Quantile Regression with Unobserved Heterogeneity: Advancing Policy Evaluation and Asset Pricing

Ruofan Xu

Bachelor of Science, Australian National University, 2018

Bachelor of Statistics (Honours), Australian National University, 2019

A thesis submitted for the degree of

Doctor of Philosophy

at Monash University in 2024

MONASH University

Department of Econometrics and Business Statistics

Monash University

Melbourne, Australia

2024

Copyright Notice

Abstract

The central theme of this thesis is developing new methods for estimating and forecasting the distributional structure for a set of high-dimensional economics and financial data. This thesis advances the literature by extending the quantile random coefficient models and quantile functional-coefficient models to panel data. We adopt a quantile latent factor structure to flexibly capture the unobserved serial and cross-sectional heterogeneities embedded in the panel data. Applying these methods, we contribute to the empirical literature on policy evaluation and asset pricing.

First, we consider a quantile random coefficient model, where the random coefficients are heterogeneous across both time and cross-sections. We model the random coefficient using a panel regression with interactive fixed effects to simultaneously explain the variation with high-level coefficients and control for the latent heterogeneities. This model is capable of documenting the observed and unobserved heterogeneities at both micro- and macro-levels, as well as the interactions between the two levels. Thus, it is widely applicable to data exhibiting a hierarchical structure; one typical example is survey data. In this thesis, we explore the model's potential in estimating the group-level policy effects, which are allowed to be heterogeneous across individuals based on given observed and unobserved characteristics. We provide the identification strategy for a number of treatment parameters. On top of that, we establish and prove the theorems on the asymptotic validity of the proposed method. As an empirical illustration, the model is applied to evaluate the effect of the minimum wage policy on earnings between 1967 and 1980 in the United States. Our results suggest that (i) minimum wage policy help improve the economic status of low-income black and female workers, and (ii) the minimum wage policy significantly reduces the inequality between sub-population groups but has little effect on narrowing the income disparity within the sub-population.

Second, we consider a quantile functional-coefficient model for high-dimensional panel data, motivated by the time-varying co-movement structure of the large-scale stock return distributions. Quantile functional-coefficient models have been extensively studied for cross-sectional or time-series data. However, the existing models suffer from the so-called "omitted-variable bias" on panel data due to the latent heterogeneities. To this end, we incorporate the quantile functional-coefficient model with a latent quantile factor structure, which controls the unobserved cross-sectional and temporal effects in a flexible way. We develop an iterative estimation procedure, and prove the theoretical and finite-sample validity via both the asymptotic theorems and simulation study. A simple specification test for the constancy of the model coefficients is constructed based on a Wald-type statistic, whose p-value is estimated through a wild bootstrap procedure in finite samples. The proposed test is applied to test the constancy and significance of functional quantile-coefficients in the Fama-French five-factor model. The estimation and testing results advocate the conditional asset pricing model over the unconditional one across different quantile levels. In addition, we find that Fama-French five factors are only sufficient to characterise the stock return when there is no unobserved idiosyncratic shocks, while a more sophisticated model needs to be considered when idiosyncratic shocks appear.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge and belief, contains no material published or written by another person, except where due reference is made in the thesis.

Ruofan Xu

January 18, 2024

Acknowledgements

Firstly, I would like to thank my supervisors Professor Jiti Gao and Professor Tatsushi Oka for being incredible supervisors and mentors who have provided me with wisdom throughout this journey. I am deeply thankful for Professor Jiti Gao for supporting me not only professionally but also at a personal level. Especially, during the COVID-19 lockdown period, he has been constantly motivating me and looking after my health and mental wellbeing.

I would like to thank my milestone evaluation panel, Professor Xibin Zhang, and Professor Xueyan Zhao, for their constructive comments to improve my research findings. I would like to extend my sincere thanks to Professor Gael Martin and Professor Catherine Forbes for being wonderful PhD directors and providing invaluable support to make my candidature fullfilling.

This research was financially supported by the Monash Graduate Scholarship, Monash International Tuition Scholarship, and Faculty of Business and Economics Dean's Postgraduate Research Excellence Award. I am grateful for the financial support that I received. I would also like to take this opportunity to express my sincere gratitude to all the academic and administrative staff members of the Department of Econometrics and Business Statistics at Monash University.

I would like to thank my dearest friends, who supported me in many ways. The last few years would not have been possible without your friendship and support. In particular, to my office mates, the many conversations and social outings we enjoyed together helped keep me sane throughout these years.

It goes without saying that I would like to thank my family members, who have wholeheartedly and unconditionally supported me in all my endeavours throughout my life. I truly believe that without their unwavering support, I would not be here today.

Thank you all.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Large-scale economic and financial data often involve intricate dynamics, demanding sophisticated modelling techniques to capture their nuances. At one time exotic, the use of economic data with both temporal and cross-sectional variation has now become common practice. The term "panel data" refers to any dataset with repeated observations over time for the same cross-sections, whether they be workers, households, firms, industries, or financial instruments, to name a few.

Since the early work of Mundlak (1961), panel data have received ample interest, and methods for the analysis of such data have generated a vast amount of literature. Most of the advances are restricted to understanding the mean structure in the past few decades, while increasing empirical evidence has shown its limitations. Taking policy evaluation as an example, it is common to witness heterogeneous treatment effects across different quantiles. For instance, a minimum wage policy may be effective only for low-income workers.

To obtain a more comprehensive understanding of the distributional structure, quantile regression (QR) methods have been gradually extended to a wide variety of data settings, including panel data, since the seminal work of Koenker and Bassett Jr (1978). Nonetheless, it has proven particularly difficult to apply quantile regression techniques to panel data analysis. One major obstacle arises from the non-linearity and non-smoothness quantile loss function in its parameters.

The main objective of the thesis is to develop methodologies to characterise the distributional structure of high-dimensional panel data, namely both dimensions of time and cross-sections are large, while flexibly controlling for the unobserved heterogeneity through a latent factor structure. This dissertation explores the empirical application in two major fields of economics: (i) micro-economics and (ii) financial economics.

With the publication of the groundbreaking work of Koenker and Bassett Jr (1978), titled "Regression Quantiles", a large body of literature extending beyond mean regression was created. Quantile regression models the relationship between the conditional quantile of the response variable with a set of covariates, providing a more informative regression picture beyond the conditional expectation. Following the induction of Koenker and Bassett Jr (1978), models are generalized and applied in a variety of scientific fields, including finance, economics, ecology, medicine, and more (e.g., Yu and Jones, 1998; Kim, 2007). Models and procedures for panel data have recently received increasing attention in the literature on quantile regression .

The econometric literature on panel data has placed considerable emphasis on modelling heterogeneity. Without controlling for the unobserved effects, the models suffer from the so-called "omitted-variable bias". The literature on panel data quantile regression is scarce until the work by Koenker (2004), who considers a quantile regression with fixed effects. Such a model is highly parametrised and suffers from the incidental-parameter problem; its asymptotic properties were not properly established until the work by Kato et al. (2012). Advances are then made to loosen the model assumptions to accommodate issues with endogeneity(Wüthrich, 2020) and covariate-high-dimensionality (Zheng et al., 2015), among others.

Recently, increasing methodological development has been focused on characterising latent heterogeneity in forms more flexible than the additive fixed effects. This follows the progress of panel factor model in the conditional mean setting (e.g., Pesaran, 2006; Bai, 2003, 2009; Moon and Weidner, 2015). In this piece of literature, the latent factor structure is also known as the interactive fixed effects (IFE). Given the multiplicative form, it allows the time-varying unobserved common shocks affecting cross-sections differently. The literature on quantile factor models remains limited until very recently due to the inherent technical difficulties of quantile estimation caused by the non-smooth quantile objective function. For example, Chen et al. (2021) consider a pure latent factor structure in the quantile setting. Ando and Bai (2020) and Ando et al. (2023) consider a panel quantile regression with an additional latent factor component in the model, and Belloni et al. (2023) then extend the linear regression to high-dimensional cases by combining the quantile loss function with ℓ_1 and nuclear norm regularisation. However, the above literature either ignores the potential covariates or only allows for a linear relationship for the covariates.

In this thesis, we extend the scope of quantile factor model to (i) quantile random-coefficient regression and (ii) quantile functional-coefficient regression, and explore the potential of the IFE in capturing the flexible unobserved heterogeneity in more complex settings.

In Chapter 2, we consider the repeated cross-sectional data at multiple time periods. It forms a hierarchical structure, where the group-level consists of a number of cross-sections across time and the individual-level comprises random samples within each group unit. This hierarchical structure is natural in many research fields, including policy evaluation, which is the empirical focus of this chapter. In such a data-rich environment, heterogeneity poses great challenge in modelling. There are at least three common sources of heterogeneity in the data: (i) cross-sectional heterogeneity due to observable group characteristics; (ii) unobservable individual heterogeneity within groups; (iii) unobservable heterogeneity across groups and time. That said, how to make full use of individual- and group-level information while controlling the three types of heterogeneity is the key focus of this chapter.

To this end, the random-coefficient regression model is prevailing in modelling hierarchical data. One of the fundamental assumptions is that the variation in the marginal effects of individual regressors can be explained by higher-level features. This general assumption not only allows the heterogeneity across groups but also links the individual- and group-level explanations. The advances are largely made for data with only one period. The development in the conditional mean paradigm can be dated back to the 1970s. Swamy (1970) considers a random-coefficient model to estimate the shared group mean and group-specific variation. Hsiao (1974, 1975) takes into account the group-level unobserved heterogeneity by including the group-level fixed effects. Amemiya (1978) and Borjas and Sueyoshi (1994) extend the model by explaining the random-coefficient with group-level covariates, known as slopes-as-outcome regression. In the quantile paradigm, following similar ideas, Chetverikov et al. (2016) study the quantile extension by modelling the individual-level quantile coefficient with group-level covariates. On the other hand, little method can be found for the hierarchical data with multiple periods. In an recent paper by Oka and Yamada (2023), the slopes-as-outcomes regression is extended to data with multiple periods. They characterise the unobserved time and group heterogeneity via two-way fixed effects in the quantile random-coefficients. In Chapter 2, we allow the unobserved temporal and cross-sectional effects to be interactive with each other, which includes the two-way fixed effects as a special case.

In Chapter 3, we relax the linearity assumption in the panel quantile regression and consider a functional-coefficient quantile regression model for panel data. Nonlinear (nonparametric and semiparametric) quantile regression models are extensively studied for cross-sectional data or time series data using sieve approximation (e.g., Kim, 2007; Su and Hoshino, 2016) or kernel method (e.g., Cai and Xu, 2008; Cai and Xiao, 2012). Nevertheless, those methods suffer from "omitted-variable bias" on panel data if there exists latent heterogeneity. To the best of our knowledge, the first attempt to control for the unobserved heterogeneity in a functional-coefficient panel quantile regression is made by Atak et al. (2023). They include an additional latent factor structure to the time-varying quantile regression, and the unobserved factors are assumed to be location-shift from the mean factors. The location-shift assumption largely simplifies the estimation approach and theoretical derivation, but is somewhat restrictive. It forces the number of factors to remain the same across different quantiles, which may not be the case in practice (Ando and Bai, 2020). In this chapter, we consider a functional-coefficient quantile regression model for panel data, and use IFE to control the flexible unobservable cross-sectional and temporal effects. We do not impose any structure on the latent factors, allowing them to be heterogeneous across quantiles and correlated with the observed covariates.

1.2 Contributions

In this thesis, we extend the quantile random coefficient models and quantile functional-coefficient models to panel data. We adopt the quantile latent factor structure, flexibly capturing the unobserved heterogeneity, to avoid the omitted-variable bias. We develop methodologies to estimate these models and establish the theoretical properties for the estimators. We assess the performance of estimation approach using simulation data. These novel econometric frameworks are used to understand the empirical questions of policy evaluation and asset pricing. The improved methods are proposed and studied in two main chapters of the thesis. Below, I summarise the main contributions of Chapters 2 and 3, respectively.

The contributions of Chapter 2 are threefold. First, from the methodological point of view, we consider the repeated cross-sectional data, and propose a quantile random coefficient model, which allows the individual-level coefficients to be heterogeneous across groups and time and explained by higher-level observed and unobserved factors. We adopt a two-step estimation approach, an individual-level quantile regression followed by the group-level panel regression with IFE, to estimate the unknown parameters. The research papers most closely related to ours are Chetverikov et al. (2016) and Oka and Yamada (2023). The former work only considers random coefficients at one particular time point. The latter is an empirical paper and can be viewed as a special case of ours as they only control for the unobservable time and group effects in an additive form. We contribute to the literature of quantile random coefficient models by (i) allowing the random coefficients to be heterogeneous across both groups and time, and (ii) characterising the unobserved group-level effects in a flexible way via IFE.

Second, from the theoretical perspective, we establish the consistency and limiting distribution of the proposed group-level coefficients estimator when the number of observations per group N_{st}, the number of cross-sections S, the number of time units T go to infinity simultaneously. The estimation error from the first-step quantile regression imposes a great challenge in deriving the asymptotic expression of the group-level estimators. However, we show in the proof that the first-step estimation error is negligible, as long as the relative growth rate between individual-level and group-level sample size is sufficiently large.

Last but not least, from the empirical perspective, the model is applied to analyse the group-level policy effects, which are allowed to be heterogeneous across individuals based on given observed and unobserved features. We propose a number of treatment parameters and provide the identification strategy using our model.

The use of policy changes to identify causal effects is widespread in empirical research in economics and other social sciences. Many empirical studies have used policy variations across time and groups, such as states, countries, and industries. For policy evaluation, the IFE approach has been employed to identify average causal effects using a synthetic control approach (Billmeier and Nannicini, 2013), difference-in-differences approach (Hsiao et al., 2012; Kim and Oka, 2014), and the matrix completion method (Athey et al., 2021). For group-level policy evaluation, Gobillon and Magnac (2016) establish identification of the average treatment on the treated, in the presence of interactive fixed effects. On the other hand, the potential of utilising latent quantile factors remains largely unexplored in analysing the distributional treatment effects. We extend the scope of these aforementioned studies by establishing the identification strategy of the distributional group-level policy effect which is heterogeneous across the individual characteristics.

In addition, applying this model, we contribute to the debate of racial income disparity in the United States. According to the census data, only one considerable drop in the income differences between white and black works was witnessed during the past 70 years, occurring during the late 1960s and early 1970s. Several explanations have been discussed in the literature; for example, the impact of federal anti-discrimination legislation (Freeman et al., 1973), improvements in education (Card and Krueger, 1992), and the minimum wage policy effect (Derenoncourt and Montialoux, 2021). Our estimation results support the core conclusion of Derenoncourt and Montialoux (2021) and provide additional quantitative findings that (i) the minimum wage policy leads to a significant reduction on the racial and gender income gap up to the conditional medium; (ii) although the significant positive policy effects concentrate on quantiles up to the medium, resulting in the compression of the conditional wage distribution to some extents, such effects are insignificant from zero.

Chapter 3 contributes to the existing literature from at least the following three aspects. First, from the methodological perspective, our proposed functional-coefficient quantile regression model with IFE enriches both the literature of semi-parametric quantile regression models and quantile factor models. Our model extends the semi-parametric quantile regression models from cross-sectional or time series data to panel data, and flexibly captures the intrinsic latent heterogeneity in panel data. Such a generic model has been extensively studied in the conditional mean setting (see, for example, Dong et al., 2021; Cai et al., 2022); our model can be viewed as the first quantile extension of Cai et al. (2022).

Second, from the theory point of view, we establish the consistency result and central limit theorem for the estimators of functional coefficients. The major technical difficulties arises from the non-smooth quantile objective function. We overcome those technical issues with a generalised function approach.

Third, from the empirical aspect, we contribute to the debate of time-varying betas in the asset pricing literature. For the expected returns or the conditional mean structure, betas have been identified to be business-cycle dependent (e.g., McQueen and Roley, 1993; Jagannathan and Wang, 1996; Celebi and Hönig, 2019) and modelled by the conditional linear factor models, which is known as the conditional asset pricing in literature. However, the potentially business-cycle dependent feature of the quantile co-movement structure of the asset return distributions, to the best of the author's knowledge, remains largely unexplored. In Chapter 3, we apply our model and find supporting evidence for a conditional asset pricing model over the unconditional model across different quantiles.

1.3 Outline of the Thesis

The rest of the thesis is organised as follows.

In Chapter 2, we study the heterogeneous effects of group-level policies. We consider the repeated cross-sectional data, which contain both the individual- and group-level information.

The panel-data-formed group-level information possesses rich group- and time- dynamics in policy evaluation. Together with the individual-level characteristics, data provides us with great potential in analysing the heterogeneous effects at both levels and interactively. In this chapter, we propose a random-coefficient quantile regression framework, using a panel regression framework with latent factor structure to characterise the marginal effects of individual-level covariates. The model is estimated via a two-step approach. We establish the identification results for a number of treatment effect parameters, and drive the consistency and asymptotic normality for the parameter of interests. In the simulation study, we evaluate the finite sample properties of our proposed model and the estimation accuracy of the treatment parameters. In the empirical application, the model is applied to evaluate the effect of the minimum wage policy on earnings between 1967 and 1980 in the United States.

In Chapter 3, we examine the time-varying feature in the distribution of asset returns. We propose a panel quantile functional-coefficient model with latent factors, motivated by the flexible co-movement structure of the large-scale stock return distributions. To accommodate the time-varying influence of the observed risk factors, we allow the regression coefficients to be unknown functions of macroeconomics variables, varying across quantiles. Moreover, a latent factor structure is incorporated to characterise the unobserved common shocks and cross-sectional dependence, which avoids the problem of omitted-variable bias. We develop an iterative estimation procedure, which alternates between the quantile sieve estimation and conventional quantile regression estimation, and establish the asymptotic properties of proposed estimators accordingly. A specification test of the functional coefficient is worthy of consideration, but would also suffer from great theoretical difficulties. We proposed a practical bootstrap estimation method and used it in the empirical study. The relevant theory will be established in the future. In the simulation study, we assess the accuracy of in-sample and out-of-sample estimation of our proposed method with its competitors. Empirically, we re-examine the Fama-French five-factor model and test for time-varying betas in the models.

The final concluding chapter briefly summarises the main findings of the research topics addressed in Chapters 2 and 3, and provides directions for future research.

Chapter 2

Quantile Random Coefficient Regression for Heterogenous Policy Evaluation

2.1 Introduction

The use of policy changes to identify causal effects is widespread in empirical research in economics and other social sciences. Many empirical studies have used policy variations across time and groups, such as states, countries, and industries. A common approach for evaluating group-level policies is to use mean regression with group and time fixed effects. However, standard fixed effects approaches are arguably restrictive, as they can control for only limited unobserved effects and are also unable to document heterogeneous policy effects across individuals. A large body of economic literature has witnessed that cross-sectional units, such as workers, households, and firms, are substantially different in observed and unobserved ways (see Heckman, 2001; Imbens, 2007). The interaction of individual heterogeneities with policy variables potentially has an important role in policy evaluation but remains "somewhat neglected" (see Koenker, 2017; Cox, 1984).

This chapter proposes a flexible yet practically simple quantile random coefficient model for estimating the heterogeneous effect of group-level policies, particularly policy effects that depend on individuals' observed and unobserved characteristics. The proposed method uses repeated cross-sectional data in a two-step estimation procedure. First, to deal with the complex interplay between individual-level covariates and latent individual-level heterogeneities, we estimate the quantile regression model, introduced by Koenker and Bassett Jr (1978), for each pair of group and time. Then, we employ a group-level panel data model to explain variations in the quantile random regression coefficients while controlling for interactive fixed effects, which parsimoniously capture the complex group and time unobserved heterogeneities. Our analysis complements and extends the work of Chetverikov et al. (2016), who first propose a two-step estimation method for a quantile random-coefficient panel regression. Their second step, based on the Hausman and Taylor (1981) approach, deals with endogenous group-level covariates using two-stage least squares. Contrarily, our second step controls for group-level unobserved heterogeneity using interactive fixed effects and can be interpreted as an extension of the difference-in-differences framework.

The advantages of our proposed method are fourfold. First, a distinguishing feature is the ability to capture heterogeneous average policy effects through the interplay of policy variables and individual observed and unobserved characteristics. Quantile regression has the capacity to handle complicated interactions between observed covariates and latent heterogeneity, and document heterogeneous responses of outcomes to variations in covariates. Thus, by using quantile estimates in the first step, our approach can study how policies affect the marginal effects differently among observationally identical individuals, and uncover policy effects depending on individual unobserved heterogeneities. In addition, we document how group-level policies affect outcomes through the interaction of changes in marginal effects with individual observed heterogeneity under the two-level regression framework.

Second, our approach provides a straightforward way to identify the policy effects on inequality measures. Many policies and programs have been introduced to address the issues of economic and social inequality. Measuring the impact of these interventions on inequality is of key importance in determining their effectiveness. Under the quantile regression framework, the conditional quantile spread at two distinctive quantile levels has been used as a within-inequality measure (see Katz and Murphy, 1992; Buchinsky, 1994). Alternatively, the difference between two quantiles conditional on different covariates can be interpreted as a between-inequality measure. Our model captures the policy effects on these inequality measures and the identification of policy effects is established under a type of difference-in-differences assumption.

Moreover, the unobserved group-level heterogeneities are controlled through interactive fixed effects, which is more flexible compared to the standard additive group and time effects. The interactive fixed effects account for time-varying unobserved common shocks with distinct impacts across groups by a factor structure and include the two-way fixed effects as a special case. In the panel data literature, a factor structure has been extensively employed to account for unobserved structures when modeling for the conditional expectation (Bai, 2003) and extended to control for interactive fixed effects (Pesaran, 2006; Bai, 2009; Hsiao et al., 2022). In addition, quantile factor models are only a recently emerging field (Chen et al., 2021; Ando and Bai, 2020). Our paper contributes to the current literature by extending the framework to the identification and estimation of distributional policy effects. For policy evaluation, the interactive fixed effects approach has been employed to identify average causal effects using a synthetic control approach (Billmeier and Nannicini, 2013), difference-in-differences approach (Hsiao et al., 2012; Kim and Oka, 2014), and matrix completion method (Athey et al., 2021). For group-level policy evaluation, Gobillon and Magnac (2016) establish identification of the average treatment on the treated, in the presence of interactive fixed effects. Our work extends the scope of these studies by estimating heterogeneous and distributional policy effects.

Lastly, our approach applies to a broad range of settings, including labour, public finance, health, and development economics. For the evaluation of distributional policy effects, various empirical studies use unconditional quantiles or their variants as a dependent variable, including Lee (1999) for minimum wages on wage, Angrist and Lang (2004) for a school desegregation program on test scores, and Bitler et al. (2006) for welfare reforms on earnings, among others. The method proposed in this study can be used to extend the scope of these analyses by considering the conditional quantiles. We make R codes publicly available to facilitate empirical studies.

We apply this methodology to study the effect of the minimum wage policy on earnings from 1967 to 1980, in the United States, using the Current Population Survey (CPS). The 1966 Fair Labor Standards Act (FLSA) extended federal minimum wage coverage to industrial sectors, where black workers were overrepresented. Using minimum wage variations across industries, Derenoncourt and Montialoux (2021) find that minimum wage policies can play a critical role in reducing racial income disparities on average. They present a variety of mean regression results to convincingly document intricate facets of the policy effects by carefully selecting dependent variables, including log annual wages and their unconditional quantiles, and sub-samples based on workers' characteristics. Our method quantifies such heterogeneous policy effects, particularly on racial income inequality, under a unified framework without the need of alternating responses and selecting sub-samples. In addition, the interactive fixed effects in our model are suitable for controlling for time-varying unobserved common shocks, such as macroeconomic shocks, which could have different effects on different industries. Our estimation results support the core conclusion of Derenoncourt and Montialoux (2021) and provide additional findings that (i) the minimum wage policy leads to a significant reduction on the racial and gender income gap up to the conditional medium, (ii) although the significant positive policy effects concentrate on quantiles up to the medium, which results in the compression of the conditional wage distribution to some extents, such effect is indistinguishable from zero.

This study falls within a broad range of research, accounting for observed and unobserved heterogeneity in the panel data. To explain the individual heterogenities with higher-level information, the classical literature considers a mean regression model with random coefficients (Swamy, 1970; Hsiao, 1975; Djebbari and Smith, 2008). The approach we developed here, building upon ideas from recent studies on quantile regression, goes a step further in enabling the researchers to explore the heterogeneous and distributional effects across individuals. In this regard, the paper most closely related to ours is Chetverikov et al. (2016) and Oka and Yamada (2023). Chetverikov et al. (2016) only consider random quantile coefficients at one particular time point and focuse on dealing with endogenous group-level covariates. On the other hand, we consider a more general setting where the quantile coefficients are heterogenouse across both time and groups. In this case, the model of Chetverikov et al. (2016) is not sufficient to capture the complex temporal and cross-sectional heterogenities, leading to the so-called "omitted-variable bias" for panel regression coefficient at the group-level. Under the same setting as ours, Oka and Yamada (2023) control the group-level latent heterogenities via two-way fixed effects in an empirical study of the recent minimum wage policy effects. Our model extends Oka and Yamada (2023) by allowing the group-level time and cross-sectional heterogeneities to be incteractive with each other. In addition, we present identification conditions for heterogeneous policy effects and first establish the asymptotic theorems for the quantile random coefficients models with random coefficients being heterogeous across both time and groups.

In the absence of the group structure in the data, the literature considers panel quantile regression models with fixed effects (Koenker, 2004; Canay, 2011), correlated random effects (Abrevaya and Dahl, 2008), and interactive fixed effects (Harding and Lamarche, 2014; Ando and Bai, 2020; Chen et al., 2021). We referred to Galvao and Kato (2017) for a recent survey. These studies focus on estimating the effect of individual-level covariates in the presence of individual-level unobservables. In contrast, our focus is on estimating group-level policy effects after controlling for group-level unobservables, and we obtain the theoretical properties of our estimator, allowing for cross-sectional and stationary temporal dependency.

The chapter proceeds as follows. Section 2.2 describes the model and discusses identification of policy effects. Section 2.3 explains estimation methods and Section 2.4 presents the asymptotic properties of the estimators. In Section 2.5, we apply the proposed method to analyse the effect of minimum wage on earning under the 1967 Labor Standards Act. Section 2.6 examines the finite sample properties of the estimator through Monte Carlo simulation. Section 2.7 concludes. Section 2.8 provides additional information about the empirical dataset and establishes the proof of the theorems in the main text, while the technical details are provided in Appendix Section S.1.

2.2 Model and Identification

In this section, we first provide the model setup. We subsequently introduce policy parameters and establish their identification. Before proceeding, we introduce some notations. Throughout the paper, let $Q_{Y|x}(u)$ denote the u^{th} conditional quantile of Y given $X = x$. Let $\|\cdot\|$ denote the Euclidean norm for vectors and the spectral norm for matrices, that is $\|a\| := \sqrt{a'a}$ and $\|A\| := \sup_{a\neq 0}\|Aa\|/\|a\|$ for a column vector a and a matrix A. Let I_p denote the p-dimensional identity matrix, whose dimension varies according to the subscript. Let $1\{\cdot\}$ denote the indicator function. Let e_k be a unit column vector having 1 at the k^{th} entry and 0 for the others, and the dimension of e_k is allowed to vary according to the context. Let $\text{diag}(\cdot)$ denote the diagonal matrix, whose diagonal entries are given in the parentheses. We denote by $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$ for scalars a, b. Let C_M, c_M be some pre-determined positive real numbers which are independent of the sample.

2.2.1 Data and Model

Given group $s = 1, \dots S$ and time $t = 1, \dots, T$, let $\{(y_{ist}, z'_{ist})\}_{i=1}^{N_{st}}$ be repeated cross-sectional observations of a scalar outcome y_{ist} and a $J \times 1$ regressor vector z_{ist} for individual i, which includes a constant 1 if needed, with the sample size N_{st}. We denote supports of y_{ist} and z_{ist} by $\mathcal{Y} \subseteq \mathbb{R}$ and $\mathcal{Z} \subseteq \mathbb{R}^J$, respectively.[1] Also, we observe a $K \times 1$ vector of group-level covariates x_{st}, whose support is $\mathcal{X} \subseteq \mathbb{R}^K$, and a dummy variable d_{st} which takes 1 when some policy is employed in group s and time t and 0 otherwise.

We assume that the repeated cross-sectional observations are randomly sampled for each pair of group and time, while allowing for dependency across the pairs, which can be characterised by both the observed and unobserved group-level information. Specifically, we assume

[1] The supports $\mathcal{Y}$ and $\mathcal{Z}$ can be allowed to depend on group and time, while we suppress the dependency for notational simplicity.

that the u^{th} quantile of the conditional distribution of y_{ist} is given by

$$Q_{y_{ist}|z_{ist},\alpha_{st}}(u) = z_{ist}'\alpha_{st}(u), \quad \alpha_{st}(u) \equiv [\alpha_{1st}(u), \ldots, \alpha_{Jst}(u)]', \tag{2.1}$$

$$\alpha_{jst}(u) = \delta_{jt}(u)d_{st} + x_{st}'\beta_j(u) + f_{jt}(u)'\lambda_{js}(u) + \eta_{jst}(u), \tag{2.2}$$

where $Q_{y_{ist}|z_{ist},\alpha_{st}}(u)$ is the u^{th} conditional quantile of y_{ist} given (z_{ist}, α_{st}) for $u \in \mathcal{U} \subseteq (0,1)$, $\alpha_{jst}(u)$ is the scalar random quantile coefficient corresponding to the j^{th} component of z_{ist}, $\delta_{jt}(u)$ is a scalar coefficient for the policy effect at time t, $\beta_j(u)$ is a $K \times 1$ vector of coefficients, $f_{jt}(u)$ is an $r \times 1$ vector of unobservable macro-level factors, which are heterogeneous across time t, $\lambda_{js}(u)$ is the corresponding factor loading vector, and $\eta_{jst}(u)$ is an idiosyncratic error satisfying $\mathbb{E}[\eta_{jst}(u)|d_{st}, x_{st}, f_{jt}(u), \lambda_{js}(u)] = 0$. For notational simplicity, we write $Q_{st}(u|z_{ist}) \equiv Q_{y_{ist}|z_{ist},\alpha_{st}}(u)$ in what follows.

It is known that, the quantile coefficient $\alpha_{st}(u)$ can be interpreted as the marginal effect of individual covariates z_{ist} on the u^{th} quantile of outcome variables. Moreover, when the underlying structural model depends on multi-dimension unobservables, Sasaki (2015) shows that the quantile regression coefficients can be interpreted as the marginal effects of z_{ist} on y_{ist} averaged over the unobserved variables that satisfy mild regularity conditions. Thus, quantile regression coefficients $\alpha_{st}(u)$ can succinctly summarise the marginal effect of observed individual characteristics on the outcome, averaged over individual unobserved heterogeneity among observationally equivalent individuals within each group and time. That said, although model (2.1) doesn't include an individual fixed effect for each given pair of (s,t), it allows for the unobserved heterogeneity across individuals within the same group-time-pair.

We are interested in studying how these marginal effects depend on the group-level information, especially the group-level policy. To this end, we impose a linear panel regression model (2.2) with interactive fixed effects on each $\alpha_{jst}(u)$. The interactive fixed effects structure $f_{jt}(u)'\lambda_{js}(u)$ accounts for unobserved group and time effects in a flexible way. For example, it captures time-varying macro shocks $f_{jt}(\cdot)$ affecting industry or regions differently via $\lambda_{js}(\cdot)$. Also, the two-way fixed effects model is included as a special case if $f_{jt}(u) = [1, \nu_{jt}(u)]'$ and

$\lambda_{js}(u) = [\phi_{js}(u), 1]'$.

As an example of where the above hierarchical modeling framework is useful, consider the empirical study in this chapter where we aim to quantify the effects of a policy d_{st}, which varies at the industry-by-year-level, on the distribution of individuals' wages y_{ist} across individuals with innate heterogeneities z_{ist}, such as race. Then, the specification of (2.2) allow the researcher to study how the gender effect of being a white/black worker on the wage distribution varies as a function of policy, together with a set of observed and unobserved macro-economic factors.

In addition, considering a special case where (2.2) only applies to the intercept in (2.1) and the group and time fixed effects are additive, model (2.1)-(2.2) is reduced to a two-way fixed effects model, which has been used for estimating distributional policy effects in empirical studies (e.g. Angrist and Lang, 2004), as follows:

$$Q_{y_{ist}|z_{ist},\alpha_{st}}(u) = \delta_{jt}(u)d_{st} + x_{st}'\beta(u) + f_t(u)'\lambda_s(u) + z_{ist}'\gamma_{st}(u).$$

2.2.2 Group-Level Policy Evaluation

To reveal the capacity of the above modelling framework in group-level policy evaluation, we first consider our model under the potential outcome framework, introduced by Rubin (1974), and then provide several policy effect parameters.

In what follows, we consider a binary group-level policy that is employed at a known time T_0 onward with $1 < T_0 < T$. Thus, the sample periods can be divided into the before-period ($t < T_0$) and after-period ($t \geq T_0$). Also, let $d_s = 1$ if group s is treated after T_0 and 0 otherwise. Then, the policy dummy $d_{st} := d_s 1\{t \geq T_0\}$. Let y_{ist}^1 and y_{ist}^0 be the potential outcomes if individual i is exposed to the group-level policy ($d_{st} = 1$) or not ($d_{st} = 0$), respectively. Then, the observed outcome is written as

$$y_{ist} = (1 - d_{st})y_{ist}^0 + d_{st}y_{ist}^1.$$

Given treatment status $d_{st} = d \in \{0, 1\}$, we assume that the u^{th} conditional quantile of the potential outcome y^d_{ist} is given by

$$Q^d_{st}(u|z_{ist}) \equiv Q_{y^d_{ist}|z_{ist},\alpha^d_{st}}(u) = z'_{ist}\alpha^d_{st}(u). \tag{2.3}$$

Moreover, the group-level treatment affects the potential conditional quantile through the potential marginal effects $\alpha^d_{st}(u) \equiv [\alpha^d_{1st}(u), ..., \alpha^d_{Jst}(u)]'$. That is, we specify the j^{th} element of $\alpha^d_{st}(u)$ given the treatment status $d_{st} = d$ as

$$\alpha^d_{jst}(u) = \Delta_{jst}(u)d + x'_{st}\beta_j(u) + f_{jt}(u)'\lambda_{js}(u) + \eta^d_{jst}(u), \quad j = 1, ..., J, \tag{2.4}$$

where $\Delta_{jst}(u)$ is a scalar random policy effect, and $\eta^d_{jst}(u)$ is an error term specific to treatment status d.

Below we introduce the treatment effect parameters. Given the policy status $d \in \{0, 1\}$, we fix individual-level regressors $z \in \mathcal{Z}$ and a probability level $u \in \mathcal{U}$. As a treatment effect parameter, we consider the average quantile treatment effect on the treated (AQTT) at time $t \geq T_0$, which is defined by

$$\Delta^{AQTT}_t(u|z) := \mathbb{E}[Q^1_{st}(u|z) - Q^0_{st}(u|z)|d_s = 1].$$

Since AQTT is a map $(u, z) \mapsto \Delta^{AQTT}_t(u|z)$, we can interpret AQTT as a measure of the policy effects that vary depending on individual-level observed and unobserved heterogeneity, z and u, respectively. AQTT shares a similar concept with Arellano and Bonhomme (2016), who measure the conditional average treatment effect of a nonlinear response by the average conditional quantile treatment effect.

Comparing to the vast literature on quantile treatment effects (e.g., Callaway and Li, 2019; Wüthrich, 2020; Ishihara, 2022), where the conditional quantiles are considered to be identical across groups, and the treatment effects are measured as the difference of the quantile functions of the treated and untreated response, our treatment parameter AQTT is defined as the time-specific difference between the same quantile of the treated and untreated, averaged over treated group, which accommodates the heterogeneity of the conditional quantile function $Q^d_{st}(u|z)$ across group and time.

As alternative measures, we consider spreads of conditional quantile functions to quantify inequality within and between collections of individuals characterised by the individual-level regressors. Given the policy status $d \in \{0, 1\}$, we fix individual-level regressors $z \in \mathcal{Z}$ and consider two probability levels of interest $u_1, u_2 \in \mathcal{U}$ with $u_2 > u_1$. Then, a within-inequality measure under the policy status $d_{st} = d$ at time $t \geq T_0$ is defined as the spread of conditional quantiles:

$$\Delta^{W,d}_{st}(u_1, u_2|z) := Q^d_{st}(u_2|z) - Q^d_{st}(u_1|z).$$

Similarly, we fix individual attributes $z_1, z_2 \in \mathcal{Z}$ and a probability level $u \in \mathcal{U}$ to define a between-inequality measure under the policy status d at time $t \geq T_0$ as

$$\Delta^{B,d}_{st}(u|z_1, z_2) := Q^d_{st}(u|z_2) - Q^d_{st}(u|z_1).$$

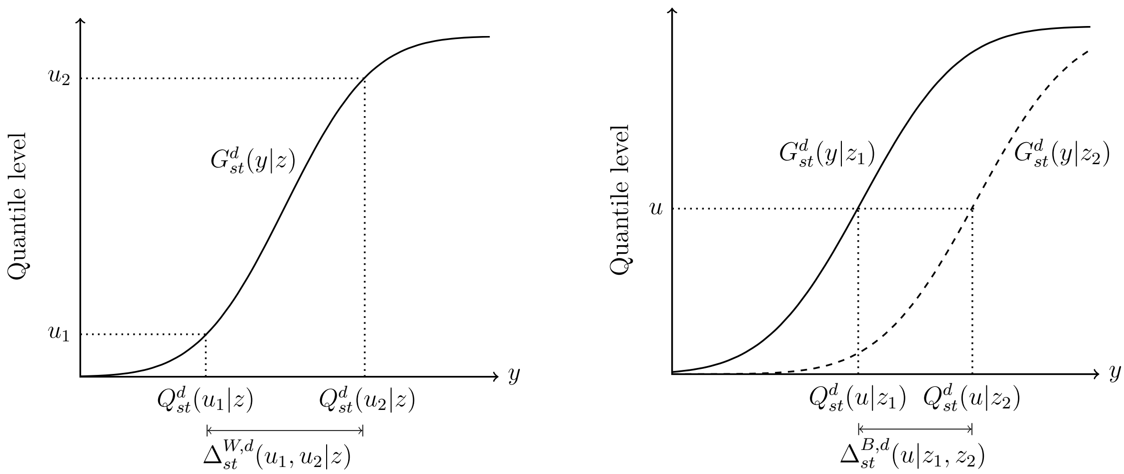
Figure 2.1 illustrates the within- and between-inequalities. The within-inequality measures the dispersion of the distribution of the outcome conditional on individual characteristics z by using two conditional quantile functions. On the other hand, the between-inequality measures the distance between two conditional distributions at a certain probability level.

Group-level policies can affect these inequality measures and their impact can be quantified as changes in the inequality measures at time t averaged over treated groups:

$$\dot{\Delta}^W_t(u_1, u_2|z) := \mathbb{E}\big[\Delta^{W,1}_{st}(u_1, u_2|z) - \Delta^{W,0}_{st}(u_1, u_2|z)|d_s = 1\big],$$
$$\dot{\Delta}^B_t(u|z_1, z_2) := \mathbb{E}\big[\Delta^{B,1}_{st}(u|z_1, z_2) - \Delta^{B,0}_{st}(u|z_1, z_2)|d_s = 1\big].$$

Figure 2.1 Inequality Measures

(a) Within-Inequality (b) Between-Inequality



Notes: Panel (a) illustrates the within-inequality measure as the speared of two conditional quantile functions at quantile levels u_1, u_2, under the treatment status d. Panel (b) shows the between-inequality measure as a distance of two distributions functions conditional on two distinct sets of individual attributes z_1, z_2, given the same quantile level u, under the treatment status d. We denote $G^d_{st}(y|z)$ as the conditional distribution function corresponding to $Q^d_{st}(u|z)$.

2.2.3 Identification

We now exhibits conditions, under which model (2.1)-(2.2) on the observed outcome allows for the identification for group-level policy effect parameters. Under a similar setup, Gobillon and Magnac (2016) prove the identification of average policy effects, using the mean regression model. Let $X_s := [x_{s1}, \ldots, x_{sT}]'$, $F_j(u) := [f_{j1}(u), \ldots, f_{jT}(u)]'$. We make the following assumptions.

Assumption 2.2.1. *For each fixed* $s, t \geq 1$*,*

(i) Individual observations $\{(y_{ist}, z_{ist})\}_{i=1}^{N_{st}}$ *are independent and identically distributed (i.i.d.). The regressor vector* z_{ist} *satisfies* $\|z_{ist}\| < C_M$ *almost surely.*

(ii) All eigenvalues of $\mathbb{E}[z_{1st}z_{1st}']$ *are bounded from below by* $c_M > 0$*.*

Assumption 2.2.1.(i) requires that cross-sectional observations are i.i.d., while allowing for dependency across groups and time, which is also considered in Chetverikov et al. (2016). Assumption 2.2.1.(ii) is a familiar identification condition in regression analysis (Koenker, 2004).

Assumption 2.2.2. *For any fixed* $u \in \mathcal{U}$*,* $s = 1, ..., S$*,* $t = 1, ..., T$*, and for all* $y \in (z_{1st}'\alpha_{st}(u) - c_M, z_{1st}'\alpha_{st}(u) + c_M)$ *with some* $c_M > 0$*,*

(i) The conditional density function $g_{st}(y)$ *is continuously differentiable with the derivative* $g_{st}'(\cdot)$ *satisfying* $|g_{st}'(y)| \leq C_M$ *and* $|g_{st}'(z_{1st}'\alpha_{st}(u))| \geq c_M$*.*

(ii) $g_{st}(y) \leq C_M$*, and* $g_{st}(z_{1st}'\alpha_{st}(u)) \geq c_M$ *for some* $c_M > 0$*.*

Assumptions 2.2.2 is a set of mild regularity conditions that are typically imposed in the quantile regression literature (Koenker, 2004; Chetverikov et al., 2016).

Assumption 2.2.3. *For all* $(s, t) \in \{1, \ldots, S\} \times \{1, \ldots, T\}$*, and each* $j = 1, ..., J$ *and* $u \in \mathcal{U}$*,*

(i) $T^{-1}F_j(u)'F_j(u) = I_r$ *and* $S^{-1}\Lambda_j(u)'\Lambda_j(u)$ *is a positive-definite diagonal matrix.*

(ii) $\mathbb{P}(d_s = 1)$ *is bounded away from zero and one, the eigenvalues of* $\mathbb{E}[x_{st}x_{st}']$ *are bounded away from zero.*

Assumption 2.2.3 provides the identification conditions of the regression coefficients, factor and loadings. Specifically, Assumption 2.2.3.(i) guarantees the identification of factor and loadings up to an orthogonal rotation matrix and column sign change. Such assumption is standard in the literature of both the mean panel factor models (Bai, 2009; Jiang et al., 2021a) and quantile factor models (Ando and Bai, 2020; Chen et al., 2021). Assumption 2.2.3.(ii) ensures the identifiability of the policy parameter $\delta_{jt}(u)$ and regression coefficients $\beta_j(u)$.

Assumption 2.2.4. *For all* $(s, t) \in \{1, \ldots, S\} \times \{1, \ldots, T\}$*, and each* $j = 1, ..., J$ *and* $u \in \mathcal{U}$*,*

(i) $\mathbb{E}[\eta_{jst}^d(u)|d_{st}, X_s, \lambda_{js}(u), F_j(u)] = \mathbb{E}[\eta_{jst}^d(u)|X_s, \lambda_{js}(u), F_j(u)] = 0$ *for* $d \in \{0, 1\}$*,*

(ii) $\mathbb{E}[\Delta_{jst}(u)|d_s = 1, X_s] = \mathbb{E}[\Delta_{jst}(u)|d_s = 1]$*.*

Assumption 2.2.4.(i) requires that the error term for the potential outcome is mean-zero and mean-independent of the treatment status conditional on group-level observed and unobserved variables. Assumption 2.2.4.(ii) is a technical assumption to restrict the relationship between the random policy effect $\Delta_{jst}(u)$ and the group-level regressors.

Assumption 2.2.4.(i) implies a type of parallel-trend assumption. For clarification, we first note that the assumption implies that, for each $j = 1, ..., J$, we have $\mathbb{E}[\eta^0_{jst}(u) - \eta^0_{js,T_0-1}(u)|d_s = 1, X_s, \lambda_{js}(u), F_j(u)] = \mathbb{E}[\eta^0_{jst}(u) - \eta^0_{js,T_0-1}(u)|d_s = 0, X_s, \lambda_{js}(u), F_j(u)]$. By the iterative law of expectation and model (2.3), this further implies $\mathbb{E}[Q^0_{st}(u|z) - Q^0_{s,T_0-1}(u|z)|d_s = 1] = \mathbb{E}[Q^0_{st}(u|z) - Q^0_{s,T_0-1}(u|z)|d_s = 0]$, which states that, on average, the changes in the conditional distribution of untreated potential outcomes does not depend on whether the individual belongs to treated groups or not.

In this chapter, we do not impose the standard yet restrictive rank invariance (Doksum, 1974; Imbens, 2007), or less restrictive rank similarity (Chernozhukov and Hansen, 2005) assumptions on the individual unobserved heterogeneities, which requires an individual's rank in the potential outcome distribution to be the same or has the same probability distribution across treatment status. Instead, our identification results directly rely on the quantile specifications of the potential outcome. However, we do note that, if the rank preservation assumption holds up, AQTT can be additionally interpreted as individual causal effect for (the same) individual at u^{th} quantile before and after treatment, and quantile treatment effect parameters can be identified accordingly.

The theorem below shows that we can identify the time-varying heterogeneous impact of a group-level policy on an individual and the inequality measures using model (2.1)-(2.2).

Theorem 2.2.1. *Suppose that Assumptions 2.2.1-2.2.4 hold. Then, for $t \geq T_0$ and for each $(u, z) \in \mathcal{U} \times \mathcal{Z}$, we have*

$$\Delta^{AQTT}_t(u|z) = z'\delta_{\cdot t}(u).$$

Here, $\delta_{\cdot t}(u) := [\delta_{1t}(u), ..., \delta_{Jt}(u)]'$, whose j^{th} element $\delta_{jt} := \mathbb{E}[\Delta_{jst}(u)|d_s = 1]$ can be identified as $\delta_{jt}(u) = \mathbb{E}[d_{st}\Pi_{st}]^{-1}\mathbb{E}[\Pi_{st}(\alpha_{jst}(u) - f_{jt}(u)'\lambda_{js}(u))]$ with $\Pi_{st} := d_{st} - \mathbb{E}[d_{st}x'_{st}]\mathbb{E}[x_{st}x'_{st}]^{-1}x_{st}$. Furthermore, for $u_1, u_2 \in \mathcal{U}$ and $z_1, z_2 \in \mathcal{Z}$,

$$\dot{\Delta}^B_t(u|z_1, z_2) = (z_2 - z_1)'\delta_{\cdot t}(u) \quad \text{and} \quad \dot{\Delta}^W_t(u_1, u_2|z) = z'\big(\delta_{\cdot t}(u_2) - \delta_{\cdot t}(u_1)\big).$$

The above result shows that we can identify the group-level policy effects which are allowed to vary according to individuals' observed and unobserved characteristics. Taking into account the interplay between a group-level policy and individuals' characteristics, our framework can explicitly identify heterogeneous impacts of the policy across individuals sharing the same observed regressors z and also the impact on the within- and between-inequalities among individuals. To simplify the proof, we treat factor and loadings as observed. When they are unobserved, the iterative estimation approach, proposed in Section 2.3, can be used to obtain their estimators.

2.3 Estimation

For estimation and inference purposes, we write model (2.2) in vector form as follows:

$$A_{js}(u) = D_s\delta_j(u) + X_s\beta_j(u) + F_j(u)\lambda_{js}(u) + \eta_{js}(u),$$

where $A_{js}(u) := [\alpha_{js1}(u), ..., \alpha_{jsT}(u)]'$, $D_s := d_s[e_{T_0}, ..., e_T]$, $\delta_j := [\delta_{jT_0}(u), ..., \delta_{jT}(u)]'$, $\Lambda_j(u) := [\lambda_{j1}(u), ..., \lambda_{jS}(u)]'$, and $\eta_{js}(u) := [\eta_{js1}(u), ..., \eta_{jsT}(u)]'$ are $T \times 1$ vectors, $F_j(u)$ is $T \times r$ matrix of unobservable factors defined above Assumption 2.2.1.

Below we propose an estimation approach for model (2.1)-(2.2). For the estimation of interactive fixed effects, we adopt a common practice of imposing the normalisation restrictions in Assumption 2.2.3.(i), which ensure the identification of $F_j(u)$ and the factor loadings $\Lambda_j(u)$ up to an orthogonal rotation matrix.

For each $u \in \mathcal{U}$ and $j = 1, ..., J$, we propose a two-step estimation procedure to obtain the estimator of $\big(\delta_j(u), \beta_j(u), F_j(u), \Lambda_j(u)\big)$. We use the superscript m to denote the number of iteration in the second step of estimation, and denote the converged estimator as $\big(\widehat{\delta}_j(u), \widehat{\beta}_j(u), \widehat{F}_j(u), \widehat{\Lambda}_j(u)\big)$. The details of the algorithm are given below.

Step 1: Using the cross-sectional data $\{(y_{ist}, z_{ist})\}_{i=1}^{N_{st}}$ for each pair of group and time (s, t) separately, we obtain the estimator $\widehat{\alpha}_{st}(u)$ of $\alpha_{st}(u)$ as the solution of the following minimisation problem:

$$\min_{a \in \mathbb{R}^J} \sum_{i=1}^{N_{st}} \varrho_u(y_{ist} - z'_{ist}a),$$

where $\varrho_u(v) := (u - 1\{v < 0\})v$ for $v \in \mathbb{R}$.

Step 2: Given a collection of the estimators $\widehat{A}_{js}(u) := [\widehat{\alpha}_{js1}(u), ..., \widehat{\alpha}_{jsT}(u)]'$, we obtain the estimator of $\big(\delta_j(u), \beta_j(u), F_j(u), \Lambda_j(u)\big)$ by minimising the following sum of squared residuals:

$$\text{SSR}_u\big(\delta_j, \beta_j, F_j, \Lambda_j\big) := \sum_{s=1}^{S} \big\|\widehat{A}_{js}(u) - D'_s\delta_j - X'_s\beta_j - F_j\lambda_{js}\big\|^2, \tag{2.5}$$

with the normalisation condition in Assumption 2.2.3.(i). The least squares estimators are obtained using an iterated procedure as follows:

(i) We obtain initial estimator $(\widehat{\delta}_j^{(0)}(u), \widehat{\beta}_j^{(0)}(u))$ using the least squares estimator without the factor components. That is, we find the solution to $\min_{(\delta_j,\beta_j)} \sum_{s=1}^{S} \|\widehat{A}_{js}(u) - D_s\delta_j - X'_s\beta_j\|^2$.

(ii) Given $\big(\widehat{\delta}_j^{(m-1)}(u), \widehat{\beta}_j^{(m-1)}(u)\big)$ for $m \geq 1$, we obtain $\big(\widehat{F}_j^{(m)}(u), \widehat{\Lambda}_j^{(m)}(u)\big)$ as the solution to $\min_{(F_j,\Lambda_j)} \text{SSR}_u\big(\widehat{\delta}_j^{(m-1)}(u), \widehat{\beta}_j^{(m-1)}(u), F_j, \Lambda_j\big)$ by applying the principle component analysis (PCA) with the normalisation conditions in Assumption 2.2.3.(i).

(iii) Given $\big(\widehat{F}_j^{(m)}(u), \widehat{\Lambda}_j^{(m)}(u)\big)$, we obtain $\big(\widehat{\delta}_j^{(m)}(u), \widehat{\beta}_j^{(m)}(u)\big)$ as the minimiser of the objective function $\text{SSR}_u\big(\delta_j, \beta_j, \widehat{F}_j^{(m)}(u), \widehat{\Lambda}_j^{(m)}(u)\big)$.

(iv) Repeat (ii)-(iii) until numerical convergence is reached. Specifically, we stop the algorithm if $\big\|\widehat{\delta}_j^{(m)}(u) - \widehat{\delta}_j^{(m-1)}(u)\big\| \leq 10^{-5}$, $\big\|\widehat{\beta}_j^{(m)}(u) - \widehat{\beta}_j^{(m-1)}(u)\big\| \leq 10^{-5}$, and $\big\|\widehat{F}_j^{(m)}(u)\widehat{\Lambda}_j^{(m)}(u)' - \widehat{F}_j^{(m-1)}(u)\widehat{\Lambda}_j^{(m-1)}(u)'\big\| \leq 10^{-5}$.

As the number of factors r is unknown in practice, we adopt a popular eigen-ratio criterion in PCA to select the number of factors. That is, for each m and j, we select the number of factors that minimises the modified eigen-ratio criterion of Casas et al. (2021) as follows:

$$\min_{1\le r\le r_{\max}}\left(\frac{\widehat{\rho}^{(m)}_{j,r+1}(u)}{\widehat{\rho}^{(m)}_{j,r}(u)}\cdot 1\left\{\frac{\widehat{\rho}^{(m)}_{j,r}(u)}{\widehat{\rho}^{(m)}_{j,1}(u)}\ge\frac{1}{\ln\big(S\vee\widehat{\rho}^{(m)}_{j,1}(u)\big)}\right\}+1\left\{\frac{\widehat{\rho}^{(m)}_{j,r}(u)}{\widehat{\rho}^{(m)}_{j,1}(u)}<\frac{1}{\ln\big(S\vee\widehat{\rho}^{(m)}_{j,1}(u)\big)}\right\}\right),$$

where $r_{\max}$ is a pre-specified integer and $\widehat{\rho}^{(m)}_{j,1}(u),...,\widehat{\rho}^{(m)}_{j,T}(u)$ are the estimated eigenvalues of the $T\times T$ matrix $\widehat{L}_j\big(\widehat{\delta}^{(m-1)}_j(u),\widehat{\beta}^{(m-1)}_j(u)\big)$ in descending order, where

$$\widehat{L}_j(\delta,\beta):=\frac{1}{ST}\sum_{s=1}^{S}\big(\widehat{A}_{js}(u)-D_s\delta-X_s\beta\big)\big(\widehat{A}_{js}(u)-D_s\delta-X_s\beta\big)'. \tag{2.6}$$

Since it suffices to set $r_{\max}$ to be relatively large, we choose $r_{\max}$ to be the cardinality of the set $\{\widehat{\rho}^{(m)}_{j,r}(u):\widehat{\rho}^{(m)}_{j,r}(u)>T^{-1}\sum_{r=1}^{T}\widehat{\rho}^{(m)}_{j,r}(u), r=1,\dots,T\}$ in Section 2.5 and Appendix 2.6.

2.4 Assumptions and Asymptotic Properties

In this section, we first introduce the assumptions and then present asymptotic properties of the estimators of the regression coefficients.

2.4.1 Assumptions

In this subsection, we provide assumptions required for deriving asymptotic properties of the recursive estimator along with necessary explanations. In what follows, we consider the case where the set $\mathcal{U}$ is finite, since our empirical application mainly focuses on multiple quantiles and their spreads, instead of the entire distribution.

Assumption 2.4.1. *For all* $(s,t)\in\{1,...,S\}\times\{1,...,T\}$ *and for each* $u\in\mathcal{U}$ *and* $j=1,...,J,$

(i) $\mathbb{E}[||\lambda_{js}(u)||^4]\le C_M$.

(ii) $\mathbb{E}[\eta_{jst}(u)|d_{gl},x_{gl},\lambda_{jg}(u),f_{jl}(u)]=0$ *for all* $(g,l)\in\{1,...,S\}\times\{1,...,T\}$. [2]

[2] We use $\sum_{t\neq l}^{T}$ as a shorthand notation of $\sum_{t=1}^{T}\sum_{l=1,l\neq t}^{T}$ throughout the thesis.

(iii) The largest eigenvalue of the $T \times T$ matrix $\mathbb{E}[\eta_{js}(u)\eta_{js}(u)']$ is bounded uniformly in s and T.

Assumption 2.4.1.(i) requires standard moment conditions for our analysis. Assumption 2.4.1.(ii) and (iii) impose weak restriction on the correlation among the idiosyncratic error components, group-level regressors and common factors. These assumptions are often imposed in the factor model literature (e.g., Bai, 2009; Jiang et al., 2021a).

Assumption 2.4.2. *For any fixed $u \in \mathcal{U}$ and $j = 1, ..., J$,*

(i) For $s = 1, ..., S$, the random sequence $\{\ell_{jst}(u) := (d_{st}, x'_{st}, f'_{jt}(u), \eta_{jst}(u)) : t \geq 1\}$ is a stationary and α-mixing process with mixing coefficient $a_s(\tau)$ and $\tau > 0$. Furthermore, there exists a positive coefficient function $a(\tau)$ such that $\sup_s a_s(\tau) \leq a(\tau)$ and $\sum_{t\neq l}^T a(|t-l|)^{\delta/(4+\delta)} = O(T)$ for such $\delta > 0$ that $\sup_{s,t} \mathbb{E}\left[\|\ell_{jst}(u)\|^{4+\delta}\right] < \infty$.

(ii) For any cross-groups s and g with $s \neq g$, the random sequence $\{(\ell_{jst}(u), \ell_{jgt}(u)) : t \geq 1\}$ is also an α-mixing process with mixing coefficient $a_{sg}(\tau)$ such that $\sum_{s\neq g}^S a_{sg}(0)^{\delta/(4+\delta)}$ $= O(S)$ and $\sum_{s\neq g}^S \sum_{t\neq l}^T a_{sg}(|t-l|)^{\delta/(4+\delta)} = O(ST)$.

(iii) For any cross-groups $s, g, k, m = 1, ..., S$ where $s \neq g \neq k \neq m$, the random sequence $\{(\ell_{jst}(u), \ell_{jgt}(u), \ell_{jkt}(u), \ell_{jmt}(u)) : t \geq 1\}$ is an α-mixing process with mixing coefficient $a_{sgkm}(\tau)$ such that $\sum_{s,g,k,m=1}^S \sum_{t\neq l}^T a_{sgkm}(|t-l|)^{\delta/(4+\delta)} = O\left(S^2T\right)$.

Assumption 2.4.2 uses the notation of "α-mixing" for panel data (e.g., Dong et al., 2015; Jiang et al., 2021a) to capture the temporal dependence exhibited in large panels and controls for the cross-sectional dependency by regulating the mixing coefficient across sections in a concise manner. Alternatively, one can assume the high-order moment conditions employed by Bai (2009).

Assumption 2.4.3. *Let $N_{\min} := \min\{N_{st}, s = 1, ..., S, t = 1, ..., T\}$. As $S, T \to \infty$, we have (i) $T/S \to \kappa > 0$ and (ii) $(ST)^{3/4}(\ln(N_{\min})/N_{\min})^{1/2} \leq C_M$.*

Assumption 2.4.3 controls the diverging rates of the number of group S, time T and individuals per group and time N_{st}. Assumption 2.4.3.(i) is standard in the panel factor model literature (Bai, 2009). Assumption 2.4.3.(ii) requires that the number of individuals per group grows sufficiently fast as S, T jointly go to infinity, such that the estimation error from the quantile estimation in the first-step is negligible. Compared with Assumption 3 of Chetverikov et al. (2016), Assumption 2.4.3.(ii) imposes a more explicit yet comparable growth rate, which is necessary in analysing the limiting property of the interactive fixed-effects estimator.

Assumption 2.4.4. *For any fixed $u \in \mathcal{U}$, as $S, T \to \infty$ jointly,*

(i) $S^{-1}\sum_{s=1}^{S} R_{js}(u)^2 > 0$, *where* $R_{js}(u) := d_s - S^{-1}\sum_{g=1}^{S}\omega_{j,sg}(u)d_g$ *with* $\omega_{j,sg}(u) := \lambda_{jg}(u)'\big(S^{-1}\Lambda_j(u)'\Lambda_j(u)\big)^{-1}\lambda_{js}(u)$.

(ii) The eigenvalues of the following quantities:

$$\text{(a) } (ST)^{-1}\sum_{s=1}^{S} X_s' X_s, \quad \text{(b) } (ST)^{-1}\Big\{\sum_{s=1}^{S} X_s' X_s - \Big(\sum_{s=1}^{S} d_s\Big)^{-1}\sum_{s,g=1}^{S} X_s' D_s D_g' X_g\Big\},$$

$$\text{(c) } \inf_{F: T^{-1}F'F = I_r} (ST)^{-1}\sum_{s=1}^{S} X_s' M_F X_s,$$

are bounded away from zero with probability one, where $M_F := I_T - T^{-1}FF'$.

Assumption 2.4.4 is a technical assumption that guarantees that the inverse matrix in the initial and recursive estimators of the regression coefficients are well-defined, so that the estimation in the second step is valid. Similar assumptions are adopted in Bai (2009) and Jiang et al. (2021a), etc.

Assumption 2.4.5. *For each $u \in \mathcal{U}$, $j = 1, ..., J$ and recursive step $m \geq 1$, there exist positive definite matrices Σ_{FF_j} and Σ_0 such that, as $S, T \to \infty$,*

(i) $\widehat{F}_j^{(m)}(u) \in \{F \in \mathbb{R}^{T\times r} : T^{-1}F'F = I_r, T^{-2}F_j(u)'FF'F_j(u) \to \Sigma_{FF_j}\}$;

(ii) $(\widehat{\delta}_j^{(m)}(u), \widehat{\beta}_j^{(m)}(u)) \in \{(\delta, \beta) \in \mathbb{R}^{T-T_0+1}\times\mathbb{R}^K : F_j^{(m)}(u)'\widehat{L}_j(\delta,\beta)F_j^{(m)}(u) \xrightarrow{p} \Sigma_0\}$, *where* $\widehat{L}_j(\delta, \beta)$ *is defined as (2.6).*

Assumption 2.4.5.(i) is required for deriving a closed-form expression for the recursive formula of $(\widehat{\delta}_j^{(m)}(u), \widehat{\beta}_j^{(m)}(u))$, so that the CLT can be established accordingly, and (ii) is a technical assumption required in the derivations, which ensures the invertibility of $\widehat{V}_j^{(m)}(u) := \text{diag}\big(\widehat{\rho}_{j,1}^{(m)}(u), ..., \widehat{\rho}_{j,r}^{(m)}(u)\big)$.

Assumption 2.4.6. *For $u \in \mathcal{U}$, the following rates hold:*

(i) $\mathbb{E}\| \sum_{s=1}^{S} X_s X_s' \|^2 = O(S^2 T^2)$,

(ii) $\mathbb{E}\| \sum_{s=1}^{S} X_s' F_j(u) \lambda_{js}(u) \|^2 = O(ST)$,

(iii) for all $t \geq T_0$, $\mathbb{E}\| \sum_{s=1}^{S} d_s f_{jt}(u)' \lambda_{js}(u) \|^2 = O(S)$ *uniformly.*

Assumption 2.4.6 guarantees the desirable rates of the regression coefficients. In particular, Condition (i) follows trivially when $\max_{s,g,t,l} |\mathbb{E}(x_{st}' x_{gt} x_{sl}' x_{gl})| \leq C < \infty$. Conditions (ii) and (iii) hold when the factor component is independent of the group-level regressors and satisfies $\mathbb{E}[\lambda_{js}(u)' f_{jt}(u)] = 0$.

In the assumption below, we impose a condition on the $J \times 1$ vector $K_t(u)$, whose j^{th} element is given by $K_{jt}(u) := \big(S^{-1} \sum_{s=1}^{S} R_{js}(u)^2\big)^{-1} S^{-1/2} \sum_{s=1}^{S} R_{js}(u) \eta_{jst}(u)$ with $R_{js}(u)$ defined in Assumption 2.4.4(i).

Assumption 2.4.7. *For any $u_1, u_2 \in \mathcal{U}$ and $t \geq T_0$, as $S \to \infty$, we have*

$$\begin{bmatrix} K_t(u_1) \\ K_t(u_2) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \Sigma_t(u_1, u_1), \Sigma_t(u_1, u_2) \\ \Sigma_t(u_2, u_1), \Sigma_t(u_2, u_2) \end{bmatrix}\right),$$

where $\mathbf{0}$ is a $(2J) \times 1$ vector and $\Sigma_t(u_1, u_2) := \lim_{S \to \infty} \mathbb{E}[K_t(u_1) K_t(u_2)']$.

Assumption 2.4.7 is required to derive the joint Central Limit Theorem (CLT) in Theorem 2.4.2 for the converged estimator of the policy parameter, given quantile levels $u_1, u_2 \in \mathcal{U}$. The assumption shares the same idea as Assumption E of Bai (2009).

2.4.2 Consistency and Limiting Distribution

In this section, we establish asymptotic properties for estimators of $(\delta_j(u), \beta_j(u))$. We first provide the convergence rate of the recursive estimators of both the policy parameter $\delta_j(u)$ and the regression coefficient $\beta_j(u)$. Then, we establish the joint central limit theorem (CLT) for the converged estimator of the policy parameter. To avoid distraction from the key parameter of interests, we do not present the CLT for $\beta_j(u)$, which can be derived in the same manner as that of $\delta_j(u)$. Finally, for empirical analysis, we derive the corresponding consistent estimator of the empirical limiting distribution.

Theorem 2.4.1. *Suppose that Assumptions 2.2.1-2.2.3, 2.4.1-2.4.6 hold. Then, for any fixed $u \in \mathcal{U}$, $j = 1, ..., J$ and $m \geq 0$, as $S, T \to \infty$, we have*

(i) $\sqrt{S}\big(\widehat{\delta}_{jt}^{(m)}(u) - \delta_{jt}(u)\big) = O_P(1)$ *for each $t \geq T_0$,*

(ii) $\sqrt{ST}\big(\widehat{\beta}_j^{(m)}(u) - \beta_j(u)\big) = O_P(1)$.

The time-varying policy effects are estimated for each time period following the policy intervention. The convergence rate of the policy effect estimator depends solely on the group size S. On the other hand, the remaining regression coefficients are estimated using the full sample, and their convergence rate depends on ST.

Next, as inference of the converged estimators is of key interest, we establish the joint CLT for the converged estimator $(\widehat{\delta}_t(u_1)', \widehat{\delta}_t(u_2)')'$ in the following theorem,

Theorem 2.4.2. *Suppose that Assumptions 2.2.1-2.2.3, 2.4.1-2.4.7 hold. Let $\widehat{\delta}_{\cdot t}(u) := [\widehat{\delta}_{1t}(u), ..., \widehat{\delta}_{Jt}(u)]'$ denote the converged estimator of $\delta_{\cdot t}(u)$. Then, for any $u_1,\ u_2 \in \mathcal{U}$ and $t \geq T_0$, we have, as $S, T \to \infty$,*

$$\sqrt{S}\begin{pmatrix} \widehat{\delta}_{\cdot t}(u_1) - \delta_{\cdot t}(u_1) \\ \widehat{\delta}_{\cdot t}(u_2) - \delta_{\cdot t}(u_2) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} B_t(u_1) \\ B_t(u_2) \end{bmatrix}, \begin{bmatrix} \Sigma_t(u_1,u_1), \Sigma_t(u_1,u_2) \\ \Sigma_t(u_2,u_1), \Sigma_t(u_2,u_2) \end{bmatrix}\right),$$

where $\Sigma_t(u_1, u_2)$ is defined in Assumption 2.4.7, and $B_t(u) := \operatorname*{plim}_{S,T\to\infty} [\widetilde{B}_{1t}(u), ..., \widetilde{B}_{Jt}(u)]'$ is the bounded asymptotic bias, whose j^{th} component is given by

$$\widetilde{B}_{jt}(u) := -\left(\frac{1}{S}\sum_{s=1}^{S} R_{js}(u)^2\right)^{-1} \frac{1}{S^{3/2}T}\sum_{s,g=1}^{S} d_s \mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)']F_j(u)\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\lambda_{js}(u),$$

for $j = 1, ..., J$, where $R_{js}(u)$ is defined in Assumption 2.4.4(i).

In view of this theorem, under general cases, the asymptotic distribution of the recursive estimator $\widehat{\delta}_{jt}(u)$ depends on: (i) the quantiles, (ii) the accuracy of the first-step estimation (i.e., $\widehat{\alpha}_{jst}(u) - \alpha_{jst}(u)$), (iii) the consistency of the initial estimation of the second-step, and (iv) the degenerating estimation error of the regression coefficient, factor and loadings carried over from the iterative steps. Therefore, although the CLT is derived per given time t, we still require $S, T \to \infty$ jointly, as the convergence of the policy parameter estimator relies on the convergence of the estimators of the factors and loadings, which only hold when $S, T \to \infty$ jointly. In addition, we note that the estimation error of the first-step $\widehat{\alpha}_{jst}(u) - \alpha_{jst}(u)$ depends on the sample size of individual observations within each group-time (N_{st}). Hence, by controlling the relative growth rate between individual-level and group-level sample size (Assumption 2.4.3.(ii)), the asymptotic first-step estimation error becomes negligible in the asymptotic representation of $\widehat{\delta}_{jt}(u)$, and the estimation error from the panel regression contributed to the asymptotic bias and covariance, similar to Bai (2009). We note that the growth rate of $N_{\min}$ relative to S and T can be relaxed, but at the expense of more complicated asymptotic expressions.

According to the identification Theorem 2.2.1, we define the estimators of the treatment parameters as $\widehat{\Delta}_t^{AQTT}(u) := z'\widehat{\delta}_{\cdot t}(u)$, $\widehat{\dot{\Delta}}_t^{B}(u|z_1, z_2) := (z_2 - z_1)'\widehat{\delta}_{\cdot t}(u)$, $\widehat{\dot{\Delta}}_t^{W}(u_1, u_2|z) := z'(\widehat{\delta}_{\cdot t}(u_2) - \widehat{\delta}_{\cdot t}(u_1))$. By virtue of Theorem 2.4.2, the limiting distribution of the treatment parameters can be established accordingly by linear combinations of the individual-level covariates z and the policy coefficient $\delta_t(u)$ as Corollary 2.4.1, whose proof is straightforward and thus omitted.

Corollary 2.4.1. *Under the assumptions of Theorem 2.4.2, for any $u, u_1,\ u_2 \in \mathcal{U}$, $z, z_1, z_2 \in \mathcal{Z}$ and $t \geq T_0$, we have, as $S, T \to \infty$,*

$$\sqrt{S}\Big(\widehat{\Delta}_t^{AQTT}(u|z) - \Delta_t^{AQTT}(u|z)\Big) \xrightarrow{d} \mathcal{N}\Big(z'B_t(u), z'\Sigma_t(u,u)z\Big),$$

$$\sqrt{S}\Big(\widehat{\dot{\Delta}}_t^{B}(u, |z_1, z_2) - \dot{\Delta}_t^{B}(u|z_1, z_2)\Big) \xrightarrow{d} \mathcal{N}\Big(\big(z_2 - z_1\big)'B_t(u), \sigma_B^2(u, z_1, z_2)\Big),$$

$$\sqrt{S}\Big(\widehat{\dot{\Delta}}_t^{W}(u_1, u_2|z) - \dot{\Delta}_t^{W}(u_1, u_2|z)\Big) \xrightarrow{d} \mathcal{N}\Big(z'\big(B_t(u_2) - B_t(u_1)\big), \sigma_W^2(u_1, u_2, z)\Big),$$

where $B_t(u)$ is defined in Theorem 2.4.2, $\sigma_B^2(u, z_1, z_2) := (z_2 - z_1)'\Sigma_t(u,u)(z_2 - z_1)$ and

$\sigma^2_W(u_1, u_2, z) := z'\big(\Sigma_t(u_1, u_1) - \Sigma_t(u_1, u_2) - \Sigma_t(u_2, u_1) + \Sigma_t(u_2, u_2)\big)z$ *with* $\Sigma_t(u_1, u_2)$ *in Assumption 2.4.7.*

Lastly, for inferential propose, we show that the asymptotic bias and covariance can be consistently estimated by their empirical counterparts in the next corollary. Attention is paid to heteroscedasticities in both time and cross-sectional dimensions, assuming no correlation in both dimensions to simplify the presentation.

Define $\widehat{B}_t(u) := [\widehat{B}_{1t}(u), ..., \widehat{B}_{Jt}(u)]'$, where, for $j = 1, ..., J$,

$$\widehat{B}_{jt}(u) := -\left(\frac{1}{S}\sum_{s=1}^{S}\widehat{R}_{js}(u)^2\right)^{-1}\cdot\frac{1}{S^{3/2}T}\sum_{s,g=1}^{S}d_s\big(\widehat{\eta}_{jgt}(u)\big)^2\widehat{f}_{jt}(u)'\left(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1}\widehat{\lambda}_{js}(u),$$

where $\widehat{R}_{js}(u) := d_s - S^{-1}\sum_{g=1}^{S}d_g\widehat{\lambda}_{jg}(u)'\big(S^{-1}\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)\big)^{-1}\widehat{\lambda}_{js}(u)$.

We construct an estimator of the asymptotic covariance matrix $\Sigma_t(u_1, u_2)$, denoted as $\widehat{\Sigma}_t(u_1, u_2)$, by its empirical counterpart. $\widehat{\Sigma}_t(u_1, u_2)$ is a $J \times J$ block matrix, whose $(j, k)^{\text{th}}$ block is given by

$$\left(\frac{1}{S}\sum_{s=1}^{S}\widehat{R}_{j,s}(u_1)^2\right)^{-1}\left(\frac{1}{S}\sum_{s=1}^{S}\widehat{R}_{k,s}(u_2)^2\right)^{-1}\frac{1}{S}\sum_{s=1}^{S}\widehat{R}_{j,s}(u_1)\widehat{R}_{k,s}(u_2)\widehat{\eta}_{jst}(u_1)\widehat{\eta}_{kst}(u_2).$$

The below corollary establishes that the proposed estimators of the asymptotic bias and covariance are consistent.

Corollary 2.4.2. *Suppose that the assumptions of Theorem 2.4.2 hold. In addition, we assume that for any fixed* $u_1, u_2 \in \mathcal{U}$ *and* $j, k = 1, ..., J$, *for* $t, l \in \{1, ..., T\}$ *and* $s, g \in \{1, ..., S\}$, $\mathbb{E}[\eta_{jst}(u_1)\eta_{kgl}(u_2)|D_s, D_g, W_s, W_g, \Lambda_j(u_1), F_j(u_1), \Lambda_k(u_2), F_k(u_2)] = 0$ *if* $s \neq g$ *or* $t \neq l$. *Then, for any given* $u_1, u_2 \in \mathcal{U}$ *and* $t \geq T_0$, *we have* $\widehat{B}_t(u) \xrightarrow{p} B_t(u)$, *and* $\widehat{\Sigma}_t(u_1, u_2) \xrightarrow{p} \Sigma_t(u_1, u_2)$ *as* $S, T \to \infty$.

Corollary 2.4.2 assumes that idiosyncratic errors are uncorrelated across groups and over time, after conditioning group-level regressors and interactive fixed effects. For correlated idiosyncratic errors, Bai (2009) provides some conjectures for bias-correction and covariance estimators using the partial sample method together with the Newey-West procedure.

Using Corollary 2.4.1, a bias-corrected estimator can be constructed as

$$\widehat{\delta}^{\dagger}_{\cdot t}(u) := \widehat{\delta}_{\cdot t}(u) - \frac{1}{\sqrt{S}}\widehat{B}_t(u). \tag{2.7}$$

We also note that there are other bias correction procedures, such as panel jackknife (Hahn and Newey, 2004) and bootstrap method (Kim and Sun, 2016) that could be use. In addition, together with Corollary 2.4.2, it is straightforward to construct the bias-corrected estimators and confidence intervals for the time-varying AQTT, $\Delta^{AQTT}(u|z)$, and changes in between- and within-inequality, $\dot{\Delta}^B(u|z_1, z_2)$ and $\dot{\Delta}^W(u_1, u_2|z)$ accordingly.

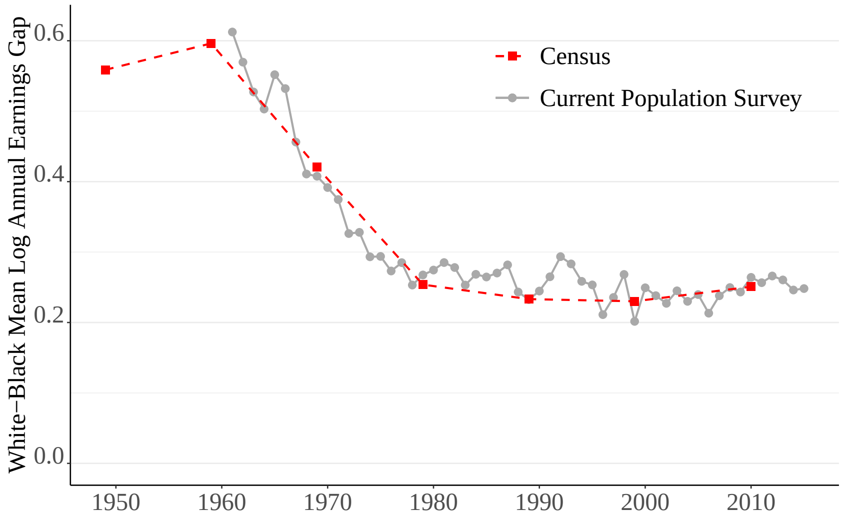
2.5 Empirical Analysis

2.5.1 Background on Racial Income Inequality

Racial economic inequalities have persisted in the United States over long periods of time. Among these inequalities, the income gap between black and white workers is evident. As in Figure 2.2, the income gap, measured by the average annual earnings, was around 20-30% for the last two decades, whereas the gap significantly dropped during the late 1960s and early 1970s. The empirical literature has explored factors that narrowed the racial income gap during those periods, including federal anti-discrimination legislation (Smith and Welch, 1984) and improvements in education (Smith and Welch, 1977; Card and Krueger, 1992).

Recently, Derenoncourt and Montialoux (2021) put forward a new explanation: the extension of the federal minimum wage to some industries. The 1966 FLAS established a federal minimum wage (effective February 1967) in previously unregulated industries, which employed about 20% of the total workforce in the US and nearly a third of all black workers. They evaluate the minimum wage policy effect on earning, using a cross-industry difference-in-differences design, in which seven treated and eight control industries were subject to the minimum wage under the 1966 and 1938 FLSA, respectively. Additional information and background about the dataset are provided in Section 2.8.1.

Figure 2.2 White-Black Unadjusted Wage Gap in the Long Run



Notes: This figure is a replication of Figure 1 in Derenoncourt and Montialoux (2021). The data sources are the Current Population Survey 1962–2016, U.S. Census from 1950 to 2000, and American Community Survey data in 2010 and 2017. Sample includes black or white adults aged 25–65, who worked more than 13 weeks last year, worked three hours last week, do not live in group quarters, are not self-employed and not unpaid family worker with no missing industry or occupation code.

Using repeated cross-sections of black and white workers aged between 25 and 55 for year 1961 and 1963–1980, extracted from March CPS,[3] they estimate the following two-way fixed effects mean regression model:

$$y_{ist} = \delta_0 + \delta_t d_{st} + z'_{ist}\beta + \phi_s + \nu_t + \eta_{ist}, \tag{2.8}$$

for worker i in industry $s = 1, \ldots, 15$ and time $t = 1961, 1963 \ldots, 1980$. Here, y_{ist} is the log annual earnings deflated by annual CPI-U-RS (\$2017)[4], and d_{st} denotes a dummy variable taking 1 if industrial sector s is subject to the federal minimum wage and 0 otherwise. Also, z_{ist} is a vector of worker's characteristics and unobserved random variables consist of industry fixed effect ϕ_s, time fixed effect ν_t and an idiosyncratic error η_{ist}. The parameter of interest is δ_t, which measures dynamic policy effects.

Their result shows that, after controlling for individual characteristics, the average wage of

[3]Since The March CPS of year t contains information in calendar year $t-1$, the data source is the 1962, 1964–1981 March CPS. The 1963 March CPS is excluded due to the lack of observations and missing demographic information.

[4]Using March CPS data the 1960s and early 1970s, we only directly observe annual earnings, but not hourly wages, whereas the CPS contains more detailed individual worker–level information the Bureau of Labor Statistics data. See Section III.B. of Derenoncourt and Montialoux (2021).

workers in the newly covered industries is around 5% higher relative to that in control industries in 1967–1980 compared with the pre-period 1961–1966, and the effect of minimum wage reform on workers' log-earning is more than twice as large for black workers as that for white workers on average. In addition to the above regression, they present several regression results to uncover intricate facets of the effects of minimum wage by taking various variables as the dependent variable in (2.8), including log annual wages or its unconditional quantiles, and also selecting sub-samples based on workers' characteristics.

2.5.2 Model and Practical Implementation

We analyse time-varying policy effects from 1967 to 1980 at quantile $u \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. Following the original model in (2.8), we consider the following u^{th} quantile regression model in (2.1) with $\alpha_{st}(u) = [\alpha_{1st}(u), \ldots, \alpha_{Jst}(u)]'$, where, for $j = 1, \ldots, J$,

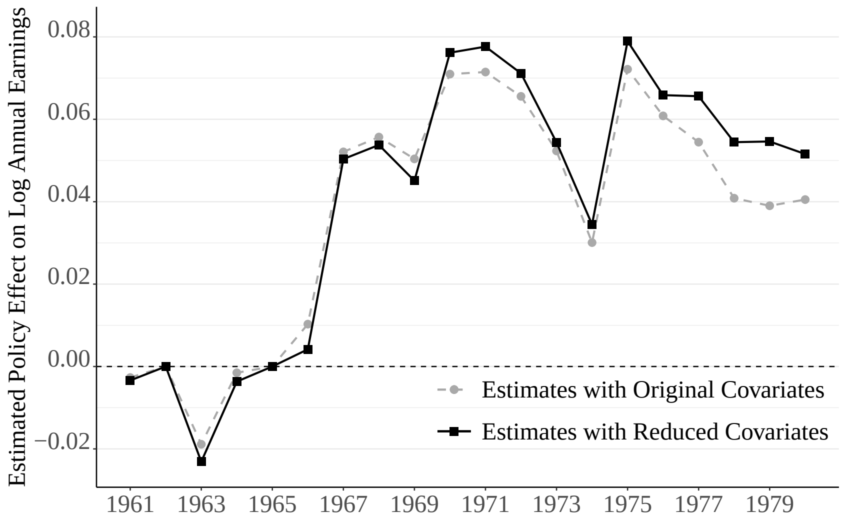
$$\alpha_{jst}(u) = \delta_{j0}(u) + \delta_{jt}(u)d_{st} + f_{jt}(u)'\lambda_{js}(u) + \eta_{jst}(u). \tag{2.9}$$

Here, a set of coefficients $\{\delta_{jt}(u)\}_{t=1967}^{1980}$ measures the time-varying policy effect.

For simplicity of interpretation, we treat some of the original covariates as ordered variables rather than dummy variables for categories. More precisely, z_{ist} includes a constant 1, dummy variables for race (white/black), gender (male/female) and work type (full-time/part-time), and ordered variables including years of schooling, experience, experience squared, the number of weeks worked in a year, and the number of hours worked in a week[5]. This selection yields a very similar result of the mean-regression in (2.8) as the one in Derenoncourt and Montialoux (2021). See Figure 2.3. In this empirical study, we are particularly interested in the heterogeneous policy-effects varies between gender and race.

[5]Derenoncourt and Montialoux (2021) use dummy variables to control for the number of weeks worked in a year and the number of hours worked in a week, because hourly wage is not available in the CPS data during the periods of interest.

Figure 2.3 Estimation Results of Mean Regression



Notes: We plot the estimates of time-varying policy effect δ_t in Model (2.8) given the original set of controlled variables (in dashed grey line) and the estimates given the reduced set of covariates that used in our empirical analysis (in solid black line).

2.5.3 Results Analysis

In Figure 2.4, we present time-varying policy effects $\delta_{jt}(u)$ in (2.9) with 95% confidence intervals for the categorical individual-level covariates. The policy effects for the continuous covariates: eduction and experience are insignificant across quantiles. That said, after taken the differentiation between gender and race into account, the policy effect is insignificant across the levels of skill (measured by education and experience). For presentation conciseness, those figures are omitted. Panel (a) reports the effects on the intercept coefficients, which correspond to white, male, full-time workers, which are insignificantly different from zero for most of the estimates. Panel (b) shows statistically significant positive policy effects for black workers in majority of the years across all quantiles. Especially, the policy effects are most significant at the 0.1th conditional quantile, which are 15–20$\%$ (0.15–0.2 log points).

Panel (c) of Figure 2.4 presents the estimated policy effects on the female dummy's coefficient. The effects in the late 1970s are positive and significant up to 0.7th quantile with a magnitude of 5$\%$. However, caution is warranted in interpreting the results, which may be an integrated impact of the 1966 Fair Labor Standards Act and two pieces of important legislation that targeted labor market discrimination against women: the Equal Pay Act of 1963 and Title

VII of the Civil Rights Act of 1964.[6] Although the gender gap of median wages for full-time, full-year workers was unchanged over the 1960s–1970s (see Blau and Kahn, 2017), Bailey et al. (2021) recently document sharp increases in women's wages relative to men's below median during the 1960s. They underscore the importance of minimum wage policy and the laws to target gender-based workplace discrimination. Our result is consistent with their findings and furthermore suggests long-term positive effects even at 0.5th and 0.7th quantiles, conditional on individual attributes.

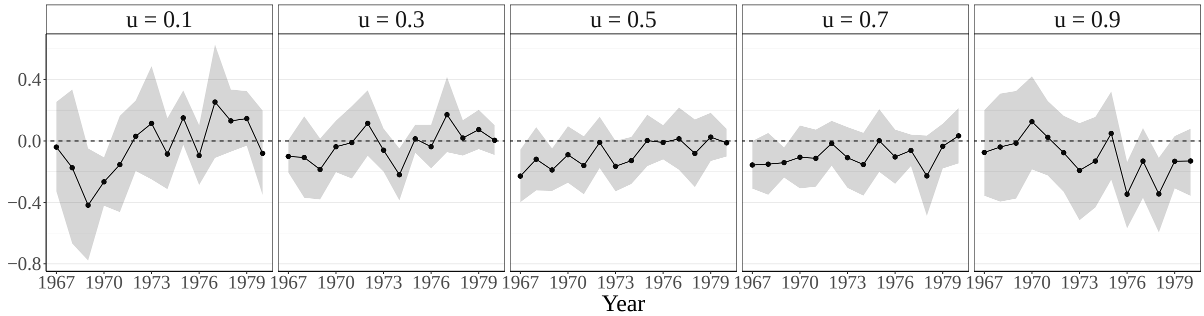
In Figure 2.5, we present estimated policy effects on changes in the within-inequality measure $\dot{\Delta}_t^W(u_1, u_2|z)$ in the year 1980. For quantile pairs $(u_1, u_2) = (0.1, 0.9)$ or $(0.1, 0.5)$, we measure how much the minimum wage policy changes the conditional quantile spread. In the following analysis, we consider the labors with average skill level (12 years of education and 20 years of experience), while reporting three pairs of categorical individual attributes, as shown in the horizontal axis. The estimates suggest that the introduction of minimum wage reduces the within-inequality, while the reduction is statistically indistinguishable from 0 for all groups. In addition, we also note that the same conclusion can be drawn when considering less-skillful or more-skillful labors within the three sub-populations considered in Figure 2.5, since the estimated $\delta_{jt}(u)$ associated with education and experience are insignificantly different from 0.

Figure 2.6 reports the policy effects on the changes in the between-inequality measure, $\dot{\Delta}_t^B(u|z_1, z_2)$, for $t = 1980$ and $u \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The baseline z_2 is fixed to include white, male workers and z_1 changes over the three pairs of categorical attributes as in Figure 2.5, while the remaining variables are the same in z_1 and z_2. Here, we note that the identification result in Theorem 2.2.1, $\dot{\Delta}_t^B(u|z_1, z_2) = (z_2 - z_1)'\delta(u)$. Thus, the common values in z_1 and z_2 cancel each other and do not affect the conclusions. Panels (a) suggests significant negative impacts on the between-inequality for black, male workers, comparing to white, male workers, with magnitudes 5–20% for all quantiles excepts the 0.7 conditional quantile. Panels (b) also shows reduction (0.05–0.20) in the between-inequality for white, female workers at the 0.7th

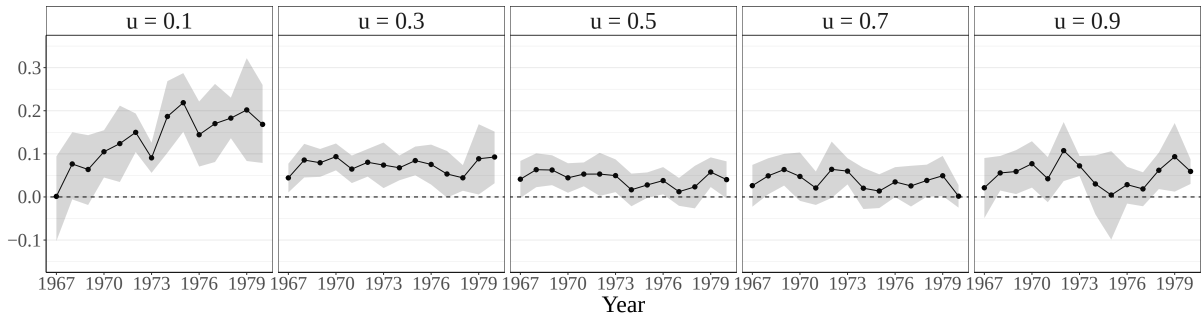
[6]The Equal Pay Act of 1963 is a federal law that amends the Fair Labor Standards Act and prohibits wage disparity based on gender. Title VII of The Civil Rights Act of 1964 more broadly prohibits discrimination in employment on the basis of race, colour, religion, national origin, and gender.

Figure 2.4 Time-Varying Policy Effect Estimates (δ_{jt})

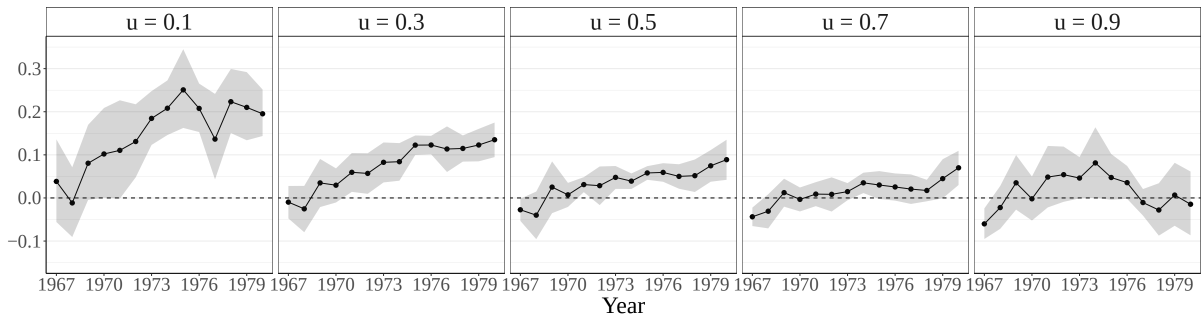
(a) Intercept (Baseline: White, Male, Full-Time Workers)



(b) Black



(c) Female



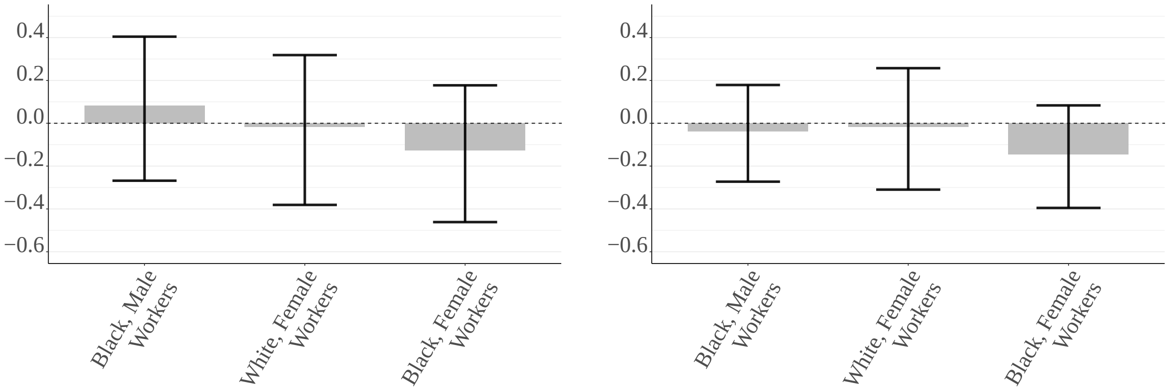
Notes: Panels (a)-(c) present the estimated time-varying marginal policy effect $\delta_{jt}(u)$ for $t = 1967, \ldots, 1980$. From left to right, figures correspond to the estimates at quantiles $u = 0.1, 0.3, 0.5, 0.7, 0.9$. Point estimates are plotted with solid black lines, and the pointwise 95% confidence interval are shown as grey shaded area.

conditional quantile and below. Panels (c) plots results for female, black workers, and the estimated changes in the between-inequality range from -0.05 to -0.35 at the 0.7th conditional quantile and below.

To illustrate the robustness of the significant policy-effects in reducing the between-equality, Figure 2.7 plots the changes in between-inequality for black, female workers from 1967 to 1980. In quantiles up to medium, the magnitude of reduction in the between-inequality in-

Figure 2.5 Changes in Within-Inequality in 1980

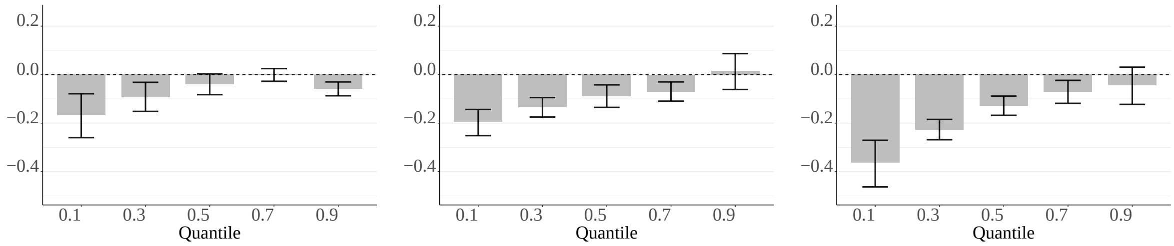
(a) $u_1 = 0.1,\ u_2 = 0.9$ (b) $u_1 = 0.1,\ u_2 = 0.5$



Notes: Panels (a)-(b) report estimated changes in the within-inequality $\dot{\Delta}_t^W(u_1, u_2|z)$ among individuals with attributes z in year $t = 1980$, with $(u_1, u_2) = (0.1, 0.9)$ in Panel (a) and $(u_1, u_2) = (0.1, 0.5)$ in Panel (b). The estimates are presented by grey bars with the 95% confidence intervals in black. The horizontal axis shows the three subpopulations based on race and gender categories, while we fix 12 years of education and 20 years of experience.

Figure 2.6 Changes in Between-Inequality in 1980

(a) Black, Male Workers (b) White, Female Workers (c) Black, Female Workers

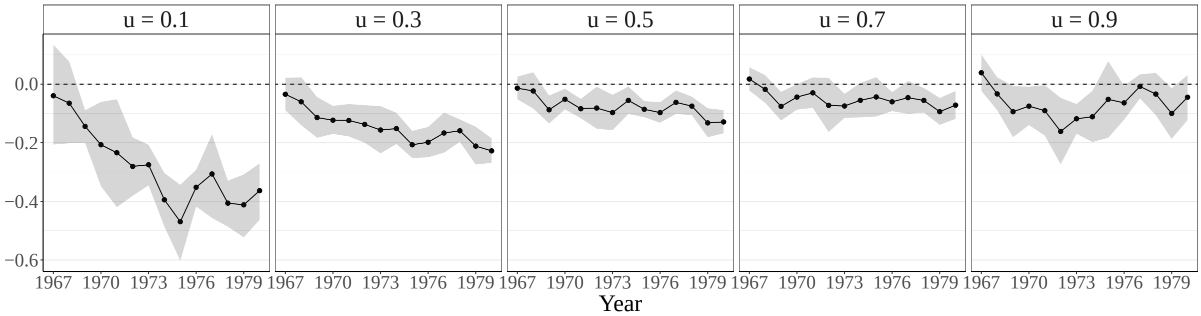


Notes: Panels (a)-(c) plot the changes in between-inequality ($\dot{\Delta}_t^B$) between multiple groups and the base-level group: white male workers (with 12 years of education and 20 years of experience) while holding other individual covariates constant. The inequality measure are considered at quantiles $u = 0.1, 0.3, 0.5, 0.7, 0.9$ at year 1980. Point estimates are presented by grey bars with point-wise 95% confidence intervals shown in black.

creases as time increase. In addition, such policy effects are significant after 1970. Similar patterns are also witnessed in the other two groups presented in Figure 2.6. We choose not to report those plots for the concise of presentation.

Overall, the results above confirm the findings of Derenoncourt and Montialoux (2021) that the reform was effective in improving the black economic status and reducing the racial income gap. In addition, we provide an empirical evidence of a compounded impact of the policy effect in reducing the racial and gender income gap, which leads to the significant reduction in the between-inequality, at least up to the medium.

Figure 2.7 Changes in Between-Inequality for Black, Female Workers from 1967 to 1980



Notes: From left to right, figures correspond to the estimates at quantiles $u = 0.1, 0.3, 0.5, 0.7, 0.9$. Point estimates are plotted with solid black lines, and the pointwise 95% confidence intervals are shown as grey shaded area.

2.6 Simulation Analysis

In this section, we investigate the accuracy of both the point estimators of the policy parameter $\delta_{jt}(u)$ and the treatment parameters $\Delta_t^{AQTT}(u|z)$, $\dot{\Delta}_t^{B}(u|z_1, z_2)$ and $\dot{\Delta}_t^{W}(u_1, u_2|z)$ through Monte Carlo simulations. We consider two data-generating processes (DGP). DGP I is a simplified scenario of our model, similar to the settings of Chetverikov et al. (2016). Using this setting, we illustrate the fast convergence of our proposed estimation method by documenting the estimation bias after given iteration steps. In addition, we consider a location-scale model as the DGP II, in which case the expressions for the treatment parameters can be derived. We investigate the estimation accuracy of the treatment parameters in DGP II. For both DGPs, we separately consider two scenarios, where the unobserved factors are either correlated or uncorrelated with the group-level regressors.

2.6.1 Data Generating Process I

In this section, we investigate the convergence rate of our proposed two-step iterative estimation approach. We generate data according to the following simplified model, similar to Chetverikov et al. (2016). That is, for $i = 1, ..., N,\ s = 1, ..., S$, and $t = 1, ..., T$,

$$y_{ist} = \alpha_{0st}(u_{ist}) + z_{ist}\alpha_{1st}(u_{ist}) + \Phi^{-1}(u_{ist}),$$
$$\alpha_{0st}(u) = \delta_0(u) + d_{st}\delta_t(u) + x_{st}\beta(u) + f_t(u)'\lambda_s(u) + \eta_{st}(u), \quad \alpha_{1st}(u) = 2 + 0.1u,$$
$$\delta_0(u) = 2 + \frac{u^2}{4}, \quad \delta_t(u) = 2 + \frac{t}{2T} + \frac{u^2}{4}, \quad \eta_{st}(u) = (\xi_{st} - 0.5)u,$$

where $\Phi(\cdot)$ is the cdf of $N(0,1)$,$\{u_{ist}\}$ and $\{\xi_{st}\}$ are i.i.d. $\mathrm{U}(0,1)$, $\{z_{ist}\}$ are i.i.d. $\mathrm{U}(0,1)$, x_{st} are i.i.d. $N(0,1)$, (f_{t1}, f_{t2}) are generated orthogonally via the SVD of a $T \times T$ random matrix whose entries are i.i.d. $N(0,1)$, and $(\lambda_{s1}, \lambda_{s2})$ are generated from i.i.d. $\mathrm{U}(0,2)$. The policy dummy variable $d_{st} = 1\{t \geq T/4\} \times 1\{s \geq S/4\}$, that is, we fixed the first one-quarter of the cross-sections as the control groups, and the policy is implemented at $T_0 = T/4$. We consider the following two scenarios in terms of whether endogeneity exists in the group-level:

1. group-level observables x_{st} are i.i.d. $N(0,1)$.

2. group-level observables $x_{st} = \zeta_{st} + 0.02f_{t1}^2 + 0.02\lambda_{s1}^2$, where $\{\zeta_{st}\}$ are i.i.d $N(0,1)$. This setting allows moderate endogeneity at group-level.

We investigate the accuracy of the point estimators of the policy parameters $\delta_T(u)$ for both scenarios. The sample bias and standard deviation at finite iteration steps $m = 2, 5$ and after the convergence criterion satisfied, are reported in Table 2.1. The simulation shows a few nice properties, which can be summarised as follows. First of all, the recursive estimator converges quite quickly. In all cases, the mean bias and standard deviation of the coefficient estimation at the second iteration are reasonably small. In addition, the estimators remain valid even if the number of observations per group (N_{st}) is relatively small in comparison to the total number of groups and time ($S \times T$), which is a particular attractive property in practice.

2.6.2 Data Generating Process II

In this section, we focus on the estimation accuracy of the policy parameters and treatment parameters after convergence. We consider a location-scale model, whose conditional quantile

Table 2.1 Sample Bias and Standard Deviation of Estimates of the Policy Parameters

				Scenario 1			Scenario 2		
N	S	T	u	$\widehat{\delta}_T^{(2)}(u)$	$\widehat{\delta}_T^{(5)}(u)$	$\widehat{\delta}_T(u)$	$\widehat{\delta}_T^{(2)}(u)$	$\widehat{\delta}_T^{(5)}(u)$	$\widehat{\delta}_T(u)$
500	40	25	0.1	0.003 (0.074)	0.001 (0.017)	0.001 (0.016)	0.005 (0.084)	0.002 (0.017)	0.001 (0.016)
			0.5	0.001 (0.031)	0.001 (0.015)	0.001 (0.014)	0.003 (0.035)	0.001 (0.015)	0.001 (0.014)
			0.9	0.005 (0.041)	0.003 (0.014)	0.001 (0.014)	−0.003 (0.085)	−0.003 (0.028)	−0.002 (0.028)
	40	50	0.1	0.003 (0.041)	0.001 (0.025)	0.001 (0.014)	0.003 (0.038)	0.001 (0.008)	0.001 (0.008)
			0.5	0.003 (0.031)	0.001 (0.014)	0.001 (0.014)	0.001 (0.012)	0.001 (0.007)	0.001 (0.007)
			0.9	0.002 (0.035)	0.003 (0.034)	−0.001 (0.014)	0.003 (0.041)	−0.001 (0.012)	−0.001 (0.011)
1000	40	50	0.1	0.006 (0.041)	0.003 (0.035)	0.002 (0.014)	−0.005 (0.076)	0.001 (0.013)	0.001 (0.012)
			0.5	0.004 (0.051)	−0.003 (0.035)	0.001 (0.014)	−0.001 (0.029)	0.001 (0.012)	0.001 (0.012)
			0.9	0.003 (0.031)	0.001 (0.015)	−0.001 (0.014)	−0.004 (0.087)	−0.001 (0.026)	−0.001 (0.025)
	60	50	0.1	0.006 (0.081)	0.002 (0.015)	0.001 (0.014)	0.004 (0.037)	0.001 (0.005)	0.001 (0.005)
			0.5	0.005 (0.031)	0.003 (0.015)	0.001 (0.014)	0.001 (0.010)	0.001 (0.005)	0.001 (0.005)
			0.9	0.002 (0.041)	−0.002 (0.015)	0.001 (0.014)	−0.001 (0.041)	0.001 (0.010)	0.001 (0.010)
2000	60	50	0.1	0.005 (0.070)	0.001 (0.010)	0.001 (0.009)	0.005 (0.070)	0.001 (0.011)	0.001 (0.009)
			0.5	0.004 (0.070)	0.003 (0.014)	0.002 (0.008)	−0.003 (0.029)	0.001 (0.010)	0.001 (0.010)
			0.9	0.002 (0.070)	−0.001 (0.011)	0.001 (0.009)	0.004 (0.084)	0.001 (0.026)	0.001 (0.025)
	80	50	0.1	0.002 (0.070)	0.001 (0.011)	0.001 (0.009)	0.001 (0.037)	0.001 (0.004)	0.001 (0.004)
			0.5	0.002 (0.064)	0.001 (0.010)	0.001 (0.007)	0.001 (0.011)	0.001 (0.005)	0.001 (0.005)
			0.9	0.001 (0.068)	0.001 (0.011)	0.001 (0.010)	0.002 (0.039)	−0.001 (0.010)	−0.001 (0.010)

Notes: The number of Monte-Carlo replications is 250. We report bias averaged over 250 replications with the standard deviation in the parentheses at finite iteration steps $m = 2, 5$ and after convergence.

is easy to derive. The setting of the factor structure is similar to that of Chen et al. (2021). Specifically, we generate

$$y_{ist} = \sum_{j=1}^{3} z_{ist(j)}\alpha_{1(j),st} + \sum_{j=1}^{3} z_{ist(j)}\alpha_{2(j),st}\epsilon_{ist},$$

$$\alpha_{1(j),st} = 3 + d_{st}\delta_{1(j),t} + x_{st}'\beta_{1(j)} + f_{1(j),t}'\lambda_{2(j),s} + \eta_{1(j),st}$$

$$\alpha_{2(j),st} = 0.5 + d_{st}\delta_{2(j),t} + x_{st}'\beta_{2(j)} + f_{2(j),t}'\lambda_{2(j),s} + \eta_{2(j),st}, \quad j = 1, 2, 3,$$

where

$$z_{ist(1)} \equiv 1, \quad z_{ist(2)} \overset{iid}{\sim} Bern(0.6), \quad z_{ist(3)} \overset{iid}{\sim} N(0,1) \text{ truncated between } [0,3],$$

$$\epsilon_{ist} \overset{iid}{\sim} N(0,1), \quad \eta_{1(j),st} \overset{iid}{\sim} N(0,2), \quad \eta_{2(j),st} \overset{iid}{\sim} N(0,0.5) \text{ truncated between } [-0.1, 0.1],$$

$$d_{st} = 1\{s \geq S/3\} \times 1\{t \geq T/3\},$$

$$\beta_{1(j)} = 5 + j/3, \quad \beta_{2(j)} = 0.1, \quad \delta_{1(j),t} = 5 + t/T, \quad \delta_{2(j),t} = 0.1,$$

$$f_{1(1),t} = 0.5 f_{1(1),t-1} + \xi_{1,t}, \quad f_{1(2),t} = 0.75 f_{1(2),t-1} + \xi_{2,t}, \quad \xi_{1,t}, \xi_{2,t} \overset{iid}{\sim} N(0,1),$$

$$f_{1(3),t} \overset{iid}{\sim} N(0,1), \quad f_{2(j),t} \overset{iid}{\sim} |N(0,0.5)|, \quad \lambda_{1(j),s} \overset{iid}{\sim} N(0,1), \quad \lambda_{2(j),s} \overset{iid}{\sim} U(0,0.5).$$

Again, we consider the following two scenarios in terms of whether exists endogeneity in the group-level:

1. Group-level observables x_{st} are i.i.d. exp$(0.1N(0,1))$.

2. Group-level observables $x_{st} = \zeta_{st} + 0.1f_{t2}^2 + 0.2\lambda_{s2}^2$, where $\{\zeta_{st}\}$ are i.i.d. $\exp\big(0.25N(0,1)\big)$. This setting allows moderate endogeneity at group-level.

For both scenarios, it is easy to check that $z_{ist}\alpha_{2(j),st} > 0$ for all $j = 1, 2, 3$. Thus, it is straightforward to obtain the conditional quantile function of y_{ist} as

$$
\begin{aligned}
Q_{y_{ist}}(u|d_{st}, x_{it}, z_t, f_t(u), \lambda_i(u)) &= \sum_{j=1}^{3} z_{ist(j)}\alpha_{jst}(u), \\
\alpha_{jst}(u) &= \alpha_{1(j),st} + \alpha_{2(j),st}Q_\epsilon(u) \\
&= \delta_{j0}(u) + d_{st}\delta_{jt}(u) + x_{st}'\beta_j(u) + f_{jt}(u)'\lambda_{js}(u) + \eta_{jst}(u),^{7}
\end{aligned}
$$

where

$$
\begin{aligned}
&\delta_{j0}(u) = 3 + 0.5Q_\epsilon(u), \quad \delta_{jt}(u) = \delta_{1(j),t} + \delta_{2(j),t}Q_\epsilon(u), \quad \beta_j(u) = \beta_{1(j)} + \beta_{2(j)}Q_\epsilon(u), \\
&f_{jt}(u) = [f_{1(j),t}, f_{2(j),t}]', \quad \lambda_{js}(u) = [\lambda_{1(j),s}, \lambda_{2(j),s}Q_\epsilon(u)]', \\
&\eta_{jst}(u) = \eta_{1(j),st} + \eta_{2(j),st}Q_\epsilon(u).
\end{aligned}
$$

We investigate the accuracy of the point estimators of the policy parameter $\delta_{jt}(u)$ for both scenarios, whose The sample bias and standard deviation at time $t = T$ are reported in Table 2.2. In addition, as an bias-corrected estimator of $\delta_{jt}(u)$ is proposed as (2.7), we report its accuracy in Table 2.3.

Table 2.2 reports the estimation accuracy of $\delta_{jt}(u)$. The simulation shows a few nice properties similar to those found in Section 2.6.1. First of all, the recursive estimator converges quite quickly at both tail and central quantiles. In all cases, the mean bias and standard deviation of the coefficient estimation at the second iteration are reasonably small. An additional appeal-

[7] The DGP is constructed such that the assumption $\mathbb{E}[\eta_{jst}(u)|x_{st}, d_{st}, f_{jt}, \lambda_{js}] = 0$ is satisfied.

ing feature is that the estimators remain accurate even if the number of observations per group ($N_{st} \equiv N$) is relatively small compared to the total number of groups and time ($S \times T$). Last but not least, we conclude that the moderate group-level endogenity, captured by the common factor structure, does not diminish the precision of the estimators. This feature is aligned with our asymptotic analysis. Table 2.3 reports the accuracy of the bias-corrected estimates, whose bias reduces in most of the cases compared with the original estimates in Table 2.2. However, the reduction is relatively minor. As the bias of the uncorrected estimates is already reasonably small, the bias-correction is not necessary in this case. Also, we note that, apart from the analytical bias-correction considered in this thesis, other biased-correction method, such as wild bootstrap (Gao et al., 2023), may lead to improved performance in finite samples.

Table 2.2 Point Estimation Accuracy of $\widehat{\delta}_{jT}(u)$

				Scenario 1			Scenario 2		
N	S	T	u	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
1000	20	20	0.1	0.046 (0.769)	−0.048 (0.882)	0.038 (0.787)	0.042 (0.766)	−0.033 (0.872)	0.041 (0.789)
			0.5	0.032 (0.900)	−0.038 (0.869)	0.063 (0.780)	0.029 (0.897)	−0.025 (0.860)	0.066 (0.782)
			0.9	0.028 (1.149)	−0.014 (0.898)	0.065 (0.821)	0.022 (1.149)	−0.001 (0.890)	0.069 (0.824)
	40	40	0.1	−0.051 (0.402)	0.096 (0.534)	−0.017 (0.436)	−0.052 (0.402)	0.096 (0.534)	−0.018 (0.436)
			0.5	−0.051 (0.394)	0.081 (0.472)	−0.029 (0.425)	−0.051 (0.394)	0.081 (0.472)	−0.030 (0.424)
			0.9	−0.052 (0.412)	0.079 (0.488)	−0.043 (0.437)	−0.052 (0.412)	0.080 (0.490)	−0.043 (0.436)
	60	60	0.1	−0.005 (0.331)	0.013 (0.381)	−0.014 (0.329)	−0.006 (0.331)	0.013 (0.381)	−0.013 (0.328)
			0.5	−0.007 (0.328)	0.012 (0.364)	−0.020 (0.306)	−0.008 (0.328)	0.012 (0.364)	−0.019 (0.306)
			0.9	−0.010 (0.356)	0.014 (0.379)	−0.027 (0.316)	−0.010 (0.355)	0.015 (0.379)	−0.027 (0.316)
2000	20	20	0.1	−0.072 (0.846)	−0.032 (0.797)	0.144 (1.340)	−0.080 (0.824)	−0.041 (0.786)	0.027 (0.818)
			0.5	−0.059 (0.831)	−0.003 (0.770)	0.055 (0.677)	−0.070 (0.797)	−0.012 (0.760)	0.018 (0.818)
			0.9	−0.044 (1.015)	0.022 (0.776)	0.034 (0.713)	−0.046 (0.911)	0.013 (0.765)	0.030 (0.707)
	40	40	0.1	−0.072 (0.424)	−0.037 (0.438)	0.033 (0.434)	−0.072 (0.424)	−0.037 (0.439)	0.033 (0.435)
			0.5	−0.066 (0.415)	−0.034 (0.425)	0.022 (0.422)	−0.066 (0.415)	−0.035 (0.426)	0.022 (0.423)
			0.9	−0.062 (0.432)	−0.032 (0.443)	0.011 (0.434)	−0.062 (0.431)	−0.032 (0.444)	0.011 (0.435)
	60	60	0.1	−0.032 (0.346)	0.007 (0.345)	−0.022 (0.348)	−0.031 (0.345)	0.006 (0.345)	−0.022 (0.349)
			0.5	−0.024 (0.322)	0.014 (0.327)	−0.032 (0.332)	−0.024 (0.323)	0.013 (0.328)	−0.032 (0.333)
			0.9	−0.019 (0.336)	0.019 (0.344)	−0.041 (0.347)	−0.019 (0.335)	0.019 (0.345)	−0.041 (0.348)

Notes: The number of Monte-Carlo replications is 250. We report the bias averaged over 250 replications with the standard deviation in the parentheses.

Furthermore, to illustrate the estimation accuracy of our proposed treatment parameters, Tables 2.4-2.6 report the estimation accuracy of $\Delta_t^{AQTT}(u|z)$, $\dot{\Delta}_t^B(u|z_1, z_2)$ and $\dot{\Delta}_t^W(u_1, u_2|z)$, respectively. Throughout the comparison, we fix $z := (1, 1, 1)'$, $z_1 := (1, 1, 0)'$ and $z_2 := (1, 0, 1)'$ to allow sufficient variation in terms of the individual characteristics. It is not surprising that the estimation results are relatively similar to those for $\delta_{jt}(u)$, as the treatment parameters are identified as linear combinations of $\delta_{jt}(u)$ according to Theorem 2.2.1.

Table 2.3 Point Estimation Accuracy of the Baised-Corrected Estimator $\widehat{\delta}^{\dagger}_{jT}(u)$

				Scenario 1			Scenario 2		
N	S	T	u	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
1000	20	20	0.1	0.044 (0.959)	−0.047 (0.893)	0.039 (0.787)	0.041 (0.785)	−0.031 (0.785)	0.040 (0.759)
			0.5	0.029 (0.980)	−0.035 (0.894)	0.065 (0.780)	0.028 (0.935)	−0.025 (0.856)	0.056 (0.794)
			0.9	0.029 (1.254)	−0.015 (0.890)	0.075 (0.780)	0.024 (1.157)	−0.002 (0.781)	0.071 (0.825)
	40	40	0.1	−0.051 (0.422)	0.094 (0.518)	−0.018 (0.459)	−0.051 (0.452)	0.098 (0.545)	−0.017 (0.483)
			0.5	−0.051 (0.495)	0.080 (0.476)	−0.027 (0.445)	−0.049 (0.345)	0.080 (0.475)	−0.031 (0.447)
			0.9	−0.050 (0.478)	0.079 (0.487)	−0.045 (0.437)	−0.058 (0.410)	0.078 (0.480)	−0.042 (0.407)
	60	60	0.1	−0.005 (0.451)	0.015 (0.399)	−0.013 (0.377)	−0.006 (0.354)	0.012 (0.315)	−0.011 (0.350)
			0.5	−0.007 (0.397)	0.015 (0.365)	−0.019 (0.350)	−0.007 (0.304)	0.015 (0.379)	−0.018 (0.310)
			0.9	−0.010 (0.356)	0.017 (0.389)	−0.028 (0.358)	−0.008 (0.365)	0.014 (0.399)	−0.025 (0.384)
2000	20	20	0.1	−0.073 (0.867)	−0.031 (0.799)	0.143 (1.347)	−0.078 (0.846)	−0.040 (0.787)	0.027 (0.868)
			0.5	−0.057 (0.835)	−0.003 (0.795)	0.055 (0.677)	−0.070 (0.797)	−0.012 (0.770)	0.017 (0.818)
			0.9	−0.043 (1.185)	0.024 (0.786)	0.035 (0.738)	−0.036 (0.801)	0.018 (0.735)	0.029 (0.727)
	40	40	0.1	−0.070 (0.487)	−0.035 (0.488)	0.032 (0.445)	−0.072 (0.428)	−0.034 (0.475)	0.035 (0.471)
			0.5	−0.074 (0.464)	−0.035 (0.415)	0.020 (0.454)	−0.064 (0.405)	−0.036 (0.475)	0.024 (0.401)
			0.9	−0.062 (0.420)	−0.035 (0.475)	0.017 (0.445)	−0.060 (0.441)	−0.031 (0.425)	0.011 (0.478)
	60	60	0.1	−0.031 (0.397)	0.007 (0.378)	−0.021 (0.387)	−0.030 (0.455)	0.007 (0.357)	−0.021 (0.385)
			0.5	−0.021 (0.378)	0.011 (0.348)	−0.034 (0.378)	−0.034 (0.353)	0.023 (0.378)	−0.031 (0.343)
			0.9	−0.018 (0.345)	0.018 (0.354)	−0.041 (0.378)	−0.018 (0.357)	0.017 (0.387)	−0.040 (0.479)

Notes: The number of Monte-Carlo replications is 250. We report the bias averaged over 250 replications with the standard deviation in the parentheses.

Table 2.4 Point Estimation Accuracy of $\widehat{\Delta}^{AQTT}_{T}(u|z)$

N	S	T	u	Scenario 1	Scenario 2
1000	20	20	0.1	0.036 (1.409)	0.050 (1.399)
			0.5	0.056 (1.497)	0.070 (1.488)
			0.9	0.079 (1.719)	0.090 (1.718)
	40	40	0.1	0.027 (0.817)	0.027 (0.816)
			0.5	0.001 (0.760)	0.001 (0.760)
			0.9	−0.016 (0.783)	−0.016 (0.783)
	60	60	0.1	−0.006 (0.606)	−0.005 (0.606)
			0.5	−0.015 (0.584)	−0.015 (0.585)
			0.9	−0.023 (0.617)	−0.022 (0.617)
2000	20	20	0.1	0.040 (1.731)	−0.094 (1.271)
			0.5	−0.006 (1.218)	−0.065 (1.223)
			0.9	0.013 (1.401)	−0.002 (1.286)
	40	40	0.1	−0.077 (0.744)	−0.077 (0.745)
			0.5	−0.079 (0.723)	−0.078 (0.724)
			0.9	−0.084 (0.748)	−0.084 (0.750)
	60	60	0.1	−0.047 (0.608)	−0.047 (0.609)
			0.5	−0.042 (0.559)	−0.042 (0.560)
			0.9	−0.041 (0.567)	−0.041 (0.568)

Notes: The number of Monte-Carlo replications is 250. We report bias averaged over 250 replications with the standard deviation in the parentheses. We consider individual covariate $z := (1, 1, 1)'$.

Table 2.5 Point Estimation Accuracy of $\widehat{\Delta}_T^B(u|z_1, z_2)$

N	S	T	u	Scenario 1	Scenario 2
1000	20	20	0.1	0.086 (1.122)	0.073 (1.115)
			0.5	0.101 (1.110)	0.091 (1.105)
			0.9	0.079 (1.163)	0.070 (1.160)
	40	40	0.1	−0.113 (0.679)	−0.114 (0.679)
			0.5	−0.110 (0.623)	−0.111 (0.623)
			0.9	−0.122 (0.640)	−0.123 (0.641)
	60	60	0.1	−0.027 (0.501)	−0.026 (0.501)
			0.5	−0.031 (0.483)	−0.031 (0.483)
			0.9	−0.041 (0.506)	−0.041 (0.505)
2000	20	20	0.1	0.176 (1.556)	0.068 (1.197)
			0.5	0.058 (1.107)	0.030 (1.200)
			0.9	0.012 (1.137)	0.017 (1.130)
	40	40	0.1	0.070 (0.625)	0.070 (0.627)
			0.5	0.056 (0.607)	0.057 (0.608)
			0.9	0.043 (0.626)	0.043 (0.626)
	60	60	0.1	−0.029 (0.490)	−0.028 (0.490)
			0.5	−0.045 (0.474)	−0.045 (0.474)
			0.9	−0.060 (0.504)	−0.060 (0.504)

Notes: The number of Monte-Carlo replications is 250. We report bias averaged over 250 replications with the standard deviation in the parentheses. We consider individual covariates $z_1 := (1, 1, 0)'$ and $z_2 := (1, 0, 1)'$.

Table 2.6 Point Estimation Accuracy of $\widehat{\Delta}_T^W(u_1, u_2|z)$

N	S	T	(u_1, u_2)	Scenario 1	Scenario 2
1000	20	20	$(0.1, 0.5)$	0.021 (0.444)	0.020 (0.444)
			$(0.1, 0.9)$	0.043 (0.894)	0.040 (0.898)
			$(0.5, 0.9)$	0.022 (0.497)	0.020 (0.499)
	40	40	$(0.1, 0.5)$	-0.027 (0.209)	-0.026 (0.207)
			$(0.1, 0.9)$	-0.043 (0.372)	-0.042 (0.371)
			$(0.5, 0.9)$	-0.016 (0.184)	-0.016 (0.184)
	60	60	$(0.1, 0.5)$	-0.009 (0.190)	-0.009 (0.189)
			$(0.1, 0.9)$	-0.017 (0.376)	-0.017 (0.374)
			$(0.5, 0.9)$	-0.008 (0.188)	-0.008 (0.187)
2000	20	20	$(0.1, 0.5)$	-0.046 (1.187)	0.029 (0.320)
			$(0.1, 0.9)$	-0.027 (1.437)	0.091 (0.909)
			$(0.5, 0.9)$	0.019 (0.596)	0.062 (0.737)
	40	40	$(0.1, 0.5)$	-0.002 (0.182)	-0.002 (0.181)
			$(0.1, 0.9)$	-0.007 (0.365)	-0.007 (0.363)
			$(0.5, 0.9)$	-0.005 (0.186)	-0.005 (0.185)
	60	60	$(0.1, 0.5)$	0.005 (0.177)	0.004 (0.176)
			$(0.1, 0.9)$	0.006 (0.352)	0.005 (0.350)
			$(0.5, 0.9)$	0.001 (0.177)	0.001 (0.176)

Notes: The number of Monte-Carlo replications is 250. We report bias averaged over 250 replications with the standard deviation in the parentheses. We consider individual covariate $z := (1, 1, 1)'$.

2.7 Conclusion

In this chapter, we introduce an estimation method for evaluating the effect of group-level policies under the quantile regression framework with interactive fixed effects. Our method can capture the heterogeneous policy effects through the interaction of policy variables and the individual observed and unobserved characteristics, while controlling the unobserved interactive fixed effects, and provides a straightforward way of identifying the policy effect on inequality measures. The consistency and limiting distribution of the proposed estimators are established. Using our proposed model, we evaluate the effect of the minimum wage policy on earnings between 1967 and 1980 in the United States. Our analysis confirms the findings of Derenoncourt and Montialoux (2021) that the policy helps reduce the racial income gap by improving the black economic status. On top of that, we provide empirical evidence of a compounded policy effect in narrowing the racial and gender gap, which contributes to the significant reduction in the between-inequality.

2.8 Additional Empirical and Theoretical Results

In this appendix, Section 2.8.1 provides the additional details about the empirical dataset, and Section 2.8.2 establish the proofs of theorems in the main text. Specifically, we first establish the identification theorem of treatment parameters (Theorems 2.2.1), followed by the proofs of the asymptotic results in Section 2.4.2. All the preliminary lemmas required are established and proved in Technical Supplement.

2.8.1 Additional Details about the March CPS Dataset

In this section, we provide additional information about the March CPS dataset.

Table 2.7 summarizes the 15 industries was covered by 1938 and 1966 FLSA, respectively. As the key interest of this empirical study is to quantity the minimum wage policy effect of 1966 FLSA, the eight industries covered by 1966 FLSA are considered as treated industries in this case study, while the eight industries covered by 1938 FLSA are classified as control industries. For each listed industry, the average number of observations per year is listed. We note that Forestry and Fishing only contains 35 individual observations per year. As the sample size is not sufficient to perform the first-step quantile estimation, we remove Forestry and Fishing from the treated industries to avoid estimation error, which leads to seven treated industries in our study.

Figure 2.8 presents the aggregate evidence of stable employment trends in the CPS. Figures 2.8.(a) and (b) show that the proportion of black employees and female employees across industry type (industries covered in 1938 v.s. covered in 1967) are relatively stable from the early 1960s to 1980. Notably, while black proportions are below 25% for both treated and control industries, the female proportions are doubled in the treated industries comparing with those in the control industries throughout the years. Figures 2.8.(c) and (d) further decompose these aggregate employment trends by gender and race. which also show a stable trend.

Figure 2.9 summaries the distribution of education and experience level of the four major groups of individuals of interest. The average and medium education level is around 12 years of schools for all groups, with heavy tails at the lower quantile, representing a large variation in education level among the society. There are vast literature analysing the effect of improving education on reducing the wage disparities, such as (Card and Krueger, 1992). However, the education factor cannot explain the dramatic decline of wage gap after 1967, as education levels remain stable over the years. In addition, the distributions of experience level are rather similar for all groups, with the first quarter and medium level of experience being 10 and 20 years, respectively.

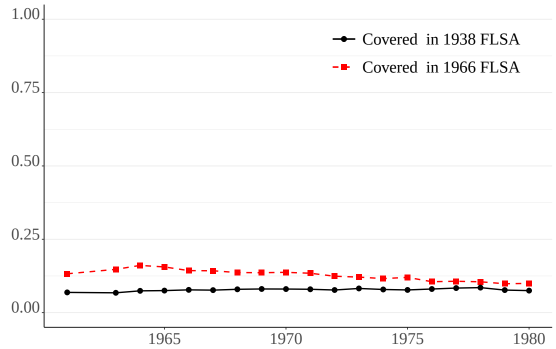
Table 2.7 List of industries used in March CPS (1962–1980)

Industry	Average Observations per Year
Control industries (industries covered by 1938 FLSA):	
Business and Repair Services	842
Durable manufacturing	5108
Finance, Insurance, and Real Estate	1520
Food manufacturing	727
Mining	305
Other non-durable manufacturing	2481
Transportation, Communication, and Other Utilities	2204
Wholesale Trade	1149
Treated industries (industries covered by 1966 FLSA):	
Agriculture	351
Entertainment and Recreation Services	229
Forestry and Fishing	35
Restaurants	649
Hospitals	1135
Hotels, laundries, and other personal services	531
Nursing homes and other professional services	1515
Schools and other educational services	2694

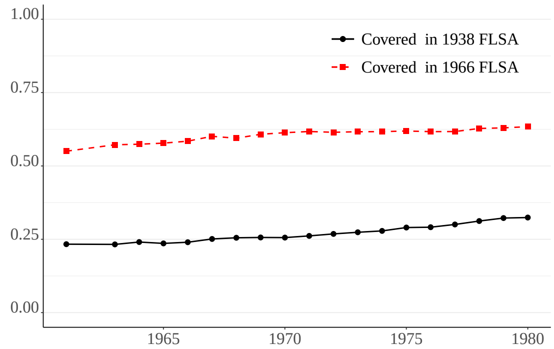
Notes: The data sources are the Current Population Survey 1962–1980. Sample includes black or white adults aged 25–65, who worked more than 13 weeks last year, worked three hours last week, do not live in group quarters, are not self-employed and not unpaid family worker with no missing industry or occupation code.

Figure 2.8 Aggregate Employment Shares

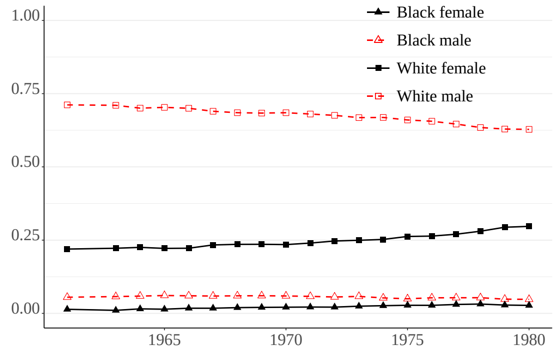
(a) Black proportions, by industries



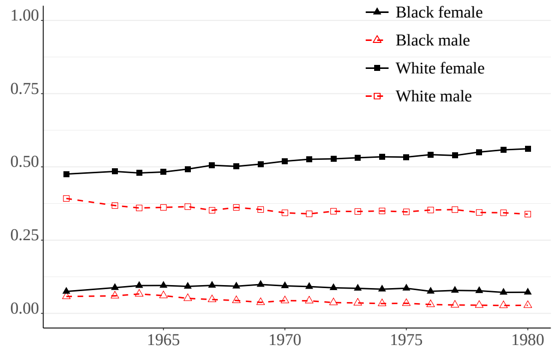
(b) Female proportions, by industries



(c) 1938 industries (control), by race and gender



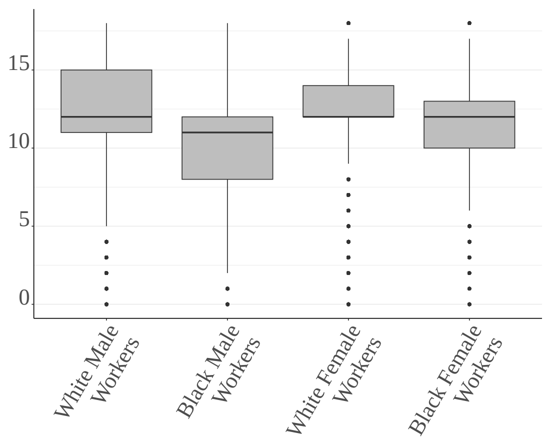
(d) 1966 industries (treated), by race and gender



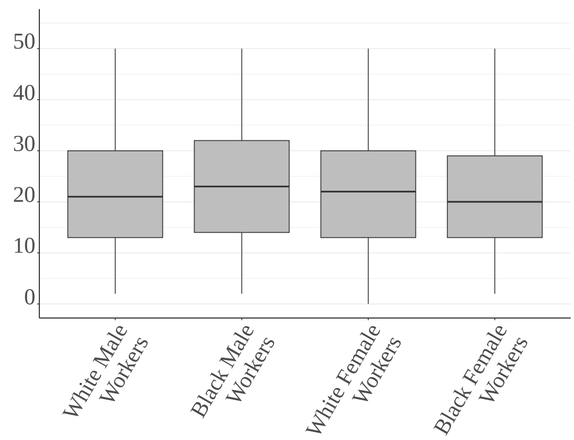
Notes: The data sources are the Current Population Survey 1962–1980. Sample includes black or white adults aged 25–65, who worked more than 13 weeks last year, worked three hours last week, do not live in group quarters, are not self-employed and not unpaid family worker with no missing industry or occupation code.

Figure 2.9 Education and Experience by Race and Gender

(a) Years of education



(b) Years of experience



Notes: The data sources are the Current Population Survey 1962–1980. Sample includes black or white adults aged 25–65, who worked more than 13 weeks last year, worked three hours last week, do not live in group quarters, are not self-employed and not unpaid family worker with no missing industry or occupation code.

2.8.2 Proofs of the Main Results

In what follows, we use a few additional notation. For a square matrix A, let $\text{tr}(A)$ denote the trace operation of A and let $\rho_{\max}(A)$ and $\rho_{\min}(A)$ denote the maximum and minimum eigenvalue of A, respectively. Notice that $\rho_{\max}(A) \leq \text{tr}(A)$ for every symmetric positive semi-definite matrix. Thus, $\|X\|^2 \leq \text{tr}(X'X)$ for matrix X, and we repeatedly use this inequality. We denote c and C as strictly positive constants that depend only on $c_M,\ C_M$, whose values can change at each appearance. Let $x_{st}, x_{s(k)}$ and $x_{st(k)}$ represent different vectors related to the regressor X. Precisely, x_{st} is a $K \times 1$ vector, $x_{s(k)}$ is a $T \times 1$ vector and $x_{st(k)}$ is a scalar for any $(s,t,k) \in \{1,...,S\} \times \{1,...,T\} \times \{1,...,K\}$. Let $\delta_{ST} \equiv \min(\sqrt{S}, \sqrt{T})$. Let e_k denote the a column vector, whose entries are all 0 except for the k-th entry; the dimension of e_k varies along with the context.

Proof of the Identification Theorem

Proof of Theorem 2.2.1. Let $u \in \mathcal{U}$, $z \in \mathcal{Z}$ and $t \geq T_0$ be fixed. Under the potential outcome framework (2.3)-(2.4), we can write $\Delta_t^{AQTT}(u|z) = z'\mathbb{E}[\alpha_{st}^1(u) - \alpha_{st}^0(u)|d_s = 1]$. Under Assumption 2.2.4.(i), it follows from (2.4) that $\mathbb{E}[\alpha_{jst}^1(u) - \alpha_{jst}^0(u)|d_s = 1] = \mathbb{E}[\Delta_{jst}(u)|d_s = 1]$, for $j = 1,...,J$. Thus, it is straightforward that $\Delta_t^{AQTT}(u|z) = z'\mathbb{E}[\Delta_{st}(u)|d_s = 1]$, $\dot{\Delta}_t^W(u_1,u_2|z) = z'\mathbb{E}[\Delta_{st}(u_2) - \Delta_{st}(u_1)|d_s = 1]$, and $\dot{\Delta}_t^B(u|z_1,z_2) = (z_2 - z_1)'\mathbb{E}[\Delta_{st}(u)|d_s = 1]$, where $\Delta_{st}(u) := [\Delta_{1st}(u),...,\Delta_{Jst}(u)]'$.

Moreover, under the potential outcome framework, we have $\alpha_{jst}(u) = (1-d_{st})\alpha_{jst}^0(u) + d_{st}\alpha_{jst}^1(u)$. Therefore, we can rewrite (2.4) as

$$\alpha_{jst}(u) = d_{st}\mathbb{E}[\Delta_{jst}(u)|d_s = 1] + x_{st}'\beta_j(u) + f_{jt}(u)'\lambda_{js}(u) + \eta_{jst}(u) \tag{2.10}$$

where $\eta_{jst}(u) := d_{st}(\Delta_{jst}(u) - \mathbb{E}[\Delta_{jst}(u)|d_s = 1]) + (1-d_{st})\eta_{jst}^0(u) + d_{st}\eta_{jst}^1(u)$. Then, it is clear that (2.10) coincides with (2.2) by defining $\delta_{jt}(u) := \mathbb{E}[\Delta_{jst}(u)|d_s = 1]$.

To prove the theorem, it remains to check that $\delta_{jt}(u)$ can be identified from model (2.1)-(2.2) for each $j = 1, \ldots, J$ and $t \geq T_0$.

It is know that, for model (2.1)-(2.2), under Assumption 2.2.1, we can estimate the quantile regression coefficients $\alpha_{jst}(u)$ from model (2.1). Also, we can identify the regression coefficients, factors and factor loadings, using the argument of Bai (2009) under Assumption 2.2.3. Thus, we treat the quantile regression coefficients, factors and loadings as known objects in the rest of the proof. Then, since $\mathbb{E}[\eta_{jst}(u)|d_{st}, X_s] = 0$ under Assumption 2.2.4, taking the expectation on both sides of (2.2) leads to the normal equations

$$\mathbb{E}[d_{st}(\alpha_{jst} - f_{jt}(u)'\lambda_{js}(u))] = \mathbb{E}[d_{st}^2]\delta_{jt}(u) + \mathbb{E}[d_{st}x'_{st}]\beta_j(u), \tag{2.11}$$

$$\mathbb{E}[x_{st}(\alpha_{jst} - f_{jt}(u)'\lambda_{js}(u))] = \mathbb{E}[d_{st}x_{st}]\delta_{jt}(u) + \mathbb{E}[x_{st}x'_{st}]\beta_j(u). \tag{2.12}$$

Solving (2.12) with respect to $\beta_j(u)$, we have $\beta_j(u) = \mathbb{E}[x_{st}x'_{st}]^{-1}\{\mathbb{E}[x_{st}(\alpha_{jst} - f_{jt}(u)'\lambda_{js}(u))] - \mathbb{E}[d_{st}x_{st}]\delta_{jt}(u)\}$ given that $\mathbb{E}[x_{st}x'_{st}]^{-1}$ is invertible under Assumption 2.2.3.(ii). Substituting the solution into (2.11), we obtain

$$\delta_{jt}(u) = \mathbb{E}[d_{st}\Pi_{st}]^{-1}\mathbb{E}[\Pi_{st}(\alpha_{jst}(u) - f_{jt}(u)'\lambda_{js}(u))],$$

where $\Pi_{st} := d_{st} - \mathbb{E}[d_{st}x'_{st}]\mathbb{E}[x_{st}x'_{st}]^{-1}x_{st}$. Moreover, since $\text{Var}(x_{st}) = \mathbb{E}[x_{st}x'_{st}] - \mathbb{E}[x_{st}]\mathbb{E}[x'_{st}] > 0$, it follows that $\mathbb{E}[d_{st}\Pi_{st}] = \mathbb{E}[\mathbb{E}[d_{st}\Pi_{st}]|d_{st}] = \mathbb{P}(d_{st} = 1)(1 - \mathbb{E}[x'_{st}]\mathbb{E}[x_{st}x'_{st}]^{-1}\mathbb{E}[x_{st}]) > 0$, that is, $\mathbb{E}[d_{st}\Pi_{st}]$ is invertible.

In summary, $\delta_{jt}(u)$ is identifiable in model (2.1)-(2.2). Consequently, the policy parameters $\Delta_t^{AQTT}(u|z)$, $\dot{\Delta}_t^B(u|z_1, z_2)$ and $\dot{\Delta}_t^W(u_1, u_2|z)$ are identifiable, following the arguments at the beginning of the proof ■

Proofs of the Asymptotic Properties

In what follows, we use the following facts: for all $(s,t) \in \{1, ..., S\} \times \{1, ..., T\}$, $\|X_s\| = O_p(\sqrt{T})$, $\|F_j(u)\| = O_p(\sqrt{T})$, $\|\widehat{F}_j^{(m)}(u)\| = O_p(\sqrt{T})$, $\|\Lambda_j(u)\| = O_p(\sqrt{S})$, $\|\widehat{\Lambda}_j^{(m)}(u)\| = O_p(\sqrt{S})$ under Assumption 2.4.1. Also, $\|M_{F_j}(u)\| = O_p(1)$ since the largest eigenvalue of $M_{F_j}(u)$ is 1 as $M_{F_j}(u)$ is a projection matrix, and similarly, $\|M_{\widehat{F}_j}^{(m)}(u)\| = O_p(1)$. In addition, we define the orthogonal projection matrix $P_{F_j}(u) := F_j(u)(F_j(u)'F_j(u))^{-1}F_j(u)$.

Before proceeding to the proof, since the convergence of $\widehat{\delta}_{jt}^{(m)}(u)$ and $\widehat{\beta}_j^{(m)}(u)$ are different, there is a need to partial-out the expression of $\widehat{\delta}_{jt}^{(m)}(u)$ and $\widehat{\beta}_j^{(m)}(u)$ from that of $\widehat{\beta}_j^{(m)}(u) := [\widehat{\delta}_{jT_0}^{(m)}(u), ..., \widehat{\delta}_{jT}^{(m)}(u), \widehat{\beta}_j^{(m)}(u)]'$ obtained in Section 2.3. Recall that the estimator $\widehat{\beta}_j^{(m)}(u)$ is the minimiser of the SSR of Equation (2.5). Therefore, by setting the first order derivative of SSR with respect to $\beta_j^{(m)}(u)$ to zero, we obtain the following relationships between $\widehat{\delta}_{jt}^{(m)}(u)$ and $\widehat{\beta}_j^{(m)}(u)$:

$$\begin{cases} \widehat{\delta}_{jt}^{(0)}(u) = \left(\sum_{s=1}^S d_s^2\right)^{-1} \sum_{s=1}^S d_s\left(\widehat{\alpha}_{jst}(u) - x_{st}'\widehat{\beta}_j^{(0)}(u)\right), \quad t = T_0, ..., T, \\ \widehat{\beta}_j^{(0)}(u) = \left(\sum_{s=1}^S X_s'X_s\right)^{-1} \sum_{s=1}^S X_s'\left(\widehat{A}_{js}(u) - D_s\widehat{\delta}_j^{(0)}(u)\right), \end{cases} \tag{2.13}$$

and, for $m \geq 1$, given $(\widehat{F}_j^{(m)}(u), \widehat{\Lambda}_j^{(m)}(u))$,

$$\begin{cases} \widehat{\delta}_{jt}^{(m)}(u) = \left(\sum_{s=1}^S d_s^2\right)^{-1} \sum_{s=1}^S d_s\left(\widehat{\alpha}_{jst}(u) - x_{st}'\widehat{\beta}_j^{(m)}(u) - \widehat{f}_{jt}^{(m)}(u)'\widehat{\lambda}_{js}^{(m)}(u)\right), \\ \widehat{\beta}_j^{(m)}(u) = \left(\sum_{s=1}^S X_s'X_s\right)^{-1} \sum_{s=1}^S X_s'\left(\widehat{A}_{js}(u) - D_s\widehat{\delta}_j^{(m)}(u) - \widehat{F}_j^{(m)}(u)\widehat{\lambda}_{js}^{(m)}(u)\right). \end{cases} \tag{2.14}$$

Using the above relationships, we present the proof of Theorem 2.4.1 as follows.

Proof of Theorem 2.4.1. Since u and j are fixed, we suppress u and j throughout the following proof. The proof is completed by induction. We first show that (i) $\sqrt{ST}\|\widehat{\beta}^{(0)} - \beta\| = O_P(1)$, $\sqrt{S}|\widehat{\delta}_t^{(0)} - \delta_t| = O_P(1)$, and $\sqrt{S/T}|\sum_{t=1}^{T_0} \widehat{\delta}_t^{(0)} - \delta_t| = O_P(1)$. Then, using the iterative formula and the rates of the previous estimators, we show that suppose the above rates hold at $(m-1)^{\text{th}}$ iteration, then the rates hold at m^{th} iteration for all $m \geq 1$.

(i) We start with $\widehat{\beta}^{(0)} - \beta$. By simple algebra on (2.13) and the linear model (2.2) for α_{st}, we obtain

$$\begin{cases} \widehat{\delta}_t^{(0)} - \delta_t = \left(\sum_{s=1}^S d_s^2\right)^{-1} \sum_{s=1}^S d_s\left((\widehat{\alpha}_{st} - \alpha_{st}) - w_{st}'(\widehat{\beta}^{(0)} - \beta) + f_t'\lambda_s + \eta_{st}\right), \\ \widehat{\beta}^{(0)} - \beta = \left(\sum_{s=1}^S X_s'X_s\right)^{-1} \sum_{s=1}^S X_s'\left((\widehat{A}_s - A_s) - D_s(\widehat{\delta}^{(0)} - \delta) + F\lambda_s + \eta_s\right). \end{cases} \tag{2.15}$$

By further substituting the expression of $\widehat{\delta}_t^{(0)} - \delta_t$ $(t = T_0, ..., T)$ into $\widehat{\beta}^{(0)} - \beta$, we obtain

$$\Big[\sum_{s=1}^{S} X_s' X_s - \Big(\sum_{s=1}^{S} d_s^2\Big)^{-1} \sum_{s,g=1}^{S} X_s' D_s D_g' X_g\Big] \big(\widehat{\beta}^{(0)} - \beta\big)$$
$$= \sum_{s=1}^{S} X_s' [(\widehat{A}_s - A_s) + F\lambda_s + \eta_s] - \Big(\sum_{s=1}^{S} d_s^2\Big)^{-1} \sum_{s,g=1}^{S} X_s' D_s D_g' [(\widehat{A}_g - A_g) + F\lambda_g + \eta_g].$$

An application of the triangle inequality together with Lemma S.1.1 yields $\big\| \sum_{s=1}^{S} X_s' \big(\widehat{A}_s - A_s\big) \big\| = O_P((ST)^{1/4})$. Also, $(ST)^{-1} \sum_{s=1}^{S} X_s' F \lambda_s = O_P\big(\sqrt{ST}\big)$ and (S.2) in Lemma S.1.2 shows that $(ST)^{-1/2} \sum_{s=1}^{S} X_s' \eta_s = O_P(1)$. In addition, recall that $D_s = d_s[e_{T_0}, ..., e_T]$ and $d_s \in \{0, 1\}$, we then have

$$\Big\| \sum_{s,g=1}^{S} X_s' D_s D_g' (\widehat{A}_g - A_g) \Big\| = \sup_{g,t} |\widehat{\alpha}_{gt} - \alpha_{gt}| \cdot \sum_{s,g=1}^{S} \sum_{t=T_0}^{T} \|x_{st} d_s d_g\| = O_P(S^{5/4} T^{1/4}),$$

following from Lemma S.1.1,

$$\Big\| \sum_{s,g=1}^{S} X_s' D_s D_g F \lambda_g \Big\| \leq \Big(\sum_{t=T_0}^{T} \Big\| \sum_{s=1}^{S} d_s w_{st} \Big\|^2\Big)^{1/2} \Big(\sum_{t=T_0}^{T} \Big\| \sum_{g=1}^{S} d_s f_t' \lambda_g \Big\|^2\Big)^{1/2} = O_P(ST),$$

due to the assumption that $\mathbb{E}\| \sum_{g=1}^{S} d_s f_t' \lambda_g \|^2 = O_P(S)$ for all $t \geq T_0$ uniformly, and,

$$\Big\| \sum_{s,g=1}^{S} X_s' D_s D_g \eta_g \Big\| = \Big\| \sum_{s,g=1}^{S} \sum_{t=T_0}^{T} d_s d_g x_{st} \eta_{gt} \Big\| = O_P(S\sqrt{T})$$

by α-mixing. Finally, under Assumption 2.4.4, we have $\|(\sum_{s=1}^{S} d_s^2)^{-1}\| = O_P(S^{-1})$, and $\|[\sum_{s=1}^{S} X_s' X_s - (\sum_{s=1}^{S} d_s^2)^{-1} \sum_{s,g=1}^{S} X_s' D_s D_g' X_g]^{-1}\| = O_P((ST)^{-1})$. Thus, we collect the rates for all terms and obtain $\sqrt{ST} \|\widehat{\beta}_j^{(0)}(u) - \beta_j(u)\| = O_P(1)$.

(ii) We then consider $\widehat{\delta}_t^{(0)} - \delta_t$ for a given $t \geq T_0$. We check all the terms of the expression of $\widehat{\delta}_t^{(0)} - \delta_t$ in (2.15) as follows: $| \sum_{s=1}^{S} d_s (\widehat{\alpha}_{st} - \alpha_{st})| \leq \sup_{s,t} |\widehat{\alpha}_{st} - \alpha_{st}| \cdot \sum_{s=1} |d_s| = O_P(S^{1/4} T^{-3/4})$ using Lemma S.1.1, $| \sum_{s=1}^{S} d_s f_t' \lambda_s| = O_P(\sqrt{S})$ by assumption, $| \sum_{s=1}^{S} d_s \eta_{st}| = O_P(\sqrt{S})$ by α-mixing, and $| \sum_{s=1}^{S} d_s x_{st}' (\widehat{\beta}^{(0)} - \beta)| \leq \|\widehat{\beta}_j^{(0)}(u) - \beta_j(u)\| \sum_{s=1}^{S} |x_{st}| = O_P(\sqrt{S/T})$. Finally, since $|(\sum_{s=1}^{S} d_s^2)^{-1}\| = O_P(S^{-1})$ under Assumption 2.4.4, collecting the rates of all terms yields $\sqrt{S}|\widehat{\delta}_{jt}^{(0)}(u)-\delta_{jt}(u)|=O_P(1)$.

In addition, we show that $|\sum_{t=T_0}^T \widehat{\delta}_{jt}^{(0)}(u)-\delta_{jt}(u)|=O_P(1)$ as follows. Since

$$\sum_{t=T_0}^T \widehat{\delta}_t^{(0)}-\delta_t=\big(\sum_{s=1}^S d_s^2\big)^{-1}\sum_{s=1}^S\sum_{t=T_0}^T d_s\big((\widehat{\alpha}_{st}-\alpha_{st})-w_{st}'(\widehat{\beta}^{(0)}-\beta)+f_t'\lambda_s+\eta_{st}\big),$$

applying similar arguments as those in (ii), it is easy to check that $|\sum_{s=1}^S\sum_{t=T_0}^T d_s(\widehat{\alpha}_{st}-\alpha_{st})|=O_P((ST)^{1/4})$, $|\sum_{s=1}^S\sum_{t=T_0}^T d_s x_{st}'(\widehat{\beta}^{(0)}-\beta)|=O_P(\sqrt{ST})$, $|\sum_{s=1}^S\sum_{t=T_0}^T d_s f_t'\lambda_s|=O_P(\sqrt{ST})$, and $|\sum_{s=1}^S\sum_{t=T_0}^T d_s\eta_{st}|=O_P(\sqrt{ST})$. Therefore, we have

$$\begin{aligned}&\Big|\sum_{t=T_0}^T \widehat{\delta}_t^{(0)}-\delta_t\Big|\\ &\qquad\le O_P(S^{-1})\cdot\Big|\sum_{s=1}^S\sum_{t=T_0}^T d_s\big((\widehat{\alpha}_{st}-\alpha_{st})-x_{st}'(\widehat{\beta}^{(0)}-\beta)+f_t'\lambda_s+\eta_{st}\big)\Big|=O_P(1).\end{aligned} \tag{2.16}$$

We also note that a similar argument yields that $\|\sum_{t=T_0}^T \widehat{f}_t^{(1)}(\widehat{\delta}_t^{(0)}-\delta_t)\|=O_P(1)$.

Now, for a given $m\ge 1$, suppose that $\sqrt{ST}\|\widehat{\beta}^{(m-1)}-\beta\|=O_P(1)$, $\sqrt{S}|\widehat{\delta}_t^{(m-1)}-\delta_t|=O_P(1)$, $|\sum_{t=T_0}^T \widehat{\delta}_t^{(m-1)}-\delta_t|=O_P(1)$ and $\|\sum_{t=T_0}^T \widehat{f}_t^{(m)}(\widehat{\delta}_t^{(m-1)}-\delta_t)\|=O_P(1)$.

(iii) We want to show $\sqrt{ST}\|\widehat{\beta}^{(m)}-\beta\|=O_P(1)$. By simple algebra on (2.14) and the linear model for α_{st}, we obtain

$$\begin{aligned}&\Big[\sum_{s=1}^S X_s'X_s-\Big(\sum_{s=1}^S d_s^2\Big)^{-1}\sum_{s,g=1}^S X_s'D_sD_g'X_g\Big]\big(\widehat{\beta}^{(m)}-\beta\big)\\ &\qquad=\sum_{s=1}^S X_s'\big[(\widehat{A}_s-A_s)+\big(F\lambda_s-\widehat{F}^{(m)}\widehat{\lambda}_s^{(m)}\big)+\eta_s\big]\\ &\qquad\quad-\Big(\sum_{s=1}^S d_s^2\Big)^{-1}\sum_{s,g=1}^S X_s'D_sD_g\big[(\widehat{A}_g-A_g)+\big(F\lambda_g-\widehat{F}^{(m)}\widehat{\lambda}_g^{(m)}\big)+\eta_g\big],\end{aligned} \tag{2.17}$$

and

$$\widehat{\delta}_t^{(m)} - \delta_t = \big(\sum_{s=1}^{S} d_s^2\big)^{-1} \sum_{s=1}^{S} d_s\big((\widehat{\alpha}_{st} - \alpha_{st}) - x_{st}'(\widehat{\beta}^{(m)} - \beta) + \big(f_t'\lambda_s - \widehat{f}_t^{(m)\prime}\widehat{\lambda}_s^{(m)}\big) + \eta_{st}\big). \tag{2.18}$$

To find the rate of $\widehat{\beta}_j^{(m)}(u) - \beta_j(u)$, we derive the rates for all terms on the right-hand side of (2.17) below. As show in part (i), we have $\| \sum_{s=1}^{S} X_s'[(\widehat{A}_s - A_s) + \eta_s]\| = O_P(\sqrt{ST})$. We now consider the term $\sum_{s=1}^{S} X_s'(F\lambda_s - \widehat{F}^{(m)}\widehat{\lambda}_s^{(m)})$. Using the expression that $\widehat{\lambda}_s^{(m)} = T^{-1}\widehat{F}^{(m)\prime}(\widehat{A}_s - D_s\widehat{\delta}^{(m)} - X_s\widehat{\beta}^{(m)})$, we obtain

$$\begin{aligned}
\sum_{s=1}^{S} X_s'(F\lambda_s - \widehat{F}^{(m)}\widehat{\lambda}_s^{(m)}) =& \sum_{s=1}^{S} X_s'(F - \widehat{F}^{(m)}H^{(m)})\lambda_s + \sum_{s=1}^{S} X_s'\widehat{F}^{(m)}(H^{(m)}\lambda_s - \widehat{\lambda}_s^{(m)}) \\
=& - \sum_{s=1}^{S} X_s'P_{\widehat{F}}^{(m)}(\widehat{A}_s - A_s) - \sum_{s=1}^{S} X_s'P_{\widehat{F}}^{(m)}\eta_s \\
&+ \sum_{s=1}^{S} X_s'P_{\widehat{F}}^{(m)}X_s(\widehat{\beta}^{(m-1)} - \beta) + \sum_{s=1}^{S} X_s'P_{\widehat{F}}^{(m)}D_s(\widehat{\delta}^{(m-1)} - \delta) \\
&+ \sum_{s=1}^{S} X_s'M_{\widehat{F}}^{(m)}(F - \widehat{F}^{(m)}H^{(m)})\lambda_s,
\end{aligned} \tag{2.19}$$

where

$$H_j^{(m)}(u) = \widehat{\Upsilon}_j^{(m)}(u)\left(K_j^{(m)}(u)\right)^{-1}, \tag{2.20}$$

in which $\widehat{\Upsilon}_j^{(m)}(u)$ is the $r \times r$ diagonal matrix with diagonal elements being the r largest eigenvalues of $\widehat{L}_j\big(\widehat{\delta}_j^{(m-1)}(u), \widehat{\beta}_j^{(m-1)}(u)\big)$, defined in (2.6), in descending order, and

$$K_j^{(m)}(u) = \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)\left(\frac{F_j(u)'\widehat{F}_j^{(m)}(u)}{T}\right).$$

Now, we consider each term in the second equation of (2.19). For the first three terms, by

Hölder's and triangle inequalities, it is straightforward to show

$$\Big\|\sum_{s=1}^{S} X_s' P_{\widehat{F}}^{(m)} (\widehat{A}_s - A_s)\Big\| \leq \sup_s \|\widehat{A}_s - A_s\| \cdot \sum_{s=1}^{S} \|X_s\| \cdot \|P_{\widehat{F}}^{(m)}\| = O_P((ST)^{1/4}),$$
$$\Big\|\sum_{s=1}^{S} X_s' P_{\widehat{F}}^{(m)} X_s(\widehat{\beta}^{(m-1)} - \beta)\Big\| \leq \sum_{s=1}^{S} \|X_s\|^2 \cdot \|P_{\widehat{F}}^{(m)}\| \cdot \|\widehat{\beta}^{(m-1)} - \beta\| = O_P(\sqrt{ST}),$$

and $\| \sum_{s=1}^{S} X_s' P_{\widehat{F}}^{(m)} \eta_s \| = O_P(\sqrt{ST})$ by α-mixing property. Using the definition of D_s, the forth term is of order $O_P(T)$ as follows

$$\begin{aligned}\Big\|\sum_{s=1}^{S} X_s' P_{\widehat{F}}^{(m)} D_s(\widehat{\delta}^{(m-1)} - \delta)\Big\| &= \Big\|\frac{1}{T}\sum_{s=1}^{S} X_s' \widehat{F}^{(m)} d_s \sum_{t=T_0}^{T} \widehat{f}_t^{(m)} (\widehat{\delta}_t^{(m-1)} - \delta_t)\Big\| \\ &\leq \frac{1}{T}\sum_{s=1}^{S} \|X_s\| \cdot \|\widehat{F}^{(m)}\| \cdot \Big\|\sum_{t=T_0}^{T} \widehat{f}_t^{(m)} (\widehat{\delta}_t^{(m-1)} - \delta_t)\Big\| \\ &= O_P(T),\end{aligned}$$

provided $\| \sum_{t=T_0}^{T} \widehat{f}_t^{(m)} (\widehat{\delta}_t^{(m-1)} - \delta_t)\| = O_P(1)$. In addition, Lemma S.1.4 yields that the last term

$$\begin{aligned}\sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)} F\lambda_s = &- \sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)} \left(I_1^{(m)} + I_2^{(m)} + I_3^{(m)}\right) \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1} \lambda_s \\ &+ O_P(\sqrt{S}) \cdot \left|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t\right|^2 + O_P(T\sqrt{S}) \cdot \left\|\widehat{\beta}^{(m-1)} - \beta\right\|^2 \\ &+ O_P(\sqrt{ST}) \cdot \left|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t\right| \cdot \left\|\widehat{\beta}^{(m-1)} - \beta\right\| \\ &+ O_P(\sqrt{T}) \cdot \left(\left\|\widehat{\beta}^{(m-1)} - \beta\right\| + \left|\sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)}\right|\right) + O_P(\sqrt{ST}),\end{aligned}$$

where, according to the proof of Lemma S.1.4, the leading term associated with $I_1^{(m)}$ is of order $O_P(S) \cdot |\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t|^2 + O_P(ST) \cdot \|\widehat{\beta}^{(m-1)} - \beta\|^2 + O_P(S\sqrt{T}) \cdot |\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t| \cdot \|\widehat{\beta}^{(m-1)} - \beta\|$, and the leading terms associated with $I_2^{(m)}$ and $I_3^{(m)}$ are of order $O_P(\sqrt{ST}) \cdot (\|\widehat{\beta}^{(m-1)} - \beta\| + |\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m)} - \delta_t|)$. Thus, given the rates: $\sqrt{ST}\|\widehat{\beta}^{(m-1)} - \beta\| = O_P(1)$,

$\sqrt{S}|\widehat{\delta}_t^{(m-1)} - \delta_t| = O_P(1)$, $|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t| = O_P(1)$ and $\| \sum_{t=T_0}^{T} \widehat{f}_t^{(m)}(\widehat{\delta}_t^{(m-1)} - \delta_t)\| = O_P(1)$, we conclude that

$$\Big\| \sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)}(F - \widehat{F}^{(m)} H^{(m)})\lambda_s \Big\| = O_P(\sqrt{ST}).$$

Finally, given the rates of all terms on the right-hand side of (2.19), we have $\| \sum_{s=1}^{S} X_s'(F\lambda_s - \widehat{F}^{(m)}\widehat{\lambda}_s^{(m)})\| = O_P(\sqrt{ST})$ since $T/S \to \kappa$.

Collecting all terms so far, we obtain that the first term on the right-hand side of (2.17)

$$\Big\| \sum_{s=1}^{S} X_s'\big[(\widehat{A}_s - A_s) + \big(F\lambda_s - \widehat{F}^{(m)}\widehat{\lambda}_s^{(m)}\big) + \eta_s\big] \Big\| = O_P(\sqrt{ST}),$$

and similar arguments yield that

$$\Big\| \sum_{s,g=1}^{S} X_s' D_s D_g\big[(\widehat{A}_g - A_g) + \big(F\lambda_g - \widehat{F}^{(m)}\widehat{\lambda}_g^{(m)}\big) + \eta_g\big] \Big\| = O_P(S^{3/2}T^{1/2}).$$

In addition, as $(ST)^{-1}\big[\sum_{s=1}^{S} X_s'X_s - \big(\sum_{s=1}^{S} d_s^2\big)^{-1} \sum_{s,g=1}^{S} X_s' D_s D_g' X_g\big]$ and $S^{-1}\sum_{s=1}^{S} d_s^2$ are invertible and their inverse are bounded above according to Assumption 2.4.4, from (2.17) we obtain that

$$\begin{aligned}
\sqrt{ST}\big(\widehat{\beta}^{(m)} - \beta\big) =& \Big[\frac{1}{ST}\sum_{s=1}^{S} X_s'X_s - \Big(\frac{1}{S}\sum_{s=1}^{S} d_s^2\Big)^{-1} \frac{1}{S^2T}\sum_{s,g=1}^{S} X_s' D_s D_g' X_g\Big]^{-1} \\
&\cdot \Big\{ \frac{1}{\sqrt{ST}} \sum_{s=1}^{S} X_s'\big[(\widehat{A}_s - A_s) + \big(F\lambda_s - \widehat{F}^{(m)}\widehat{\lambda}_s^{(m)}\big) + \eta_s\big] \\
&\quad - \Big(\frac{1}{S}\sum_{s=1}^{S} d_s^2\Big)^{-1} \frac{1}{S^{3/2}T^{1/2}} \sum_{s,g=1}^{S} X_s' D_s D_g\big[(\widehat{A}_g - A_g) + \big(F\lambda_g - \widehat{F}^{(m)}\widehat{\lambda}_g^{(m)}\big) + \eta_g\big] \Big\} \\
=& O_P(1).
\end{aligned}$$

(iv) Finally, we consider the terms associated with $\widehat{\delta}_t^{(m)} - \delta_t$ for a given $m \geq 1$. We start by deriving the rates for all terms in (2.18). Recall from part (ii) that $|\sum_{s=1}^{S} d_s[(\widehat{\alpha}_{st} - \alpha_{st})| = O_P(S^{1/4}T^{-3/4})$, $|\sum_{s=1}^{S} d_s\eta_{st}| = O_P(\sqrt{S})$ and $|\sum_{s=1}^{S} d_s x_{st}'(\widehat{\beta}^{(m)} - \beta)| = O_P(\sqrt{S/T})$

given that $\|\widehat{\beta}^{(m)} - \beta\| = O_P(\sqrt{ST})$. Below we consider consider the term $|\sum_{s=1}^S d_s(f_t'\lambda_s - \widehat{f}_t^{(m)\prime}\widehat{\lambda}_s^{(m)})|$.

Using the estimation expression of $\widehat{\lambda}_s^{(m)}$, we obtain that

$$\begin{aligned}\sum_{s=1}^S d_s(f_t'\lambda_s - \widehat{f}_t^{(m)\prime}\widehat{\lambda}_s^{(m)}) = &- \frac{1}{T}\sum_{s=1}^S d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}(\widehat{A}_s - A_s) - \frac{1}{T}\sum_{s=1}^S d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}\eta_s \\ &+ \frac{1}{T}\sum_{s=1}^S d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}X_s(\widehat{\beta}^{(m-1)} - \beta) \\ &+ \frac{1}{T}\sum_{s=1}^S d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}D_s(\widehat{\delta}^{(m-1)} - \delta) \\ &+ \sum_{s=1}^S d_s e_t' M_{\widehat{F}}^{(m)}(F - \widehat{F}^{(m)}H^{(m)})\lambda_s. \end{aligned} \tag{2.21}$$

Using triangle- and Cauchy–Schwartz inequalities, it is easy to check the rates of the first four terms of (2.21) as follows

$$\left|\frac{1}{T}\sum_{s=1}^S d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}(\widehat{A}_s - A_s)\right| \le \frac{1}{T}\sup_s(\widehat{A}_s - A_s)\sum_{s=1}^S |d_s|\cdot\|\widehat{f}_t^{(m)}\|\cdot\|\widehat{F}^{(m)}\| = O_P(S^{-1/4}T^{-3/4}),$$

$$\left|\frac{1}{T}\sum_{s=1}^S d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}\eta_s\right| = \frac{1}{T}\left|\sum_{s=1}^S\sum_{l=1}^T d_s\widehat{f}_t^{(m)\prime}\widehat{f}_l^{(m)}\eta_{sl}\right| = O_P\left(\sqrt{\frac{S}{T}}\right),$$

$$\begin{aligned}&\left|\frac{1}{T}\sum_{s=1}^S d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}X_s(\widehat{\beta}^{(m-1)} - \beta)\right| \\ &\quad\le \frac{1}{T}\left(\sum_{s=1}^S |d_s|^2\right)^{1/2}\cdot\left(\sum_{s=1}^S \|X_s\|^2\right)^{1/2}\cdot\|\widehat{f}_t^{(m)}\|\cdot\|\widehat{F}^{(m)}\|\cdot\|\widehat{\beta}^{(m-1)} - \beta)\| = O_P\left(\sqrt{\frac{S}{T}}\right),\end{aligned}$$

$$\left|\frac{1}{T}\sum_{s=1}^S d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}D_s(\widehat{\delta}^{(m-1)} - \delta)\right| \le \frac{1}{T}\sum_{s=1}^S |d_s|\cdot\|\widehat{f}_t^{(m)}\|\cdot\left\|\sum_{l=T_0}^T \widehat{f}_l^{(m)}(\widehat{\delta}_l^{(m-1)} - \delta_l)\right| = O_P\left(\frac{S}{T}\right),$$

given that $\sup_s \widehat{A}_s - A_s = O_P(S^{-3/4}T^{-1/4})$ from Lemma S.1.1, $\|\sum_{l=T_0}^T \widehat{f}_l^{(m)}(\widehat{\delta}_l^{(m-1)} - \delta_l)| = O_P(1)$ and $\|\widehat{\beta}^{(m-1)} - \beta\| = O_P((ST)^{-1/2})$ from induction assumption. Lastly, according to Lemma S.1.5, we have

$$\sum_{s=1}^S d_s e_t' M_{\widehat{F}}^{(m)}(F - \widehat{F}^{(m)}H^{(m)})\lambda_s = \frac{1}{S}\sum_{s,g=1}^S\sum_{l=T_0}^T \omega_{sg}d_s d_g(\widehat{\delta}_t^{(m-1)} - \delta_t) - \frac{1}{S}\sum_{s,g=1}^S \omega_{sg}d_s\eta_{gt}$$

$$- \frac{1}{ST} \sum_{s,g=1}^{S} d_s \mathbb{E}[\eta_{gt}\eta_g'] \widehat{F}^{(m)} (K^{(m)})^{-1} \lambda_s + O_P\Big(\sqrt{\frac{S}{T}}\Big),$$

where the remaining terms on the right-hand side have been shown to be $O_P(\sqrt{S})$ in Lemma S.1.5. Therefore, given the rates of all terms of (2.18), we finally obtain that

$$\begin{aligned}
\sqrt{S}\big(\widehat{\delta}_t^{(m)} - \delta_t\big) = & \Big(\frac{1}{S}\sum_{s=1}^{S} d_s^2\Big)^{-1} \cdot \frac{1}{\sqrt{S}} \Big[\sum_{s=1}^{S} \Big(d_s - \frac{1}{S}\sum_{g=1}^{S} \omega_{sg} d_g\Big) \eta_{st} \\
& - \frac{1}{S} \sum_{s,g=1}^{S} \sum_{l=T_0}^{T} \omega_{sg} d_s d_g (\widehat{\delta}_t^{(m-1)} - \delta_t) \\
& - \frac{1}{ST} \sum_{s,g=1}^{S} d_s \mathbb{E}[\eta_{gt}\eta_g'] \widehat{F}^{(m)} (K^{(m)})^{-1} \lambda_s \Big] + o_P(1),
\end{aligned} \tag{2.22}$$

and the leading terms on the right-hand side have shown to be $O_P(1)$, and thus, we conclude that $\sqrt{S}|\widehat{\delta}_t^{(m)} - \delta_t| = O_P(1)$.

To complete the proof of induction, it requires to further show $|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m)} - \delta_t| = O_P(1)$ and $|\sum_{t=T_0}^{T} \widehat{f}_t^{(m+1)\prime}(\widehat{\delta}_t^{(m)} - \delta_t)| = O_P(1)$, which can be proof in the same manner as above, and thus is omitted here. Finally, combining parts (i) to (iv), we complete the proof of Theorem 2.15.

■

Proof of Theorem 2.4.2. To complete the proof, we first show that, for any given $u \in \mathcal{U}$, $j = 1, ..., J$ and $t \geq T_0$, the converged estimator $\widehat{\delta}_{jt}(u)$ has an asymptotic linear expansion around the true parameter, and then establish the asymptotic normality for the joint policy parameter $(\widehat{\delta}_t(u_1)', \widehat{\delta}_t(u_2)')'$.

Recall from the proof of Theorem 2.4.1, we derived an asymptotic representation for the iterative estimator $\widehat{\delta}_{jt}^{(m)}(u)$ for $m \geq 1$ as (2.22), where the subscript j and quantile u is omitted. Therefore, when the estimators converges, we have

$$\sqrt{S}\big(\widehat{\delta}_{jt}(u) - \delta_{jt}(u)\big)$$

$$=\left(\frac{1}{S}\sum_{s=1}^{S}d_s^2\right)^{-1}\cdot\frac{1}{\sqrt{S}}\Bigg[\sum_{s=1}^{S}R_{js}(u)\eta_{jst}(u)-\frac{1}{S}\sum_{s,g=1}^{S}\sum_{l=T_0}^{T}\omega_{j,sg}(u)d_sd_g\big(\widehat{\delta}_{jt}(u)-\delta_{jt}(u)\big)$$
$$-\frac{1}{ST}\sum_{s,g=1}^{S}d_s\mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)']\widehat{F}_j(u)\left(\frac{F_j(u)'\widehat{F}_j(u)}{T}\right)^{-1}\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\lambda_{js}(u)\Bigg]+o_P(1),$$

where $R_{js}(u)$ and $\omega_{j,sg}(u)$ are defined above Assumption 2.4.4. Equivalently, we have

$$\sqrt{S}\big(\widehat{\delta}_{jt}(u)-\delta_{jt}(u)\big)$$
$$=\left(\frac{1}{S}\sum_{s=1}^{S}R_{js}(u)^2\right)^{-1}\cdot\frac{1}{\sqrt{S}}\Bigg[\sum_{s=1}^{S}R_{js}(u)\eta_{jst}(u)$$
$$-\frac{1}{ST}\sum_{s,g=1}^{S}d_s\mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)']\widehat{F}_j(u)\left(\frac{F_j(u)'\widehat{F}_j(u)}{T}\right)^{-1}\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\lambda_{js}(u)\Bigg]+o_P(1),$$

given the fact that

$$\frac{1}{S}\sum_{s=1}^{S}d_s^2-\frac{1}{S^2}\sum_{s,g=1}^{S}\sum_{l=T_0}^{T}\omega_{j,sg}(u)d_sd_g=\frac{1}{S}\sum_{s=1}^{S}R_{js}(u)^2>0$$

according to the definition of $\omega_{j,sg}(u)$ and Assumption 2.4.4.(i).

Next, we claim that

$$\frac{1}{S^{3/2}T}\sum_{s,g=1}^{S}d_s\mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)']\widehat{F}_j(u)\left(\frac{F_j(u)'\widehat{F}_j(u)}{T}\right)^{-1}\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\lambda_{js}(u)$$
$$\xrightarrow{p}\frac{1}{S^{3/2}T}\sum_{s,g=1}^{S}d_s\mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)']F_j(u)\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\lambda_{js}(u).$$

To prove this claim, it is sufficient to show that

$$\left\|\widehat{F}_j(u)\left(\frac{F_j(u)'\widehat{F}_j(u)}{T}\right)^{-1}\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}-F_j(u)\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\right\|$$
$$\leq\left\|\widehat{F}_j(u)-F_j(u)\left(\frac{F_j(u)'\widehat{F}_j(u)}{T}\right)^{-1}\right\|\cdot\left\|\left(\frac{F_j(u)'\widehat{F}_j(u)}{T}\right)^{-1}\right\|\cdot\left\|\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\right\|$$
$$\leq\left\|\widehat{F}_j(u)-P_{F_j}(u)\widehat{F}_j(u)\right\|\cdot\left\|\left(\frac{F_j(u)'\widehat{F}_j(u)}{T}\right)^{-1}\right\|\cdot\left\|\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\right\|$$

$$
\begin{aligned}
&\leq \sqrt{T}\|P_{\widehat{F}_j}(u) - P_{F_j}(u)\| \cdot \left\| \left(\frac{F_j(u)'\widehat{F}_j(u)}{T} \right)^{-1} \right\| \cdot \left\| \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S} \right)^{-1} \right\| \\
&= o_P(\sqrt{T}),
\end{aligned}
$$

since $\|P_{\widehat{F}_j}(u) - P_{F_j}(u)\| = o_P(1)$, which can be shown similar to (S.15) of Lemma S.1.10. Thus,

$$
\begin{aligned}
&\left\| \frac{1}{S^{3/2}T} \sum_{s,g=1}^{S} d_s \mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)'] \widehat{F}_j(u) \left(\frac{F_j(u)'\widehat{F}_j(u)}{T} \right)^{-1} \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S} \right)^{-1} \lambda_{js}(u) \right. \\
&\qquad \left. - \frac{1}{S^{3/2}T} \sum_{s,g=1}^{S} d_s \mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)'] F_j(u) \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S} \right)^{-1} \lambda_{js}(u) \right\| \\
&\leq \frac{1}{S^{3/2}T} \cdot \left\| \sum_{s=1} d_s \lambda_{js}(u) \right\| \cdot \left\| \sum_{g=1}^{S} \mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)'] \right\| \\
&\quad \cdot \left\| \widehat{F}_j(u) \left(\frac{F_j(u)'\widehat{F}_j(u)}{T} \right)^{-1} \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S} \right)^{-1} - F_j(u) \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S} \right)^{-1} \right\| \\
&= \frac{1}{S^{3/2}T} \cdot O_P(S) \cdot O_P(S) \cdot o_P(\sqrt{T}) = o_P(1).
\end{aligned}
$$

Therefore, we obtain the asymptotic expansion of $\widehat{\delta}_{jt}(u)$ as

$$
\begin{aligned}
&\sqrt{S}\big(\widehat{\delta}_{jt}(u) - \delta_{jt}(u)\big) \\
&= - \left(\frac{1}{S} \sum_{s=1}^{S} R_{js}(u)^2 \right)^{-1} \frac{1}{S^{3/2}T} \sum_{s,g=1}^{S} d_s \mathbb{E}[\eta_{j,gt}(u)\eta_{jg}(u)'] F_j(u) \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S} \right)^{-1} \lambda_{js}(u) \\
&\quad + \left(\frac{1}{S} \sum_{s=1}^{S} R_{js}(u)^2 \right)^{-1} \cdot \frac{1}{\sqrt{S}} \sum_{s=1}^{S} R_{js}(u)\eta_{jst}(u) + o_P(1), \qquad (2.23)
\end{aligned}
$$

where the first term on the right-hand side has the probability limit $B_{jt}(u)$ defined in Theorem 2.4.2

Combining (2.23) with the definition of $\widehat{\delta}_t(u)$, we obtain that

$$
\sqrt{S}\big(\widehat{\delta}_t(u) - \delta_t(u)\big) = B_t(u) + K_t(u) + o_P(1),
$$

where $B_t(u)$ is defined in Theorem 2.4.2, and $K_t(u)$ is defined above Assumption 2.4.7. Finally, combining the representation with Assumption 2.4.7, we establish the joint CLT in Theorem 2.4.2.

■

Proof of Corollary 2.4.2. **(i)** To show $\widehat{B}_t(u) \xrightarrow{p} B_t(u)$, it is sufficient to show that $\widehat{B}_{jt}(u) \xrightarrow{p} B_{jt}(u)$ for any given $j = 1, ..., J$ and $u \in \mathcal{U}$.

We first note that, under the assumption of no cross-sectional dependence, the asymptotic bias $B_{jt}(u)$ is reduced to

$$B_{jt}(u) = \operatorname*{plim}_{S,T\to\infty} -\left(\frac{1}{S}\sum_{s=1}^{S} R_{js}(u)^2\right)^{-1} \cdot \frac{1}{S^{3/2}T}\sum_{s,g=1}^{S} d_s\mathbb{E}[\eta_{j,gt}(u)^2]f_{jt}(u)'\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\lambda_{js}(u).$$

To shown the consistency of $\widehat{B}_{jt}(u)$, it is sufficient to prove the following two claims:

$$\left\|\frac{1}{S}\sum_{s=1}^{S} R_{js}(u)^2 - \frac{1}{S}\sum_{s=1}^{S}\widehat{R}_{js}(u)^2\right\| = o_P(1), \tag{2.24}$$

$$\left\|\frac{1}{S^{3/2}T}\sum_{s,g=1}^{S} d_s\mathbb{E}[\eta_{j,gt}(u)^2]f_{jt}(u)'\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\lambda_{js}(u) - \frac{1}{S^{3/2}T}\sum_{s,g=1}^{S} d_s\big(\widehat{\eta}_{j,gt}(u)\big)^2\widehat{f}_{jt}(u)'\left(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1}\widehat{\lambda}_{js}(u)\right\| = o_P(1). \tag{2.25}$$

We start with the proof of (2.24). Using the identity $a^2 - b^2 = (a+b)(a-b)$ and Hölder's inequality, we obtain that

$$\left\|\frac{1}{S}\sum_{s=1}^{S} R_{js}(u)^2 - \frac{1}{S}\sum_{s=1}^{S}\widehat{R}_{js}(u)^2\right\| \leq \frac{1}{S}\left[\sum_{s=1}^{S}\left\|\frac{1}{S}\sum_{g=1}^{S}\big(\widehat{\omega}_{j,sg}(u) + \omega_{j,sg}(u)\big)d_g - 2d_s\right\|^2\right]^{1/2} \cdot \left[\sum_{s=1}^{S}\left\|\frac{1}{S}\sum_{g=1}^{S}\big(\widehat{\omega}_{j,sg}(u) - \omega_{j,sg}(u)\big)d_g\right\|^2\right]^{1/2}. \tag{2.26}$$

It is straightforward to show that

$$\begin{aligned}
&\Big[\sum_{s=1}^{S}\Big\|\frac{1}{S}\sum_{g=1}^{S}\big(\widehat{\omega}_{j,sg}(u)+\omega_{j,sg}(u)\big)d_g-2d_s\Big\|^2\Big]^{1/2}\\
=&\frac{1}{S}\Big\|\sum_{g=1}^{S}\Big[\widehat{\Lambda}_j(u)\Big(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\Big)^{-1}\widehat{\lambda}_{jg}(u)+\Lambda_j(u)\Big(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\Big)^{-1}\lambda_{jg}(u)\Big]d_g-2D_s\Big\|\\
=&\frac{1}{S}\Big\{\Big\|\widehat{\Lambda}_j(u)\Big(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\Big)^{-1}\sum_{g=1}^{S}\widehat{\lambda}_{jg}(u)d_g\Big\|+\Big\|\Lambda_j(u)\Big(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\Big)^{-1}\sum_{g=1}^{S}\lambda_{jg}(u)d_g\Big\|\Big\}\\
=&O_P(\sqrt{S}), \qquad (2.27)
\end{aligned}$$

since

$$\begin{aligned}
\Big\|\Lambda_j(u)\Big(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\Big)^{-1}\sum_{g=1}^{S}\lambda_{jg}(u)d_g\Big\| &\leq \|\Lambda_j(u)\|\cdot\Big\|\Big(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\Big)^{-1}\Big\|\cdot\Big\|\sum_{g=1}^{S}\lambda_{jg}(u)d_g\Big\|\\
&=O_P(S^{3/2})
\end{aligned}$$

and the same rate holds for the first term in the second last equation.

Given the expression of $\omega_{j,sg}(u)$ and the identity that $\widehat{a}\widehat{b}\widehat{c}-abc=(\widehat{a}-a)\widehat{b}\widehat{c}+a(\widehat{b}-b)\widehat{c}+ab(\widehat{c}-c)$, we write

$$\begin{aligned}
&\Big[\sum_{s=1}^{S}\Big\|\frac{1}{S}\sum_{g=1}^{S}\big(\widehat{\omega}_{j,sg}(u)+\omega_{j,sg}(u)\big)d_g-2d_s\Big\|^2\Big]^{1/2}\\
=&\frac{1}{S}\Big\|\sum_{g=1}^{S}\Big[\widehat{\Lambda}_j(u)\Big(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\Big)^{-1}\widehat{\lambda}_{jg}(u)-\Lambda_j(u)\Big(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\Big)^{-1}\lambda_{jg}(u)\Big]d_g\Big\|\\
\leq&\frac{1}{S}\Big\{\Big\|\Big(\widehat{\Lambda}_j(u)-\Lambda_j(u)H_j(u)'\Big)\Big(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\Big)^{-1}\sum_{g=1}^{S}\widehat{\lambda}_{jg}(u)d_g\Big\|\\
&+\Big\|\Lambda_j(u)H_j(u)'\Big[\Big(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\Big)^{-1}\\
&\qquad-\big(H_j(u)'\big)^{-1}\Big(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\Big)^{-1}\big(H_j(u)\big)^{-1}\Big]\sum_{g=1}^{S}\widehat{\lambda}_{jg}(u)d_g\Big\|\\
&+\Big\|\Lambda_j(u)\Big(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\Big)^{-1}\big(H_j(u)\big)^{-1}\Big[\sum_{g=1}^{S}\widehat{\lambda}_{jg}(u)-H_j(u)\lambda_{jg}(u)\Big]d_g\Big\|\Big\}.
\end{aligned}$$

Combining (S.18) of Lemma S.1.11 with (S.9) and Theorem 2.4.1, we obtain that $\|\widehat{\Lambda}_j(u) - \Lambda_j(u)H_j(u)'\| = O_P(1)$, and $\|(S^{-1}\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u))^{-1} - (H_j(u)')^{-1}(S^{-1}\Lambda_j(u)'\Lambda_j(u))^{-1}(H_j(u))^{-1}\| = O_P(\delta_{ST}^{-1})$. Thus, by Cauchy–Schwartz inequality, we obtain that the first term in the last inequality is

$$
\begin{aligned}
&\left\|\left(\widehat{\Lambda}_j(u) - \Lambda_j(u)H_j(u)'\right)\left(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1}\sum_{g=1}^{S}\widehat{\lambda}_{jg}(u)d_g\right\| \\
&\quad=\left\|\widehat{\Lambda}_j(u) - \Lambda_j(u)H_j(u)'\right\| \cdot \left\|\left(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1}\right\| \cdot \left\|\sum_{g=1}^{S}\widehat{\lambda}_{jg}(u)d_g\right\| = O_P(S),
\end{aligned}
$$

and the same rate applies to the second term. For the third term,

$$
\begin{aligned}
&\left\|\Lambda_j(u)\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\left(H_j(u)\right)^{-1}\left[\sum_{g=1}^{S}\widehat{\lambda}_{jg}(u) - H_j(u)\lambda_{jg}(u)\right]d_g\right\| \\
&\le \|\Lambda_j(u)\| \cdot \left\|\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\right\| \cdot \left\|\left(H_j(u)\right)^{-1}\right\| \\
&\qquad \cdot \left(\sum_{g=1}^{S}\|\widehat{\lambda}_{jg}(u) - H_j(u)\lambda_{jg}(u)\|^2\right)^{1/2} \cdot \left(\sum_{g=1}^{S}|d_g|^2\right)^{1/2} \\
&\le \|\Lambda_j(u)\| \cdot \left\|\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\right\| \cdot \left\|\left(H_j(u)\right)^{-1}\right\| \cdot \left\|\widehat{\Lambda}_j(u) - \Lambda_j(u)H_j(u)'\right\|^2 \cdot \left(\sum_{g=1}^{S}|d_g|^2\right)^{1/2} \\
&= O_P(S).
\end{aligned}
$$

Collecting all terms, we have

$$
\left[\sum_{s=1}^{S}\left\|\frac{1}{S}\sum_{g=1}^{S}\left(\widehat{\omega}_{j,sg}(u) + \omega_{j,sg}(u)\right)d_g - 2d_s\right\|^2\right]^{1/2} = O_P(1).
$$

And thus, together with (2.26) and (2.27), we prove the claim (2.24).

To prove the second claim (2.25), we consider two terms

$$
\frac{1}{S^{3/2}T}\sum_{s,g=1}^{S} d_s\mathbb{E}[\eta_{j,gt}(u)^2]\left(f_{jt}(u)'\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1}\lambda_{js}(u) - \widehat{f}_{jt}(u)'\left(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1}\widehat{\lambda}_{js}(u)\right),
$$

and

$$\frac{1}{S^{3/2}T}\sum_{s,g=1}^{S} d_s\Big(\mathbb{E}[\eta_{j,gt}(u)^2]-\widehat{\eta}_{j,gt}(u)^2\Big)\widehat{f}_{jt}(u)'\left(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1}\widehat{\lambda}_{js}(u).$$

The first term is $o_P(1)$ by expanding the terms related to factor and loadings and apply Lemma S.1.10 and S.1.11. For the second term, it is easy to show that $\sum_{s=1}^{S}\left(\eta_{j,st}(u)^2-\widehat{\eta}_{j,st}^{(m)}(u)^2\right)=O_P(\sqrt{S})$ and $\sum_{s=1}^{T}\mathbb{E}[\eta_{j,st}(u)^2]-\eta_{j,st}(u)^2=O_P(\sqrt{S})$, which leads to

$$\begin{aligned}
&\left\|\frac{1}{S^{3/2}T}\sum_{s,g=1}^{S} d_s\Big(\mathbb{E}[\eta_{j,gt}(u)^2]-\widehat{\eta}_{j,gt}(u)^2\Big)\widehat{f}_{jt}(u)'\left(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1}\widehat{\lambda}_{js}(u)\right\|\\
&\quad\leq\frac{1}{S^{3/2}T}\cdot\left\|\sum_{g=1}^{S}\mathbb{E}[\eta_{j,gt}(u)^2]-\widehat{\eta}_{j,gt}(u)^2\right\|\cdot\|\widehat{f}_{jt}(u)\|\cdot\left\|\left(\frac{\widehat{\Lambda}_j(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1}\right\|\cdot\left\|\sum_{s=1}^{S}d_s\widehat{\lambda}_{js}(u)\right\|\\
&\quad=o_P(1).
\end{aligned}$$

Then, given these two claims, it is straightforward that $\widehat{B}_t(u)\xrightarrow{p}B_t(u)$.

(ii) We first note that under the assumption of no cross-sectional correlation, the $(j,k)^{\text{th}}$ entry of $\Sigma_t(u_1,u_2)$ is given by

$$\begin{aligned}
\sigma_{t,jk}(u_1,u_2)=\operatorname*{plim}_{S,T\to\infty}&\left(\frac{1}{S}\sum_{s=1}^{S}R_{js}(u_1)^2\right)^{-1}\left(\frac{1}{S}\sum_{s=1}^{S}R_{ks}(u_2)^2\right)^{-1}\\
&\cdot\frac{1}{S}\sum_{s=1}^{S}R_{js}(u_1)R_{ks}(u_2)\eta_{jst}(u_1)\eta_{kst}(u_2).
\end{aligned}$$

Thus, to show $\widehat{\Sigma}_t(u_1,u_2)\xrightarrow{p}\Sigma_t(u_1,u_2)$, it is sufficient to show $\widehat{\sigma}_{t,jk}(u_1,u_2)\xrightarrow{p}\sigma_{t,jk}(u_1,u_2)$, where $\widehat{\sigma}_{t,jk}(u_1,u_2)$ is the $(j,k)^{\text{th}}$ entry of $\widehat{\Sigma}_t(u_1,u_2)$. Given (2.24), it remains to show

$$\frac{1}{S}\sum_{s=1}^{S}\widehat{R}_{js}(u_1)\widehat{R}_{ks}(u_2)\widehat{\eta}_{j,st}(u_1)\widehat{\eta}_{k,st}(u_2)\xrightarrow{p}\frac{1}{S}\sum_{s=1}^{S}\mathbb{R}_{js}(u_1)R_{ks}(u_2)\eta_{j,st}(u_1)\eta_{k,st}(u_2).\qquad(2.28)$$

Using the identity that $\widehat{a}\widehat{b} - ab = (\widehat{a} - a)\widehat{b} + a(\widehat{b} - b)$, we first note that

$$\left|\frac{1}{S}\sum_{s=1}^{S}\left[\widehat{R}_{js}(u_1)\widehat{R}_{ks}(u_2) - R_{js}(u_1)R_{ks}(u_2)\right]\eta_{jst}(u_1)\eta_{kst}(u_2)\right| = o_P(1),$$

whose proof is similar to (2.24). It remains to consider the term

$$\left|\frac{1}{S}\sum_{s=1}^{S}R_{js}(u_1)R_{ks}(u_2)\big(\widehat{\eta}_{jst}(u_1)\widehat{\eta}_{kst}(u_2) - \eta_{jst}(u_1)\eta_{kst}(u_2)\big)\right|.$$

From

$$\begin{aligned}&\left|\frac{1}{S}\sum_{s=1}^{S}d_s^2\big(\widehat{\eta}_{jst}(u_1)\widehat{\eta}_{k,st}(u_2) - \eta_{jst}(u_1)\eta_{kst}(u_2)\big)\right|\\ \leq&\left|\frac{1}{S}\sum_{s=1}^{S}d_s\big(\widehat{\eta}_{jst}(u_1) - \eta_{jst}(u_1)\big)\widehat{\eta}_{kst}(u_2)\right| + \left|\frac{1}{S}\sum_{s=1}^{S}d_s\big(\widehat{\eta}_{kst}(u_2) - \eta_{k,st}(u_2)\big)\eta_{jst}(u_1)\right|,\end{aligned}$$

and the expression of $\widehat{\eta}_{jst}(u)$, we write

$$\begin{aligned}&\left|\sum_{s=1}^{S}d_s\big(\widehat{\eta}_{kst}(u_2) - \eta_{kst}(u_2)\big)\eta_{j,st}(u_1)\right|\\ \leq&\left|\sum_{s=1}^{S}d_s\eta_{jst}(u_1)(\widehat{\alpha}_{jst}(u) - \alpha_{jst}(u))\right| + \left|\sum_{s=1}^{S}d_s\eta_{jst}(u_1)d_s(\delta_{jt}(u) - \widehat{\delta}_{jt}(u))\right|\\ &+\left|\sum_{s=1}^{S}d_s\eta_{jst}(u_1)x_{st}'(\beta_j(u) - \widehat{\beta}_j(u))\right| + \left|\sum_{s=1}^{S}d_s\eta_{jst}(u_1)(f_{jt}(u)'\lambda_{js}(u) - \widehat{f}_{jt}(u)'\widehat{\lambda}_{js}(u)])\right|\\ =&o_P(1),\end{aligned}$$

whose proof are similar to part (iv) in the proof of Theorem 2.4.1. Given a similar reason, we also have $|S^{-1}\sum_{s=1}^{S}d_s(\widehat{\eta}_{jst}(u_1) - \eta_{jst}(u_1))\widehat{\eta}_{kst}(u_2)| = o_P(1)$. In addition, replacing d_s with $S^{-1}\sum_{g=1}^{S}\omega_{j,sg}(u)d_g$ does not affect the convergence rate. Therefore, we obtain that

$$\left|\frac{1}{S}\sum_{s=1}^{S}R_{js}(u_1)R_{ks}(u_2)\big(\widehat{\eta}_{jst}(u_1)\widehat{\eta}_{kst}(u_2) - \eta_{jst}(u_1)\eta_{kst}(u_2)\big)\right| = o_P(1),$$

which completes the proof of (2.28). Thus, the proof of (ii) is complete. ■

Chapter 3

Quantile Functional-Coefficient Regression for Conditional Asset Pricing

3.1 Introduction

Financial time series have strong common movements, which are powerful in returns forecast of securities, and thus have received intensive research interest both empirically and methodologically. To model the co-movement, linear factor models are the cornerstone of the modern portfolio theory, among which the capital asset pricing model (CAPM) (Sharpe, 1964) is the most fundamental. Building upon the idea of CAPM, a number of market- and firm-specific factors are identified (e.g., Fama and French, 1993; Kogan and Papanikolaou, 2013, 2014; Fama and French, 2015).

Although linear factor models have dominated the literature for decades, there are at least three limitations. First, increasing empirical studies have shown that factor loadings and risk premia are subject to time variation due to changes in the economic cycle and asset characteristics. To this end, conditional linear factor models, which assume functional specifications (linear forms in general) of financial and macroeconomic variables for the factor loadings, have been extensively studied (see e.g., Shanken, 1990; Ferson and Schadt, 1996; Petkova and Zhang, 2005; Connor et al., 2012; Gagliardini et al., 2016). We refer to Gagliardini et al. (2020)

for a recent survey. Second, there is a limited access to all common factors in reality. As such, an approximate (unobservable) factor structure under the theoretical framework of the Arbitrage Pricing Theory (APT) is first considered for small panels (Roll and Ross, 1980), and then generated to large panels thanks to the recent development in panel factor models (e.g. Pesaran, 2006; Bai, 2009; Bai and Li, 2012; Cai et al., 2022). Last but not least, the prevailing works on the expected financial asset returns are built upon the restrictive finite-fourth-moment assumption, which has shown to be violated empirically in many cases (Bradley and Taqqu, 2003). In contrast, quantile regression (QR) has emerged as a promising alternative to obtain robust estimation for data with heavy-tails and outliers since the seminal work of Koenker and Bassett Jr (1978). For example, the median regression estimators are comparable to those of Gagliardini et al. (2016), which can be used to test asset pricing theories under comparable quantile restrictions, and to design investment strategies. On top of that, quantile asset pricing methods provide comprehensive insights about the conditional distribution away from the center, which is of its own importance. One possible interpretation of the estimates at bottom and upper quantiles is that they represent performance of firms or portfolios that unperformed or outperformed in the sense that the conditional mean under- or over-predicts the return due to receiving negative (or positive) idiosyncratic shocks or bad (or good) news of varying magnitude.

In view of the above discussion, we introduce a flexible functional-coefficients panel quantile model with an unobservable factor structure:

$$Q_{y_{it}}(\tau|x_{it}, v_{it}, f_t(\tau)) = x_{it}'\beta_i(\tau, v_{it}) + f_t'(\tau)\lambda_i(\tau),$$

where y_{it} is the excess stock return at time t of stock i, x_{it} is a set of firm-specific or common factors. We consider β_i to be a vector of smooth functions of v_{it}, which include constant, simple linear and quadratic forms as special cases. Instead of considering deterministic time trend, which is known as time-varying regression, we consider v_{it} as a set of stationary firm-level or macroeconomic factors to better explain the variation in β_i. We refer to Avramov and Chordia (2006) for convincing theoretical and empirical arguments in favour of the chosen conditional specification. The use of functional-coefficients generalises the conditional linear factor models by allowing non-linear β's, and provides insight of how the impact of firm characteristics on stock returns varies under different economic conditions. The unobservable factor $f_t(\tau)$ aims to capture the unobservable market common shocks, and the corresponding loadings $\lambda_i(\tau)$ represent heterogeneous responses across individual stocks. In addition, the beta-coefficients, unobserved factors and loadings are all allowed to be quantile-dependent.

We apply this model to re-examine the Fama-French (FF) five factor models using the FF 25 and 100 portfolios sorted by size and book-to-market ratio from January 1983 to July 2022, to gain insight into the quantile co-movement structure of asset return distribution. There is a growing body of literature on quantile linear factor models for financial time series. To name a few, Ma and Pohlman (2008) identify a set of observable factors with quantile regression. Galvao et al. (2018) consider a linear panel quantile regression model with individual fixed effects, and multiple sets of widely considered firm characteristic variables are tested for the slope homogeneity. They find that the effect of firm characteristics to the stock excess return is heterogeneous across stocks at the tail quantiles, while the effect is homogeneous at the mean and central quantiles. They argue that the heterogeneous slope coefficients in the tails correspond to the distress and boom period and provide the conjecture that investors' overreaction produces the under-pricing of distressed stocks and over-pricing of the boom stocks. Under the mean regression framework, this phenomenon has been identified as the business-cycle dependent feature (e.g., McQueen and Roley, 1993; Jagannathan and Wang, 1996; Celebi and Hönig, 2019) and modelled by the conditional linear factor models. However, as far as the author is awared, there is a lack of explicit modelling of the business-cycle dependent feature when considering the quantile co-movement structure of the stock returns. With this model, this chapter seeks to answer the following empirical questions:

(i) Do risk exposures of the observed factors vary across the business cycle?

(ii) Are risk exposures different in the tail and central quantiles?

We point out that this generic framework not only has empirical interest in terms of quantile co-movement in large financial panels, but is also of methodological importance. Our study is closely related to the literature of functional-coefficient models and panel factor models. Although there have been a few papers in the integration of these two fields under the mean regression framework (e.g. Liu et al., 2018; Dong et al., 2021; Casas et al., 2021), this general model lacks research under the quantile paradigm. Below we review these two fields respectively.

Suppose that the unobservable factor structure is negligible, then the estimator of the heterogeneous and functional-coefficient regression coefficients can be found by running the semi-parametric quantile regression for each cross-section. Two major strands of the estimation approaches include the sieve estimation method (Kim, 2007; Wang et al., 2009; Su and Hoshino, 2016), which converts the nonparametric function into an approximate parametric form that is easy to estimate, and the kernel method (Cai and Xu, 2008; Cai and Xiao, 2012), which estimates the functional-coefficient via a local approximation. In order to obtain a clear view of how the regression coefficient depends on the macroeconomic conditions, the sieve approach is adopted in our study. Note that our model allows the explanatory variables to depend on the unobservable effects. Therefore, if the unobservable factor exists but is ignored, indicating the ignorance of possible endogenity, then the standard functional-coefficient quantile regression approach produces biased results.

The use of panel factor models have been common practice in capturing the heterogeneous impact of unobserved macro-economic factors. In the mean regression paradigm, estimation and inference procedure is well studied (to name a few, Pesaran, 2006; Bai, 2009; Moon and Weidner, 2015; Jiang et al., 2021b). However, literature on quantile interactive fixed effect models remains limited until very recently due to the inherent technical difficulties of quantile estimation caused by the non-smooth quantile objective function, for example, Ando and Bai (2020); Chen et al. (2021); Belloni et al. (2023); Ma et al. (2021) and Ando et al. (2023). However, the above literature either ignores the potential covariates or only allows for a static relationship for the covariates.

In the light of the above discussion, our contributions are threefold. Firstly, the model we developed here extends the semiparametric quantile regression models from cross-sectional or time series data to panel data, and flexibly captures the intrinsic latent heterogeneity in panel data. Nonlinear (nonparametric and semiparametric) quantile regression models are extensively studied for cross-sectional data or time series data (e.g., Kim, 2007; Wang et al., 2009; Su and Hoshino, 2016; Cai and Xu, 2008; Cai and Xiao, 2012). Nevertheless, those methods suffer from "omitted-variable bias" on panel data if there exists latent heterogeneity. To the best of our knowledge, the first attempt to control for the unobserved heterogeneity in a functional-coefficient panel quantile regression was made by Atak et al. (2023). They include an additional latent factor structure to the time-varying quantile regression, and the unobserved factors are assumed to be location shift from the mean factors. The location-shift assumption largely simplifies the estimation approach and theoretical derivation, but is somewhat restrictive, given the empirical evidence that unobserved factors can act distinctively and the number of common factors may vary across quantiles (Ando and Bai, 2020). In this chapter, we consider a functional-coefficient quantile regression model for panel data, and use a latent factor structure to control the flexible unobservable cross-sectional and temporal effects. We do not impose any structure on the latent factors, allowing them to be heterogeneous across quantiles and correlated with the observed covariates. Such a generic model has been extensively studied in the conditional mean setting (see, for example, Dong et al., 2021; Cai et al., 2022); our model can be viewed as the first quantile extension of Cai et al. (2022).

Second, we develop an iterative procedure, which alternates between the quantile sieve estimation and conventional quantile regression estimation, to obtain the estimates of the varying coefficients, factors and loadings. This procedure is straightforward to implement and produces quick convergence, which is justified by several simulated examples. Unlike the existing literature on quantile factor models, the asymptotic properties for the recursive estimators are established by approximating the standard quantile loss function with a series of generalised sequences (Gel'fand and Shilov, 1964), which overcome the theoretical difficulties caused by the non-smooth quantile loss function and greatly simplify the proof.

Last but not least, we construct the specification test for the constancy of betas and significance of alphas using a Wald-type statistic, whose empirical distribution is estimated through a wild bootstrap procedure in finite samples. Using the estimation and tests, our empirical study on Fama-French 25 and 100 portfolios provides both visualisation and quantitative evidence advocating the time-varying FF five factor models. In addition, we first provide evidence that the pricing error (α) in the conditional asset pricing model are of opposite signs at the tail quantiles and passes through zero at some point around the center of the distribution of the conditional returns. This finding leads us to the conclusion that the FF five-factor model is suitable for the medium portfolios but not for the under-performing and over-performing ones, underscoring the fact that looking at just the conditional mean can sweep a lot of interesting economic relationships under the carpet.

We also note that our proposed model can be applied to a broad range of settings, where data exhibits time-varying, nonlinear and cross-sectional dependency features. For example, apart from the financial applications, our model is also applicable to macroeconomics and health economics data. Under the mean setting, Dong et al. (2021) consider economies of scale for commercial banks in the U.S., and Casas et al. (2021) examine the time-varying income elasticity of health care expenditure in the OECD and the Eurozone with a similar model to ours. Our model can be adopted to contemplate their studies by exploring the quantile structure and distributional effects.

The chapter proceeds as follows. Section 3.2 describes the model and our proposed estimation method, Section 3.3 discusses the model assumptions and establish the asymptotic properties for the estimates. Section 3.4 constructs a specification test for the time-varying betas, and provide a wild bootstrap method to estimate the p-value in finite samples. Section 3.5 re-examines the Fama-French five-factor models and provides distributional evidence in favour of time-varying alphas and betas. Section 3.6 examines the finite sample properties of the estimator through Monte Carlo simulation. Section 3.7 concludes the chapter. Finally, Section 3.8 provides the additional empirical results and the main proofs of the theorems, while the technical details are provided in the Appendix Section S.2.

3.2 Methodology

In this section, we first provide the model setup, and then introduce our proposed estimation approach. Before proceeding, we introduce some notations. Let $\|\cdot\|$ denote the Euclidean norm for vectors and the spectral norm for matrices, that is $\|a\| := \sqrt{a'a}$ and $\|A\| := \sup_{a\neq 0}\|Aa\|/\|a\|$ for a column vector a and a matrix A. Let I_p denote the p-dimensional identity matrix, whose dimension varies according to the subscript. Let $\varrho_i(A)$ denote the i-th largest eigenvalue of matrix A, and specifically, we denote the largest and smallest eigenvalues by $\varrho_{\max}$ and $\varrho_{\min}$ respectively. Let $1\{\cdot\}$ denote the indicator function. Let $\otimes$ denote the Kronecker product. We denote by $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$ for scalars a, b.

3.2.1 Model

Let $\{(y_{it}, x_{it}', v_{it}), i = 1, ..., N,\ t = 1, ..., T\}$ be observations at cross-section i and time t. We denote the supports of y_{it}, x_{it} and v_{it} by $\mathcal{Y} \subseteq \mathbb{R}$, $\mathcal{X} \subseteq \mathbb{R}^p$, and $\mathcal{V} \subseteq \mathbb{R}$ respectively.[8] For any random variables Y and X, Let $G(y|x) := \mathbb{P}\{Y \leq y|X = x\}$ be the conditional distribution function of Y given $X = x$, and we define the corresponding conditional quantile function as $Q_Y(\tau|x) := \inf\{y : G(y|x) \geq \tau\}$ for $\tau \in \mathcal{U} \subset (0, 1)$. Then, we formally introduce the heterogeneous and time-varying regression quantile model with interactive fixed effects:

$$Q_{y_{it}}(\tau|x_{it}, v_t, f_t(\tau), \lambda_i(\tau)) = x_{it}'\beta_i(\tau, v_{it}) + f_t'(\tau)\lambda_i(\tau), \quad i = 1, ..., N,\ t = 1, ..., T, \tag{3.1}$$

where, for a given quantile level $\tau \in (0, 1)$, $\beta_i(\tau, v)$ is an unknown deterministic function of observable variable v, $F(\tau) := (f_1(\tau), ..., f_T(\tau))'$ and $\Lambda(\tau) := (\lambda_1(\tau), ..., \lambda_N(\tau))'$ are the unobservable factor and loadings, whose supports are give by $\mathcal{F} \subseteq \mathbb{R}^r$ and $\mathcal{A} \subseteq \mathbb{R}^r$, respectively. Equivalently, we have

$$y_{it} = x_{it}'\beta_i(\tau, v_{it}) + f_t'(\tau)\lambda_i(\tau) + \epsilon_{it}(\tau), \tag{3.2}$$

[8]The supports $\mathcal{V}$ can be extended to low-dimensional vector case. However, for the simplicity of exposition, we consider the scalar case in this chapter.

where $\mathbb{P}(\epsilon_{it}(\tau) \leq 0 | x_{it}, v_{it}, f_t, \lambda_i) = \tau$ almost surely.

It is well-known in the literature on factor models that $F(\tau)$ and $\Lambda(\tau)$ cannot be identified separately without imposing normalisation constraints. Without the loss of generality, we choose the following constraints (Bai and Li, 2012; Chen et al., 2021):

$$F(\tau)'F(\tau)/T = I_r, \text{ and } \Lambda(\tau)'\Lambda(\tau)/N \text{ is a diagonal matrix,} \tag{3.3}$$

such that the factor and factor loadings are identified up to a column sign change.

3.2.2 Estimation Method: Iterative Quantile Regression

To estimate the unknown function $\beta_i(\tau, \cdot)$, the sieve estimation method is adopted in this chapter, considering its advantage in converting a nonparametric function into an approximate parametric form that is easy to interpret and estimate. To accommodate most of the regression functions including linear parametric forms, we suppose that for each given $\tau \in (0, 1)$, all elements of $\beta_i(\tau, \cdot)$ belong to a Hilbert space $L^2(\mathcal{R}, \pi(w)) := \{h(\cdot) | \int_{\mathcal{R}} h^2(w)\pi(w)dw < \infty\}$, where $\pi(\cdot)$ is a known probability weight function. The use of the density $\pi(\cdot)$ makes the Hilbert space sufficiently large to include almost all kinds of functions, such as powers and polynomials. Here, we allow the support $\mathcal{R}$ to be either bounded or unbounded. When $\mathcal{R}$ is a compact interval on $\mathbb{R}$, conventional orthonormal sequence such as Fourier series or polynomial sequence can be used. When $\mathcal{R}$ is unbounded, such as $\mathcal{R} = (-\infty, \infty)$, with $\pi(w) = \exp(-w^2)$, the sequence of physicists' Hermite polynomials is an orthonormal basis in $L^2(\mathcal{R}, \pi(w))$.

We briefly introduce some properties of the Hilbert space $L^2(\mathcal{R}, \pi(w))$ and the associated orthonormal systems. Define inner product $\langle g, h\rangle := \int_{\mathcal{R}} g(w)h(w)\pi(w)dw$ for $g, h \in L^2(\mathcal{R}, \pi(w))$ that induces norm $\|h\|_{L^2} := \sqrt{\langle h, h\rangle}$. For all $g \in L^2(\mathcal{R}, \pi(w))$, we have an infinite orthogonal series expansion $g(w) = \sum_{j=0}^{\infty} c_j h_j(w)$, where $\{h_j(\cdot), j = 0, 1, ...\}$ is an orthonormal sequence in $L^2(\mathcal{R}, \pi(w))$, and $c_j = \langle g, h_j\rangle$. By the Parseval equality, $\|g\|_{L^2} = \sum_{j=0}^{\infty} c_j^2$. Throughout the study, for any vector of functions $G(\cdot) := (g_1(\cdot), ..., g_k(\cdot))'$, its norm is defined as $\|G\|_{L^2} := \big\{ \sum_{j=0=1}^{k} \|g_j\|_{L^2}^2 \big\}^{1/2}$.

For the sake of simplicity, for each $i \in \{1, ..., N\}$, we truncate all elements of $\beta_i(\tau, \cdot)$ using the same m-dimensional sieve basis, that is

$$\beta_i(\tau, v) = [H_m(v)' \otimes I_p] b_{m,i}(\tau) + \Delta_{\beta_i},$$

where $H_m(v) := (h_0(v), ..., h_{m-1}(v))'$ is an $m \times 1$ vector of orthogonal sieve basis functions evaluated at v, $b_{m,i}(\tau)$ is the sieve coefficient defined conformably, and Δ_{β_i} is the truncation error.

For notational simplicity, we suppress the dependence of coefficients on quantile $\tau \in \mathcal{U}$, unless necessary, in the rest of the chapter.

For model (3.1), we estimate the unknown coefficients by solving the objective function

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(y_{it} - [H_m(v_{it}) \otimes x_{it}]' b_{m,i} - f_t' \lambda_i), \tag{3.4}$$

where $\rho_\tau(u) = (\tau - 1\{u < 0\})u$ for $u \in \mathcal{R}$, through an iterative estimation procedure as follows:

(i) Initiate $\widehat{f}_t^{(0)}$.

(ii) For $k \geq 1$, Given $\widehat{f}_t^{(k-1)}$, for each $i = 1, ..., N$, estimate $\widehat{\beta}_i^{(k)}(v)$ and $\widetilde{\lambda}_i^{(k)}$ simultaneously by

$$\left(\widehat{b}_{m,i}^{(k)}, \widetilde{\lambda}_i^{(k)}\right) = \arg\min_{b_i, \lambda_i} \frac{1}{T} \sum_{t=1}^{T} \rho_\tau \left(y_{it} - [H_m(v_{it}) \otimes x_{it}]' b_i - \widehat{f}_t^{(k-1)\prime} \lambda_i\right),$$

and compute $\widehat{\beta}_{m,i}^{(k)}(v) = [H_m(v)' \otimes I_p] \widehat{b}_{m,i}^{(k)}$.

(iii) Given $\widehat{\beta}_{m,i}^{(k)}(v)$ and $\widetilde{\lambda}_i^{(k)}$, for each $t = 1, ..., T$, estimate $\widetilde{f}_t^{(k)}$ by

$$\widetilde{f}_t^{(k)} = \arg\min_{f_t} \frac{1}{N} \sum_{i=1}^{N} \rho_\tau \left(y_{it} - x_{it}' \widehat{\beta}_{m,i}^{(k)}(v_{it}) - f_t' \widetilde{\lambda}_i^{(k)}\right),$$

(iv) Repeat steps (ii)-(iii) until convergence, and then normalise the converged factor and loadings according to constraint (3.3).[9] Denote the final estimators by $\widehat{\beta}_{m,i}$, $\widehat{F}$ and $\widehat{\Lambda}$.

There are a number of possible ways to determine the initial factor estimators. For example, we can first treat β_i as time-invariant and use the proposed approach of Ando and Bai (2020) to obtain the initial $\widehat{f}_t^{(0)}$. Alternatively, we may first run the quantile sieve regression for β_i and then obtain $\widehat{f}_t^{(0)}$ through principal component analysis on the residuals. Whether such iterative method guarantees finding the global minimum remains an open question which is hard to address. However, all of simulation examples illustrates that (i) our proposed algorithm always converges to the true parameters regardless of the choice of the initial estimators; (ii) the number of iterations varies depending on the choice of the initial estimators.

3.3 Assumptions and Asymptotic Theory

In this section, we first introduce the assumptions and then present asymptotic properties of the estimators of regression coefficients at any given quantile τ in a compact subset $\mathcal{U} \subset (0, 1)$. In what follows, the subscript 0 is used to denote the true value of an estimator. For example, $\beta_{0,i}(\cdot)$, $b_{0m,i}$, $f_{0,t}$ and $\lambda_{0,i}$ represent the true values for $\beta_i(\cdot)$, $b_{m,i}$, f_t and λ_i, respectively.

3.3.1 Assumptions

Assumption 3.3.1. *For all $i = 1, ..., N$ and $t = 1, ..., T$,*

(i) There exists $\delta > 0$ such that $\mathbb{E}\left[\|x_{it}\|^{4+\delta}\right] < \infty$, $\mathbb{E}\left[\|f_{0,t}\|^{4+\delta}\right] < \infty$, $\mathbb{E}[\|\lambda_{0,i}\|^{4+\delta}] < \infty$ and $\mathbb{E}\left[|h_j(v_{it})|^{4+\delta}\right] < \infty$ uniformly in j.

(ii) The eigenvalues of $\mathbb{E}\left[(H_m(v_{it})H_m(v_{it})') \otimes (x_{it}x_{it}')\right]$ and $\mathbb{E}[f_{0,t}f_{0,t}']$ are bounded away from 0 and ∞ uniformly in m.

[2]Let $\widetilde{F}^{(k)} := (\widetilde{f}_1^{(k)}, ..., \widetilde{f}_T^{(k)})'$, and $\widetilde{\Lambda}^{(k)} := (\widetilde{\lambda}_1^{(k)}, ..., \widetilde{\lambda}_N^{(k)})'$. Define $\widetilde{\Sigma}_F := \widetilde{F}^{(k)\prime}\widetilde{F}^{(k)}/T$, and $\widetilde{\Sigma}_\Lambda := \widetilde{\Lambda}^{(k)\prime}\widetilde{\Lambda}^{(k)}/N$. We find the SVD matrices U, S, V such that $\widetilde{\Sigma}_F^{\frac{1}{2}}\widetilde{\Sigma}_\Lambda\widetilde{\Sigma}_F^{\frac{1}{2}} = USV'$. Then, the normalized factor and factor loadings are defined by $\widehat{F}^{(k)} := \widetilde{F}^{(k)}\widetilde{\Sigma}_F^{-\frac{1}{2}}U$ and $\widehat{\Lambda}^{(k)} := \widetilde{\Lambda}^{(k)}\widetilde{\Sigma}_\Lambda^{\frac{1}{2}}U$.

Assumption 3.3.1 imposes some standard moment conditions used in quantile regression literature. Comparing to the existing literature of quantile factor models, see Ando and Bai (2020) and Chen et al. (2021), we do not restrict the supports of x_{it}, $\lambda_{0,i}$ and $f_{0,t}$ to be compact.

Assumption 3.3.2. *Define the filtration* $\mathcal{F}_t := \sigma(\epsilon_{s-1}, X_s, V_s, F_{0,s}, \Lambda_0, s \leq t+1)$*, where* $\epsilon_t := (\epsilon_{1t}, ..., \epsilon_{Nt})'$*. We assume*

(i) $\{\epsilon_{it}, i = 1, ..., N\}$ *are mutually independent conditional on* $\mathcal{F}_{t-1}$ *for each* $t \geq 1$.

(ii) $\psi(\epsilon_t)$ *forms a martingale difference sequence, that is* $\mathbb{E}[\psi(\epsilon_t)|\mathcal{F}_{t-1}] = 0$ *almost surely, where* $\psi(u) \equiv \psi_\tau(u) := 1(u < 0) - \tau$ *is the subgradient of the quantile loss function* $\rho_\tau(\cdot)$.

(iii) The conditional density function of ϵ_{it} *given* $\mathcal{F}_{t-1}$*, denoted as* $g_{it}(\cdot)$*, is continuous and uniformly bounded away from 0 almost surely, and* $g_{it}(0) = \tau$ *almost surely. In addition,* $\sup_{i,t} g_{it}(u) \leq g(u)$*, where* $g(u)$ *belongs to the test function space* $\mathcal{S}$ *defined in Appendix 3.8.3. .*

Assumption 3.3.2 allows for the heteroscedastic and serially-correlated errors, which relaxed the conditional independence assumption in the literature of quantile factor models (Ando and Bai, 2020; Chen et al., 2021).Condition 3.3.2.(ii) automatically holds if $\{\epsilon_t\}$ are independently distributed. Condition 3.3.2.(iii) is a technical requirement in related calculations.

Assumption 3.3.3. *For any fixed* $\tau \in \mathcal{U}$*, let* $\ell_{it} := [x'_{it}, v_{it}, f_{0,t}(\tau)']'$ *and* $\ell_i := [\ell'_{i1}, ..., \ell'_{iT}]'$*. We assume that*

(i) For each $i = 1, ..., N$*, the random sequence* $\{\ell_{it} : t \geq 1\}$ *is a strictly stationary and* α*-mixing process with mixing coefficient* $a_i(u)$ *and* $u > 0$*. Furthermore, there exists a positive coefficient function* $a(u)$ *such that* $\sup_i a_i(u) \leq a(u)$ *and* $\sum_{t \neq s}^{T} a(|t-s|)^{\delta/(4+\delta)} = O(T)$*, where* $\delta > 0$ *is specified in Assumption 3.3.1.*

(ii) For any cross groups i *and* j *with* $i \neq j$*, the random sequence* $\{(\ell_{it}, \ell_{jt}) : t \geq 1\}$ *is also an* α*-mixing process with mixing coefficient* $a_{ij}(u)$ *such that* $\sum_{i \neq j}^{N} a_{ij}(0)^{\delta/(4+\delta)} = O(N)$ *and* $\sum_{i \neq j}^{N} \sum_{t \neq s}^{T} a_{ij}(|t-s|)^{\delta/(4+\delta)} = O(NT)$.

Assumption 3.3.3 imposes weak restriction on the correlation among the observed regressors and unobserved common factors. Using the notation of "α-mixing" for panel data (e.g., Jiang et al., 2021b), the assumption captures both the temporal and cross-sectional dependence exhibited in large panels in a concise manner. Alternatively, one can assume the high-order moment conditions employed by Bai (2009). In addition, since $h_j(\cdot)$ is a Borel measurable function, $h_j(v_{it})$ is also a stationary and α-mixing process whose mixing coefficient satisfies the same property as $\{v_{it}\}$.

Assumption 3.3.4. *Let $\beta_{i(l)}(v)$ denote the l^{th} element of $\beta_i(v)$, and define $b_{i(l),k} := \langle \beta_{i,(l)}, H_k \rangle$. Then, we assume*

(i) $\max_{1\leq l\leq p}\{\sum_{k=m}^{\infty} b_{i(l),k}^2\}^{1/2} = O(m^{-\mu})$ *for a positive constant μ.*

(ii) $\sup_{v\in\mathcal{V}} \int_{\mathcal{X}} g_{xv}(x,v)/\pi(v)x'xdx < \infty$, *where $g_{xv}(x,v)$ is the joint density function of x_{it} and v_{it}.*

(iii) $Tm^{-2\mu} \to 0$.

Assumption 3.3.4 is similar to Assumption 2.3 of Dong et al. (2021). Assumption 3.3.4.(i) is similar to Assumption 3 of Newey (1997). Assumption 3.3.4.(ii) generalises Assumption 2.3.(b) of Dong et al. (2021) in the sense that variables x_{it} and v_{it} are allowed to be correlated. When x_{it} and v_{it} are independent, Assumption 2.3.(b) of Dong et al. (2021) implies our assumption. Assumption 3.3.4.(iii) imposes restrictions on the truncation parameters.

Assumption 3.3.5. *For the initial factor estimator, we assume a linear expansion*

$$\widehat{f}_t^{(0)} - f_{0,t} = \frac{1}{N}\sum_{i=1}^{N}\xi_{it} + o_P\left(\frac{1}{\sqrt{N}}\right), \quad t = 1, ..., T,$$

where $\xi_{it} \equiv \xi(\epsilon_{it})$ is an $r \times 1$ vector such that $\mathbb{E}\left[\xi_{it}|\mathcal{F}_{t-1}\right] = 0$.

Assumption 3.3.5 assumes that there exists a consistent initial estimator for the unknown factors, which admits a linear expression. This condition can be verified under many classical cases. For example, when there exists no covariate or simple linear covariates, Chen et al.

(2021) and Ando and Bai (2020) both show that the estimator of unobservable factors admits the above asymptotic linear expansion. As such, we may first treat β_i as time-invariant and use the proposed approach of Ando and Bai (2020) to obtain the initial estimator $f_t^{(0)}$.

3.3.2 Asymptotic Results

In this section, we establish the main asymptotic properties of $\widehat{\beta}_{m,i}$, $\widehat{\lambda}_i$ and $\widehat{f}_t$. The key underlying difficulties in establishing the asymptotic theories arise from the non-smooth check function $\rho_\tau(\cdot)$ and the incidental-parameters problem due to the unobserved factor structure. To this end, we adopt the generalised function approach to dealing with the non-smooth loss function.

The concept of generalised functions was first introduced in physics and then formalized by mathematicians by defining them as functional acting on tempered test function space $\mathcal{S}$. We introduce the preliminaries about the generalised functions in Section 3.8.3, while we refer the readers to Gel'fand and Shilov (1964) for additional mathematical details. We note that a couple of research papers have applied this approximation approach to study the asymptotic properties of M-estimators (e.g., Phillips, 1995; Dong et al., 2023b). In addition, this approximation technique is recently considered in QR literature, known as convolutional-type smoothed QR (Fernandes et al., 2021). From then on, to the best of our knowledge, there have been no attempts for using generalised functions to deal with nonlinear panel QR.

Since our proof strategy is substantially different from Ando and Bai (2020) and Chen et al. (2021), we first briefly sketch the main underlying ideas. We first approximate the non-smooth loss function $\rho_\tau(\cdot)$ with the corresponding generalised sequence $\{\rho_{h,\tau}(\cdot), h = 1, 2, ...\}$ defined as

$$\rho_{h,\tau}(u) := \int_{-\infty}^{+\infty} \rho_\tau(x)\phi_h(x-u)dx, \quad \phi_h(u) := \sqrt{\frac{h}{\pi}}\exp(-hx^2),\ h = 1, 2, ... \tag{3.5}$$

As $\rho_{h,\tau}(\cdot)$ is infinitely-differentiable, we are able to construct an approximate quadratic-formed loss function from the second-order Taylor expansion of $\rho_{h,\tau}(\cdot)$. Then, we show that the difference between the minimiser of the original QR loss function L in (3.4) and the approximated one is negligible, so is that in the limit. As such, our attention is reduced to analysing the quadratic-formed loss function, whose minimiser has an explicit expression. We note that, as the estimates are obtained recursively, we prove the claims through induction. Below, we provide the rate of convergence and asymptotic distribution of $\widehat{\beta}_{m,i}$, $\widehat{\lambda}_i$ and $\widehat{f}_t$, whose proofs are given in Section 3.8.2.

Theorem 3.3.1. *Under Assumptions 3.3.1-3.3.5, for each $i = 1, ..., N$ and $t = 1, ..., T$,*

(i) $\|\widehat{\beta}_{m,i} - \beta_{0,i}\|_{L^2} = O_P(m^{1/2}(T \wedge N)^{-1/2})$,

(ii) $\|\widehat{\lambda}_i - \lambda_{0,i}\| = O_P((T \wedge N)^{-1/2})$,

(iii) $\|\widehat{f}_t - f_{0,t}\| = O_P((T \wedge N)^{-1/2})$.

Similar to the other semiparametric regression, all orders of the estimators consist of two parts, the dominant part coming from the estimation bias and another part arising from the truncation residuals, where the later one is shown to be $O_P(m^{1/2-\mu})$ in (3.24), which is negligible under Assumption 3.3.4.(iii). A similar rate of the truncation residuals is also derived in example Newey (1997, Theorem 1) and Dong et al. (2021, Lemma 2.1) for example.

Theorem 3.3.2. *Let $z_{it} := \left[H_m(v_{it})' \otimes x'_{it}, f'_{0,t}\right]$, and S_β and S_λ be the selection matrices for $b_{m,i}$ and λ_i from $\theta_i := (b'_{m,i}, \lambda'_i)'$, that is $b_{m,i} = S_\beta \theta_i$ and $\lambda_i = S_\lambda \theta_i$. Suppose $T/N \to \kappa < \infty$ as $N, T \to \infty$. Then, under Assumptions 3.3.1-3.3.5, for $\tau \in \mathcal{U}$, and a given $i \in \{1, ..., N\}$, we have*

$$\sqrt{\frac{T}{m}}\left(\widehat{\beta}_{m,i}(v) - \beta_{0,i}(v)\right) \xrightarrow{d} N\left(0, \Sigma^{\beta}_{i,\tau}(v)\right), \tag{3.6}$$

$$\sqrt{T}\left(\widehat{\lambda}_i - \lambda_{0,i}\right) \xrightarrow{d} N\left(0, \Sigma^{\lambda}_{i,\tau}\right), \tag{3.7}$$

where the positive-definite asymptotic covariance matrices are given by

$$\Sigma^{\beta}_{i,\tau}(v) := \lim_{m\to\infty} \frac{\tau(1-\tau)}{m}[H_m(v)' \otimes I_p] S_\beta \left(\Gamma^Z_{i,\tau,m}\right)^{-1} \sigma^Z_{i,\tau,m} \left(\Gamma^Z_{i,\tau,m}\right)^{-1} S'_\beta [H_m(v) \otimes I_p],$$

$$\Sigma_{i,\tau}^{\lambda} := \lim_{m\to\infty} \tau(1-\tau)S_\lambda\left(\Gamma_{i,\tau,m}^{Z}\right)^{-1}\sigma_{i,\tau,m}^{Z}\left(\Gamma_{i,\tau,m}^{Z}\right)^{-1}S_\lambda',$$

where $\Gamma_{i,\tau,m}^{Z} := \mathrm{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[z_{it}z_{it}']$, *and* $\sigma_{i,\tau,m}^{Z} := \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbb{E}\left[z_{it}z_{it}'\right]$.

Furthermore, for a given $t \in \{1,...,T\}$, *we have*

$$\sqrt{N}\left(\widehat{f}_t - f_{0,t}\right) \xrightarrow{d} N\left(0, \Sigma_{t,\tau}^{f}\right). \tag{3.8}$$

where $\Sigma_{t,\tau}^{f} := \tau(1-\tau)\left(\Gamma_{t,\tau}^{\Lambda}\right)^{-1}\sigma_{\tau}^{\Lambda}\left(\Gamma_{t,\tau}^{\Lambda}\right)^{-1}$ *with* $\Gamma_{t,\tau}^{\Lambda} := \mathrm{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} g_{it}(0)\mathbb{E}[\lambda_{0,i}\lambda_{0,i}']$, *and* $\sigma_{\tau}^{\Lambda} := \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathbb{E}\left[\lambda_{0,i}\lambda_{0,i}'\right]$.

The results in Theorem 3.3.2 derive the asymptotic distribution of our proposed estimators. In order to perform inference, an estimator of the asymptotic covariance matrices, $\Sigma_{i,\tau}^{\beta}(v)$, $\Sigma_{i,\tau}^{\lambda}$ and $\Sigma_{t,\tau}^{f}$, are to be proposed. Given their expressions, we construct the corresponding consistent estimators via their empirical counterparts

$$\widehat{\Sigma}_{i,\tau}^{\beta}(v) := \frac{\tau(1-\tau)}{m}[H_m(v)'\otimes I_p]S_\beta\left(\widehat{\Gamma}_{i,\tau,m}^{Z}\right)^{-1}\widehat{\sigma}_{i,\tau,m}^{Z}\left(\widehat{\Gamma}_{i,\tau,m}^{Z}\right)^{-1}S_\beta'[H_m(v)\otimes I_p], \tag{3.9}$$

$$\widehat{\Sigma}_{i,\tau}^{\lambda} := \tau(1-\tau)S_\lambda\left(\widehat{\Gamma}_{i,\tau,m}^{Z}\right)^{-1}\widehat{\sigma}_{i,\tau,m}^{Z}\left(\widehat{\Gamma}_{i,\tau,m}^{Z}\right)^{-1}S_\lambda', \tag{3.10}$$

$$\widehat{\Sigma}_{t,\tau}^{f} := \tau(1-\tau)\left(\widehat{\Gamma}_{t,\tau}^{\Lambda}\right)^{-1}\widehat{\sigma}_{\tau}^{\Lambda}\left(\widehat{\Gamma}_{t,\tau}^{\Lambda}\right)^{-1}. \tag{3.11}$$

As σ_{τ}^{Λ} and $\sigma_{i,\tau,m}^{Z}$ depend only on X_i, F_0 and Λ_0, in light of the consistency results in Theorem 3.3.1, it can be shown that a consistent estimator of these matrices are

$$\widehat{\sigma}_{\tau}^{\Lambda} := N^{-1}\sum_{i=1}^{N}\widehat{\lambda}_i\widehat{\lambda}_i', \text{ and } \widehat{\sigma}_{i,\tau,m}^{Z} := T^{-1}\sum_{t=1}^{T}\widehat{z}_{it}\widehat{z}_{it}',$$

where $\widehat{z}_{it} := [w_{it}', \widehat{f}_t']'$.

On the other hand, the consistent estimator of $\Gamma_{t,\tau}^{\Lambda}$ and $\Gamma_{i,\tau,m}^{Z}$ requires more detailed analyse as they both depend on the conditional density $g_{it}(0)$, which is unknown in practice. Under the assumptions of Theorem 3.3.2, the time and cross-sectional specific density $g_{it}(0)$ can be consistently estimated using the information from the neighboring quantiles of τ. One possible approach is making use of the derivative of the inverse function, see Sagner (2020). This estimation approach relies on the numerical approximation of the derivative function in a small neighborhood of τ, but the estimation performs poorly in practice not only due to the numerical approximation error, but also the potential quantile crossing issue. Alternatively, to obtain a more accurate estimator, a stronger assumption on $\{\epsilon_{it}\}$ needs to be imposed. Apart from the conditional cross-sectional independence assumption, we further assume that $\{\epsilon_{it}\}$ are identically distributed across i. Recall that $g_{it}(0) \equiv g_t(0) = \mathbb{E}[\ddot{\rho}_\tau(\epsilon_{it})|\mathcal{F}_{t-1}]$, where $\ddot{\rho}_\tau(\cdot)$ is the second-derivative of $\rho_\tau(\cdot)$. As such, since $\ddot{\rho}_\tau(\cdot)$ can be approximated by the corresponding generalised sequence $\ddot{\rho}_{h,\tau}(\cdot)$, we simply define the estimator of $g_t(0)$ by the corresponding sample averages as

$$\widehat{g}_t(0) := \frac{1}{N}\sum_{i=1}^{N} \ddot{\rho}_{h,\tau}(\widehat{\epsilon}_{it}),$$

where $\widehat{\epsilon}_{it} := y_{it} - x_{it}'\widehat{\beta}_{m,i}(v_{it}) - \widehat{f}_t'\widehat{\lambda}_i$, and $\ddot{\rho}_{h,\tau}(\cdot)$ is the second derivative of $\rho_{h,\tau}(\cdot)$. Correspondingly, we define the estimators

$$\widehat{\Gamma}^{\Lambda}_{t,\tau} := N^{-1}\sum_{i=1}^{N}\widehat{g}_t(0)\widehat{\lambda}_i\widehat{\lambda}_i' \text{ , and } \widehat{\Gamma}^{Z}_{i,\tau,m} := T^{-1}\sum_{t=1}^{T}\widehat{g}_t(0)\widehat{z}_{it}\widehat{z}_{it}'.$$

We then have the following corollary.

Corollary 3.3.1. *Let the conditions of Theorem 3.3.2 hold. Furthermore, suppose $\{\epsilon_{it}\}$ are identically distributed across i and t. Then, $\widehat{\Sigma}^{\beta}_{i,\tau}(v) \xrightarrow{p} \Sigma^{\beta}_{i,\tau}(v)$, $\widehat{\Sigma}^{\lambda}_{i,\tau} \xrightarrow{p} \Sigma^{\lambda}_{i,\tau}$, and $\widehat{\Sigma}^{f}_{t,\tau} \xrightarrow{p} \Sigma^{f}_{t,\tau}$ as $N, T \to \infty$.*

3.4 Constancy Test

3.4.1 Hypothesis and Test Statistics

For any fixed $i = 1, ..., N$ and $\tau \in \mathcal{U}$, we consider the following null and alternative hypotheses:

$$H_0 : \beta_i(\tau, v) = \beta_{0,i}(\tau) \text{ a.s. for some } \beta_{0,i} \in \mathbb{R}^p \text{ v.s. } H_A : \beta_i(\tau, v) \neq \beta_{0,i}(\tau) \text{ a.s.}$$

In this chapter, we consider a Wald-type statistic, proposed by Su and Hoshino (2016), that requires only consistent estimation of the unrestricted model. The test statistics is specified as

$$\widehat{L}_T = \sum_{t=1}^{T} \left\| \widehat{\beta}_{m,i}(v_{it}) - \overline{\widehat{\beta}}_{m,i} \right\|^2 W(v_{it})$$

where $\overline{\widehat{\beta}}_{m,i} = T^{-1} \sum_{t=1}^{T} \widehat{\beta}_{m,i}(v_{it})$, and $W(\cdot)$ is a uniformly bounded non-negative weight function defined on the support $\mathcal{V}$ of v_{it}. In addition, the weight matrix is allowed to be taken as a constant, i.e., $W(v_{it}) \equiv 1$, which leads to an unweighted version of the test.

3.4.2 A Bootstrap Testing Procedure

It is generally acknowledged that nonparametric tests using asymptotic normal null distributions perform poorly in finite samples. Therefore, we suggest adopting a bootstrap method to obtain the finite-sample approximation to distribution of the test statistic under the null. Motivated by Feng et al. (2011), we propose the following wild bootstrap testing procedure:

(i) Obtain the unrestricted estimates $\widehat{\beta}_{m,i}(v)$, $\widehat{\lambda}_i$, $\widehat{f}_t$, and calculate the unrestricted residuals $\widehat{\epsilon}_{it} = Y_{it} - x_{it}'\widehat{\beta}_{m,i}(v_{it}) - \widehat{f}_t'\widehat{\lambda}_i$, as well as the test statistic $\widehat{L}_T$.

(ii) For $i = 1, ..., N$ and $t = 1, ..., T$, generate the wild bootstrap residuals $\epsilon_{it}^* := |\widehat{\epsilon}_{it}|\eta_i$, where $\{\eta_i\}_{i=1}^{N}$ are independently generated from a two-point mass distribution with probabilities $1 - \tau$ and τ at $\eta = 2(1 - \tau)$ and $\eta = -2\tau$, respectively.

(iii) Generate the bootstrapped sample under H_0. Generate y_{it}^* via $y_{it}^* := x_{it}'\overline{\widehat{\beta}}_{it} + \widehat{f}_t'\widehat{\lambda}_i + \epsilon_{it}^*$. Then, regard $\{y_{it}^*, x_{it}, v_{it}, i = 1, ..., N, t = 1..., T\}$ as the bootstrap sample.

(iv) Refit the model on the bootstrapping sample, and compute the bootstrap test statistic $\widehat{L}_{T(b)}^*$ in the same way as $\widehat{L}_T$.

(v) Repeat steps (ii)-(iv) B times to obtain B bootstrapping test statistics $\{\widehat{L}_{T(b)}^*\}_{b=1}^{B}$. Calculate the bootstrapping p-value via $p^* := B^{-1} \sum_{b=1}^{B} 1\{\widehat{L}_{T(b)}^* \geq \widehat{L}_T\}$. Reject the null hypothesis H_0 if p^* is smaller than the prescribed nominal level of significance.

3.4.3 Practical Implementations

In this section, we address some practical issues that may arise with the estimation and testing procedure.

- Selection of optimal truncation parameter m and the number of factors r:

  As the number of factors r and the dimension of the sieve basis m are unknown, we need to consider the selection of factors jointly with the selection of the basis dimensions. In the simulation and empirical study, we consider a data-driven approach. Specifically, we adopt the "leave-one(-cross-section)-out" cross-validation method to estimate the optimal r and m in step (i), that is,

$$(\widehat{r}, \widehat{m}) := \underset{r,m\geq 1}{\arg\min} \sum_{j=1} \left\{ \frac{1}{NT} \sum_{i\neq j}^{N} \sum_{t=1}^{T} \rho_\tau (y_{it} - [H_m(v_{it}) \otimes x_{it}]' \widehat{b}_{m,i} - \widehat{f}_t' \widehat{\lambda}_i) \right\}. \quad (3.12)$$

  In addition, to ease the computational burden, we do not re-estimate r and m when refitting the model on bootstrap samples in step (iv).

- Hypothesis test on a subvector of $\beta_i(\tau, \cdot)$:

  We may also test whether a subvector of $\beta_i(\tau, \cdot)$ is a constant. The test statistic is similar to $\widehat{L}_T$ above, but only evaluated at the tested subvector. Additionally, in the bootstrap procedure, when generate bootstrap response y_{it}^*, we allow the untested beta components to vary but only restrict the tested subvector under the null. The rest of the testing procedure remains.

- Cross-sectionally invariant x_t and v_t case:

  In many applications, the observed covariates x and v are cross-sectionally invariant. For example, in the asset pricing literature, risk factors are typically constructed from the market portfolios, which are invariant across individual stocks. Additionally, in the conditional asset pricing literature, the conditional variable v can be either firm-level characteristics, such as book-to-market equity (Gagliardini et al., 2016) or macro-level factors, such as term spread and default spread. Nevertheless, the estimation and inference for the heterogeneous β_i discussed before are still applicable in this case.

3.5 Empirical Analysis

3.5.1 Data and Model

There is evidence that there exist quantile-dependent common factor structures in the financial market (Ma and Pohlman, 2008; Galvao et al., 2018; Ando and Bai, 2020). In addition, it has been widely identified that the factor structure of the expected excess return is business-cycle dependent (e.g., McQueen and Roley, 1993; Jagannathan and Wang, 1996; Gagliardini et al., 2016; Celebi and Hönig, 2019). In this chapter, we seek to explore the economic-cycle-dependent quantile common factor structures in the excess stock returns.

We re-examine the Fama and French (2015) (FF hereafter) five-factor model using the proposed functional-coefficient quantile panel data model. When attention is restricted to the conditional expected return, there are abundant empirical evidences suggesting that betas in the pricing model are generally time-varying. For example, Ferson and Harvey (1999) and Wang (2003) provide strong evidences that betas in the FF model vary with lagged macroeconomic and financial instruments using the times series data, and Cai et al. (2022) estimate and test for conditional betas with panel data, which achieve substantial efficiency gain comparing to those performed on time series data. In contrast with existing studies, to the best of our knowledge, this is the first attempt to examine the conditional asset pricing model on the conditional return distribution. Similar to Cai et al. (2022), we estimate the quantile-dependent conditional betas using the panel data to allow for possible cross-sectional dependence among portfolio returns via the unobserved factor structure. Finally, the nonparametric Wald statistic is applied to testing (i) the constancy of betas and (ii) the significance of pricing errors alphas.

We collect monthly returns of the Fama-French 25 and 100 portfolios which are sorted by size (market equity, "ME") and book-to-market ratio ("BM") for the period from January 1983 to July 2022. The data of portfolio returns, the monthly risk-free rate, and the FF five factors are all downloaded from the Kenneth French Data Library.

The following heterogeneous vary-coefficient panel data model is adopted for estimating the conditional quantile FF five-factor model:

$$Q_{R_{it}}(\tau|x_t, v_{t-1}, f_t) = \alpha_i(\tau, v_{t-1}) + x_t'\beta_i(\tau, v_{t-1}) + f_t(\tau)'\lambda_i(\tau)$$

where $x_t := [\text{SMB}_t, \text{HML}_t, \text{RMW}_t, \text{CMA}_t, \text{Mkt.RF}_t]'$ are the FF five-factors. Following Ferson and Harvey (1999) and Cai et al. (2022), various lagged conditional variables v_t are considered, including the spread between a ten-year and one-year Treasury bond yields (r10m1), the spread between the returns of the three-month and the one-month Treasury bill (r3m1), and the spread between Moody's BAA and AAA corporate bond yield (BmA). Given the domain of v_t, Hermite polynomials are employed as the sieve basis in the study, whose dimensions are chosen to be 5 according to the cross-validation method given as (3.12).

3.5.2 Results

For presentation conciseness, we report the results for FF 25 portfolios with conditional variable v_t being r10m1. The estimation and testing results conducted on FF 100 portfolios and the other choices of conditional variables (r3m1, BmA) are provided in Section 3.8.1 for completion; similar conclusions are drawn using different dataset and conditional variables.

Figure 3.1 plots the estimates of functional regression coefficients for FF 25 on the range of observed v_t from -0.5 to 3.5. The top panel reports the α-estimates, which reflect the performance of underlying assets compared to the market, or known as the pricing error. The regression coefficients are almost flat over v_t, except for the range close to the boundaries. This suggests that when the market condition changes significantly, the pricing error varies. In addition, for portfolios that perfectly track the market (i.e., at 0.5th conditional quantile), the estimated intercept α_i's fluctuates around zero, while they show opposite signs at opposite ends of the distribution of conditional returns. Specifically, for the underperforming portfolios at 0.1th quantile, the estimated α_i's are negative, ranging between -6 and 0. On the other hand, those for the outperforming portfolios are between 0 and 6 at 0.9th quantile. For the β-estimates for the FF five-factors, some similar patterns are spotted. Firstly, β-estimates are heterogeneous across portfolios with dispersion roughly the same across quantiles. In addition, the estimates show the non-constant feature at all quantiles, with large variation at the boundary of v_t.

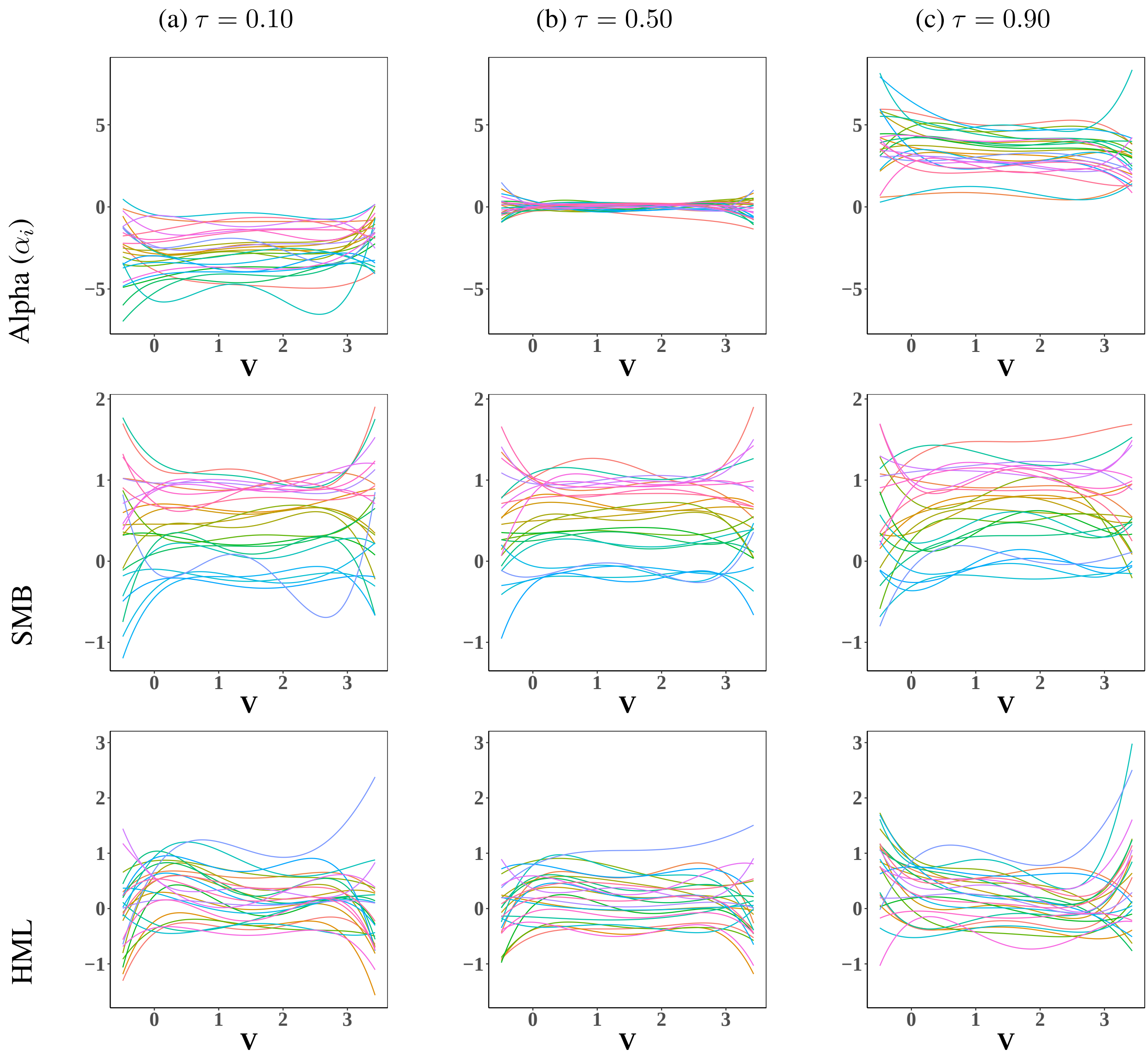
To demonstrate the potential relationship between regression coefficient and business cycle, Figure 3.2 plots the estimates of functional regression coefficients for FF 25 over time, and with the recession period, identified by the National Bureau of Economic Research (NBER) highlighted as vertical shaded areas. For all plots, we witness large positive and negative strays from the time-invariant estimates around recession periods, which suggest that the pricing error and risk factors follow the macroeconomic cycle. Similar patterns are spotted on FF 100, whose figures are omitted for the sake of conciseness.

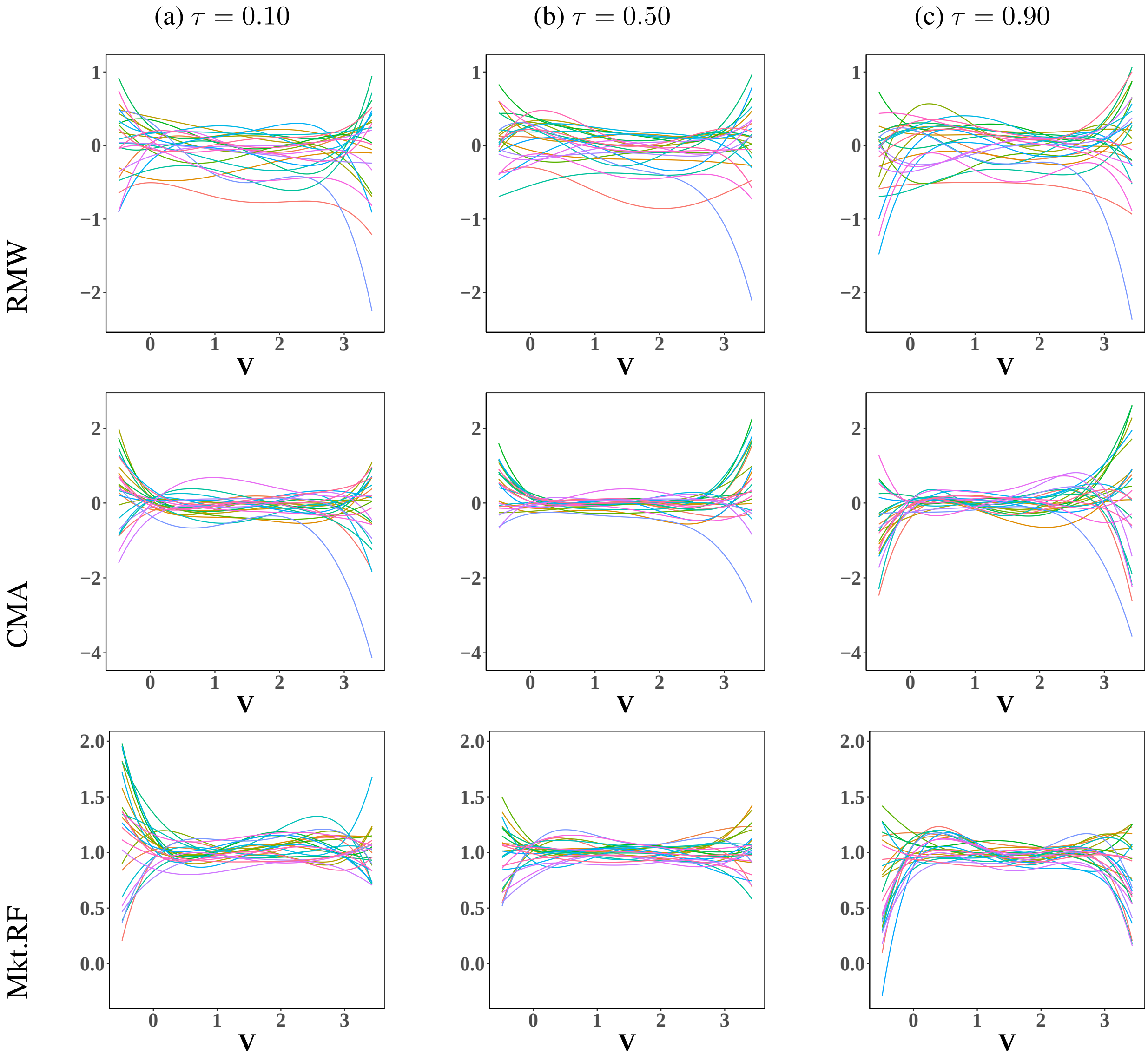
Next, we test for the joint constancy of alphas and betas. The null hypothesis is given as, for each given $i = 1, ..., N$,

$$H_0 : \alpha_i(v) = \alpha_i, \beta_i(v) = \beta_i, \text{ for certain } (\alpha_i, \beta_i) \in \mathbb{R} \times \mathbb{R}^p,$$

and the alternative hypothesis is specified as its complement. Table 3.1 reports the Bootstrapping p-values for testing H_0 based on 200 Bootstrapping samples. The notations ME1 through ME5 and BM1 through BM5 stand for the FF (increasing) quantiles on size and book-to-market ratio. The numbers in the row of BM1 and the column of ME5, for example, reports the p-values for testing the constancy for the portfolio of stocks in the smallest quantile of book-to-market quantile and the highest quantile of size. From Table 3.1, one can find the rejection ratios at 5% significance level for r10m1 are 0.28 (7/25, which means 7 out of 25 portfolios reject the null hypothesis), 0.80 (20/25), 0.84 (21/25), 0.76 (19/25) , 0.44 (11/25) at quantiles $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, respectively, suggesting that the majority of the portfolios relatively tracking the market (with returns between 25% and 75% quantile) are in favour of the time-varying alphas and betas. These results in line with the constancy testing results of Cai et al. (2022), who consider a similar model under the conditional mean framework, using FF

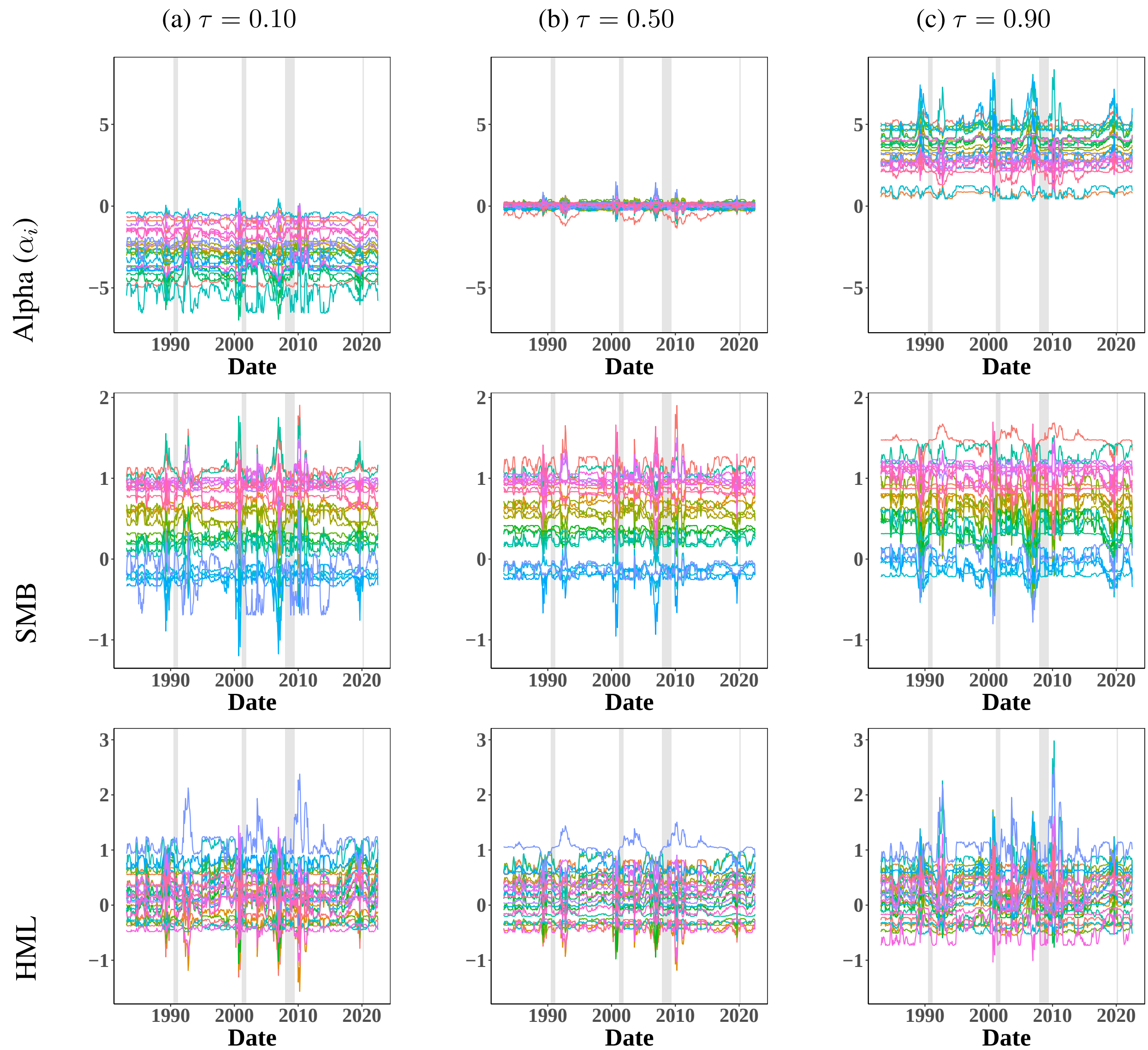
Figure 3.1 Functional-Coefficient Regression Coefficients Estimates

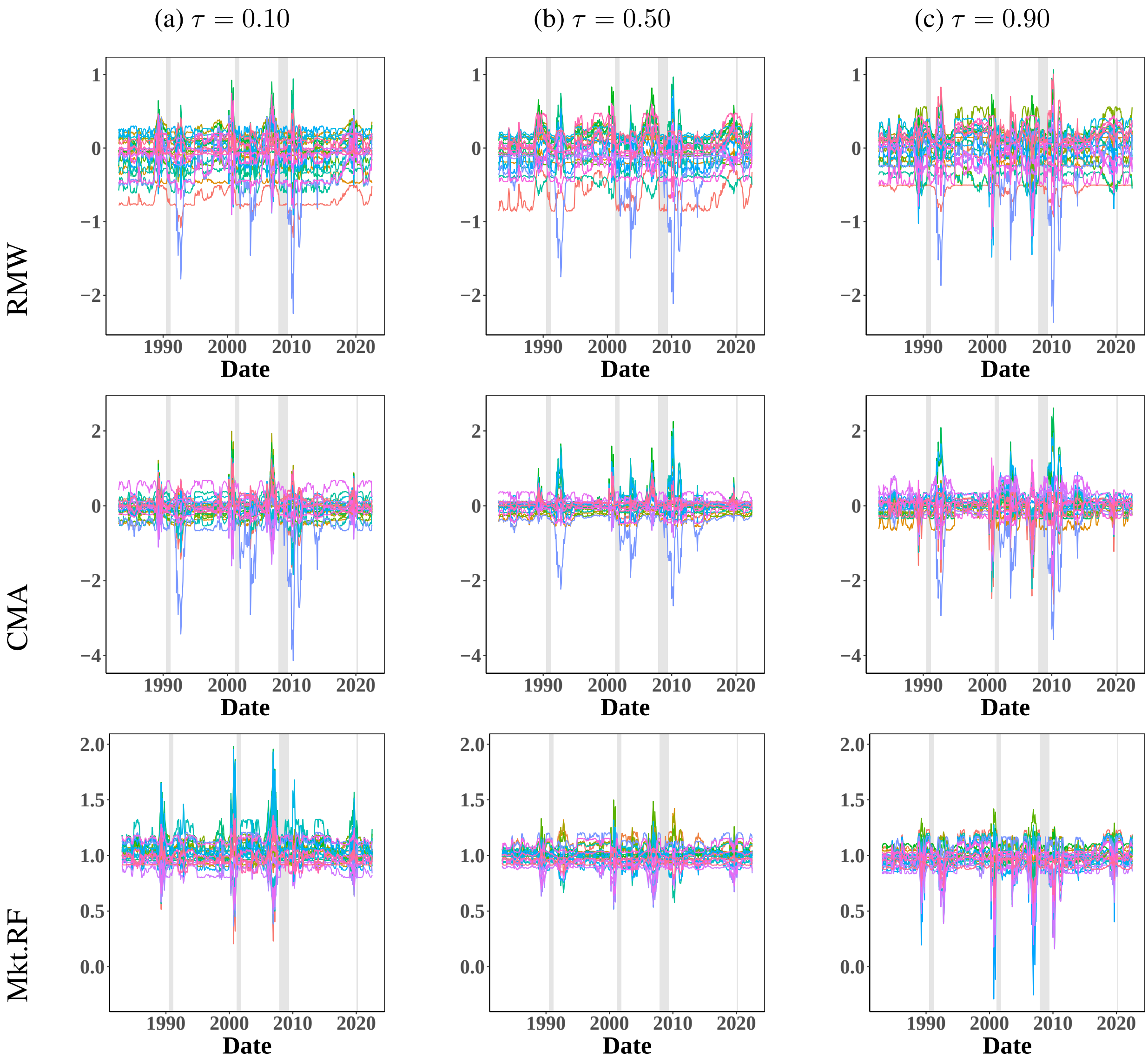




Notes: Panels (a)-(c) correspond to estimates at quantiles $\tau = 0.1, 0.5, 0.9$, respectively. From the top panel to the bottom are the coefficients estimates for the intercept and five FF variables. In each sub-figure, the rainbow-colored lines represent estimates for the 25 FF portfolios. Figures are plotted over the range of observed v_t between -0.5 and 3.5.

Figure 3.2 Functional-Coefficient Regression Coefficients Estimates over Time





Notes: Panels (a)-(c) correspond to estimates at quantiles $\tau = 0.1, 0.5, 0.9$, respectively. From the top panel to the bottom are the coefficients estimates for the intercept and five FF variables. In each sub-figure, the rainbow-colored lines represent estimates for the 25 FF portfolios. Vertical shaded grey areas label recessions determined by the National Bureau of Economic Research (NBER). The recessions start at the peak of a business cycle and ends at the trough.

25 portfolios from July 1963 to July 2018 and r10m1 as the conditional variable. We note that the low rejection rates at the tail quantiles $\tau = 0.1$ and 0.9 are likely due to the insufficient number of cross-sections ($N = 25$). We should be cautious about viewing it as a supporting evidence for the time-invariant betas. The rejection rates at the tails are comparable to those in the central when the same tests are performed on FF 100 portfolios, documented in the Section 3.8.1.

In order to avoid any potential structural break during the recession, we also conduct the constancy test over four blocks of periods free from recession that cover different turmoil periods in financial markets. As shown in Figure 3.3, regardless of the periods we considered, the majority of portfolios whose excess returns are between the 25th and 75th quantiles provide evidence that supports the time-varying alphas and betas.

Next, we test the significance of the pricing error alpha. If FF five-factor is sufficient to characterise the return distribution, one would expect that the estimated alphas are insignificant. Therefore, the following test is considered:

$$H_0 : \alpha_i(\tau, v_{it}) = 0 \text{ v.s. } H_A : \alpha_i(\tau, v_{it}) \neq 0$$

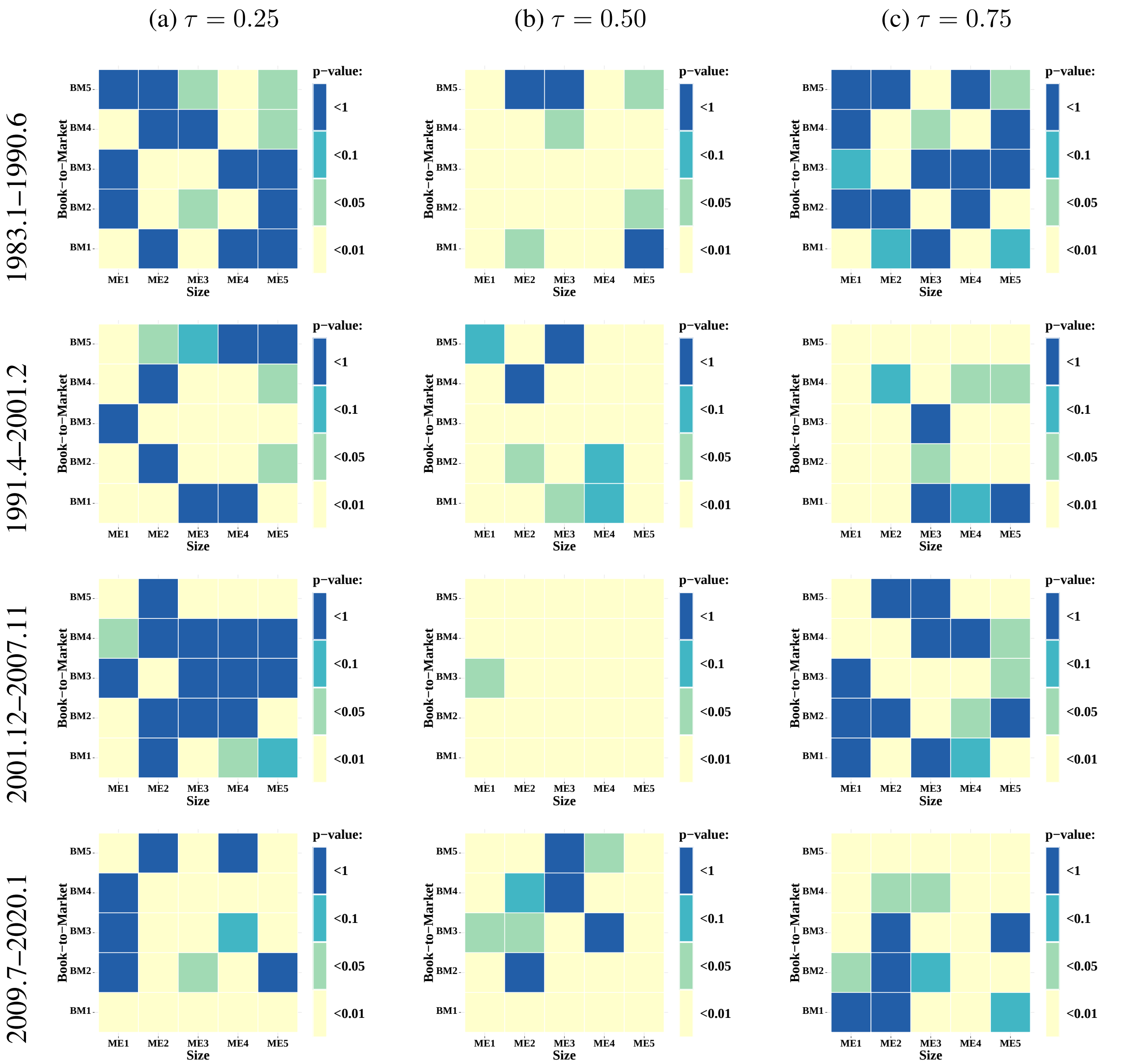
Here, restrictions are only on the alphas, but the betas are allowed to vary with the conditioning variables. Table 3.2 reports the estimated p-values based on 200 bootstrap samples. The rejection ratios at 5% significance level for r10m1 are 0.80 (20/25), 0.76 (19/25), 0.04 (1/25), 0.88 (22/25) , 0.88 (22/25) at quantiles $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, respectively. These results strongly reject the FF five-factor model at the quantile other than the medium, while providing strong evidence in favour of the model at the central quantile. In addition, as alphas are estimated to be opposite signs at the opposite end of the distribution (see Figure 3.2), the large negative and positive impacts in the tails can cancel each other out, which may explain the insignificant alpha from the conditional mean estimates.

Table 3.1 Bootstrap p-values for $H_0 : \alpha_i(\tau, v_t) = \alpha_i, \beta_i(\tau, v) = \beta_i(\tau)$ on FF25 Portfolios

	ME1	ME2	ME3	ME4	ME5
$\tau = 0.10$					
BM1	0.260	0.070	0.000	0.765	0.485
BM2	0.560	0.005	0.280	0.735	0.115
BM3	0.890	0.000	0.460	0.005	0.330
BM4	0.285	0.360	0.655	0.045	0.610
BM5	0.875	0.890	0.720	0.000	0.000
$\tau = 0.25$					
BM1	0.000	0.000	0.000	0.000	0.115
BM2	0.005	0.000	0.020	0.000	0.015
BM3	0.460	0.000	0.045	0.000	0.000
BM4	0.000	0.280	0.000	0.000	0.020
BM5	0.610	0.015	0.010	0.135	0.000
$\tau = 0.50$					
BM1	0.000	0.000	0.000	0.490	0.000
BM2	0.000	0.000	0.005	0.000	0.000
BM3	0.395	0.000	0.025	0.000	0.000
BM4	0.000	0.645	0.000	0.000	0.000
BM5	0.145	0.000	0.000	0.000	0.000
$\tau = 0.75$					
BM1	0.080	0.020	0.000	0.010	0.000
BM2	0.600	0.000	0.015	0.000	0.000
BM3	0.110	0.150	0.000	0.000	0.000
BM4	0.000	0.000	0.000	0.060	0.155
BM5	0.015	0.000	0.000	0.000	0.000
$\tau = 0.90$					
BM1	0.785	0.140	0.315	0.000	0.010
BM2	0.580	0.520	0.780	0.030	0.920
BM3	0.780	0.015	0.050	0.080	0.000
BM4	0.310	0.02	0.000	0.905	0.005
BM5	0.530	0.465	0.010	0.000	0.000

Notes: The constancy test is perform on the whole dataset, from January 1983 to July 2022, with p-values estimated based on 200 bootstrap repetitions. The conditional variable v_t is chosen to be r10m1.

Figure 3.3 Bootstrap p-values for $H_0 : \alpha_i(\tau, v_t) = \alpha_i, \beta_i(\tau, v) = \beta_i(\tau)$ on FF25 Portfolios During Non-Recession Period



Notes: Panels (a)-(c) correspond to estimates at quantiles $\tau = 0.25, 0.5, 0.75$, respectively. From the top panel to the bottom are heat maps of Bootstrap p-values for constancy test based on 200 bootstrap samples, during four non-recession periods between January 1983 and January 2020.

Table 3.2 Bootstrap p-values for $H_0 : \alpha_i(\tau, v_t) = 0$ on FF25 portfolios

	ME1	ME2	ME3	ME4	ME5
$\tau = 0.10$					
BM1	0.000	0.000	0.000	0.000	0.205
BM2	0.000	0.000	0.000	0.000	0.000
BM3	0.000	0.000	0.000	0.000	0.000
BM4	0.590	0.005	0.000	0.000	0.000
BM5	0.060	0.105	0.000	0.000	0.175
$\tau = 0.25$					
BM1	0.000	0.000	0.000	0.000	1.000
BM2	0.000	0.490	0.000	0.000	0.000
BM3	0.000	0.135	0.000	0.000	0.000
BM4	0.460	0.270	0.000	0.000	0.000
BM5	0.010	0.220	0.005	0.000	0.000
$\tau = 0.50$					
BM1	0.325	1.000	1.000	1.000	1.000
BM2	1.000	1.000	1.000	1.000	1.000
BM3	1.000	1.000	1.000	1.000	1.000
BM4	1.000	1.000	1.000	1.000	1.000
BM5	1.000	1.000	1.000	1.000	0.000
$\tau = 0.75$					
BM1	0.000	0.000	0.000	0.000	0.615
BM2	0.000	0.000	0.000	0.000	0.000
BM3	0.000	0.000	0.000	0.000	0.000
BM4	0.010	0.080	0.000	0.000	0.005
BM5	0.000	0.230	0.000	0.000	0.000
$\tau = 0.90$					
BM1	0.000	0.000	0.000	0.000	0.000
BM2	0.000	0.000	0.000	0.000	0.000
BM3	0.000	0.000	0.000	0.000	0.000
BM4	0.000	0.000	0.000	0.000	0.000
BM5	0.395	0.160	0.000	0.000	0.000

Notes: The constancy test is performed on the whole FF25 dataset, from January 1983 to July 2022, with p-values estimated based on 200 bootstrap repetitions. The conditional variable v_t is chosen to be r10m1.

3.6 Simulation Analysis

In this section, we compare the in-sample and out-of-sample estimation accuracy of our proposed estimators using the simulated data.

3.6.1 Data Generating Process

To illustrate the finite sample performance of our proposed estimation method, we consider the following data generating process.[10] For $i = 1, ..., N$, and $t = 1, ..., T$,

$$y_{it} = x_{it}\beta_i(u_{it}, v_t) + f_t(u_{it})'\lambda_i(u_{it}), \quad \beta_i(u, v) = 1 + \frac{i}{N} + 0.1u - v + 0.5v^2,$$
$$x_{it} = \exp(0.25z_{it}), \quad f_t(u_{it}) = (f_{t1}\sqrt{u_{it}}, f_{t2}\sqrt{u_{it}})', \quad \lambda_i(u_{it}) = (\lambda_{i1} + 0.5\sqrt{u_{it}}, \lambda_{i2} + 0.1\sqrt{u_{it}})',$$

where $\{u_{it}\}$ are i.i.d. $U(0, 1)$, $\{f_{t1}, f_{t2}\}$ are generated via the SVD of a $T \times T$ random matrix whose entries are i.i.d. $N(2, 1)$, and $\{\lambda_{i1}, \lambda_{i2}\}$ are generated from i.i.d. $\mathrm{U}(2, 3)$. We consider the following two constructions of z_{it} and v_t:

(i) z_{it} is generated as i.i.d. $N(0, 1)$, and v_t is generated as i.i.d. N(0,1).

(ii) z_{it} is generated as i.i.d. $N(0, 1)$, and v_t is generated as AR(1) with the AR coefficient being 0.5.

3.6.2 In-Sample Performance of Estimators

We approximate the non-linear beta-coefficients by the the Hermite polynomials. We compare the in-sample estimation accuracy of our method with the following two benchmark models: (i) a heterogeneous quantile regression with interactive fixed effects (Ando and Bai, 2020), that is model (3.1) with $\beta_i(\tau, v_{it}) \equiv \beta_i(\tau)$, and (ii) a heterogeneous functional-coefficient quantile regression for time series (Kim, 2007), that is model (3.1) with the number of factor $r = 0$.

[3] We can show that $Q_{y_{it}}(\tau|x_{it}, v_t, f_t(\tau), \lambda_i(\tau)) = x_{it}\beta_i(\tau, v_t) + f_t(\tau)'\lambda_i(\tau)$.

As the QR estimators are L_1-norm based, we consider two error measurements that are based on L_1-norm. Specifically, we evaluate the estimation accuracy of parameter $\beta_i(\tau, z_t)$ by the mean absolute error (MAE)

$$\text{MAE}(\tau) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| \widehat{\beta}_i(\tau, v_t) - \beta_i(\tau, v_t) \right|.$$

To measure the fitted accuracy of the model, we employ the τ-average weighted absolute error (TAWE)

$$\text{TAWE}(\tau) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau \big(y_{it} - \widehat{Q}_{y_{it}}(\tau | x_{it}, v_t, f_t(\tau), \lambda_i(\tau)) \big),$$

where $\widehat{Q}_{y_{it}}(\tau | x_{it}, v_t, \widehat{f}_t(\tau), \widehat{\lambda}_i(\tau))$ is the fitted conditional quantile of y_{it}.

Tables 3.3-3.4 report the averages of the error measurements MAE and TAWE over 200 replications with the standard deviation reported in parentheses. Compared to the benchmarks, our proposed estimation method produces smaller error in the estimations of both the regression coefficient and conditional quantile of y_{it} in both settings, regardless of the sample size. Comparing the estimation results of our method with those of Kim (2007), we confirm that including the latent factor structure significantly reduces the "omitted-variable bias" in panel data.

Table 3.3 Mean and Standard Deviation of the MAE(τ)

			Scenario 1			Scenario 2		
N	T	τ	Model (3.1)	Ando and Bai (2020)	Kim (2007)	Model (3.1)	Ando and Bai (2020)	Kim (2007)
25	50	0.1	0.078 (0.008)	0.805 (0.158)	0.202 (0.023)	0.077 (0.007)	0.960 (0.276)	0.202 (0.022)
		0.5	0.039 (0.004)	0.832 (0.162)	0.141 (0.010)	0.038 (0.004)	0.951 (0.274)	0.141 (0.009)
		0.9	0.082 (0.011)	0.831 (0.157)	0.127 (0.019)	0.083 (0.009)	0.930 (0.255)	0.128 (0.018)
50	75	0.1	0.073 (0.008)	0.809 (0.114)	0.132 (0.014)	0.073 (0.006)	0.955 (0.249)	0.133 (0.012)
		0.5	0.022 (0.002)	0.818 (0.127)	0.097 (0.004)	0.022 (0.002)	0.970 (0.222)	0.097 (0.005)
		0.9	0.075 (0.007)	0.819 (0.123)	0.096 (0.007)	0.074 (0.008)	0.940 (0.220)	0.097 (0.008)
75	100	0.1	0.073 (0.009)	0.814 (0.096)	0.101 (0.010)	0.072 (0.011)	0.975 (0.188)	0.100 (0.012)
		0.5	0.015 (0.001)	0.817 (0.114)	0.076 (0.003)	0.015 (0.001)	0.958 (0.179)	0.076 (0.003)
		0.9	0.073 (0.009)	0.816 (0.105)	0.086 (0.005)	0.073 (0.007)	0.973 (0.195)	0.086 (0.004)

Notes: We report the average of the MAE over 200 repetitions with standard deviation in the parentheses.

Table 3.4 Mean and Standard Deviation of the TAWE(τ)

			Scenario 1			Scenario 2		
N	T	τ	Model (3.1)	Ando and Bai (2020)	Kim (2007)	Model (3.1)	Ando and Bai (2020)	Kim (2007)
25	50	0.1	0.013 (0.001)	0.085 (0.026)	0.233 (0.033)	0.013 (0.001)	0.096 (0.038)	0.234 (0.027)
		0.5	0.039 (0.001)	0.273 (0.077)	0.136 (0.004)	0.039 (0.001)	0.302 (0.108)	0.136 (0.003)
		0.9	0.016 (0.001)	0.089 (0.027)	0.036 (0.042)	0.016 (0.001)	0.097 (0.036)	0.033 (0.034)
50	75	0.1	0.012 (0.001)	0.107 (0.021)	0.197 (0.032)	0.012 (0.000)	0.126 (0.036)	0.199 (0.025)
		0.5	0.035 (0.001)	0.295 (0.059)	0.114 (0.002)	0.035 (0.001)	0.348 (0.088)	0.114 (0.002)
		0.9	0.015 (0.000)	0.109 (0.021)	0.025 (0.019)	0.015 (0.000)	0.123 (0.034)	0.026 (0.022)
75	100	0.1	0.010 (0.001)	0.117 (0.016)	0.174 (0.026)	0.010 (0.001)	0.137 (0.031)	0.172 (0.033)
		0.5	0.031 (0.000)	0.306 (0.052)	0.101 (0.001)	0.031 (0.000)	0.355 (0.076)	0.101 (0.001)
		0.9	0.013 (0.000)	0.118 (0.019)	0.023 (0.022)	0.013 (0.000)	0.136 (0.033)	0.022 (0.017)

Notes: We report the average of the TAWE over 200 repetitions with standard deviation in the parentheses.

3.6.3 Out-of-Sample Performance of Estimators

In this section, we report the comparisons of out-of-sample performance among the three models in the previous section. We fixed the forecasting horizons $h = 1,\ 3,\ 5,\ 10$. The comparisons are based on the following two evaluation criteria: (i) forecasting mean absolute error (FMAE) and (ii) forecasting τ-average weighted absolute error (FTAWE). Specifically, for a given forecasting horizon h, we define

$$\text{FMAE}_h(\tau) = \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{\beta}_i(\tau, \upsilon_{T+h}) - \beta_i(\tau, \upsilon_{T+h})\right|,$$

$$\text{FTAWE}_h(\tau) = \frac{1}{N}\sum_{i=1}^{N}\rho_\tau\left(y_{i,T+h} - \widehat{Q}_{y_{i,T+h}}(\tau | x_{i,T+h}, \upsilon_{T+h}, \widehat{f}_{T+h}(\tau), \widehat{\lambda}_i(\tau))\right),$$

where $\widehat{Q}_{y_{i,T+h}}(\tau | x_{i,T+h}, \upsilon_{T+h}, \widehat{f}_{T+h}(\tau), \widehat{\lambda}_i(\tau))$ is the fitted conditional quantile of $y_{i,T+h}$. Since f_{T+h} is unobservable, we obtain the h-step-ahead forecast for each component of f_{T+h} by fitting a univariate ARMA(p, q) model, where p, q are automatically determined by the AIC criterion.

Tables 3.5-3.6 report the averages of the error measurements FMAE and FTAWE over 200 replications with the standard deviation reported in parentheses. As the forecasting horizon increases, the forecasting accuracy slightly decreases. However, it is worth mentioning that, at any given forecasting horizon h, our proposed model greatly outperformed the benchmark models at all quantiles.

Table 3.5 Mean and Standard Deviation of the FMAE$_h(\tau)$

N	T	τ	Scenario 1 Model (3.1)	Scenario 1 Ando and Bai (2020)	Scenario 1 Kim (2007)	Scenario 2 Model (3.1)	Scenario 2 Ando and Bai (2020)	Scenario 2 Kim (2007)
$h=1$								
25	50	0.1	0.078 (0.023)	0.749 (0.729)	0.201 (0.104)	0.080 (0.030)	1.030 (1.210)	0.212 (0.107)
		0.5	0.042 (0.036)	0.816 (1.198)	0.149 (0.085)	0.043 (0.031)	1.047 (1.268)	0.152 (0.085)
		0.9	0.085 (0.038)	0.761 (0.913)	0.134 (0.067)	0.088 (0.048)	0.992 (1.395)	0.140 (0.092)
50	75	0.1	0.073 (0.011)	0.763 (0.831)	0.135 (0.078)	0.073 (0.009)	0.960 (1.108)	0.137 (0.072)
		0.5	0.023 (0.018)	0.756 (0.934)	0.101 (0.063)	0.023 (0.013)	0.982 (1.102)	0.100 (0.041)
		0.9	0.075 (0.011)	0.867 (0.934)	0.097 (0.028)	0.075 (0.018)	1.005 (1.148)	0.100 (0.041)
75	100	0.1	0.072 (0.009)	0.918 (1.051)	0.108 (0.054)	0.072 (0.012)	0.971 (1.068)	0.098 (0.034)
		0.5	0.014 (0.007)	0.771 (0.906)	0.075 (0.022)	0.015 (0.008)	0.916 (1.180)	0.077 (0.030)
		0.9	0.074 (0.012)	0.812 (1.030)	0.091 (0.038)	0.074 (0.009)	1.108 (1.387)	0.089 (0.022)
$h=3$								
25	50	0.1	0.080 (0.028)	0.848 (0.826)	0.212 (0.109)	0.085 (0.050)	1.099 (1.500)	0.235 (0.171)
		0.5	0.040 (0.027)	0.697 (0.779)	0.147 (0.080)	0.045 (0.049)	0.977 (1.184)	0.156 (0.124)
		0.9	0.085 (0.036)	0.863 (1.097)	0.137 (0.080)	0.090 (0.050)	1.026 (1.227)	0.140 (0.089)
50	75	0.1	0.073 (0.009)	0.783 (0.833)	0.134 (0.061)	0.073 (0.009)	0.925 (1.003)	0.136 (0.063)
		0.5	0.023 (0.015)	0.866 (0.971)	0.100 (0.040)	0.025 (0.017)	1.139 (1.686)	0.107 (0.054)
		0.9	0.077 (0.017)	0.960 (1.155)	0.104 (0.040)	0.075 (0.016)	1.005 (1.257)	0.098 (0.033)
75	100	0.1	0.073 (0.010)	0.777 (0.938)	0.102 (0.053)	0.072 (0.010)	1.007 (1.112)	0.101 (0.045)
		0.5	0.017 (0.010)	0.997 (1.223)	0.083 (0.036)	0.015 (0.009)	0.919 (1.194)	0.077 (0.030)
		0.9	0.074 (0.012)	0.807 (1.088)	0.087 (0.021)	0.073 (0.008)	0.996 (1.188)	0.086 (0.015)
$h=5$								
25	50	0.1	0.081 (0.050)	0.796 (1.007)	0.219 (0.178)	0.083 (0.046)	1.019 (1.186)	0.231 (0.199)
		0.5	0.042 (0.027)	1.001 (1.131)	0.149 (0.073)	0.043 (0.038)	0.998 (1.395)	0.154 (0.096)
		0.9	0.085 (0.034)	0.792 (0.968)	0.132 (0.064)	0.089 (0.039)	1.022 (1.187)	0.139 (0.076)
50	75	0.1	0.072 (0.010)	0.830 (1.049)	0.135 (0.071)	0.074 (0.015)	1.109 (1.490)	0.154 (0.118)
		0.5	0.022 (0.013)	0.847 (1.029)	0.096 (0.034)	0.023 (0.013)	0.905 (0.970)	0.100 (0.043)
		0.9	0.078 (0.019)	0.899 (1.242)	0.105 (0.047)	0.076 (0.018)	1.033 (1.405)	0.100 (0.040)
75	100	0.1	0.073 (0.013)	0.830 (1.203)	0.104 (0.046)	0.072 (0.012)	0.904 (1.049)	0.105 (0.064)
		0.5	0.016 (0.010)	0.868 (1.165)	0.081 (0.038)	0.015 (0.010)	1.010 (1.300)	0.078 (0.035)
		0.9	0.073 (0.011)	0.989 (1.191)	0.087 (0.019)	0.074 (0.009)	0.994 (1.314)	0.088 (0.021)
$h=10$								
25	50	0.1	0.079 (0.025)	0.836 (0.937)	0.207 (0.107)	0.083 (0.037)	1.030 (1.196)	0.226 (0.156)
		0.5	0.039 (0.023)	0.834 (0.926)	0.144 (0.060)	0.042 (0.033)	1.008 (1.212)	0.151 (0.082)
		0.9	0.085 (0.038)	0.839 (1.006)	0.135 (0.075)	0.090 (0.043)	1.101 (1.380)	0.144 (0.094)
50	75	0.1	0.072 (0.010)	0.860 (1.023)	0.133 (0.066)	0.074 (0.012)	1.021 (1.249)	0.150 (0.097)
		0.5	0.023 (0.013)	0.890 (1.110)	0.099 (0.041)	0.022 (0.013)	0.963 (1.166)	0.100 (0.040)
		0.9	0.076 (0.014)	0.934 (1.163)	0.098 (0.030)	0.075 (0.015)	0.838 (0.919)	0.098 (0.033)
75	100	0.1	0.073 (0.009)	0.855 (1.043)	0.099 (0.040)	0.072 (0.012)	0.909 (0.962)	0.103 (0.050)
		0.5	0.014 (0.005)	0.718 (0.771)	0.073 (0.016)	0.015 (0.008)	0.829 (0.923)	0.077 (0.028)
		0.9	0.073 (0.010)	0.797 (0.953)	0.087 (0.019)	0.073 (0.008)	0.980 (1.251)	0.085 (0.013)

Notes: We report the average of the FMAE over 200 repetitions with standard deviation in the parentheses.

Table 3.6 Mean and Standard Deviation of the $\text{FTAWE}_h(\tau)$

N	T	τ	Scenario 1			Scenario 2		
			Model (3.1)	Ando and Bai (2020)	Kim (2007)	Model (3.1)	Ando and Bai (2020)	Kim (2007)
$h = 1$								
25	50	0.1	0.028 (0.014)	0.336 (0.760)	0.285 (0.064)	0.030 (0.034)	0.604 (1.377)	0.295 (0.063)
		0.5	0.058 (0.014)	0.472 (0.982)	0.158 (0.029)	0.058 (0.013)	0.541 (0.763)	0.160 (0.029)
		0.9	0.029 (0.012)	0.299 (0.802)	0.046 (0.050)	0.030 (0.018)	0.489 (1.426)	0.047 (0.055)
50	75	0.1	0.020 (0.008)	0.241 (0.372)	0.227 (0.052)	0.019 (0.005)	0.418 (1.162)	0.225 (0.042)
		0.5	0.045 (0.008)	0.347 (0.396)	0.126 (0.018)	0.045 (0.007)	0.525 (1.284)	0.128 (0.015)
		0.9	0.020 (0.004)	0.278 (0.438)	0.030 (0.024)	0.021 (0.006)	0.353 (0.961)	0.031 (0.029)
75	100	0.1	0.016 (0.004)	0.313 (0.711)	0.196 (0.036)	0.016 (0.004)	0.306 (0.463)	0.190 (0.041)
		0.5	0.038 (0.005)	0.358 (0.520)	0.109 (0.010)	0.038 (0.005)	0.439 (0.612)	0.109 (0.010)
		0.9	0.017 (0.004)	0.229 (0.325)	0.027 (0.030)	0.017 (0.004)	0.377 (0.861)	0.025 (0.019)
$h = 3$								
25	50	0.1	0.029 (0.016)	0.342 (0.737)	0.282 (0.066)	0.031 (0.031)	0.649 (1.396)	0.295 (0.082)
		0.5	0.058 (0.014)	0.433 (0.924)	0.161 (0.030)	0.059 (0.019)	0.525 (0.766)	0.165 (0.051)
		0.9	0.029 (0.012)	0.334 (0.831)	0.047 (0.056)	0.031 (0.015)	0.504 (1.429)	0.045 (0.046)
50	75	0.1	0.019 (0.006)	0.237 (0.348)	0.225 (0.047)	0.020 (0.006)	0.416 (1.144)	0.228 (0.043)
		0.5	0.045 (0.007)	0.394 (0.433)	0.127 (0.015)	0.045 (0.007)	0.595 (1.388)	0.128 (0.016)
		0.9	0.021 (0.006)	0.303 (0.474)	0.031 (0.025)	0.020 (0.006)	0.368 (0.975)	0.030 (0.024)
75	100	0.1	0.016 (0.004)	0.288 (0.732)	0.192 (0.036)	0.016 (0.004)	0.294 (0.463)	0.191 (0.041)
		0.5	0.038 (0.005)	0.455 (0.621)	0.111 (0.012)	0.038 (0.004)	0.447 (0.617)	0.109 (0.010)
		0.9	0.017 (0.003)	0.226 (0.318)	0.026 (0.024)	0.017 (0.003)	0.365 (0.862)	0.024 (0.019)
$h = 5$								
25	50	0.1	0.081 (0.050)	0.796 (1.007)	0.219 (0.178)	0.083 (0.046)	1.019 (1.186)	0.231 (0.199)
		0.5	0.042 (0.027)	1.001 (1.131)	0.149 (0.073)	0.043 (0.038)	0.998 (1.395)	0.154 (0.096)
		0.9	0.085 (0.034)	0.792 (0.968)	0.132 (0.064)	0.089 (0.039)	1.022 (1.187)	0.139 (0.076)
50	75	0.1	0.072 (0.010)	0.830 (1.049)	0.135 (0.071)	0.074 (0.015)	1.109 (1.490)	0.154 (0.118)
		0.5	0.022 (0.013)	0.847 (1.029)	0.096 (0.034)	0.023 (0.013)	0.905 (0.970)	0.100 (0.043)
		0.9	0.078 (0.019)	0.899 (1.242)	0.105 (0.047)	0.076 (0.018)	1.033 (1.405)	0.100 (0.040)
75	100	0.1	0.073 (0.013)	0.830 (1.203)	0.104 (0.046)	0.072 (0.012)	0.904 (1.049)	0.105 (0.064)
		0.5	0.016 (0.010)	0.868 (1.165)	0.081 (0.038)	0.015 (0.010)	1.010 (1.300)	0.078 (0.035)
		0.9	0.073 (0.011)	0.989 (1.191)	0.087 (0.019)	0.074 (0.009)	0.994 (1.314)	0.088 (0.021)
$h = 10$								
25	50	0.1	0.079 (0.025)	0.836 (0.937)	0.207 (0.107)	0.083 (0.037)	1.030 (1.196)	0.226 (0.156)
		0.5	0.039 (0.023)	0.834 (0.926)	0.144 (0.060)	0.042 (0.033)	1.008 (1.212)	0.151 (0.082)
		0.9	0.085 (0.038)	0.839 (1.006)	0.135 (0.075)	0.090 (0.043)	1.101 (1.380)	0.144 (0.094)
50	75	0.1	0.072 (0.010)	0.860 (1.023)	0.133 (0.066)	0.074 (0.012)	1.021 (1.249)	0.150 (0.097)
		0.5	0.023 (0.013)	0.890 (1.110)	0.099 (0.041)	0.022 (0.013)	0.963 (1.166)	0.100 (0.040)
		0.9	0.076 (0.014)	0.934 (1.163)	0.098 (0.030)	0.075 (0.015)	0.838 (0.919)	0.098 (0.033)
75	100	0.1	0.073 (0.009)	0.855 (1.043)	0.099 (0.040)	0.072 (0.012)	0.909 (0.962)	0.103 (0.050)
		0.5	0.014 (0.005)	0.718 (0.771)	0.073 (0.016)	0.015 (0.008)	0.829 (0.923)	0.077 (0.028)
		0.9	0.073 (0.010)	0.797 (0.953)	0.087 (0.019)	0.073 (0.008)	0.980 (1.251)	0.085 (0.013)

Notes: We report the average of the FTAWE over 200 repetitions with standard deviation in the parentheses.

3.7 Conclusion

In this chapter, we introduce a general framework of a functional-coefficient panel quantile regression model with interactive fixed effects. Our model can capture of the time-varying quantile co-movement of large financial time series. An iterative estimation approach is proposed, and has shown to work well, in terms of both in-sample and out-of-sample accuracy, using extensive simulation examples. The asymptotic properties for the recursive estimators are established through an approximation of the generalised sequences, which greatly facilitate the derivation. We have also constructed a simple specification test for testing the constancy of the model coefficients based on a Wald-type statistic, whose p-value is estimated through a wild bootstrap procedure in finite samples. The proposed test is applied to test the constancy and significance of functional quantile-coefficients in the Fama–French five-factor model.

Using our model, we re-examined the Fama–French five-factor models using the FF25 and FF100 portfolios. The estimation and testing results suggest that the quantile co-movement of stock returns follows the macroeconomic cycle, which provides empirical evidence in favour of the conditional asset pricing model over the unconditional model across different quantile levels. In addition, the pricing error is insignificant at the central quantile, suggesting that the proposed conditional asset pricing model is suitable for the medium excess return, which is comparable to the expected return. On the other hand, the significant pricing error at the tail quantiles suggests additional factors or a more sophisticated model needs to be considered at the tails, as the tail returns are, in general, more volatile. The distinct performance at the central and tail quantiles provides a more comprehensive view of the return dynamics beyond the mean, and can be further used in terms of risk management and portfolio allocation.

3.8 Additional Empirical Evidence and Theoretical Results

In this appendix, Section 3.8.1 provides the additional empirical results. Section 3.8.2 establishes the proofs of theorems in the main text. The necessary preliminary lemmas are stated and proven in Section S.2. Section 3.8.3 introduces the fundamental knowledge of generalised functions.

3.8.1 Additional Empirical Evidence

In this section, we provide the additional testing results on FF25 and FF100 portfolios. Tables 3.7 and 3.8 summarise the proportion of FF25 and FF100 portfolios that rejects the constancy test given different conditional variables, including “r10m1”, “r3m1” and “BmA”, respectively. Regardless of the choice of the conditional variable and underlying portfolios, more than half of the portfolios reject the null hypothesis for quantiles between 0.25 and 0.75. However, the rejection proportion is much lower at the tail quantiles. Cautious is necessary in interpreting this result as the evidence in favour of the constant betas, as the results may not be effective due to the large estimation variance at the tails.

Tables 3.9 and 3.10 report the the proportion of FF25 and FF100 portfolios that rejects the null hypothesis of an insignificant alpha given different conditional variables. Testing results based on FF25 and FF100 portfolios reveal similar patterns. When the underlying assets receive unexpected idiosyncratic shocks (i.e., $\tau \neq 0.5$), the FF five factors are not sufficient to capture the return dynamics in almost all cases, regardless of the choice of the conditional variable. Contradictorily, when the underlying assets do not expose to any idiosyncratic shock (i.e. $\tau = 0.5$), the FF five factors adequately capture the return variation if considering the term spread (either r10m1 or r3m1) as the conditional variable.

Table 3.7 The Proportion of FF25 Portfolios that Rejects $H_0 : \alpha_i(\tau, v_t) = \alpha_i(\tau), \beta_i(\tau, v_t) = \beta_i(\tau)$

τ	significance level 1%	significance level 5%	significance level 10%
v_t: r10m1			
0.10	0.240	0.280	0.320
0.25	0.680	0.920	0.920
0.50	0.800	0.840	0.840
0.75	0.600	0.760	0.840
0.90	0.240	0.440	0.520
v_t: r3m1			
0.10	0.360	0.560	0.720
0.25	0.840	0.840	0.840
0.50	0.560	0.640	0.720
0.75	0.680	0.760	0.920
0.90	0.200	0.360	0.440
v_t: BmA			
0.10	0.080	0.200	0.240
0.25	0.400	0.520	0.600
0.50	0.760	0.800	0.800
0.75	0.360	0.520	0.640
0.90	0.040	0.040	0.040

Notes: The constancy test is performed on the FF25 dataset, from January 1983 to July 2022, with p-values estimated based on 200 bootstrap repetitions.

Table 3.8 The Proportion of FF100 Portfolios that Rejects $H_0 : \alpha_i(\tau, v_t) = \alpha_i(\tau), \beta_i(\tau, v_t) = \beta_i(\tau)$

τ	significance level 1%	significance level 5%	significance level 10%
v_t: r10m1			
0.10	0.323	0.385	0.448
0.25	0.708	0.802	0.812
0.50	0.875	0.917	0.948
0.75	0.635	0.719	0.750
0.90	0.396	0.458	0.469
v_t: r3m1			
0.10	0.333	0.438	0.490
0.25	0.729	0.781	0.802
0.50	0.729	0.750	0.781
0.75	0.667	0.740	0.750
0.90	0.323	0.427	0.469
v_t: BmA			
0.10	0.365	0.490	0.531
0.25	0.615	0.698	0.719
0.50	0.833	0.854	0.854
0.75	0.646	0.719	0.750
0.90	0.312	0.417	0.469

Notes: The constancy test is performed on the FF100 dataset, from January 1983 to July 2022, with p-values estimated based on 200 bootstrap repetitions.

Table 3.9 The Proportion of FF25 Portfolios that Rejects $H_0 : \alpha_i(\tau, v_t) = 0$

τ	significance level 1%	significance level 5%	significance level 10%
v_t: r10m1			
0.10	0.800	0.800	0.840
0.25	0.720	0.760	0.760
0.50	0.040	0.040	0.040
0.75	0.840	0.880	0.920
0.90	0.840	0.880	0.880
v_t: r3m1			
0.10	0.320	0.640	0.720
0.25	0.040	0.080	0.680
0.50	0.000	0.000	0.000
0.75	0.680	0.680	0.680
0.90	0.440	0.520	0.600
v_t: BmA			
0.10	0.800	0.800	0.840
0.25	0.840	0.840	0.840
0.50	0.280	0.280	0.320
0.75	0.800	0.800	0.800
0.90	0.800	0.920	0.920

Notes: The constancy test is performed on the FF25 dataset, from January 1983 to July 2022, with p-values estimated based on 200 bootstrap repetitions.

Table 3.10 The Proportion of FF100 Portfolios that Rejects $H_0 : \alpha_i(\tau, v_t) = 0$

τ	significance level 1%	significance level 5%	significance level 10%
v_t: r10m1			
0.10	0.844	0.865	0.896
0.25	0.750	0.771	0.771
0.50	0.000	0.000	0.000
0.75	0.740	0.771	0.792
0.90	0.438	0.656	0.708
v_t: r3m1			
0.10	0.604	0.875	0.948
0.25	0.719	0.729	0.750
0.50	0.000	0.000	0.000
0.75	0.625	0.740	0.760
0.90	0.146	0.396	0.562
v_t: BmA			
0.10	0.500	0.625	0.698
0.25	0.875	0.896	0.917
0.50	0.427	0.448	0.469
0.75	0.844	0.885	0.927
0.90	0.542	0.667	0.740

Notes: The constancy test is performed on the FF100 dataset, from January 1983 to July 2022, with p-values estimated based on 200 bootstrap repetitions.

3.8.2 Proofs of the Main Results

In what follows, we use a few additional notations. For a function $g : \mathbb{R} \to \mathbb{R}$, we use $\dot{g}$, $\ddot{g}$ and $g^{(k)}$ to denote the first, second, and k-th ($k \geq 3$) derivatives of g, respectively, in generalised function context, which agrees with the conventional second derivative whenever it exists. Let $\psi_\tau(\cdot)$ denote the subgradient of the check function $\rho_\tau(\cdot)$. Though the value of $\psi_\tau(\cdot)$ is not unique at $u = 0$, we allow $\psi_\tau(\cdot)$ to take any of them, while the analysis in the sequel remains unchanged. For notational simplicity, we suppress the dependency of quantile τ in the subscript unless necessary in what follows.

Let C represent an absolute constant, whose value may vary at each appearance. Let $\to$ denote convergence in generalised sequence sense, which contains the convergence of ordinary functions as a special case.

Denote $w_{it} := H_m(v_{it}) \otimes x_{it}$, $z_{it} := [w_{it}', f_{0,t}']'$, and $z_{it}^{(k)} := [w_{it}', \widehat{f}^{(k)\prime}]'$. Let $\beta_{0,mi}(v_{it}) := [H_m(v_{it})' \otimes I_p] b_{0,mi}$, $\theta_i := [b_{m,i}', \lambda_i']'$, and use $\theta_{0,i}$ to denote its true value.

Define the following loss functions

$$L_i(\theta_i|F) := \frac{1}{T}\sum_{t=1}^{T} \rho(y_{it} - w_{it}' b_{m,i} - f_t'\lambda_i), \text{ and } L_t(f_t|\Theta) := \frac{1}{N}\sum_{i=1}^{N} \rho(y_{it} - w_{it}' b_{m,i} - f_t'\lambda_i),$$

where $\theta := [\theta_1', ..., \theta_i']'$. Correspondingly, define the centered version

$$\widetilde{L}_i(\theta_i|F) := L_i(\theta_i|F) - L_i(\theta_{0,i}|F), \text{ and } \widetilde{L}_t(f_t|\theta) := L_t(f_t|\theta) - L_t(f_{0,t}|\theta).$$

It is straightforward that, in the k-th iteration, $(\widehat{b}_{m,i}^{(k)}, \widehat{\lambda}_i^{(k)}) = \text{argmin}_{\theta_i} L_i(\theta_i|\widehat{F}^{(k-1)})$, and $\widehat{f}_t^{(k)} = \text{argmin}_{f_t} L_t(f_t|\widehat{\Theta}^{(k)})$. Moreover, those estimators are also the minimiser of the corresponding centred version.

Define

$$L_{h,i}(\theta_i|F) := \frac{1}{T}\sum_{t=1}^{T} \rho_h(y_{it} - w_{it}' b_{m,i} - f_t'\lambda_i), \quad h = 1, 2, ...,$$

$$L_{h,t}(f_t|\Theta) := \frac{1}{N}\sum_{i=1}^{N}\rho_h(y_{it} - w_{it}'b_{m,i} - f_t'\lambda_i),$$

where $\rho_h(u)$ is a regular sequence of $\rho(u)$, defined as (3.5), and $\Theta := [\theta_1', ..., \theta_N']'$ whose estimator is denoted as $\widehat{\Theta}$. Analogously, we define the centered smoothed loss function $\widetilde{L}_{h,i}(\theta_i|F) := L_{h,i}(\theta_i|F) - L_{h,i}(\theta_{0,i}|F)$, and $\widetilde{L}_{h,t}(f_t|\Theta) := L_{h,t}(f_t|\Theta) - L_{h,t}(f_{0,t}|\Theta)$. Since $\rho_h(\cdot)$ is twice-differentiable, we consider the Taylor expansion for both $\rho_h(y_{it} - z_{it}^{(k)\prime}\theta_i)$ and $\rho_h(y_{it} - z_{it}^{(k)\prime}\theta_{0,i})$ around ϵ_{it} to the second order, which yields

$$\widetilde{L}_{h,i}\big(\theta_i|\widehat{F}^{(k)}\big) = S_{h,i}(\theta_i|\widehat{F}^{(k)}) + O_P\Big(\frac{1}{T}\sum_{t=1}^{T}\big[z_{it}^{(k)\prime}(\theta_i - \theta_{0,i})\big]^3\Big).$$

where

$$S_{h,i}(\theta_i|\widehat{F}^{(k)}) := \frac{1}{T}\sum_{t=1}^{T}\Big\{-\Big[\dot{\rho}_h\big(\epsilon_{it}\big) + \Delta_{it}^{(k)}\ddot{\rho}_h(\epsilon_{it})\big)\Big]z_{it}^{(k)\prime}(\theta_i - \theta_{0,i}) + \frac{1}{2}\ddot{\rho}_h\big(\epsilon_{it}\big)\big[z_{it}^{(k)\prime}(\theta_i - \theta_{0,i})\big]^2\Big\},$$

where $\Delta_{it}^{(k)} := x_{it}'(\beta_{0,i}(v_{it}) - \beta_{0,mi}(v_{it})) + (f_{0,t} - \widehat{f}_t^{(k)})'\lambda_{0,i}$. Similarly, given $\widehat{\Theta}^{(k)} := \{(\widehat{b}_{m,i}^{(k)\prime}, \widehat{\lambda}_i^{(k)\prime})', i = 1, ..., N\}$, Taylor expansion for both $\rho_h\big(y_{it} - w_{it}'\widetilde{b}_{m,i}^{(k)} - f_t'\widetilde{\lambda}_i^{(k)}\big)$ and $\rho_h\big(y_{it} - w_{it}'\widetilde{b}_{m,i}^{(k)} - f_{0,t}'\widetilde{\lambda}_i^{(k)}\big)$ around ϵ_{it}, and yields

$$\widetilde{L}_{h,t}\big(f_t|\widehat{\Theta}^{(k)}\big) = S_{h,t}\big(f_t|\widehat{\Theta}^{(k)}\big) + O_P\Big(\frac{1}{N}\sum_{i=1}^{N}\Big[\widehat{\lambda}_i^{(k)\prime}(f_t - f_{0,t})\Big]^3\Big),$$

where

$$\begin{aligned}
&S_{h,t}\big(f_t|\widehat{\Theta}^{(k)}\big) \\
&:= \frac{1}{N}\sum_{i=1}^{N}\Big\{-\Big[\dot{\rho}_h(\epsilon_{it}) + \Big(x_{it}'\big(\beta_{0,i}(v_{it}) - \widehat{\beta}_{m,i}^{(k)}(v_{it})\big) + f_{0,t}'\big(\lambda_{0,i} - \widehat{\lambda}_i^{(k)}\big)\Big)\ddot{\rho}_h(\epsilon_{it})\Big]\widehat{\lambda}_i^{(k)\prime}(f_t - f_{0,t}) \\
&\quad + \frac{1}{2}\ddot{\rho}_h(\epsilon_{it})\Big[\widehat{\lambda}_i^{(k)\prime}(f_t - f_{0,t})\Big]^2\Big\}.
\end{aligned}$$

Now, let $\widetilde{\theta}_i^{(k)}$ and $\widetilde{f}_t^{(k)}$ be the minimizers of $S_{h,i}(\theta_i|\widehat{F}^{(k-1)})$ and $S_{h,t}(\theta_i|\widehat{\Theta}^{(k)})$, respectively. Lemma 3.8.1 provides the limiting expressions for both $\widetilde{\theta}_i^{(k)}$ and $\widetilde{f}_t^{(k)}$ given the previous estimators. Then, we further show that $\|\widehat{\theta}_i^{(k)} - \widetilde{\theta}_i^{(k)}\|$ and $\|\widehat{f}_t^{(k)} - \widetilde{f}_t^{(k)}\|$ are sufficiently small, which completes the proof of Theorem 3.3.1, and the CLT in Theorem 3.3.2 can be derived accordingly.

Lemma 3.8.1. *Under Assumption 3.3.1-3.3.5, suppose $T/N \to \kappa > 0$ as $N, T \to \infty$ and $h = O(T^{(1+c)})$. Then, given the initial estimator $\widehat{F}^{(0)}$, we have*

$$
\begin{aligned}
&\sqrt{\frac{T}{m}}\Big(\widetilde{\beta}_{m,i}^{(1)}(v) - \beta_{0,i}(v)\Big) \\
&\qquad = [H_m(v)' \otimes I_p]\big(R_{1,i}\big)^{-1} \frac{1}{\sqrt{mT}} \sum_{t=1}^{T} \psi(\epsilon_{it})\Big(w_{it} - K_{1,i} f_{0,t}\Big) + o_P(1), \\
&\sqrt{T}\Big(\widetilde{\lambda}_i^{(1)} - \lambda_{0,i}\Big) = \big(R_{2,i}\big)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi(\epsilon_{it})\Big(f_{0,t} - K_{2,i} w_{it}\Big) + o_P(1),
\end{aligned}
$$

where

$$
\begin{aligned}
K_{1,i} &:= \left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[w_{it} f_{0,t}']\right)\left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[f_{0,t} f_{0,t}']\right)^{-1}, \\
K_{2,i} &:= \left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[f_{0,t} w_{it}']\right)\left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[w_{it} w_{it}']\right)^{-1}, \\
R_{1,i} &:= \left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[w_{it} w_{it}']\right) - K_{1,i}\left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[f_{0,t} w_{it}']\right), \\
R_{2,i} &:= \left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[f_{0,t} f_{0,t}']\right) - K_{2,i}\left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[w_{it} f_{0,t}']\right).
\end{aligned}
$$

Proof. Given the expression of $S_{h,i}(\theta_i|\widehat{F}^{(0)})$, it is immediately to obtain

$$
\widetilde{\theta}_i^{(1)} - \theta_{0,i} = \left(\frac{1}{T}\sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) z_{it}^{(0)} z_{it}^{(0)\prime}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T}\Big[\dot{\rho}_h(\epsilon_{it}) + \Delta_{it}^{(0)} \ddot{\rho}_h(\epsilon_{it})\Big] z_{it}^{(0)}. \tag{3.13}
$$

Firstly, Lemma S.2.4 implies $T^{-1}\sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) z_{it}^{(0)} z_{it}^{(0)\prime} = T^{-1}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[z_{it} z_{it}'] + o_P(1)$. Moreover, the block-diagonal elements of $T^{-1}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[z_{it} z_{it}']$ are invertible under Assumptions 3.3.1.(ii) and 3.3.2.(ii). That said, $T^{-1}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[z_{it} z_{it}']$ is invertible, and admits the decomposition [11]

$$\Big(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\Big)^{-1} = \begin{bmatrix} R_{1,i} & \mathbf{0} \\ \mathbf{0} & R_{2,i} \end{bmatrix}^{-1} \begin{bmatrix} I_{mp} & -K_{1,i} \\ -K_{2,i} & I_r \end{bmatrix}, \tag{3.14}$$

where $K_{j,i}$ and $R_{j,i}$ ($j = 1,2$) are given in Lemma 3.8.1.

Since $\widetilde{b}_i^{(0)}$ and $\widetilde{\lambda}_i^{(0)}$ possess distinct convergence rates, to have a clear picture, we explicitly consider the limiting expression for both variables. That is, using the above decomposition, we derive the expression of $\widetilde{b}_i^{(0)}$ and $\widetilde{\lambda}_i^{(0)}$ from (3.13) that

$$\begin{aligned}
\big(R_{1,i} + o_P(1)\big)\Big(\widetilde{b}_{m,i}^{(0)} - b_{0,mi}\Big) &= \frac{1}{T}\sum_{t=1}^{T}\Big(\dot{\rho}_h(\epsilon_{it}) + \Delta_{it}^{(0)}\ddot{\rho}_h(\epsilon_{it})\Big)\Big(w_{it} - K_{1,i}\widehat{f}_t^{(0)}\Big),\\
\big(R_{2,i} + o_P(1)\big)\Big(\widetilde{\lambda}_{i}^{(0)} - \lambda_{0,i}\Big) &= \frac{1}{T}\sum_{t=1}^{T}\Big(\dot{\rho}_h(\epsilon_{it}) + \Delta_{it}^{(0)}\ddot{\rho}_h(\epsilon_{it})\Big)\Big(\widehat{f}_t^{(0)} - K_{2,i}w_{it}\Big).
\end{aligned}$$

Below, we first consider the order of $\widetilde{b}_{m,i}^{(0)} - b_{0,mi}$ by examining each term in $\sum_{t=1}^{T}\big(\dot{\rho}_h(\epsilon_{it}) + \Delta_{it}^{(0)}\ddot{\rho}_h(\epsilon_{it})\big)\big(w_{it} - K_{1,i}\widehat{f}_t^{(0)}\big)$. Specifically, recall the definition of $\Delta_{it}^{(0)}$, we want to show

$$\sum_{t=1}^{T}\big(\dot{\rho}_h(\epsilon_{it}) - \psi(\epsilon_{it})\big)(w_{it} - K_{1,i}f_{0,t}) = O_P\Big(\sqrt{\frac{T^2 m}{h}}\Big), \tag{3.15}$$

$$\sum_{t=1}^{T}\psi(\epsilon_{it})(w_{it} - K_{1,i}f_{0,t}) = O_P(\sqrt{mT}), \tag{3.16}$$

$$\sum_{t=1}^{T}\dot{\rho}_h(\epsilon_{it})K_{1,i}\big(\widehat{f}_t^{(0)} - f_{0,t}\big) = O_P\Big(\sqrt{\frac{T^2 m}{Nh}}\Big) + O_P\Big(\sqrt{\frac{mT}{N}}\Big), \tag{3.17}$$

$$\sum_{t=1}^{T}\ddot{\rho}_h(\epsilon_{it})\big(\beta_{0,i}(v_{it}) - \beta_{0,mi}(v_{it})\big)'x_{it}\big(w_{it} - K_{1,i}\widehat{f}_t^{(0)}\big) = O_P(Tm^{1/2-\mu}), \tag{3.18}$$

$$\sum_{t=1}^{T}\ddot{\rho}_h(\epsilon_{it})\big(f_{0,t} - \widehat{f}_t^{(0)}\big)'\lambda_{0,i}\big(w_{it} - K_{1,i}\widehat{f}_t^{(0)}\big) = O_P\Big(\sqrt{\frac{mT}{N}}\Big) + O_P\Big(\frac{\sqrt{m}T}{N}\Big). \tag{3.19}$$

First of all, (3.15) follows immediately from Lemma S.2.1.(iii) that $\dot{\rho}_h(\epsilon_{it}) - \psi(\epsilon_{it}) = O_P(h^{-1/2})$ uniformly, and $\mathbb{E}[\|w_{it}\|] = O(\sqrt{m})$, $\mathbb{E}\|K_{1,i}\| = O(\sqrt{m})$ under Assumption 3.3.1.

[11] It is known that $\left[\begin{smallmatrix} A & B \\ C & D\end{smallmatrix}\right]^{-1} = \left[\begin{smallmatrix} A-BD^{-1}C & 0 \\ 0 & D-CA^{-1}B\end{smallmatrix}\right]^{-1}\left[\begin{smallmatrix} I & -BD^{-1} \\ -CA^{-1} & I\end{smallmatrix}\right]$, suppose A and D invertible.

Additionally, as $\psi(\epsilon_{it})$ is a martingale sequence according to Assumption 3.3.2, and , it is easy to show $\sum_{t=1}^{T}\psi(\epsilon_{it})(w_{it}-K_{1,i}f_{0,t})=O_P(\sqrt{mT})$, which completes the proof of (3.16).

Next, we consider (3.17). Firstly, Hölder's inequality yields

$$\begin{aligned}
&\left\|\sum_{t=1}^{T}\big(\dot{\rho}_h(\epsilon_{it})-\psi(\epsilon_{it})\big)K_{1,i}\big(\widehat{f}_t^{(0)}-f_{0,t}\big)\right\| \\
&\quad\le\left(\sum_{t=1}^{T}\left\|\dot{\rho}_h(\epsilon_{it})-\psi(\epsilon_{it})\right\|^2\right)^{1/2}\cdot\left\|K_{1,i}\right\|\cdot\left(\sum_{t=1}^{T}\left\|\widehat{f}_t^{(0)}-f_{0,t}\right\|^2\right)^{1/2} \\
&\quad=O_P\left(\sqrt{\frac{T}{h}}\right)\cdot O_P(\sqrt{m})\cdot O_P(1)\cdot O_P\left(\sqrt{\frac{T}{N}}\right)=O_P\left(\sqrt{\frac{T^2m}{Nh}}\right).
\end{aligned}$$

And furthermore,

$$\begin{aligned}
\mathbb{E}\left\|\sum_{t=1}^{T}\psi(\epsilon_{it})K_{1,i}\big(\widehat{f}_t^{(0)}-f_{0,t}\big)\right\|^2=&\mathbb{E}\left\|\frac{1}{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\psi(\epsilon_{it})K_{1,i}\xi_{jt}\right\|^2 \\
=&\frac{1}{N^2}K_{1,i}\sum_{t_1,t_2=1}^{T}\sum_{j_1,j_2=1}^{N}\mathbb{E}\Big[\psi(\epsilon_{i,t_1})\psi(\epsilon_{i,t_2})\xi_{j_1,t_1}'\xi_{j_2,t_2}\Big]K_{1,i}'.
\end{aligned}$$

Recall that $\xi_{it}\equiv\xi(\epsilon_{it})$ is some function of ϵ_{it}, and $\mathbb{E}[\psi(\epsilon_{it})|\mathcal{F}_{t-1}]=\mathbb{E}[\xi(\epsilon_{it})|\mathcal{F}_{t-1}]=0$. Therefore, under Assumptions 3.3.2 and 3.3.5, it is easy to check

$$\begin{aligned}
\sum_{t_1,t_2=1}^{T}\sum_{j_1,j_2=1}^{N}\mathbb{E}\Big[\psi(\epsilon_{i,t_1})\psi(\epsilon_{i,t_2})\xi_{j_1,t_1}'\xi_{j_2,t_2}\Big]=&\sum_{t=1}^{T}\sum_{j=1}^{N}\mathbb{E}\Big[\psi(\epsilon_{it})^2\xi_{jt}'\xi_{jt}\Big]+\sum_{t=1}^{T}\sum_{j_1\neq j_2}^{N}\mathbb{E}\Big[\psi(\epsilon_{it})^2\xi_{j_1,t}'\xi_{j_2,t}\Big] \\
&+2\sum_{t_1\le t_2}^{T}\sum_{j_1,j_2=1}^{N}\mathbb{E}\Big[\psi(\epsilon_{i,t_1})\xi_{j_1,t_1}'\mathbb{E}\big[\psi(\epsilon_{i,t_2})\xi_{j_2,t_2}|\mathcal{F}_{t-2}\big]\Big] \\
=&O(NT).
\end{aligned}$$

In addition, as $\|K_{1,i}\|=O_P(\sqrt{m})$, we conclude that $\sum_{t=1}^{T}\psi(\epsilon_{it})K_{1,i}\big(\widehat{f}_t^{(0)}-f_{0,t}\big)=O_P(\sqrt{mT/N})$. Thus, (3.17) follows naturally.

Next, we consider the claim (3.18), whose rate is dominated controlled by the truncation error of $\beta_{0,i}(v)$. Before we proceed to the claim, we first note that under Assumption 3.3.4, Lemma S.2.3 shows that $\sum_{t=1}^{T}\left\|\big(\beta_{0,i}(v_{it})-\beta_{0,mi}(v_{it})\big)'x_{it}\right\|^2=O_p\big(Tm^{-2\mu}\big)$. Thus, Hölder's inequality leads to

$$\begin{aligned}
&\Big\| \sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) \big(\beta_{0,i}(v_{it}) - \beta_{0,mi}(v_{it})\big)' x_{it} w_{it} \Big\| \\
&\quad \le \Big(\sum_{t=1}^{T} \big\| \big(\beta_{0,i}(v_{it}) - \beta_{0,mi}(v_{it})\big)' x_{it} \big\|^2 \Big)^{1/2} \Big(\sum_{t=1}^{T} \big\| \ddot{\rho}_h(\epsilon_{it}) w_{it} \big\|^2 \Big)^{1/2} \\
&\quad = O_p\big(\sqrt{Tm^{-2\mu}}\big) \cdot O_P(\sqrt{Tm}) = O_P(Tm^{1/2-\mu}) = o_P(\sqrt{mT}),
\end{aligned}$$

under Assumptions 3.3.1 and 3.3.4. In addition, it is easy to check that replacing w_{it} with $K_{1,i}\widehat{f}_t^{(0)}$ leads to the same rate, which completes the proof of (3.18).

Finally, we consider

$$\begin{aligned}
\sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) \big(f_{0,t} - \widehat{f}_t^{(0)}\big)' \lambda_{0,i} \big(w_{it} - K_{1,i}\widehat{f}_t^{(0)}\big) =& \sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) \big(f_{0,t} - \widehat{f}_t^{(0)}\big)' \lambda_{0,i} \big(w_{it} - K_{1,i} f_{0,t}\big) \\
&- \sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) \big(f_{0,t} - \widehat{f}_t^{(0)}\big)' \lambda_{0,i} K_{1,i} \big(\widehat{f}_t^{(0)} - f_{0,t}\big),
\end{aligned}$$

which is dominant by the first term on the right-hand side, whose rate can be derived by analogous arguments of (3.17) under Assumptions 3.3.2 and 3.3.5. Taking the first element as an example,

$$\begin{aligned}
&\mathbb{E} \Big\| \sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) \big(\widehat{f}_t^{(0)} - f_{0,t}\big)' \lambda_{0,i} w_{it} \Big\|^2 \\
&= \frac{1}{N^2} \sum_{t_1,t_2=1}^{T} \sum_{j_1,j_2=1}^{N} \mathbb{E}\Big[\ddot{\rho}_h(\epsilon_{i,t_1}) \ddot{\rho}_h(\epsilon_{i,t_2}) \xi_{j_1,t_1}' \lambda_{0,i} w_{i,t_1}' w_{i,t_2} \lambda_{0,i} \xi_{j_2,t_2} \Big] \\
&= \frac{1}{N^2} \sum_{t=1}^{T} \sum_{j_1,j_2=1}^{N} \mathbb{E}\Big[w_{i,t}' w_{i,t} \lambda_{0,i}' \mathbb{E}\big[\ddot{\rho}_h(\epsilon_{i,t})^2 \xi_{j_1,t} \xi_{j_2,t}' | \mathcal{F}_{t-1} \big] \lambda_{0,i} \Big] \\
&\quad + \frac{2}{N^2} \sum_{t_1 \le t_2} \sum_{j_1,j_2=1}^{N} \mathbb{E}\Big[\ddot{\rho}_h(\epsilon_{i,t_1}) w_{i,t_1}' w_{i,t_2} \xi_{j_1,t_1}' \lambda_{0,i} \mathbb{E}\big[\ddot{\rho}_h(\epsilon_{i,t_2}) \xi_{j_2,t_2}' | \mathcal{F}_{t_2-1} \big] \lambda_{0,i} \Big] \\
&= O(m) \Big(O(T/N) + O(T^2/N^2) \Big).
\end{aligned}$$

where the last line holds since $\mathbb{E}\big[\ddot{\rho}_h(\epsilon_{i,t})^2 \xi_{j_1,t} \xi_{j_2,t}' | \mathcal{F}_{t-1}\big] \neq 0$ only if $j_1 = j_2$, and $\mathbb{E}\big[\ddot{\rho}_h(\epsilon_{it}) \xi_{jt}' | \mathcal{F}_{t-1}\big] \neq$

0 only if $j = i$.

Given (3.15)-(3.19), by choosing $h = O(T^{1+c})$, we conclude

$$\widetilde{b}^{(1)}_{m,i} - b_{0,mi} = \big(R_{1,i}\big)^{-1} \frac{1}{T} \sum_{t=1}^{T} \psi(\epsilon_{it}) \Big(w_{it} - K_{1,i} f_{0,t} \Big) + O_P(m^{1/2-\mu}) + o_P\Bigg(\sqrt{\frac{m}{T}}\Bigg), \tag{3.20}$$

and $\|\widetilde{b}^{(1)}_{m,i} - b_{0,mi}\| = O_P(\sqrt{m/T})$, since $O_P(m^{1/2-\mu}) = o_P(\sqrt{m/T})$ under Assumption 3.3.4.(iii). Then, by the Parseval equality, $\|\widetilde{\beta}^{(1)}_{m,i} - \beta_{0,mi}\|_{L^2} = O_P(\sqrt{m/T})$. Furthermore, Assumption 3.3.4 implies that $\big\|\beta_{0,i} - \beta_{0,mi}\big\|_{L^2} = O_P(m^{-\mu}) = o_P(\sqrt{m/T})$. Thus, by triangle inequality we have

$$\big\|\widetilde{\beta}^{(1)}_{m,i} - \beta_{0,i}\big\|_{L^2} \leq \big\|\widetilde{\beta}^{(1)}_{m,i} - \beta_{0,mi}\big\|_{L^2} + \big\|\beta_{0,i} - \beta_{0,mi}\big\|_{L^2} = O_P\Bigg(\sqrt{\frac{m}{T}}\Bigg),$$

and

$$\begin{aligned}
\widetilde{\beta}^{(1)}_{m,i}(v) - \beta_{0,i}(v) &= \Big(\widetilde{\beta}^{(1)}_{m,i}(v) - \beta_{0,mi}(v)\Big) + \Big(\beta_{0,mi}(v) - \beta_{0,i}(v)\Big) \\
&= [H_m(v)' \otimes I_p] \big(R_{1,i}\big)^{-1} \frac{1}{T} \sum_{t=1}^{T} \psi(\epsilon_{it}) \Big(w_{it} - K_{1,i} f_{0,t} \Big) + o_P\Bigg(\sqrt{\frac{m}{T}}\Bigg).
\end{aligned}$$

Next, we consider the limiting expression of $\widetilde{\lambda}^{(1)}_i$. By replacing w_{it} with $f_{0,t}$, analogous arguments yields

$$\widetilde{\lambda}^{(1)}_i - \lambda_{0,i} = \big(R_{2,i}\big)^{-1} \frac{1}{T} \sum_{t=1}^{T} \psi(\epsilon_{it}) \Big(f_{0,t} - K_{2,i} w_{it} \Big) + o_P\Bigg(\frac{1}{\sqrt{T}}\Bigg), \tag{3.21}$$

and $\|\widetilde{\lambda}^{(1)}_i - \lambda_{0,i}\| = O_P(T^{-1/2})$.

■

Proof of Theorem 3.3.1. Before proceeding to the details, we first briefly sketch the main steps. Since the estimation is based on an iterative algorithm, we derive the Bahadur-Kiefer Representation by induction. We first derive the representation of $\widetilde{\beta}^{(1)}_i(v)$ and $\widetilde{\gamma}^{(1)}_i$ given the initial estimator $\widehat{F}^{(0)}$, followed by that of $\widehat{f}^{(k)}_t$ for $k \geq 1$ given $\widehat{\beta}^{(k)}_i(v)$ and $\widehat{\gamma}^{(k)}_i$. Finally, we complete the proof by obtaining the general representation of $\widehat{\beta}_i^{(k+1)}(v)$ and $\widehat{\gamma}_i^{(k+1)}$ given $\widehat{f}_t^{(k)}$ for $k \geq 1$.

Step 1: We start from $k = 1$. By virtue of Lemma S.2.1.(ii) that $\sup_u |\rho_h(u) - \rho(u)| \leq Ch^{-1/2}$, we have

$$\sup_\theta \left| \widetilde{L}_i\big(\theta|\widehat{F}^{(0)}\big) - \widetilde{L}_{h,i}\big(\theta|\widehat{F}^{(0)}\big)\right| = O(h^{-1/2})$$

almost surely. In addition, recall that $S_{h,i}(\theta|\widehat{F}^{(0)})$ is the second-order Taylor expansion of $\widetilde{L}_{h,i}\big(\theta|\widehat{F}^{(0)}\big)$, we thus have

$$\left|\widetilde{L}_i\big(\theta_i|\widehat{F}^{(0)}\big) - S_{h,i}\big(\theta_i|\widehat{F}^{(0)}\big)\right| = O_P(h^{-1/2}) + O_P(\|\theta_i - \theta_{0,i}\|^3). \tag{3.22}$$

Recall that $\widehat{\theta}_i^{(1)} = (\widehat{b}_{m,i}^{(1)\prime}, \widehat{\lambda}_i^{(1)\prime})'$ and $\widetilde{\theta}_i^{(1)} = (\widetilde{b}_{m,i}^{(1)\prime}, \widetilde{\lambda}_i^{(1)\prime})'$ are the minimizers of $\widetilde{L}_i\big(\theta_i|\widehat{F}^{(0)}\big)$ and $S_{h,i}\big(\theta_i|\widehat{F}^{(0)}\big)$, respectively. It is shown by (3.20)-(3.21) in the proof of Lemma 3.8.1 that $\|\widetilde{b}_{m,i}^{(1)} - b_{0,mi}\| = O_P(\sqrt{m/T})$ and $\|\widetilde{\lambda}_i^{(1)} - \lambda_{0,i}\| = O_P(\sqrt{1/T})$. Then, to show the convergence rate of $\widehat{b}_{m,i}^{(1)}$ and $\widehat{\lambda}_i^{(1)}$, we shall show

$$\sqrt{\frac{T}{m}}\big(\widehat{\theta}_i^{(1)} - \widetilde{\theta}_i^{(1)}\big) = o_P(d_{NT}), \tag{3.23}$$

where $d_{NT} := T^{-1/4}\log(\log(T))$. To this end, for any given $\varepsilon > 0$, we shall show $\mathbb{P}(\sqrt{T/m}\|\widehat{\theta}_i^{(1)} - \widetilde{\theta}_i^{(1)}\| > d_{NT}\varepsilon) \to 0$ as $(N,T) \to (\infty,\infty)$. Denote $D_i := \{\theta_i : \sqrt{T/m}\|\theta_i - \widetilde{\theta}_i^{(1)}\| \leq d_{NT}\varepsilon\}$ be a ball with center $\widetilde{\theta}_i^{(1)}$ and radium $d_{NT}\varepsilon$. Since $\sqrt{T/m}(\widetilde{\theta}_i^{(1)} - \theta_{0,i})$ is bounded in probability according to Lemma 3.8.1, D_i can be covered by some compact set that satisfies $\|\sqrt{T/m}(\theta_i - \theta_{0,i})\| \leq c$. Then, by choosing $h = O(T^{1+c})$, (3.22) implies that

$$r_i := \sup_{\theta_i \in D_i} |\widetilde{L}_{h,i}(\theta_i|\widehat{F}^{(0)}) - S_{h,i}(\theta_i|\widehat{F}^{(0)})| = O_P(T^{-(1+c)/2}) + O_P\big((m/T)^{3/2}\big). \tag{3.24}$$

Suppose $\theta_i \notin D_i$, then there exists $c_0 > \varepsilon$ such that $\theta_i = \widetilde{\theta}_i^{(1)} + c_0 d_{NT}\sqrt{m/T}e$ where e is a unit vector. Let $\theta_i^\dagger \in D_i$ be the boundary point on the segment joining θ_i and $\widetilde{\theta}_i^{(1)}$, that is

$\theta_i^\dagger = \widetilde{\theta}_i^{(1)} + \varepsilon d_{NT}\sqrt{m/T}e.$

Using the expression of $\widetilde{\theta}_i^{(1)}$, we reformulate $S_{h,i}(\theta|\widehat{F}^{(0)})$ as follows:

$$\begin{aligned} S_{h,i}(\theta_i|\widehat{F}^{(0)}) &= \frac{1}{2}\big(\theta_i - \widetilde{\theta}_i^{(1)}\big)'\Sigma_i^{(0)}\big(\theta_i - \widetilde{\theta}_i^{(1)}\big) - \frac{1}{2}\big(\widetilde{\theta}_i^{(1)} - \theta_{0,i}\big)'\Sigma_i^{(0)}\big(\widetilde{\theta}_i^{(1)} - \theta_{0,i}\big), \\ &= \frac{1}{2}\big(\theta_i - \widetilde{\theta}_i^{(1)}\big)'\Sigma_i^{(0)}\big(\theta_i - \widetilde{\theta}_i^{(1)}\big) + S_{h,i}(\widetilde{\theta}_i^{(1)}|\widehat{F}^{(0)}) \end{aligned} \tag{3.25}$$

where $\Sigma_i^{(0)} := T^{-1}\sum_{t=1}^T \ddot{\rho}_h(\epsilon_{it})z_{it}^{(0)}z_{it}^{(0)\prime}$. Then, it follows from the convexity of $\widetilde{L}_i(\theta_i|\widehat{F}^{(0)})$ and (3.25) that

$$\begin{aligned} \frac{\varepsilon}{c_0}\widetilde{L}_i(\theta_i|\widehat{F}^{(0)}) + \left(1 - \frac{\varepsilon}{c_0}\right)\widetilde{L}_i(\widetilde{\theta}_i^{(1)}|\widehat{F}^{(0)}) \geq & S_{h,i}(\theta_i^\dagger|\widehat{F}^{(0)}) + \big[\widetilde{L}_i(\theta_i^\dagger|\widehat{F}^{(0)}) - S_{h,i}(\theta_i^\dagger|\widehat{F}^{(0)})\big] \\ \geq & \frac{m(d_{NT}\varepsilon)^2}{2T}e'\Sigma_i^{(0)}e + S_{h,i}(\widetilde{\theta}_i^{(1)}|\widehat{F}^{(0)}) - r_i \\ \geq & \frac{m(d_{NT}\varepsilon)^2}{2T}e'\Sigma_i^{(0)}e + \widetilde{L}_i(\widetilde{\theta}_i^{(1)}|\widehat{F}^{(0)}) - 2r_i \\ = & \frac{m(d_{NT}\varepsilon)^2}{2T}\varrho_{\min}\big(\Sigma_i^{(0)}\big) + \widetilde{L}_i(\widetilde{\theta}_i^{(1)}|\widehat{F}^{(0)}) - 2r_i. \end{aligned}$$

Rearranging to have

$$\inf_{\theta_i \notin D_i} \widetilde{L}_i(\theta_i|\widehat{F}^{(0)}) \geq \widetilde{L}_i(\widetilde{\theta}_i^{(1)}|\widehat{F}^{(0)}) + \frac{mc_0 d_{NT}^2}{2T\varepsilon}\left[\varepsilon^2\varrho_{\min}\big(\Sigma_i^{(0)}\big) - \frac{4r_iT}{d_{NT}^2 m}\right].$$

Since $\Sigma_i^{(0)} - T^{-1}\sum_{t=1}^T g_{it}(0)\mathbb{E}[z_{it}z_{it}'] = o_P(1)$ according to Lemma S.2.4, and the latter term is positive definite under Assumptions 3.3.1.(ii) and 3.3.2.(ii). It follows that $\varepsilon^2\varrho_{\min}\big(\Sigma_i^{(0)}\big)$ converges in probability to a constant $c_1 > 0$. Moreover, it is easy to check that $r_iT/(d_{NT}^2 m) = o_P(1)$, implying that eventually $4r_iT/(d_{NT}^2 m) < c_1/2$ in probability. Thus,

$$\inf_{\theta_i \notin D_i} \widetilde{L}_{h,i}(\theta_i|\widehat{F}^{(0)}) \geq \widetilde{L}_{h,i}(\widetilde{\theta}_i^{(1)}|\widehat{F}^{(0)}) + \frac{mc_0c_1d_{NT}^2}{4T\varepsilon} \geq \widetilde{L}_{h,i}(\widetilde{\theta}_i^{(1)}|\widehat{F}^{(0)})$$

for each N and T sufficiently large. Consequently, it is impossible that $\widetilde{L}_i(\theta_i|\widehat{F}^{(0)})$ attains its minimum outside D_i, which completes the proof of (3.23).

Therefore, combining (3.20)-(3.21) with (3.23), we immediately obtain

$$\widehat{b}^{(1)}_{m,i} - b_{0,mi} = \big(R_{1,i}\big)^{-1}\frac{1}{T}\sum_{t=1}^{T}\psi(\epsilon_{it})\Big(w_{it} - K_{1,i}f_{0,t}\Big) + o_P\bigg(\sqrt{\frac{m}{T}}\bigg), \tag{3.26}$$

$$\widehat{\lambda}^{(1)}_{i} - \lambda_{0,i} = \big(R_{2,i}\big)^{-1}\frac{1}{T}\sum_{t=1}^{T}\psi(\epsilon_{it})\Big(f_{0,t} - K_{2,i}w_{it}\Big) + o_P\bigg(\frac{1}{\sqrt{T}}\bigg), \tag{3.27}$$

and, $\|\widehat{b}^{(1)}_{m,i} - b_{0,mi}\| = O_P(\sqrt{m/T})$ and $\|\widehat{\lambda}^{(1)}_{i} - \lambda_{0,i}\| = O_P(1/\sqrt{T})$.

Step 2: Next, we analyze $\widehat{f}^{(1)}_t$ by showing (i) $\|\widetilde{f}^{(1)}_t - f_{0,t}\| = O_P(\sqrt{N})$ and (ii) $\|\widehat{f}^{(1)}_t - \widetilde{f}^{(1)}_t\| = o_P(\sqrt{N})$.

Given the definition of $\widetilde{S}_{h,t}(f_t|\widehat{\Theta}^{(1)})$, its minimizer $\widetilde{f}^{(1)}_t$ satisfies

$$\begin{aligned}\widetilde{f}^{(1)}_t - f_{0,t} = &\bigg(\frac{1}{N}\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})\widehat{\lambda}^{(1)}_i\widehat{\lambda}^{(1)\prime}_i\bigg)^{-1} \\ &\cdot\frac{1}{N}\sum_{i=1}^{N}\Big(\dot{\rho}_h(\epsilon_{it}) + \ddot{\rho}_h(\epsilon_{it})\Big[x'_{it}\big(\beta_{0,i}(v_{it}) - \widehat{\beta}^{(1)}_{m,i}(v_{it})\big) + f'_{0,t}(\lambda_{0,i} - \widehat{\lambda}^{(1)}_i)\Big]\Big)\widehat{\lambda}^{(1)}_i.\end{aligned}$$

To obtain its limiting expression and convergence rate, we show the following claims

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\ddot{\rho}_h(\epsilon_{it})\widehat{\lambda}^{(1)}_i\widehat{\lambda}^{(1)\prime}_i - g_{it}(0)\mathbb{E}\big[\lambda_{0,i}\lambda'_{0,i}\big]\Big) = o_P(1), \tag{3.28}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\dot{\rho}_h(\epsilon_{it})\widehat{\lambda}^{(1)}_i - \psi(\epsilon_{it})\lambda_{0,i}\Big) = O_P\bigg(\frac{1}{\sqrt{NT}}\bigg) + O_P\bigg(\frac{1}{T}\bigg) + O_P\bigg(\frac{1}{\sqrt{h}}\bigg), \tag{3.29}$$

$$\frac{1}{N}\sum_{i=1}^{N}\psi(\epsilon_{it})\lambda_{0,i} = O_P\bigg(\frac{1}{\sqrt{N}}\bigg), \tag{3.30}$$

$$\frac{1}{N}\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x'_{it}\big(\beta_{0,i}(v_{it}) - \widehat{\beta}^{(1)}_{m,i}(v_{it})\big)\widehat{\lambda}^{(1)}_i = O_P\bigg(\frac{1}{\sqrt{NT}}\bigg) + O_P\bigg(\frac{1}{T}\bigg), \tag{3.31}$$

$$\frac{1}{N}\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})f'_{0,t}(\lambda_{0,i} - \widehat{\lambda}^{(1)}_i)\widehat{\lambda}^{(1)}_i = O_P\bigg(\frac{1}{\sqrt{NT}}\bigg) + O_P\bigg(\frac{1}{T}\bigg), \tag{3.32}$$

Note that (3.28) can be shown in the same way as Lemma S.2.4, thus the proof is omitted. Also, (3.30) is straightforward as $\psi(\epsilon_{it})$ is a martingale-difference sequence. The proofs of the remaining terms are tedious but similar to those of claims (3.15)-(3.19), thus details are provided later in Lemma S.2.5.

Then, given (3.28)-(3.32), and the fact that $N^{-1}\sum_{i=1}^{N} g_{it}(0)\mathbb{E}[\lambda_{0,i}\lambda_{0,i}']$ is invertible under Assumptions 3.3.1.(ii) and 3.3.2.(ii), we obtain the asymptotic expression of $\widetilde{f}_t^{(1)}$ as

$$\widetilde{f}_t^{(1)} - f_{0,t} = \left(\frac{1}{N}\sum_{i=1}^{N} g_{it}(0)\mathbb{E}[\lambda_{0,i}\lambda_{0,i}']\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\psi(\epsilon_{it})\lambda_{0,i} + o_P(N^{-1/2}),$$

and $\|\widetilde{f}_t^{(1)} - f_{0,t}\| = O_P(N^{-1/2})$.

Furthermore, a similar arguments to the proof of (3.23) immediately yields that $\sqrt{N}(\widehat{f}_t^{(1)} - \widetilde{f}_t^{(1)}) = o_P(1)$, thus, we conclude

$$\begin{aligned}\widehat{f}_t^{(1)} - f_{0,t} =& (\widehat{f}_t^{(1)} - \widetilde{f}_t^{(1)}) + (\widetilde{f}_t^{(1)} - f_{0,t})\\ =& \left(\frac{1}{N}\sum_{i=1}^{N} g_{it}(0)\mathbb{E}[\lambda_{0,i}\lambda_{0,i}']\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\psi(\epsilon_{it})\lambda_{0,i} + o_P(N^{-1/2}).\end{aligned}$$

Step 3: Finally, we extend the above analysis to $k > 1$. Since $\widehat{f}_t^{(1)} - f_{0,t}$ has a linear expression that satisfies Assumption 3.3.5. Therefore, the arguments for $\widehat{\beta}_{m,i}^{(1)}(v)$ and $\widehat{\lambda}_i^{(1)}$ can be extended to $\widehat{\beta}_{m,i}^{(k)}(v)$ and $\widehat{\lambda}_i^{(k)}$ for $k \geq 2$ naturally, and so is that for the factor $\widehat{f}_t^{(k)}$. Hence, after the estimation algorithm converges, we have

$$\begin{aligned}\sqrt{\frac{T}{m}}\Big(\widehat{\beta}_{m,i}(v) - \beta_{0,i}(v)\Big) =& \sqrt{\frac{T}{m}}[H_m(v)'\otimes I_p]\Big(\widehat{b}_{m,i}(v) - b_{0,mi}(v)\Big) + \sqrt{\frac{T}{m}}\Big(\beta_{0,mi}(v) - \beta_{0,i}(v)\Big)\\ =&[H_m(v)'\otimes I_p]\big(R_{1,i}\big)^{-1}\frac{1}{\sqrt{mT}}\sum_{t=1}^{T}\psi(\epsilon_{it})\Big(w_{it} - K_{1,i}f_{0,t}\Big) + o_P(1),\\ \sqrt{T}\Big(\widehat{\lambda}_i - \lambda_{0,i}\Big) =& \big(R_{2,i}\big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi(\epsilon_{it})\Big(f_{0,t} - K_{2,i}w_{it}\Big) + o_P(1),\\ \sqrt{N}\Big(\widehat{f}_t - f_{0,t}\Big) =& \left(\frac{1}{N}\sum_{i=1} g_{it}(0)\mathbb{E}[\lambda_{0,i}\lambda_{0,i}']\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi(\epsilon_{it})\lambda_{0,i} + o_P(1).\end{aligned} \tag{3.33}$$

In addition, we have shown the rate of convergence for $\widehat{\beta}_i$, $\widehat{\lambda}_i$ and $\widehat{f}_t$ along the line in deriving the above expressions, which completes the proof.

■

Proof of Theorem 3.3.2. First of all, let S_β and S_λ be the selection matrices drawing β_i and λ_i from θ_i, that is $\beta_i = S_{\beta_i}\theta_i$, and a similarly definition applies to S_{λ_i}.

First of all, the limiting expression of $\widehat{\beta}_{m,i}(v)$ and $\widehat{\lambda}_i$ in (3.33) can be written more compactly as

$$\sqrt{\frac{T}{m}}\Big(\widehat{\beta}_{m,i}(v)-\beta_{0,i}(v)\Big) = [H_m(v)'\otimes I_p]S_\beta\left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\right)^{-1}\frac{1}{\sqrt{mT}}\sum_{t=1}^{T}\psi(\epsilon_{it})z_{it}+o_P(1),$$
$$\sqrt{T}\Big(\widehat{\lambda}_i-\lambda_{0,i}\Big) = S_\lambda\left(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi(\epsilon_{it})z_{it}+o_P(1),$$

making use of the relationship of $K_{j,i}$ and $\Sigma_{j,i}$ $(j=1,2)$ in (3.14).

Since $\{\psi(\epsilon_{it}), t=1,...,T\}$ is a martingale difference sequence, it is easy to verify $\mathbb{E}\big[\widehat{\beta}_{m,i}(v)-\beta_{0,i}(v)\big]=0$ and $\mathbb{E}\big[\widehat{\lambda}_i-\lambda_{0,i}\big]=0$.

Furthermore, we compute the asymptotic covariance. Again, using the property of martingale difference sequence, we have

$$\begin{aligned}\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t,s=1}^{T}\psi(\epsilon_{it})\psi(\epsilon_{is})z_{it}z_{is}'\right] &= \lim_{T\to\infty}\frac{1}{mT}\sum_{t=1}^{T}\mathbb{E}\Big[\psi(\epsilon_{it})^2 z_{it}z_{it}'\Big]\\ &= \lim_{T\to\infty}\frac{\tau(1-\tau)}{T}\sum_{t=1}^{T}\mathbb{E}\Big[z_{it}z_{it}'\Big] =: \tau(1-\tau)\sigma_{i,\tau,m}^{Z}.\end{aligned}$$

Additionally, $\Gamma_{i,\tau,m}^{Z} := \text{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}[z_{it}z_{it}']$ is uniformly bounded away from 0 for all m under Assumption 3.3.1.(ii) and 3.3.2.(ii). Thus, we obtain the asymptotic covariance

$$\begin{aligned}\Sigma_{i,\tau}^{\lambda} &= \lim_{T,m\to\infty}\frac{1}{T}S_\lambda\mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T}g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\right)^{-1}\sum_{t,s=1}^{T}\psi(\epsilon_{it})\psi(\epsilon_{is})z_{it}z_{is}'\right.\\ &\qquad\left.\cdot\left(\frac{1}{T}\sum_{t=1}^{T}g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\right)^{-1}\right]S_\lambda'\\ &= \lim_{m\to\infty}\tau(1-\tau)S_\lambda\big(\Gamma_{i,\tau,m}^{Z}\big)^{-1}\sigma_{i,\tau,m}^{Z}\big(\Gamma_{i,\tau,m}^{Z}\big)^{-1}S_\lambda',\end{aligned}$$

and

$$\Sigma^{\beta}_{i,\tau}(v) = \lim_{T,m\to\infty} \frac{1}{mT}[H_m(v)' \otimes I_p]S_\beta \mathbb{E}\Bigg[\Bigg(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\Bigg)^{-1} \sum_{t,s=1}^{T} \psi(\epsilon_{it})\psi(\epsilon_{is})z_{it}z_{is}' \cdot \Bigg(\frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\Bigg)^{-1}\Bigg] S_\beta'[H_m(v)\otimes I_p]$$

$$= \lim_{m\to\infty} \frac{\tau(1-\tau)}{m}[H_m(v)'\otimes I_p]S_\beta\big(\Gamma^Z_{i,\tau,m}\big)^{-1}\sigma^Z_{i,\tau,m}\big(\Gamma^Z_{i,\tau,m}\big)^{-1}S_\beta'[H_m(v)\otimes I_p].$$

Finally, we show they converges to normal distributions with the aid of Cramer-Wold device. Taking $\widehat{\lambda}_i$ as an example, to show $\sqrt{T}(\widehat{\lambda}_i - \lambda_{0,i}) \xrightarrow{d} N(0, \Sigma^{\lambda}_{i,\tau})$, it is sufficient to show $\sqrt{T}a'(\widehat{\lambda}_i - \lambda_{0,i}) \xrightarrow{d} N(0, a'\Sigma^{\lambda}_{i,\tau}a)$ for any arbitrary vector $a \in \mathbb{R}^r$.

Let $a \in \mathbb{R}^r$. Define $r_{it} := a'S_\lambda\big(T^{-1}\sum_{t=1}^{T} g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\big)^{-1}T^{-1/2}z_{it}\psi(\epsilon_{it})$, $R_i := \sum_{t=1}^{T} r_{it} = \sqrt{T}a'\big(\widehat{\lambda}_i - \lambda_{0,i}\big)$, and $\Phi_i^2 := a'\Sigma^{\lambda}_{i,\tau}a$. Since $\mathbb{E}[r_{it}|\mathcal{F}_{t-1}] = 0$, we can apply the martingale central limit theorem by Brown (1971), which states that $\Phi_i^{-1}R_i \xrightarrow{d} N(0,1)$ if

$$\Phi_i^{-2}\sum_{t=1}^{T}\mathbb{E}\Big[r_{it}^2 1\{|r_{it}| > \varepsilon\Phi_i\}\big|\mathcal{F}_{t-1}\Big] \xrightarrow{p} 0, \quad \varepsilon > 0, \tag{3.34}$$

$$\Phi_i^{-2}\sum_{t=1}^{T}\mathbb{E}\Big[r_{it}^2\big|\mathcal{F}_{t-1}\Big] \xrightarrow{p} 1. \tag{3.35}$$

We start with (3.34). Let $\varepsilon > 0$ be given, we have

$$\begin{aligned}\Phi_i^{-2}\sum_{t=1}^{T}\mathbb{E}\Big[r_{it}^2 1\{|r_{it}| > \varepsilon\Phi_i\}\big|\mathcal{F}_{t-1}\Big] &= \varepsilon^{-2}\Phi_i^{-4}\sum_{t=1}^{T}\mathbb{E}\Big[\varepsilon^2\Phi_i^2 r_{it}^2 1\{|r_{it}| > \varepsilon\Phi_i\}\big|\mathcal{F}_{t-1}\Big]\\ &< \varepsilon^{-2}\Phi_i^{-4}\sum_{t=1}^{T}\mathbb{E}\Big[r_{it}^4 1\{|r_{it}| > \varepsilon\Phi_i\}\big|\mathcal{F}_{t-1}\Big]\\ &\le \varepsilon^{-2}\Phi_i^{-4}\sum_{t=1}^{T}\mathbb{E}\Big[r_{it}^4\big|\mathcal{F}_{t-1}\Big].\end{aligned}$$

It follows that, to verify condition (3.34), it is sufficient to show $\Phi_i^{-4}\sum_{t=1}^{T}\mathbb{E}\big[r_{it}^4\big|\mathcal{F}_{t-1}\big] \xrightarrow{p} 0$.

Since Φ_i is invertible, which implies $|\Phi_i^{-1}| = O_P(1)$, and

$$\begin{aligned}
\sum_{t=1}^{T}\mathbb{E}\big[r_{it}^4|\mathcal{F}_{t-1}\big] &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\Bigg[\Bigg(a'S_\lambda\Big(\frac{1}{T}\sum_{t=1}^{T}g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\Big)^{-1}z_{it}\psi(\epsilon_{it})\Bigg)^4\Bigg|\mathcal{F}_{t-1}\Bigg] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\Bigg(a'S_\lambda\Big(\frac{1}{T}\sum_{t=1}^{T}g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\Big)^{-1}z_{it}\Bigg)^4\mathbb{E}\big[\psi(\epsilon_{it})^4|\mathcal{F}_{t-1}\big] \\
&= O_P\Big(\frac{1}{T}\Big),
\end{aligned}$$

under Assumption 3.3.1. We conclude $\Phi_i^{-4}\sum_{t=1}^{T}\mathbb{E}\big[r_{it}^4|\mathcal{F}_{t-1}\big] \xrightarrow{p} 0$, and thus Condition (3.34) follows.

Next, we verify Condition (3.35). Denote $K_i^2 := \sum_{t=1}^{T}\mathbb{E}[r_{it}^2|\mathcal{F}_{t-1}]$. In order to show $\Phi_i^{-2}K_i^2 \xrightarrow{p} 1$, it suffices to show $\mathbb{E}\big[(K_i^2-\Phi_i^2)^2\big] = o(1)$. We first note that, as $\Phi_i = \mathbb{E}[R_i^2] = \mathbb{E}\big[\sum_{t=1}^{T}\mathbb{E}[r_{it}^2|\mathcal{F}_{t-1}]\big] = \mathbb{E}[K_i^2]$, it is straightforward that $\mathbb{E}\big[(K_i^2-\Phi_i^2)^2\big] = \mathbb{E}[K_i^4]-\Phi_i^4$.

Note that

$$\begin{aligned}
K_i^2 &= \sum_{t=1}^{T}\mathbb{E}\big[r_{it}^2|\mathcal{F}_{t-1}\big] \\
&= \frac{1}{T}\sum_{t=1}^{T}\Bigg(a'S_\lambda\Big(\frac{1}{T}\sum_{t=1}^{T}g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\Big)^{-1}z_{it}\Bigg)^2\mathbb{E}\big[\psi(\epsilon_{it})^2|\mathcal{F}_{t-1}\big] \\
&= \frac{\tau(1-\tau)}{T}\sum_{t=1}^{T}\Bigg(a'S_\lambda\Big(\frac{1}{T}\sum_{t=1}^{T}g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\Big)^{-1}z_{it}\Bigg)^2 \\
&= \tau(1-\tau)\psi_i'\Big(\frac{1}{T}\sum_{t=1}^{T}z_{it}z_{it}'\Big)\psi_i,
\end{aligned}$$

where $\psi_i := \Big(\frac{1}{T}\sum_{t=1}^{T}g_{it}(0)\mathbb{E}\big[z_{it}z_{it}'\big]\Big)^{-1}S_\lambda'a$. Thus, by re-writing $\Phi_i^2 = \tau(1-\tau)\psi_i'\mathbb{E}[T^{-1}\sum_{t=1}^{T}z_{it}z_{it}']\psi_i$, we obtain

$$\begin{aligned}
&\mathbb{E}\big[K_i^4\big]-\Phi_i^4 = \mathbb{E}\big[K_i^4-\Phi_i^4\big] \\
&=\tau^2(1-\tau)^2\mathbb{E}\Bigg[\psi_i'\Big(\frac{1}{T}\sum_{t=1}^{T}z_{it}z_{it}'\Big)\psi_i\psi_i'\Big(\frac{1}{T}\sum_{s=1}^{T}z_{is}z_{is}'\Big)\psi_i \\
&\qquad - \psi_i'\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[z_{it}z_{it'}\big]\Big)\psi_i\psi_i'\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[z_{it}z_{it'}\big]\Big)\psi_i\Bigg]
\end{aligned}$$

$$
\begin{aligned}
=&\tau^2(1-\tau)^2\psi_i'\mathbb{E}\Big[\Big(\frac{1}{T}\sum_{t=1}^T z_{it}z_{it}'\Big)\psi_i\psi_i'\Big(\frac{1}{T}\sum_{s=1}^T z_{is}z_{is}'\Big)\\
&\qquad\qquad -\Big(\frac{1}{T}\sum_{t=1}^T \mathbb{E}\big[z_{it}z_{it'}\big]\Big)\psi_i\psi_i'\Big(\frac{1}{T}\sum_{t=1}^T \mathbb{E}\big[z_{it}z_{it}'\big]\Big)\Big]\psi_i\\
=&\tau^2(1-\tau)^2\psi_i'\Big\{\mathbb{E}\Big[\frac{1}{T}\sum_{t=1}^T\Big(z_{it}z_{it}'-\mathbb{E}\big[z_{it}z_{it}'\big]\Big)\psi_i\psi_i'\Big(\frac{1}{T}\sum_{s=1}^T z_{is}z_{is}'\Big)\Big]\\
&+\Big(\frac{1}{T}\sum_{t=1}^T \mathbb{E}\big[z_{it}z_{it'}\big]\Big)\psi_i\psi_i'\mathbb{E}\Big[\frac{1}{T}\sum_{t=1}^T\Big(z_{it}z_{it}'-\mathbb{E}\big[z_{it}z_{it}'\big]\Big)\Big]\Big\}\psi_i=o(1),
\end{aligned}
$$

where the last equation follows from the law of large number. So far, both conditions have been verify. Therefore, by the martingale CLT of Brown (1971), we conclude $\Phi_i^{-1}R_i \xrightarrow{d} N(0,1)$. Equivalently, we have $\sqrt{T}a'(\widehat{\lambda}_i-\lambda_{0,i}) \xrightarrow{d} N(0, a'\Sigma_{i,\tau}^{\lambda}a)$. Since $a\in\mathbb{R}^r$ is arbitrary, due to Cramer-Wald device, we conclude

$$
\sqrt{T}(\widehat{\lambda}_i-\lambda_{0,i}) \xrightarrow{d} N(0,\Sigma_{i,\tau}^{\lambda}).
$$

Given the asymptotic mean and variance of $\widehat{\beta}_{m,i}(v)$ derived at the beginning of the proof, Cramer-Wald device and martingale CLT yield

$$
\sqrt{\frac{T}{m}}\Big(\widehat{\beta}_{m,i}(v)-\beta_{0,i}(v)\Big) \xrightarrow{d} N\big(0,\Sigma_{i,\tau}^{\beta}(v)\big),
$$

whose proof are similar to those for $\widehat{\lambda}_i$, and thus omitted for conciseness.

(ii) Recall the asymptotic expression of $\widehat{f}_t$ in (3.33), Since ϵ_{it} are ϵ_{jt} are independent conditional on $\mathcal{F}_{t-1}$, we have

$$
\begin{aligned}
\lim_{N\to\infty}\mathbb{E}\Big[\frac{1}{N}\sum_{i,j=1}^N\psi(\epsilon_{it})\psi(\epsilon_{jt})\lambda_{0,i}\lambda_{0,j}'\Big]=&\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\mathbb{E}\big[\psi(\epsilon_{it})^2\lambda_{0,i}\lambda_{0,i}'\big]\\
=&\tau(1-\tau)\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\mathbb{E}\big[\lambda_{0,i}\lambda_{0,i}'\big]=:\tau(1-\tau)\sigma_\tau^{\Lambda}.
\end{aligned}
$$

Then, by applying the martingale CLT and Cramer-Wald device, we immediately have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi(\epsilon_{it})\lambda_{0,i}\xrightarrow{d}N\Big(0,\tau(1-\tau)\sigma_{\tau}^{\Lambda}\Big),$$

as $\psi(\epsilon_{it})$ is a martingale-difference sequence. Furthermore, under Assumptions 3.3.1.(ii) and 3.3.2.(ii), $\Gamma_{t,\tau}^{\Lambda}:=\text{plim}_{N\to\infty}N^{-1}\sum_{i=1}^{N}g_{it}(0)\mathbb{E}[\lambda_{0,i}\lambda_{0,i}]$ is invertible. Thus, we can conclude

$$\sqrt{N}\Big(\widehat{f}_t-f_{0,t}\Big)\xrightarrow{d}N\Big(0,\tau(1-\tau)\big(\Gamma_{t,\tau}^{\Lambda}\big)^{-1}\sigma_{\tau}^{\Lambda}\big(\Gamma_{t,\tau}^{\Lambda}\big)^{-1}\Big).$$

■

Proof of Corollary 3.3.1. We start with the simplest term (3.11). In order to show $\widehat{\Sigma}_{t,\tau}^{f}\xrightarrow{p}\Sigma_{t,\tau}^{f}$, it is sufficient to show $\widehat{\Gamma}_{t,\tau}^{\Lambda}\xrightarrow{p}\Gamma_{t,\tau}^{\Lambda}$ and $\widehat{\sigma}_{\tau}^{\Lambda}\xrightarrow{p}\sigma_{\tau}^{\Lambda}$, respectively. First of all, the matrix σ_{τ}^{Λ} depends only on $\lambda_{0,i}$. Thus, in light of the consistency results in Theorem 3.3.1, the empirical counterpart $\widehat{\sigma}_{\tau}^{\Lambda}$ is a consistent estimator of σ_{τ}^{Λ}. More specifically, we have

$$\widehat{\sigma}_{\tau}^{\Lambda}-\sigma_{\tau}^{\Lambda}=\frac{1}{N}\sum_{i=1}^{N}\big(\widehat{\lambda}_i\widehat{\lambda}_i'-\lambda_{0,i}\lambda_{0,i}'\big)+\frac{1}{N}\sum_{i=1}^{N}\big(\lambda_{0,i}\lambda_{0,i}'-\mathbb{E}[\lambda_{0,i}\lambda_{0,i}']\big)+o_P(1)=o_P(1),$$

where the second term in the first equation is $o_P(1)$ by the law of large numbers, and the first term can be analyzed by plugging-in the asymptotic expression of $\widehat{\lambda}_i-\lambda_{0,i}$ from (3.33) as follows

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\big(\widehat{\lambda}_i\widehat{\lambda}_i'-\lambda_{0,i}\lambda_{0,i}'\big)=&\frac{1}{N}\sum_{i=1}^{N}\big(\widehat{\lambda}_i-\lambda_{0,i}\big)\widehat{\lambda}_i'+\frac{1}{N}\sum_{i=1}^{N}\lambda_{0,i}\big(\widehat{\lambda}_i-\lambda_{0,i}\big)'\\ =&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(R_{2,i}\big)^{-1}\psi(\epsilon_{it})\big(f_{0,t}-K_{2,i}w_{it}\big)\widehat{\lambda}_i'\\ &+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{0,i}\psi(\epsilon_{it})\big(f_{0,t}-K_{2,i}w_{it}\big)'\big(R_{2,i}\big)^{-1}+o_P(1)=o_P(1).\end{aligned}$$

For illustration, we consider

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda_{0,i}\psi(\epsilon_{it})f'_{0,t}\big(R_{2,i}\big)^{-1}\right\|^{2}\right]\\
&=\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathbb{E}\Big[\mathbb{E}\Big[\psi(\epsilon_{it})\psi(\epsilon_{js})\lambda_{0,i}f'_{0,t}\big(R_{2,i}\big)^{-1}\big(R_{2,j}\big)^{-1}f_{0,s}\lambda_{0,j}\Big|\mathcal{F}_{t-1}\Big]\Big]\\
&=\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t=1}^{T}\mathbb{E}\Big[\mathbb{E}\big[\psi(\epsilon_{it})\psi(\epsilon_{jt})\big|\mathcal{F}_{t-1}\big]\lambda_{0,i}f'_{0,t}\big(R_{2,i}\big)^{-1}\big(R_{2,j}\big)^{-1}f_{0,t}\lambda_{0,j}\Big]\\
&\quad+\frac{2}{N^2T^2}\sum_{i,j=1}^{N}\sum_{s<t}^{T}\mathbb{E}\Big[\mathbb{E}\big[\psi(\epsilon_{it})\big|\mathcal{F}_{t-1}\big]\psi(\epsilon_{js})\lambda_{0,i}f'_{0,t}\big(R_{2,i}\big)^{-1}\big(R_{2,j}\big)^{-1}f_{0,s}\lambda_{0,j}\Big]\\
&=\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t=1}^{T}\mathbb{E}\Big[\mathbb{E}\big[\psi(\epsilon_{it})\psi(\epsilon_{jt})\big|\mathcal{F}_{t-1}\big]\lambda_{0,i}f'_{0,t}\big(R_{2,i}\big)^{-1}\big(R_{2,j}\big)^{-1}f_{0,t}\lambda_{0,j}\Big]=O_P\left(\frac{1}{T}\right),
\end{aligned}$$

under Assumption 3.3.1 and Assumption 3.3.2.(ii) that $\mathbb{E}[\psi(\epsilon_{it})|\mathcal{F}_{t-1}]=0$ almost surely. The remaining terms in $N^{-1}\sum_{i=1}^{N}\big(\widehat{\lambda}_i\widehat{\lambda}'_i-\lambda_{0,i}\lambda'_{0,i}\big)$ can be analyzed accordingly.

Next, we consider the term $\widehat{\Gamma}^{\Lambda}_{t,\tau}\xrightarrow{p}\Gamma^{\Lambda}_{t,\tau}$. Under the assumption that ϵ_{it} are identically distributed across i and t, we have $g_{it}(0)\equiv g_t(0)$ for all i, and thus, $\Gamma^{\Lambda}_{t,\tau}=g_t(0)\sigma^{\Lambda}_{\tau}$. It remains to show $\widehat{g}_t(0)\xrightarrow{p}g_t(0)$. Recall that $g_t(0)=\mathbb{E}[\ddot{\rho}(\epsilon_{it})|\mathcal{F}_{t-1}]$, we have

$$\begin{aligned}
\big\|\widehat{g}_t(0)-g_t(0)\big\|=&\left\|\frac{1}{N}\sum_{i=1}^{N}\Big(\ddot{\rho}_h(\widehat{\epsilon}_{it})-\mathbb{E}[\ddot{\rho}(\epsilon_{it})|\mathcal{F}_{t-1}]\Big)\right\|\\
\leq&\left\|\frac{1}{N}\sum_{i=1}^{N}\Big(\ddot{\rho}_h(\widehat{\epsilon}_{it})-\ddot{\rho}_h(\epsilon_{it})\Big)\right\|+\left\|\frac{1}{N}\sum_{i=1}^{N}\Big(\ddot{\rho}_h(\epsilon_{it})-\mathbb{E}[\ddot{\rho}_h(\epsilon_{it})|\mathcal{F}_{t-1}]\Big)\right\|\\
&+\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\big[\ddot{\rho}_h(\epsilon_{it})-\ddot{\rho}(\epsilon_{it})|\mathcal{F}_{t-1}\big]\right\|=o_P(1),
\end{aligned}$$

where the first term in the inequality is $o_P(1)$ due to the differentiability of $\ddot{\rho}_h$ and the fact that $\widehat{\epsilon}_{it}-\epsilon_{it}=x'_{it}\big(\widehat{\beta}_{m,i}(v_{it})-\beta_{0,i}(v_{it})\big)+\big(\widehat{f}'_t\widehat{\lambda}_i-f'_{0,t}\lambda_{0,i}\big)=O_P\big(m^{1/2}T^{-1/2},N^{-1/2}\big)=o(1)$, the second term is $o_P(1)$ due to ergodicity of $\{\epsilon_{it},i=1,...,N\}$, and the last term is $O_P(h^{-1/2})=o_P(1)$ according to Lemma S.2.1.(iv). Accordingly, $\widehat{\Gamma}^{\Lambda}_{t,\tau}\xrightarrow{p}\Gamma^{\Lambda}_{t,\tau}$.

Similar arguments yield $\widehat{\Sigma}^{\beta}_{i,\tau}(v)\xrightarrow{p}\Sigma^{\beta}_{i,\tau}(v)$, $\widehat{\Sigma}^{\lambda}_{i,\tau}\xrightarrow{p}\Sigma^{\lambda}_{i,\tau}$, whose proofs are omitted to avoid unnecessary repetition.

■

3.8.3 Preliminaries on Generalized Functions

As a generalised function approach is adopted in this chapter to prove the main results in Section 3.3.2, we introduce some preliminaries in this section. Note that, in mathematical context, generalised functions are referred as distributions or tempered distributions according to the spaces of test functions. Additional advancing contents can be found in Gel'fand and Shilov (1964) and Kanwal (1998).

Definition 3.8.1 (Space D)**.** *The space of all functions $\phi(x)$ defined on the real line is called Space D if satisfying the following properties:*

(i) $\phi(x)$ is an infinitely differentiable function defined at every point on $\mathbb{R}$;

(ii) $\phi(x)$ has a compact support.

Definition 3.8.2 (Convergence in D)**.** *A sequence $\{\phi_m, m = 1, 2, ...\}$ in D is said to converge to a function ϕ_0 if the following conditions are satisfied:*

(i) all ϕ_m as well as ϕ_0 vanish outside a common region;

(ii) $\phi_m^{(k)} \to \phi_0^{(k)}$ uniformly over $\mathbb{R}$ as $m \to \infty$ for all $k > 0$.

Definition 3.8.3 (Distribution)**.** *A continuous linear functional on the space D is called a distribution. The space of all distributions on D is denoted by D'.*

All distributions can be classified into two categories. The set of distributions that are mostly useful are those generated by locally integrable functions: regular distributions.

Definition 3.8.4 (Regular and singular distribution)**.** *A distribution $f \in D'$ is called a regular distribution if the distribution $f : \phi \to \mathbb{R}$ is generated through $\langle f, \phi \rangle = \int f(x)\phi(x)dx$, for all $\phi \in D$, where $f(x)$ is a locally integrable function on $\mathbb{R}$. Any distributions other than regular ones are called singular distribution.*

Remarkably, it is proven that two continuous functions generating the same regular distribution are identical (Kanwal, 1998, p. 27). Moreover, if two locally integrable functions produce the same regular distributions, they are identical almost everywhere. These enable the function to be identified from the distributions they generate.

It is wellknown that not all ordinary functions are differentiable. In contrast, generalised functions always have derivatives, which are also generalised functions, and consequently have derivatives of any order. See Gel'fand and Shilov (1964, p. 18).

Definition 3.8.5 (Derivative of distributions)**.** *Let $f \in D'$. A functional g defined on D given by $\langle g, \phi\rangle = -\langle f, \phi'\rangle$, for all $\phi \in D$, is called the derivative of f, denoted as f'.*

Definition 3.8.6 (Convergence in D')**.** *A sequence of distributions $\{f_m,\ m = 1, 2, ...\}$ in D', is said to converge to a distribution $f \in D'$ if $\lim_{m\to\infty}\langle f_m, \phi\rangle = \langle f, \phi\rangle$, for all $\phi \in D$. A set of distributions $\{f_\epsilon\}$ indexed by real ϵ is said to converge to f when $\epsilon \to \epsilon_0$, if for all $\phi \in D$, $\lim_{\epsilon\to\epsilon_0}\langle f_\epsilon, \phi\rangle = \langle f, \phi\rangle$. In addition, a series of distributions $\sum_{m=1}^{\infty} f_m$ converges to a distribution f if the sequence of partial sum $s_M = \sum_{s=1}^{M} f_m$ converges to f as $M \to \infty$.*

These definitions contain the convergence of ordinary functions as a special case. Indeed, if all members of distribution sequence $\{f_m\}$ are regular, and $f_m(x)$ converge to $f(x)$ uniformly on any compact interval, then

$$\lim_{m\to\infty}\langle f_m, \phi\rangle = \lim_{m\to\infty}\int f_m(x)\phi(x)dx = \int f(x)\phi(x)dx = \langle f, \phi\rangle, \quad \forall\phi \in D,$$

by uniform convergence theorem.

A consequence of the definition is that if $f_m \to f$ then $f_m^{(k)} \to f^{(k)}$ for any $k > 0$; if $\sum_{m=0}^{\infty} f_m \to f$, then the series can be differentiated term by term as many times as required. See Kanwal (1998, p. 59).

The following lemma (Kanwal, 1998, p. 59) provides a method to construct delta sequence, which can be regarded as a bridge between regular and singular distributions.

Lemma 3.8.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be a non-negative function such that $\int_{\mathbb{R}} f(x)dx = 1$. Put*

$f_\epsilon(x) = \epsilon^{-1} f(x/\epsilon)$ *where* $\epsilon > 0$. *Then* $\lim_{\epsilon \to 0} f_\epsilon(x) = \delta(x)$, *where* $\delta(\cdot)$ *is the Dirac delta function.*

Note that Dirac delta $\delta(\cdot) : \phi \to \phi(x_0),\ \forall \phi \in D$ with a fixed $x_0 \in \mathbb{R}$, is a singular distribution, as shown in Gel'fand and Shilov (1964, p. 5).

In some cases, in order to extend the compact support of test functions in D to the entire real line, we consider a space larger than D as the test function space.

Definition 3.8.7 (Space S)**.** *The space S of test functions of rapid decay contain all functions ϕ on $\mathbb{R}$ that satisfy*

(i) $\phi(x)$ *is infinitely differentiable;*

(ii) $\phi(x)$ *and its derivatives of all orders, vanishes at infinity faster than any power of* $1/|x|$. *That is, for any* $p, k \geq 0$, $|x^p \phi^{(k)}(x)| \leq C_{pk}$ *where the constant* C_{pk} *only depends on* p, k, ϕ,

Similar to D and D', we may define the convergence of sequence and derivative of distributions in S and S'. One may find these in Gel'fand and Shilov (1964, p. 17) and Kanwal (1998, p. 138). Given the above definitions, one may show that S is linear, and $D \subset S \subset S' \subset D'$.

Chapter 4

Conclusions and Future Directions

4.1 Conclusions

Understanding the inherent dynamics in the data beyond the mean has become increasingly important; a large body of quantile regression literature has emerged in the past half century. When data forms a panel, modelling the heterogeneities across time and cross-sections becomes a crucial yet challenging task. Using a latent factor structure is one of the most promising ways to control the unobserved heterogeneities in the conditional mean setting, while the development in the quantile setting is still at its early stage. In this thesis, motivated from the modelling challenges arising in (i) policy evaluation and (ii) asset pricing, we adapt the latent factor structure into the existing quantile regression framework to improve methodologies in characterising the distributional structure of the data.

In this conclusion chapter, the objective and main findings of the two main chapters are briefly stated and discussed. The directions for potential future research topics is also provided in the next section.

In Chapter 2, in order to characterise the heterogeneous group-level policy effects on individuals with distinct characteristics, we proposed a random-coefficient quantile regression model, whose regression coefficient is modelled by a panel regression framework with latent factor structure. This chapter establishes the identification results for a number of treatment effect parameters, and drives the consistency and asymptotic normality for the parameter of interests. In the simulation study, we find two attractive properties of our proposed two-stage estimation procedure in practice: (i) the estimation converges within a limited number of iterations, and (ii) the estimation remains valid even when the number of individuals per group and time is relatively small compared to the total number of group and time. In the empirical application, we quantify the policy effect of the 1966 Fair Labor Standard Act on earnings of different sub-populations between 1967 and 1980 in the United States. Our results suggest that (i) the minimum wage policy helped improve the economic status of black workers and female workers, and (ii) it had a significant negative impact on the between-inequality but little effect on the within-inequality.

In Chapter 3, motivated by the time-varying co-movement structure of the large-scale stock return distributions, we propose a panel quantile functional-coefficient model with latent factors. We establish the consistency and central limit theorem for the proposed estimators. In addition, we proposed a practical bootstrap method to perform a consistency test of the regression coefficients. In the simulation study, we find our proposed estimation method has favourable finite sample properties under moderate sample size. In addition, we justify that our model outperforms its competitor in terms of both in-sample and out-of-sample estimation accuracy. Empirically, we consider a quantile FF five factor model (Fama and French, 2015), and find that: (i) the estimation and testing results advocate the conditional asset pricing model over the unconditional model across different quantile levels, (ii) FF five factors is only sufficient when there is no unobserved idiosyncratic shocks, while a more sophisticated model needs to be considered at the tails when receiving idiosyncratic shocks.

4.2 Future Directions

From our experience on working on the methodological developments and empirical applications studied in this thesis, we present some research topics that could provide directions for further research.

In Chapter 2, we impose a mean restriction on the group-level random coefficient, which can be replaced with a more natural quantile restriction. This will bring two major improvements: (i) all the unknown parameters can be estimated within one stage, utilising the individual information more efficiently, and thus improving the convergence rate for the policy parameters, and (ii) quantile treatment effects can be identified with a set of modified identification conditions (see, Athey and Imbens, 2006; Callaway and Li, 2019).

In Chapter 3, we consider the regression coefficient β as a functional form of a scalar variable v for simplicity. To adapt for more general cases, it is worthwhile to consider the extension of multivariate or even high-dimensional v. On the other hand, it is well-known that sieve or kernel methods often suffer from the so-called "curse-of-dimensionality" and are only capable of recovering functions with low-dimensional arguments. To tackle this issue, dimension reduction is essential. One efficient way of doing so is to use index models. Single-index models have been studied in the past few decades by using a general function form $h(z'\theta)$. For mean regression, Dong et al. (2016) study the single-index model for nonstationary time series, Dong et al. (2015) investigate a single-index model with fixed effects, which first establishes the consistent closed-form estimates under panel data set-up. For quantile regression, to the best of our knowledge, the efforts are made for i.i.d. (Kong and Xia, 2012; Jiang and Yu, 2023) or time-series data (Dong et al., 2023b), but not panel data. When considering a quantile single-index model with IFE, there are two major challenges. First, close-form estimates are not available due to the non-smooth quantile loss function. Second, the convergence rate of the final estimators relies on the quality of initial estimators due to the existence of non-separable quantile factor structure. Dong et al. (2023b) derive an asymptotic approximation for the quantile single-index estimators under the time series setting using the genearlised sequence approach, which shed some light on this matter. However, it is worthwhile to further explore its practicality in the panel data set-up. When the dimensionality of v is high, more flexible multiple-index models may be considered as alternative approaches, which can be estimated via the kernel method (Xia, 2008) or a more advanced deep neural network (Dong et al., 2023a).

Inference is one of the primary interests in nonlinear regression. In Chapter 3, we propose a bootstrap testing procedure for the constancy test without theoretical justification. The corresponding theory will be established in the future. In addition to the constancy test, other forms of the specification test are also worth considering. Inference problem in quantile panel data model is not new in the literature. However the existing literature either ignores the time-dependency or cross-sectional dependency (see, e.g., Kapetanios, 2008; Gonçalves and Perron, 2020; Gonçalves, 2011). Although there is no directly applicable method, existing bootstrap methods in the panel data literature may be extendable to our model. Recently, Gao et al. (2023) develop a dependent wild bootstrap (DWB) method for data with both cross-sectional and serial dependency. A modified version of the DWB method might be suitable for the quantile setting.

Bibliography

Abrevaya, J. and Dahl, C. M. (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics*, 26(4):379–397.

Amemiya, T. (1978). A note on a random coefficients model. *International Economic Review*, pages 793–796.

Ando, T. and Bai, J. (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association*, 115(529):266–279.

Ando, T., Li, K., and Lu, L. (2023). A spatial panel quantile model with unobserved heterogeneity. *Journal of Econometrics*, 232(1):191–213.

Angrist, J. D. and Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's metco program. *American Economic Review*, 94(5):1613–1634.

Arellano, M. and Bonhomme, S. (2016). Nonlinear panel data estimation via quantile regressions. *Econometrics Journal*, 19(3):61–94.

Atak, A., Montes-Rojas, G., and Olmo, J. (2023). Functional coefficient quantile regression model with time-varying loadings. *Journal of Applied Economics*, 26(1):2167151.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497.

Avramov, D. and Chordia, T. (2006). Asset pricing models and financial market anomalies. *The Review of Financial Studies*, 19(3):1001–1040.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

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

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1):436–465.

Bailey, M. J., Helgerman, T., and Stuart, B. A. (2021). Changes in the U.S. gender gap in wages in the 1960s. *AEA Papers and Proceedings*, 111:143–148.

Belloni, A., Chen, M., Madrid Padilla, O. H., and Wang, Z. (2023). High-dimensional latent panel quantile regression with an application to asset pricing. *The Annals of Statistics*, 51(1):96–121.

Billmeier, A. and Nannicini, T. (2013). Assessing economic liberalization episodes: a synthetic control approach. *Review of Economics and Statistics*, 95(3):983–1001.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2006). What mean impacts miss: distributional effects of welfare reform experiments. *American Economic Review*, 96(4):988–1012.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Borjas, G. J. and Sueyoshi, G. T. (1994). A two-stage estimator for probit models with structural group effects. *Journal of Econometrics*, 64(1-2):165–182.

Bradley, B. O. and Taqqu, M. S. (2003). Financial risk and heavy tails. In *Handbook of Heavy Tailed Distributions in Finance*, pages 35–103. Elsevier.

Brown, B. M. (1971). Martingale central limit theorems. *The Annals of Mathematical Statistics*, pages 59–66.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963–1987: application of quantile regression. *Econometrica*, 62(2):405–458.

Cai, Z., Fang, Y., and Xu, Q. (2022). Testing capital asset pricing models using functional-coefficient panel data models with cross-sectional dependence. *Journal of Econometrics*, 227(1):114–133.

Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, 167(2):413–425.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103(484):1595–1608.

Callaway, B. and Li, T. (2019). Quantile treatment effects in difference in differences models with panel data. *Quantitative Economics*, 10(4):1579–1618.

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386.

Card, D. and Krueger, A. B. (1992). School quality and black-white relative earnings: A direct assessment. *The quarterly journal of Economics*, 107(1):151–200.

Casas, I., Gao, J., Peng, B., and Xie, S. (2021). Time-varying income elasticities of healthcare expenditure for the oecd and eurozone. *Journal of Applied Econometrics*, 36(3):328–345.

Celebi, K. and Hönig, M. (2019). The impact of macroeconomic factors on the german stock market: Evidence for the crisis, pre-and post-crisis periods. *International Journal of Financial Studies*, 7(2):18.

Chen, L., Dolado, J. J., and Gonzalo, J. (2021). Quantile factor models. *Econometrica*, 89(2):875–910.

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261.

Chetverikov, D., Larsen, B., and Palmer, C. (2016). Iv quantile regression for group-level treatments, with an application to the distributional effects of trade. *Econometrica*, 84(2):809–833.

Connor, G., Hagmann, M., and Linton, O. (2012). Efficient semiparametric estimation of the fama–french model and extensions. *Econometrica*, 80(2):713–754.

Cox, D. R. (1984). Interaction. *International Statistical Review*, 52(1):1–24.

Derenoncourt, E. and Montialoux, C. (2021). Minimum wages and racial inequality. *The Quarterly Journal of Economics*, 136(1):169–228.

Djebbari, H. and Smith, J. (2008). Heterogeneous impacts in progresa. *Journal of Econometrics*, 145(1-2):64–80.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The annals of statistics*, pages 267–277.

Dong, C., Gao, J., and Peng, B. (2015). Semiparametric single-index panel data models with cross-sectional dependence. *Journal of Econometrics*, 188(1):301–312.

Dong, C., Gao, J., and Peng, B. (2021). Varying–coefficient panel data models with nonstationarity and partially observed factor structure. *Journal of Business & Economic Statistics*, 39(3):700–711.

Dong, C., Gao, J., Peng, B., and Yan, Y. (2023a). Estimation of semiparametric multi-index models using deep neural networks. *arXiv preprint arXiv:2311.02789*.

Dong, C., Gao, J., and TjØstheim, D. (2016). Estimation for single-index and partially linear single-index integrated models. *The Annals of Statistics*, 44(1):425–453.

Dong, C., Gao, J., Tu, Y., and Peng, B. (2023b). Robust M-estimation for additive single-index cointegrating time series models. *arXiv preprint arXiv:2301.06631*.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, 98(4):995–999.

Fernandes, M., Guerre, E., and Horta, E. (2021). Smoothing quantile regressions. *Journal of Business & Economic Statistics*, 39(1):338–357.

Ferson, W. E. and Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4):1325–1360.

Ferson, W. E. and Schadt, R. W. (1996). Measuring fund strategy and performance in changing economic conditions. *The Journal of Finance*, 51(2):425–461.

Freeman, R. B., Gordon, R., Bell, D., and Hall, R. E. (1973). Changes in the labor market for black americans, 1948-72. *Brookings Papers on Economic Activity*, 1973(1):67–131.

Gagliardini, P., Ossola, E., and Scaillet, O. (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica*, 84(3):985–1046.

Gagliardini, P., Ossola, E., and Scaillet, O. (2020). Estimation of large dimensional conditional factor models in finance. *Handbook of Econometrics*, 7:219–282.

Galvao, A. F., Juhl, T., Montes-Rojas, G., and Olmo, J. (2018). Testing slope homogeneity in quantile regression panel data with an application to the cross-section of stock returns. *Journal of Financial Econometrics*, 16(2):211–243.

Galvao, A. F. and Kato, K. (2017). Quantile regression methods for longitudinal data. In *Handbook of Quantile Regression*, pages 363–380. Chapman and Hall, CRC.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. CRC Press.

Gao, J., Peng, B., and Yan, Y. (2023). Higher-order expansions and inference for panel data models. Taylor & Francis.

Gel'fand, I. M. and Shilov, G. E. (1964). *Generalized Functions*. Academic Press.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.

Gonçalves, S. (2011). The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory*, 27(5):1048–1082.

Gonçalves, S. and Perron, B. (2020). Bootstrapping factor models with cross sectional dependence. *Journal of Econometrics*, 218(2):476–495.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

Harding, M. and Lamarche, C. (2014). Estimating and testing a quantile regression model with interactive effects. *Journal of Econometrics*, 178(1):101–113.

Hausman, J. A. and Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, 49(6):1377–1398.

Heckman, J. J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4):673–748.

Hsiao, C. (1974). Statistical inference for a model with both random cross-sectional and time effects. *International Economic Review*, 15(1):12–30.

Hsiao, C. (1975). Some estimation methods for a random coefficient model. *Econometrica: Journal of the Econometric Society*, 43(2):305–325.

Hsiao, C., Shi, Z., and Zhou, Q. (2022). Transformed estimation for panel interactive effects models. *Journal of Business & Economic Statistics*, 40(4):1831–1848.

Hsiao, C., Steve Ching, H., and Ki Wan, S. (2012). A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics*, 27(5):705–740.

Imbens, G. W. (2007). Nonadditive models with endogenous regressors. In *Advances in Economics and Econometrics*, pages 17–46. Cambridge University Press.

Ishihara, T. (2022). Panel data quantile regression for treatment effect models. *Journal of Business & Economic Statistics*, 41(3):720–736.

Jagannathan, R. and Wang, Z. (1996). The conditional capm and the cross–section of expected returns. *The Journal of Finance*, 51(1):3–53.

Jiang, B., Yang, Y., Gao, J., and Hsiao, C. (2021a). Recursive estimation in large panel data models: theory and practice. *Journal of Econometrics*, 224(3):439–465.

Jiang, B., Yang, Y., Gao, J., and Hsiao, C. (2021b). Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics*, 224(2):439–465.

Jiang, R. and Yu, K. (2023). No-crossing single-index quantile regression curve estimation. *Journal of Business & Economic Statistics*, 41(2):309–320.

Kanwal, R. P. (1998). *Generalized Functions Theory and Technique: Theory and Technique*. Springer Science & Business Media.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *The Econometrics Journal*, 11(2):377–395.

Kato, K., Galvao Jr, A. F., and Montes-Rojas, G. V. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170(1):76–91.

Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The Quarterly Journal of Economics*, 107(1):35–78.

Kim, D. and Oka, T. (2014). Divorce law reforms and divorce rates in the USA: an interactive fixed-effects approach. *Journal of Applied Econometrics*, 29(2):231–245.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics*, 35(1):92–108.

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

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics*, 9(1):155–176.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Kogan, L. and Papanikolaou, D. (2013). Firm characteristics and stock returns: the role of investment-specific shocks. *The Review of Financial Studies*, 26(11):2718–2759.

Kogan, L. and Papanikolaou, D. (2014). Growth opportunities, technology shocks, and asset prices. *The Journal of Finance*, 69(2):675–718.

Kong, E. and Xia, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory*, 28(4):730–768.

Lee, D. S. (1999). Wage inequality in the United States during the 1980s: rising dispersion or falling minimum wage? *The Quarterly Journal of Economics*, 114(3):977–1023.

Liu, F., Gao, J., and Yang, Y. (2018). Nonparametric estimation in panel data models with heterogeneity and time-varyingness. *Available at SSRN 3214046*.

Ma, L. and Pohlman, L. (2008). Return forecasts and optimal portfolio construction: a quantile regression approach. *The European Journal of Finance*, 14(5):409–425.

Ma, S., Linton, O., and Gao, J. (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics*, 222(1):295–323.

McQueen, G. and Roley, V. V. (1993). Stock prices, news, and business conditions. *The Review of Financial Studies*, 6(3):683–707.

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

Mundlak, Y. (1961). Empirical production function free of management bias. *Journal of Farm Economics*, 43(1):44–56.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168.

Oka, T. and Yamada, K. (2023). Heterogeneous impact of the minimum wage: implications for changes in between-and within-group inequality. *Journal of Human Resources*, 58(1):335–362.

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

Petkova, R. and Zhang, L. (2005). Is value riskier than growth? *Journal of Financial Economics*, 78(1):187–202.

Phillips, P. C. (1995). Robust nonstationary regression. *Econometric Theory*, 11(5):912–951.

Roll, R. and Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance*, 35(5):1073–1103.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

Sagner, A. (2020). High dimensional quantile factor analysis. Working papers central bank of chile, Central Bank of Chile.

Sasaki, Y. (2015). What do quantile regressions identify for general structural functions? *Econometric Theory*, 31(5):1102–1116.

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, 45(1-2):99–120.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Smith, J. P. and Welch, F. (1984). Affirmative action and labor markets. *Journal of Labor Economics*, 2(2):269–301.

Smith, J. P. and Welch, F. R. (1977). Black–white male wage ratios: 1960-70. *The American Economic Review*, 67(3):323–338.

Su, L. and Hoshino, T. (2016). Sieve instrumental variable quantile regression estimation of functional coefficient models. *Journal of Econometrics*, 191(1):231–254.

Swamy, P. A. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38(2):311–323.

Wang, H. J., Zhu, Z., and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *The Annals of Statistics*, 37(6B):3841–3866.

Wang, K. Q. (2003). Asset pricing with conditioning information: a new test. *The Journal of Finance*, 58(1):161–196.

Wüthrich, K. (2020). A comparison of two quantile models with endogeneity. *Journal of Business & Economic Statistics*, 38(2):443–456.

Xia, Y. (2008). A multiple-index model and dimension reduction. *Journal of the American Statistical Association*, 103(484):1631–1640.

Yu, K. and Jones, M. (1998). Local linear quantile regression. *Journal of the American statistical Association*, 93(441):228–237.

Zheng, Q., Peng, L., and He, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Annals of Statistics*, 43(5):2225.

Technical Supplement

This supplementary material contains the technical details required for the theoretical development of Chapters 2 and 3. Specifically, all the preliminary lemmas required in Section 2.8.2 are established and proved in Section S.1, while the preliminary lemmas required in Section 3.8.2 are given in Section S.2.

S.1 Technical Details for Proofs in Chapter 2

This section contains the technical details of some theoretical results derived in Chapter 2. Section S.1.1 establishes the intermediate results required in the proofs in Section 2.8.2. Section S.1.2 presents the preliminary lemmas and the proofs that are required in Section S.1.1.

S.1.1 Intermediate Results

Lemma S.1.1. *Under Assumption 2.2.1, 2.2.2 and 2.4.3, for fixed $u \in \mathcal{U}$, and all $(s,t) \in \{1,...,S\} \times \{1,...,T\}$,*

$$\sup_{s,t} \|\widehat{\alpha}_{st}(u) - \alpha_{st}(u)\| = O_P\left((ST)^{-3/4}\right).$$

Proof. To prove the claim, it is enough to show that, for some $c > 0$, there exists a fixed and sufficiently large $M > 0$, such that

$$\mathbb{P}\left(\sup_{s,t} \|\widehat{\alpha}_{st}(u) - \alpha_{st}(u)\| > \frac{M}{(ST)^{3/4}}\right) \to 0.$$

Under Assumption 2.4.3 that $(ST)^{3/4}(\ln(N_{\min})/N_{\min})^{1/2} \leq C_M$, we choose $M > \left(2C_M^2/(3c)\right)^{1/2}$, where c is the constant in Lemma S.1.6 independent of $S, T, N_{\min}$. Then, we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{s,t} \|\widehat{\alpha}_{st}(u) - \alpha_{st}(u)\| > \frac{M}{(ST)^{3/4}}\right) &\leq \sum_{s=1}^{S}\sum_{t=1}^{T} \mathbb{P}\left(\|\widehat{\alpha}_{st}(u) - \alpha_{st}(u)\| > \frac{M}{(ST)^{3/4}}\right) \\
&\leq \sum_{s=1}^{S}\sum_{t=1}^{T} C e^{-cM^2 N_{st}/(ST)^{3/2}} \\
&\leq CST e^{-cM^2 N_{\min}/(ST)^{3/2}} \\
&\leq C \cdot \left(\frac{C_M^2 N_{\min}}{\ln(N_{\min})}\right)^{2/3} N_{\min}^{-cM^2/C_M^2} \to 0,
\end{aligned}$$

where the second inequality follows from the non-asymptotic upper bound given in Lemma S.1.6, the last inequality follows from Assumption 2.4.3, which converges to 0 since $2/3 - cM^2/C_M^2 < 0$.

■

Lemma S.1.2. *Under Assumption 2.4.1 and 2.4.2, we have the following estimations, for any fixed $u \in \mathcal{U}$, $k = 1, ..., K$ and $j = 1, ..., J$,*

$$\left\| \frac{1}{ST} \sum_{s=1}^{S} \eta_{js}(u) \eta_{js}'(u) \right\| = O_P\left(\max\left(\frac{1}{\sqrt{S}}, \frac{1}{\sqrt{T}} \right) \right), \tag{S.1}$$

$$\left\| \frac{1}{ST} \sum_{s=1}^{S} x_{s(k)} \eta_{js}'(u) \right\| = O_P\left(\frac{1}{\sqrt{ST}} \right), \tag{S.2}$$

$$\left\| \frac{1}{ST} \sum_{s=1}^{S} F_j(u) \lambda_{js}(u) \eta_{js}'(u) \right\| = O_P\left(\frac{1}{\sqrt{S}} \right). \tag{S.3}$$

Proof. Since u and j are fixed, we suppress u and j throughout the following proof. For (S.1),

$$\begin{aligned}
\mathbb{E}\left[\left\| \frac{1}{ST} \sum_{s=1}^{S} \eta_s \eta_s' \right\|^2 \right] &\leq \mathbb{E}\left[\frac{1}{S^2T^2} \sum_{s,g=1}^{S} \sum_{t,l=1}^{T} \eta_{st}\eta_{gt}\eta_{sl}\eta_{gl} \right] \\
=& \frac{1}{S^2T^2} \sum_{s=1}^{S} \sum_{t,l=1}^{T} \mathbb{E}\left[\eta_{st}^2 \eta_{sl}^2 \right] + \frac{1}{S^2T^2} \sum_{s\neq g}^{S} \sum_{t=1}^{T} \mathbb{E}\left[\eta_{st}^2 \eta_{gt}^2 \right] + \frac{1}{S^2T^2} \sum_{s\neq g}^{S} \sum_{t\neq l}^{T} \mathbb{E}\left[\eta_{st}\eta_{gt}\eta_{sl}\eta_{gl} \right] \\
=& O\left(\frac{1}{S} \right) + O\left(\frac{1}{T} \right) + O\left(\frac{1}{ST} \right) = O\left(\max\left(\frac{1}{S}, \frac{1}{T} \right) \right),
\end{aligned}$$

where the second last equality is obtained using Assumption 2.4.1.(iv) and the fact that

$$\begin{aligned}
&\left| \frac{1}{S^2T^2} \sum_{s\neq g}^{S} \sum_{t\neq l}^{T} \mathbb{E}\left[\eta_{st}\eta_{gt}\eta_{sl}\eta_{gl} \right] \right| \\
&\leq \frac{1}{S^2T^2} \sum_{s\neq g}^{S} \sum_{t\neq l}^{T} \left| \mathbb{E}\left[\eta_{st}\eta_{gt}\eta_{sl}\eta_{gl} \right] - \mathbb{E}[\eta_{st}]\mathbb{E}[\eta_{gt}]\mathbb{E}[\eta_{sl}]\mathbb{E}[\eta_{gl}] \right| \\
&\leq C \sum_{s\neq g}^{S} \sum_{t\neq l}^{T} a_{sg}(|t-l|)^{\frac{\delta}{\delta+4}} \left(\mathbb{E}\left[|\eta_{st}|^{\delta+4} \right] \right)^{\frac{1}{\delta+4}} \left(\mathbb{E}\left[|\eta_{gt}|^{\delta+4} \right] \right)^{\frac{1}{\delta+4}} \\
&\qquad \cdot \left(\mathbb{E}\left[|\eta_{sl}|^{\delta+4} \right] \right)^{\frac{1}{\delta+4}} \left(\mathbb{E}\left[|\eta_{gl}|^{\delta+4} \right] \right)^{\frac{1}{\delta+4}} \\
&\leq C \sum_{s\neq g}^{S} \sum_{t\neq l}^{T} \cdot a_{sg}(|t-l|)^{\frac{\delta}{\delta+4}} = O\left(\frac{1}{ST} \right),
\end{aligned}$$

under Assumption 2.4.2.(i) and (ii). Therefore,

$$\left\| \frac{1}{ST} \sum_{s=1}^{S} \eta_s \eta_s' \right\| = O_P\left(\max\left(\frac{1}{\sqrt{S}}, \frac{1}{\sqrt{T}} \right) \right).$$

For (S.2),

$$\mathbb{E}\left[\left\|\frac{1}{ST}\sum_{s=1}^{S}x_{s(k)}\eta_s'\right\|^2\right] \leq \frac{1}{S^2T^2}\sum_{s,g=1}^{S}\sum_{t,l=1}^{T}\mathbb{E}\left[\eta_{sl}x_{st(k)}x_{gt(k)}\eta_{gl}\right]$$

$$=\frac{1}{S^2T^2}\sum_{s=1}^{S}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{st}^2x_{st(k)}^2\right] + \frac{1}{S^2T^2}\sum_{s\neq g}^{S}\sum_{t\neq l}^{T}\mathbb{E}\left[\eta_{sl}x_{st(k)}x_{gt(k)}\eta_{gl}\right]$$

$$+ \frac{1}{S^2T^2}\sum_{s=1}^{S}\sum_{t\neq l}^{T}\mathbb{E}\left[\eta_{sl}x_{st(k)}x_{st(k)}\eta_{sl}\right] + \frac{1}{S^2T^2}\sum_{s\neq g}^{S}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{st}x_{st(k)}x_{gt(k)}\eta_{gt}\right]$$

$$=O\left(\frac{1}{ST}\right), \tag{S.4}$$

where the last equality comes from the following facts that the first term in the first equality is

$$\frac{1}{S^2T^2}\sum_{s=1}^{S}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{st}^2x_{st(k)}^2\right] \leq \frac{1}{S^2T^2}\sum_{s=1}^{S}\sum_{t=1}^{T}\left(\mathbb{E}\left[\eta_{st}^4\right]\right)^{1/2}\cdot\left(\mathbb{E}\left[x_{st(k)}^4\right]\right)^{1/2} = O\left(\frac{1}{ST}\right),$$

under Assumption 2.4.1.(i) and (iv), and the second term in the first equality of (S.4) is

$$\frac{1}{S^2T^2}\sum_{s\neq g}^{S}\sum_{t\neq l}^{T}\mathbb{E}\left[\eta_{sl}x_{st(k)}x_{gt(k)}\eta_{gl}\right]$$

$$=\frac{1}{S^2T^2}\sum_{s\neq g}^{S}\sum_{t\neq l}^{T}\left(\mathbb{E}\left[\eta_{st}x_{st(k)}x_{gt(k)}\eta_{gl}\right] - \mathbb{E}\left[\eta_{sl}x_{st(k)}\right]\cdot\mathbb{E}\left[x_{gt(k)}\eta_{gl}\right]\right)$$

$$\leq\frac{1}{S^2T^2}\sum_{s\neq g}^{S}\sum_{t\neq l}^{T}10a_{sg}(|t-l|)^{\frac{\delta}{\delta+2}}\left(\mathbb{E}\left[|\eta_{sl}x_{st(k)}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}\left(\mathbb{E}\left[|x_{gt(k)}\eta_{gl}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}$$

$$=O\left(\frac{1}{ST}\right),$$

where the first equality holds since $\mathbb{E}\left[\eta_{st}x_{st(k)}\right] = \mathbb{E}[x_{st(k)}]\mathbb{E}(\eta_{st}|x_{st(k)})] = 0$ under Assumption 2.4.1.(iv), the last inequality holds due to Lemma S.1.7 and Assumption 2.4.1.(iv), and the last equality follows from Assumption 2.4.2(i) and 2.4.2(ii). The third term in the first equality of (S.4) is

$$\frac{1}{S^2T^2}\sum_{s=1}^{S}\sum_{t\neq l}^{T}\mathbb{E}\left[\eta_{st}x_{st(k)}x_{sl(k)}\eta_{sl}\right]$$

$$=\frac{1}{S^2T^2}\sum_{s=1}^{S}\sum_{t\neq l}^{T}\left(\mathbb{E}\left[\eta_{st}x_{st(k)}x_{sl(k)}\eta_{sl}\right] - \mathbb{E}\left[\eta_{st}x_{st(k)}\right]\cdot\mathbb{E}\left[x_{sl(k)}\eta_{sl}\right]\right)$$

$$\leq\frac{1}{S^2T^2}\sum_{s=1}^{S}\sum_{t\neq l}^{T}10a(|t-l|)^{\frac{\delta}{\delta+2}}\left(\mathbb{E}\left[|\eta_{st}x_{st(k)}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}\left(\mathbb{E}\left[|x_{sl(k)}\eta_{sl}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}$$

$$=O\left(\frac{1}{ST}\right),$$

where the last equality follows from Assumption 2.4.2.(i). And lastly, the fourth term in the first equality of (S.4) is

$$\begin{aligned}
&\frac{1}{S^2T^2}\sum_{s\neq g}^{S}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{st}x_{st(k)}x_{gl(k)}\eta_{gl}\right]\\
&=\frac{1}{S^2T^2}\sum_{s\neq g}^{S}\sum_{t=1}^{T}\left(\mathbb{E}\left[\eta_{st}x_{st(k)}x_{gt(k)}\eta_{gt}\right]-\mathbb{E}\left[\eta_{st}x_{st(k)}\right]\cdot\mathbb{E}\left[x_{gt(k)}\eta_{gt}\right]\right)\\
&\leq\frac{1}{S^2T^2}\sum_{s\neq g}^{S}\sum_{t=1}^{T}10a_{sg}(0)^{\frac{\delta}{\delta+2}}\left(\mathbb{E}\left[|\eta_{st}x_{st(k)}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}\left(\mathbb{E}\left[|x_{gt(k)}\eta_{gt}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}\\
&=O\left(\frac{1}{ST}\right)
\end{aligned}$$

where the last equality follows from Assumption 2.4.2(i) and (ii).

Therefore, we have

$$\left\|\frac{1}{ST}\sum_{s=1}^{S}x_{s(k)}\eta_s'\right\|=O\left(\frac{1}{ST}\right).$$

For (S.3), we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{ST}\sum_{s=1}^{S}F\lambda_s\eta_s'\right\|^2\right]\leq\frac{1}{S^2T^2}\sum_{s,g=1}^{S}\mathbb{E}\left[\text{tr}\left(\eta_s\lambda_s'F'F\lambda_g\eta_g'\right)\right]\\
&=\frac{1}{S^2T}\sum_{i=1}^{r}\sum_{s=1}^{S}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{st}^2\lambda_{si}^2\right]+\frac{1}{S^2T}\sum_{i=1}^{r}\sum_{s\neq g}^{S}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{gt}\eta_{st}\lambda_{si}\lambda_{gi}\right]=O\left(\frac{1}{S}\right),
\end{aligned}$$

where the last equality holds due to the facts that

$$\frac{1}{S^2T}\sum_{i=1}^{r}\sum_{s=1}^{S}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{st}^2\lambda_{si}^2\right]\leq\frac{1}{S^2T}\sum_{i=1}^{r}\sum_{s=1}^{S}\sum_{t=1}^{T}\left(\mathbb{E}\left[\eta_{st}^4\right]\right)^{1/2}\left(\mathbb{E}\left[\lambda_{si}^4\right]\right)^{1/2}=O\left(\frac{1}{S}\right),$$

under Assumption 2.4.1, and

$$\begin{aligned}
&\frac{1}{S^2T}\sum_{i=1}^{r}\sum_{s\neq g}^{S}\sum_{t=1}^{T}\mathbb{E}\left[\eta_{gt}\eta_{st}\lambda_{si}\lambda_{gi}\right]\\
&=\frac{1}{S^2T}\sum_{i=1}^{r}\sum_{s\neq g}^{S}\sum_{t=1}^{T}\left(\mathbb{E}\left[\eta_{st}\lambda_{si}\eta_{gt}\lambda_{gi}\right]-\mathbb{E}\left[\eta_{st}\lambda_{si}\right]\cdot\mathbb{E}\left[\eta_{gt}\lambda_{gi}\right]\right)\\
&\leq\frac{1}{S^2T}\sum_{i=1}^{r}\sum_{s\neq g}^{S}\sum_{t=1}^{T}10a_{sg}(0)^{\frac{\delta}{\delta+2}}\left(\mathbb{E}\left[|\eta_{st}\lambda_{si}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}\left(\mathbb{E}\left[|\eta_{gt}\lambda_{gi}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}=O\left(\frac{1}{S}\right),
\end{aligned}$$

under Assumption 2.4.2.(i) and (ii).

Therefore,

$$\left\| \frac{1}{ST} \sum_{s=1}^{S} F \lambda_s \eta_s' \right\| = O_P\left(\frac{1}{\sqrt{S}} \right).$$

■

Lemma S.1.3. *Under assumptions of Theorem 2.4.1, we have for any fixed $u \in \mathcal{U}$, $j = 1, ..., J$ and $m \geq 1$,*

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \left(\widehat{F}_j^{(m)}(u) H_j^{(m)}(u) - F_j(u) \right) \\
&= \frac{1}{\sqrt{T}} \left[I_1^{(m)}(u) + I_2^{(m)}(u) + I_3^{(m)}(u) \right] \widehat{F}_j^{(m)}(u) \left(K_j^{(m)}(u) \right)^{-1} + O_P(\delta_{ST}^{-1}) \\
&\quad + o_P\left(\frac{1}{\sqrt{ST}} \right) \cdot \left(\left\| \widehat{\beta}_j^{(m-1)}(u) - \beta_j(u) \right\| + \left| \sum_{t=T_0}^{T} \widehat{\delta}_{jt}^{(m-1)}(u) - \delta_{jt}(u) \right| \right),
\end{aligned}$$

where

$$\begin{aligned}
I_1^{(m)} := \frac{1}{ST} \sum_{s=1}^{S} \Big[& D_s(\widehat{\delta}_j^{(m-1)}(u) - \delta_j(u))(\widehat{\delta}_j^{(m-1)}(u) - \delta_j(u))' D_s' \\
& + D_s(\widehat{\delta}_j^{(m-1)}(u) - \delta_j(u))(\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u))' X_s' \\
& + X_s(\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u))(\widehat{\delta}_j^{(m-1)}(u) - \delta_j(u))' D_s' \\
& + X_s(\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u))(\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u))' X_s' \Big],
\end{aligned}$$

$$\begin{aligned}
I_2^{(m)}(u) := \frac{1}{ST} \sum_{s=1}^{S} \Big[& D_s \left(\delta_j(u) - \widehat{\delta}_j^{(m-1)}(u) \right) \lambda_{js}(u)' F_j(u)' \\
& + X_s \left(\beta_j(u) - \widehat{\beta}_j^{(m-1)}(u) \right) \lambda_{js}(u)' F_j(u)' \\
& + F_j(u) \lambda_{js}(u) \left(\delta_j(u) - \widehat{\delta}_j^{(m-1)}(u) \right)' D_s' \\
& + F_j(u) \lambda_{js}(u) \left(\beta_j(u) - \widehat{\beta}_j^{(m-1)}(u) \right)' X_s' \Big],
\end{aligned}$$

$$\begin{aligned}
I_3^{(m)}(u) := \frac{1}{ST} \sum_{s=1}^{S} \Big[& D_s \left(\delta_j(u) - \widehat{\delta}_j^{(m-1)}(u) \right) \eta_{js}(u)' + X_s \left(\beta_j(u) - \widehat{\beta}_j^{(m-1)}(u) \right) \eta_{js}(u)' \\
& + \eta_s \left(\delta - \widehat{\delta}^{(m-1)} \right)' D_s' + \eta_s \left(\beta - \widehat{\beta}^{(m-1)} \right)' X_s' \Big],
\end{aligned}$$

and

$$K_j^{(m)}(u) := \left(\frac{\Lambda_j(u)' \Lambda_j(u)}{S} \right) \left(\frac{F_j(u)' \widehat{F}_j^{(m)}(u)}{T} \right), \quad H_j^{(m)}(u) := \widehat{\Upsilon}_j^{(m)}(u) \left(K_j^{(m)}(u) \right)^{-1},$$

in which $\widehat{\Upsilon}_j^{(m)}(u)$ is the $r \times r$ diagonal matrix with diagonal elements being the r largest eigenvalues of $\widehat{L}_j\big(\widehat{\delta}_j^{(m-1)}(u), \widehat{\beta}_j^{(m-1)}(u)\big)$, defined in (2.6), in descending order.

Proof. Since u and j are fixed in Lemma S.1.3, we suppress u and j throughout the following proof.

It is easy to show that a square matrix A is invertible if $\det(AA') = \det(A)\det(A') \neq 0$, and thus, $T^{-1}F'\widehat{F}^{(m)}$ is invertible and bounded under Assumption 2.4.5.(i). Additionally, $S^{-1}\Lambda'\Lambda$ is invertible and bounded under Assumption 2.4.1.(ii), and thus, $K^{(m)}$ is invertible, and

$$\left\| \left(K^{(m)}\right)^{-1} \right\| \leq \left\| \left(\frac{\Lambda'\Lambda}{S}\right)^{-1} \right\| \cdot \left\| \left(\frac{F'\widehat{F}^{(m)}}{T}\right)^{-1} \right\| = O_P(1). \tag{S.5}$$

Then, given the definition of $H^{(m)}$, to obtain the asymptotic expression of $\widehat{F}^{(m)}H^{(m)} - F$, it is enough to analyze $\widehat{F}^{(m)}\widehat{\Upsilon}^{(m)} - FK^{(m)}$, and then $\widehat{F}^{(m)}H^{(m)} - F = \big(\widehat{F}^{(m)}\widehat{\Upsilon}^{(m)} - FK^{(m)}\big)\big(K^{(m)}\big)^{-1}$.

With $(\widehat{\delta}^{(m-1)}, \widehat{\beta}^{(m-1)})$ obtained at the $(m-1)^{\text{th}}$ step, we have the estimator $\big(\widehat{F}^{(m)}, \widehat{\Lambda}^{(m)}\big)$ via the PCA. Thus, $\widehat{F}^{(m)}$ statisfies

$$\widehat{F}^{(m)}\widehat{\Upsilon}^{(m)} = \widehat{L}^{(m)}\widehat{F}^{(m)}. \tag{S.6}$$

Moreover, $\widehat{L}^{(m)} = (ST)^{-1}\sum_{s=1}^{S}\big(\widehat{A}_s - X_s\widehat{\beta}^{(m-1)}\big)\big(\widehat{A}_s - X_s\widehat{\beta}^{(m-1)}\big)'$ can be deposed into eight terms by substituting $\widehat{A}_s$ in $\widehat{L}^{(m)}$ with $\widehat{A}_s = X_s\beta + F\lambda_s + \eta_s + (\widehat{A}_s - A_s)$ as follows:

$$\begin{aligned}
I_1^{(m)} = \frac{1}{ST}\sum_{s=1}^{S}\Big[&D_s(\widehat{\delta}^{(m-1)} - \delta)(\widehat{\delta}^{(m-1)} - \delta)'D_s' + D_s(\widehat{\delta}^{(m-1)} - \delta)(\widehat{\beta}^{(m-1)} - \beta)'X_s' \\
&+ X_s(\widehat{\beta}^{(m-1)} - \beta)(\widehat{\delta}^{(m-1)} - \delta)'D_s' + X_s(\widehat{\beta}^{(m-1)} - \beta)(\widehat{\beta}^{(m-1)} - \beta)'X_s'\Big],
\end{aligned}$$

$$\begin{aligned}
I_2^{(m)} = \frac{1}{ST}\sum_{s=1}^{S}\Big[&D_s\left(\delta - \widehat{\delta}^{(m-1)}\right)\lambda_s'F' + X_s\left(\beta - \widehat{\beta}^{(m-1)}\right)\lambda_s'F' \\
&+ F\lambda_s\left(\delta - \widehat{\delta}^{(m-1)}\right)'D_s' + F\lambda_s\left(\beta - \widehat{\beta}^{(m-1)}\right)'X_s'\Big],
\end{aligned}$$

$$\begin{aligned}
I_3^{(m)} = \frac{1}{ST}\sum_{s=1}^{S}\Big[&D_s\left(\delta - \widehat{\delta}^{(m-1)}\right)\eta_s' + X_s\left(\beta - \widehat{\beta}^{(m-1)}\right)\eta_s' \\
&+ \eta_s\left(\delta - \widehat{\delta}^{(m-1)}\right)'D_s' + \eta_s\left(\beta - \widehat{\beta}^{(m-1)}\right)'X_s'\Big],
\end{aligned}$$

$$I_4^{(m)} = \frac{1}{ST}\sum_{s=1}^{S}\Big[F\lambda_s\eta_s' + \eta_s\lambda_s'F' + \eta_s\eta_s'\Big],$$

$$\begin{aligned}
I_5^{(m)} = \frac{1}{ST}\sum_{s=1}^{S}\Big[&\left(\widehat{A}_s - A_s\right)\left(\delta - \widehat{\delta}^{(m-1)}\right)'D_s' + \left(\widehat{A}_s - A_s\right)\left(\beta - \widehat{\beta}^{(m-1)}\right)'X_s' \\
&+ D_s\left(\delta - \widehat{\delta}^{(m-1)}\right)\left(\widehat{A}_s - A_s\right)' + X_s\left(\beta - \widehat{\beta}^{(m-1)}\right)\left(\widehat{A}_s - A_s\right)'\Big],
\end{aligned}$$

$$I_6^{(m)} = \frac{1}{ST}\sum_{s=1}^{S}\left[\left(\widehat{A}_s - A_s\right)\eta_s' + \eta_s\left(\widehat{A}_s - A_s\right)' + \left(\widehat{A}_s - A_s\right)\lambda_s' F' + F\lambda_s\left(\widehat{A}_s - A_s\right)'\right],$$
$$I_7^{(m)} = \frac{1}{ST}\sum_{s=1}^{S}\left(\widehat{A}_s - A_s\right)\left(\widehat{A}_s - A_s\right)',$$
$$I_8^{(m)} = \frac{1}{ST}\sum_{s=1}^{S} F\lambda_s\lambda_s' F'.$$

Then, since $I_8^{(m)}\widehat{F}^{(m)} = FK^{(m)}$ according to the definition of $K^{(m)}$, we have

$$\widehat{F}^{(m)}\widehat{\Upsilon}^{(m)} - FK^{(m)} = \sum_{h=1}^{7} I_h^{(m)}\widehat{F}^{(m)}, \tag{S.7}$$

where $I_1^{(m)}$, $I_2^{(m)}$ and $I_3^{(m)}$ are the leading terms, while the rest of the terms are negligible in the limit. To show this, we start by showing the rate of covergence for $I_1^{(m)}$. By triangle and Cauchy-Schwartz inequalities, we have

$$\begin{aligned}
\|I_1^{(m)}\| \leq & \frac{1}{ST}\Bigg[\Big\|\sum_{s=1}^{S} D_s(\widehat{\delta}^{(m-1)} - \delta)(\widehat{\delta}^{(m-1)} - \delta)' D_s'\Big\| \\
& + 2\Big\|\sum_{s=1}^{S} D_s(\widehat{\delta}^{(m-1)} - \delta)(\widehat{\beta}^{(m-1)} - \beta)' X_s'\Big\| \\
& + \Big\|\sum_{s=1}^{S} W_s(\widehat{\beta}^{(m-1)} - \beta)(\widehat{\beta}^{(m-1)} - \beta)' X_s'\Big\|\Bigg] \\
\leq & \frac{1}{ST}\Bigg[\Big\|\sum_{s=1}^{S}\sum_{t,l=T_0}^{T} d_s e_t(\widehat{\delta}_t^{(m-1)} - \delta_t)(\widehat{\delta}_l^{(m-1)} - \delta_l) e_l'\Big\| \\
& + 2\Big\|\sum_{s=1}^{S} X_s\Big\| \cdot \Big\|\sum_{t=T_0}^{T} e_t(\widehat{\delta}_t^{(m-1)} - \delta_t)\Big\| \cdot \|\widehat{\beta}^{(m-1)} - \beta\| \\
& + \sum_{s=1}^{S}\|X_s\|^2 \cdot \|\widehat{\beta}^{(m-1)} - \beta\|^2\Bigg] \\
\leq & \frac{1}{ST}\Bigg[\sum_{s=1}^{S}|d_s| \cdot \Big|\sum_{t=T_0}^{T}\widehat{\delta}_t^{(m-1)} - \delta_t\Big|^2 + \sum_{s=1}^{S}\|X_s\|^2 \cdot \|\widehat{\beta}^{(m-1)} - \beta\|^2 \\
& + 2\Big\|\sum_{s=1}^{S} X_s\Big\| \cdot \Big|\sum_{t=T_0}^{T}\widehat{\delta}_t^{(m-1)} - \delta_t\Big| \cdot \|\widehat{\beta}^{(m-1)} - \beta\|\Bigg] \\
= & O_P\left(\frac{1}{T}\right) \cdot \Big|\sum_{t=T_0}^{T}\widehat{\delta}_t^{(m-1)} - \delta_t\Big|^2 + O_P(1) \cdot \|\widehat{\beta}^{(m-1)} - \beta\|^2
\end{aligned}$$

$$+ O_P\left(\frac{1}{\sqrt{T}}\right) \cdot \left| \sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t \right| \cdot \|\widehat{\beta}^{(m-1)} - \beta\|,$$

For $I_2^{(m)}$, a similar argument yields that

$$\begin{aligned}
\|I_2^{(m)}\| \leq & \frac{1}{ST}\left[\left\| \sum_{s=1}^{S} D_s(\widehat{\delta}^{(m-1)} - \delta)\lambda_s' F' \right\| + \left\| \sum_{s=1}^{S} W_s(\widehat{\beta}^{(m-1)} - \beta)\lambda_s' F' \right\|\right] \\
\leq & \frac{1}{ST}\left[\left\| \sum_{s=1}^{S} d_s \lambda_s' F' \right\| \cdot \left| \sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t \right| + \left\| \sum_{s=1}^{S} W_s \lambda_s' F' \right\| \cdot \|\widehat{\beta}^{(m-1)} - \beta\|\right] \\
= & O_P\left(\frac{1}{\sqrt{ST}}\right) \cdot \left(\left\|\widehat{\beta}^{(m-1)} - \beta\right\| + \left| \sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)} \right| \right),
\end{aligned}$$

under Assumption 2.4.6 that $\mathbb{E}\| \sum_{s=1}^{S} W_s \lambda_s' F' \|^2 = O_P(ST)$ and $\mathbb{E}\| \sum_{s=1}^{S} d_s \lambda_s' F' \|^2 = O_P(ST)$.

For $I_3^{(m)}$ and $I_4^{(m)}$, by triangle and Cauchy-Schwartz inequalities and the property of α-mixing varaibles (see Lemma S.1.2) we have

$$\begin{aligned}
\left\|I_3^{(m)}\right\| \leq & 2\left(\sum_{k=1}^{K} \left\| \frac{1}{ST} \sum_{s=1}^{S} x_{s(k)} \eta_s' \right\|^2 \right)^{1/2} \left(\sum_{k=1}^{K} \left| \beta_k - \widehat{\beta}_k^{(m-1)} \right|^2 \right)^{1/2} \\
& + \frac{2}{ST} \left\| \sum_{s=1}^{S} d_s \eta_s' \right\| \cdot \left| \sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)} \right| \\
\leq & O_P\left(\frac{1}{\sqrt{ST}}\right) \cdot \left(\left\|\widehat{\beta}^{(m-1)} - \beta\right\| + \left| \sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)} \right| \right), \\
\left\|I_4^{(m)}\right\| \leq & 2\left\| \frac{1}{ST} \sum_{s=1}^{S} F \lambda_s \eta_s' \right\| + \left\| \frac{1}{ST} \sum_{s=1}^{S} \eta_s \eta_s' \right\| = O_P\left(\delta_{ST}^{-1}\right).
\end{aligned}$$

For $I_5^{(m)}$ to $I_7^{(m)}$, we apply Cauchy-Schwartz inequality and have

$$\begin{aligned}
\left\|I_5^{(m)}\right\| & \leq \frac{2}{ST} \sup_s \left\|\widehat{A}_s - A_s\right\| \cdot \left(\sum_{s=1}^{S} \|X_s\| \cdot \left\|\beta - \widehat{\beta}^{(m-1)}\right\| + \left\| \sum_{s=1}^{S} d_s \right\| \cdot \left| \sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)} \right| \right) \\
& = O_P\left((ST)^{-3/4}\right) \cdot \left(\left\|\widehat{\beta}^{(m-1)} - \beta\right\| + \left| \sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)} \right| \right), \\
\left\|I_6^{(m)}\right\| & \leq \frac{2}{ST} \sup_s \left\|\widehat{A}_s - A_s\right\| \cdot \sum_{s=1}^{S} \left(\|\eta_s\| + \|\lambda_s' F\| \right) = O_P\left((ST)^{-3/4}\right), \qquad \text{(S.8)} \\
\left\|I_7^{(m)}\right\| & \leq \frac{1}{ST} \sum_{s=1}^{S} \left\|\widehat{A}_s - A_s\right\|^2 = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} \|\widehat{\alpha}_{st} - \alpha_{st}\|^2 = O_P\left((ST)^{-3/2}\right),
\end{aligned}$$

since $\sup_s \|\widehat{A}_s - A_s\|^2 = \sum_{t=1}^{T} \sup_{s,t} \|\widehat{\alpha}_{st} - \alpha_{st}\|^2 = O_P(S^{-3/2} T^{-1/2})$ according to Lemma S.1.1

and $\sum_{s=1}^{S}\|X_s\| = O_P(S\sqrt{T})$, $\sum_{s=1}^{S}\|\eta_s\| = O_P(S\sqrt{T})$ and $\sum_{s=1}^{S}\|\lambda_s' F'\| = O_P(S\sqrt{T})$ under Assumption 2.4.1.

Finally, combining (S.5)-(S.7) with $\widehat{F}^{(m)} = O_P(\sqrt{T})$ and $\left(K^{(m)}\right)^{-1} = O_P(1)$, we obtain

$$\begin{aligned}\frac{1}{\sqrt{T}}\left(\widehat{F}^{(m)}H^{(m)} - F\right) &= \frac{1}{\sqrt{T}}\sum_{h=1}^{7} I_h^{(m)}\widehat{F}^{(m)}\left(K^{(m)}\right)^{-1} \\ &= \frac{1}{\sqrt{T}}\left[I_1^{(m)} + I_2^{(m)} + I_3^{(m)}\right]\widehat{F}^{(m)}\left(K^{(m)}\right)^{-1} + O_P(\delta_{ST}^{-1}) \\ &\quad + o_P\left(\frac{1}{\sqrt{ST}}\right)\cdot\left(\left\|\widehat{\beta}^{(m-1)} - \beta\right\| + \left|\sum_{t=T_0}^{T}\widehat{\delta}_t^{(m-1)} - \delta_t\right|\right),\end{aligned}$$

and furthermore,

$$\begin{aligned}\left\|\widehat{F}^{(m)}H^{(m)} - F\right\| = O_P\left(\frac{1}{\sqrt{T}}\right)\cdot\left|\sum_{t=T_0}^{T}\widehat{\delta}_t^{(m-1)} - \delta_t\right|^2 + O_P(\sqrt{T})\cdot\|\widehat{\beta}^{(m-1)} - \beta\|^2 \\ + O_P(1)\cdot\left|\sum_{t=T_0}^{T}\widehat{\delta}_t^{(m-1)} - \delta_t\right|\cdot\|\widehat{\beta}^{(m-1)} - \beta\| \\ + O_P\left(\frac{1}{\sqrt{S}}\right)\cdot\left(\left\|\widehat{\beta}^{(m-1)} - \beta\right\| + \left|\sum_{t=T_0}^{T}\delta_t - \widehat{\delta}_t^{(m-1)}\right|\right) + O_P(1). \quad \text{(S.9)}\end{aligned}$$

■

Lemma S.1.4. *Under assumptions of Theorem 2.4.2, for any fixed $u \in \mathcal{U}$, $j = 1, ..., J$ and recursive step $m \geq 1$,*

$$\begin{aligned}&\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u)\left(F_j(u) - \widehat{F}_j^{(m)}(u)H_j^{(m)}(u)\right)\lambda_{js}(u) \\ &= -\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u)\left(I_1^{(m)} + I_2^{(m)}\right)\widehat{F}_j^{(m)}(u)\left(K_j^{(m)}(u)\right)^{-1}\lambda_{js}(u) \\ &\quad + O_P(\sqrt{S})\cdot\left|\sum_{t=T_0}^{T}\widehat{\delta}_{jt}^{(m-1)}(u) - \delta_{jt}(u)\right|^2 + O_P(T\sqrt{S})\cdot\left\|\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u)\right\|^2 \\ &\quad + O_P(\sqrt{ST})\cdot\left|\sum_{t=T_0}^{T}\widehat{\delta}_{jt}^{(m-1)}(u) - \delta_{jt}(u)\right|\cdot\left\|\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u)\right\| \\ &\quad + O_P(\sqrt{T})\cdot\left(\left\|\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u)\right\| + \left|\widehat{\delta}_{jt}^{(m-1)}(u) - \delta_{jt}(u)\right|\right) + O_P(\sqrt{ST})\end{aligned}$$

where $I_1^{(m)}$, $I_2^{(m)}$, $I_3^{(m)}$, $H_j^{(m)}(u)$ and $K_j^{(m)}(u)$ are defined in Lemma S.1.3.

Proof. Since u and j are fixed, for notational simplicity, we suppress u and j throughout the following proof.

By definition of $M_{\widehat{F}}^{(m)}$, we have $\sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)} F\lambda_s = \sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)} \big(F - \widehat{F}^{(m)} H^{(m)}\big)\lambda_s$. Also, recall from the proof of Lemma S.1.3, we have $\widehat{F}^{(m)} H^{(m)} - F = \sum_{h=1}^{7} I_h^{(m)} \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1}$ with $I_h^{(m)}$, $h = 1, ..., 7$ defined above (S.7), so that we can write

$$\sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)} \big(F - \widehat{F}^{(m)} H^{(m)}\big)\lambda_s = -\sum_{h=1}^{7}\sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)} I_h^{(m)} \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1} \lambda_s$$
$$=: \sum_{h=1}^{7} J_h^{(m)}, \tag{S.10}$$

where $J_1^{(m)}$, $J_2^{(m)}$ and $J_3^{(m)}$ are the leading terms in the asymptotic expression, while the remaining terms are negligible. We first compute the norm of the leading terms. Given the rates $I_1^{(m)}$, $I_2^{(m)}$ and $I_3^{(m)}$ of given in the proof of Lemma S.1.3, by Cauchy-Swartz and Hölder's inequality, we have

$$\begin{aligned}\left\|J_1^{(m)}\right\| &\leq 2\Big(\sum_{s=1}^{S}\|X_s\|^2\Big)^{1/2} \cdot \Big(\sum_{s=1}^{S}\|\lambda_s\|^2\Big)^{1/2} \cdot \left\|M_{\widehat{F}}^{(m)}\right\| \cdot \left\|I_1^{(m)}\right\| \cdot \left\|\widehat{F}^{(m)}\right\| \cdot \left\|\left(K^{(m)}\right)^{-1}\right\| \\ &= O_P(S) \cdot \left|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t\right|^2 + O_P(ST) \cdot \|\widehat{\beta}^{(m-1)} - \beta\|^2 \\ &\quad + O_P(S\sqrt{T}) \cdot \left|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t\right| \cdot \|\widehat{\beta}^{(m-1)} - \beta\|,\end{aligned}$$

where $\|(K^{(m)})^{-1}\| = O_P(1)$ according to (S.5), $\sum_{s=1}^{S}\|X_s\|^2 = O_P(ST)$, $\sum_{s=1}^{S}\|\lambda_s\|^2 = O_P(S)$ and $\|\widehat{F}^{(m)}\| = O_P(\sqrt{T})$ under Assumption 2.4.1. Similar arguments yield that $J_2^{(m)}$ and $J_3^{(m)}$ are of rate $O_P(\sqrt{ST}) \cdot \big(\|\widehat{\beta}^{(m-1)} - \beta\| + |\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m)} - \delta_t|\big)$.

For $J_4^{(m)}$, we plug-in the expression of $I_4^{(m)}$ to obtain

$$J_4^{(m)} = -\frac{1}{ST}\sum_{s,g=1}^{S} X_s' M_{\widehat{F}}^{(m)} \Big\{F\lambda_g\eta_g' + \eta_g\lambda_g' F' + \big(\eta_g\eta_g' - \mathbb{E}[\eta_g\eta_g']\big) + \mathbb{E}[\eta_g\eta_g']\Big\} \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1} \lambda_s,$$

and below we evaluate the asymptotic bound for each of the terms expanding by terms in $I_4^{(m)}$. The first term has order $O_P(\sqrt{ST}) \cdot \|\widehat{F}^{(m)} H^{(m)} - F\|$ due to

$$\begin{aligned}&\left\|\frac{1}{ST}\sum_{s=1}^{S} M_{\widehat{F}}^{(m)} F\lambda_s\eta_s' \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1}\right\| \\ &\leq \left\|M_{\widehat{F}}^{(m)}\right\| \cdot \left\|\widehat{F}^{(m)} H^{(m)} - F\right\| \cdot \left\|\frac{1}{ST}\sum_{s=1}^{S} F\lambda_s\eta_s'\right\| \cdot \left\|\left(K^{(m)}\right)^{-1}\right\| \\ &= \left\|\widehat{F}^{(m)} H^{(m)} - F\right\| \cdot O_P\left(\frac{1}{\sqrt{S}}\right) \cdot O_P(1),\end{aligned}$$

where the first inequality follows from Cauchy-Swartz inequality and the property that $M_{\widehat{F}}^{(m)} F =$

$M_{\widehat{F}}^{(m)}\left(F - \widehat{F}^{(m)}H^{(m)}\right)$, and the second last equality follows from (S.3) in Lemma S.1.2, (S.5), and (S.9). For the second term, we plug-in the expression of $K^{(m)}$ and then apply an α-mixing argument similar to (S.2) in Lemma S.1.2 to obtain

$$\frac{1}{ST}\sum_{s,g=1}^{S} X_s' M_{\widehat{F}}^{(m)} \eta_g \lambda_g' F' \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1} \lambda_s = \sum_{g=1}^{S}\left(\frac{1}{S}\sum_{s=1}^{S} \omega_{sg} X_s' M_{\widehat{F}}\right)\eta_g = O_P\left(\sqrt{ST}\right).$$

The third term is shown to be $O_P(\sqrt{T}) \cdot (\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\|^2 + \|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\|) + O_P(\delta_{ST})$ in Lemma S.1.9 and the fourth term has order $O_P(S)$ given that $\sum_{s=1}^{S}\mathbb{E}[\eta_s\eta_s'] = O(S)$ according to Assumption 2.4.1.(ii). Collecting the terms in $J_4^{(m)}$ and using the rates of $\|\widehat{F}^{(m)}H^{(m)} - F\|$ obtained from (S.9), we have

$$\begin{aligned}
\left\|J_4^{(m)}\right\| =& O_P(\sqrt{S}) \cdot \left|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t\right|^2 + O_P(T\sqrt{S}) \cdot \left\|\widehat{\beta}^{(m-1)} - \beta\right\|^2 \\
&+ O_P(\sqrt{ST}) \cdot \left|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t\right| \cdot \left\|\widehat{\beta}^{(m-1)} - \beta\right\| \\
&+ O_P(\sqrt{T}) \cdot \left(\left\|\widehat{\beta}^{(m-1)} - \beta\right\| + \left|\sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)}\right|\right) + O_P(\sqrt{ST}).
\end{aligned}$$

For $\sum_{h=5}^{7} J_h^{(m)}$, it is easy to check that

$$\begin{aligned}
\sum_{h=5}^{7}\left\|J_h^{(m)}\right\| \leq& \left(\sum_{s=1}^{S}\|X_s\|^2\right)^{1/2} \cdot \left(\sum_{s=1}^{S}\|\lambda_s\|^2\right)^{1/2} \cdot \sum_{h=5}^{7}\left\|I_h^{(m)}\right\| \cdot \left\|\widehat{F}^{(m)}\right\| \cdot \left\|M_{\widehat{F}}^{(m)}\right\| \cdot \left\|\left(K^{(m)}\right)^{-1}\right\| \\
=& O_P\left((ST)^{1/4}\right) \cdot \left(1 + \left|\sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)}\right| + \left\|\widehat{\beta}^{(m-1)} - \beta\right\|\right),
\end{aligned}$$

given the rates of $I_h^{(m)}$ for $h = 5, 6, 7$ from (S.8). Finally, collecting the rate of convergence for $J_h^{(m)}$, $h = 1, ..., 9$, we obtain

$$\begin{aligned}
\sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)} F \lambda_s =& -\sum_{s=1}^{S} X_s' M_{\widehat{F}}^{(m)}\left(I_1^{(m)} + I_2^{(m)} + I_3^{(m)}\right)\widehat{F}^{(m)}\left(K^{(m)}\right)^{-1}\lambda_s \\
&+ O_P(\sqrt{S}) \cdot \left|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t\right|^2 + O_P(T\sqrt{S}) \cdot \left\|\widehat{\beta}^{(m-1)} - \beta\right\|^2 \\
&+ O_P(\sqrt{ST}) \cdot \left|\sum_{t=T_0}^{T} \widehat{\delta}_t^{(m-1)} - \delta_t\right| \cdot \left\|\widehat{\beta}^{(m-1)} - \beta\right\| \\
&+ O_P(\sqrt{T}) \cdot \left(\left\|\widehat{\beta}^{(m-1)} - \beta\right\| + \left|\sum_{t=T_0}^{T} \delta_t - \widehat{\delta}_t^{(m-1)}\right|\right) + O_P(\sqrt{ST}),
\end{aligned} \tag{S.11}$$

which completes the proof.

■

Lemma S.1.5. *Suppose the assumptions of Theorem 2.4.2 hold, and* $\sqrt{ST}\|\widehat{\beta}^{(m-1)}-\beta\| = O_P(1)$, $\sqrt{S}|\widehat{\delta}_t^{(m-1)}-\delta_t| = O_P(1)$, $|\sum_{t=T_0}^{T}\widehat{\delta}_t^{(m-1)}-\delta_t| = O_P(1)$ *and* $\|\sum_{t=T_0}^{T}\widehat{f}_t^{(m)}(\widehat{\delta}_t^{(m-1)}-\delta_t)\| = O_P(1)$. *Then, as* $T, S\to\infty$, *we have, for any fixed* $u\in\mathcal{U}$, $j=1,...,J$, $m\geq 0$,

$$\begin{aligned}&\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}_j}^{(m)}(u)(F_j(u)-\widehat{F}_j^{(m)}(u)H_j^{(m)}(u))\lambda_{js}(u)\\ &=\frac{1}{S}\sum_{s,g=1}^{S}\sum_{l=T_0}^{T}\omega_{sg}d_sd_g(\widehat{\delta}_{jt}^{(m-1)}(u)-\delta_{jt}(u))-\frac{1}{S}\sum_{s,g=1}^{S}\omega_{sg}d_s\eta_{jgt}(u)\\ &\quad-\frac{1}{ST}\sum_{s,g=1}^{S}d_s\mathbb{E}[\eta_{jgt}(u)\eta_{jg}(u)']\widehat{F}_j^{(m)}(u)(K_j^{(m)}(u))^{-1}\lambda_{js}(u)+O_P\left(\sqrt{\frac{S}{T}}\right),\end{aligned}$$

where $\omega_{j,sg}(u)$ *is defined in Assumption 2.4.4, and* $H_j^{(m)}(u)$ *and* $K_j^{(m)}(u)$ *are defined in Lemma S.1.3. Moreover, the leading terms on the right-hand side are* $O_P(\sqrt{S})$.

***Proof*.** Since u and j are fixed, for notational simplicity, we suppress u and j throughout the following proof.

Similar to Lemma S.1.4, we can write

$$\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)}(F-\widehat{F}^{(m)}H^{(m)})\lambda_s = -\sum_{h=1}^{7}\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_h^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s.$$

For similar reason to the proof of Lemma S.1.4, the terms associated with $I_5^{(m)}$, $I_6^{(m)}$, $I_7^{(m)}$ are negligible, proof omitted for concise. Below, we consider the terms associated with $I_1^{(m)}$, ..., $I_4^{(m)}$, respectively.

Using $e_t' M_{\widehat{F}}^{(m)} = e_t' - T^{-1}\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}$, we write

$$\begin{aligned}\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_1^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s =&\sum_{s=1}^{S} d_s e_t' I_1^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s\\ &-\frac{1}{T}\sum_{s=1}^{S} d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}I_1^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s.\end{aligned}$$

For the first term on the right-hand side of the equation, we substitute-in the expression of $I_1^{(m)}$ and further obtain that

$$\begin{aligned}\sum_{s=1}^{S} d_s e_t' I_1^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s = \frac{1}{ST}\Bigg[&\sum_{s,g=1}^{S}\sum_{l=T_0}^{T} d_s(\widehat{\delta}_t^{(m-1)}-\delta_t)(\widehat{\delta}_l^{(m-1)}-\delta_l)\widehat{f}_l^{(m)\prime}(K^{(m)})^{-1}\lambda_s\\ &+\sum_{s,g=1}^{S} d_sd_g(\widehat{\delta}_t^{(m-1)}-\delta_t)(\widehat{\beta}^{(m-1)}-\beta)'X_g'\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s\end{aligned}$$

$$+\sum_{s,g=1}^{S}\sum_{l=T_0}^{T} d_s w_{gt}'(\widehat{\beta}^{(m-1)}-\beta)(\widehat{\delta}_l^{(m-1)}-\delta_l)\widehat{f}_l^{(m)\prime}(K^{(m)})^{-1}\lambda_s$$
$$+\sum_{s,g=1}^{S} d_s w_{gt}'(\widehat{\beta}^{(m-1)}-\beta)(\widehat{\beta}^{(m-1)}-\beta)'X_g'\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s\Bigg],$$

where all terms are of $O_P(S/T)$ under the assumptions that $\sqrt{ST}\|\widehat{\beta}^{(m-1)}-\beta\|=O_P(1)$, $\sqrt{S}|\widehat{\delta}_t^{(m-1)}-\delta_t|=O_P(1)$, $|\sum_{t=T_0}^{T}\widehat{\delta}_t^{(m-1)}-\delta_t|=O_P(1)$ and $\|\sum_{t=T_0}^{T}\widehat{f}_t^{(m)}(\widehat{\delta}_t^{(m-1)}-\delta_t)\|=O_P(1)$. Furthermore, using the rate of $I_1^{(m)}$ from Lemma S.1.3, the rate of the second term

$$\begin{aligned}
&\left|\frac{1}{T}\sum_{s=1}^{S} d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}I_1^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s\right|\\
\le&\frac{1}{T}\left(\sum_{s=1}^{S}|d_s|^2\right)^{1/2}\cdot\left(\sum_{s=1}^{S}\|\lambda_s\|^2\right)^{1/2}\cdot\|\widehat{f}_t^{(m)}\|\cdot\|\widehat{F}^{(m)}\|\cdot\|I_1^{(m)}\|\cdot\|\widehat{F}^{(m)}\|\cdot\|(K^{(m)})^{-1}\|\\
=&O_P\left(\frac{S}{T}\right).
\end{aligned}$$

Therefore, we obtain that

$$\left|\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_1^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s\right| = O_P\left(\frac{S}{T}\right).$$

Under similar derivations, we shown that $|\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_3^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s| = O_P(\sqrt{S/T})$.

For the term associated with $I_2^{(m)}$, again, we expand it as

$$\begin{aligned}
\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_2^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s =& \sum_{s=1}^{S} d_s e_t' I_2^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s\\
&-\frac{1}{T}\sum_{s=1}^{S} d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)\prime}I_2^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s,
\end{aligned}$$

and the second term is $O_P(\sqrt{S/T})$ given the rate of $I_2^{(m)}$ from Lemma S.1.3. Now we expand the first term by the expression of $I_2^{(m)}$:

$$\begin{aligned}
&\sum_{s=1}^{S} d_s e_t' I_2^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s\\
&\quad=-\frac{1}{S}\sum_{s,g=1}^{S}\sum_{l=T_0}^{T}\omega_{sg}d_s d_g(\widehat{\delta}_t^{(m-1)}-\delta_t)-\frac{1}{S}\sum_{s,g=1}^{S}\omega_{sg}d_s x_{gt}'(\widehat{\beta}^{(m-1)}-\beta)\\
&\qquad-\frac{1}{ST}\sum_{s,g=1}^{S}\sum_{l=T_0}^{T} d_s f_t'\lambda_g(\widehat{\delta}_l^{(m-1)}-\delta_l)\widehat{f}_l^{(m)\prime}(K^{(m)})^{-1}\lambda_s
\end{aligned}$$

$$- \frac{1}{ST} \sum_{s,g=1}^{S} d_s f_t' \lambda_g (\widehat{\beta}^{(m-1)} - \beta)' X_g' \widehat{F}^{(m)} (K^{(m)})^{-1} \lambda_s.$$

Using the induction assumptions and Hölder's and Cauchy-Schwartz inequalities, we obtain

$$\left| \frac{1}{S} \sum_{s,g=1}^{S} \sum_{l=T_0}^{T} \omega_{sg} d_s d_g (\widehat{\delta}_t^{(m-1)} - \delta_t) \right| = O_P(\sqrt{S}),$$

$$\left| \frac{1}{S} \sum_{s,g=1}^{S} \omega_{sg} d_s x_{gt}' (\widehat{\beta}^{(m-1)} - \beta) \right\| = O_P\left(\sqrt{\frac{S}{T}} \right),$$

$$\left| \frac{1}{ST} \sum_{s,g=1}^{S} \sum_{l=T_0}^{T} d_s f_t' \lambda_g (\widehat{\delta}_l^{(m-1)} - \delta_l) \widehat{f}_l^{(m)\prime} (K^{(m)})^{-1} \lambda_s \right| = O_P\left(\frac{S}{T} \right),$$

$$\left| \frac{1}{ST} \sum_{s,g=1}^{S} d_s f_t' \lambda_g (\widehat{\beta}^{(m-1)} - \beta)' X_g' \widehat{F}^{(m)} (K^{(m)})^{-1} \lambda_s \right| = O_P\left(\sqrt{\frac{S}{T}} \right).$$

Therefore, we conclude

$$\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_2^{(m)} \widehat{F}^{(m)} (K^{(m)})^{-1} \lambda_s = - \frac{1}{S} \sum_{s,g=1}^{S} \sum_{l=T_0}^{T} \omega_{sg} d_s d_g (\widehat{\delta}_t^{(m-1)} - \delta_t) + O_P\left(\sqrt{\frac{S}{T}} \right).$$

For the term associated with $I_4^{(m)}$, we expand the terms using the expression of $I_4^{(m)}$ and obtain

$$\begin{aligned} &\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_4^{(m)} \widehat{F}^{(m)} (K^{(m)})^{-1} \lambda_s \\ =& \frac{1}{ST} \sum_{s,g=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} \left\{ F \lambda_g \eta_g' + \eta_g \lambda_g' F' + \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g'] \right) + \mathbb{E}[\eta_g \eta_g'] \right\} \widehat{F}^{(m)} \left(K^{(m)} \right)^{-1} \lambda_s, \end{aligned}$$

where, by standard arguments, we show that $|(ST)^{-1} \sum_{s,g=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} (\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']) \widehat{F}^{(m)}$ $(K^{(m)})^{-1} \lambda_s| = O_P(S/T)$, and $|(ST)^{-1} \sum_{s,g=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} F \lambda_g \eta_g' \widehat{F}^{(m)} (K^{(m)})^{-1} \lambda_s| = O_P(\sqrt{S/T})$. In addition,

$$\frac{1}{ST} \sum_{s,g=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} \eta_g \lambda_g' F' \widehat{F}^{(m)} \left(K^{(m)} \right)^{-1} \lambda_s = \frac{1}{S} \sum_{s,g=1}^{S} \omega_{sg} d_s \eta_{gt} - \frac{1}{ST} \sum_{s,g=1}^{S} \omega_{sg} d_s \widehat{f}_t' \widehat{F} \eta_g,$$

where the first term is $O_P(\sqrt{S})$, while the second term is $O_P(\sqrt{S/T})$. The last term

$$\frac{1}{ST} \sum_{s,g=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} \mathbb{E}[\eta_g \eta_g'] \widehat{F}^{(m)} (K^{(m)})^{-1} \lambda_s$$

$$= \frac{1}{ST}\sum_{s,g=1}^{S} d_s\mathbb{E}[\eta_{gt}\eta_g']\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s - \frac{1}{ST^2}\sum_{s,g=1}^{S} d_s\widehat{f}_t^{(m)\prime}\widehat{F}^{(m)}\mathbb{E}[\eta_g\eta_g']\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s,$$

where the first term is $O_P(S/\sqrt{T})$ and the second term is $O_P(S/T)$. Therefore, collecting all terms of $\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_4^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s$, we obtain

$$\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)} I_4^{(m)}\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s$$
$$= \frac{1}{S}\sum_{s,g=1}^{S}\omega_{sg}d_s\eta_{gt} + \frac{1}{ST}\sum_{s,g=1}^{S} d_s\mathbb{E}[\eta_{gt}\eta_g']\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s + O_P\Big(\sqrt{\frac{S}{T}}\Big),$$

where the first two terms on the right-hand side are $O_P(\sqrt{S})$ and $O_P(S/\sqrt{T})$, respectively.

Collecting all terms associated with $I_h^{(m)}$, we obtain

$$\sum_{s=1}^{S} d_s e_t' M_{\widehat{F}}^{(m)}(F - \widehat{F}^{(m)}H^{(m)})\lambda_s = \frac{1}{S}\sum_{s,g=1}^{S}\sum_{l=T_0}^{T}\omega_{sg}d_s d_g(\widehat{\delta}_t^{(m-1)} - \delta_t) - \frac{1}{S}\sum_{s,g=1}^{S}\omega_{sg}d_s\eta_{gt}$$
$$- \frac{1}{ST}\sum_{s,g=1}^{S} d_s\mathbb{E}[\eta_{gt}\eta_g']\widehat{F}^{(m)}(K^{(m)})^{-1}\lambda_s + O_P\Big(\sqrt{\frac{S}{T}}\Big),$$

where the remaining terms on the right-hand side have been shown to be $O_P(\sqrt{S})$ given that $T/S \to \kappa$. ■

S.1.2 Preliminary Lemmas

Lemma S.1.6 (Theorem 3 of Chetverikov et al. (2016))**.** *Under Assumption 2.2.1 and 2.2.2, there exist constants $\bar{c}, c, C > 0$, which are independent of $s, t, N_{st}, S, T, N_{ST}$, such that for all $(s,t) \in \{1, ..., S\} \times \{1, ..., T\}$, $u \in \mathcal{U}$ and $x \in (0, \bar{c})$,*

$$\mathbb{P}(\|\widehat{\alpha}_{st}(u) - \alpha_{st}(u)\| > x) \le Ce^{-cx^2 N_{st}}.$$

Lemma S.1.7 (Lemma A.1 of Gao (2007))**.** *Suppose that $\{M_m^n : -\infty < m \le n < +\infty\}$ are the σ-fields generated by a stationary and α-mixing process $\{\xi_i\}_{-\infty}^{+\infty}$ with the mixing coefficient $a(i)$. For some positive integers m, let $\delta_i \in M_{s_i}^{t_i}$ where $s_1 < t_1 < s_2 < t_2 < \cdots < s_m < t_m$ and assume that $t_i - s_i \ge \tau$ for all i and some $\tau > 0$. Assume further that, for some $p_i > 1, \mathbb{E}\,|\delta_i|^{p_i} < +\infty$, for which $Q := \sum_{i=1}^{\ell}\frac{1}{p_i} < 1$. Then we have*

$$\left|\mathbb{E}\left(\Pi_{i=1}^{\ell}\delta_i\right) - \Pi_{i=1}^{\ell}\mathbb{E}\left(\delta_i\right)\right| < 10(\ell - 1)a(\tau)^{1-Q}\Pi_{i=1}^{\ell}\left(\mathbb{E}\,|\delta_i|^{p_i}\right)^{\frac{1}{p_i}}.$$

Lemma S.1.8. *If $\mathbb{E}\|x_{st}\|^4 \le C$ for any $1 \le s \le S$ and $1 \le t \le T$, then $\max_{1\le s\le S}\|X_s\| = o_p\big((ST)^{1/2}\big)$.*

Proof. Let $M > 0$ be an arbitrary constant.

By Markov inequality, we have

$$P\Big(\max_{1\le s\le S}\|X_s\|>M(ST)^{1/2}\Big)\le\sum_{s=1}^{S}P\Big(\rho_{\max}(X_sX_s')>M^2ST\Big)\le\frac{\sum_{s=1}^{S}\mathbb{E}[\rho_{\max}(X_sX_s')^2]}{M^4(ST)^2}.$$

In addition, under Assumption 2.4.1.(i), we have

$$\sum_{s=1}^{S}\mathbb{E}\big[\rho_{\max}(X_sX_s')^2\big]\le\sum_{s=1}^{S}\mathbb{E}\big[\mathrm{tr}(X_s'X_s)^2\big]\le\sum_{s=1}^{S}\mathbb{E}\Bigg[\Bigg(\sum_{t=1}^{T}\|x_{st}\|^2\Bigg)^2\Bigg]=O(ST^2),$$

which implies $(ST)^{-1/2}\max_{1\le s\le S}\sqrt{\rho_{\max}(X_sX_s')}=o_P(1)$. ■

Lemma S.1.9. *Under the assumptions of Theorem 2.4.1, for any fixed $u\in\mathcal{U}$, $j=1,...,J$ and $m\ge 1$,*

$$\begin{aligned}&\frac{1}{ST}\sum_{s,g=1}^{S}X_s'M_{\widehat{F}_j}^{(m)}(u)\left(\eta_{js}(u)\eta_{js}(u)'-\mathbb{E}[\eta_{js}(u)\eta_{js}(u)']\right)\widehat{F}_j^{(m)}(u)\left(K_j^{(m)}(u)\right)^{-1}\lambda_{js}(u)\\ &=O_P(\sqrt{T})\cdot\left(\left\|\widehat{F}_j^{(m)}(u)-F_j(u)\left(H^{(m}(u)\right)^{-1}\right\|^2+\left\|\widehat{F}_j^{(m)}(u)-F_j(u)\left(H^{(m}(u)\right)^{-1}\right\|\right)\\ &\quad+O_P\left(\delta_{ST}\right).\end{aligned}$$

Proof. Since u and j are fixed, we suppress u and j throughout the following proof.

We first plug-in $M_{\widehat{F}}^{(m)}=I_T-T^{-1}\widehat{F}^{(m)}\widehat{F}^{(m)\prime}$ to obtain

$$\begin{aligned}&\frac{1}{ST}\sum_{s,g=1}^{S}X_s'M_{\widehat{F}}^{(m)}\left(\eta_s\eta_s'-\mathbb{E}[\eta_s\eta_s']\right)\widehat{F}^{(m)}\left(K^{(m)}\right)^{-1}\lambda_s\\ &=\frac{1}{ST}\sum_{s,g=1}^{S}X_s'\left(\eta_g\eta_g'-\mathbb{E}[\eta_g\eta_g']\right)\widehat{F}^{(m)}\left(K^{(m)}\right)^{-1}\lambda_s\\ &\quad-\frac{1}{ST}\sum_{s=1}^{S}\frac{X_s'\widehat{F}^{(m)}}{T}\left(\sum_{g=1}^{S}\widehat{F}^{(m)\prime}\left(\eta_g\eta_g'-\mathbb{E}[\eta_g\eta_g']\right)\widehat{F}^{(m)}\left(K^{(m)}\right)^{-1}\right)\lambda_s\\ &=\frac{1}{ST}\sum_{s,g=1}^{S}X_s'\left(\eta_g\eta_g'-\mathbb{E}[\eta_g\eta_g']\right)\left(\widehat{F}^{(m)}-F\left(H^{(m)}\right)^{-1}\right)\left(K^{(m)}\right)^{-1}\lambda_s\\ &\quad+\frac{1}{ST}\sum_{s,g=1}^{S}X_s'\left(\eta_g\eta_g'-\mathbb{E}[\eta_g\eta_g']\right)F\left(H^{(m)}\right)^{-1}\left(K^{(m)}\right)^{-1}\lambda_s\\ &\quad-\frac{1}{ST}\sum_{s=1}^{S}\frac{X_s'\widehat{F}^{(m)}}{T}\left(\sum_{g=1}^{S}\widehat{F}^{(m)\prime}\left(\eta_g\eta_g'-\mathbb{E}[\eta_g\eta_g']\right)\widehat{F}^{(m)}\left(K^{(m)}\right)^{-1}\right)\lambda_s,\end{aligned}$$

and then we evaluate the three terms in the last equality separately.

Apply Cauchy-Schwartz inequality to the first term, we obtain

$$
\begin{aligned}
&\left\| \frac{1}{ST} \sum_{s,g=1}^{S} X_s' \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \left(\widehat{F}^{(m)} - F \left(H^{(m)}\right)^{-1}\right) \left(K^{(m)}\right)^{-1} \lambda_s \right\| \\
\leq & \frac{1}{\sqrt{S}} \left\| \widehat{F}^{(m)} - F \left(H^{(m)}\right)^{-1} \right\| \cdot \left\| \left(K^{(m)}\right)^{-1} \right\| \\
& \cdot \left(\sum_{s=1}^{S} \left\| \frac{1}{T\sqrt{S}} \sum_{g=1}^{S} X_s' \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \right\|^2 \right)^{1/2} \left(\sum_{s=1}^{S} \|\lambda_s\|^2 \right)^{1/2} \\
= & \frac{1}{\sqrt{S}} \cdot O_P\left(\left\| \widehat{F}^{(m)} - F \left(H^{(m)}\right)^{-1} \right\| \right) \cdot O_P(1) \cdot O_P(1) \cdot O_P(\sqrt{S}) \\
= & O_P\left(\left\| \widehat{F}^{(m)} - F \left(H^{(m)}\right)^{-1} \right\| \right),
\end{aligned}
$$

due to Assumption 2.4.1.(ii), (S.9), (S.5), and the fact that, for any fixed $s = 1, ..., S$,

$$
\mathbb{E}\left[\left\| \frac{1}{T\sqrt{S}} \sum_{g=1}^{S} X_s' \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \right\|^2 \right] = \frac{C}{ST^2} \cdot O(ST^2) = O(1), \tag{S.12}
$$

according to Lemma S.1.7 under Assumption 2.4.1 and Assumption 2.4.2.

For the second term, similarly we have

$$
\begin{aligned}
&\left\| \frac{1}{ST} \sum_{s,g=1}^{S} X_s' \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) F \left(H^{(m)}\right)^{-1} \left(K^{(m)}\right)^{-1} \lambda_s \right\| \\
\leq & \frac{1}{\sqrt{S}} \cdot \left\| \left(H^{(m)}\right)^{-1} \right\| \cdot \left\| \left(K^{(m)}\right)^{-1} \right\| \\
& \cdot \left(\sum_{s=1}^{S} \left\| \frac{1}{T\sqrt{S}} \sum_{g=1}^{S} \sum_{t,l=1}^{T} x_{st} \left(\eta_{gt} \eta_{gl} - \mathbb{E}[\eta_{gt} \eta_{gl}]\right) f_l' \right\|^2 \right)^{1/2} \left(\sum_{s=1}^{S} \|\lambda_s\|^2 \right)^{1/2} \\
= & O_P\left(\sqrt{S}\right).
\end{aligned}
$$

We consider the third term as follows

$$
\begin{aligned}
&\left\| \frac{1}{ST} \sum_{s=1}^{S} \frac{X_s' \widehat{F}^{(m)}}{T} \left(\sum_{g=1}^{S} \widehat{F}^{(m)\prime} \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1} \right) \lambda_s \right\| \\
\leq & \left(\sum_{s=1}^{S} \|X_s\|^2 \right)^{1/2} \cdot \left(\sum_{s=1}^{S} \|\lambda_s\|^2 \right)^{1/2} \cdot \left\| \widehat{F}^{(m)} \right\| \cdot \left\| \left(K^{(m)}\right)^{-1} \right\| \\
& \cdot \left\| \frac{1}{ST^2} \sum_{g=1}^{S} \widehat{F}^{(m)\prime} \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \widehat{F}^{(m)} \right\| \\
= & O_P(ST) \cdot \left\| \frac{1}{ST^2} \sum_{g=1}^{S} \widehat{F}^{(m)\prime} \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \widehat{F}^{(m)} \right\|
\end{aligned}
$$

$$=O_P(\sqrt{T}) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\|^2 + O_P(\sqrt{T}) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\| + O_P\left(\delta_{ST}\right),$$

where the last equality follows from the following result

$$\frac{1}{ST^2} \sum_{g=1}^{S} \widehat{F}^{(m)\prime} \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \widehat{F}^{(m)}$$
$$= O_P\left(\frac{1}{T\sqrt{S}}\right) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\|^2 + O_P\left(\frac{1}{T\sqrt{S}}\right) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\| + O_P\left(\frac{1}{T\sqrt{S}}\right),$$

whose proof is given as follows. By adding and subtracting $F(H^{(m)})^{-1}$ we have

$$\begin{aligned}
&\frac{1}{ST^2} \sum_{g=1}^{S} \widehat{F}^{(m)\prime} \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \widehat{F}^{(m)} \\
&= \frac{1}{ST^2} \sum_{g=1}^{S} \left(\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right)' \left(\eta_g' \eta_g - \mathbb{E}[\eta_g' \eta_g]\right) \left(\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right) \\
&\quad + \frac{1}{ST^2} \sum_{g=1}^{S} \left(H^{(m)}\right)^{-1} F' \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) \left(\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right) \\
&\quad + \frac{1}{ST^2} \sum_{g=1}^{S} \left(\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right)' \left(\eta_g' \eta_g - \mathbb{E}[\eta_g' \eta_g]\right) F\left(H^{(m)\prime}\right)^{-1} \\
&\quad + \frac{1}{ST^2} \sum_{g=1}^{S} \left(H^{(m)}\right)^{-1} F' \left(\eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g']\right) F\left(H^{(m)}\right)^{-1}.
\end{aligned}$$

The first term

$$\begin{aligned}
&\left\| \frac{1}{ST^2} \sum_{g=1}^{S} \left(\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right)' \left(\eta_g' \eta_g - \mathbb{E}[\eta_g' \eta_g]\right) \left(\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right) \right\| \\
&\leq \frac{1}{T\sqrt{S}} \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\|^2 \cdot \left\| \frac{1}{T\sqrt{S}} \sum_{g=1}^{S} \left(\eta_g' \eta_g - \mathbb{E}[\eta_g' \eta_g]\right) \right\| \\
&= O_P\left(\frac{1}{T\sqrt{S}}\right) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\|^2,
\end{aligned}$$

using the fact that $\left\| \sum_{g=1}^{S} \eta_g \eta_g' - \mathbb{E}[\eta_g \eta_g'] \right\| = O_P\left(T\sqrt{S}\right)$, which can be shown in a similar way to the proof of (S.12). Similar arguments yield that the second and third term are $O_P(\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \|) + O_P(\max(S^{-1}T^{-1/2}, S^{-1/2}T^{-1}))$, and the fourth term is $O_P(S^{-1/2}T^{-1})$.

Collecting all the terms so far, we obtain

$$\frac{1}{ST} \sum_{s,g=1}^{S} X_s' M_{\widehat{F}}^{(m)} \left(\eta_s \eta_s' - \mathbb{E}[\eta_s \eta_s']\right) \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1} \lambda_s$$

$$= O_P(\sqrt{T}) \cdot \left(\left\| \widehat{F}^{(m)} - F \left(H^{(m)} \right)^{-1} \right\|^2 + \left\| \widehat{F}^{(m)} - F \left(H^{(m)} \right)^{-1} \right\| \right) + O_P \left(\delta_{ST} \right).$$

■

Lemma S.1.10. *Under assumptions of Theorem 2.4.1, for any fixed $u \in \mathcal{U}$, $j = 1, ..., J$ and $m \geq 1$,*

$$\left\| \frac{1}{T} F_j'(u) \left(\widehat{F}_j^{(m)}(u) H_j^{(m)}(u) - F_j(u) \right) \right\| = O_P \left(\frac{1}{\sqrt{ST}} \right) \cdot \left\| \widehat{F}^{(m)} - F \big(H^{(m)} \big)^{-1} \right\|, \quad \text{(S.13)}$$

$$\left\| I_r - \big(H_j^{(m)}(u)' \big)^{-1} \big(H_j^{(m)}(u) \big)^{-1} \right\| \quad \text{(S.14)}$$
$$= \frac{1}{T} \left\| \widehat{F}^{(m)} - F \big(H^{(m)} \big)^{-1} \right\|^2 + O_P \Big(\frac{1}{\sqrt{ST}} \Big) \cdot \left\| \widehat{F}^{(m)} - F \big(H^{(m)} \big)^{-1} \right\|,$$

$$\left\| P_{\widehat{F}_j}^{(m)} - P_{F_j} \right\| \quad \text{(S.15)}$$
$$= \frac{1}{T} \left\| \widehat{F}^{(m)} - F \big(H^{(m)} \big)^{-1} \right\|^2 + O_P \Big(\frac{1}{\sqrt{ST}} \Big) \cdot \left\| \widehat{F}^{(m)} - F \big(H^{(m)} \big)^{-1} \right\|,$$

where $H_j^{(m)}$ is defined in Lemma S.1.3, and $\|\widehat{F}^{(m)} - F \big(H^{(m)} \big)^{-1}\|$ is given by (S.9).

Proof. Since u and j are fixed, we suppress u and j throughout the following proof.

For (S.13), according to the proof of Lemma S.1.3, we have the following decomposition:

$$\frac{1}{T} F' \left(\widehat{F}^{(m)} H^{(m)} - F \right) = \sum_{h=1}^{7} \frac{1}{T} F' I_h^{(m)} \widehat{F}^{(m)} \left(K^{(m)} \right)^{-1}$$

where $I_h^{(m)}$, $h = 1, ..., 7$ are defined above (S.7) Using the rate of $\|I_h^{(m)}\|$ from the proof of Lemma S.1.3 and (S.5), it is straightforward to check that for $h = 1$,

$$\left\| \frac{1}{T} F' I_h^{(m)} \widehat{F}^{(m)} \left(K^{(m)} \right)^{-1} \right\| \leq \left\| \frac{F'}{\sqrt{T}} \right\| \cdot \left\| I_1^{(m)} \right\| \cdot \left\| \frac{\widehat{F}^{(m)}}{\sqrt{T}} \right\| \cdot \left\| \left(K^{(m)} \right)^{-1} \right\| = o_P(1) \cdot \left\| \widehat{\beta}^{(m-1)} - \beta \right\|.$$

Applying similar arguments, we obtain $\|T^{-1} F' I_h^{(m)} \widehat{F}^{(m)} \big(K^{(m)} \big)^{-1}\| = O_P(\|\widehat{\beta}^{(m-1)} - \beta\|)$ for $h = 2, 3, 5, 6, 7$. Thus, it remains to check the rate of

$$\begin{aligned} \frac{1}{T} F' I_4^{(m)} \widehat{F}^{(m)} \left(K^{(m)} \right)^{-1} =& \frac{1}{ST^2} \sum_{s=1}^{S} F' \Big[F \lambda_s \eta_s' + \eta_s \lambda_s' F' + \eta_s \eta_s' \Big] \widehat{F}^{(m)} \left(K^{(m)} \right)^{-1} \\ =& \frac{1}{ST} \sum_{s=1}^{S} \lambda_s \eta_s' \widehat{F}^{(m)} \left(K^{(m)} \right)^{-1} + \frac{1}{ST} \sum_{s=1}^{S} F' \eta_s \lambda_s' \left(\frac{\Lambda' \Lambda}{S} \right)^{-1} \\ &+ \frac{1}{ST^2} \sum_{s=1}^{S} F' \eta_s \eta_s' \widehat{F}^{(m)} \cdot \left(K^{(m)} \right)^{-1}. \end{aligned} \quad \text{(S.16)}$$

The first term in the last equality of (S.16)

$$
\begin{aligned}
&\left\| \frac{1}{ST} \sum_{s=1}^{S} \lambda_s \eta_s' \widehat{F}^{(m)} \left(K^{(m)}\right)^{-1} \right\| \\
&\le \left[\left\| \frac{1}{ST} \sum_{s=1}^{S} \lambda_s \eta_s' \left(\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right) \right\| + \left\| \frac{1}{ST} \sum_{s=1}^{S} \lambda_s \eta_s' F \left(H^{(m)}\right)^{-1} \right\| \right] \cdot \left\| \left(K^{(m)}\right)^{-1} \right\| \\
&\le \left[\left\| \frac{1}{ST} \sum_{s=1}^{S} \lambda_s \eta_s' \right\| \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\| + \left\| \frac{1}{ST} \sum_{s=1}^{S} \lambda_s \eta_s' F \right\| \cdot \left\| \left(H^{(m)}\right)^{-1} \right\| \right] \cdot O_P(1) \\
&= O_P\left(\frac{1}{\sqrt{ST}} \right) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\| + O_P\left(\frac{1}{\sqrt{ST}} \right) \cdot O_P(1) \\
&= O_P\left(\frac{1}{\sqrt{ST}} \right) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\|,
\end{aligned}
$$

where the second last equality follows from (S.9) and Lemma S.1.2. Following Lemma S.1.2, the second term in the last equality of (S.16) becomes

$$
\left\| \frac{1}{ST} \sum_{s=1}^{S} F' \eta_s \lambda_s' \left(\frac{\Lambda'\Lambda}{S} \right)^{-1} \right\| \le \left\| \frac{1}{ST} \sum_{s=1}^{S} F' \eta_s \lambda_s' \right\| \cdot \left\| \left(\frac{\Lambda'\Lambda}{S} \right)^{-1} \right\| = O_P\left(\frac{1}{\sqrt{ST}} \right).
$$

The third term in the last equality of (S.16) is

$$
\begin{aligned}
&\left\| \frac{1}{ST^2} \sum_{s=1}^{S} F' \eta_s \eta_s' \widehat{F}^{(m)} \cdot \left(K^{(m)}\right)^{-1} \right\| \\
&\le \left\| \frac{1}{ST^2} \sum_{s=1}^{S} F' \eta_s \eta_s' \left(\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right) \right\| \cdot \left\| \left(K^{(m)}\right)^{-1} \right\| \\
&\quad + \left\| \frac{1}{ST^2} \sum_{s=1}^{S} F' \eta_s \eta_s' F \left(H^{(m)}\right)^{-1} \right\| \cdot \left\| \left(K^{(m)}\right)^{-1} \right\| \\
&\le \left\{ \left\| \frac{F'}{\sqrt{T}} \right\| \cdot \left\| \frac{1}{ST} \sum_{s=1}^{S} \eta_s \eta_s' \right\| \cdot \frac{1}{\sqrt{T}} \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\| \right. \\
&\quad \left. + \left\| \frac{1}{ST^2} \sum_{s=1}^{S} F' \eta_s \eta_s' F \right\| \cdot \left\| \left(H^{(m)}\right)^{-1} \right\| \right\} \cdot \left\| \left(K^{(m)}\right)^{-1} \right\| \\
&= O_P\left(\frac{1}{\sqrt{ST}} \right) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\| + O_P\left(\frac{1}{T} \right) \\
&= O_P\left(\frac{1}{\sqrt{ST}} \right) \cdot \left\| \widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1} \right\|,
\end{aligned}
$$

due to Lemma S.1.2 and the following fact that

$$
\mathbb{E}\left[\left\| \frac{1}{ST^2} \sum_{s=1}^{S} F' \eta_s \eta_s' F \right\| \right] \le \mathbb{E}\left[\frac{1}{ST^2} \left\| \sum_{s=1}^{S} F' \eta_s \right\|^2 \right] \le \frac{1}{ST^2} \sum_{s=1}^{S} \sum_{t,l=1}^{T} \mathbb{E}\left[\eta_{st} f_t' f_l \eta_{sl} \right]
$$

$$
\begin{aligned}
=&\frac{1}{ST^2}\sum_{s=1}^{S}\sum_{t=l}^{T}\mathbb{E}\left[\eta_{st}^2 f_t' f_t\right] + \frac{1}{ST^2}\sum_{s=1}^{S}\sum_{t\neq l}^{T}\mathbb{E}\left[\eta_{st} f_t' f_l \eta_{sl}\right] \\
\leq & O_P\left(\frac{1}{T}\right) + \frac{C}{ST^2}\sum_{i=1}^{r}\sum_{s=1}^{S}\sum_{t\neq l}^{T} a_s(|t-l|)^{\frac{\delta}{\delta+2}}\left(\mathbb{E}\left[|\eta_{st} f_{ti}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}}\left(\mathbb{E}\left[|\eta_{sl} f_{li}|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}} \\
=& O_P\left(\frac{1}{T}\right),
\end{aligned}
$$

by Lemma S.1.7. Collecting all the terms, we have

$$
\left\|\frac{1}{T}F'\left(\widehat{F}^{(m)}H^{(m)} - F\right)\right\| = O_P\Big(\frac{1}{\sqrt{ST}}\Big)\cdot\left\|\widehat{F}^{(m)} - F\big(H^{(m)}\big)^{-1}\right\|.
$$

For (S.14), by multiplying $\big(H^{(m)\prime}\big)^{-1}$ and $\big(H^{(m)}\big)^{-1}$ on both sides of $T^{-1}F'(\widehat{F}^{(m)}H^{(m)} - F)$, respectively, we obtain

$$
\left\|\frac{1}{T}\big(H^{(m)\prime}\big)^{-1}F'\widehat{F}^{(m)} - \big(H^{(m)\prime}\big)^{-1}\big(H^{(m)}\big)^{-1}\right\| = \left\|\widehat{F}^{(m)} - F\big(H^{(m)}\big)^{-1}\right\|,
$$

since $\|F\| = O_P(\sqrt{T})$ under Assumption 2.4.1.(ii), and $\|\big(H^{(m)}\big)^{-1}\| \leq \|\big(\Upsilon^{(m)}\big)^{-1}\|\cdot\|K^{(m)}\| = O_P(1)$. Moreover, we have

$$
\begin{aligned}
\left\|I_r - \frac{1}{T}\big(H^{(m)\prime}\big)^{-1}F'\widehat{F}^{(m)}\right\| &= \frac{1}{T}\left\|\widehat{F}^{(m)\prime}\left(\widehat{F}^{(m)} - F\big(H^{(m)\prime}\big)^{-1}\right)\right\| \\
&\leq \frac{1}{T}\left\|\widehat{F}^{(m)} - F\big(H^{(m)\prime}\big)^{-1}\right\|^2 + \left\|\big(H^{(m)}\big)^{-1}\right\|\cdot\left\|\frac{1}{T}F'\left(\widehat{F}^{(m)} - F\big(H^{(m)\prime}\big)^{-1}\right)\right\| \\
&= \frac{1}{T}\left\|\widehat{F}^{(m)} - F\big(H^{(m)}\big)^{-1}\right\|^2 + O_P\Big(\frac{1}{\sqrt{ST}}\Big)\cdot\left\|\widehat{F}^{(m)} - F\big(H^{(m)}\big)^{-1}\right\|.
\end{aligned}
$$

Therefore, by triangle inequality, we show that

$$
\begin{aligned}
&\left\|I_r - \big(H^{(m)\prime}\big)^{-1}\big(H^{(m)}\big)^{-1}\right\| \\
&\quad\leq \left\|I_r - \frac{1}{T}\big(H^{(m)\prime}\big)^{-1}F'\widehat{F}^{(m)}\right\| + \left\|\frac{1}{T}\big(H^{(m)\prime}\big)^{-1}F'\widehat{F}^{(m)} - \big(H^{(m)\prime}\big)^{-1}\big(H^{(m)}\big)^{-1}\right\| \\
&\quad= \frac{1}{T}\left\|\widehat{F}^{(m)} - F\big(H^{(m)}\big)^{-1}\right\|^2 + O_P\Big(\frac{1}{\sqrt{ST}}\Big)\cdot\left\|\widehat{F}^{(m)} - F\big(H^{(m)}\big)^{-1}\right\|.
\end{aligned}
$$

For (S.15), since

$$
\left\|P_{\widehat{F}}^{(m)} - P_F\right\|^2 \leq \text{tr}\left(P_{\widehat{F}}^{(m)} - P_F\right)^2 = 2\text{tr}\left(I_r - \frac{1}{T}F'P_{\widehat{F}^{(m)}}F\right),
$$

it is suffices to examine

$$
I_r - \frac{1}{T}F'P_{\widehat{F}^{(m)}}F = I_r - \frac{F'\widehat{F}^{(m)}}{T}\frac{\widehat{F}^{(m)\prime}F}{T}
$$

$$
\begin{aligned}
=& I_r - \left[\frac{1}{T}\widehat{F}^{(m)\prime}\left(F - \widehat{F}^{(m)}\left(H^{(m)}\right)^{-1}\right) + \left(H^{(m)}\right)^{-1}\right]' \\
&\cdot \left[\frac{1}{T}\widehat{F}^{(m)\prime}\left(F - \widehat{F}^{(m)}\left(H^{(m)}\right)^{-1}\right) + \left(H^{(m)}\right)^{-1}\right] \\
=& I_r - \left(H^{(m)\prime}\right)^{-1}\left(H^{(m)}\right)^{-1} - \frac{1}{T^2}\left(F - \widehat{F}^{(m)}\left(H^{(m)}\right)^{-1}\right)'\widehat{F}^{(m)}\widehat{F}^{(m)\prime}\left(F - \widehat{F}^{(m)}\left(H^{(m)}\right)^{-1}\right) \\
&- \left(H^{(m)\prime}\right)^{-1}\cdot\frac{1}{T}\widehat{F}^{(m)\prime}\left(F - \widehat{F}^{(m)}\left(H^{(m)}\right)^{-1}\right) - \frac{1}{T}\left(F - \widehat{F}^{(m)}\left(H^{(m)}\right)^{-1}\right)'\widehat{F}^{(m)}\cdot\left(H^{(m)}\right)^{-1} \\
=& \frac{1}{T}\left\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right\|^2 + O_P\left(\frac{1}{\sqrt{ST}}\right)\cdot\left\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right\|,
\end{aligned}
$$

where the second equality follows from the fact that $T^{-1}\widehat{F}^{(m)\prime}\widehat{F}^{(m)} = I_r$, and the last equality follows from (S.13)-(S.14) and $\left\|\left(H^{(m)}\right)^{-1}\right\| = O_P(1)$.

■

Lemma S.1.11. *Under the assumptions of Theorem 2.4.1, we have for any fixed $u \in \mathcal{U}$, $j = 1, ..., J$, $m \geq 1$, $s = 1, ..., S$,*

$$
\begin{aligned}
&\left\|\widehat{\lambda}_{js}^{(m)}(u) - H_j^{(m)}(u)\lambda_{js}(u)\right\| \qquad\qquad \text{(S.17)}\\
&\quad = O_P\left(\frac{1}{\sqrt{T}}\right)\cdot\left|\sum_{t=T_0}^{T}\delta_t - \widehat{\delta}_t^{(m-1)}\right| + O_P\left(\left\|\beta - \widehat{\beta}^{(m-1)}\right\|\right) + O_P\left(\frac{1}{\sqrt{T}}\right)\cdot\left\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right\|,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{S}\left\|\widehat{\Lambda}_j^{(m)}(u) - \Lambda_j(u)H_j^{(m)}(u)'\right\|^2 \qquad\qquad \text{(S.18)}\\
&= \quad O_P\left(\frac{1}{T}\right)\cdot\left|\sum_{t=T_0}^{T}\delta_t - \widehat{\delta}_t^{(m-1)}\right|^2 + O_P\left(\left\|\beta - \widehat{\beta}^{(m-1)}\right\|^2\right) + O_P\left(\frac{1}{T}\right)\cdot\left\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right\|^2,
\end{aligned}
$$

$$
\begin{aligned}
&\left\|\left(\frac{\widehat{\Lambda}_j^{(m)}(u)'\widehat{\Lambda}_j(u)}{S}\right)^{-1} - \left(H_j^{(m)}(u)'\right)^{-1}\left(\frac{\Lambda_j^{(m)}(u)'\Lambda_j(u)}{S}\right)^{-1}\left(H_j^{(m)}(u)\right)^{-1}\right\| \qquad \text{(S.19)}\\
&\quad = O_P\left(\frac{1}{T}\right)\cdot\left|\sum_{t=T_0}^{T}\delta_t - \widehat{\delta}_t^{(m-1)}\right|^2 + O_P\left(\left\|\beta - \widehat{\beta}^{(m-1)}\right\|^2\right) + O_P\left(\frac{1}{T}\right)\cdot\left\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right\|^2,
\end{aligned}
$$

where $H_j^{(m)}$ is defined in Lemma S.1.3.

Proof. Since u and j are fixed in Lemma S.1.11, we suppress u and j throughout the following proof.

As $\widehat{\lambda}_s^{(m)}$ is estimated via PCA, we have $\widehat{\lambda}_s^{(m)} = T^{-1}\widehat{F}^{(m)\prime}(\widehat{A}_s - D_s\widehat{\delta}^{(m-1)} - X_s\widehat{\beta}^{(m-1)})$. Thus, for (S.17), substituting $\widehat{A}_s = D_s\delta + X_s\beta + F\lambda_s + \eta_s + (\widehat{A}_s - A_s)$ into the expression of $\widehat{\lambda}_s^{(m)}$, we obtain

$$
\begin{aligned}
\widehat{\lambda}_s^{(m)} &= \frac{1}{T}\widehat{F}^{(m)\prime}\left(\widehat{A}_s - D_s\widehat{\delta}^{(m-1)} - X_s\widehat{\beta}^{(m-1)}\right) \\
&= \frac{1}{T}\widehat{F}^{(m)\prime}D_s\left(\delta - \widehat{\delta}^{(m-1)}\right) + \frac{1}{T}\widehat{F}^{(m)\prime}X_s\left(\beta - \widehat{\beta}^{(m-1)}\right) \\
&\quad + \frac{1}{T}\widehat{F}^{(m)\prime}F\lambda_s + \frac{1}{T}\widehat{F}^{(m)\prime}\eta_s + \frac{1}{T}\widehat{F}^{(m)\prime}\left(\widehat{A}_s - A_s\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{T}\widehat{F}^{(m)\prime}D_s\left(\delta - \widehat{\delta}^{(m-1)}\right) + \frac{1}{T}\widehat{F}^{(m)\prime}X_s\left(\beta - \widehat{\beta}^{(m-1)}\right) + \frac{1}{T}\widehat{F}^{(m)\prime}\left[F - \widehat{F}^{(m)}H^{(m)}\right]\lambda_s \\
&\quad + \frac{1}{T}\widehat{F}^{(m)\prime}\widehat{F}^{(m)}H^{(m)}\lambda_s + \frac{1}{T}\left[\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right]'\eta_s \\
&\quad + \frac{1}{T}\left[F\left(H^{(m)}\right)^{-1}\right]'\eta_s + \frac{1}{T}\widehat{F}^{(m)\prime}\left(\widehat{A}_s - A_s\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\left\|\widehat{\lambda}_s^{(m)} - H^{(m)}\lambda_s\right\| \\
&\quad \le \left\|\frac{1}{T}\widehat{F}^{(m)\prime}D_s\left(\delta - \widehat{\delta}^{(m-1)}\right)\right\| + \left\|\frac{1}{T}\widehat{F}^{(m)\prime}X_s\left(\beta - \widehat{\beta}^{(m-1)}\right)\right\| \\
&\qquad + \left\|\frac{1}{T}\widehat{F}^{(m)\prime}\left[F - \widehat{F}^{(m)}H^{(m)}\right]\lambda_s\right\| + \left\|\frac{1}{T}\left[\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right]'\eta_s\right\| \\
&\qquad + \left\|\frac{1}{T}\left[F\left(H^{(m)}\right)^{-1}\right]'\eta_s\right\| + \left\|\frac{1}{T}\widehat{F}^{(m)\prime}\left(\widehat{A}_s - A_s\right)\right\| \\
&\quad \le \left\|\frac{\widehat{F}^{(m)\prime}}{T}\right\| \cdot |d_s| \cdot \left\|\sum_{t=T_0}^{T}\delta_t - \widehat{\delta}_t^{(m-1)}\right\| + \left\|\frac{\widehat{F}^{(m)\prime}}{\sqrt{T}}\right\| \cdot \left\|\frac{X_s}{\sqrt{T}}\right\| \cdot \left\|\beta - \widehat{\beta}^{(m-1)}\right\| \\
&\qquad + \left\|\frac{1}{T}\widehat{F}^{(m)\prime}\left[F - \widehat{F}^{(m)}H^{(m)}\right]\right\| \cdot \|\lambda_s\| + \left\|\frac{1}{\sqrt{T}}\left[\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right]'\right\| \cdot \left\|\frac{\eta_s}{\sqrt{T}}\right\| \\
&\qquad + \left\|\left(H^{(m)\prime}\right)^{-1}\right\| \cdot \left\|\frac{1}{T}\sum_{t=1}^{T}f_t\eta_{st}\right\| + \frac{1}{\sqrt{T}}\left\|\frac{\widehat{F}^{(m)\prime}}{\sqrt{T}}\right\| \cdot \left\|\widehat{A}_s - A_s\right\| \\
&\quad = O_P\left(\frac{1}{\sqrt{T}}\right) \cdot \left|\sum_{t=T_0}^{T}\delta_t - \widehat{\delta}_t^{(m-1)}\right| + O_P\left(\left\|\beta - \widehat{\beta}^{(m-1)}\right\|\right) + O_P\left(\frac{1}{\sqrt{ST}}\right) \cdot \left\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right\| \\
&\qquad + O_P\left(\frac{1}{\sqrt{T}}\right) \cdot \left\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right\| + O_P\left(\frac{1}{\sqrt{T}}\right) + O_P\left(\frac{1}{S\sqrt{T}}\right) \\
&\quad = O_P\left(\frac{1}{\sqrt{T}}\right) \cdot \left|\sum_{t=T_0}^{T}\delta_t - \widehat{\delta}_t^{(m-1)}\right| + O_P\left(\left\|\beta - \widehat{\beta}^{(m-1)}\right\|\right) + O_P\left(\frac{1}{\sqrt{T}}\right) \cdot \left\|\widehat{F}^{(m)} - F\left(H^{(m)}\right)^{-1}\right\|,
\end{aligned}
$$

where the second last equality follows from Assumption 2.4.1, Lemma S.1.1, Lemma S.1.10 and the fact that

$$
\begin{aligned}
\mathbb{E}\left[\left\|\frac{1}{T}F'\eta_s\right\|^2\right] &= \frac{1}{T^2}\sum_{t,l=1}^{T}\sum_{k=1}^{r}\mathbb{E}[\eta_{st}f_{t(k)}f_{l(k)}\eta_{sl}] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{k=1}^{r}\mathbb{E}[\eta_{st}^2 f_{t(k)}^2] + \frac{1}{T^2}\sum_{t\neq l}^{T}\sum_{k=1}^{r}\mathbb{E}[\eta_{st}f_{t(k)}f_{l(k)}\eta_{sl}] \\
&\le O_P\left(\frac{1}{T}\right) + \frac{1}{T^2}\sum_{t\neq l}^{T}\sum_{k=1}^{r}10a(|t-l|)^{\frac{\delta}{\delta+2}}\left(\mathbb{E}\left[\left|\eta_{st}f_{t(k)}\right|^{\delta+2}\right]\right)^{\frac{1}{\delta+2}} = O\left(\frac{1}{T}\right),
\end{aligned}
$$

following Lemma S.1.7.

Apply similar argument to (S.17), we can show

$$\frac{1}{S}\left\|\widehat{\Lambda}^{(m)}-\Lambda H^{(m)\prime}\right\|^2=\frac{1}{S}\sum_{s=1}^{S}\left\|\widehat{\lambda}_s^{(m)}-H^{(m)}\lambda_{js}(u)\right\|^2$$
$$=O_P\left(\frac{1}{T}\right)\cdot\left|\sum_{t=T_0}^{T}\delta_t-\widehat{\delta}_t^{(m-1)}\right|^2+O_P\left(\left\|\beta-\widehat{\beta}^{(m-1)}\right\|^2\right)+O_P\left(\frac{1}{T}\right)\cdot\left\|\widehat{F}^{(m)}-F\left(H^{(m)}\right)^{-1}\right\|^2.$$

For the last claim, we write

$$\left\|\left(\frac{\widehat{\Lambda}^{(m)\prime}\widehat{\Lambda}^{(m)}}{S}\right)^{-1}-\left(H^{(m)\prime}\right)^{-1}\left(\frac{\Lambda'\Lambda}{S}\right)^{-1}\left(H^{(m)}\right)^{-1}\right\|$$
$$=\left\|\left(\frac{\widehat{\Lambda}^{(m)\prime}\widehat{\Lambda}^{(m)}}{S}\right)^{-1}-\left(\frac{(\Lambda H^{(m)\prime})'(\Lambda H^{(m)\prime})}{S}\right)^{-1}\right\|$$
$$\leq\left\|\frac{\widehat{\Lambda}^{(m)\prime}\widehat{\Lambda}^{(m)}}{S}-\frac{(\Lambda H^{(m)\prime})'(\Lambda H^{(m)\prime})}{S}\right\|\cdot\left\|\left(\frac{\widehat{\Lambda}^{(m)\prime}\widehat{\Lambda}^{(m)}}{S}\right)^{-1}\right\|\cdot\left\|\left(H^{(m)\prime}\right)^{-1}\left(\frac{\Lambda'\Lambda}{S}\right)^{-1}\left(H^{(m)}\right)^{-1}\right\|$$
$$\leq O_P\left(\frac{1}{T}\right)\cdot\left|\sum_{t=T_0}^{T}\delta_t-\widehat{\delta}_t^{(m-1)}\right|^2+O_P\left(\left\|\beta-\widehat{\beta}^{(m-1)}\right\|^2\right)+O_P\left(\frac{1}{T}\right)\cdot\left\|\widehat{F}^{(m)}-F\left(H^{(m)}\right)^{-1}\right\|^2,$$

where the last line follows from (S.18). ■

S.2 Technical Details for Proofs in Chapter 3

Lemma S.2.1. *Let $\rho(u)=u(\tau-1\{u<0\})$, where $\tau\in(0,1)$. Define a regular sequence of $\rho(u)$ as*

$$\rho_h(u):=\int_{-\infty}^{+\infty}\rho(x)\phi_h(x-u)dx,\quad \phi_h(u):=\sqrt{h/\pi}\exp(-hx^2),$$

for $h=1,2,....$

Let $g(\cdot)$ be the density of a random variable e. Suppose that $g(u)$ is differentiable, $\int|\dot{g}(u)|du<\infty$, and $\rho(u)g(u)\to 0$ as $|u|\to\infty$. Suppose $\mathbb{E}[\ddot{\rho}(e)]$ exists where $\ddot{\rho}(u)$ is considered as a generalized function; suppose also $\int|\ddot{g}(u)|du<\infty$, $\psi(u)\dot{g}(u)\to 0$ when $|u|\to\infty$. Then,

(i) $\rho_h(\cdot),\ h=1,2,...,$ *are differentiable with any order, and in particular,*

$$\dot{\rho}_h(u)=-\int_{-\infty}^{+\infty}\rho(x)\dot{\phi}_h(x-u)dx,\quad \ddot{\rho}_h(u)=-\int_{-\infty}^{+\infty}\rho(x)\ddot{\phi}_h(x-u)dx,$$

and $\dot{\rho}_h(u)\to\psi(u)$ and $\ddot{\rho}_h(u)\to\ddot{\rho}(u)$ as $n\to\infty$ except on a set of measure zero.

(ii) $\sup_{u\in\mathbb{R}}|\rho_h(u)-\rho(u)|\leq Ch^{-1/2}$ *where $C\geq 0$. Consequently, $\rho_h(e)-\rho(e)=O(h^{-1/2})$ almost surely for any random variable e.*

(iii) $\dot{\rho}_h(e)-\psi(e)=O_P(h^{-1/2})$,

(iv) $\mathbb{E}[\ddot{\rho}_h(e) - \ddot{\rho}(e)] = O(h^{-1/2})$.

(v) $\mathbb{E}[\ddot{\rho}_h(e+\epsilon) - \ddot{\rho}_h(e)] = O(h^{-1/2} + |\epsilon|)$ *for any given small* ϵ.

(vi) $\mathbb{E}[\dot{\rho}_h(e+\epsilon) - \dot{\rho}_h(e)] = O(h^{-1/2} + |\epsilon|)$ *for any given small* ϵ.

(vii) $\ddot{\rho}_h(\cdot)$, $h = 1, 2, ...$, *are Lipschitz continuous functions.*

***Proof*.** Lemma S.2.1.(i)-(v) are given as Theorem 2.1 of Dong et al. (2023b), and thus the proof is omitted.

(vi) We first consider

$$\begin{aligned}\left|\mathbb{E}\left[\dot{\rho}(e+\epsilon) - \dot{\rho}(e)\right]\right| &= \left|\int_{-\infty}^{+\infty}\left[\dot{\rho}(u+\epsilon) - \dot{\rho}(u)\right]g(u)du\right| \\ &= \left|\int_{-\infty}^{+\infty}\left[\rho(u+\epsilon) - \rho(u)\right]\dot{g}(u)du\right| \\ &\le \left|\int_{-\infty}^{+\infty}\left|\rho(u+\epsilon) - \rho(u)\right|\cdot\left|\dot{g}(u)\right|du\right| \\ &= |\epsilon|\int_{-\infty}^{+\infty}\left|\dot{g}(u)\right|du = C|\epsilon|,\end{aligned}$$

then, by triangle-inequality, we have

$$\begin{aligned}\left|\mathbb{E}[\dot{\rho}_h(e+\epsilon) - \dot{\rho}_h(e)]\right| \le& \left|\mathbb{E}[\dot{\rho}(e+\epsilon) - \dot{\rho}(e)]\right| + \left|\mathbb{E}[\dot{\rho}_h(e+\epsilon) - \dot{\rho}(e+\epsilon)]\right| + \left|\mathbb{E}[\dot{\rho}_h(e) - \dot{\rho}(e)]\right| \\ =& O(h^{-1/2} + |\epsilon|)|\end{aligned}$$

(vii) Let $u, e \in \mathbb{R}$, by definition of $\ddot{\rho}_h(\cdot)$, we have

$$\begin{aligned}|\ddot{\rho}_h(u+e) - \ddot{\rho}_h(u)| &= \left|\int_{-\infty}^{+\infty}\big(\rho(x+u+e) - \rho(x+u)\big)\ddot{\phi}_h(x)dx\right| \\ &\le \int_{-\infty}^{+\infty}\left|\rho(x+u+e) - \rho(x+u)\right|\cdot\left|\ddot{\phi}_h(x)\right|dx \\ &\le C_1|e|\int_{-\infty}^{+\infty}\left|\ddot{\phi}_h(x)\right|dx = C_2|e|,\end{aligned}$$

where the first inequality follows from Jensen's inequality, and the second one holds as $\rho(\cdot)$ is Lipschitz continuous. Thus, $\ddot{\rho}_h(\cdot)$ is a Lipschitz continuous function. ■

Lemma S.2.2 (Lemma A.1 of Gao (2007))**.** *Suppose that* $\{M_m^n : -\infty < m \le n < +\infty\}$ *are the σ-fields generated by a stationary and α-mixing process* $\{\xi_i\}_{-\infty}^{+\infty}$ *with the mixing coefficient* $a(i)$. *For some positive integers* m, *let* $x_i \in M_{s_i}^{t_i}$ *where* $s_1 < t_1 < s_2 < t_2 < \cdots < s_m < t_m$ *and assume that* $t_i - s_i \ge \tau$ *for all* i *and some* $\tau > 0$. *Assume further that, for some* $p_i > 1, \mathbb{E}\,|x_i|^{p_i} < +\infty,$ *for which* $Q := \sum_{i=1}^{\ell}\frac{1}{p_i} < 1$. *Then we have*

$$\left|\mathbb{E}\left(\Pi_{i=1}^{\ell}x_i\right) - \Pi_{i=1}^{\ell}\mathbb{E}\left(x_i\right)\right| < 10(\ell-1)a(\tau)^{1-Q}\Pi_{i=1}^{\ell}\left(\mathbb{E}\,|x_i|^{p_i}\right)^{\frac{1}{p_i}}.$$

Lemma S.2.3. *Under Assumption 3.3.4, we have*

$$\sum_{t=1}^{T} \big\| \big(\beta_{0,i}(v_{it}) - \beta_{0,mi}(v_{it})\big)' x_{it} \big\|^2 = O_p\big(Tm^{-2\mu}\big).$$

Proof. Let $\Delta_{\beta_i}(v) := \beta_{0,i}(v) - \beta_{0,mi}(v) = \sum_{k=m}^{\infty} b_{i,k}\mathcal{H}_k(v)$ and $\Delta\beta_{0,i(l)}(v) := \sum_{k=m}^{\infty} b_{i(l),k}\mathcal{H}_k(v)dv$ denote the l^{th} element of $\Delta_{\beta_i}(v)$ for $l = 1, ...p$, and $x_{it(l)}$ denote the the l^{th} element of x_{it}. We first note that,

$$\begin{aligned}
\max_{1\le l_1, l_2 \le p} \iint_{\mathcal{R}} \Delta_{\beta,i(l_1)}(v)\Delta_{\beta,i(l_2)}(v)\pi(v)dv &= \sum_{k=m}^{\infty} b_{i(l_1),k} b_{i(l_2),k} \\
&\le \left(\sum_{k=m}^{\infty} b^2_{i(l_1),k}\right)^{1/2} \left(\sum_{k=m}^{\infty} b^2_{i(l_2),k}\right)^{1/2} = O(m^{-2\mu})
\end{aligned}$$

under Assumption 3.3.4. Thus,

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E} \big\| \big(\beta_{0,i}(v_{it}) - \beta_{0,mi}(v_{it})\big)' x_{it} \big\|^2 \\
&\quad = \sum_{t=1}^{T} \iint_{-\infty}^{+\infty} x'\big(\beta_{0,i}(v) - \beta_{0,mi}(v)\big)\big(\beta_{0,i}(v) - \beta_{0,mi}(v)\big)' x' g_{xv}(x,v) dx dv \\
&\quad = \sum_{t=1}^{T} \int_{-\infty}^{+\infty} \sup_v \frac{g_{xv}(x,v)}{\pi(v)} x' \left(\int_{-\infty}^{+\infty} \pi(v)\big(\beta_{0,i}(v) - \beta_{0,mi}(v)\big)\big(\beta_{0,i}(v) - \beta_{0,mi}(v)\big)' dv\right) x dx \\
&\quad = O(m^{-2\mu}) \sum_{t=1}^{T} \int_{-\infty}^{+\infty} \sup_v \frac{g_{xv}(x,v)}{\pi(v)} x' x dx = O(Tm^{-2\mu}).
\end{aligned}$$

■

Lemma S.2.4. *Under the Assumptions of Theorem 3.3.1, we have*

(i) $\left\| T^{-1} \sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) \left(z_{it}^{(0)} z_{it}^{(0)\prime} - z_{it} z_{it}'\right) \right\| = O_P(\sqrt{m/N}) = o_P(1)$,

(ii) $\left\| T^{-1} \sum_{t=1}^{T} \left(\ddot{\rho}_h(\epsilon_{it}) - \mathbb{E}[\ddot{\rho}_h(\epsilon_{it}) | \mathcal{F}_{t-1}]\right) z_{it} z_{it}' \right\| = O_P(m/\sqrt{T}) = o_P(1)$,

(iii) $\left\| T^{-1} \sum_{t=1}^{T} \left(\mathbb{E}[\ddot{\rho}_h(\epsilon_{it}) | \mathcal{F}_{t-1}] z_{it} z_{it}' - g_{it}(0)\mathbb{E}[z_{it} z_{it}']\right) \right\| = O_P(m/\sqrt{T}) = o_P(1)$,

by choosing $h = O_P(T)$.

Proof. (i) We write $z_{it}^{(0)} = z_{it} + (z_{it}^{(0)} - z_{it})$, and it is straightforward that

$$\begin{aligned}
&\sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) \left(z_{it}^{(0)} z_{it}^{(0)\prime} - z_{it} z_{it}'\right) \\
&\qquad = \sum_{t=1}^{T} \ddot{\rho}_h(\epsilon_{it}) \Big[z_{it}\big(z_{it}^{(0)} - z_{it}\big)' + \big(z_{it}^{(0)} - z_{it}\big) z_{it}' + \big(z_{it}^{(0)} - z_{it}\big)\big(z_{it}^{(0)} - z_{it}\big)'\Big].
\end{aligned}$$

The first term can be evaluated as follows:

$$\begin{aligned}
&\left\|\sum_{t=1}^{T}\ddot{\rho}_h(\epsilon_{it})z_{it}\big(z_{it}^{(0)}-z_{it}\big)'\right\|^2\\
&\quad=\sum_{j=0}^{m-1}\left\|\sum_{t=1}^{T}h_j(v_{it})x_{it}\big(\widehat{f}_t^{(0)}-f_{0,t}\big)'\right\|^2+\left\|\sum_{t=1}^{T}\ddot{\rho}_h(\epsilon_{it})f_{0t}\big(\widehat{f}_t^{(0)}-f_{0,t}\big)'\right\|^2\\
&\quad\leq\sum_{j=0}^{m-1}\left(\sum_{t=1}^{T}\left\|h_j(v_{it})x_{it}\right\|^2\sum_{t=1}^{T}\left\|\widehat{f}_t^{(0)}-f_{0,t}\right\|^2\right)+\sum_{t=1}^{T}\left\|\ddot{\rho}_h(\epsilon_{it})f_{0t}\right\|^2\sum_{t=1}^{T}\left\|\widehat{f}_t^{(0)}-f_{0,t}\right\|^2\\
&\quad=O_P(T^2/N)(m+1)=o_P(T^2),
\end{aligned}$$

where the second last equality can be easily obtained by Hölder's inequality and Assumptions 3.3.1 and 3.3.5, while the last equality follows from Assumption 3.3.4. The rest two terms can be be shown in similar manners, and thus the first results of this lemma follows.

(ii) Let $z_{it(j)}$ denotes the j-th element of z_{it} $(j=1,...,mp+r)$, then

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}\Big(\ddot{\rho}_h(\epsilon_{it})-\mathbb{E}[\ddot{\rho}_h(\epsilon_{it})|\mathcal{F}_{t-1}]\Big)z_{it}z_{it}'\right\|^2\\
&=\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\Big\{\mathbb{E}\Big[\big(\ddot{\rho}_h(\epsilon_{it})-\mathbb{E}[\ddot{\rho}_h(\epsilon_{it})|\mathcal{F}_{t-1}]\big)^2\Big|\mathcal{F}_{t-1}\Big]z_{it}z_{it}'z_{it}z_{it}'\Big\}\\
&\quad+\frac{2}{T^2}\sum_{s<t}^{T}\mathbb{E}\Big\{\mathbb{E}\Big[\ddot{\rho}_h(\epsilon_{it})-\mathbb{E}\big[\ddot{\rho}_h(\epsilon_{it})|\mathcal{F}_{t-1}\big]\Big|\mathcal{F}_{t-1}\Big]\big(\ddot{\rho}_h(\epsilon_{is})-\mathbb{E}\big[\ddot{\rho}_h(\epsilon_{is})|\mathcal{F}_{s-1}\big]\big)z_{it}z_{it}'z_{is}z_{is}'\Big\}\\
&\leq\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\Big\{\mathbb{E}\Big[\big(\ddot{\rho}_h(\epsilon_{it})\big)^2\Big|\mathcal{F}_{t-1}\Big]-\big(\mathbb{E}[\ddot{\rho}_h(\epsilon_{it})|\mathcal{F}_{t-1}]\big)^2\Big\}\cdot\sup_i\mathbb{E}\Big[\|z_{it}\|^4\Big]=O\Big(\frac{m^4}{T}\Big),
\end{aligned}$$

as $\sup_i\mathbb{E}[|z_{it}|^4]<\infty$ under Assumption 3.3.1, and

$$\begin{aligned}
\mathbb{E}\Big[\big(\ddot{\rho}_h(\epsilon_{it})\big)^2\Big|\mathcal{F}_{t-1}\Big]&=\int_{-\infty}^{+\infty}\big[\ddot{\rho}_h(u)\big]^2g_{it}(u)du=\int\left[\int\rho(x)\ddot{\phi}_h(x-u)dx\right]^2g_{it}(u)du\\
&=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}\rho(x+u)\ddot{\phi}_h(u)dx\right]^2g_{it}(u)du\\
&=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}\rho\Big(\frac{x}{\sqrt{h}}+u\Big)2h\frac{1}{\sqrt{\pi}}(2x^2-1)e^{-x^2}dx\right]g_{it}(u)du\\
&=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}\Big(\ddot{\rho}(u)\frac{x^2}{h}+O(h^{-2})\Big)2h\frac{1}{\sqrt{\pi}}(2x^2-1)e^{-x^2}dx\right]g_{it}(u)du\\
&=\int_{-\infty}^{+\infty}\left[\ddot{\rho}(u)\int_{-\infty}^{+\infty}\frac{2}{\sqrt{\pi}}x^2(2x^2-1)e^{-x^2}dx\right]g_{it}(u)du+O(h^{-1})\\
&=\int_{-\infty}^{+\infty}\big[\ddot{\rho}(u)\big]^2g_{it}(u)du+O(h^{-1})=\mathbb{E}\big[\big(\ddot{\rho}(\epsilon_{it})\big)^2|\mathcal{F}_{t-1}\big]+O(h^{-1}),
\end{aligned}$$

by Taylor expansion and integrals $\int_{-\infty}^{+\infty}x^j(2x^2-1)e^{-x^2}dx=0$ for $j=0,1,3,...$, and similarly

$\mathbb{E}[\ddot{\rho}_h(\epsilon_{it})|\mathcal{F}_{t-1}] = \mathbb{E}[\ddot{\rho}(\epsilon_{it})|\mathcal{F}_{t-1}] + O(h^{-1})$.

(iii). We note that $\mathbb{E}[\ddot{\rho}(\epsilon_{it})|\mathcal{F}_{t-1}] = g_{it}(0)$, then Lemma S.2.1.(iv) immediately yields that

$$\left\| \frac{1}{T}\sum_{t=1}^{T}\Big(\mathbb{E}[\ddot{\rho}_h(\epsilon_{it})|\mathcal{F}_{t-1}] - g_{it}(0)\Big) z_{it}z_{it}' \right\| = O_P(m/\sqrt{h}) = o_P(1).$$

Furthermore, since $\{z_{it}, t = 1, ..., T\}$ is an α-mixing sequence under Assumption 3.3.3. Using the upper bound provided by Lemma S.2.2, we obtain

$$\mathbb{E}\left\| \frac{1}{T}\sum_{t=1}^{T} g_{it}(0)\Big(z_{it}z_{it}' - \mathbb{E}[z_{it}z_{it}']\Big)\right\|^2 = O\left(\frac{m^2}{T}\right) = o_P(1),$$

which completes the proof. ■

Lemma S.2.5. *Under the Assumptions of Theorem 3.3.1, suppose $\widehat{b}_{m,i}$ and $\widehat{\lambda}_i$ satisfy (3.26) and (3.27), respectively. Then,*

(i) $\sum_{i=1}^{N}\Big(\dot{\rho}_h(\epsilon_{it})\widehat{\lambda}_i - \psi(\epsilon_{it})\lambda_{0,i}\Big) = O_P(\sqrt{N/T} \vee N/T \vee N/\sqrt{h})$,

(ii) $\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x_{it}'\big(\beta_{0,i}(v_{it}) - \widehat{\beta}_{m,i}(v_{it})\big)\widehat{\lambda}_i = O_P(Nm^{-\mu} \vee \sqrt{N/T} \vee N/T)$,

(iii) $\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})f_{0,t}'(\lambda_{0,i} - \widehat{\lambda}_i)\widehat{\lambda}_i = O_P(\sqrt{N/T} \vee N/T)$,

where $\widehat{\beta}_{m,i}(v) = [H_m(v)' \otimes I_p]\widehat{b}_{m,i}$.

Proof. (i) To prove the claim, we consider

$$\sum_{i=1}^{N}\Big(\dot{\rho}_h(\epsilon_{it})\widehat{\lambda}_i - \psi(\epsilon_{it})\lambda_{0,i}\Big) = \sum_{i=1}^{N}\Big(\dot{\rho}_h(\epsilon_{it}) - \psi(\epsilon_{it})\Big)\widehat{\lambda}_i + \sum_{i=1}^{N}\psi(\epsilon_{it})\Big(\widehat{\lambda}_i - \lambda_{0,i}\Big),$$

where the first term on the right-hand side is $O_P(N/\sqrt{h})$ due to Lemma S.2.1.(iii), while the second term

$$\sum_{i=1}^{N}\psi(\epsilon_{it})\Big(\widehat{\lambda}_i - \lambda_{0,i}\Big) = \frac{1}{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\psi(\epsilon_{it})\big(R_{2,i}\big)^{-1}\psi(\epsilon_{is})\Big(f_{0,s} - K_{2,i}w_{is}\Big) + o_P\left(\frac{N}{\sqrt{T}}\right),$$

is $O_P(\sqrt{N/T} \vee N/T)$. To illustrate this, we evaluate the first term

$$\begin{aligned}
&\mathbb{E}\left\| \sum_{i=1}^{N}\sum_{s=1}^{T}\psi(\epsilon_{it})\big(R_{2,i}\big)^{-1}\psi(\epsilon_{is})f_{0,s}\right\|^2 \\
&\quad = \sum_{i_1,i_2=1}^{N}\sum_{s=1}^{T}\mathbb{E}\big[\psi(\epsilon_{i_1,t})\psi(\epsilon_{i_2,t})\psi(\epsilon_{i_1,s})\psi(\epsilon_{i_2,s})f_{0,s}'\big(\Sigma_{i_1,2}\big)^{-1}\big(\Sigma_{i_2,2}\big)^{-1}f_{0,s}\big] \\
&\qquad + \sum_{i_1,i_2=1}^{N}\sum_{s_1\neq s_2}^{T}\mathbb{E}\big[\psi(\epsilon_{i_1,t})\psi(\epsilon_{i_2,t})\psi(\epsilon_{i_1,s_1})\psi(\epsilon_{i_2,s_2})f_{0,s_1}'\big(\Sigma_{i_1,2}\big)^{-1}\big(\Sigma_{i_2,2}\big)^{-1}f_{0,s_2}\big],
\end{aligned}$$

$$=O(NT)+O(N^2),$$

where the last equality is obtained by considering all scenarios given the relationship among s_1, s_2, t. Take the first term in the second last equality as an example

$$\begin{aligned}
&\sum_{i_1,i_2=1}^{N}\sum_{s=1}^{T}\mathbb{E}\big[\psi(\epsilon_{i_1,t})\psi(\epsilon_{i_2,t})\psi(\epsilon_{i_1,s})\psi(\epsilon_{i_2,s})f_{0,s}'\big(\Sigma_{i_1,2}\big)^{-1}\big(\Sigma_{i_2,2}\big)^{-1}f_{0,s}\big]\\
&\quad=\sum_{i_1,i_2=1}^{N}\sum_{s=1}^{t-1}\mathbb{E}\Big[\mathbb{E}\big[\psi(\epsilon_{i_1,t})\psi(\epsilon_{i_2,t})|\mathcal{F}_{t-1}\big]\psi(\epsilon_{i_1,s})\psi(\epsilon_{i_2,s})f_{0,s}'\big(\Sigma_{i_1,2}\big)^{-1}\big(\Sigma_{i_2,2}\big)^{-1}f_{0,s}\Big]\\
&\qquad+\sum_{i_1,i_2=1}^{N}\mathbb{E}\big[\psi(\epsilon_{i_1,t})^2\psi(\epsilon_{i_2,t})^2f_{0,t}'\big(\Sigma_{i_1,2}\big)^{-1}\big(\Sigma_{i_2,2}\big)^{-1}f_{0,t}\big]\\
&\qquad+\sum_{i_1,i_2=1}^{N}\sum_{s=t+1}^{T}\mathbb{E}\Big[\mathbb{E}\big[\psi(\epsilon_{i_1,s})\psi(\epsilon_{i_2,s})|\mathcal{F}_{s-1}\big]\psi(\epsilon_{i_1,t})\psi(\epsilon_{i_2,t})f_{0,s}'\big(\Sigma_{i_1,2}\big)^{-1}\big(\Sigma_{i_2,2}\big)^{-1}f_{0,s}\Big]\\
&\qquad=O(NT)+O(N^2)+O(NT),
\end{aligned}$$

where the last line follows from Assumption 3.3.2. Collecting all the rates, claim (i) follows.

(ii) To show claim (ii), we consider the following two terms $\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x_{it}'\big(\beta_{0,i}(v_{it})-\beta_{0,mi}(v_{it})\big)\widehat{\lambda}_i$ and $\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x_{it}'\big(\beta_{0,mi}(v_{it})-\widehat{\beta}_{m,i}(v_{it})\big)\widehat{\lambda}_i$ separately. For the first term, by Lemma S.2.3,

$$\begin{aligned}
&\Big\|\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x_{it}'\big(\beta_{0,i}(v_{it})-\beta_{0,mi}(v_{it})\big)\widehat{\lambda}_i\Big\|\\
&\quad\leq\Big(\sum_{i=1}^{N}\big\|\big(\beta_{0,i}(v_{it})-\beta_{0,mi}(v_{it})\big)'x_{it}\big\|^2\Big)^{1/2}\Big(\sum_{i=1}^{N}\Big\|\ddot{\rho}_h(\epsilon_{it})\widehat{\lambda}_i\Big\|^2\Big)^{1/2}\\
&\quad=O_P(\sqrt{Nm^{-2\mu}})\cdot O_P(\sqrt{N})=O_P(Nm^{-\mu})
\end{aligned}$$

where the last line can be obtain similar to Lemma S.2.3 under Assumption 3.3.4. Next, we consider the second term

$$\begin{aligned}
&\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x_{it}'\big(\beta_{0,mi}(v_{it})-\widehat{\beta}_{m,i}(v_{it})\big)\widehat{\lambda}_i\\
&\quad=\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x_{it}'\big(\beta_{0,mi}(v_{it})-\widehat{\beta}_{m,i}(v_{it})\big)\big(\widehat{\lambda}_i-\lambda_{0,i}\big)+\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x_{it}'\big(\beta_{0,mi}(v_{it})-\widehat{\beta}_{m,i}(v_{it})\big)\lambda_{0,i}.
\end{aligned}$$

Given (3.20), (3.21) and their rates, it is straightforward to show that the first term on the right-hand side is a negligible term. On the other hand, the second term

$$\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})x_{it}'\big(\beta_{0,mi}(v_{it})-\widehat{\beta}_{m,i}(v_{it})\big)\lambda_{0,i}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\ddot{\rho}_h(\epsilon_{it})\psi(\epsilon_{is})\lambda_{0,i}w_{it}'\big(R_{1,i}\big)^{-1}\Big(w_{is} - K_{1,i}f_{0,s}\Big) \\
&= O_P(\sqrt{N/T}) + O_P(N/T),
\end{aligned}
$$

under Assumption 3.3.2. Collecting the rates for all terms, the proof of (ii) is complete.

(iii) Similar to the proof of (ii), the results of (iii) follows immediately.

■