

4-26-2018

Identification and Estimation in Panel Models with Overspecified Number of Groups

Ruiqi Liu

Binghamton University--SUNY, rliu14@binghamton.edu

Follow this and additional works at: https://orb.binghamton.edu/dissertation_and_theses



Part of the [Statistics and Probability Commons](#)

Recommended Citation

Liu, Ruiqi, "Identification and Estimation in Panel Models with Overspecified Number of Groups" (2018). *Graduate Dissertations and Theses*. 96.

https://orb.binghamton.edu/dissertation_and_theses/96

This Dissertation is brought to you for free and open access by the Dissertations, Theses and Capstones at The Open Repository @ Binghamton (The ORB). It has been accepted for inclusion in Graduate Dissertations and Theses by an authorized administrator of The Open Repository @ Binghamton (The ORB). For more information, please contact ORB@binghamton.edu.

Identification and estimation in panel models with overspecified number of groups

BY

RUIQI LIU

BS, Sun Yat-sen University, 2013

MA, Binghamton University, 2015

DISSERTATION

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Mathematical Sciences
in the Graduate School of
Binghamton University
State University of New York
2018

Accepted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Mathematical Sciences
in the Graduate School of
Binghamton University
State University of New York
2018

April 26th, 2018

Anton Schick, Chair and Faculty Advisor
Department of Mathematical Sciences, Binghamton University

Qiqing Yu, Member
Department of Mathematical Sciences, Binghamton University

Xingye Qiao, Member
Department of Mathematical Sciences, Binghamton University

Solomon Polachek, Outside Examiner
Department of Economics, Binghamton University

Abstract

In this thesis, we provide a simple approach to identify and estimate group structure in panel models by adapting the M-estimation method. We consider both linear and nonlinear panel models where the regression coefficients are heterogeneous across groups but homogeneous within a group and the group membership is unknown to researchers. The main result of the thesis is that under certain assumptions, our approach is able to provide uniformly consistent group parameter estimator as long as the number of groups used in estimation is not smaller than the true number of groups. We also show that, with probability approaching one, our method can partition some true groups into further subgroups, but cannot mix individuals from different groups. When the true number of groups is used in estimation, all the individuals can be categorized correctly with probability approaching one, and we establish the limiting distribution for the estimates of the group parameters. In addition, we provide an information criterion to choose the number of group and established its consistency under some mild conditions. Monte Carlo simulations are conducted to examine the finite sample performance of our proposed method. Findings in the simulation confirm our theoretical results in the paper. Application to labor force participation also highlights the necessity to take into account of individual heterogeneity and group heterogeneity.

Acknowledgement

Firstly, I would like to express my special appreciation and thanks to my advisor Professor Anton Schick, he has been a tremendous mentor for me. Dr. Schick not only taught me how to conduct mathematical proof, but also helped to make mathematics fun for me. Secondly, I would like to thank my co-advisor Dr. Zuofeng Shang for the support of my research, for his patience, motivation, and immense knowledge. Next, I would like to appreciate Professor Qiqing Yu for his kindness and help. Without him, I may not be able to become a Ph.D student in Department of Mathematical Sciences. I also would like to thank other committee members, Dr Xingye Qiao and Professor Solomon Polachek for their suggestions and comments. Finally, I would like to say Thank You to my Parents. I do not think I can complete my Ph.D program without their support.

Contents

1	Introduction	1
2	Methodology	4
2.1	Estimation	4
2.2	Numerical Algorithm	6
3	Asymptotic Theory	7
3.1	Estimation Consistency	7
3.2	Detection of Group Structure among Individuals	10
3.3	Determination of Number of Groups	12
3.4	Asymptotic Normality	13
4	Numerical Performance	16
4.1	Simulations	16
4.2	Empirical Application	20
5	Technical Proofs	23
5.1	Proof of Theorem 3.1	23
5.2	Proof of Theorem 3.2	31
5.3	Proof of Theorem 3.3	34
5.4	Proof of Theorem 3.4	40
5.5	Proof of Theorem 3.6	44
	Bibliography	57

Chapter 1

Introduction

Panel data models are widely used in empirical research of both economics and finance. An important feature to use panel data is that it allows researchers to control individual-level heterogeneity. Unfortunately, most of this heterogeneity, however, is unobservable, e.g., willingness to pay for education, impact of economic policy, personal innate ability, etc. In practice, there are two opposite approaches to deal with this individual level heterogeneity. The first one is to completely ignore the heterogeneity among individuals by assuming common parameters across individuals, see, e.g., [18], [12], [2]. Indeed, this approach reduces the model complexity and facilitates statistical inference. However, this common parameters assumption might be too strong in practice and may lead to model misspecification: see, e.g., [13]. Moreover, this assumption has also been found to be too restrictive in many empirical studies, see, for example, [15] and [19], among others. The other approach is to allow cross-sectional slope heterogeneity, e.g., [14], [4]. This assumption helps avoid the misspecification problem; however, it may lose latent connections between individuals and efficiency of estimation. To be more specific, if part of the individuals share a common parameter, it sacrifices this essential connection and leads to estimators with larger variance.

To allow such a possibility that part of the individuals shares a common parameter, a mild and reasonable assumption is to impose group structure in panels. Group structure in panels refers to the regression parameters that are the same within each group but differ across groups. Recently, group structure in panels has received lots of attention in the literature both empirically and theoretically. To name a few, for linear model,

[21] consider linear panel model with group structure on both intercept and slope. When there are only two groups and one regressor, they propose a threshold based estimation method to identify the latent group structure and show that the estimator is consistent. Under the same setup of [21], [23] propose a modified k-means algorithm to determine the number of clusters and estimate parameters. [6] consider the linear panel data models with a latent group structure on the time-varying individual-specific effects and propose a group fixed effects estimator. The work of [6] has been extended to models with interactive fixed effects and nonlinear panel models by [1] and [5], respectively. More recently, [24] propose a classifier Lasso (C-Lasso) penalized procedure to identify and estimate panels with group structure.

Following the work of [21] and [24], this thesis proposes a simple and straightforward method to identify and estimate panels with group structure when the true number of groups and the membership are unknown. The method we propose can be applied to both linear and nonlinear panels. Besides the simplicity, the proposed method has several advantages as follows.

First, the major theoretical contribution of this paper is that we show, under certain assumptions, the consistency of our proposed estimation is independent of the number of groups used as long as this number is not underestimated. The important practical implication of this result is that for estimation of the regression coefficients, one does not necessarily need to estimate the number of groups correctly as long as this number is not underestimated. The implication of this result is that a safe way in estimating the panel model with an unknown group structure is to set a slight large number of groups. This is of crucial importance to researchers since generally speaking, the number of groups in the data is usually unknown. We also show that, with probability approaching one, our method can partition some true group into further subgroups, but cannot misclassify individuals from different groups into the same group. When the true number of groups is used in estimation, all the individuals can be categorized correctly with probability approaching one.

Second, once the group membership is correctly identified and estimated, our proposed estimation performs similarly to the estimation based on true (or oracle) group membership. This oracle property allows one to combine existing estimation and in-

ference technique with our method, for instance, for the classified group units, one can adapt the jackknife method in [12] or [9] to reduce the bias for fixed effects estimation in both linear and nonlinear panels.

Finally, unlike the C-Lasso approach proposed by [24], which relies on the choice of tuning parameter, our approach is penalty free if the number of groups is correctly specified, which is a significant advantage for empirical application. It is well known in the literature that Lasso type methods are able to consistently select variables. However, the consistency of variable selection highly depends on the right choice of the tuning parameter (e.g., [8] and [17]). Therefore, in empirical applications, the estimation results may be sensitive to the choice of tuning parameters, and how to choose the optimal tuning parameter in C-Lasso is still an open question. Consequently, it would be convenient to have a penalty free approach to identify the group structure in panels, and our proposed method serves this purpose.

The rest of the paper is organized as follows. In Chapter 2, we first introduce a fixed effects model with unknown group structure, and then propose an estimation and classification procedure. Asymptotic properties of our estimator are established in Chapter 3. Chapter 4 carries out a set of Monte Carlo simulations to investigate the finite sample performance of our method and apply our method to labor force participation. All mathematical derivation of main theorems and lemmas are provided in the Chapter 5.

Notation: The Frobenius norm of matrix A is given by $\|A\|_2 = \sqrt{\text{tr}(AA')}$. For square matrix A , $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues respectively. For positive integer k , define $[k] \equiv \{1, 2, \dots, k\}$. The symbols \xrightarrow{P} and \xrightarrow{D} denote convergence in probability and in distribution, respectively. Finally, $(N, T) \rightarrow \infty$ denotes N and T go to infinity jointly.

Chapter 2

Methodology

2.1 Estimation

Let Y_{it} be a real-valued observation and $X_{it} \in \mathbb{R}^p$ be a real vector of observed covariates, both collected on the i th individual at time t for $i \in [N] := \{1, 2, \dots, N\}$, $t \in [T] := \{1, 2, \dots, T\}$. Assume that the N individuals are actually belonging to G^0 underlying groups where G^0 is unknown. In particular, $G^0 = 1$ corresponds to the traditional fixed effect model without group structure (see [12]). To identify group structure, a common practice is to predetermine the number of groups, denoted G , and classify the N individuals into G groups. In practice, correctly specifying G , i.e., $G = G_0$, is difficult due to the unobservability of group pattern. A more realistic way is to pick G relatively large so that $G \geq G^0$. Obviously, such misspecification brings more challenges into both theoretical study and practical applications. In this paper, we propose a method for identifying group patterns under this misspecification and investigate its asymptotic property.

For individual i , let $g_i \in [G] := \{1, 2, \dots, G\}$ denote the group membership variable, $\beta_{g_i} \in \mathbb{K} \subset \mathbb{R}^p$ denote the unobservable group-specific parameter, and $\alpha_i \in \mathbb{A} \subset \mathbb{R}$ denote the unobservable individual-specific parameter, where both \mathbb{K} and \mathbb{A} are compact subsets. If individuals i, j belong to the same group, then $\beta_{g_i} = \beta_{g_j}$, i.e., they share a common group parameter, but α_i and α_j might still be different due to individual-level heterogeneity. Let $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_G) \in \mathbb{K}^G$ denote the tuple of G group-specific parameters, $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N) \in \mathbb{A}^N$ denote the N -vector of individual parameters,

and $\gamma_N = (g_1, g_2, \dots, g_N) \in \Gamma_N$ denote the N -vector of group membership variables, where $\Gamma_N = [G]^N$ is the class of all possible group assignments. Our aim is to estimate the triplet $\theta_N = (\underline{\beta}, \underline{\alpha}, \gamma_N)$ which can be performed through the following M -estimation:

$$\hat{\theta}_N = \arg \max_{\theta_N = (\underline{\beta}, \underline{\alpha}, \gamma_N) \in \Theta_N} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \psi(X_{it}, Y_{it}, \beta_{g_i}, \alpha_i), \quad (2.1)$$

where $\Theta_N = \mathbb{K}^G \times \mathbb{A}^N \times \Gamma_N$ denotes the entire parameter space, $\psi(X_{it}, Y_{it}, \beta_{g_i}, \alpha_i)$ denotes the logarithm of the pseudo likelihood function of Y_{it} given X_{it} under parameters β_{g_i}, α_i .

Unlike the C-Lasso approach proposed by [24], our M -estimation procedure (2.1) requires optimizing the objective function over the pre-regularized parameter space Θ_N where the parameters β_{g_i} therein naturally incorporate group constraint. This important feature avoids the delicate choice of penalty parameters as required by penalization-based methods. Various choices of the function ψ will be provided in the following three examples.

Example 2.1. Linear panel model: $Y_{it} = \beta'_{g_i} X_{it} + \alpha_i + \epsilon_{it}$, $1 \leq i \leq N$, $1 \leq t \leq T$, where ϵ_{it} 's represent the idiosyncratic error. In this case, one chooses

$$\psi(x, y, \beta, \alpha) = -(y - \beta'x - \alpha)^2.$$

Example 2.2. Binary choice panel model: $Y_{it} = 1(\beta'_{g_i} X_{it} + \alpha_i \geq \epsilon_{it})$, $1 \leq i \leq N$, $q \leq t \leq T$, where ϵ_{it} 's represent the idiosyncratic error with common distribution function F , and $1(\cdot)$ denotes the indicator. In this case, we choose

$$\psi(x, y, \beta, \alpha) = y \log F(\beta'x + \alpha) + (1 - y) \log(1 - F(\beta'x + \alpha)).$$

Example 2.3. Poisson panel model: Given X_{it} and under β_{g_i}, α_i , Y_{it} follows Poisson distribution with mean $\exp(\beta'_{g_i} X_{it} + \alpha_i)$. In this case, we can choose $\psi(x, y, \beta, \alpha)$ as the logarithm of Poisson density function with mean $\exp(\beta'x + \alpha)$ or as

$$\psi(x, y, \beta, \alpha) = y(\beta'x + \alpha) - \exp(\beta'x + \alpha).$$

2.2 Numerical Algorithm

Due to the complex structure of the parameter space Θ_N , it is challenging to directly solve (2.1). Instead, we introduce an efficient iterative algorithm. Before that, let us introduce some notation to simplify writing. Define

$$H_i(\beta, \alpha) = E(\psi(X_{i1}, Y_{i1}, \beta, \alpha)), \quad \widehat{H}_i(\beta, \alpha) = \frac{1}{T} \sum_{t=1}^T \psi(X_{it}, Y_{it}, \beta, \alpha),$$

$$\Psi_N(\theta_N) = \Psi_N(\underline{\beta}, \underline{\alpha}, \gamma_N) = \frac{1}{N} \sum_{i=1}^N H_i(\beta_{g_i}, \alpha_i),$$

$$\widehat{\Psi}_N(\theta_N) = \widehat{\Psi}_N(\underline{\beta}, \underline{\alpha}, \gamma_N) = \frac{1}{N} \sum_{i=1}^N \widehat{H}_i(\beta_{g_i}, \alpha_i).$$

Here H_i is the expected objective function for individual i , $\Psi_N(\theta_N)$ is the expected pooled objective function taking into account the group variables, \widehat{H}_i and $\widehat{\Psi}_N(\theta_N)$ are their respective sample versions. with this notation, (2.1) can be rewritten as

$$\widehat{\theta}_N = \arg \max_{\theta_N \in \Theta_N} \frac{1}{N} \sum_{i=1}^N \widehat{H}_i(\beta_{g_i}, \alpha_i). \quad (2.2)$$

We propose the following iterative algorithm to solve (2.2):

- (a) Choose initial estimators $(\underline{\beta}^{(0)}, \underline{\alpha}^{(0)})$.
- (b) For each $i \in [N]$, in the s th iteration, find $g_i^{(s+1)} = \arg \max_{g \in [G]} \widehat{H}_i(\beta_g^{(s)}, \alpha_i^{(s)})$. Then set $\gamma_N^{(s+1)} = (g_1^{(s+1)}, \dots, g_N^{(s+1)})$ and compute

$$(\underline{\beta}^{(s+1)}, \underline{\alpha}^{(s+1)}) = \arg \max_{(\underline{\beta}, \underline{\alpha}) \in \mathbb{K}^G \times \mathbb{A}^N} \widehat{\Psi}_N(\underline{\beta}, \underline{\alpha}, \gamma_N^{(s+1)}).$$

- (c) Repeat step (b) until the solution converges.

The following simple procedure is recommended to choose the initial estimators. For each $i \in [N]$, let $\widehat{\beta}_i^{\text{ML}}$'s and $\widehat{\alpha}_i^{\text{ML}}$'s be the pseudo maximum likelihood estimators of β_i^0 's and α_i^0 's based on $\{X_{it}, Y_{it}\}_{t=1}^T$, i.e., $(\widehat{\beta}_i^{\text{ML}}, \widehat{\alpha}_i^{\text{ML}}) = \arg \max_{\beta \in \mathbb{K}, \alpha \in \mathbb{A}} \widehat{H}_i(\beta, \alpha)$. Firstly, we choose $\underline{\alpha}^{(0)} = (\widehat{\alpha}_1, \widehat{\alpha}_2, \dots, \widehat{\alpha}_N)$. Next, one applies the standard k -means algorithm with $k = G$ to $\widehat{\beta}_i^{\text{ML}}$'s to get G clustering centers, say, $(\beta_1^{(0)}, \dots, \beta_G^{(0)})$. Finally, let $\underline{\beta}^{(0)} = (\beta_1^{(0)}, \dots, \beta_G^{(0)})$ to be initial estimators for iteration. In Monte Carlo simulations, we find this initial estimator works well and leads to a very fast convergence.

Chapter 3

Asymptotic Theory

In this section, we present several asymptotic results such as estimation consistency (Theorems 3.1 and 3.2) and classification consistency (Theorem 3.3). It is worthwhile to point out that such results hold under a misspecified G with $G \geq G^0$. As a byproduct, we provide a consistent procedure to determine the number of groups. Moreover, asymptotic normality for the estimators is established with a correctly specified G . Throughout this section, let $\theta_N^0 = (\underline{\beta}^0, \underline{\alpha}^0, \gamma_N^0)$ denote the true parameters under which the observations X_{it}, Y_{it} are generated, where $\underline{\beta}^0 = (\beta_1^0, \beta_2^0, \dots, \beta_{G^0}^0)$, $\underline{\alpha}^0 = (\alpha_1^0, \alpha_2^0, \dots, \alpha_N^0)$, and $\gamma_N^0 = (g_1^0, g_2^0, \dots, g_N^0)$.

3.1 Estimation Consistency

The main result of this section is to show that the proposed M -estimation method is consistent. Before stating our main theorems, let us introduce some technical conditions.

To start, for each $g \in [G^0]$, we let

$$N_g = \sum_{i=1}^N I(g_i^0 = g),$$

denote the true number of individuals from group g .

Assumption A1. (a) $\{X_{it}, Y_{it}\}_{t=1}^T$ are mutually independent across $i \in [N]$.

(b) For each $i \in [N]$, $\{X_{it}, Y_{it} : t \in [T]\}$ is stationary and α -mixing with mixing coefficients $\alpha_{[i]}(\cdot)$. Moreover, $\alpha(\tau) := \max_{1 \leq i \leq N} \alpha_{[i]}(\tau)$ satisfies $\alpha(\tau) \leq \exp(-C_0 \tau^{b_0})$, where $C_0 > 0$ and $b_0 > 0$ are constants.

(c) For each $i \geq 1$, $H_i(\beta, \alpha)$ is uniquely maximized at $(\beta_{g_i}^0, \alpha_i^0)$ and, for each $\epsilon > 0$,

$$\chi(\epsilon) := \inf_{i \geq 1} \inf_{\|\beta - \beta_{g_i}^0\|_2^2 + |\alpha - \alpha_i^0|^2 \geq \epsilon} [H_i(\beta_{g_i}^0, \alpha_i^0) - H_i(\beta, \alpha)] > 0.$$

(d) $d_0 \equiv \inf_{\check{g} \neq g} \|\beta_g^0 - \beta_{\check{g}}^0\|_2 > 0$.

(e) There is a non-negative function $Q(x, y)$ such that for all $(\beta, \alpha), (\check{\beta}, \check{\alpha}) \in \mathbb{K} \times \mathbb{A}$,

$$|\psi(x, y, \beta, \alpha)| \leq Q(x, y),$$

and

$$|\psi(x, y, \beta, \alpha) - \psi(x, y, \check{\beta}, \check{\alpha})| \leq Q(x, y)(\|\beta - \check{\beta}\|_2^2 + |\alpha - \check{\alpha}|^2)^{1/2}.$$

Furthermore, there exist $b_1 \in (0, \infty]$ and $B_1 > 0$ such that

$$\sup_{i \in [N]} P(Q(X_{i1}, Y_{i1}) > v) \leq \exp\left(1 - (v/B_1)^{b_1}\right), \text{ for all } v > 0.$$

Particularly, when $b_1 = \infty$, it means ψ is a bounded function and one can choose $B_1 = 2\|\psi\|_\infty$.

(f) For all $g \in [G^0]$, there exists a positive constant π_g such that $N_g/N \rightarrow \pi_g$ as $(N, T) \rightarrow \infty$.

Remark 1. Assumption A1.(a) assumes cross-sectional independence among the individuals which is standard for panel data, e.g., [20] and [24]. Assumption A1.(b) imposes weak dependence for the observations along the time dimension with the level of dependence controlled by an exponential bound with parameter b_0 . The stationarity assumption can be relaxed at cost of introducing more notation. A similar α -mixing condition can be found in [24] and [6]. Assumption A1.(c) is an identification condition, which can be verified case by case under certain mild conditions. The same condition was also assumed by [12] and [11]. Assumption A1.(d) says that the pairwise differences between the group parameters are bounded from below. This condition is needed to guarantee the identification of the group parameters. Similar conditions were also assumed by [6] and [24]. Assumption A1.(e) states that ψ is smooth satisfying certain exponential tail condition with decay rate of the tail probability characterized by b_1 .

Similar tail conditions are also assumed by [6] for the error term. Compared with other conditions such as moment conditions, the exponential tail condition can lead to better convergence results and is still valid in commonly used models such as Examples 2.1, 2.2 and 2.3. Assumption A1.(f) excludes the groups with zero proportion. This condition is standard and necessary for panel models with finite number of groups, e.g., see [6] and [24].

Let $d = b_0 b_1 / (b_0 + b_1)$. Since b_0 and b_1 characterize the weak dependence of the observations and decay rate of the tail probability, respectively, as discussed in Remark 1, d can be viewed as a quantity jointly controlling both. A special case is $b_1 = \infty$, i.e., ψ is bounded, where we have $d = b_0$.

Assumption A2. $\log N = o(T^{\frac{d}{1+d}})$.

Remark 2. For theoretical consideration, compared to the standard assumption on the rate of N and T in the literature where the ratio of T/N being a nonzero constant (e.g., [12] among others), Assumption A2 is a relatively weak condition, since Assumption A2 allows N to diverge exponentially faster than T , where the ratio of T/N goes to zero. Furthermore, Assumption A2 is also quite reasonable in practice, since most microeconomic datasets are with moderate large T and very large N .

In order to prove the consistency of $\hat{\theta}_N$, we introduce the following pseudo metric d_N on Θ_N . For any $\theta_N = (\underline{\beta}, \underline{\alpha}, \gamma_N)$, $\tilde{\theta}_N = (\underline{\tilde{\beta}}, \underline{\tilde{\alpha}}, \tilde{\gamma}_N) \in \Theta_N$, define

$$d_N(\theta_N, \tilde{\theta}_N) = \frac{1}{N} \sum_{i=1}^N \left(\|\beta_{g_i} - \tilde{\beta}_{\tilde{g}_i}\|_2 + |\alpha_i - \tilde{\alpha}_i| \right).$$

Specifically, $d_N(\theta_N, \tilde{\theta}_N)$ measures the average discrepancy of (β_{g_i}, α_i) 's and $(\tilde{\beta}_{\tilde{g}_i}, \tilde{\alpha}_i)$'s. Theorem 3.1 below proves consistency for $\hat{\theta}_N$ under this pseudo metric.

Theorem 3.1. *Suppose $G \geq G^0$ and Assumptions A1 and A2 hold. Then*

$$d_N(\hat{\theta}_N, \theta_N^0) \xrightarrow{P} 0 \quad \text{as} \quad (N, T) \rightarrow \infty.$$

Theorem 3.1 establishes the consistency results for the parameter set θ_N including the slope coefficients and fixed effects. If the parameters of interest are slope coefficients,

then it is easy to see that

$$\frac{1}{N} \sum_{i=1}^N \|\widehat{\beta}_{\widehat{g}_i} - \beta_{g_i^0}\|_2 \xrightarrow{P} 0 \text{ as } (N, T) \rightarrow \infty.$$

In other words, the estimators of the group parameters are consistent only in an ‘‘average’’ sense, and it is possible that a small proportion of the estimators $\widehat{\beta}_{\widehat{g}_i}$ ’s may still be inconsistent. In Theorem 3.2, we strengthen this result by showing that the estimators $\widehat{\beta}_{\widehat{g}_i}$ ’s are uniformly consistent without any additional assumption.

Theorem 3.2. *Suppose $G \geq G^0$ and Assumptions A1 and A2 hold. Then*

$$\sup_{1 \leq i \leq N} \|\widehat{\beta}_{\widehat{g}_i} - \beta_{g_i^0}\|_2 \xrightarrow{P} 0 \text{ as } (N, T) \rightarrow \infty.$$

Theorem 3.2 states that the estimators of all group parameters uniformly converge to the true group parameters. Again, both Theorems 3.1 and 3.2 only require $G \geq G^0$. If $G < G^0$, then the above results will be invalid since in this scenario, individuals from different groups need to be classified into the same group, and this will lead to inconsistency.

3.2 Detection of Group Structure among Individuals

Detection of group structure in panel data is a fundamentally important problem. The popular C-LASSO approach recently proposed by [24] requires the use of penalty for effectively classifying the individuals. In this section, we study our penalty-free grouping method and investigate its asymptotic property. Our theory and method are valid under $G \geq G^0$.

Recall that $\widehat{\gamma}_N = (\widehat{g}_1, \widehat{g}_2, \dots, \widehat{g}_N)$ is the estimator of the group membership variables obtained in (2.2). Our grouping method is simply based on \widehat{g}_i ’s as follows. For $g \in [G]$, define $\widehat{\mathcal{C}}_g = \{i \in [N] : \widehat{g}_i = g\}$, i.e., $\widehat{\mathcal{C}}_g$ is the collection of the individuals belonging to the g -th estimated group. Also define $\mathcal{C}_g^0 = \{i \in [N] : g_i^0 = g\}$ for $g \in [G^0]$, i.e., \mathcal{C}_g^0 is the population analogy based on the true group membership variables. It is important to provide the conditions under which such a simple grouping method is valid, that is,

for any $g \in [G]$, there exists a $\tilde{g} \in [G^0]$ such that $\widehat{\mathcal{C}}_g \subseteq \mathcal{C}_g^0$ with probability approaching one. Formal statement of this result is provided in Theorem 3.3. Such property implies that the individuals are correctly grouped.

To prove this result, we need stronger assumptions on the smoothness of ψ . In order to deal with partial derivatives of a multivariate function, we introduce the following multi-index notation. Let $\mathbf{k} = (k_1, k_2, \dots, k_{p+1})$ denote a multi-index, where k_l 's are non-negative integers. For any $\beta \in \mathbb{K} \subset \mathbb{R}^p$, denote $\beta = (\beta_{[1]}, \beta_{[2]}, \dots, \beta_{[p]})$, where $\beta_{[l]}$ is the l th coordinate of β . Define the \mathbf{k} th order partial derivative of $\psi(x, y, \beta, \alpha)$ with respect to β, α as follows:

$$D^{\mathbf{k}}\psi(x, y, \beta, \alpha) = \frac{\partial^{|\mathbf{k}|}\psi(x, y, \beta, \alpha)}{\partial\beta_{[1]}^{k_1} \dots \partial\beta_{[p]}^{k_p} \partial\alpha^{k_{p+1}}},$$

where $|\mathbf{k}| = k_1 + k_2 + \dots + k_{p+1}$. Also denote the Hessian of ψ and H_i (with respect to β, α) by

$$\ddot{\psi}(x, y, \beta, \alpha) = \begin{pmatrix} \frac{\partial^2 \psi(x, y, \beta, \alpha)}{\partial\beta\partial\beta'} & \frac{\partial^2 \psi(x, y, \beta, \alpha)}{\partial\beta'\partial\alpha} \\ \frac{\partial^2 \psi(x, y, \beta, \alpha)}{\partial\beta\partial\alpha} & \frac{\partial^2 \psi(x, y, \beta, \alpha)}{\partial\alpha^2} \end{pmatrix}, \quad \ddot{H}_i(\beta, \alpha) = E(\ddot{\psi}(X_{i1}, Y_{i1}, \beta, \alpha)).$$

We require the following conditions on the partial derivatives of ψ and Hessian of H_i 's.

Let $\mathcal{B}_i = \{(\beta, \alpha) \in \mathbb{K} \times \mathbb{A} : \|\beta - \beta_{g_i}^0\|_2 + |\alpha - \alpha_i^0| \leq a_0\}$ for $i \geq 1$, and $\mathcal{B} = \cup_{i \geq 1} \mathcal{B}_i$.

Assumption A3. (a) There exist some function $J(x, y)$, constant $a_0 > 0$ and integer $q_0 \geq 4$ such that for any \mathbf{k} with $|\mathbf{k}| \leq 4$ and $(\beta, \alpha) \in \mathcal{B}$,

$$|D^{\mathbf{k}}\psi(x, y, \beta, \alpha)| \leq J(x, y), \quad \sup_{i \geq 1} E J^{q_0}(X_{i1}, Y_{i1}) < \infty.$$

(b) The Hessian matrices $\{\ddot{H}_i(\beta_{g_i}^0, \alpha_i^0), i \geq 1\}$ are negative definite with the largest eigenvalues uniformly bounded by zero, i.e., $\sup_{i \geq 1} \lambda_{\max}(\ddot{H}_i(\beta_{g_i}^0, \alpha_i^0)) < 0$.

Remark 3. Assumption A3.(a) requires the partial derivatives of ψ with respect to (β, α) have bounded moment up to order of 4. Similar assumption has been made by [12] and [24]. Assumption A3.(b) requires the Hessian matrices of the expected objective function to be uniformly negative definite. It can be compared to the conditions on the Hessian matrices of the profiled objective function in [24].

Below is the main result of this section which provides the classification consistency of our grouping method under $G \geq G^0$.

Theorem 3.3. *Suppose $G \geq G^0$ and Assumptions A1-A3 hold. Then for each $g \in [G]$, there exists a $\tilde{g} \in [G^0]$ such that $\lim_{(N,T) \rightarrow \infty} P\left(\widehat{\mathcal{C}}_g \subseteq \mathcal{C}_{\tilde{g}}^0\right) = 1$ holds.*

Remark 4. Theorem 3.3 demonstrates that the proposed grouping method is valid under misspecification in the sense that, with probability approaching one, any grouped individuals asymptotically belong to a population group. This implies that any population group is either identical to a selected group or is partitioned into subgroups without any misclassification, which is possibly the best result one can expect under $G \geq G^0$. In the special case $G = G^0$, Theorem 3.3 naturally leads to classification consistency, i.e., upto a proper relabeling, with probability approaching one, $\widehat{\mathcal{C}}_g = \mathcal{C}_g^0$ for any $g \in [G^0]$. Classification consistency was also established by [24] when $G = G^0$.

Intuitively, Theorem 3.3 implies that under Assumptions A1-A3 and if $G > G^0$, then with probability approaching one: (i) individuals from the same group may be divided into different subgroups; (ii) individuals from different groups can not be categorized into the same group.

The condition $G \geq G^0$ in Theorem 3.3 is of great practical importance. Since the true number of groups is unknown in practice, it is safe to use a relative large number of groups to classify the data and to obtain consistent estimation. Otherwise, if $G < G_0$, different from both Theorems 3.2 and 3.3, neither the estimation nor the classification is consistent.

3.3 Determination of Number of Groups

Though our estimation and classification results are valid for misspecified G , it is still of interest to estimate the number of groups. In this section, we propose an efficient approach based on penalization to address this problem and establish its theoretical validity. Let $\widehat{\theta}_N^G$ be the estimator in (2.2) using G as the number of groups. To estimate G^0 , we define a penalized criterion function

$$PC(G) = \widehat{\Psi}_N(\widehat{\theta}_N^G) - \eta_{NT}G,$$

where $\eta_{NT} > 0$ is a penalty parameter that is used to exclude the extremely large and unlikely choice of G . We estimate G^0 based on following procedure:

$$\widehat{G} = \arg \max_{G \in [G_{\max}]} PC(G), \quad (3.1)$$

where G_{\max} is a predetermined upper bound for G . The following theorem shows that \widehat{G} is consistent, i.e., $\widehat{G} = G_0$ with high probability.

Theorem 3.4. *Suppose Assumptions A1 and A3 hold. Suppose also that N, T and η_{NT} satisfy $\log N = o(T^{\frac{d}{2(1+d)}})$, $\eta_{NT} T^{\frac{1}{4(1+d)}} \rightarrow \infty$ and $\eta_{NT} \rightarrow 0$. Then we have*

$$\lim_{(N,T) \rightarrow \infty} P(\widehat{G} = G^0) = 1.$$

Note that the rate condition $\log N = o(T^{\frac{d}{2(1+d)}})$ in Theorem 3.4 is slightly stronger than Assumption A2, though both conditions allow N to grow exponentially with T .

3.4 Asymptotic Normality

In this section, we study the asymptotic normality of $\widehat{\underline{\beta}}$ under $G = G^0$. For this, we introduce the following ‘‘oracle’’ estimator $\widetilde{\underline{\beta}}$ of $\underline{\beta}$ when the true group assignment γ_N^0 is known. Specifically, let

$$\widetilde{\underline{\beta}} = \arg \max_{\underline{\beta} \in \mathbb{K}^{G^0}} \max_{\alpha_i \in \mathbb{A}} \frac{1}{N} \sum_{i=1}^N \widehat{H}_i(\beta_{g_i^0}, \alpha_i).$$

Of course, $\widetilde{\underline{\beta}}$ is infeasible since γ_N^0 is practically unavailable. Interestingly, $\widetilde{\underline{\beta}}$ and $\widehat{\underline{\beta}}$ are in fact asymptotically equivalent as summarized in the following lemma.

Lemma 3.5. *Suppose Assumptions A1-A3 and $G = G^0$ hold. Under appropriate relabeling, we have $\lim_{(N,T) \rightarrow \infty} P(\widehat{\underline{\beta}} = \widetilde{\underline{\beta}}) = 1$.*

It can be seen from Lemma 3.5 that, to derive the asymptotic normality of $\widehat{\underline{\beta}}$, it is sufficient to derive the asymptotic normality of $\widetilde{\underline{\beta}}$. To achieve the latter, we make an additional Assumption A4. Before that, let us introduce some notation. Define

$$\begin{aligned} \rho_i &= E^{-1}\left(\frac{\partial^2 \psi}{\partial \alpha \partial \alpha}(X_{i1}, Y_{i1}, \beta_{g_i^0}^0, \alpha_i^0)\right) E\left(\frac{\partial^2 \psi}{\partial \beta \partial \alpha}(X_{i1}, Y_{i1}, \beta_{g_i^0}^0, \alpha_i^0)\right), \\ U_i(x, y, \beta, \alpha) &= \frac{\partial \psi}{\partial \beta}(x, y, \beta, \alpha) - \rho_i \frac{\partial \psi}{\partial \alpha}(x, y, \beta, \alpha), \quad R_i(x, y, \beta, \alpha) = \frac{\partial \psi}{\partial \alpha}(x, y, \beta, \alpha), \\ V_i(x, y, \beta, \alpha) &= \frac{\partial U_i}{\partial \beta'}(x, y, \beta, \alpha), \quad \mathcal{I}_i = E(V_i(X_{i1}, Y_{i1}, \beta_{g_i^0}^0, \alpha_i^0)). \end{aligned}$$

The above notation are standard in the literature of nonlinear panel models, e.g., [12] and [3]. To simplify writing, we introduce the following notation: $U_i^\alpha = \partial U_i / \partial \alpha$, $U_i^{\alpha\alpha} = \partial^2 U_i / \partial \alpha^2$, $U_{it} = U_i(X_{it}, Y_{it}, \beta_{g_i}^0, \alpha_i^0)$ and $U_{it}^\alpha = U_i^\alpha(X_{it}, Y_{it}, \beta_{g_i}^0, \alpha_i^0)$. We define R_{it}, R_{it}^α analogically. For each $i \geq 1$, let Λ_i denote the asymptotic covariance matrix of $\sum_{t=1}^T U_{it} / \sqrt{T}$ as $T \rightarrow \infty$, which has an expression

$$\Lambda_i = E(U_{it}U_{it}') + 2 \sum_{t=1}^{\infty} E(U_{i1}U_{i,1+t}').$$

Convergence of the above series holds uniformly for i due to Assumptions A1 and A3.

Assumption A4. (a) There exists a constant $0 < B_3 < 1$ such that

$$B_3 \leq \inf_{i \geq 1} \lambda_{\min}(\Lambda_i) \leq \sup_{i \geq 1} \lambda_{\max}(\Lambda_i) \leq 1/B_3.$$

Moreover, for each $g \in [G^0]$, there exist invertible matrices D_g and W_g such that

$$\lim_{N \rightarrow \infty} \sum_{i: g_i^0 = g} \Lambda_i / N_g = D_g \text{ and } \lim_{N \rightarrow \infty} \sum_{i: g_i^0 = g} \mathcal{I}_i / N_g = W_g.$$

(b) For each $g \in [G^0]$, there exists a vector $\Delta_g \in \mathbb{R}^p$ such that

$$\lim_{(N, T) \rightarrow \infty} \frac{1}{N_g T} \sum_{i: g_i^0 = g} E \left\{ \left(\frac{\sum_{t=1}^T R_{it}}{\sqrt{T} E(R_{i1}^\alpha)} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\frac{E(U_{i1}^{\alpha\alpha})}{2E(R_{it}^\alpha)} R_{it} - U_{it}^\alpha \right] \right) \right\} = \Delta_g.$$

Remark 5. Assumption A4.(a) requires that the eigenvalues of the covariance matrices Λ_i are bounded away from zero and infinity. Assumption A4.(b) is a common condition for handling asymptotic bias (see [12] and [3] for similar conditions).

As the main result of this section, Theorem 3.6 shows that the elements of $\widehat{\beta}$ are asymptotically normally distributed.

Theorem 3.6. *Suppose Assumptions A1, A3, A4, $G = G^0$ hold and $N/T \rightarrow \kappa$ for some $\kappa \geq 0$. Then under appropriate relabeling, as $(N, T) \rightarrow \infty$, for each $g \in [G^0]$,*

$$\sqrt{NT}(\widehat{\beta}_g - \beta_g^0) + \sqrt{N/T} W_g^{-1} \Delta_g \xrightarrow{D} N(0, \pi_g^{-1} W_g^{-1} D_g W_g^{-1}),$$

As a consequence, under appropriate relabeling, for each $g \in [G^0]$,

$$\sqrt{NT}(\widehat{\beta}_g - \beta_g^0) \xrightarrow{D} N(-\sqrt{\kappa} W_g^{-1} \Delta_g, \pi_g^{-1} W_g^{-1} D_g W_g^{-1}),$$

Remark 6. Theorem 3.6 is closely related to a number of work on panel data models with fixed effects. First, the asymptotic bias of $\widehat{\beta}_g$ is of order $O(\sqrt{N/T})$. For fixed effects model, [12] derived the same order for the asymptotic bias of the fixed effects estimator. In particular, $\widehat{\beta}$ becomes asymptotically unbiased when $N = o(T)$. Second, when $\{X_{it}, Y_{it} : i \in [N], t \in [T]\}$ are independent, the bias term has an expression:

$$\sqrt{\kappa\pi_g^{-1}}W_g^{-1}\Delta_g = \sqrt{\kappa\pi_g^{-1}}W_g^{-1} \lim_{N \rightarrow \infty} \frac{1}{N_g} \sum_{i:g_i^0=g} \left(\frac{E(U_{i1}^{\alpha\alpha})E(|R_{i1}|^2)}{2E^2(R_{i1}^\alpha)} - \frac{E(R_{i1}U_{i1}^\alpha)}{E(R_{i1}^\alpha)} \right).$$

For fixed effects model without group structure, i.e., $\pi_g = N_g/N = 1$, the above expression coincides with [3]. When $T = o(N)$, the bias of $\widehat{\beta}_g$ tends to infinity. This issue can be resolved by adapting the jackknife procedure proposed by [12] and [9] into our M-estimation procedure.

Chapter 4

Numerical Performance

In this chapter, we evaluate finite sample performance of our estimator by simulation and apply our estimation procedure to a real data set.

4.1 Simulations

In order to evaluate the finite-sample performance of the classification and estimation procedure, following [24], we consider three data generating processes (DGPs) that cover both linear and nonlinear panels of static and dynamic models. Throughout these DGPs, we generate the fixed effect α_i and the idiosyncratic error u_{it} are I.I.D $N(0, 1)$ across i and t . Moreover u_{it} is also independent of all regressors. We set the number of groups to be three (e.g., $G_0 = 3$), and the number of elements in each group are given by $N_1 = \lfloor 0.3N \rfloor$, $N_2 = \lfloor 0.3N \rfloor$ and $N_3 = N - N_1 - N_2$, where N is the total number of cross-sectional units and $\lfloor \cdot \rfloor$ denotes the integer part of ”.”.

DGP 1 (Linear panel model): The data is generated as

$$y_{it} = \alpha_i + X'_{it}\beta_{g_i} + u_{it}, \quad (4.1)$$

where $X_{it} = (0.2\alpha_i + e_{it,1}, 0.2\alpha_i + e_{it,2})'$ and $e_{it,1}, e_{it,2} \sim \text{I.I.D}N(0, 1)$ across i, t and are independent of α_i . The true coefficients are $(0.4, 1.6)$, $(1, 1)$, $(1.6, 0.4)$ for the three groups, respectively.

DGP 2 (Linear dynamic panel model): The data is generated as

$$y_{it} = \alpha_i(1 - \gamma_{g_i}) + \gamma_{g_i}y_{it-1} + X'_{it}\beta_{g_i} + u_{it}, \quad (4.2)$$

where X_{it} is a 2×1 vector of exogenous variables following two dimensional standard normal distribution. The true coefficients are $(0.4, 1.6, 1.6)$, $(0.6, 1, 1)$, $(0.8, 0.4, 0.4)$ for the three groups, respectively.

DGP 3 (Dynamic Panel Probit model):

$$y_{it} = 1(\gamma_{g_i} y_{it-1} + x_{it} \beta_{1,g_i} + \beta_{2,g_i} + \alpha_i > u_{it}), \quad (4.3)$$

where $x_{it} = 0.1\alpha_i + e_{it}$ with $e_{it} \sim \text{I.I.D } N(0, 1)$ and is independent of all other variables. The true coefficients are $(1, -1, 0.5)$, $(0.5, 0, -0.25)$, and $(0, 1, 0)$. It should be noted that γ_{g_i} and β_{1,g_i} are identifiable in this model, whereas β_{2,g_i} is unidentifiable because it is absorbed into the individual specified effects α_i .

For all the three DGPs, we consider the combinations of (N, T) with $N = (100, 200)$ and $T = (15, 25, 50)$. During the replication, the group membership is held fixed. The number of replication is set to be $R = 1000$. Since the goal of this paper is consistently estimate the regression coefficients, group membership and number of groups, we follow [24] to consider the following three criteria to examine the finite sample performance of the proposed M-estimation.

(1) We use the algorithm to determine the number of groups and then estimate the parameters through M-estimation procedure. For the estimation of parameters, the estimators are evaluated using the root mean squared error (RMSE) for each estimated group number G defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\hat{\beta}_{\hat{g}_i} - \beta_{g_i^0}\|_2^2}.$$

When $G = G_0$, we also consider another type RMSE similar to [24], defined as

$$\text{Group RMSE} = \sqrt{\frac{1}{G^0} \sum_{g=1}^{G^0} \|\hat{\beta}_g - \beta_g^0\|_2^2}.$$

(2) Frequency or empirical percentage of selecting the number of groups for a given true number of groups ($G^0 = 3$ in our designs).

(3) Classification, which is the percentage of correct classification. It is calculated as the percentage of correct classification of the N units, calculated as $\sum_{i=1}^N I(\hat{g}_i = g_i^0)/N$ under appropriate relabelling, averaged over the Monte Carlo replications.

Simulation results of DPGs 1-3 are summarized in Tables 4.1-4.4 and several interesting findings can be observed. First, Table 4.1 provides the RMSE for the proposed M-estimation using different number of groups with $G^0 = 3$. As we show in Theorem 3.2, our M-estimation procedure can lead to consistent estimator as long as $G \geq G^0$. From Table 4.1, we can observe that the RMSE decreases rapidly with the increase of either N or T , which is evident that the M-estimation is consistent. Moreover, as shown by Table 4.2, the group RMSE also decreases with the increase of N and T , and performs similarly to the oracle estimator (e.g., knowing the true group membership), which is consistent with our findings in Theorems 3.1-3.2. Second, Table 4.3 summarizes the accuracy of determination of number of groups using the criterion $PC(G)$ proposed in Section 3.3. We note that throughout all our designs of both linear and nonlinear panels, the determination of number of groups using the proposed algorithm is very accurate in the sense that the percentage of choosing the true number of groups is quite close to 1 with the increase of either N or T . Finally, Table 4.4 presents the simulation results of correct classification and group RMSE. For the correctness of classification, we can observe that with the increase of N and T , the algorithm we proposed is able to provide very accurate classification for both linear and nonlinear panels, which is evident that the classification is consistent as shown in Theorem 3.3. In all, we can claim that the simulation results confirm our theoretical findings in this paper regarding the identification and estimation for panels with unknown group structure under misspecification.

Table 4.1: RMSE under $G = 3, 4, 5$ with $G^0 = 3$.

N	T	DGP1			DGP2			DGP3		
		3	4	5	3	4	5	3	4	5
100	15	0.190	0.217	0.234	0.141	0.166	0.184	0.296	0.512	0.571
100	25	0.113	0.140	0.157	0.078	0.104	0.118	0.190	0.256	0.277
100	50	0.036	0.068	0.083	0.035	0.052	0.064	0.119	0.173	0.182
200	15	0.188	0.214	0.233	0.136	0.158	0.174	0.286	0.381	0.399
200	25	0.109	0.136	0.153	0.076	0.098	0.112	0.185	0.240	0.261
200	50	0.032	0.065	0.080	0.027	0.048	0.060	0.116	0.162	0.176

Table 4.2: Bias and GRMSE of DGPs 1-3 with $G^0 = 3$

N	T	DGP1				DGP2				DGP3			
		Bias		GRMSE		Bias		GRMSE		Bias		GRMSE	
		estimate	oracle	estimate	oracle	estimate	oracle	estimate	oracle	estimate	oracle	estimate	oracle
100	15	0.023	0.012	0.070	0.048	0.066	0.029	0.060	0.040	0.020	0.008	0.180	0.135
100	25	0.016	0.009	0.042	0.037	0.055	0.021	0.039	0.031	0.006	0.003	0.110	0.091
100	50	0.007	0.004	0.034	0.025	0.032	0.013	0.033	0.020	0.006	0.003	0.077	0.066
200	15	0.025	0.014	0.050	0.036	0.073	0.031	0.043	0.038	0.013	0.003	0.125	0.094
200	25	0.016	0.009	0.031	0.026	0.059	0.023	0.031	0.029	0.006	0.003	0.080	0.068
200	50	0.006	0.003	0.021	0.018	0.033	0.013	0.022	0.018	0.003	0.002	0.054	0.047

Note: "estimate" refers to estimation based estimated group membership,
"oracle" refers to estimation using the true group membership, i.e., g_i^0 .

Table 4.3: Percentage of choosing $G = 1, 2, \dots, 5$ with $G^0 = 3$.

N	T	DGP1					DGP2					DGP3				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
100	15	0	0.004	0.976	0.02	0	0	0	0.484	0.366	0.15	0	0.081	0.612	0.262	0.045
100	25	0	0	0.996	0.004	0	0	0	0.94	0.058	0.002	0	0.058	0.810	0.128	0.004
100	50	0	0	0.988	0.012	0	0	0	0.984	0.016	0	0	0.007	0.895	0.098	0
200	15	0	0	0.996	0.004	0	0	0	0.658	0.246	0.096	0	0.063	0.705	0.221	0.011
200	25	0	0	1	0	0	0	0	0.942	0.058	0	0	0.011	0.881	0.106	0.002
200	50	0	0	1	0	0	0	0	0.996	0.004	0	0	0.002	0.932	0.066	0

Table 4.4: Percentage of correct classification with $G^0 = 3$

N	T	DGP1	DGP2	DGP3
100	15	0.902	0.926	0.883
100	25	0.934	0.978	0.949
100	50	0.966	0.989	0.979
200	15	0.903	0.932	0.883
200	25	0.967	0.980	0.949
200	50	0.995	0.998	0.980

4.2 Empirical Application

In this section, we apply the above estimation and classification method to study the women's labor force participation. The dataset comes from Panel Study of Income Dynamics (PSID) and contains 1461 married women for 10 calendar years 1979-1988. We consider the following dynamic panel binary choice model with fixed effects

$$y_{it} = 1 (\alpha_i + \gamma_i y_{it-1} + x'_{it} \beta_{g_i} + \varepsilon_{it} > 0)$$

where y_{it} takes value one if woman i participate in period t and zero otherwise, α_i and δ_t represent individual specific effects and time effects, respectively. Other independent variables are $x_{it} = (\#children_{it}, \logincome_{it}, race, eduwife, agewife$ and $agewife^2)$, where $\#children_{it}$ is the number of children aged between 0 and 17, \logincome is the log of husband's labor income deflated by Consumer Price Index, $race$ is an indicator function and takes value 1 for black, $eduwife$ is the years of education of woman, $agewife$ is the age of women (divided by 10) and $agewife^2$ is squared age. Similar variables are also considered by [16] and [7].

Using the classification method in the previous section, we are able to divide the original sample into two groups, i.e., $G = 2$. The summary statistics for the original sample and two groups are provided in Table 4.5. From Table 4.5, we can observe that these two groups have quite distinct observations for some variables. For example, comparatively, individuals in group 2 have more children, lower percentage of black race and younger age, while, individuals in group 1 have more years of education. The difference in these two groups make a lot of difference in the estimation. Furthermore, based on the grouping, we can note that, on average, individuals from group 2 have much higher tendency to join the labor market comparing with individuals from group 1, e.g., the mean of labor force participation rate is 0.7898 for individuals from group 2 and is 0.3982 for group 1.

For the estimated group membership, we apply the fixed effects logit regression for each group and the whole sample. The estimation results are summarized in Table 4.6. Several interesting findings can be observed in the above estimation. First of all, we note that the effects of variables of previous year's labor force participation, husband's

income and wife's age remain the same across the whole sample and two groups, even if the effects are quite different across different groups. Second, we note that race has negative effects on the labor force participation in the whole sample and group 1, while race is no longer significant in group 2. From the summary statistics, we note that group 2 has relative low percentage of race black, which indicates that effects of race is offset by other variables in this group. Finally, we observe that education of wife is not significant in the whole sample and group 1, while it is significant in group 2, which indicates that education indeed has positive significant effect on the labor force participation for individuals in group 2. In all, we can conclude that, in order to capture the individual heterogeneity and group heterogeneity, it would be of crucial importance to classify individuals into different groups instead of pooling all individuals in the same group.

Table 4.5: Summary statistics for the original sample and two groups

Variables	Whole sample			Group1			Group2		
	min	mean	max	min	mean	max	min	mean	max
yit	0	0.5743	1	0	0.3982	1	0	0.7898	1
#children	0	1.76	7	0	1.691	6	0	1.841	7
logincome	5.806	10.471	13.846	5.806	10.483	12.995	6.64	10.46	13.85
Race	0	0.1642	1	0	0.1788	1	0	0.1471	1
eduwife	5	12.05	18	5	12.13	18	5	11.95	18
agewife	1.8	3.557	6.3	1.9	3.671	6.2	1.8	3.424	6.3
agewife2	3.24	13.41	39.69	3.61	14.25	38.44	3.24	12.43	39.69

Table 4.6: Logit estimation for the whole sample and two groups

Variable	Whole sample	Group1	Group2
yit-1	2.0504*** (0.0683)	2.1779*** (0.0841)	0.8835*** (0.1044)
#children	0.00001 (0.0295)	-0.0395 (0.0401)	-0.0915** (0.047)
logincome	-0.1933*** (0.0472)	-0.2408*** (0.0615)	-0.1787** (0.0814)
Race	-0.1735** (0.0842)	-0.1938* (0.1135)	0.1215 (0.1412)
eduwife	0.0096 (0.0082)	0.0088 (0.0118)	0.0257** (0.0126)
agewife	1.2635*** (0.2977)	1.8465*** (0.4133)	2.6434*** (0.4523)
agewife2	-0.161*** (0.0386)	-0.2297*** (0.0534)	-0.3097*** (0.0598)

Note: *, **, *** refer to significance at 10%, 5% and 1% level, respectively.

Chapter 5

Technical Proofs

In this chapter, we provide proofs of the theorems and relevant lemmas.

5.1 Proof of Theorem 3.1

Lemma 5.1. *Under Assumption A1 and A3, there exists a non negative function $\tilde{J}(x, y)$ such that for all $(\beta_1, \alpha_1), (\beta_2, \alpha_2) \in \mathcal{B}_i, i \geq 1$ and all $|\mathbf{k}| \leq 3$,*

$$\begin{aligned} |D^{\mathbf{k}}\psi(x, y, \beta, \alpha) - D^{\mathbf{k}}\psi(x, y, \beta, \alpha)| &\leq \tilde{J}(x, y)(\|\beta_1 - \beta_2\|_2^2 + |\alpha_1 - \alpha_2|^2)^{1/2}, \\ |D^{\mathbf{k}}\psi(x, y, \beta, \alpha)| &\leq \tilde{J}(x, y), \end{aligned}$$

and

$$\sup_{i \geq 1} E(\tilde{J}^{q_0}(X_{i1}, Y_{i1})) < \infty.$$

Proof. This is a consequence of Assumption A1.(a) and mean value theorem. ■

Lemma 5.2. *Under Assumption A1, the inequality*

$$\inf_{d_N(\theta_N, \theta_N^0) \geq \epsilon} [\Psi_N(\theta_N^0) - \Psi_N(\theta_N)] \geq \frac{\epsilon}{2R} \chi(\epsilon^2/8)$$

holds for every $0 < \epsilon < R$, with $R = \sup_{\beta_1, \beta_2 \in \mathbb{K}, \alpha_1, \alpha_2 \in \mathbb{A}} \|\beta_1 - \beta_2\|_2 + |\alpha_1 - \alpha_2|$.

Proof. Fix $0 < \epsilon < R$, let θ_N and θ_N^0 satisfy

$$d_N(\theta_N, \theta_N^0) = \frac{1}{N} \sum_{i=1}^N [\|\beta_{g_i} - \beta_{g_i^0}\|_2 + |\alpha_i - \alpha_i^0|] \geq \epsilon.$$

Then the cardinality k of the set of indices $A = \{i \in [N] : \|\beta_{g_i} - \beta_{g_i^0}\|_2 + |\alpha_i - \alpha_i^0| \geq \epsilon/2\}$ satisfies the inequality $(N - k)\epsilon/2 + kR \geq Nd_N(\theta_N, \theta_N^0) \geq N\epsilon$. From this we conclude $k \geq N\epsilon/(2R - \epsilon) \geq N\epsilon/(2R)$. The inequality $(a + b)^2 \leq 2a^2 + 2b^2$ and Assumption A1.(c) yield

$$[\Psi_N(\theta_N^0) - \Psi_N(\theta_N)] \geq \frac{1}{N} \sum_{i \in A} [H_i(\beta_{g_i^0}, \alpha_i^0) - H_i(\beta_{g_i}, \alpha_i)] \geq \frac{k}{N} \chi(\epsilon^2/8) \geq \frac{\epsilon}{2R} \chi(\epsilon^2/8).$$

By taking infimum on above inequality, the desired result follows. \blacksquare

Lemma 5.3. *Under Assumption A1, the following Lipchitz condition holds*

$$\sup_{i \geq 1} \sup_{(\beta_1, \alpha_1) \neq (\beta_2, \alpha_2) \in \mathbb{K} \times \mathbb{A}} \frac{|H_i(\beta_1, \alpha_1) - H_i(\beta_2, \alpha_2)|}{(\|\beta_1 - \beta_2\|_2^2 + |\alpha_1 - \alpha_2|^2)^{1/2}} \leq B_2, \quad (5.1)$$

with $B_2 = \int_0^\infty \exp(1 - (t/B_1)^{b_1}) dt$ if $0 < b_1 < \infty$ and $B_2 = B_1$ if $b_1 = \infty$.

Proof. The desired result is valid when $b_1 = \infty$. Now it suffices to show the case when $b_1 < \infty$. In the view of Assumption A1.(e), we have

$$\sup_{i \geq 1} E(Q(X_{i1}, Y_{i1})) \leq \int_0^\infty \sup_{i \geq 1} P(Q(X_{i1}, Y_{i1}) > t) dt \leq \int_0^\infty \exp(1 - (t/B_1)^{b_1}) dt < \infty.$$

In view of A1.(e), we can bound the left-hand side of (5.1) by

$$\sup_{i \geq 1} \sup_{(\beta_1, \alpha_1) \neq (\beta_2, \alpha_2) \in \mathbb{K} \times \mathbb{A}} E\left(\frac{|\psi(X_{i1}, Y_{i1}, \beta_1, \alpha_1) - \psi(X_{i1}, Y_{i1}, \beta_2, \alpha_2)|}{(\|\beta_1 - \beta_2\|_2^2 + |\alpha_1 - \alpha_2|^2)^{1/2}}\right) \leq \sup_{i \geq 1} E(Q(X_{i1}, Y_{i1}))$$

which gives the desired result. \blacksquare

Lemma 5.4. *Let $\{Z_t, t \geq 1\}$ be a sequence of mean zero stationary random variables with alpha-mixing coefficients satisfying*

$$\alpha(t) \leq \exp(-C_0 t^{b_0}), \quad t \geq 0,$$

with tail probabilities satisfying

$$P(|Z_t| > z) \leq \exp^{1-(z/B_1)^{b_1}}, \quad z > 0,$$

and satisfying

$$E(Z_1^2) + 2 \sum_{t=1}^{\infty} E(Z_1 Z_{1+t}) \leq M$$

for some finite M . Then there exist positive constants L_1 and L_2 such that the inequality

$$\left(\frac{1}{T} \left| \sum_{t=1}^T Z_t \right| > z\right) \leq 4 \left(1 + \frac{T^{\frac{1}{1+d}} z^2}{16M}\right)^{-T^{\frac{d}{1+d}}/2} + \frac{16L_1}{z} \exp\left(-L_2 T^{\frac{d}{1+d}} z^d\right),$$

holds for all $z > 0$ and all positive integers T , where $d = b_0 b_1 / (b_0 + b_1)$.

Proof. Evaluating equation (1.7) in [22] at $\lambda = Tz/4$ and $r = T^{\frac{d}{1+d}}$ yields the result. \blacksquare

Lemma 5.5. *Suppose Assumption A1 holds. Then there exist positive constants C_3, C_4 and C_5 such that the inequality*

$$\begin{aligned} & \sup_{1 \leq i \leq N} P \left(\sup_{(\beta, \alpha) \in \mathbb{K} \times \mathbb{A}} \left| \widehat{H}_i(\beta, \alpha) - H_i(\beta, \alpha) \right| > 10z \right) \\ \leq & C_4 \left[1 + \frac{1}{z^{2(p+2)}} \right] \\ & \times \left[\left(1 + \frac{T^{\frac{1}{1+d}} z^2}{C_5} \right)^{-T^{\frac{d}{1+d}}/4} + \frac{1}{d} \exp\left(-C_3 T^{\frac{d}{1+d}} z^d\right) + \frac{\exp(-C_3 d T^{\frac{d}{1+d}} z^d)}{1 - \exp(-C_3 d T^{\frac{d}{1+d}} z^d)} \right] \end{aligned}$$

holds for all $z > 0, N \geq 1$ and $T^{\frac{d}{1+d}} \geq 4(p+2)$. Furthermore, the condition $\log N = o(T^{\frac{d}{1+d}})$ implies

$$\sup_{1 \leq i \leq N} \sup_{(\beta, \alpha) \in \mathbb{K} \times \mathbb{A}} \left| \widehat{H}_i(\beta, \alpha) - H_i(\beta, \alpha) \right| = o_P(1)$$

and

$$\sup_{\theta_N \in \Theta_N} \left| \widehat{\Psi}_N(\theta_N) - \Psi_N(\theta_N) \right| = o_P(1).$$

Proof of Lemma 5.5. For distinct elements $\tau_1 = (\beta_1, \alpha_1)$ and $\tau_2 = (\beta_2, \alpha_2)$ of $\mathbb{K} \times \mathbb{A}$, define

$$l_{it} = \psi(X_{it}, Y_{it}, \beta_1, \alpha_1) - \psi(X_{it}, Y_{it}, \beta_2, \alpha_2).$$

By Assumption A1.(e) and Lemma 5.3, we have

$$|l_{it} - E(l_{it})| \leq (Q(X_{it}, Y_{it}) + B_2) \|\tau_1 - \tau_2\|_2. \quad (5.2)$$

Assumption A1.(e) and the inequality $(t - B_2)/B_1 \geq t/(B_1 + B_2) \geq 1$ valid for all $t \geq B_1 + B_2$ imply

$$P(Q(X_{i1}, Y_{i1}) + B_2 > t) \leq \exp\left(1 - (t/C)^{b_1}\right), \quad t > 0, \quad (5.3)$$

with $C = B_1 + B_2$. Using this inequality we find

$$\begin{aligned}\Lambda \equiv \sup_{i \geq 1} E(|Q(X_{i1}, Y_{i1}) + B_2|^3) &= \int_0^\infty P(|Q(X_{i1}, Y_{i1}) + B_2|^3 > t) dt \\ &\leq \int_0^\infty \exp\left(1 - \frac{t^{b_1/3}}{C^{b_1}}\right) dt\end{aligned}$$

is finite. In the view of [10][Proposition 2.5] and (5.2), we conclude

$$|\text{Cov}(l_{it}, l_{is})| \leq 8\alpha_{[i]}^{1/3}(t-s)E^{2/3}(|l_{i1} - E(l_{i1})|^3) \leq 8 \exp(-C_0|t-s|^{b_0}/3)\Lambda^{2/3}\|\tau_1 - \tau_2\|_2^2,$$

for ipositive integers s and t . Combing the above, we find the bound

$$\sup_{i \geq 1} \left[\text{Cov}(l_{i1}, l_{i1}) + 2 \sum_{t>1} \text{Cov}(l_{i1}, l_{it}) \right] \leq M\|\tau_1 - \tau_2\|_2^2, \quad (5.4)$$

with the finite constant $M = 16\Lambda^{2/3} \sum_{t=0}^\infty \exp(-C_0t^{b_0}/3)$. In the view of (5.2), (5.3) and (5.4), we can apply Lemma 5.4 with $Z_t = [l_{it} - E(l_{it})]/\|\tau_1 - \tau_2\|_2$ to obtain the inequality

$$P\left(\left|\frac{1}{T} \sum_{t=1}^T (l_{it} - E(l_{it}))\right| > z\right) \leq 4\left(1 + \frac{T_1 z^2}{16M}\right)^{-T_1^{d/2}} + \frac{16c_1}{z} \exp\left(-c_2(T_1 z)^d\right), \quad (5.5)$$

for all positive integers i and T and all $z > 0$, with $T_1 = T^{1/(1+d)}$ and positive constants c_1, c_2 . In the following, we will apply chaining argument to prove the concentration inequality.

For $\tau = (\beta, \alpha)$, define processes

$$X_i(\tau) = \frac{1}{T} \sum_{t=1}^T [\psi(X_{it}, Y_{it}, \beta, \alpha) - E(\psi(X_{it}, Y_{it}, \beta, \alpha))]$$

for all $i \in [N]$. Now fix i and abbreviate $X_i(\tau)$ by $X(\tau)$. We next construct a sequence of nested sets $T_0 \subset T_1 \subset \dots \subset \mathbb{K} \times \mathbb{A}$ such that

$$\|\tau - \tilde{\tau}\|_2 > 4^{-j},$$

for every distinct points $\tau, \tilde{\tau} \in T_j$, and that each T_j is "maximal" in the sense that no additional points can be added to T_j without violating above inequality. Therefore, by construction, the cardinality $|T_j|$ of T_j is bounded by $D4^j$, with $D = \max(\text{diam}(\mathbb{K} \times$

$\mathbb{A}), 2^{1/d}$). For $j \geq 0$, we link every element $\tau_{j+1} \in T_{j+1}$ to one and only one $\tau_j \in T_j$ such that

$$\|\tau_{j+1} - \tau_j\|_2 \leq 4^{-j}, \quad (5.6)$$

which can be done by the construction of T_{j+1} and T_j . Continue this process to link all points in T_j with points in T_{j-1} , and so on, to obtain for every $\tau_{j+1} \in T_{j+1}$ a chain $\tau_{j+1}, \tau_j, \dots, \tau_0$ that connects to a point in T_0 . To explicitly formulate such a link, we define link function $\gamma_{j+1}(\tau_{j+1}) = \tau_j$, where τ_{j+1} and τ_j satisfy (5.6) and recursively define $\gamma_{j,q}(\tau_q) = \gamma_{j+1} \circ \dots \circ \gamma_{q-1} \circ \gamma_q(\tau_q)$, for $q > j$. Also, let $\gamma_{q,q}$ idenote the identity map on T_q . For every positive integer k , each point $\tau \in T_k$ is linked to an element $\gamma_{0,k}(\tau) \in T_0$. Therefore, by the triangle inequality, we have for each $\tau \in T_k$,

$$|X(\tau) - X(\gamma_{0,k}(\tau))| \leq \sum_{j=1}^k |X(\gamma_{j,k}(\tau)) - X(\gamma_{j-1,k}(\tau))| \leq \sum_{j=0}^k \max_{\tau \in T_j} |X(\tau) - X(\gamma_j(\tau))|$$

where we used the fact $\gamma_{j,k}(\tau) \in T_j$ in the last step. Since, for fixed j , the maximum is taken over at most $|T_j| \leq D4^{j(p+1)}$ elements and $\|t - \gamma_j(t)\| \leq 4^{-(j-1)}$ for $t \in T_j$, we have, for $z > 0$,

$$\begin{aligned} \pi_k(4z) &= P\left(\sup_{\tau \in T_k} |X(\tau) - X(\gamma_{0,k}(\tau))| > 4z\right) \\ &\leq P\left(\sum_{j=1}^k \max_{\tau \in T_j} |X(\tau) - X(\gamma_j(\tau))| > 4z\right) \\ &\leq P\left(\sum_{j=1}^k \max_{\tau \in T_j} |X(\tau) - X(\gamma_j(\tau))| > 4 \sum_{j=1}^k 2^{-j} z\right) \\ &\leq \sum_{j=1}^k P\left(\max_{\tau \in T_j, \tau \neq \gamma_j(\tau)} \frac{|X(\tau) - X(\gamma_j(\tau))| 4^{-(j-1)}}{\|\tau - \gamma_j(\tau)\|_2} > 4 \times 2^{-j} z\right) \\ &\leq \sum_{j=1}^k \max_{\tau \in T_j, \tau \neq \gamma_j(\tau)} P\left(\frac{|X(\tau) - X(\gamma_j(\tau))|}{\|\tau - \gamma_j(\tau)\|_2} > 2^j z\right) D4^{j(p+1)}. \end{aligned}$$

Thus, with the aid of (5.5) we obtain the inequality

$$\pi_k(4z) \leq 4^{p+2} DK_1(z) + \frac{16Dc_1}{z} K_2(z), \quad z > 0, \quad (5.7)$$

with

$$\begin{aligned} K_1(z) &= \sum_{j=0}^{\infty} 4^{j(p+1)} \left(1 + \frac{T_1 4^j z^2}{4M}\right)^{-T_1^d/2}, \\ K_2(z) &= \sum_{j=1}^{\infty} 4^{j(p+1)} 2^{-j} \exp(-c_2 2^{jd} (zT_1)^d). \end{aligned}$$

With $a = T_1 z^2 / (4M)$, we calculate

$$a^{p+1} K_1(z) = \sum_{j=0}^{\infty} (1 + a4^j)^{-T_1^d/2} (4^j a)^{p+1} \leq \sum_{j=0}^{\infty} (1 + 4^j a)^{-T_1^d/2 + p+1}.$$

Since $T_1^{\frac{d}{1+d}} \geq 4(p+2)$ is assumed and $4^j \geq \max(1, j)$ holds for $j \geq 0$, we conclude

$$a^{p+1} K_1(z) \leq \sum_{j=0}^{\infty} (1 + 4^j a)^{-T_1^d/4-1} = 2(1+a)^{-T_1^d/4-1} + \sum_{j=2}^{\infty} (1 + ja)^{-T_1^d/4-1}$$

Next we use the inequality

$$\sum_{j=2}^{\infty} (1 + aj)^{-r-1} \leq \int_1^{\infty} (1 + ax)^{-r-1} dx = \frac{1}{ar} (1+a)^{-r},$$

valid for $r > 0$, to conclude

$$K_1(z) \leq \frac{1}{a^{p+1}} (1+a)^{-T_1^d/4} \left[2 + \frac{4}{aT_1^d}\right] \leq 2 \left[\frac{1}{a^{p+1}} + \frac{1}{a^{p+2}}\right] (1+a)^{-T_1^d/4}. \quad (5.8)$$

With $b = c_2(T_1 z)^d$, we obtain

$$K_2(z) = \sum_{j=1}^{\infty} \exp(-b2^{jd}) 2^{j(2p+1)} = \left(\frac{4p+2}{db}\right)^{(2p+1)/d} \sum_{j=1}^{\infty} \exp(-b2^{jd}) \left(\frac{db2^{jd}}{4p+2}\right)^{(2p+1)/d}$$

Using the inequality $x^y \leq \exp(xy)$ valid for positive x and y , we find

$$K_2(z) \leq \left(\frac{4p+2}{db}\right)^{(2p+1)/d} \sum_{j=1}^{\infty} \exp(-2^{jd}b/2).$$

In view of the inequalities $2^{jd} \geq 1$ for $1 \leq j \leq 1/d$ and $2^{jd} \geq jd$ for $j > 1/d$, we obtain the bound

$$K_2(z) \leq \left(\frac{4p+2}{db}\right)^{(2p+1)/d} \left[\frac{1}{d} \exp(-b/2) + \frac{\exp(-db/2)}{1 - \exp(-db/2)}\right]. \quad (5.9)$$

Therefore, in the view of (5.7), (5.8) and (5.9), we conclude that for $T^{\frac{d}{1+d}} \geq 4(p+2)$ and all $z > 0$ we have the bound

$$\begin{aligned} \pi_k(4z) &\leq C_{41} \left[\left(\frac{1}{T_1 z^2} \right)^{p+1} + \left(\frac{1}{T_1 z^2} \right)^{p+2} \right] \\ &\quad \times \left[\left(1 + \frac{T_1 z^2}{4M} \right)^{-T_1^d/4} + \frac{1}{d} \exp(-c_2(T_1 z)^d/2) + \frac{\exp(-c_2 d(T_1 z)^d/2)}{1 - \exp(-c_2 d(T_1 z)^d/2)} \right], \end{aligned} \quad (5.10)$$

where $C_{41} = \max\{2D4^{2p+4}(1+M)^{p+2}, 16Dc_1[(4p+2)/(dc_2)]^{2(p+1)/d}\}$. The triangle inequality yields the bound

$$\sup_{\tau, \tilde{\tau} \in T_k} |X(\tau) - X(\tilde{\tau})| \leq 2 \sup_{\tau \in T_k} |X(\tau) - X(\gamma_{0,k}(\tau))| + \sup_{\tau_0, \tilde{\tau}_0 \in T_0} |X(\tau_0) - X(\tilde{\tau}_0)|. \quad (5.11)$$

Here we also used the fact that $\gamma_{0,k}(\tau)$ and $\gamma_{0,k}(\tilde{\tau})$ are elements of T_0 . The maximum of the last term on the right-hand side of (5.11) is taken over at most $|T_0|^2 \leq D^2$ terms.

It follows that for all $z > 0$, we have the bound

$$\pi_o(z) = P\left(\sup_{\tau_0, \tilde{\tau}_0 \in T_0} |X(\tau_0) - X(\tilde{\tau}_0)| > z \right) \leq D^2 \sup_{\tau_0 \neq \tilde{\tau}_0 \in T_0} P\left(\frac{|X(\tau_0) - X(\tilde{\tau}_0)|D}{\|\tau_0 - \tilde{\tau}_0\|_2} > z \right)$$

and therefore by (5.5),

$$\pi_o(z) \leq 4D^2 \left(1 + \frac{T_1 z^2}{16MD^2} \right)^{-T_1^d/4} + \frac{16c_1 D^3}{z} \exp(-c_2(T_1 z/D)^d). \quad (5.12)$$

Combining (5.10), (5.11), (5.12) and the fact $D^d \geq 2$, we conclude that

$$p_k(9z) \equiv P\left(\sup_{\tau, \tilde{\tau} \in T_k} |X(\tau) - X(\tilde{\tau})| > 9z \right) \leq \pi_k(4z) + \pi_o(z)$$

and therefore $p_k(9z) \leq p(9z)$, where

$$\begin{aligned} p(9z) &= C_{42} \left[1 + \left(\frac{1}{T_1 z^2} \right)^{p+1} + \left(\frac{1}{T_1 z^2} \right)^{p+2} + \frac{1}{z} \right] \\ &\quad \times \left[\left(1 + \frac{T_1 z^2}{16MD^2} \right)^{-T_1^d/4} + \frac{1}{d} \exp(-c_2(T_1 z/D)^d) + \frac{\exp(-c_2 d(T_1 z/D)^d)}{1 - \exp(-c_2 d(T_1 z/D)^d)} \right], \end{aligned} \quad (5.13)$$

where $C_{42} = \max\{C_{41}, 4D^2, 16c_1 D^3\}$. By the continuity of the process $X(\tau)$, we have the identity

$$\sup_{\tau, \tilde{\tau} \in \mathbb{K} \times \mathbb{A}} |X(\tau) - X(\tilde{\tau})| = \sup_{\tau, \tilde{\tau} \in \bigcup_{k=1}^{\infty} T_k} |X(\tau) - X(\tilde{\tau})|$$

and thus obtain

$$P\left(\sup_{\tau, \tilde{\tau} \in \mathbb{K} \times \mathbb{A}} |X(\tau) - X(\tilde{\tau})| > 9z\right) \leq p(9z). \quad (5.14)$$

Using similar argument as in proving (5.5) we have for $\tau_0 \in \mathbb{K} \times \mathbb{A}$,

$$P(|X(\tau_0)| > z) \leq 4\left(1 + \frac{T_1 z^2}{16M}\right)^{-T_1^{d/2}} + \frac{16c_1}{z} \exp(-c_2(T_1 z)^d) \quad (5.15)$$

As a result, in view of (5.13), (5.14), (5.15) and $D^d \geq 2$, we conclude that for all $z > 0$

$$\pi(z) \equiv P\left(\sup_{\tau \in \mathbb{K} \times \mathbb{A}} |X(\tau)| > 10z\right) \leq P\left(\sup_{\tau \in \mathbb{K} \times \mathbb{A}} |X(\tau) - X(\tau_0)| > 9z\right) + P(|X(\tau_0)| > z)$$

satisfies

$$\pi(z) \leq C_4 \left[1 + \frac{1}{z^{2(p+2)}}\right] \left[\left(1 + \frac{T_1 z^2}{C_5}\right)^{-T_1^{d/4}} + \frac{1}{d} \exp(-C_3(T_1 z)^d) + \frac{\exp(-C_3 d(T_1 z)^d)}{1 - \exp(-C_3 d(T_1 z)^d)}\right]$$

with $C_3 = c_2/D^d$, $C_4 = \max\{C_{42}, 4 + 16C_{21}\}$ and $C_5 = 16MD^2$. This yields the first conclusion of the lemma. In view of the inequality

$$\begin{aligned} \sup_{\theta_N \in \Theta_N} |\widehat{\Psi}_N(\theta_N) - \Psi_N(\theta_N)| &\leq \sup_{\theta_N \in \Theta_N} \left| \frac{1}{N} \sum_{i=1}^N \left[\widehat{H}_i(\beta_{g_i}, \alpha_i) - H_i(\beta_{g_i}, \alpha_i) \right] \right| \\ &\leq \sup_{1 \leq i \leq N} \sup_{(\beta, \alpha) \in \mathbb{K} \times \mathbb{A}} |\widehat{H}_i(\beta, \alpha) - H_i(\beta, \alpha)| \end{aligned}$$

the other conclusions follow from the first conclusion and the condition $\log(N) = o(T^{\frac{d}{1+d}})$. \blacksquare

Define

$$S_{NT} = \sup_{1 \leq i \leq N} \sup_{\beta \in \mathbb{K}, \alpha \in \mathbb{A}} |\widehat{H}_i(\beta, \alpha) - H_i(\beta, \alpha)|.$$

Under certain assumptions, Lemma 5.5 shows that $S_{NT} = o_P(1)$, which plays an important role in our proof. Moreover, the definition of $\widehat{\Psi}$ suggests the following inequality:

$$\begin{aligned} \sup_{\theta_n \in \Theta_N} |\widehat{\Psi}(\theta_N) - \Psi(\theta_N)| &= \sup_{(\underline{\beta}, \underline{\alpha}, \gamma_N) \in \mathbb{K} \times \mathbb{A} \times \Gamma_N} \left| \frac{1}{N} \sum_{i=1}^N \left(\widehat{H}_i(\beta_{g_i}, \alpha_i) - H_i(\beta_{g_i}, \alpha_i) \right) \right| \\ &\leq S_{NT}. \end{aligned}$$

Proof of Theorem 3.1. By the definition of $\widehat{\theta}_N$, we have the inequality

$$\Psi_N(\theta_N^0) - S_{NT} \leq \widehat{\Psi}_N(\theta_N^0) \leq \widehat{\Psi}_N(\widehat{\theta}_N) \leq \Psi_N(\widehat{\theta}_N) + S_{NT} \leq \Psi_N(\theta_N^0) + S_{NT}.$$

This inequality and Lemma 5.5 establish the convergence result

$$\Psi_N(\widehat{\theta}_N) - \Psi_N(\theta_N^0) = o_P(1). \quad (5.16)$$

Fix ϵ in the interval $(0, R)$ with $R = \sup_{\beta_1, \beta_2 \in \mathbb{K}, \alpha_1, \alpha_2 \in \mathbb{A}} \|\beta_1 - \beta_2\|_2 + |\alpha_1 - \alpha_2|$. Lemma 5.2 yields the lower bound

$$\Psi_N(\theta_N^0) - \Psi_N(\widehat{\theta}_N) \geq [\Psi_N(\theta_N^0) - \Psi_N(\widehat{\theta}_N)]I(d_N(\widehat{\theta}_N, \theta_N^0) \geq \epsilon) \geq c_0 I(d_N(\widehat{\theta}_N, \theta_N^0) \geq \epsilon)$$

with $c_0 = \epsilon \chi(\epsilon^2/8)/(4R)$. From this and (5.16) we derive that $P(d_N(\widehat{\theta}_N, \theta_N^0) \geq \epsilon)$ converges to zero. This implies the desired result $d_N(\widehat{\theta}_N, \theta_N^0) = o_P(1)$. \blacksquare

5.2 Proof of Theorem 3.2

This section contains proofs of the main theorems. We begin by introducing some notation. For fixed $\beta \in \mathbb{K}$, define

$$\widehat{\alpha}_i(\beta) \equiv \arg \max_{\alpha \in \mathbb{A}} \widehat{H}_i(\beta, \alpha),$$

and for $\underline{\beta} \equiv (\beta_1, \beta_2, \dots, \beta_G) \in \mathbb{K}^G$, define

$$\widehat{\gamma}_N(\underline{\beta}) \equiv \arg \max_{\gamma_N \in \Gamma_N} \max_{\alpha \in \mathbb{A}^N} \widehat{\Psi}_i(\underline{\beta}, \alpha),$$

with $(\widehat{g}_1(\underline{\beta}), \widehat{g}_2(\underline{\beta}), \dots, \widehat{g}_N(\underline{\beta}))$ being the elements in $\widehat{\gamma}_N(\underline{\beta})$. To measure the difference between $(\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_G)$ and $(\beta_1^0, \beta_2^0, \dots, \beta_{G^0}^0)$ with possibly $G \neq G^0$, we define a map $\sigma : [G^0] \rightarrow [G]$ by:

$$\sigma(g) = \arg \min_{\tilde{g} \in [G]} \|\widehat{\beta}_{\tilde{g}} - \beta_g^0\|_2, \text{ for } g \in [G^0]. \quad (5.17)$$

If there are multiple minimizers, we just pick one of them.

Lemma 5.6. *Suppose Assumptions A1, A2 and $G \geq G^0$ hold. Then we have*

$$\max_{g \in [G^0]} \|\widehat{\beta}_{\sigma(g)} - \beta_g^0\|_2 = o_P(1).$$

Proof of Lemma 5.6. Recall that $N_g = \sum_{i=1}^N I(g_i^0 = g)$ denotes the size of group g . By the definition of σ , we have

$$\|\widehat{\beta}_{\sigma(g)} - \beta_g^0\|_2 = \frac{1}{N_g} \sum_{i=1}^N I(g = g_i^0) \|\widehat{\beta}_{\sigma(g_i^0)} - \beta_{g_i^0}^0\|_2 \leq \frac{1}{N_g} \sum_{i=1}^N I(g = g_i^0) \|\widehat{\beta}_{g_i^0} - \beta_{g_i^0}^0\|_2$$

for $g \in [G^0]$, and therefore obtain the bound

$$\max_{g \in [G^0]} \|\widehat{\beta}_{\sigma(g)} - \beta_g^0\|_2 \leq \max_{g \in [G^0]} \frac{N}{N N_g} \sum_{i=1}^N \|\widehat{\beta}_{g_i^0} - \beta_{g_i^0}^0\|_2 \leq \max_{g \in [G^0]} \frac{N}{N_g} d_N(\widehat{\theta}_N, \theta_N^0).$$

Thus the desired result follows from Assumption A1.(f) and Theorem 3.1. \blacksquare

Lemma 5.7. *Assumptions A1 and A2 imply*

$$\sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0| = o_P(1).$$

Furthermore, if $\{\beta_{Ti}, i \in [N]\}$ are random vectors satisfying $\sup_{1 \leq i \leq N} \|\beta_{Ti} - \beta_{g_i^0}^0\|_2 = o_P(1)$, then we have

$$\sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{Ti}) - \alpha_i^0| = o_P(1).$$

Proof of Lemma 5.7. By definition of S_{NT} and Assumption A1.(c)

$$\begin{aligned} 0 &\geq \widehat{H}_i(\beta_{g_i^0}^0, \alpha_i^0) - \widehat{H}_i(\beta_{g_i^0}^0, \widehat{\alpha}_i(\beta_{g_i^0}^0)) \\ &\geq H_i(\beta_{g_i^0}^0, \alpha_i^0) - H_i(\beta_{g_i^0}^0, \widehat{\alpha}_i(\beta_{g_i^0}^0)) - 2S_{NT} \\ &\geq \chi(|\widehat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0|^2) - 2S_{NT}. \end{aligned}$$

By this inequality and the fact that χ is nondecreasing, it follows from Lemma 5.5 that

$$\chi\left(\sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0|^2\right) = \sup_{1 \leq i \leq N} \chi(|\widehat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0|^2) \leq 2S_{NT} = o_P(1).$$

Noticing $\chi(0) = 0$ and $\chi(\epsilon) > 0$ for all $\epsilon > 0$, the above inequality implies

$$\sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0|^2 = o_P(1), \quad (5.18)$$

which is the first assertion.

The definition of S_{NT} and Lemma 5.3 imply

$$\begin{aligned} \sup_{1 \leq i \leq N} \sup_{\alpha \in \mathbb{A}} |\widehat{H}_i(\beta_{Ti}, \alpha) - \widehat{H}_i(\beta_{g_i^0}^0, \alpha)| &\leq \sup_{1 \leq i \leq N} \sup_{\alpha \in \mathbb{A}} |H_i(\beta_{Ti}, \alpha) - H_i(\beta_{g_i^0}^0, \alpha)| + 2S_{NT} \\ &\leq B_2 \sup_{1 \leq i \leq N} \|\beta_{Ti} - \beta_{g_i^0}^0\|_2 + 2S_{NT}. \end{aligned} \quad (5.19)$$

We have the lower bound

$$\hat{D}_i \equiv \sup_{\alpha \in \mathbb{A}} |\hat{H}_i(\beta_{Ti}, \alpha) - \hat{H}_i(\beta_{g_i^0}, \alpha)| \geq |\sup_{\alpha \in \mathbb{A}} \hat{H}_i(\beta_{Ti}, \alpha) - \sup_{\alpha \in \mathbb{A}} \hat{H}_i(\beta_{g_i^0}, \alpha)|.$$

Using this inequality and the definitions of S_{NT} and $\hat{\alpha}_i(\beta)$ we derive the inequality

$$\begin{aligned} \hat{D}_i &\geq |\hat{H}_i(\beta_{Ti}, \hat{\alpha}_i(\beta_{Ti})) - \hat{H}_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{Ti}))| \\ &\geq |\hat{H}_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{Ti})) - \hat{H}_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{g_i^0}))| - |\hat{H}_i(\beta_{Ti}, \hat{\alpha}_i(\beta_{Ti})) - \hat{H}_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{g_i^0}))| \\ &\geq |H_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{Ti})) - H_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{g_i^0}))| - 2S_{NT} - \hat{D}_i \end{aligned}$$

Combining the above inequalities yields

$$\sup_{1 \leq i \leq N} |H_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{Ti})) - H_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{g_i^0}))| \leq 2B_2 \sup_{1 \leq i \leq N} \|\beta_{Ti} - \beta_{g_i^0}\|_2 + 6S_{NT}. \quad (5.20)$$

In the view of (5.18), (5.20), Lemma 5.3 and Lemma 5.5, we have

$$\begin{aligned} &\sup_{1 \leq i \leq N} |H_i(\beta_{g_i^0}, \alpha_i^0) - H_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{Ti}))| \\ &\leq \sup_{1 \leq i \leq N} |H_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{Ti})) - H_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{g_i^0}))| + \sup_{1 \leq i \leq N} |H_i(\beta_{g_i^0}, \alpha_i^0) - H_i(\beta_{g_i^0}, \hat{\alpha}_i(\beta_{g_i^0}))| \\ &\leq 2B_2 \sup_{1 \leq i \leq N} \|\beta_{Ti} - \beta_{g_i^0}\|_2 + 6S_{NT} + B_2 \sup_{1 \leq i \leq N} |\hat{\alpha}_i(\beta_{g_i^0}) - \alpha_i^0| = o_P(1). \end{aligned}$$

By above inequality and the argumentt used to prove (5.18) imply

$$\sup_{1 \leq i \leq N} |\hat{\alpha}_i(\beta_{Ti}) - \alpha_i^0| = o_P(1),$$

which is the second result. \blacksquare

Proof of Theorem 3.2. From the inequality $\hat{H}_i(\hat{\beta}_{\sigma(g_i^0)}, \hat{\alpha}(\hat{\beta}_{\sigma(g_i^0)})) \leq \hat{H}_i(\hat{\beta}_{\hat{g}_i}, \hat{\alpha}(\hat{\beta}_{\hat{g}_i}))$

and the definition of S_{NT} , we derive

$$\begin{aligned} H_i(\hat{\beta}_{\sigma(g_i^0)}, \hat{\alpha}(\hat{\beta}_{\sigma(g_i^0)})) - S_{NT} &\leq \hat{H}_i(\hat{\beta}_{\sigma(g_i^0)}, \hat{\alpha}(\hat{\beta}_{\sigma(g_i^0)})) \\ &\leq \hat{H}_i(\hat{\beta}_{\hat{g}_i}, \hat{\alpha}(\hat{\beta}_{\hat{g}_i})) \leq H_i(\hat{\beta}_{\hat{g}_i}, \hat{\alpha}(\hat{\beta}_{\hat{g}_i})) + S_{NT} \end{aligned}$$

and therefore

$$C_{NT} \equiv \sup_{1 \leq i \leq N} \left(H_i(\hat{\beta}_{\sigma(g_i^0)}, \hat{\alpha}(\hat{\beta}_{\sigma(g_i^0)})) - H_i(\hat{\beta}_{\hat{g}_i}, \hat{\alpha}(\hat{\beta}_{\hat{g}_i})) \right) \leq 2S_{NT}. \quad (5.21)$$

Fix $\epsilon > 0$ and set $A_{NT} = \{\sup_{1 \leq i \leq N} \|\widehat{\beta}_{g_i} - \beta_{g_i}^0\|_2 \geq \epsilon\}$. By Assumption A1.(c), we have

$$\begin{aligned} & \inf_{1 \leq i \leq N} \left(H_i(\beta_{g_i}^0, \alpha_i^0) - H_i(\widehat{\beta}_{g_i}, \widehat{\alpha}(\widehat{\beta}_{g_i})) \right) \\ & \geq \inf_{1 \leq i \leq N} \left(H_i(\beta_{g_i}^0, \alpha_i^0) - H_i(\widehat{\beta}_{g_i}, \widehat{\alpha}(\widehat{\beta}_{g_i})) \right) I(A_{NT}) \\ & \geq \chi(\epsilon^2) I(A_{NT}). \end{aligned} \quad (5.22)$$

Lemma 5.3 and the triangle inequality imply

$$\begin{aligned} D_{NT} & \equiv \sup_{1 \leq i \leq N} \left(H_i(\beta_{g_i}^0, \alpha_i^0) - H_i(\widehat{\beta}_{\sigma(g_i^0)}, \widehat{\alpha}(\widehat{\beta}_{\sigma(g_i^0)})) \right) \\ & \leq \sup_{1 \leq i \leq N} B_2 \left(\|\widehat{\beta}_{\sigma(g_i^0)} - \beta_{g_i}^0\|_2 + |\widehat{\alpha}(\widehat{\beta}_{\sigma(g_i^0)}) - \alpha_i^0| \right). \end{aligned} \quad (5.23)$$

By Lemma 5.6, we have

$$\sup_{1 \leq i \leq N} \|\widehat{\beta}_{\sigma(g_i^0)} - \beta_{g_i}^0\|_2 \leq \max_{g \in [G^0]} \|\widehat{\beta}_{\sigma(g)} - \beta_g\|_2 = o_P(1). \quad (5.24)$$

Thus Lemma 5.7 applied with $\beta_{T_i} = \widehat{\beta}_{\sigma(g_i^0)}$ yields

$$\sup_{1 \leq i \leq N} |\widehat{\alpha}(\widehat{\beta}_{\sigma(g_i^0)}) - \alpha_i^0| = o_P(1). \quad (5.25)$$

The inequalities (5.23), (5.24) and (5.25) further imply

$$D_{NT} = o_P(1) \quad (5.26)$$

Combining (5.21), (5.22), (5.26) and Lemma 5.5, we have

$$\chi(\epsilon^2) I(A_{NT}) \leq C_{NT} + D_{NT} \leq 2S_{NT} + D_{NT} = o_P(1).$$

This implies $P(A_{NT}) \rightarrow 0$ as $(N, T) \rightarrow \infty$, which is the desired result. \blacksquare

5.3 Proof of Theorem 3.3

To compare $\underline{\beta} = (\beta_1, \dots, \beta_G) \in \mathbb{K}^G$ and $\widetilde{\underline{\beta}} = (\widetilde{\beta}_1, \dots, \widetilde{\beta}_{\widetilde{G}}) \in \mathbb{K}^{\widetilde{G}}$ for possible $G \neq \widetilde{G}$, we introduce following Hausdorff distance:

$$d_H(\underline{\beta}, \widetilde{\underline{\beta}}) = \max \left\{ \max_{g \in [G]} \min_{\widetilde{g} \in [\widetilde{G}]} \|\beta_g - \widetilde{\beta}_{\widetilde{g}}\|_2, \max_{\widetilde{g} \in [\widetilde{G}]} \min_{g \in [G]} \|\beta_g - \widetilde{\beta}_{\widetilde{g}}\|_2 \right\}. \quad (5.27)$$

Lemma 5.8. *Suppose Assumptions A1, A2 and $G \geq G^0$ hold. Then we have*

$$d_H(\widehat{\underline{\beta}}, \underline{\beta}^0) = o_P(1).$$

Proof. Since Lemma 5.6 implies $\max_{g \in [G^0]} \min_{\tilde{g} \in [G]} \|\widehat{\beta}_{\tilde{g}} - \beta_g^0\|_2 = o_P(1)$, it remains to show

$$\max_{\tilde{g} \in [G]} \min_{g \in [G^0]} \|\widehat{\beta}_{\tilde{g}} - \beta_g^0\|_2 = o_P(1).$$

To see this, we consider the set $A_{\tilde{g}} = \{i \in [N] : \widehat{g}_i = \tilde{g}\}$. For all $i \in A_{\tilde{g}}$, it holds that

$$\min_{g \in [G^0]} \|\widehat{\beta}_{\tilde{g}} - \beta_g^0\|_2 \leq \|\widehat{\beta}_{\tilde{g}} - \beta_{g_i^0}^0\|_2 = \|\widehat{\beta}_{\widehat{g}_i} - \beta_{g_i^0}^0\|_2.$$

Taking supremum over $A_{\tilde{g}}$ and \tilde{g} , leads to

$$\max_{\tilde{g} \in [G]} \min_{g \in [G^0]} \|\widehat{\beta}_{\tilde{g}} - \beta_g^0\|_2 \leq \max_{\tilde{g} \in [G]} \sup_{i \in A_{\tilde{g}}} \|\widehat{\beta}_{\widehat{g}_i} - \beta_{g_i^0}^0\|_2 = \sup_{1 \leq i \leq N} \|\widehat{\beta}_{\widehat{g}_i} - \beta_{g_i^0}^0\|_2 = o_P(1),$$

where the last step is due to Theorem 3.2. ■

Lemma 5.9. *Let x_0 be a point in \mathbb{R}^m and h be a function defined on the ball $B = \{x \in \mathbb{R}^m : \|x - x_0\| < r\}$ with derivative \dot{h} and Hessian matrix \ddot{h} which satisfies $\|\ddot{h}(x) - \ddot{h}(x_0)\| \leq L\|x - x_0\|$ for $x \in B$. If h is maximized at x_0 , then*

$$h(x_0) - h(x) \geq (1/2)\|x - x_0\|^2(\lambda - L\delta), \quad \|x - x_0\| < \delta,$$

holds for all $0 < \delta < r$ and with λ the smallest eigenvalue of the matrix $-\ddot{h}(x_0)$.

Proof. Since x_0 is a stationary point, the desired result follows from the Taylor expansion

$$h(x_0 + t) - h(x_0) = t'\dot{h}(x_0) + \int_0^1 (1-s)t'\ddot{h}(x_0 + st)t ds$$

and the fact that the integral is bounded by $[t'\ddot{h}(x_0)t + L\|t\|^3]/2$. ■

Lemma 5.10. *Suppose Assumptions A1 and A3 hold. Then there exist positive constants C_6 and C_7 such that for any $(\beta_i, \alpha_i) \in \mathbb{K} \times \mathbb{A}, 1 \leq i \leq N$ satisfying*

$$\sup_{1 \leq i \leq N} (\|\beta_i - \beta_{g_i^0}^0\|_2^2 + |\alpha_i - \alpha_i^0|^2) \leq C_7^2$$

the following inequalities hold,

$$H_i(\beta_{g_i^0}, \alpha_i^0) - H_i(\beta_i, \alpha_i) \geq C_6(\|\beta_i - \beta_{g_i^0}\|_2^2 + |\alpha_i - \alpha_i^0|^2), \quad 1 \leq i \leq N,$$

and

$$\frac{1}{N} \sum_{i=1}^N \left(H_i(\beta_{g_i^0}, \alpha_i^0) - H_i(\beta_i, \alpha_i) \right) \geq \frac{C_6}{N} \sum_{i=1}^N (\|\beta_i - \beta_{g_i^0}\|_2^2 + |\alpha_i - \alpha_i^0|^2).$$

Proof. This follows from Lemma 5.9 applied with x_0 equal to $(\beta_{g_i^0}, \alpha_i)$ and h equal to the restriction of H_i to \mathcal{B}_i . Indeed, by Lemma 5.1, it follows that \ddot{H}_i is Lipschitz in \mathcal{B}_i with Lipschitz constant $L_i = (p+1)(E(\tilde{J}^2(X_{i1}, Y_{i1})))^{1/2}$. Note that $M = \inf_{i \geq 1} \lambda_{\min}(-\ddot{H}_i(\beta_{g_i^0}, \alpha_i^0))$ is positive and $\Lambda = \sup_{i \geq 1} L_i$ is finite. Thus the choices $C_6 = M/4$ and $C_7 = \min(a_0, M/(2\Lambda))$ work. \blacksquare

Lemma 5.11. *Under Assumptions A1 and A3, there exist positive constants C_8 and C_9 such that for $\epsilon > 0$ small enough,*

$$\inf_{i \geq 1} \inf_{\|\beta - \beta_{g_i^0}\|_2^2 + |\alpha - \alpha_i^0|^2 \geq \epsilon} [H_i(\beta_{g_i^0}, \alpha_i^0) - H_i(\beta, \alpha)] \geq \min(C_8, C_9\epsilon).$$

Proof. Assumption A1.(c) and Lemma 5.10 imply

$$\begin{aligned} \inf_{1 \leq i \leq N} \inf_{\|\beta - \beta_{g_i^0}\|_2^2 + |\alpha - \alpha_i^0|^2 \geq C_7^2} [H_i(\beta_{g_i^0}, \alpha_i^0) - H_i(\beta, \alpha)] &\geq \chi(C_7^2) \\ \inf_{1 \leq i \leq N} \inf_{C_7^2 > \|\beta - \beta_{g_i^0}\|_2^2 + |\alpha - \alpha_i^0|^2 \geq \epsilon} [H_i(\beta_{g_i^0}, \alpha_i^0) - H_i(\beta, \alpha)] &\geq C_6\epsilon \end{aligned}$$

Therefore, the result holds with $C_8 = \chi(C_7^2)$ and $C_9 = C_6$. \blacksquare

Lemma 5.12. *Suppose Assumptions A1-A3 hold. Let $\{\beta_{g_i}, i \in [N]\}$ be a random vectors such that the probability of the event $B_{NT} = \{\sup_{1 \leq i \leq N} \|\beta_{g_i} - \beta_{g_i^0}\|_2 \leq \eta\}$ converges to one for some small enough $\eta > 0$. Then there exists a constant $C_{10} > 0$ such that with probability approaching one*

$$\sup_{1 \leq i \leq N} |\hat{\alpha}_i(\beta_{g_i}) - \alpha_i^0| \leq C_{10}\sqrt{\eta}.$$

Moreover, if $\sup_{1 \leq i \leq N} \|\delta_i\|_2 = o_P(1)$, where $\delta_i = \beta_{g_i} - \beta_{g_i^0}$, then we have

$$\sup_{0 \leq s \leq 1} \sup_{1 \leq i \leq N} |\hat{\alpha}_i(\beta_{g_i^0} + s\delta_i) - \alpha_i^0| = o_P(1).$$

Proof of Lemma 5.12. The definition of S_{NT} and Lemma 5.3 imply

$$\begin{aligned} \sup_{1 \leq i \leq N} \sup_{\alpha \in \mathbb{A}} |\widehat{H}_i(\beta_{g_i}, \alpha) - \widehat{H}_i(\beta_{g_i}^0, \alpha)| &\leq \sup_{1 \leq i \leq N} \sup_{\alpha \in \mathbb{A}} |H_i(\beta_{g_i}, \alpha) - H_i(\beta_{g_i}^0, \alpha)| + 2S_{NT} \\ &\leq B_2 \sup_{1 \leq i \leq N} \|\beta_{g_i} - \beta_{g_i}^0\|_2 + 2S_{NT}. \end{aligned} \quad (5.28)$$

As in the proof of Lemma 5.7 we derive the inequality

$$\begin{aligned} \widehat{D}_i &\equiv \sup_{\alpha \in \mathbb{A}} |\widehat{H}_i(\beta_{g_i}, \alpha) - \widehat{H}_i(\beta_{g_i}^0, \alpha)| \geq |\widehat{H}_i(\beta_{g_i}, \widehat{\alpha}_i(\beta_{g_i})) - \widehat{H}_i(\beta_{g_i}^0, \widehat{\alpha}_i(\beta_{g_i}))| \\ &\geq |\widehat{H}_i(\beta_{g_i}^0, \widehat{\alpha}_i(\beta_{g_i})) - \widehat{H}_i(\beta_{g_i}^0, \widehat{\alpha}_i(\beta_{g_i}^0))| - |\widehat{H}_i(\beta_{g_i}, \widehat{\alpha}_i(\beta_{g_i})) - \widehat{H}_i(\beta_{g_i}^0, \widehat{\alpha}_i(\beta_{g_i}^0))| \\ &\geq |H_i(\beta_{g_i}^0, \widehat{\alpha}_i(\beta_{g_i})) - H_i(\beta_{g_i}^0, \widehat{\alpha}_i(\beta_{g_i}^0))| - 2S_{NT} - \widehat{D}_i. \end{aligned}$$

The above inequalities imply

$$\sup_{1 \leq i \leq N} |H_i(\beta_{g_i}^0, \widehat{\alpha}_i(\beta_{g_i})) - H_i(\beta_{g_i}^0, \widehat{\alpha}_i(\beta_{g_i}^0))| \leq 2B_2 \sup_{1 \leq i \leq N} \|\beta_{g_i} - \beta_{g_i}^0\|_2 + 6S_{NT}.$$

and hence in view of Lemma 5.11

$$\min(C_8, C_9 |\widehat{\alpha}_i(\beta_{g_i}) - \widehat{\alpha}_i(\beta_{g_i}^0)|^2) \leq 2B_2 \sup_{1 \leq i \leq N} \|\beta_{g_i} - \beta_{g_i}^0\|_2 + 6S_{NT},$$

Now we introduce $A_{NT} = B_{NT} \cap \{S_{NT} < \eta, \sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{g_i}^0) - \alpha_i^0|^2 < \eta\}$ with $0 < \eta < C_8/(2B_2 + 6)$. On the event A_{NT} we have

$$\sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{g_i}) - \widehat{\alpha}_i(\beta_{g_i}^0)|^2 \leq \frac{2B_2 + 6}{C_9} \eta \quad (5.29)$$

and therefore by the triangle inequality

$$\sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{g_i}) - \alpha_i^0| \leq \sqrt{(2B_2 + 6)\eta/C_9} + \sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{g_i}^0) - \alpha_i^0| \leq C_{10}\sqrt{\eta} \quad (5.30)$$

with $C_{10} = \sqrt{(2B_2 + 6)/c_9} + 1$. This proves the first assertion as $P(A_{NT})$ converges to one if $P(B_{NT})$ does in view of Lemmas 5.5 and 5.7. The second assertion follows as the inequality (5.30) is also valid if β_{g_i} is replaced by $\beta_{g_i}^0 + s\delta_i$ for each $0 \leq s \leq 1$. \blacksquare

Before proving Theorem 3.3, we will introduce some notation. Define a neighborhood of $\underline{\beta}^0$ by $\mathcal{N}_\eta = \{\underline{\beta} \in \mathbb{K}^G : d_H(\underline{\beta}, \underline{\beta}^0) < \eta\}$ for $\eta > 0$. Also for each $\underline{\beta} \in \mathcal{N}_\eta$, we define sets $\mathcal{A}_\eta(\underline{\beta}, g) = \{\tilde{g} \in [G] : \|\beta_{\tilde{g}} - \beta_g^0\|_2 < \eta\} \subset [G]$, for all $g \in [G^0]$. Here $\mathcal{A}_\eta(\underline{\beta}, \cdot)$ plays a role of relabelling that connects labels in $[G^0]$ with labels in $[G]$.

Lemma 5.13. *Suppose Assumption A1-A3 and $G \geq G^0$ hold. Then for $\eta > 0$ small enough, we have the following:*

- (i) *For all $\underline{\beta} \in \mathcal{N}_\eta$, $\{\mathcal{A}_\eta(\underline{\beta}, g), g \in [G^0]\}$ is a partition of $[G]$ and each $\mathcal{A}_\eta(\underline{\beta}, g)$ is non empty for all $g \in [G^0]$.*
- (ii) $\lim_{(N,T) \rightarrow \infty} P\left(\sup_{\underline{\beta} \in \mathcal{N}_\eta} \sup_{1 \leq i \leq N} I(\widehat{g}_i(\underline{\beta}) \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)) > 0\right) = 0.$
- (iii) *If $G = G^0$, then each $\mathcal{A}_\eta(\underline{\beta}, g)$ contains exactly one element for all $g \in [G^0]$ and thus $\mathcal{A}_\eta(\underline{\beta}, \cdot)$ is a permutation of $[G^0]$. Under this permutation,*

$$\lim_{(N,T) \rightarrow \infty} P\left(\sup_{\underline{\beta} \in \mathcal{N}_\eta} \sup_{1 \leq i \leq N} I(\widehat{g}_i(\underline{\beta}) \neq g_i^0) > 0\right) = 0.$$

Proof of Lemma 5.13. (i) For $\underline{\beta} \in \mathcal{N}_\eta$, by definition, each $\mathcal{A}_\eta(\underline{\beta}, g)$ is not empty. Moreover, definition of \mathcal{N}_η and \mathcal{A}_η shows that $\cup_{g=1}^{G^0} \mathcal{A}_\eta(\underline{\beta}, g) = [G]$. Now we remain to show that $\{\mathcal{A}_\eta(\underline{\beta}, g), g \in [G^0]\}$ is a partition of $[G]$. Assume there exist some $g_{12} \in [G], g_1, g_2 \in [G^0]$ such that $g_{12} \in \mathcal{A}_\eta(\underline{\beta}, g_1) \cap \mathcal{A}_\eta(\underline{\beta}, g_2)$, then by Assumption A1.(d) and for $\eta < d_0/2$, it follows that

$$d_0 \leq \|\beta_{g_1}^0 - \beta_{g_2}^0\|_2 \leq \|\beta_{g_{12}} - \beta_{g_1}^0\|_2 + \|\beta_{g_{12}} - \beta_{g_2}^0\|_2 < 2\eta < d_0,$$

which is a contradiction.

- (ii) By definition of $\widehat{g}_i(\underline{\beta})$, we have for all $g \in [G]$ and $\tilde{g} \in [G]$:

$$I(\widehat{g}_i(\underline{\beta}) = g) \leq I(\widehat{H}_i(\beta_{\tilde{g}}, \widehat{\alpha}_i(\beta_{\tilde{g}})) \leq \widehat{H}_i(\beta_g, \widehat{\alpha}_i(\beta_g))),$$

Therefore, for any $\tilde{g}_i \in \mathcal{A}_\eta(\underline{\beta}, g_i^0)$, it implies that

$$\begin{aligned} I(\widehat{g}_i(\underline{\beta}) \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)) &= \sum_{g=1}^G I(g \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)) I(\widehat{g}_i(\underline{\beta}) = g) \\ &\leq \sum_{g=1}^G I(g \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)) I(\widehat{H}_i(\beta_{\tilde{g}_i}, \widehat{\alpha}_i(\beta_{\tilde{g}_i})) \leq \widehat{H}_i(\beta_g, \widehat{\alpha}_i(\beta_g))) \\ &\equiv \sum_{g=1}^G W_{ig}(\underline{\beta}). \end{aligned} \tag{5.31}$$

Since for all $\underline{\beta} \in \mathcal{N}_\eta$, $\{\mathcal{A}_\eta(\underline{\beta}, g), g \in [G^0]\}$ is a partition of $[G]$. Therefore, for all $g \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)$, we have $g \in \mathcal{A}_\eta(\underline{\beta}, g_j^0)$ with $g_j^0 \in [G^0]$ and $g_j^0 \neq g_i^0$. By definition, for small enough η and by Assumption A1.(d), we have

$$\|\beta_{g_i^0}^0 - \beta_g\|_2 \geq \|\beta_{g_i^0}^0 - \beta_{g_j^0}^0\|_2 - \|\beta_g - \beta_{g_j^0}^0\| \geq d_0 - \eta > 0.$$

According to Lemma 5.5 and above inequality, it follows that for all $\underline{\beta} \in \mathcal{N}_\eta$ and $g \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)$,

$$H_i(\beta_{g_i^0}^0, \alpha_i^0) - H_i(\beta_g, \hat{\alpha}_i(\beta_g)) \geq \chi(|d_0 - \eta|^2) \quad (5.32)$$

On the other hand, by Lemma 5.3, for all $\tilde{g}_i \in \mathcal{A}_\eta(\underline{\beta}, g_i^0)$, we have

$$\begin{aligned} H_i(\beta_{g_i^0}^0, \alpha_i^0) - H_i(\beta_{\tilde{g}_i}, \hat{\alpha}_i(\beta_{\tilde{g}_i})) &\leq B_2 \|\beta_{g_i^0}^0 - \beta_{\tilde{g}_i}\|_2 + B_2 |\alpha_i^0 - \hat{\alpha}_i(\beta_{\tilde{g}_i})| \\ &\leq B_2 \eta + B_2 |\alpha_i^0 - \hat{\alpha}_i(\beta_{\tilde{g}_i})| \\ &\leq B_2 \eta + B_2 \sup_{1 \leq i \leq N} |\alpha_i^0 - \hat{\alpha}_i(\beta_{\tilde{g}_i})|. \end{aligned} \quad (5.33)$$

Next define event $A_{NT} = \{\sup_{1 \leq i \leq N} |\alpha_i^0 - \hat{\alpha}_i(\beta_{\tilde{g}_i})| \leq C_{10} \sqrt{\eta}\}$, then $P(A_{NT}^c) = o(1)$ by Lemma 5.12. To proceed further, choose sufficiently small η such that $\epsilon_\eta \equiv \chi(|d_0 - \eta|^2) - B_2 \eta - C_{10} \sqrt{\eta} > 0$ for all (N, T) is large enough and this can be done by Assumption A1.(c). Hence by (5.32) and (5.33), for all $\underline{\beta} \in \mathcal{N}_\eta$ and $i \in [N]$, on the event A_{NT} , it holds that

$$H_i(\beta_{\tilde{g}_i}, \hat{\alpha}_i(\beta_{\tilde{g}_i})) - H_i(\beta_g, \hat{\alpha}_i(\beta_g)) \geq \epsilon_\eta. \quad (5.34)$$

As a consequence of (5.34), it yields that

$$\begin{aligned} W_{ig}(\underline{\beta}) &= I(g \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)) I(\hat{H}_i(\beta_{\tilde{g}_i}, \hat{\alpha}_i(\beta_{\tilde{g}_i})) \leq \hat{H}_i(\beta_g, \hat{\alpha}_i(\beta_g))) \\ &\leq I(\epsilon_\eta \leq \hat{H}_i(\beta_g, \hat{\alpha}_i(\beta_g)) - \hat{H}_i(\beta_{\tilde{g}_i}, \hat{\alpha}_i(\beta_{\tilde{g}_i})) - H_i(\beta_g, \hat{\alpha}_i(\beta_g)) + H_i(\beta_{\tilde{g}_i}, \hat{\alpha}_i(\beta_{\tilde{g}_i}))) \\ &\quad \times I(g \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)) \times I(A_{NT}) + I(A_{NT}^c) \\ &\leq 2I(\epsilon_\eta/2 \leq \sup_{\beta \in \mathbb{K}, \alpha \in \mathbb{A}} |\hat{H}_i(\beta, \alpha) - H_i(\beta, \alpha)|) + I(A_{NT}^c). \end{aligned} \quad (5.35)$$

By Lemma 5.5, (5.31) and (5.35), it yields that

$$\begin{aligned}
& P(\sup_{\underline{\beta} \in \mathcal{N}_\eta} \sup_{1 \leq i \leq N} I(\widehat{g}_i(\underline{\beta}) \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)) > 0) \\
&= P(\sup_{\underline{\beta} \in \mathcal{N}_\eta} \sup_{1 \leq i \leq N} I(\widehat{g}_i(\underline{\beta}) \notin \mathcal{A}_\eta(\underline{\beta}, g_i^0)) > 0.5) \\
&\leq \sum_{g=1}^G P\{\sup_{\underline{\beta} \in \mathcal{N}_\eta} \sup_{1 \leq i \leq N} W_{i,g}(\underline{\beta}) > 0.5\} \\
&\leq \sum_{g=1}^G P(\sup_{1 \leq i \leq N} 2I(\epsilon_\eta/2 \leq \sup_{\beta \in \mathbb{K}, \alpha \in \mathbb{A}} |\widehat{H}_i(\beta, \alpha) - H_i(\beta, \alpha)|) > 0.25) \\
&\quad + P(I(A_{NT}^c) > 0.25) \\
&= \sum_{g=1}^G P(\sup_{1 \leq i \leq N} \sup_{\beta \in \mathbb{K}, \alpha \in \mathbb{A}} |\widehat{H}_i(\beta, \alpha) - H_i(\beta, \alpha)| \geq \epsilon_\eta/2) + P(A_{NT}^c) = o(1).
\end{aligned}$$

(iii) If $G = G^0$, since $\{\mathcal{A}_\eta(\underline{\beta}, g), g \in [G^0]\}$ is a partition of $[G] = [G^0]$ and each $\mathcal{A}_\eta(\underline{\beta}, g)$ is non empty, so $\mathcal{A}_\eta(\underline{\beta}, g)$ only contains exactly on elements in $[G] = [G^0]$. We are able to define the permutation $\mathcal{A}_\eta(\underline{\beta}, \cdot) : [G^0] \rightarrow [G^0]$. Therefore, under this permutation, the second result is a specially case of first one. Proof completed. \blacksquare

Proof of Theorem 3.3. From the definition of d_H and Lemma 5.8, we conclude that for $0 < \eta < d_0/2$, with probability approaching one, $\{\mathcal{A}_\eta(\widehat{\underline{\beta}}, g), g \in [G^0]\}$ is a partition of $[G]$ and each $\mathcal{A}_\eta(\widehat{\underline{\beta}}, g)$ is non empty for all $g \in [G^0]$.

Next by Lemma 5.13 and the fact that with probability approaching one, $\widehat{\underline{\beta}} \in \mathcal{N}_\eta$, we have

$$\lim_{(N,T) \rightarrow \infty} P(\widehat{g}_i \in \mathcal{A}_\eta(\widehat{\underline{\beta}}, g_i^0), \forall i \in [N]) = 1.$$

Finally, suppose $i, j \in \widehat{\mathcal{C}}_g$ for some $g \in [G]$, then $\widehat{g}_i = \widehat{g}_j = g$. From argument above, we can see, with probability approaching one, $g \in \mathcal{A}_\eta(\widehat{\underline{\beta}}, g_i^0)$ and $g \in \mathcal{A}_\eta(\widehat{\underline{\beta}}, g_j^0)$. Notice with probability approaching one, $\{\mathcal{A}_\eta(\widehat{\underline{\beta}}, g), g \in [G^0]\}$ is a partition of $[G]$, so it follows that $g_i^0 = g_j^0$. Now define $\widetilde{g} = g_i^0 = g_j^0 \in [G^0]$, then $i, j \in \mathcal{C}_{\widetilde{g}}$. Therefore, with probability approaching one, for each $g \in [G]$, there exist $\widetilde{g} \in [G^0]$, such that $\widehat{\mathcal{C}}_g \subset \mathcal{C}_{\widetilde{g}}$. \blacksquare

5.4 Proof of Theorem 3.4

Lemma 5.14. For each $i \geq 1$, let $\{\zeta_{it}, t \in [T]\}$ be a stationary process with mean 0 and α -mixing coefficients satisfying $\alpha_{[i]}(t) \leq \exp(-C_0 t^{b_0})$ for all $t \geq 1$. Furthermore if

$\sup_i E(|\zeta_{it}|^{q_0}) \leq K$ holds for some positive constant K , then there is a constant C for which the inequalities

$$\sup_{i \geq 1} E\left(\left|\sum_{t=1}^T \zeta_{it}\right|^{q_0}\right) \leq CT^{q_0/2},$$

and

$$P\left(\sup_{1 \leq i \leq N} \left|\frac{1}{T} \sum_{t=1}^T \zeta_{it}\right| > \epsilon\right) \leq C\epsilon^{-q_0} NT^{-q_0/2}$$

hold. As a consequence, the condition $N = o(T^{q_0/2})$ implies

$$\sup_{1 \leq i \leq N} \left|\frac{1}{T} \sum_{t=1}^T \zeta_{it}\right| = o_P(1).$$

Proof of Lemma 5.14. By [10, Theorem 2.17 and Proposition 2.7], we have

$$E\left(\left|\sum_{t=1}^T \zeta_{it}\right|^{q_0}\right) \leq CT^{q_0/2}, \text{ for all } i \geq 1, \quad (5.36)$$

where $C > 0$ is a constant only relying on C_0, b_0, K, q_0 . Chebyshev's inequality and (5.36) imply

$$P\left(\sup_{1 \leq i \leq N} \left|\frac{1}{T} \sum_{t=1}^T \zeta_{it}\right| > \epsilon\right) \leq \sum_{i=1}^N P\left(\left|\frac{1}{T} \sum_{t=1}^T \zeta_{it}\right| > \epsilon\right) \leq \frac{NCT^{q_0/2}}{\epsilon^q T_0^q} = \frac{CN}{\epsilon_0^q T^{q_0/2}}.$$

for all $\epsilon > 0$. ■

Lemma 5.15. *Under Assumptions A1, A3 and $G < G^0$, there exists a constant B_4 satisfying*

$$\liminf_{(N,T) \rightarrow \infty} [\Psi_N(\theta_N^0) - \Psi_N(\hat{\theta}_N)] \geq B_4 > 0.$$

Proof. In view of the definition of d_0 , $G < G^0$ implies

$$\max_{g \in [G^0]} \min_{\tilde{g} \in [G]} \|\hat{\beta}_{\tilde{g}} - \beta_g^0\|_2 \geq d_0/2,$$

and this yields the lower bound

$$\sum_{i=1}^N \|\hat{\beta}_{\hat{g}_i} - \beta_{g_i^0}^0\|_2 = \sum_{g=1}^{G^0} \sum_{i=1}^N I(g_i^0 = g) \|\hat{\beta}_{\hat{g}_i} - \beta_g^0\|_2 \geq \sum_{g=1}^{G^0} N_g \min_{\tilde{g} \in [G]} \|\hat{\beta}_{\tilde{g}} - \beta_g^0\|_2 \geq \min_{g \in [G^0]} N_g d_0/2.$$

By Assumption A1.(f), for sufficient large N , we have the lower bound

$$\frac{1}{N} \sum_{i=1}^N \|\hat{\beta}_{\hat{g}_i} - \beta_{g_i^0}^0\|_2 \geq \epsilon_0 \equiv \min_{g \in [G^0]} \pi_g d_0/3 > 0.$$

and thus by Lemma 5.2

$$\Psi_N(\theta_N^0) - \Psi_N(\widehat{\theta}_N) \geq \frac{\epsilon_0}{2R} \chi(\epsilon_0^2/8) > 0.$$

Therefore, the results follow with $B_4 = \epsilon_0 \chi(\epsilon_0^2/8)/(2R)$. \blacksquare

Lemma 5.16. *Let ν be a constant such that $0 < \nu \leq \frac{1}{2(1+d)}$. Suppose Assumptions A1-A3 and $G \geq G^0$ hold. Then the condition $\log N = o(T^{(\frac{1}{1+d}-\nu)d})$ implies*

$$\frac{1}{N} \sum_{i=1}^N (\|\widehat{\beta}_{g_i} - \beta_{g_i}^0\|_2^2 + |\widehat{\alpha}_i - \alpha_i^0|^2) = O_p(T^{-\nu}).$$

Proof. For notational simplicity, define $\widetilde{d}_N(\theta_N, \theta_N^0) = \sum_{i=1}^N (\|\beta_{g_i} - \beta_{g_i}^0\|_2^2 + |\alpha_i - \alpha_i^0|^2)/N$ and $A_N = \{\theta_N \in \Theta_N : \sup_{1 \leq i \leq N} (\|\beta_{g_i} - \beta_{g_i}^0\|_2^2 + |\alpha_i - \alpha_i^0|^2) \leq C_7^2\}$. By Lemma 5.12 and Theorem 3.2, we have $\widetilde{d}_N(\widehat{\theta}_N, \theta_N^0) = o_P(1)$ and $\lim_{(N,T) \rightarrow \infty} P(\widehat{\theta}_N \in A_N) = 1$. For a positive number R_T define $S_{N,j} = \{\theta_N \in \Theta_N : 2^{j-1} \leq R_T \widetilde{d}_N(\theta_N, \theta_N^0) \leq 2^j\} \cap A_N$. If $R_T \widetilde{d}_N(\widehat{\theta}_N, \theta_N^0) > 2^k$ for some sufficient large integer k , then $\widehat{\theta}_N$ is in one of the sets $S_{N,j}$ for some $j > k$. So it follows that, for all $\eta > 0$

$$P(R_T \widetilde{d}_N(\widehat{\theta}_N, \theta_N^0) > 2^k, \widehat{\theta}_N \in A_N) \leq P(\widetilde{d}_N(\widehat{\theta}_N, \theta_N^0) > \eta) + P_{NT} \quad (5.37)$$

with

$$P_{NT} = \sum_{j>k, 2^j \leq \eta R_T} \pi_{Nj} \quad \text{and} \quad \pi_{Nj} = P\left(\sup_{\theta_N \in S_{N,j}} [\widehat{\Psi}_N(\theta_N) - \widehat{\Psi}_N(\theta_N^0)] \geq 0\right).$$

The first term on the right-hand side of (5.37) converges to zero by the argument above. Thus it suffices to show that P_{NT} converges to zero. Notice if $\theta_N \in S_{N,j}$, and choosing sufficiently small η , we have $\widetilde{d}_N(\theta_N, \theta_N^0) \leq \eta$. As a consequence, by Lemma 5.10, it holds for all $\theta_N \in S_{N,j}$ that

$$\Psi_N(\theta_N^0) - \Psi_N(\theta_N) \geq C_6 \widetilde{d}_N(\theta_N, \theta_N^0) \geq C_6 2^{j-1}/R_T.$$

From this we derive the inequality

$$\begin{aligned} \pi_{Nj} &\leq P\left(\sup_{\theta_N \in S_{N,j}} [\widehat{\Psi}_N(\theta_N) - \Psi_N(\theta_N) - \widehat{\Psi}_N(\theta_N^0) + \Psi_N(\theta_N^0)] \geq C_6 2^{j-1}/R_T\right) \\ &\leq 2P\left(\sup_{\theta_N \in S_{N,j}} |\widehat{\Psi}_N(\theta_N) - \Psi_N(\theta_N)| \geq C_6 2^{j-2}/R_T\right) \end{aligned}$$

Now we take $R_T = C_6 T^\nu / 20$ and use the inequality

$$H_{NT} \equiv \sup_{1 \leq i \leq N} \sup_{(\beta, \alpha) \in \mathbb{K} \times \mathbb{A}} |\widehat{H}_i(\beta, \alpha) - H_i(\beta, \alpha)| \geq \sup_{\theta_N \in \Theta_N} |\widehat{\Psi}_N(\theta_N) - \Psi_N(\theta_N)|$$

to obtain the bound

$$P_{NT} \leq 2 \sum_{j > k, 2^j \leq C_6 \eta T^\nu / 20} P(H_{NT} > 10 T^{-\nu} 2^k).$$

Next, we apply Lemma 5.5 to bound P_{NT} by

$$2C_4 N M \left[1 + \frac{T^{2\nu(p+2)}}{2^{2k(p+2)}} \right] \left[\left(1 + \frac{T^{\frac{1}{1+d} - 2\nu} 4^k}{C_5} \right)^{-T^{\frac{d}{1+d}/4}} + \Phi_d(C_3 2^{kd} T^{(\frac{1}{1+d} - \nu)d}) \right]$$

where $M = \log(C_6 \eta T^\nu / 20) / \log(2)$ is a bound on the cardinality of $\{j : j > k, 2^j \leq c_6 \eta T^\nu / 20\}$. Taking k large enough so that $4^k \geq C_5(e-1)$ and $dC_3 2^{kd} \geq \max\{1, -\log(d)\}$ hold, we conclude

$$P_N \leq 2C_4 N M \left[1 + T^{2\nu(p+2)} \right] \left[\exp(-T^{\frac{d}{1+d}/4}) + \frac{2}{d} \exp(-T^{(\frac{1}{1+d} - \nu)d}) \right].$$

This shows that P_{NT} converges zero in view of the rate $\log N = o(T^{(\frac{1}{1+d} - \nu)d})$. \blacksquare

Lemma 5.17. *Suppose Assumption A1, A3 and $G > G^0$ hold. Then the condition $\log N = o(T^{\frac{d}{2(1+d)}})$ implies*

$$|\widehat{\Psi}_N(\widehat{\theta}_N) - \widehat{\Psi}_N(\theta_N^0)| = O_P(T^{-\frac{1}{4(1+d)}}).$$

Proof. Assumption A1.(e) and the Cauchy-Schwarz inequality yield

$$\begin{aligned} & |\widehat{\Psi}_N(\widehat{\theta}_N) - \widehat{\Psi}_N(\theta_N^0)| \\ & \leq \sum_{i=1}^N \frac{\sum_{t=1}^T Q(X_{it}, Y_{it})}{TN} (\|\widehat{\beta}_{\widehat{g}_i} - \beta_{g_i}^0\|_2^2 + |\widehat{\alpha}_i - \alpha_i^0|^2)^{1/2} \\ & \leq \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{\sum_{t=1}^T Q(X_{it}, Y_{it})}{T} \right)^2} \sqrt{\frac{1}{N} \sum_{i=1}^N (\|\widehat{\beta}_{\widehat{g}_i} - \beta_{g_i}^0\|_2^2 + |\widehat{\alpha}_i - \alpha_i^0|^2)}. \end{aligned}$$

The term under the second square root is of order $O_P(T^{-\frac{1}{4(1+d)}})$ by Lemma 5.16 with $\nu = \frac{1}{2(1+d)}$, while the term under the first square root is of order $O_P(1)$ by Assumption A1.(b), A1.(e) and Lemma 5.14, \blacksquare

Proof of Theorem 3.4. It suffices to show

$$\lim_{(N,T) \rightarrow \infty} P(PC(G) \leq PC(G^0)) = 1. \quad (5.38)$$

Now we consider two cases, namely $G < G^0$ and $G > G^0$.

Under fitting case, $G < G^0$: By direct examination and Lemma 5.15, for (N, T) is large enough, it follows that

$$\begin{aligned} PC(G^0) - PC(G) &= \widehat{\Psi}_N(\widehat{\theta}_N^{G^0}) - \widehat{\Psi}_N(\theta_N^0) - \widehat{\Psi}_N(\widehat{\theta}_N^G) + \widehat{\Psi}_N(\theta_N^0) - \eta_{NT}(G^0 - G) \\ &\geq \widehat{\Psi}_N(\theta_N^0) - \widehat{\Psi}_N(\widehat{\theta}_N^G) - \eta_{NT}(G^0 - G) \\ &= \Psi_N(\theta_N^0) - \Psi_N(\widehat{\theta}_N^G) - \eta_{NT}(G^0 - G) + o_p(1) \\ &\geq B_4/2 + o_p(1). \end{aligned} \quad (5.39)$$

Since $\eta_{NT} \rightarrow 0$, it follows from (5.39) that (5.38) holds for the case $G < G^0$.

Over fitting case, $G > G^0$: By Lemma 5.17, it follows that

$$\begin{aligned} PC(G^0) - PC(G) &= \widehat{\Psi}_N(\widehat{\theta}_N^{G^0}) - \widehat{\Psi}_N(\theta_N^0) - \widehat{\Psi}_N(\widehat{\theta}_N^G) + \widehat{\Psi}_N(\theta_N^0) + \eta_{NT}(G - G^0) \\ &= O_p(T^{-\frac{1}{4(1+d)}}) + \eta_{NT}(G - G^0). \end{aligned} \quad (5.40)$$

Since $\eta_{NT} T^{\frac{1}{4(1+d)}} \rightarrow \infty$ and $G > G^0$, so (5.38) holds for the case when $G > G^0$. \blacksquare

5.5 Proof of Theorem 3.6

In this section, we abbreviate (X_{it}, Y_{it}) as Z_{it} for convenience.

Proof of Lemma 3.5. Suppose $G = G^0$, then by Theorem 3.3, under appropriate relabelling, it follows that for each $g \in [G^0]$,

$$\lim_{(N,T) \rightarrow \infty} P(\widehat{\mathcal{C}}_g = \mathcal{C}_g) = 1.$$

And above equation implies that

$$\lim_{(N,T) \rightarrow \infty} P(\widehat{g}_i = g_i^0, \forall i \in [N]) = 1.$$

Since on the event $\{\widehat{g}_i = g_i^0, \forall i \in [N]\}$, we have $\underline{\widehat{\beta}} = \underline{\widetilde{\beta}}$. Therefore, we finish the proof. \blacksquare

Recall the following terms defined in Section 3.4:

$$\begin{aligned}\rho_i &= \left\{ E \left(\frac{\partial^2 \psi}{\partial \alpha \partial \alpha} (Z_{i1}, \beta_{g_i^0}, \alpha_i^0) \right) \right\}^{-1} E \left(\frac{\partial^2 \psi}{\partial \beta \partial \alpha} (Z_{i1}, \beta_{g_i^0}, \alpha_i^0) \right), \\ U_i(x, y, \beta, \alpha) &= \frac{\partial \psi}{\partial \beta} (x, y, \beta, \alpha) - \rho_i \frac{\partial \psi}{\partial \alpha} (x, y, \beta, \alpha), \\ \Lambda_i &= E(U_{it} U_{it}') + 2 \sum_{t=1}^{\infty} E(U_{i1} U_{i,1+t}') \quad \text{with } U_{it} = U_i(Z_{it}, \beta_{g_i^0}, \alpha_i^0)\end{aligned}$$

Lemma 5.18. *Under Assumptions A1 and A3, then the averages $\bar{U}_i = \sum_{i=1}^T U_{it}/T$ satisfy*

$$\lim_{T \rightarrow \infty} \sup_{i \geq 1} \|TE(\bar{U}_i \bar{U}_i') - \Lambda_i\|_2 = 0.$$

Proof. For a unit vector in \mathbb{R}^p , define $\zeta_{it} = u' U_i(Z_{it}, \beta_{g_i^0}, \alpha_i^0)$ and autocovariance function $r_i(\tau) = \text{Cov}(\zeta_{it}, \zeta_{i,t+\tau})$, for $\tau \geq 0$. Since the random variables ζ_{it} are centered, it suffices to show

$$\lim_{T \rightarrow \infty} \sup_{i \geq 1} \left| \frac{1}{T} \text{Var} \left(\sum_{t=1}^T \zeta_{it} \right) - \left(r_i(0) + 2 \sum_{\tau=1}^{\infty} r_i(\tau) \right) \right| = 0 \quad (5.41)$$

Assumption A3.(b) implies that

$$\lambda \equiv \inf_{i \geq 1} \left| E \left(\frac{\partial^2 \psi}{\partial \alpha \partial \alpha} (Z_{i1}, \beta_{g_i^0}, \alpha_i^0) \right) \right| > 0.$$

This and Lemma 5.1 yield

$$\rho_* \equiv \sup_{i \geq 1} \|\rho_i\|_2 \leq \frac{1}{\lambda} \sqrt{p} E(\tilde{J}(Z_{i1})) \leq \frac{1}{\lambda} \sqrt{p} E^{1/q_0}(\tilde{J}^{q_0}(Z_{i1})) < \infty$$

and

$$|\zeta_{it}| \leq \|U_{it}\|_2 \leq (\sqrt{p} + \rho_*) \tilde{J}(Z_{it}).$$

By [10][Proposition 2.5] and Assumption A1.(b), we have

$$|r_i(\tau)| \leq 8\alpha_{[i]}^{1/3}(\tau) E^{2/3}(|\zeta_{it}|^3) \leq C\alpha_{[i]}^{1/3}(\tau)$$

with $C = 8((\sqrt{p} + \rho_*)^2 \sup_{i \geq 1} E^{2/3}(\tilde{J}^3(Z_{i1}))$ being finite due to Lemma 5.1. The above inequality and Assumption A1.(b) imply

$$\sup_{i \geq 1} \sum_{\tau=1}^{\infty} |r_i(\tau)| \leq C \sum_{\tau=1}^{\infty} \sup_{i \geq 1} \alpha_{[i]}^{1/3}(\tau) \leq C \sum_{\tau=1}^{\infty} \exp(-C_0 \tau^{b_0}/3) < \infty. \quad (5.42)$$

We calculate

$$\frac{1}{T} \text{Var} \left(\sum_{t=1}^T \zeta_{it} \right) = r_i(0) + 2 \sum_{\tau=1}^{T-1} \left(1 - \frac{\tau}{T}\right) r_i(\tau).$$

Thus the left-hand side of (5.41) is bounded by

$$2 \limsup_{T \rightarrow \infty} \sup_{i \geq 1} \left(\sum_{\tau=T}^{\infty} |r_i(\tau)| + \sum_{\tau=1}^{T-1} \frac{\tau}{T} |r_i(\tau)| \right) \leq 2 \lim_{T \rightarrow \infty} \sum_{\tau=1}^{\infty} \min\{1, \tau/T\} \exp(-C_0 \tau^{b_0})$$

which is zero in view of the Lebesgue dominated convergence theorem. \blacksquare

Lemma 5.19. *Suppose Assumptions A1 and A3 hold and the condition $N = o(T^{q_0/2})$ is met. Then for random vectors $\{\beta_{Ti}, i \in [N]\}$ satisfying $\sup_{1 \leq i \leq N} \|\beta_{Ti} - \beta_{g_i}^0\|_2 = o_P(1)$ and for all $|\mathbf{k}| \leq 3$, we have*

$$\sup_{0 \leq s \leq 1} \sup_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^T D^{\mathbf{k}} \psi(Z_{it}, \beta_{g_i}^0 + s\delta_i, \hat{\alpha}_i(\beta_{g_i}^0 + s\delta_i)) - E(D^{\mathbf{k}} \psi(Z_{it}, \beta_{g_i}^0, \alpha_i^0)) \right| = o_P(1),$$

with $\delta_i = \beta_{Ti} - \beta_{g_i}^0$. If also $\sup_{1 \leq i \leq N} |\alpha_{Ti} - \alpha_i^0| = o_P(1)$ holds, then we have

$$\sup_{0 \leq s \leq 1} \sup_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^T D^{\mathbf{k}} \psi(Z_{it}, \beta_{g_i}^0 + s\delta_i, \alpha_i^0 + sr_i) - E(D^{\mathbf{k}} \psi(Z_{it}, \beta_{g_i}^0, \alpha_i^0)) \right| = o_P(1),$$

with $r_i = \alpha_{Ti} - \alpha_i^0$.

Proof. Fix $|\mathbf{k}| \leq 3$ and set $K = D^{\mathbf{k}} \psi$. In view of the triangle inequality, the first assertion follows if we show

$$T_1 = \sup_{0 \leq s \leq 1} \sup_{1 \leq i \leq N} \left| \sum_{t=1}^T [K(Z_{it}, \beta_{g_i}^0 + s\delta_i, \hat{\alpha}_i(\beta_{g_i}^0 + s\delta_i)) - K(Z_{it}, \beta_{g_i}^0, \alpha_i^0)] / T \right| = o_P(1)$$

and

$$T_2 = \sup_{1 \leq i \leq N} \left| \sum_{t=1}^T K(Z_{it}, \beta_{g_i}^0, \alpha_i^0) / T - E(K(Z_{it}, \beta_{g_i}^0, \alpha_i^0)) \right| = o_P(1).$$

The second statement is an immediate consequence of Lemma 5.14 and the properties of K , see Lemma 5.1. Lemma 5.14 and the properties of \tilde{J} imply

$$J_{NT} = \sup_{1 \leq i \leq N} \frac{1}{T} \sum_{t=1}^T \tilde{J}(Z_{it}) = O_p(1)$$

According to Lemma 5.12, we have

$$R_{NT} := \sup_{1 \leq i \leq N} \|\delta_i\|_2 + \sup_{0 \leq s \leq 1} \sup_{1 \leq i \leq N} |\hat{\alpha}_i(\beta_{g_i}^0 + s\delta_i) - \alpha_i^0| = o_P(1).$$

Recall a_0 defined in Assumption A3.(a). For $a < a_0$, Lemma 5.1 implies

$$I[R_{NT} \leq a]T_1 \leq I(R_{NT} \leq a)J_{NT}R_{NT}$$

and therefore

$$P(T_1 > \epsilon) \leq P(R_{NT} > a) + P(R_{NT}J_{NT} > \epsilon) \rightarrow 0.$$

for every $\epsilon > 0$. This completes the proof of the first assertion. The second one is proved in a similar way. \blacksquare

Lemma 5.20. *Suppose Assumptions A1 and A3 and the condition $N = o(T^{\frac{90}{2}})$ hold. Then the following statements are true.*

$$(i) \inf_{i \geq 1} E(-R_{i1}^\alpha) > 0.$$

$$(ii) \sup_{i \geq 1} \|\rho_i\|_2 < \infty.$$

$$(iii) \frac{1}{N} \sum_{i=1}^N \left| \frac{1}{T} \sum_{t=1}^T U_{it}^\alpha \right|^2 = O_P(T^{-1}).$$

$$(iv) \frac{1}{N} \sum_{i=1}^N \frac{1}{E^2(R_{i1}^\alpha)} \left(\frac{1}{T} \sum_{t=1}^T R_{it} \right)^2 = O_P(T^{-1}).$$

Proof. For (i), we notice $E(-R_{i1}^\alpha)$ is a diagonal element of $-\ddot{H}_i(\beta_{g_i^0}^0, \alpha_i^0)$. So the result follows from Assumption A3.(b).

By Lemma 5.1, we can see

$$\sup_{i \geq 1} \left\| E \left(\frac{\partial^2 \psi}{\partial \beta \partial \alpha} (Z_{i1}, \beta_{g_i^0}^0, \alpha_i^0) \right) \right\|_2 < \infty.$$

Now (ii) follows from (i) and above inequality.

By definition, we can see $E(U_{it}^\alpha) = 0$ for all i, t . Moreover, Lemma 5.1 and (i) implies $\sup_{i \geq 1} E(\|U_{it}^\alpha\|_2) < \infty$. So one applies Lemma 5.14 to each element of the vector U_{it}^α and it follows that

$$E \left(\frac{1}{N} \sum_{i=1}^N \left| \frac{1}{T} \sum_{t=1}^T U_{it}^\alpha \right|^2 \right) \leq CT^{-1},$$

where C is some constant free of N, T . This proves (iii).

Proof of (iv) is similar to that of (iii). \blacksquare

Lemma 5.21. *Suppose Assumption A1 and A3 and the condition $N = o(T^{q_0/2})$ hold. Then for random vectors $\{\beta_{Ti}, i \in [N]\}$ satisfying $\sup_{1 \leq i \leq N} \|\beta_{Ti} - \beta_{g_i^0}\|_2 = o_P(1)$, the following statements hold.*

- (i) $\frac{1}{N} \sum_{i=1}^N |\widehat{\alpha}_i(\beta_{Ti}) - \widehat{\alpha}_i(\beta_{g_i^0})|^2 = O_p(\sup_{1 \leq i \leq N} \|\beta_{Ti} - \beta_{g_i^0}\|_2^2)$.
- (ii) $\left| \widehat{\alpha}_i(\beta_{g_i^0}) - \alpha_i^0 + E^{-1}(R_{i1}^\alpha) \frac{1}{T} \sum_{t=1}^T R_{it} \right| \leq \Delta_{NT} |\widehat{\alpha}_i(\beta_{g_i^0}) - \alpha_i^0|$, where Δ_{NT} is some error term with stochastic rate $o_P(1)$.
- (iii) $\frac{1}{N} \sum_{i=1}^N |\widehat{\alpha}_i(\beta_{g_i^0}) - \alpha_i^0|^2 = O_P(T^{-1})$.

As a consequence, it holds that

$$\frac{1}{N} \sum_{i=1}^N |\widehat{\alpha}_i(\beta_{Ti}) - \alpha_i^0|^2 = O_p\left(\sup_{1 \leq i \leq N} \|\beta_{Ti} - \beta_{g_i^0}\|_2^2 + T^{-1}\right).$$

Proof. The triangular inequality yields

$$\sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{Ti}) - \alpha_i^0| \leq \sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{Ti}) - \widehat{\alpha}_i(\beta_{g_i^0})| + \sup_{1 \leq i \leq N} |\widehat{\alpha}_i(\beta_{g_i^0}) - \alpha_i^0|. \quad (5.43)$$

In the following, we will bound two terms in (5.43) respectively. For first term, by the definition of $\widehat{\alpha}_i(\beta)$, we have $\sum_{t=1}^T \frac{\partial \psi}{\partial \alpha}(Z_{it}, \beta, \widehat{\alpha}_i(\beta)) = 0$. By implicit function differential theorem, we have

$$\frac{\partial \widehat{\alpha}_i}{\partial \beta}(\beta) = \left[\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \psi}{\partial \alpha \partial \alpha}(Z_{it}, \beta, \widehat{\alpha}_i(\beta)) \right]^{-1} \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \psi}{\partial \beta \partial \alpha}(Z_{it}, \beta, \widehat{\alpha}_i(\beta)).$$

Therefore, by Taylor expansion and Lemma 5.1, it holds that

$$\begin{aligned} \widehat{\alpha}_i(\beta_{Ti}) - \widehat{\alpha}_i(\beta_{g_i^0}) - E^{-1} \left[\frac{\partial^2 \psi}{\partial \alpha \partial \alpha}(Z_{i1}, \beta_{g_i^0}, \alpha_i^0) \right] E \left[\frac{\partial^2 \psi}{\partial \beta \partial \alpha}(Z_{i1}, \beta_{g_i^0}, \alpha_i^0) \right] (\beta_{Ti} - \beta_{g_i^0}) \\ = M_{NT,i}(\beta_{Ti} - \beta_{g_i^0}), \end{aligned} \quad (5.44)$$

with

$$\begin{aligned} M_{NT,i} = & \int_0^1 \left\{ \frac{\partial \widehat{\alpha}_i}{\partial \beta}(\beta_{g_i^0} + s(\beta_{Ti} - \beta_{g_i^0})) \right. \\ & \left. - E^{-1} \left[\frac{\partial^2 \psi}{\partial \alpha \partial \alpha}(Z_{i1}, \beta_{g_i^0}, \alpha_i^0) \right] E \left[\frac{\partial^2 \psi}{\partial \beta \partial \alpha}(Z_{i1}, \beta_{g_i^0}, \alpha_i^0) \right] \right\} ds. \end{aligned}$$

By Lemma 5.20.(i), we have

$$\begin{aligned} \inf_{i \geq 1} E \left[- \frac{\partial^2 \psi}{\partial \alpha \partial \alpha} (Z_{i1}, \beta_{g_i^0}^0, \alpha_i^0) \right] &= \inf_{i \geq 1} E(-R_{i1}^\alpha) > 0, \\ \sup_{i \geq 1} \left\| E \left[\frac{\partial^2 \psi}{\partial \beta \partial \alpha} (Z_{i1}, \beta_{g_i^0}^0, \alpha_i^0) \right] \right\|_2 &\leq \infty, \end{aligned} \quad (5.45)$$

which further leads to

$$\sup_{i \geq 1} \left\| E^{-1} \left[\frac{\partial^2 \psi}{\partial \alpha \partial \alpha} (Z_{i1}, \beta_{g_i^0}^0, \alpha_i^0) \right] E \left[\frac{\partial^2 \psi}{\partial \beta \partial \alpha} (Z_{i1}, \beta_{g_i^0}^0, \alpha_i^0) \right] \right\|_2 < \infty. \quad (5.46)$$

Therefore, according to Lemma 5.19, it yields that $\sup_{1 \leq i \leq N} |M_{NT,i}| = o_P(1)$. Combining above and (5.44), we have

$$\frac{1}{N} \sum_{i=1}^N |\hat{\alpha}_i(\beta_{T_i}) - \hat{\alpha}_i(\beta_{g_i^0}^0)|^2 \leq \sup_{1 \leq i \leq N} |\hat{\alpha}_i(\beta_{T_i}) - \hat{\alpha}_i(\beta_{g_i^0}^0)|^2 = O_p \left(\sup_{1 \leq i \leq N} \|\beta_{T_i} - \beta_{g_i^0}^0\|_2^2 \right).$$

which is (i).

For (ii), by Taylor expansion, we have

$$\begin{aligned} & -\frac{1}{T} \sum_{t=1}^T \frac{\partial \psi}{\partial \alpha} (Z_{it}, \beta_{g_i^0}^0, \alpha_i^0) - E \left[\frac{\partial^2 \psi}{\partial \alpha \partial \alpha} (Z_{i1}, \beta_{g_i^0}^0, \alpha_i^0) \right] (\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0) \\ &= \frac{1}{T} \sum_{t=1}^T \left[\frac{\partial \psi}{\partial \alpha} (Z_{it}, \beta_{g_i^0}^0, \hat{\alpha}_i(\beta_{g_i^0}^0)) - \frac{\partial \psi}{\partial \alpha} (Z_{it}, \beta_{g_i^0}^0, \alpha_i^0) \right] \\ & \quad - E \left[\frac{\partial^2 \psi}{\partial \alpha \partial \alpha} (Z_{i1}, \beta_{g_i^0}^0, \alpha_i^0) \right] (\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0) \\ &= \Delta_{NT,i} (\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0), \end{aligned} \quad (5.47)$$

with

$$\Delta_{NT,i} = \int_0^1 \left\{ \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \psi}{\partial \alpha \partial \alpha} (Z_{it}, \beta_{g_i^0}^0, \alpha_i^0 + s(\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0)) - E \left[\frac{\partial^2 \psi}{\partial \alpha \partial \alpha} (Z_{i1}, \beta_{g_i^0}^0, \alpha_i^0) \right] \right\} ds.$$

Lemma 5.19 implies $\sup_{1 \leq i \leq N} |\Delta_{NT,i}| = o_P(1)$. So by Assumption A3.(b) and (5.47), it yields that

$$\left| \hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0 + E^{-1}(R_{i1}^\alpha) \frac{1}{T} \sum_{t=1}^T R_{it} \right| \leq \left| \frac{\Delta_{NT,i}}{E(R_{i1}^\alpha)} \right| |\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0| \leq \Delta_{NT} |\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0|, \quad (5.48)$$

where $\Delta_{NT} = \sup_{1 \leq i \leq N} \left| \frac{\Delta_{NT,i}}{E(R_{i1}^\alpha)} \right| = o_P(1)$ due to Lemma 5.20.(i).

Next let us prove (iii). By (5.48), we have

$$\frac{1}{N} \sum_{i=1}^N |\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0|^2 \leq \frac{2}{N} \sum_{i=1}^N |E^{-1}(R_{i1}^\alpha)| \frac{1}{T} \sum_{t=1}^T R_{it} + \frac{\Delta_{NT}^2}{N} \sum_{i=1}^N |\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0|^2,$$

where the first term on right side of above inequality is $O_P(T^{-1})$ due to Lemma 5.20.(iv).

As a consequence, we find

$$\frac{1}{N} \sum_{i=1}^N |\hat{\alpha}_i(\beta_{g_i^0}^0) - \alpha_i^0|^2 = O_P(T^{-1}).$$

■

Lemma 5.22. *Suppose Assumptions A1 and A3 and the condition $N = O(T^{q_0/2})$ hold.*

Then, for each $g \in [G^0]$, we have following stochastic expansion:

$$\begin{aligned} & \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T [U_i(Z_{it}, \tilde{\beta}_g, \hat{\alpha}_i(\tilde{\beta}_g)) - U_i(Z_{it}, \beta_g^0, \alpha_i^0)] \\ &= \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T V_i(Z_{it}, \beta_g^0, \alpha_i^0) (\tilde{\beta}_g - \beta_g^0) \\ & \quad + \frac{1}{N_g T} \sum_{i:g_i^0=g} \left(\frac{\sum_{t=1}^T R_{it}}{\sqrt{T} E(R_{i1}^\alpha)} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\frac{E(U_{i1}^{\alpha\alpha})}{2E(R_{i1}^\alpha)} R_{it} - U_{it}^\alpha \right] \right) \\ & \quad + o_P(\|\tilde{\beta}_g - \beta_g^0\|_2) + o_P(T^{-1}). \end{aligned}$$

Proof. For notational simplicity, we assume $\tilde{\beta}_g$ is a scalar. Extension to the case when $\tilde{\beta}_g$ is multi-dimensional can be easily done by similar technique. By Theorem 3.2 and Theorem 3.6, we can see $\|\tilde{\beta}_g - \beta_g^0\|_2 = o_p(1)$. Let $\delta_\beta = \tilde{\beta}_g - \beta_g^0$, $\delta_{\alpha,i} = \hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0$. By Lemma 5.12, we have $\sup_{i:g_i^0=g} |\delta_{\alpha,i}| = o_P(1)$. Taylor expansion yields

$$\begin{aligned} & \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T [U_i(Z_{it}, \tilde{\beta}_g, \hat{\alpha}_i(\tilde{\beta}_g)) - U_i(Z_{it}, \beta_g^0, \alpha_i^0) - V_i(Z_{it}, \beta_g^0, \alpha_i^0) (\tilde{\beta}_g - \beta_g^0)] \\ &= T_1 + T_2 + T_3 + T_4 + M_1 (\tilde{\beta}_g - \beta_g^0)^2 \\ & \quad + \frac{1}{N_g} \sum_{i:g_i^0=g} M_{2i} (\tilde{\beta}_g - \beta_g^0) (\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0) + \frac{1}{N_g} \sum_{i:g_i^0=g} M_{3i} (\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0)^2, \end{aligned} \tag{5.49}$$

with

$$\begin{aligned}
T_1 &= \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T \frac{\partial U_i}{\partial \alpha}(Z_{it}, \beta_g^0, \alpha_i^0)(\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0), \\
T_2 &= \frac{1}{2N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T \frac{\partial^2 U_i}{\partial \beta \partial \beta}(Z_{it}, \beta_g^0, \alpha_i^0)(\tilde{\beta}_g - \beta_g^0)^2 \\
T_3 &= \frac{1}{2N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T \frac{\partial^2 U_i}{\partial \beta \partial \alpha}(Z_{it}, \beta_g^0, \alpha_i^0)(\tilde{\beta}_g - \beta_g^0)(\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0) \\
T_4 &= \frac{1}{2N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T \frac{\partial^2 U_i}{\partial \alpha \partial \alpha}(Z_{it}, \beta_g^0, \alpha_i^0)(\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0)^2 \\
M_1 &= \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T \int_0^1 (1-s) \left[\frac{\partial^2 U_i}{\partial \beta \partial \beta}(Z_{it}, \beta_g^0 + s\delta_\beta, \alpha_i^0 + s\delta_\alpha) \right. \\
&\quad \left. - \frac{\partial^2 U_i}{\partial \beta \partial \beta}(Z_{it}, \beta_g^0, \alpha_i^0) \right] ds, \\
M_{2i} &= \frac{1}{T} \sum_{t=1}^T \int_0^1 (1-s) \left[\frac{\partial^2 U_i}{\partial \beta \partial \alpha}(Z_{it}, \beta_g^0 + s\delta_\beta, \alpha_i^0 + s\delta_\alpha) - \frac{\partial^2 U_i}{\partial \beta \partial \alpha}(Z_{it}, \beta_g^0, \alpha_i^0) \right] ds, \\
M_{3i} &= \frac{1}{T} \sum_{t=1}^T \int_0^1 (1-s) \left[\frac{\partial^2 U_i}{\partial \alpha \partial \alpha}(Z_{it}, \beta_g^0 + s\delta_\beta, \alpha_i^0 + s\delta_\alpha) - \frac{\partial^2 U_i}{\partial \alpha \partial \alpha}(Z_{it}, \beta_g^0, \alpha_i^0) \right] ds,
\end{aligned}$$

By Lemma 5.19, we can see $\|M_1\|_2$, $\sup_{1 \leq i \leq N} \|M_{2i}\|_2$ and $\sup_{1 \leq i \leq N} \|M_{3i}\|_2$ are of order $o_P(1)$. So by Lemma 5.21, the last three items in (5.49) is of order $o_P(\|\tilde{\beta}_g - \beta_g^0\|_2 + T^{-1})$.

Next we will bound T_1, T_2, T_3, T_4 respectively. Direct examination yields

$$\begin{aligned}
T_1 &= \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T \frac{\partial U_i}{\partial \alpha}(Z_{it}, \beta_g^0, \alpha_i^0)(\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0) \\
&= \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T U_{it}^\alpha(\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0) \\
&= T_{11} + T_{12},
\end{aligned}$$

with

$$T_{11} = \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T U_{it}^\alpha(\hat{\alpha}_i(\tilde{\beta}_g) - \hat{\alpha}_i(\beta_g^0)), \quad T_{12} = \frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T U_{it}^\alpha(\hat{\alpha}_i(\beta_g^0) - \alpha_i^0).$$

Concerning T_{11} , by Cauchy's inequality, we have

$$\begin{aligned} |T_{11}|^2 &\leq \frac{1}{N_g^2} \sum_{i:g_i^0=g} \left| \frac{1}{T} \sum_{t=1}^T U_{it}^\alpha \right|^2 \sum_{i:g_i^0=g} |\widehat{\alpha}_i(\widetilde{\beta}_g) - \widehat{\alpha}_i(\beta_g^0)|^2 \\ &= O_p(T^{-1}) O_p(\|\widetilde{\beta}_g - \beta_g^0\|_2^2), \end{aligned}$$

where the last equality comes from Lemma 5.20.(iii) and Lemma 5.21.(i). For T_{12} , by Lemma 5.21.(ii), it holds that

$$\begin{aligned} &\left| T_{12} + \frac{1}{N_g} \sum_{i:g_i^0=g} E^{-1}(R_{i1}^\alpha) \left(\frac{1}{T} \sum_{t=1}^T U_{it}^\alpha \right) \left(\frac{1}{T} \sum_{t=1}^T R_{it} \right) \right| \\ &\leq \frac{\Delta_{NT}}{N_g} \sum_{i:g_i^0=g} \left| \frac{1}{T} \sum_{t=1}^T U_{it}^\alpha \right| |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0|. \end{aligned} \quad (5.50)$$

Since by Cauchy's inequality, Lemma 5.20.(iii) and Lemma 5.21.(iii)., we have

$$\begin{aligned} \left| \frac{1}{N_g} \sum_{i:g_i^0=g} \left(\frac{1}{T} \sum_{t=1}^T U_{it}^\alpha \right) (\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0) \right|^2 &\leq \frac{1}{N_g^2} \sum_{i:g_i^0=g} \left| \frac{1}{T} \sum_{t=1}^T U_{it}^\alpha \right|^2 \sum_{i:g_i^0=g} |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0|^2 \\ &= O_p(T^{-1}) O_p(T^{-1}). \end{aligned}$$

As a consequence, the term in right side of (5.50) is $o_P(T^{-1})$. Therefore, combining above, it follows that

$$\begin{aligned} T_1 &= -\frac{1}{N_g} \sum_{i:g_i^0=g} E^{-1}(R_{i1}^\alpha) \left(\frac{1}{T} \sum_{t=1}^T U_{it}^\alpha \right) \left(\frac{1}{T} \sum_{t=1}^T R_{it} \right) \\ &\quad + O_p(\|\widetilde{\beta}_g - \beta_g^0\|_2 T^{-1/2}) + o_p(T^{-1}). \end{aligned} \quad (5.51)$$

Secondly, let $c_{iT} = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 U_i}{\partial \beta \partial \beta} (Z_{it}, \beta_g^0, \alpha_i^0)$, then by Lemma 5.19, we have

$$\sup_{1 \leq i \leq N} |c_{iT} - E[\frac{\partial^2 U_i}{\partial \beta \partial \beta} (Z_{i1}, \beta_g^0, \alpha_i^0)]| = o_P(1).$$

Therefore, it follows that

$$\begin{aligned} |T_2| &= \left| \frac{1}{N_g} \sum_{i:g_i^0=g} c_{iT} \right| \|\widetilde{\beta}_g - \beta_g^0\|_2^2 \\ &\leq \left| \frac{1}{N_g} \sum_{i:g_i^0=g} E[\frac{\partial^2 U_i}{\partial \beta \partial \beta} (Z_{i1}, \beta_g^0, \alpha_i^0)] \right| + o_P(1) \|\widetilde{\beta}_g - \beta_g^0\|_2^2 \\ &= O_P(\|\widetilde{\beta}_g - \beta_g^0\|_2^2) = o_P(\|\widetilde{\beta}_g - \beta_g^0\|_2). \end{aligned} \quad (5.52)$$

Similarly, let $d_{iT} = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 U_i}{\partial \beta \partial \alpha}(Z_{it}, \beta_g^0, \alpha_i^0)$, then by Lemma 5.19, we have

$$\sup_{1 \leq i \leq N} |d_{iT} - E[\frac{\partial^2 U_i}{\partial \beta \partial \alpha}(Z_{i1}, \beta_g^0, \alpha_i^0)]| = o_P(1).$$

Therefore, it follows that

$$\begin{aligned} |T_3| &= \left| \frac{1}{N_g} \sum_{i: g_i^0 = g} d_{iT} (\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0) \right| \|\tilde{\beta}_g - \beta_g^0\|_2 \\ &\leq \frac{1}{N_g} \sum_{i: g_i^0 = g} |d_{iT}| \|\tilde{\beta}_g - \beta_g^0\|_2 \sup_{1 \leq i \leq N} |\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0| \\ &\leq \left(\frac{1}{N_g} \sum_{i: g_i^0 = g} \left| E[\frac{\partial^2 U_i}{\partial \beta \partial \alpha}(Z_{i1}, \beta_g^0, \alpha_i^0)] \right| + o_P(1) \right) \|\tilde{\beta}_g - \beta_g^0\|_2 \sup_{1 \leq i \leq N} |\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0| \\ &\leq \left(\sup_{i \geq 1} E(\tilde{J}(Z_{i1})) + o_P(1) \right) \|\tilde{\beta}_g - \beta_g^0\|_2 \sup_{1 \leq i \leq N} |\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0| \\ &= o_P(\|\tilde{\beta}_g - \beta_g^0\|_2), \end{aligned} \tag{5.53}$$

where the last equality is due to Lemma 5.1 and Lemma 5.7. Lastly, we will deal with

T_4 . Let $e_{iT} = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 U_i}{\partial \alpha \partial \alpha}(Z_{it}, \beta_g^0, \alpha_i^0)$ By Lemma 5.19, we have

$$\sup_{1 \leq i \leq N} |e_{iT} - E(U_{i1}^{\alpha\alpha})| = o_P(1).$$

As a consequence, Lemma 5.21 implies

$$\begin{aligned} &\left| T_4 - \frac{1}{N_g} \sum_{i: g_i^0 = g} E(U_{i1}^{\alpha\alpha}) (\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0)^2 \right| \\ &= \left| \frac{1}{N_g} \sum_{i: g_i^0 = g} \left(e_{iT} - E(U_{i1}^{\alpha\alpha}) \right) (\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0)^2 \right| \\ &\leq \sup_{1 \leq i \leq N} |e_{iT} - E(U_{i1}^{\alpha\alpha})| \left| \frac{1}{N_g} \sum_{i: g_i^0 = g} (\hat{\alpha}_i(\tilde{\beta}_g) - \alpha_i^0)^2 \right| \\ &= o_P(\|\tilde{\beta}_g - \beta_g^0\|_2^2 + T^{-1}), \end{aligned} \tag{5.54}$$

By Assumption A3.(b) and Lemma 5.1, it follows that $\sup_{i \geq 1} |E(U_{i1}^{\alpha\alpha})| < \infty$. Further-

more, in the view of Lemma 5.21, we have

$$\begin{aligned}
& \left| \frac{1}{N_g} \sum_{i:g_i^0=g} E(U_{i1}^{\alpha\alpha})(\widehat{\alpha}_i(\widetilde{\beta}_g) - \alpha_i^0)^2 - \frac{1}{N_g} \sum_{i:g_i^0=g} E(U_{i1}^{\alpha\alpha})(\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0)^2 \right| \\
&= \left| \frac{1}{N_g} \sum_{i:g_i^0=g} E(U_{i1}^{\alpha\alpha})[\widehat{\alpha}_i(\widetilde{\beta}_g) - \widehat{\alpha}_i(\beta_g^0) + \widehat{\alpha}_i(\beta_g^0) - \alpha_i^0]^2 \right. \\
&\quad \left. - \frac{1}{N_g} \sum_{i:g_i^0=g} E(U_{i1}^{\alpha\alpha})(\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0)^2 \right| \\
&\leq \sup_{i:g_i^0=g} |E(U_{i1}^{\alpha\alpha})| \frac{1}{N_g} \sum_{i:g_i^0=g} \left(|\widehat{\alpha}_i(\widetilde{\beta}_g) - \widehat{\alpha}_i(\beta_g^0)|^2 + 2|\widehat{\alpha}_i(\widetilde{\beta}_g) - \widehat{\alpha}_i(\beta_g^0)||\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0| \right) \\
&\leq \sup_{i:g_i^0=g} |E(U_{i1}^{\alpha\alpha})| \times \left(\frac{1}{N_g} \sum_{i:g_i^0=g} |\widehat{\alpha}_i(\widetilde{\beta}_g) - \widehat{\alpha}_i(\beta_g^0)|^2 \right. \\
&\quad \left. + 2\sqrt{\frac{1}{N_g} \sum_{i:g_i^0=g} |\widehat{\alpha}_i(\widetilde{\beta}_g) - \widehat{\alpha}_i(\beta_g^0)|^2} \sqrt{\frac{1}{N_g} \sum_{i:g_i^0=g} |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0|^2} \right) \\
&= O_p(\|\widetilde{\beta}_g - \beta_g^0\|_2^2 + \|\widetilde{\beta}_g - \beta_g^0\|_2 T^{-1/2}) = o_P(\|\widetilde{\beta}_g - \beta_g^0\|_2). \tag{5.55}
\end{aligned}$$

Direct examination shows

$$\begin{aligned}
& \left| \frac{1}{N_g} \sum_{i:g_i^0=g} E(U_{i1}^{\alpha\alpha})(\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0)^2 - \frac{1}{N_g} \sum_{i:g_i^0=g} \frac{E(U_{i1}^{\alpha\alpha})}{E^2(R_{i1}^\alpha)} \left(\frac{1}{T} \sum_{t=1}^T R_{it} \right)^2 \right| \\
&\leq \sup_{1 \leq i \leq N} |E(U_{i1}^{\alpha\alpha})| \left| \frac{1}{N_g} \sum_{i:g_i^0=g} \left((\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0)^2 - \frac{1}{E^2(R_{i1}^\alpha)} \left(\frac{1}{T} \sum_{t=1}^T R_{it} \right)^2 \right) \right|.
\end{aligned}$$

The second factor on right side of above inequality has bound,

$$\begin{aligned}
& \left| \frac{1}{N_g} \sum_{i:g_i^0=g} \Delta_{NT} |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0| \left(\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0 - \frac{1}{TE(R_{i1}^\alpha)} \sum_{t=1}^T R_{it} \right) \right| \\
&\leq \frac{1}{N_g} \sum_{i:g_i^0=g} \Delta_{NT} |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0| \left| \widehat{\alpha}_i(\beta_g^0) - \alpha_i^0 - \frac{1}{TE(R_{i1}^\alpha)} \sum_{t=1}^T R_{it} \right| \\
&\leq \frac{1}{N_g} \sum_{i:g_i^0=g} \Delta_{NT} |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0| \left(\left| \frac{2}{TE(R_{i1}^\alpha)} \sum_{t=1}^T R_{it} \right| + \Delta_{NT} |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0| \right) \\
&\leq \frac{\Delta_{NT}^2}{N_g} \sum_{i:g_i^0=g} |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0|^2 \\
&\quad + 2\Delta_{NT} \sqrt{\frac{1}{N_g} \sum_{i:g_i^0=g} |\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0|^2} \sqrt{\frac{1}{N_g} \sum_{i:g_i^0=g} \frac{1}{E^2(R_{i1}^\alpha)} \left(\frac{1}{T} \sum_{t=1}^T R_{it} \right)^2}.
\end{aligned}$$

Combining above, we have

$$\left| \frac{1}{N_g} \sum_{i:g_i^0=g} E(U_{i1}^{\alpha\alpha})(\widehat{\alpha}_i(\beta_g^0) - \alpha_i^0)^2 - \frac{1}{N_g} \sum_{i:g_i^0=g} \frac{E(U_{i1}^{\alpha\alpha})}{E^2(R_{i1}^\alpha)} \left(\frac{1}{T} \sum_{t=1}^T R_{it} \right)^2 \right| = o_P(T^{-1}), \quad (5.56)$$

the last rate follow sfrom Lemma 5.21.(iii) and Lemma 5.20.(iv). Combining (5.54), (5.55) and (5.56), we have

$$T_4 = \frac{1}{N_g} \sum_{i:g_i^0=g} \frac{E(U_{i1}^{\alpha\alpha})}{E^2(R_{i1}^\alpha)} \left(\frac{1}{T} \sum_{t=1}^T R_{it} \right)^2 + o_P(\|\widetilde{\beta}_g - \beta_g^0\|_2) + o_P(T^{-1}). \quad (5.57)$$

The desired result follows from (5.49), (5.51), (5.52), (5.53) and (5.57). Proof completed. \blacksquare

Lemma 5.23. *Suppose Assumptions A1, A3, A4 and $N = O(T)$ hold. Then for all $g \in [G^0]$*

$$\sqrt{N_g T}(\widetilde{\beta}_g - \beta_g^0) + \sqrt{N_g/T} W_g^{-1} \Delta_g \xrightarrow{D} N(0, W_g^{-1} D_g W_g^{-1}).$$

Proof. By Lemma 5.14, it follows that

$$\frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T V_i(Z_{it}, \beta_g^0, \alpha_i^0) = \frac{1}{N_g T} \sum_{i:g_i^0=g} \mathcal{I}_i + o_P(1). \quad (5.58)$$

By definition of $\widetilde{\beta}_g$, we find that

$$\frac{1}{N_g T} \sum_{i:g_i^0=g} \sum_{t=1}^T U_i(Z_{it}, \widetilde{\beta}_g, \widehat{\alpha}_i(\widetilde{\beta}_g)) = 0. \quad (5.59)$$

Apply the same argument in Lemma 5.18 and Assumption A4.(b), we can establish following convergence result:

$$\frac{1}{N_g} \sum_{i:g_i^0=g} \left(\frac{\sum_{t=1}^T R_{it}}{\sqrt{T} E(R_{i1}^\alpha)} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\frac{E(U_{i1}^{\alpha\alpha})}{2E(R_{it}^\alpha)} R_{it} - U_{it}^\alpha \right] \right) = \Delta_g + o_P(1). \quad (5.60)$$

Combining Assumption A4.(b), (5.58), (5.59) and Lemma 5.22, we have following equation,

$$\begin{aligned} & -\frac{1}{\sqrt{N_g T}} \sum_{i:g_i^0=g} \sum_{t=1}^T U_i(Z_{it}, \beta_g^0, \alpha_i^0) \\ &= \left[\frac{1}{N_g} \sum_{i:g_i^0=g} \mathcal{I}_i \right] \sqrt{N_g T}(\widetilde{\beta}_g - \beta_g^0) + \sqrt{N_g/T} [\Delta_g + o_P(1)] \\ & \quad + o_P(\sqrt{N_g T} \|\widetilde{\beta}_g - \beta_g^0\|_2) + o_P(\sqrt{N_g/T}). \end{aligned} \quad (5.61)$$

Since $N = O(T)$ by assumption, (5.61) shows the asymptotic distribution of $\tilde{\beta}_g$ is contributed by

$$\left[-\frac{1}{N_g} \sum_{i:g_i^0=g} \mathcal{I}_i \right]^{-1} \frac{1}{\sqrt{N_g T}} \sum_{i:g_i^0=g} \sum_{t=1}^T U_i(Z_{it}, \beta_g^0, \alpha_i^0). \quad (5.62)$$

Next we will derive the asymptotic distribution of (5.62) by Lyapunov C.L.T and Cramer-Wold device. For any $u \in \mathbb{R}^p$, define $\zeta_{Ti} = \sum_{t=1}^T u' U_i(Z_{it}, \beta_g^0, \alpha_i^0) / \sqrt{T}$. By Lemma 5.14 and Lemma 5.1, for some constant $C_u \geq 0$ depending on u , we have

$$\sum_{i:g_i^0=g} E(\zeta_i^3) \leq N_g C_u. \quad (5.63)$$

Direct examination implies

$$\begin{aligned} s_{N_g}^2 &\equiv \sum_{i:g_i^0=g} E(\zeta_i^2) \\ &= \sum_{i:g_i^0=g} u' E \left(\sum_{t=1}^T U_i(Z_{it}, \beta_g^0, \alpha_i^0) \sum_{t=1}^T U_i'(Z_{it}, \beta_g^0, \alpha_i^0) \right) u / T \end{aligned} \quad (5.64)$$

Thanks to Lemma 5.18 and Assumption A4.(a), we can show that

$$\lim_{(N,T) \rightarrow \infty} s_{N_g}^2 / N_g = u' D_g u. \quad (5.65)$$

Combining (5.63), (5.64), (5.65) and Assumption A4.(a), we have

$$\begin{aligned} \lim_{(N,T) \rightarrow \infty} \frac{\sum_{i:g_i^0=g} E(\zeta_i^3)}{s_{N_g}^3} &= \lim_{(N,T) \rightarrow \infty} \frac{N_g C_u}{\left(N_g u' D_g u \right)^{3/2}} \\ &\leq \lim_{(N,T) \rightarrow \infty} \frac{N_g C_u}{\left(N_g B_3 \|u\|_2^2 \right)^{3/2}} = 0. \end{aligned} \quad (5.66)$$

By (5.65), (5.66) and Lyapunov C.L.T., for any $u \in \mathbb{R}^p$, we have

$$\frac{u'}{\sqrt{N_g T}} \sum_{i:g_i^0=g} \sum_{t=1}^T U_i(Z_{it}, \beta_g^0, \alpha_i^0) \xrightarrow{D} N(0, u' D_g u).$$

Since u is arbitrary, by above equation and Assumption A4.(a), it follows that

$$\sqrt{N_g T} (\tilde{\beta}_g - \beta_g^0) + \sqrt{N_g / T} W_g^{-1} \Delta_g \xrightarrow{D} N(0, W_g^{-1} D_g W_g^{-1}).$$

■

Proof of Theorem 3.6. The asymptotic distribution follows from asymptotic equivalence in Lemma 3.5 and Lemma 5.23. ■

Bibliography

- [1] T. Ando and J. Bai. Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*, 31(1):163–191, 2016.
- [2] M. Arellano and S. Bonhomme. Robust priors in nonlinear panel data models. *Econometrica*, 77(2):489–536, 2009.
- [3] M. Arellano and J. Hahn. Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, 43:381, 2007.
- [4] B. H. Baltagi, G. Bresson, and A. Pirotte. To pool or not to pool? In Springer Berlin Heidelberg, editor, *The econometrics of panel data*, pages 517–546. 2008.
- [5] C. A. Bester and C. B. Hansen. Grouped effects estimators in fixed effects models. *Journal of Econometrics*, 190(1):197–208, 2016.
- [6] S. Bonhomme and E. Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184, 2015.
- [7] J.M. Carro. Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, 140(2):503–528, 2007.
- [8] S. Chand. On tuning parameter selection of lasso-type methods-a monte carlo study. pages 120–129. In *Applied Sciences and Technology*, 2012.
- [9] G. Dhaene and K. Jochmans. Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030, 2015.
- [10] J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Science Business Media, LLC, 2003.
- [11] J. Hahn and H. R. Moon. Panel data models with finite number of multiple equilibria. *Econometric Theory*, 26(03):863–881, 2010.
- [12] J. Hahn and W. Newey. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319, 2004.
- [13] C. Hsiao. *Analysis of panel data (No. 54)*. University Press, Cambridge, 2014.
- [14] C. Hsiao and H. Pesaran. Random coefficient panel data models. In *The Econometrics of Panel Data*, pages 187–216. Springer Berlin Heidelberg, 2008.
- [15] C. Hsiao and A. K. Tahmiscioglu. A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association*, 92(438):455–465, 1997.

- [16] Dean R. Hyslop. State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, 67(6):1255–1294, 1999.
- [17] L. A. Kirkland, F. Kanfer, and S. Millard. Lasso tuning parameter selection. pages 49–56. Proceedings of the South African Statistical Association Conference, 2015.
- [18] T. Lancaster. Orthogonal parameters and panel data. *The Review of Economic Studies*, 69(3):647–666, 2002.
- [19] K. Lee, MH. Pesaran, and R. Smith. Growth and convergence in a multi-country empirical stochastic growth model. *Journal of Applied Econometrics*, 12(2):357–392, 1997.
- [20] Y. Lee and P. C. Phillips. Model selection in the presence of incidental parameters. *Journal of Econometrics*, 188(2):474–489, 2015.
- [21] C. C. Lin and S. Ng. Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1(1):42–55, 2012.
- [22] F. Merlevde, M. Peligrad, and E. Rio. A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3):435–474, 2011.
- [23] V. Sarafidis and N. Weber. A partially heterogeneous framework for analyzing panel data. *Oxford Bulletin of Economics and Statistics*, 77(2):274–296, 2015.
- [24] L. Su, Z. Shi, and P. C. Phillips. Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264, 2016.