MONASH University

MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

Estimation of Heterogeneous Treatment Effects Using Quantile Regression with Interactive Fixed Effects

Ruofan Xu
Jiti Gao
Tatsushi Oka
Yoon-Jae Whang

Working Paper No.13/22
August 2022

Estimation of Heterogeneous Treatment Effects Using Quantile Regression with Interactive Fixed Effects*

Ruofan Xu[a], Jiti Gao[a], Tatsushi Oka[b] and Yoon–Jae Whang[c]

[a]Monash University, [b]AI Lab CyberAgent and [c]Seoul National University

August 4, 2022

Abstract

We study the estimation of heterogeneous effects of group-level policies, using quantile regression with interactive fixed effects. Our approach can identify distributional policy effects, particularly effects on inequality, under a type of difference-in-differences assumption. We provide asymptotic properties of our estimators and an inferential method. We apply the model to evaluate the effect of the minimum wage policy on earnings between 1967 and 1980 in the United States. Our results suggest that the minimum wage policy has a significant negative impact on the between-inequality but little effect on the within-inequality.

Keywords: Heterogeneous policy effects; quantile regression; interactive fixed effects.

JEL Codes: C13, C31, J15, J31, J38.

*This paper benefited greatly from our discussions with Dukpa Kim. We are also grateful for comments from Martin Huber, Rustam Ibragimov, Artem Prokhorov, and participants at the Center for Econometrics and Business Analytics (CEBA) talk and seminars. We would also like to thank Ellora Derenoncourt and Claire Montialoux for sharing the data and code for the empirical application. Gao and Oka gratefully acknowledge financial support from the Australian Government through the Australian Research Council's Discovery Projects. All errors are our own.

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 paper proposes a flexible yet practically simple method 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, using cross-sectional data separately, we estimate the quantile regression model, introduced by Koenker and Bassett (1978), to obtain regression coefficient estimators for each pair of groups and time. Second, we employ a group-level panel data model to explain variations in the quantile regression estimators while controlling for interactive fixed effects, which parsimoniously capture the complex group-and time-unobserved effects. Our analysis complements and extends the work of Chetverikov, Larsen and Palmer (2016), who first propose a two-step estimation method for a quantile 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 proposed estimation method has several advantages. First, a distinguishing feature is the ability to capture heterogeneous policy effects through the interplay of policy variables and individual observed and unobserved characteristics. Quantile regression allows the marginal effects of covariates to vary depending on the quantile levels. Thus, our approach can study how group-level policies affect the marginal effects differently among observationally identical individuals, and uncover policy effects depending on individual unobserved heterogeneities. In addition, we can document how policies affect outcomes through the interaction of changes in marginal effects with individual observed heterogeneity under the 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; Juhn et al., 1993; Buchinsky, 1994). Alternatively, the difference between two quantiles conditional on different covariates can be interpreted as a between-inequality measure. We show that our model captures the policy effects on these inequality measures and the identification of policy effects is established under a type of difference-in-difference assumption. We also provide the asymptotic property of our estimator of policy effects on within- and between-inequality.

Finally, the second step of our approach controls for unobserved group and time effects through interactive fixed effects. The interactive fixed effects account for time-varying unobserved common shocks with distinct impacts across groups by a factor structure and include the standard additive time and group effects as a special case. In the panel data literature, a factor structure has been employed to account for unobserved structures in panel data (Bai and Ng, 2002; Bai, 2003) and extended to control for interactive fixed effects (Pesaran, 2006; Bai, 2009; Moon and Weidner, 2015, 2017). The interactive fixed effects approach has been employed to identify average causal effects using a synthetic control approach (Abadie et al., 2010; 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.

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 1967 Labor Standards Act 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. 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 using sub-samples based on workers' characteristics. Our method estimates such heterogeneous policy effects, particularly on racial income inequality, under a unified framework. 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 affect industries differently. Our estimation results support the core conclusion of Derenoncourt and Montialoux (2021) and provide additional finding that the minimum wage policy has a significant negative impact on between-inequality but little effect on within-inequality.

This study falls within a broad range of research, accounting for observed and unobserved heterogeneity in the panel data. The classical literature considers a mean regression model with random coefficients (Swamy, 1970; Hsiao, 1974, 1975; Amemiya, 1978; Wooldridge, 2005; 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). Applying a two-step approach with a two-way fixed effects model, Oka and Yamada (2021) empirically studied recent minimum wage policy effects. Our study provides a way to control for more complex unobserved heterogeneity and presents identification conditions for heterogeneous policy effects. In the absence of the group structure in the data, the literature considers panel quantile regression models with fixed effects (Koenker, 2004; Kato et al., 2012; Galvao and Kato, 2016), correlated random effects (Abrevaya and Dahl, 2008; Arellano and Bonhomme, 2016), and interactive fixed effects (Harding and Lamarche, 2014; Ando and Bai, 2020; Chen et al., 2021; Ma 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 an α-mixing-type dependency across groups and time.

Our approach applies to a broad range of settings, including labor, 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, Bitler et al. (2006) for welfare reforms on earnings, and Finkelstein and McKnight (2008) for the social insurance system on out-of-pocket spending, among others. The method proposed in this study can be used for analysis. We make R codes publicly available to facilitate empirical studies.

The paper proceeds as follows. Section 2 describes the model and discusses identification of policy effects. Section 3 explains estimation methods and Section 4 presents the asymptotic properties of the estimators. Section 5 applies the proposed method to analyse the effect of minimum wage on earning under the 1967 Labor Standards Act. Section 6 examines the finite sample properties of the estimator through Monte Carlo simulation. Section 7 concludes. Appendix A provides the proofs of the main results and technical conditions and results, while all the preliminary lemmas and their proofs are given in the online appendix available from the first author.

2 Model and Identification

In this section, we first provide the model setup. We subsequently introduce policy parameters and establish their identification. Before preceding, we introduce some notations. Throughout the paper, $\|\cdot\|$ denotes 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. We denote by $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$ for scalars a, b. Let C_f, c_f, C_M, c_M be some pre-determined positive real numbers which are independent of the sample.

2.1 Data and Model

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 with the sample size N_{st}, given group $s = 1, \dots S$ and time $t = 1, \dots, T$. 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, suppose that we observe a $d_w \times 1$ vector of group-level covariates w_{st} and d_{st} is a dummy variable taking 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. Let $G_{st}(y|z_{ist}) := \Pr\{y_{ist} \leq y|z_{ist}\}$ be the conditional distribution function of y_{ist} given z_{ist} for each pair of group and time (s, t), and we define the corresponding conditional quantile function as $Q_{st}(u|z_{ist}) := \inf\{y : G_{st}(y|z_{ist}) \geq u\}$ for $u \in \mathcal{U} \subset (0, 1)$. Following Koenker and Bassett (1978), we consider the u^{th} quantile regression, given by

$$Q_{st}(u|z_{ist}) = z_{ist}'\alpha_{st}(u), \tag{1}$$

[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.

where $\alpha_{st}(u)$ is a $J \times 1$ vector of quantile regression coefficients.[2] The coefficients $\alpha_{st}(u)$ can capture heterogeneous marginal effects depending on individual unobserved heterogeneity, in that the marginal effects can vary depending on u among observationally equivalent individuals. Moreover, the marginal effects are allowed to vary across groups and time.

Even when the model in (1) is miss-specified, Angrist et al. (2006) show that the quantile regression is the best predictor in the L_2 sense. Moreover, when the underlying structural model depends on multi-dimention unobservables, Sasaki (2015) shows that the quantile regression coefficients can be interpreted as the marginal effects averaged over the unobserved variables under some regularity conditions. Thus, quantile regression coefficients can succinctly summarize the heterogeneous marginal effect of regressors on the outcome.

To characterize group and time variations in the coefficients $\alpha_{st}(u) \equiv [\alpha_{1st}(u), \ldots, \alpha_{Jst}(u)]'$, we consider a linear panel regression model with interactive fixed effects, for each $j = 1, \ldots, J$,

$$\alpha_{jst}(u) = \delta_{jt}(u)d_{st} + w_{st}'\gamma_j(u) + f_{jt}(u)'\lambda_{js}(u) + \eta_{jst}(u), \tag{2}$$

where $\delta_{jt}(u)$ is a scalar coefficient for the policy effect at time t, $\gamma_j(u)$ is a $(K-1) \times 1$ vector of coefficients, $f_{jt}(u)$ is an $r \times 1$ vector of unobservable group-level factors, $\lambda_{js}(u)$ is the corresponding factor loading vector, and $\eta_{jst}(u)$ is an idiosyncratic error. In the above regression, $\delta_{jt}(u)$, $\gamma_j(u)$, $f_{jt}(u)$, and $\lambda_{js}(u)$ are unknown coefficients to be estimated. The interactive fixed effects structure $f_{jt}(u)'\lambda_{js}(u)$ can account for unobserved group and time effects. For example, this model structure can capture 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 $\lambda_{js}(u) = [\phi_{js}(u), 1]'$.

If individual covariate z_{ist} only contains a constant term and the group and time fixed effects are additive, then model (1)-(2) is reduced to the two-way fixed effects model for unconditional quantile, which has been used for estimating distributional policy effects in empirical studies (e.g. Angrist and Lang, 2004).

[2] For notational simplicity, we write $Q_{st}(u|z_{ist}) \equiv Q_{st}(u|z_{ist}, \alpha_{st}(u))$.

2.2 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. Given a binary group-level policy, 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

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

In what follows, we assume that a group-level policy 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 ($T_0 \leq t$). Also, let $d_s = 1$ if group s is treated after T_0 and 0 otherwise.

Given treatment status $d_{st} = d \in \{0, 1\}$, we define $Q_{st}^d(u|z)$ as the u^{th} conditonal quantile of the potential outcome y_{ist}^d given $z_{ist} = z$. For fixed u and z, the conditional quantile $Q_{st}^d(u|z)$ can vary across groups and time. 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

$$\Delta_t^{AQTT}(u|z) := \mathbb{E}[Q_{st}^1(u|z) - Q_{st}^0(u|z)|d_s = 1].$$

Since AQTT is a map $(u, z) \mapsto \Delta_t^{AQTT}(u|z)$, we can interpret AQTT as a measure of policy effects that vary depending on individual-level observed and unobserved heterogeneity. Also, AQTT is time-varying policy effects averaged over groups. Arellano and Bonhomme (2016) use a similar concept to measure averaged derivative of nonlinear model.

As an alternative measure, we consider spreads of conditional quantile functions to quantify inequality within and between collections of individuals characterized 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 at time $t \geq T_0$ is defined as the spread of conditional quantiles:

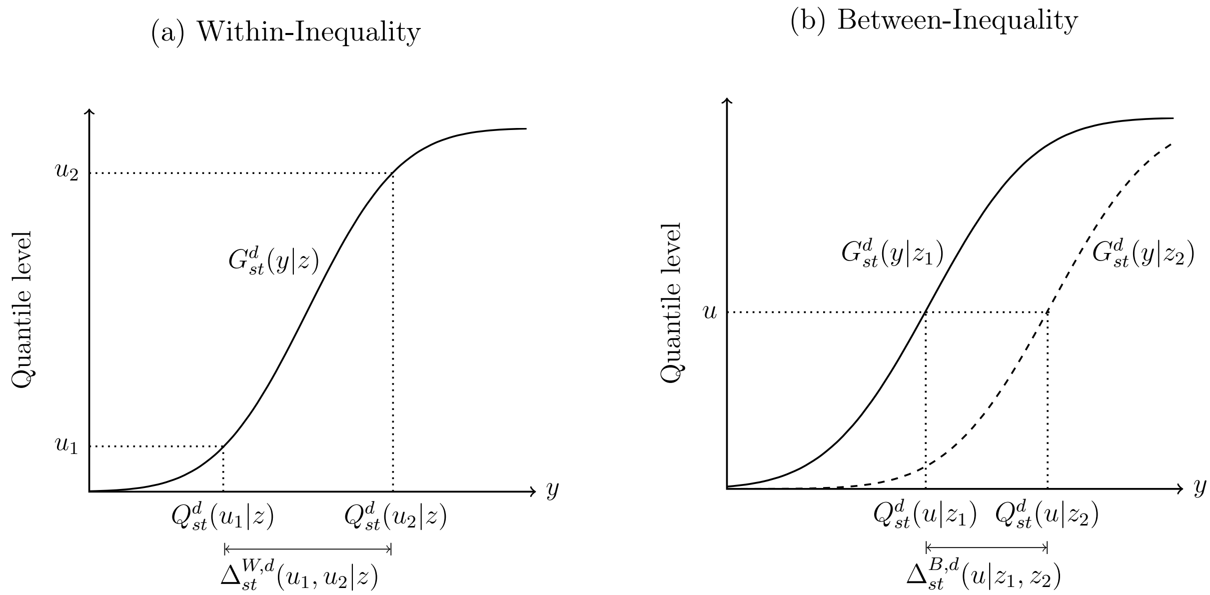
$$\Delta_{st}^{W,d}(u_1, u_2|z) := Q_{st}^d(u_2|z) - Q_{st}^d(u_1|z).$$

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

$$\Delta_{st}^{B,d}(u|z_1, z_2) := Q_{st}^d(u|z_2) - Q_{st}^d(u|z_1).$$

Figure 1 illustrates the within- and between-inequality. 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.

Figure 1 Inequality Measures



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 set of individual attributes z_1, z_2, given the same quantile level u, under the treatment status d.

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:

$$\begin{aligned}\dot{\Delta}_t^W(u_1,u_2|z) &:= \mathbb{E}\big[\Delta_{st}^{W,1}(u_1,u_2|z)-\Delta_{st}^{W,0}(u_1,u_2|z)|d_s=1\big],\\ \dot{\Delta}_t^B(u|z_1,z_2) &:= \mathbb{E}\big[\Delta_{st}^{B,1}(u|z_1,z_2)-\Delta_{st}^{B,0}(u|z_1,z_2)|d_s=1\big].\end{aligned}$$

2.3 Identification

For each $d \in \{0,1\}$, the u^{th} conditonal quantile of the potential outcome y_{ist}^d is given by

$$Q_{st}^d(u|z_{ist}) = z_{ist}'\alpha_{st}^d(u).$$

Here, the regression coefficients $\alpha_{st}^d(u)$ can be interpreted as potential marginal effects under treatment status d. To relate the treatment to the potential marginal effects, we consider a linear regression model for the j^{th} element of $\alpha_{st}^d(u)$, given $d_{st}=d$:

$$\alpha_{jst}^d(u) = d\Delta_{jst}(u) + w_{st}'\gamma_j(u) + f_{jt}(u)'\lambda_{js}(u) + \eta_{jst}^d(u), \tag{3}$$

where $\Delta_{jst}(u)$ is a scalar random policy effects, $\gamma_j(u)$ is a coefficients vector, and $\eta_{jst}^d(u)$ is an error term specific to treatment status d.

We now present the identification result for group-level policy effects for finite T, while the identification result extends to the case in which T goes to infinity by simply taking limits. Under a similar setup, Gobillon and Magnac (2016) prove the identification of average policy effects, using the mean regression model. Let $W_s := [w_{s1},\ldots,w_{sT}]'$, $F_j(u) := [f_{j1}(u),\ldots,f_{jT}(u)]'$, and $M_{F_j}(u) := I_T - T^{-1}F_j(u)F_j(u)'$. We make the following assumptions:

Assumption 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 $\|z_{ist}\| < C_M$.

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

(iii) In the neighbourhood of the conditional quantiles of interest $\mathcal{U}$, for all $z \in \mathcal{Z}$, the conditional quantile $Q_{st}(\cdot|z)$ of interest satisfies the quantile regression form in (1) and the conditional distribution function $G_{st}(\cdot|z)$ has a conditional density $g_{st}(\cdot|z)$, which is uniformly bounded away from 0 and ∞.

Assumption 2.2. *For each $u \in \mathcal{U}$, $j = 1, ..., J$ and $s \geq 1$,*

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

(ii) $\mathbb{E}[V_s' M_{F_j}(u) V_s]$ and $\mathbb{E}[W_s' M_{F_j}(u) W_s]$ are positive definite, where $V_s := [v_{s1}, ..., v_{sT}]'$ and $v_{st} := [d_{st}, w_{st}']'$.

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

(ii) $\mathbb{E}[\Delta_{jst}(u)|d_s = 1, W_s, \lambda_{js}(u), F_j(u)] = \mathbb{E}[\Delta_{jst}(u)|d_s = 1]$.

Assumption 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 2.1.(ii)-(iii) are standard in the quantile regression literature. Assumption 2.2 provides the identification conditions of the regression coefficients, factor and loadings. Specifically, Assumption 2.2.(i) guarantees the identification of factor and the loadings up to an orthogonal rotation matrix (Bai, 2009; Bai and Li, 2014; Jiang et al., 2021). Assumption 2.2.(ii) ensures the identifiability of the regression coefficients. Assumption 2.3.(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. This condition implies a type of parallel-trend assumption. Assumption 2.3 requires that the random policy effect $\Delta_{jst}(u)$ is mean-independent of group-level regressors and the factor structure.

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.

Theorem 2.1. *Let* $\delta_t(u) := [\delta_{1t}(u), \ldots, \delta_{Jt}(u)]'$ *with* δ_{jt} *in Equation (2) for* $j = 1, \ldots, J$. *Suppose that Assumptions 2.1-2.3 hold. Then, for* $t \geq T_0$ *and for each* $(u, z) \in \mathcal{U} \times \mathcal{Z}$,

$$\Delta_t^{AQTT}(u|z) = z'\delta_t(u), \tag{4}$$

and, for $u_1, u_2 \in \mathcal{U}$ *and* $z_1, z_2 \in \mathcal{Z}$,

$$\dot{\Delta}_t^W(u_1, u_2|z) = z'\big(\delta_t(u_2) - \delta_t(u_1)\big) \qquad \text{and} \qquad \dot{\Delta}_t^B(u|z_1, z_2) = (z_2 - z_1)'\delta_t(u).$$

The result in Theorem 2.1 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.

3 Estimation

For estimation and inference purpose, we consider a general regression model of (2) as follows:

$$A_{js}(u) = X_s\beta_j(u) + F_j(u)\lambda_{js}(u) + \eta_{js}(u), \tag{5}$$

where $A_{js}(u) := [\alpha_{js1}(u), ..., \alpha_{jsT}(u)]'$ and $\eta_{js}(u) := [\eta_{js1}(u), ..., \eta_{jsT}(u)]'$ are $T \times 1$ vectors, $X_s := [x_{s1}, ..., x_{sT}]'$ is a $T \times K$ matrix of group-level covariates, $F_j(u)$ is $T \times r$ matrix of unobservable factors defined above Assumption 2.1.

Below we propose an estimation approach for model (1) and (5). For the estimation of interactive fixed effects, we adopt a common practice of imposing the normalization restrictions in Assumption 2.2.(i), which ensure the identification of $F_j(u)$ and the factor loadings $\Lambda_j(u) := [\lambda_{j1}(u), ..., \lambda_{jS}(u)]'$ up to an orthogonal rotation matrix.

For each $u \in \mathcal{U}$ and $j = 1, ..., J$, we obtain the estimator of $\big(\beta_j(u), F_j(u), \Lambda_j(u)\big)$, using the following two-step estimation procedure.

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 minimization problem:

$$\min_{a \in \mathbb{R}^J} \sum_{i=1}^{N_{st}} \varrho_u(y_{ist} - z_{ist}' a), \tag{6}$$

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

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

$$\mathrm{SSR}_u\big(\beta_j, F_j, \Lambda_j\big) := \sum_{s=1}^{S} \big\|\widehat{A}_{js}(u) - X_s\beta_j - F_j\lambda_{js}\big\|^2,$$

with the normalization condition in Assumption 2.2. The least squares estimators are obtained using an iterated procedure as follows:

1. As an initial estimator of $\beta_j(u)$, we use the least squares estimator without the factor components: $\widehat{\beta}_j^{(0)}(u) := (\sum_{s=1}^S X_s' X_s)^{-1} \sum_{s=1}^S X_s' \widehat{A}_{js}(u)$.

2. Given $\widehat{\beta}_j^{(m-1)}(u)$ for $m \geq 1$, we solve $\min_{(F_j, \Lambda_j)} \mathrm{SSR}_u\big(\widehat{\beta}_j^{(m-1)}(u), F_j, \Lambda_j\big)$, applying the principle component analysis (PCA) with the normalization condition in Assumption 2.2, and obtain the estimator $\big(\widehat{F}_j^{(m)}(u), \widehat{\Lambda}_j^{(m)}(u)\big)$.

3. Given $\big(\widehat{F}_j^{(m)}(u), \widehat{\Lambda}_j^{(m)}(u)\big)$, we solve $\min_{\beta_j} \mathrm{SSR}_u\big(\beta_j, \widehat{F}_j^{(m)}(u), \widehat{\Lambda}_j^{(m)}(u)\big)$, and obtain $\widehat{\beta}_j^{(m)}(u) := (\sum_{s=1}^S X_s' M_{\widehat{F}_j}^{(m)}(u) X_s)^{-1} \sum_{s=1}^S X_s' M_{\widehat{F}_j}^{(m)}(u) \widehat{A}_{js}(u)$, where $M_{\widehat{F}_j}(u) := I_T - \widehat{F}_j(u)\widehat{F}_j(u)'/T$.

4. Repeat 2-3 until numerical convergence is reached. Specifically, we stop the algorithm if $\big\|\widehat{\beta}_j^{(m)}(u) - \widehat{\beta}_j^{(m-1)}(u)\big\| \leq 10^{-5}$ 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 which minimizes the modified eigen-ratio criterion of Casas et al. (2021) as follows:

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

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

$$\widehat{L}^{(m)}_j:=\frac{1}{ST}\sum_{s=1}^{S}\left(\widehat{A}_{js}(u)-X_s\widehat{\beta}^{(m-1)}_j(u)\right)\left(\widehat{A}_{js}(u)-X_s\widehat{\beta}^{(m-1)}_j(u)\right)',\tag{7}$$

in descending order. 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 Sections 5-6.

4 Asymptotic Properties

In this section, we first introduce some assumptions and then present asymptotic properties of our estimators.

4.1 Assumptions

In this subsection, we provide key assumptions required for deriving asymptotic properties of the recursive estimator along with necessary explanations. Additional technical assumptions required for the asymptotic distribution are presented in Appendix A.1. Also, 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.

Assumption 4.1. *For all $(s,t)\in\{1,...,S\}\times\{1,...,T\}$ and for each $u\in\mathcal{U}$ and $j=1,...,J$,*

(i) $\mathbb{E}[||x_{st}||^4]\le C_M$, $\mathbb{E}[||f_{jt}(u)||^4]\le C_M$, $\mathbb{E}[||\lambda_{js}(u)||^4]\le C_M$, *and* $\mathbb{E}[|\eta_{jst}(u)|^4]\le C_M$.

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

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

(iv) 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 4.1.(i) requires standard moment conditions for our analysis. Assumption 4.1.(ii) states the identification assumption for factors and loadings. Assumption 4.1.(iii) and (iv) 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, 2003, 2009; Jiang et al., 2021).

Assumption 4.2. *For any fixed $u \in \mathcal{U}$ and $j = 1, ..., J$,*

(i) for $s = 1, ..., S$, the random sequence $\{\ell_{jst}(u) := (x'_{st}, f'_{jt}(u), \eta_{jst}(u)) : t \geq 1\}$ 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 $\sum_{s,g,k,m=1}^{S} \sum_{t \neq l}^{T} a_{sgkm}(|t-l|)^{\delta/(4+\delta)} = O\left(S^2 T\right)$.

Assumption 4.2 uses the notation of "α-mixing" for panel data (e.g., Chen et al., 2012; Jiang et al., 2021) to capture 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).

Assumption 4.3. *For any fixed $u \in \mathcal{U}$ and for $s = 1, ..., S$ and $t = 1, ..., T$, define the support $B_{st}(u, c_f) := (z'_{1st}\alpha_{st}(u) - c_f, z'_{1st}\alpha_{st}(u) + c_f)$ for some $c_f > 0$. Then,*

(i) the conditional density function $g_{st}(\cdot)$ is continuously differentiable on $B_{st}(u, c_f)$ with the derivative g'_{st} satisfying $|g'_{st}(y)| \leq C_f$ for all $y \in B_{st}(u, c_f)$ and $|g'_{st}(z'_{1st}\alpha_{st}(u))| \geq c_f$.

(ii) $g_{st}(y) \leq C_f$ *for all* $y \in B_{st}(u, c_f)$ *and* $g_{st}(z'_{1st}\alpha_{st}(u)) \geq c_f$ *for some* $c_f > 0$*.*

Assumption 4.4. *As* $S, T \to \infty$*, we have* $(ST)^{3/4}(\ln(N_{\min})/N_{\min})^{1/2} \leq C_M$*, where* $N_{\min} := \min\{N_{st}, s = 1, ..., S, t = 1, ..., T\}$*.*

Assumption 4.5. *For any fixed* $u \in \mathcal{U}$ *and* $j = 1, ..., J$*, there exist some positive definite matrics* Σ_x *and* Σ_{xF}*, such that, as* $S, T \to \infty$ *jointly, (i)* $(ST)^{-1}\sum_{s=1}^{S} X'_s X_s \xrightarrow{p} \Sigma_x$*, and (ii)* $\inf_{F:T^{-1}F'F=I_r}(ST)^{-1}\sum_{s=1}^{S} X'_s M_F X_s \xrightarrow{p} \Sigma_{xF}$*.*

Assumptions 4.3 is a set of mild regularity conditions that are typically imposed in the quantile regression literature. Assumption 4.4 requires that the number of observations 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), this assumption imposes a more explicit yet comparable growth rate, which is necessary in analyzing the limiting property of the interactive-effects estimator. Assumption 4.5 guarantees that the inverse matrix in the initial estimator $\widehat{\beta}_j^{(0)}(u)$ and the recursive estimator at each positive step $\widehat{\beta}_j^{(m)}(u)$ are well-defined, so that the estimation in the second step is valid.

4.2 Consistency and Limiting Distribution

In this section, we establish asymptotic properties for the recursive estimator $\widehat{\beta}_j^{(m)}(u)$. We derive the consistency of the regression coefficient estimator, and establish the joint central limit theorem for the recursive regression coefficient estimator. For empirical analysis, we derive the corresponding consistent estimator of the empirical limiting distribution; detailed expressions are given in Corollary A.1 of Appendix A.2.

Proposition 4.1. *Suppose that Assumptions 2.1, 4.1-4.5 and A.1 hold. Define*

$$\vartheta_j(u) := \frac{1}{S^2T^2}\mathbb{E}\left\|\sum_{s=1}^{S} X'_s F_j(u)\lambda_{js}(u)\right\|^2 \quad \textit{and} \quad \xi_{ST} := \frac{1}{S^2T^2}\mathbb{E}\left\|\sum_{s=1}^{S} X_s X'_s\right\|^2.$$

Suppose $\vartheta_j(u) = o(1)$ *and* $\xi_{ST} = O(1)$*, then, for any fixed* $u \in \mathcal{U}$*,* $j = 1, ..., J$*, and* $m \geq 1$*, we have, as* $S, T \to \infty$*,*

(i) $\|\widehat{\beta}_j^{(0)}(u)-\beta_j(u)\| = O_P\big(\max\{(ST)^{-1/2},(\vartheta_j(u))^{1/2}\}\big) = o_P(1)$,

(ii) $\|\widehat{\beta}_j^{(m)}(u)-\beta_j(u)\| = O_P\Big(\max\Big((ST)^{-1/2},S^{-1},T^{-1},(\vartheta_j(u))^{2^m}\Big)\Big) = o_P(1)$.

Proposition 4.1 implies that the asymptotic order of $\|\widehat{\beta}_j^{(m)}(u)-\beta_j(u)\|$ not only depends on the sample size, but also the level of interdependence between the factor component and the individual-level regressors. In particular, the initial estimator $\widehat{\beta}_j^{(0)}(u)$ is $\sqrt{ST}$-consistent when $\vartheta_j(u)=O\big((ST)^{-1}\big)$, which holds when the factor component is independent of the regressors X and satisfies $\mathbb{E}[\lambda_{js}(u)'f_{jt}(u)]=0$. The condition $\xi_{ST}=O(1)$ follows trivially when $\max_{s,g,t,l}|\mathbb{E}(x'_{st}x_{gt}x'_{sl}x_{gl})|\le C<\infty$.

We now present the joint asymptotic distribution of the estimator after some iterations. Let $\widehat{\beta}^{(m)}(u)$ be the estimator of $\beta(u):=[\beta'_1(u),...,\beta'_J(u)]'$ at the m^{th} iteration.

Theorem 4.1. *Suppose that Assumptions 2.1, 4.1-4.5, A.1 and A.2 hold, and we additionally assume that $\vartheta_j(u)=O\big((ST)^{-1}\big)$, $\xi_{ST}=O(1)$, and $T/S\to\kappa>0$. Then, for any $u_1,u_2\in\mathcal{U}$ and $m\ge 1$, we have, as $T,S\to\infty$,*

$$\sqrt{ST}\begin{pmatrix}\widehat{\beta}^{(m)}(u_1)-\beta(u_1)\\ \widehat{\beta}^{(m)}(u_2)-\beta(u_2)\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(\begin{bmatrix}B^{(m)}(u_1)\\ B^{(m)}(u_2)\end{bmatrix},\Sigma^{(m)}(u_1,u_2)\right),$$

where $B^{(m)}(u)$ is the bounded asymptotic bias given by (A.1) in Appendix A.2, and $\Sigma^{(m)}(u_1,u_2)$ is the positive definite, asymptotic covariance matrix, defined in Assumption A.2 in Appendix A.1.

In view of this theorem, under general cases, the asymptotic properties of the recursive estimator $\widehat{\beta}^{(m)}(u)$ depend on the quantiles, and the relative sample size between individual-level and group-level, in addition to the consistency of the initial estimator $\widehat{\beta}^{(0)}(u)$ and the information from the iterative steps, on which the recursive estimator in the mean regression model with interactive fixed effects depends (Jiang et al., 2021).

An immediate corollary of the above result is the asymptotic joint distribution for the estimators of the time-varying policy effect parameters $\big(\delta_t(u_1),\delta_t(u_2)\big)$. The limit distribution is also useful for inference of the AQTT and inequality measures. Letting S_t be the selection matrix such that $\big(\delta_t(u_1)', \delta_t(u_2)'\big)' = S_t\big(\beta(u)', \beta(u_2)')\big)'$, we easily obtain the asymptotic joint distribution of the estimated policy effect from Theorem 4.1 as in the corollary below, and thus its proof is omitted.

Corollary 4.1. *Suppose that the assumptions of Theorem 4.1 hold. Then, for any $u_1, u_2 \in \mathcal{U}$, $t = 1, ..., T$ and $m \geq 1$, as $S, T \to \infty$,*

$$\sqrt{ST}\begin{pmatrix} \widehat{\delta}_t^{(m)}(u_1) - \delta_t(u_1) \\ \widehat{\delta}_t^{(m)}(u_2) - \delta_t(u_2) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(S_t \begin{bmatrix} B^{(m)}(u_1) \\ B^{(m)}(u_2) \end{bmatrix}, S_t\Sigma^{(m)}(u_1, u_2)S_t'\right).$$

Using Theorem 4.1 and Corollary 4.1, we can easily obtain the asymptotic distribution of the estimator of the AQTT, $\Delta_t^{AQTT}(u|z)$, and the estimator of changes in between- and within-inequality, $\dot{\Delta}_t^W(u_1, u_2|z)$ and $\dot{\Delta}_t^B(u|z_1, z_2)$.

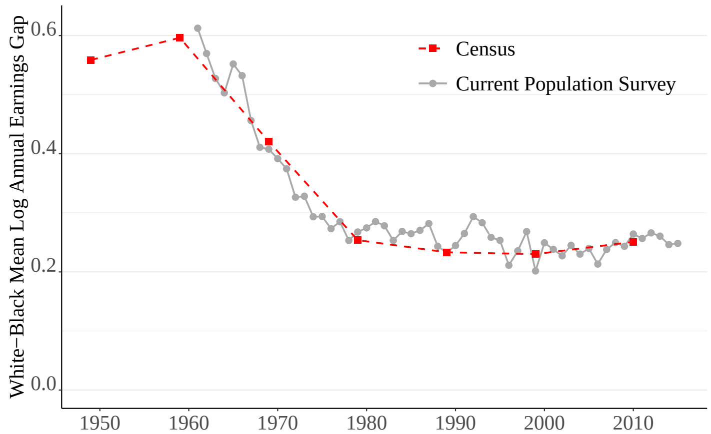
Corollary A.1 in Appendix A.2 provides the estimator of the asymptotic bias $B(u)$ and the asymptotic covariance matrix $\Sigma(u_1, u_2)$ and establishes their consistency. To simplify the presentation, attention is paid to heteroskedasticity in both space and time dimensions, assuming no correlation in either dimensions when the number of iterations $m \to \infty$.

5 Empirical Analysis

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, 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 (Freeman et al., 1973; Smith and Welch, 1984) and improvements in education (Smith and Welch, 1977, 1989; Card and Krueger, 1992).

Figure 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.

Recently, Derenoncourt and Montialoux (2021) put forward a new explanation: the extension of the federal minimum wage to some industries. The Fair Labor Standards Act of 1966 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 7 treated and 8 control industries were subject to the minimum wage under the 1966 and 1938 Fair Labor Standards Act, respectively.

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:

$$y_{ist} = \delta_0 + \delta_t d_{st} + z_{ist}'\beta + \phi_s + \nu_t + \eta_{ist}, \tag{8}$$

for worker i in industry $s = 1, \dots, 15$ and time $t = 1961, 1963 \dots, 1980$. Here, y_{ist} is the log

[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.

annual earning 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 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 (8), including log annual wages or its unconditional quantiles, and also selecting sub-samples based on workers' characteristics.

5.2 Model and Estimation Results

We analyze 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 (8), we consider the following u^{th} quantile regression model in (1) with $\alpha_{st}(u) = [\alpha_{1st}(u), \ldots, \alpha_{Jst}(u)]'$, where, for $j = 1, \ldots, J$,

$$\alpha_{jst}(u) = \delta_{j0}(u) + \delta_{jt}(u)d_{st} + f_{jt}(u)'\lambda_{js}(u) + \eta_{jst}(u). \tag{9}$$

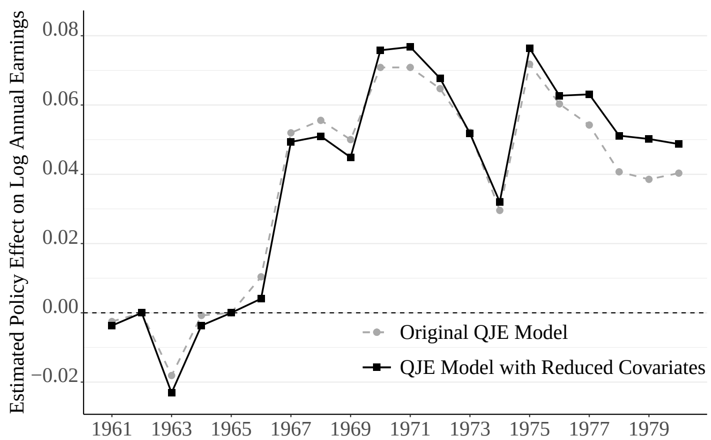
Here, a set of coefficients $\{\delta_{jt}(u)\}_{t=1967}^{1980}$ measures the time-varying policy effect.

For simplicity of interpretation, our covariates z_{ist} is a subset of the original covariates as we treat some of them 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

[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).

schooling and experience and their squares, 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 (8) as the one in Derenoncourt and Montialoux (2021). See Figure 3.

Figure 3 Estimation Results of Mean Regression



Notes: We plot the estimates of time-varying policy effect δ_t in Model (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).

In Figure 4, we present time-varying policy effects $\delta_{jt}(u)$ in (9) with 95% confidence intervals. Panel (a) reports the effects on the intercept coefficients, which correspond to white, male, full-time workers. The estimates that are statistically different from zero are found at quantiles of 0.3 and 0.7 with their magnitude ranging from 30-70% (0.3-0.7 log points), while most of the other estimates are statistically insignificant. Panel (b) shows statistically significant positive policy effects for black workers, which are 20-25% (0.2-0.25 log points) and 10-15% (0.1-0.15 log points) at the 0.1th and 0.3th conditional quantiles, respectively. For other quantiles in Panel (b), the estimated policy effects are much smaller or statistically insignificant. Panel (d) reports the estimated policy effects on the coefficient of the part-time dummy. Most of the estimates are positive but not statistically different from zero.[6]

Panel (c) of Figure 4 presents the estimated policy effects on the female dummy's coefficient. The effects in the late 1970s are positive with a magnitude of 15-20%, 15%, 10% and 10% at

[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.

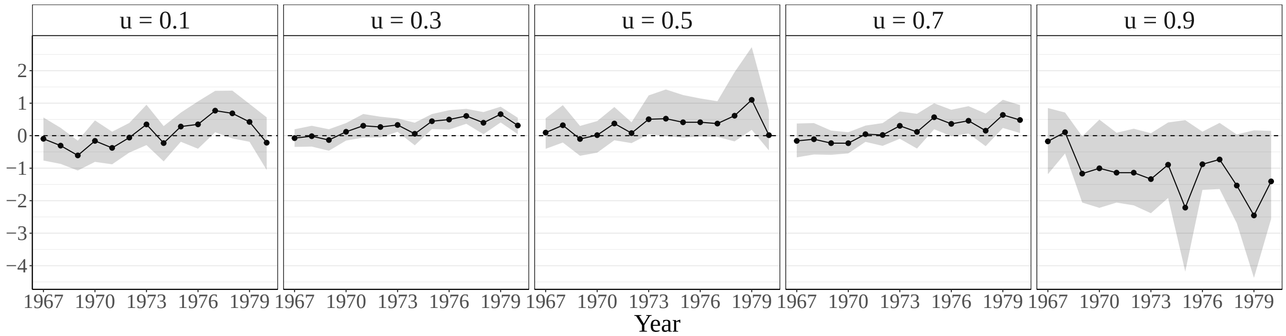
[6]This is partly because part-time workers only account for 5% to 30% of the sample by industry and year.

quantiles of 0.1, 0.3, 0.5, and 0.7, respectively, and the estimates are statistically significant. 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.[7] Although the gender gap of median wages for full-time, full-year workers was unchanged over the 1960-1970s (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-run positive effects even at 0.5th and 0.7th quantiles conditional on individual attributes.

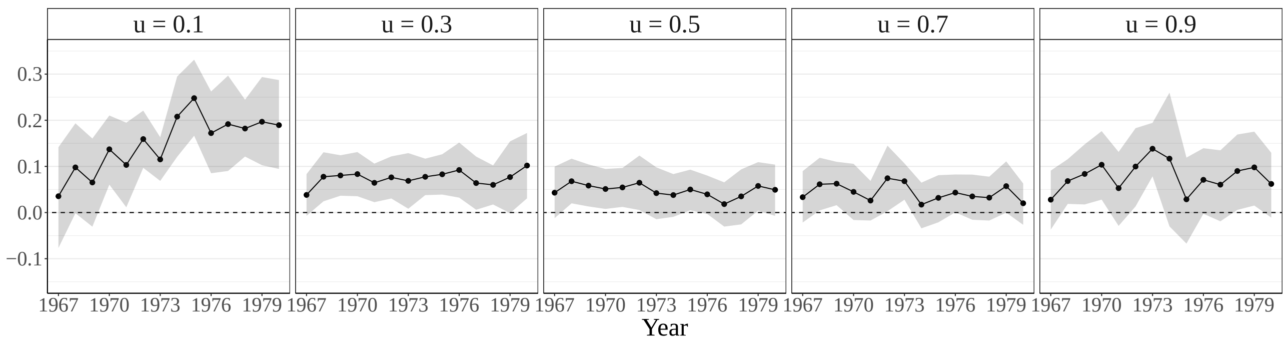
[7]The Equal Pay Act of 1963 is a federal law that amends the Fair Labor Standards Act and prohibits wage disparity based on gender. Title VII of The Civil Rights Act of 1964 more broadly prohibits discrimination in employment on the basis of race, color, religion, national origin, and gender.

Figure 4 Time-Varying Policy Effect Estimates (δ_{jt})

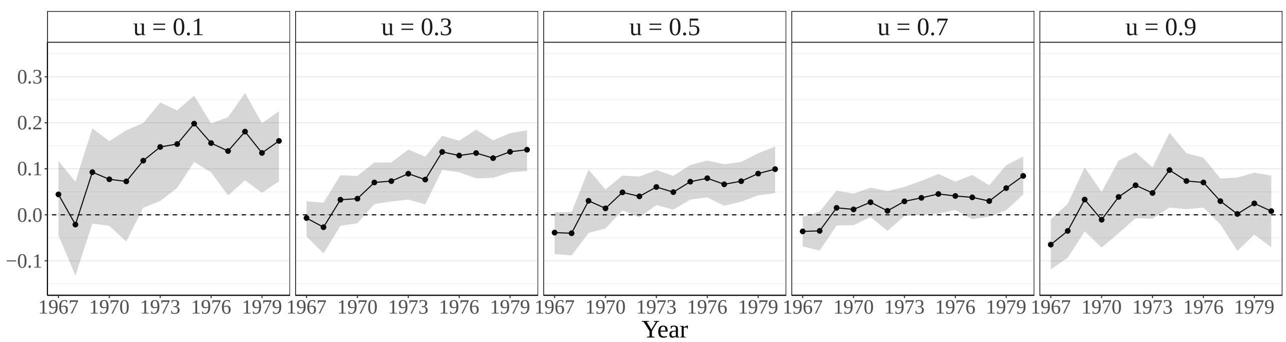
(a) Intercept (Base: White, Male, Full-Time Workers)



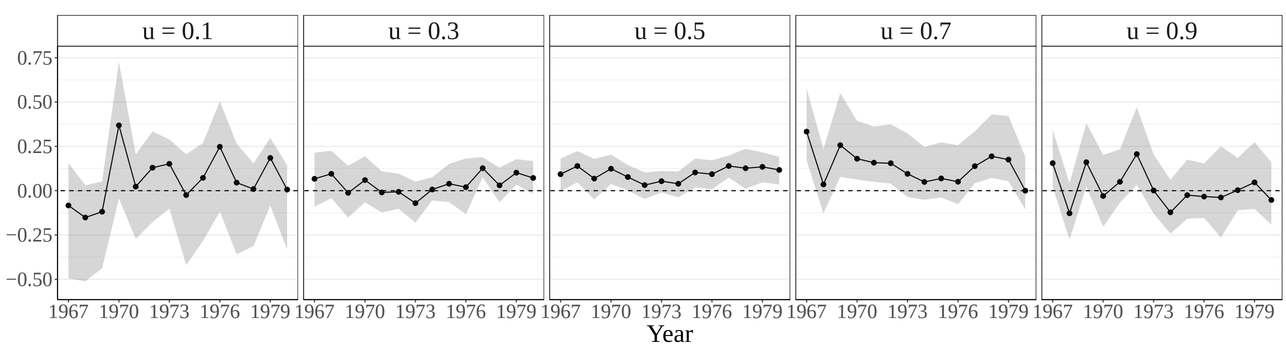
(b) Black



(c) Female



(d) Part Time

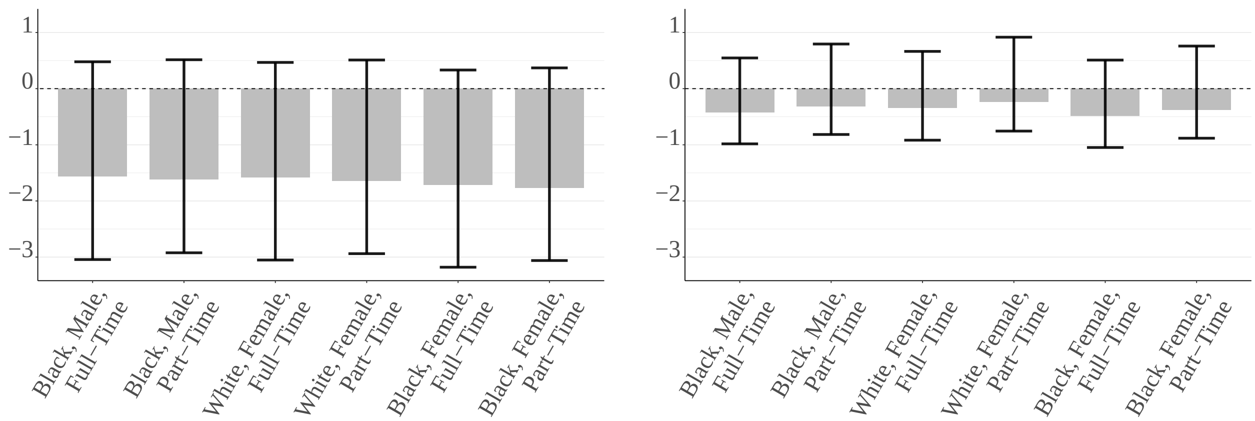


Notes: Panels (a)-(d) present the estimated time-varying marginal policy effect $\delta_{jt}(u)$ for $t = 1967, \ldots, 1980$. From left to right, figures correspond to the estimates at quantiles $u = 0.1, 0.3, 0.5, 0.7, 0.9$. Point estimates are plotted with solid black lines, and the pointwise 95%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

In Figure 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. For conditional variables z, we fix 12 years of education and 10 years of experience, while reporting six pairs of categorical individual attributes, as shown in the horizontal axis. All estimates suggests that the introduction of minimum wage reduces the within-inequality, while all estimates are statistically indistinguishable from 0.

Figure 5 Changes in Within-Inequality in 1980

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



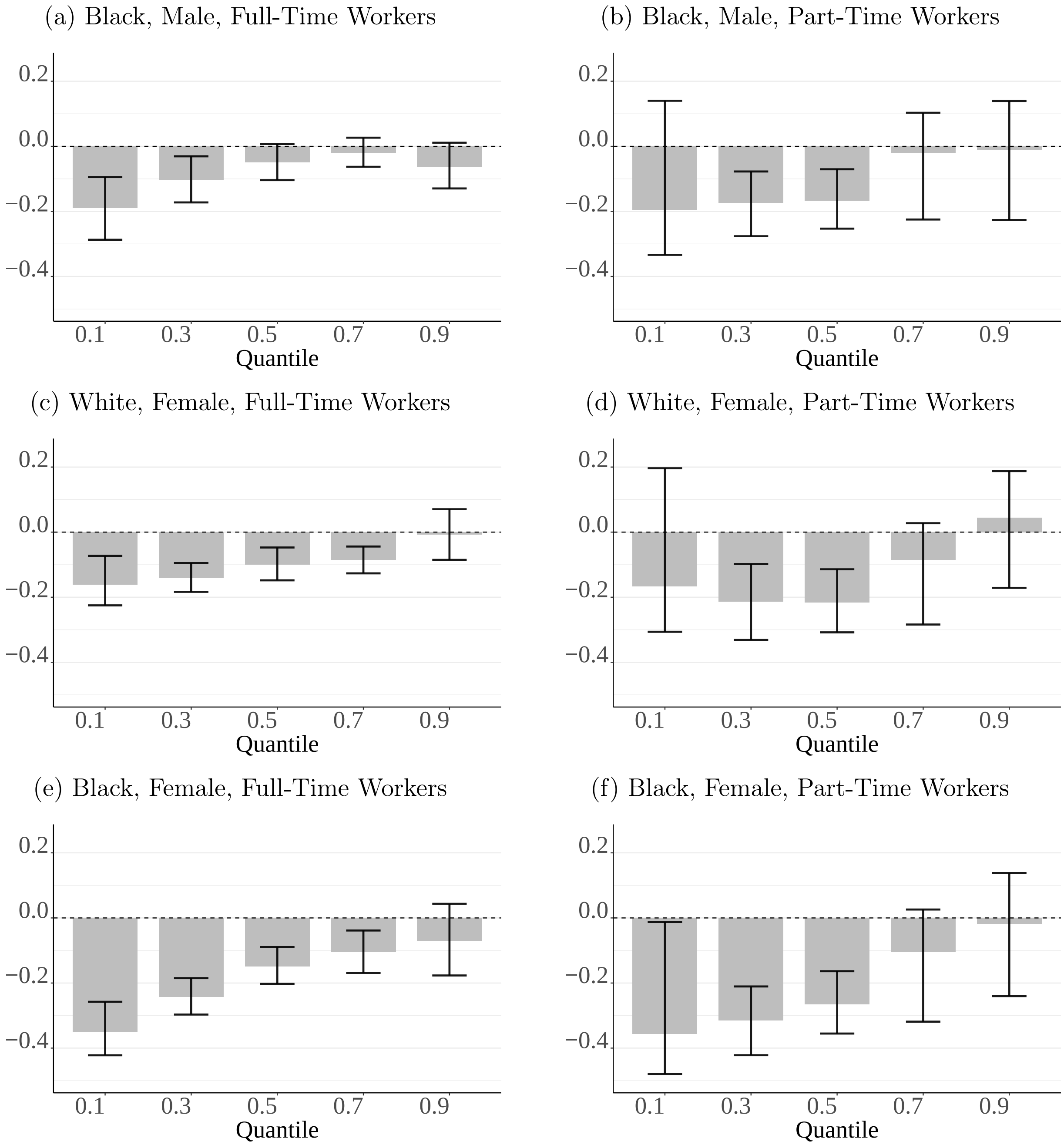
Notes: Panels (a)-(b) report estimated changes in the within-inequality $\dot{\Delta}_t^W(u_1, u_2|z)$ among individuals with attributes z in year $t = 1980$, with $(u_1, u_2) = (0.1, 0.9)$ in Panel (a) and $(u_1, u_2) = (0.1, 0.5)$ in Panel (b). The estimates are presented by grey bars with the 95% confidence intervals in black. The horizontal axis shows six pairs of black/white, female/male, and full/part-time, while we fix 12 years of education and 10 years of experience.

Figure 6 reports 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 and full-time and z_1 changes over the six pairs of categorical attributes as in Figure 5, while all ordered variables are the same in z_1 and z_2.[8] Panels (a)-(b) suggests negative impacts on the between-inequality for black, male workers with magnitudes 10-20% below the conditional median. Panels (c)-(d) also show reduction (0.05-0.20) in the between-inequality for white, female workers at the 0.7th conditional quantile and below. Panels (e)-(f) plot results for female, black

[8]From the identification result in Theorem 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 out each other.

workers, and the estimated changes in the between-inequality range from -0.05 to 0.35 at the 0.7th conditional quantile and below.

Figure 6 Changes in Between-Inequality in 1980



Notes: Panels (a)-(f) plot the changes in between-inequality ($\dot{\Delta}_t^B$) between multiple groups and the base-level group: full-time white male workers while holding other continuous 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.

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.

6 Simulation Study

To demonstrate the usefulness of our proposed method, we investigate the accuracy of both the point and interval estimators of $\beta_j(u)$ through Monte Carlo simulations. We first compare the bias and standard deviation of $\widehat{\beta}_j^{(m)}(u)$ at a number of quantiles and different iterations with an initial estimator being either consistent or not. The comparison aims to illustrate that the algorithm can converge at a relatively quick speed regardless of the choice of the initial estimator. In addition, we calculate the coverage rate of confidence interval estimated according to Corollary A.1.

6.1 Data Generating Process

We consider the following data generating process, where the two-level structure are similar to Chetverikov et al. (2016). We generate data according to the following model, for $i = 1, ..., N,\ s = 1, ..., S$, and $t = 1, ..., T$,

$$y_{ist} = \alpha_{st0}(u_{ist}) + z_{ist}\delta(u_{ist}), \quad \alpha_{st0}(u_{ist}) = x_{st}\beta(u_{ist}) + f_t(u_{ist})'\lambda_s(u_{ist}) + \eta_{st}(u_{ist}),$$

$$\delta(u) = 2 + 0.1u, \quad \beta(u) = 3 + 0.25u^2, \quad \eta_{st}(u) = \xi_{st}u - 0.5u,$$

$$f_t(u_{ist}) = (f_{t1}\sqrt{u_{ist}}, f_{t2}\sqrt{u_{ist}})', \quad \lambda_s(u_{ist}) = (\lambda_{s1} + 0.5\sqrt{u_{ist}}, \lambda_{s2} + 0.5\sqrt{u_{ist}})',$$

where $\{u_{ist}\}$ and $\{\xi_{st}\}$ are i.i.d $U(0,1)$, $\{z_{ist}\}$ are i.i.d $U(0,1)$, (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 $\mathrm{U}(0,2)$.

We consider the following two scenarios in terms of whether exists endogenity in the group-level:

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

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

6.2 Performance of the Point Estimator

We investigate the accuracy of the point estimator $\widehat{\beta}^{(m)}(u)$ under both consistent and inconsistent initials within finite iterations. It is easy to check that, according to Proposition 4.1.(i), the initial estimator $\widehat{\beta}^{(0)}_{OLS}(u) = (\sum_{s=1}^{S} X_s' X_s)^{-1} \sum_{s=1}^{S} X_s' \widehat{A}_s(u)$ is consistent. Thus, in the simulation study of each scenario, we compare the consistent initial estimator $\widehat{\beta}^{(0)}(u) = \widehat{\beta}^{(0)}_{OLS}(u)$ with an inconsistent estimator $\widehat{\beta}^{(0)}(u) = 0$.

We report the average bias and standard deviation of the estimated $\widehat{\beta}^{(m)}(u)$ for quantiles $u = 0.1, 0.5, 0.9$ and iteration steps $m = 0, 1, 2$ for both scenarios in Table 1 and 2, respectively. The simulation shows a few nice properties, which can be summarized as follows. First of all, the recursive estimator converge quite fast regardless of the choice of the initial estimator, as in all cases, the mean bias and standard deviation of the coefficient estimation at the second iteration are reasonably small. In addition, the consistent initial estimator leads to a quicker numerical convergence of the algorithm in comparison to the inconsistent one. The inconsistency initial $\widehat{\beta}^{(0)} = 0$ leads to inconsistent estimator in the first iteration, however, it becomes consistency from the second iteration. On the other hand, the consistent initial leads to consistent recursive estimator from the first round. Last but not least, the estimators remain valid even if the number of observations per group (N) is relatively small comparing to the number of groups ($S \times T$), which is a particular attractive property in practice.

Table 1 Bias and Standard Deviation of $\widehat{\beta}^{(m)}(u)$ for Scenario 1

		$\widehat{\beta}^{(0)}(u)=0$			$\widehat{\beta}^{(0)}(u)=\widehat{\beta}^{(0)}_{OLS}(u)$		
$(N,\ S,\ T)$	u	$m=0$	$m=1$	$m=2$	$m=0$	$m=1$	$m=2$
(200, 15, 20)	0.1	−3.002 (0)	−0.168 (0.087)	−0.011 (0.018)	0.003 (0.074)	0.001 (0.017)	0.001 (0.016)
	0.5	−3.062 (0)	−0.086 (0.072)	−0.006 (0.015)	0.001 (0.031)	0.001 (0.015)	0.001 (0.014)
	0.9	−3.203 (0)	−0.191 (0.095)	−0.017 (0.029)	−0.003 (0.085)	−0.004 (0.028)	−0.004 (0.028)
(200, 30, 40)	0.1	−3.002 (0)	−0.082 (0.020)	−0.003 (0.008)	0.003 (0.038)	0.001 (0.008)	0.001 (0.008)
	0.5	−3.062 (0)	−0.050 (0.043)	−0.001 (0.007)	0.001 (0.012)	0.001 (0.007)	0.001 (0.007)
	0.9	−3.203(0)	−0.091 (0.025)	−0.004 (0.012)	0.001 (0.041)	−0.001 (0.012)	−0.001 (0.011)
(500, 15, 20)	0.1	−3.002 (0)	−0.172 (0.096)	−0.012 (0.015)	−0.005 (0.076)	0.001 (0.013)	0.001 (0.012)
	0.5	−3.062 (0)	−0.080 (0.070)	−0.005 (0.013)	−0.001 (0.029)	0.001 (0.012)	0.001 (0.012)
	0.9	−3.203 (0)	−0.189 (0.092)	−0.015 (0.027)	−0.004 (0.087)	−0.001 (0.026)	−0.001 (0.025)
(500, 30, 40)	0.1	−3.002 (0)	−0.084 (0.020)	−0.003 (0.005)	0.004 (0.037)	0.001 (0.005)	0.001 (0.005)
	0.5	−3.062 (0)	−0.041 (0.037)	−0.002 (0.005)	0.001 (0.010)	0.001 (0.005)	0.001 (0.005)
	0.9	−3.203 (0)	−0.090 (0.024)	−0.004 (0.011)	−0.001 (0.041)	0.001 (0.010)	0.001 (0.010)
(1000, 15, 20)	0.1	−3.002 (0)	−0.166 (0.076)	−0.012 (0.012)	0.005 (0.070)	0.001 (0.011)	0.001 (0.009)
	0.5	−3.062 (0)	−0.076 (0.063)	−0.005 (0.011)	−0.003 (0.029)	0.001 (0.010)	0.001 (0.010)
	0.9	−3.203 (0)	−0.182 (0.093)	−0.012 (0.026)	0.004 (0.084)	0.001 (0.026)	0.001 (0.025)
(1000, 30, 40)	0.1	−3.002 (0)	−0.084 (0.019)	−0.003 (0.004)	0.001 (0.037)	0.001 (0.004)	0.001 (0.004)
	0.5	−3.062 (0)	−0.039 (0.036)	−0.001 (0.005)	0.001 (0.011)	0.001 (0.005)	0.001 (0.005)
	0.9	−3.203 (0)	−0.091 (0.023)	−0.004 (0.010)	0.002 (0.039)	−0.001 (0.010)	−0.001 (0.010)

Notes: The number of Monte-Carlo repetitions is 500. We report bias averaged over 500 repetitions with the standard deviation in the parenthesis.

Table 2 Bias and Standard Deviation of $\widehat{\beta}^{(m)}(u)$ for Scenario 2

		$\widehat{\beta}^{(0)}(u)=0$			$\widehat{\beta}^{(0)}(u)=\widehat{\beta}^{(0)}_{OLS}(u)$		
$(N,\ S,\ T)$	u	$m=0$	$m=1$	$m=2$	$m=0$	$m=1$	$m=2$
(200, 15, 20)	0.1	−3.002 (0)	−0.345 (0.177)	−0.024 (0.024)	−0.173 (0.081)	−0.010 (0.018)	0.001 (0.016)
	0.5	−3.062 (0)	−0.099 (0.077)	−0.010 (0.017)	0.001 (0.038)	0.001 (0.015)	0.001 (0.015)
	0.9	−3.203 (0)	−0.120 (0.047)	−0.024 (0.028)	0.175 (0.094)	−0.002 (0.029)	−0.015 (0.028)
(200, 30, 40)	0.1	−3.002 (0)	−0.147 (0.048)	−0.004 (0.008)	−0.164 (0.047)	−0.005 (0.008)	0.001 (0.008)
	0.5	−3.062 (0)	−0.048 (0.037)	−0.002 (0.008)	0.001 (0.016)	0.001 (0.007)	0.001 (0.007)
	0.9	−3.203 (0)	−0.063 (0.018)	−0.010 (0.012)	0.169 (0.051)	−0.001 (0.012)	−0.008 (0.011)
(500, 15, 20)	0.1	−3.002 (0)	−0.339 (0.176)	−0.024 (0.024)	−0.180 (0.079)	−0.010 (0.014)	0.002 (0.012)
	0.5	−3.062 (0)	−0.089 (0.070)	−0.009 (0.014)	0.001 (0.036)	0.001 (0.012)	0.001 (0.012)
	0.9	−3.203 (0)	−0.123 (0.050)	−0.021 (0.026)	0.173 (0.097)	0.001 (0.027)	−0.012 (0.025)
(500, 30, 40)	0.1	−3.002 (0)	−0.147 (0.042)	−0.004 (0.006)	−0.167 (0.045)	−0.005 (0.006)	0.001 (0.005)
	0.5	−3.062 (0)	−0.042 (0.033)	−0.003 (0.005)	0.001 (0.014)	0.001 (0.005)	0.001 (0.005)
	0.9	−3.203 (0)	−0.062 (0.018)	−0.009 (0.011)	0.167 (0.048)	−0.001 (0.011)	−0.007 (0.011)
(1000, 15, 20)	0.1	−3.002 (0)	−0.337 (0.168)	−0.025 (0.021)	−0.169 (0.077)	−0.011 (0.012)	0.001 (0.009)
	0.5	−3.062 (0)	−0.082 (0.059)	−0.008 (0.012)	0.004 (0.033)	0.001 (0.010)	0.001 (0.010)
	0.9	−3.203 (0)	−0.116 (0.046)	−0.019 (0.026)	0.177 (0.090)	0.002 (0.027)	−0.001 (0.025)
(1000, 30, 40)	0.1	−3.002 (0)	−0.145 (0.042)	−0.004 (0.004)	−0.168 (0.043)	−0.005 (0.004)	0.001 (0.004)
	0.5	−3.062 (0)	−0.042 (0.032)	−0.003 (0.005)	0.001 (0.014)	0.001 (0.005)	0.001 (0.005)
	0.9	−3.203 (0)	−0.063 (0.016)	−0.010 (0.010)	0.171 (0.046)	−0.001 (0.011)	−0.007 (0.010)

Notes: See Table 1.

6.3 Performance of the Interval Estimator

We investigate the coverage rate of the confidence interval when m is sufficiently large. We reduce our attention to the estimation with consistent initiation $\widehat{\beta}^{(0)}(u) = \widehat{\beta}^{(0)}_{OLS}(u)$ and set the number of iteration $m = 50$. The coverage rate is calculated by the rate that $\beta(u)$ lies in the 95% confidence interval within 500 simulation repetitions. The estimated asymptotic bias and covariance are computed according to Corollary A.1. Table 3 shows when there is no endogeneity (i.e. Scenario 1), the coverage rate is around 0.95 when the sample size is sufficiently large. While the coverage rate is slightly underestimated when endogenity exists. In addition, the accuracy increases as the sample sizes N, S and T increase, and the group-level sample size (S and T) is more pronounced.

Table 3 Coverage Rate of the 95% Confidence Interval for $\beta(u)$ in Scenario 1–2

	N	$S = 15,\ T = 20$			$S = 30,\ T = 40$		
		$u = 0.1$	$u = 0.5$	$u = 0.9$	$u = 0.1$	$u = 0.5$	$u = 0.9$
Scenario 1	200	0.936	0.928	0.934	0.938	0.936	0.950
	500	0.936	0.916	0.922	0.940	0.956	0.952
	1000	0.942	0.930	0.910	0.942	0.956	0.948
Scenario 2	200	0.940	0.924	0.890	0.928	0.936	0.906
	500	0.926	0.912	0.896	0.944	0.952	0.888
	1000	0.938	0.928	0.890	0.938	0.958	0.872

Notes: Coverage rate based on 500 simulation repetitions.

7 Conclusion

In this paper, 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.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

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.

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica*, 74(2):539–563.

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.

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.

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. (2014). Theory and methods of panel data models with interactive effects. *The Annals of Statistics*, 42(1):142–170.

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

Bailey, M. J., Helgerman, T., and Stuart, B. A. (2021). Changes in the U.S. gender gap in wages in the 1960s. In *AEA Papers and Proceedings*, volume 111, pages 143–48.

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.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963–1987: application of quantile regression. *Econometrica*, pages 405–458.

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.

Chen, J., Gao, J., and Li, D. (2012). Semiparametric trending panel data models with cross–sectional dependence. *Journal of Econometrics*, 171(1):71–85.

Chen, L., Dolado, J. J., and Gonzalo, J. (2021). Quantile factor models. *Econometrica*, 89(2):875–910.

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.

Cox, D. R. (1984). Interaction. *International Statistical Review/Revue Internationale de Statistique*, pages 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.

Finkelstein, A. and McKnight, R. (2008). What did medicare do? The initial impact of medicare on mortality and out of pocket medical spending. *Journal of Public Economics*, 92(7):1644–1668.

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.

Galvao, A. F. and Kato, K. (2016). Smoothed quantile regression for panel data. *Journal of Econometrics*, 193(1):92–112.

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.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.

Harding, M. and Lamarche, C. (2014). Estimating and testing a quantile regression model with interactive effects. *Journal of Econometrics*, 178:101–113.

Hausman, J. A. and Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, pages 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*, 43(2):305–325.

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. *Econometric Society Monographs*, 43:17.

Jiang, B., Yang, Y., Gao, J., and Hsiao, C. (2021). Recursive estimation in large panel data models: theory and practice. *Journal of Econometrics*, 224(3):439–465.

Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3):410–442.

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.

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, J. G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

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.

Ma, S., Linton, O., and Gao, J. (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics*, 222(1):295–323.

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

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

Oka, T. and Yamada, K. (2021). Heterogeneous impact of the minimum wage: implications for changes in between-and within-group inequality. *Journal of Human Resources*, pages 0719–10339R1.

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

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

Sasaki, Y. (2015). What do quantile regressions identify for general structural functions? *Econometric Theory*, 31(5):1102–1116.

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.

Smith, J. P. and Welch, F. R. (1989). Black economic progress after Myrdal. *Journal of Economic Literature*, 27(2):519–564.

Swamy, P. A. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38(2):311–323.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random–coefficient and treatment–effect panel data models. *Review of Economics and Statistics*, 87(2):385–390.

Appendix A

In this appendix, Section A.1 presents the relevant technical assumptions for deriving the asymptotic distribution, and then Section A.2 provides explicit expressions for asymptotic bias and covariance and their consistent estimators. Finally, Section A.3 provides the proofs of the theorems and propositions in the main text and Appendix A.2. All the preliminary lemmas and their proofs are in the online supplementary Appendix B.

In what follows, we use a few additional notation. For a square matrix A, let $\mathrm{tr}(A)$ 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 $\rho_{\max}(A) \leq \mathrm{tr}(A)$ for every symmetric positive semi-definite matrix. Thus, $\|X\|^2 \leq \mathrm{tr}(X'X)$ for matrix X, and we repeatedly use this inequality. 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 $\mathrm{diag}(\cdot)$ denotes the diagonal matrix, whose diagonal entries are given in the parenthesis. Let $\odot$ and $\otimes$ denote the Hadamard and Kronecker product, respectively. We use C as a generic constant that varies according to the context and is independent of $(i, s, t, N_{\min}, S, T)$.

A.1 Additional Technical Assumptions

Assumption A.1. *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_β 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) $\widehat{\beta}_j^{(m)}(u) \in \{\beta \in \mathbb{R}^K : F_j^{(m)}(u)'L(\beta)F_j^{(m)}(u) \xrightarrow{p} \Sigma_\beta\}$, *where*

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

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

We define $Z^{(m)}(u) := [Z_1^{(m)}(u)', \ldots, Z_J^{(m)}(u)']'$, where, for $j = 1, \ldots, J$,

$$
\begin{aligned}
Z_j^{(m)}(u) \quad := \quad & \big[I_K - E_j(u)\big]^{-1}\big[I_K - (E_j(u))^m\big] D_j^{-1}(u) \frac{1}{\sqrt{ST}} \sum_{s=1}^{S} R_{js}(u)' M_{F_j}(u) \eta_{js}(u) \\
& + \big(E_j(u)\big)^m \left(\frac{1}{ST} \sum_{g=1}^{S} X_g' X_g\right)^{-1} \frac{1}{\sqrt{ST}} \sum_{s=1}^{S} X_s' \eta_{js}(u),
\end{aligned}
$$

with $D_j(u) := (ST)^{-1} \sum_{s=1}^{S} X_s' M_{F_j}(u) X_s$, $\omega_{j,sg}(u) := \lambda_{jg}(u)' \big(S^{-1} \Lambda_j(u)' \Lambda_j(u)\big)^{-1} \lambda_{js}(u)$,

$$
R_{js}(u) := X_s - \frac{1}{S} \sum_{g=1}^{S} \omega_{j,sg}(u) X_g, \quad \text{and} \quad E_j(u) := D_j^{-1}(u) \frac{1}{S^2 T} \sum_{s,g=1}^{S} \omega_{j,sg}(u) X_s' M_{F_j}(u) X_g.
$$

Assumption A.2. *For each $u \in \mathcal{U}$, $\|E_j(u)\| < 1$ with probability one. For any $u_1, u_2 \in \mathcal{U}$ and for each step $m > 0$, we have*

$$
\begin{bmatrix} Z^{(m)}(u_1) \\ Z^{(m)}(u_2) \end{bmatrix} \xrightarrow{d} \mathcal{N}\big(\mathbf{0}, \Sigma^{(m)}(u_1, u_2)\big),
$$

where $\mathbf{0}$ is a $(2JK) \times 1$ vector and $\Sigma^{(m)}(u_1, u_2)$ is the limit of the covariance matrix of $[Z^{(m)}(u_1)', Z^{(m)}(u_2)']'$ when $S, T \to \infty$.

Assumption A.2 is used to derive the joint Central Limit Theorem (CLT) in Theorem 4.1 for the m^{th} recursive estimator given quantile levels $u_1, u_2 \in \mathcal{U}$. The assumption is essentially same as Assumption E of Bai (2009) and Assumption 4 of Jiang et al. (2021). The term $E_j(u)$ is the decaying rate of the estimation error in each iteration, which requires to be less then one to ensure the convergence of the asymptotic bias and covariance. A simple justification of the assumption is provided in Appendix B.1.

A.2 Asymptotic Bias and Covariance

The asymptotic bias $B^{(m)}(u)$ in Theorem 4.1 can be written as

$$
B^{(m)}(u) := \big[B_1^{(m)}(u)', ..., B_J^{(m)}(u)'\big]', \tag{A.1}
$$

where, for $j = 1, \dots, J$,

$$B_j^{(m)}(u) := \operatorname*{plim}_{S,T\to\infty} \Big\{ \big[I_K - E_j(u)\big]^{-1} \cdot \left[I_K - (E_j(u))^m\right] D_j^{-1}(u) \left(\kappa^{1/2}\psi_j(u) + \kappa^{-1/2}\zeta_j(u)\right) + \big(E_j(u)\big)^m \left(\frac{1}{ST}\sum_{g=1}^{S} X_g' X_g\right)^{-1} \left(\frac{1}{\sqrt{ST}}\sum_{s=1}^{S} X_s' F_j(u)\lambda_{js}(u)\right)\Big\},$$

with $D_j(u)$, $E_j(u)$, $R_{js}(u)$ and $\omega_{j,sg}(u)$ defined right above Assumption A.2, and

$$\psi_j(u) := -\frac{1}{ST}\sum_{s,g=1}^{S} \frac{R_{js}(u)' F_j(u)}{T} \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1} \lambda_{jg}(u)\mathbb{E}[\eta_{js}(u)'\eta_{jg}(u)],$$

$$\zeta_j(u) := -\frac{1}{ST}\sum_{s=1}^{S} X_s' M_{F_j}(u)\Omega_j(u)F_j(u) \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1} \lambda_{js}(u),$$

$$\Omega_{jk,s}(u_1,u_2) := \mathbb{E}[\eta_{js}(u_1)\eta_{ks}(u_2)'] \quad \Omega_j(u) := \frac{1}{S}\sum_{s=1}^{S}\Omega_{jj,s}(u,u).$$

Below we present the simplified asymptotic bias and covariance, when the number of iterations m goes to ∞ and there is no correlation in either time series or cross–sectional dimensions.

For $u_1, u_2 \in \mathcal{U}$, let $u \in \{u_1, u_2\}$. Define $B(u) := \lim_{m\to\infty} B^{(m)}(u) = [B_1(u)', ..., B_J(u)']'$, and $Z_j(u) := \lim_{m\to\infty} Z_j^{(m)}(u)$. When $m \to \infty$, the term $\big(E_j(u)\big)^m \xrightarrow{p} 0$ as $\|E_j(u)\| < 1$ in probability under Assumption A.2. Then, for $j = 1, ..., J$, the asymptotic bias

$$B_j(u) = \operatorname*{plim}_{S,T\to\infty} [I_K - E_j(u)]^{-1} D_j(u)^{-1}\big(\kappa^{1/2}\psi_j(u) + \kappa^{-1/2}\zeta_j(u)\big),$$

where the term $\psi_j(u)$ is reduced to

$$\psi_j(u) = -\frac{1}{ST}\sum_{s=1}^{S}\left(\frac{R_{js}(u)'F_j(u)}{T}\right)\left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1} \lambda_{js}(u)\mathbb{E}[\eta_{js}(u)'\eta_{js}(u)],$$

under the assumption that there is no correlation in both dimensions. While the expression for $\zeta_j(u)$ is unchanged, the component $\Omega_j(u)$ becomes a diagonal matrix as $\Omega_{jk,s}(u_1,u_2)$ is reduced to $\operatorname{diag}\big(\mathbb{E}[\eta_{js}(u_1) \odot \eta_{ks}(u_2)]\big)$ under the assumption of no correlation in time series and cross-sectional dimensions.

In addition, the asymptotic covariance matrix is reduced to

$$\Sigma(u_1,u_2) := \lim_{m\to\infty} \Sigma^{(m)}(u_1,u_2) = \begin{bmatrix} \sigma(u_1,u_1) & \sigma(u_1,u_2) \\ \sigma(u_2,u_1) & \sigma(u_2,u_2) \end{bmatrix},$$

where $\sigma(u_1,u_2)$ is a $J\times J$ block matrix, whose $(j,k)^{\text{th}}$ entry $\sigma_{jk}(u_1,u_2) := \lim_{S,T\to\infty}\mathbb{E}[Z_j(u_1)Z_k(u_2)']$ is a $K\times K$ matrix, given by

$$\sigma_{jk}(u_1,u_2) = \operatorname*{plim}_{S,T\to\infty} \big(I_K - E_j(u_1)\big)^{-1} D_j(u_1)^{-1}\Xi_{jk}(u_1,u_2)\big[\big(I_K - E_k(u_2)\big)^{-1}D_k(u_2)^{-1}\big]', \tag{A.2}$$

where

$$\Xi_{jk}(u_1,u_2) := \frac{1}{ST}\sum_{s=1}^{S} R_{js}(u_1)' M_{F_j}(u_1)\Omega_{jk,s}(u_1,u_2) M_{F_k}(u_2) R_{ks}(u_2).$$

We now provide a way to estimate the asymptotic bias and covariance given above, and subsequently, Corollary A.1 below establishes their consistency.

We define an estimator of the asymptotic bias $\widehat{B}^{(m)}(u) := \big[\widehat{B}_1^{(m)}(u)', ..., \widehat{B}_J^{(m)}(u)'\big]'$, where, for $j = 1,\ldots,J$,

$$\widehat{B}_j^{(m)}(u) := \big[I_K - \widehat{E}_j^{(m)}(u)\big]^{-1}\big(\widehat{D}_j^{(m)}(u)\big)^{-1}\left(\sqrt{\frac{T}{S}}\widehat{\psi}_j^{(m)}(u) + \sqrt{\frac{S}{T}}\widehat{\zeta}_j^{(m)}(u)\right).$$

Here, we estimate each component of the asymptotic bias by its sample analog as follows:

$$\widehat{D}_j^{(m)}(u) := \frac{1}{ST}\sum_{g=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u) X_g,$$

$$\widehat{E}_j^{(m)}(u) := \big(\widehat{D}_j^{(m)}(u)\big)^{-1}\frac{1}{S^2T}\sum_{s,g=1}^{S}\widehat{\omega}_{j,sg}^{(m)}(u) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g,$$

$$\widehat{\psi}_j^{(m)}(u) := -\frac{1}{ST}\sum_{s=1}^{S}\frac{\widehat{R}_{js}(u)'\widehat{F}_j^{(m)}(u)}{T}\left(\frac{\widehat{\Lambda}_j^{(m)}(u)'\widehat{\Lambda}_j^{(m)}(u)}{S}\right)^{-1}\widehat{\lambda}_{js}^{(m)}(u)\widehat{\eta}_{js}^{(m)}(u)'\widehat{\eta}_{js}^{(m)}(u),$$

$$\widehat{\zeta}_j^{(m)}(u) := -\frac{1}{ST}\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u)\widehat{\Omega}_j^{(m)}(u)\widehat{F}_j^{(m)}(u)\left(\frac{\widehat{\Lambda}_j^{(m)}(u)'\widehat{\Lambda}_j^{(m)}(u)}{S}\right)^{-1}\widehat{\lambda}_{js}^{(m)}(u),$$

with $\widehat{\eta}_{js}^{(m)}(u) := \widehat{A}_{js}(u) - X_s\widehat{\beta}_j^{(m)}(u) - \widehat{F}_j^{(m)}(u)'\widehat{\lambda}_{js}^{(m)}(u)$, $\widehat{R}_{js}(u) := X_s - S^{-1}\sum_{g=1}^{S}\widehat{\omega}_{j,sg}^{(m)}(u)X_g$,

$\widehat{\Omega}^{(m)}_{jk,s}(u_1,u_2) := \text{diag}(\widehat{\eta}^{(m)}_{js}(u_1) \odot \widehat{\eta}^{(m)}_{ks}(u_2))$, $\widehat{\Omega}^{(m)}_{j}(u) := S^{-1}\sum_{s=1}^{S}\widehat{\Omega}^{(m)}_{jj,s}(u,u)$, and

$$\widehat{\omega}^{(m)}_{j,sg}(u) := \widehat{\lambda}^{(m)}_{jg}(u)'\left(\frac{\widehat{\Lambda}^{(m)}_{j}(u)'\widehat{\Lambda}^{(m)}_{j}(u)}{S}\right)^{-1}\widehat{\lambda}^{(m)}_{js}(u).$$

We construct an estimator of the asymptotic covariance, using the 2×2 block matrix

$$\widehat{\Sigma}^{(m)}(u_1,u_2) := \begin{bmatrix} \widehat{\sigma}(u_1,u_1), \widehat{\sigma}(u_1,u_2) \\ \widehat{\sigma}(u_2,u_1), \widehat{\sigma}(u_2,u_2) \end{bmatrix},$$

where $\widehat{\sigma}(u_1,u_2)$ is a $J \times J$ block matrix, whose $(j,k)^{\text{th}}$ block is given by

$$\widehat{\sigma}_{jk}(u_1,u_2) := \left[I_K - \widehat{E}^{(m)}_{j}(u_1)\right]^{-1}\left(\widehat{D}^{(m)}_{j}(u_1)\right)^{-1}\widehat{\Xi}^{(m)}_{jk}(u_1,u_2)\left(\widehat{D}^{(m)}_{k}(u_2)'\right)^{-1}\left[I_K - \widehat{E}^{(m)}_{k}(u_2)'\right]^{-1},$$

where

$$\widehat{\Xi}^{(m)}_{jk}(u_1,u_2) = \frac{1}{ST}\sum_{s=1}^{S}\widehat{R}_{js}(u_1)'M^{(m)}_{\widehat{F}_j}(u_1)\widehat{\Omega}^{(m)}_{jk,s}(u_1,u_2)M^{(m)}_{\widehat{F}_k}(u_2)'\widehat{R}_{ks}(u_2).$$

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

Corollary A.1. *Suppose that the assumptions of Theorem 4.1 hold. In addition, we assume that, for any fixed $u_1,u_2 \in \mathcal{U}$ and $j,k = 1,...,J$, for all $s,g \in \{1,...,S\}$, the following conditions holds:*

1. $\mathbb{E}[\eta_{js}(u_1)\eta_{kg}(u_2)'|X_k, X_j, \Lambda_j(u_1), F_j(u_1), \Lambda_k(u_2), F_k(u_2)] = \mathbf{0}$ *if* $s \neq g$,
2. $\mathbb{E}[\eta_{js}(u_1)\eta_{ks}(u_2)'|X_k, X_j, \Lambda_j(u_1), F_j(u_1), \Lambda_k(u_2), F_k(u_2)]$ *is a* $T \times T$ *diagonal matrix.*

Then, $\widehat{B}^{(m)}(u) \xrightarrow{p} B(u)$ *for each* $u = u_1, u_2$ *and* $\widehat{\Sigma}^{(m)}(u_1,u_2) \xrightarrow{p} \Sigma(u_1,u_2)$ *as* $S,T \to \infty$ *simultaneously and* $m \to \infty$.

In practice, we are interested in the inference when the estimators are converged. It is shown above that the asymptotic bias is an decreasing function of m. Therefore, Corollary A.1 considers the estimators for asymptotic bias and covariance when the number of iterations $m \to \infty$. In addition, the corollary above 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.

A.3 Proofs of the Main Results

Proof of Theorem 2.1. Let $u \in \mathcal{U}$ and $z \in \mathcal{Z}$ be fixed. Under the potential framework, we can write $\Delta_t^{AQTT}(u|z) = z'\mathbb{E}[\alpha_{st}^1(u) - \alpha_{st}^0(u)|d_s = 1]$. Under Assumption 2.3, it follows from (3) that $\mathbb{E}[\alpha_{jst}^1(u) - \alpha_{jst}^0(u)|d_s = 1] = \mathbb{E}[\Delta_{jst}(u)|d_s = 1]$. Thus, to prove (4), it suffices to show that $\mathbb{E}[\Delta_{jst}(u)|d_s = 1] = \delta_{jt}(u)$ for each $j = 1, \ldots, J$ and for $t \geq T_0$.

Under Assumption 2.1, we can identify the quantile regression coefficients $\alpha_{jst}(u)$. Also, we can identify the regression coefficients, factors and factor loadings, using the argument of Bai (2009) under Assumption 2.2. Thus, we treat the quantile regression coefficients and factors as known objects in the rest of the proof. We consider the matrix form of equation (2) and apply the projection matrix $M_{F_j}(u)$ onto the orthogonal complement of the column space of $F_j(u)$. Then, we have

$$M_{F_j}(u)A_{js}(u) = d_s M_{F_j}(u)\delta_{j\cdot}(u) + M_{F_j}(u)W_s\gamma_j(u) + M_{F_j}(u)\eta_{js}(u),$$

where $\delta_{j\cdot}(u) := [0, \ldots, 0, \delta_{jT_0}(u), ..., \delta_{jT}(u)]'$ is a $T \times 1$ vector. This leads to the normal equations, given by

$$\mathbb{E}[d_s M_{F_j}(u)A_{js}(u)] = \mathbb{E}[d_s M_{F_j}(u)]\delta_{j\cdot}(u) + \mathbb{E}[d_s M_{F_j}(u)W_s]\gamma_j(u), \tag{A.3}$$

$$\mathbb{E}[W_s' M_{F_j}(u)A_{js}(u)] = \mathbb{E}[d_s W_s' M_{F_j}(u)]\delta_{j\cdot}(u) + \mathbb{E}[W_s' M_{F_j}(u)W_s]\gamma_j(u). \tag{A.4}$$

Solving (A.4) with respect to $\gamma_j(u)$, we obtain $\gamma_j(u) = \left(\mathbb{E}[W_s' M_{F_j}(u)W_s]\right)^{-1}\left\{\mathbb{E}[W_s' M_{F_j}(u)A_{js}(u)] - \mathbb{E}[d_s W_s' M_{F_j}(u)]\delta_{j\cdot}(u)\right\}$. Substituting the solution into (A.3), we obtain

$$\mathbb{E}\left[\Pi_s M_{F_j}(u)A_{js}(u)\right] = \mathbb{E}[d_s \Pi_s M_{F_j}(u)]\delta_{j\cdot}(u), \tag{A.5}$$

where $\Pi_s := d_s I_T - \mathbb{E}[d_s M_{F_j}(u)W_s]\left(\mathbb{E}[W_s' M_{F_j}(u)W_s]\right)^{-1}W_s'$.

Under the potential outcome framework, we have $\alpha_{jst}(u) = \alpha^0_{jst}(u)$ for $t < T_0$ and $\alpha_{jst}(u) = (1-d_s)\alpha^0_{jst}(u) + d_s a^1_{jst}(u)$ for $t \geq T_0$. Then, we can rewrite (3) in the following matrix form:

$$A_{js}(u) = d_s \Delta_{js\cdot}(u) + W_s' \gamma_j(u) + F_j(u)' \lambda_{js}(u) + (1-d_s)\eta^0_{js}(u) + d_s \eta^1_{js}(u),$$

where $\Delta_{js\cdot}(u) := [0, \ldots, 0, \Delta_{jsT_0}, \ldots, \Delta_{jsT}]'$. Under Assumption 2.3.(i), $\mathbb{E}[\Pi_s M_{F_j}(u)\eta^d_{js}(u)] = 0$ for $d = 0, 1$ and also simple algebra shows that $\mathbb{E}[\Pi_s M_{F_j}(u) W_s] = 0$. It follows that $\mathbb{E}\big[\Pi_s M_{F_j}(u) A_{js}(u)\big] = \mathbb{E}\big[\Pi_s M_{F_j}(u) d_s \Delta_{js\cdot}(u)\big]$, which, together with the law of iterated expectation, yields

$$\mathbb{E}\big[\Pi_s M_{F_j}(u) A_{js}(u)\big] = \mathbb{E}\big[d_s \Pi_s M_{F_j}(u) \mathbb{E}\big(\Delta_{js\cdot}(u) | W_s, F_j, d_s = 1\big)\big] = \mathbb{E}[d_s \Pi_s M_{F_j}(u)] \mathbb{E}[\Delta_{js\cdot}(u) | d_s = 1],$$

where the last equality is due to Assumption 2.3.(ii). It is easy to check that $\mathbb{E}[d_s \Pi_s M_{F_j}(u)]$ is invertible under Assumption 2.2.(ii), by writing V_s as a block matrix of d_s and W_s. Then, it follows from (A.5) that $\delta_{j\cdot}(u) = \mathbb{E}[\Delta_{js\cdot}(u) | d_s = 1]$ or $\delta_{jt}(u) = \mathbb{E}[\Delta_{jst}(u) | d_s = 1]$ for $t \geq T_0$. Therefore, we can show the desired result. Similar argument yields the identification result for changes in the within- and between-inequality measures. ■

In what follows, we use the following facts: for all $(s,t) \in \{1, ..., S\} \times \{1, ..., T\}$, $\|X_s\| = O_p(\sqrt{T})$, $\|F_j(u)\| = O_p(\sqrt{T})$, $\|\widehat{F}^{(m)}_j(u)\| = O_p(\sqrt{T})$, $\|\Lambda_j(u)\| = O_p(\sqrt{S})$, $\|\widehat{\Lambda}^{(m)}_j(u)\| = O_p(\sqrt{S})$ under Assumption 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, $\|M^{(m)}_{\widehat{F}_j}(u)\| = O_p(1)$.

Proof of Proposition 4.1. **(i)** By simple algebra and the linear model for $A_{js}(u)$, we can write the initial estimator $\widehat{\beta}^{(0)}_j(u) = (\sum_{s=1}^S X_s' X_s)^{-1} \sum_{s=1}^S X_s' \widehat{A}_{js}(u)$ as

$$\widehat{\beta}^{(0)}_j(u) - \beta_j(u) = \left(\sum_{s=1}^S X_s' X_s\right)^{-1} \sum_{s=1}^S X_s' \big\{\big(\widehat{A}_{js}(u) - A_{js}(u)\big) + F_j(u)\lambda_{js}(u) + \eta_{js}(u)\big\}. \tag{A.6}$$

Under Assumption 4.5(i), $(ST)^{-1} \sum_{s=1}^S X_s' X_s \xrightarrow{p} \Sigma_x$, which is positive definite. It follows that

$$\big\|\widehat{\beta}^{(0)}_j(u) - \beta_j(u)\big\| \leq \big(\rho_{\min}(\Sigma_x)^{-1} + o_P(1)\big) \left\| \frac{1}{ST} \sum_{s=1}^S X_s' \big\{\big(\widehat{A}_{js}(u) - A_{js}(u)\big) + F_j(u)\lambda_{js}(u) + \eta_{js}(u)\big\} \right\|.$$

An application of the triangle inequality together with Lemma B.1 yields $\left\| \sum_{s=1}^{S} X_s'\big(\widehat{A}_{js}(u) - A_{js}(u)\big)\right\| = O_P(1)$.

Also, $(ST)^{-1}\sum_{s=1}^{S} X_s' F_j(u)\lambda_{js}(u) = O_P\big(\sqrt{\vartheta_j(u)}\big)$ and Lemma B.2(ii) shows that $(ST)^{-1/2}\sum_{s=1}^{S} X_s'\eta_{js}(u) = O_P(1)$. Collecting the results so far, we obtain that, under the assumption that $\vartheta_j(u) = o(1)$,

$$\big\|\widehat{\beta}_j^{(0)}(u) - \beta_j(u)\big\| = O_P\left(\frac{1}{ST}\right) + O_P\left(\sqrt{\vartheta_j(u)}\right) + O_P\left(\frac{1}{\sqrt{ST}}\right) = o_P(1).$$

(ii) Let $m \geq 1$ be fixed. Given the linear model for $A_{js}(u)$ in (5), we can write

$$\widehat{\beta}_j^{(m)}(u) - \beta_j(u) = \left(\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u) X_s\right)^{-1} \sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u)\big\{\big(\widehat{A}_{js}(u) - A_{js}(u)\big) + F_j(u)\lambda_{js}(u) + \eta_{js}(u)\big\}.$$

By Cauchy-Swartz inequality, we can show that

$$\left\|\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u)\big(\widehat{A}_{js}(u) - A_{js}(u)\big)\right\| \leq \sup_s \big\|\widehat{A}_{js}(u) - A_{js}(u)\big\| \cdot \textstyle\sum_{s=1}^{S} \|X_s\| \cdot \big\|M_{\widehat{F}_j}^{(m)}(u)\big\| = O_P\big((ST)^{1/4}\big),$$

where the last equality holds because Lemma B.1 implies that $\sup_s \|\widehat{A}_{js}(u) - A_{js}(u)\| = O_P(S^{-3/4}T^{-1/4})$. It follows that,

$$\begin{aligned} &\widehat{\beta}_j^{(m)}(u) - \beta_j(u) \\ &= \left(\frac{1}{ST}\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u) X_s\right)^{-1} \frac{1}{ST}\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u)\big(F_j(u)\lambda_{js}(u) + \eta_{js}(u)\big) + o_P\left(\frac{1}{\sqrt{ST}}\right). \end{aligned} \tag{A.7}$$

We can show that $\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u)\eta_{js}(u) = O_P(\sqrt{ST})$ by the upper bound of α-mixing sequence given in Lemma B.9. See also Lemma B.2. Then, together with Lemma B.4, which analyses the term $\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u) F_j(u)\lambda_{js}(u)$, we obtain the recursive expression of $\widehat{\beta}_j^{(m)}(u)$ that

$$\begin{aligned} &\widehat{\beta}_j^{(m)}(u) - \beta_j(u) \\ &= \left(\frac{1}{ST}\sum_{g=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u) X_g\right)^{-1} \Bigg\{\frac{1}{S^2T}\sum_{s,g=1}^{S} \omega_{j,sg}(u) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g\big(\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u)\big) \\ &\quad + \frac{1}{S^2T^2}\sum_{s,g=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u) F_j(u)\lambda_{jg}(u)\big(\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u)\big)' X_g' \widehat{F}_j^{(m)}(u) K_j^{(m)}(u)\lambda_{js}(u) \end{aligned} \tag{A.8}$$

$$+\frac{1}{S^2T^2}\sum_{s,g=1}^{S}X_s'M_{\widehat{F}_j}^{(m)}(u)X_g\big(\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\big)\big(\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\big)'X_g'\widehat{F}_j^{(m)}(u)K_j^{(m)}(u)\lambda_{js}(u)\Big\}$$
$$+O_P\Big(\max\Big\{\frac{1}{\sqrt{ST}},\frac{1}{S},\frac{1}{T}\Big\}\Big)+o_P\Big(\Big\|\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\Big\|\Big)+o_P\Big(\Big\|\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\Big\|^2\Big),$$

where $K_j^{(m)}(u):=\big(S^{-1}\Lambda_j(u)'\Lambda_j(u)\big)\big(T^{-1}F_j(u)'\widehat{F}_j^{(m)}(u)\big)$, and $\omega_{j,sg}(u)$ is defined in Assumption A.2. By Hölder's and triangle inequalities, the third term on the right-hand side of (A.8) satisfies

$$\begin{aligned}
&\Big\|\frac{1}{S^2T^2}\sum_{s,g=1}^{S}X_s'M_{\widehat{F}_j}^{(m)}(u)X_g\big(\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\big)\big(\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\big)'X_g'\widehat{F}_j^{(m)}(u)K_j^{(m)}(u)\lambda_{js}(u)\Big\|\\
&\quad\le\Big(\frac{1}{S}\sum_{s=1}^{S}\Big\|\frac{X_s}{\sqrt{T}}\Big\|^2\Big)^{1/2}\cdot\Big(\frac{1}{S}\sum_{s=1}^{S}\|\lambda_{js}(u)\|^2\Big)^{1/2}\cdot\|M_{\widehat{F}_j}^{(m)}(u)\|\cdot\Big\|\frac{\widehat{F}_j^{(m)}(u)K_j^{(m)}(u)}{\sqrt{T}}\Big\|\\
&\qquad\cdot\Big\|\frac{1}{ST}\sum_{g=1}^{S}X_g\big(\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\big)\big(\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\big)'X_g'\Big\|\\
&\quad=O_P\Big(\Big\|\frac{1}{ST}\sum_{s=1}^{S}X_sX_s'\Big\|\Big)\cdot\|\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\|^2=O_P\big(\sqrt{\xi_{ST}}\big)\cdot\|\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\|^2,
\end{aligned}$$

where the first three terms in the first inequality are $O_P(1)$ under Assumption 4.1, and the term $K_j^{(m)}(u)=O_P(\sqrt{T})$ under Assumptions 4.1.(ii) and A.1.(i). Similar argument yields that the first term of (A.8) is bounded by $O_P\big((\vartheta_j(u))^{1/2}\big)\|\widehat{\beta}^{(m-1)}(u)-\beta(u)\|$, and the second term is bounded by $O_P\big((S\wedge T)^{-1/2}\big)\|\widehat{\beta}^{(m-1)}(u)-\beta(u)\|+O_P(1)\|\widehat{\beta}^{(m-1)}-\beta\|^2$. Thus, we have

$$\begin{aligned}
\|\widehat{\beta}_j^{(m)}(u)-\beta_j(u)\|=&O_P\big(\sqrt{\xi_{ST}}\big)\cdot\|\widehat{\beta}_j^{(m-1)}(u)-\beta_j(u)\|^2\\
&+O_P\Big(\max\Big\{\sqrt{\vartheta_j(u)},\frac{1}{\sqrt{S}},\frac{1}{\sqrt{T}}\Big\}\Big)\cdot\|\widehat{\beta}^{(m-1)}-\beta\|+O_P\Big(\max\Big\{\frac{1}{\sqrt{ST}},\frac{1}{S},\frac{1}{T}\Big\}\Big)\\
=&O_P\Big(\prod_{i=0}^{m-1}\sqrt{\xi_{ST}}\Big)\cdot\|\widehat{\beta}_j^{(0)}(u)-\beta_j(u)\|^{2^m}+O_P\Big(\max\Big\{\frac{1}{\sqrt{ST}},\frac{1}{S},\frac{1}{T}\Big\}\Big)\\
=&O_P\Big(\max\Big\{\frac{1}{\sqrt{ST}},\frac{1}{S},\frac{1}{T},\big(\vartheta_j(u)\big)^{2^m}\Big\}\Big)=o_P(1),
\end{aligned}$$

where the second equality follows from the recursive procedure. ■

Proof of Theorem 4.1. The proof consists of two steps. First, we show that the estimator $\widehat{\beta}^{(m)}(u)$ has an asymptotic linear expansion around the true parameter. Next, we establish an asymptotic normality.

Let $u\in\mathcal{U}$, $j\in\{1,...,J\}$ and $m\ge1$ be fixed. Combining (A.7) with Lemma B.5, which analyzes the term $\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)} F_j(u)\lambda_{js}(u)$ given the consistency of the recursive estimator $\widehat{\beta}_j^{(m)}(u)$, we show that $\widehat{\beta}_j^{(m)}(u)$ has the recursive formula

$$\begin{aligned}
&\sqrt{ST}\big(\widehat{\beta}_j^{(m)}(u) - \beta_j(u)\big)\\
&=\left(\frac{1}{ST}\sum_{s=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u) X_s\right)^{-1}\Bigg\{\frac{1}{\sqrt{ST}}\sum_{s=1}^{S} R_{js}(u)' M_{\widehat{F}_j}^{(m)}(u)\eta_{js}(u)\\
&\quad - \frac{1}{(ST)^{3/2}}\sum_{s,g=1}^{S} X_s' M_{\widehat{F}_j}^{(m)}(u)\mathbb{E}[\eta_{jg}(u)\eta_{jg}'(u)]\widehat{F}_j^{(m)}(u)\big(K_j^{(m)}(u)\big)^{-1}\lambda_s\\
&\quad + \left[\frac{1}{S^2T}\sum_{s,g=1}^{S}\omega_{j,sg}(u) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g\right]\sqrt{ST}\big(\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u)\big)\Bigg\} + o_P(1).
\end{aligned}$$

which in essence is the same as Jiang et al. (2021) (equations (C.29)-(C.30) in the proof of Theorem 4). We expand the estimated quantities in the right-side of the above expression around the true quantities using Lemma B.6, and obtain the following asymptotic representation:

$$\begin{aligned}
&\sqrt{ST}\big(\widehat{\beta}_j^{(m)}(u) - \beta_j(u)\big) \qquad\qquad \text{(A.9)}\\
&=D_j^{-1}(u)\left[\frac{1}{\sqrt{ST}}\sum_{s=1}^{S} R_{js}(u)' M_{F_j}(u)\eta_{js}(u) + \sqrt{\frac{T}{S}}\psi_j(u) + \sqrt{\frac{S}{T}}\zeta_j(u)\right]\\
&\quad + E_j(u)\sqrt{ST}\big(\widehat{\beta}_j^{(m-1)}(u) - \beta_j(u)\big) + o_P(1)\\
&= [I_K - E_j(u)]^{-1}\,[I_K - (E_j(u))^m]\,D_j^{-1}(u)\left(\sqrt{\frac{T}{S}}\psi_j(u) + \sqrt{\frac{S}{T}}\zeta_j(u)\right)\\
&\quad + [I_K - E_j(u)]^{-1}\,[I_K - (E_j(u))^m]\,D_j^{-1}(u)\frac{1}{\sqrt{ST}}\sum_{s=1}^{S} R_{js}(u)' M_{F_j}(u)\eta_{js}(u)\\
&\quad + \big(E_j(u)\big)^m\sqrt{ST}\left(\widehat{\beta}_j^{(0)}(u) - \beta_j(u)\right) + o_P(1),
\end{aligned}$$

where the first equality follows from Lemma B.6, the second equality follows from iterating the recursive formula, $D_j(u)$, $E_j(u)$ and $R_{js}(u)$ are defined in Assumption A.2, and $\psi_j(u)$ and $\zeta_j(u)$ are defined in Section A.2.

Also, we have that $(ST)^{-1}\sum_{s=1}^{S} X_s' X_s = O_P(1)$ under Assumption 4.5.(i) and $\sum_{s=1}^{S} X_x\big(\widehat{A}_{js}(u) - A_{js}(u)\big) = O_P(1)$ by Lemma B.1. It follows from (A.6) that

$$\widehat{\beta}_j^{(0)}(u) - \beta_j(u) = \left(\sum_{s=1}^{S} X_s' X_s\right)^{-1}\sum_{s=1}^{S} X_s'\big(F_j(u)\lambda_{js}(u) + \eta_{js}(u)\big) + O_P\big((ST)^{-1}\big). \qquad \text{(A.10)}$$

Combining (A.9)-(A.10) with the definition of $\widehat{\beta}^{(m)}(u)$, we obtain

$$\sqrt{ST}\big(\widehat{\beta}^{(m)}(u) - \beta(u)\big) = \widetilde{B}^{(m)}(u) + Z^{(m)}(u) + o_P(1), \tag{A.11}$$

where $Z^{(m)}(u)$ is defined in Assumption A.2, and $\widetilde{B}^{(m)}(u) := [\widetilde{B}_1^{(m)}(u)', ..., \widetilde{B}_J^{(m)}(u)']'$ with

$$\begin{aligned}\widetilde{B}_j^{(m)}(u) :=& [I_K - E_j(u)]^{-1}\left[I_K - \big(E_j(u)\big)^m\right] D_j^{-1}(u)\left(\sqrt{\frac{T}{S}}\psi_j(u) + \sqrt{\frac{S}{T}}\zeta_j(u)\right) \\ &+ \big(E_j(u)\big)^m \left(\frac{1}{ST}\sum_{s=1}^{S} X_s' X_s\right)^{-1} \frac{1}{\sqrt{ST}}\sum_{s=1}^{S} X_s' F_j(u)\lambda_{js}(u).\end{aligned}$$

Under the assumption that $\vartheta_j(u) = O((ST)^{-1})$ and $T/S \to \kappa > 0$, Lemma B.6 implies that $\widetilde{B}^{(m)}(u)$ is bounded, and $\widetilde{B}^{(m)}(u) \xrightarrow{p} B^{(m)}(u)$. Combining with Assumption A.2, we obtain the desired result. ■

Proof of Corollary A.1. **(i).** To show $\widehat{B}^{(m)}(u) \xrightarrow{p} B(u)$, it is sufficient to show $\widehat{D}_j^{(m)}(u)$, $\widehat{E}_j^{(m)}(u)$, $\widehat{\psi}_j^{(m)}(u)$, and $\widehat{\zeta}_j^{(m)}(u)$ are all consistent as $S, T \to \infty$ and $m \to \infty$. The consistency of $\widehat{D}_j^{(m)}(u)$ is shown in Lemma B.6.(i). Then, the consistency of $\widehat{E}_j^{(m)}(u)$ can be derived by showing that $\widehat{D}_j^{(m)}(u)\widehat{E}_j^{(m)}(u) \xrightarrow{p} D_j(u)E_j(u)$. By adding and subtracting terms, we have

$$\begin{aligned}\widehat{D}_j^{(m)}(u)\widehat{E}_j^{(m)}(u) - D_j(u)E_j(u) =& \frac{1}{S^2T}\sum_{s,g=1}^{S} \{\widehat{\omega}_{j,sg}^{(m)}(u) - \omega_{j,sg}(u)\} X_s' M_{\widehat{F}_j}^{(m)}(u) X_g \\ &+ \frac{1}{S^2T}\sum_{s,g=1}^{S} \omega_{j,sg}(u) X_s' \{M_{\widehat{F}_j}^{(m)}(u) - M_{F_j}(u)\} X_g = o_P(1),\end{aligned} \tag{A.12}$$

where the second term is of order $o_P(1)$ according to Lemma B.6.(iv). Considering the first term above, there are three items being estimated in $\widehat{\omega}_{j,st}^{(m)}(u)$, namely $\widehat{\lambda}_{js}^{(m)}(u)$, $S^{-1}\widehat{\Lambda}_j^{(m)}(u)'\widehat{\Lambda}_j^{(m)}(u)$ and $\widehat{\lambda}_{jg}^{(m)}(u)$. Using the identity $\hat{a}\hat{b}\hat{c} - abc = (\hat{a} - a)\hat{b}\hat{c} + a(\hat{b} - b)\hat{c} + ab(\hat{c} - c)$, we write

$$\begin{aligned}&\frac{1}{S^2T}\sum_{s,g=1}^{S} \big(\widehat{\omega}_{j,sg}^{(m)}(u) - \omega_{j,sg}(u)\big) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g \\ =& \frac{1}{S^2T}\sum_{s,g=1}^{S} \big(\widehat{\lambda}_{jg}^{(m)}(u) - H_j^{(m)}(u)\lambda_{jg}(u)\big)' \left(\frac{\widehat{\Lambda}_j^{(m)}(u)'\widehat{\Lambda}_j^{(m)}(u)}{S}\right)^{-1} \widehat{\lambda}_{js}^{(m)}(u) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g \\ &+ \frac{1}{S^2T}\sum_{s,g=1}^{S} \Big(H_j^{(m)}(u)\lambda_{jg}(u)\Big)' \left[\left(\frac{\widehat{\Lambda}_j^{(m)}(u)'\widehat{\Lambda}_j^{(m)}(u)}{S}\right)^{-1}\right.\end{aligned} \tag{A.13}$$

$$
- \left(H_j^{(m)}(u)'\right)^{-1} \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1} \left(H_j^{(m)}(u)\right)^{-1} \Bigg] \widehat{\lambda}_{js}^{(m)}(u) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g
$$

$$
+ \frac{1}{S^2 T} \sum_{s,g=1}^{S} \lambda_{jg}(u)' \left(\frac{\Lambda_j(u)'\Lambda_j(u)}{S}\right)^{-1} \left(H_j^{(m)}(u)\right)^{-1} \left(\widehat{\lambda}_{js}^{(m)}(u) - H_j^{(m)} \lambda_{js}(u)\right) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g,
$$

where $H_j^{(m)}(u) := \widehat{\Upsilon}_j^{(m)}(u)\big(K_j^{(m)}(u)\big)^{-1}$, and $\Upsilon_j^{(m)}(u)$ is the diagnoal matrix consisting the r largest eigenvalues of $\widehat{L}_j^{(m)}(u)$ defined in (7), in descending order. Applying Hölder's inequality for the first term on the right-hand side above, we can show that

$$
\begin{aligned}
&\left\| \frac{1}{S^2 T} \sum_{s,g=1}^{S} \left(\widehat{\lambda}_{jg}^{(m)}(u) - H_j^{(m)}(u)\lambda_{jg}(u)\right)' \left(\frac{\widehat{\Lambda}_j^{(m)}(u)'\widehat{\Lambda}_j^{(m)}(u)}{S}\right)^{-1} \widehat{\lambda}_{js}^{(m)}(u) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g \right\| \\
&\quad \leq \left(\frac{1}{S}\sum_{g=1}^{S} \left\|\widehat{\lambda}_{jg}^{(m)}(u) - H_j^{(m)}(u)\lambda_{jg}(u)\right\|^2\right)^{1/2} \left(\frac{1}{S}\sum_{g=1}^{S}\left\|\frac{X_g}{\sqrt{T}}\right\|^2\right)^{1/2} \\
&\qquad \cdot \left(\frac{1}{S}\sum_{s=1}^{S}\left\|\frac{X_s}{\sqrt{T}}\right\|^2\right)^{1/2} \left(\frac{1}{S}\sum_{s=1}^{S}\left\|\widehat{\lambda}_{js}^{(m)}(u)\right\|^2\right)^{1/2} \left\|M_{\widehat{F}_j}^{(m)}(u)\right\| \left\|\left(\frac{\widehat{\Lambda}_j^{(m)}(u)'\widehat{\Lambda}_j^{(m)}(u)}{S}\right)^{-1}\right\|,
\end{aligned}
$$

Here, the first term on the right-side is $o_P(1)$ by Lemma B.14.(iii), and the rest terms are all $O_P(1)$ under Assumption 4.1, thereby yielding that the right-hand side is $o_P(1)$. Applying similar arguments to the remaining two terms of (A.13), we show that $(S^2T)^{-1}\sum_{s,g=1}^{S}\big(\widehat{\omega}_{j,sg}^{(m)}(u) - \omega_{j,sg}(u)\big) X_s' M_{\widehat{F}_j}^{(m)}(u) X_g = o_P(1)$, and thus, $\|\widehat{E}_j^{(m)}(u) - E_j(u)\| = o_P(1)$.

Similar arguments yield that $\|\widehat{\psi}_j^{(m)}(u) - \psi_j(u)\| = o_P(1)$ and $\|\widehat{\zeta}_j^{(m)}(u) - \zeta_j(u)\| = o_P(1)$, as shown in Lemma B.7.

(ii). To show $\widehat{\Sigma}^{(m)}(u_1,u_2) \xrightarrow{p} \Sigma(u_1,u_2)$, it remains to show the consistency of $\widehat{\Xi}_{jk}(u_1,u_2)$. Let $i,j \in \{1,...,J\}$ be given. We have

$$
\begin{aligned}
&\widehat{\Xi}_{jk}(u_1,u_2) - \Xi_{jk}(u_1,u_2) \qquad \text{(A.14)} \\
&\quad = \frac{1}{ST}\sum_{s=1}^{S}\left(\widehat{R}_{js}(u_1)' M_{\widehat{F}_j}^{(m)}(u_1) - R_{js}(u_1)' M_{F_j}(u_1)\right) \widehat{\Omega}_{ik,s}^{(m)}(u_1,u_2) M_{F_k}(u_2) R_{ks}(u_2) \\
&\qquad + \frac{1}{ST}\sum_{s=1}^{S} R_{js}(u_1)' M_{F_j}(u_1)\big(\widehat{\Omega}_{ik,s}^{(m)}(u_1,u_2) - \Omega_{ik.s}(u_1,u_2)\big) M_{F_k}(u_2) R_{ks}(u_2) \\
&\qquad + \frac{1}{ST}\sum_{s=1}^{S} \widehat{R}_{js}(u_1)' M_{\widehat{F}_j}^{(m)}(u_1) \widehat{\Omega}_{ik,s}^{(m)}(u_1,u_2) \left(M_{\widehat{F}_k}^{(m)}(u_2)\widehat{R}_{ks}(u_2) - M_{F_k}(u_2) R_{ks}(u_2)\right).
\end{aligned}
$$

The first and third terms on the right-hand side are $o_P(1)$, whose proofs are similar to Lemma B.7.(i). For the second term, by a further decomposition, we can see that

$$
\begin{aligned}
&\frac{1}{ST}\sum_{s=1}^{S} R_{js}(u_1)' M_{F_j}(u_1)\big(\widehat{\Omega}_{ik,s}^{(m)}(u_1,u_2) - \Omega_{ik.s}(u_1,u_2)\big) M_{F_k}(u_2) R_{ks}(u_2) \\
&=\frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T} R_{js}(u_1)' M_{F_j}(u_1) e_t e_t' M_{F_k}(u_2) R_{ks}(u_2) \{\widehat{\eta}_{jst}^{(m)}(u_1)\widehat{\eta}_{kst}^{(m)}(u_2) - \eta_{jst}(u_1)\eta_{kst}(u_2)\} \\
&\quad + \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T} R_{js}(u_1)' M_{F_j}(u_1) e_t e_t' M_{F_k}(u_2) R_{ks}(u_2) \{\eta_{jst}(u_1)\eta_{kst}(u_2) - \mathbb{E}[\eta_{jst}(u_1)\eta_{kst}(u_2)]\},
\end{aligned}
$$

where the first term is $o_P(1)$ due to the covergence of $\widehat{\eta}_{jst}^{(m)}(u_1)\widehat{\eta}_{kst}^{(m)}(u_2)$, and the second term is $o_P(1)$ due to the law of large numbers. Thus, $\widehat{\Xi}_{jk}(u_1,u_2) \xrightarrow{p} \Xi_{jk}(u_1,u_2)$. Moreover, since $\widehat{D}_j^{(m)}(u) \xrightarrow{p} D_j(u)$ and $\widehat{E}_j^{(m)}(u) \xrightarrow{p} E_j(u)$ are shown above. Together with (A.2), we conclude that $\widehat{\sigma}_{jk}(u_1,u_2) \xrightarrow{p} \sigma_{jk}(u_1,u_2)$, and thus $\widehat{\Sigma}(u_1,u_2) \xrightarrow{p} \Sigma(u_1,u_2)$. ■