GROUPED PATTERNS IN PANEL DATA

by

BINZHI CHEN

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics
Birmingham Business School
College of Social Sciences
University of Birmingham
October 2025

Contents

List of Tables

List of Figures

Acknowledgements

It is finally here. This PhD journey has been a remarkable experience, and I owe a deep debt of gratitude to many individuals who have supported me all along the way.

First and foremost, I would like to express my deepest appreciation to my supervisors, Dr.Marco Barassi and Professor Yiannis Karavias, and my teaching mentor, Yi Liu. Their support and constructive feedback have been invaluable throughout my research. Your belief in my work has inspired me to push the boundaries of my knowledge and to strive for excellence. You have not just been a supervision team, but we have also become friends. I would like to express my sincere gratitude to my internal examiner, Dr.Marco Ercolani, and my external examiner, Professor Vasilis Sarafidis, for their insightful feedback. Their careful reading and thoughtful suggestions have significantly improved the quality and clarity of this thesis. I am deeply appreciative of the time and effort they dedicated to this examination process.

I would also like to thank the University of Birmingham and the Department of Economics. It was my honour to receive a full scholarship for my PhD after graduating from the MSc Financial Economics programme at the University of Birmingham. The day I accepted the offer was a life-changing decision for me. Otherwise, I would have been on a different life path I would definitely regret.

To my parents. Thanks for their support for my entire academic journey. They have loved me unconditionally since I was born. They are everything to me. Thanks them for giving me life. Thanks them for being my parents.

To Li Xiang, she is my best friend. I thank her for the invaluable companionship since the day we met. Thank her for enriching my life with friendship and for always being there when I needed it most. She is more than a friend to me. She is my family. I shall protect her till the end of the world.

To all my friends and family members, I would like to express my gratitude to the following people for their support:

- Di Chen
- Yiqing Gao
- Yi Han
- Botong Liu
- Zihui Ma
- Xi Niu
- Jin Ren
- Xueqing Sun
- Dongliang Wang
- Fangqing Wu
- Jinye Wu
- Jiani Yan
- Jiaqi Yang
- Dawei Zhao
- Jiani Zhao
- Lei Zhao
- Fanchao Zheng

To Yan Gu. Thank you for having accompanied me. This is the thing that I promised you. I owe you so much.

Thank you all for your contributions to this work and for being a part of this journey with me.

Chapter 1: Introduction

1.1 Research Background

Due to the increasing availability of large datasets, many studies can obtain data across two dimensions: the number of cross-sectional units (N) and the number of time periods (T). This type of data, known as panel data (Hsiao 2014), offers a rich source of information by combining cross-sectional and time-series components. Panel data allows researchers to account for both inter-individual and intra-individual variations, offering advantages over purely cross-sectional or time-series data. However, one of the main challenges in panel data econometrics is the assumption of homogeneity across units, which may not hold in real-world settings (Pesaran & Yamagata 2008). In many applications, the literature assumes that the relationships between variables are the same for all units in the dataset, leading to potential biases when this assumption does not hold. One of the main sources of heterogeneity in panel data is the presence of unobserved factors that, although not directly measured, still affect the data-generating process. These unobserved factors, when not properly accounted for, can lead to omitted variable bias, distorting the estimated relationships between the variables of interest. To address this issue, researchers commonly employ fixed effects models. These models are designed to capture unobserved heterogeneity by allowing intercepts to vary across either cross-sectional units (i.e., individual fixed effects that vary across N) or time periods (i.e., time fixed effects that vary across T). Fixed-effects models are particularly useful useful when there are unobserved factors that remain constant over time or across individuals but differ between them. For instance, in studies involving countries, individual fixed effects can control for country-specific characteristics that remain constant over time. Examples include geography or cultural traits. In contrast, time fixed effects account for global shocks that affect all countries simultaneously. However, individual fixed effects models become increasingly inefficient as the number of cross-sectional units grows large. Estimating heterogeneity at the individual level assumes that each unit possesses unique unobserved characteristics, which requires the model to estimate a separate intercept for each individual. This approach becomes computationally intensive with large datasets and may lead to overfitting, where the model captures noise instead of meaningful patterns. Additionally, individual-level fixed effects may fail to identify broader patterns that emerge when groups of units share similar characteristics or experiences.

Instead of estimating fixed effects for each individual firm, the group-level fixed effects approach allows researchers to model heterogeneity at the group level, where groups consist of units with similar characteristics. This not only reduces the number of parameters to estimate but also provides a more structured and interpretable model. Group structures are particularly useful in large datasets, where individual-level heterogeneity may obscure broader trends. For example, in cross-country studies at the macroeconomic level, countries in different regions or with similar income levels may experience comparable economic shocks, such as a global financial crisis or commodity price fluctuations. Grouping these countries together allows for a more parsimonious model that accounts for shared economic conditions. Similarly, in labour economics, datasets containing millions of individuals may involve workers employed in related industries or sectors, sharing similar wage dynamics or labour market trends. Group-level fixed effects can capture these common patterns, enabling more efficient estimation while preserving the model's ability to account for individual differences. Thus, a substantial body of literature has discussed the possibility of group structures in panel data models (Bester & Hansen 2013, Bonhomme & Manresa 2015). In this context, the Grouped Fixed Effects (hereafter GFE) framework proposed by Bonhomme & Manresa (2015) offers an attractive modelling strategy. By endogenously identifying a small number of latent groups that share fixed effects, GFE captures unobserved heterogeneity in a parsimonious yet interpretable form. This approach strikes a balance between fully pooled models and models with unit-specific fixed effects, making it especially suitable for large N settings where structural similarities among units are plausible. Importantly, the GFE framework also facilitates the detection of economically meaningful clusters, which is particularly relevant in macroeconomic datasets.

The GFE estimator, proposed by Bonhomme & Manresa (2015), originates from the first paper to discuss the unknown grouped structure and its inference in panel data models. Unlike traditional fixed effects models, which estimate unit-specific intercepts to capture heterogeneity, the GFE framework introduces a different structure. It assumes that cross-sectional units, such as countries or firms, can be partitioned into a finite number of latent groups. Each group then shares a common fixed effect. This structure reflects the plausible economic intuition that certain units may exhibit similar structural behaviours over time. The GFE estimator further allows for group-specific effects to vary over time ($\alpha_{g_i,t}$) enabling dynamic modelling of heterogeneity at the group level. Estimation proceeds by minimising the sum of squared residuals across the panel using an iterative procedure that alternates between estimating group-specific parameters and reassigning units to groups, typically via a K-means clustering algorithm. This approach not only reduces the computational burden associated with classical fixed effects in high-dimensional panels but also offers a more structured and parsimonious way to address unobserved heterogeneity. By grouping units with similar characteristics, GFE improves the model's capacity to detect common patterns and trends that might be obscured in fully heterogeneous settings. Moreover, under regularity conditions and as both the cross-sectional dimension N and the time dimension T grow large, the GFE estimator achieves consistent group classification and asymptotically normal parameter estimates. It is also important to note that the GFE approach does not directly address the Nickell bias (Nickell 1981) commonly associated with dynamic panel models. Instead, it avoids the issue by assuming that unit-specific fixed effects can be replaced by group-specific intercepts. Since the correlation between lagged dependent variables and individual fixed effects is the source of Nickell bias, the absence of such individual-specific terms in the GFE framework effectively sidesteps the problem, rather than correcting it. This method forms the foundation for both the empirical and methodological analyses in this thesis. In particular, it is extended to dynamic panel settings to uncover grouped patterns in macroeconomic relationships, such as the inequality-growth nexus. It is also applied to a Panel Vector Autoregressive (PVAR) model with unknown group structures to examine innovation dynamics.

In applied econometrics, the increasing availability of large panel datasets poses both opportunities and challenges for modelling heterogeneity. The growing recognition of the advantages of group structures has led to their widespread adoption in empirical studies across various fields. There has been substantial empirical work utilising grouped fixed effects to accurately identify group structures within the data when the true group membership is unknown or unclear. For instance, Gómez-Puig et al. (2022) applied group structures to analyse the heterogeneous relationship between public debt and economic growth, finding that the effect of public debt on growth differs significantly across groups of countries with similar economic characteristics. Similarly, Grunewald et al. (2017) examined the relationship between income inequality and carbon dioxide emissions using grouped fixed effects, identifying distinct group-level patterns that explain variations in emissions across countries with similar income levels. These examples highlight the potential of GFE models to provide more nuanced insights into economic relationships by accounting for group-level heterogeneity. Alongside empirical applications, substantial progress has been made in the theoretical econometrics literature to further explore the potential of group structures in panel data models. Beyond the Grouped Fixed Effects (GFE) framework, a growing body of econometric research has developed alternative approaches to modelling grouped slope-parameter heterogeneity in panel data settings. These methods can be broadly classified into clustering-based and penalisation-based techniques. Clustering-based methods, such as those proposed by Sarafidis & Weber (2015) and Bonhomme & Manresa (2015), employ algorithms like K-means to assign units to latent groups based on the similarity of their estimated coefficients. Penalisation-based approaches, most notably the Classifier-Lasso (C-Lasso) developed by Su et al. (2016), achieve joint estimation and classification by shrinking individual coefficients towards group-specific targets via penalised likelihood criteria. Further developments in this literature have incorporated interactive fixed effects and factor structures (Ando & Bai 2016, Su & Ju 2018), threshold dynamics (Miao, Su & Wang 2020), and structural breaks (Okui & Wang 2021), thereby enabling these models to address unobserved common shocks alongside grouped slope heterogeneity. Although these approaches differ in estimation strategies and underlying assumptions, they share the goal of uncovering latent heterogeneity that may be obscured in traditional fixed effects models. This thesis builds upon this literature by adopting and extending the GFE framework, owing to its computational tractability and flexibility in capturing grouped structures in both static and dynamic macroeconomic panel contexts.[3]

1.2 Research Objectives

This PhD thesis consists of three chapters related to group structures. It explores both the empirical and theoretical dimensions of group structures in econometrics, aiming to provide a comprehensive understanding of their application and theoretical inference. Chapter 2 provides an empirical analysis on how group structures affect the relationship between economic growth and income inequality across countries. Chapters 3 and 4 respond to these empirical findings by developing new theoretical frameworks that address the limitations identified in the empirical analysis. Chapter 3 proposes a generalised multivariate model for grouped patterns, offering theoretical improvements that enhance the flexibility and interdependence of existing methods. Chapter 4 further focuses on robustness and the application of the Granger causality test in grouped structures. Together, these chapters contribute to a more comprehensive understanding of group structures in econometrics, demonstrating how empirical evidence can inform theoretical advancements and, conversely, how improved theoretical models can enhance empirical applications.

In line with the broader literature using grouped fixed effects models, Chapter 2

[3] A more detailed review of the relevant grouping literature is provided in subsequent chapters.

makes an empirical contribution by re-examining a well-known economic relationship under a grouped structure framework. Chapter 2: Grouped Patterns in Economic Growth and Income Inequality revisits the long-debated nexus between economic growth and income inequality through the lens of grouped heterogeneity. Since the seminal work of Kuznets (1955), who proposed that the relationship between inequality and growth follows an inverted U-shaped curve (the Kuznets curve), researchers have sought to confirm or refute this hypothesis. The Kuznets curve suggests that in the early stages of economic development, inequality tends to rise as growth increases, but as economies mature, inequality declines as growth continues. Despite substantial empirical investigation, findings remain inconclusive and contested. Some studies have found a negative relationship between income inequality and economic growth (Alesina & Rodrik 1994, Castelló-Climent 2010, Khalifa et al. 2010), while others have reported a positive one (Li & Zou 1998, Forbes 2000). More recent research has suggested that the relationship might be non-linear, potentially reconciling the conflicting findings (Barro 2000, Cho et al. 2014, Abebe & Ratbek 2020). In this Chapter, we aim to contribute to this ongoing debate by re-examining the growth-inequality nexus using grouped heterogeneity to account for differences across countries. We employ the Grouped Fixed Effects (GFE) estimator proposed by Bonhomme & Manresa (2015), which allows for the identification of latent grouped patterns and group-specific effects within the data. The GFE model is particularly suited to addressing the issue of unobserved heterogeneity, as it enables the grouping of units that share similar characteristics while maintaining flexibility in capturing group-specific dynamics. Our analysis is based on a worldwide panel dataset consisting of 92 economies, covering the period from 1998 to 2017. The contributions of this chapter are threefold. First, we explore the non-linear relationship between inequality and growth, testing the validity of the Kuznets curve hypothesis across groups of countries. Second, we identify latent group structures within the dataset, revealing distinct patterns of growth and inequality dynamics across different groups of countries. Third, we assess coefficient heterogeneity across these groups, showing how the magnitude and direction of the growth-inequality relationship vary depending on latent group membership. Our findings suggest that the GFE model successfully identifies group-specific heterogeneity and uncovers the presence of three distinct groups within the dataset. Interestingly, two countries in the dataset do not belong to any of these groups and instead form their own individual clusters. Moreover, our results provide support for the Kuznets curve hypothesis: we observe a non-linear relationship between inequality and growth, consistent with an inverted U-shape. The relationship between inequality and growth is positive at lower levels of development but turns negative as countries grow wealthier, suggesting that while inequality may spur growth in the early stages of development, it becomes detrimental at higher levels of inequality. The results are robust across different specifications of the model and hold when using two different Gini indices to measure inequality. Additionally, our findings align with previous studies that suggest a positive short-term relationship between inequality and growth (Forbes 2000), while also supporting the notion of a negative long-term effect, as shown in Castelló-Climent (2010), Khalifa et al. (2010), Lee & Son (2016), Vo et al. (2019). Thus, this study provides a nuanced view of the growth-inequality nexus, emphasising the importance of considering group-specific dynamics and heterogeneous effects across different countries.

Building on the analysis of grouped heterogeneity in the growth-inequality relationship from Chapter 2, Chapter 3 shifts the focus towards a more advanced econometric model that addresses group-based heterogeneity in a multivariate setting. In Chapter 3, we make a significant methodological contribution by developing a Panel Vector Autoregression (PVAR) model with Grouped Fixed Effects (PVAR-GFE). While Chapter 2 dealt with single equation models, group-based heterogeneity is also prevalent in multivariate systems where multiple variables interact dynamically over time. PVAR models have been widely used in macroeconomic applications since their introduction by Holtz-Eakin et al. (1988). It treats each variable as endogenous and models their simultaneous interactions. However, they often fail to account for cross-sectional dependence between units, which can lead to biased estimates and inconsistent inference. The existing extensions of PVAR, such as Factor PVAR, Large Scale PVAR, and Bayesian PVAR (Canova & Ciccarelli 2013), have largely overlooked the potential for cross-sectional dependence arising from shared, unobserved group-level characteristics. This chapter addresses this limitation by incorporating group-based heterogeneity via grouped fixed effects (GFE) into the PVAR framework. The PVAR-GFE model introduced in this Chapter provides a flexible approach that does not require pre-specifying the number of groups or their memberships. It can be extended to accommodate group-specific heterogeneous coefficients, which allows the model to capture distinct group dynamics within the multivariate system. By clustering units based on shared characteristics, the PVAR-GFE model effectively reduces the dimensionality of the estimation problem, making it computationally efficient while maintaining flexibility in modelling group-specific time-varying factors. The GFE estimator is asymptotically consistent and $\sqrt{NT}$-consistent, making it well-suited for large panel datasets with many cross-sectional units (N) and time periods (T). In particular, the PVAR-GFE model provides unbiased estimates in large panel settings and offers a robust framework for estimating group-specific dynamics. Monte Carlo simulations show that the bias shrinks as N and T increase, and the group assignment misspecification rate is negligible in large panels even with more groups. Additionally, the iterative least squares method used to estimate the model, combined with the K-means clustering algorithm, ensures that group memberships are determined endogenously, offering flexibility when the true number of groups is unknown. This chapter extends the econometric toolkit by providing a parsimonious structural model that addresses cross-sectional dependence through the introduction of grouped fixed effects. The theoretical results derived in this chapter establish the consistency of the PVAR-GFE estimator under large N and T, while empirical evidence and simulations confirm its performance in terms of accuracy and efficiency, especially in the presence of unobserved group structures. This makes the PVAR-GFE model an ideal choice for macroeconomic studies where dynamic interactions between variables are influenced by latent group characteristics.

Chapter 4 introduces a novel Wald-type test designed specifically for grouped panels with group-specific heterogeneous coefficients and grouped fixed effects. The Granger causality framework has been widely used to assess the predictive relationships between time series variables, but existing methods often fail to account for heterogeneity across groups of cross-sectional units. This chapter extends the standard Granger causality test by incorporating group-specific heterogeneity and grouped fixed effects into the testing procedure. The proposed Wald-type test addresses the limitations of previous studies by allowing for heterogeneous coefficients across groups of units while retaining flexibility in estimating group-specific fixed effects. Under the null hypothesis of Granger non-causality, the group-specific parameters are assumed to reduce to zero, and the model treats them as homogeneous. This structure allows the homogeneous component of the model to be estimated using ordinary least squares (OLS) after projecting out the other parameters. An advantage of this approach is that the model structure assumes that there is only grouped fixed effects instead of individual fixed effects hence we address it implicitly via the model structure. Additionally, this test serves as a special type of homogeneity test, allowing researchers to test whether parameters are truly homogeneous across groups or exhibit significant heterogeneity. This chapter contributes to the econometric literature in several ways. First, it provides a robust method for testing Granger non-causality in panels with group-specific heterogeneity, offering a flexible and efficient approach for handling the complexities of cross-sectional dependence. Second, the heterogeneous coefficients testing could be a valuable tool for empirical research, particularly in macroeconomics and finance, where panel data often exhibit dynamic relationships between variables. Finally, the test can be used to assess parameter homogeneity, further extending its applicability beyond traditional Granger causality testing.

This thesis comprises three substantive chapters, each addressing the role of grouped heterogeneity in macroeconomic panel data analysis. Chapter 2 revisits the long-debated relationship between economic growth and income inequality through the lens of grouped fixed effects, offering fresh empirical evidence in support of the Kuznets curve within a heterogeneous cross-country framework. Chapter 3 advances the methodology by developing a PVAR model that accommodates group-specific dynamics, thereby capturing richer interactions among macroeconomic variables while addressing cross-sectional dependence. Chapter 4 further extends this line of enquiry by proposing a novel Wald-type test for Granger non-causality in grouped panels, which is robust to Nickell bias and enables testing for parameter heterogeneity. Chapter 5 synthesises the key insights and outlines directions for future research.

Chapter 2: Grouped Patterns in Economic Growth and Income Inequality

Abstract

The relationship between income inequality and economic growth has been the subject of extensive debate. This study revisits the growth-inequality nexus by accounting for latent grouped patterns across countries. The study employs the grouped fixed effects (GFE) estimator, which estimates the unobserved group structure. The results indicate a non-linear relationship between inequality and growth, consistent with the predictions of the Kuznets curve, but the relationship varies across groups. The results are robust to alternative model specifications and to the use of two different Gini indices.

2.1 Introduction

Kuznets (1955) was the first to study the relationship between income inequality and economic growth. His most influential contribution was the hypothesis that the relationship between growth and inequality follows an inverted U-shaped curve. An inverted U implies an initially positive relationship between inequality and growth; however, once a certain level of development is reached, the relationship becomes negative. Since then, a significant body of empirical research has examined the medium- and long-term linear effects of inequality on growth. The literature presents mixed findings. Several studies have reported a negative relationship between income inequality and economic growth (Alesina & Rodrik 1994, Castelló-Climent 2010, Khalifa et al. 2010), while others have found evidence of a positive relationship (Li & Zou 1998, Forbes 2000). More recent studies (Barro 2000, Cho et al. 2014, Abebe & Ratbek 2020) question these earlier conclusions, identifying the relationship between growth and inequality as non-linear.

This paper investigates the growth-inequality nexus using a worldwide panel dataset of 92 economies with annual observations from 1998 to 2017. Using the grouped fixed effects (GFE) estimator proposed by Bonhomme & Manresa (2015), the analysis focuses on three key features of the data: the non-linear relationship between inequality and growth, the latent group patterns, and the heterogeneity in the coefficients of different groups. Our findings suggest that the GFE model helps identify both heterogeneous effects and latent group patterns. The data-driven approach reveals that units in the panel can be broadly classified into three groups, with distinct forms of heterogeneity in the inequality-growth nexus. Notably, two economies do not belong to any of the identified groups; instead, each forms its own singleton group. Importantly, the results support the Kuznets hypothesis of a non-linear relationship between inequality and growth in the short term. This suggests that inequality may promote economic growth in the early stages of development but becomes detrimental at higher levels. It is important to note that this paper does not necessarily overturn previous findings. In contrast, it actually supports the positive relationship in the short term, as found in Forbes (2000) and may be viewed as complementing the negative long-term relationship reported in studies using alternative model specifications (Castelló-Climent 2010, Khalifa et al. 2010, Lee & Son 2016, Vo et al. 2019). Furthermore, the non-monotonic inverted U-shaped relationship aligns with the Kuznets curve hypothesis (Kuznets 1955).

This paper is structured as follows. Section 2 reviews previous studies on the growth-inequality nexus. Section 3 presents the data and general descriptive statistics. Section 4 outlines the different methods employed. Section 5 presents the results from the baseline model and compares them with those obtained under alternative specifications and control variables. Section 6 concludes with a summary of the main findings [4].

[4]For this paper, we employ both Stata 18.0 and MATLAB 2024b as the primary software platforms. Stata is used for data cleaning and the presentation of empirical results, while MATLAB is used to implement the grouped fixed effects (GFE) estimation procedures. The main Stata commands used include: xtreg, regife, xtregcluster, xtdcce2, classifylasso, and esttab. In MATLAB, the key function is parfor, which enables parallel computation across multiple starting values to improve the efficiency of the GFE estimation.

2.2 Literature Review

Kuznets (1955) was the first to explore the possibility of an empirical relationship between economic growth and income inequality. His framework assumes that, in the early stages of economic development, the onset of industrialisation, particularly in urban areas, increases labour productivity in cities relative to productivity levels in the predominantly agricultural rural sector. This, in turn, widens the income gap between the two sectors and, more broadly, increases overall income inequality [5].

As the economy reaches higher levels of development, individuals in the labour force migrate from low-productivity rural areas to high-productivity urban centres. In more developed economies, the agricultural sector becomes relatively smaller, while the industrial sector expands and achieves higher productivity. This transition generates two main effects: (i) an increase in average earnings, and (ii) a reduction in inequality accompanied by welfare redistribution. According to this prediction, the relationship between growth and income inequality follows an inverted U-shape, known as the Kuznets Curve (KC hereafter). It is characterised by a positive relationship between inequality and growth in the early stages of development, followed by a negative relationship once a certain threshold is surpassed.

In his seminal paper, Kuznets (1955) relied on data from only three developed countriesthe United States, the United Kingdom, and Germanyto conduct a simple descriptive analysis in support of his hypothesis, due to limited data availability. Since then, a substantial body of literature has developed on the topic. Unlike Kuznets's original study (Kuznets 1955), which does not explicitly address the direction of empirical causality in the growth-inequality nexus, subsequent research has primarily focused on the impact of income inequality on economic growth,

[5] The widening of the income distribution is attributed not to the size of the agricultural sector, but to the disproportionate prosperity of emerging urban industries, which intensify overall inequality.

thereby implying a causal direction from inequality to growth.

Prior to 2000, the literature utilised simple cross-sectional and basic panel regressions. After 2000, with the development of new panel data techniques and the increased availability of data, researchers shifted the focus of economic development studies to include both developed and developing countries, while also addressing issues such as reverse causality from growth to inequality and the potential endogeneity of inequality. These developments are critically reviewed in the following subsections.

2.2.1 Literature Before 2000

The meta-analyses conducted by De Dominicis et al. (2008) and Neves et al. (2016) argue that discrepancies in the empirical findings on the relationship between income inequality and economic growth may stem from differences in estimation methods and data quality. For example, a reduced-form negative relationship between income inequality and economic growth has been documented in several studies using cross-sectional OLS regression (Persson & Tabellini 1991, Alesina & Rodrik 1994, Perotti 1996), even though the data were originally collected in panel format. By collapsing panel data into a cross-sectional format, researchers attempted to address the potential endogeneity of inequality arising from reverse causality. In these models, the dependent variable is typically defined as the average annual growth rate of GDP per capita, while the key independent variable is the initial level of income inequality. However, these cross-sectional studies were more susceptible to omitted variable bias and significantly reduced the number of observations used for estimation. Additionally, scholars have criticised cross-sectional approaches for capturing only cross-country historical variation, while failing to account for within-country changes over time (Deininger & Squire 1996).

To address concerns about data quality, Deininger & Squire (1996) compiled a new dataset aimed at improving the reliability of empirical research on income inequality. First, they argue that the Gini coefficient (commonly referred to as the Gini index) is an appropriate proxy for income distribution. This is not only because it is widely used in empirical research, but also because results derived from alternative summary measures of inequality are generally comparable to those obtained using the Gini coefficient. [6]

Additionally, they discussed the following data selection criteria in order to improve data quality:

1. The unit of observations should come from households surveys' individual units instead of national accounts.

2. The observations obtained from households surveys can be used as a representative of all the population.

3. The measurement of income inequality should be assessed using a range of income sources and representative groups.

The dataset compiled by Deininger & Squire (1996) substantially expanded the time dimension of global income data and laid the groundwork for the use of panel econometric methods in its analysis. Despite the advent of this new panel dataset, numerous studies have produced contradictory results, largely due to the differing econometric techniques employed.

Li & Zou (1998) was the first to use the data supplied by (Deininger & Squire 1996) and, using data for 46 countries from 1960 to 1985, estimated fixed effects

[6] Using the Lorentz curve, the Gini coefficient measures the fairness of annual income distribution. It is a proportional value that ranges from 0 to 1. As a percentage, the Gini index is the Gini coefficient multiplied by 100. When the Gini coefficient equals 1, the annual income distribution among residents is completely unequal (all income is concentrated in the hands of one person, and the rest of the population has no income), while when the Gini coefficient equals 0, the annual income distribution among residents is completely even, that is, income is completely distributed evenly among people. The Gini coefficient can only have a value between these two extremes, i.e. between 0 and 1. The lower the Gini coefficient, the more evenly distributed annual income is; the higher the Gini coefficient, the more unequal the annual income distribution is.

and random effects panel models to study the link between income inequality and growth, reporting a positive relationship between the two variables.

Forbes (2000) examined the relationship between inequality and growth using first-difference GMM estimation in order to address the potential endogeneity of inequality and the inconsistency of panel OLS estimators.

According to Forbes (2000), the conflicting findings in the literature are due to differences in data quality and estimation approach. To support this argument, Forbes (2000) also employs the panel dataset from Deininger & Squire (1996), constructing a model similar to that of Perotti (1996) but estimated using GMM. The findings indicate a positive association between inequality and growth over short- and medium-term horizons, thereby contradicting conventional findings about the direction of the relationship between inequality and growth. This result illustrates how exploiting the time dimension of panel data can yield different outcomes from those obtained through simple cross-sectional analysis, which relies on time-invariant variables and is more prone to omitted variable bias. Barro (2000) employed three-stage least squares (3SLS) and found little evidence supporting the Kuznets Curve in the global panel.

2.2.2 Literature After 2000

Following the introduction of improved estimation techniques and increased access to data after 2000, many researchers have provided further insights into inequality-growth nexus. As shown in Table 1, which summarises the pre-2000 literature, even studies using the most recent datasets typically include fewer than 50 countries, with developed economies overrepresented relative to developing ones. This pattern, however, does not hold in the post-2000 literature. Specifically, researchers have not only examined the full-sample relationship across all countries in the Deininger & Squire (1996) dataset, but have also shifted the focus of analysis, at times concentrating exclusively on developing countries, on both developed and developing economies, or on the effects of membership in one or more economic organisations. [7]

Since 2000, a growing volume of global data has become available through the World Bank and other sources. This increased data availability, combined with advances in econometric techniques, has led researchers to reassess the global relationship between inequality and economic growth. Castelló-Climent (2010) was among the first to highlight the role of heterogeneity when countries are at different stages of development. Castelló-Climent (2010) also cautions that employing difference GMM estimation, as done in the study by Forbes (2000), may be misleading when model variables exhibit high persistence. Therefore, they, along with later studies by Halter et al. (2014) and Lee & Son (2016), employed system GMM estimation and found that, globally, the relationship between income inequality and economic growth is negative, whereas advanced economies display a stable, positive association. Since then, a growing body of literature has emphasised the importance of accounting for heterogeneity in the inequality-growth relationship.

Khalifa et al. (2010) presented the first systematic study on the relationship between inequality and development using threshold methods. Using data from 70 countries over a 29-year period, they confirmed that the relationship between income inequality and economic growth depends on a country's stage of economic development. Herzer & Vollmer (2012) examined the long-run relationship between income inequality and economic development using several heterogeneous panel cointegration methods. Their results indicate that the relationship is negative in the long run, even after accounting for heterogeneity. A panel structural VAR (PSVAR) model, used in Grigoli et al. (2016), further demonstrates that, even when accounting for cross-country heterogeneity in coefficients, the relationship

[7]Globally, countries are grouped using a variety of classification schemes. As proposed by the International Monetary Fund, countries are divided into three categories: developed countries, developing countries, and transition countries. Several sources below categorize it according to the World Bank definition, which is high-income countries, lower middle income countries and low income countries.

between income inequality and growth remains negative.

Recent literature continues to explore the possibility of a non-linear relationship between income inequality and economic growth. Several studies have replicated the analysis in Barro (2000), which examined the non-linear relationship between inequality and growth. Chen (2003) analysed long-run annual economic growth as a function of polynomial terms of initial inequality, using the theoretical framework of Barro (1991). Cross-sectional long-run results seem to be consistent with a Kuznets's curve, however short-run results are still unclear. Similarly, non-parametric methods have also identified evidence of a non-linear relationship. One of the first studies to relax the assumption of a linear functional form was conducted by Banerjee & Duflo (2003). Their findings suggest that both increases and decreases in inequality lead to slower economic growth in the subsequent decade.

More recent studies address both non-linearity and heterogeneity in the inequality-growth relationship. Cho et al. (2014) employed Panel Smooth Transition Regression (PSTR) to corroborate earlier findings on non-linear effects. Khalifa et al. (2010), followed by Abebe & Ratbek (2020), advanced the development of a non-linear threshold framework to analyse the effect of income inequality on economic growth. In contrast to the static panel threshold model employed by Khalifa et al. (2010), Abebe & Ratbek (2020) uses both a dynamic panel threshold model and a dynamic heterogeneous pooled mean group estimator to confirm the presence of heterogeneity and non-linearity in the inequality-growth relationship.

As more data became available from a wider range of countries, the number of studies examining the relationship between income inequality and economic growth continued to grow, increasingly incorporating developing and emerging economies. Fawaz et al. (2014) criticised earlier studies, such as Barro (2000) and Forbes (2000), for focusing exclusively on developed countries while overlooking the implications for developing economies. In contrast, Fawaz et al. (2014) focused exclusively on a selection of developing countries. Using the World Bank's classification of developing countries into high- and low-income groups, they analysed the relationship between inequality and growth within each category using Fixed Effects (FE) and the dynamic Generalised Method of Moments (GMM). They argue that the World Bank's classification is arbitrary, as the criteria and cluster memberships have changed since the initial 1978 report. To address this, they employed a threshold model to endogenously classify countries into low- and high-development groups and re-estimated the relationship using FE and System GMM. Notably, Fawaz et al. (2014) was the first study to implement endogenous sample classification, in contrast to the more common practice of subjective grouping. Their results indicate a negative relationship between inequality and growth in low-income developing economies, but a positive relationship in high-income developing countries. These findings are robust to whether the initial classification is based on the World Bank grouping or the threshold model. Furthermore, Vo et al. (2019) examine how income inequality affects economic growth among middle-income developing countries. Using a System GMM approach, they report a negative relationship between inequality and growth.

Using a panel of OECD countries, Cingano (2014) report a negative relationship between income inequality and economic growth. Ncube et al. (2014) examine the inequality-growth-poverty nexus and identify a negative association in the Middle East but a positive one in North Africa (MENA region). Yang & Greaney (2017) expanded the sample used in Birdsall et al. (1995) to cover the Asia-Pacific region and found that the relationship between inequality and income resembles an S-shaped curve. This constitutes one of the first empirical challenges to the traditional Kuznets curve since the original study by Kuznets (1955). In addition, a growing number of studies examine the relationship between income distribution and economic growth at the subnational or state level across different regions (see Atems & Jones (2015), Frank (2009), Kennedy et al. (2017), Litschig & Lombardi

(2019)).

2.2.3 Literature About Group Patterns

The emergence of grouped patterns in panel data analysis represents a significant advancement in modelling unobserved heterogeneity, particularly in macroeconomic contexts such as the inequality-growth nexus. Traditional panel data estimatorssuch as fixed effects (FE) and random effects (RE) modelstypically assume either unit-specific effects or complete homogeneity across cross-sectional units. However, a growing body of empirical evidence suggests that countries or economic units may exhibit latent group structures, in which units within the same group share common dynamics or structural characteristics that differ across groups (Lee et al. 1997). This recognition has prompted the development of econometric techniques capable of endogenously identifying such latent structures. These techniques can broadly be classified into clustering-based and penalisation-based approaches.

2.2.3.1 Clustering-based Method

One of the earliest contributions in this field is by Lin & Ng (2012), who propose two clustering-based methods to estimate panel data models with unknown group-specific coefficients. The first is a pseudo-threshold method that groups units based on time-series estimates of individual slope coefficients, drawing on the logic of panel threshold regressions. The second adapts the K-means clustering algorithm to a conditional framework, grouping units by minimising deviations between unit-specific and group-level conditional means. These approaches aim to strike a balance between flexibility and interpretability, enabling consistent group estimation without requiring prior knowledge of group membership. Empirical applicationssuch as international growth regressions and regional housing marketsdemonstrate their practical relevance.

Building on this line of work, Sarafidis & Weber (2015) introduce a partially heterogeneous panel framework, allowing for homogeneous slope coefficients within groups but heterogeneity across them. Their methodology is based on partitioning the panel into clusters by minimising the residual sum of squares (RSS), combined with an information-based criterion to select the number of groups. Importantly, the method remains consistent in identifying the correct group structure even when the time dimension is fixedmaking it particularly valuable for short panels in macroeconomic research.

A major innovation in this literature is the Grouped Fixed Effects (GFE) estimator developed by Bonhomme & Manresa (2015), which extends the traditional fixed effects model by incorporating group-specific, time-varying fixed effects. This approach relaxes the assumption of unit-specific fixed effects and instead estimates a set of time-varying profiles shared within latent groups. Estimation is conducted by minimising a least-squares objective over all possible groupings, typically operationalised using clustering algorithms such as K-means. The GFE approach is particularly well-suited to macroeconomic applications in which unitssuch as countriesmay exhibit similar time paths due to comparable development trajectories or shared policy regimes. The flexibility of the GFE framework has motivated numerous extensions, including quantile-based GFE models (Zhang et al. 2019, Gu & Volgushev 2019), threshold-based GFE specifications (Miao, Su & Wang 2020, Miao, Li & Su 2020), and models that accommodate structural breaks across groups (Lumsdaine et al. 2020). Further methodological advances in grouping algorithms are proposed by Bonhomme et al. (2022), Chetverikov & Manresa (2022), and Mugnier (2025).

2.2.3.2 Unknown group patterns: Penalisation-based approaches

A complementary approach to uncovering latent group structures relies on penalisation techniques that enable simultaneous estimation and classification. A key contribution in this domain is the Classifier-Lasso (C-Lasso) estimator proposed by Su et al. (2016). This method imposes a mixed additive-multiplicative penalty that shrinks unit-specific coefficients toward unknown group-level targets. Crucially, it does not require prior group membership information and can be applied in both linear and nonlinear settings. Estimation is carried out using penalised profile likelihood (PPL) or penalised GMM, depending on whether endogeneity is present. C-Lasso exhibits desirable theoretical properties, including uniform classification consistency and, under PPL, the oracle property. The number of groups is determined endogenously using a BIC-type criterion, and the method demonstrates strong performance in both simulated and empirical settings. Su & Ju (2018) extend this framework to dynamic panel models with interactive fixed effects, accommodating both unobserved slope heterogeneity and cross-sectional dependence. They propose the Penalised Principal Components (PPC) estimator, which combines elements of factor modelling with the group shrinkage structure of C-Lasso. This integration enables the model to capture latent common factors while simultaneously identifying grouped slope heterogeneity. Extending the C-Lasso framework further, Huang et al. (2020) adapt the method to cointegrated panel models, accommodating both stationary and non-stationary regressors. Their penalised least squares estimator identifies latent group structures in long-run relationships while allowing for full heterogeneity in short-run dynamics. The estimator maintains uniform classification consistency and retains the oracle property asymptotically, even in the presence of endogeneity.

In parallel with these penalisation-based approaches, Ando & Bai (2016, 2017) propose a model that incorporates grouped factor structures, allowing for both group-specific coefficients and group-specific unobserved common factors. The estimation procedure employs least squares combined with non-convex shrinkage penalties, such as SCAD, for variable selection. The method jointly estimates group memberships, factor structures, and regression parameters, thereby nesting both grouped heterogeneity and interactive fixed effects within a unified framework.

2.2.4 Contributions

Although the impact of income inequality on economic growth has been well studied, to the best of our knowledge, this is the first time that the GFE framework has been used to model income inequality and economic growth. A key motivation for applying the GFE estimator lies in its ability to uncover latent group structures that are economically, rather than statistically, determined. While traditional empirical studies often divide countries into groups based on income levels or development status, such classifications are externally imposed and do not necessarily reflect underlying economic behaviour. In contrast, we use the GFE clustering algorithm to determine endogenously both the number of groups and group memberships. GFE groups countries based on similarity in economically meaningful classification.

One of the key advantages of the GFE approach is its ability to uncover policy relevant groupings that go beyond traditional classifications based on income or region. While existing categories (e.g., high-income vs. low-income countries) are useful for simple comparisons, they may less effective to show how countries respond to structural changes such as a set of economical factors. The GFE estimator clusters countries based on shared underlying economic behaviour, particularly in how inequality influences growth. This means that countries within the same latent group, regardless of geographic location or development status, are likely to exhibit similar policy responsiveness. As a result, policy prescriptions can be better targeted and more effective, since they are aligned with the empirically observed structural dynamics rather than imposed categorisations. For instance, a tax reform that reduces income inequality may be expected to promote growth in one group of countries but not in another. Such distinctions are made possible by the data-driven clustering inherent in GFE.

[8]Note: This table shows the literature about the relationship between income inequality and economic growth before 2000. One thing we should note is that the time period T in the literature stands for the time span of the dataset. However, most of literature average the time periods over five or ten years.

Table 1: Literature Review of Chapter 1 Before 2000[8]

Author & Year	Literature Detail			
	Estimation Methods	Results	Number of Units	Time Periods
Kuznets (1955)	Descriptive	Non-linear	$N = 3$	Time Varies (Due to data availability)
Persson & Tabellini (1991)	Cross-Sectional OLS Regression	Negative	$N = 56$	$T = 5$ (1960-1985)
Alesina & Rodrik (1994)	Cross-Sectional 2SLS Regression	Negative	$N = 70$	$T = 5$ (1960-1985)
Li & Zou (1998)	Fixed & Random effects Panel Regression	Positive	$N = 46$	$T = 5$ (1960-1985)
Forbes (2000)	Difference GMM Panel Regression	Positive	$N = 45$	$T = 6$ (1965-1995)
Barro (2000)	3SLS Regression	Non-linear	$N = 84$	$T = 7$ (1960-1995)

2.3 Data

We use data for 92 economies worldwide, covering the period from 1998 to 2017. The data sources include the World Development Indicators (WDI) from the World Bank, the Human Development Reports from the United Nations Development Programme (UNDP), and the Standardized World Income Inequality Database (SWIID). Unlike most previous studies, we use annual data rather than five-year averages, as higher-frequency data is generally preferable, particularly when using GFE estimators, which benefit from a longer time dimension that improves the identification of latent group structures.

In previous studies, researchers typically extract the Gini coefficient from either the Luxembourg Income Study (LIS) or the World Income Inequality Database (WIID). However, issues of data comparability and consistency can reduce data quality and negatively affect empirical results. Specifically, combining data from

[9]Note: This table shows the literature about the relationship between income inequality and economic growth after 2000. The time periods T in the literature stand for the 5-year average time span.

Table 2: Literature Review of Chapter 1 After 2000[9]

Author & Year	Literature Detail			
	Estimation Methods	Results	Number of Units	Time Periods
Castelló-Climent (2010)	System GMM Dynamic Panel Regression	Negative	$N = 102$	$T = 7$ (1965-1995)
Khalifa et al. (2010)	Panel Threshold Regression	Negative	$N = 70$	$T = 6$ (1970-1999)
Herzer & Vollmer (2012)	Heterogeneous Panel Cointegration	Negative	$N = 46$	$T = 5$ (1970-1995)
Lee & Son (2016)	System GMM Dynamic Panel Regression	Negative	$N = 107$	$T = 6$ (1980-2010)
Abebe & Ratbek (2020)	Dynamic Panel Threshold Pool Mean Group	Nonlinear	$N = 100$	$T = 10$ (1965-2014)
Cho et al. (2014)	Panel Smooth Transition Regression	Nonlinear	$N = 77$	$T = 6$ (1980-2007)
Fawaz et al. (2014)	System GMM	Positive Negative	$N = 56$ $N = 55$	$T = 10$ (1960-2010)
Grigoli et al. (2016)	Panel Structural VAR	Negative	$N = 77$	$T = 11$ (1960-2014)
Vo et al. (2019)	System GMM	Negative	$N = 86$	$T = 11$ (1960-2014)
Sbaouelgi et al. (2018)	Quantile Regression	Negative	$N = 117$	$T = 10$ (1960-2011)
Atems & Jones (2015)	Panel VAR Regression	Negative	$N = 48$	$T = 15$ (1930-2005)

the LIS and WIID may not yield a proper panel dataset, as variable definitions often differ between the two sources. We therefore use two Gini indices from the Standardized World Income Inequality Database (SWIID), constructed by Solt (2020), specifically: the Gini index of inequality in household disposable income (post-tax, post-transfer) and the Gini index of inequality in household market income (pre-tax, pre-transfer). As expected, the market income Gini index is always higher than the disposable income Gini index, since the former excludes taxes and government transfers (i.e., redistribution) [10]. Compared to other Gini index sources used in the literature, SWIID 9.1 is better suited for cross-country income inequality research. Accordingly, both Gini indices are used as independent variables in the empirical analysis.

As the dependent variable, we use the annual growth rate of GDP per capita, obtained from the World Development Indicators (WDI)[11]. Compared to the five-year average GDP per capita used in many previous studies, annual data allow for the analysis of fluctuations over the entire time span and are more effective in detecting latent group structures when applying the GFE estimator.

As additional covariates, we include the investment share of GDP from the World Bank's World Development Indicators (WDI) and the price level of investment from version 10.0 of the Penn World Tables (PWT). [12]. Both are commonly used control variables in the literature, and the latter serves as a proxy for market distortions.

Table 3 presents descriptive statistics for the variables of interest across two groups of countries, as classified by the International Monetary Fund (IMF) in the World Economic Outlook (WEO). These are categorised into Developed Economies and Emerging & Developing Economies, based on criteria including per capita income, export diversification, and the degree of integration into the global financial system

[10]Market household income is defined as the sum of labour income (paid employment and self-employment income) and capital income. Disposable household income is the sum of labour income (paid employment and self-employment income), capital income, transfer income social security transfers (work-related insurance transfers, universal benefits, and assistance benefits) and private transfers minus income taxes and social security contributions

[11]This variable is calculated as real GDP at constant 2010 national prices divided by total population.

[12]The latest version of PWT updated on June 18, 2021. More details in Feenstra et al. (2015)

[13].

From Table 3 and Figure 1, we observe that, on average, real per capita GDP (RGDP) is higher in developed economies than in emerging and developing economies. The developed group reports a mean of 10.285 and a standard deviation of 0.951, compared to a mean of 8.127 and standard deviation of 1.061 in the latter over the period 1998-2017. Notably, during the period of the Subprime and Global Financial Crises (2007-2008), there is a slight decline in both real GDP per capita and its growth rate across both groups of countries.

Table 3: Descriptive Statistics categorized by Developed economies, and Emerging economies & Developing economies

Descriptive Statistics	Developed economies		Emerging economies & Developing economies	
	Mean	Std.	Mean	Std.
Real GDP per capita (log)	10.285	0.951	8.127	1.061
Gini disposable income	30.430	4.197	41.128	7.309
Gini market income	47.131	4.137	45.959	6.385
Price level of investment	0.729	0.200	0.465	0.159
Investment to GDP ratio	23.743	4.635	23.791	7.409
Population Growth	0.565	0.853	1.215	1.170

[13] WEO only classifies the countries only into two groups. The difference between Emerging Economies and Developing Economies is not considered in the website, for more details see https://www.imf.org/en/Publications/WEO

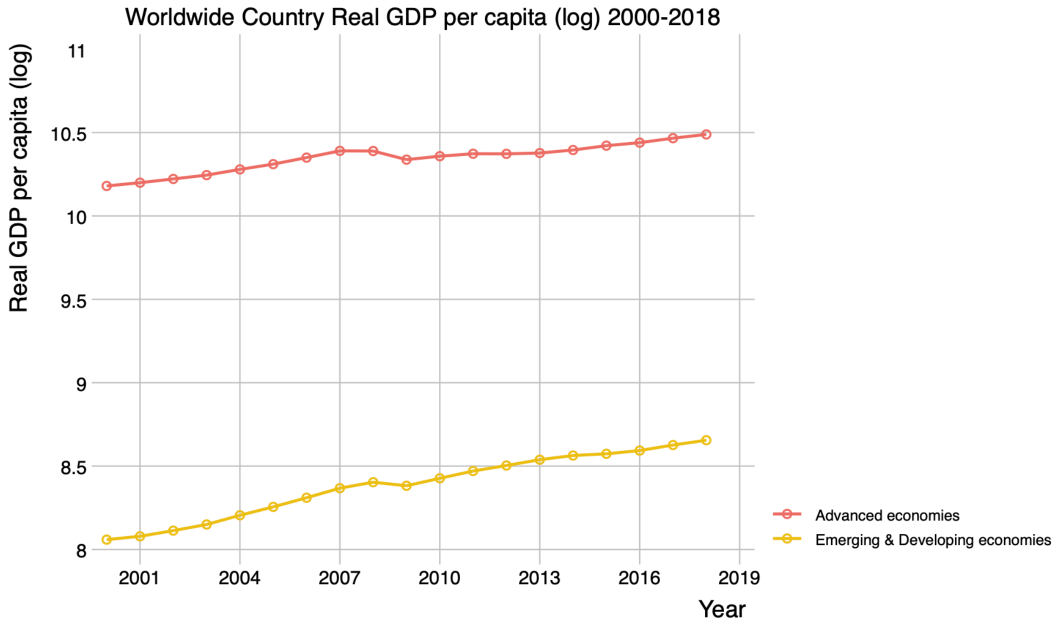

Figure 1: Observations classify as Developed economies, and Emerging economies & Developing economies in terms of real GDP per capita (log)

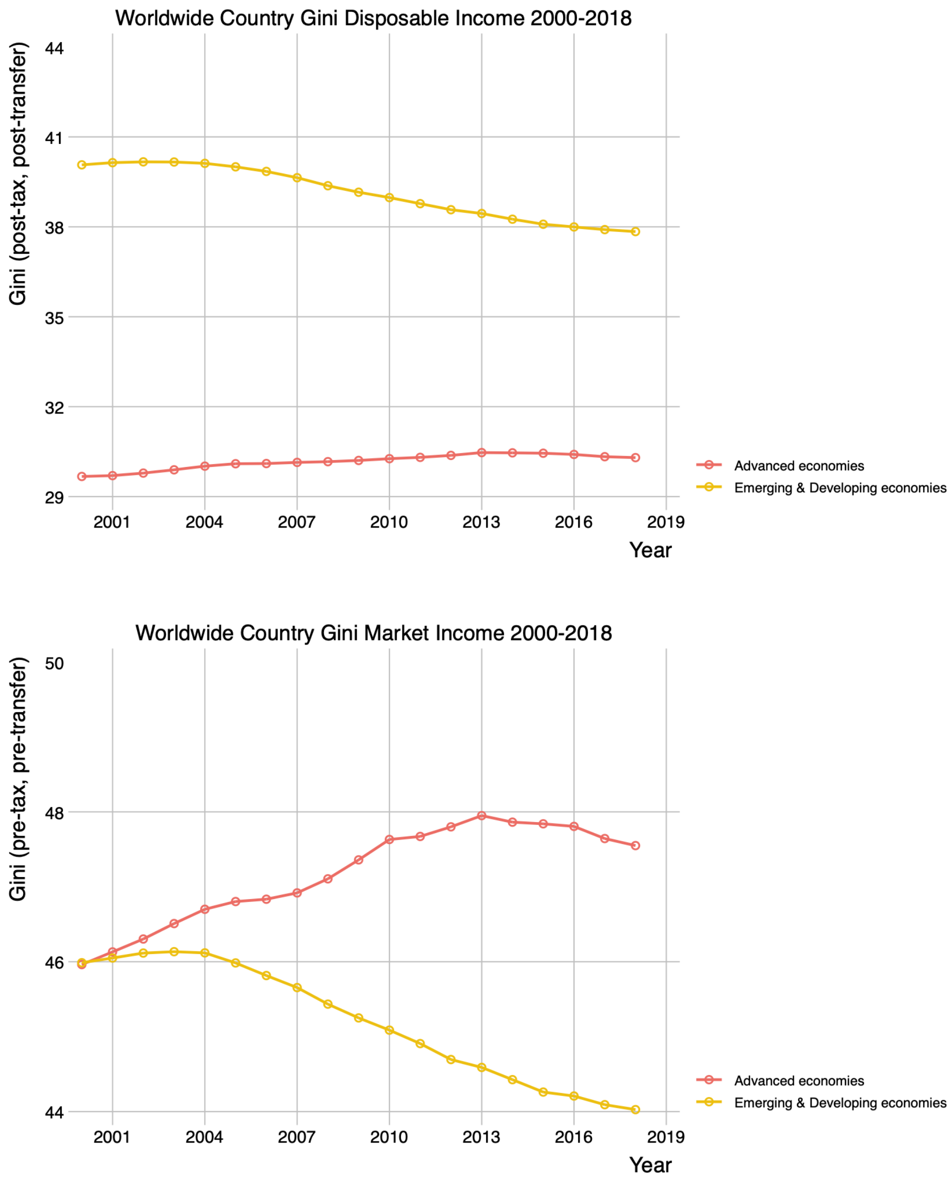

Figure 2: Observations classify as Developed economies, and Emerging economies & Developing economies in terms of Gini disposable income index (on the top) and Gini market income index (on the bottom)

The top and bottom panels of Figure 2 display the time paths of the Gini indices for disposable income and market income, respectively. From 2002 onward, the Gini index for market income in developed economies becomes consistently higher than that in emerging and developing economies, with the gap widening over time. However, after accounting for redistribution measures such as taxes and government transfers, developed economies exhibit lower Gini indices for disposable income than emerging and developing economies throughout the entire 20-year period.

The World Bank (WB) classifies countries into four income groups: high income, upper-middle income, lower-middle income, and low income. Table 4 presents descriptive statistics for the variables of interest across the four income groups, as classified by the WB. The average values of the two measures of inequality differ across the four groups, depending on whether inequality is measured before redistribution (Gini Market Income, GMI) or after (Gini Disposable Income, GDI). In high- and upper-middle-income countries, initial inequality is substantially reduced through the use of effective redistribution policies.

Figure 3 displays the evolution of GDP per capita for the four income groups, showing that all groups experienced a relatively steady increase over the 20-year period. Specifically, high-income countries recorded the highest average real GDP per capita (in logs) over the entire period, with a mean of 10.212 and a standard deviation of 0.872. The log of GDP per capita in upper-middle- and lower-middle-income countries increases sharply after 2000, with means of 8.622 and 7.430, respectively.

Figure 4 shows that the Gini index for market income in high-income countries increased significantly after 2000 and gradually declined beginning in 2015. However, after accounting for redistribution measures, high-income countries consistently maintained the lowest Gini index for disposable income throughout the period.

Table 4: Descriptive Statistics categorised by High income, Upper middle income, Lower middle income and Low income

Descriptive Statistics	High income		Upper middle income		Lower middle income		Low income	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Real GDP per capita (log)	10.212	0.872	8.622	0.477	7.430	0.590	6.242	0.330
Gini disposable income	31.408	5.595	41.387	7.897	41.122	6.002	44.113	4.320
Gini market income	47.340	4.038	47.600	7.150	42.382	4.757	46.469	4.510
Price level of investment	0.709	0.198	0.478	0.144	0.402	0.166	0.465	0.114
Investment to GDP ratio	23.202	4.382	24.450	6.586	25.299	7.904	19.849	10.499
Population Growth	0.530	0.824	0.795	1.048	1.619	0.874	2.846	0.845

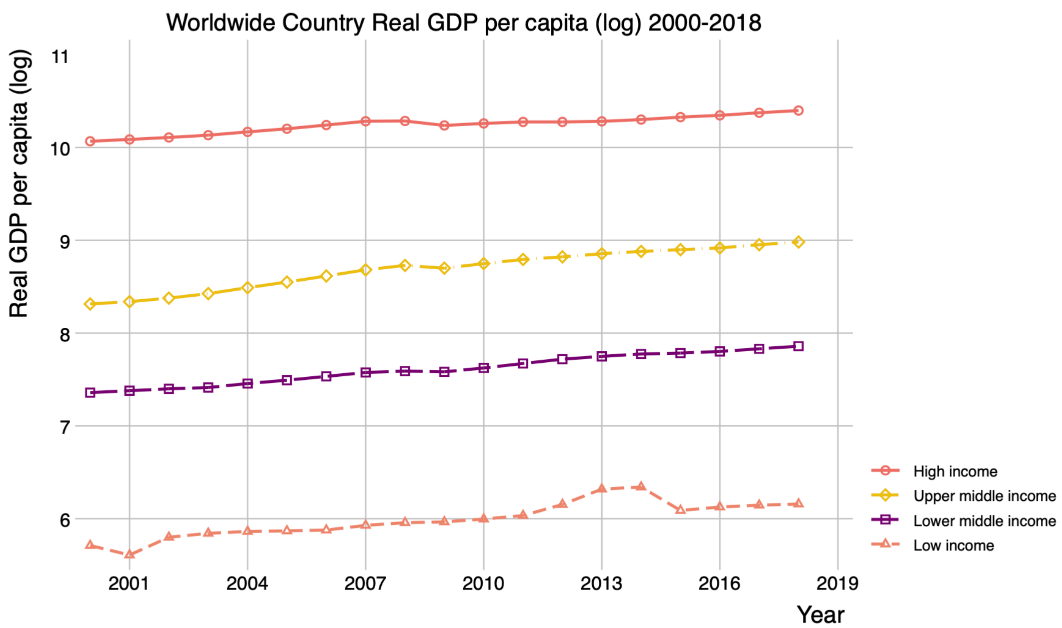

Figure 3: Observations classify by income levels, namely High income, Upper middle income, Lower middle income and Low income in terms of real GDP per capita (log)

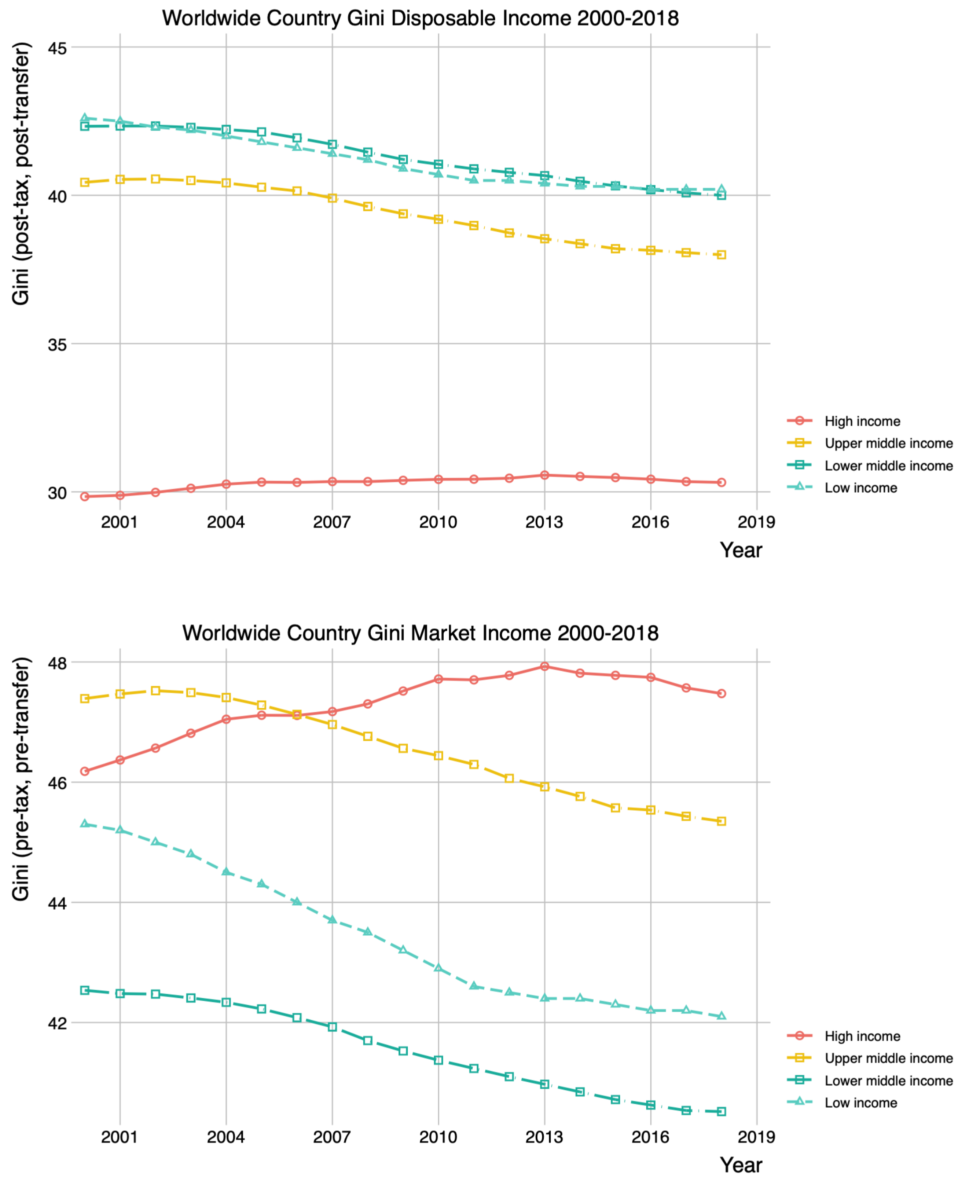

Figure 4: Observations classify as High income, Upper middle income, Lower middle income and Low income in terms of Gini disposable income index (on the top) and Gini market income index (on the bottom)

2.4 Methodology

This section outlines the baseline model, the Grouped Fixed Effects (GFE) model, and its extensions to be estimated. As discussed above, we use the growth rate of GDP per capita as the dependent variable, since the non-stationarity of real GDP per capita would violate the assumptions necessary for valid inference under the GFE estimator and its extensions.

2.4.1 Baseline Models

Equations 1, 2 and 3 below present the baseline models, which will be estimated using Pooled Ordinary Least Squares (OLS), Fixed Effects (FE), and Two-Way Fixed Effects (TWFE), respectively.

$$y_{it} = \beta_1 x_{it-1} + \beta_2 x^2_{it-1} + \beta_3 y_{it-1} + \mathbf{z}'_{it-1}\boldsymbol{\theta} + v_{it} \tag{1}$$

$$y_{it} = \beta_1 x_{it-1} + \beta_2 x^2_{it-1} + \beta_3 y_{it-1} + \mathbf{z}'_{it-1}\boldsymbol{\theta} + \eta_i + v_{it} \tag{2}$$

and

$$y_{it} = \beta_1 x_{it-1} + \beta_2 x^2_{it-1} + \beta_3 y_{it-1} + \mathbf{z}'_{it-1}\boldsymbol{\theta} + \eta_i + \lambda_t + v_{it} \tag{3}$$

Here, $i = 1, \ldots, N$ indexes the cross-sectional units, and $t = 1, \ldots, T$ denotes the time periods. The variable y_{it} represents the annual growth rate of real GDP per capita for country i in time period t, and y_{it-1} denotes its lagged value. The variable x_{it-1} refers to the lagged measure of income inequality, captured by either of the two Gini indices [14]. We include a quadratic term to capture potential non-linear effects of income inequality on economic growth. All additional control variables, including the price level of investment and the investment-to-GDP ratioare incorporated in the vector $\mathbf{z}_{it-1}$.[15] In both equations 1 and 3, λ_t represents time fixed effects, which are invariant across units but vary over time. In Equation 2 and Equation 3, η_i denotes unit-specific fixed effects, which may be arbitrarily correlated with the covariates in $\mathbf{z}_{it-1}$ and are constant over time. The term v_{it} captures the idiosyncratic error component.

[14] There has long been debate regarding the appropriate measure of income inequality. Some authors argue against using the Gini coefficient, as a single-index measure may fail to capture the full complexity of income distribution (Voitchovsky 2005). Bartak & Jabłoński (2020), for example, propose alternative measures such as the Atkinson index. In this study, however, we employ both the Gini indices of disposable income and market income from Solt (2020), in line with previous literature, to facilitate comparability of results.

[15] Further control variables will be introduced in later sections as part of robustness checks.

2.4.2 The Basic Grouped Fixed Effects Model (GFE)

The motivation for the Grouped Fixed Effects (GFE) approach stems from the limitations posed by the Nickell bias (Nickell 1981) in short panel data settings. In dynamic panel models with unit-specific fixed effects, the inclusion of a large number of nuisance parameters relative to a small number of time periods (T) introduces a bias of order $O(1/T)$, known as the Nickell bias (Nickell 1981). This bias undermines the reliability of parameter estimates, especially when T is small. Unlike standard fixed-effects, Bonhomme & Manresa (2015) propose the GFE that individuals can be grouped into a small number of unobserved categories where each group shares a common time pattern of unobserved heterogeneity. By modelling unobserved heterogeneity as group-specific and time-varying rather than individual-specific and time-invariant, GFE eliminates the source of the bias by design. It is important to note that the Grouped Fixed Effects (GFE) estimator, introduced by Bonhomme & Manresa (2015), does not suffer from Nickell bias, a common issue in dynamic fixed effects panel models. This is not because the estimator explicitly corrects for the bias, but rather because it avoids it by construction. In the GFE framework, unit-specific fixed effects are replaced with a small number of group-specific intercepts. This structural shift eliminates the correlation between lagged dependent variables and individual fixed effectsa key source of Nickell bias introduced by the within transformation in traditional fixed effects models. This simplification enhances the GFE estimator's performance, particularly in panel data settings where both the cross-sectional dimension (N) and the time dimension (T) are large. Empirical studies suggest that the GFE model not only outperforms standard fixed effects estimators in terms of bias and efficiency but also demonstrates greater robustness in the presence of heteroskedasticity. Therefore, the GFE approach represents a powerful tool for panel data analysis when addressing challenges related to unobserved heterogeneity and Nickell bias. Moreover, Bonhomme & Manresa (2015) highlight the potential for the dynamic one-way fixed effects estimator to yield misleading results when latent group-specific patterns are ignored. To address this, they introduce a novel component into the panel framework: a time-varying group-specific fixed effect, denoted by $\alpha_{g_i t}$. This feature allows the model to account for unobserved heterogeneity that evolves over time within latent groups, and it underpins the flexibility and effectiveness of the GFE model. Incorporating the Grouped Fixed Effects (GFE) coefficient offers several important advantages. First, it enables the model to capture unobserved heterogeneity by allowing units that share similar structural characteristics to be grouped, thereby improving the model's ability to reflect latent cross-sectional patterns. Second, it promotes model parsimony, a key objective in econometric modelling, by reducing the number of parameters to be estimated. This not only enhances computational efficiency but also mitigates the risk of overfitting, particularly in high-dimensional panel settings.

Before the seminal work of Bonhomme & Manresa (2015), most literature on estimating unknown group structures primarily described algorithms and estimation procedures, as exemplified by Lin & Ng (2012). Bonhomme & Manresa (2015) marks a significant turning point in this field, as it was the first to present not only a data-driven grouping algorithm and its estimation procedure, but also a rigorous analysis of the asymptotic properties of the resulting estimators. This comprehensive approach filled a crucial gap in the literature by offering theoretical guarantees concerning consistency and inference, thereby enhancing the credibility and applicability of grouped fixed effects models.

Following Bonhomme & Manresa (2015), in order to investigate the relationship between growth and inequality, we parameterise our empirical model as:

$$y_{it} = \beta_1 x_{it-1} + \beta_2 x_{it-1}^2 + \beta_3 y_{it-1} + \mathbf{z}_{it-1}'\boldsymbol{\theta} + \alpha_{g_i t} + v_{it} \tag{4}$$

Let $i = 1, \ldots, N$ denote the cross-sectional units and $t = 1, \ldots, T$ the time periods. Specifically, y_{it} represents the annual growth rate of real GDP per capita for unit i in period t, and y_{it-1} denotes its one-period lag. The variable x_{it-1} denotes the lagged income inequality index, measured using two versions of the Gini index: the Gini Disposable Income Index (GDI) and the Gini Market Income Index (GMI). We also include the squared term of x_{it-1} to capture potential non-linear effects of income inequality on real GDP per capita growth. The vector $\mathbf{z}_{it-1}$ contains the control variables, initially including the price level of investment and the investment-to-GDP ratio.Additional control variables will be included in later specifications to assess the robustness of the results. The variable g_i indicates the group membership of unit i, assigning each unit to one of G groups, such that $g_i \in G_1, G_2, \ldots, G_G$ and is fixed over time. Finally, v_{it} denotes the idiosyncratic error term. The number of groups is determined by an algorithm analogous to the K-means clustering algorithm used in machine learning. [16]. The estimation process of the GFE is obtained from:

$$(\widehat{\theta}, \widehat{\alpha}) = \underset{(\theta,\alpha)}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \mathbf{x}'_{it}\theta - \alpha_{\widehat{g}_i(\theta,\alpha)t} \right)^2 \tag{5}$$

Where $\mathbf{x}_{it}$ is a vector that now includes all the independent variables in equation 4. Group membership of the individual units $\widehat{g}_i(\theta, \alpha)$ is obtained from:

$$\widehat{g}_i(\theta, \alpha) = \underset{g \in \{G_1, G_2, \ldots, G_n\}}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_{it} - \mathbf{x}'_{it}\theta - \alpha_{gt} \right)^2 \tag{6}$$

The determination of group membership follows a series of iterative steps. Initially, the algorithm randomly assigns each unit to one of the G groups and sets initial values for both θ and α. Subsequently, the algorithm employs least squares estimation to iteratively update the model parameters and group assignments. This iterative process continues until numerical convergence is achieved, typically assessed using the L_2 norm. In essence, the algorithm repeatedly reassigns units to groups until it identifies the combination of θ and α, denoted as $\widehat{g}_i(\theta, \alpha)$, that minimises the sum of squared residuals across all possible values of $\alpha_{g_i t}$. It is important to note that this process may need to be repeated multiple times to ensure convergence to the global minimum, thereby facilitating the identification of parameter values and group structures that minimise the objective function.

[16] In the original paper from Bonhomme & Manresa (2015), there are 2 algorithms that could be used to decide the number of groups in the objective function, namely the iterative (K-means) and the Variable Neighborhood Search (VNS). The VNS method is more suitable when the size of observations is sufficiently large. In this paper, we will use the iterative algorithm to decide the number of groups.

Researchers should be mindful of the necessity of multiple repetitions to obtain robust results. As a result, the procedure can become computationally intensive, particularly in large-scale datasets.

Attaining the global minimum becomes increasingly challenging as the dimensionality of the problem increases, particularly when the number of units, time periods, covariates, and groups expands. It is important to note that researchers must determine either the maximum number of groups, denoted as $G_{\max}$, or the total number of groups, G, based on subjective judgment or established model selection criteria such as the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC). In this study, we set the maximum number of groups, $G_{\max}$, to 8. To ensure robustness and accuracy, we conducted an extensive number of simulations, running the iterative process 10 million times for each group specification, ranging from $G = 2$ to $G = 8$. This extensive simulation exercise enhances the reliability and robustness of our results. The BIC is calculated as follows:

$$BIC(G) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{it}'\widehat{\theta}^{(G)} - \widehat{\alpha}_{\widehat{g}_i t}^{(G)}\right)^2 + \widehat{\sigma}^2\frac{GT+N+K}{NT}\ln(NT) \qquad (7)$$

The AIC is calculated as:

$$AIC(G) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - x_{it}'\hat{\theta}^{(G)} - \hat{\alpha}_{g_{it}}^{(G)}\right)^2 + \widehat{\sigma}^2\frac{2(GT+N+K)}{NT} \qquad (8)$$

Where $\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{it}'\widehat{\theta}^{(G)} - \widehat{\alpha}_{\widehat{g}_i t}^{(G)}\right)^2$ is the objective function, the total number of parameter are estimated is the time-varying GFE estimators (GT), total number of group membership $\sum_{i=1}^{N} g_i = N$ and the number of covariates K. $\widehat{\sigma}^2$ is the consistent estimator of the variance of the idiosyncratic term and it is calculated as:

$$\widehat{\sigma}^2 = \frac{1}{NT - G_{\max}T - N - K}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{it}'\widehat{\theta} - \widehat{\alpha}_{\widehat{g}_i t}\right)^2$$

Where $G_{\max}$ is the largest number of groups.

In this study, we compare the results obtained using the AIC and the BIC to determine the optimal number of groups in the GFE estimator, as shown in Figure 5. While the AIC is designed to minimise prediction error and tends to favour more complex models by selecting a larger number of groups, it does not guarantee consistency in estimating the true number of latent groups. In our case, the AIC selects eight groups, which merely suggests that the group assignment may involve eight or more distinct categories. In contrast, the BIC imposes a stronger penalty that increases with sample size, ensuring parsimony and reducing the risk of overfitting. Theoretical considerations suggest that the BIC is consistent (Sarafidis & Weber 2015, Bonhomme & Manresa 2015), meaning that as the sample size increases, it identifies the true number of groups with high probability. Given the large panel dataset used in this study where both N and T are sufficiently large, we prioritise model consistency and interpretability over pure predictive accuracy. Therefore, we rely on the BIC to determine the final number of groups, ensuring that the identified structure is both robust and theoretically justified. In our estimation, the BIC selects four groups when $G_{\max} = 8$.

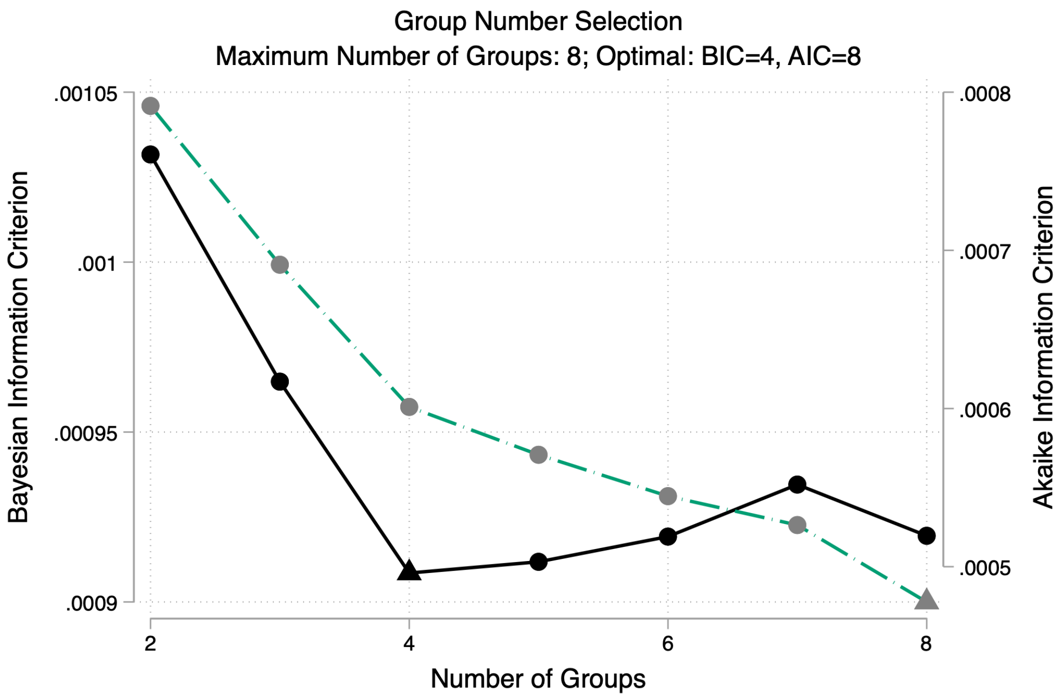

Figure 5: GFE Selected by AIC and BIC Values

2.4.3 GFE Model with Individual Fixed Effects

The model can be extended by including individual fixed effects. This will allow us to model any further unobserved heterogeneity in equation 4. The GFE model will now be expressed in terms of deviations from country/unit means, i.e. including the time-invariant individual fixed effects and time-varying group fixed effects simultaneously:

$$y_{it} = \beta_1 x_{it-1} + \beta_2 x_{it-1}^2 + \beta_3 y_{it-1} + \mathbf{z}'_{it-1}\boldsymbol{\theta} + \alpha_{g_i t} + \eta_i + v_{it} \tag{9}$$

Where the variables are denoted in the same way as equation 4, but we now include the extra individual fixed effects term η_i. We estimate equation 9 by demeaning the data to remove the individual fixed effects:

$$\begin{aligned} y_{i,t} - \bar{y}_i = {} & \beta_1(x_{it-1} - \bar{x}_i) + \beta_2(x_{it-1}^2 - \bar{x}_i^2) \\ & + (\alpha_{g_i,t} - \bar{\alpha}_{g_i}) + (\mathbf{z}_{i,t} - \mathbf{z}_i)'\boldsymbol{\theta} + (v_{i,t} - \bar{v}_i) \end{aligned} \tag{10}$$

Where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}$, $\bar{y}_{i,t-1} = \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}$, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{i,t}$, $\bar{\alpha}_{g_i} = \frac{1}{T}\sum_{t=1}^{T} \alpha_{g_i,t}$ and $\bar{v}_i = \frac{1}{T}\sum_{t=1}^{T} v_{i,t}$. Since now equation 10 has the same model structure as 4, we can now use the GFE iterative algorithm to obtain group memberships and estimate the remaining parameters in the model.

2.4.4 GFE with Heterogeneous Coefficients

The majority of previous research on the relationship between inequality and economic growth has primarily focused on addressing issues related to unobserved endogeneity and reverse causality. These approaches typically rely on the assumption of parameter homogeneity across cross-sectional units. However, it is plausible to depart from this assumption, as coefficients may exhibit heterogeneityincluding time-varying characteristicsas demonstrated in studies such as Burnside (1996), Lee et al. (1997), and Hsiao & Pesaran (2004). To address the need for flexibility and accommodate potential parameter heterogeneity, we introduce an extended version of the Grouped Fixed Effects (GFE) model. This extended model relaxes the homogeneity assumption, allowing for a more nuanced and realistic representation of the relationship between inequality and economic growth. In particular, we allow income inequality to exert heterogeneous effects on economic growth across different latent groups:

$$y_{it} = \beta_{g_i1}x_{it-1} + \beta_2 x_{it-1}^2 + \beta_3 y_{it-1} + \mathbf{z}_{it-1}'\boldsymbol{\theta}_+ \alpha_{g_it} + \eta_i + v_{it} \tag{11}$$

In equation 11, we incorporate a multifaceted In equation 11, we employ a multifaceted modelling framework that simultaneously accounts for non-linear effects, individual-specific characteristics, latent group-specific unobservables, and heterogeneous effects. To relax the assumption of parameter homogeneity, we introduce a heterogeneous coefficient associated with the independent variable, the Gini index. This coefficient captures the variation in the impact of income inequality across different latent group structures. Additionally, we include the quadratic term x_{it-1}^2 to capture the non-linear relationship between income inequality and economic growth. Furthermore, our model incorporates both group fixed effects, denoted α_{g_it}, and individual fixed effects, denoted η_i, to account for unobserved heterogeneity. These fixed effects control for latent characteristics at both the individual and group levels, thereby enhancing the model's capacity to capture and explain the complex relationship between inequality and economic growth.

2.5 Estimation Results

2.5.1 Baseline Regression

In Table 5, we present the estimates of the parameters for the equations from 1 to equation 4 respectively by means of Pooled OLS (POLS), Fixed Effects (FE), two-way Fixed Effects (TWFE), GFE and GFE with individual fixed effects (GFE-FE) using the two different measures of income inequality, namely, the Gini Disposable Index and the Gini Market Index (GDI and GMI hereafter). Columns (1)-(3) and Column (6)-(8) in table 5 report the estimated parameters for equation 1 to equation 3, respectively by POLS, FE and TWFE, while columns (4)-(5) and Column (9)-(10) display the results of the basic GFE and GFE with FE for the regressions using the Gini disposable index and Gini market index, respectively.

Note that the optimal number of groups in the GFE and GFE-FE in Table 5, is obtained using the Bayesian Information Criterion (BIC) as displayed in Table 7. Bonhomme & Manresa (2015) indicate that increasing the number of units (N), time periods (T), groups (G), and covariates (K) may cause the objective function to converge only to a local minimum rather than a global one. To mitigate this issue, a maximum number of groups, denoted $G_{\max}$, must be specified. In this study, we set $G_{\max} = 8$ and conduct 10,000,000 simulations for each group specification. As noted earlier, because the objective function typically decreases with an increasing number of groups, we determine the optimal number of groups as the value that minimises the BIC, shown in Figure 5. This yields $G = 4$ for both the GFE and GFE-FE specifications.Specifically, Columns (1) to (3) show that the coefficients on the GDI and its squared term are both statistically significant, indicating that the relationship between income inequality and economic growth is non-linear. This result partially aligns with Forbes (2000), who report a positive growth-inequality relationship; however, it further suggests that higher levels of inequality are detrimental to economic growth. A similar result holds when using the GMI as the inequality measure (see Columns 6 and 8). Columns (4)-(5) of Table 5 further show that, in both the GFE and GFE-FE specifications, the coefficients on the GDI (0.162 and 0.671, respectively) and its square (-0.023 and -0.092) remain statistically significant, reinforcing the evidence of a non-linear relationship between inequality and GDP growth. A comparable result holds when using the GMI, as shown in Columns (9)-(10). Notably, when comparing the GFE results in Columns (4) and (9) with the Pooled OLS estimates in Columns (1) and (6), the Gini coefficients from the GFE models are consistently smaller in magnitude. A similar pattern holds when comparing the GFE-FE estimates to those from the standard TWFE model. Finally, comparing the estimated effects across inequality definitions reveals that a one-percentage-point increase in the GMI leads to a 0.263 percentage point increase in growth under the GFE model (2.178 under GFE-FE). This effect is more moderate when using the GDI0.126 under GFE and 0.671 under GFE-FE, reflecting the redistributive role of taxes and transfers. Furthermore, the threshold beyond which inequality becomes

Table 5: Regression results of the baseline models

Dependent variable:	Gini Disposable Index					Gini Market Index				
$\text{Growth}_{i,t}$	POLS	FE	TWFE	GFE	GFE-FE	POLS	FE	TWFE	GFE	GFE-FE
$\text{Growth}_{i,t-1}$	0.271***	0.206***	0.211***	0.391***	0.301***	0.272***	0.201***	0.206***	0.392***	0.294***
	(0.037)	(0.048)	(0.055)	(0.037)	(0.039)	(0.037)	(0.048)	(0.054)	(0.037)	(0.039)
$\text{Gini\_Index}_{i,t-1}$	0.215**	0.956**	0.814**	0.162**	0.671**	0.390*	2.572***	2.443***	0.263*	2.178**
	(0.098)	(0.376)	(0.354)	(0.069)	(0.312)	(0.201)	(0.861)	(0.921)	(0.150)	(0.929)
$\text{Gini\_Index}^2_{i,t-1}$	-0.030**	-0.130**	-0.111**	-0.023**	-0.092**	-0.052**	-0.338***	-0.322***	-0.035*	-0.288**
	(0.014)	(0.052)	(0.049)	(0.010)	(0.043)	(0.026)	(0.113)	(0.120)	(0.019)	(0.121)
Population $\text{Growth}_{i,t-1}$	-0.005***	-0.010***	-0.009***	-0.005***	-0.009***	-0.006***	-0.010***	-0.009***	-0.005***	-0.009***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
Price Level of $\text{Investment}_{i,t-1}$	-0.027***	-0.032***	-0.026***	-0.021***	-0.022***	-0.027***	-0.031***	-0.028***	-0.022***	-0.024***
	(0.002)	(0.005)	(0.006)	(0.002)	(0.006)	(0.002)	(0.005)	(0.006)	(0.002)	(0.006)
Investment to GDP $\text{Ratio}_{i,t-1}$	0.011**	0.005	0.009	0.010***	-0.003	0.011**	0.004	0.009	0.010***	-0.004
	(0.005)	(0.011)	(0.009)	(0.004)	(0.007)	(0.005)	(0.011)	(0.010)	(0.004)	(0.007)
N	1748	1748	1748	1748	1748	1748	1748	1748	1748	1748
Individual	No	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
Grouped	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Year	No	Yes	Yes	No	No	No	No	Yes	No	No
Turning Point	3.545***	3.664***	3.500***	3.588***	3.646***	3.774***	3.805***	3.800***	3.763***	3.788***
Confidence Interval of Turning Point	(3.399, 3.691)	(3.493, 3.836)	(3.482, 3.845)	(3.453, 3.724)	(3.440, 3.853)	(3.628, 3.920)	(3.740, 3.869)	(3.735, 3.863)	(3.590, 3.937)	(3.714, 3.862)

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

harmful to growth is approximately 3.588 for the GDI (3.763 for the GMI) under GFE, and 3.646 (GDI) and 3.788 (GMI) under GFE-FE. Beyond these thresholds, the marginal effect of inequality turns negative, reducing growth by approximately 0.023-0.092 (GDI) and 0.035-0.288 (GMI) percentage points, respectively.

2.5.2 Heterogeneity

To further investigate heterogeneity, we relax the parameter homogeneity assumption by employing the GFE model with heterogeneous coefficients, as specified in equation 11, to explore the growth-inequality relationship across different latent group structures. This specification simultaneously accounts for non-linear effects, latent group patterns, individual time-invariant fixed effects, and coefficient heterogeneity. The optimal number of groups is again determined by the minimum BIC value, and the number of simulations is set to 10,000,000. For both Gini indices, the number of groups is four, which is consistent with the results obtained from equations 4 and 9. Group membership remains consistent with the previous assignment. Table 6 presents the estimation results for equation 11, which incorporates GFE with heterogeneous coefficients and individual fixed effects. In Table 6, we focus on Groups 1 and 4, which exhibit clear group-specific patterns. In both groups, we observe Kuznets-curve-like patterns, though with differing magnitudes.

Table 6: GFE with Heterogeneous Coefficient

	Gini Disposable Index	Gini Market Index
$y_{i,t-1}$	0.298***	0.292***
	(0.038)	(0.039)
Investment to GDP Ratio	-0.003	-0.004
	(0.007)	(0.007)
Price Level of Investment	-0.023***	-0.024***
	(0.005)	(0.005)
Population Growth	-0.009***	-0.009***
	(0.002)	(0.002)
Heterogeneous Coefficient	Gini Disposable Index	Gini Market Index
Group 1 Gini_Index$_{i,t-1}$	0.646*	2.053**
	(0.366)	(0.902)
Group 1 Gini_Index$^2_{i,t-1}$	-0.085*	-0.269**
	(0.051)	(0.118)
Group 4 Gini_Index$_{i,t-1}$	1.001*	3.228***
	(0.541)	(1.558)
Group 4 Gini_Index$^2_{i,t-1}$	-0.145*	-0.433***
	(0.075)	(0.205)
N	1710	1710
Individual Fixed Effects	Yes	Yes
Grouped Fixed Effects	Yes	Yes

Notes: This table shows the regression results for GFE with individual fixed effects and heterogeneous coefficient (equation 9). Dependent variable is the annual growth rate of GDP per capita. In column (1), we use Gini disposable index and its squared term as independent variables while we use Gini market index and its squared term as independent variables in Column (2). The table only shows the heterogeneous coefficient of groups with more than one members.

[a]Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5.3 Memberships in GFE

Table 7: GFE-Based Country Group Membership

Group 1	**Income Group**	**Development Group**	**Group 1 (cont.)**	**Income Group**	**Development Group**
Albania	Upper Middle Income	Advanced Economies	Australia	High Income	Advanced Economies
Austria	High Income	Advanced Economies	Bangladesh	Lower Middle Income	Emerging & Developing
Barbados	High Income	Emerging & Developing	Belgium	High Income	Advanced Economies
Bolivia	Lower Middle Income	Emerging & Developing	Brazil	Upper Middle Income	Emerging & Developing
Canada	High Income	Advanced Economies	China	Upper Middle Income	Emerging & Developing
Colombia	Upper Middle Income	Emerging & Developing	Costa Rica	Upper Middle Income	Emerging & Developing
Cyprus	High Income	Advanced Economies	Czech Republic	High Income	Advanced Economies
Denmark	High Income	Advanced Economies	Dominican Republic	Upper Middle Income	Emerging & Developing
Egypt	Lower Middle Income	Emerging & Developing	El Salvador	Lower Middle Income	Emerging & Developing
Estonia	High Income	Advanced Economies	Fiji	Upper Middle Income	Emerging & Developing
France	High Income	Advanced Economies	Gambia	Low Income	Emerging & Developing
Germany	High Income	Advanced Economies	Ghana	Lower Middle Income	Emerging & Developing
Greece	High Income	Advanced Economies	Hong Kong	High Income	Emerging & Developing
Hungary	High Income	Emerging & Developing	India	Lower Middle Income	Emerging & Developing
Indonesia	Lower Middle Income	Emerging & Developing	Iran	Lower Middle Income	Emerging & Developing
Ireland	High Income	Advanced Economies	Israel	High Income	Advanced Economies
Italy	High Income	Advanced Economies	Jamaica	Upper Middle Income	Emerging & Developing
Japan	High Income	Advanced Economies	Jordan	Upper Middle Income	Emerging & Developing
Kazakhstan	Upper Middle Income	Emerging & Developing	Kenya	Lower Middle Income	Emerging & Developing
Luxembourg	High Income	Advanced Economies	Malaysia	Upper Middle Income	Emerging & Developing
Mauritius	Upper Middle Income	Emerging & Developing	Mexico	Upper Middle Income	Emerging & Developing
Mozambique	Low Income	Emerging & Developing	Netherlands	High Income	Advanced Economies
New Zealand	High Income	Advanced Economies	Norway	High Income	Advanced Economies
Pakistan	Lower Middle Income	Emerging & Developing	Paraguay	Upper Middle Income	Emerging & Developing
Peru	Upper Middle Income	Emerging & Developing	Philippines	Lower Middle Income	Emerging & Developing
Poland	High Income	Emerging & Developing	Portugal	High Income	Advanced Economies
Korea (Rep.)	High Income	Advanced Economies	Romania	Upper Middle Income	Emerging & Developing
Senegal	Lower Middle Income	Emerging & Developing	Singapore	High Income	Advanced Economies
South Africa	Upper Middle Income	Emerging & Developing	Spain	High Income	Advanced Economies
Sweden	High Income	Advanced Economies	Switzerland	High Income	Advanced Economies
Tanzania	Lower Middle Income	Emerging & Developing	Thailand	Upper Middle Income	Emerging & Developing
Uganda	Low Income	Emerging & Developing	United Kingdom	High Income	Advanced Economies
United States	High Income	Advanced Economies	Uruguay	High Income	Emerging & Developing
Vietnam	Lower Middle Income	Emerging & Developing			
Group 4	**Income Group**	**Development Group**	**Group 4 (cont.)**	**Income Group**	**Development Group**
Argentina	Upper Middle Income	Emerging & Developing	Armenia	Upper Middle Income	Emerging & Developing
Belarus	Upper Middle Income	Emerging & Developing	Bulgaria	Upper Middle Income	Emerging & Developing
Chile	High Income	Emerging & Developing	Croatia	High Income	Emerging & Developing
Ecuador	Upper Middle Income	Emerging & Developing	Finland	High Income	Advanced Economies
Honduras	Lower Middle Income	Emerging & Developing	Iceland	High Income	Advanced Economies
Latvia	High Income	Advanced Economies	Lithuania	High Income	Advanced Economies
Mongolia	Lower Middle Income	Emerging & Developing	Niger	Low Income	Emerging & Developing
Panama	Upper Middle Income	Emerging & Developing	Moldova	Upper Middle Income	Emerging & Developing
Russia	Upper Middle Income	Emerging & Developing	Serbia	Upper Middle Income	Emerging & Developing
Slovakia	High Income	Advanced Economies	Slovenia	High Income	Advanced Economies
Sri Lanka	Lower Middle Income	Emerging & Developing	Turkey	Upper Middle Income	Emerging & Developing
Ukraine	Lower Middle Income	Emerging & Developing			

Table 8: Descriptive Statistics by GFE Group Membership

Descriptive Statistics	Group 1		Group 4	
	Mean	Std. Dev.	Mean	Std. Dev.
GDP Growth Rate	0.024	0.031	0.033	0.045
Gini Disposable Index	3.601	0.214	3.540	0.231
Gini Market Index	3.837	0.117	3.801	0.135
Population growth rate (%)	1.098	0.991	0.488	1.270
Price Level of Investment (log)	-0.625	0.396	-0.705	0.358
Investment to GDP ratio	3.131	0.251	3.188	0.271

The classification generated by the GFE model is not only statistically validated via the BIC minimisation but also economically meaningful. Table 7 lists countries grouped by GFE and other statistical indicators and Table 8 shows the descriptive statistics for Group 1 and Group 4. Group 1 includes a broad range of high-income and upper-middle-income economies, such as the United States, Germany, Japan, Brazil, and China. In contrast, Group 4 is primarily composed of Eastern European and post-socialist economies such as Bulgaria, Romania, and the Russian Federation. These countries have undergone significant structural transformations from planned to market economies and face distinct challenges including redistributive system disruptions and labour market volatility. This implies that economies in Group 4 are more sensitive to increases in inequality, potentially due to weaker redistribution systems, more fragile political institutions, or underdeveloped credit markets, suggested by Kornai (1994), Rodrik (2000), Acemoglu & Robinson (2001), Acemoglu et al. (2003) As reported in Table 6, the model allows for heterogeneous slope coefficients across groups, capturing structural differences in how income inequality affects economic growth. Both Group 1 and Group 4 exhibit a statistically significant inverted-U shaped relationship between inequality and growth, consistent with the theoretical framework proposed by Barro (2000), Banerjee & Duflo (2003), Forbes (2000).

However, the estimated coefficients differ in magnitude: Group 4 shows a stronger curvature, indicating that the negative effects of inequality emerge at lower levels of inequality compared to Group 1. This suggests that the countries in Group 4 may have more limited institutional capacity to mitigate inequality's adverse impacts, such as credit market imperfections or weaker redistributive systems.

These findings are further supported by descriptive statistics. Group 4 economies, on average, display higher but more volatile growth rates, lower investment prices, and higher investment-to-GDP ratios. Such patterns are consistent with countries undergoing capital accumulation and structural transformation. In contrast, Group 1 includes a larger set of high- and upper-middle-income economies with more stable growth and higher inequality levels. These differences help explain the observed heterogeneity in the inequality-growth nexus across groups. Importantly, the GFE-based grouping differs from standard classifications by income or development stage. It identifies countries based on similarity in estimated economic behaviours, rather than on exogenous characteristics. Compared with models grouped by region or development indicators (see Table 10), the GFE specification achieves a better fit, as indicated by lower BIC values. This supports the interpretation that the GFE method captures latent heterogeneity that is both statistically relevant and economically interpretable.

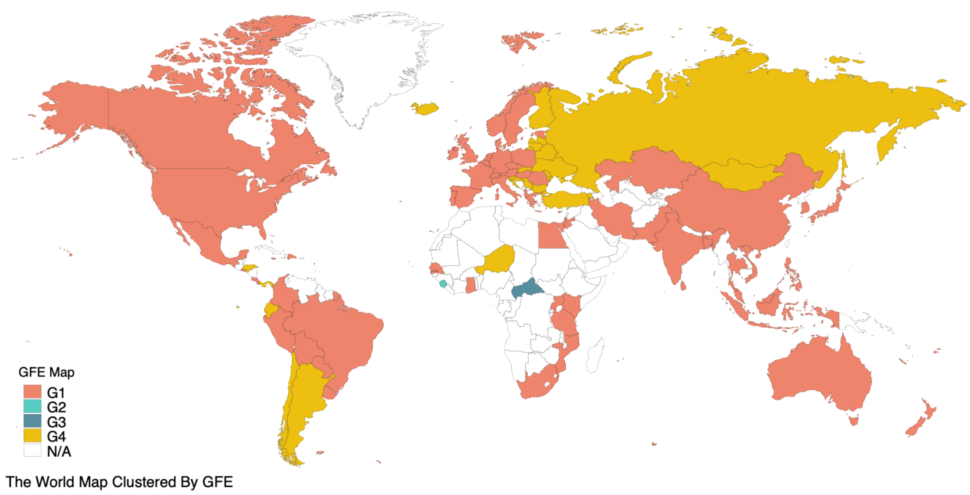

Figure 6: The World Map Clustered By GFE

2.6 Robustness Check

To enhance the robustness of our findings, we augment the baseline regression by incorporating additional covariates and exploring alternative model specifications.

2.6.1 Additional Covariates

Previous empirical studies have commonly employed both the human capital index and measures of educational attainment as additional control variables. In this study, we obtain the human capital index from the Penn World Table (PWT). As a proxy for educational attainment, we use the mean years of schooling, which reflects the average number of completed years of education for individuals aged 25 and older. While earlier research frequently relied on the Barro-Lee Educational Attainment Dataset, its coverage ends in 2015. To access more up-to-date data, we instead use the Human Development Reports from the United Nations Development Programme, which integrate information from multiple sources, including the UNESCO Institute for Statistics, Barro and Lee, ICF Macro Demographic and Health Surveys, UNICEF Multiple Indicator Cluster Surveys, and the OECD. As part of the robustness analysis, we re-estimate the model using both the standard Grouped Fixed Effects (GFE) estimator and its extension with individual fixed effects (GFE-FE). The non-linear effects of both Gini indices on economic growth are still evident and statistically significant, as shown in Columns (1)-(4) of Table 9. The signs of the additional control variables are consistent with theoretical expectations. Moreover, the number of latent groups and their composition remain stable across specifications, confirming the robustness of the estimated group structure.

Table 9: Robustness Check with Homogeneous Coefficient

	Robustness Check 1				Robustness Check 2		Robustness Check 3	
	Gini Disposable Index		Gini Market Index		Gini Disposable Index	Gini Market Index	Gini Disposable Index	Gini Market Index
$\text{Growth}_{i,t-1}$	0.386***	0.296***	0.388***	0.289***	0.290***	0.281***	0.090**	0.106***
	(0.035)	(0.038)	(0.037)	(0.039)	(0.041)	(0.042)	(0.038)	(0.037)
$\text{Gini\_Index}_{i,t-1}$	0.185***	0.699**	0.318*	2.249**	0.630**	2.011***	9.127	7.305
	(0.071)	(0.322)	(0.166)	(0.940)	(0.258)	(0.770)	(5.650)	(5.186)
$\text{Gini\_Index}^2_{i,t-1}$	-0.025***	-0.096**	-0.042**	-0.297**	-0.087**	-0.266***	-1.231	-0.973
	(0.010)	(0.044)	(0.021)	(0.123)	(0.036)	(0.101)	(0.804)	(0.679)
Investment to GDP $\text{Ratio}_{i,t-1}$	0.010***	-0.003	0.010***	-0.004	-0.001	-0.002	-0.072***	-0.083***
	(0.004)	(0.007)	(0.004)	(0.007)	(0.005)	(0.005)	(0.016)	(0.016)
Price Level of $\text{Investment}_{i,t-1}$	-0.022***	-0.022***	-0.024***	-0.024***	-0.020***	-0.022***	-0.056***	-0.049***
	(0.003)	(0.006)	(0.002)	(0.006)	(0.004)	(0.004)	(0.012)	(0.012)
Population $\text{Growth}_{i,t-1}$	-0.004***	-0.009***	-0.004***	-0.009***	-0.010***	-0.010***	-0.002	0.003
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.013)	(0.014)
Human Capital $\text{Index}_{i,t-1}$	0.000	-0.001	-0.001	-0.005				
	(0.005)	(0.009)	(0.004)	(0.010)				
Educational $\text{Attainment}_{i,t-1}$	0.001	-0.003	0.001	-0.002				
	(0.001)	(0.002)	(0.001)	(0.002)				
N	1691	1691	1691	1691	1710	1710	1710	1710
Individual	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Group	Yes	Yes	Yes	Yes	No	No	No	No
Year	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Interactive	No	No	No	No	Yes	Yes	No	No
Mean Group	No	No	No	No	No	Yes	Yes	

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6.2 Interactive Fixed Effects

In this section, we employ the interactive fixed effects model developed by Bai (2009) and extended by Moon & Weidner (2017), which captures the interaction between individual units and time periods in a linear panel framework. Unlike models that control only for time-varying group-level heterogeneity, this approach allows for heterogeneous responses to common shocks by incorporating unit-specific factor loadings. It thereby accounts for multidimensional structural shocks in the real economy and varying sensitivities of individual units to these shocks. The model specification is as follows:

$$Y_{it} = \beta_1 x_{it-1} + \beta_2 x_{it-1}^2 + \beta_3 y_{it-1} + \mathbf{z}'_{it-1}\boldsymbol{\theta} + \lambda'_i F_t + \varepsilon_{it} \tag{12}$$

Where F is the common factor and λ stands for factor loadings. The result in Table 9 Column (5) to Column (6) indicates that the non-linear effects still hold for both index

2.6.3 Mean Group Estimator

To validate the robustness of our findings based on the Grouped Fixed Effects (GFE) estimator, we conduct an additional robustness check using the Dynamic Common Correlated Effects Mean Group (DCCE-MG) estimator, as proposed by Chudik & Pesaran (2015). This approach addresses both cross-sectional dependence through unobserved common factors and heterogeneous slope coefficients, thereby providing a flexible framework that is particularly well-suited for macro panel settings in which omitted common shocks may lead to biased estimates in conventional models. The corresponding model specification is presented below:

$$y_{it} = \alpha_i + \lambda_i y_{i,t-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \sum_{l=0}^{p_T} \boldsymbol{\delta}_{il}' \overline{\mathbf{z}}_{t-l} + e_{it} \tag{13}$$

where $\overline{\mathbf{z}_t} = (\bar{y}_{t-1}, \overline{\mathbf{x}}_t)$. Consider λ_i and $\boldsymbol{\beta}_i$ as stacked into $\boldsymbol{\pi}_i = (\lambda_i, \boldsymbol{\beta}_i)$; then the MG estimates are

$$\widehat{\boldsymbol{\pi}}_{\mathrm{MG}} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\pi}}_i \tag{14}$$

As part of our robustness checks, we compare the results obtained from the Grouped Fixed Effects (GFE) estimator (Bonhomme & Manresa 2015) with those produced by the Dynamic Common Correlated Effects Mean Group (DCCE-MG) estimator (Chudik & Pesaran 2015) and the Interactive Fixed Effects (IFE) estimator (Bai 2009). While the DCCE-MG estimator allows for unrestricted heterogeneity in slope coefficients across units, its estimates in our context fail to achieve statistical significance. This likely reflects the estimator's high degree of flexibility, whichwhen combined with a limited time dimension or weak signalscan result in imprecise estimates and over-dispersion. In contrast, the IFE model performs substantially better, yielding statistically significant and robust estimates. This discrepancy underscores the nature of heterogeneity in the data: rather than being entirely idiosyncratic across countries, the heterogeneity appears to be driven by unobserved common factors with unit-specific loadings (as captured by the IFE estimator), or by latent group structures (as identified by the GFE approach). Thus, the superior performance of the IFE model relative to DCCE-MG reinforces the central claim of this studythat cross-country heterogeneity in the inequality-growth relationship is more accurately characterised by structured patterns than by purely individual-level variation.

2.6.4 Grouped by Different Specifications

2.6.4.1 Grouped by Statistical Index

Table 10 presents results from robustness check3, in which countries are grouped according to their development stage (as defined by the IMF) and income level (as defined by the World Bank). Columns (1) to (4) use the Gini Disposable and Market Index with grouping based on development status, while columns (5) to (8) apply grouping based on income level.Across all specifications, covering both the Gini Disposable Index and the Gini Market Index, the results continue to support a nonlinear relationship between income inequality and economic growth. Specifically, the coefficients on both the linear and squared Gini terms are statistically significant, confirming an inverted-U pattern whereby inequality initially stimulates growth but turns harmful beyond a threshold. Importantly, while these exogenous groupings yield statistically significant results, they perform worse in terms of model fittings. The BIC values across Columns (1) to (8) in Table 10 are consistently higher than those obtained under the GFE specification (see Figure 5), indicating that the models with traditional classifications are less preferred under information criteria. This provides formal evidence that GFE, which determines groups endogenously from the data, better captures the underlying heterogeneity in the inequality-growth relationship. Therefore, it is important to note that these classifications are externally imposed and do not necessarily reflect countries' structural behaviour. While useful for robustness, they may fail to capture the true latent heterogeneity in the way inequality affects growth. While traditional classifications are useful benchmarks, GFE remains the preferred framework for exploring the inequality-growth nexus in a way that aligns with the underlying economic mechanisms.

Table 10: Robustness Check 4

	Grouped by Development Stage				Grouped by Income Level			
	Gini Disposable Index		Gini Market Index		Gini Disposable Index		Gini Market Index	
$\text{Growth}_{i,t-1}$	0.276***	0.184***	0.295***	0.206***	0.285***	0.198***	0.288***	0.194***
	(0.042)	(0.052)	(0.042)	(0.054)	(0.042)	(0.052)	(0.042)	(0.052)
$\text{Gini\_Index}_{i,t-1}$	0.235***	1.008**	0.337**	2.443***	0.229***	0.954**	0.315*	2.556**
	(0.089)	(0.395)	(0.168)	(0.921)	(0.083)	(0.392)	(0.172)	(1.083)
$\text{Gini\_Index}^2_{i,t-1}$	-0.033***	-0.136**	-0.045**	-0.322***	-0.032***	-0.129**	-0.042*	-0.335**
	(0.013)	(0.054)	(0.022)	(0.120)	(0.012)	(0.053)	(0.022)	(0.141)
Investment to GDP $\text{Ratio}_{i,t-1}$	0.014***	0.006	0.014***	0.009	0.016***	0.008	0.016***	0.008
	(0.005)	(0.010)	(0.005)	(0.010)	(0.005)	(0.010)	(0.005)	(0.010)
Price Level of $\text{Investment}_{i,t-1}$	-0.024***	-0.023***	-0.025***	-0.028***	-0.023***	-0.023***	-0.024***	-0.026***
	(0.003)	(0.007)	(0.002)	(0.006)	(0.003)	(0.007)	(0.003)	(0.007)
Population $\text{Growth}_{i,t-1}$	-0.005***	-0.008***	-0.005***	-0.009***	-0.005***	-0.009***	-0.005***	-0.009***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
N	1748	1748	1748	1748	1748	1748	1748	1748
BIC	.0015	.0013	.0015	.0014	.0014	.0013	.0014	.0013
Individual	No	Yes	No	Yes	No	Yes	No	Yes
Grouped	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6.4.2 Grouped by Sarafidis & Weber (2015)

As an additional robustness exercise, we implement the grouping algorithm proposed by Sarafidis & Weber (2015), which partitions countries into clusters based on the similarity of slope coefficients using a least squares minimisation procedure (Table 11). This approach enables us to assess the sensitivity of the GFE results to alternative, externally specified clustering schemes. The number of groups is selected using the Model Information Criterion (MIC), which in our case suggests a two-group structure (see Figure 7). This aligns closely with the core results of the GFE estimator, which identifies two dominant latent groups out of a total of four, with the remaining two capturing small outlier clusters. When comparing group memberships, we find that approximately 63.04% of countries are classified consistently across the two methods. The minor discrepancies arise primarily from the treatment of extreme outliers such as Sierra Leone (SLE) and the Central African Republic (CAF), which are allocated to their own clusters under the GFE framework due to their distinct dynamics. In contrast, the method of Sarafidis & Weber (2015) tends to assign such outliers to the nearest large group in order to maintain model parsimony. These differences reflect the underlying modelling philosophies of the two approaches: whereas the GFE method permits the formation of highly specific latent groups, the alternative approach favours broader clusters that absorb outliers into existing structures. Overall, the high degree of concordance in group structure affirms the robustness of the GFE estimator and reinforces the empirical relevance of latent grouped heterogeneity in capturing cross-country differences in the inequality-growth nexus.

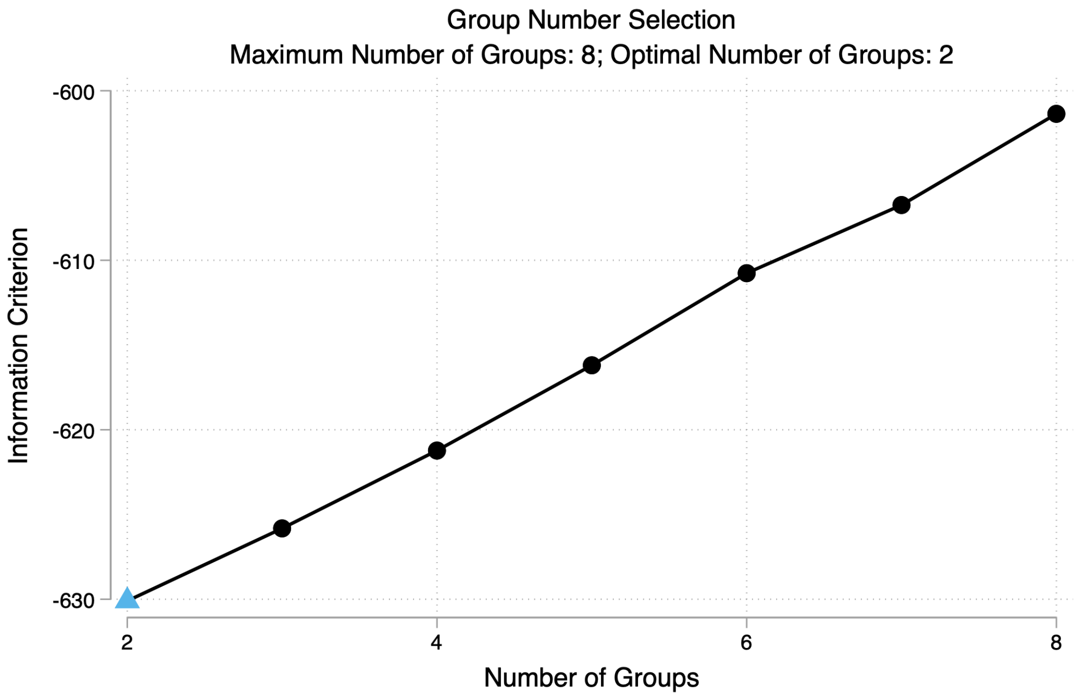

Figure 7: GFE Selected by Sarafidis & Weber (2015) MIC values

2.6.4.3 Grouped by Su et al. (2016)

To assess the robustness of our results based on the Grouped Fixed Effects (GFE) estimator, we apply the latent group classification method introduced by Su et al. (2016), which offers an alternative approach to modelling heterogeneous slope structures in panel data (Table 11). Their method employs the Classifier-Lasso (C-Lasso) estimator, which simultaneously performs classification and estimation by shrinking unit-specific slope coefficients toward latent group-specific values. This is achieved through a penalised profile likelihood or penalised GMM procedure, depending on the presence of endogeneity. A key innovation of the C-Lasso approach lies in its mixed additive-multiplicative penalty structure, which allows each unit's coefficient vector to be simultaneously attracted to multiple, unidentified group means. This design avoids the need for prior knowledge of group membership and ensures that the resulting classification is uniformly consistent and, under certain conditions, satisfies the oracle property. When applied to our dataset on economic growth and income inequality, the C-Lasso method identifies two latent groups based on estimated slope heterogeneity (See Figure 8). Comparing these classifications to those generated by the GFE model, we observe a moderate degree of overlap, approximately 56.5%, suggesting partial agreement between the two approaches. The divergence, however, reflects underlying differences in estimation strategy. While both models capture heterogeneity in slope coefficients, they do so via distinct mechanisms: the GFE estimator identifies group structures in both slopes and time-varying fixed effects by minimising residual variance, whereas the method of Su et al. (2016) focuses exclusively on slope-based clustering without modelling time-varying intercept heterogeneity.

In light of these methodological differences, the GFE estimator may offer advantages in our empirical setting, as it jointly captures heterogeneity in both slope and intercept termsan important feature when both structural and unobserved factors vary across latent groups. Consequently, the lower classification overlap with C-Lasso should not be interpreted as inconsistency, but rather as evidence of multidimensional latent heterogeneity that may be more comprehensively captured by GFE. While C-Lasso remains a valuable robustness check, we maintain GFE as our preferred specification due to its broader and more flexible treatment of grouped heterogeneity.

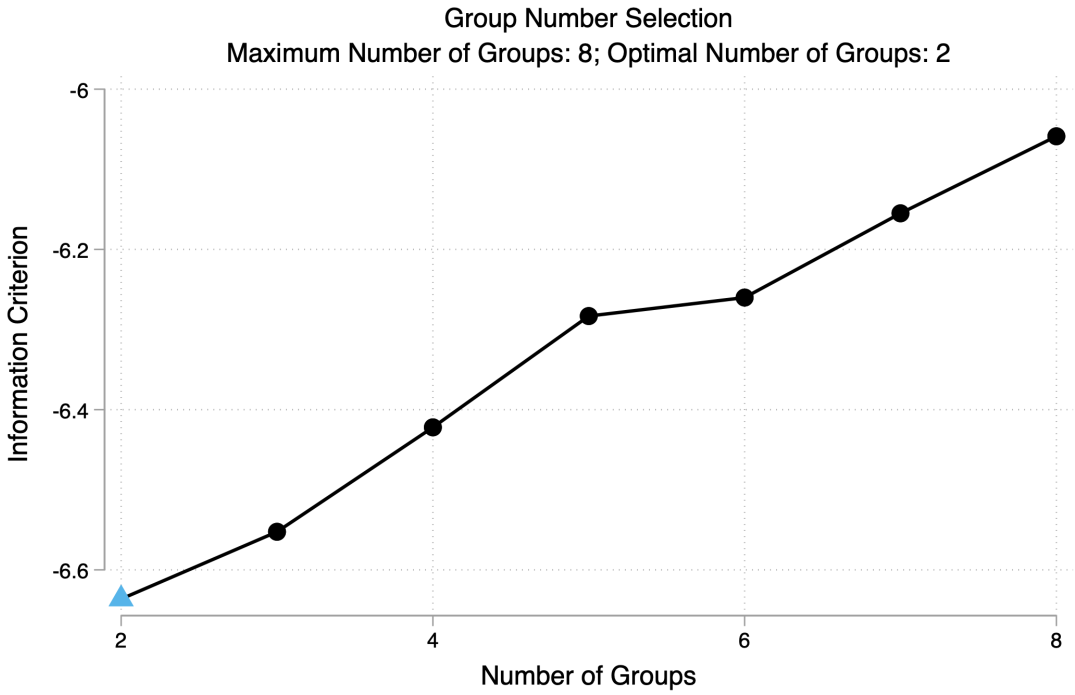

Figure 8: GFE Selected by C-lasso Information Criterion

Table 11: Robustness Check 5-6

	Grouped by Su et al. (2016)		Grouped by Su et al. (2016)	
	Gini Disposable Index	Gini Market Index	Gini Disposable Index	Gini Market Index
Group 1 Gini_Index$_{i,t-1}$	4.044***	3.259*	1.285**	2.345*
	(0.702)	(1.821)	(0.522)	(1.181)
Group 1 Gini_Index$^2_{i,t-1}$	-0.547***	-0.434*	-0.181**	-0.310**
	(0.097)	(0.239)	(0.074)	(0.155)
Group 2 Gini_Index$_{i,t-1}$	0.366	2.279**	2.115***	3.077**
	(0.295)	(0.935)	(0.827)	(1.171)
Group 2 Gini_Index$^2_{i,t-1}$	-0.052	-0.296**	-0.277***	-0.401**
	(0.041)	(0.122)	(0.109)	(0.154)
N	1748	1748	1748	1748
Individual	Yes	Yes	Yes	Yes
Grouped	Yes	Yes	Yes	Yes
Year	Yes	Yes	No	No
CLasso	No	No	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.7 Conclusions

This paper examines the short-term relationship between income inequality and economic growth using a broad, balanced panel of 92 economies covering the period 1998-2017. By applying the Grouped Fixed Effects (GFE) estimator, the study accounts for latent group-specific heterogeneity and simultaneously captures non-linearity, multiple unobserved effects, and slope heterogeneity. The GFE model identifies four latent groups; however, only two contain more than one member. Notably, Sierra Leone and the Central African Republic are not assigned to any broader group, instead forming their own distinct clusters. The results reveal a non-linear relationship between inequality and economic growth, as measured by both the post-tax and post-transfer Gini Disposable Index and the pre-tax and pre-transfer Gini Market Index. Specifically, inequality appears to promote economic growth at lower levels but becomes detrimental beyond a certain thresholdconsistent with the inverted-U shape predicted by the Kuznets Curve. Rather than contradicting previous studies, these findings complement the existing literature by focusing on short-run dynamics and highlighting the importance of latent grouped heterogeneity. This paper contributes to the literature by shedding light on the structural group patterns inherent in the inequality-growth nexus. Future research may build upon this work using extended GFE models, including quantile-based GFE estimators (Zhang et al. 2019, Gu & Volgushev 2019), threshold GFE models (Miao, Su & Wang 2020, Miao, Li & Su 2020), and models accommodating structural breaks with group-varying specifications (Lumsdaine et al. 2020).

Chapter 3: Panel VAR Model with Latent Group Structures

Abstract

Previous research has thoroughly examined univariate panel models with interactive fixed effects. This study explores the multivariate panel vector autoregression (PVAR) model incorporating group-based factors. Specifically, we consider a Panel VAR framework with heterogeneous coefficients, allowing for an unknown number of latent groups and group memberships. Additionally, it is a parsimonious structural model that is computationally straightforward. We derive the asymptotic distribution and prove the estimator's consistency as N and T approach infinity. The empirical application suggests that the group patterns may yield different outcomes than those of a conventional Panel VAR model.

3.1 Introduction

Since the introduction of the Panel VAR (PVAR) model by Holtz-Eakin et al. (1988), it has become a widely used tool in macroeconomic research. In contrast to reduced-form panel models, the PVAR framework treats all variables as endogenous and employs a dynamic, systematic approach to capture their simultaneous interactions. Extensions of the original PVAR model have been developed to enhance flexibility and address growing model complexity, including Factor PVARs, Large-Scale PVARs, and Bayesian PVARs (see Canova & Ciccarelli (2013) for a comprehensive review). Recent studies, particularly Tuğan (2021), have addressed cross-sectional dependence by employing interactive fixed effects, resulting in the development of the Panel VAR model with interactive fixed effects (PVAR-IFE). While this framework captures unit-specific, time-varying heterogeneity, it may suffer from over-parameterisation and lacks the structural interpretability of group-level patterns.

These models generally assume homogeneous dynamics across cross-sectional units or rely on observable characteristics to account for heterogeneity. However, in the context of macroeconomic panels, assuming homogeneity is often unrealistic. Countries, regions, or provinces may exhibit latent grouping structures shaped by their stages of development, institutional frameworks, or geographic and cultural proximityall of which influence their dynamic responses to shocks. Ignoring these unobserved groupings can lead to inconsistent estimates, biased impulse responses, and reduced forecast accuracy.

In this paper, we propose a Panel Vector Autoregression model with Grouped Fixed Effects (PVAR-GFE) that addresses these limitations by allowing for latent, endogenous group structures in both the intercepts and dynamic coefficients. Building on the grouped fixed effects methodology of Bonhomme & Manresa (2015) for univariate cases, our framework extends this approach to a multivariate Panel VAR setting. It eliminates the need for prior knowledge of group assignments and can effectively uncover hidden clusters through a computationally efficient iterative algorithm. Compared to the PVAR model with interactive fixed effects (PVAR-IFE), our model is more parsimonious and yields interpretable group-level heterogeneity, making it particularly well suited for macroeconomic applications.

The implications of capturing group-level heterogeneity in this manner are substantial. By recovering group-specific impulse response functions (IRFs), the PVAR-GFE framework enables researchers and policymakers to assess how different latent clusters respond to economic shocks. This has direct relevance for the design of regionally tailored macroeconomic policiesfor example, targeting fiscal incentives to groups of economies with similar spending multipliers, or analysing how energy consumption responds to financial development within structurally similar clusters of provinces. Furthermore, forecasts generated by the PVAR-GFE model are inherently more robust in heterogeneous panels, as they are conditioned on group-specific dynamics rather than on an average response that may not accurately represent any individual unit.

This modelling innovation is particularly relevant in modern macroeconomics, where increasing global integration coexists with persistent structural divergence. Countries and regions often respond differently to monetary policy shocks, fiscal interventions, or external disturbancesnot merely due to measurement error, but because of genuine differences in structural parameters across latent macroeconomic regimes. By uncovering and modelling these group-specific dynamics, the PVAR-GFE framework equips researchers with a sharper empirical lens to capture heterogeneity in transmission mechanisms. For instance, fiscal multipliers may be larger in high-debt regimes, while monetary responses may differ systematically between commodity exporting and manufacturing based economies. The ability to estimate group-specific impulse responses enhances not only explanatory power but also the formulation of data-driven, targeted policies. From a forecasting perspective, the PVAR-GFE model mitigates bias arising from misspecified homogeneity assumptions and enables more accurate, group-informed projections. In sum, this modelling approach supports a macroeconomic research agenda that is both structurally grounded and empirically responsive to the complexities of the contemporary global economy.

The model accounts for unobserved cross-sectional heterogeneity by allowing units to be partitioned into a finite number of latent groups, each with distinct intercepts. Estimation is carried out using an iterative least squares algorithm combined with a clustering procedure (K-means) to recover group membership. Under standard regularity conditions and assuming both the cross-sectional dimension N and time dimension T tend to infinity, We establish the NT consistency of the estimator and derive its asymptotic distribution, demonstrating that parameter estimates converge in probability to their true values. We provide detailed Monte Carlo simulations to demonstrate that the PVAR-GFE estimator is consistent under large N and large T, and the group assignment error is negligible when the number of groups is correctly specified. Compared to the PVAR-IFE and PVAR-GMM models, the PVAR-GFE approach provides a better balance between flexibility, bias correction, and computational simplicity.

The remainder of the paper is organised as follows. Section 2 introduces the PVAR-GFE model and its extensions. Section 3 outlines the estimation procedure under both infeasible and feasible scenarios. In Section 4, we present the key assumptions and derive the asymptotic properties of the estimators for the coefficients in the PVAR-GFE model. Section 5 reports Monte Carlo simulations based on various data-generating processes. Section 6 provides an empirical application examining the relationship between financial development, economic growth, and energy consumption in China. Section 7 concludes. All technical proofs are provided in Appendix A [17].

Throughout the paper, we adopt the following notation. For an $m \times n$ real matrix A, we denote its transpose by A', and its Frobenius norm by $|A|_F$. For an $m \times 1$ real vector a, we denote its Euclidean norm by $|a|_2$. The operator $\xrightarrow{P}$ denotes convergence in probability, and $\xrightarrow{D}$ denotes convergence in distribution. We write $(N,T) \to \infty$ to indicate that N and T jointly tend to infinity.

3.2 The Model

In this section, we provide the setup and identification for PVAR-GFE model with grouped heterogeneous coefficient[18].

[17] For this paper, we use MATLAB 2024b version as the main programming language for Monte Carlo Simulations and replication studies. In MATLAB, the key function is parfor, which enables parallel computation across multiple starting values to improve the efficiency of the PVAR-GFE estimation.

[18] For simplicity, we only demonstrate the PVAR-GFE model with grouped heterogeneous coefficient in the main content. For more details about homogeneous coefficient please refer to Appendix A

3.2.1 Heterogeneous Coefficient Set Up

The setup of PVAR-GFE(P) with heterogeneous coefficient as follows:

$$y_{m,i,t} = \sum_{p=1}^{P} \psi_{m,p,g_i}^{m} y_{m,i,t-p} + \sum_{p=1}^{P} \sum_{n \neq m}^{M} \phi_{n,p,g_i}^{m} y_{n,i,t-p} + \boldsymbol{x}_{i,t}' \boldsymbol{\gamma}_{m,g_i} + \alpha_{m,g_i,t} + \epsilon_{m,i,t} \tag{15}$$

In compact form, we have:

$$\boldsymbol{y}_{i,t} = \boldsymbol{z}_{i,t} \boldsymbol{\beta}_{g_i} + \boldsymbol{\alpha}_{g_i,t} + \boldsymbol{\epsilon}_{i,t} \tag{16}$$

Where

$$\underset{M\times 1}{\boldsymbol{y}_{i,t}} = \begin{bmatrix} y_{1,i,t} \\ \vdots \\ y_{M,i,t} \end{bmatrix}, \quad \underset{M\times \mathcal{F}}{\boldsymbol{z}_{i,t}} = \begin{bmatrix} \boldsymbol{z}_{1,i,t}' \\ \vdots \\ \boldsymbol{z}_{M,i,t}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_M \otimes \mathcal{Y}_{i,t-1}', & \boldsymbol{I}_M \otimes \mathcal{X}_{i,t-1}' \end{bmatrix}, \underset{\mathcal{F}_g\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\psi}_{g_i} \\ \boldsymbol{\phi}_{g_i} \\ \boldsymbol{\gamma}_{g_i} \end{bmatrix}$$

$$\underset{MP\times 1}{\mathcal{Y}_{i,t}} = \begin{bmatrix} \boldsymbol{y}_{i,t} \\ \vdots \\ \boldsymbol{y}_{i,t-p+1} \end{bmatrix}', \quad \underset{MK\times 1}{\mathcal{X}_{i,t}} = \begin{bmatrix} \boldsymbol{x}_{i,t} \\ \vdots \\ \boldsymbol{x}_{i,t-p+1} \end{bmatrix}', \underset{M\times 1}{\boldsymbol{\alpha}_{g_i,t}} = \begin{bmatrix} \alpha_{1,g_i,t} \\ \vdots \\ \alpha_{M,g_i,t} \end{bmatrix}, \quad \underset{M\times 1}{\boldsymbol{\epsilon}_{i,t}} = \begin{bmatrix} \boldsymbol{\epsilon}_{i,t,1} \\ \vdots \\ \boldsymbol{\epsilon}_{i,t,M} \end{bmatrix},$$

Let $i \in \{1, 2, \ldots, N\}$ index the cross-sectional units, $t \in \{1, 2, \ldots, T\}$ the time periods, $p \in \{1, 2, \ldots, P\}$ the time lags, and $m \in \{1, 2, \ldots, M\}$ the number of endogenous (dependent) variables in the PVAR system. The dependent variables $y_{m,i,t}$ correspond to the m-th equation of the VAR for unit i at time t. Each unit also includes a $(K \times 1)$ vector of exogenous variables $\boldsymbol{x}_{i,t}$, which appears in every equation for all dependent variables, with equation-specific coefficients $\boldsymbol{\gamma}_m \in \mathbb{R}^K$, assumed to be heterogeneous across units. In compact form, $\boldsymbol{\beta}_{g_i}$ denotes the vector of coefficients to be estimated. This includes all autoregressive and cross-lagged coefficients for each lag p, while $\epsilon_{m,i,t}$ represents the idiosyncratic errors. Assume there are G total groups, and let g index an element of the set $1, 2, \ldots, G$. Define g_i as the group membership index of unit i, indicating the group to which unit i belongs. We thus allow for group-specific intercepts $\alpha_{m,g_i,t}$, which may vary over time to capture dynamic heterogeneity at the group level across equations. Let $\mathbb{G}_g$ denote the number of cross-sectional units in group g. Group memberships are assumed to be exclusive in structure, such that for any distinct units i and j,

$\alpha_{m,g_i,t} \neq \alpha_{m,g_j,t}$ if $i \neq j$. One key assumption of the model is that both the number of groups and group memberships remain constant, as discussed in Bonhomme & Manresa (2015), Su et al. (2016). For simplicity, let $\mathcal{F}$ denote the total number of parameters, defined as $\mathcal{F} = M^2P + MK$. Define $\boldsymbol{z}_{i,t} \in \mathbb{R}^{M \times \mathcal{F}}$ to stack all lagged endogenous regressors and common covariates. The vector $\boldsymbol{\beta}_{g_i} \in \mathbb{R}^{\mathcal{F}}$ collects all slope parameters.

In matrix form, a simple bivariate PVAR-GFE(1) model with heterogeneous coefficient is:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \begin{bmatrix} y_{1,i,t-1} & y_{2,i,t-1} & 0 & 0 \\ 0 & 0 & y_{1,i,t-1} & y_{2,i,t-1} \end{bmatrix} \begin{bmatrix} \psi^1_{1,1,g_i} \\ \phi^1_{2,1,g_i} \\ \psi^2_{2,1,g_i} \\ \phi^2_{1,1,g_i} \end{bmatrix}$$
$$+ \begin{bmatrix} x_{i,t-1} & 0 \\ 0 & x_{i,t-1} \end{bmatrix} \begin{bmatrix} \gamma_{1,1,g_i} \\ \gamma_{2,1,g_i} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,i,t} \\ \epsilon_{2,i,t} \end{bmatrix}$$

3.2.2 Identification

The model in Equation 16 includes two sets of parameters. The first comprises the slope coefficients $\boldsymbol{\beta}_{g_i} \in \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ denotes the sample space comprising subsets of $\mathbb{R}^{\mathcal{F}}$. The second comprises the group-specific fixed effects $\boldsymbol{\alpha}_{g_i} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ denotes the sample space of fixed effects, defined as subsets of $\mathbb{R}^{MT}$. Let τ denote a specific partitioning scheme selected from the set of possible group membership assignments g_i, which allocates the N cross-sectional units into G distinct groups. Formally, $\tau \in \boldsymbol{\Gamma}_G$, where $\boldsymbol{\Gamma}_G$ represents the set of all possible partitions of the units into G groups. We denote $\hat{\boldsymbol{\beta}}_{g_i}$, $\hat{\boldsymbol{\alpha}}_{g_i}$, and $\hat{\tau}$ as the estimators of the slope coefficients, grouped fixed effects, and group membership assignment, respectively. Equation 16 is estimated by least squares (LS) through an iterative procedure:

$$\left(\hat{\boldsymbol{\beta}}_{g_i}, \hat{\boldsymbol{\alpha}}_{g_i}, \hat{\tau}\right) = \underset{(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\in\boldsymbol{\Omega}\times\boldsymbol{\Theta}^{G\times TM}\times\boldsymbol{\Gamma}_G}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i;g_i=g} \sum_{t=1}^{T} \|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta}_{g_i} - \boldsymbol{\alpha}_{g_i,t}\|_2^2 \tag{17}$$

Following Bai (2009), our analysis begins by considering the case in which the group structure is known. Specifically, assume that the true number of groups G is known and that the membership g_i for each unit is correctly identified such that $g_i = g_i^0$. The model becomes:

$$\boldsymbol{y}_{i,t} = \boldsymbol{z}_{i,t}\boldsymbol{\beta}_{g_i} + \boldsymbol{\alpha}_{g_i^0,t} + \boldsymbol{\epsilon}_{i,t} \tag{18}$$

All that remains is to estimate Equation 16 and obtain $\hat{\boldsymbol{\beta}}_{g_i}$ and $\hat{\boldsymbol{\alpha}} g_i^0, t$. This is achieved directly by ordinary least squares, incorporating interaction time dummies t for the given group assignment g_i^0:

$$(\hat{\boldsymbol{\beta}_{g_i}}, \hat{\boldsymbol{\alpha}}_{g_i^0,t}) = \underset{\boldsymbol{\beta} \in \boldsymbol{\Omega}}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i;g_i=g} \sum_{t=1}^{T} \left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta}_{g_i} - \boldsymbol{\alpha}_{g_i^0,t} \right\|_2^2$$

However, in practice, neither the number of groups nor the group memberships is known. In this case, the estimation of Equation 16 must be conducted using an iterative algorithm:

Algorithm 1 PVAR-GFE Algorithm

1: Initialisation.

Choose the total number of groups G and the number of replications in the K-means algorithm, which randomly assigns each unit into the G groups by OLS without grouped fixed effects. Then, set the initial values of $\left(\boldsymbol{\beta}_{g_i}^{(0)}, \boldsymbol{\alpha}_{g_i}, \tau^{(0)}\right)$. Let $s = \{1, 2, \ldots, S\}$ denote the iteration index within a single simulation.

2: Update the group members

Update $\tau^{(s)}$ within each group based on the current values of $\boldsymbol{\beta}_{g_i}^{(s)}$ and $\boldsymbol{\alpha}_{g_i}^{(s)}$

$$g_i^{(s+1)} = \underset{g \in \{1,\ldots,G\}}{\operatorname{argmin}} \sum_{t=1}^{T} \left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta}_{g_i}^{(s)} - \boldsymbol{\alpha}_{g_i,t}^{(s)} \right\|_2^2$$

3: Update the parameter of interest

Update $\boldsymbol{\beta}^{(s)}$ based on the current values of $\boldsymbol{\alpha}_{g_i}^{(s+1)}$ and $\tau^{(s+1)}$ using OLS:

$$\boldsymbol{\beta}_{g_i}^{(s+1)} = \underset{\boldsymbol{\beta}_{g_i} \in \boldsymbol{\Omega}^G}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i; g_i = g} \sum_{t=1}^{T} \left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta}_{g_i} - \boldsymbol{\alpha}_{g_i,t^{s+1}}^{(s+1)} \right\|_2^2$$

4: Check numerical convergence.

Set $s = s + 1$ and go to Step 2 till the norm of $\boldsymbol{\beta}_{g_i}$ is lower than the convergence error set by researcher.

3.3 Asymptotic Theory

3.3.1 Consistency

In this section we will present a set of assumptions that make consistency holds.

Assumption 1 *Stationarity. Define the companion matrix* F *as*

$$\mathrm{F} = \begin{bmatrix} \boldsymbol{\theta}_1 & \boldsymbol{\theta}_2 & \boldsymbol{\theta}_3 & \cdots & \boldsymbol{\theta}_{p-1} & \boldsymbol{\theta}_p \\ \mathbf{I}_M & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_M & \mathbf{0} \end{bmatrix}$$

The eigenvalues of the matrix F *satisfy the characteristic equation:*

$$\left|\mathbf{I}_M \lambda^p - \boldsymbol{\theta}_1 \lambda^{p-1} - \boldsymbol{\theta}_2 \lambda^{p-2} - \cdots - \boldsymbol{\theta}_p\right| = 0.$$

Hence, the PVAR-GFE model is covariance-stationary as long as $|\lambda| < 1$ *for all values of* λ*. Equivalently, it is stationary if all values of* z *satisfying*

$$\left|\mathbf{I}_M - \boldsymbol{\theta}_1 z - \boldsymbol{\theta}_2 z^2 - \cdots - \boldsymbol{\theta}_p z^p\right| = 0$$

lie outside the unit circle.

In the PVAR framework, stationarity is a crucial assumption for the theoretical results presented later in the paper. While the grouped fixed effects estimator can also be applied in a non-stationary setting, this extension lies beyond the scope of the present study.

Assumption 2 *There exists a constant* $\mathbb{M} > 0$ *such that:*

a *Compactness. Parameter space* $\boldsymbol{\Omega}$ *and* $\boldsymbol{\Theta}$ *are compact subsets of* $\mathbb{R}^{\mathcal{K}}$ *and* $\mathbb{R}$ *with* $\sup_{\boldsymbol{\beta} \in \boldsymbol{\Omega}} \|\boldsymbol{\theta}\| \leq M$

b *Strict exogeneity.*$\mathbb{E}\left[\boldsymbol{\epsilon}_{i,t} \mid \mathbf{x}_{i,s}\right] = 0$ *for all* $s, t,$ *and all* i*.*

c $\mathbb{E}(\boldsymbol{\epsilon}) = 0$ *and* $E\left\|\boldsymbol{\epsilon}\right\|^8 < \infty$*.*

d $\mathbb{E}(\boldsymbol{\epsilon}_{i,t}' \boldsymbol{\epsilon}_{j,t}) = 0$ *if* $i \neq j$*.*

e $\mathbb{E}(\boldsymbol{\epsilon}_{i,t}' \boldsymbol{\epsilon}_{i,s}) = 0$ *if* $t \neq s$*.*

f *The number of equations in the model, namely* M *is small and fixed.*

g $\left|\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}\left\|\mathbf{z}_{i,t}' \boldsymbol{\epsilon}_{i,t}\right\|^2\right| \leq \mathbb{M}$

h *Let* $\bar{\mathbf{z}}_{g\wedge\widetilde{g},t}$ *denote* $\bar{\mathbf{z}}_{g\wedge\widetilde{g},t} = \frac{\sum_{i=1}^{N}\mathbf{1}\{g_i^0 = g\}\mathbf{1}\{g_i = \widetilde{g}\}\mathbf{z}_{it}}{\sum_{i=1}^{N}\mathbf{1}\{g_i^0 = g\}\mathbf{1}\{g_i = \widetilde{g}\}}$*. For all groupings* $\tau = \{g_1, \ldots, g_N\} \in \Gamma_G$*, we define* $\widehat{\rho}(\tau)$ *as the minimum eigenvalue of the following matrix:*

$$\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0\wedge g_i,t}\right)\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0\wedge g_i,t}\right)'.$$

Then $\operatorname{plim}_{N,T\to\infty}\min_{\tau\in\Gamma_G}\widehat{\rho}(\tau) = \rho > 0$.

Assumption (a) ensures that no closed-form optimisation problem yields an optimal solution outside a closed and bounded parameter space. Assumption (b) states that the exogenous variables are independent of the idiosyncratic error terms. Assumption (c) requires that the idiosyncratic terms have zero mean and are bounded by a finite eighth moment. Assumption (d) imposes the absence of cross-sectional dependence among the idiosyncratic terms, such that any dependence is fully captured by the grouped fixed effects $\boldsymbol{\alpha}_{g_i,t}$. Assumption (e) assumes no serial correlation within the idiosyncratic terms, which is equivalent to stating that the number of lagged dependent variables, M, is sufficient to account for serial dependence in the model. Assumption (f) asserts that the number of equations does not affect the asymptotic properties of the estimator. Assumption (g) ensures that the second moment of the cross-product $\mathbf{z}_{i,t}'\boldsymbol{\epsilon}_{i,t}$ is uniformly bounded in N and T. Assumption (h) imposes a full-rank condition, requiring sufficient variation across individuals within the panel.

We first define the objective function as:

$$\begin{aligned}\hat{\mathrm{Q}}(\boldsymbol{\beta}_g, \boldsymbol{\alpha}, \tau) &= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta}_g - \boldsymbol{\alpha}_{g_i,t}\right\|^2 \\ &= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t} + \boldsymbol{z}_{i,t}(\boldsymbol{\beta}_g^0 - \boldsymbol{\beta}_g) + \boldsymbol{\alpha}_{g_i^0,t}^0 - \boldsymbol{\alpha}_{g_i,t}\right\|^2\end{aligned} \tag{19}$$

As stated in Newey & McFadden (1994), the consistency of extremum estimators requires $\operatorname*{plim}_{N,T\to\infty}\sup_{(\boldsymbol{\beta}_g,\boldsymbol{\alpha},\tau)\in\boldsymbol{\Omega}\times\boldsymbol{\Theta}^{GTM}\times\boldsymbol{\Gamma}_G}\left|\hat{\mathrm{Q}}(\boldsymbol{\beta}_g, \boldsymbol{\alpha}, \tau) - \mathrm{Q}(\boldsymbol{\beta}_g, \boldsymbol{\alpha}, \tau)\right| = o_p(1)$ when N and T go to infinity. However, the grouped fixed effects parameter $\boldsymbol{\alpha}$ grows as T increases. The limit function could be unavailable as there would be infinite number of parameters within $\boldsymbol{\alpha}$ when T go to infinity. Therefore, we construct an auxiliary objective function $\tilde{\mathrm{Q}}_{NT}(\beta_g, \alpha, \gamma)$ which is also dependent on N and T and it could reach the unique minimum to the true value of parameters. We first show that $\hat{\mathrm{Q}}(\beta_g, \alpha, \gamma)$ converges uniformly to $\tilde{\mathrm{Q}}(\beta_g, \alpha, \gamma)$ and then $\tilde{\mathrm{Q}}(\beta_g, \alpha, \gamma)$ is minimised uniquely at the true value of parameters.

By Lemma 1 and Lemma 2 we have the following Theorem:

Theorem 1

$$\hat{\boldsymbol{\beta}}_g \xrightarrow{p} \boldsymbol{\beta}_g \quad as\ (N, T) \to \infty$$
$$\frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \boldsymbol{\alpha}^0_{g_i^0, t} - \hat{\boldsymbol{\alpha}}_{\hat{g}_i, t} \right\|^2 = o_p(1)$$

3.3.2 Asymptotic Equivalence of Heterogeneous Coefficient Model

Assumption 3 *a Define*

$$M(\gamma, g, \tilde{g}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\} \mathbf{1}\left\{g_i = \tilde{g}\right\} \cdot$$
$$\begin{pmatrix} \frac{1}{MT}\sum_{t=1}^{T} \mathbf{z}_{it}\mathbf{z}'_{it} & \frac{1}{\sqrt{MT}}\mathbf{z}_{i1} & \frac{1}{\sqrt{MT}}\mathbf{z}_{i2} & \cdots & \frac{1}{\sqrt{MT}}\mathbf{z}_{iT} \\ \frac{1}{\sqrt{MT}}\mathbf{z}'_{i1} & 1 & 0 & \cdots & 0 \\ \frac{1}{\sqrt{MT}}\mathbf{z}'_{i2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{MT}}\mathbf{z}'_{iT} & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Let $\hat{\rho}(\gamma, g, \tilde{g})$ denote the minimum eigenvalue of $M(\gamma, g, \tilde{g})$. There exist $\hat{\rho}$ and $\rho > 0$ such that $\hat{\rho} \to \rho$ and $\forall g$, $\min_{\tau \in \boldsymbol{\Gamma}_G} \max_{\tilde{g} \in \mathbb{G}} \hat{\rho}(\gamma, g, \tilde{g}) > \hat{\rho}$

b For all $g \neq \widetilde{g}$ there exists a constant $c_{g,\widetilde{g}} > 0$ such that $\operatorname{plim}_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N} D^0_{g\widetilde{g}_i} \geq c_{g,\widetilde{g}}$ and, for all $i \in \{1, \ldots, N\}$, $\operatorname{plim}_{T\to\infty} D^0_{g\widetilde{g}_i} \geq c_{g,\widetilde{g}}$, where

$$D^0_{g\widetilde{g}_i} = \frac{1}{MT} \sum_{t=1}^{T} \left\| \mathbf{z}'_{it}\left(\boldsymbol{\beta}^0_g - \boldsymbol{\beta}^0_{\widetilde{g}}\right) + \boldsymbol{\alpha}^0_{gt} - \boldsymbol{\alpha}^0_{\widetilde{g}t} \right\|^2$$

.

c There exists a constant $M^* > 0$ such that:

$$\sup_{i \in \{1,\ldots,N\}} \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} \left\|\mathbf{z}_{it}\right\|^4 \geq M^*\right) = o\left(T^{-\delta}\right) \ \textit{for all } \delta > 0.$$

d For all constants $c > 0$:

$$\sup_{i\in\{1,\ldots,N\}} \Pr\left(\left\|\frac{1}{MT}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{it}\mathbf{z}_{it}\right\|^2 > c\right) = o\left(T^{-\delta}\right) \textit{ for all } \delta > 0$$

Assumption 3 a is an identification assumption that is commonly used in previous studies (Bonhomme & Manresa (2015), Okui & Wang (2021), Leng et al. (2021)), which rules out the possibility of multicollinearity. When the model includes only heterogeneous coefficients without GFE estimators, the identification matrix reduces to:

$$M(\gamma, g, \tilde{g}) = \frac{1}{MN}\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0 = g\right\}\mathbf{1}\left\{g_i = \tilde{g}\right\}\begin{pmatrix} \mathbf{z}_{i1}\mathbf{z}_{i1}' & 0 & \ldots & 0 \\ 0 & \mathbf{z}_{i2}\mathbf{z}_{i2}' & \ldots & \ldots \\ \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & 0 & \mathbf{z}_{iT}\mathbf{z}_{iT}' \end{pmatrix}$$

. Assumption 3 b is similar as Assumption 7 b

By Lemma 10 and Lemma 11 we have the following Theorem:

Theorem 2 *For all $\delta > 0$,*

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{g}} = \tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} + o_p(T^{-\delta}) \quad \textit{as } (N, T) \to \infty \tag{20}$$

By Lemma 10, Lemma 11 and Theorem 2, we have:

Theorem 3 *For all $\delta > 0$,*

$$\hat{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}} + o_p(T^{-\delta}) \quad \textit{as } (N, T) \to \infty \tag{21}$$

Next we proof that the group is estimated sup-consistently.

Theorem 4 *for all $\delta > 0$ and as $NT^{-\delta}$ tends to zero:*

$$\Pr\left(\sup_{i\in\{1,\ldots,N\}}\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) - g_i^0\right| \neq 0\right) = o(1) \tag{22}$$

3.3.3 Asymptotic Distribution of Heterogeneous Coefficient Model

In this section we derive the asymptotic distribution based on the asymptotic equivalence between feasible estimators and infeasible estimators.

Assumption 4 *a For all i, j, t, and g we have:* $\mathbb{E}\left(\mathbf{1}\{g_i^0 = g\}\mathbf{z}'_{jt}\boldsymbol{\epsilon}_{it}\right) = 0$.

b There exist positive definite matrices Σ_{β_g} and Ω_{β_g} such that

$$\Sigma_{\beta_g} = \text{plim}_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{g_i^0 = g\}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0}\right)',$$

$$\Omega_{\beta_g} = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[\mathbf{1}\{g_i^0 = g\}\mathbf{1}\{g_j^0 = g\}\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}'_{js}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)\left(\mathbf{z}_{js} - \bar{\mathbf{z}}_{g_j^0 s}\right)'\right].$$

c As N and T tend to infinity: $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{g_i^0 = g\}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)'\boldsymbol{\epsilon}_{it} \xrightarrow{d} \mathbb{N}\left(0, \Omega_{\beta_g}\right)$.

d For all (g, t): $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left(\mathbf{1}\left\{g_i^0 = g\right\}\mathbf{1}\left\{g_j^0 = g\right\}\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{jt}\right) = \omega_{gt} > 0$

e For all (g, t), and as N and T tend to infinity: $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0 = g\right\}\boldsymbol{\epsilon}_{it} \xrightarrow{d} \mathbb{N}\left(0, \omega_{gt}\right)$.

Where $\bar{\mathbf{z}}_{g_i^0 t} = \frac{\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0 = g\right\}\mathbf{z}_{it}}{\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0 = g\right\}}$ is the mean of $\mathbf{z}_{it}$ across the true group assignment g_i^0. Thus we have the following Corollary:

Corollary 1 *Let Assumption 6, 3 and 4 holds, when N and T tends to infinity and for all $\delta > 0$, $NT^{-\delta} \to 0$, we have:*

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}_{\boldsymbol{g}} - \boldsymbol{\beta}^{\mathbf{0}}_{\boldsymbol{g}}) \xrightarrow{d} \mathbb{N}\left(0, \Sigma_{\beta_g}^{-1}\Omega_{\beta}\Sigma_{\beta_g}^{-1}\right) \tag{23}$$

3.4 Impulse Response Functions

In most VAR models, deriving impulse response functions (IRFs) is essential for understanding how shocks propagate through the system via specific variables. IRFs provide important insights in many empirical macroeconomic applications under various types of shocks, such as productivity shocks and monetary policy shocks (see Romer & Romer (2017), Acemoglu et al. (2019)). Previous studies on the construction of confidence bands for IRFs in panel VAR settings typically rely on Monte Carlo simulations or bootstrap resampling methods. However, these approaches can be computationally intensive in models with grouped fixed effects, due to the added dimension introduced by the number of groups, G. To address this challenge, we adopt the local projections (LP) method proposed by Jordà (2005) to estimate the IRFs. LP is now widely used in multivariate analysis due to its computational simplicity and flexibility. By regressing future values of the dependent variables on current independent variables using ordinary least squares (OLS) at each forecast horizon, IRFs can be derived in a straightforward manner. Recent theoretical work demonstrates that LP is both numerically and asymptotically equivalent to VAR-based IRF estimation under general conditions (see Plagborg-Møller & Wolf (2021), Li et al. (2022)). However, Mei et al. (2023) show that LP may suffer from an implicit Nickell bias in panel VAR settings, particularly when estimated using a univariate fixed effects model. This arises due to violations of the strict exogeneity assumption, and the resulting asymptotic bias can distort inference by affecting the size of hypothesis tests that rely on fixed effects estimators. We show that this bias does not arise in the PVAR-GFE framework, as the inclusion of grouped fixed effects controls for unobserved heterogeneity across units. The h-step-ahead LP specification for the PVAR model is given as follows:

$$y_{i,t+h} = \alpha_{g_i,t,h} + x_{i,t}\beta_{g_i,h} + \varepsilon_{i,t+h}, \quad h = 0, 1, \ldots, H-1, \tag{24}$$

Where y is the univariate dependent variable that we are interested, β_h is the parameter of interest which forms the IRF.

3.5 Remarks

3.5.1 Determine the Number of Groups

Following Bai & Ng (2002), Bai (2009), Bonhomme & Manresa (2015), we use an iterative BIC method to decide the optimal model:

$$\begin{aligned} BIC(G,P) = & \frac{1}{MNT}\sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}^{(P)}\hat{\boldsymbol{\beta}}_{\boldsymbol{g_i}}^{(G)} - \hat{\boldsymbol{\alpha}}_{\hat{g}_i,t}^{(G)} \right\|_2^2 \\ & + \widehat{\sigma}^2 \frac{GTM + N + M^2P + MK}{NT}\ln(NT) \end{aligned} \tag{25}$$

where $\widehat{\sigma}^2$ is an estimate of the error variance. In each iteration, we fix the number of lagged dependent variables and compute the Bayesian Information Criterion (BIC) for a range of values for the number of groups. The specification that minimises the BIC is selected. We then compare the BIC values obtained across iterations. The optimal number of lagged dependent variables and the number of groups are jointly determined as follows:

$$(\hat{G},\hat{P}) = \underset{G\in\{1,\ldots,G_{\max}\},P\in\{1,\ldots,P_{\max}\}}{\operatorname{argmin}} BIC(G,P) \tag{26}$$

While the proposed PVAR-GFE model offers substantial flexibility in capturing unobserved heterogeneity through latent group structures. We do acknowledge the challenges of practical implementation. The most diffcult thing is to correctly determine the number of groups G when group memberships are unknown. Since the estimation is non-convex, it requires iterative procedures that are sensitive to initialisation. To mitigate this, multiple starting values must be explored to avoid local minima, and the estimation must be repeated across a grid of candidate group numbers. Furthermore, this process must be nested within the selection of the lag order P, as both dimensions jointly influence model fit. The result is a highly computationally intensive procedure, particularly in settings with large N, T, or when $G_{\max}$ and $P_{\max}$ are large. These complexities underscore the trade-off between model richness and computational tractability, and they motivate future research on more efficient algorithms.

3.6 Monte Carlo Simulation

To evaluate the performance of the proposed PVAR-GFE model, we conduct a series of Monte Carlo simulations under various data-generating processes (DGPs). We propose the following DGP model for DGP 1-3:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} \theta_1, g & \theta_2, g \\ \theta_4, g & \theta_3, g \end{bmatrix}}_{\Theta_g^0} \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix}, \tag{27}$$

Where we set Θ_g^0 according with different number of groups g. The coefficient of Θ_g^0 confirms this model is stationary. The errors are i.i.d and it follows a standard normal distribution. We replicate 2,000 times for the simulation. We assume the DGP of grouped fixed effects follows $\mathbf{1}_{\{g_i=g\}}\boldsymbol{\alpha_{g_i,t}} \overset{i.i.d.}{\sim} N(g_i, 1)$ for different g. Hence, we have the following MC[19]:

DGP 1 (2 Groups): This scenario includes one informative group and one null group:

$$\Theta_g^0 = \underbrace{\begin{pmatrix} 0.6 & 0.3 \\ 0.2 & 0.6 \end{pmatrix}}_{\Theta_1^0} \text{and} \underbrace{\begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}}_{\Theta_2^0}$$

DGP 2 (3 Groups): In this DGO, a third group is added with a distinct dynamic structure:

$$\Theta_g^0 = \underbrace{\begin{pmatrix} 0.6 & 0.3 \\ 0.2 & 0.6 \end{pmatrix}}_{\Theta_1^0}, \underbrace{\begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}}_{\Theta_2^0} \text{and} \underbrace{\begin{pmatrix} -0.2 & 0.3 \\ 0.5 & 0.8 \end{pmatrix}}_{\Theta_3^0}$$

DGP 3 (4 Groups): A fourth group with its unique autoregressive dynamics is introduced:

$$\Theta_g^0 = \underbrace{\begin{pmatrix} 0.6 & 0.3 \\ 0.2 & 0.6 \end{pmatrix}}_{\Theta_1^0}, \underbrace{\begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}}_{\Theta_2^0}, \underbrace{\begin{pmatrix} -0.2 & 0.3 \\ 0.5 & 0.8 \end{pmatrix}}_{\Theta_3^0} \text{and} \underbrace{\begin{pmatrix} 0.4 & 0.3 \\ 0.6 & 0.5 \end{pmatrix}}_{\Theta_4^0}$$

In each case, the proportion of group membership is set to be evenly distributed across groups unless otherwise specified. Table 12 reports the results from the

[19] For simplicity, we only show the mean, bias and RMSE of Θ_1^0

Monte Carlo simulations for DGPs 1 to 3, focusing on the estimation performance of the group-specific coefficient matrix Θ_1^0. The simulations were conducted under varying panel dimensions, with the number of cross-sectional units $N \in \{50, 100, 300\}$ and time periods $T \in \{50, 100, 300\}$. For each DGP, we present the average estimated coefficient matrix, the bias, and the root mean squared error (RMSE) across 2,000 replications. The results clearly demonstrate the strong finite-sample performance of the proposed PVAR-GFE estimator. Across all data-generating processes, the estimated mean values closely approximate the true coefficient matrix, indicating negligible bias. Notably, the bias terms are consistently close to zero, even in relatively small panels (e.g., $N = 50$, $T = 50$). This provides empirical support for the unbiasedness of the estimator. As expected, both bias and RMSE decline with increases in either N or T. This behaviour aligns with the asymptotic properties, wherein consistency of the estimator is guaranteed as $N, T \to \infty$. Moreover, the results are robust to different groups settings in the data-generating process. The introduction of additional latent groups in DGPs 2 and 3 does not compromise the accuracy of the estimator. While marginal increases in bias and RMSE are observed with additional groups and in smaller panels, these deviations remain minimal and diminish when N and T increase.

Next we demonstrate the effectiveness if we increase both the number of dependent variables m and the number of groups g. We propose the following DGP model for DGP 4-6:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \\ y_{3,i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} \theta_1, g & \theta_2, g & \theta_3, g \\ \theta_4, g & \theta_5, g & \theta_6, g \\ \theta_7, g & \theta_8, g & \theta_9, g \end{bmatrix}}_{\Theta_g^0} \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \\ y_{3,i,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \\ \alpha_{3,g_i,t} \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \\ e_{3,i,t} \end{bmatrix},$$

Where we set Θ_g^0 according with different number of groups g and the number of dependent variables m. The coefficient of Θ_g^0 confirms this model is stationary. The errors are i.i.d and it follows a standard normal distribution. We replicate 2,000 times for the simulation. We assume the DGP of grouped fixed effects follows

$\mathbf{1}_{\{g_i=g\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}} \overset{i.i.d.}{\sim} N(g_i, 1)$ for different g. We propose the following DGP 4 to 6:

DGP 4 (2 Groups): This scenario includes one informative group and one null group:

$$\Theta_g^0 = \underbrace{\begin{pmatrix} 0.6 & 0.3 & -0.5 \\ 0.2 & 0.6 & -0.7 \\ 0.2 & 0.4 & 0.6 \end{pmatrix}}_{\Theta_1^0} \text{ and } \underbrace{\begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}}_{\Theta_2^0}$$

DGP 5 (3 Groups): In this DGO, a third group is added with a distinct dynamic structure:

$$\Theta_g^0 = \underbrace{\begin{pmatrix} 0.6 & 0.3 & -0.5 \\ 0.2 & 0.6 & -0.7 \\ 0.2 & 0.4 & 0.6 \end{pmatrix}}_{\Theta_1^0}, \underbrace{\begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}}_{\Theta_2^0} \text{ and } \underbrace{\begin{pmatrix} -0.2 & 0.3 & 0.4 \\ 0.5 & 0.8 & -0.4 \\ 0.1 & 0.6 & -0.3 \end{pmatrix}}_{\Theta_3^0}$$

DGP 6 (4 Groups): A fourth group with its unique autoregressive dynamics is introduced:

$$\Theta_g^0 = \underbrace{\begin{pmatrix} 0.6 & 0.3 & -0.5 \\ 0.2 & 0.6 & -0.7 \\ 0.2 & 0.4 & 0.6 \end{pmatrix}}_{\Theta_1^0}, \underbrace{\begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}}_{\Theta_2^0}, \underbrace{\begin{pmatrix} -0.2 & 0.3 & 0.4 \\ 0.5 & 0.8 & -0.4 \\ 0.1 & 0.6 & -0.3 \end{pmatrix}}_{\Theta_3^0} \text{ and } \underbrace{\begin{pmatrix} 0.4 & 0.3 & 0.2 \\ 0.6 & 0.5 & 0.4 \\ 0.3 & -0.7 & 0.3 \end{pmatrix}}_{\Theta_4^0}$$

Table 13 presents the Monte Carlo simulation results corresponding to DGPs 4 through 6, which were designed to evaluate the performance of the proposed estimator in settings with an increased number of dependent variables ($m = 3$) and latent groups ($g = 2, 3, 4$). The simulations were conducted over varying panel dimensions, specifically $N, T \in \{50, 100, 300\}$, and replicated 2,000 times. Across all DGPs, the mean estimates of the autoregressive matrices Θ_g^0 closely approximate their true values, with consistently low bias and RMSE across different cases. For DGP 4, which contains one dynamic and one null group, the bias is virtually negligible even in the smallest sample setting ($N = 50$, $T = 50$). The RMSE values are likewise extremely small. DGP 5 contains a third group exhibiting a different dynamic structure. Despite this complexity, the estimator maintains strong performance in terms of mean accuracy, bias, and RMSE. Importantly, we observe that results improve steadily as either N or T increases, demonstrating the consistency of the estimator in both dimensions. For instance, at $N = 300$, $T = 300$, the estimator's performance approaches nearly perfectly recovers the true parameter values. DGP 6 further increases model complexity by introducing a fourth latent group. Even under this challenging case, the estimator continues to perform robustly. While minor increases in bias and RMSE are evident in smaller panels, the overall magnitude of these errors remains limited. Overall, these simulation results offer strong evidence for the practical efficacy of the PVAR-GFE estimator, especially in applications involving latent heterogeneity and multivariate dynamics.

3.6.1 BIC values

In this section, we assess the capability of the Bayesian Information Criterion (BIC) to correctly identify the true model across various configurations. The analysis is based on Monte Carlo simulations conducted under different combinations of time periods $T = \{50, 100, 300\}$, cross-sectional units $N = \{50, 100, 300\}$, and group numbers $G = \{2, 3, 4, 5\}$[20]. Table 14 presents the accuracy ratesexpressed as percentages with which BIC selects the correct model. As seen in Table 14, BIC demonstrates excellent performance in identifying the correct number of groups when the number of groups is small (e.g., $G = 2$ or $G = 3$), achieving nearly perfect selection rates across all settings. However, its accuracy decreases as the true number of groups increases. For instance, when $G = 5$, the correct model selection rate falls significantly, especially for small T and N but improves markedly as both dimensions grow. This highlights the sensitivity of BIC to the complexity of the data structure and emphasises the importance of large sample sizes for consistent model selection in more intricate scenarios.

[20]In this simulation, we set $G\_max = 8$

Table 12: Monte Carlo Simulation

		DGP 1			DGP 2			DGP 3		
N	T	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
50	50	$\begin{bmatrix} 0.599 & 0.300 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.000 \\ 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.001 \\ 0.000 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.598 & 0.300 \\ 0.202 & 0.594 \end{bmatrix}$	$\begin{bmatrix} -0.002 & 0.000 \\ 0.002 & -0.006 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.001 \\ 0.001 & 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.592 & 0.312 \\ 0.226 & 0.555 \end{bmatrix}$	$\begin{bmatrix} -0.008 & 0.012 \\ 0.026 & -0.045 \end{bmatrix}$	$\begin{bmatrix} 0.002 & 0.002 \\ 0.004 & 0.008 \end{bmatrix}$
	100	$\begin{bmatrix} 0.599 & 0.300 \\ 0.199 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & -0.000 \\ -0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.598 & 0.301 \\ 0.201 & 0.598 \end{bmatrix}$	$\begin{bmatrix} -0.002 & 0.001 \\ 0.001 & -0.002 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.597 & 0.306 \\ 0.213 & 0.578 \end{bmatrix}$	$\begin{bmatrix} -0.003 & 0.006 \\ 0.013 & -0.022 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.001 \\ 0.002 & 0.004 \end{bmatrix}$
	300	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.599 & 0.300 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.000 \\ 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.599 & 0.302 \\ 0.203 & 0.594 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.002 \\ 0.003 & -0.006 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.001 & 0.001 \end{bmatrix}$
100	50	$\begin{bmatrix} 0.599 & 0.300 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & -0.000 \\ -0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.598 & 0.300 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.002 & 0.000 \\ 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.597 & 0.303 \\ 0.205 & 0.592 \end{bmatrix}$	$\begin{bmatrix} -0.003 & 0.003 \\ 0.005 & -0.008 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.001 \\ 0.001 & 0.001 \end{bmatrix}$
	100	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 \\ -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.599 & 0.300 \\ 0.200 & 0.598 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.000 \\ 0.000 & -0.002 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$
	300	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 \\ -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 \\ 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 \\ 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$
300	50	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 \\ 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 \\ -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$
	100	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$
	300	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 \\ 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 \\ -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 \\ 0.000 & 0.000 \end{bmatrix}$

Table 13: Monte Carlo Simulation

N	T	DGP 4 Mean	DGP 4 Bias	DGP 4 RMSE	DGP 5 Mean	DGP 5 Bias	DGP 5 RMSE	DGP 6 Mean	DGP 6 Bias	DGP 6 RMSE
50	50	$\begin{bmatrix} 0.599 & 0.300 & -0.500 \\ 0.201 & 0.599 & -0.700 \\ 0.199 & 0.400 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.000 & 0.000 \\ 0.001 & -0.001 & -0.000 \\ -0.001 & 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.596 & 0.296 & -0.480 \\ 0.197 & 0.595 & -0.680 \\ 0.196 & 0.395 & 0.618 \end{bmatrix}$	$\begin{bmatrix} -0.004 & -0.004 & 0.020 \\ -0.003 & -0.005 & 0.020 \\ -0.004 & -0.005 & 0.018 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.001 & 0.004 \\ 0.001 & 0.001 & 0.004 \\ 0.001 & 0.001 & 0.004 \end{bmatrix}$	$\begin{bmatrix} 0.588 & 0.289 & -0.442 \\ 0.189 & 0.587 & -0.642 \\ 0.189 & 0.387 & 0.656 \end{bmatrix}$	$\begin{bmatrix} -0.012 & -0.011 & 0.058 \\ -0.011 & -0.013 & 0.058 \\ -0.011 & -0.013 & 0.056 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.001 & 0.010 \\ 0.001 & 0.001 & 0.010 \\ 0.001 & 0.001 & 0.009 \end{bmatrix}$
	100	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & 0.000 \\ -0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.597 & 0.299 & -0.489 \\ 0.198 & 0.597 & -0.689 \\ 0.198 & 0.397 & 0.611 \end{bmatrix}$	$\begin{bmatrix} -0.003 & -0.001 & 0.011 \\ -0.002 & -0.003 & 0.011 \\ -0.002 & -0.003 & 0.011 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.002 \\ 0.000 & 0.000 & 0.002 \\ 0.000 & 0.000 & 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.592 & 0.291 & -0.454 \\ 0.193 & 0.589 & -0.655 \\ 0.192 & 0.389 & 0.643 \end{bmatrix}$	$\begin{bmatrix} -0.008 & -0.009 & 0.046 \\ -0.007 & -0.011 & 0.045 \\ -0.008 & -0.011 & 0.043 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.001 & 0.007 \\ 0.001 & 0.001 & 0.007 \\ 0.001 & 0.001 & 0.007 \end{bmatrix}$
	300	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 & -0.000 \\ 0.000 & -0.000 & -0.000 \\ 0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.599 & 0.299 & -0.494 \\ 0.199 & 0.599 & -0.694 \\ 0.199 & 0.399 & 0.606 \end{bmatrix}$	$\begin{bmatrix} -0.001 & -0.001 & 0.006 \\ -0.001 & -0.001 & 0.006 \\ -0.001 & -0.001 & 0.006 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.001 \\ 0.000 & 0.000 & 0.001 \\ 0.000 & 0.000 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.594 & 0.295 & -0.472 \\ 0.195 & 0.594 & -0.672 \\ 0.195 & 0.394 & 0.626 \end{bmatrix}$	$\begin{bmatrix} -0.006 & -0.005 & 0.028 \\ -0.005 & -0.006 & 0.028 \\ -0.005 & -0.006 & 0.026 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.000 & 0.005 \\ 0.001 & 0.000 & 0.004 \\ 0.001 & 0.000 & 0.004 \end{bmatrix}$
100	50	$\begin{bmatrix} 0.599 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.000 & 0.000 \\ 0.000 & -0.000 & 0.000 \\ -0.000 & 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.599 & 0.300 & -0.496 \\ 0.199 & 0.599 & -0.696 \\ 0.199 & 0.399 & 0.603 \end{bmatrix}$	$\begin{bmatrix} -0.001 & -0.000 & 0.005 \\ -0.001 & -0.001 & 0.004 \\ -0.001 & -0.001 & 0.003 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.001 \\ 0.000 & 0.000 & 0.001 \\ 0.000 & 0.000 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.596 & 0.295 & -0.478 \\ 0.196 & 0.595 & -0.678 \\ 0.196 & 0.395 & 0.621 \end{bmatrix}$	$\begin{bmatrix} -0.004 & -0.005 & 0.022 \\ -0.004 & -0.005 & 0.022 \\ -0.004 & -0.005 & 0.021 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.003 \\ 0.000 & 0.000 & 0.003 \\ 0.000 & 0.000 & 0.003 \end{bmatrix}$
	100	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 & 0.000 \\ 0.000 & -0.000 & -0.000 \\ -0.000 & 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 & -0.499 \\ 0.200 & 0.600 & -0.699 \\ 0.200 & 0.399 & 0.601 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 & 0.001 \\ 0.000 & -0.000 & 0.001 \\ 0.000 & -0.001 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.598 & 0.298 & -0.490 \\ 0.198 & 0.598 & -0.689 \\ 0.198 & 0.398 & 0.610 \end{bmatrix}$	$\begin{bmatrix} -0.002 & -0.002 & 0.010 \\ -0.002 & -0.002 & 0.011 \\ -0.002 & -0.002 & 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.002 \\ 0.000 & 0.000 & 0.002 \\ 0.000 & 0.000 & 0.002 \end{bmatrix}$
	300	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 & -0.000 \\ 0.000 & -0.000 & -0.000 \\ 0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 & -0.000 \\ 0.000 & -0.000 & -0.000 \\ -0.000 & 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.599 & 0.299 & -0.497 \\ 0.199 & 0.599 & -0.697 \\ 0.199 & 0.400 & 0.603 \end{bmatrix}$	$\begin{bmatrix} -0.001 & -0.001 & 0.003 \\ -0.001 & -0.001 & 0.003 \\ -0.001 & -0.001 & 0.003 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.001 \\ 0.000 & 0.000 & 0.001 \\ 0.000 & 0.000 & 0.001 \end{bmatrix}$
300	50	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} 0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & -0.000 \\ -0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 & 0.000 \\ -0.000 & -0.000 & 0.000 \\ -0.000 & -0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$
	100	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} 0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & -0.000 \\ -0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 & 0.000 \\ -0.000 & -0.000 & 0.000 \\ -0.000 & -0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$
	300	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} 0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & -0.000 \\ -0.000 & -0.000 & 0.000 \\ 0.000 & -0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.600 & 0.300 & -0.500 \\ 0.200 & 0.600 & -0.700 \\ 0.200 & 0.400 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & -0.000 & 0.000 \\ -0.000 & -0.000 & 0.000 \\ -0.000 & -0.000 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \end{bmatrix}$

Table 14: BIC Accuracy in Percentage

		Percentage Rate			
T	N	$G = 2$	$G = 3$	$G = 4$	$G = 5$
50	50	100.00	98.65	79.15	28.10
	100	100.00	99.40	85.40	43.00
	300	100.00	99.85	92.60	57.50
100	50	100.00	99.70	94.15	54.95
	100	100.00	99.95	97.40	71.25
	100	100.00	100.00	99.65	86.90
300	50	100.00	100.00	100.00	97.20
	100	100.00	100.00	100.00	99.90
	300	100.00	100.00	100.00	100.00

3.6.2 Misclassification Rate

Table 15 presents the average misclassification rates for six models (DGP 1 to DGP 6) across different panel sizes, characterised by varying numbers of cross-sectional units (N) and time periods (T). The misclassification rate reflects the proportion of incorrectly assigned group memberships when estimating the grouped structure using the PVAR-GFE model. As N increases, the misclassification rate decreases across all models, indicating improved accuracy in group classification. For instance, when $T = 50$, the misclassification rate for DGP 1 decreases from 0.012 when $N = 50$ to 0.001 when $N = 300$, demonstrating that larger panels with more cross-sectional units enhance classification performance. Similarly, an increase in the time dimension T also leads to lower misclassification rates. For example, with $N = 50$, the misclassification rate for DGP 1 declines from 0.012 when $T = 50$ to 0.000 when $T = 300$. Across all models, panels with a larger time dimension (e.g., $T = 300$) achieve very low misclassification rates, particularly when the number of cross-sectional units is also large. When comparing models, DGP 1 consistently shows the lowest misclassification rates across most panel settings, closely followed by DGP 2, 3, 4 and 5 while DGP 6 generally exhibits higher misclassification rates, especially when N is small. For example, with $N = 50$ and $T = 50$, DGP 6 has a misclassification rate of 0.152 , whereas DGP 1 has a much lower rate. Overall, the models perform best when both N and T are large, with near-perfect group classification achieved when $N = 100$ or $N = 300$ and $T = 300$, where the misclassification rate across all models approaches zero. This suggests that larger panels, both in terms of cross-sectional units and time periods, significantly improve the accuracy of the PVAR-GFE model's group classification, with DGP 1 consistently outperforming the others, while DGP 6 tends to exhibit higher misclassification rates, particularly in smaller panels.

Table 15: Average Missclassification Rate

		Missclassification Rate					
T	N	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6
50	50	0.012	0.021	0.104	0.007	0.072	0.152
	100	0.011	0.010	0.054	0.006	0.044	0.123
	300	0.001	0.005	0.020	0.005	0.030	0.084
100	50	0.002	0.002	0.018	0.002	0.020	0.067
	100	0.001	0.001	0.003	0.001	0.008	0.035
	100	0.001	0.000	0.001	0.001	0.005	0.017
300	50	0.000	0.000	0.000	0.000	0.000	0.001
	100	0.000	0.000	0.000	0.000	0.000	0.000
	300	0.000	0.000	0.000	0.000	0.000	0.000

3.6.3 Empirical Application: The Nexus of Financial Development, Economic Growth, and Energy Consumption in China

In this section, we apply the Panel Vector Autoregressive model with Grouped Fixed Effects (PVAR-GFE) to examine the dynamic relationships among financial development, economic growth, and energy consumption in China. Following the methodology of Ouyang & Li (2018), we investigate how these key macroeconomic variables interact across different Chinese provinces, accounting for group-specific heterogeneity that captures unobserved structural patterns. The dataset comprises quarterly observations for 30 provinces in China over the period 1996Q1 to 2015Q4.

The relationship between financial development, economic growth, and energy consumption has become increasingly important in emerging economies, particularly in rapidly industrialising countries such as China. As the world's largest energy consumer and one of the fastest-growing economies, China faces significant challenges in balancing sustainable economic development with rising energy demand and the expansion of its financial sector. Understanding the dynamic interplay among these factors is essential for designing policies that promote long-term growth while addressing the environmental pressures associated with energy consumption. The nexus between financial development and economic growth has been widely examined in the economic literature. Foundational studies, including King & Levine (1993) and Levine (1997), established a positive link between well-functioning financial systems and sustained economic growth. However, more recent research highlights the complexity of this relationship, particularly when energy consumption is taken into account. As financial development advances, it frequently stimulates increased investment in energy-intensive industrial sectors, thereby raising energy consumptionan effect documented in both advanced and developing economies (Sadorsky 2010).

China offers a unique context for examining the dynamics among financial development, economic growth, and energy consumption due to pronounced regional disparities in economic development, financial infrastructure, and energy usage patterns. Several studies have explored how energy consumption affects economic growth in China. For instance, Zhang (2011) found that energy consumption is a key driver of growth, but the magnitude and direction of its impact vary across regions, depending on differences in industrial structure and development stage. Similarly, Shahbaz et al. (2013) examined the influence of financial development on energy consumption and found that, while financial markets may stimulate investment in energy-intensive industries, they can also support the adoption of green technologiesresulting in complex, region-specific dynamics. A particularly relevant study by Ouyang & Li (2018) reported that financial development has a significantly negative effect on economic growth but can substantially reduce energy consumption across all regions. Their findings underscore the heterogeneity in how financial development influences growth and energy demand, shaped by regional characteristics. For example, financially advanced provinces demonstrated stronger linkages between financial development and improvements in energy efficiency, whereas less developed regions exhibited weaker or qualitatively different relationships. Despite these valuable insights, the study by Ouyang & Li (2018) employed a traditional Panel Vector Autoregression (PVAR) model, which assumes homogeneity across regions. This assumption may obscure underlying group-specific dynamics that help explain the observed variation in regional responses.

To address these limitations, this paper replicates the study by Ouyang & Li (2018) using a more advanced econometric framework: the Panel Vector Autoregression with Grouped Fixed Effects (PVAR-GFE) model. This approach offers several advantages over traditional PVAR models by allowing for group-specific heterogeneity and capturing cross-sectional dependence. By incorporating grouped fixed effects, the model can identify latent group structures that reveal how regions in Chinaclassified by levels of financial development, industrial structure, and energy consumption patternsrespond similarly to economic and financial shocks. This methodology is particularly valuable in the context of China's large and diverse economy, where provinces differ significantly in development stages and energy dependencies. The empirical application not only contributes to the literature by enhancing the methodological framework used to study the finance-energy-growth nexus, but also provides deeper insights into regional heterogeneity within China's economic system.

This study uses quarterly data from 30 Chinese provinces over the period 1996Q1 to 2015Q4, consistent with the dataset used by Ouyang & Li (2018) [21]. The key variables of interest include:

- Financial Development: Measured using a composite index derived from six indicators, including M2 (money supply), credit-to-GDP ratio, insurance revenue-to-GDP ratio, stock market value-to-GDP ratio, stock turnover ratio, and foreign direct investment (FDI)[22].
- Economic Growth: Measured by real GDP per capita (in logarithmic form) for each province.
- Energy Consumption: Proxied by total energy consumption per province, reflecting the energy needs of both industrial and residential sectors.

To estimate the dynamic relationships between these variables, they are ordered as follows:

$$y_{i,t} = \left|\Delta ln_{m2_{i,t}}, \Delta ln_{EC_{i,t}}, \Delta ln_{GDP_{i,t}}\right|'$$

We assume that financial development is more exogenous than energy consumption and economic growth, based on two considerations. First, financial development is often viewed as a technological or institutional factor. Given that our model focuses solely on China, shocks to financial development are more likely to originate from external sources, such as global financial markets. Second, we assume that current shocks to energy consumption have an immediate effect on economic growth, whereas shocks to economic growth affect energy consumption with a lag.

[21]30 Provinces are generally classified into 3 regions: The eastern region includes the most developed provinces with developed financial system, namely Beijing, Fujian, Guangdong, Hainan, Hebei, Jiangsu, Liaoning, Shanghai, Tianjin, Zhejiang and Shandong. The central region contains Anhui, Heilongjiang, Henan, Hubei, Hunan, Jiangxi, Jilin and Shanxi. Finally, the western region consists of Chongqing, Gansu, Guangxi, Guizhou, Inner Mongolia, Ningxia, Qinghai, Shaanxi, Sichuan, Xinjiang and Yunnan.

[22]In this empirical application, we only use M2 as the indicator of financial development for simplicity. The results are consistent with the rest of indicators.

In contrast to the findings of Ouyang & Li (2018), which suggest a more uniform response across Chinese provinces, the PVAR-GFE model reveals heterogeneous impacts that vary considerably across groups of provinces. As shown in Figure 9, both models indicate that financial development has a positive effect on economic growth (GDP), consistent with the established literature on the role of financial markets in driving economic expansion. However, the PVAR-GFE model uncovers important differences in the magnitude and duration of this effect. In the model of Ouyang & Li (2018), the response of GDP to a shock in financial development is moderate but persistent. The initial impact is positive, and the effect dissipates gradually over time, suggesting that the benefits of financial development are spread over an extended period. In contrast, the PVAR-GFE model shows a more immediate and pronounced response. The initial impact is larger, indicating that GDP responds more quickly to changes in financial development when latent group structures are taken into account. Regarding the effect of energy consumption (EC) shocks, Ouyang & Li (2018) report a positive and relatively sharp response of GDP, peaking early at around 0.006. This suggests that energy consumption shocks have a strong and immediate effect on economic output, particularly in provinces with energy-intensive industries such as manufacturing or resource extraction. However, in the PVAR-GFE model, the initial response of GDP to an EC shock is much smaller and fluctuates within a narrow range of approximately -0.2% to 0.2%. This muted and heterogeneous response suggests that provinces differ in their reliance on energy consumption for short-term growth, a pattern that is revealed by allowing for latent grouped heterogeneity.

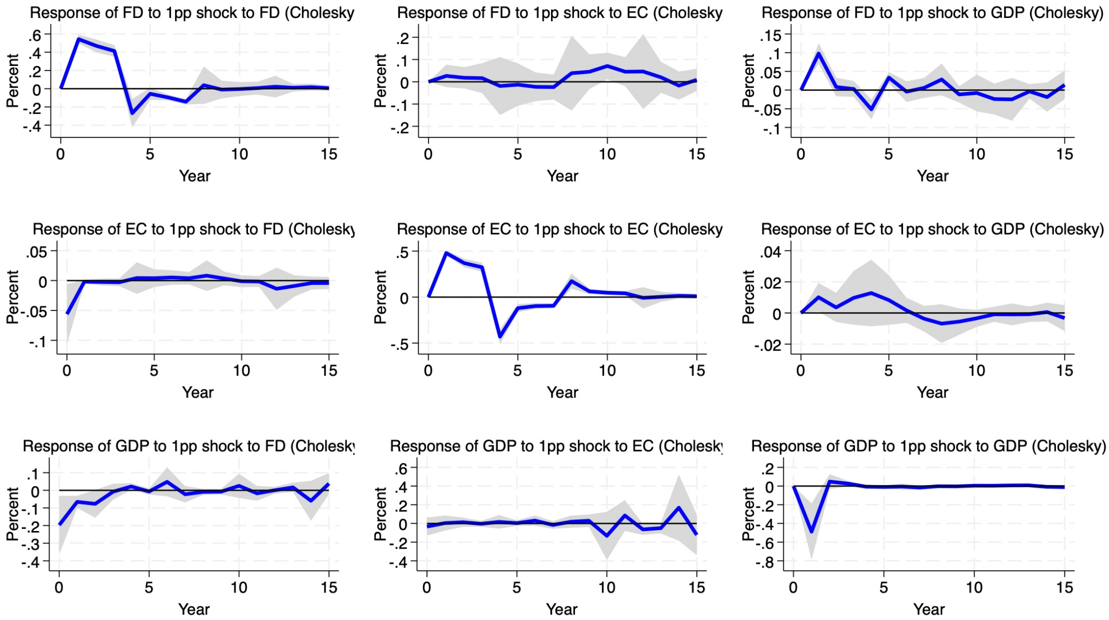

Figure 9: Local Projection Results of Application in financial development, economic growth and energy consumption.

3.7 Group Membership

One of the key findings of the regional classification derived from the PVAR-GFE model is the emergence of four distinct groups of provinces, in contrast to the traditional three-region division of Eastern, Central, and Western China. Unlike conventional classifications based on geography or administrative boundaries, the PVAR-GFE approach groups provinces according to shared economic and financial characteristics, uncovering more nuanced dynamics that may be overlooked in standard regional frameworks. Group 1 consists of the most economically and financially advanced provinces, including Beijing, Shanghai, Tianjin, and Guangdong. These provinces are characterised by highly developed financial systems and strong integration with global markets. Their impulse response functions (IRFs) display faster recoveries from financial development shocks, consistent with their roles as national financial hubs. This group broadly overlaps with the traditional Eastern region. Group 2 includes provinces primarily from the Central and Northeastern regions, such as Henan, Hubei, Hunan, Jilin, and

Heilongjiang. These provinces exhibit a mix of industrial and agricultural economic structures and maintain moderate levels of financial development. Group 3 encompasses many provinces in Western China, including Gansu, Guangxi, Qinghai, and Xinjiang. These regions are generally less developed both economically and financially, which is reflected in their more muted responses to financial and energy-related shocks. Group 4 is particularly noteworthy, comprising provinces traditionally classified as part of both the Eastern (e.g., Jiangsu) and Western (e.g., Sichuan, Hainan) regions. The inclusion of Jiangsu alongside Sichuan and Hainan suggests that, despite clear geographical and administrative differences, these provinces exhibit similar economic and financial dynamics. The impulse response functions (IRFs) indicate that their responses to shocks differ from those of their geographic neighbours, implying that this grouping is driven by latent structural similarities rather than spatial proximity. This finding challenges conventional regional classifications in China and highlights the importance of data-driven approaches for identifying economically coherent groups. It also suggests that these provinces may benefit from more targeted policy interventions that address their specific structural characteristics rather than relying on broad, region-based strategies.

3.8 Conclusion

In this paper, we introduce a Panel VAR model with group-specific heterogeneous coefficients to address cross-sectional grouped heterogeneity. The model is flexible in that the number of groups, group memberships, and the number of lagged dependent variables are all treated as unknown and estimated from the data. Instead of using conventional impulse response functions (IRFs), we employ local projections to estimate dynamic responses. The group-based local projection method outperforms standard fixed effects local projections by avoiding the implicit Nickell bias, thereby improving the reliability of IRF estimation. We also derive asymptotic inference for the PVAR-GFE model and its extensions under large N and T asymptotics. Monte Carlo simulations demonstrate that the proposed procedure performs well: bias decreases with larger N and T, and the group assignment error becomes negligible even in high-dimensional PVAR settings.

In the empirical application, we use the PVAR-GFE model to examine the dynamic relationships among financial development, economic growth, and energy consumption across Chinese provinces. By accounting for group-specific heterogeneity, the model identifies distinct regional patterns that deviate from conventional geographic classifications. This application highlights the value of incorporating grouped fixed effects to capture complex regional dynamics, thereby offering more accurate and policy-relevant insights into macroeconomic relationships.

3.9 Appendix A: PVAR-GFE with Homogeneous Coefficient

We illustrate the PVAR-GFE(P) model as follows: In this case, Equation 29 can be presented as:

$$y_{m,i,t} = \sum_{p=1}^{P} \psi_{m,p}^{m} y_{m,i,t-p} + \sum_{p=1}^{P} \sum_{n \neq m}^{M} \phi_{n,p}^{m} y_{n,i,t-p} + \boldsymbol{x}_{i,t}' \boldsymbol{\gamma}_m + \alpha_{m,g_i,t} + \epsilon_{m,i,t} \tag{28}$$

$$\boldsymbol{y}_{i,t} = \boldsymbol{z}_{i,t} \boldsymbol{\beta} + \boldsymbol{\alpha}_{g_i,t} + \boldsymbol{\epsilon}_{i,t} \tag{29}$$

Where

$$\underset{M\times 1}{\boldsymbol{y}_{i,t}} = \begin{bmatrix} y_{1,i,t} \\ \vdots \\ y_{M,i,t} \end{bmatrix}, \quad \underset{M\times \mathcal{F}}{\boldsymbol{z}_{i,t}} = \begin{bmatrix} \boldsymbol{z}_{1,i,t}' \\ \vdots \\ \boldsymbol{z}_{M,i,t}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_M \otimes \mathcal{Y}_{i,t-1}', & \boldsymbol{I}_M \otimes \mathcal{X}_{i,t-1}' \end{bmatrix}, \underset{\mathcal{F}\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\gamma} \end{bmatrix}$$

$$\underset{MP\times 1}{\mathcal{Y}_{i,t}} = \begin{bmatrix} \boldsymbol{y}_{i,t} \\ \vdots \\ \boldsymbol{y}_{i,t-p+1} \end{bmatrix}', \quad \underset{MK\times 1}{\mathcal{X}_{i,t}} = \begin{bmatrix} \boldsymbol{x}_{i,t} \\ \vdots \\ \boldsymbol{x}_{i,t-p+1} \end{bmatrix}', \underset{M\times 1}{\boldsymbol{\alpha}_{g_i,t}} = \begin{bmatrix} \alpha_{1,g_i,t} \\ \vdots \\ \alpha_{M,g_i,t} \end{bmatrix}, \quad \underset{M\times 1}{\boldsymbol{\epsilon}_{i,t}} = \begin{bmatrix} \boldsymbol{\epsilon}_{i,t,1} \\ \vdots \\ \boldsymbol{\epsilon}_{i,t,M} \end{bmatrix},$$

Let $i \in \{1, 2, \ldots, N\}$ index the cross-sectional units, $t \in \{1, 2, \ldots, T\}$ the time periods, $p \in \{1, 2, \ldots, P\}$ the lag orders, and $m \in \{1, 2, \ldots, M\}$ the number of endogenous (dependent) variables in the PVAR system. The variables $y_{m,i,t}$ denote the dependent variables for each equation of the unit-specific VAR. Each unit includes a $(K \times 1)$ vector of exogenous variables $\boldsymbol{x}_{i,t}$, which appears in every equation for all dependent variables, with equation-specific coefficients $\boldsymbol{\gamma}_m \in \mathbb{R}^K$, assumed to be homogeneous across units. In compact form, $\boldsymbol{\beta}$ denotes the vector of coefficients to be estimated. It includes all autoregressive and cross-lagged coefficients for lag order p, while $\epsilon_{m,i,t}$ denotes the idiosyncratic errors. Suppose there are G groups, and let $g \in \{1, 2, \ldots, G\}$ denote a group index. Define g_i as the group membership indicator for unit i, which identifies the group to which unit i belongs. We allow for group-specific intercepts $\alpha_{m,g_i,t}$, which may vary over time to capture dynamic heterogeneity at the group level across equations. Let $\mathbb{G}_g$ denote the number of cross-sectional units in group g. Group memberships are mutually exclusive, such that for any units $i \neq j$, $\alpha_{m,g_i,t} \neq \alpha_{m,g_j,t}$. One key assumption in this model is that the number of groups and group memberships remains constant, as discussed in Bonhomme & Manresa (2015), Su et al. (2016). For simplicity, let $\mathcal{F}$ denote the total number of parameters, defined as $\mathcal{F} = M^2P + MK$. Define $\boldsymbol{z}_{i,t} \in \mathbb{R}^{M \times \mathcal{F}}$ as the stacked vector of lagged endogenous regressors and common covariates. The vector $\boldsymbol{\beta} \in \mathbb{R}^{\mathcal{F}}$ collects all slope parameters, where $\boldsymbol{\theta}$ is an $(M^2P) \times 1$ vector and $\boldsymbol{\gamma}$ is an $(MK) \times 1$ vector.

Below, we show a simple bivariate PVAR-GFE(1) model with one additional common covariate and one group-specific factor for each equation, that is:

$$\begin{aligned} y_{1,i,t} &= \theta^1_{1,1} y_{1,i,t-1} + \theta^1_{2,1} y_{i,t-1,2} + \gamma_{1,1} x_{i,t-1} + \alpha_{1,g_i,t} + \epsilon_{i,t,1} \\ y_{2,i,t} &= \theta^2_{2,1} y_{1,i,t-1} + \theta^2_{1,1} y_{i,t-1,2} + \gamma_{2,1} x_{i,t-1} + \alpha_{2,g_i,t} + \epsilon_{i,t,2} \end{aligned} \tag{30}$$

In matrix form:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \begin{bmatrix} y_{1,i,t-1} & y_{2,i,t-1} & 0 & 0 \\ 0 & 0 & y_{1,i,t-1} & y_{2,i,t-1} \end{bmatrix} \begin{bmatrix} \theta^1_{1,1} \\ \theta^1_{2,1} \\ \theta^2_{2,1} \\ \theta^2_{1,1} \end{bmatrix}$$
$$+ \begin{bmatrix} x_{i,t-1} & 0 \\ 0 & x_{i,t-1} \end{bmatrix} \begin{bmatrix} \gamma_{1,1} \\ \gamma_{2,1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,i,t} \\ \epsilon_{2,i,t} \end{bmatrix}$$

The PVAR-GFE(p) model can accommodate different extensions. For example, if we include a vector of 1 in $\boldsymbol{z}_{i,t}$, we obtain:

$$\boldsymbol{z}_{i,t} = \begin{bmatrix} \boldsymbol{z}'_{1,i,t} \\ \vdots \\ \boldsymbol{z}'_{M,i,t} \end{bmatrix} = \left[\begin{array}{ccc} \boldsymbol{I}_M \otimes \mathcal{Y}'_{i,t-1}, & \boldsymbol{I}_M \otimes \mathcal{X}'_{i,t-1}, \boldsymbol{I}_M \otimes d'_i \end{array} \right], \quad \underset{\mathcal{F}\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\gamma} \\ \boldsymbol{\eta} \end{bmatrix}$$

the model in 29 becomes the PVAR-GFE(p) with individual fixed effects.

3.9.1 Estimation

3.9.1.1 Identification

The model in Equation 29 includes two sets of parameters. The first consists of the slope coefficients $\boldsymbol{\beta} \in \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is defined as the sample space comprising subsets of $\mathbb{R}^{\mathcal{F}}$. The second consists of the group-specific fixed effects $\boldsymbol{\alpha}_{g_i} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ denotes the sample space of fixed effects, defined as subsets of $\mathbb{R}^{MT}$. Let τ denote a specific partitioning scheme, selected from the set of possible group membership assignments g_i, which allocates the N cross-sectional units into G distinct groups. Formally, $\tau \in \boldsymbol{\Gamma}_G$, where $\boldsymbol{\Gamma}_G$ represents the set of all possible partitions of the units into G groups. We denote $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}_{g_i}$, and $\hat{\tau}$ as the estimators of the slope coefficients, grouped fixed effects, and group membership assignment, respectively. Equation 29 is estimated by least squares (LS) through an iterative procedure:

$$\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}_{g_i}, \hat{\tau}\right) = \underset{(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\in\boldsymbol{\Omega}\times\boldsymbol{\Theta}^{G\times TM}\times\boldsymbol{\Gamma}_G}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i;g_i=g} \sum_{t=1}^{T} \left\|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i,t}\right\|_2^2 \tag{31}$$

Following Bai (2009), our analysis begins by considering the casein which the group structure is known. Specifically, assume that the true number of groups G is known and that the membership g_i to each group is correctly identified such that $g_i = g_i^0$. The model becomes:

$$\boldsymbol{y}_{i,t} = \boldsymbol{z}_{i,t}\boldsymbol{\beta} + \boldsymbol{\alpha}_{g_i^0,t} + \boldsymbol{\epsilon}_{i,t} \tag{32}$$

All that is left is to estimate 29 and obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}_{g_i^0,t}$. This is achieved directly by ordinary least squares with interaction time dummies t for given group assignment g_i^0:

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}_{g_i^0,t}) = \underset{\boldsymbol{\beta}\in\boldsymbol{\Omega}}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i;g_i=g} \sum_{t=1}^{T} \left\|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i^0,t}\right\|_2^2$$

However, in practice, neither the number of groups nor the group memberships is known. In this case, the estimation of Equation 29 must be conducted using an iterative algorithm:

Algorithm 2 PVAR-GIFE Algorithm

1: Initialisation.

Choose the maximum number of groups G, assign each unit into the G groups using K-means method by residual sum of squares from OLS without grouped fixed effects. Then we have initial values of $\left(\boldsymbol{\beta}^{(0)}, \boldsymbol{\alpha}_{g_i}^{(0)}, \tau^{(0)}\right)$. Set $s = \{1, 2, \ldots, S\}$ for the number of iterations within estimation process.

2: Update the group members

Update $\tau^{(s)}$ in each group based on current values of $\boldsymbol{\beta}^{(s)}$ and $\boldsymbol{\alpha}_{g_i}^{(s)}$. Calculate the RSS for each unit across time in different groups and assign it into the group with minimum RSS.

$$g_i^{(s+1)} = \underset{g \in \{1,\ldots,G\}}{\operatorname{argmin}} \sum_{t=1}^{T} \left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta}^{(s)} - \boldsymbol{\alpha}_{g_i,t}^{(s)} \right\|_2^2$$

3: Update the parameter of interest

Update $\boldsymbol{\beta}^{(s)}$ based on current values of $\boldsymbol{\alpha}_{g_i}^{(s+1)}$ and $\tau^{(s+1)}$ using OLS:

$$\boldsymbol{\beta}^{(s+1)} = \underset{\boldsymbol{\beta} \in \boldsymbol{\Omega}}{\operatorname{argmin}} \sum_{g=1}^{G} \sum_{i; g_i = g} \sum_{t=1}^{T} \left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i^{s+1},t}^{(s+1)} \right\|_2^2$$

4: Check numerical convergence.

Set $s = s + 1$ and go to Step 2 till the norm of $\boldsymbol{\beta}$ is lower than the convergence error set by researchers.

3.9.2 Asymptotic Theory

In this section we will present a set of assumptions that make consistency holds.

Assumption 5 *Stationarity. Define the companion matrix* F *as*

$$
\mathrm{F} = \begin{bmatrix} \boldsymbol{\theta}_1 & \boldsymbol{\theta}_2 & \boldsymbol{\theta}_3 & \cdots & \boldsymbol{\theta}_{p-1} & \boldsymbol{\theta}_p \\ \mathbf{I}_M & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_M & \mathbf{0} \end{bmatrix}
$$

The eigenvalues of the matrix F *satisfy the characteristic equation:*

$$
\left| \mathbf{I}_M \lambda^p - \boldsymbol{\theta}_1 \lambda^{p-1} - \boldsymbol{\theta}_2 \lambda^{p-2} - \cdots - \boldsymbol{\theta}_p \right| = 0.
$$

Hence, the PVAR-GFE is covariance-stationary as long as $|\lambda| < 1$ *for all values of* λ*. Equivalently, it is stationary if all values of* z *satisfying*

$$
\left| \mathbf{I}_M - \boldsymbol{\theta}_1 z - \boldsymbol{\theta}_2 z^2 - \cdots - \boldsymbol{\theta}_p z^p \right| = 0
$$

lie outside the unit circle.

In the PVAR framework, stationarity is a crucial assumption for the theoretical results presented later in the paper. While the grouped fixed effects estimator can also be applied in a non-stationary setting, this extension lies beyond the scope of the present study.

Assumption 6 *There exists a constant* $\mathbb{M} > 0$ *such that:*

a *Compactness. Parameter space* $\boldsymbol{\Omega}$ *and* $\boldsymbol{\Theta}$ *are compact subsets of* $\mathbb{R}^{\mathcal{K}}$ *and* $\mathbb{R}$ *with* $\sup_{\boldsymbol{\beta} \in \boldsymbol{\Omega}} \|\boldsymbol{\theta}\| \leq M$

b *Strict exogeneity.* $\mathbb{E}\left[\boldsymbol{\epsilon}_{i,t} \mid \mathbf{x}_{i,s}\right] = 0$ *for all* s, t*, and all* i*.*

c $\mathbb{E}(\boldsymbol{\epsilon}) = 0$ *and* $E\left\|\boldsymbol{\epsilon}\right\|^8 < \infty$*.*

d $\mathbb{E}(\boldsymbol{\epsilon}'_{i,t}\boldsymbol{\epsilon}_{j,t}) = 0$ *if* $i \neq j$*.*

e $\mathbb{E}(\boldsymbol{\epsilon}'_{i,t}\boldsymbol{\epsilon}_{i,s}) = 0$ *if* $t \neq s$*.*

f *The number of equations in the model, namely* M *is small and fixed.*

g $\left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E} \left\| \mathbf{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} \right\|^2 \right| \leq \mathbb{M}$

h Let $\bar{\mathbf{z}}_{g\wedge\widetilde{g},t}$ *denote* $\bar{\mathbf{z}}_{g\wedge\widetilde{g},t} = \dfrac{\sum_{i=1}^{N}\mathbf{1}\{g_i^0 = g\}\mathbf{1}\{g_i = \widetilde{g}\}\mathbf{z}_{it}}{\sum_{i=1}^{N}\mathbf{1}\{g_i^0 = g\}\mathbf{1}\{g_i = \widetilde{g}\}}$. *For all groupings* $\tau = \{g_1, \ldots, g_N\} \in \Gamma_G$, *we define* $\widehat{\rho}(\tau)$ *as the minimum eigenvalue of the following matrix:*

$$\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0\wedge g_i,t}\right)\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0\wedge g_i,t}\right)'.$$

Then $\operatorname{plim}_{N,T\to\infty}\min_{\tau\in\Gamma_G}\widehat{\rho}(\tau) = \rho > 0$.

Assumption (a) ensures that no closed-form optimisation problem yields an optimal solution outside a closed and bounded parameter space. Assumption (b) states that the exogenous variables are independent of the idiosyncratic error terms. Assumption (c) requires that the idiosyncratic terms have zero mean and are bounded by a finite eighth moment. Assumption (d) imposes the absence of cross-sectional dependence among the idiosyncratic terms, such that any dependence is fully captured by the grouped fixed effects $\boldsymbol{\alpha}_{g_i,t}$. Assumption (e) assumes no serial correlation within the idiosyncratic terms, which is equivalent to stating that the number of lagged dependent variables, M, is sufficient to account for serial dependence in the model. Assumption (f) asserts that the number of equations does not affect the asymptotic properties of the estimator. Assumption (g) ensures that the second moment of the cross-product $\mathbf{z}i,t'\boldsymbol{\epsilon}i,t$ is uniformly bounded in N and T. Assumption (h) imposes a full-rank condition, requiring sufficient variation across individuals within the panel.

We first define the objective function as:

$$\begin{aligned}\hat{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau) &= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i,t}\|^2 \\ &= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t} + \boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{g_i^0,t} - \boldsymbol{\alpha}_{g_i,t}\right\|^2\end{aligned} \tag{33}$$

As stated in Newey & McFadden (1994), the consistency of extremum estimators requires $\underset{N,T\to\infty}{\operatorname{plim}}\ \underset{(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\in\boldsymbol{\Omega}\times\boldsymbol{\Theta}^{GTM}\times\boldsymbol{\Gamma}_G}{\sup}\left|\hat{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau) - \mathrm{Q}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\right| = o_p(1)$ when N and T go to infinity. However, the grouped fixed effects estimator $\boldsymbol{\alpha}$ grows as T increases. The limit function could be unavailable as there would be infinite number of parameters within $\boldsymbol{\alpha}$ when T go to infinity. Therefore, we construct an auxiliary objective function $\tilde{\mathrm{Q}}_{NT}(\beta,\alpha,\gamma)$ which is also dependent on N and T and it could reach the unique minimum to the true value of parameters. We first show that $\hat{\mathrm{Q}}(\beta,\alpha,\gamma)$ converges uniformly to $\tilde{\mathrm{Q}}(\beta,\alpha,\gamma)$ and then $\tilde{\mathrm{Q}}(\beta,\alpha,\gamma)$ is minimised uniquely at the true value of parameters.

By Lemma 1 and Lemma 2 we have the following Theorem:

Theorem 5

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta} \quad \textit{as } (N,T) \to \infty$$

$$\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{g_i^0,t} - \hat{\boldsymbol{\alpha}}_{\hat{g}_i,t}\right\|^2 = o_p(1)$$

3.9.3 Asymptotic Equivalence

In this section, we show that the PVAR-GFE estimator and the infeasible PVAR-GFE estimator are asymptotically equivalent, which forms the basis for deriving their asymptotic distribution. The assumptions used in the proof of asymptotic equivalence are stated as follows:

Assumption 7 *a All units are divided into a finite number of groups denoted by G, each of them containing N_g number of group memberships that increase as N increases. When N goes to infinity, we have:*

$$\operatorname{plim}_{N\to\infty} \quad 0 < \pi_g < 1 \quad \textit{where } \pi_g = \frac{N_g}{N}$$

b For all $(g,\widetilde{g}) \in \{1,\ldots,G\}^2$ and $g \neq \widetilde{g}$: $\operatorname{plim}_{T\to\infty} \frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt} - \boldsymbol{\alpha}^0_{\tilde{g}t}\right\|^2 = c_{g,\tilde{g}}$, $\infty > c_{g,\tilde{g}} > 0$.

c There exist constants $a > 0$ and $d_1 > 0$ and a sequence $\alpha[t] \leq e^{-at^{d_1}}$ such that, for all $i \in \{1,\ldots,N\}$ and $(g,\widetilde{g}) \in \{1,\ldots,G\}^2$ such that $g \neq \widetilde{g}, \{\boldsymbol{\epsilon}_{it}\}_t$, $\left\{\boldsymbol{\alpha}^0_{gt} - \boldsymbol{\alpha}\tilde{g}t^0\right\}_t$, and $\left\{\left(\boldsymbol{\alpha}gt^0 - \boldsymbol{\alpha}\tilde{g}t^0\right)\boldsymbol{\epsilon}_{it}\right\}_t$ are strongly mixing processes with mixing coefficients $\boldsymbol{\alpha}[t]$. Moreover, $\mathbb{E}\left(\left(\boldsymbol{\alpha}^0_{gt} - \boldsymbol{\alpha}^0_{\tilde{g}t}\right)'\boldsymbol{\epsilon}_{it}\right) = 0$

d There exist constants $b > 0$ and $d_2 > 0$ such that $\Pr\left(|\boldsymbol{\epsilon}_{it}| > m\right) \leq e^{1-(m/b)^{d_2}}$ for all i, t, and $m > 0$.

e There exists a constant $M^* > 0$ such that, as N, T tend to infinity:

$$\sup_{i\in\{1,\ldots,N\}} \Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\left\|\mathbf{z}_{it}\right\|^2 \geq M^*\right) = o\left(T^{-\delta}\right) \qquad \textit{for all } \delta > 0.$$

Next we proof that the group is estimated sup-consistently.

Theorem 6 *for all $\delta > 0$ and as $NT^{-\delta}$ tends to zero:*

$$\Pr\left(\sup_{i\in\{1,\dots,N\}}\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right)=o(1) \tag{34}$$

By Lemma 5 and Lemma 6 we have the following Theorem:

Theorem 7 *For all $\delta > 0$,*

$$\hat{\boldsymbol{\beta}}=\tilde{\boldsymbol{\beta}}+o_p(T^{-\delta}) \quad \textit{as } (N,T)\to\infty \tag{35}$$

By Lemma 5, Lemma 6 and Theorem 7, we have:

Theorem 8 *For all $\delta > 0$,*

$$\hat{\boldsymbol{\alpha}}=\tilde{\boldsymbol{\alpha}}+o_p(T^{-\delta}) \quad \textit{as } (N,T)\to\infty \tag{36}$$

3.9.4 Asymptotic Distribution

In this section, we derive the asymptotic distribution of the estimators, based on their asymptotic equivalence to the corresponding infeasible estimators.

Assumption 8 *a For all i, j, and t : $\mathbb{E}\left(\mathbf{z}_{jt}'\boldsymbol{\epsilon}_{it}\right)=0$.*

b There exist positive definite matrices Σ_β and Ω_β such that

$$\Sigma_\beta=\operatorname{plim}_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0}\right)',$$

$$\Omega_\beta=\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\left[\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{js}'\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)\left(\mathbf{z}_{js}-\bar{\mathbf{z}}_{g_j^0s}\right)'\right].$$

c As N and T tend to infinity: $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)'\boldsymbol{\epsilon}_{it}\xrightarrow{d}\mathbb{N}\left(0,\Omega_\theta\right)$.

d For all (g,t): $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left(\mathbf{1}\left\{g_i^0=g\right\}\mathbf{1}\left\{g_j^0=g\right\}\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{jt}\right)=\omega_{gt}>0$

e For all (g,t), and as N and T tend to infinity: $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0=g\right\}\boldsymbol{\epsilon}_{it}\xrightarrow{d}\mathbb{N}\left(0,\omega_{gt}\right)$.

Where $\bar{\mathbf{z}}_{g_i^0 t} = \frac{\sum_{i=1}^{N} \mathbf{1}\{g_i^0 = g\} \mathbf{z}_{it}}{\sum_{i=1}^{N} \mathbf{1}\{g_i^0 = g\}}$ is the mean of $\mathbf{z}_{it}$ across the true group assignment g_i^0. Assumption 8 (a) is implied by Assumption 6 (b) as all variables in $\mathbf{z}$ are exogenous and predetermined. Assumption 8 (a) - (c) implies that the infeasible estimator $\tilde{\boldsymbol{\beta}}$ is normal distributed. Assumption 8 (d) and (e) ensures the idiosyncratic term follows normal distribution asymptotically. Thus we have the following Corollary:

Corollary 2 *Let Assumption 6, 7 and 8 holds, when N and T tends to infinity and for all $\delta > 0$, $NT^{-\delta} \to 0$, we have:*

$$\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\mathbf{0}}) \xrightarrow{d} \mathbb{N}\left(0, \Sigma_\beta^{-1} \Omega_\beta \Sigma_\beta^{-1}\right) \tag{37}$$

And we also have the Corollary for the grouped fixed effects estimator $\hat{\alpha}_{g,t}$:

Corollary 3 *Let Assumption 6, 7 and 8 holds, when N and T tends to infinity and for all $\delta > 0$, $NT^{-\delta} \to 0$, for all g and t we have:*

$$\sqrt{N}\left(\widehat{\boldsymbol{\alpha}}_{gt} - \boldsymbol{\alpha}^0_{g^0 t}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega_{gt}}{\pi_g^2}\right) \tag{38}$$

3.9.5 Monte Carlo Simulation

3.9.5.1 Model 1: Bivariate PVAR-GFE(1) Model without Constant

We propose the following DGP model:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0.6 & 0.3 \\ 0.2 & 0.6 \end{bmatrix}}_{\Theta_1^0} \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix},$$

Where the lag of dependent variable equals 1, the number of groups equals 2, and the number of dependent variables equals 2. Also, the coefficient of Θ_1^0 confirms this model is stationary. The errors are i.i.d and it follows a standard normal distribution. We replicate 2,000 times for the simulation.

Regarding the DGP of grouped fixed effects $\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}}$ and $\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}}$:

$$\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}} \overset{i.i.d.}{\sim} N(1,1)$$

$$\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}} \overset{i.i.d.}{\sim} N(2,1)$$

Table 16: Monte Carlo Simulation

		PVAR			OLS		
N	T	Mean	Bias	RMSE	Mean	Bias	RMSE
50	50	$\begin{bmatrix} 0.612 & 0.320 \\ 0.221 & 0.609 \end{bmatrix}$	$\begin{bmatrix} 0.012 & 0.020 \\ 0.021 & 0.009 \end{bmatrix}$	$\begin{bmatrix} 0.040 & 0.042 \\ 0.042 & 0.040 \end{bmatrix}$	$\begin{bmatrix} 0.674 & 0.384 \\ 0.283 & 0.672 \end{bmatrix}$	$\begin{bmatrix} 0.074 & 0.084 \\ 0.083 & 0.073 \end{bmatrix}$	$\begin{bmatrix} 3.320 & 3.770 \\ 3.751 & 3.261 \end{bmatrix}$
	100	$\begin{bmatrix} 0.616 & 0.322 \\ 0.224 & 0.613 \end{bmatrix}$	$\begin{bmatrix} 0.016 & 0.022 \\ 0.024 & 0.013 \end{bmatrix}$	$\begin{bmatrix} 0.038 & 0.041 \\ 0.042 & 0.038 \end{bmatrix}$	$\begin{bmatrix} 0.675 & 0.383 \\ 0.283 & 0.674 \end{bmatrix}$	$\begin{bmatrix} 0.075 & 0.083 \\ 0.083 & 0.074 \end{bmatrix}$	$\begin{bmatrix} 3.360 & 3.707 \\ 3.727 & 3.297 \end{bmatrix}$
	300	$\begin{bmatrix} 0.621 & 0.324 \\ 0.227 & 0.617 \end{bmatrix}$	$\begin{bmatrix} 0.021 & 0.024 \\ 0.027 & 0.017 \end{bmatrix}$	$\begin{bmatrix} 0.036 & 0.039 \\ 0.039 & 0.035 \end{bmatrix}$	$\begin{bmatrix} 0.036 & 0.039 \\ 0.039 & 0.035 \end{bmatrix}$	$\begin{bmatrix} 0.076 & 0.082 \\ 0.082 & 0.075 \end{bmatrix}$	$\begin{bmatrix} 3.392 & 3.676 \\ 3.655 & 3.371 \end{bmatrix}$
100	100	$\begin{bmatrix} 0.600 & 0.303 \\ 0.203 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.003 \\ 0.003 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.020 & 0.022 \\ 0.020 & 0.021 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.374 \\ 0.281 & 0.671 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.074 \\ 0.081 & 0.071 \end{bmatrix}$	$\begin{bmatrix} 3.459 & 3.331 \\ 3.605 & 3.169 \end{bmatrix}$
	100	$\begin{bmatrix} 0.600 & 0.303 \\ 0.203 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.003 \\ 0.003 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.018 & 0.020 \\ 0.019 & 0.020 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.375 \\ 0.280 & 0.671 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.075 \\ 0.080 & 0.071 \end{bmatrix}$	$\begin{bmatrix} 3.455 & 3.343 \\ 3.597 & 3.167 \end{bmatrix}$
	300	$\begin{bmatrix} 0.600 & 0.304 \\ 0.204 & 0.600 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.004 \\ 0.004 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.017 & 0.019 \\ 0.018 & 0.019 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.375 \\ 0.280 & 0.671 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.075 \\ 0.080 & 0.071 \end{bmatrix}$	$\begin{bmatrix} 3.438 & 3.353 \\ 3.589 & 3.183 \end{bmatrix}$
300	50	$\begin{bmatrix} 0.599 & 0.301 \\ 0.201 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.011 \\ 0.010 & 0.012 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.279 & 0.669 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.079 & 0.069 \end{bmatrix}$	$\begin{bmatrix} 3.490 & 3.147 \\ 3.545 & 3.082 \end{bmatrix}$
	100	$\begin{bmatrix} 0.599 & 0.301 \\ 0.201 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.011 \\ 0.009 & 0.011 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.280 & 0.668 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.080 & 0.068 \end{bmatrix}$	$\begin{bmatrix} 3.494 & 3.140 \\ 3.570 & 3.053 \end{bmatrix}$
	300	$\begin{bmatrix} 0.599 & 0.301 \\ 0.201 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.010 \\ 0.009 & 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.280 & 0.668 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.080 & 0.068 \end{bmatrix}$	$\begin{bmatrix} 3.500 & 3.141 \\ 3.565 & 3.058 \end{bmatrix}$

In Model 1, which excludes a constant term, PVAR-GFE consistently delivers parameter estimates that are close to the true values with minimal bias (e.g., the Bias for N = 50 and T = 50 is only 0.012 for the coefficient in Θ_1^0), while the OLS estimator shows substantial bias (e.g., 0.074 for the same parameter) and a significantly higher RMSE (e.g., 3.320 for OLS compared to 0.040 for PVAR-GFE). This pattern is replicated across various sample sizes, confirming that PVAR-GFE is particularly well-suited for panel data with time lags and fixed effects.

3.9.5.2 Model 2: Bivariate PVAR-GFE(1) Model with constant

We propose the following DGP model:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \underbrace{\begin{bmatrix} 0.6 & 0.3 \\ 0.2 & 0.6 \end{bmatrix}}_{\Theta_1^0} \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix},$$

Where the lag of dependent variable equals 1, the number of group equals 2, and the number of dependent variables equals 2. Also, the coefficient of Θ_1^0 confirms this model is stationary. The errors are i.i.d and it follows a standard normal distribution. We replicate 2,000 times for the simulation.

Regarding the DGP of grouped fixed effects $\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}}$ and $\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}}$:

$$\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}} \overset{i.i.d.}{\sim} N(1,1)$$

$$\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}} \overset{i.i.d.}{\sim} N(2,1)$$

Table 17: Monte Carlo Simulation

		PVAR			OLS		
N	T	Mean	Bias	RMSE	Mean	Bias	RMSE
50	50	$\begin{bmatrix} 0.613 & 0.325 \\ 0.226 & 0.609 \end{bmatrix}$	$\begin{bmatrix} 0.013 & 0.025 \\ 0.026 & 0.009 \end{bmatrix}$	$\begin{bmatrix} 0.040 & 0.045 \\ 0.046 & 0.039 \end{bmatrix}$	$\begin{bmatrix} 0.665 & 0.378 \\ 0.277 & 0.662 \end{bmatrix}$	$\begin{bmatrix} 0.065 & 0.078 \\ 0.077 & 0.062 \end{bmatrix}$	$\begin{bmatrix} 2.912 & 3.481 \\ 3.480 & 2.786 \end{bmatrix}$
	100	$\begin{bmatrix} 0.614 & 0.327 \\ 0.228 & 0.610 \end{bmatrix}$	$\begin{bmatrix} 0.014 & 0.027 \\ 0.028 & 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.037 & 0.043 \\ 0.044 & 0.036 \end{bmatrix}$	$\begin{bmatrix} 0.663 & 0.379 \\ 0.277 & 0.662 \end{bmatrix}$	$\begin{bmatrix} 0.063 & 0.079 \\ 0.077 & 0.062 \end{bmatrix}$	$\begin{bmatrix} 2.848 & 3.543 \\ 3.464 & 2.789 \end{bmatrix}$
	300	$\begin{bmatrix} 0.618 & 0.330 \\ 0.232 & 0.614 \end{bmatrix}$	$\begin{bmatrix} 0.018 & 0.030 \\ 0.032 & 0.014 \end{bmatrix}$	$\begin{bmatrix} 0.035 & 0.043 \\ 0.043 & 0.034 \end{bmatrix}$	$\begin{bmatrix} 0.664 & 0.378 \\ 0.277 & 0.662 \end{bmatrix}$	$\begin{bmatrix} 0.064 & 0.078 \\ 0.077 & 0.062 \end{bmatrix}$	$\begin{bmatrix} 2.865 & 3.519 \\ 3.459 & 2.804 \end{bmatrix}$
100	50	$\begin{bmatrix} 0.597 & 0.304 \\ 0.204 & 0.597 \end{bmatrix}$	$\begin{bmatrix} -0.002 & 0.004 \\ 0.004 & -0.002 \end{bmatrix}$	$\begin{bmatrix} 0.020 & 0.022 \\ 0.020 & 0.021 \end{bmatrix}$	$\begin{bmatrix} 0.668 & 0.372 \\ 0.275 & 0.664 \end{bmatrix}$	$\begin{bmatrix} 0.068 & 0.072 \\ 0.075 & 0.064 \end{bmatrix}$	$\begin{bmatrix} 3.076 & 3.233 \\ 3.396 & 2.888 \end{bmatrix}$
	100	$\begin{bmatrix} 0.597 & 0.305 \\ 0.204 & 0.597 \end{bmatrix}$	$\begin{bmatrix} -0.002 & 0.005 \\ 0.004 & -0.002 \end{bmatrix}$	$\begin{bmatrix} 0.019 & 0.021 \\ 0.019 & 0.021 \end{bmatrix}$	$\begin{bmatrix} 0.669 & 0.372 \\ 0.276 & 0.664 \end{bmatrix}$	$\begin{bmatrix} 0.069 & 0.072 \\ 0.076 & 0.064 \end{bmatrix}$	$\begin{bmatrix} 3.107 & 3.225 \\ 3.404 & 2.864 \end{bmatrix}$
	300	$\begin{bmatrix} 0.598 & 0.306 \\ 0.205 & 0.597 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.006 \\ 0.005 & -0.002 \end{bmatrix}$	$\begin{bmatrix} 0.017 & 0.020 \\ 0.018 & 0.019 \end{bmatrix}$	$\begin{bmatrix} 0.669 & 0.371 \\ 0.277 & 0.662 \end{bmatrix}$	$\begin{bmatrix} 0.069 & 0.071 \\ 0.077 & 0.062 \end{bmatrix}$	$\begin{bmatrix} 3.118 & 3.200 \\ 3.450 & 2.810 \end{bmatrix}$
300	50	$\begin{bmatrix} 0.599 & 0.301 \\ 0.201 & 0.598 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.011 \\ 0.010 & 0.011 \end{bmatrix}$	$\begin{bmatrix} 0.673 & 0.367 \\ 0.275 & 0.665 \end{bmatrix}$	$\begin{bmatrix} 0.073 & 0.067 \\ 0.075 & 0.065 \end{bmatrix}$	$\begin{bmatrix} 3.281 & 3.028 \\ 3.370 & 2.926 \end{bmatrix}$
	100	$\begin{bmatrix} 0.598 & 0.301 \\ 0.201 & 0.598 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.010 & 0.011 \\ 0.010 & 0.011 \end{bmatrix}$	$\begin{bmatrix} 0.673 & 0.367 \\ 0.276 & 0.664 \end{bmatrix}$	$\begin{bmatrix} 0.073 & 0.067 \\ 0.076 & 0.064 \end{bmatrix}$	$\begin{bmatrix} 3.290 & 3.020 \\ 3.398 & 2.892 \end{bmatrix}$
	300	$\begin{bmatrix} 0.598 & 0.300 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.000 \\ 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.010 \\ 0.009 & 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.673 & 0.366 \\ 0.275 & 0.665 \end{bmatrix}$	$\begin{bmatrix} 0.073 & 0.066 \\ 0.075 & 0.065 \end{bmatrix}$	$\begin{bmatrix} 3.301 & 2.996 \\ 3.394 & 2.907 \end{bmatrix}$

In Model 2, which includes a constant term, the PVAR-GFE estimator continues to outperform the OLS estimator. For instance, when $N = 50$ and $T = 50$, the bias of the PVAR-GFE estimator for one of the coefficients, Θ_1^0, is 0.013, while the corresponding bias under OLS is substantially larger at 0.065. In terms of root mean squared error (RMSE), PVAR-GFE also performs better, with an RMSE of 0.040 compared to an RMSE of 2.912 for OLSnearly a tenfold difference. These results demonstrate that the inclusion of a constant term does not diminish the superiority of PVAR-GFE relative to OLS; the grouped fixed effects approach remains both more efficient and more accurate in estimating the parameters. As N and T increase, the efficiency gain becomes more pronounced. For example, when $N = 100$ and $T = 50$, the RMSE for PVAR-GFE decreases to 0.037, while OLS still exhibits a high RMSE of 2.848.

3.9.5.3 Model 3: Bivariate PVAR-GFE(1) Model with No Lagged Dependent Variable

We propose the following DGP model:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\Theta_1^0} \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix},$$

Where the lag of dependent variable equals 0, the number of groups equals 2, and the number of dependent variables equals 2. Also, the coefficient of Θ_1^0 confirms this model is stationary. The errors are i.i.d and it follows a standard normal distribution. We replicate 2,000 times for the simulation.

Regarding the DGP of grouped fixed effects $\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}}$ and $\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}}$:

$$\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}} \overset{i.i.d.}{\sim} N(1,1)$$

$$\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha}_{\boldsymbol{g_i,t}} \overset{i.i.d.}{\sim} N(2,1)$$

Table 18: Monte Carlo Simulation

		PVAR			OLS		
N	T	Mean	Bias	RMSE	Mean	Bias	RMSE
50	50	$\begin{bmatrix} 0.629 & 0.332 \\ 0.235 & 0.625 \end{bmatrix}$	$\begin{bmatrix} 0.029 & 0.032 \\ 0.035 & 0.025 \end{bmatrix}$	$\begin{bmatrix} 0.048 & 0.050 \\ 0.052 & 0.045 \end{bmatrix}$	$\begin{bmatrix} 0.676 & 0.381 \\ 0.282 & 0.674 \end{bmatrix}$	$\begin{bmatrix} 0.076 & 0.081 \\ 0.082 & 0.074 \end{bmatrix}$	$\begin{bmatrix} 3.389 & 3.642 \\ 3.665 & 3.328 \end{bmatrix}$
	100	$\begin{bmatrix} 0.632 & 0.335 \\ 0.239 & 0.626 \end{bmatrix}$	$\begin{bmatrix} 0.032 & 0.035 \\ 0.039 & 0.026 \end{bmatrix}$	$\begin{bmatrix} 0.047 & 0.048 \\ 0.051 & 0.043 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.381 \\ 0.284 & 0.672 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.081 \\ 0.084 & 0.072 \end{bmatrix}$	$\begin{bmatrix} 3.423 & 3.604 \\ 3.758 & 3.208 \end{bmatrix}$
	300	$\begin{bmatrix} 0.635 & 0.337 \\ 0.241 & 0.629 \end{bmatrix}$	$\begin{bmatrix} 0.035 & 0.037 \\ 0.041 & 0.029 \end{bmatrix}$	$\begin{bmatrix} 0.045 & 0.048 \\ 0.050 & 0.042 \end{bmatrix}$	$\begin{bmatrix} 0.676 & 0.382 \\ 0.282 & 0.674 \end{bmatrix}$	$\begin{bmatrix} 0.076 & 0.082 \\ 0.082 & 0.074 \end{bmatrix}$	$\begin{bmatrix} 3.389 & 3.667 \\ 3.655 & 3.298 \end{bmatrix}$
100	50	$\begin{bmatrix} 0.605 & 0.306 \\ 0.208 & 0.604 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.006 \\ 0.008 & 0.004 \end{bmatrix}$	$\begin{bmatrix} 0.024 & 0.025 \\ 0.025 & 0.024 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.374 \\ 0.280 & 0.671 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.074 \\ 0.080 & 0.071 \end{bmatrix}$	$\begin{bmatrix} 3.435 & 3.309 \\ 3.568 & 3.187 \end{bmatrix}$
	100	$\begin{bmatrix} 0.605 & 0.308 \\ 0.208 & 0.604 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.008 \\ 0.008 & 0.004 \end{bmatrix}$	$\begin{bmatrix} 0.022 & 0.023 \\ 0.023 & 0.022 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.374 \\ 0.280 & 0.670 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.074 \\ 0.080 & 0.070 \end{bmatrix}$	$\begin{bmatrix} 3.434 & 3.332 \\ 3.584 & 3.146 \end{bmatrix}$
	300	$\begin{bmatrix} 0.606 & 0.309 \\ 0.209 & 0.605 \end{bmatrix}$	$\begin{bmatrix} 0.006 & 0.009 \\ 0.009 & 0.005 \end{bmatrix}$	$\begin{bmatrix} 0.019 & 0.021 \\ 0.020 & 0.020 \end{bmatrix}$	$\begin{bmatrix} 0.676 & 0.375 \\ 0.280 & 0.671 \end{bmatrix}$	$\begin{bmatrix} 0.076 & 0.075 \\ 0.080 & 0.071 \end{bmatrix}$	$\begin{bmatrix} 3.417 & 3.343 \\ 3.571 & 3.169 \end{bmatrix}$
300	50	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.011 \\ 0.009 & 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.279 & 0.669 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.079 & 0.069 \end{bmatrix}$	$\begin{bmatrix} 3.499 & 3.109 \\ 3.525 & 3.068 \end{bmatrix}$
	100	$\begin{bmatrix} 0.599 & 0.301 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.011 \\ 0.009 & 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.279 & 0.668 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.079 & 0.068 \end{bmatrix}$	$\begin{bmatrix} 3.484 & 3.119 \\ 3.536 & 3.059 \end{bmatrix}$
	300	$\begin{bmatrix} 0.599 & 0.301 \\ 0.201 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.008 & 0.010 \\ 0.008 & 0.009 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.279 & 0.668 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.079 & 0.068 \end{bmatrix}$	$\begin{bmatrix} 3.474 & 3.135 \\ 3.551 & 3.049 \end{bmatrix}$

Model 3, which removes the lagged dependent variable, reveals that PVAR-GFE continues to provide more accurate estimates even when the autoregressive component is removed. For instance, with $N = 100$ and $T = 50$, the Bias for PVAR-GFE is only 0.032 for the coefficients in Θ_1^0. while OLS again shows a higher Bias of 0.077. The RMSE values further reinforce PVAR-GFE's superiority, with PVAR-GFE showing an RMSE of 0.048 compared to OLS's 3.423. This indicates that PVAR-GFE is highly effective even when dealing with models that lack a lag structure.

3.9.5.4 Model 4: Bivariate PVAR-GFE(1) Model with No Granger Causality

We propose the following DGP model:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}}_{\Theta_1^0} \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix},$$

Where the lag of dependent variable equals 1, the number of groups equals 2, and the number of dependent variables equals 2. Also, the coefficient of Θ_1^0 confirms this model is stationary. The errors are i.i.d and it follows a standard normal distribution. We replicate 2,000 times for the simulation.

Regarding the DGP of grouped fixed effects $\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha_{g_i,t}}$ and $\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha_{g_i,t}}$:

$$\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha_{g_i,t}} \overset{i.i.d.}{\sim} N(1,1)$$

$$\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha_{g_i,t}} \overset{i.i.d.}{\sim} N(2,1)$$

Table 19: Monte Carlo Simulation

		PVAR			OLS		
T	N	Mean	Bias	RMSE	Mean	Bias	RMSE
50	50	$\begin{bmatrix} 0.629 & 0.040 \\ 0.040 & 0.630 \end{bmatrix}$	$\begin{bmatrix} 0.029 & 0.040 \\ 0.040 & 0.030 \end{bmatrix}$	$\begin{bmatrix} 0.060 & 0.067 \\ 0.066 & 0.061 \end{bmatrix}$	$\begin{bmatrix} 0.775 & 0.185 \\ 0.186 & 0.775 \end{bmatrix}$	$\begin{bmatrix} 0.175 & 0.185 \\ 0.186 & 0.175 \end{bmatrix}$	$\begin{bmatrix} 7.806 & 8.270 \\ 8.292 & 7.813 \end{bmatrix}$
	100	$\begin{bmatrix} 0.630 & 0.041 \\ 0.043 & 0.630 \end{bmatrix}$	$\begin{bmatrix} 0.030 & 0.041 \\ 0.043 & 0.030 \end{bmatrix}$	$\begin{bmatrix} 0.055 & 0.061 \\ 0.062 & 0.055 \end{bmatrix}$	$\begin{bmatrix} 0.773 & 0.186 \\ 0.186 & 0.774 \end{bmatrix}$	$\begin{bmatrix} 0.173 & 0.186 \\ 0.186 & 0.174 \end{bmatrix}$	$\begin{bmatrix} 7.767 & 8.317 \\ 8.325 & 7.798 \end{bmatrix}$
	300	$\begin{bmatrix} 0.631 & 0.042 \\ 0.042 & 0.631 \end{bmatrix}$	$\begin{bmatrix} 0.031 & 0.042 \\ 0.042 & 0.031 \end{bmatrix}$	$\begin{bmatrix} 0.049 & 0.055 \\ 0.056 & 0.049 \end{bmatrix}$	$\begin{bmatrix} 0.775 & 0.185 \\ 0.186 & 0.774 \end{bmatrix}$	$\begin{bmatrix} 0.175 & 0.185 \\ 0.186 & 0.174 \end{bmatrix}$	$\begin{bmatrix} 7.840 & 8.279 \\ 8.312 & 7.785 \end{bmatrix}$
100	50	$\begin{bmatrix} 0.600 & 0.007 \\ 0.006 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.007 \\ 0.006 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.025 & 0.026 \\ 0.026 & 0.024 \end{bmatrix}$	$\begin{bmatrix} 0.776 & 0.182 \\ 0.182 & 0.775 \end{bmatrix}$	$\begin{bmatrix} 0.176 & 0.182 \\ 0.182 & 0.175 \end{bmatrix}$	$\begin{bmatrix} 7.862 & 8.153 \\ 8.147 & 7.815 \end{bmatrix}$
	100	$\begin{bmatrix} 0.600 & 0.006 \\ 0.005 & 0.600 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.006 \\ 0.005 & 0.000 \end{bmatrix}$	$\begin{bmatrix} 0.021 & 0.022 \\ 0.022 & 0.021 \end{bmatrix}$	$\begin{bmatrix} 0.776 & 0.182 \\ 0.181 & 0.776 \end{bmatrix}$	$\begin{bmatrix} 0.176 & 0.182 \\ 0.181 & 0.176 \end{bmatrix}$	$\begin{bmatrix} 7.892 & 8.114 \\ 8.107 & 7.879 \end{bmatrix}$
	300	$\begin{bmatrix} 0.599 & 0.006 \\ 0.005 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.006 \\ 0.005 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.018 & 0.018 \\ 0.018 & 0.017 \end{bmatrix}$	$\begin{bmatrix} 0.776 & 0.182 \\ 0.182 & 0.776 \end{bmatrix}$	$\begin{bmatrix} 0.176 & 0.182 \\ 0.182 & 0.176 \end{bmatrix}$	$\begin{bmatrix} 7.845 & 8.149 \\ 8.127 & 7.880 \end{bmatrix}$
300	50	$\begin{bmatrix} 0.599 & 0.001 \\ 0.001 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.011 & 0.010 \\ 0.010 & 0.011 \end{bmatrix}$	$\begin{bmatrix} 0.777 & 0.180 \\ 0.180 & 0.777 \end{bmatrix}$	$\begin{bmatrix} 0.177 & 0.180 \\ 0.180 & 0.177 \end{bmatrix}$	$\begin{bmatrix} 7.906 & 8.029 \\ 8.029 & 7.911 \end{bmatrix}$
	100	$\begin{bmatrix} 0.599 & 0.001 \\ 0.001 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.009 \\ 0.009 & 0.009 \end{bmatrix}$	$\begin{bmatrix} 0.777 & 0.179 \\ 0.179 & 0.777 \end{bmatrix}$	$\begin{bmatrix} 0.177 & 0.179 \\ 0.179 & 0.177 \end{bmatrix}$	$\begin{bmatrix} 7.934 & 7.996 \\ 8.001 & 7.946 \end{bmatrix}$
	300	$\begin{bmatrix} 0.599 & 0.001 \\ 0.001 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.009 \\ 0.009 & 0.009 \end{bmatrix}$	$\begin{bmatrix} 0.777 & 0.179 \\ 0.179 & 0.777 \end{bmatrix}$	$\begin{bmatrix} 0.177 & 0.179 \\ 0.179 & 0.177 \end{bmatrix}$	$\begin{bmatrix} 7.919 & 7.999 \\ 8.010 & 7.945 \end{bmatrix}$

In Model 4, which introduces a no-Granger-causality assumption (i.e., the off-diagonal elements are set to zero), PVAR-GFE continues to outperform OLS. For example, with $N = 50$ and $T = 50$, the Bias for PVAR-GFE in estimating is 0.029 for coefficients in Θ_1^0, while OLS shows a much larger Bias of 0.175. Similarly, the RMSE for PVAR-GFE remains relatively low at 0.060, while OLS shows a dramatically larger RMSE of 7.806. These results indicate that PVAR-GFE performs well even in situations where there is no interdependence between variables, while OLS is severely biased and inefficient.

3.9.5.5 Model 5: Bivariate PVAR-GFE(1) Model with Heteroskedasticity Errors

We propose the following DGP model:

$$\begin{bmatrix} y_{1,i,t} \\ y_{2,i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0.6 & 0.3 \\ 0.2 & 0.6 \end{bmatrix}}_{\Theta_1^0} \begin{bmatrix} y_{1,i,t-1} \\ y_{2,i,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{1,g_i,t} \\ \alpha_{2,g_i,t} \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix},$$

Where the lag of dependent variable equals 2, the number of groups equals 2, and the number of dependent variables equals 2. Also, the coefficient of Θ_1^0 confirms this model is stationary. The errors follows settings from Ando & Bai (2016):

$$\mathbf{e}_{it} = 0.9\boldsymbol{\varepsilon}_{it}^1 + \delta_t 0.9\boldsymbol{\varepsilon}_{it}^2 \tag{39}$$

Where $\delta_t = 1$ if t is an odd number while $\delta_t = 0$ if t is an even number. We replicate 2,000 times for the simulation.

Regarding the DGP of grouped fixed effects $\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha}_{\boldsymbol{g_i},\boldsymbol{t}}$ and $\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha}_{\boldsymbol{g_i},\boldsymbol{t}}$:

$$\mathbf{1}_{\{g_i=1\}}\boldsymbol{\alpha}_{\boldsymbol{g_i},\boldsymbol{t}} \overset{i.i.d.}{\sim} N(1,1)$$

$$\mathbf{1}_{\{g_i=2\}}\boldsymbol{\alpha}_{\boldsymbol{g_i},\boldsymbol{t}} \overset{i.i.d.}{\sim} N(2,1)$$

Table 20: Monte Carlo Simulation

		PVAR			OLS		
T	N	Mean	Bias	RMSE	Mean	Bias	RMSE
50	50	$\begin{bmatrix} 0.629 & 0.332 \\ 0.235 & 0.625 \end{bmatrix}$	$\begin{bmatrix} 0.029 & 0.032 \\ 0.035 & 0.025 \end{bmatrix}$	$\begin{bmatrix} 0.048 & 0.050 \\ 0.052 & 0.045 \end{bmatrix}$	$\begin{bmatrix} 0.676 & 0.381 \\ 0.282 & 0.674 \end{bmatrix}$	$\begin{bmatrix} 0.076 & 0.081 \\ 0.082 & 0.074 \end{bmatrix}$	$\begin{bmatrix} 3.389 & 3.642 \\ 3.665 & 3.328 \end{bmatrix}$
	100	$\begin{bmatrix} 0.632 & 0.335 \\ 0.239 & 0.626 \end{bmatrix}$	$\begin{bmatrix} 0.032 & 0.035 \\ 0.039 & 0.026 \end{bmatrix}$	$\begin{bmatrix} 0.047 & 0.048 \\ 0.051 & 0.043 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.381 \\ 0.284 & 0.672 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.081 \\ 0.084 & 0.072 \end{bmatrix}$	$\begin{bmatrix} 3.423 & 3.604 \\ 3.758 & 3.208 \end{bmatrix}$
	300	$\begin{bmatrix} 0.635 & 0.337 \\ 0.241 & 0.629 \end{bmatrix}$	$\begin{bmatrix} 0.035 & 0.037 \\ 0.041 & 0.029 \end{bmatrix}$	$\begin{bmatrix} 0.045 & 0.048 \\ 0.050 & 0.042 \end{bmatrix}$	$\begin{bmatrix} 0.676 & 0.382 \\ 0.282 & 0.674 \end{bmatrix}$	$\begin{bmatrix} 0.076 & 0.082 \\ 0.082 & 0.074 \end{bmatrix}$	$\begin{bmatrix} 3.389 & 3.667 \\ 3.655 & 3.298 \end{bmatrix}$
100	50	$\begin{bmatrix} 0.605 & 0.306 \\ 0.208 & 0.604 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.006 \\ 0.008 & 0.004 \end{bmatrix}$	$\begin{bmatrix} 0.024 & 0.025 \\ 0.025 & 0.024 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.374 \\ 0.280 & 0.671 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.074 \\ 0.080 & 0.071 \end{bmatrix}$	$\begin{bmatrix} 3.435 & 3.309 \\ 3.568 & 3.187 \end{bmatrix}$
	100	$\begin{bmatrix} 0.605 & 0.308 \\ 0.208 & 0.604 \end{bmatrix}$	$\begin{bmatrix} 0.005 & 0.008 \\ 0.008 & 0.004 \end{bmatrix}$	$\begin{bmatrix} 0.022 & 0.023 \\ 0.023 & 0.022 \end{bmatrix}$	$\begin{bmatrix} 0.677 & 0.374 \\ 0.280 & 0.670 \end{bmatrix}$	$\begin{bmatrix} 0.077 & 0.074 \\ 0.080 & 0.070 \end{bmatrix}$	$\begin{bmatrix} 3.434 & 3.332 \\ 3.584 & 3.146 \end{bmatrix}$
	300	$\begin{bmatrix} 0.606 & 0.309 \\ 0.209 & 0.605 \end{bmatrix}$	$\begin{bmatrix} 0.006 & 0.009 \\ 0.009 & 0.005 \end{bmatrix}$	$\begin{bmatrix} 0.019 & 0.021 \\ 0.020 & 0.020 \end{bmatrix}$	$\begin{bmatrix} 0.676 & 0.375 \\ 0.280 & 0.671 \end{bmatrix}$	$\begin{bmatrix} 0.076 & 0.075 \\ 0.080 & 0.071 \end{bmatrix}$	$\begin{bmatrix} 3.417 & 3.343 \\ 3.571 & 3.169 \end{bmatrix}$
300	50	$\begin{bmatrix} 0.600 & 0.300 \\ 0.200 & 0.600 \end{bmatrix}$	$\begin{bmatrix} -0.000 & 0.000 \\ 0.000 & -0.000 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.011 \\ 0.009 & 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.279 & 0.669 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.079 & 0.069 \end{bmatrix}$	$\begin{bmatrix} 3.499 & 3.109 \\ 3.525 & 3.068 \end{bmatrix}$
	100	$\begin{bmatrix} 0.599 & 0.301 \\ 0.200 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.000 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.009 & 0.011 \\ 0.009 & 0.010 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.279 & 0.668 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.079 & 0.068 \end{bmatrix}$	$\begin{bmatrix} 3.484 & 3.119 \\ 3.536 & 3.059 \end{bmatrix}$
	300	$\begin{bmatrix} 0.599 & 0.301 \\ 0.201 & 0.599 \end{bmatrix}$	$\begin{bmatrix} -0.001 & 0.001 \\ 0.001 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.008 & 0.010 \\ 0.008 & 0.009 \end{bmatrix}$	$\begin{bmatrix} 0.678 & 0.370 \\ 0.279 & 0.668 \end{bmatrix}$	$\begin{bmatrix} 0.078 & 0.070 \\ 0.079 & 0.068 \end{bmatrix}$	$\begin{bmatrix} 3.474 & 3.135 \\ 3.551 & 3.049 \end{bmatrix}$

Finally, Model 5, which incorporates heteroskedasticity in the error terms, demonstrates PVAR-GFE's robustness in handling violations of the homoskedasticity assumption. For instance, with $N = 100$ and $T = 50$, PVAR-GFE's Bias for is 0.032, whereas OLS has a Bias of 0.076. The RMSE values again support the efficiency of PVAR-GFE, with an RMSE of 0.047 for PVAR-GFE compared to OLS's 3.423. Even under heteroskedastic errors, PVAR-GFE provides more accurate and efficient parameter estimates than OLS.

3.9.5.6 Misclassification Rate

The table presents the average misclassification rates for five models (Model 1 to Model 5) across different panel sizes, characterised by varying numbers of cross-sectional units N and time periods T. The misclassification rate reflects the proportion of incorrectly assigned group memberships when estimating the grouped structure using the PVAR-GFE model. As N increases, the misclassification rate decreases across all models, indicating improved accuracy in group classification. For instance, when $T = 50$, the misclassification rate for Model 1 decreases from 0.009 when $N = 50$ to 0.007 when $N = 300$, demonstrating that larger panels with more cross-sectional units enhance classification performance. Similarly, an increase in the time dimension T also leads to lower misclassification rates. For example, with $N = 50$, the misclassification rate for Model 1 declines from 0.009 when $T = 50$ to 0.000 when $T = 300$. Across all models, panels with a larger time dimension (e.g., $T = 300$) achieve very low misclassification rates, particularly when the number of cross-sectional units is also large. When comparing models, Model 1 consistently shows the lowest misclassification rates across most panel settings, closely followed by Models 2, 3, and 4, while Model 5 generally exhibits higher misclassification rates, especially when N is small. Overall, the models perform best when both N and T are large, with near-perfect group classification achieved when $N = 100$ or $N = 300$ and $T = 300$, where the misclassification rate across all models approaches zero. This suggests that larger panels, both in terms of cross-sectional units and time periods, significantly improve the accuracy of the PVAR-GFE model's group classification, with Model 1 consistently outperforming the others, while Model 5 tends to exhibit higher misclassification rates, particularly in smaller panels.

Table 21: Average Missclassification Rate

		Missclassification Rate				
T	N	Model 1	Model 2	Model 3	Model 4	Model 5
50	50	0.009	0.011	0.011	0.013	0.028
	100	0.008	0.009	0.009	0.011	0.021
	300	0.007	0.008	0.008	0.009	0.019
100	50	0.001	0.001	0.001	0.001	0.003
	100	0.000	0.000	0.000	0.001	0.002
	100	0.000	0.000	0.000	0.000	0.002
300	50	0.000	0.000	0.000	0.000	0.000
	100	0.000	0.000	0.000	0.000	0.000
	300	0.000	0.000	0.000	0.000	0.000

3.10 Appendix B: Proofs

3.10.1 Proofs of the Homogeneous Coefficient Model

Lemma 1

$$\operatorname*{plim}_{N,T\to\infty} \sup_{(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\in\boldsymbol{\Omega}\times\boldsymbol{\Theta}^{GTM}\times\boldsymbol{\Gamma}_G} \left|\hat{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau) - \tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\right| = o_p(1) \tag{40}$$

Proof 1 (Proof of Lemma 1) *Define the following auxiliary objective function as:*

$$\begin{aligned}\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau) &= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0-\boldsymbol{\beta})+\boldsymbol{\alpha}^0_{g_i^0,t}-\boldsymbol{\alpha}_{g_i,t}\right\|^2 \\ &+\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t}\right\|^2\end{aligned} \tag{41}$$

Then we have:

$$
\begin{aligned}
\mathbb{Q} &= \hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) \\
&= \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \boldsymbol{\epsilon}_{i,t} + \boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{g_i^0,t} - \boldsymbol{\alpha}_{g_i,t} \right\|^2 \\
&- \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{g_i^0,t} - \boldsymbol{\alpha}_{g_i,t} \right\|^2 \\
&- \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \boldsymbol{\epsilon}_{i,t} \right\|^2 \\
&= \frac{2}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\boldsymbol{\epsilon}_{i,t} \left(\boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{g_i^0,t} - \boldsymbol{\alpha}_{g_i,t} \right)' \right] \\
&= (\boldsymbol{\beta}^0 - \boldsymbol{\beta}) \frac{2}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} + \frac{2}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{\alpha}^{0\prime}_{g_i^0,t} \boldsymbol{\epsilon}_{i,t} - \frac{2}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{\alpha}'_{g_i,t} \boldsymbol{\epsilon}_{i,t}
\end{aligned} \tag{42}
$$

For term $(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) \frac{2}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{z}'_{i,t} \boldsymbol{\epsilon}_{i,t}$*, it can be viewed as:*

$$
(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) \frac{2}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} = (\boldsymbol{\beta}^0 - \boldsymbol{\beta}) \frac{2}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \begin{bmatrix} \boldsymbol{\epsilon}_{i,t} \otimes \mathcal{Y}_{i,t-1} \\ \boldsymbol{\epsilon}_{i,t} \otimes x_{i,t} \end{bmatrix} \tag{43}
$$

From Assumption 6 (c), (d), and (e), we obtain the following inequality via the Cauchy-Schwarz (CS) inequality:

$$
\begin{aligned}
\frac{2}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} &= \frac{1}{MN} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} \\
&\leq \frac{1}{MN} \left(\sum_{i=1}^{N} 1^2 \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} \right\|^2 \right)^{1/2} \\
&= \frac{1}{M} \left(\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} \right\|^2 \right)^{1/2} \\
&\leq O_p\left(T^{-1/2}\right)
\end{aligned} \tag{44}
$$

The last inequality from Assumption 6 (g) that:

$$
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E} \left\| \mathbf{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} \right\|^2 = \mathbb{E} \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \frac{1}{T} \mathbf{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} \right\|^2 \right) \leq \frac{\mathbb{M}}{T} \tag{45}
$$

Therefore, we conclude that:

$$
\frac{2}{TNM} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{z}'_{i,t} \boldsymbol{\epsilon}_{i,t} = O_p\left(\frac{1}{\sqrt{T}}\right) = o_p(1) \tag{46}
$$

Next we show that $\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}'_{g_i,t}\boldsymbol{\epsilon}_{i,t}$ *is uniformly* $o_p(1)$

$$\begin{aligned}\frac{1}{NTM}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}'_{g_i,t}\boldsymbol{\epsilon}_{i,t} &= \frac{1}{M}\sum_{g=1}^{G}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\left\{g_i=g\right\}\boldsymbol{\alpha}'_{g,t}\boldsymbol{\epsilon}_{i,t}\right] \\ &= \frac{1}{M}\sum_{g=1}^{G}\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\alpha}_{g,t}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{g_i=g\right\}\boldsymbol{\epsilon}_{i,t}\right)\right]\end{aligned} \tag{47}$$

Where by Cauchy-Schwarz inequality:

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\alpha}_{g,t}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{g_i=g\right\}\boldsymbol{\epsilon}_{i,t}\right)\right\|^2 \leq \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{g,t}\right\|^2\right)\times\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{g_i=g\right\}\boldsymbol{\epsilon}_{i,t}\right\|^2\right) \tag{48}$$

Since $\frac{1}{T}\sum_{t=1}^{T}\left|\boldsymbol{\alpha}_{g,t}\right|^2$ *is uniformly bounded by Assumption (a), we now proceed to show that the second term is also uniformly bounded.*

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{g_i=g\right\}\boldsymbol{\epsilon}_{i,t}\right\|^2 \\ &= \frac{1}{TN^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}\left\{g_i=g\right\}\mathbf{1}\left\{g_j=g\right\}\sum_{t=1}^{T}\boldsymbol{\epsilon}'_{i,t}\boldsymbol{\epsilon}_{j,t} \\ &\leq \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}'_{i,t}\boldsymbol{\epsilon}_{j,t}\right| \\ &= O_p(N^{-1/2}T^{-1/2})\end{aligned} \tag{49}$$

The last equality holds by Cauchy-Schwarz inequality:

$$\begin{aligned}&\mathbb{E}\left\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}'_{i,t}\boldsymbol{\epsilon}_{j,t}\right|\right\|^2 \\ &\leq \mathbb{E}\left(\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}'_{i,t}\boldsymbol{\epsilon}_{j,t}\right\|^2\right) \\ &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\left\|\boldsymbol{\epsilon}'_{i,t}\boldsymbol{\epsilon}_{j,t}\right\|^2\right) \\ &= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\left\|\boldsymbol{\epsilon}_{i,t}\right\|^4\right) \\ &= O_p(N^{-1}T^{-1})\end{aligned} \tag{50}$$

Thus we have

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{g_i=g\right\}\boldsymbol{\epsilon}_{i,t}\right\|^2 = o_p(1) \tag{51}$$

The uniformly bounded term $\frac{1}{NTM}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}'_{g_i,t}\boldsymbol{\epsilon}_{i,t}$ *also implies that* $\frac{1}{NTM}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}^{0\prime}_{g_i^0,t}\boldsymbol{\epsilon}_{i,t}$ *is uniformly bounded. We therefore proof Lemma 1.*

Next we show that $\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)$ is uniquely minimised at true values of $\boldsymbol{\beta^0},\boldsymbol{\alpha^0},\tau^0$ by the following lemma.

Lemma 2 *For* $\boldsymbol{\beta}\neq\boldsymbol{\beta^0}$, $\boldsymbol{\alpha}\neq\boldsymbol{\alpha^0}$ *and* $\tau\neq\tau^0$,

$$\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)-\tilde{\mathrm{Q}}(\boldsymbol{\beta^0},\boldsymbol{\alpha^0},\tau^0)\geq\rho\left\|\boldsymbol{\beta}^0-\boldsymbol{\beta}\right\|^2 \tag{52}$$

.

Proof 2 (Proof of Lemma 2) *Recall the auxiliary objective function* $\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)$, *it reduces to*

$$\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t}\right\|^2 \tag{53}$$

When it reaches at its true value $\tilde{\mathrm{Q}}(\boldsymbol{\beta^0},\boldsymbol{\alpha^0},\tau^0)$. *Therefore, we have:*

$$\begin{aligned}
\mathbb{Q}&=\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)-\tilde{\mathrm{Q}}(\boldsymbol{\beta^0},\boldsymbol{\alpha^0},\tau^0)\\
&=\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0-\boldsymbol{\beta})+\boldsymbol{\alpha}^0_{g_i,t}-\boldsymbol{\alpha}_{g_i,t}\right\|^2\\
&=\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\left[(\boldsymbol{z}_{i,t}-\boldsymbol{z}_{g\wedge\tilde{g},t})(\boldsymbol{\beta}^0-\boldsymbol{\beta})+\boldsymbol{B}\right]\right\|^2\\
&\geq(\boldsymbol{\beta}^0-\boldsymbol{\beta})'\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{z}_{i,t}-\boldsymbol{z}_{g\wedge\tilde{g},t})'(\boldsymbol{z}_{i,t}-\boldsymbol{z}_{g\wedge\tilde{g},t})(\boldsymbol{\beta}^0-\boldsymbol{\beta})\\
&\overset{1}{=}(\boldsymbol{\beta}^0-\boldsymbol{\beta})'\varSigma(\gamma)(\boldsymbol{\beta}^0-\boldsymbol{\beta})\\
&\geq(\boldsymbol{\beta}^0-\boldsymbol{\beta})'\min_{\gamma\in\boldsymbol{\Gamma}_G}\varSigma(\gamma)(\boldsymbol{\beta}^0-\boldsymbol{\beta})\\
&\overset{2}{\geq}\min_{\gamma\in\boldsymbol{\Gamma}_G}\hat{\rho}(\gamma)\left\|\boldsymbol{\beta}^0-\boldsymbol{\beta}\right\|^2
\end{aligned} \tag{54}$$

Where $\overset{1}{=}$ *comes from Assumption 6 (h) and* $\overset{2}{\geq}$ *comes from eigenvalue inequalities* *for quadratic forms*[23]*. The intuition behind the equation is that the difference of objective function values is bounded away from zero when parameters are not equal to their true values.*

Proof 3 (Proof of Theorem 5) *The consistency can be shown by:*

$$\begin{aligned}\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) &\stackrel{1}{=} \hat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) + o_p(1) \\ &\stackrel{2}{\leq} \hat{\mathrm{Q}}(\boldsymbol{\beta^0}, \boldsymbol{\alpha}^0, \tau^0) + o_p(1) \\ &\stackrel{3}{=} \tilde{\mathrm{Q}}(\boldsymbol{\beta^0}, \boldsymbol{\alpha}^0, \tau^0) + o_p(1)\end{aligned} \tag{55}$$

Where $\stackrel{1}{=}$ *and* $\stackrel{3}{=}$ *is from Lemma 1,* $\stackrel{2}{\leq}$ *is the definition of least squares where the estimator are estimated by the argmin of objective function. However,* $\tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ *is minimised at its true value. Then we have:*

$$\left[\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta^0}, \boldsymbol{\alpha}^0, \tau^0)\right] \geq 0 \tag{56}$$

Combine equation 55 and equation 56 we have:

$$\left[\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta^0}, \boldsymbol{\alpha}^0, \tau^0)\right] = o_p(1) \tag{57}$$

Since $\tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ *includes grouped fixed effects estimators whose dimension increases with* T*, it is important to establish the consistency of these estimators. In what follows, we show that the grouped fixed effects estimators are consistent in an*

[23] suppose $A = A^T, A = Q\Lambda Q^T$ with eigenvalues sorted so $\lambda_1 \geq \cdots \geq \lambda_n$

$$\begin{aligned}x^T A x &= x^T Q \Lambda Q^T x \\ &= \left(Q^T x\right)^T \Lambda \left(Q^T x\right) \\ &= \sum_{i=1}^{n} \lambda_i \left(q_i^T x\right)^2 \\ &\geq \lambda_n \sum_{i=1}^{n} \left(q_i^T x\right)^2 \\ &= \lambda_1 \|x\|^2\end{aligned}$$

i.e., we have $x^T A x \geq \lambda_1 x^T x$

averaged norm:

$$
\begin{aligned}
\mathbb{Q} &= \left|\frac{1}{MNT}\left[\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \hat{\boldsymbol{\alpha}}, \hat{\tau})\right]\right| \\
&= \left|\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}) + \boldsymbol{\alpha}^0_{g_i^0,t} - \hat{\boldsymbol{\alpha}}_{\hat{g}_i,t}\right\|^2 - \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{g_i^0,t} - \hat{\boldsymbol{\alpha}}_{\hat{g}_i,t}\right\|^2\right| \\
&\leq \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left|[\boldsymbol{A}+\boldsymbol{B}]'[\boldsymbol{A}+\boldsymbol{B}] - \boldsymbol{B}'\boldsymbol{B}\right| \\
&= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\boldsymbol{A}'\boldsymbol{A} + \boldsymbol{A}'\boldsymbol{B} + \boldsymbol{B}'\boldsymbol{A}| \\
&= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\boldsymbol{A}'\boldsymbol{A} + 2\boldsymbol{A}'\boldsymbol{B}| \\
&\overset{1}{\leq} \|\boldsymbol{A}\|\,\|\boldsymbol{A} + 2\boldsymbol{B}\| \\
&\overset{2}{\leq} \|\boldsymbol{A}\|\left(\|\boldsymbol{A}\| + \|2\boldsymbol{B}\|\right) \\
&= \|\boldsymbol{A}\|^2 + 2\|\boldsymbol{A}\|\,\|\boldsymbol{B}\| \\
&\overset{3}{=} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\boldsymbol{z}_{i,t}\|^2 \times \left\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\right\|^2 + \left(2\sup_{\boldsymbol{\alpha}\in\mathcal{A}}|\boldsymbol{\alpha}_t|\right) \times \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\boldsymbol{z}_{i,t}\| \times \left\|\boldsymbol{\beta}^0 - \widehat{\boldsymbol{\beta}}\right\|
\end{aligned} \tag{58}
$$

For simplicity, we denote $\boldsymbol{A}$ *is* $\left(\boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})\right)$ *and* $\boldsymbol{B}$ *is* $(\boldsymbol{\alpha}^0_{g_i,t} - \hat{\boldsymbol{\alpha}}_{\hat{g}_i,t})$*.* $\overset{1}{\leq}$ *is Schwarz Inequality.* $\overset{2}{\leq}$ *is Triangle Inequality.* $(2\sup_{\boldsymbol{\alpha}\in\mathcal{A}}|\boldsymbol{\alpha}_t|)$ *in* $\overset{3}{=}$ *comes from Assumption 6 (a) that* $\boldsymbol{\alpha}$ *is bounded. Therefore, we can see that* $\left|\frac{1}{MNT}\left[\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \hat{\boldsymbol{\alpha}}, \hat{\tau})\right]\right|$ *is* $o_p(1)$ *by Assumption 6 (a) and Equation 55. Then we have:*

$$
\begin{aligned}
\mathbb{Q} &= \left|\frac{1}{MNT}\left[\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \hat{\boldsymbol{\alpha}}, \hat{\tau})\right]\right| \\
&= \left|\frac{1}{MNT}\left[\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) + \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^0, \tau^0) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^0, \tau^0) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \hat{\boldsymbol{\alpha}}, \hat{\tau})\right]\right| \\
&\overset{1}{=} \left|\frac{1}{MNT}\left[\tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^0, \tau^0) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \hat{\boldsymbol{\alpha}}, \hat{\tau})\right]\right| \\
&= o_p(1)
\end{aligned} \tag{59}
$$

Where $\overset{1}{=}$ *comes from equation 57. From* $\left|\frac{1}{MNT}\left[\tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^0, \tau^0) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \hat{\boldsymbol{\alpha}}, \hat{\tau})\right]\right| = o_p(1)$ *we have:*

$$
\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{g_i,t} - \hat{\boldsymbol{\alpha}}_{\hat{g}_i^0,t}\right\|^2 = o_p(1) \tag{60}
$$

We now proof a set of lemmas that will be used in the following Theorem about asymptotic distribution. We now proof that $\hat{\boldsymbol{\alpha}}$ is consistent for $\boldsymbol{\alpha}^0$ regarding Hausdorff distance d_H in $\mathbb{R}^{GTM}$. We define Hausdorff distance as:

$$d_H(A,B)^2 = \max\{H(A,B)^2, H(B,A)^2\} \tag{61}$$

Where

$$\begin{aligned} H(A,B)^2 &= \max_{a\in A}\left\{\min_{b\in B}\|a-b\|^2\right\} \\ H(B,A)^2 &= \max_{b\in B}\left\{\min_{a\in A}\|b-a\|^2\right\} \end{aligned} \tag{62}$$

As the objective function is invariant of relabelling of groups, thus we have the Hausdorff distance of $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}^0$ as:

$$\begin{aligned} d_H(\hat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}^0)^2 &= \max\left\{H(A,B)^2, H(B,A)^2\right\} \\ &= \max\left\{ \max_{g\in 1,2\ldots G}\{\min_{\tilde{g}\in\{1,2\ldots G\}} \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\tilde{g},t}, -\boldsymbol{\alpha}^0_{g,t}\|^2\}, \right. \\ &\left. \max_{\tilde{g}\in 1,2\ldots G}\{\min_{g\in\{1,2\ldots G\}} \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\tilde{g},t}, -\boldsymbol{\alpha}^0_{g,t}\|^2\} \right\} \end{aligned} \tag{63}$$

Lemma 3 *$\hat{\boldsymbol{\alpha}}$ is consistent for $\boldsymbol{\alpha}^0$ regarding Hausdorff distance: $d_H(\hat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}^0) = o_p(1)$*

Proof 4 (Proof of Lemma 3) *We first show that the inside term $\min_{\tilde{g}\in\{1,2\ldots G\}} \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\tilde{g},t}, -\boldsymbol{\alpha}^0_{g,t}\|^2 = o_p(1)$ for all $g \in 1,2\ldots G$. By Assumption 7 (a) and (b) is it clearly that*

$$\begin{aligned} \min_{\tilde{g}\in\{1,2\ldots G\}} \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\tilde{g},t}, -\boldsymbol{\alpha}^0_{g,t}\|^2 &= \left(\frac{1}{MN}\sum_{n=1}^{N}\mathbf{1}\left\{g_i^0 = g\right\}\right)\left(\min_{\tilde{g}\in\{1,2\ldots G\}}\frac{1}{T}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\tilde{g},t}, -\boldsymbol{\alpha}^0_{g,t}\|^2\right) \\ &\overset{1}{\leq} \left(\frac{1}{MN}\sum_{n=1}^{N}\mathbf{1}\left\{g_i^0 = g\right\}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\hat{g}_i,t}, -\boldsymbol{\alpha}^0_{g,t}\|^2\right) \\ &\overset{2}{\leq} \frac{1}{MNT}\left(\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\hat{g}_i,t}, -\boldsymbol{\alpha}^0_{g_i^0,t}\|^2\right) \\ &\overset{3}{=} o_p(1) \end{aligned} \tag{64}$$

Where $\overset{1}{\leq}$ comes from the definition of the minimum function. Therefore, $\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\hat{g}_i,t}, -\boldsymbol{\alpha}^0_{g,t}\|^2\right)$ could only be equal or greater than

$\left(\min_{\tilde g\in\{1,2...G\}} \frac{1}{T}\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\tilde g,t}, -\boldsymbol\alpha^0_{g,t}\|^2\right)$. $\overset{2}{\leq}$ is from the fact that the sum of the units belongs to g of the term $\left(\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\hat g_i,t}, -\boldsymbol\alpha^0_{g_i^0,t}\|^2\right)$ should only be equal or lower than the sum of all units of that term. $\overset{3}{=}$ is proofed by Theorem 5. This completes the first part of Hausdorff distance $\min_{\tilde g\in\{1,2...G\}} \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\tilde g,t}, -\boldsymbol\alpha^0_{g,t}\|^2$ is $o_p(1)$.

We next prove that the second part of Hausdorff distance $\min_{g\in\{1,2...G\}} \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\tilde g,t}, -\boldsymbol\alpha^0_{g,t}\|^2$ is $o_p(1)$ Next we define function $\sigma(g)$ as:

$$\sigma(g) = \underset{\tilde g\in\{1,\ldots,G\}}{\operatorname{argmin}} \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\tilde g,t} - \boldsymbol\alpha^0_{g,t}\|^2 \tag{65}$$

Where the function selects the optimal $\widetilde{g} \in \{1,\ldots,G\}$ based on the minimised function $\frac{1}{MT}\sum_{t=1}^{T}\left(\widehat{\alpha}_{\widetilde{g}t} - \alpha^0_{gt}\right)^2$. We next show that $\sigma : \{1,\ldots,G\} \to \{1,\ldots,G\}$ is an one-to-one function w.p.a.1. Let $g \neq \tilde g$, we have:

$$\begin{aligned}
&\left\|\frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\sigma(g),t} - \hat{\boldsymbol\alpha}_{\sigma(\tilde g),t}\|^2\right\| + \underbrace{\left\|\frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\sigma(\tilde g),t} - \boldsymbol\alpha^0_{\tilde g,t}\|^2\right\|}_{a.1} + \underbrace{\left\|\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol\alpha^0_{g,t} - \hat{\boldsymbol\alpha}_{\sigma(g),t}\|^2\right\|}_{a.2}\\
&\overset{1}{\geq} \left\|\frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\sigma(g),t} - \hat{\boldsymbol\alpha}_{\sigma(\tilde g),t}\|^2 + \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol\alpha}_{\sigma(\tilde g),t} - \boldsymbol\alpha^0_{\tilde g,t}\|^2 + \frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol\alpha^0_{g,t} - \hat{\boldsymbol\alpha}_{\sigma(g),t}\|^2\right\|\\
&\underbrace{\overset{2}{\geq} \left\|\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol\alpha^0_{g,t} - \boldsymbol\alpha^0_{\tilde g,t}\|^2\right\|}_{a.3}\\
&= O_p(c_{g,\tilde g}^{-1/2})
\end{aligned} \tag{66}$$

Where by triangle inequality we have $\overset{1}{\geq}$ and $\overset{2}{\geq}$, a.1 and a.2 are $o_p(1)$ by Equation 64, a.3 is by Assumption 7 (b). Thus we have the conclusion that $\sigma : \{1,\ldots,G\} \to \{1,\ldots,G\}$ is an one-to-one mapping function w.p.a.1 for $g \neq \tilde g$ and the inverse function σ^{-1} is well defined. Thus we have:

$$\tilde g = \sigma(\sigma^{-1}(\tilde g)) \tag{67}$$

and for all $\tilde{g} \in \{1, \ldots, G\}$*:*

$$\begin{aligned}
\min_{g \in \{1,2\ldots G\}} \frac{1}{MT} \sum_{t=1}^{T} \|\hat{\boldsymbol{\alpha}}_{\tilde{g},t}, -\boldsymbol{\alpha}^0_{g,t}\|^2 &\overset{1}{\leq} \frac{1}{MT} \sum_{t=1}^{T} \|\hat{\boldsymbol{\alpha}}_{\tilde{g},t}, -\boldsymbol{\alpha}^0_{\sigma^{-1}(\tilde{g}),t}\|^2 \\
&\overset{2}{=} \min_{h \in \{1,2\ldots G\}} \frac{1}{MT} \sum_{t=1}^{T} \|\hat{\boldsymbol{\alpha}}_{h,t}, -\boldsymbol{\alpha}^0_{\sigma^{-1}(\tilde{g}),t}\|^2 \\
&\overset{3}{=} o_p(1)
\end{aligned} \tag{68}$$

Where $\overset{3}{=}$ *comes from Equation 64. Combine Equation 64 and Equation 68 we complete the proof of Lemma 3 that the Hausdorff distance is* $o_p(1)$

We now prove that the estimated group membership $\hat{g}_i$ is consistent with the true group assignment g_i^0 at a rate of $T^{-\delta}$, for sufficiently small η, as both N and T tend to infinity.

Lemma 4 *Let* $\hat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \underset{g \in \{1,\ldots,G\}}{\operatorname{argmin}} \frac{1}{MT} \sum_{t=1}^{T} \|(\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i,t})\|^2$*, for some* $\eta > 0$ *but small enough while* N *and* T *go to infinity, we have for all* $\delta > 0$ *such that:*

$$\sup_{(\boldsymbol{\beta},\boldsymbol{\alpha}) \in \mathcal{N}_\eta} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left\{\widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \neq g_i^0\right\} = o_p\left(T^{-\delta}\right) \tag{69}$$

Where $\mathcal{N}_\eta$ *is the set of parameter satisfies:*

$$\mathcal{N}_\eta = \left\{\boldsymbol{\beta}, \boldsymbol{\alpha} \in \boldsymbol{\Omega} \times \boldsymbol{\Theta}^{GTM} : \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\|^2 < \eta, \frac{1}{MT} \sum_{t=1}^{T} \left\|\boldsymbol{\alpha}^0_{g,t} - \boldsymbol{\alpha}_{g,t}\right\|^2 < \eta, g \in \{1, 2, \ldots, G\}\right\} \tag{70}$$

Proof 5 (Proof of Lemma 4) *For all* $g \in \{1, 2, \ldots, G\}$*, we have the following inequality:*

$$\mathbf{1}\left\{\widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = g\right\} \leq \mathbf{1}\left\{\sum_{t=1}^{T} \|\mathbf{y}_{it} - \mathbf{z}_{it}\boldsymbol{\beta} - \boldsymbol{\alpha}_{gt}\|^2 \leq \sum_{t=1}^{T} \left\|\mathbf{y}_{it} - \mathbf{z}_{it}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i^0 t}\right\|^2\right\} \tag{71}$$

Sum across units N *we have:*

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left\{\widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \neq g_i^0\right\} &= \sum_{g=1}^{G} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 \neq g\right\} \mathbf{1}\left\{\widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = g\right\} \\
&\leq \sum_{g=1}^{G} \frac{1}{N} \sum_{i=1}^{N} Z_{ig}(\boldsymbol{\beta}, \boldsymbol{\alpha})
\end{aligned} \tag{72}$$

Where

$$
\begin{aligned}
Z_{ig}(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \mathbf{1}\left\{g_i^0 \neq g\right\} \times \mathbf{1}\left\{\sum_{t=1}^{T} \|\mathbf{y}_{it} - \mathbf{z}_{it}\boldsymbol{\beta} - \boldsymbol{\alpha}_{gt}\|^2 \leq \sum_{t=1}^{T} \|\mathbf{y}_{it} - \mathbf{z}_{it}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g^0 t}\|^2\right\} \\
&= \mathbf{1}\left\{g_i^0 \neq g\right\} \times \mathbf{1}\left\{\sum_{t=1}^{T} \left\|\boldsymbol{\epsilon}_{it} + \mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{g^0,t} - \boldsymbol{\alpha}_{gt}\right\|^2\right. \\
&\left. - \sum_{t=1}^{T} \left\|\boldsymbol{\epsilon}_{it} + \mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{g_i^0,t} - \boldsymbol{\alpha}_{g_i^0 t}\right\|^2 \leq 0\right\} \\
&= \mathbf{1}\left\{g_i^0 \neq g\right\} \times \mathbf{1}\left\{\sum_{t=1}^{T} (\boldsymbol{\alpha}^2_{gt} - \boldsymbol{\alpha}^2_{g_i^0 t}) + 2\boldsymbol{\epsilon}_{it}\boldsymbol{\alpha}_{g^0 t} + 2\mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta})\boldsymbol{\alpha}_{g^0 t}\right. \\
&\left. + 2\boldsymbol{\alpha}^0_{g_i^0 t}\boldsymbol{\alpha}_{g_i^0 t} - 2\boldsymbol{\epsilon}_{it}\boldsymbol{\alpha}_{gt} - 2\mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta})\boldsymbol{\alpha}_{gt} - 2\boldsymbol{\alpha}^0_{g^0 t}\boldsymbol{\alpha}_{gt} \leq 0\right\} \\
&= \mathbf{1}\left\{g_i^0 \neq g\right\} \times \mathbf{1}\left\{\sum_{t=1}^{T} (\boldsymbol{\alpha}_{gt} - \boldsymbol{\alpha}_{g^0 t})(\boldsymbol{\alpha}_{gt} + \boldsymbol{\alpha}_{g^0 t}) + 2\boldsymbol{\epsilon}_{it}(\boldsymbol{\alpha}_{g^0 t} - \boldsymbol{\alpha}_{gt})\right. \\
&\left. + 2\mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta})(\boldsymbol{\alpha}_{g^0 t} - \boldsymbol{\alpha}_{gt}) + 2\boldsymbol{\alpha}^0_{g_i^0 t}(\boldsymbol{\alpha}_{g^0 t} - \boldsymbol{\alpha}_{gt}) \leq 0\right\} \\
&= \mathbf{1}\left\{g_i^0 \neq g\right\} \times \mathbf{1}\left\{\sum_{t=1}^{T} (\boldsymbol{\alpha}_{g^0 t} - \boldsymbol{\alpha}_{gt}) \left[\boldsymbol{\epsilon}_{it} + \mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{g^0 t} - \frac{(\boldsymbol{\alpha}_{g^0 t} + \boldsymbol{\alpha}_{gt})}{2}\right] \leq 0\right\} \\
&\overset{1}{\leq} \max_{\tilde{g}\neq g} \mathbf{1}\left\{\sum_{t=1}^{T} (\boldsymbol{\alpha}_{\tilde{g}t} - \boldsymbol{\alpha}_{gt}) \left[\boldsymbol{\epsilon}_{it} + \mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{\tilde{g}t} - \frac{(\boldsymbol{\alpha}_{\tilde{g}t} + \boldsymbol{\alpha}_{gt})}{2}\right] \leq 0\right\} \\
&\overset{2}{=} \max_{\tilde{g}\neq g} \mathbf{1}\left\{\sum_{t=1}^{T} (\boldsymbol{\alpha}^0_{\tilde{g}t} - \boldsymbol{\alpha}^0_{gt}) \left[\boldsymbol{\epsilon}_{it} + \boldsymbol{\alpha}^0_{\tilde{g}t} - \frac{(\boldsymbol{\alpha}^0_{\tilde{g}t} + \boldsymbol{\alpha}^0_{gt})}{2}\right]\right. \\
&\leq \left|\sum_{t=1}^{T} (\boldsymbol{\alpha}_{\tilde{g}t} - \boldsymbol{\alpha}_{gt}) \left[\boldsymbol{\epsilon}_{it} + \mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{\tilde{g}_i t} - \frac{(\boldsymbol{\alpha}_{\tilde{g}t} + \boldsymbol{\alpha}_{gt})}{2}\right]\right. \\
&\left.\left. - \sum_{t=1}^{T} (\boldsymbol{\alpha}^0_{\tilde{g}t} - \boldsymbol{\alpha}^0_{gt}) \left[\boldsymbol{\epsilon}_{it} + \boldsymbol{\alpha}^0_{\tilde{g}t} - \frac{(\boldsymbol{\alpha}^0_{\tilde{g}t} + \boldsymbol{\alpha}^0_{gt})}{2}\right]\right|\right\}
\end{aligned} \tag{73}
$$

Where $\overset{1}{\leq}$ *holds as when* $\tilde{g} \neq g$, $\mathbf{1}\left\{g_i^0 \neq g\right\}$ *holds. Then the maximum of the indicator function*

$$
\max_{\tilde{g}\neq g} \mathbf{1}\left\{\sum_{t=1}^{T} (\boldsymbol{\alpha}_{\tilde{g}_i t} - \boldsymbol{\alpha}_{gt}) \left[\boldsymbol{\epsilon}_{it} + \mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{\tilde{g}t} - \frac{(\boldsymbol{\alpha}_{\tilde{g}t} + \boldsymbol{\alpha}_{gt})}{2}\right] \leq 0\right\}
$$

will always greater than the indicator function itself. Where $\overset{2}{=}$ *holds as* $\sum_{t=1}^{T} (\boldsymbol{\alpha}_{\tilde{g}_i t} - \boldsymbol{\alpha}_{gt}) \left[\boldsymbol{\epsilon}_{it} + \mathbf{z}_{it}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{\tilde{g}t} - \frac{(\boldsymbol{\alpha}_{\tilde{g}t}+\boldsymbol{\alpha}_{gt})}{2}\right] \leq 0$, *then then absolute value* *of the difference between*

$$\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}_i t}-\boldsymbol{\alpha}_{gt})\left[\boldsymbol{\epsilon}_{it}+\mathbf{z}_{it}(\boldsymbol{\beta}^0-\boldsymbol{\beta})+\boldsymbol{\alpha}^0_{\tilde{g}t}-\frac{(\boldsymbol{\alpha}_{\tilde{g}t}+\boldsymbol{\alpha}_{gt})}{2}\right]$$

and

$$\sum_{t=1}^{T}(\boldsymbol{\alpha}^0_{\tilde{g}t}-\boldsymbol{\alpha}^0_{gt})\left[\boldsymbol{\epsilon}_{it}+\boldsymbol{\alpha}^0_{\tilde{g}t}-\frac{(\boldsymbol{\alpha}^0_{\tilde{g}t}+\boldsymbol{\alpha}^0_{gt})}{2}\right]$$

will be greater than

$$\sum_{t=1}^{T}(\boldsymbol{\alpha}^0_{\tilde{g}t}-\boldsymbol{\alpha}^0_{gt})\left[\boldsymbol{\epsilon}_{it}+\boldsymbol{\alpha}^0_{\tilde{g}t}-\frac{(\boldsymbol{\alpha}^0_{\tilde{g}t}+\boldsymbol{\alpha}^0_{gt})}{2}\right]$$

Now we let:

$$\begin{aligned}\mathbb{A}_T=&\left|\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}-\boldsymbol{\alpha}_{gt})\left[\boldsymbol{\epsilon}_{it}+\mathbf{z}_{it}(\boldsymbol{\beta}^0-\boldsymbol{\beta})+\boldsymbol{\alpha}^0_{\tilde{g}_i t}-\frac{(\boldsymbol{\alpha}_{\tilde{g}t}+\boldsymbol{\alpha}_{gt})}{2}\right]\right.\\&\left.-\sum_{t=1}^{T}(\boldsymbol{\alpha}^0_{\tilde{g}t}-\boldsymbol{\alpha}^0_{gt})\left[\boldsymbol{\epsilon}_{it}+\boldsymbol{\alpha}^0_{\tilde{g}t}-\frac{(\boldsymbol{\alpha}^0_{\tilde{g}t}+\boldsymbol{\alpha}^0_{gt})}{2}\right]\right|\\&\overset{1}{\leq}\underbrace{\left|\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}-\boldsymbol{\alpha}_{gt})\boldsymbol{\epsilon}_{it}-\sum_{t=1}^{T}(\boldsymbol{\alpha}^0_{\tilde{g}t}-\boldsymbol{\alpha}^0_{gt})\boldsymbol{\epsilon}_{it}\right|}_{a.1}\\&+\underbrace{\left|\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}-\boldsymbol{\alpha}_{gt})\mathbf{z}_{it}(\boldsymbol{\beta}^0-\boldsymbol{\beta})\right|}_{a.2}\\&+\underbrace{\left|\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}-\boldsymbol{\alpha}_{gt})(\boldsymbol{\alpha}^0_{\tilde{g}t}-\frac{(\boldsymbol{\alpha}_{\tilde{g}t}+\boldsymbol{\alpha}_{gt})}{2})-\sum_{t=1}^{T}(\boldsymbol{\alpha}^0_{\tilde{g}t}-\boldsymbol{\alpha}^0_{gt})(\boldsymbol{\alpha}^0_{\tilde{g}t}-\frac{(\boldsymbol{\alpha}^0_{\tilde{g}t}+\boldsymbol{\alpha}^0_{gt})}{2})\right|}_{a.3}\end{aligned} \tag{74}$$

Where $\overset{1}{\leq}$ *holds by triangle inequality. By Cauchy-Schwarz inequality, we have a.1:*

$$
\begin{aligned}
\left|\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}-\boldsymbol{\alpha}_{\tilde{g}t}^{0})\boldsymbol{\epsilon}_{it}-\sum_{t=1}^{T}(\boldsymbol{\alpha}_{gt}^{0}-\boldsymbol{\alpha}_{gt})\boldsymbol{\epsilon}_{it}\right| &\overset{1}{\leq}\left(\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^{2}\sum_{t=1}^{T}\left\|(\boldsymbol{\alpha}_{\tilde{g}t}-\boldsymbol{\alpha}_{\tilde{g}t}^{0})\right\|^{2}\right)^{1/2}\\
&+\left(\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^{2}\sum_{t=1}^{T}\left\|(\boldsymbol{\alpha}_{gt}-\boldsymbol{\alpha}_{gt}^{0})\right\|^{2}\right)^{1/2}\\
&\leq MT\left(\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^{2}\frac{1}{MT}\sum_{t=1}^{T}\left\|(\boldsymbol{\alpha}_{\tilde{g}t}-\boldsymbol{\alpha}_{\tilde{g}t}^{0})\right\|^{2}\right)^{1/2}\\
&+MT\left(\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^{2}\frac{1}{MT}\sum_{t=1}^{T}\left\|(\boldsymbol{\alpha}_{gt}-\boldsymbol{\alpha}_{gt}^{0})\right\|^{2}\right)^{1/2}\\
&\overset{2}{=}2MT\sqrt{\eta}\left(\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^{2}\right)^{1/2}
\end{aligned}
\tag{75}
$$

Where $\overset{1}{\leq}$ *is by Cauchy-Schwarz inequality.* $\overset{2}{=}$ *is by the definition of equation 70 and Assumption 6 (c). Similarly, we have a.2:*

$$
\left|\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}-\boldsymbol{\alpha}_{gt})\mathbf{z}_{it}(\boldsymbol{\beta}^{0}-\boldsymbol{\beta})\right|\overset{1}{\leq}MT\left(\frac{1}{MT}\sum_{t=1}^{T}\|\mathbf{z}_{it}\|^{2}\frac{1}{MT}\sum_{t=1}^{T}\left\|(\boldsymbol{\alpha}_{gt}-\boldsymbol{\alpha}_{gt}^{0})(\boldsymbol{\beta}^{0}-\boldsymbol{\beta})\right\|^{2}\right)^{1/2}
\tag{76}
$$

Where $\overset{1}{\leq}$ *and* $\overset{2}{\leq}$ *are by Cauchy-Schwarz inequality. Therefore, we have:*

$$
\mathbb{A}_{T}=MTC_{1}\sqrt{\eta}\left(\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^{2}\right)^{1/2}+MTC_{2}\sqrt{\eta}\left(\frac{1}{MT}\sum_{t=1}^{T}\|\mathbf{z}_{it}\|^{2}\right)^{1/2}+MTC_{3}\sqrt{\eta}
\tag{77}
$$

Where C_1, C_2 and C_3 are positive constants that are independent with η and T.

Therefore, we have

$$
\begin{aligned}
Z_{ig}(\boldsymbol{\beta},\boldsymbol{\alpha}) &\leq \max_{\tilde{g}\neq g}\mathbf{1}\left\{\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0})\left[\boldsymbol{\epsilon}_{it}+\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\frac{(\boldsymbol{\alpha}_{\tilde{g}t}^{0}+\boldsymbol{\alpha}_{gt}^{0})}{2}\right]\leq\mathbb{A}_{T}\right\}\\
&=\max_{\tilde{g}\neq g}\mathbf{1}\left\{\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0})\boldsymbol{\epsilon}_{it}+\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0})\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\frac{1}{2}\sum_{t=1}^{T}(\left\|\boldsymbol{\alpha}_{\tilde{g}t}^{0}\right\|^{2}-\left\|\boldsymbol{\alpha}_{gt}^{0}\right\|^{2})\leq\mathbb{A}_{T}\right\}\\
&=\max_{\tilde{g}\neq g}\mathbf{1}\left\{\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0})\boldsymbol{\epsilon}_{it}+\frac{1}{2}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0}\right\|^{2}\leq\mathbb{A}_{T}\right\}\\
&\equiv\tilde{Z}_{ig}
\end{aligned}
\tag{78}
$$

Where the right side of the inequality in $\tilde{Z}_{ig}$ does not dependent on $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. Thus we have:

$$\sup_{(\boldsymbol{\beta},\boldsymbol{\alpha})\in\mathcal{N}_\eta} Z_{ig}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \leq \tilde{Z}_{ig} \tag{79}$$

Thus

$$\sup_{(\boldsymbol{\beta},\boldsymbol{\alpha})\in\mathcal{N}_\eta} \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left\{\widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \neq g_i^0\right\} \leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N} Z_{ig}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N} \tilde{Z}_{ig} \tag{80}$$

We now show that $Pr(\tilde{Z}_{ig} = 1)$ is bounded. Denote $\tilde{M} > \max(M, M^*)$ Where M and M^* are some constant defined in Assumption 6 (a) and 7 (e). Note that by Assumption 6 (c) and Cauchy-Schwarz inequality, we also have:

$$\begin{aligned} E(\|\boldsymbol{\epsilon}_{it}\|^2) &= \sqrt{E(\|\boldsymbol{\epsilon}_{it}\|^2 \times 1^2)^2} \\ &\leq \sqrt{E(\|\boldsymbol{\epsilon}_{it}\|^4)} \\ &\leq \sqrt{M} \end{aligned} \tag{81}$$

By Boole's inequality we have:

$$\begin{aligned} \Pr\left(\tilde{Z}_{ig} = 1\right) \leq &\sum_{\tilde{g}\neq g} \Pr\left(\sum_{t=1}^{T}(\boldsymbol{\alpha}_{\tilde{g}t}^0 - \boldsymbol{\alpha}_{gt}^0)\boldsymbol{\epsilon}_{it} \leq -\frac{1}{2}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{\tilde{g}t}^0 - \boldsymbol{\alpha}_{gt}^0\right\|^2\right. \\ &+ MTC_1\sqrt{\eta}\left(\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^2\right)^{1/2} \\ &\left.+MTC_2\sqrt{\eta}\left(\frac{1}{MT}\sum_{t=1}^{T}\|\mathbf{z}_{it}\|^2\right)^{1/2} + MTC_3\sqrt{\eta}\right) \\ &\leq \sum_{\tilde{g}\neq g}\left[\underbrace{\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\|\mathbf{z}_{it}\|^2 \geq \tilde{M}\right)}_{a.1}\right. \\ &+\underbrace{\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{\tilde{g}t}^0 - \boldsymbol{\alpha}_{gt}^0\right\|^2 \leq \frac{c_{g,\tilde{g}}}{2}\right)}_{a.2} + \underbrace{\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^2 \geq \tilde{M}\right)}_{a.3} \\ &\left.+\underbrace{\Pr\left(\sum_{t=1}^{T}\left(\boldsymbol{\alpha}_{\tilde{g}t}^0 - \boldsymbol{\alpha}_{gt}^0\right)\boldsymbol{\epsilon}_{it} \leq -T\frac{c_{g,\tilde{g}}}{4} + TC_1\sqrt{\eta}\sqrt{\tilde{M}} + TC_2\sqrt{\eta}\tilde{M} + TC_3\sqrt{\eta}\right)}_{a.4}\right] \end{aligned} \tag{82}$$

Where a.1 *is directly indicated by Assumption 7 (e).* a.2 *a direct proof by Rio (1999), Bonhomme & Manresa (2015). By Assumption 7 (b), for* $g \neq \tilde{g}$ *we have:*

$$\text{plim}_{T\to\infty} \frac{1}{MT}\sum_{t=1}^{T} \left\|\boldsymbol{\alpha}_{gt}^0 - \boldsymbol{\alpha}_{\tilde{g}t}^0\right\|^2 = c_{g,\tilde{g}}, \quad \infty > c_{g,\tilde{g}} > 0$$

Which means that when T *is large enough, we have: have:*

$$\text{plim}_{T\to\infty} \frac{1}{MT}\sum_{t=1}^{T} \left\|\boldsymbol{\alpha}_{gt}^0 - \boldsymbol{\alpha}_{\tilde{g}t}^0\right\|^2 \geq \frac{2}{3}c_{g,\tilde{g}}$$

Set the strongly mixing process $z_t = \left\|\boldsymbol{\alpha}_{gt}^0 - \boldsymbol{\alpha}_{\tilde{g}t}^0\right\|^2 - \mathbb{E}\left[\left\|\boldsymbol{\alpha}_{gt}^0 - \boldsymbol{\alpha}_{\tilde{g}t}^0\right\|^2\right]$ *and* $z = \frac{1}{6}c_{g,\tilde{g}}$. *By Assumption 7 (c), for all* $\delta > 0$ *and* T *goes to infinity, we have:*

$$\begin{aligned}
\Pr\left(\left|\frac{1}{MT}\sum_{t=1}^{T} z_t\right| \geq z\right) &= \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} z_t \geq z\right) + \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} z_t \leq -z\right) \\
&= \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} z_t \geq z\right) + \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} \left\|\boldsymbol{\alpha}_{gt}^0 - \boldsymbol{\alpha}_{\tilde{g}t}^0\right\|^2 \leq \frac{1}{2}c_{g,\tilde{g}}\right) \\
&= o_p(T^{-\delta})
\end{aligned} \tag{83}$$

Thus we have a.2*:* $\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{gt}^0 - \boldsymbol{\alpha}_{\tilde{g}t}^0\right\|^2 = \frac{1}{2}c_{g,\tilde{g}}\right) \leq o_p(T^{-\delta})$. *Similarly,* a.3 *set the strongly mixing process* $z_t = \left\|\boldsymbol{\epsilon}_{it}\right\|^2 - E\left[\left\|\boldsymbol{\epsilon}_{it}\right\|^2\right]$ *and* $z = \tilde{M} - \sqrt{M}$. *Under Assumption7 (d), for all* $\delta > 0$ *and* T *goes to infinity, we have:*

$$\begin{aligned}
\Pr\left(\left|\frac{1}{MT}\sum_{t=1}^{T} z_t\right| \geq z\right) &= \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} z_t \geq z\right) + \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} z_t \leq -z\right) \\
&= \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} \left\|\boldsymbol{\epsilon}_{it}\right\|^2 \geq \tilde{M}\right) + \Pr\left(\frac{1}{MT}\sum_{t=1}^{T} z_t \leq -z\right) \\
&= o_p(T^{-\delta})
\end{aligned} \tag{84}$$

Thus we have a.3*:* $\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{it}\right\|^2 \geq \tilde{M}\right) \leq o_p(T^{-\delta})$. a.4 *is the directly result by* a.1*,* a.2 *and* a.3*.*

Take η *such that*

$$\eta \leq \left(\frac{c}{8\left(C_1\sqrt{\tilde{M}} + C_2\tilde{M} + C_3\right)}\right)^2 \tag{85}$$

Where c is the minimum $c_{g,\tilde{g}}$ across groups, for all $g \neq \tilde{g}$, we have:

$$\begin{aligned}
&\Pr\left(\sum_{t=1}^{T}\left(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0}\right)\boldsymbol{\epsilon}_{it} \leq -T\frac{c_{g,\tilde{g}}}{4}+TC_1\sqrt{\eta}\sqrt{\tilde{M}}+TC_2\sqrt{\eta}\tilde{M}+TC_3\sqrt{\eta}\right)\\
&\leq \Pr\left(\sum_{t=1}^{T}\frac{1}{T}\left(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0}\right)\boldsymbol{\epsilon}_{it} \leq -\frac{c_{g,\tilde{g}}}{8}\right)\\
&\leq \Pr\left(\left|\sum_{t=1}^{T}\frac{1}{T}\left(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0}\right)\boldsymbol{\epsilon}_{it}\right| \geq \frac{c_{g,\tilde{g}}}{8}\right)\\
&= o_p(T^{-\delta})
\end{aligned} \tag{86}$$

Where $\left(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0}\right)\boldsymbol{\epsilon}_{it}$ is also a strongly mixing process and can apply the exponential inequalities by Rio (1999), Bonhomme & Manresa (2015) Thus we have the bounded $\Pr\left(\tilde{Z}_{ig}=1\right)$:

$$\begin{aligned}
\Pr\left(\tilde{Z}_{ig}=1\right) \leq \sum_{\tilde{g}\neq g}\Bigg[&\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\|\mathbf{z}_{it}\|^2 \geq \tilde{M}\right)\\
&+\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0}\right\|^2 \leq \frac{c_{g,\tilde{g}}}{2}\right)+\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\|^2 \geq \tilde{M}\right)\\
&\Pr\left(\sum_{t=1}^{T}\left(\boldsymbol{\alpha}_{\tilde{g}t}^{0}-\boldsymbol{\alpha}_{gt}^{0}\right)\boldsymbol{\epsilon}_{it} \leq -T\frac{c_{g,\tilde{g}}}{4}+TC_1\sqrt{\eta}\sqrt{\tilde{M}}+TC_2\sqrt{\eta}\tilde{M}+TC_3\sqrt{\eta}\right)\Bigg]\\
&\leq \sum_{\tilde{g}\neq g} o_p(T^{-\delta})\\
&= (G-1)o_p(T^{-\delta})
\end{aligned} \tag{87}$$

Which indicates that

$$\begin{aligned}
E\left(\sup_{\boldsymbol{\beta},\boldsymbol{\alpha}\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha})\neq g_i^0\right\}\right) &\leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}E\left(\tilde{Z}_{ig}\right)\\
&= \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}\Pr\left(\tilde{Z}_{ig}=1\right)\\
&= G(G-1)o\left(T^{-\delta}\right)=o\left(T^{-\delta}\right)
\end{aligned} \tag{88}$$

By Markov inequality we have:

$$\begin{aligned}
\Pr(\sup_{\boldsymbol{\beta},\boldsymbol{\alpha}\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha})\neq g_i^0\right\} > \epsilon T^{-\delta}) &\leq \frac{E\left(\sup_{\boldsymbol{\beta},\boldsymbol{\alpha}\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha})\neq g_i^0\right\}\right)}{\epsilon T^{-\delta}}\\
&= o_p(1)
\end{aligned} \tag{89}$$

Which ends the proof of Lemma 4.

Now denote

$$\hat{Q}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \hat{\tau}) = \hat{Q}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{\hat{g}_i,t}\|^2 \tag{90}$$

and

$$\tilde{Q}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau^0) = \tilde{Q}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i^0,t}\right\|^2 \tag{91}$$

Under Lemma 4, we can show the following Lemma:

Lemma 5

$$\underset{N,T\to\infty}{\text{plim}} \sup_{(\boldsymbol{\beta},\boldsymbol{\alpha})\in\mathcal{N}_n} \left|\hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})\right| = o_p(T^{-\delta}) \tag{92}$$

Proof 6 (Proof of Lemma 5) *It is easy to show that for $(\boldsymbol{\beta}, \boldsymbol{\alpha}) \in \mathcal{N}_n$ and N T go to infinity:*

$$\begin{aligned}
\mathbb{Q} &= \hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \\
&= \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \neq g_i^0\right\} \left(\|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{\hat{g}_i,t}\|^2 - \left\|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i^0,t}\right\|^2\right) \\
&\overset{1}{=} o_p(T^{-\delta})
\end{aligned} \tag{93}$$

Where $\overset{1}{=}$ is indicated by Lemma 4

Hence, by the definition of $\mathcal{N}_n$, the consistent estimator $\hat{\boldsymbol{\beta}}$ under Theorem 5 and the consistent estimator $\hat{\boldsymbol{\alpha}}$ under Theorem 5 in objective function $\hat{\mathcal{Q}}$, we have:

$$\Pr\left((\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) \notin \mathcal{N}_\eta\right) \to 0 \quad (N, T) \to \infty \tag{94}$$

Similarly, by the definition of least square estimator, $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\alpha}}$ are still consistent estimators under Theorem 5 and Theorem 5 in objective function $\tilde{\mathcal{Q}}$. Thus:

$$\Pr\left((\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \notin \mathcal{N}_\eta\right) \to 0 \quad (N, T) \to \infty \tag{95}$$

By basic probability algebra, we have the following inequality:

$$\begin{aligned}
&\Pr\left[\left|\hat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})\right| > \varepsilon T^{-\delta}\right] \\
&\leq \underbrace{\Pr\left((\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) \notin \mathcal{N}_\eta\right)}_{a.1:\to 0} + \underbrace{\Pr\left[\sup_{(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}})\in\mathcal{N}_n} \left|\hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})\right| > \varepsilon T^{-\delta}\right]}_{a.2:o_p(1)} \\
&= o_p(1)
\end{aligned} \tag{96}$$

Thus we have:

$$\hat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = o_p(T^{-\delta}) \tag{97}$$

Similarly, we have

$$\begin{aligned} &\Pr\left[\left|\hat{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})\right| > \varepsilon T^{-\delta}\right] \\ &\quad \leq \underbrace{\Pr\left((\tilde{\boldsymbol{\beta}}, \tilde{\alpha}) \notin \mathcal{N}_\eta\right)}_{a.1:\to 0} + \underbrace{\Pr\left[\sup_{(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \in \mathcal{N}_n} \left|\tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})\right| > \varepsilon T^{-\delta}\right]}_{a.2: o_p(1)} \\ &\quad = o_p(1) \end{aligned} \tag{98}$$

Thus we have:

$$\hat{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) = o_p(T^{-\delta}) \tag{99}$$

By the definition of least square estimators, equation 163 and equation 165, we have the following inequality:

$$0 \leq \underbrace{\widetilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \widetilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})}_{a.1} = \underbrace{\widehat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \widehat{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})}_{a.2} + o_p\left(T^{-\delta}\right) \leq o_p\left(T^{-\delta}\right) \tag{100}$$

Where $a.1 \geq 0$ as $\widetilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})$ minimised the objective function $\widetilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ by least square definition and $a.2 \leq 0$ as $\widehat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$ minimised the objective function $\hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ by least square definition. Thus, for all $\delta > 0$, we have:

$$\widetilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \widetilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) = o_p\left(T^{-\delta}\right) \tag{101}$$

Next we show that $\tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is uniquely minimised at $\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}$ by the following lemma.

Lemma 6 *For $\hat{\boldsymbol{\beta}} \neq \tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}} \neq \tilde{\boldsymbol{\alpha}}$*

$$\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \geq \rho \left\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}\right\|^2 \tag{102}$$

$$
\begin{aligned}
\mathbb{Q} &= \tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \\
&= \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\alpha}}_{g_i^0,t} \right\|^2 - \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t} \tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\alpha}}_{g_i^0,t} \right\|^2 \\
&= \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t} \tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\alpha}}_{g_i^0,t} + \mathbf{A} \right\|^2 - \left\| \boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t} \tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\alpha}}_{g_i^0,t} \right\|^2 \right] \\
&= \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left\| \boldsymbol{z}_{i,t}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\alpha}}_{g_{i,t}^0} - \hat{\boldsymbol{\alpha}}_{g_{i,t}^0}) \right\|^2 \right. \\
&\left. + 2(\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t} \tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\alpha}}_{g_i^0,t})(\boldsymbol{z}_{i,t}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\alpha}}_{g_{i,t}^0} - \hat{\boldsymbol{\alpha}}_{g_{i,t}^0})) \right] \\
&\overset{1}{=} \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left\| \boldsymbol{z}_{i,t}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\alpha}}_{g_{i,t}^0} - \hat{\boldsymbol{\alpha}}_{g_{i,t}^0}) \right\|^2 \right] \\
&= \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left\| (\boldsymbol{z}_{i,t} - \boldsymbol{z}_{g \wedge \tilde{g},t})(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + \mathbf{B} \right\|^2 \right] \\
&\geq \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left\| (\boldsymbol{z}_{i,t} - \boldsymbol{z}_{g \wedge \tilde{g},t})(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) \right\|^2 \right] \\
&\overset{1}{=} (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})' \Sigma(\gamma) (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) \\
&\geq (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})' \min_{\gamma \in \boldsymbol{\Gamma}_G} \Sigma(\gamma) (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) \\
&\overset{2}{\geq} \min_{\gamma \in \boldsymbol{\Gamma}_G} \hat{\rho}(\gamma) \left\| \tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} \right\|^2
\end{aligned}
\tag{103}
$$

For simplicity, we denote $\mathbf{A} = \left(\boldsymbol{z}_{i,t}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\alpha}}_{g_{i,t}^0} - \hat{\boldsymbol{\alpha}}_{g_{i,t}^0}) \right)$ *and* $\mathbf{B} = (\boldsymbol{z}_{g \wedge \tilde{g},t})(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\alpha}}_{g_{i,t}^0} - \hat{\boldsymbol{\alpha}}_{g_{i,t}^0})$*. Where* $\overset{1}{=}$ *holds by the least square estimators' properties,* $\overset{2}{\geq}$ *holds by Assumption 6 (g). The intuition behind the equation is that the difference of objection function values is bounded away from zero when parameters are not equal to* $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})$*.*

Proof 7 (Proof of Theorem 6) *By Boole's inequality, we have:*

$$
\begin{aligned}
\Pr\left(\sup_{i\in\{1,\dots,N\}}\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right) &= \Pr\left(\bigcup_{i\in\{1,\dots,N\}}\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right)\\
&\leq \sum_{i=1}^{N}\Pr\left(\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right)\\
&\leq N\sup_{i\in\{1,\dots,N\}}\Pr\left(\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right)\\
&\leq \underbrace{\Pr\left((\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\notin\mathcal{N}_\eta\right)}_{a.1}\\
&\quad \underbrace{+N\sup_{i\in\{1,\dots,N\}}\Pr\left((\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\in\mathcal{N}_\eta,\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\neq g_i^0\right)}_{a.2}
\end{aligned}
\tag{104}
$$

$a.1$ *towards to zero by Equation 160, for* $a.2$*:*

$$
\begin{aligned}
N\sup_{i\in\{1,\dots,N\}}\Pr\left((\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\in\mathcal{N}_\eta,\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\neq g_i^0\right) &\overset{1}{=} N\sup_{i\in\{1,\dots,N\}}\mathrm{E}\left(\mathbf{1}\left\{(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\in\mathcal{N}_\eta\right\}\mathbf{1}\left\{\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\neq g_i^0\right\}\right)\\
&\overset{2}{\leq} N\sup_{i\in\{1,\dots,N\}}\mathrm{E}\left(\mathbf{1}\left\{(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\in\mathcal{N}_\eta\right\}\sum_{g=1}^{G}\tilde{Z}_{ig}\right)\\
&\leq N\sup_{i\in\{1,\dots,N\}}\mathrm{E}\left(\sum_{g=1}^{G}\tilde{Z}_{ig}\right)\\
&= N\sup_{i\in\{1,\dots,N\}}\sum_{g=1}^{G}\Pr\left(\tilde{Z}_{ig}=1\right)\\
&= NG(G-1)o_p(T^{-\delta})\\
&= o_p(NT^{-\delta})
\end{aligned}
\tag{105}
$$

Where $\overset{1}{=}$ *is the definition of indicator function,* $\overset{2}{\leq}$ *comes from Lemma 4. When* $NT^{-\delta}$ *tends to zero, Equation 174 is* $o_p(1)$*. Thus we end the proof of the theroem.*

Proof 8 (Proof of Theorem 8) *It is easy to show that under the Lemma 5,* *Lemma 6 and Theorem 7, we have:*

$$\begin{aligned}\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\hat{\boldsymbol{\alpha}}_{g_i^0,t}-\tilde{\boldsymbol{\alpha}}_{g_i^0,t}\right\|^2 &= \frac{1}{MNT}\sum_{g=1}^{G}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\left\{g_i^0=g\right\}\left\|\hat{\boldsymbol{\alpha}}_{g^0t}-\tilde{\boldsymbol{\alpha}}_{g^0t}\right\|^2\\ &\overset{1}{=}\sum_{g=1}^{G}\pi_g\frac{1}{MT}\sum_{t=1}^{T}\mathbf{1}\left\{g_i^0=g\right\}\left\|\hat{\boldsymbol{\alpha}}_{g^0t}-\tilde{\boldsymbol{\alpha}}_{g^0t}\right\|^2\\ &= o_p(T^{-\delta})\end{aligned} \tag{106}$$

Where $\overset{1}{=}$ comes from Assumption 7 (b), Thus, for all g:

$$\frac{1}{MT}\sum_{t=1}^{T}\left\|\hat{\boldsymbol{\alpha}}_{g^0t}-\tilde{\boldsymbol{\alpha}}_{g^0t}\right\|^2 = o_p(T^{-\delta}) \tag{107}$$

Which implies that, for all t, we have for all $\delta>0$:

$$\frac{1}{M}\left\|\hat{\boldsymbol{\alpha}}_{g^0t}-\tilde{\boldsymbol{\alpha}}_{g^0t}\right\|^2 \leq o_p(T^{1-\delta}) \tag{108}$$

Proof 9 (Proof of Corollary 2) *Under the known group, we have the following algebraic expression for $\tilde{\boldsymbol{\alpha}}$:*

$$\begin{aligned}\widetilde{\boldsymbol{\alpha}}_{gt} &= \bar{\mathbf{y}}_{g_i^0t}-\bar{\mathbf{z}}'_{g_i^0t}\tilde{\boldsymbol{\beta}}\\ &= \frac{\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0=g\right\}\left(\mathbf{y}_{it}-\mathbf{z}'_{it}\tilde{\boldsymbol{\beta}}\right)}{\sum_{i=1}^{N}\mathbf{1}\left\{g_i^0=g\right\}}\end{aligned} \tag{109}$$

and $\tilde{\boldsymbol{\beta}}$:

$$\tilde{\boldsymbol{\beta}} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)'\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)\left(\mathbf{y}_{it}-\bar{\mathbf{y}}_{g_i^0t}\right)\right) \tag{110}$$

By using equation 175 we can easily show that:

$$\begin{aligned}\sqrt{NT}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right) &= \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)'\right)^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}-\bar{\mathbf{z}}_{g_i^0t}\right)\boldsymbol{\epsilon}_{it}\right)\\ &\sim N(0,\Sigma_\beta^{-1}\Omega_\beta\Sigma_\beta^{-1})\end{aligned} \tag{111}$$

Since the infeasible estimatoe $\widetilde{\boldsymbol{\beta}}$ is asymptotic equivalent to $\hat{\boldsymbol{\beta}}$, we have:

$$\sqrt{NT}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right)\sim N(0,\Sigma_\beta^{-1}\Omega_\beta\Sigma_\beta^{-1}) \tag{112}$$

Proof 10 (Proof of Corollary 3) *From equation 175 we have:*

$$\begin{aligned}\widetilde{\boldsymbol{\alpha}}_{gt} &= \frac{\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}\left(\mathbf{y}_{it} - \mathbf{z}_{it}'\tilde{\boldsymbol{\beta}}\right)}{\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}} \\ &= \boldsymbol{\alpha}_{g^0 t}^0 + \underbrace{\frac{\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}\mathbf{z}_{it}'}{\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}}\left(\boldsymbol{\beta}^0 - \tilde{\boldsymbol{\beta}}\right)}_{a.1} + \frac{\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}\boldsymbol{\epsilon}_{it}}{\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}} \\ \sqrt{N}\left(\widetilde{\boldsymbol{\alpha}}_{gt} - \boldsymbol{\alpha}_{g^0 t}^0\right) &= O_p\left(\frac{1}{\sqrt{NT}}\right) + \underbrace{\frac{\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}\boldsymbol{\epsilon}_{it}}{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}}}_{a.2}\end{aligned} \tag{113}$$

Where term a.1 is bounded by $O_p\left(\frac{1}{\sqrt{NT}}\right)$ *by Assumption 6 and Assumption 7. Term a.2 follows normal distribution by Assumption 8. Thus we have:*

$$\sqrt{N}\left(\widehat{\boldsymbol{\alpha}}_{gt} - \boldsymbol{\alpha}_{g^0 t}^0\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega_{gt}}{\pi_g^2}\right) \tag{114}$$

3.10.2 Proofs of the Heterogeneous Coefficient Model

We first define the objective function as:

$$\begin{aligned}\hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) &= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta}_{g_i} - \boldsymbol{\alpha}_{g_i,t}\right\|^2 \\ &= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t} + \boldsymbol{z}_{i,t}(\boldsymbol{\beta}_{g_i}^0 - \boldsymbol{\beta}_{g_i}) + \boldsymbol{\alpha}_{g_i,t}^0 - \boldsymbol{\alpha}_{g_i,t}\right\|^2\end{aligned} \tag{115}$$

Lemma 7

$$\underset{N,T\to\infty}{\mathrm{plim}} \sup_{(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\in\boldsymbol{\Omega}^G\times\boldsymbol{\Theta}^{GTM}\times\boldsymbol{\Gamma}_G}\left\|\hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)\right\| = o_p(1) \tag{116}$$

Proof 11 (Proof of Lemma 7) *Define the following auxiliary objective function as:*

$$\begin{aligned}\tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = &\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{z}_{i,t}(\boldsymbol{\beta}_{g_i}^0 - \boldsymbol{\beta}_{g_i}) + \boldsymbol{\alpha}_{g_i,t}^0 - \boldsymbol{\alpha}_{g_i,t}\right\|^2 \\ &+ \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t}\right\|^2\end{aligned} \tag{117}$$

Then we have:

$$
\begin{aligned}
\mathbb{Q} &= \hat{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau) - \tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\\
&= \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t} + \boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0_{g_i} - \boldsymbol{\beta}_{g_i}) + \boldsymbol{\alpha}^0_{g_i,t} - \boldsymbol{\alpha}_{g_i,t}\right\|^2\\
&- \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}) + \boldsymbol{\alpha}^0_{g_i,t} - \boldsymbol{\alpha}_{g_i,t}\right\|^2\\
&- \frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t}\right\|^2\\
&= \frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\boldsymbol{\epsilon}_{i,t}\left(\boldsymbol{z}_{i,t}(\boldsymbol{\beta}^0_{g_i} - \boldsymbol{\beta}_{g_i}) + \boldsymbol{\alpha}^0_{g_i,t} - \boldsymbol{\alpha}_{g_i,t}\right)'\right]\\
&= (\boldsymbol{\beta}^0_{g_i} - \boldsymbol{\beta}_{g_i})\frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t} + \frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}^{0\prime}_{g_i,t}\boldsymbol{\epsilon}_{i,t} - \frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}'_{g_i,t}\boldsymbol{\epsilon}_{i,t}\\
&= \frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}^0_{g_i} - \frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}_{g_i} + \frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}^{0\prime}_{g_i,t}\boldsymbol{\epsilon}_{i,t}\\
&- \frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}'_{g_i,t}\boldsymbol{\epsilon}_{i,t}
\end{aligned}
\tag{118}
$$

For term $\frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}^0_{g_i}$*, it can be viewed as:*

$$
\frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}^0_{g_i} = \frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\begin{bmatrix}\boldsymbol{\epsilon}_{i,t}\otimes\mathcal{Y}_{i,t-1}\\ \boldsymbol{\epsilon}_{i,t}\otimes x_{i,t}\end{bmatrix}
\tag{119}
$$

From Assumption 6 (c), (d) and (e) we have the following Cauchy-Schwarz (CS) inequality:

$$
\begin{aligned}
\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}_{g_i} &= \sum_{g=1}^{G}\frac{1}{MN}\sum_{i=1}^{N}\mathbf{1}\{g_i = g\}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}_{g_i}\\
&= \sum_{g=1}^{G}\frac{1}{MN}\boldsymbol{\beta}_g\sum_{i=1}^{N}\mathbf{1}\{g_i = g\}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}
\end{aligned}
\tag{120}
$$

For $\frac{1}{MN}\boldsymbol{\beta}_g\sum_{i=1}^{N}\mathbf{1}\{g_i = g\}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}$ *we have:*

$$
\begin{aligned}
\frac{1}{MN}\boldsymbol{\beta}_g\sum_{i=1}^{N}\mathbf{1}\{g_i = g\}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t} &\le \frac{1}{MN}\left(\sum_{i=1}^{N}\|\boldsymbol{\beta}_g\|^2\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\{g_i = g\}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}\right\|^2\right)^{1/2}\\
&= \frac{1}{M}\left(\|\boldsymbol{\beta}_g\|^2\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\{g_i = g\}\boldsymbol{z}'_{i,t}\boldsymbol{\epsilon}_{i,t}\right\|^2\right)^{1/2}
\end{aligned}
\tag{121}
$$

The term $\|\boldsymbol{\beta}_g\|^2$ *is bounded by Assumption 6 (a), the second term is bounded by* $O_p(T^{-1/2})$*by Assumption 6 (g):*

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\{g_i=g\}\boldsymbol{z}_{i,t}'\boldsymbol{\epsilon}_{i,t}\right\|^2 &= \frac{1}{T^2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}\{g_i=g\}\mathbf{1}\{g_j=g\}\sum_{t=1}^{T}\boldsymbol{z}_{i,t}'\boldsymbol{z}_{j,t}\boldsymbol{\epsilon}_{i,t}'\boldsymbol{\epsilon}_{j,t}\\ &\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{1}{T}\mathbf{z}_{i,t}'\boldsymbol{\epsilon}_{i,t}\right\|^2\\ &\leq O_p(T^{-1/2})\end{aligned} \tag{122}$$

The last inequality from Assumption 6 (g) that:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left\|\mathbf{z}_{i,t}'\boldsymbol{\epsilon}_{i,t}\right\|^2=\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{1}{T}\mathbf{z}_{i,t}'\boldsymbol{\epsilon}_{i,t}\right\|^2\right)\leq\frac{\mathbb{M}}{T} \tag{123}$$

Therefore, we conclude that:

$$\frac{2}{TNM}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{i,t}'\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}_{g_i}=O_p\left(\frac{1}{\sqrt{T}}\right)=o_p(1) \tag{124}$$

The term $\frac{2}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{i,t}'\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}_{g_i}^0$ *is also* $o_p(1)$ *by* $\frac{2}{TNM}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{i,t}'\boldsymbol{\epsilon}_{i,t}\boldsymbol{\beta}_{g_i}=o_p(1)$.

Next we show that $\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}_{g_i,t}'\boldsymbol{\epsilon}_{i,t}$ *is uniformly* $o_p(1)$

$$\begin{aligned}\frac{1}{NTM}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}_{g_i,t}'\boldsymbol{\epsilon}_{i,t} &= \frac{1}{M}\sum_{g=1}^{G}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{g_i=g\}\boldsymbol{\alpha}_{g,t}'\boldsymbol{\epsilon}_{i,t}\right]\\ &= \frac{1}{M}\sum_{g=1}^{G}\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\alpha}_{g,t}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\boldsymbol{\epsilon}_{i,t}\right)\right]\end{aligned} \tag{125}$$

Where by Cauchy-Schwarz inequality:

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\alpha}_{g,t}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\boldsymbol{\epsilon}_{i,t}\right)\right\|^2\leq\left(\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{\alpha}_{g,t}\|^2\right)\times\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\boldsymbol{\epsilon}_{i,t}\right\|^2\right) \tag{126}$$

Where $\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{\alpha}_{g,t}\|^2$ *is uniformly bounded by Assumption a, we next show that the* *second term is uniformly bounded.*

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\boldsymbol{\epsilon}_{i,t}\right\|^2 \\
&=\frac{1}{TN^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}\{g_i=g\}\mathbf{1}\{g_j=g\}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{i,t}'\boldsymbol{\epsilon}_{j,t} \\
&\leq\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{i,t}'\boldsymbol{\epsilon}_{j,t}\right| \\
&=O_p(N^{-1/2}T^{-1/2})
\end{aligned} \tag{127}$$

The last equality holds by Cauchy-Schwarz inequality:

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{i,t}'\boldsymbol{\epsilon}_{j,t}\right|\right\|^2 \\
&\leq\mathbb{E}\left(\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{\epsilon}_{i,t}'\boldsymbol{\epsilon}_{j,t}\right\|^2\right) \\
&=\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\left\|\boldsymbol{\epsilon}_{i,t}'\boldsymbol{\epsilon}_{j,t}\right\|^2\right) \\
&=\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\|\boldsymbol{\epsilon}_{i,t}\|^4\right) \\
&=O_p(N^{-1}T^{-1})
\end{aligned} \tag{128}$$

Thus we have

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{g_i=g\}\boldsymbol{\epsilon}_{i,t}\right\|^2=o_p(1) \tag{129}$$

The uniformly bounded term $\frac{1}{NTM}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}_{g_i,t}'\boldsymbol{\epsilon}_{i,t}$ also implies that $\frac{1}{NTM}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\alpha}_{g_i,t}^{0\prime}\boldsymbol{\epsilon}_{i,t}$ is bounded. We therefore proof Lemma 7.

We now proof that $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}^0$ regarding Hausdorff distance d_H in $\boldsymbol{\Theta}^G$ and $\hat{\boldsymbol{\alpha}}$ is consistent for $\boldsymbol{\alpha}^0$ regarding Hausdorff distance d_H in $\mathcal{R}^{GTM}$. We define Hausdorff distance as:

$$d_H(A,B)^2=\max\{H(A,B)^2,H(B,A)^2\} \tag{130}$$

Where

$$\begin{aligned}
H(A,B)^2&=\max_{a\in A}\left\{\min_{b\in B}\|a-b\|^2\right\} \\
H(B,A)^2&=\max_{b\in B}\left\{\min_{a\in A}\|b-a\|^2\right\}
\end{aligned} \tag{131}$$

As the objective function is invariant of relabelling of groups, thus we have the Hausdorff distance of the estimators as:

$$
\begin{aligned}
d_H(\{\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^0\}, \{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\alpha}}^0\})^2 &= \max\left\{H(A,B)^2, H(B,A)^2\right\} \\
&= \max\Bigg\{\max_{g\in 1,2\ldots G}\{\min_{\tilde{g}\in\{1,2\ldots G\}}\|\hat{\boldsymbol{\beta}}_{\tilde{g}}, -\boldsymbol{\beta}_g^0\|^2 + \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\tilde{g},t} - \boldsymbol{\alpha}_{g,t}^0\|^2\}, \\
&\max_{\tilde{g}\in 1,2\ldots G}\{\min_{g\in\{1,2\ldots G\}}\|\hat{\boldsymbol{\beta}}_{\tilde{g}}, -\boldsymbol{\beta}_g^0\|^2 + \frac{1}{MT}\sum_{t=1}^{T}\|\hat{\boldsymbol{\alpha}}_{\tilde{g},t} - \boldsymbol{\alpha}_{g,t}^0\|^2\}\Bigg\}
\end{aligned} \tag{132}
$$

Lemma 8 *$\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}^0$ regarding Hausdorff distance:*

$$d_H(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^0) = o_p(1)$$

and $\hat{\boldsymbol{\alpha}}$ is consistent for $\boldsymbol{\alpha}^0$ regarding Hausdorff distance:

$$d_H(\hat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}^0) = o_p(1)$$

Proof 12 (Proof of Lemma 8)

$$
\begin{aligned}
\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) &\overset{1}{=} \hat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) + o_p(1) \\
&\overset{2}{\leq} \hat{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^0, \tau^0) + o_p(1) \\
&\overset{3}{=} \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^0, \tau^0) + o_p(1)
\end{aligned} \tag{133}
$$

Where $\overset{1}{=}$ and $\overset{3}{=}$ is from Lemma 7, $\overset{2}{\leq}$ is the definition of least squares where the estimator are estimated by the argmin of objective function. However, by Lemma 2 we know that $\tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ is minimised at its true value. Then we have:

$$\left[\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^0, \tau^0)\right] \geq 0 \tag{134}$$

Combine equation 133 and equation 134 we have:

$$\left[\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{\tau}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^0, \tau^0)\right] = o_p(1) \tag{135}$$

On the other hand, we have:

$$
\begin{aligned}
&\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)-\tilde{\mathrm{Q}}(\boldsymbol{\beta^0},\boldsymbol{\alpha}^0,\tau^0)=\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\mathbf{z}'_{it}\left(\boldsymbol{\beta}^0_{g_i^0}-\boldsymbol{\beta}_{g_i}\right)+\boldsymbol{\alpha}^0_{g_i^0t}-\boldsymbol{\alpha}_{g_it}\right\|^2\\
&\overset{1}{=}\sum_{g=1}^{G}\sum_{\widetilde{g}=1}^{G}\begin{pmatrix}\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\\ \frac{1}{\sqrt{MT}}\left(\boldsymbol{\alpha}^0_g-\boldsymbol{\alpha}_{\widetilde{g}}\right)\end{pmatrix}' M(\gamma,g,\widetilde{g})\begin{pmatrix}\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\\ \frac{1}{\sqrt{MT}}\left(\boldsymbol{\alpha}^0_g-\boldsymbol{\alpha}_{\widetilde{g}}\right)\end{pmatrix}\\
&\overset{2}{\geq}\sum_{g=1}^{G}\sum_{\widetilde{g}=1}^{G}\widehat{\rho}(\gamma,g,\widetilde{g})\left[\left\|\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\right\|^2+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt}-\boldsymbol{\alpha}_{\widetilde{g}t}\right\|^2\right]\\
&\overset{3}{\geq}\sum_{g=1}^{G}\left(\sum_{\widetilde{g}=1}^{G}\widehat{\rho}(\gamma,g,\widetilde{g})\right)\min_{\widetilde{g}\in\{1,\dots,G\}}\left[\left\|\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\right\|^2+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt}-\boldsymbol{\alpha}_{\widetilde{g}t}\right\|^2\right]\\
&\overset{4}{\geq}\sum_{g=1}^{G}\left(\max_{\widetilde{g}\in\{1,\dots,G\}}\widehat{\rho}(\gamma,g,\widetilde{g})\right)\min_{\widetilde{g}\in\{1,\dots,G\}}\left[\left\|\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\right\|^2+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt}-\boldsymbol{\alpha}_{\widetilde{g}t}\right\|^2\right]\\
&\overset{5}{\geq}\sum_{g=1}^{G}\widehat{\rho}\times\min_{\widetilde{g}\in\{1,\dots,G\}}\left[\left\|\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\right\|^2+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt}-\boldsymbol{\alpha}_{\widetilde{g}t}\right\|^2\right]\\
&\overset{6}{\geq}\widehat{\rho}\times\max_{g\in\{1,\dots,G\}}\left\{\min_{\widetilde{g}\in\{1,\dots,G\}}\left[\left\|\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\right\|^2+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt}-\boldsymbol{\alpha}_{\widetilde{g}t}\right\|^2\right]\right\}
\end{aligned}
\tag{136}
$$

Where $\overset{1}{=}$ *comes from the definition of quadratic matrix form,* $\overset{2}{\geq}$ *is the eigenvalue inequalities for quadratic forms,* $\overset{3}{\geq}$ *is the definition of minimum,* $\overset{4}{\geq}$ *and* $\overset{6}{\geq}$ *is the definition of maximum and* $\overset{5}{\geq}$ *is stated by Assumption 3 a. The equation shows that the difference between* $\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)$ *and* $\tilde{\mathrm{Q}}(\boldsymbol{\beta^0},\boldsymbol{\alpha}^0,\tau^0)$ *is bounded away from zero. Combine with Equation 135 we have:*

$$
\max_{g\in\{1,\dots,G\}}\left\{\min_{\widetilde{g}\in\{1,\dots,G\}}\left[\left\|\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\right\|^2+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt}-\boldsymbol{\alpha}_{\widetilde{g}t}\right\|^2\right]\right\}=o_p(1) \tag{137}
$$

We proof the first half definition of the Hausdorff distance. Next we define function $\sigma(g)$ *as:*

$$
\sigma(g)=\arg\min_{\widetilde{g}\in\{1,\dots,G\}}\left[\left\|\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\right\|^2+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt}-\boldsymbol{\alpha}_{\widetilde{g}t}\right\|^2\right]
$$

Where the function selects the optimal $\widetilde{g}\in\{1,\dots,G\}$ *based on the minimised function* $\left\|\boldsymbol{\beta}^0_g-\boldsymbol{\beta}_{\widetilde{g}}\right\|^2+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}^0_{gt}-\boldsymbol{\alpha}_{\widetilde{g}t}\right\|^2$*. We next show that* $\sigma:\{1,\dots,G\}\to\{1,\dots,G\}$ *is an one-to-one function w.p.a.1. By triangle* *inequalities, we have: We have, for all $\widetilde{g} \neq g$:*

$$\left(\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}'\left(\widehat{\boldsymbol{\beta}}_{\sigma(g)}-\widehat{\boldsymbol{\beta}}_{\sigma(\widetilde{g})}\right)+\widehat{\boldsymbol{\alpha}}_{\sigma(g)t}-\widehat{\boldsymbol{\alpha}}_{\sigma(\widetilde{g})t}\right)^{2}\right)^{\frac{1}{2}} \geq$$
$$\underbrace{\left(\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}'\left(\boldsymbol{\beta}_{g}^{0}-\boldsymbol{\beta}_{\widetilde{g}}^{0}\right)+\boldsymbol{\alpha}_{gt}^{0}-\boldsymbol{\alpha}_{\widetilde{g}t}^{0}\right)^{2}\right)^{\frac{1}{2}}}_{a.1}$$
$$-\underbrace{\left(\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}'\left(\widehat{\boldsymbol{\beta}}_{\sigma(g)}-\boldsymbol{\beta}_{g}^{0}\right)+\boldsymbol{\alpha}_{\sigma(g)t}-\boldsymbol{\alpha}_{gt}^{0}\right)^{2}\right)^{\frac{1}{2}}}_{a.2}$$
$$-\underbrace{\left(\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{z}_{it}'\left(\widehat{\boldsymbol{\beta}}_{\sigma(\widetilde{g})}-\boldsymbol{\beta}_{\widetilde{g}}^{0}\right)+\widehat{\boldsymbol{\alpha}}_{\sigma(\widetilde{g})t}-\boldsymbol{\beta}_{\widetilde{g}t}^{0}\right)^{2}\right)^{\frac{1}{2}}}_{a.3}$$

Where $a.1$ is bounded away zero by Assumption 3 b, terms $a.2$ and $a.3$ is $o_p(1)$ by Equation 135. Therefore, we have $\sigma(g) \neq \sigma(\tilde{g})$ with probability approaching one, which implies that with probability approaching one σ is bijective and has the inverse which is denoted as σ^{-1}. Thus, for all $\tilde{g} \in \{1, 2, \dots, G\}$:

$$\begin{aligned}\min_{g\in\{1,\dots,G\}}\left\|\boldsymbol{\beta}_{g}^{0}-\hat{\boldsymbol{\beta}}_{\tilde{g}}\right\|^{2}+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{gt}^{0}-\hat{\boldsymbol{\alpha}}_{\tilde{g}t}\right\|^{2} &\leq \left\|\boldsymbol{\beta}_{\sigma^{-1}(\tilde{g})}^{0}-\hat{\boldsymbol{\beta}}_{\tilde{g}}\right\|^{2}+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{\sigma^{-1}(\tilde{g})t}^{0}-\hat{\boldsymbol{\alpha}}_{\tilde{g}t}\right\|^{2}\\ &= \min_{h\in\{1,\dots,G\}}\left\|\boldsymbol{\beta}_{\sigma^{-1}(\tilde{g})}^{0}-\hat{\boldsymbol{\beta}}_{h}\right\|^{2}+\frac{1}{MT}\sum_{t=1}^{T}\left\|\boldsymbol{\alpha}_{\sigma^{-1}(\tilde{g})t}^{0}-\hat{\boldsymbol{\alpha}}_{ht}\right\|^{2}\\ &\overset{1}{=} o_p(1)\end{aligned} \tag{138}$$

Where $\overset{1}{=}$ is from Equation 135. Thus we complete the second half definition of Hausdorff distance by Equation 138.

The proof of Lemma 8 shows that there exists a permutation $\sigma : \{1, \dots, G\} \to \{1, \dots, G\}$ such that:

$$\left\|\widehat{\boldsymbol{\beta}}_{\sigma(g)}-\boldsymbol{\beta}_{g}^{0}\right\|^{2}+\frac{1}{MT}\sum_{t=1}^{T}\left(\widehat{\boldsymbol{\alpha}}_{\sigma(g)t}-\boldsymbol{\alpha}gt^{0}\right)^{2}\xrightarrow{p}0$$

By relabelling we may take $\sigma(g) = g$.

Now, we define the set:

$$\mathcal{N}_\eta = \left\{ \boldsymbol{\beta}, \boldsymbol{\alpha} \in \boldsymbol{\Omega}^G \times \boldsymbol{\Theta}^{GTM} : \left\| \boldsymbol{\beta}_g - \boldsymbol{\beta}_g^0 \right\|^2 < \eta, \frac{1}{MT} \sum_{t=1}^{T} \left\| \boldsymbol{\alpha}_{g,t}^0 - \boldsymbol{\alpha}_{g,t} \right\|^2 < \eta, g \in \{1, 2, \ldots, G\} \right\} \tag{139}$$

Lemma 9 *Let* $\hat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \underset{g \in \{1,\ldots,G\}}{\operatorname{argmin}} \frac{1}{MT} \sum_{t=1}^{T} \left\| (\boldsymbol{y}_{i,t} - \boldsymbol{z}_{i,t}\boldsymbol{\beta} - \boldsymbol{\alpha}_{g_i,t}) \right\|^2$*, for some* $\eta > 0$ *but small enough while* N *and* T *go to infinity, we have for all* $\delta > 0$ *such that:*

$$\sup_{(\boldsymbol{\beta}, \boldsymbol{\alpha}) \in \mathcal{N}_\eta} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1} \left\{ \widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \neq g_i^0 \right\} = o_p\left(T^{-\delta}\right) \tag{140}$$

Proof 13 (Proof of Lemma 9) *For all* $g \in \{1, 2, \ldots, G\}$*, we have the following inequality:*

$$\mathbf{1} \left\{ \widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = g \right\} \leq \mathbf{1} \left\{ \sum_{t=1}^{T} \left\| \mathbf{y}_{it} - \mathbf{z}_{it}\boldsymbol{\beta}_g - \boldsymbol{\alpha}_{gt} \right\|^2 \leq \sum_{t=1}^{T} \left\| \mathbf{y}_{it} - \mathbf{z}_{it}\boldsymbol{\beta}_g - \boldsymbol{\alpha}_{g_i^0 t} \right\|^2 \right\} \tag{141}$$

For simplicity, we take $\boldsymbol{\beta}_g = \boldsymbol{\beta}_{g_i,t}$ *for* $g_i = g$*,* $t = 1, 2 \ldots, T$*. Then we can write the model compactly without loss of generalisation. Sum across units* N *we have:*

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1} \left\{ \widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \neq g_i^0 \right\} &= \sum_{g=1}^{G} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1} \left\{ g_i^0 \neq g \right\} \mathbf{1} \left\{ \widehat{g}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = g \right\} \\ &\leq \sum_{g=1}^{G} \frac{1}{N} \sum_{i=1}^{N} Z_{ig}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \end{aligned} \tag{142}$$

Where

$$\begin{aligned} Z_{ig}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{1} \left\{ g_i^0 \neq g \right\} \times \mathbf{1} \left\{ \sum_{t=1}^{T} \left\| \mathbf{y}_{it} - \mathbf{z}_{it}\boldsymbol{\beta}_{g_i,t} \right\|^2 \leq \sum_{t=1}^{T} \left\| \mathbf{y}_{it} - \mathbf{z}_{it}\boldsymbol{\beta}_{g_i^0,t} \right\|^2 \right\} \\ \max_{\tilde{g} \in \mathbb{G} \setminus \{g\}} 1 \left\{ \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} - \boldsymbol{\beta}_{g,t} \right) \left(\mathbf{z}_{it}' \boldsymbol{\beta}_{\tilde{g},t}^0 + \boldsymbol{\epsilon}_{it} - \frac{\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} + \boldsymbol{\beta}_{g,t} \right)}{2} \right) \leq 0 \right\} \end{aligned} \tag{143}$$

Where $\overset{1}{\leq}$ *holds as when* $\tilde{g} \neq g$*,* $\mathbf{1} \left\{ g_i^0 \neq g \right\}$ *holds. Let*

$$\begin{aligned} \mathbb{A}_T = \Bigg| &\sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} - \boldsymbol{\beta}_{g,t} \right) \left(\mathbf{z}_{it}' \boldsymbol{\beta}_{\tilde{g},t}^0 + \boldsymbol{\epsilon}_{it} - \frac{x_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} + \boldsymbol{\beta}_{g,t} \right)}{2} \right) \\ &- \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^0 - \boldsymbol{\beta}_{g,t}^0 \right) \left(\mathbf{z}_{it}' \boldsymbol{\beta}_{\tilde{g},t}^0 + \boldsymbol{\epsilon}_{it} - \frac{\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^0 + \boldsymbol{\beta}_{g,t}^0 \right)}{2} \right) \Bigg| \end{aligned} \tag{144}$$

Then,

$$\begin{aligned}
\mathbb{A}_T \leq & \left| \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} - \boldsymbol{\beta}_{g,t}\right) \boldsymbol{\epsilon}_{it} - \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} - \boldsymbol{\beta}_{g,t}^{0}\right) \boldsymbol{\epsilon}_{it} \right| \\
& + \left| \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} - \boldsymbol{\beta}_{g,t}\right) \mathbf{z}_{it}'\boldsymbol{\beta}_{\tilde{g},t}^{0} - \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} - \boldsymbol{\beta}_{g,t}^{0}\right) \mathbf{z}_{it}'\boldsymbol{\beta}_{\tilde{g},t}^{0} \right| \\
& + \left| \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} - \boldsymbol{\beta}_{g,t}\right) \frac{\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} + \boldsymbol{\beta}_{g,t}\right)}{2} - \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} - \boldsymbol{\beta}_{g,t}\right) \frac{\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} + \boldsymbol{\beta}_{g,t}^{0}\right)}{2} \right| \\
& + \left| \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} - \boldsymbol{\beta}_{g,t}\right) \frac{\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} + \boldsymbol{\beta}_{g,t}^{0}\right)}{2} - \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} - \boldsymbol{\beta}_{g,t}^{0}\right) \frac{\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} + \boldsymbol{\beta}_{g,t}^{0}\right)}{2} \right|
\end{aligned} \tag{145}$$

Thus, when $\boldsymbol{\beta} \in \mathcal{N}_\eta$, the CS inequality implies that

$$\begin{aligned}
\mathbb{A}_T \leq & 2MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left\| \boldsymbol{\epsilon}_{it} \mathbf{z}_{it} \right\|^2 \right)^{1/2} \sqrt{\eta} + 2MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left(\mathbf{z}_{it}' \boldsymbol{\beta}_{\tilde{g},t}^{0} \right)^2 \left\| \mathbf{z}_{it} \right\|^2 \right)^{1/2} \sqrt{\eta} \\
& + MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left(\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t} - \boldsymbol{\beta}_{g,t} \right) \right)^2 \left\| \mathbf{z}_{it} \right\|^2 \right)^{1/2} \sqrt{\eta} + MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left(\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} + \boldsymbol{\beta}_{g,t}^{0} \right) \right)^2 \left\| \mathbf{z}_{it} \right\|^2 \right)^{1/2}
\end{aligned} \tag{146}$$

As $\mathcal{B}$ is bounded, we have, for $\boldsymbol{\beta} \in \mathcal{N}_\eta$,

$$\mathbb{A}_T \leq C_1 \sqrt{\eta} MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left\| \boldsymbol{\epsilon}_{it} \mathbf{z}_{it} \right\|^2 \right)^{1/2} + C_2 \sqrt{\eta} MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left\| \mathbf{z}_{it} \right\|^4 \right)^{1/2} \tag{147}$$

where C_1 and C_2 are constants that are independent of η and T. Thus, we have

$$\begin{aligned}
Z_{ig}(\boldsymbol{\beta}) \leq \max_{\tilde{g} \in \mathbb{G} \setminus \{g\}} \mathbf{1} \Bigg\{ & \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} - \boldsymbol{\beta}_{g,t}^{0} \right) \left(\mathbf{z}_{it}' \boldsymbol{\beta}_{\tilde{g},t}^{0} + \boldsymbol{\epsilon}_{it} - \frac{\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} + \boldsymbol{\beta}_{g,t}^{0} \right)}{2} \right) \\
& \leq C_1 \sqrt{\eta} MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left\| \boldsymbol{\epsilon}_{it} \mathbf{z}_{it} \right\|^2 \right)^{1/2} + C_2 \sqrt{\eta} MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left\| \mathbf{z}_{it} \right\|^4 \right)^{1/2} \Bigg\}.
\end{aligned} \tag{148}$$

Let

$$\begin{aligned}
\tilde{Z}_{ig} = \max_{\tilde{g} \in \mathbb{G} \setminus \{g\}} \mathbf{1} \Bigg\{ & \sum_{t=1}^{T} \mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} - \boldsymbol{\beta}_{g,t}^{0} \right) \left(\mathbf{z}_{it}' \boldsymbol{\beta}_{\tilde{g},t}^{0} + \boldsymbol{\epsilon}_{it} - \frac{\mathbf{z}_{it}' \left(\boldsymbol{\beta}_{\tilde{g},t}^{0} + \boldsymbol{\beta}_{g,t}^{0} \right)}{2} \right) \\
& \leq C_1 \sqrt{\eta} MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left\| \boldsymbol{\epsilon}_{it} \mathbf{z}_{it} \right\|^2 \right)^{1/2} + C_2 \sqrt{\eta} MT \left(\frac{1}{MT} \sum_{t=1}^{T} \left\| \mathbf{z}_{it} \right\|^4 \right)^{1/2} \Bigg\}
\end{aligned} \tag{149}$$

Thus, we have

$$\sup_{\boldsymbol{\beta}\in\mathcal{N}_\eta} \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta}) \neq g_i^0\right\} \leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N} Z_{ig}(\boldsymbol{\beta}) \leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}\tilde{Z}_{ig}.$$

Note that $\tilde{Z}_{ig}$ does not depend on $\boldsymbol{\beta}$.

We now show that $Pr(\tilde{Z}_{ig} = 1)$ is bounded. Denote $\tilde{M} > \max(M, M^*)$ Where M and M^* are some constant defined in Assumption 6 (a) and 7 (e). Note that by Assumption 6 (c) and Cauchy-Schwarz inequality, we also have:

$$\begin{aligned} E(\|\boldsymbol{\epsilon}_{it}\|^2) &= \sqrt{E(\|\boldsymbol{\epsilon}_{it}\|^2 \times 1^2)^2} \\ &\leq \sqrt{E(\|\boldsymbol{\epsilon}_{it}\|^4)} \\ &\leq \sqrt{M} \end{aligned} \tag{150}$$

By Boole's inequality we have:

$$\begin{aligned} &\Pr\left(\tilde{Z}_{ig} = 1\right) \\ \leq &\sum_{\tilde{g}\in\mathbb{G}\setminus\{g\}}\left[\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\|\boldsymbol{\epsilon}_{it}\mathbf{z}_{it}\|^2 \geq M^*\right) + \Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\|\mathbf{z}_{it}\|^4 \geq M^*\right)\right. \\ &+\Pr\left(\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{z}_{it}'\left(\boldsymbol{\beta}_{\tilde{g},t}^0 - \boldsymbol{\beta}_{g,t}^0\right)\right)^2 \leq \frac{c_{g,\tilde{g}}}{2}\right) \\ &\left.+\Pr\left(\sum_{t=1}^{T}\mathbf{z}_{it}'\left(\boldsymbol{\beta}_{\tilde{g},t}^0 - \boldsymbol{\beta}_{g,t}^0\right)\boldsymbol{\epsilon}_{it} \leq -T\frac{c_{g,\tilde{g}}}{4} + TC_3\sqrt{\eta}\sqrt{M^*}\right)\right], \end{aligned} \tag{151}$$

where C_3 is a constant that is independent of η and T. Take η such that

$$\eta \leq \left(\frac{\min_{\tilde{g}\in\mathbb{G}\setminus\{g\}} c_{g,\tilde{g}}}{8C_3\sqrt{M^*}}\right). \tag{152}$$

We then have

$$\begin{aligned} &\Pr\left(\sum_{t=1}^{T}\mathbf{z}_{it}'\left(\boldsymbol{\beta}_{\tilde{g},t}^0 - \boldsymbol{\beta}_{g,t}^0\right)\boldsymbol{\epsilon}_{it} \leq -T\frac{c_{g,\tilde{g}}}{4} + TC_3\sqrt{\eta}\sqrt{M^*}\right) \\ \leq &\Pr\left(\frac{1}{MT}\sum_{t=1}^{T}\mathbf{z}_{it}'\left(\boldsymbol{\beta}_{\tilde{g},t}^0 - \boldsymbol{\beta}_{g,t}^0\right)\boldsymbol{\epsilon}_{it} \leq -\frac{c_{g,\tilde{g}}}{8}\right) \\ \leq &\Pr\left(\left|\frac{1}{MT}\sum_{t=1}^{T}\mathbf{z}_{it}'\left(\boldsymbol{\beta}_{\tilde{g},t}^0 - \boldsymbol{\beta}_{g,t}^0\right)\boldsymbol{\epsilon}_{it}\right| > \frac{c_{g,\tilde{g}}}{8}\right) \end{aligned} \tag{153}$$

Then we have the following equations:

$$\Pr\left(\left|\frac{1}{MT}\sum_{t=1}^{T}\left(\mathbf{z}_{it}'\left(\boldsymbol{\beta}_{\tilde{g}}^0 - \boldsymbol{\beta}_{g}^0\right) + \boldsymbol{\alpha}_{\tilde{g}}^0 - \boldsymbol{\alpha}_{g}^0\right)\boldsymbol{\epsilon}_{it}\right| > \frac{c_{g,\tilde{g}}}{8}\right) = O\left(T^{-\delta}\right) \tag{154}$$

We thus have

$$\Pr\left(\tilde{Z}_{ig}=1\right)\leq (G-1)O\left(T^{-\delta}\right).$$

This implies that

$$\begin{aligned}
E\left(\sup_{\beta\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha})\neq g_i^0\right\}\right) &\leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}E\left(\tilde{Z}_{ig}\right)\\
&=\frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}\Pr\left(\tilde{Z}_{ig}=1\right)\\
&=G(G-1)O\left(T^{-\delta}\right)=O\left(T^{-\delta}\right)
\end{aligned}$$

By Markov inequality we have:

$$\begin{aligned}
\Pr(\sup_{\boldsymbol{\beta},\boldsymbol{\alpha}\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha})\neq g_i^0\right\}>\epsilon T^{-\delta}) &\leq \frac{E\left(\sup_{\boldsymbol{\beta},\boldsymbol{\alpha}\in\mathcal{N}_\eta}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha})\neq g_i^0\right\}\right)}{\epsilon T^{-\delta}}\\
&=o_p(1)
\end{aligned} \tag{155}$$

Which ends the proof of Lemma 9.

Now denote

$$\hat{Q}(\boldsymbol{\beta},\boldsymbol{\alpha},\hat{\tau})=\hat{Q}(\boldsymbol{\beta},\boldsymbol{\alpha})=\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\boldsymbol{y}_{i,t}-\boldsymbol{z}_{i,t}\boldsymbol{\beta}_g-\boldsymbol{\alpha}_{\hat{g}_i,t}\|^2 \tag{156}$$

and

$$\tilde{Q}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau^0)=\tilde{Q}(\boldsymbol{\beta},\boldsymbol{\alpha})=\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\boldsymbol{y}_{i,t}-\boldsymbol{z}_{i,t}\boldsymbol{\beta}_{g^0}-\boldsymbol{\alpha}_{g_i^0,t}\right\|^2 \tag{157}$$

Under Lemma 9, we can show the following Lemma:

Lemma 10

$$\operatorname*{plim}_{N,T\to\infty}\sup_{(\boldsymbol{\beta},\boldsymbol{\alpha})\in\mathcal{N}_n}\left|\hat{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha})-\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha})\right|=o_p(T^{-\delta}) \tag{158}$$

Proof 14 (Proof of Lemma 10) *It is easy to show that for $(\boldsymbol{\beta},\boldsymbol{\alpha})\in\mathcal{N}_n$ and N T goes to infinity:*

$$\begin{aligned}
\mathbb{Q}&=\hat{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha})-\tilde{\mathrm{Q}}(\boldsymbol{\beta},\boldsymbol{\alpha})\\
&=\frac{1}{MNT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\left\{\hat{g}_i(\boldsymbol{\beta},\boldsymbol{\alpha})\neq g_i^0\right\}\left(\|\boldsymbol{y}_{i,t}-\boldsymbol{z}_{i,t}\boldsymbol{\beta}_g-\boldsymbol{\alpha}_{\hat{g}_i,t}\|^2-\left\|\boldsymbol{y}_{i,t}-\boldsymbol{z}_{i,t}\boldsymbol{\beta}_{g^0}-\boldsymbol{\alpha}_{g_i^0,t}\right\|^2\right)\\
&\stackrel{1}{=}o_p(T^{-\delta})
\end{aligned} \tag{159}$$

Where $\stackrel{1}{=}$ is indicated by Lemma 9

Hence, by the definition of $\mathcal{N}_n$, the consistent estimator $\hat{\boldsymbol{\beta}}$ under Theorem 5 and the consistent estimator $\hat{\boldsymbol{\alpha}}$ under Theorem 5 in objective function $\hat{\mathcal{Q}}$, we have:

$$\Pr\left((\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) \notin \mathcal{N}_\eta\right) \to 0 \quad (N,T) \to \infty \tag{160}$$

Similarly, by the definition of least square estimator, $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\alpha}}$ are still consistent estimators under Theorem 5 and Theorem 5 in objective function $\tilde{\mathcal{Q}}$. Thus:

$$\Pr\left((\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \notin \mathcal{N}_\eta\right) \to 0 \quad (N,T) \to \infty \tag{161}$$

By basic probability algebra, we have the following inequality:

$$\begin{aligned}
&\Pr\left[\left|\hat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})\right| > \varepsilon T^{-\delta}\right] \\
&\quad \le \underbrace{\Pr\left((\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) \notin \mathcal{N}_\eta\right)}_{a.1:\to 0} + \underbrace{\Pr\left[\sup_{(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) \in \mathcal{N}_n} \left|\hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})\right| > \varepsilon T^{-\delta}\right]}_{a.2:o_p(1)} \\
&\quad = o_p(1)
\end{aligned} \tag{162}$$

Thus we have:

$$\hat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = o_p(T^{-\delta}) \tag{163}$$

Similarly, we have

$$\begin{aligned}
&\Pr\left[\left|\hat{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})\right| > \varepsilon T^{-\delta}\right] \\
&\quad \le \underbrace{\Pr\left((\tilde{\boldsymbol{\beta}}, \tilde{\alpha}) \notin \mathcal{N}_\eta\right)}_{a.1:\to 0} + \underbrace{\Pr\left[\sup_{(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \in \mathcal{N}_n} \left|\tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})\right| > \varepsilon T^{-\delta}\right]}_{a.2:o_p(1)} \\
&\quad = o_p(1)
\end{aligned} \tag{164}$$

Thus we have:

$$\hat{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) = o_p(T^{-\delta}) \tag{165}$$

By the definition of least square estimators, equation 163 and equation 165, we have the following inequality:

$$0 \le \underbrace{\widetilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \widetilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})}_{a.1} = \underbrace{\widehat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \widehat{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})}_{a.2} + o_p\left(T^{-\delta}\right) \le o_p\left(T^{-\delta}\right) \tag{166}$$

Where $a.1 \geq 0$ as $\widetilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})$ minimised the objective function $\widetilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ by least square definition and $a.2 \leq 0$ as $\widehat{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$ minimised the objective function $\hat{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ by least square definition. Thus, for all $\delta > 0$, we have:

$$\widetilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \widetilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) = o_p\left(T^{-\delta}\right) \tag{167}$$

Next we show that $\tilde{\mathrm{Q}}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is uniquely minimised at $\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}$ by the following lemma.

Lemma 11 *For $\hat{\boldsymbol{\beta}} \neq \tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}} \neq \tilde{\boldsymbol{\alpha}}$*

$$\tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \geq \rho \left\|\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} - \hat{\boldsymbol{\beta}}_g\right\|^2 \tag{168}$$

$$\begin{aligned}
\mathbb{Q} &= \tilde{\mathrm{Q}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \tilde{\mathrm{Q}}(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}) \\
&\geq \frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\left\| (\boldsymbol{z}_{i,t} - \boldsymbol{z}_{g \wedge \tilde{g},t})(\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{g}}) \right\|^2 \right] \\
&\stackrel{1}{=} (\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{g}})' \Sigma(\gamma) (\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{g}}) \\
&\geq (\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{g}})' \min_{\gamma \in \boldsymbol{\Gamma}_G} \Sigma(\gamma) (\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{g}}) \\
&\stackrel{2}{\geq} \min_{\gamma \in \boldsymbol{\Gamma}_G} \hat{\rho}(\gamma) \left\| \tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} - \hat{\boldsymbol{\beta}}_{\boldsymbol{g}} \right\|^2
\end{aligned} \tag{169}$$

Where $\stackrel{1}{=}$ holds by the least square estimators' properties, $\stackrel{2}{\geq}$ holds by Assumption 6 (g). The intuition behind the equation is that the difference of objection function values is bounded away from zero when parameters are not equal to $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})$.

Proof 15 (Proof of Theorem 3) *It is easy to show that under the Lemma 5, Lemma 6 and Theorem 7, we have:*

$$\begin{aligned}
\frac{1}{MNT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \hat{\boldsymbol{\alpha}}_{g_i^0,t} - \tilde{\boldsymbol{\alpha}}_{g_i^0,t} \right\|^2 &= \frac{1}{MNT} \sum_{g=1}^{G} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}\left\{g_i^0 = g\right\} \left\| \hat{\boldsymbol{\alpha}}_{g^0 t} - \tilde{\boldsymbol{\alpha}}_{g^0 t} \right\|^2 \\
&\stackrel{1}{=} \sum_{g=1}^{G} \pi_g \frac{1}{MT} \sum_{t=1}^{T} \mathbf{1}\left\{g_i^0 = g\right\} \left\| \hat{\boldsymbol{\alpha}}_{g^0 t} - \tilde{\boldsymbol{\alpha}}_{g^0 t} \right\|^2 \\
&= o_p(T^{-\delta})
\end{aligned} \tag{170}$$

Where $\stackrel{1}{=}$ comes from Assumption 7 (b), Thus, for all g:

$$\frac{1}{MT} \sum_{t=1}^{T} \left\| \hat{\boldsymbol{\alpha}}_{g^0 t} - \tilde{\boldsymbol{\alpha}}_{g^0 t} \right\|^2 = o_p(T^{-\delta}) \tag{171}$$

Which implies that, for all t, we have for all $\delta > 0$:

$$\frac{1}{M}\left\|\hat{\boldsymbol{\alpha}}_{g^0t} - \tilde{\boldsymbol{\alpha}}_{g^0t}\right\|^2 \leq o_p(T^{1-\delta}) \tag{172}$$

Proof 16 (Proof of Theorem 4) *By Boole's inequality, we have:*

$$\begin{aligned}\Pr\left(\sup_{i\in\{1,\dots,N\}}\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right) &= \Pr\left(\bigcup_{i\in\{1,\dots,N\}}\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right)\\ &\leq \sum_{i=1}^{N}\Pr\left(\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right)\\ &\leq N\sup_{i\in\{1,\dots,N\}}\Pr\left(\left|\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})-g_i^0\right|\neq 0\right)\\ &\leq \underbrace{\Pr\left((\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\notin\mathcal{N}_\eta\right)}_{a.1}\\ &\quad\underbrace{+N\sup_{i\in\{1,\dots,N\}}\Pr\left((\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\in\mathcal{N}_\eta,\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\neq g_i^0\right)}_{a.2}\end{aligned} \tag{173}$$

$a.1$ towards to zero by Equation 160, for $a.2$:

$$\begin{aligned}N\sup_{i\in\{1,\dots,N\}}\Pr\left((\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\in\mathcal{N}_\eta,\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\neq g_i^0\right) &\overset{1}{=} N\sup_{i\in\{1,\dots,N\}}\mathrm{E}\left(\mathbf{1}\left\{(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\in\mathcal{N}_\eta\right\}\mathbf{1}\left\{\widehat{g}_i(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\neq g_i^0\right\}\right)\\ &\overset{2}{\leq} N\sup_{i\in\{1,\dots,N\}}\mathrm{E}\left(\mathbf{1}\left\{(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\alpha}})\in\mathcal{N}_\eta\right\}\sum_{g=1}^{G}\tilde{Z}_{ig}\right)\\ &\leq N\sup_{i\in\{1,\dots,N\}}\mathrm{E}\left(\sum_{g=1}^{G}\tilde{Z}_{ig}\right)\\ &= N\sup_{i\in\{1,\dots,N\}}\sum_{g=1}^{G}\Pr\left(\tilde{Z}_{ig}=1\right)\\ &= NG(G-1)o_p(T^{-\delta})\\ &= o_p(NT^{-\delta})\end{aligned} \tag{174}$$

Where $\overset{1}{=}$ is the definition of indicator function, $\overset{2}{\leq}$ comes from Lemma 4. When $NT^{-\delta}$ tends to zero, Equation 174 is $o_p(1)$. Thus we end the proof of the theroem.

Proof 17 (Proof of Corollary 1) *Under the known group, we have the following* *algebraic expression for $\tilde{\boldsymbol{\alpha}}$:*

$$\widetilde{\boldsymbol{\alpha}}_{gt} = \bar{\mathbf{y}}_{g_i^0 t} - \bar{\mathbf{z}}'_{g_i^0 t}\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}}$$
$$= \frac{\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}\left(\mathbf{y}_{it} - \mathbf{z}'_{it}\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}}\right)}{\sum_{i=1}^{N} \mathbf{1}\left\{g_i^0 = g\right\}} \tag{175}$$

and $\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}}$:

$$\tilde{\boldsymbol{\beta}}_{\boldsymbol{g}} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{g_i^0 = g\}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)'\right)^{-1} \tag{176}$$
$$\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{g_i^0 = g\}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)\left(\mathbf{y}_{it} - \bar{\mathbf{y}}_{g_i^0 t}\right)\right) \tag{177}$$

By using equation 175 we can easily show that:

$$\begin{aligned}\sqrt{NT}\left(\widetilde{\boldsymbol{\beta}_{\boldsymbol{g}}} - \boldsymbol{\beta}^0_{\boldsymbol{g}}\right) &= \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{g_i^0 = g\}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)'\right)^{-1} \\ &\quad \left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}\{g_i^0 = g\}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)\boldsymbol{\epsilon}_{it}\right) \\ &\sim N(0, \Sigma_\beta^{-1}\Omega_\beta\Sigma_\beta^{-1})\end{aligned} \tag{178}$$

Since the infeasible estimate $\widetilde{\boldsymbol{\beta}_{\boldsymbol{g}}}$ is asymptotic equivalent to $\hat{\boldsymbol{\beta}}_{\boldsymbol{g}}$, we have:

$$\sqrt{NT}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{g}} - {\boldsymbol{\beta}_{\boldsymbol{g}}}^0\right) \sim N(0, \Sigma_{\beta_g}^{-1}\Omega_{\beta_g}\Sigma_{\beta_g}^{-1}) \tag{179}$$

Chapter 4: Testing for Granger Causality in Grouped Panels

Abstract

This paper proposes a Wald type test for Granger non-causality in grouped panels with group-specific heterogeneous coefficient and grouped fixed effects. Under the null hypothesis, grouped heterogeneous parameters are reduced to zero then they are treated as homogeneous parameters. The homogeneous part then could be estimated by standard OLS projecting out other parameters. Monte Carlo simulations show that it has desirable results for both the size of the test and the power of the test under large N and T.

4.1 Introduction

The study of causal relationships between economic variables is a fundamental aspect of econometrics, and understanding these relationships often relies on the concept of Granger causality. Originally developed by Granger (1969), Granger causality provides a framework for determining whether one time series contains predictive information about another. This method has been widely adopted across disciplines, particularly in macroeconomics and finance, to examine directional relationships between variables such as GDP, interest rates, and consumption.

In the context of panel data, where observations include both cross-sectional units (N) and time-series periods (T), Granger non-causality testing introduces additional complexities. Traditional panel Granger causality tests often impose the assumption of homogeneous relationships across unitsimplying that all units (e.g., countries or firms) respond in a similar fashion. This assumption, however, overlooks the potential for heterogeneous dynamics, in which different units may exhibit distinct causal patterns. For instance, in cross-country analyses, nations with varying levels of economic development or institutional structures may respond differently to similar shocks in financial markets or energy demand.

Ignoring such heterogeneity can result in biased or misleading conclusions regarding the nature and direction of causal relationships.

A key attempt to address this issue was made by Dumitrescu & Hurlin (2012), who developed a Granger non-causality test for heterogeneous panels by averaging individual Granger test statistics across units. While this method accounts for slope heterogeneity, it still assumes that unit-specific fixed effects are uncorrelated with the lagged explanatory variables. This assumption can lead to biased inference in dynamic panel models, particularly when T is relatively small compared to N.

Recent developments in econometrics have aimed to address heterogeneity more explicitly. A notable contribution is by Juodis et al. (2021), who introduced a Half-Panel Jackknife (HPJ) procedure to correct for bias in dynamic panel settings. Their method mitigates the so-called Nickell bias by employing a split-panel jackknife estimator, which centres the bias around zero. However, the inclusion of such bias correction techniques adds complexity to the testing framework. In contrast to the HPJ approach, this chapter proposes a new Wald-type test for Granger non-causality within a Grouped Fixed Effects (GFE) framework, as introduced by Bonhomme & Manresa (2015), adapted to the Panel VAR context. The key innovation lies in modelling both the slope coefficients and fixed effects as group-specific, allowing for flexible heterogeneity across latent clusters. Unlike conventional dynamic panel estimators, our approach avoids the need for jackknife-type bias correction by leveraging the asymptotic properties of the grouped fixed effects estimator. As both N and T increase, the model mitigates Nickell bias (Nickell 1981) naturally through its structure. It is important to clarify that while Bonhomme & Manresa (2015) proposed a novel method for estimating grouped fixed effects in static panel settings, their framework does not explicitly address the Nickell bias associated with dynamic models. Their focus was on parsimoniously modelling unobserved heterogeneity via latent group structures rather than on bias correction in dynamic contexts.

Our model builds on their latent grouping framework and extends it to a dynamic PVAR structure, offering a robust alternative for Granger causality testing in heterogeneous panel environments.

The importance of addressing heterogeneity in Granger causality testing becomes even more apparent in contexts where units naturally cluster into groups based on shared characteristics. For example, regions within a country may share similar economic structures, or firms within the same industry may follow comparable investment patterns. Recognising such grouped structures and accounting for group-specific dynamics can significantly enhance the accuracy of causal inference in panel data settings. In many empirical applications, however, the true group memberships are unknown or latent, complicating the process of modelling group-specific effects.

To address these challenges, this chapter introduces a Wald-type test for Granger non-causality in panels with grouped fixed effects and heterogeneous coefficients. This approach extends traditional Granger causality tests by allowing both the coefficients and fixed effects to vary across latent groups, enabling the simultaneous testing of causal relationships within and across groups. The key innovation lies in the method's flexibility to accommodate group-specific heterogeneity without requiring prior knowledge of group memberships. Under the null hypothesis, group-specific coefficients are restricted to zero, implying homogeneity in causal relationships. The test then identifies statistically significant group-level patterns while correcting for biases that arise when heterogeneity is ignored. This contribution is important for several reasons. First, it offers a new framework for Granger non-causality testing in settings where cross-sectional units may follow distinct structural dynamics. In many empirical contextssuch as those involving countries, industries, or sectorsthe assumption of homogeneity across units is invalid. By allowing for group-level heterogeneity, the proposed method offers a more nuanced understanding of causal interactions across clusters. Second, the test provides an efficient and computationally feasible way to estimate causal effects in large panel datasets, particularly when both N and T are large. This is especially relevant in macroeconomic and financial applications, where latent group structures are common and panel dimensions are high. The Wald-type test proposed in this chapter fills a gap in the econometrics literature by offering a tractable and robust solution to the problem of grouped heterogeneity in panel Granger causality testing. Unlike existing methods such as Dumitrescu & Hurlin (2012), the proposed framework accommodates unobserved group membership, yielding more reliable inference in heterogeneous panel settings.

The structure of this chapter is as follows. Section 2 introduces the theoretical model and the null hypothesis underlying the Wald-type test. Section 3 reports the results of Monte Carlo simulations, which demonstrate the effectiveness of the proposed test in capturing varying degrees of group-level heterogeneity. Section 4 concludes the chapter[24].

Throughout the paper, we adopt the following notation. For an $m \times n$ real matrix A, we denote its transpose by A' and its Frobenius norm by $|A|_F$. For an $m \times 1$ real vector a, we denote its Euclidean norm by $|a|_2$. The operator $\xrightarrow{P}$ denotes convergence in probability, and $\xrightarrow{D}$ denotes convergence in distribution. We write $(N, T) \to \infty$ to indicate that N and T tend to infinity jointly.

4.2 Testing Model Setup

We propose the setup of PVAR-GFE(P) testing model as follows[25]:

$$y_{m,i,t} = \sum_{p=1}^{P} y_{m,i,t-p}\theta^{m}_{g_i,p} + \sum_{q=1}^{Q}\sum_{n\neq m}^{M} y_{n,i,t-q}\phi^{n}_{g_i,q} + \alpha_{m,g_i,t} + \epsilon_{m,i,t} \tag{180}$$

Let $i \in 1, 2, \ldots, N$ index cross-sectional units, $t \in 1, 2, \ldots, T$ denote time periods, $p \in 1, 2, \ldots, P$ represent lag orders, and $m \in 1, 2, \ldots, M$ index the equations in the

[24] For this paper, we use MATLAB 2024b version as the main programming language for Monte Carlo Simulations. The key function is the PVAR-GFE with heterogeneous coefficients model based on the study of Chapter 3. In MATLAB, we use parfor to enable parallel computation across multiple starting values to improve the efficiency of the PVAR-GFE estimation.

[25] The model could also include exogenous variables but we remove it for simplicity.

PVAR system. The dependent variable for each equation is denoted by $y_{m,i,t}$. The parameters $\theta^m g_i, p$ and $\phi^n g_i, q$ are group-specific heterogeneous coefficients to be estimated. This model includes two types of shocks. The term $\epsilon_{m,i,t}$ represents the idiosyncratic error, while $\alpha_{m,g_i,t}$ denotes the grouped fixed effects, which vary over time for each group and each equation. Here, $g \in 1, 2, \ldots, G$ indexes the latent groups, and $g_i \in 1, 2, \ldots, G$ denotes the group membership of unit i. We define $\mathbb{G}_g$ as the set of cross-sectional units belonging to group g. The group membership structure is mutually exclusive and exhaustive. That is, $\alpha m, g_j, t \neq \alpha m, g_k, t$ for any $j \neq k$, $\bigcup_{g=1}^{G} \mathbb{G}_g = 1, 2, \ldots, N$, and $\mathbb{G}_j \cap \mathbb{G}_k = \varnothing$ for any $j \neq k$. A key assumption of the model is that both the number of groups G and the group memberships g_i remain fixed over time.

For example, a simple bivariate PVAR-GFE(P) model for each equation is [26]:

$$
\begin{aligned}
y_{1,i,t} &= \sum_{p=1}^{P} \theta^1_{g_i,p} y_{1,i,t-p} + \sum_{q=1}^{Q} \phi^1_{g_i,q} y_{2,i,t-q} + \alpha_{1,g_i,t} + \epsilon_{1,i,t} \\
y_{2,i,t} &= \sum_{p=1}^{P} \theta^2_{g_i,p} y_{1,i,t-p} + \sum_{q=1}^{Q} \phi^2_{g_i,q} y_{2,i,t-q} + \alpha_{2,g_i,t} + \epsilon_{2,i,t}
\end{aligned}
\tag{181}
$$

In the Panel VAR system, if not all coefficients $\phi^1_{g_i,q}$ are equal to zero, then y_2 is said to Granger-cause y_1. A feedback effect is present if y_2 Granger-causes y_1 and y_1 Granger-causes y_2 simultaneously. The null hypothesis for the Granger non-causality test is formulated as follows:

$$
H_0: \quad \phi^1_{g_i,q} = 0, \qquad \text{for all } g \text{ and } q, \tag{182}
$$

against the alternative:

$$
H_1: \quad \phi^1_{g_i,q} \neq 0, \qquad \text{for some } g \text{ and } q, \tag{183}
$$

The hypothesis represents a generalisation of various Grouped Fixed Effects (GFE) model-based frameworks. If $y_{1,i,t-1}$ has group-specific heterogeneous coefficients,

[26] Without loss of generality, we use bivariate model in the rest of the paper. It can be extended to a m-dimensional multivariate system for analysis

then under the null hypothesis, the model reduces to a PVAR-GFE in which $y_{1,i,t-1}$ is influenced only by its own lagged values within each group. Conversely, if $y_{1,i,t-1}$ has homogeneous coefficients, the model under the null reduces to a PVAR-GFE with common coefficients across groups, as described in Bonhomme & Manresa (2015) and Chapter 2.

Theorem 9 *Let Assumption 6, 3 and 4 from Chapter 3 holds, when N and T tends to infinity and for all $\delta > 0$, $NT^{-\delta} \to 0$, we have:*

$$\sqrt{NT}(\hat{\boldsymbol{\phi}}_g - \boldsymbol{\phi}_{\boldsymbol{g}}^{\mathbf{0}}) \xrightarrow{d} \mathbb{N}\left(0, \Sigma_{\phi_g}^{-1}\Omega_{\phi_g}\Sigma_{\phi_g}^{-1}\right) \tag{184}$$

In this Theroem, we assume there exist positive definite matrices Σ_{β_g} and Ω_{β_g} such that

$$\Sigma_{\phi_g} = \text{plim}_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{1}\{g_i^0 = g\}\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0}\right)',$$

$$\Omega_{\phi_g} = \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathbb{E}\left[\mathbf{1}\{g_i^0 = g\}\mathbf{1}\{g_j^0 = g\}\boldsymbol{\epsilon}_{it}\boldsymbol{\epsilon}_{js}'\left(\mathbf{z}_{it} - \bar{\mathbf{z}}_{g_i^0 t}\right)\left(\mathbf{z}_{js} - \bar{\mathbf{z}}_{g_j^0 s}\right)'\right].$$

An advantage of the GFE-PVAR framework is that its bias is centred around zero by assuming away Nickell Bias. Consequently, there is no need to apply bias correction using procedures such as the Half-Panel Jackknife (HPJ), as proposed by Dhaene & Jochmans (2015) and Juodis et al. (2021). Accordingly, we propose a conventional Wald test procedure for the PVAR-GFE model with heterogeneous coefficients:

Corollary 4 *Let Assumption 6, 3 and 4 from Chapter 3 holds, when N and T tends to infinity and for all $\delta > 0$, $NT^{-\delta} \to 0$, we have:*

$$\hat{W}_{GFE} = NT\hat{\boldsymbol{\phi}}_g'\left({\Sigma_{\phi_g}}^{-1}\hat{\Omega}_{\phi_g}{\Sigma_{\phi_g}}^{-1}\right)^{-1}\hat{\boldsymbol{\phi}}_g \xrightarrow{d} \chi^2(Q), \tag{185}$$

In this model, one important thing we need to solve is that the test should be performed under the correct number of groups estimated by the iterative BIC procedure proposed in Chapter 3, The procedure begins by holding the number of lags P constant and evaluating BIC values across varying numbers of groups G. The Minimised value of the BIC is retained. Then process is iterated over different values of P, allowing us to jointly determine the optimal pair $(\hat{G}, \hat{P})$ via:

$$(\hat{G}, \hat{P}) = \underset{G\in\{1,\ldots,G_{\max}\}, P\in\{1,\ldots,P_{\max}\}}{\text{argmin}} BIC(G, P) \tag{186}$$

While the proposed PVAR-GFE model offers substantial flexibility in capturing unobserved heterogeneity through latent group structures. We do acknowledge the challenges of computational burdens. It could be a highly computationally intensive procedure, particularly in settings with large N, T, or when $G_{\max}$ and $P_{\max}$ are large. [27].

4.3 Monte Carlo Simulations: Case with 2 Groups

We assume the data generating process from a bivariate PVAR-GFE(1) model:

$$\begin{aligned} y_{1,i,t} &= \theta^1_{g_i,1} y_{1,i,t-1} + \phi^1_{g_i,1} y_{2,i,t-1} + \alpha_{1,g_i,t} + \epsilon_{1,i,t} \\ y_{2,i,t} &= \theta^2_{g_i,1} y_{1,i,t-1} + \phi^2_{g_i,1} y_{2,i,t-1} + \alpha_{2,g_i,t} + \epsilon_{2,i,t} \end{aligned} \tag{187}$$

Where

g	$\theta^1_{g_i,1}$	$\theta^2_{g_i,1}$	$\phi^2_{g_i,1}$
1	0.6	0.5	0.4
2	0.3	0.2	0.6

(188)

In this case, we set up the number of cross-sectional units and time periods as:

$$N = \{25; 50; 100; 200\}; \quad T = \{25; 50; 100; 200\}. \tag{189}$$

In the simple case, we set the number of group in the DGP is 2 and we split the sample into even half. We set up $\phi^1_{g_i,1}$ as:

$$\begin{aligned} &\text{Case 1: } \phi^1_{1,1} = \phi^1_{2,1} = 0 && \text{for g = 1, 2} \\ &\text{Case 2: } \phi^1_{1,1} = 0.02; \phi^1_{2,1} = 0.05 && \text{for g = 1, 2} \\ &\text{Case 3: } \phi^1_{1,1} = 0.07; \phi^1_{2,1} = 0.1 && \text{for g = 1, 2} \end{aligned} \tag{190}$$

If we setup $\phi^1_{g_i,1}$ homogeneously equals zero (refer to the case 1 in Equation 190). Then the rejection rate of the test is equivalent to the size of the test. The power of the test refers to the rejection rate of case 2 and 3 in Equation 190. In all cases,

[27] For simplicity, we perform the Monte Carlo simulations under the assumption that the number of groups are known in advance. The number of group memberships is still unknown in the simulations. As seen in Chapter 3, Table 14, BIC demonstrates excellent performance in identifying the correct number of groups when the number of groups is small (e.g., $G = 2$ or $G = 3$), achieving nearly perfect selection rates across all settings.

we set the number of simulations as 2,000 times.

To see if idiosyncratic terms could distort the test, we perform the Monte Carlo simulation under both homoskedasticity and heteroskedasticity [28]. The errors follow settings from Ando & Bai (2016):

$$\begin{cases} \text{Homoskedasticity} & \text{Var}(\varepsilon_{1,i,t}) = \text{Var}(\varepsilon_{2,i,t}) = 0.09 \\ \text{Heteroskedasticity} & \boldsymbol{\varepsilon}_{it} = 0.9\boldsymbol{e}_{it}^{1} + \delta_t 0.9\boldsymbol{e}_{it}^{2} \end{cases} \tag{191}$$

Where $\delta_t = 1$ if t is an odd number while $\delta_t = 0$ if t is an even number.

To assess the performance and robustness of the proposed Wald-type test under group-specific heterogeneity, we compare it against the Half-Panel Jackknife (HPJ)-based test developed by Juodis et al. (2021). While our method explicitly accounts for latent group structures in both the slope coefficients and fixed effects, the HPJ approach assumes individual-specific heterogeneity and corrects for the Nickell bias using a jackknife procedure. This comparison serves to evaluate how the HPJ test behaves when its assumptions are violatedspecifically, when the true data-generating process (DGP) follows a grouped structure rather than individual-level heterogeneity. We conduct Monte Carlo simulations using the same grouped DGP and examine the empirical size and power of both tests under the null and alternative hypotheses. By doing so, we aim to highlight the relative advantages and limitations of each test in finite samples, particularly when the model is misspecified in terms of slope and fixed-effect heterogeneity. The HPJ test is given by:

$$\hat{W}_{HPJ} = NT\tilde{\boldsymbol{\beta}}' \left(\hat{\boldsymbol{J}}^{-1}\hat{\boldsymbol{V}}\hat{\boldsymbol{J}}^{-1}\right)^{-1} \tilde{\boldsymbol{\beta}} \tag{192}$$

Where $\tilde{\boldsymbol{\beta}}$ is the HPJ estimator.

[28] Autocorrelation is not considered in the simulation, we assume that the first lagged dependent variable could fully capture the autocorrelation hence it does not violate the assumption of the model itself.

Table 22: Monte Carlo Simulations Table for $G = 2$

N	T	Homoskedasticity Errors			Heteroskedasticity Errors		
		Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
25	25	13.7	5.3	55.7	13.8	5.4	56.3
	50	9.2	8.0	65.5	9.0	7.9	67.2
	100	7.8	24.4	90.3	7.7	25.6	89.9
	200	6.0	33.4	97.0	6.0	32.5	97.3
50	25	11.9	6.7	69.6	11.7	6.4	70.1
	50	8.4	15.3	78.6	8.7	15.0	80.1
	100	7.0	27.3	92.1	7.2	28.8	93.7
	200	5.8	35.4	97.9	5.5	36.2	98.1
100	25	10.5	23.7	77.2	10.3	21.8	77.9
	50	9.6	30.5	82.2	9.1	21.3	86.0
	100	7.0	42.6	95.3	6.6	40.6	97.7
	200	5.6	58.3	100.0	5.8	55.7	100.0
200	25	16.2	25.8	75.2	10.0	27.5	77.4
	50	9.0	36.8	95.2	8.8	37.5	96.1
	100	7.1	66.4	99.8	6.9	65.2	99.9
	200	5.2	89.6	100.0	5.1	83.1	100.0

Table 23: Monte Carlo Simulations Table for $G = 2$ by HPJ Test

N	T	Homoskedasticity Errors			Heteroskedasticity Errors		
		Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
25	25	39.9	77.1	96.3	57.0	90.3	99.1
	50	22.1	72.2	98.6	32.4	88.0	99.7
	100	12.8	65.0	99.6	17.4	83.2	100.0
	200	10.3	69.1	100.0	10.0	84.3	100.0
50	25	52.5	95.0	100.0	73.7	98.9	100.0
	50	23.5	92.5	100.0	41.4	98.7	100.0
	100	11.5	91.1	100.0	18.1	98.5	100.0
	200	7.4	94.8	100.0	8.7	98.8	100.0
100	25	71.2	99.8	100.0	90.8	100.0	100.0
	50	35.1	99.8	100.0	60.3	100.0	100.0
	100	11.8	99.7	100.0	21.3	100.0	100.0
	200	7.4	99.9	100.0	7.6	100.0	100.0
200	25	92.0	100.0	100.0	99.2	100.0	100.0
	50	51.5	100.0	100.0	84.0	100.0	100.0
	100	15.9	100.0	100.0	31.4	100.0	100.0
	200	7.9	100.0	100.0	9.9	100.0	100.0

Table 22 and Table 23 reports the Monte Carlo simulation results for the proposed Wald-type test under both homoskedastic and heteroskedastic error structures, across varying combinations of cross-sectional units (N) and time periods (T) by grouped heterogeneity and individual heterogeneity, respectively. Case 1 reflects the empirical size of the test, that is, the probability of incorrectly rejecting the null hypothesis when it is true. Cases 2 and 3 correspond to the power of the test under two degrees of deviation from the null hypothesis, with Case 3 representing a stronger signal than Case 2.

In Table 22, the test demonstrates good size control: rejection rates in Case 1 remain close to the nominal 5% level across most sample sizes, with modest inflation in smaller panels (e.g., $N = 25$, $T = 25$), and convergence toward correct size as N and T grow. Importantly, these results are nearly identical under both homoskedastic and heteroskedastic specifications, indicating robustness of the test to variance heterogeneity. In terms of power (Cases 2 and 3), the Wald test performs increasingly well as either dimension of the panel expands. For moderate deviations from the null (Case 2), power improves substantially with larger N and T, reaching above 50% in many configurations and exceeding 85% in large panels (e.g., $N = 200$, $T = 200$). Under stronger alternatives (Case 3), the test exhibits excellent performance, achieving power near or equal to 100% in panels of moderate to large size. Notably, the power remains consistently high even under heteroskedastic error terms, further confirming the robustness of the test in more realistic settings. These results validate the theoretical properties of the test and suggest it is well suited for detecting Granger causality in panel data with grouped heterogeneity.

In contrast, Table 23 presents the results from the HPJ-based test of Juodis et al. (2021), applied to the same data-generating process featuring grouped heterogeneity. The findings reveal significant size distortions under the null hypothesis (Case 1), particularly for small and moderate panels. For instance, when $N = 25$ and $T = 25$, the empirical size exceeds 39% under homoskedasticity and rises above 57% under heteroskedasticity. Even with increasing T, the test continues to over-reject under the null, with size remaining well above the nominal 5% level across most scenarios. The test incorrectly attributes the structured group variation to spurious individual effects, inflating the rejection rate. Furthermore, the observed power in Cases 2 and 3 is uniformly high, often exceeding 90% or reaching 100% even in small samples. However, this apparent power is misleading, as it coincides with the severe size distortion. That is, the test tends to reject the null hypothesis indiscriminately, regardless of whether it is true or false. This phenomenon indicates that the test statistic is driven more by systematic bias due to model mis-specification than by genuine signal detection. Therefore, the HPJ test fails to maintain validity when applied to data with grouped heterogeneity panels.

4.4 Monte Carlo Simulations: Case with 4 Groups

We assume the data generating process from a bivariate PVAR-GFE(1) model:

$$\begin{aligned} y_{1,i,t} &= \theta^1_{g_i,1} y_{1,i,t-1} + \phi^1_{g_i,1} y_{2,i,t-1} + \alpha_{1,g_i,t} + \epsilon_{1,i,t} \\ y_{2,i,t} &= \theta^2_{g_i,1} y_{1,i,t-1} + \phi^2_{g_i,1} y_{2,i,t-1} + \alpha_{2,g_i,t} + \epsilon_{2,i,t} \end{aligned} \tag{193}$$

Where

g	$\theta^1_{g_i,1}$	$\theta^2_{g_i,1}$	$\phi^2_{g_i,1}$
1	0.6	0.5	0.4
2	0.3	0.2	0.6
3	0.2	0.5	0.8
4	0.4	0.6	0.5

(194)

In this case, we set up the number of cross-sectional units and time periods as:

$$N = \{25; 50; 100; 200\}; \quad T = \{25; 50; 100; 200\}. \tag{195}$$

In the simple case, we set the number of group in the DGP is 4 and we split the sample into evenly sample. We set up $\phi^1_{g_i,1}$ as:

$$\begin{aligned} &\text{Case 1: } \phi^1_{1,1} = \phi^1_{2,1} = \phi^1_{3,1} = \phi^1_{4,1} = 0 && \text{for g = 1, 2, 3, 4} \\ &\text{Case 2: } \phi^1_{1,1} = 0.02; \phi^1_{2,1} = 0.05; \phi^1_{3,1} = 0.03; \phi^1_{4,1} = 0.04 && \text{for g = 1, 2, 3, 4} \\ &\text{Case 3: } \phi^1_{1,1} = 0.07; \phi^1_{2,1} = 0.1; \phi^1_{3,1} = 0.13; \phi^1_{4,1} = 0.16 && \text{for g = 1, 2, 3, 4} \end{aligned} \tag{196}$$

If we setup $\phi^1_{g_i,1}$ homogeneously equals zero (refer to the case 1 in Equation 196). Then the rejection rate of the test is equivalent to the size of the test. The power of the test refers to the rejection rate of case 2 and 3 in Equation 196. Particularly, Case 2 mimics cases where some units respond to the predictor while others do not respond to the dependent variable. In all cases, we set the number of simulations as 2,000 times.

To see if idiosyncratic terms could distort the test, we perform the Monte Carlo simulation under both homoskedasticity and heteroskedasticity [29]. The errors follow settings from Ando & Bai (2016):

$$\begin{cases} \text{Homoskedasticity} & \operatorname{Var}(\varepsilon_{1,i,t}) = \operatorname{Var}(\varepsilon_{2,i,t}) = 0.09 \\ \text{Heteroskedasticity} & \boldsymbol{\varepsilon}_{it} = 0.9\boldsymbol{e}_{it}^{1} + \delta_t 0.9\boldsymbol{e}_{it}^{2} \end{cases} \tag{197}$$

Where $\delta_t = 1$ if t is an odd number while $\delta_t = 0$ if t is an even number.

[29] Autocorrelation is not considered in the simulation, we assume that the first lagged dependent variable could fully capture the autocorrelation hence it does not violate the assumption of the model itself.

Table 24: Monte Carlo Simulations Table for $G = 4$

N	T	Homoskedasticity Errors			Heteroskedasticity Errors		
		Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
25	25	14.5	4.8	52.6	14.7	5.1	54.2
	50	9.6	7.1	62.9	9.0	7.0	64.5
	100	7.6	20.1	89.2	7.5	21.5	88.7
	200	6.3	30.7	96.5	6.0	29.9	96.8
50	25	12.7	6.0	66.3	12.5	5.8	67.8
	50	8.6	8.1	76.1	8.5	7.9	77.3
	100	7.0	22.5	91.3	7.1	24.9	92.8
	200	5.6	33.9	97.5	5.5	34.1	97.8
100	25	10.2	16.5	75.4	10.1	22.7	76.3
	50	9.1	19.5	81.0	9.0	26.1	83.9
	100	6.5	36.4	94.9	6.6	40.7	97.0
	200	5.3	51.8	99.9	5.4	55.6	100.0
200	25	10.0	23.5	83.7	9.9	24.9	81.6
	50	8.9	31.2	90.7	8.8	34.5	94.1
	100	6.1	58.1	99.7	6.0	61.7	99.9
	200	5.1	82.3	100.0	5.2	86.5	100.0

Table 25: Monte Carlo Simulations Table for $G = 4$ by HPJ Test

N	T	Homoskedasticity Errors			Heteroskedasticity Errors		
		Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
25	25	80.7	98.5	100.0	86.7	98.3	100.0
	50	69.5	99.8	100.0	78.8	99.6	100.0
	100	47.1	100.0	100.0	58.9	99.7	100.0
	200	27.3	100.0	100.0	39.7	99.9	100.0
50	25	93.7	100.0	100.0	95.9	100.0	100.0
	50	83.1	100.0	100.0	88.8	100.0	100.0
	100	55.0	100.0	100.0	69.8	100.0	100.0
	200	30.9	100.0	100.0	42.5	100.0	100.0
100	25	99.3	100.0	100.0	99.7	100.0	100.0
	50	85.1	100.0	100.0	98.0	100.0	100.0
	100	72.0	100.0	100.0	85.1	100.0	100.0
	200	36.7	100.0	100.0	51.9	100.0	100.0
200	25	100.0	100.0	100.0	100.0	100.0	100.0
	50	99.4	100.0	100.0	100.0	100.0	100.0
	100	89.1	100.0	100.0	95.9	100.0	100.0
	200	47.2	100.0	100.0	70.1	100.0	100.0

Table 24 and Table 25 presents the Monte Carlo simulation results for the proposed Wald-type test under the presence of four latent groups ($G = 4$) by grouped heterogeneity and individual heterogeneity. The table reports the empirical rejection rates across various panel sizes, under both homoskedastic and heteroskedastic error structures. As before, Case 1 represents the null hypothesis where no Granger causality is present ($\phi^1_{g_i,1} = 0$ for all groups), while Cases 2 and 3 correspond to alternatives with moderate and strong Granger-causal effects, respectively. In Case 2, only a subset of groups (two out of four) exhibit nonzero values of $\phi^1_{g_i,1}$, introducing partial heterogeneity under the alternative. In Case 3, all groups are assumed to exhibit nonzero and increasingly strong values of $\phi^1_{g_i,1}$, representing a fully heterogeneous and high-power setting.

Regarding test size (Case 1) of Table 24, the results indicate that the Wald-type test is reasonably well-sized across the range of N and T values. For small panels ($N = 25$, $T = 25$), the test tends to over-reject the null hypothesis, with empirical sizes reaching up to approximately 14.5%. However, this over-rejection diminishes as either the cross-sectional dimension or the time dimension increases. In larger panels ($N \geq 100$, $T \geq 100$), the rejection rates approach the nominal 5% level, indicating appropriate finite-sample performance. Notably, the size performance is consistent across homoskedastic and heteroskedastic error structures, suggesting that the test is robust to the presence of time-varying variance. The power results (Cases 2 and 3) further demonstrate the efficacy of the test under alternatives. In Case 2, where only half of the groups violate the null, the power increases steadily with larger N and T. For example, the power rises from roughly 4.8% in the smallest setting ($N = 25$, $T = 25$) to over 80% when $N = 200$ and $T = 200$, under both error specifications. The heteroskedasticity case even shows slightly higher rejection rates, likely due to the added variability accentuating the signal. In Case 3, where all four groups deviate from the null, the test demonstrates excellent power. Rejection rates exceed 90% for most configurations with $T \geq 100$, and reach or approximate 100% for large panels. These results confirm the theoretical expectation that the Wald-type test becomes highly reliable in detecting Granger-causal relationships when the signal is sufficiently strong and widespread across groups.

Table 25 reports the HPJ-based test results using the same $G = 4$ grouped DGP. Across all cases, the test also exhibits severe size distortions under the null (Case 1), with empirical rejection rates far exceeding the nominal level, especially in smaller panels. This over-rejection persists even as T increases, highlighting the incompatibility of the HPJ test with grouped slope heterogeneity. While the test shows near-perfect power in Cases 2 and 3, this should be interpreted cautiously, as much of the power reflects inflated size rather than true discriminatory ability. Consequently, the high power of the HPJ test in this context does not reflect genuine discriminative capacity but rather its tendency to reject excessively regardless of the underlying hypothesis.

4.5 Conclusion

In this chapter, we introduced a Wald-type test for Granger non-causality in grouped panels with heterogeneous coefficients and grouped fixed effects. By allowing for group-specific heterogeneity, the proposed test addresses the limitations of traditional Granger causality tests, which often assume homogeneity across units. Through theoretical development and Monte Carlo simulations, we demonstrate that the PVAR-GFE model provides a robust and flexible framework for testing causal relationships in panels. Unlike methods that require explicit bias correction, such as the Half Panel Jackknife (HPJ) procedure; the PVAR-GFE approach mitigates bias (including Nickell-type bias) through its group-structured formulation and the use of large-T asymptotics, thereby avoiding the need for post-estimation correction. The Monte Carlo results showed that the test is well-sized and maintains high power across different sample sizes and time periods, even under heteroskedastic error structures and different group structures.

Chapter 5: Thesis Conclusion

5.1 Summary of Key Findings

This thesis provides new insights into the role of grouped heterogeneity in panel data econometrics, focusing on three key areas: the relationship between economic growth and income inequality, the development of a multivariate dynamic model, and the testing of causal relationships in grouped panels. Each Chapter addresses a specific aspect of this broader theme.

5.1.1 Key Findings in Chapter 2: Grouped Patterns in Economic Growth and Income Inequality

This chapter revisits the long-debated relationship between economic growth and income inequality by explicitly modelling grouped heterogeneity. Using the Grouped Fixed Effects (GFE) estimator on a panel of 92 countries from 1998 to 2017, the study uncovers significant cross-country heterogeneity consistent with the Kuznets curve hypothesis, revealing a non-linear, inverted-U shaped relationship between inequality and growth. Importantly, this relationship varies across latent groups, providing robust empirical evidence that the inequality-growth nexus is conditional on group-specific development stages. It shows that the GFE method identifies latent clusters of countries not based on conventional income or regional classifications, but on shared economic behaviours. Countries in Group 1, comprising many advanced and large emerging economies, exhibit a moderate, statistically significant non-linear relationship, whereas countries in Group 4 (primarily transitional economies) demonstrate that inequality becomes harmful to growth at relatively lower levels. These results remain robust across alternative model specifications and grouping algorithms.

5.1.2 Key Findings in Chapter 3: Panel VAR Model with Latent Group Structures

In this study, we propose a Panel Vector Autoregression model with Grouped Fixed Effects (PVAR-GFE) that addresses key limitations of existing PVAR frameworks by allowing for latent, group-specific heterogeneity in both intercepts and dynamic coefficients. The model offers a computationally efficient and parsimonious structure, avoiding the over-parameterisation issues commonly associated with interactive fixed effects. It captures both the dynamic interdependencies among endogenous variables and the cross-sectional grouped heterogeneity present in panel data. Unlike traditional PVAR models, which often ignore unobserved heterogeneity or cross-sectional dependence, the PVAR-GFE framework flexibly identifies latent group structures without requiring them to be pre-specified. Monte Carlo simulations confirm that the estimator performs well in large samples, exhibiting negligible group misspecification and decreasing bias as N and T increase. This methodological advancement provides a robust tool for researchers and policymakers to analyse heterogeneous dynamic responses in multivariate panels. It is especially valuable in macroeconomic applications, where regional or institutional differences are often unobserved but crucially shape group structures and dynamic behaviours.

5.1.3 Chapter 4: Testing for Granger Causality in Grouped Panels

In Chapter 3, the thesis proposes a Wald-type test for Granger non-causality in grouped panels, addressing limitations in standard tests that assume homogeneity across units. The testing framework allows for group-specific heterogeneous coefficients and grouped fixed effects, making it suitable for panels where cross-sectional units may belong to different latent groups. Monte Carlo simulations under various conditions confirm that the test is both size-accurate and powerful, even in the presence of heteroskedasticity.

This chapter advances the literature on panel Granger causality testing by providing a robust test that accommodates grouped heterogeneity without additional bias correction. The test also doubles as a homogeneity test, offering a diagnostic for the presence of group-specific dynamics. The proposed method improves the reliability of causal inference in panels where unobserved heterogeneity plays a significant role, such as in regional economic analyses or sectoral studies.

5.2 Future Research Directions

While this thesis has made significant contributions to the understanding of grouped heterogeneity in panel data models and has introduced new methods for dynamic modelling and causality testing, several avenues remain open for future research. The methods developed in this thesis can be further extended and applied in different contexts, providing opportunities to enhance both the empirical and theoretical literature in econometrics.

- Time-Varying Group Membership. One of the key assumptions in the Grouped Fixed Effects (GFE) and PVAR-GFE models is that group membership remains constant over time. However, in reality, the economic conditions of countries, firms, or individuals may change over time, leading to time-varying group memberships. Future research could extend these models by allowing for dynamic group transitions, where units may shift from one group to another as their characteristics evolve. Developing a framework for modelling such transitions would be particularly useful for understanding how countries or regions respond differently to economic shocks over time, as well as for applications in other domains where group membership is fluid.

- Machine Learning Approaches for Group Classification. The current methodology relies on algorithms like K-means clustering to determine group membership. However, advances in machine learning techniques could improve the identification of latent group structures by leveraging more sophisticated classification algorithms. Future research could explore how machine learning methods, such as hierarchical clustering, random forests, or neural networks, could be integrated into the GFE and PVAR-GFE frameworks to better capture complex grouped patterns in large and high-dimensional datasets. These methods could also be employed to estimate the optimal number of groups, which remains a challenge in many empirical applications.

- Time-Varying Coefficients in PVAR-GFE Models. The PVAR-GFE model developed in this thesis assumes that group-specific coefficients are constant over time. An important extension of this work would be to introduce time-varying coefficients, allowing for more flexibility in modeling dynamic relationships that change over time. This would be especially useful in contexts where the economic environment is rapidly evolving, such as during financial crises or major policy shifts. Incorporating time-varying coefficients would enable the model to better capture short-term fluctuations in group-specific dynamics and could provide deeper insights into the temporal evolution of economic relationships across groups.

- Generalizing to Nonlinear Models. The methods developed in this thesis, particularly the GFE and PVAR-GFE models, focus on linear relationships between variables. However, many economic and financial relationships are inherently nonlinear. Future work could extend the GFE and PVAR-GFE frameworks to nonlinear models, such as threshold models, nonlinear autoregressive models, or regime-switching models. These extensions would allow researchers to capture more complex dynamics and interactions between variables, particularly in datasets where linear models are insufficient to describe the underlying relationships. This would be an important contribution to both applied econometrics and the broader literature on dynamic modelling.

References

Abebe, H. & Ratbek, D. (2020), 'Income inequality and economic growth: Heterogeneity and nonlinearity', *Studies in Nonlinear Dynamics & Econometrics* **24**(3).

Acemoglu, D., Johnson, S., Robinson, J. & Thaicharoen, Y. (2003), 'Institutional causes, macroeconomic symptoms: volatility, crises and growth', *Journal of monetary economics* **50**(1), 49–123.

Acemoglu, D., Naidu, S., Restrepo, P. & Robinson, J. A. (2019), 'Democracy does cause growth', *Journal of political economy* **127**(1), 47–100.

Acemoglu, D. & Robinson, J. A. (2001), 'A theory of political transitions', *American Economic Review* **91**(4), 938–963.

Alesina, A. & Rodrik, D. (1994), 'Distributive politics and economic growth', *The quarterly journal of economics* **109**(2), 465–490.

Ando, T. & Bai, J. (2016), 'Panel data models with grouped factor structure under unknown group membership', *Journal of Applied Econometrics* **31**(1), 163–191.

Ando, T. & Bai, J. (2017), 'Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures', *Journal of the American Statistical Association* **112**(519), 1182–1198.

Atems, B. & Jones, J. (2015), 'Income inequality and economic growth: a panel var approach', *Empirical Economics* **48**(4), 1541–1561.

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

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

Banerjee, A. V. & Duflo, E. (2003), 'Inequality and growth: What can the data say?', *Journal of economic growth* **8**(3), 267–299.

Barro, R. J. (1991), 'Economic growth in a cross section of countries', *The quarterly journal of economics* **106**(2), 407–443.

Barro, R. J. (2000), 'Inequality and growth in a panel of countries', *Journal of economic growth* **5**(1), 5–32.

Bartak, J. & Jabłoński, Ł. (2020), 'Inequality and growth: What comes from the different inequality measures?', *Bulletin of Economic Research* **72**(2), 185–212.

Bester, C. A. & Hansen, C. B. (2013), 'Grouped effects estimators in fixed effects models', *Journal of Econometrics* **190**(1), 197–208.

Birdsall, N., Ross, D. & Sabot, R. (1995), 'Inequality and growth reconsidered: lessons from east asia', *The World Bank Economic Review* **9**(3), 477–508.

Bonhomme, S., Lamadon, T. & Manresa, E. (2022), 'Discretizing unobserved heterogeneity', *Econometrica* **90**(2), 625–643.

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

Burnside, C. (1996), 'Production function regressions, returns to scale, and externalities', *Journal of monetary Economics* **37**(2), 177–201.

Canova, F. & Ciccarelli, M. (2013), *Panel Vector Autoregressive Models: A Survey The views expressed in this article are those of the authors and do not necessarily reflect those of the ECB or the Eurosystem.*, Emerald Group Publishing Limited.

Castelló-Climent, A. (2010), 'Inequality and growth in advanced economies: an empirical investigation', *The Journal of Economic Inequality* **8**(3), 293–321.

Chen, B.-L. (2003), 'An inverted-u relationship between inequality and long-run growth', *Economics Letters* **78**(2), 205–212.

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

Cho, D., Kim, B. M. & Rhee, D.-E. (2014), 'Inequality and growth: nonlinear evidence from heterogeneous panel data', *KIEP Research Paper No. Working Papers-14-01* .

Chudik, A. & Pesaran, M. H. (2015), 'Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors', *Journal of econometrics* **188**(2), 393–420.

Cingano, F. (2014), 'Trends in income inequality and its impact on economic growth'.

De Dominicis, L., Florax, R. J. & De Groot, H. L. (2008), 'A meta-analysis on the relationship between income inequality and economic growth', *Scottish Journal of Political Economy* **55**(5), 654–682.

Deininger, K. & Squire, L. (1996), 'A new data set measuring income inequality', *The World Bank Economic Review* **10**(3), 565–591.

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

Dumitrescu, E.-I. & Hurlin, C. (2012), 'Testing for granger non-causality in heterogeneous panels', *Economic modelling* **29**(4), 1450–1460.

Fawaz, F., Rahnama, M. & Valcarcel, V. J. (2014), 'A refinement of the relationship between economic growth and income inequality', *Applied Economics* **46**(27), 3351–3361.

Feenstra, R. C., Inklaar, R. & Timmer, M. P. (2015), 'The next generation of the penn world table', *American economic review* **105**(10), 3150–82.

Forbes, K. J. (2000), 'A reassessment of the relationship between inequality and growth', *American economic review* **90**(4), 869–887.

Frank, M. W. (2009), 'Inequality and growth in the united states: Evidence from a new state-level panel of income inequality measures', *Economic Inquiry* **47**(1), 55–68.

Gómez-Puig, M., Sosvilla-Rivero, S. & Martínez-Zarzoso, I. (2022), 'On the heterogeneous link between public debt and economic growth', *Journal of International Financial Markets, Institutions and Money* **77**, 101528.

Granger, C. W. (1969), 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica: journal of the Econometric Society* pp. 424–438.

Grigoli, F., Paredes, E. & Di Bella, G. (2016), *Inequality and growth: A heterogeneous approach*, International Monetary Fund.

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

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

Halter, D., Oechslin, M. & Zweimüller, J. (2014), 'Inequality and growth: the neglected time dimension', *Journal of economic growth* **19**(1), 81–104.

Herzer, D. & Vollmer, S. (2012), 'Inequality and growth: evidence from panel cointegration', *The Journal of Economic Inequality* **10**(4), 489–503.

Holtz-Eakin, D., Newey, W. & Rosen, H. S. (1988), 'Estimating vector autoregressions with panel data', *Econometrica: Journal of the econometric society* pp. 1371–1395.

Hsiao, C. (2014), *Analysis of panel data*, number 54, Cambridge university press.

Hsiao, C. & Pesaran, M. H. (2004), 'Random coefficient panel data models'.

Huang, W., Jin, S. & Su, L. (2020), 'Identifying latent grouped patterns in cointegrated panels', *Econometric Theory* **36**(3), 410–456.

Jordà, Ò. (2005), 'Estimation and inference of impulse responses by local projections', *American economic review* **95**(1), 161–182.

Juodis, A., Karavias, Y. & Sarafidis, V. (2021), 'A homogeneous approach to testing for granger non-causality in heterogeneous panels', *Empirical economics* **60**(1), 93–112.

Kennedy, T., Smyth, R., Valadkhani, A. & Chen, G. (2017), 'Does income inequality hinder economic growth? new evidence using australian taxation statistics', *Economic Modelling* **65**, 119–128.

Khalifa, S., El Hag, S. et al. (2010), 'Income disparities, economic growth and development as a threshold', *Journal of Economic Development* **35**(2), 23.

King, R. G. & Levine, R. (1993), *Financial intermediation and economic development*, Vol. 15689, Cambridge: Cambridge University Press.

Kornai, J. (1994), 'Transformational recession: the main causes', *Journal of comparative economics* **19**(1), 39–63.

Kuznets, S. (1955), 'Economic growth and income inequality', *The American economic review* **45**(1), 1–28.

Lee, D. J. & Son, J. C. (2016), 'Economic growth and income inequality: evidence from dynamic panel investigation', *Global Economic Review* **45**(4), 331–358.

Lee, K., Pesaran, M. H. & Smith, R. (1997), 'Growth and convergence in a multi-country empirical stochastic solow model', *Journal of applied Econometrics* **12**(4), 357–392.

Leng, X., Chen, H. & Wang, W. (2021), 'Multi-dimensional latent group structures with heterogeneous distributions', *Journal of Econometrics* .

Levine, R. (1997), 'Financial development and economic growth: views and agenda', *Journal of economic literature* **35**(2), 688–726.

Li, D., Plagborg-Møller, M. & Wolf, C. K. (2022), Local projections vs. vars: Lessons from thousands of dgps, Technical report, National Bureau of Economic Research.

Li, H. & Zou, H.-f. (1998), 'Income inequality is not harmful for growth: theory and evidence', *Review of development economics* **2**(3), 318–334.

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.

Litschig, S. & Lombardi, M. (2019), 'Which tail matters? inequality and growth in brazil', *Journal of Economic Growth* **24**(2), 155–187.

Lumsdaine, R. L., Okui, R. & Wang, W. (2020), 'Estimation of panel group structure models with structural breaks in group memberships and coefficients', *Available at SSRN 3617416* .

Mei, Z., Sheng, L. & Shi, Z. (2023), 'Implicit nickell bias in panel local projection', *arXiv preprint arXiv:2302.13455* .

Miao, K., Li, K. & Su, L. (2020), 'Panel threshold models with interactive fixed effects', *Journal of Econometrics* **219**(1), 137–170.

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

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

Mugnier, M. (2025), 'A simple and computationally trivial estimator for grouped fixed effects models', *Journal of Econometrics* **250**, 106011.

Ncube, M., Anyanwu, J. C. & Hausken, K. (2014), 'Inequality, economic growth and poverty in the middle east and north africa (mena)', *African Development Review* **26**(3), 435–453.

Neves, P. C., Afonso, Ó. & Silva, S. T. (2016), 'A meta-analytic reassessment of the effects of inequality on growth', *World Development* **78**, 386–400.

Newey, W. K. & McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of econometrics* **4**, 2111–2245.

Nickell, S. (1981), 'Biases in dynamic models with fixed effects', *Econometrica: Journal of the econometric society* pp. 1417–1426.

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

Ouyang, Y. & Li, P. (2018), 'On the nexus of financial development, economic growth, and energy consumption in china: New perspective from a gmm panel var approach', *Energy economics* **71**, 238–252.

Perotti, R. (1996), 'Growth, income distribution, and democracy: What the data say', *Journal of Economic growth* **1**(2), 149–187.

Persson, T. & Tabellini, G. (1991), Is inequality harmful for growth? theory and evidence, Technical report, National Bureau of Economic Research.

Pesaran, M. H. & Yamagata, T. (2008), 'Testing slope homogeneity in large panels', *Journal of econometrics* **142**(1), 50–93.

Plagborg-Møller, M. & Wolf, C. K. (2021), 'Local projections and vars estimate the same impulse responses', *Econometrica* **89**(2), 955–980.

Rio, E. (1999), *Théorie asymptotique des processus aléatoires faiblement dépendants*, Vol. 31, Springer Science & Business Media.

Rodrik, D. (2000), 'Institutions for high-quality growth: what they are and how to acquire them', *Studies in comparative international development* **35**(3), 3–31.

Romer, C. D. & Romer, D. H. (2017), 'New evidence on the aftermath of financial crises in advanced countries', *American Economic Review* **107**(10), 3072–3118.

Sadorsky, P. (2010), 'The impact of financial development on energy consumption in emerging economies', *Energy policy* **38**(5), 2528–2535.

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

Sbaouelgi, J. et al. (2018), 'Income inequality and economic growth: Application of quantile regression', *Asian Development Policy Review* **6**(1), 1–14.

Shahbaz, M., Khan, S. & Tahir, M. I. (2013), 'The dynamic links between energy consumption, economic growth, financial development and trade in china: fresh evidence from multivariate framework analysis', *Energy economics* **40**, 8–21.

Solt, F. (2020), 'Measuring income inequality across countries and over time: The standardized world income inequality database', *Social Science Quarterly* **101**(3), 1183–1199.

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

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

Tuğan, M. (2021), 'Panel var models with interactive fixed effects', *The Econometrics Journal* **24**(2), 225–246.

Vo, D. H., Nguyen, T. C., Tran, N. P. et al. (2019), 'What factors affect income inequality and economic growth in middle-income countries?', *Journal of Risk and Financial Management* **12**(1), 40.

Voitchovsky, S. (2005), 'Does the profile of income inequality matter for economic growth?', *Journal of Economic growth* **10**(3), 273–296.

Yang, Y. & Greaney, T. M. (2017), 'Economic growth and income inequality in the asia-pacific region: A comparative study of china, japan, south korea, and the united states', *Journal of Asian Economics* **48**, 6–22.

Zhang, Y.-J. (2011), 'The impact of financial development on carbon emissions: An empirical analysis in china', *Energy policy* **39**(4), 2197–2203.

Zhang, Y., Wang, H. J. & Zhu, Z. (2019), 'Quantile-regression-based clustering for panel data', *Journal of Econometrics* **213**(1), 54–67.