable Neighborhood Search Based Algorithm (see Algorithm 6 in Appendix A.1). The first main addition of this algorithm is a local search that ensures a local optimum is found. It is important to note that the local optimum found here is not necessarily a local minimum found in Algorithm 1. The other addition is the neighbourhood search which explores the optimised function more rigorously.

After using an algorithm of choice for different group sizes, the optimal number of groups is chosen by minimising the information criterion in equation 5 (Bonhomme and Manresa (2015)).

$$\text{BIC}(G) = \log \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x'_{it}\hat{\theta}^{(G)} - \hat{\alpha}_{g_i t}^{(G)})^2 \right) + \hat{\sigma}^2 \frac{GT + N + K}{NT} \log(NT) \tag{5}$$

3.1.1 Model Extensions

To extend this model, Bonhomme and Manresa (2015) discuss allowing for unit-specific heterogeneity as shown in equation 6. Applying a simple with-in transformation to equation 6 yields a similar specification to the basic GFE model, allowing for an identical estimation process to the basic GFE model. The second model extension proposed by the authors allows for heterogeneous coefficients as seen in equation 7. To perform statistical inference on equation 7, an augmented version of Algorithm 1 must be used. The final model of interest would be to apply both extensions of unit-specific heterogeneity and heterogeneous coefficients (equation 8). Once again, using a simple transformation, equation 8 can be rewritten in a similar form to equation 7.

$$y_{it} = x'_{it}\theta + \alpha_{g_i t} + \eta_i + \epsilon_{it} \tag{6}$$

$$y_{it} = x'_{it}\theta_{g_i} + \alpha_{g_i t} + \epsilon_{it} \tag{7}$$

$$y_{it} = x'_{it}\theta_{g_i} + \alpha_{g_i t} + \eta_i + \epsilon_{it} \tag{8}$$

One important note for the interested researcher is that estimation of model specifications 7 and 8 require Algorithm 1 or Algorithm 6 to be repurposed. However, the general procedure remains unchanged. One final point is that the sandwich estimator is used to calculate the standard errors of this method (as discussed in Bonhomme and Manresa (2015)).

3.1.2 Three-step Iterative GFE Algorithm: Algorithm Proposed in this Paper

In this section, an original algorithm, the Three Stage Iterative GFE estimator, is developed to estimate the GFE-type models. This algorithm fuses the methodologies of Algorithm 1 as discussed by Bonhomme and Manresa (2015) and the Two-Step GFE as discussed by Bonhomme et al. (2022). An interesting point to note is that both methods individually have strong asymptotic consistency properties.

The core innovation here is the use of coefficient estimates of Algorithm 1 from Bonhomme and Manresa (2015) to define the $h_i(y_i, x'_i)$ function of the Two-Step GFE of Bonhomme et al. (2022). Then, the Two-Step GFE is used to conduct a neighbourhood search. Whenever there is an improvement in the objective value, the algorithm reverts to the first step which applies the regular GFE model to find a "local solution". Then the updated coefficient estimates reinform the $h_i(y_i, x'_i)$ function to augment the "revealed" information regarding the grouping structure of the individuals. One hyperparameter to specify here is the *neigh_max* value. This value specifies how many times the k-means algorithm is applied (steps 2 through 4) for a single neighbourhood search. If no better group structure is found within the specified *neigh_max* number of iterations the algorithm stops. The researcher should note that the Three-Step Iterative GFE Algorithm is used to augment and improve the estimates of the basic GFE algorithm.

The Three-Step Iterative GFE algorithm can also be adapted for the specification with heterogeneous coefficients (equation 7). Please check the appendix for this algorithm (Appendix A.2). Note that this algorithm consistently enhances the estimates of the regular GFE algorithm.

3.1.3 Monte Carlo Simulation to test the relative performance of the Algorithms

A set of Monte Carlo simulations is performed to evaluate the relative performance of the 3-step GFE against Algorithms 1 and 2 of Bonhomme and Manresa (2015) (only simplest case considered). The main points to consider are the time to convergence per iteration, the bias of the coefficients of interest, and the misclassification rate. Please look at Appendix 3.1.3 for further details on the specific settings of each algorithm in this Monte Carlo Simulation. To this end, this paper generates data from DGP in equation 9.

Algorithm 2 Three-Step Iterative GFE Algorithm, Heterogeneous case

1. **Apply Algorithm 1 of Bonhomme and Manresa (2015):**
   (a) Initialize based on random or specific starting points and estimate:
   $$y_{it} = x'_{it}\theta + \alpha_{g_i t} + \epsilon_{it}$$
   (b) After numerical convergence, store the sum of squared errors (SSE), parameters (θ, α), and group memberships g_i for all $i = 1, \ldots, N$.

2. **Iterative Two-Step GFE (Bonhomme and Manresa, 2021):**
   (a) Define function h_i for mapping underlying groupings:
   $$h_i(y_i, x'_i) = y_i - x'_i\theta$$
   Where:
   $$y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}, \quad x'_i = \begin{bmatrix} x'_{i1} \\ \vdots \\ x'_{iT} \end{bmatrix}$$
   (b) Set iteration counter $iter = 0$ and neigh_max to set the number of iterations.

3. **Classification and Estimation Steps:**
   (a) Perform the classification step as per equation 4 from Bonhomme and Manresa (2021) to update groupings g_i for all i (this is Lloyd's/K-means algorithm).
   (b) Minimize SSE in the estimation step analogous to their equation 5:
   $$(\theta, \alpha) = \arg \min_{\theta, \alpha \in \Theta \times A^{GT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x'_{it}\theta - \alpha_{g_i t})^2$$

4. **Check for Improved SSE:**
   (a) If SSE is lower than the previous best, update stored (θ, α) and g_i for all i, and go back to step 1 with these new values as initialization.
   (b) Otherwise, if $iter \geq$ neigh_max, end the search; else increment $iter$ by 1, retain the previous best (θ, α), and return to step 3.

$$y_{it} = x_{it}'\theta^0 + \alpha^0_{g_i t} + \epsilon_{it}, \quad i = 1, \ldots, N \quad \text{and} \quad t = 1, \ldots, T \tag{9}$$

The specifications detailed in the following paragraph apply to all permutations of $G = 3$, $N = 50, 100$, and $T = 10, 20$. Additionally, the values of σ_{it} are specified using the formulas $(0.4 + 0.1\mathbf{x}_{it} \cdot \mathbf{x}_{it})^{1/2}$ and $(1 + 0.1\mathbf{x}_{it} \cdot \mathbf{x}_{it})^{1/2}$:

This paper allows for two regressors, each being sampled from an independent and identically distributed normal distribution $N(0, 1)$. The slope coefficients θ^0 are specified to be $(1.5, 2)$. Due to the high dimensionality of the α parameters, they are also sampled from an independent and identically distributed normal distribution $N(0, 10)$ for all g and t. ϵ_{it} is distributed according to $N(0, \sigma^2_{it})$. The proportion of data distribution to each group is $[0.5, 0.3, 0.2]$ when $G = 3$. The proportions here are chosen such that no group has only a single element (which would lead to issues with rank conditions in OLS). Please note that this paper does not report extra tables for the 2 different σ_{it} specifications. Alternatively, this paper averages over the σ_{it} specifications for all combinations of (G, N, T). The hyperparameters chosen for the VNS algorithm are 10 and 10 for the maximum iterations and neighbourhood search parameters respectively. The maximum number of iterations set for the Three Step Iterative GFE is 30. Please note that this paper does not report separate tables for the 2 different σ_{it} specifications. Alternatively, this paper averages over the σ_{it} specifications for all combinations of (G, N, T).

	Basic	**VNS**	**Three Step**
N = 50, T = 10	0.0503	0.0459	0.0454
N = 50, T = 20	0.0144	0.0049	0.0054
N = 100, T = 10	0.0447	0.0342	0.0306
N = 100, T = 20	0.0044	0.0044	0.0035

Table 1: Average Misclassification rates over iterations when G = 3 (averaged over both cases of σ_{it})

From table 1 it is clear that the Three-step Iterative GFE performs equally well (or better) than the other algorithms in classifying the individuals into the correct groups. Particularly, the Three Step iterative Algorithm performs best when N and T are larger. Furthermore, it is important to note that the Three-step Iterative GFE algorithm significantly outperforms the VNS algorithm by having significantly lower run times. However, the Basic GFE model, although more inaccurate in its group estimation, has much faster run times due to the lack of neighbourhood search elements. Therefore, from the standpoint of runtimes and the ability to correctly classify elements into their respective underlying groups, it would make sense to use either the Basic GFE algorithm (for greater speed) or the Three-step Iterative GFE Algorithm (both faster and better in classifying than the VNS).

Looking more closely at the statistics for the parameter estimate of the different algorithms, it is clear that the algorithms perform either faster or make superior parameter and group membership estimates. VNS and Three Step GFE marginally outperform the Basic GFE, but are much more computationally intensive. An important point to note here is that the Three-

Algorithm	N/T Combination	Theta_1			Theta_2		
		Bias	RMSE	CP	Bias	RMSE	CP
Basic GFE	N = 50, T = 10	0.0026	0.0237	0.9550	-0.0044	0.0452	0.9450
	N = 50, T = 20	0.0044	0.0471	0.9400	0.0014	0.0438	0.9350
	N = 100, T = 10	-0.0010	0.0296	0.9550	0.0015	0.0285	0.9500
	N = 100, T = 20	0.0010	0.0194	0.9500	-0.0010	0.0219	0.9450
VNS	N = 50, T = 10	0.0023	0.0398	0.9400	0.0015	0.0389	0.9500
	N = 50, T = 20	0.0021	0.0261	0.9450	0.0013	0.0274	0.9300
	N = 100, T = 10	0.0021	0.0290	0.9550	-0.0030	0.0280	0.9550
	N = 100, T = 20	0.0001	0.0188	0.9450	0.0023	0.0192	0.9400
Three Step GFE	N = 50, T = 10	0.0015	0.0421	0.9500	0.0001	0.0396	0.9350
	N = 50, T = 20	0.0005	0.0269	0.9500	-0.0034	0.0282	0.9400
	N = 100, T = 10	-0.0008	0.0261	0.9500	-0.0003	0.0255	0.9550
	N = 100, T = 20	-0.0017	0.0190	0.9500	0.0013	0.0199	0.9300

Table 2: Table showing averaged Bias, RMSE, and CP for Theta_1 and Theta_2 across different N/T combinations and algorithms (average over both specifications of σ_{it})

step GFE tends to obtain lower sum square residual values than the other two algorithms. For example, when N, T = 100, 20, the Basic GFE and the VNS achieve average SSR values of 1939 and 1931. On the other hand, the Three Step Algorithm achieves a lower average SSR of 1925. Therefore, in some cases, the main reason for its relatively high bias (of the Three Step GFE) is due to noise in the data rather than its ability to reach the lowest Sum Squared Residual Value. With this extra information, the three-step GFE makes better estimates than the basic algorithm and equally good estimates as the VNS. Do note that the runtimes of the Three-Step GFE are significantly better than those of the VNS. To this end, when estimating the GFE model of equation 1, this paper suggests either using the Basic GFE estimator as in Algorithm 1 if quicker runtime is the priority or Algorithm 2 if accurate parameter and group membership estimates are the priority. One final note is that increasing N and T leads to lower misclassification rates and biases. This reaffirms the asymptotic theory discussed by Bonhomme and Manresa (2015).

3.2 (Group) Interactive Fixed Effects Models

This section of the paper develops a hybrid of the Interactive Fixed Effects and Grouped Fixed Effects models. Although variations of this idea have been explored in previous literature, for example by Ando and Bai (2016) and Su and Ju (2018), no literature has studied a model which considers group-specific factor loadings and universal time-specific common shocks. Ando and Bai (2016) comes the closest but considers both group-specific heterogeneous common shocks and group-specific factor loadings rather than common shocks which remain homogeneous across groups.

The rest of this section will be structured as follows. First, the simple Interactive Fixed effects model will be discussed in 3.2.1 to provide a basis for the model structure, necessary restrictions and estimation process. Then in Section 3.2.2, the specifics of the Group Interactive Fixed Effects model will be discussed. Finally, a Monte Carlo Simulation will be applied to test the finite sample properties of the different estimators.

3.2.1 Interactive Fixed Effects

Before moving to the more difficult group-based setting, first, it is pertinent to consider the regular interactive fixed effects model as discussed by Bai (2009). The model specification (equation 10) and restrictions are detailed as follows in the rest of this subsection.

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\theta} + \lambda_i'\mathbf{F}_t + \epsilon_{it}, \\ \epsilon_{it} &\sim N(0, \sigma_{it}^2), \end{aligned} \tag{10}$$

Where y_{it} is the dependent variable for individual i at time t, $\mathbf{x}_{it}$ is a vector of regressors, $\boldsymbol{\theta}$ is a vector of slope coefficients, λ_i is an individual-specific factor loading vector, $\mathbf{f}_t$ is a time-specific factor vector, and ϵ_{it} is the error term, assumed to be normally distributed with mean zero and variance σ_{it}.

To estimate the IFE as in equation 10, it is helpful to consider the matrix notation of the model. Subsequently, the above is rewritten as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{F}\boldsymbol{\Lambda}' + \boldsymbol{\epsilon} \tag{11}$$

where $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_N]$, $\boldsymbol{\Lambda}' = [\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N]$, and

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{i1}' \\ \vdots \\ \mathbf{x}_{iT}' \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_T \end{bmatrix}^\top, \quad \boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{iT} \end{bmatrix}.$$

The model above follows a factor structure. This implies that certain identification restrictions must be applied to attain parameter estimates. This is due to the lack of unique solutions for $\mathbf{F}$ and $\boldsymbol{\Lambda}'$. Without restrictions, the following holds: $\mathbf{F}\boldsymbol{\Lambda}' = \mathbf{F}AA^{-1}\boldsymbol{\Lambda}' = \mathbf{F}^*\boldsymbol{\Lambda}^{*'}$, where A is any $r \times r$ invertible matrix. Such invertible matrices have r^2 free elements, implying that r^2 restrictions are needed to identify $\mathbf{F}$ and $\boldsymbol{\Lambda}'$. To this end, two different identification restrictions are placed: $\frac{1}{T}\mathbf{F}'\mathbf{F} = \mathbf{I}$ and $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ is diagonal. The first restriction leads to $\frac{r(r+1)}{2}$ restrictions while the latter leads to $\frac{r(r-1)}{2}$ restrictions. The two sets of restrictions jointly lead to sufficient restrictions to identify $\boldsymbol{\Lambda}'$ and $\mathbf{F}$.

The overarching least squares optimisation which must be solved is detailed in equation 12.

$$(\hat{\theta}, \hat{F}, \hat{\Lambda}) = argmin_{\theta,F,\Lambda}\{\sum_{i=1}^{N}(Y_i - X_i\theta - F\lambda_i)'(Y_i - X_i\theta - F\lambda_i)\} \tag{12}$$

Due to the complicated nature of the problem, Bai (2009) suggests an iterative procedure to solve this least squares problem. Do note, that the optimisation in equation 12 is not universally convex. This implies that there is no guarantee that a global optimum will be found from any arbitrary initial β. To this end, Bai (2009) suggest using the within-group parameter estimate as a reasonable initialisation for the β estimation. Please refer to Appendix A.3 for the specific steps of the algorithm. It is also interesting to note that the GFE is a specific case of the IFE.

As a final point, Bai (2009) propose the information criterion in equation 13 to choose the number of factors the IFE model should use.

$$\mathrm{IC}(p) = \log\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}'\hat{\boldsymbol{\theta}} - \hat{\lambda}_i'\hat{\mathbf{F}}_t)^2\right) + p\frac{N+T}{NT}\log(NT) \tag{13}$$

3.2.2 Group Interactive Fixed Effects

Having discussed the simpler Interactive Fixed Effects model, now this paper details the Grouped Interactive Fixed Effects Model. As discussed at the start of this section, this model is suitable for a situation where latent common shocks affect the whole sample space and all individuals in the population are affected based on their group status. Such a model would have the specification of equation 14. Notice that this model resembles the simple GFE model in equation 1. The only difference lies in the fact that $\alpha_{g_i t}$ has been replaced by $\lambda_{g_i}'\mathbf{f}_t$. Another intricacy to note here is that the GIFE model as in equation 14 is equivalent to the IFE model discussed in 3.2.1 where $\lambda_i = \lambda_j$ if $G_i = G_j$ $\forall i, j = 1, ..., N$.

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\theta} + \lambda_{g_i}'\mathbf{F}_t + \epsilon_{it}, \\ \epsilon_{it} &\sim N(0, \sigma_{it}), \end{aligned} \tag{14}$$

Here, y_{it} is the dependent variable for individual i at time t, $\mathbf{x}_{it}$ is a vector of regressors, $\boldsymbol{\theta}$ is a vector of slope coefficients, λ_{g_i} is a group-specific factor loading vector, $\mathbf{F}_t$ is a time-specific factor vector, and ϵ_{it} is the error term, assumed to be normally distributed with mean zero and variance σ_{it}.

To estimate the model above, the same identification restrictions are required as in Section 3.2.1. However, one must notice the increased difficulty in the estimation of this process due to the grouping structure of the factor loadings. Keeping the previous identification restrictions, this paper proposes an augmented version of Algorithm 8 (for IFE model estimation) to account for the grouped structure of the factor loadings. This leads to Algorithm 3 where Group Membership is assumed to be known. The significant difference applied in Algorithm 3 (for the GIFE model with known group membership) in comparison to Algorithm 8 (for the regular IFE model) is in step 2. To enforce the grouping structure in the factor loadings, instead of applying PCA in equation 21, W_{it} is replaced with a new term, $W_{g_i t}$ (equation 15). Since $W_{g_i t}$ is aggregated over all individuals within a group, individuals in the same group have the same $W_{g_i t}$ value and are therefore forced to have the same factor loadings.

The estimation process of this method requires the simple GFE estimation (algorithm 1), the VNS algorithm (algorithm 6) or the Three-step Iterative GFE Algorithm (algorithm 2) to be augmented. One can incorporate both the parameter and group membership estimation procedure of equation 14 into the estimation process. The adaption of the simplest GFE algorithm

Algorithm 3 Infeasible Grouped Interactive Fixed Effects Estimation Algorithm with known Group Membership

1. **Initialization:**
   - Set $s = 0$.
   - Initialize $\beta^{(0)}$ using the within-group estimator.
2. **Compute Residuals:**
   - Set $W_{it} = y_{it} - x'_{it}\beta^{(s)}$.
   - Now define:

$$W_{g_i t} = \sum_{j \in g_i} W_{jt} \tag{15}$$

   Estimation of the following yields a pure factor model:

$$W_{g_i t} = \lambda'_{g_i} F_t + \epsilon_{it} \tag{16}$$

3. **Principal Component Analysis:**
   - Apply Principal Component Analysis (PCA) on the factor structure to find the factor loadings $\hat{\lambda}_{g_i}^{(s)}$ and common shocks $\hat{F}_t^{(s)}$.
4. **Update β:**
   - Considering $\hat{\lambda}_{g_i}^{(s)}$ and $\hat{F}_t^{(s)}$ as "observable" at this stage, update β by applying Ordinary Least Squares (OLS):

$$\beta^{(s+1)} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} X'_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} \left(Y_{it} - \hat{\lambda}_{g_i}^{(s)} \hat{F}_t^{(s)\prime} \right) \right)$$

5. **Check Convergence:**
   - Stop if numerical convergence is achieved. Otherwise, set $s = s + 1$ and go back to step 2.

for the GIFE estimation can be found in Algorithm 4. An augmented version of the 3-step GFE algorithm is used to conduct the Monte Carlo of the GIFE model in Section 3.2.4.

The main way Algorithm 4 (GIFE estimation) differs from its counterpart for the GFE algorithm is in step 6. The main difference here lies in the estimation of the parameters. Here, the iterative process of Algorithm 3 (algorithm for GIFE with known group membership) is used to make parameter estimates rather than the analytical solution for the known group membership as in the regular GFE algorithm. This also causes the estimation process to become much more computationally intensive than the regular simple GFE algorithm (an iterative process within an iterative process). Also, one should note that parameter estimates change marginally for the same group membership in algorithm 3 (algorithm stops with numerical convergence). To this end, algorithm 4 stops when group membership experiences no change and coefficient values either remain the same or change insignificantly across iterations. Similar changes can be made to the other algorithms (VNS and 3-Step Iterative GFE) to find parameter estimates.

Algorithm 4 Basic Estimation Process Repurposed for GIFE

1: **Initialization:**
- Let $(\theta^{(0)}, \alpha^{(0)}) \in \Theta \times A_{GT}$ be some starting value.
- Set $s = 0$.

2: **while** not converged **do**
3: **for** all $i \in \{1, \ldots, N\}$ **do**
4: Compute:

$$g_i^{(s+1)} = \arg \min_{g \in \{1,\ldots,G\}} \sum_{t=1}^{T} (y_{it} - x'_{it}\theta^{(s)} - \alpha_{gt}^{(s)})^2$$

5: **end for**
6: Update parameters using Algorithm 3 to estimate the GIFE model with known group structure:

$$(\theta^{(s+1)}, \Lambda^{(s+1)}, F^{(s+1)}) = argmin_{(\theta,\Lambda,F)\in\Theta\times\Lambda^G\times F^T} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x'_{it}\theta - \lambda'_{g_i^{(s+1)}} f_t)^2$$

7: Let $\alpha_{g_i t}^{(s+1)} = \lambda'_{g_i^{(s+1)}} f_t$ and increment s:

$$s = s + 1$$

8: **end while**

One final note is that no information criterion has been developed for the GIFE. For the sake of simplicity, this paper suggests using a simple elbow plot to decide the number of groups and the number of factors. Finding a proper information criterion is left for future research.

3.2.3 Extensions to the Group Interactive Fixed Effects

Analogously to the regular GFE model, the GIFE model can also be generalised to consider individual-specific intercepts and heterogeneous coefficients. The process followed to generalise the regular GIFE model to the more complicated cases is very similar to the same generalisations in the GFE Model.

The first extension to consider is the GIFE with unit-specific heterogeneity (equation 17). Similar to the GFE case, this model can be rewritten as a simple GIFE using a with-in transformation. The second model extension is in equation 18 where heterogeneous coefficients are accounted for.

$$y_{it} = x_{it}'\theta + \lambda_{g_i}'F_t + \eta_i + \epsilon_{it} \tag{17}$$

$$y_{it} = x_{it}'\theta_{g_i} + \lambda_{g_i}'F_t + \epsilon_{it} \tag{18}$$

A model with both heterogeneous coefficients and unit-specific heterogeneity can also be considered. Using a with-in transformation, such a model can be estimated similarly to equation 18.

3.2.4 Monte Carlo Simulation to test properties of the regular GIFE estimator

In this subsection, this paper aims to test the finite sample properties of the parameter estimates in equation 14 using the estimation process as in Algorithm 4 in a Monte Carlo Simulation.

Firstly, this paper tests the coefficient estimates of the **infeasible** estimator where group membership is known beforehand and compares these to the ones made by the regular algorithm. This will allow for a proper evaluation of algorithm 3 by itself in making parameter estimates. Furthermore, the infeasible estimation process with unknown group membership is also tested to check the validity of the overall algorithm (algorithm 6).

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\boldsymbol{\theta}^{\mathbf{0}} + \lambda_{g_i}^{0}{}'\mathbf{F}_t^0 + \epsilon_{it}, \\ \epsilon_{it} &\sim N(0, \sigma_{it}^2), \end{aligned} \tag{19}$$

The experimental setup is designed with several key specifications. There are three groups ($G = 3$) and the sample sizes are set at $N = 50$ and 100, across periods $T = 10$ and 50, with each setup involving three parameters ($p = 3$). The model incorporates two regressors, each following an independent and identically distributed normal distribution $N(0, 1)$. The slope coefficients, θ^0, are specified as $(1.5, 2)$. Given the high dimensionality of the F_t and λ_{g_i} parameters, their elements are sampled from a normal distribution $N(0, 10)$ for all g and t. The error terms, ϵ_{it}, are distributed according to a normal distribution $N(0, \sigma_{it}^2)$, where σ_{it} is defined as $(0.4 + 0.1\mathbf{x}_{it} \cdot \mathbf{x}_{it})^{1/2}$ and $(1 + 0.1\mathbf{x}_{it} \cdot \mathbf{x}_{it})^{1/2}$. Lastly, the proportion of data allocated to each group is set at $[0.5, 0.3, 0.2]$. Please note that this paper does not report separate tables for the 2 different σ_{it} specifications. Alternatively, this paper averages over the σ_{it} specifications for all combinations of (G, N, T). Note that here only the basic GFE and GIFE algorithms are used rather than the VNS or three-step counterparts.

From table 3 it is clear that the infeasible GIFE (known group membership) itself makes proper estimates of the actual coefficients implying that algorithm 3 works as intended. Furthermore, algorithm 4 also performs very well in the test cases considered, classifying individuals into the correct groups in almost all cases. Therefore, it performs equivalently well to the infeasible GIFE in making its parameter estimates. Given this information, the estimation process discussed throughout 3.2.2 is appropriate for making proper coefficient estimates and finding correct grouping structures. Also notice that with a larger T (when N = 50), the misclassification rate reduces. Furthermore, the parameter becomes more accurate and the confidence

Algorithm	N/T Combination	Theta_1			Theta_2			MR
		Bias	RMSE	CP	Bias	RMSE	CP	
Inf. GIFE	N = 50, T = 10	0.0045	0.0381	0.935	0.0045	0.0381	0.955	N/A
	N = 50, T = 50	0.0027	0.0211	0.950	0.0027	0.0211	0.945	N/A
	N = 100, T = 10	0.0021	0.0262	0.935	0.0021	0.0262	0.935	N/A
	N = 100, T = 50	0.0001	0.0115	0.970	0.0001	0.0115	0.950	N/A
Reg. GIFE	N = 50, T = 10	0.0088	0.0384	0.935	0.0088	0.0384	0.950	0.0042
	N = 50, T = 50	0.0059	0.0212	0.950	0.0059	0.0212	0.945	0.0011
	N = 100, T = 10	0.0042	0.0262	0.935	0.0042	0.0262	0.935	0.0000
	N = 100, T = 50	0.0003	0.0116	0.975	0.0003	0.0116	0.965	0.0049

Table 3: Aggregated Table showing Bias, RMSE, CP, and Misclassification Rate for Theta_1 and Theta_2 across different N/T combinations in the Infeasible Estimation Case and the Real Estimation Based Case when G = 3 and p = 3

intervals also have a higher probability of containing the actual parameter values at the 5% significance level.

Subsequently, this paper evaluates the effect of choosing other model specifications in making parameter estimates when DGP in equation 19 is the correct specification. A few different (in)correct model specifications are considered. Firstly, the model is estimated using the simple IFE model as discussed in Section 3.2.1. Note that this is not a misspecification considering the GIFE is a nested model (specific case) of the IFE. However, the IFE is expected to make worse estimates due to its general form in comparison to the GIFE. Furthermore, no group estimates are made by the IFE. Afterwards, this paper also attempts the incorrect specification of the regular GFE. The GIFE model is a baseline model to evaluate the estimates of the other specifications. In this experiment, the number of loadings (p) and number of groups (G) are assumed to be known.

In this Monte Carlo experiment, this paper utilizes a setup comprising three groups ($G = 3$), with sample sizes $N = 100$, and periods $T = 10$ and 50. Each setup involves three factors ($p = 3$). The model includes two regressors, each sampled from an independent and identically distributed normal distribution $N(0, 1)$. The slope coefficients, θ^0, are established at $(1.5, 2)$. Given the high dimensionality, the elements of F_t and λ_{g_i} are also drawn from an independent and identically distributed normal distribution $N(0, 10)$ for all groups g and periods t. The error terms, ϵ_{it}, follow a normal distribution $N(0, \sigma_{it}^2)$, where σ_{it} is calculated as $(0.4+0.1\mathbf{x}_{it}\cdot\mathbf{x}_{it})^{1/2}$ and adjusted to $(1+0.1\mathbf{x}_{it}\cdot\mathbf{x}_{it})^{1/2}$. The distribution of data across the groups is set at $[0.5, 0.3, 0.2]$. Please note that this paper does not report extra tables for the 2 different σ_{it} specifications. Alternatively, this paper averages over the σ_{it} specifications for all combinations of (G, N, T). Note that here only the basic GFE and GIFE algorithms are used rather than the VNS or three-step counterparts.

From table 4, it is first important to note that IFE does not make any group membership estimates and therefore is not informative about the underlying group structure of the data.

Specification	N/T Combination	Theta_1			Theta_2			MR
		Bias	RMSE	CP	Bias	RMSE	CP	
GFE	N = 100, T = 10	-0.0001	0.0271	0.965	0.0017	0.0271	0.965	0.0194
	N = 100, T = 50	-0.0010	0.0117	0.955	-0.0002	0.0120	0.950	0.0823
IFE	N = 100, T = 10	0.0006	0.0340	0.945	0.0012	0.0334	0.960	N/A
	N = 100, T = 50	-0.0011	0.0112	0.945	-0.0010	0.0131	0.945	N/A
GIFE	N = 100, T = 10	0.0021	0.0262	0.935	-0.0001	0.0262	0.935	0.0000
	N = 100, T = 50	0.0001	0.0116	0.975	0.0012	0.0123	0.965	0.0049

Table 4: Averaged results for Bias, RMSE, and Confidence Probability (CP) for Theta_1 and Theta_2, along with Misclassification Rate (MR) for each specification and N/T combination. All numerical values, including confidence probabilities, are formatted to four decimal places.

First, it is pertinent to discuss the T = 10 case. Here, it seems that the GIFE, IFE and GFE make similarly (un)biased estimates. However, GIFE makes much better group membership estimates. The GIFE's confidence intervals on average contain less of the actual parameter values (5% level). However, considering the GIFE is a nested model of the IFE and the GFE is a nested model of the IFE and GIFE (Section 3.4), it is quite surprising that the GFE makes better estimates than the IFE. The main reason for the IFE's relatively biased estimates is probably due to the model overfitting noise.

When comparing the case of T = 10 to T = 50, it can be observed that the parameter estimates become more accurate due to the longer time dimension (RMSEs reduce) for all specifications, correct or incorrect. One interesting note is that the misclassification rate significantly increases in the GFE algorithm. This would imply that with greater T, the factor structure of the model becomes more prevalent. However, the bias seems to remain relatively unchanged for the GFE. On the other hand, the estimates from IFE and GIFE become more accurate and unbiased, exhibiting statistical consistency properties. GIFE continues making good use of this extra information, keeping its misclassification rate relatively low. This makes its estimates nearly equivalent to the infeasible estimator with known group membership. Furthermore, the GIFE's prediction intervals also outperform the prediction intervals of the other two methods.

3.3 Bootstrapped Standard Errors

Here, this paper details the method used to calculate the standard errors in all the model specifications discussed in the methodology. In the empirical study conducted in this paper, bootstrapping is used to calculate the Standard Errors of the coefficients (Efron & Tibshirani, 1986). Both papers by Bai (2009) and Bonhomme and Manresa (2015) detail analytical formulae for the standard errors of their coefficient estimates. In this paper, the analytical standard error formula is considered only for the GFE. Bootstrapped standard errors are applied in all cases to maintain consistency. The general steps necessary are as follows:

Algorithm 5 Steps for Bootstrapped Standard Errors

1: **Create Bootstrap Sample:** For each iteration, draw a sample of individuals from the original dataset with replacement. This simulates drawing from the population multiple times and helps in approximating the sampling distribution. Make the bootstrapped sample as large as the actual dataset.

2: **Run Full Estimation Process:** For each bootstrapped sample, run the full estimation algorithm. This includes any initialisation and iterative steps required by the specific algorithm being used.

3: **Store Estimates:** Collect the estimated parameters from each bootstrapped sample. These estimates will be used to calculate the variability of the estimator.

4: **Calculate Standard Errors:** Compute the standard errors as the standard deviation of the collected bootstrap estimates.

3.4 Relationship between GFE and GIFE

As a final step in the methodology, it is pertinent to address the relationship between the GFE and GIFE. Specifically in the empirical application considered in Section 4, it is found that the GFE and GIFE, where p is set to high values, p = 4,...,7, yield identical estimates, group memberships and SSR. This is quite interesting because it also implies that when $p \to T$, this may result in $\alpha_{g_i t,GFE} \approx \lambda_{g_i t,GIFE} F_{t,GIFE}$. The rationale behind this can be explained intuitively.

When the θ estimate converges to the same value in the iterative IFE-like Algorithm (Algorithm 3 for known group membership in GIFE) to the coefficient estimate found by the regular GFE (and the factor model is applied in equation 16) the process becomes equivalent to applying a factor structure to $\alpha_{g_i t}$ in equation 1. This is because with the same theta θ and $p \to T$, $W_{g_i,t} \approx \alpha_{g_i t}$. However, the reader should note that setting $p << T$ (e,g: p = 1, 2, 3) leads to different group membership and parameter estimates between the GFE and the GIFE. This also has some important implications for researchers looking to apply these models. The final important point to take into account is that it cannot be guaranteed that setting $p \to T$ leads to the same estimates.

4 Empirical Application

Having analyzed the finite sample statistical properties of the estimators and algorithms, I now consider an empirical application regarding underlying heterogeneity in the study of the effect of income on democracy. The effect of income on democracy has received extensive attention in the literature. Two overarching schools of thought exist in this domain; some believe that democratization is determined by income while others believe that historical events completely determine the same. To provide some background information, I first conduct a literature review. Subsequently, an empirical study is considered where estimates are made with the main models with homogeneous coefficients being considered. An important note is that all results of the GFE model (homogeneous or heterogeneous case) in Section 4.2 are a replication of Bonhomme and Manresa (2015).

4.1 Literature Review for Empirical Application

The application of the effect of income on democracy received widespread exposure in Acemoglu, Johnson, Robinson and Yared (2008). Their main hypothesis was that the strong relationship between income and democracy, found by previous studies, reflected a case of high correlation rather than any causality. By including country-specific fixed effects, this relationship became statistically insignificant. This led them to the conclusion that the observed correlation between income and democracy did not imply any causation. Their paper also accounts for other estimation processes and effects, reinforcing their analysis.

Treisman (2011) and Heid, Langer and Larch (2012) take different approaches to examine this problem using more sophisticated econometric models. Heid et al. (2012) uses a GMM estimation process, finding a significant positive relationship between income and democracy. Their findings are robust across various model specifications and instrument sets. To solidify this result, Treisman (2011) finds that higher income leads to greater democracy in the medium term (10-20 years). However, they find that in the short term, greater income leads to reinforcement of the current political regime. Acemoglu, Naidu, Restrepo and Robinson (2019) also looks at the reverse effect, the effect of democracy on income. After controlling for country-fixed effects and various dynamics of GDP, they find a positive effect of democracy on GDP per capita.

To further the research in this empirical application, Bonhomme and Manresa (2015) propose and utilise the GFE model and its variants to study the effect of income on democracy (models discussed in Section 3.1). Their model specifications allow for a nuanced analysis by applying the GFE estimation process. Their main contribution is modelling heterogeneity in the panel by finding the underlying groups of countries while making consistent parameter estimates. Ultimately, they provide an economic interpretation for the case with G = 4 groups. The first of the 4 regimes is the high democracy group, including countries like the US, Western Europe and Japan. The second group is the low democracy group including many parts of Africa and Asia. The last two groups are the early and late transition groups who transitioned to democracy early or late in the panel. This case of the GFE with 4 groups is considered further in the next section. Therefore, their research played an important role in informing policy as well. It is insufficient to only promote income growth to increase democracy levels. Rather, policymakers must also consider many contextual factors like geography, politics and social factors when promoting democracy. Lu and Su (2017) also look for latent group structures in this empirical application. They identified significant heterogeneity in slope coefficients when accounting for three distinct latent groups.

To conduct the study in the following section, this research utilises the data available on Bonhomme's website (Bonhomme, n.d.), which is the same data used by Acemoglu et al. (2008). This includes data used to perform the GFE, IFE and GIFE estimation process and also data on covariates used to explain the group memberships of the GFE and GIFE (both homogeneous and heterogeneous cases).

4.2 Empirical Analysis

Having discussed a short literature review regarding the empirical application at hand, now this paper introduces the different model specifications which will be studied and compared in this section. The first model specification considered is:

$$demo_{i,t} = \theta_1 demo_{i,t-1} + \theta_2 inc_{i,t-1} + f(it) + \nu_{it} \tag{20}$$

Depending on which model specification is being used, $f(i,t)$ is defined accordingly:

$$f_{\text{IFE}}(i,t) : \lambda_i' F_t \quad \text{(I)}$$
$$f_{\text{GFE}}(i,t) : \alpha_{g_i t} \quad \text{(II)}$$
$$f_{\text{GIFE}}(i,t) : \lambda_{g_i}' F_t \quad \text{(III)}$$

Components	**Objective**	**IC**	**Lag. Dem.**	**Lag Income**	**Cumulative Income**
1	19.8873	**-2.4734**	0.7970 (0.0319)	0.0164 (0.0028)	0.0807 (0.0046)
2	13.8623	-1.8726	0.8674 (0.0213)	0.0112 (0.0019)	0.0841 (0.0053)
3	8.8134	-1.3842	0.9188 (0.0199)	0.0074 (0.0018)	0.0869 (0.0076)
4	5.3391	-0.9646	0.9545 (0.0189)	0.0049 (0.0019)	0.0140 (0.0843)
5	2.8183	-0.7032	0.9850 (0.0156)	0.0025 (0.0015)	0.0232 (1.8322)
6	1.1226	-0.7438	1.0040 (0.0139)	0.0010 (0.0014)	-0.0136 (1.3196)
7	0.0000	-63.7750	1.3684 (0.1667)	-0.2435 (0.0251)	-0.0187 (5.4132)

Table 5: Summary stats for Interactive Fixed Effects. Also, note that all standard errors are bootstrapped.

First, let's discuss the results of the IFE specification. As seen in table 5, parameter estimates change significantly based on the number of factors. Depending on this choice, the effect of income on democracy changes significantly. Firstly, it should be noted that allowing as many factors as there are periods leads to an overfitted model with 0 objective value (p = 7). Therefore, its well-performing Information Criterion (IC) value should be disregarded. Among the specifications with 1 to 6 factors, the specification with only 1 factor performs the best according to the IC. Here, the effect of cumulative income on democracy is highly significant at a 5% level. Therefore, the cumulative income coefficient value implies that an increase of one unit in income leads to a 0.0807 increase in the democracy index. However, the IFE model may not perform well in shorter panels (as discussed in Bonhomme and Manresa (2015)). Also, there is strong evidence of group structures in this empirical application (in the literature). Thus, this paper applies the group-based models (GFE and GIFE) to perform a more nuanced analysis.

Having discussed the results of the IFE model and the shortcomings of its estimates, now it is pertinent to consider the GFE model. From table 6, it is clear that different group sizes lead to varying coefficient estimates. Here, an information criterion can again be applied to decide the optimal number of groups. Considering the BIC plateaus to a relatively stable level between

Groups	Objective	BIC	Lag. Dem.	Lag Income	Cumulative Income
1	24.3012	0.056	0.6654 (0.049)	0.0831 (0.014)	0.247 (0.018)
2	19.8469	0.0432	0.6006 (0.0406, 0.0656)	0.0607 (0.0111, 0.0193)	0.1529 (0.0210, 0.0563)
3	16.5987	0.039	0.4064 (0.0508, 0.1069)	0.0894 (0.0111, 0.0167)	0.1506 (0.0128, 0.0323)
4	14.3187	0.0364	0.3016 (0.0530, 0.1008)	0.0823 (0.0092, 0.0150)	0.1178 (0.0110, 0.0268)
5	12.5933	0.0346	0.2546 (0.0489, 0.0972)	0.0793 (0.0093, 0.0132)	0.1064 (0.0091, 0.0264)
6	11.1317	0.03333	0.4652 (0.0416, 0.1071)	0.0638 (0.0070, 0.0120)	0.1193 (0.0109, 0.0328)
7	10.0589	0.0327	0.4030 (0.0416, 0.1097)	0.0647 (0.0081, 0.0110)	0.1084 (0.0105, 0.0266)
8	9.2514	0.0325	0.3334 (0.0420, 0.1075)	0.0696 (0.0080, 0.0135)	0.1044 (0.0099, 0.0294)
9	8.4260	0.0323	0.3122 (0.0429, 0.1170)	0.0694 (0.0076, 0.0153)	0.1009 (0.0090, 0.0362)
10	7.7491	0.0323	0.2772 (0.0456, 0.1087)	0.0753 (0.0075, 0.0160)	0.1042 (0.0082, 0.0355)

Table 6: Summary stats for Group Fixed Effects. The Group Fixed Effects is applied for Group sizes between 1 to 10. For the estimation of the BIC, this paper applies $G_{max} = 15$. The estimation process uses the 3-Step Iterative GFE Algorithm with a neighbourhood search parameter set at 30. Also note, that the standard errors (in brackets) first use a large T normal approximation using the sandwich estimator and then a bootstrapping method of Section 3.3. Analytical standard errors for cumulative income were found using the delta method. The G = 1 uses only the simple OLS Standard error. To make estimates, 10000 iterations of the 3-step iterative algorithm are used to reach the optimal estimates. The hyperparameter, the number of neighbourhood searches, is set to 10.

G = 4 and 10, G = 4 is chosen as the correct specification (similar to Bonhomme and Manresa (2015)). The G = 4 case is also interesting because the groups formed by this specification can be easily interpreted (discussed later). Furthermore, the estimate here differs significantly from the IFE(p=1) estimate. Even though the GFE is a nested model of the IFE, the GFE model captures a different underlying heterogeneity relative to the IFE model. A final interesting point to make note of is the higher bootstrapped standard error (SE) values relative to the large-T normal approximation SEs (in all cases). When bootstrapped SEs are significantly larger than the analytical SEs, it can indicate the possibility of many issues in the estimation process. The two main reasons for this occurrence in this empirical application may be the following. Firstly, the sample may be too short to assume a large T approximation while calculating the analytical SEs. The other reason may be the dependence of the estimated group structures on specific individual countries. Finally, the reader should notice that when G = 4, the cumulative income estimate is highly significant at the 5% level. An increase of 1 in cumulative income leads to an increase of 0.1178 in democracy.

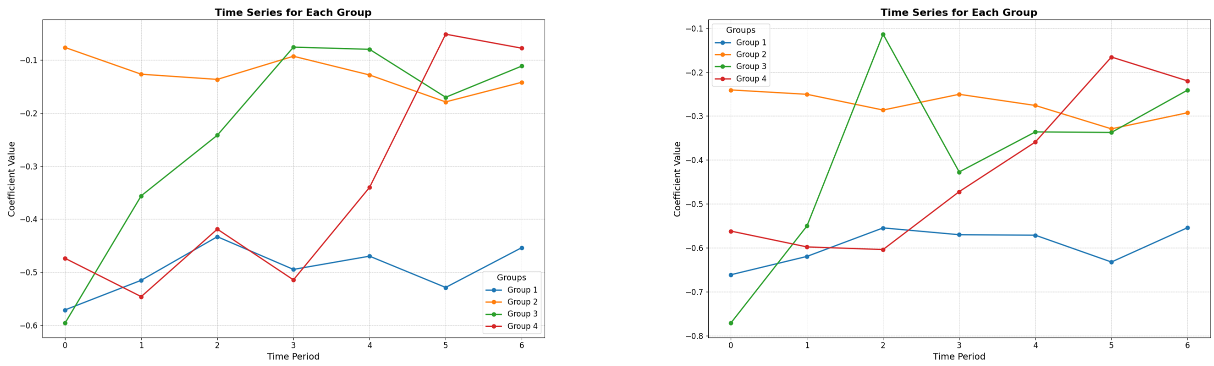
Considering the extra parameters of the IFE model may cause incidental parameter bias and previous literature details the existence of regimes in democracy development, it seems likely that the IFE model is insufficient in its analysis of the problem. Equally, many time-based worldwide shocks occurred between 1960 and 2000 (data collection period). These underlying shocks should be taken into account as well (Fayad, Bates & Hoeffler, 2012). To this end, even the GFE model may be insufficient in its model specifications. Therefore, to further investigate this empirical application, I consider the GIFE model which simultaneously considers the effects of common shocks and underlying groupings on parameter estimates. For this study, this paper only considers the case with 4 groups.

Components	**Objective**	**Lag. Dem.**	**Lag Income**	**Cumulative Income**
1	18.6536	0.4351	0.0552	0.0977
		(0.0557)	(0.0048)	(0.0037)
2	15.7153	0.3884	0.0686	0.1122
		(0.0570)	(0.0041)	(0.0071)
3	14.4129	0.3656	0.0923	0.1455
		(0.0543)	(0.0041)	(0.0074)
4	14.3187	0.3017	0.0822	0.1178
		(0.0554)	(0.0042)	(0.0065)
5	14.3187	0.3017	0.0822	0.1178
		(0.0594)	(0.0049)	(0.0066)
6	14.3187	0.3017	0.0822	0.1178
		(0.0556)	(0.0045)	(0.0065)
7	14.3187	0.3017	0.0822	0.1178
		(0.0567)	(0.0046)	(0.0069)

Table 7: Regression Stats for Group Interactive Fixed Effects. A repurposed version of the 3-Step iterative GFE is used to make estimates here. To make estimates, 10000 iterations of the 3-step iterative algorithm are used to reach the optimal estimates. The hyper-parameter, the number of neighborhood searches is set to 50.

In table 7, the first notable point is that for the GIFE specifications with 4 or more factors (section 3.4), the objective value (SSR) and the parameter estimates do not change, achieving the same value as the regular GFE. Therefore, the GIFE with fewer components may be considered more parsimonious than the GFE, as it fits less to noise. Do note that the computational intensity increases monotonically with more and more factors. Considering the GIFE with $p >= 4$ is equivalent to the GFE in this empirical case, it implies that the GIFE is preferred to the GFE model. To decide on the number of factors, this paper uses an elbow plot (of the objective values). To this end, this paper suggests using p = 3 as the correct specification. For this specific case, it is clear that the cumulative income coefficient is highly significant; one increase in cumulative income leads to a 0.1455 increase in democracy.

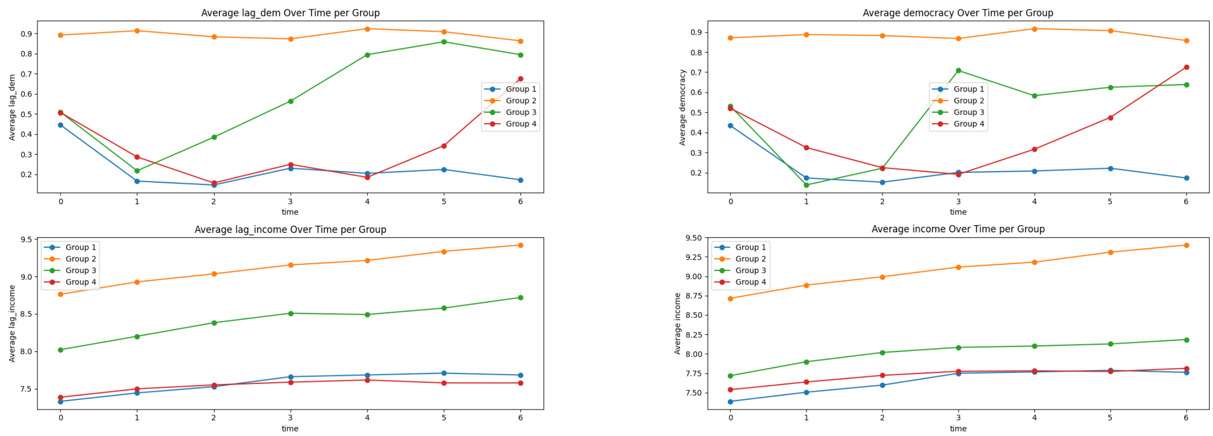
To better understand the nuances between the GFE (G = 3) and GIFE (G = 3, p = 3), this paper presents the $\alpha_{g_i t}$s from the GFE and the $\lambda'_{g_i} F_t$s (which can be considered as the $\alpha_{g_i t}$s from the GIFE) in Figure 1. It is evident that reducing the number of factors equally affects the coefficients of the independent variables and the intercept terms. However, it is important to note that the alphas of both methods seem to follow similar patterns.



(a) GFE for Homogeneous Coefficients

(b) GIFE for Homogeneous Coefficients

Figure 1: Comparison of GFE and GIFE for Non-Heterogeneous Data



(a) Average Democracy and Income GFE Homogeneous Coefficients

(b) Average Democracy and Income GIFE Homogeneous Coefficients

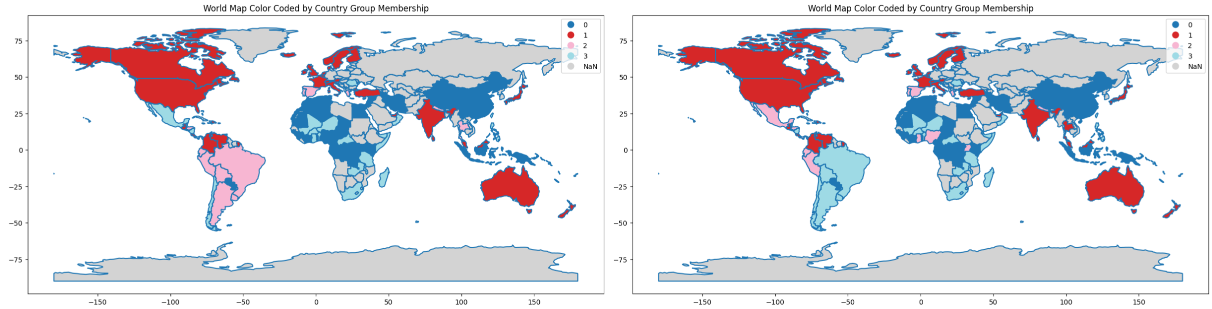
Figure 2: Comparison of Average Democracy (upper panel) and Income (lower panels) for GFE and GIFE Homogeneous Models

From figure 2, it can be observed that the GFE groups exhibit clear regime structures. Group 1 shows a low democratic regime and Group 2 reflects a highly democratic regime. Groups 3 and 4 consist of countries that are early transition countries or late transition countries respectively.

In the GIFE, groups 1 and 2 exhibit the same patterns as in the GFE. Furthermore, groups 3 and 4 follow similar patterns as in the GFE with some nuanced differences. Firstly, Group 3 doesn't seem to reach the same high levels of democracy as in the GFE. Additionally, the late transition group (Group 4) has a more concave structure in its democratic patterns. It also overtakes the early transition group at the end of the panel in terms of democracy. To summarise, Groups 1 and 2 do not change much in their composition while Groups 3 and 4 exchange some members (Figure 3), predominantly in South America.

4.2.1 Heterogeneous Coefficients (GFE and GIFE)

In this section, cases with heterogeneous coefficients are considered for the GFE and GIFE. Please note, only analytical standard errors are considered for the GFE and no standard errors are considered for the GIFE. Calculating bootstrapped standard errors is more complicated



(a) Group Membership GFE (b) Group Membership GIFE

Figure 3: Group Memberships of GFE and GIFE. Dark blue identifies low democracy. Red identifies high democracy. Pink Identifies early transition and Light Blue Identifies Late Transition

with heterogeneous coefficients because matching groups across bootstrapped samples is also necessary here (this is left for further research). Furthermore, only the case with G = 4 is considered for both the GFE and GIFE. Firstly, this paper considers the GFE (heterogeneous) parameter estimates in table 8. The reader should notice that all coefficients are significant at the 5% significance level (for all groups). Do note, that cumulative income increases democracy levels the most in high democracy countries, followed by transition countries and the least in low democracy countries.

Group	**lag_dem**	**lag_income**	**cum_income**
1 (Low Democracy)	0.3192	0.0414	0.0608
	(0.0994)	(0.0153)	(0.0242)
2 (High Democracy)	0.6436	0.0695	0.1950
	(0.0691)	(0.0144)	(0.0553)
3 (Early Transition)	0.0155	0.1216	0.1235
	(0.1109)	(0.0236)	(0.0277)
4 (Late Transition)	0.2479	0.0903	0.1201
	(0.0933)	(0.0185)	(0.0288)

Table 8: Parameter Estimates and Sandwich Estimator Standard Errors for the GFE with Heterogeneous Coefficients. To make estimates, 10000 iterations of the 3-step iterative algorithm are used to reach the optimal estimates. The hyperparameter, the number of neighborhood searches is set to 10.

Next, this paper presents the SSEs for the GIFE (p = 1,..., 7). in table 9. Using the elbow plot (of the objective values), this paper chooses p = 3 as the correct specification. Another point to note is that the SSR value for the GIFE with $p \geq 4$ is the same as the regular GFE, as are the estimates. Also, for $p \geq 3$, the GIFE and the GFE produce the same group membership estimates.

Components	1	2	3	4	5	6	7
Best SSR	15.3925	14.0243	13.7383	13.5395	13.5395	13.5395	13.5395

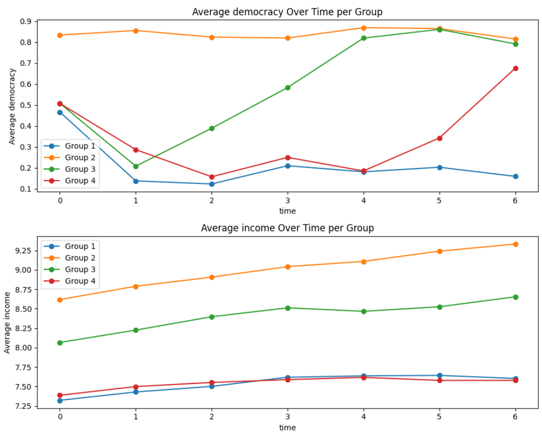
Table 9: Sum of Squared Residuals (SSR) for Different Number of Components

In table 10, this paper presents the coefficient estimates of the GIFE(p = 3). First, it is important to mention that the group membership estimate remains the same in GIFE(p = 3) and the GFE. Notice that no standard errors are included. The reasoning for this is given in earlier in this section. Another important point is that for high democracy and late transition groups, the coefficient estimates of lagged democracy and lagged income remain quite similar between the GFE and GIFE. Furthermore, all but group 1 experience similar coefficient estimates for cumulative income as well. Finally, the groups also seem well separated when considering the different coefficient estimates. Once again, cumulative income increases democracy levels the most in high democracy countries, followed by transition countries and the least in low democracy countries.

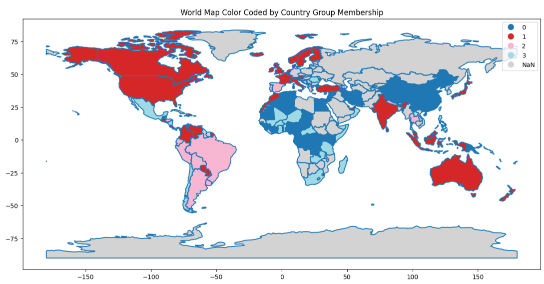
Group	**lag_dem**	**lag_income**	**cum_income**
1 (Low Democracy)	0.2470	0.0334	0.0444
2 (High Democracy)	0.6453	0.0685	0.1931
3 (Early Transition)	0.0387	0.1240	0.1289
4 (Late Transition)	0.2463	0.0898	0.1192

Table 10: Parameter Estimates for the GIFE with Heterogeneous Coefficients. Note that no standard errors have been considered. Analytical standard errors have not been derived in this paper and bootstrapped standard errors necessitate matching groups over iterations which can be quite difficult and is considered out of the scope of this paper. To make estimates, 10000 iterations of the 3-step iterative algorithm are used to reach the optimal estimates. The hyperparameter, the number of neighbourhood searches is set to 10.

As a final note, in figure 4, it is observed that the group memberships change marginally in the heterogeneous case relative to the homogeneous cases (refer to figure 3). This is reflected by comparing the average income and average democracy time series plots and the group membership maps in tables 3 and 4.



(a) Average democracy (upper panel) and income (lower panel) in



(b) GFE and GIFE Group 4 Heterogeneous Countries

Figure 4: Group membership in the heterogeneous case

4.2.2 Explaining Group Memberships (Using Multinomial Logit Models)

A final point of exploration in the empirical application is to explain the underlying drivers of group membership in the different models. As discussed in Bonhomme and Manresa (2015), applying probability models (multinomial logit or ordered probit) to the groups found in the GFE is asymptotically equivalent to estimating the same probability models with the actual underlying groups. Although a similar proof has not been derived for the GIFE, this paper still applies the multinomial logit models to the groups estimated (in the GIFE). Once again, data from Bonhomme's (Bonhomme, n.d.) personal website is used. The covariates used are detailed in table 11. This reflects the third model specification in the logistic regressions applied in Bonhomme and Manresa (2015). Most of the covariates used have self-explanatory names. The variable "Constraints" denotes the constraints placed on the executive at independence.

Regressor	GFE Homogeneous	GIFE Homogeneous	GFE/GIFE Heterogeneous
Group 2 (High Democracy)			
log GDP (1500)	0.6979	−0.8000	1.2715
	(1.757)	(1.782)	(1.409)
Independence Year	−0.0444	−0.0496	−0.0311
	(0.013)	(0.013)	(0.009)
Constraints	7.1244	8.1204	5.0699
	(2.058)	(2.222)	(1.520)
Catholic Share (1980)	0.6114	0.6034	0.5401
	(1.202)	(1.193)	(1.224)
Protestant Share (1980)	6.8106	7.2699	4.1903
	(4.375)	(4.590)	(3.693)
Group 3 (Early Transition)			
log GDP (1500)	−0.5038	−3.5386	−0.3917
	(1.872)	(1.765)	(2.138)
Independence Year	−0.0332	−0.0297	−0.0429
	(0.012)	(0.011)	(0.020)
Constraints	2.2263	3.7385	2.3871
	(2.335)	(2.013)	(2.914)
Catholic Share (1980)	1.0034	1.8197	1.7326
	(1.218)	(1.299)	(1.559)
Protestant Share (1980)	−0.5518	3.1657	−23.2209
	(7.875)	(5.715)	(27.725)
Group 4 (Late Transition)			
log GDP (1500)	−0.7508	−1.2711	−0.6130
	(1.138)	(1.268)	(1.127)
Independence Year	−0.0078	−0.0119	−0.0077
	(0.008)	(0.007)	(0.008)
Constraints	0.8476	1.4156	0.7985
	(1.389)	(1.529)	(1.406)
Catholic Share (1980)	0.8877	1.0345	1.0565
	(1.188)	(1.151)	(1.258)
Protestant Share (1980)	5.3988	5.1700	3.6080
	(3.874)	(4.105)	(3.489)

Table 11: Parameter Estimates and Standard Errors for Different Models. In all cases, coefficients are relative to Group 1, the Low-Democracy groups produced by each of the models. Please note that intercepts were also included in all cases, but estimates are not included in the results

The main point to observe in table 11 is that coefficient signs stay consistent across all models. This implies that all (comparable) regimes in all models show similar relative patterns and reactions to the underlying determinants of their estimated groups. Another important point that should be noted is that religion seems to play a negligible role in explaining the estimated groups in all cases (not statistically significant estimates). This is also the case for log GDP. Furthermore, constraints on the executive at independence only seem to play a relatively important role in explaining the group membership in the high democracy groups (relative to the corresponding low democracy groups). Therefore, earlier independence years lead to a higher chance of being in the high democracy group relative to the low democracy group. Independence Year also has a considerable (relative) effect on the high democracy and early transition groups. However, this result may be somewhat surprising since earlier independence seems to lead to a lower chance of being in the higher democracy or early transition groups (relative to being a part of the low democracy group). A final interesting point to take note of is that the groups created by the heterogeneous model and the heterogeneous GIFE have more extreme coefficients in their regression. This implies that the covariates used in the regression can better explain the underlying grouping structures in those models. This may imply that these models are better able to take underlying signals into account in comparison to the regular GFE.

4.3 Conclusion and Discussion of the Empirical Study

To conclude this empirical study, this paper finds that greater cumulative income has a significant effect on democracy in all specifications studied. By considering group-based models, meaningful underlying regimes have been found in this paper. In the cases where 4 groups are considered, the regimes found are (1) low democracy, (2) high democracy, (3) early transition and (4) late transition.

As found in the previous sections, the GFE is a special case of the GIFE. However, by setting $p << T$ number of factors, the GIFE fits less to the noise, allowing for more parsimonious model estimations. Future research should attempt to find covariates which explain the common shocks and the heterogeneous factor loadings found by the GIFE variants. It is also important to note that considering heterogeneous coefficients allows a more specific analysis of how cumulative income affects democracy in each group. It is also notable that group membership changes significantly in the transition groups when considering the GFE and GIFE in the homogeneous case. On the other hand, group membership stays the same when considering the GFE and the GIFE heterogeneous case. By considering the GIFE and GFE, small nuances change.

Furthermore, by considering external covariates, underlying drivers of group membership are also found. In all cases, income and religion are found to be statically insignificant. On the other hand, independence year and constraints placed on the executive at independence seem to explain the underlying grouping structures quite well. Considering the significant coefficients of cumulative income and the underlying determinants of the groups, both are important determinants of democracy.

5 Conclusion and Discussion

To conclude, this paper has studied the theory and algorithms of various models, discussed their various uses and applied the methods to an empirical application. It considered the least squares estimation of the GFE, IFE and the novel GIFE. The models all aim to model underlying heterogeneity in panel data using unconventional methods to make better coefficient estimates. The GFE considers underlying groups and the IFE considers common shocks and heterogeneous reactions. The GIFE considers both types of effects in its estimations. The GIFE model aims to add to the literature by considering both underlying groups and common shocks which affect the whole sample space (with heterogeneous group-based reactions).

In Section 3, the methodology is discussed. First, the Group Fixed Effects model is detailed in Section 3.1. Here, the formulae and model specifications are detailed. Then, the iterative algorithms used to estimate this model are also specified. Here the Three-Step iterative GFE is proposed as an algorithmic extension. A Monte Carlo simulation is also considered to test the statistical properties of all the estimation models. Here, it is found that if speed is desired (in cases where G, the number of groups is small), the GFE should be used. In the case where a large G (number of groups) is postulated, the Three-step GFE is both quicker and makes equally good (if not better) group membership estimates. In the following part of the methodology (Section 3.2), the Interactive Fixed Effects model and the Grouped Interactive Fixed Effects models are considered. Once again, the model specifications, restrictions and estimation processes are considered. To this end, this paper modifies the IFE estimation process to make estimates for the GIFE (with known group membership). Integrating this process into a modified GFE algorithm (this can be any one of the basic GFE, VNS or Three-Step Algorithms) leads to a complete GIFE estimation process. Finally, a Monte Carlo simulation is considered to test the estimates of the infeasible estimator (when group membership is known beforehand), where the method is shown to be valid. These estimates are compared to the estimates of the complete GIFE estimation algorithm (iterative process) to consider the accuracy of the complete estimation process. This paper also studies the effect of model misspecification when making parameter estimates when the underlying DGP follows the GIFE.

Finally, an empirical study is considered to apply the different methods considered in this paper (Section 4). The main finding is that cumulative income does affect democracy heterogeneously in different groups (significant coefficients found in all cases). Furthermore, it is also found that historical events (independence year and constraints at independence) can explain the underlying groups which are found. Therefore, historical events indirectly also affect democracy.

The final point of discussion in this paper is the areas for further research. First of all, an information criterion (to specify the number of groups and factors) and an analytical standard error formula should be developed for the GIFE. An information criterion like Ando and Bai (2016) should be considered for this use. Furthermore, covariates should be considered to explain the common shocks and factor loadings of the GIFE (specifically in empirical studies like the one considered in this paper). Further research should also develop more efficient algorithms to estimate the GIFE.

References

Acemoglu, D., Johnson, S., Robinson, J. A. & Yared, P. (2008). Income and democracy. *The American Economic Review*, *98*(3), 808–842. Retrieved from https://www.jstor.org/stable/29730096 doi: 10.1257/aer.98.3.808

Acemoglu, D., Naidu, S., Restrepo, P. & Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, *127*(1), 47–100. doi: 10.1086/700936

Ando, T. & Bai, J. (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, *31*(1), 163–191. Retrieved from https://onlinelibrary.wiley.com/doi/epdf/10.1002/jae.2467 doi: 10.1002/jae.2467

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, *77*(4), 1229–1279. doi: 10.3982/ECTA6135

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

Bonhomme, S. (n.d.). *Stephane bonhomme's research.* https://sites.google.com/site/stephanebonhommeresearch/. (Accessed: 1/7/2024)

Bonhomme, S., Lamadon, T. & Manresa, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, *90*(2), 625–643. doi: 10.3982/ECTA15238

Bonhomme, S. & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, *83*(3), 1147–1184. doi: 10.3982/ECTA12363

Browning, M. & Carro, J. M. (2014). Dynamic binary outcome models with maximal heterogeneity. *Journal of Econometrics*, *178*(3), 805–823. doi: 10.1016/j.jeconom.2013.11.005

Cieslak, A. & Pang, H. (2021). Common shocks in stocks and bonds. *Journal of Financial Economics*, *142*(2), 880–904.

Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*(1), 54-75. Retrieved from https://www.jstor.org/stable/2245500

Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. doi: 10.1016/0304-405X(93)90023-5

Fayad, G., Bates, R. H. & Hoeffler, A. (2012). *Income and democracy: Lipset's law inverted* (Research Paper No. 61). University of Oxford: Oxford Centre for the Analysis of Resource Rich Economies. Retrieved from https://scholar.harvard.edu/files/robertbates/files/income$_a$*nd*$_d$*emocracy.pdf*

Hahn, J. & Moon, H. R. (2010). Panel data models with finite number of multiple equilibria. *Econometric Theory*, *26*, 863–881. doi: 10.1017/S0266466609990132

Heid, B., Langer, J. & Larch, M. (2012). Income and democracy: Evidence from system gmm estimates. *Economics Letters*, *116*(2), 166–169. doi: 10.1016/j.econlet.2012.02.022

Hsiao, C. (2003). *Analysis of panel data.* Cambridge: Cambridge University Press.

Hsiao, C. & Tahmiscioglu, A. K. (1997). A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association*, *92*(438), 455–465. Retrieved from https://www.jstor.org/stable/2965693

Huang, W., Jin, S. & Su, L. (2020). Identifying latent grouped patterns in cointegrated panels. *Econometric Theory*, *36*(3), 410–456. Retrieved from https://ink.library.smu.edu.sg/soe$_r$*esearch*/2381 doi: 10.1017/S0266466619000197

Kasahara, H. & Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, *77*, 135–175. doi: 10.3982/ECTA6763

Lai, T. L. & Xing, H. (2008). *Statistical models and methods for financial markets.* New York, NY: Springer. Retrieved from https://link.springer.com/book/10.1007/978-0-387-77827-3 doi: 10.1007/978-0-387-77827-3

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

Lu, X. & 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. & Wang, W. (2023). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, *233*, 45–65. Retrieved from https://www.elsevier.com/locate/jeconom

Okui, R. & Wang, W. (2018). Heterogeneous structural breaks in panel data models. *Journal of Econometrics*, *220*(2), 447–473. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304407619300875 doi: 10.1016/j.jeconom.2019.04.009

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

Shen, X. & Huang, H. C. (2010). Grouping pursuit through a regularization solution surface. *Journal of the American Statistical Association*, *105*(490), 727–739.

Su, L. & Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics*, *206*(2), 554–573. doi: 10.1016/j.jeconom.2018.02.009

Su, L., Shi, Z. & Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, *84*(6), 2215–2264. doi: 10.3982/ECTA12731

Su, L., Wang, X. & Jin, S. (2018). Sieve estimation of time-varying panel data models with latent group structures. *Journal of Econometrics*, *206*(1), 189–215.

Treisman, D. (2011). Income, democracy, and the cunning of reason. *ERN: Institutions & Political Power in Transitional Economies (Topic)*(17132). doi: 10.3386/w17132

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

Wang, Y., Phillips, P. C. B. & Su, L. (2023). Panel data models with time-varying latent group structures. *arXiv preprint arXiv:2307.15863*. Retrieved from https://arxiv.org/abs/2307.15863v1

Xu, R., Gao, J., Oka, T. & Whang, Y.-J. (2022). Estimation of heterogeneous treatment effects using quantile regression with interactive fixed effects. *Monash Econometrics and Business Statistics Working Papers, 13/22*.

A Algorithms

A.1 Variable Neighborhood Search Algorithm

Algorithm 6 Variable Neighborhood Search

1: **Initialization:**

- Let $(\theta, \alpha) \in \Theta \times A^{GT}$ be some starting value.
- Perform one assignment step of Algorithm 1 and obtain an initial grouping γ_{init}.
- Set iter_{max} and $\text{neigh}_{\text{max}}$ to some desired values.
- Set $j = 0$.
- Set $\gamma^* = \gamma_{\text{init}}$.
- Set $n = 1$.

2: **while** $n \leq \text{neigh}_{\text{max}}$ **do**
3: (Neighborhood jump) Relocate n randomly selected units to n randomly selected groups, and obtain a new grouping γ'.
4: Perform one update step of Algorithm 1 and obtain new parameter values (θ', α').
5: Set $(\theta^{(0)}, \alpha^{(0)}) = (\theta', \alpha')$, and apply Algorithm 1.
6: (Local search) Starting from the grouping $\gamma = \{g_1, \dots, g_N\}$ obtained in Step 4, systematically check all reassignments of units $i \in \{1, \dots, N\}$ to groups $g \in \{1, \dots, G\}$ (for $g \neq g_i$), updating g_i when the objective function decreases; stop when no further reassignment improves the objective function.
7: Let the resulting grouping be γ''.
8: **if** the objective function using γ'' improves relative to the one using γ^* **then**
9: Set $\gamma^* = \gamma''$ and go to Step 2.
10: **else**
11: Set $n = n + 1$ and go to Step 7.
12: **end if**
13: **end while**
14: Set $j = j + 1$. If $j > \text{iter}_{\text{max}}$, then Stop; otherwise go to Step 2.

A.2 Three-Step GFE Heterogeneous Case

Algorithm 7 Three-Step Iterative GFE Algorithm, Heterogeneous coefficients case

1. **Apply Algorithm 1 of Bonhomme and Manresa (2015):**

   (a) Initialize based on random or specific starting points and estimate:

   $$y_{it} = x_{it}'\theta_{g_i} + \alpha_{g_i t} + \epsilon_{it}$$

   (b) After numerical convergence, store the sum of squared errors (SSE), parameters $(\theta_1, ..., \theta_G, \alpha)$, and group memberships g_i for all $i = 1, \ldots, N$.

2. **Iterative Two-Step GFE (Bonhomme and Manresa, 2021):**

   (a) Define function h_i for mapping underlying groupings:

   $$h_i(y_i, x_i') = y_i - x_i'\theta$$

   Where:

   $$y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}, \quad x_i' = \begin{bmatrix} x_{i1}' \\ \vdots \\ x_{iT}' \end{bmatrix}, \quad \theta = \sum_{i=1}^{G} \frac{\text{number in } g_i}{N} \theta_{g_i}$$

   (b) Set iteration counter $iter = 0$ and neigh_max to set the number of iterations.

3. **Classification and Estimation Steps:**

   (a) Perform the classification step as per equation 4 from Bonhomme and Manresa (2021) to update groupings g_i for all i (this is Lloyd's/K-means algorithm).

   (b) Minimize SSE in the estimation step analogous to their equation 5:

   $$(\theta_1, ..., \theta_G, \alpha) = \arg \min_{\theta, \alpha \in \Theta \times A^{GT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}'\theta_{g_i} - \alpha_{g_i t})^2$$

4. **Check for Improved SSE:**

   (a) If SSE is lower than the previous best, update stored $(\theta_1, ..., \theta_G, \alpha)$ and g_i for all i, and go back to step 1 with these new values as initialization.

   (b) Otherwise, if $iter \geq$ neigh_max, end the search; else increment $iter$ by 1, retain the previous best $(\theta_1, ..., \theta_G, \alpha)$, and return to step 3.

A.3 Interactive Fixed Effects Algorithm

Algorithm 8 Interactive Fixed Effects Estimation Algorithm

1. **Initialization:**
   - Set $s = 0$.
   - Initialize $\beta^{(0)}$ using the within-group estimator.
2. **Compute Residuals:**
   - Set $W_{it} = y_{it} - x'_{it}\beta^{(s)}$.

   This yields a pure factor model:
   $$W_{it} = \lambda'_i F_t + \epsilon_{it} \tag{21}$$
3. **Principal Component Analysis:**
   - Apply Principal Component Analysis (PCA) on the factor structure to find the factor loadings $\hat{\lambda}_i^{(s)}$ and common shocks $\hat{F}_t^{(s)}$.
4. **Update β:**
   - Considering $\hat{\lambda}_i^{(s)}$ and $\hat{F}_t^{(s)}$ as "observable" at this stage, update β by applying Ordinary Least Squares (OLS):
   $$\beta^{(s+1)} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X'_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\left(Y_{it} - \hat{\lambda}_i^{(s)}\hat{F}_t^{(s)\prime}\right)\right)$$
5. **Check Convergence:**
   - Stop if numerical convergence is achieved. Otherwise, set $s = s + 1$ and go back to step 2.

B Simulation Settings

B.1 Simulation Specifics For Algorithm Monte Carlo Simulation in Section 3.1.3

- for each Monte Carlo simulation of the 8 different experimental settings and each algorithm, 100 new datasets are generated.
- for each of the Monte Carlo iterations (new data set), the simple algorithm applies only 10 iterations of each algorithm. The best estimate of these 10 iterations (lowest SSR) is used to collect all the necessary data.
  - Ideally, at least 100 iterations would be performed per newly generated dataset for each algorithm such that each algorithm is given a better chance to achieve its potential. However, due to limitations in computational resources, these specifications were used.
- For the pre-specified parameters of the algorithms, the VNS algorithm uses a constant value of neighborhood_search = 10 and a maximum of 10 iterations per neighbourhood search. For the 3-step Iterative GFE, I allow for the neighborhood_search parameter to equal 30.

C Programming code

C.1 GroupedInteractiveFixedEffects.ipynb

This file first downloads the data and preprocesses it (certain dictionaries and lists which are inputs to following methods). The main methods defined here are the IFE (which is retrieved from a package) and a bootstrap for it. Then a intialisation process is programmed for the GIFE. Afterwards, the GIFE for known group membership is coded using the GroupedInteractiveFixedEffects class. This class contains an important fit method which repurposes the IFE code from the package to make its estimations. Afterwards, I also have methods to run a repurposed version of Basic GFE algorithm and the Three Step Iterative algorithm (called iterative_part() method) for the GIFE estimation. Finally some plots are made for the GIFE method in the in the empirical application and a bootstrapping is used for standard errors. Finally, a Monte Carlo Simulation is also considered.

C.2 GIFEHeterogeneous.ipynb

This does the same thing as the regular grouped interactive fixed effects method, but with heterogeneous coefficients.

C.3 MyAlgoWithImprovementsHeterogeneous.ipynb

Estimates the heterogeneous counterpart to the GIFE (similar code).

Also includes the Monte Carlo Simulation of the GIFE when GIFE is the right DGP.

C.4 MyAlgoWithImprovements.ipynb

Estimates the homogeneous counterpart of the GFE to the GIFE (similar code).

Also includes the Monte Carlo Simulation of the GFE when GIFE is the right DGP.

C.5 IncorrectIFE.ipynb

Monte Carlo simulation of the IFE model when the GIFE is the correctly specified underlying DGP.

C.6 Algorithm2ImprovedEfficiency.ipynb

Variable neighborhood search algorithm for the homogeneous GFE case.

C.7 Logit.ipynb

In this program, I estimate the multinomial logit regressions of the empirical application. I have hard-coded the group memberships after estimating these from their respective files. The data used here are the covariates collected from Bonhomme's website.