Quantile on Quantiles*

Martina Pons†

Job Market Paper

October 23, 2024

Link to the newest version

Abstract

Distributional effects provide interesting insight into how a given treatment impacts inequality. This paper extends this notion in two ways. First, it recognizes that inequality spans multiple dimensions, for example, within and between groups, with treatments potentially influencing and creating trade-offs between both. Second, the paper addresses the nontrivial challenge of ranking heterogeneous groups, which heavily depends on the social welfare function of the policymaker. To this end, I introduce a model to simultaneously study distributional effects within and between groups while remaining agnostic about this social welfare function. The model consists of a quantile function with two indices, the first capturing heterogeneity within groups and the second addressing the between-group dimension. I propose a two-step quantile regression estimator involving within-group regressions in the first stage and between-group regressions in the second stage. I show that the estimator is consistent and asymptotically normal when the number of observations per group and the number of groups diverge to infinity. In an empirical application, I study the effect of training on the distribution of firms' performance within and between markets in Kenya. The results show large positive effects among the successful firms in the best-performing markets, suggesting potential complementarities between firms and market performance.

1 Introduction

Ever since Koenker and Bassett (1978), quantile regression has been widely used for policy evaluation. Quantile treatment effects are particularly valuable when the policymaker aims to prioritize specific segments of the population (e.g., low-income individuals) or address inequality rather than simply maximizing aggregate outcomes. With grouped data, modeling welfare as a function of the unconditional distribution of the outcome ignores the role of inequalities within and between these groups and the potential trade-offs between these two components of inequality. For instance, if geographical regions define the groups, it is possible to reduce within-region inequality by moving people across space into a more segregated spatial allocation while keeping the unconditional income distribution constant. At the same time, the applied literature

*I am grateful to Blaise Melly and Anna Mikusheva for their support. I also thank Alberto Abadie, Isaiah Andrews, Jean-Michel Benkert, Ivan Fernandez-Val, Bo Honoré, Frédéric Kluser, Jonas Meier, Maximilian von Ehrlich, Kaspar Wüthrich, and seminar participants for helpful comments and discussions. I gratefully acknowledge the support from a UniBE Doc.Mobility Grant to visit the Massachusetts Institute of Technology.

†Department of Economics, University of Bern, Schanzeneckstrasse 1, 3001 Bern, Switzerland. martina.pons@unibe.ch.

has predominantly focused on treatment effect heterogeneity along a single dimension, either within or between groups. For example, there is evidence that trade shocks have heterogeneous effects both within and between regions (see, for example, Chetverikov et al., 2016; Antràs et al., 2017; Galle et al., 2023). Similarly, place-based policies, while seeming to have been effective at promoting growth and jobs in targeted regions (Becker et al., 2010; Busso et al., 2013; Ehrlich and Seidel, 2018), have also increased within-region inequality (Lang et al., 2023; Albanese et al., 2023).[1] Taken together, these findings suggest that these two dimensions of inequality are interconnected and that there are substantial trade-offs that should be considered.

This paper suggests a novel model and estimator to study treatment effect heterogeneity and inequality both within and between groups simultaneously. Geographical regions, firms, or industries could define such groups. To this end, I introduce a two-dimensional quantile model where one dimension captures heterogeneity within groups, and the other addresses heterogeneity between groups. The conditional quantile function of each group models the within-group heterogeneity. Then, to aggregate the results across groups, I model the conditional quantile function of these group-level quantile functions. This yields a quantile function of group-level quantile functions, offering insights, for instance, into how the group-level conditional medians change across groups. This approach provides a framework that allows us to model within-group and between-group heterogeneity together, enhancing our understanding of the complex dynamics of inequality.

The estimation is performed in two stages. The first stage consists of group-by-group quantile regressions of the outcome on the variables that vary within groups. Similar first-stage regressions are used, for example, in Chetverikov et al. (2016) and Galvao and Wang (2015). In the second step, for each group and quantile, the first-stage fitted values are regressed on all variables using quantile regression. This estimator is flexible, allowing coefficients to vary without restriction along both dimensions and permitting the groups' ranks to evolve freely over the within distribution.

To establish the asymptotic results of the estimator, I have to deal with the non-smoothness of the objective function, a generated dependent variable in the second stage, and the different rates of convergence of the first-stage estimator. The first stage, which uses only observations for one group at a time, converges at a rate proportional to the square root of the number of observations per group n. In comparison, the second stage, which identifies the heterogeneity between groups, converges at a rate proportional to the square root of the number of groups m. Chen, Linton, and Van Keilegom (2003) study pointwise consistency and asymptotic normality of estimators with non-smooth and non-differentiable objective functions that depend on a non-parametric first step estimator, while Volgushev, Chao, and Cheng (2019) and Galvao, Gu, and Volgushev (2020) provide a thorough analysis of the remainder of the Bahadur representation for

[1] Place-based policies are policies implemented by government or supranational organizations with the aim to promote economic development within a given area (see Neumark and Simpson, 2015 for an overview on the topic)

quantile regression.[2] Building on these results, as well as on the process results in Angrist et al. (2006), I show asymptotic normality in a framework where the number of observations per group n and the number of groups m diverge to infinity, satisfying the weak condition $\frac{\sqrt{m}\log n}{n} \to 0$. Further, I suggest an inference procedure for uniform hypotheses and demonstrate its validity.

I show that the model is useful for policy evaluation when the policymaker wants to consider a trade-off between different dimensions of inequality. Specifically, the model provides a flexible tool for analyzing how policies impact the outcome distribution over multiple dimensions. As a byproduct, the method yields valuable insights for descriptive analyses of inequalities within- and between groups – a matter of considerable policy significance.[3] Compared to variance decompositions and comparison of median (or mean) outcomes over groups, the two-level quantile function provides a more comprehensive picture of the two-dimensional inequality. For instance, it reveals which parts of the within-group distribution drive inequality between groups. Further, this method addresses the challenges of ranking heterogeneous groups, given that when groups contain heterogeneous agents, there is no single definition of high- and low-performing groups. For instance, Chetty and Hendren (2018a,b) find that the same commuting zone can be associated with high mobility at the upper (lower) tail of the parents' distribution and with low mobility at a lower (upper) tail of the parents' distribution. Hence, this suggests that the definition of a good or bad neighborhood should depend on the within rank.

Furthermore, under the assumption of rank invariance over treatment states, the individual treatment effects are identified. This makes the framework suggested in this paper particularly useful for optimal treatment assignment when the policymaker maximizes a rank-dependent social welfare function, and no baseline outcomes are available (see, e.g., Manski, 2004; Kitagawa and Tetenov, 2018, 2021). Instead, the treatment assignment exploits treatment effect heterogeneity over the distribution of the outcome, both within and between groups. This model is relevant when the treatment is assigned at the group or individual levels. Group-level treatment assignments are common in economics; for example, place-based policies and infrastructure projects—-such as highways, railways, and sanitation systems—-affect all individuals in a locality, and educational policies are often implemented at the school level.

In an empirical application, I extend the findings of McKenzie and Puerto (2021) by assessing the impact of business training on firm performance in Kenya, considering distributional effects within and between markets. The goal of the experiment is to improve the outcomes of small businesses measured by indicators such as sales and profit. A quantile regression of sales on the treatment dummy identifies the treatment effect at different points of the sales distribution. Yet, it might be different to be a median business (in the unconditional distribution) operating

[2]Consistency and asymptotic normality of quantile regression with generated regressors and/or dependent variables have also been studied in Ma and Koenker (2006); Chen, Galvao, and Song (2021) and Bhattacharya (2020). However, compared to these papers, I consider a case where the dimension of the first stage increases with the number of groups, and the first-stage estimator converges at a different rate than the second-step estimator.

[3]For instance, one of the United Nations' sustainable development goals is to reduce inequalities within and between countries.

in a high-performing market or in a lower-performing one, and therefore, the treatment likely has a different effect on these hypothetical firms. For instance, a poor-performing market might suffer from a poor location or low consumer traffic. Another perspective is that there might be complementarities between individual abilities and market quality, which a traditional quantile regression model fails to capture. Hence, allowing for heterogeneity over multiple dimensions with a two-dimensional quantile model can provide interesting insights. The results indicate larger effects for firms that perform well within their successful markets. Specifically, the effects increase with both the firm's rank within its market and the market's overall rank, providing evidence of complementarities between individual and group ranks. This finding is useful for policymakers in deciding which firms and markets should be targeted, depending on their utility function. For instance, depending on whether the policy objective is to promote overall economic growth, reduce inequality, or any combination of the two, interventions can be directed toward different parts of the distribution.

Distributional effects and inequalities within groups are studied both in the applied and theoretical literature. For example, Galvao and Wang (2015), Chetverikov et al. (2016), and Melly and Pons (2024) suggest methods to model heterogeneity in treatment responses on the within-group distribution.[4] In the applied literature, Autor et al. (2021) and Friedrich (2022) investigate the impact of import competition and trade shocks on the wage distribution within local labor markets and within firms, respectively. Similarly, Helpman et al. (2017) document that most of the increase in wage inequality arises within sectors and occupations but is driven by differences between firms. Differently, Haltiwanger et al. (2024) find that the between-industry component accounts for most of the increase in earnings inequality in the last decades. Additionally, Autor et al. (2016) and Engbom and Moser (2022) explore the effect of minimum wages on within-state inequality in the US and Brazil, respectively. In contrast, papers studying the effectiveness of place-based policies in supporting laggard or underdeveloped regions provide examples focusing on disparities between groups (see, e.g., Busso et al., 2013; Ehrlich and Seidel, 2018; Ehrlich and Overman, 2020). Only a few papers in the economic literature focus on both within- and between-group inequality and mostly use a descriptive approach. For example, Bourguignon and Morrisson (2002) analyze the historical evolution of income inequality both within and between countries, and Akerman et al. (2013) study wage inequality between and within different groups, including firms, sectors, and occupations.[5] To this end, they decompose the variance into a within and a between component and, therefore, do not examine both dimensions simultaneously.

This paper also contributes to the theoretical literature focusing on multidimensional unobserved heterogeneity where the coefficients can vary along multiple dimensions. For example,

[4]Galvao and Wang (2015) focus on a traditional panel data setting, where the groups are their individuals and the individuals are their time periods.

[5]Inequality within and between countries has been intensively studied in the sociological literature (see, for example, Firebaugh, 2000; Goesling, 2001; Hung, 2021).

Fernández-Val, Gao, Liao, and Vella (2022) introduce a model that allows for heterogeneity within and between groups. The within-group heterogeneity is modeled by allowing the coefficient to vary over the outcome levels in a distribution regression framework, and group-specific coefficients capture the between-group heterogeneity. However, in this model, the identification of heterogeneities in both dimensions relies on within-group variation of the variable of interest.[6] Arellano and Bonhomme (2016) study a fixed effects model where the group effects are modeled as latent variables using a correlated random effects approach. The treatment effects can be heterogeneous through dependence on an individual rank variable and the latent group effects. Differently, the model in Frumento, Bottai, and Fernández-Val (2021) allows studying the effect of individual-level variables on the within distribution and the effect of group-level variables on the between distribution, and Liu (2021) considers a panel data model where the effect of the individual-level variables depends on a group-level rank variable, and the individual-level error enters additively. Hence, this last model identifies the effects of individual-level variables on the outcome distribution between groups. In contrast, the model in this paper allows the effect of both individual-level and group-level variables to vary along both dimensions.

The remainder of the paper is structured as follows. Section 2 introduces the model and Section 3 explains how this model can be used for policy evaluation and for optimal treatment assignment. Section 4 presents the estimator and Section 5 the asymptotic properties of the estimator. Section 6 analyzes the finite sample performance of the estimator in a Monte Carlo study. Section 7 presents the empirical application, and Section 8 concludes.

2 Model

Consider a dataset with two dimensions where $j = 1, \ldots, m$ indexes the groups and $i = 1, \ldots, n$ denote the individuals.[7] I start by considering a naive version of the model imposing strong assumptions on the evolution of the group ranks over the within distribution. Later, I relax these assumptions and present the more general model considered in this paper.

Simple model - A naive attempt to construct a model specifies the following structural function for the outcome variable y_{ij} given the individual-level variables x_{1ij}, and the group-level variables x_{2j}:

$$y_{ij} = q(x_{1ij}, x_{2j}, v_j, u_{ij}), \quad i = 1, \ldots, n, \quad j = 1, \ldots, m, \tag{1}$$

where $q(\cdot)$ is strictly increasing in the rank variables u_{ij} and v_j. The individual-level rank variable u_{ij} is responsible for differences in outcomes between individuals with the same observable characteristics, including group membership. Conversely, v_j is responsible for differences across

[6] Coefficients on variables that vary only between groups are identified using projections of the individual-level coefficient. Therefore, this model does not identify both dimensions of the heterogeneity for regressors that vary only between groups.

[7] For ease of notation, I assume a balanced dataset. However, the results generalize to heterogeneous group sizes.

groups. Further, for the sake of this illustration let

$$u_{ij}|x_{1ij}, x_{2j}, v_j \sim U(0,1),$$
$$v_j|x_{1ij}, x_{2j} \sim U(0,1).$$

Since v_j varies only between groups and u_{ij} is standard uniform distributed within each group, u_{ij} and v_j are independent conditional on the covariates:

$$u_{ij} \perp\!\!\!\perp v_j|x_{1ij}, x_{2j}.$$

Conditional on $x_{ij} = (x'_{1ij}, x'_{2j})'$ and v_j, $q(x_{1ij}, x_{2j}, v_j, u_{ij})$ is strictly monotonic with respect to u_{ij} so that

$$Q(\tau_1, y_{ij}|x_{1ij}, x_{2j}, v_j) = q(x_{1ij}, x_{2j}, v_j, \tau_1) \tag{2}$$

is the τ_1-conditional quantile function of the outcome y_{ij} conditional on x_{1ij}, x_{2j}, and v_j. If there are no x_{1ij} variables, the τ_1-conditional quantile function of y_{ij} reduces to the unconditional percentiles of the outcome in group j. Further, as $q(\cdot)$ is strictly monotonic with respect to v_j, we can construct the τ_2-conditional quantile function of $Q(\tau_1, y_{ij}|x_{1ij}, x_{2j}, v_j)$,

$$Q(\tau_2, Q(\tau_1, y_{ij}|x_{1ij}, x_{2j}, v_j)|x_{1ij}, x_{2j}) = q(x_{1ij}, x_{2j}, \tau_2, \tau_1). \tag{3}$$

The outer quantile function in equation (3) is the τ_2-conditional quantile function of the τ_1-conditional quantile function of the outcome within each group. Thus, τ_2 ranks the groups (conditional on the covariates) according to their conditional quantile functions. A caveat of this model is that it imposes strong restrictions on the evolution of the group ranks at different values of τ_1. More precisely, the ranks are assumed to be constant over τ_1. Suppose for a moment that there are no covariates and take groups $j = \{h, l\}$ with v_h and v_l such that $v_h > v_l$. Strict monotonicity of $q(v_j, \tau_1)$ with respect to v_j implies

$$q(v_h, \tau_1) > q(v_l, \tau_1)$$

for all $\tau_1 \in (0, 1)$. Hence, if a group has a higher first decile, it must also have a higher ninth decile. This would be satisfied, for example, if the outcome is generated by $y_{ij} = h(x_{1ij}, x_{2j}, u_{ij}) + f(x_{1ij}, x_{2j}, v_j)$, where $h(\cdot)$ and $f(\cdot)$ are strictly increasing in their third argument.[8] That is, if conditional on the covariates, all groups share the same outcome distribution up to a location parameter. Essentially, this requires that v_j enter as a pure location shifter conditional on (x_{1ij}, x_{2j}).

The restriction on the evolution of the ranks over the distribution of τ_1 is a consequence of the strict monotonicity assumption on $q(\cdot)$ with respect to the *scalar* rank variable v_j. Given that this assumption is not satisfied in most real-world scenarios, in this paper, I want to allow

[8]This assumption could also be satisfied if there was no overlap between groups. I rule out this possibility since this is not satisfied in most economic applications.

for the possibility that conditional on covariates, groups can differ in more moments than their mean. In this way, groups can be at different ranks at different values of τ_1.

A straightforward extension would include a bivariate v_j, where one element determines the mean and the other the variance. This corresponds to $y_{ij} = h(x_{1ij}, x_{2j}, v_j^{(1)}, u_{ij}) + f(x_{1ij}, x_{2j}, v_j^{(2)})$. Hence, conditional on the covariates x_{1ij} and x_{2ij}, the outcome has the same distribution across groups, but different locations and variances. The heterogeneity in the variances arises due to the interaction between the individual rank variable u_{ij} and the group rank variable $v_j^{(1)}$. In this example, it is not feasible to completely separate u_{ij} and v_j, and the group rank can vary over τ_1. The τ_1-conditional quantile function in each group is $q(x_{1ij}, x_{2j}, v_j, \tau_1) = h(x_{1ij}, x_{2j}, v_j^{(1)}, \tau_1) + f(x_{1ij}, x_{2j}, v_j^{(2)})$. Yet, we can still construct a τ_2-conditional quantile function by noting that for each τ_1, there exist a scalar-valued function $v_j(\tau_1)$ such that $q(x_{1ij}, x_{2j}, v_j, \tau_1) = q(x_{1ij}, x_{2j}, v_j(\tau_1), \tau_1)$. With proper normalization and imposing monotonicity with respect to this scalar rank variable, we can construct the τ_2-conditional quantile function. To give an illustration, let $y_{ij} = u_{ij}(v_j^{(1)} + \epsilon) + v_j^{(2)}$ for some scalar ϵ. Then, $q(v_j, \tau_1) = \tau_1(v_j^{(1)} + \epsilon) + v_j^{(2)} = \tau_1\epsilon + \tau_1 v_j^{(1)} + v_j^{(2)}$ is the τ_1-conditional quantile function. It follows directly that $v_j(\tau_1) = \tau_1 v_j^{(1)} + v_j^{(2)}$ is the scalar valued function that ranks groups at τ_1. Clearly, the model can be further generalized. For instance, with a trivariate v_j, we could allow groups to be heterogeneous with respect to their skewness. Similarly, with an infinitely dimensional v_j, it would be possible to allow for unrestricted heterogeneity between groups, which is the approach that I take in the paper.

A more general model - In this paper, I do not restrict the heterogeneity between groups and allow v_j to be a possibly infinite-dimensional term. In this way, I allow the group-level conditional quantile functions to vary unrestricted with respect to τ_1. This enables groups to be at different ranks for different values of τ_1. For instance, the groups in the upper tail at $\tau_1 = 0.1$ might differ from groups in the upper tail of the distribution at $\tau_1 = 0.9$. At the same time, I maintain the assumptions on the scalar u_{ij}.[9] Thus, the τ_1-conditional quantile function in equation (2) remains unchanged.

To make the problem concrete, I consider the following linear specification:

$$y_{ij} = x_{1ij}'\beta(u_{ij}, v_j) + x_{2j}'\gamma(u_{ij}, v_j) + \alpha(u_{ij}, v_j), \tag{4}$$

where $\alpha(u_{ij}, v_j)$ is the intercept. It follows that the τ_1-conditional quantile function can be equivalently written as

$$Q(\tau_1, y_{ij}|x_{1ij}, x_{2j}, v_j) = x_{1ij}'\beta(\tau_1, v_j) + x_{2j}'\gamma(\tau_1, v_j) + \alpha(\tau_1, v_j), \tag{5}$$

where only the sum of the last two terms is identified since x_{2j} does not exhibit variation within groups.

[9]It is not possible to identify the structural function with a multidimensional u_{ij}. In such a case, the quantile function identifies the individuals that have an outcome equal to the corresponding quantile (see Hoderlein and Mammen, 2007).

Modeling the heterogeneity between groups still requires restricting the relationship between the τ_1-conditional quantile function and the possibly infinite dimensional vector $v_j = (v_j^{(1)}, v_j^{(2)}, \dots)$. As in the bivariate example, for each $\tau_1 \in (0,1)$, I assume that there exists a scalar-valued function $v_j(\tau_1)$ such that

$$q(x_{1ij}, x_{2j}, v_j, \tau_1) = q(x_{1ij}, x_{2j}, v_j(\tau_1), \tau_1).$$

Imposing strict monotonicity of $q(x_{1ij}, x_{2j}, v_j(\tau_1), \tau_1)$ with respect to $v_j(\tau_1)$, yields the τ_2-conditional quantile function of $Q(\tau_1, y_{ij}|x_{1ij}, x_{2j}, v_j)$,

$$Q(\tau_2, Q(\tau_1, y_{ij}|x_{1ij}, x_{2j}, v_j)|x_{1ij}, x_{2j}) = x_{1ij}'\beta(\tau_1, \tau_2) + x_{2j}'\gamma(\tau_1, \tau_2) + \alpha(\tau_1, \tau_2), \tag{6}$$

which I refer to as the (τ_1, τ_2)-conditional quantile function. Model (6) allows for substantial heterogeneity as all coefficients have two quantile indices: one for the heterogeneity across groups (τ_2) and one for the heterogeneity within groups (τ_1). The outer quantile function is the conditional quantile function of the conditional quantile function of the outcome within each group, $Q(\tau_1, y_{ij}|x_{1ij}, x_{2j}, v_j)$. Thus, τ_2 ranks the groups (conditional on the covariates) according to *their* τ_1-conditional quantile function, and group ranks can change over the within distribution. It is important to note that the assumptions of a scalar-valued function $v_j(\tau_1)$, a scalar rank variable u_{ij}, and strict monotonicity of $q(\cdot)$ are necessary primarily for providing a structural interpretation of the conditional quantile function in equation (6) (see Matzkin, 2003; Torgovitsky, 2015). These assumptions imply rank invariance along both dimensions—that is, the ranking of units within groups and the ranking of groups at specific values of τ_1 remain unaffected by changes in x_{1ij} or x_{2j}. However, even without these assumptions, we can still construct model (6), and it continues to identify well-defined parameters. Below, I discuss the interpretation of the parameters in both cases.

The main advantage of this way of ordering groups is that it remains agnostic with respect to the social welfare function of the policymaker. When groups contain heterogeneous agents, ranking them is a non-trivial task without specifying a social welfare function. A utilitarian policymaker would rank the groups according to their mean. However, using the mean (or median) outcome to rank groups is unsatisfactory for at least two reasons. First, an equality-minded policymaker is not indifferent over two allocations with the same mean but different variances. On the contrary, it is possible to find an allocation with a smaller mean that is strictly preferred to an alternative assignment with a higher variance. Second, this ranking does not provide information about which part of the within distribution is driving the differences between groups, and a few outlying observations could have a large effect on this measure of between-group inequality. Later, I provide an example where comparing averages across regions shows substantial income differences across regions. However, a large part of these differences are driven by high top wages in a few regions. While comparing regional medians does not suffer from the former problem, it compares regions at a single point of the within distribution and might fail to capture differential labor market situations for a large portion of the workers. I show that this is the case mostly for low-income workers. These weaknesses also extend to other methods used to assess within and between heterogeneity, such as variance decompositions. Instead, with the two-dimensional quantile function, I can provide information about which part of the within distribution is driving the between heterogeneity. Specifically, if groups were heterogenous only due to different location parameters, then the group ranks would remain stable over the distribution of τ_1. By contrast, if the shape of the conditional distribution varies over groups, we expect group ranks to change over τ_1. This implies that the between heterogeneity depends on the within dimension τ_1 in an unrestricted way, which also implies that a decomposition is no longer possible. Clearly, a unified notion of group ranks can also be constructed. For instance, in Section 3, I show that a social welfare function can be used to assign welfare weights to each group, enabling the construction of a unified measure of group order.

The price to pay for this flexibility is that the interpretation of the coefficients becomes more complicated as the groups' ranks vary over τ_1. Further, with individual-level covariates, the ranks may vary even within the groups.[10] Yet, this last point is common with quantile models. The coefficient vectors $\beta(\tau_1, \tau_2)$ and $\gamma(\tau_1, \tau_2)$ tell how the (τ_1, τ_2)-conditional quantile function responds to a change in x_{1ij} or x_{2j} by one unit. To facilitate the interpretation, it is helpful to fix τ_1. For example, $\beta(0.5, \tau_2)$ gives the effect of x_{1ij} on the τ_2-conditional quantile function of the group (conditional) medians. Hence, it allows us to assess the effect of x_{1ij} on the distribution of group medians, with groups with the highest medians positioned at the top and those with the lowest medians at the bottom of the distribution.

Interpreting these coefficients as the effects for individuals at a specific point of the distribution requires rank invariance over treatment states.[11] Given the multi-dimensionality of the model, rank invariance must hold both within groups and between groups at a given within rank. Rank invariance within groups requires that within-group ranks do not change over treatment states. Instead, rank invariance between groups requires that for each τ_1, the ranks between groups remain stable over treatment states. While this is a strong assumption, there are cases where unconditional rank invariance is violated but still holds within and between groups. For example, effect heterogeneity over the distribution of groups could violate rank invariance in the population. With rank invariance, the coefficients can be interpreted as individual effects, and $\beta(\tau_1, \tau_2)$ (or $\gamma(\tau_1, \tau_2)$) gives the quantile effects for individuals at the τ_1 percentile of their groups, belonging to a group at the τ_2 percentile, where this second distribution is viewed from their perspective. Clearly, if an individual is in the lower tail of the within-group distribution,

[10] For example, a group (e.g., region) might have a different rank for high- and low-educated individuals.

[11] A rank invariance (or rank preservation) assumption is used, for example, in Heckman and Clements (1997) and Chernozhukov and Hansen (2005). Without a rank stability assumption, individual treatment effects are not identified. Chernozhukov et al. (2023) suggest conditional prediction intervals that can be obtained with a relaxation of this assumption.

she will prefer groups with relatively high low wages and a compressed wage distribution. Differently, individuals at the top of the within-group wage distribution will favor groups with high top wages.

The two-dimensional quantile function of the outcome is directly related to the conditional cdf of the outcome y by the following transformation:

$$F_{Y|X}(y|x) = \int_0^1 \int_0^1 1\{q(x, \tau_1, \tau_2) \leq y\} d\tau_2 d\tau_1. \tag{7}$$

Inverting the conditional cdf in equation (7) yields the one-dimensional conditional quantile function. Hence, the one-dimensional quantile function is a function of the two-dimensional one, and no information is lost when modeling both dimensions. Further, this implies that it is always possible to present the results with a different and/or unified rank variable. For example, one might be interested in looking at treatment effect heterogeneity over two dimensions, where the second dimension ranks groups according to their median (or any other percentile). This requires estimating the ranks of each individual so as to construct individualized treatment effects. Then, the groups can be sorted by their median rank, and the treatment effect at different values of τ_1 can be plotted against their group rank.

Example 1. Without covariates

I now consider a special case of model (6) where there are no regressors, and the model provides a quantile function of the outcome over two dimensions. The τ_1-conditional quantile function in group j simplifies to

$$Q(\tau_1, y_{ij}|v_j) = \alpha(\tau_1, v_j),$$

where $Q(\tau_1, y_{ij}|v_j)$ is τ_1th-percentile of the outcome y_{ij} in group j. It follows directly that

$$Q(\tau_2, Q(\tau_1, y_{ij}|v_j)) = \alpha(\tau_1, \tau_2)$$

is the τ_2th percentile, over all groups, of the τ_1th group percentiles.

Imagine a scenario where groups are defined by geographical regions, and the outcome y_{ij} represents the income earned by individual i in region j. This model enhances our understanding of inequality within and between these regions and sheds light on the variation of the within percentiles of the outcome over groups. For example, if differences are predominantly within regions, we would observe significant variations along the τ_1 dimension and relatively smaller differences along the τ_2 dimension. Additionally, this model enables us to determine whether heterogeneity between regions becomes more pronounced for higher values of τ_1, providing insights into how the higher end of the wage distribution varies across groups. It thus offers a nuanced perspective on the dynamics of inequality within and between geographical regions.

These heterogeneous coefficients are identified by two-step quantile regression. (i) The conditional quantile function in each group is identified by τ_1 quantile regressions of y_{ij} on x_{1ij} for each group separately. (ii) The second dimension is identified by τ_2 quantile regressions of the fitted values from the first-stage on x_{1ij} and x_{2j}.

Remark 1 (Within versus between distributions). The model discussed in this paper focuses on simultaneously estimating the effect on the distribution of the outcome within and between groups. Melly and Pons (2024) consider a similar model where the heterogeneity arises from the individual rank variable u_{ij} and the focus is on the within distribution. Starting from equation (6) and assuming that $(x_{1ij}, x_{2j}) \perp\!\!\!\perp v_j$, it is possible to obtain their model by integrating over v_j:[12]

$$\begin{aligned}\mathbb{E}\left[Q\left(\tau_1, y_{ij}|x_{1ij}, x_{2j}, v_i\right)|x_{1ij}, x_{2j}\right] =& x_{1ij}'\int \beta(\tau_1, v)dF_V(v) + x_{2j}'\int \gamma(\tau_1, v)dF_V(v)\\ &+ \int \alpha(\tau_1, v)dF_V(v)\\ =& x_{1ij}'\bar{\beta}(\tau_1) + x_{2j}'\bar{\gamma}(\tau_1) + \bar{\alpha}(\tau_1).\end{aligned}$$

Hence, when model (6) holds, they identify the average effects over groups at the τ_1 quantile of the within distribution. The parameters of this model are identified by a first-stage group-by-group quantile regression followed by a least squares (or GMM) second stage. Hence, the first step estimator models the within distribution, and the second step averages the results over groups.

If only the heterogeneity of average outcomes between groups is of interest, one could consider the conditional quantile function of the conditional expectation function in each group. Starting from equation (4) and assuming that $(x_{1ij}, x_{2j}) \perp\!\!\!\perp u_{ij}$, we attain

$$Q(\tau_2, \mathbb{E}_{i|j}[y_{ij}|x_{1ij}, x_{2j}]|x_{1ij}, x_{2j}) = x_{1ij}'\bar{\beta}(\tau_2) + x_{2j}'\bar{\gamma}(\tau_2) + \bar{\alpha}(\tau_2),$$

with

$$\begin{aligned}\mathbb{E}_{i|j}[y_{ij}|x_{1ij}, x_{2j}] =& x_{1ij}'\mathbb{E}_{i|j}[\beta(u_{ij}, v_j)|x_{1ij}, x_{2j}] + x_{2j}'\mathbb{E}_{i|j}[\gamma(u_{ij}, v_j)|x_{1ij}, x_{2j}] + \mathbb{E}_{i|j}[\alpha(u_{ij}, v_j)|x_{1ij}, x_{2j}]\\ =& x_{1ij}'\bar{\beta}(v_j) + x_{2j}'\bar{\gamma}(v_j) + \bar{\alpha}(v_j),\end{aligned}$$

where the notation $\mathbb{E}_{i|j}$ stresses that the expectation is taken conditional on the group. This setting is common in empirical research where only aggregated data is available.

If the primary focus is on heterogeneity between groups, one may prefer to study heterogeneities in the median outcome rather than the average. This choice aligns with the framework suggested in this paper, where the specific quantile of $\tau_1 = 0.5$ is considered.

3 Application to Policy Evaluation and Optimal Treatment Assignment

In this section, I explore two conceptual frameworks that align with the model proposed in this paper. These examples parallel the interpretation of quantile regression results. The first

[12] Note that Melly and Pons (2024) include the intercept in x_{2j}.

framework focuses on the impact on the distribution, while the second requires rank invariance to recover individual effects.

3.1 Distributional Effects: Policy Evaluation

In settings with an inequality-minded policymaker (see e.g. Kitagawa and Tetenov, 2021), researchers often consider a rank-dependent welfare function where welfare is a function of a weighted average of the outcomes with weights that depend on the rank of an individual in the population:

$$W := \int\int Y_{ij} \cdot w(\text{Rank}(Y_{ij}))di\ dj. \tag{8}$$

Such a welfare function can be written as a function of the *cdf* of the outcome, $F(y)$:

$$W_{\Lambda}(F) := \int_{0}^{\infty} \Lambda(F(y))dy \tag{9}$$

where $\Lambda(\cdot) : [0,1] \to [0,1]$ is a nonincreasing, nonnegative convex function with $\Lambda(0) = 1$ and $\Lambda(1) = 0$. Given that $\Lambda(\cdot)$ is convex, we can equivalently write the social welfare function in equation (9) as a weighted average of the outcomes:

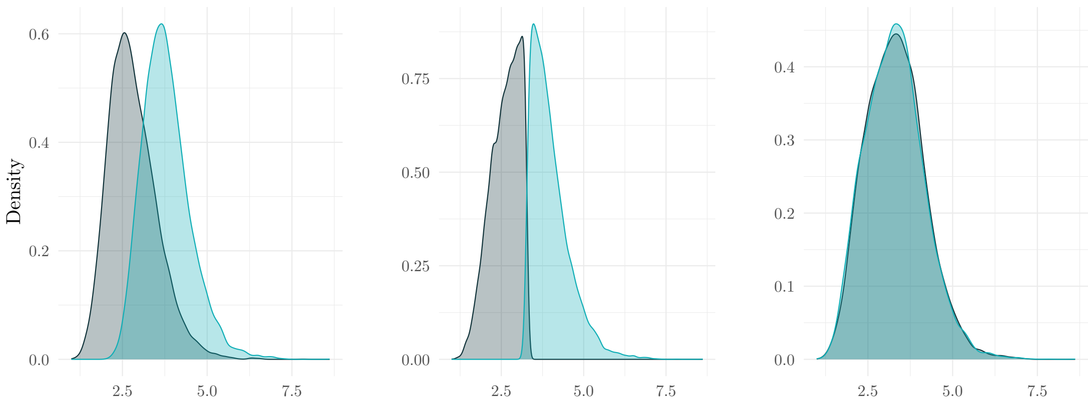
$$W_{\Lambda}(F) = \int_{0}^{1} F^{-1}(\theta)w(\theta)d\theta, \tag{10}$$

where $F^{-1}(\theta)$ is the unconditional quantile function of y and the weights $w(\theta) := \frac{d(1-\Lambda(\theta))}{d\theta}$ depend on the population quantiles θ. This social welfare function is quite general and comprises, for example, the extended Gini family. However, when dealing with grouped data, this welfare function does not consider the structure of the data and ignores the role of inequality within and between groups. For instance, all allocations with the same marginal distribution of the outcome yield identical welfare, irrespective of the distribution within and between the groups; thus, the policymaker should be indifferent between any two such allocations. Figure 1 shows three different allocations with generated data for two groups. In Allocation A, both groups have the same outcome distribution, with the exception that the blue group has a higher mean. Allocation B considers a scenario with high segregation where high-outcome individuals belong to the blue group, and low-outcome individuals are in the grey group. This allocation yields a lower within-group inequality, yet between-group inequality is larger. Allocation C minimizes inequality between groups by randomly assigning individuals to either group. However, this allocation has the highest within-group inequality. Despite the visible difference in within and between inequality, these allocations have the same marginal distribution of the outcome; hence, if welfare is measured by equation (10), these allocations are welfare equivalent.

To allow for potential trade-offs over different dimensions of inequality, I consider a welfare function that is a weighted average of the outcomes with weights that depend on both within and between ranks:

$$W = \int_{0}^{1}\int_{0}^{1} q(\tau_1, \tau_2) \cdot w(\tau_1, \tau_2)d\tau_2 d\tau_1 \tag{11}$$

Figure 1: Allocations with the Same Marginal Outcome Distribution



(a) Allocation A (b) Allocation B (c) Allocation C

Notes: The figure shows the kernel densities of two groups in three different allocations with the same outcome distribution but different redistribution between the groups.

where $q(\tau_1, \tau_2)$ is the two-level quantile function and $w(\tau_1, \tau_2)$ are the welfare weights assigned to units at the τ_1 percentile in their group belonging to a group at the τ_2 percentile.

The main advantage of this welfare function is its ability to account for the interdependencies between within-group and between-group inequality. For instance, when regions define groups, "*local inequality is actually the inverse of area-level income segregation*" (Glaeser et al., 2009). Whether reducing inequality in one dimension at the cost of increasing it in the other is desirable from a policy perspective and how to evaluate changes in both dimensions depends on the welfare function. The social welfare in equation (11) allows us to consider these trade-offs explicitly. Moreover, both dimensions of inequality are directly relevant to welfare.

Within-group inequality can have welfare effects through multiple channels. First, individuals might include a relative component in their utility functions, comparing themselves with their peers, neighbors, and co-workers (see, for example, Galí, 1994; Luttmer, 2005; Card et al., 2012). As a result, their relative rank within the group matters. Using this welfare function, policymakers can assign a higher weight to the lower tail of the within-group distribution to account for these externalities. Second, within-group inequality might cause other externalities. To give an illustration, Breza et al. (2018) find that pay inequality among production workers does, in some circumstances, reduce output, and Fehr et al. (2020) document that inequality reduces trust.[13] Additionally, inequality within groups is associated with negative outcomes. For instance, Glaeser et al. (2009) document that more unequal cities have higher murder rates, and Chetty and Hendren (2018b) show that areas with higher income inequality are linked to lower outcomes for children from low-income families.[14]

[13] Støstad and Cowell (2024) model inequality as an externality. However, inequality is modeled only using the unconditional distribution.

[14] Conversely, a certain degree of local inequality or heterogeneity might be beneficial by providing employment opportunities for individuals in the lower tail of the distribution (Mazzolari and Ragusa, 2013) or fostering empathy

Between-group inequality, or regional inequality, is central to the goals of many political institutions. Federal states or supranational entities (e.g., the European Union) rely on fiscal federalism to transfer resources across jurisdictions. For instance, the EU Regional Development fund aims at *correcting imbalances between regions* and after the German reunification, the concept of *equivalent living conditions*[15] played a major role in German regional policy, aiming to mitigate regional disparities. Similar to within-group inequality, between-group inequality can have adverse effects. The quality of the neighborhood in which children grow up significantly impacts their future outcomes (Chetty et al., 2016; Chetty and Hendren, 2018a,b).[16] Therefore, modeling welfare using equation (11) allows us to consider the welfare effects along both dimensions and the trade-offs between the two.

Starting from the welfare function (11), I now show that depending on the weights $w(\tau_1, \tau_2)$, it simplifies to welfare functions where only the mean of the outcome or only the within or between component matters. For example, let $w(\tau_1, \tau_2) = 1$; then we obtain a Benthamite (or utilitarian) welfare function:

$$W = \mathbb{E}[y_{ij}].$$

If $w(\tau_1, \tau_2) = w(\tau_1)$, the welfare function simplifies to a weighted average of the expectation of the group quantiles:

$$W = \int_0^1 \mathbb{E}_j[q(\tau_1, v_j)]w(\tau_1)d\tau_1,$$

where only the within distribution matters. This welfare function would be relevant if the focus was solely on studying (the effects on) the within distribution (see, for example, Autor et al., 2021; Friedrich, 2022; Autor et al., 2016; Engbom and Moser, 2022; Lang et al., 2023).

Differently, if $w(\tau_1, \tau_2) = w(\tau_2)$ the welfare function (11) simplifies to the weighted average of the conditional expectation in each group:

$$W = \mathbb{E}_j\left[\int_0^1 q(\tau_1, v_j)d\tau_1 w(v_j)\right] = \mathbb{E}_j\left[\mathbb{E}_{i|j}[y_{ij}]w(v_j)\right].$$

In this case, the welfare function assigns different weights to different groups based on their mean, and the outcome distribution within the group does not matter. This welfare function is relevant if the aim is to study inequality in average outcomes or regional GDP per capita (see, for example, Becker et al., 2010; Busso et al., 2013). An argument for considering only inequality within regions is that, theoretically, it is always possible to redistribute within the region. However, this is not the case, as redistribution is costly. Further, as long as a policymaker is not indifferent between two allocations where the second is a mean preserving spread of the within-group distribution of the first, a function considering both components of inequality better captures welfare.

for the poor (Glaeser, 2000).

[15]Gleichwertige Lebensverhältnisse.

[16]Poor regional or labor market performance is associated with different adverse outcomes such as an increase in fatal drug overdoses (Pierce and Schott, 2016), and single-mother families (Autor et al., 2021).

Compared to equation (11), these welfare functions do not allow for any trade-off between different dimensions of inequality and may, therefore, capture an incomplete picture of welfare. However, given their tight interconnectedness and since most policies are likely to affect both dimensions, it is surprising that they are rarely studied together.

The weighting function is crucial as it determines the extent to which a policymaker is willing to accept an increase in within-group inequality to decrease between-group inequality while keeping welfare constant. Different weighting functions can be considered. For instance, a possible extension of the Gini social welfare function to two dimensions considers $w(\tau_1, \tau_2) = 2(1 - \omega_1\tau_1 - \omega_2\tau_2)$ with $\omega_1 + \omega_2 = 1$. If $\omega_2 = 0$, the problem reduces to[17]

$$W = \int_0^1 E_j[q(\tau_1, v_j)]2(1-\tau_1)d\tau_1 = E[y]\left[\frac{\int_0^1 E_j[q(\tau_1, v_j)]2(1-\tau_1)d\tau_1}{E[y]}\right] = E[y](1 - I_{Gini}),$$

where $\int_0^1 E_j[q(\tau_1, v_j)]d\tau_1 = E[y]$ and $I_{Gini} = 1 - \frac{\int_0^1 E_j[q(\tau_1, v_j)]2(1-\tau_1)d\tau_1}{E[y]}$ is the Gini index in the average group (see Kitagawa and Tetenov (2021) for more details on the one-dimensional case). With this weighting function, the welfare weights only decay linearly in the ranks within and between ranks. Alternatively, a weighting function that is not additively separable in τ_1 and τ_2 would allow for a more complex evolution of the weights where the effect of one rank on the welfare weights depends on the other dimension.

3.2 Individual Effects: Optimal Treatment Assignment

This subsection considers a different setting that requires the stronger assumption of rank invariance over treatment states. Under this additional assumption, the individual treatment effects are identified and can be used to pinpoint optimal treatment rules. I use the conventional potential outcome framework with a treatment variable D. Let $Y_{ij}(d)$ denote the potential outcome under treatment state $d \in \{0, 1\}$. The object of interest in this paper is the two-level quantile function of the potential outcomes, conditional on observed characteristics $x_{ij} = (x'_{1ij}, x'_{2j})'$:

$$q(d, x_{ij}, \tau_1, \tau_2),$$

which can be used to compute the conditional quantile treatment effects over both dimensions:[18]

$$q(1, x_{ij}, \tau_1, \tau_2) - q(0, x_{ij}, \tau_1, \tau_2).$$

Consider a policymaker who observes data from a *sample* population with a given group structure and has to decide whom to treat in a *target* population (subject to some capacity/budget constraint) by maximizing a rank-dependent social welfare function. I consider a

[17]For the one dimensional Gini social welfare function see, for example, Blackorby and Donaldson (1978) and Weymark (1981).

[18]Integrating the conditional quantile treatment effects over both τ_1 and τ_2 yields average treatment effects:

$$ATE = \int_0^1 \int_0^1 [q(1, x_{ij}, \tau_1, \tau_2) - q(0, x_{ij}, \tau_1, \tau_2)]\, d\tau_2 d\tau_1.$$

static setting where the policy-maker chooses whom to treat out of a pool of individuals or groups based on their *unobserved* ranks. This is in contrast to a dynamic setting (e.g., Adusumilli et al., 2019), where the policymaker has to make sequential decisions, as well as to the one in Kitagawa and Tetenov (2021), where the goal is to assign optimally individuals to treatment based on *observable* covariates. Baseline outcomes can be in the set of covariates; however, these are not always available (see, e.g., Tarozzi et al., 2015). Further, this setting also differs from the one considered in Kaji and Cao (2023), which allows for heterogeneity only across one dimension. Instead, with grouped data, one might want to exploit treatment effect heterogeneity within and between groups to more efficiently allocate the treatment. As in the one-dimensional case, targeting the most deprived individuals or groups is not necessarily optimal (see Haushofer et al., 2022). Instead, there could be a trade-off between targeting the most deprived units and targeting the units that profit the most from the treatment. This trade-off is particularly relevant in this setting as group membership could be an important determinant of impact.

For simplicity, I consider a case without covariates; however, the framework can be easily extended to include other variables.[19] The policymaker maximizes social welfare over a class of feasible policies $\mathcal{G} \in \{u_{ij}, v_j \in (0,1) \times (0,1)\}$. Hence, the goal is to select a treatment rule that assigns individuals to treatment based on their ranks (u_{ij}, v_j).[20] If these ranks were observed, it would be possible to include them in the set of covariates, and the problem would coincide with the one in Kitagawa and Tetenov (2021). However, here, we have to rely on distributional methods to estimate these ranks.

In this example, the model discussed in the paper remains relevant even if the welfare function depends only on the unconditional distribution of the outcome.[21] Hence, for this illustration, I consider a welfare function as in equation (9). However, the results can be generalized to other welfare functions. When a treatment rule G is applied to the target population, the social welfare is proportional to:

$$W_\Lambda(F_G) = \int_0^\infty \Lambda(F_G(y))dy, \tag{12}$$

where F_G is the distribution of the outcome y_{ij} under treatment rule G:

$$y_{ij} = 1\{(u_{ij}, v_j) \in G\} y_{ij}(1) + 1\{(u_{ij}, v_j) \notin G\} y_{ij}(0),$$

and the optimal treatment rule solves[22]

$$G^* \in \arg\max_{G \in \mathcal{G}} W(G). \tag{13}$$

[19] If the inclusion of additional variables is necessary to identify the distribution of potential outcomes, it is straightforward to recover the unconditional distribution by integrating out the covariates.

[20] I write v_j for ease of notation. However, v_j should be regarded as the rank variable ranking groups at a specific point of the within distribution. A version of the problem could consider a treatment rule of the type $G = 1(u_{ij} \le \tilde{u}, v_j \le \tilde{v})$ for some fixed $\tilde{u}$ and $\tilde{v}$. A similar setting is considered in Kaji and Cao (2023). Yet, this class of decision rules might be restrictive in the presence of a trade-off between deprivation and impact, as discussed in Haushofer et al. (2022).

[21] Recall that the advantage of the method in this setting is that it allows the exploitation of treatment effect heterogeneity over both dimensions simultaneously and, therefore, can more efficiently allocate the treatment.

[22] Solving problem (13) is nontrivial as it lacks a closed-form solution even if we knew the distribution of the

To make the problem operational, we need to identify individual treatment effects and assign welfare weights to each observation under each policy rule. Recall that under rank invariance, $q(1, \tau_1, \tau_2) - q(0, \tau_1, \tau_2)$ gives the treatment effect for an individual at quantiles (τ_1, τ_2). Further, using equation (7), we can identify the conditional quantile function of the potential outcomes in the population under a given policy rule G. These objects can then be used to assign ranks and welfare weights w_{ij} to each observation. Notably, individuals at different τ_1 percentiles may share the same welfare weight due to their placement in different groups. However, individuals with the same y_{ij} have identical welfare weights. Summing the welfare weights within groups gives the weights assigned to group j:[23]

$$w_j = \sum_{i=1}^{n} w_{ij}.$$

While the rank of a group changes over τ_1, these groups' weights can offer a unified and welfare-based measure of a group's rank or deprivation.

To find the optimal treatment assignment rule that maximizes the social welfare function in the target population, we need to impose some assumptions on the individual treatment effects in the sample and target populations. For instance, Kitagawa and Tetenov (2018) assume that the joint distribution of the potential outcome and covariates is the same in both populations. In this setting, I need that the joint distribution of $(y_{ij}(1), y_{ij}(0), v_j, u_{ij})$ is the same in both populations. Since the ranks are normalized, u_{ij} follows the same distribution in both populations by construction. Therefore, one can equivalently assume that for all u_{ij}, the joint distribution of $y_{ij}(1), y_{ij}(0), v_j | u_{ij}$ is the same in the sample and target populations.[24]

In summary, the two-level conditional quantile function of potential outcomes in the sample population enables the identification of treatment effects for an individual at a given individual and group rank. With this information, we can identify $y_{ij}(1)$ for all i and j. Subsequently, for each $G \in \mathcal{G}$, we can construct the counterfactual outcome $y_{ij} = 1\{(u_{ij}, v_j) \in G\} y_{ij}(1) + 1\{(u_{ij}, v_j) \notin G\} y_{ij}(0)$ along with the corresponding outcome distribution and welfare. The optimal treatment rule then solves equation (13).

potential outcomes (Kitagawa and Tetenov, 2021). One difficulty arises because the welfare weights assigned to an individual might depend on the treatment assignment of other agents. Intuitively, the welfare weight assigned to an individual is weakly increasing in the outcomes of the other individuals.

[23] In the case of unbalanced groups, larger groups are more likely to have a higher welfare weight. This feature can be desirable if the cost of assigning a group to the treatment does not depend on the number of observations in this group. Alternatively, it is possible to compute the average weights.

[24] This assumption is more restrictive than necessary. For example, if the distributions of the potential outcomes in the target population are the same as in the sample population up to a location parameter, it would not be a problem in this setting. Further, I do not need to impose the restriction on the entire infinite dimensional vector v_j. Instead after conditioning on u_{ij} one could equivalently assume that $y_{ij}(1), y_{ij}(0), v_j(u_{ij}) | u_{ij}$ follow the same distribution on both populations.

4 Estimator

For simplicity of notation, I consider the same set of quantiles to model both dimensions, although this is not a requirement. For each dimension, I approximate the function using $\#\tau$ different quantiles. I propose a two-step quantile regression estimator for model (6). The first stage consists of group-by-group quantile regressions. For each group j and quantile τ_1, the outcome is regressed on the individual level variables x_{1ij} using quantile regression. Then, for each group and τ_1, the fitted values are saved. In the second stage, for each τ_1, the first-stage fitted values $\hat{y}_{ij}(\tau_1)$ are regressed on all variables using all observations. This is again done with quantile regression for each τ_2. Thus, estimation comprises $\#\tau \times m$ first stage regression and $\#\tau \times \#\tau$ second stages. Formally, the first-stage quantile regression solves the following minimization problem for each group j and quantile τ_1 separately:

$$\hat{\beta}_j(\tau_1) := \left(\hat{\beta}_{1,j}(\tau_1), \hat{\beta}_{2,j}(\tau_1)'\right)' = \underset{(b_1,b_2)\in\mathbb{R}^{dim(x_1)+1}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \rho_{\tau_1}(y_{ij} - b_1 - x'_{1ij}b_2), \tag{14}$$

where $\rho_\tau(x) = (\tau - 1\{x < 0\})x$ for $x \in \mathbb{R}$ is the check function. For group j, the true vector of first stage coefficients is given by $\beta_j(\tau_1) := \beta(\tau_1, v_j) = (\alpha(\tau_1, v_i) + x'_{2j}\gamma(\tau_1, v_j), \beta(\tau_1, v_j)')'$ and the fitted values $\hat{y}_{ij}(\tau_1) = \hat{\beta}_{1,j}(\tau_1) + x'_{1ij}\hat{\beta}_{2,j}(\tau_1)$ are estimators of the τ_1-conditional quantile function $Q(\tau_1, y_{ij}|x_{ij}, v_j)$.

The second stage quantile regression then solves for all (τ_1, τ_2):

$$\hat{\delta}(\hat{\beta}(\tau_1), \tau_2) = \underset{(a,b,g)\in\mathbb{R}^{dim(x)+1}}{\arg\min} \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} \rho_{\tau_2}(\hat{y}_{ij}(\tau_1) - x'_{2j}g - x'_{1ij}b - a), \tag{15}$$

where the notation stresses the dependence on the first step and $\delta = (\alpha, \beta', \gamma')'$.

Implementing the estimator is straightforward, requiring only programs for quantile regression. The lack of a closed-form solution for quantile regression might increase computing time, but recent algorithms enable simultaneous estimation of numerous quantiles, significantly improving computational speed. Moreover, the first stage is easily parallelizable, as all first-stage quantile regressions run independently across the groups, and the second stage is also parallelizable with respect to τ_1.

Ensuring the monotonicity of the estimated two-level quantile functions across both dimensions might require a rearrangement operation, as suggested in Chernozhukov et al. (2009, 2010). Due to the nested structure of the problem, rearrangement along the τ_1 dimension should be performed after the first stage. Monotonicity of the first stage in all groups guarantees that the second stage quantile regression remains monotonic along the τ_1 dimension. Rearrangement along the τ_2 dimension can be implemented subsequent to the second stage.

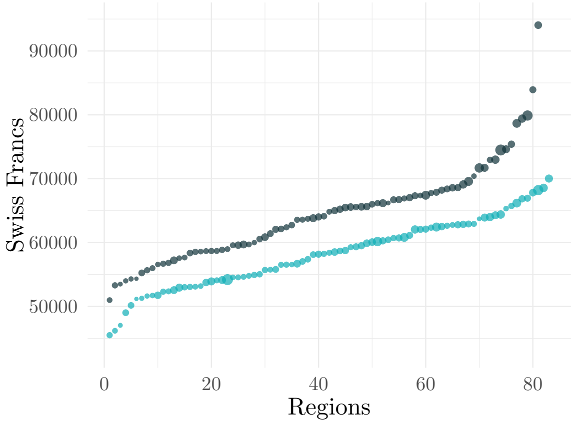
Remark 2 (Alternative estimators - instrumental variables). Model (6) assumes that the variation of both the x_{1ij} and x_{2j} are exogenous so that quantile regression in both stages yields consistent estimates. If this is not the case, the estimator suggested here can be easily extended to accommodate instrumental variables. Depending on which variables are assumed to be endogenous, either the second stage or both stages could be estimated using an instrumental variable quantile regression estimator (see, e.g., Chernozhukov and Hansen, 2005). Future research should explore this possibility.

Example 2 (Continuation of Example 1)**.** Consider the setting of Example 1, where the goal is to analyze income heterogeneity between and within geographical regions, and there are no covariates. One possibility for analyzing income heterogeneity across regions is to consider differences in median or average wages. Using administrative data from the Federal Statistical Office of Switzerland, I show that these two measures fail to capture important features of income heterogeneity between regions. Groups are defined by 2-digit ZIP codes. These groups are on a finer resolution than Swiss cantons and offer a more precise measure of labor markets. The dataset comprises information on 4.2 million individuals aged between 30 and 63, divided into 83 groups in the year 2021. Since there are no covariates, the first-step estimation consists of calculating the τ_1 sample quantiles in the groups. Subsequently, in the second stage, I compute the τ_2 quantiles, over groups, of the group-level quantiles. I consider the set of quantiles $\{0.01, 0.02, \dots, 0.99\}$ in both stages. Due to the absence of individual-level covariates, the first-stage fitted values are constant within the groups so that it is possible to collapse the dataset at the group level after the first step to increase computation speed. In this case, weights should be included in the second step.

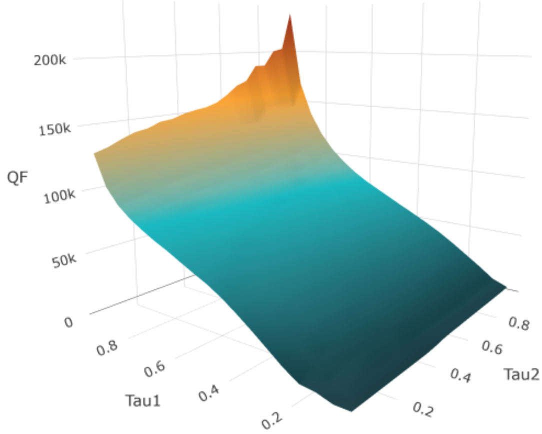
Figure 2 shows the regional averages and medians of yearly income in panel (a) and the two-dimensional quantile function of the same variable in panel (b). Both regional averages and medians are independently arranged from low to high. The darker dots in Figure 2a reveal substantial differences in average income across regions. However, Figure 2b shows that a large portion of these differences in mean income can be attributed to high top incomes in a few regions. Across most of the τ_1 distribution, the differences across regions are substantially smaller compared to the right tail of the within distribution. Thus, the differences in average wages shown in Figure 2a, not only mask substantial within-region income heterogeneity but are predominantly driven by differences in top incomes.

The lighter dots in Figure 2a show that the heterogeneity in median wages across regions is substantially smaller than the heterogeneity in average income. However, this measure solely reflects the heterogeneity at one point of the distribution, potentially overlooking the labor market situation of a considerable portion of workers. More specifically, median wages within a region might poorly relate to the labor market situation of low earners. To see this, we can look at the correlation of the group ranks for different values of τ_1. Table 1 shows that the ranks at the very low end of the distribution exhibit only a small correlation with the ranks at other deciles of the distribution. On the other hand, there is a high correlation between the ranks at the middle and the top of the distribution. This suggests the presence of a different mechanism influencing the lower tail of the within distribution.

Figure 2: Income Heterogeneity Within and Between Regions



(a) Average and Median Income by Region



(b) Two-Dimensional Quantile Function of Income

Notes: Figure 2a shows the heterogeneity in average (dark blue) and median (light blue) yearly income across regions defined by 2-digit ZIP codes. Figure 2b shows the two-dimensional quantile function of yearly income within and between regions.

5 Asymptotic Theory

Notation - Let $\tau = (\tau_1, \tau_2)$ and denote the true parameter vectors $\beta_{j,0}(\tau_1)$ and $\delta_0(\beta_0, \tau) := \delta_0(\tau_2, \beta_0(\tau_1))$. To simplify notation, I suppress the dependency of δ and β_j on τ_1 and τ_2, unless necessary. For a random variable h_{ij}, $\mathbb{E}_{i|j}[h_{ij}]$ denotes the expectation over i in group j. Let K_1 be the dimension of x_{1ij} and K_2 be the number of regressors in x_2. Furthermore, let $K = K_1 + K_2 + 1$ be the total number of regressors. Finally, denote the $(K_1 + 1)$-dimensional vector of first stage regressors as $\tilde{x}_{ij} = (1, x'_{1ij})'$. I prove weak convergence of the whole quantile regression process for $\tau_1, \tau_2 \in \mathcal{T} \times \mathcal{T}$, where $\mathcal{T} \in (0, 1)$ is a compact set of quantile indices of interest. The symbol $\ell^\infty(\mathcal{T} \times \mathcal{T})$ denotes the set of component-wise bounded vector-valued functions of $\mathcal{T} \times \mathcal{T}$ and $\rightsquigarrow$ denotes weak convergence.

5.1 Consistency and Asymptotic Normality

The derivation of asymptotic results faces two primary challenges: the non-smoothness of the quantile regression objective function and the increasing dimension of the first stage as the number of groups diverges to infinity. Several studies have addressed the asymptotic properties of estimators with non-smooth objective functions, leveraging the smoothness of the limiting objective function (see, for example, Newey and McFadden, 1994). Notably, Pakes and Pollar (1989) study the properties of Z-estimators without imposing smoothness conditions on the sample equations. Building on this work, Chen et al. (2003) broadens the scope to two-step estimators,

Table 1: Correlation of Regions' Ranks over τ_1

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1.00								
0.2	0.73	1.00							
0.3	0.54	0.95	1.00						
0.4	0.48	0.90	0.98	1.00					
0.5	0.44	0.85	0.95	0.98	1.00				
0.6	0.37	0.80	0.91	0.95	0.98	1.00			
0.7	0.26	0.73	0.86	0.90	0.94	0.97	1.00		
0.8	0.18	0.67	0.81	0.86	0.89	0.94	0.99	1.00	
0.9	0.09	0.58	0.73	0.78	0.81	0.87	0.94	0.98	1.00

Note:
The table shows the correlation matrix of the regions' ranks at different values of τ_1.

where the parameter of interest depends on an infinite-dimensional preliminary parameter.

To derive the asymptotic results, I rely on results of Chen et al. (2003) and work within the framework of Z-estimators. Similarly to their paper, my second stage parameter vector depends on a preliminary first stage whose dimension increases with the sample size. To this end, I start by making the assumptions necessary to ensure that the first-stage quantile regression is well-behaved. For this first analysis, I build on the work of Volgushev et al. (2019) and Galvao et al. (2020) and make the following assumptions:

Assumption 1 **(Sampling).** *The observations $(y_{ij}, x_{ij})_{i=1,\dots,n,\ j=1,\dots,m}$ are i.i.d. across i and j.*

Assumption 2 **(Covariates).** *(i) For all $j = 1,\dots,m$ and all $i = 1,\dots,n$, $\|x_{ij}\| \leq C$ almost surely. (ii) The eigenvalues of $\mathbb{E}_{i|j}[\tilde{x}_{ij}\tilde{x}'_{ij}]$ and $\mathbb{E}[x_{ij}x'_{ij}]$ are bounded away from zero and infinity uniformly across j.*

Assumption 3 **(Individual level heterogeneity).** *The conditional distribution $F_{y_{ij}|x_{1ij},v_j}(y|x,v)$ is twice differentiable w.r.t. y, with the corresponding derivatives $f_{y_{ij}|x_{1ij},v_j}(y|x,v)$ and $f'_{y_{ij}|x_{1ij},v_j}(y|x,v)$. Further, assume that*

$$f_y^{max} := \sup_j \sup_{y\in\mathbb{R},\ x\in\mathcal{X}_1} |f_{y_{ij}|x_{1ij},v_j}(y|x,v)| < \infty,$$

and

$$\bar{f}'_y := \sup_j \sup_{y\in\mathbb{R},\ x\in\mathcal{X}_1} |f'_{y_{ij}|x_{1ij},v_j}(y|x,v)| < \infty.$$

where $\mathcal{X}_1$ is the support of x_{1ij}.

Assumption 4 **(Bounded density I).** *There exists a constant $f_y^{min} < f_y^{max}$ such that*

$$0 < f_{min} \leq \inf_j \inf_{\tau_1\in\mathcal{T}} \inf_{x\in\mathcal{X}_1} f_{y_{ij}|x_{1ij},v_j}(Q(\tau_1, y_{ij}|x_{ij}, v_j)|x,v).$$

These are standard assumptions in the quantile regression literature. Assumption 1, assumes that the observations are i.i.d. within and between groups. Assumption 2 requires that the regressors are bounded and that both matrices $\mathbb{E}_{i|j}[\tilde{x}_{ij}\tilde{x}'_{ij}]$ and $\mathbb{E}[x_{ij}x'_{ij}]$ are invertible. Assumptions 3 and 4 require smoothness and boundedness of the conditional distribution of the outcome variable y_{ij} given (x_{ij}, v_j), the density, and its derivatives. This first set of assumptions focuses primarily on establishing the behavior of the first-stage estimator and allows to apply Lemma 3 in Galvao et al. (2020).

Further, to ensure that the second-step quantile regression is well-behaved, I make the following assumptions:

Assumption 5 (**Group level heterogeneity**). *The conditional distribution $F_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(q|x)$ is twice continuously differentiable w.r.t. q, with the corresponding derivatives $f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(q|x)$ and $f'_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(q|x)$. Further, assume that*

$$f_Q^{max} := \sup_{\tau_1\in\mathcal{T},\ q\in\mathbb{R},\ x\in\mathcal{X}} |f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(q|x)| < \infty$$

and

$$\bar{f}'_Q := \sup_{\tau_1\in\mathcal{T},\ q\in\mathbb{R},\ x\in\mathcal{X}} |f'_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(q|x)| < \infty.$$

where $\mathcal{X}$ is the support of x_{ij}.

Assumption 6 (**Bounded density II**). *There exists a constant $f_Q^{min} < f_Q^{max}$ such that*

$$0 < f_{min} \leq \inf_{\tau_1,\tau_2\in\mathcal{T}\times\mathcal{T}} \inf_{x\in\mathcal{X}} f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x'_{ij}\delta_0(\tau)|x).$$

Assumption 7 (**Compact parameter space**). *For all τ, $\beta_{j,0}(\tau_1) \in int(\mathcal{B}_j)$ and $\delta_0(\beta_0,\tau) \in int(\mathcal{D})$, where $\mathcal{B}_j$ and $\mathcal{D}$ are compact subsets of $\mathbb{R}^{K_1+1}$ and $\mathbb{R}^K$, respectively.*

Assumption 8 (**Coefficients**). *For all $\tau_1, \tau'_1 \in \mathcal{T}$ and $j = 1, \ldots, m$, $\|\beta_j(\tau_1) - \beta_j(\tau'_1)\| \leq C|\tau_1 - \tau'_1|$. Further, for all $\tau, \tau' \in \mathcal{T}\times\mathcal{T}$ and $\|\delta(\tau) - \delta(\tau')\| \leq C|\tau_1 - \tau'_1| + \leq C|\tau_2 - \tau'_2|$.*

Assumptions 5 and 6 are the second stage counterpart of of assumptions 3 and 4, with the difference that the conditional distribution $F_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(q|x)$ is required to be *continuously* differentiable. This additional assumption on the distribution of the second stage dependent variable is sufficient to ensure that its second derivative is Lipschitz continuous. An implication of this assumption is that groups must be sufficiently heterogeneous and precludes the case where, conditional on the covariates, there are only a few types of groups at each τ_1. Assumption 7 requires the parameter spaces to be compact. Compactness of the parameter space is a common assumption in the quantile regression literature, see e.g., Honoré et al. (2002); Chernozhukov and Hansen (2006), and Zhang et al. (2019). Compactness of $\mathcal{D}$ is necessary to use the results in Chen et al. (2003), while compactness of $\mathcal{B}_j$ is useful as it directly implies that the covering integral is finite but could easily be relaxed. Assumption 8 ensures that the coefficients are continuous functions of the quantile indices.

Since quantile regression is consistent but not unbiased, we need the number of observations per group to diverge to infinity. At the same time, the second-stage quantile regression exploits the heterogeneity between groups, which is determined by the heterogeneity of the group-level quantile functions, a group-specific term. Thus, also the number of groups must diverge. The following assumption states two different growth rates of the number of observations per group relative to the number of groups:

Assumption 9 (Growth rates). *As $m \to \infty$, we have*

(a) $\frac{\log m}{n} \to 0$,

(b) $\frac{\sqrt{m}\log n}{n} \to 0$.

I show that the relative growth rate in Assumption 9(a) is sufficient for consistency of the estimator. This condition is exceptionally weak, as the number of observations per group can increase at an almost arbitrarily slow rate. Differently, weak convergence requires assumption 9(b). This second growth rate is relatively mild as the number of observations per group must grow faster than the square root of the number of groups. The first result of this paper states weak uniform consistency of the two-step estimator.

Theorem 1 (Uniform Consistency). *Let assumptions 1-9 be satisfied. Then as $m \to \infty$*

$$\sup_{\tau \in \mathcal{T} \times \mathcal{T}} ||\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)|| \xrightarrow{p} 0.$$

To establish asymptotic normality, I begin by showing that $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)$ can be approximated by a linear function of two terms, each accounting for the estimation errors arising from different steps of the estimation. If the first stage parameter vector $\beta_0(\tau_1) = (\beta_{0,1}(\tau_1)', \ldots, \beta_{0,m}(\tau_1)')'$ was known, the true second-stage parameter vector $\delta_0(\beta_0, \tau)$ uniquely[25] satisfies:

$$\mathbb{E}[m(\delta_0, \beta_0, \tau)] = 0 \tag{16}$$

with

$$m(\delta, \beta, \tau) = x_{ij}'[\tau_2 - 1(\tilde{x}_{ij}'\beta_j(\tau_1) \leq x_{ij}'\delta(\beta, \tau))]. \tag{17}$$

Let $M(\delta, \beta, \tau) = \mathbb{E}[m(\delta, \beta, \tau)]$, denote the sample counterpart $M_{mn}(\delta, \beta, \tau) = \frac{1}{mn}\sum_{j=1}^m \sum_{i=1}^n m(\delta, \beta, \tau)$, and note that while $M(\delta, \beta, \tau)$ is a smooth function, this property does not extend to $M_{mn}(\delta, \beta, \tau)$. Two expressions are central to establishing asymptotic normality. (i) The sample moment evaluated at the true parameters:

$$M_{mn}(\delta_0, \beta_0, \tau) = \frac{1}{mn}\sum_{j=1}^m \sum_{i=1}^n m(\delta_0, \beta_0, \tau), \tag{18}$$

[25]Under weak regularity conditions.

(ii) and the pathwise derivative of $M(\delta, \beta_0, \tau)$ in the direction $(\beta - \beta_0)$:

$$\Gamma_2(\delta, \beta_0, \tau)[\beta - \beta_0] = \frac{1}{m}\sum_{j=1}^{m} \bar{\Gamma}_{2j}(\delta, \beta_0, \tau)[\beta_j - \beta_{j,0}], \tag{19}$$

where $\Gamma_2(\delta, \beta_0, \tau)$ is $K \times ((K_1 + 1) \cdot m)$, $\Gamma_{2j}(\tau, \delta, \beta_0)$ is the jth $K \times (K_1 + 1)$ submatrix of $\Gamma_2(\delta, \beta_0, \tau)$ and $\frac{1}{m}\bar{\Gamma}_{2j}(\tau, \delta, \beta_0) := \Gamma_{2j}(\tau, \delta, \beta_0)$ with

$$\Gamma_{2j}(\delta, \beta_0, \tau) = \frac{\partial}{\partial \beta_j} M(\delta, \beta_0, \tau) = -\frac{1}{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1, y_{ij}|x_{ij}, v_j)|x_{ij}}(x_{ij}'\delta(\beta_0, \tau)|x_i) x_{ij}\tilde{x}_{ij}'\right]. \tag{20}$$

The expression in equation (18) is directly related to the leading term of a Bahadur expansion of the unfeasible estimator $\hat{\delta}(\beta_0, \tau)$:

$$\hat{\delta}(\beta_0, \tau) - \delta(\beta_0, \tau) = \Gamma_1(\delta_0, \beta_0, \tau)^{-1} \cdot \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} m(\delta_0, \beta_0, \tau) + o_p\left(m^{-1/2}\right), \tag{21}$$

where $\Gamma_1(\delta_0, \beta_0, \tau) = \frac{\partial M(\delta, \beta_0, \tau_2)}{\partial \delta} = \mathbb{E}[f_{Q(\tau_1, y_{ij}|x_{ij}, v_j)|x_{ij}}(x_{ij}'\delta_0(\beta_0, \tau)|x_{ij}) x_{ij} x_{ij}']$. Thus, equation (18) captures the estimation error that would arise due to random variation in the second stage if we knew the true first stage and equation (19) captures the effect of the first estimation error on the second-step estimates $\hat{\delta}(\hat{\beta}, \tau)$.

Heuristically, the idea is to approximate the asymptotic distribution of $\sqrt{m}(\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau))$ with the asymptotic distribution of $\Gamma_1(\delta_0, \beta_0, \tau)^{-1}\sqrt{m}\big(\sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau)\ [\hat{\beta}_j(\tau_1) - \beta_{j,0}(\tau_1)] + M_{mn}(\delta_0, \beta_0, \tau)\big)$. To this end, I show that the two expressions are asymptotically equivalent up to a term converging to zero fast enough. Then, if we can show that

$$\sqrt{m}\left(\frac{1}{m}\sum_{j=1}^{m} \bar{\Gamma}_{2j}(\delta_0, \beta_0, \tau)[\hat{\beta}_j(\tau_1) - \beta_{j,0}(\tau_1)] + M_{mn}(\delta_0, \beta_0, \tau)\right) \xrightarrow{d} N(0, \Omega_2(\tau)),$$

for some $\Omega_2(\tau)$, asymptotic normality follows.

The following Lemma establishes the asymptotic properties of equations (18) and (19).

Lemma 1. *Let the model in equation (6) and assumptions 1-7 hold. Then*

(i) Under assumption 9(b):

$$\sup_{\tau \in \mathcal{T} \times \mathcal{T}} \left\| \frac{1}{m}\sum_{j=1}^{m} \bar{\Gamma}_{2,j}(\delta_0, \beta_0, \tau)\left(\hat{\beta}_j(\tau_1) - \beta_{j,0}(\tau_1)\right)\right\| = o_p\left(\frac{1}{\sqrt{m}}\right), \tag{22}$$

where $\bar{\Gamma}_{2j}(\delta_0, \beta_0, \tau) = \mathbb{E}_{i|j}\left[f_{Q(\tau_1, y_{ij}|x_{ij}, v_j)|x_{ij}}(x_{ij}'\delta_0(\tau)|x_{ij}) x_{ij}\tilde{x}_{ij}'\right]$.

(ii)

$$\sqrt{m}\left(M_{mn}(\delta_0, \beta_0, \tau)\right) \rightsquigarrow \mathbb{G}(\cdot), \text{ in } \ell^{\infty}(\mathcal{T} \times \mathcal{T}), \tag{23}$$

where $\mathbb{G}$ is a mean-zero Gaussian process with a uniformly continuous sample path and covariance function

$$\begin{aligned}\Omega_2(\tau, \tau') &= \mathbb{E}\left[[\tau_2 - 1(\tilde{x}_{ij}'\beta_{j,0}(\tau_1) \le x_{ij}'\delta_0(\beta_0, \tau)][\tau_2' - 1(\tilde{x}_{ij}'\beta_{j,0}(\tau_1') \le x_{ij}'\delta_0(\beta_0, \tau')] x_{ij} x_{ij}'\right] \\ &= (\min(\tau_2, \tau_2') - \tau_2\tau_2')\mathbb{E}[x_{ij} x_{ij}'].\end{aligned}$$

(iii) Under assumption 9(b):

$$\sup_{\tau \in \mathcal{T} \times \mathcal{T}} \left\| \mathrm{Cov}\left(M_{mn}(\delta_0, \beta_0, \tau), \frac{1}{m} \sum_{j=1}^{m} \bar{\Gamma}_{2,j}(\delta_0, \beta_0, \tau) \left(\hat{\beta}_j(\tau_1) - \beta_{j,0}(\tau_1) \right) \right) \right\| = o_p\left(\frac{1}{\sqrt{m}} \right).$$

Lemma 1 shows that the first-stage error converges to zero at a rate faster than the rate at which the standard deviation of the second stage converges as long as n diverges to infinity, no matter how slowly. In the special case where only one quantile is of interested, equation (23) implies $\sqrt{m}\,(M_{mn}(\delta_0, \beta_0, \tau)) \xrightarrow{d} N\left(0, \Omega_2(\tau)\right)$ with $\Omega_2(\tau) = \mathbb{E}\left[[\tau_2 - 1(\tilde{x}_{ij}'\beta_{j,0}(\tau_1) \le x_{ij}'\delta_0(\beta_0, \tau)]^2 x_{ij}x_{ij}' \right] = \tau_2(1-\tau_2)\mathbb{E}[x_{ij}x_{ij}']$.

The asymptotic distribution of the estimator can be approximated by a linear function of the sum of two terms converging at a different rate so that the asymptotic behavior will be determined by the term converging at a slower rate. Hence, in the first-order asymptotic distribution, only the second stage matters. Similar results are documented in Chetverikov et al. (2016) and in Melly and Pons (2024).

To show results for the entire quantile regression process, I first show that the linearization holds uniformly in $\tau_1, \tau_2 \in \mathcal{T} \times \mathcal{T}$. Then, I show that uniformly over τ_1, τ_2 the first stage error is $o_p(m^{-1/2})$ so that the estimator $\hat{\delta}(\hat{\beta}, \tau)$ has the same asymptotic distribution as the unfeasible estimator $\hat{\delta}(\beta_0, \tau)$. Hence, weak converges of $\sqrt{m}\left(\hat{\delta}(\beta_0, \tau) - \delta_0(\beta_0, \tau) \right)$ is sufficient to show weak converges of $\sqrt{m}\left(\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau) \right)$.

Theorem 2 (**Weak Convergence**)**.** *Let assumptions 1-9 be satisfied. Then*

$$\sqrt{m}\left(\hat{\delta}(\hat{\beta}, \cdot) - \delta_0(\beta_0, \cdot) \right) \rightsquigarrow \Gamma_1^{-1}(\cdot)\mathbb{G}(\cdot) \quad \textit{in } \ell^\infty(\mathcal{T} \times \mathcal{T}),$$

$\Gamma_1 = \Gamma_1(\delta_0, \beta_0, \cdot)$*, where* $\mathbb{G}(\cdot)$ *is a mean-zero Gaussian process with uniformly continuous sample paths and covariance function*

$$\begin{aligned} \Omega_2(\tau, \tau') &= \mathbb{E}\left[[\tau_2 - 1(\tilde{x}_{ij}'\beta_{j,0}(\tau_1) \le x_{ij}'\delta_0(\beta_0, \tau)][\tau_2' - 1(\tilde{x}_{ij}'\beta_{j,0}(\tau_1') \le x_{ij}'\delta_0(\beta_0, \tau')]x_{ij}x_{ij}' \right] \\ &= (\min(\tau_2, \tau_2') - \tau_2\tau_2')\mathbb{E}[x_{ij}x_{ij}']. \end{aligned}$$

As before, if we fix τ, the result of the previous theorem implies that

$$\sqrt{m}\left(\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau) \right) \xrightarrow{d} N(0, \Gamma_1^{-1}\Omega_2(\tau)\Gamma_1'^{-1}). \tag{24}$$

Theorem 2 shows that the entire coefficient vector converges at the $1/\sqrt{m}$ rate despite mn observations being used for the estimation. This is neither specific nor surprising to this quantile regression method. Instead, it is a consequence of modeling heterogeneities between groups. If I imposed equality of $\beta(\tau_1, \tau_2)$ over groups, it would be possible to estimate this coefficient at the $1/\sqrt{mn}$ rate. However, since $\beta(\tau_1, \tau_2)$ is allowed to vary over groups through the dependency on τ_2, variation between groups is necessary for identification. Similarly, in the least squares case, it is always possible to estimate the coefficient on the individual level variable at the

$1/\sqrt{mn}$ rate, for instance, by implementing a fixed effects estimator. However, this estimator only exploits the within-group variation and, therefore, cannot identify heterogeneities between groups. Ultimately, the between variation, which slows down the convergence rate, has to be used to identify between-group heterogeneity.

The results in Theorem 2 relies on the assumption that the number of observations per group grows faster than the square root of the number of groups. Until recently, there has been a substantial gap in the required rate of growth of n relative to m of nonlinear estimators with a smooth objective function and those with a non-smooth one, such as quantile regression. For unbiased asymptotic normality, the bias has to decrease more quickly than $1/\sqrt{m}$. Using new results in Volgushev et al. (2019) and Galvao et al. (2020), I show that $m(\log n)/\sqrt{n} \to 0$ is enough to ensure unbiased asymptotic normality of the estimator.[26] Closely related, it is worth mentioning that in empirical applications, groups might be defined by industries and geographical regions, such as counties, which tend to contain many individuals so that the results in Theorem 2 provide a useful approximation. Similarly, one might ask whether it would be possible to relax the required growth rate by using smoothed quantile regression and/or bias correction after the first stage. Given the results in the literature, smoothing does not help relaxing the assumptions for the results. It could help improve the finite sample performance at the cost of choosing a smoothing parameter. At the same time, the bias-corrected smoothed quantile regression estimator for panel data suggested in Galvao and Kato (2016) is not applicable in this setting, as it assumes homogeneity of the coefficients over groups. In ongoing work, Franguridi et al. (2024) derive an explicit formula for the bias of the leading term of the expansion. However, making the bias correction feasible remains a major challenge since it requires estimating some components involving higher-order derivatives. Furthermore, choosing tuning parameters is even more difficult. Exploring these possibilities is left for future research.

Remark 3 (Degree of heterogeneity and growth condition)**.** Assumption 5 implies that there is heterogeneity between groups. Without group-level heterogeneity, the second step quantile regression using the true first stage would be deterministic (degenerate). This suggests a faster rate of convergence with an asymptotic distribution where all the variance would come from the first stage. With a linear second step regression, Melly and Pons (2024) derive the adaptive asymptotic distribution of their estimator under the stronger growth condition that $m(\log n)^2/n \to 0$. This growth condition would be sufficient to adapt the proof of Lemma 1 to ensure that $\sqrt{mn}\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2,j}(\delta_0,\beta_0,\tau)\left(\hat{\beta}_j(\tau_1)-\beta_{j,0}(\tau_1)\right) \xrightarrow{d} N(0,\Omega_1(\tau))$ with $\Omega_1(\tau) = \mathbb{E}_j\left[\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)V_j(\tau_1)\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)'\right]$, where $V_j(\tau_1)$ is the asymptotic covariance matrix of $\hat{\beta}_j(\tau_1)$. It is likely that the inference procedure suggested here is valid adaptively. For instance, simulations without group-level heterogeneity suggest that the confidence bands have the

[26] Galvao et al. (2020) require the stronger condition $m(\log(n))^2/n \to 0$ because, in their case, the bias converges more quickly than $1/\sqrt{mn}$ to establish unbiased asymptotic normality of their estimator.

correct coverage even without group-level heterogeneity. In this case, however, it is not possible to use the same proof strategy because quantile regression is nonlinear, and the proof relies on a linearization which holds only under heterogeneity. Intuitively, I need strong separation of the groups since without heterogeneity between groups, the estimated group-level conditional quantile functions are identical up to the first stage error, and the estimator should converge at the faster $\sqrt{mn}$ rate.

5.2 Inference

To perform inference, I suggest a clustered bootstrap procedure, where entire groups are resampled with replacement. In a similar setting, Liao and Yang (2018); Lu and Su (2022) and Fernández-Val et al. (2022) show that the procedure is uniformly valid in the rate of convergence of the estimator,[27] and Melly and Pons (2024) show adaptive validity of the clustered variance estimator over the different degrees of between-group heterogeneity. Based on simulations, however, the bootstrap method appears to outperform the clustered covariance matrix estimator with this estimator. Hence, I suggest using the bootstrap in this paper. Since entire groups are resampled, the first stage is unaffected and does not need to be recomputed. Consequently, the procedure is equivalent to resampling the first-stage fitted values. The following algorithm describes how to compute inference using the clustered bootstrap. Since the randomness comes from which group are sampled, I denote $\hat{\beta}_j(z^*, \tau)$ the $(K_1 + 1)$-dimensional vector containing first-stage estimates of group j in the bootstrap world.

Algorithm 1 (Boostrap Variance Estimator)**.** *Draw a random sample with replacement* $\{(\hat{y}^*_{1j}(\tau_1), \ldots, \hat{y}^*_{nj}(\tau_1)), (x^*_{1j}, \ldots, x^*_{nj}) : j = 1, \ldots, m,\ \tau_1 \in \mathcal{T}\}$ *from* $\{(\hat{y}_{1j}(\tau_1), \ldots, \hat{y}_{nj}(\tau_1)), (x_{1j}, \ldots, x_{nj}) : j = 1, \ldots, m,\ \tau_1 \in \mathcal{T}\}$*, and run the second step estimator (Equation (15)) using the resampled data and obtain* $\hat{\delta}^*(\hat{\beta}(z^*), \tau)$*. Repeat the previous step for each bootstrap replication* $b = 1, \ldots, B$*, to obtain* $\{\hat{\delta}^*(\hat{\beta}(z^*), \tau)\}_{b=1}^B$ *for each* τ*. Compute a bootstrap estimate of* $\Gamma_1^{-1}\Omega_2(\tau)\Gamma_1^{-1}$ *such as the variance of the bootstrap estimates or the interquartile range rescaled with the normal distribution* $\hat{\Gamma}_1^{*-1}\hat{\Omega}_2^*(\tau)\hat{\Gamma}_1^{*-1} = (q_{0.75}(\tau) - q_{0.25}(\tau))/(z_{0.75} - z_{0.25})$ *for* $\tau \in \mathcal{T} \times \mathcal{T}$*, where* $q_p(\tau)$ *is the pth percentile of the bootstrap estimates.*[28]

I show that the asymptotic distribution of $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)$ can be approximated with the distribution of $\hat{\delta}^*(\hat{\beta}(z^*), \tau) - \hat{\delta}(\hat{\beta}, \tau)$.

Theorem 3 (Validity of the Bootstrap). *Assume that the condition for Theorem 5 are satisfied. Then,*

$$\sqrt{m}\left(\hat{\delta}^*(\hat{\beta}(z^*), \cdot) - \hat{\delta}(\hat{\beta}, \cdot)\right) \rightsquigarrow^* \Gamma_1^{-1}(\cdot)\mathbb{G}(\cdot) \quad \text{in } \ell^\infty(\mathcal{T} \times \mathcal{T}).$$

[27] These papers refer to this bootstrap procedure as cross-sectional bootstrap since the focus is on traditional panel data models.

[28] The interquartile range rescaled with the normal distribution is used, for example, in Chernozhukov et al. (2013).

Given the large number of coefficients estimated in the model, researchers might be interested in testing hypotheses involving many parameters. For instance, one might want to test the hypothesis that the effect of the kth regressor is constant at all quantiles:

$$H_0 : \delta_k(\tau) = \bar{\delta}_k, \; \forall \; \tau_1, \tau_2 \in \mathcal{T} \times \mathcal{T}.$$

Kolmogorov-Smirnov and Cramér-von Mises type tests are suitable in these settings. For instance, the Kolmogorov-Smirnov test statistics can be constructed by:

$$t^{KS} = \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \sqrt{\left(\hat{\delta}_k(\tau) - \bar{\delta}_k\right)' \hat{V}_k(\tau) \left(\hat{\delta}_k(\tau) - \bar{\delta}_k\right)}, \tag{25}$$

with $\bar{\delta}_k = \int_{\tau_2} \int_{\tau_1} \hat{\delta}(\tau_1, \tau_2) d\tau_1 d\tau_2$ and where $\hat{V}_k(\tau)$ is a bootstrap estimate of the asymptotic variance of $\hat{\delta}_k(\tau)$.

To obtain the critical values, I follow Chernozhukov and Fernández-Val (2005) and use the bootstrap to mimic the test statistic. To impose the null, I use the parametric bootstrap based on the estimated quantile regression process.[29] Then, for each bootstrap iteration, construct the test statistics:

$$t_b^{KS} = \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \sqrt{\left(\hat{\delta}_k^{*b}(\tau) - \hat{\delta}_k^{*b}\right)' \hat{V}_k(\tau) \left(\hat{\delta}_k^{*b}(\tau) - \hat{\delta}_k^{*b}\right)}, \tag{26}$$

where $\hat{\delta}_k^{*b} = \int_{\tau_2} \int_{\tau_1} \hat{\delta}^{*b}(\tau_1, \tau_2) d\tau_1 d\tau_2$. The critical values of a test with size α are the $1 - \alpha$ quantile of $\{t_b^{KS} : 1 \leq b \leq B\}$.

Following Chernozhukov et al. (2013), it is possible to construct functional confidence intervals that cover the entire function with a pre-specified rate by inverting the acceptance region of the KS statistics

$$t_b^{KS} = \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \sqrt{\left(\hat{\delta}_k^{*b}(\tau) - \hat{\delta}_k(\tau)\right)' \hat{V}_k(\tau) \left(\hat{\delta}_k^{*b}(\tau) - \hat{\delta}_k(\tau)\right)}. \tag{27}$$

The $(1 - \alpha)$ functional confidence bands for a coefficient $\hat{\delta}_k(\tau)$ can be constructed by

$$\hat{\delta}_k(\tau) \pm \hat{t}^*_{1-\alpha} \cdot \sqrt{\hat{V}_k(\tau)},$$

where $\hat{t}^*_{1-\alpha}$ is the $1 - \alpha$ quantile of $\{t_b^{KS} : 1 \leq b \leq B\}$. For more information, see Chernozhukov et al. (2013).

6 Simulations

To analyze the small sample performance of the estimator, I perform a Monte Carlo simulation with the following data generating process:

$$y_{ij} = 1 + \beta \cdot x_{1ij} + \gamma \cdot x_{2j} + \eta_j(1 - 0.1 \cdot x_{1ij} - 0.1 \cdot x_{2j}) + \nu_{ij}(1 + 0.1 \cdot x_{1ij} + 0.1 \cdot x_{2j})$$

[29] Compared to the nonparametric bootstrap, the parametric one showed better performance in simulations as the latter was conservative.

Table 2: Bias and Standard Deviation

$\tau_1 \setminus \tau_2$	β 0.25	β 0.5	β 0.75	γ 0.25	γ 0.5	γ 0.75
			(m, n) = (25,25)			
0.25	-0.023	0.004	0.034	-0.030	-0.006	0.018
	(0.119)	(0.110)	(0.117)	(0.243)	(0.222)	(0.239)
0.5	-0.021	-0.001	0.027	-0.029	-0.010	0.014
	(0.114)	(0.106)	(0.111)	(0.240)	(0.219)	(0.235)
0.75	-0.029	-0.005	0.024	-0.031	-0.012	0.014
	(0.114)	(0.112)	(0.119)	(0.246)	(0.222)	(0.236)
			(m, n) = (25,200)			
0.25	-0.010	0.000	0.007	-0.004	0.006	0.019
	(0.071)	(0.067)	(0.072)	(0.237)	(0.215)	(0.232)
0.5	-0.010	-0.002	0.005	-0.004	0.004	0.018
	(0.067)	(0.066)	(0.070)	(0.237)	(0.215)	(0.235)
0.75	-0.010	-0.004	0.006	-0.007	0.004	0.017
	(0.070)	(0.069)	(0.072)	(0.237)	(0.217)	(0.238)
			(m, n) = (200,25)			
0.25	-0.023	0.004	0.030	-0.018	0.003	0.022
	(0.043)	(0.040)	(0.042)	(0.082)	(0.072)	(0.078)
0.5	-0.024	-0.001	0.023	-0.018	0.001	0.019
	(0.041)	(0.037)	(0.040)	(0.078)	(0.072)	(0.077)
0.75	-0.032	-0.007	0.020	-0.020	-0.002	0.018
	(0.043)	(0.038)	(0.042)	(0.079)	(0.072)	(0.078)
			(m, n) = (200,200)			
0.25	-0.005	0.001	0.006	-0.004	0.001	0.003
	(0.028)	(0.026)	(0.028)	(0.076)	(0.073)	(0.079)
0.5	-0.005	0.000	0.006	-0.004	0.000	0.003
	(0.028)	(0.025)	(0.028)	(0.076)	(0.073)	(0.079)
0.75	-0.006	0.000	0.006	-0.005	0.001	0.002
	(0.028)	(0.026)	(0.028)	(0.077)	(0.073)	(0.079)
			(m, n) = (200,400)			
0.25	-0.003	0.000	0.003	-0.004	-0.003	0.002
	(0.026)	(0.023)	(0.026)	(0.077)	(0.073)	(0.079)
0.5	-0.003	0.000	0.003	-0.004	-0.003	0.002
	(0.025)	(0.023)	(0.025)	(0.077)	(0.073)	(0.079)
0.75	-0.004	-0.001	0.003	-0.005	-0.004	0.002
	(0.026)	(0.024)	(0.026)	(0.077)	(0.073)	(0.079)

Notes: Results based on 2,000 Monte Carlo simulations. The table provides bias and standard deviation.

with $x_{1ij} = 1+h_j+w_{ij}$, where $h_j \sim U[0,1]$ and $w_{ij}, x_{2j}, \eta_j, \nu_{ij}$ are $N(0,1)$. This is a location-scale-shift model over both quantile indices. I set $\beta = \gamma = 1$. The true coefficients on the individual-level variable are $\beta(\tau_1, \tau_2) = 1+0.1 \cdot F^{-1}(\tau_1)+0.1 \cdot F^{-1}(\tau_2)$ and the true coefficients on the group-level regressor equal $\gamma(\tau_1, \tau_2) = 1-0.1 \cdot F^{-1}(\tau_1)-0.1 \cdot F^{-1}(\tau_2)$, where F is the standard normal cdf. I consider the sample sizes $(m, n) = \{(25, 25), (200, 25), (25, 200), (200, 200), (200, 400)\}$ and focus on the set of quantiles $\{0.25, 0.5, 0.75\}$ using 2'000 Monte Carlo simulations.

Table 2 shows the bias and standard deviation. Table 3 shows the bootstrap standard errors relative to the standard deviation, and Table 4 shows the coverage probability of the 95% confidence intervals. Bootstrap standard errors are computed using 200 repetitions.

Table 3: Standard Errors Relative to Standard Deviation

$\tau_1 \setminus \tau_2$	β			γ		
	0.25	0.5	0.75	0.25	0.5	0.75
	(m, n) = (25,25)					
0.25	1.180	1.146	1.200	1.148	1.098	1.261
0.5	1.191	1.133	1.242	1.190	1.115	1.327
0.75	1.213	1.119	1.230	1.167	1.107	1.357
	(m, n) = (25,200)					
0.25	1.321	1.231	1.401	1.275	1.138	1.649
0.5	1.381	1.229	1.457	1.332	1.138	1.720
0.75	1.358	1.199	1.443	1.352	1.126	1.724
	(m, n) = (200,25)					
0.25	1.028	1.031	1.048	1.002	1.056	1.043
0.5	1.017	1.052	1.069	1.028	1.052	1.063
0.75	1.025	1.063	1.050	1.027	1.053	1.052
	(m, n) = (200,200)					
0.25	1.089	1.081	1.080	1.056	0.995	1.021
0.5	1.064	1.081	1.081	1.052	1.000	1.014
0.75	1.075	1.082	1.095	1.036	1.004	1.018
	(m, n) = (200,400)					
0.25	1.081	1.111	1.078	1.044	1.003	1.011
0.5	1.089	1.092	1.088	1.039	1.004	1.009
0.75	1.092	1.092	1.082	1.037	1.005	1.008

Notes: Results based on 2,000 Monte Carlo simulations. The table provides standard errors relative to standard deviation.

While $\hat{\beta}$ and $\hat{\gamma}$ have the same asymptotic behavior, we see differences in their finite sample properties. The simulations show that the bias of $\hat{\beta}$ decreases both as n or m increases, while the bias of $\hat{\gamma}$ decreases only when m increases. Similarly, the variance of $\hat{\gamma}$ is only minimally affected by an increase in the number of observations per group. On the other hand, the variance of $\hat{\beta}$ shows a larger improvement when n increases. Still, an increase in the number of groups yields the largest improvement in the variance. As n becomes larger, further increases in the number of observations per group will not improve the variance of $\hat{\beta}$. More precisely, the relatively large decrease in the bias and variance of $\hat{\beta}$ quickly converges to zero as n increases, and it is mostly driven by the decline in the first-stage error and increasing n from 200 to 400 only minimally affects the results.

Table 3 shows the bootstrap standard errors relative to the standard deviation. The simulation suggests that the bootstrap standard errors are conservative. The standard errors are particularly large when the number of groups m is small, and the ratio converges to 1 as m increases. The coverage probabilities of the 95% confidence bands in Table 9 are close to 95%. There are some minor discrepancies which, however, disappear as the number of groups and observations per group increase. In some instances with $(m, n) = (200, 25)$, the confidence interval tends to undercover, but this is likely driven by the large bias arising in the first stage (see Table 2). In Appendix C, I include simulation results for the clustered standard errors and

Table 4: Coverage Probability

$\tau_1 \setminus \tau_2$	β 0.25	β 0.5	β 0.75	γ 0.25	γ 0.5	γ 0.75
			(m, n) = (25,25)			
0.25	0.970	0.973	0.969	0.948	0.954	0.953
0.5	0.972	0.973	0.970	0.949	0.951	0.948
0.75	0.971	0.968	0.972	0.949	0.958	0.946
			(m, n) = (25,200)			
0.25	0.985	0.987	0.985	0.957	0.959	0.965
0.5	0.986	0.985	0.981	0.956	0.956	0.964
0.75	0.988	0.988	0.987	0.955	0.953	0.954
			(m, n) = (200,25)			
0.25	0.916	0.948	0.899	0.929	0.943	0.928
0.5	0.905	0.955	0.925	0.936	0.954	0.932
0.75	0.878	0.952	0.931	0.941	0.959	0.943
			(m, n) = (200,200)			
0.25	0.964	0.965	0.954	0.948	0.938	0.940
0.5	0.955	0.961	0.956	0.945	0.940	0.944
0.75	0.961	0.963	0.961	0.947	0.942	0.947
			(m, n) = (200,400)			
0.25	0.957	0.958	0.961	0.948	0.936	0.939
0.5	0.963	0.961	0.961	0.946	0.938	0.934
0.75	0.959	0.963	0.959	0.944	0.940	0.931

Notes: Results based on 2,000 Monte Carlo simulations. The table provides the coverage probability of the 95% confidence intervals.

the corresponding coverage probabilities for completeness. The clustered standard errors are smaller than the bootstrap standard errors, and the confidence intervals have lower coverage, mostly in cases with fewer groups.

Table 5 shows the rejection probabilities of 5% Kolmogorov-Smirnov and Cramér-von Mises tests for the null hypotheses that $\beta(\tau) = \bar{\beta}$ and that $\gamma(\tau) = \bar{\gamma}$.[30] For this set of simulations, I consider a variation of the data-generating process above. The data is generated by:

$$y_{ij} = 1 + x_{1ij} + x_{2j} + \eta_j(1 - \psi(x_{1ij} + x_{2j})) + \nu_{ij}(1 + \phi(x_{1ij} + x_{2j})),$$

where all variables' distribution are unchanged. The parameter ϕ regulate effect heterogeneity over τ_1 and ψ determines heterogeneity over τ_2. I consider five cases of this data-generating process and test for effect heterogeneity of both x_{1ij} and x_{2j}. The rejection probabilities are computed using 400 simulations, and each of these simulations includes 100 bootstrap replications. I conduct the simulations on the set of quantiles $0.1, 0.2, \ldots, 0.9$. In the first column, where the null hypotheses are true, we should observe a rejection probability of 5% (empirical size). In the other cases, the rejection probability indicates the power of the test. In smaller samples, the empirical size tends to be conservative. Additionally, when heterogeneity is limited and

[30] The Cramér-von Mises test statistics is $t^{CvM} = \int_{\mathcal{T}} \int_{\mathcal{T}} \left(\hat{\delta}_k(\tau) - \bar{\delta}_k\right)' \hat{V}_k(\tau) \left(\hat{\delta}_k(\tau) - \bar{\delta}_k\right) d\tau_1 d\tau_2$, and the critical value are computed by bootsrapping the test statistic as with the Kolmogorov-Smirnov test.

Table 5: Rejection Probability of the KS and CvM Tests

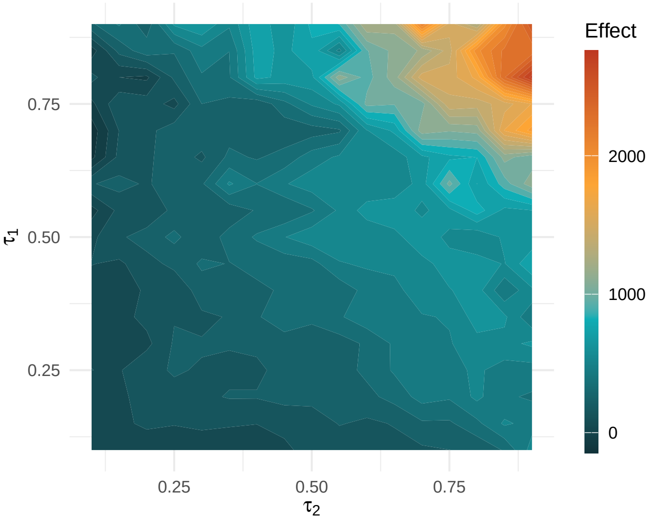
(ϕ, ψ)	(0, 0)	(0, 0.1)	(0.1, 0)	(0.1, 0.1)	(0.2, 0.2)
Panel (a): Kolmogorov-Smirnov Test					
$H_0 : \gamma(\tau) = \bar{\gamma}$					
(m, n) = (25,25)	0.007	0.005	0.007	0.009	0.034
(m, n) = (25,200)	0.015	0.013	0.020	0.032	0.173
(m, n) = (200,25)	0.026	0.209	0.251	0.469	0.996
(m, n) = (200,200)	0.046	0.307	0.397	0.826	1.000
$H_0 : \beta(\tau) = \bar{\beta}$					
(m, n) = (25,25)	0.026	0.108	0.101	0.156	0.537
(m, n) = (25,200)	0.056	0.536	0.548	0.885	1.000
(m, n) = (200,25)	0.026	0.767	0.822	0.970	1.000
(m, n) = (200,200)	0.057	1.000	1.000	1.000	1.000
Panel (b): Cramér-von Mises Test					
$H_0 : \gamma(\tau) = \bar{\gamma}$					
(m, n) = (25,25)	0.014	0.026	0.022	0.027	0.165
(m, n) = (25,200)	0.023	0.030	0.035	0.047	0.381
(m, n) = (200,25)	0.044	0.381	0.414	0.789	1.000
(m, n) = (200,200)	0.061	0.446	0.430	0.895	1.000
$H_0 : \beta(\tau) = \bar{\beta}$					
(m, n) = (25,25)	0.038	0.223	0.231	0.373	0.921
(m, n) = (25,200)	0.068	0.728	0.844	0.988	1.000
(m, n) = (200,25)	0.048	0.937	0.995	1.000	1.000
(m, n) = (200,200)	0.056	1.000	1.000	1.000	1.000

Notes: The table shows the empirical size of a 5% test testing the null hypothesis of effects homogeneity over both dimensions. The test is performed using the Kolmogorov-Smirnov and the Cramér von Mises tests statistics. The results are based on 1,000 Monte Carlo simulations using 100 bootstrap replications.

the sample size is small, the test seems to have difficulties detecting the heterogeneity. However, columns (2)-(5) demonstrate that as the degree of heterogeneity or the sample size increases, the test becomes more powerful. There are also notable differences in the performance of the test for β compared to γ. When the number of groups is small, the test on γ shows significantly lower power. Nonetheless, this discrepancy diminishes as m increases, leading to a substantial improvement in the power of the test. In contrast, heterogeneity in β appears easier to detect in small samples, with the power of this test increasing as the number of observations per group grows. In comparison, for γ, increases in n yield only marginal improvements in power.

Comparing the two tests, we see that in these simulations, the Cramér-von Mises test appears to be more powerful than the Kolmogorov-Smirnov test. However, it is important to note that different data-generating processes may lead to different results, as each test is better at detecting certain types of deviations. Consequently, there is no uniformly more powerful test, as one may outperform the other depending on the nature of the deviation.

Figure 3: Effect of Training Assignment on Income from Work



Notes: The figure shows the effect of the treatment on the within market τ_1 and the between market distribution τ_2 of income from work.

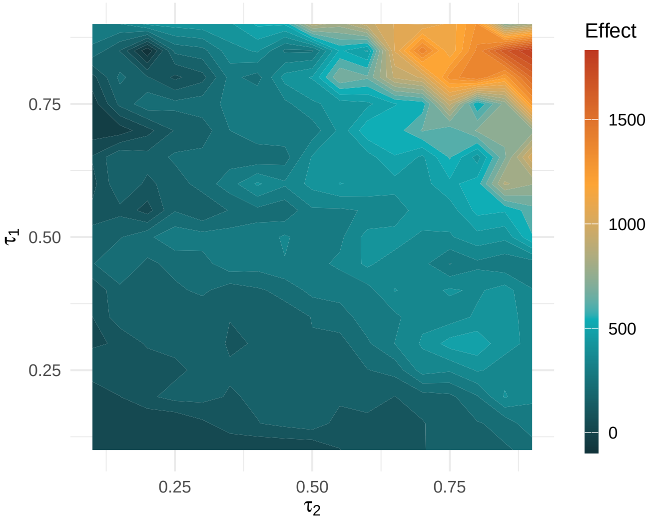
7 Empirical Application: Distributional Impacts of Business Training within and between Markets

In an empirical application, I complement the findings of McKenzie and Puerto (2021) by offering new insights into distributional effects. Their study aims to analyze the impact of business training on the outcomes of female-owned businesses and the spillover effects of such training. The sample comprises 3,537 female-owned businesses operating in 157 different rural markets in Kenya. The training program is randomly assigned to firms through a two-stage randomization process. The first stage involves market-level randomization, where 93 markets are treatment markets, and the remaining 64 serve as control markets. In the second stage, individual-level randomization assigns businesses in the treatment markets to training or control. Randomization is stratified by geographical region, market size, and quartiles of weekly profits to ensure a balanced sample. This results in 1,172 individual firms assigned to training and 2,365 firms assigned to the control group.

The training program spans five days and covers topics such as bookkeeping, recordkeeping, marketing, financial concepts, costing and pricing, and the development of new business ideas. Moreover, it specifically addresses challenges faced by women in business. For more detailed information about the program's structure or the experimental setting, refer to McKenzie and Puerto (2021) and their appendix.

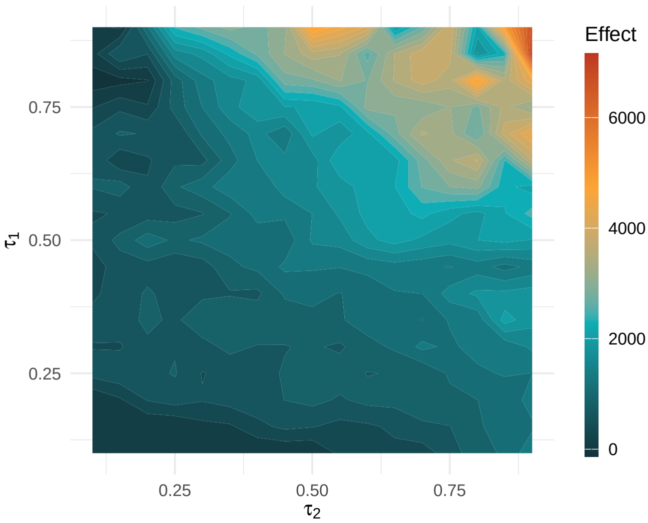
McKenzie and Puerto (2021) find a positive effect of training on the business survival after three years. Further, the training increases average weekly sales and profits by 18 and 15 percent, respectively, and firm owners assigned to the training report better mental health and a higher subjective standard of living. However, the spillover effects on businesses in treatment

Figure 4: Effect of Training Assignment on Weekly Profits



Notes: The figure shows the effect of the treatment on the within market τ_1 and the between market distribution τ_2 of weekly profits.

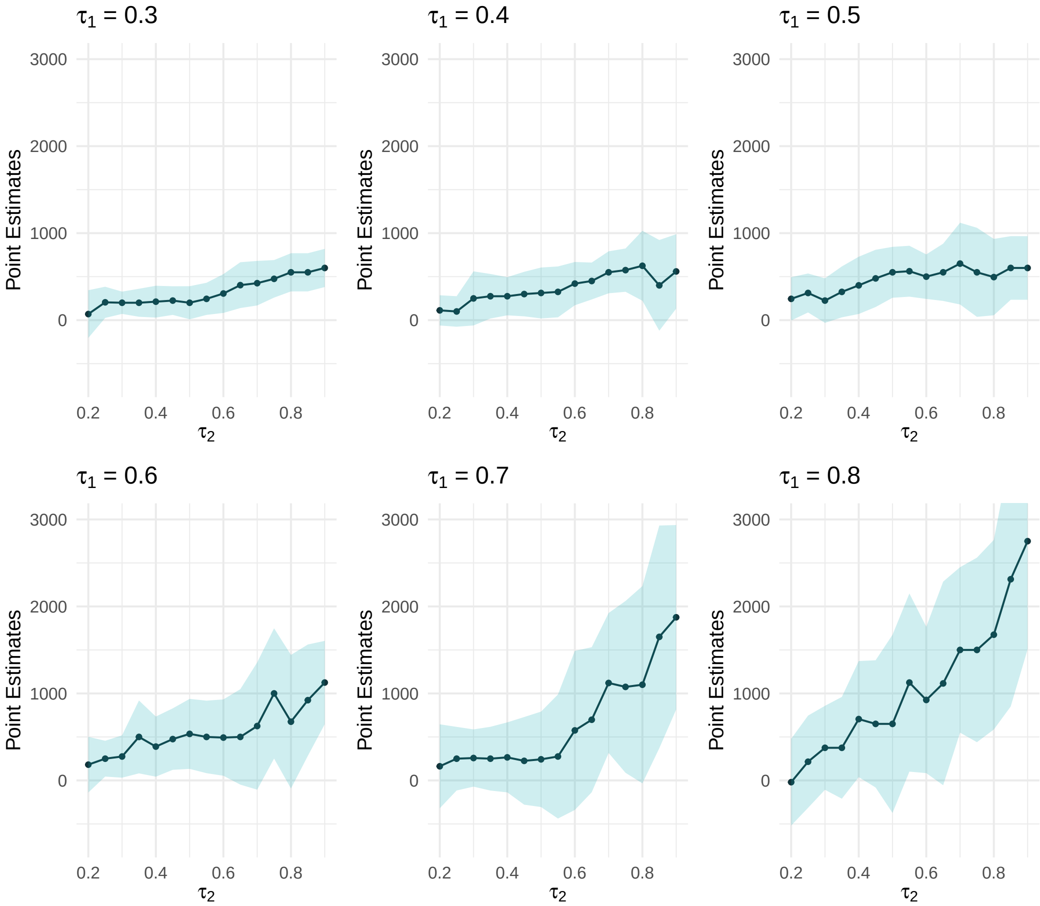
Figure 5: Effect of Training Assignment on Weekly Sales



Notes: The figure shows the effect of the treatment on the within market τ_1 and the between market distribution τ_2 of weekly sales.

markets not assigned to the program remain unclear, with point estimates being small and not statistically significant. Further, in the original paper, they estimate the distributional effects of training on profits and sales. This analysis uses data collected in two waves three years after the training program, and they document larger effects in the upper tail of the outcome distribution. In my analysis, I use data from the same two waves and compute the mean of the variables over the two waves, if a business is observed in both waves. Further, I define groups based on markets. To ensure that I have enough observations for the estimation, I drop markets

Figure 6: Effect of Training Assignment on Income From Work



Notes: The figure shows the effect of the treatment on the between market distribution τ_2 of income from work for selected quantiles of the within market distribution τ_1. The shaded areas show the 95% confidence intervals estimated using clustered bootstrap standard errors computed with 1000 replications.

with fewer than 15 businesses.[31] The final dataset includes 2,922 firms operating in 116 markets. On average, there are 27 observations per market. I consider three different outcome variables: weekly sales, weekly profits, and income from work with averages of 6,000, 1,500, and 2,300 Kenyan shillings.[32]

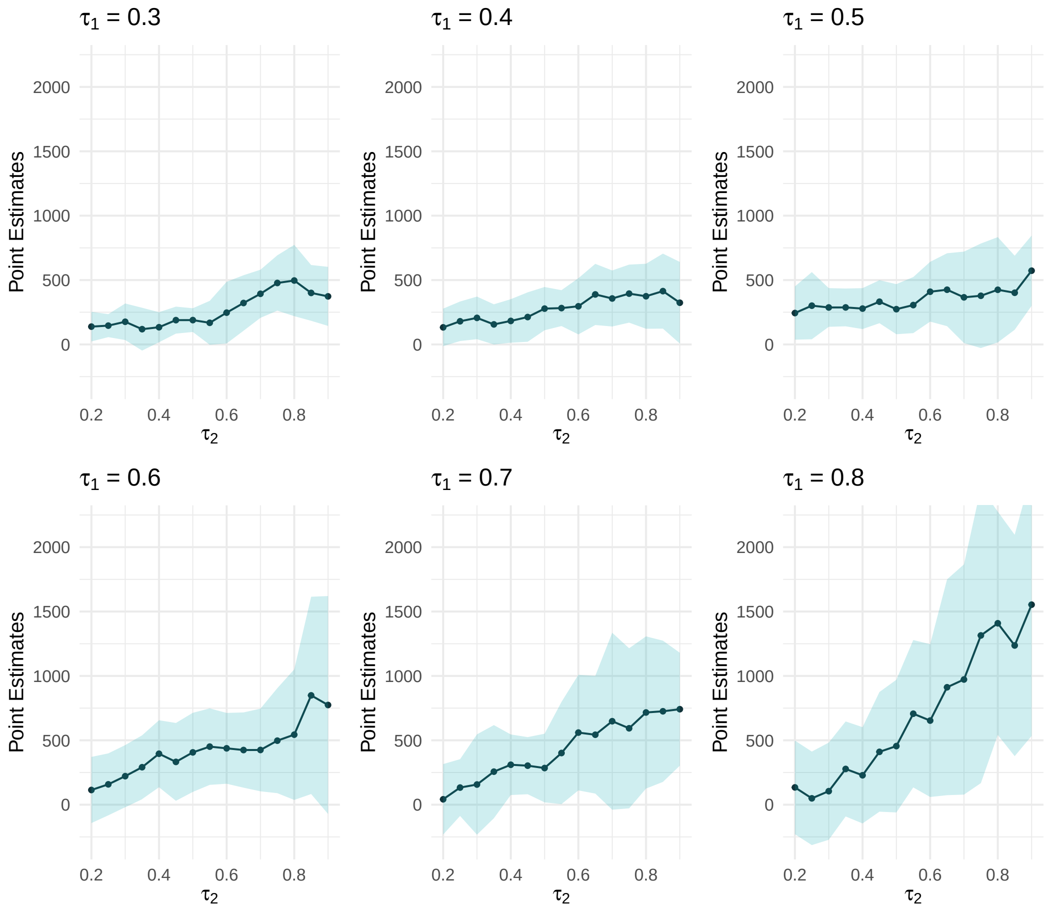
All three variables have mass points at zero. One reason is that these variables are coded to zero for firms that exit the market. More precisely, around 11% of the firms in the final dataset did not survive after three years, and profits and sales are zero in around 13% of the observations. Further, 10% of people report having no income. Due to this censoring issue, I refrain from computing the effects too far in the lower tail. Since these mass points could invalidate inference, I do not report confidence intervals for quantiles affected by the problem.[33]

[31] The results remain similar when using a different cutoff.

[32] In April 2023, 1,000 Kenyan Shillings are around 7.5 USD.

[33] If the second stage fitted values for at least one observation equals zero, I will consider the cell affected by the mass point. Fitted values of zero suggest a perfect fit, at least for some observations.

Figure 7: Effect of Training Assignment on Profits



Notes: The figure shows the effect of the treatment on the between market distribution τ_2 of weekly profits for selected quantiles of the within market distribution τ_1. The shaded areas show the 95% confidence intervals estimated using clustered bootstrap standard errors computed with 1000 replications.

Further away from the lower tail, this problem does neither affect the results nor invalidate inference.

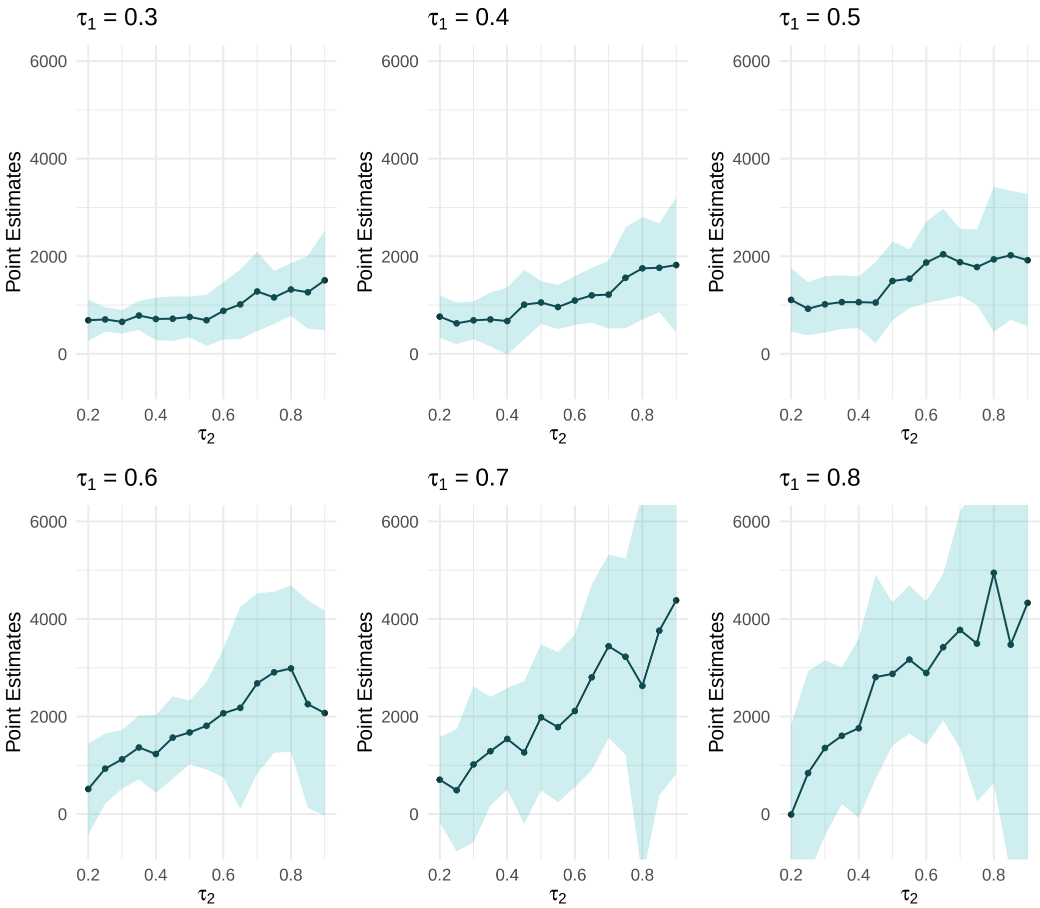
I estimate the following model:

$$Q(\tau_2, Q(\tau_1, y_{ij}|d_{ij}, s_{ij}, v_j)|d_{ij}, s_{ij}) = \beta_1(\tau) \cdot d_{ij} + \beta_2(\tau) \cdot s_{ij} + \alpha(\tau), \tag{28}$$

where y_{ij} is the outcome of firm i operating in market j, d_{ij} is the treatment dummy, and s_{ij} is a binary variable that accounts for potential spillover effects and takes value 1 for firms in the treatment markets that are assigned to the control group.

Figures 3-5 show the treatment effects estimates for sales, profits, and income from work over the two dimensions for the quantiles indices $\{0.2, 0.3, \ldots, 0.9\}$. Figures 6-8 plot the point estimates and confidence intervals over the distribution of markets τ_2 when fixing the within-market quantile τ_1. The results for all three variables show a similar pattern, where both within-group and between-group heterogeneities play essential roles, resulting in larger positive treatment effects in the upper tail of their respective distributions. For instance, at $\tau_1 = 0.5$,

Figure 8: Effect of Training Assignment on Sales



Notes: The figure shows the effect of the treatment on the between market distribution τ_2 of weekly sales for selected quantiles of the within market distribution τ_1. The shaded areas show the 95% confidence intervals estimated using clustered bootstrap standard errors computed with 1000 replications.

Table 6: P-Values of Cramér-von Mises and Kolmogorov-Smirnov Tests

	Income	Profits	Sales
Cramér-von Mises	0.024	0.027	0.024
Kolmogorov-Smirnov	0.006	0.009	0.012

Notes: The table shows the p-values of the Cramér-von Mises and Kolmogorov-Smirnov tests for the null hypothesis that the coefficients are homogeneous over both dimension. The test is performed with the parametric bootstrap with 1000 replications.

the effect on profits increases from 100 Kenyan Shillings in the lower tail of the distribution to over 600 in the upper tail. Simultaneously, the within-market rank plays a major role, even for firms operating in the most prosperous markets, where the effect on profit goes from 400 Kenyan Shillings at $\tau_1 = 0.3$ to 800 at $\tau_1 = 0.7$. The larger effect in the upper tail of both distributions could indicate the presence of complementarities between individual ability and market quality,

Table 7: Correlation of Ranks over τ_1 for Income from Work

	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.2	1							
0.3	0.74	1						
0.4	0.65	0.87	1					
0.5	0.53	0.76	0.85	1				
0.6	0.49	0.66	0.72	0.82	1			
0.7	0.42	0.6	0.66	0.69	0.83	1		
0.8	0.36	0.51	0.58	0.62	0.77	0.88	1	
0.9	0.32	0.44	0.42	0.47	0.59	0.6	0.69	1

Notes: The table shows the correlation matrix of the ranks at different values of τ_1.

suggesting that both are necessary to benefit from the treatment.

I perform a Kolmogorov-Smirnov and Carmér-von-Mises tests to test for treatment effect heterogeneity over (τ_1, τ_2). Table 6 shows the p-value of the test statistic. As shown, for all outcomes, we reject the null of treatment effect homogeneity at the 5% level, hence providing evidence of treatment effect heterogeneity.

Finally, to assess the extent to which groups in the upper tail of the between distribution at low values of τ_1 are also in the upper tail for high values of τ_1, Table 7 presents the correlation matrix of group ranks across the τ_1 dimension for the variable income from work. The table provides the correlation of group ranks at any two points of the within distribution. If heterogeneity between groups is solely due to a location shift, the ranks do not change over τ_1, and we would observe a correlation of 1 for any two values of τ_1. A lower correlation suggests that groups rank differently at different points of the within distribution, indicating that there is no unique notion of poor-performing or good-performing markets across τ_1. The correlation matrix shows that the ranks vary across the τ_1, providing evidence that a univariate rank variable would miss important characteristics of the data. Yet, it is also noticeable that the rank at $\tau_1 = 0.5$ is correlated with the ranks at different points in the distribution, suggesting that some underlying mechanism might affect the entire group distribution.

8 Conclusion

Distributional effects are particularly interesting when analyzing treatment effect heterogeneity. In economics, heterogeneity manifests itself across various dimensions, encompassing within and between groups, where groups can be, among others, geographical regions, industries, or firms. I argue that both these dimensions are welfare relevant and might be interdependent in many settings. At the same time, papers looking at both dimensions are rare and often implicitly impose strong restrictions on the function of social welfare. This paper provides a method to analyze heterogeneity and distributional effects within and between groups simultaneously while at the same time remaining agnostic about the objective function of the policymaker. To this end, I introduce a quantile model with two quantile indices: one capturing heterogeneity within groups and the other addressing heterogeneity between groups. The conditional quantile function of each group models the within-group heterogeneity. Then, to aggregate the results over the distribution of groups, I model the conditional quantile function of these group-level quantile functions. I show that constructing the two-level quantile function involves a trade-off between a simple model with a unique group rank and a more flexible model that allows for unrestricted heterogeneity between groups and allows group ranks to change over the within distribution. This paper follows the second approach as this offers a more realistic model. I show that the method can be used for policy evaluation when the interest is on the effect on the distribution and the policymaker wants to consider a trade-off between different dimensions of inequality. Further, under the stronger condition of rank invariance, the model identifies individual effects and can, therefore, be used for optimal treatment assignment, where the optimal policy rule exploits multiple dimensions of treatment effect heterogeneity to assign the treatment.

I suggest estimating the model using a two-step quantile regression estimator with within-group regressions in the first stage and between-group regressions in the second stage. I show uniform consistency and weak convergence when the number of observations per group and the number of groups grows to infinity. I show that the estimator can provide new insights about inequalities in grouped data. In a descriptive illustration, I study income heterogeneity within and between the labor markets in Switzerland. The results show that a large portion of the group-level heterogeneities are driven by high top wages in a few regions, whereas for most of the within distribution, differences between regions are less marked. A result that conventional methods, such as variance decompositions or simple comparisons of mean or median wages, would fail to capture. Furthermore, the data show that group ranks change substantially over the within distribution, suggesting that differences in median wages between regions do not provide a meaningful picture of the labor market situation of low-income individuals.
Finally, in an empirical application, I extend the findings of McKenzie and Puerto (2021) by assessing the impact of business training on firm performance in Kenya, allowing for treatment effect heterogeneity within and between markets. I find large positive effects on sales, profits, and income from work with stronger effects for good-performing firms (in their markets) operating in thriving markets, indicating that there might be complementarities between individual and group ranks.

As pointed out in Remark 2, it would be possible to extend these results to include instrumental variables. Future researchers could combine the computational advancement with new estimation methods enabling the extension of these results to account for endogenous treatments and alternative identification strategies.

References

Adusumilli, K., F. Geiecke, and C. Schilter (2019): "Dynamically optimal treatment allocation using Reinforcement Learning," .

Akerman, A., E. Helpman, O. Itskhoki, M. A. Muendler, and S. Redding (2013): "Sources of wage inequality," *American Economic Review*, 103, 214–219.

Albanese, G., G. Barone, and G. de Blasio (2023): "The impact of place-based policies on interpersonal income inequality," *Economica*, 90, 508–530.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Antràs, P., A. de Gortari, and O. Itskhoki (2017): "Globalization, inequality and welfare," *Journal of International Economics*, 108, 387–412.

Arellano, M. and S. Bonhomme (2016): "Nonlinear panel data estimation via quantile regressions," *Econometrics Journal*, 19, C61–C94.

Autor, D. H., D. Dorn, and G. H. Hanson (2021): "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men," *American Economic Review: Insights*, 1, 161–178.

Autor, D. H., A. Manning, and C. L. Smith (2016): "The contribution of the minimum wage to US wage inequality over three decades: A reassessment," *American Economic Journal: Applied Economics*, 8, 58–99.

Becker, S. O., P. H. Egger, and M. von Ehrlich (2010): "Going NUTS: The effect of EU Structural Funds on regional performance," *Journal of Public Economics*, 94, 578–590.

Bhattacharya, J. (2020): "Quantile regression with generated dependent variable and covariates," .

Blackorby, C. and D. Donaldson (1978): "Measures of relative equality and their meaning in terms of social welfare," *Journal of Economic Theory*, 18, 59–80.

Bourguignon, F. and C. Morrisson (2002): "Inequality among world citizens: 1820-1992," *American Economic Review*, 92, 727–744.

Breza, E., S. Kaur, and Y. Shamdasani (2018): "The morale effects of pay inequality," *Quarterly Journal of Economics*, 133, 611–663.

Busso, M., J. Gregory, and P. Kline (2013): "Assessing the incidence and efficiency of a prominent place based policy," *American Economic Review*, 103, 897–947.

Card, D., A. Mas, E. Moretti, and E. Saez (2012): "Inequality at work: The effect of peer salaries on job satisfaction," *American Economic Review*, 102, 2981–3003.

Chen, L., A. F. Galvao, and S. Song (2021): "Quantile regression with generated regressors," *Econometrics*, 9.

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71, 1591–1608.

Chernozhukov, V. and I. Fernández-Val (2005): "Subsampling inference on quantile regression processes," *Sankhya: The Indian Journal of Statistics*, 67, 253–276.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2009): "Improving point and interval estimators of monotone functions by rearrangement," *Biometrika*, 96, 559–575.

——— (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132, 491–525.

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2023): "Toward personalized inference on individual treatment effects," *Proceedings of the National Academy of Sciences of the United States of America*, 120, 2–4.

Chetty, R. and N. Hendren (2018a): "The Impact of Neighborhoods on Intergenerational Mobility I: County-Level Estimates," *Quarterly Journal of Economics*, 133, 1107–1162.

——— (2018b): "The impacts of neighborhoods on intergenerational mobility II: County-level How are children," *The Quarterly Journal of Economics*, 133, 1163–1228.

Chetty, R., N. Hendren, and L. F. Katz (2016): "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment," *American Economic Review*, 106, 855–902.

Chetverikov, D., B. Larsen, and C. Palmer (2016): "IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade," *Econometrica*, 84, 809–833.

Ehrlich, M. V. and H. G. Overman (2020): "Place-based policies and spatial disparities across European Cities," *Journal of Economic Perspectives*, 34, 128–149.

Ehrlich, M. V. and T. Seidel (2018): "The persistent effects of place-based policy: Evidence from the West-German Zonenrandgebiet," *American Economic Journal: Economic Policy*, 10, 344–374.

Engbom, N. and C. Moser (2022): "Earnings Inequality and the Minimum Wage: Evidence from Brazil," *American Economic Review*, 112, 3803–3847.

Fehr, D., H. Rau, S. T. Trautmann, and Y. Xu (2020): "Inequality, fairness and social capital," *European Economic Review*, 129, 103566.

Fernández-Val, I., W. Y. Gao, Y. Liao, and F. Vella (2022): "Dynamic Heterogeneous Distribution Regression Panel Models, with an Application to Labor Income Processes," 1–45.

FIREBAUGH, G. (2000): "The trend in between-nation income inequality," *Annual Review of Sociology*, 26, 323–339.

FRANGURIDI, G., B. GAFAROV, AND K. WÜTHRICH (2024): "Bias correction for quantile regression estimators," .

FRIEDRICH, B. U. (2022): "Trade shocks, firm hierarchies, and wage inequality," *Review of Economics and Statistics*, 104, 652–667.

FRUMENTO, P., M. BOTTAI, AND I. FERNÁNDEZ-VAL (2021): "Parametric Modeling of Quantile Regression Coefficient Functions With Longitudinal Data," *Journal of the American Statistical Association*, 116, 783–797.

GALÍ, J. (1994): "Keeping up with the Joneses : Consumption Externalities , Portfolio Choice , and Asset Prices," *Journal of Money , Credit and Banking*, 26, 1–8.

GALLE, S., A. RODRÍGUEZ-CLARE, AND M. YI (2023): "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade," *Review of Economic Studies*, 90, 331–375.

GALVAO, A. F., J. GU, AND S. VOLGUSHEV (2020): "On the unbiased asymptotic normality of quantile regression with fixed effects," *Journal of Econometrics*, 218, 178–215.

GALVAO, A. F. AND K. KATO (2016): "Smoothed quantile regression for panel data," *Journal of Econometrics*, 193, 92–112.

GALVAO, A. F. AND L. WANG (2015): "Efficient Minimum Distance Estimator for Quantile Regression Fixed Effects Panel Data," *Journal of Multivariate Analysis*, 133, 1–26.

GLAESER, E. L. (2000): "The Future of Urban Research: Non-Market Interactions," *Brookings-Wharton Papers on Urban Affairs*, 2000, 101–138.

GLAESER, E. L., M. RESSEGER, AND K. TOBIO (2009): "Inequality in cities," *Journal of Regional Science*, 49, 617–646.

GOESLING, B. (2001): "Changing income inequalities within and between nations: New evidence," *American Sociological Review*, 66, 745–761.

HALTIWANGER, B. J., H. R. HYATT, AND J. R. SPLETZER (2024): "Rising Top, Falling Bottom : Industries and Rising Wage Inequality," *American Economic Review*, 114, 3250–3283.

HAUSHOFER, J., P. NIEHAUS, C. PARAMO, E. MIGUEL, AND M. W. WALKER (2022): "Targeting Impact Versus Deprivation," *SSRN Electronic Journal*.

HECKMAN, J. J. AND N. CLEMENTS (1997): "Making The Most Out Of Programme Evaluations and Social Experiments : Accounting For Heterogeneity in Programme Impacts," .

HELPMAN, E., O. ITSKHOKI, M. A. MUENDLER, AND S. J. REDDING (2017): "Trade and inequality: From theory to estimation," *Review of Economic Studies*, 84, 357–405.

HODERLEIN, S. AND E. MAMMEN (2007): "Identification of marginal effects in nonseparable models without monotonicity," *Econometrica*, 75, 1513–1518.

HONORÉ, B., S. KHAN, AND J. L. POWELL (2002): "Quantile regression under random censoring," *Journal of Econometrics*, 109, 67–105.

Hung, H. F. (2021): "Recent Trends in Global Economic Inequality," *Annual Review of Sociology*, 47, 349–367.

Kaji, T. and J. Cao (2023): "Assessing Heterogeneity of Treatment Effects," 1–22.

Kitagawa, T. and A. Tetenov (2018): "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

——— (2021): "Equality-Minded Treatment Choice," *Journal of Business and Economic Statistics*, 39, 561–574.

Koenker, R. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33.

Lang, V., N. Redeker, and D. Bischof (2023): "Place-based Policies and Inequality Within Regions," *OSF Reprints*.

Liao, Y. and X. Yang (2018): "Uniform Inference for Characteristic Effects of Large Continuous-Time Linear Models," 1–52.

Liu, X. (2021): "Panel Quantile Regression with Time-Invariant Rank," 1–68.

Lu, X. and L. Su (2022): "Uniform inference in linear panel data models with two-dimensional heterogeneity," *Journal of Econometrics*.

Luttmer, E. F. (2005): "Neighbors as negatives: Relative earnings and well-being," *Quarterly Journal of Economics*, 120, 963–1002.

Ma, L. and R. Koenker (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134, 471–506.

Manski, C. F. (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246.

Matzkin, R. L. (2003): "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71, 1339–1375.

Mazzolari, F. and G. Ragusa (2013): "Spillovers from High-Skill Consumption to Low-Skill Labor Markets," *Review of Economics and Statistics*, 95, 74–86.

McKenzie, D. and S. Puerto (2021): "Growing Markets through Business Training for Female Entrepreneurs: A Market-Level Randomized Experiment in Kenya," *American Economic Journal: Applied Economics*, 13, 297–332.

Melly, B. and M. Pons (2024): "Minimum Distance Estimation of Quantile Panel Data Models," *Working Paper*.

Neumark, D. and H. Simpson (2015): *Place-Based Policies*, vol. 5, Elsevier B.V., 1 ed.

Newey, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167.

Newey, W. K. and D. McFadden (1994): "Chapter 36 Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

Pakes, A. and D. Pollar (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Pierce, J. R. and P. K. Schott (2016): "Trade Liberalization and Mortality: Evidence from U.S. Counties," *Finance and Economics Discussion Series*, 2016, 47–64.

Støstad, M. N. and F. A. Cowell (2024): "Inequality as an Externality: Consequences for Tax Design," *Journal of Public Economics*, 235.

Tarozzi, A., J. Desai, and K. Johnson (2015): "The impacts of microcredit: Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 7, 54–89.

Torgovitsky, A. (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83, 1185–1197.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press, 3 ed.

van der Vaart, A. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*, Springer New York.

Volgushev, S., S.-K. Chao, and G. Cheng (2019): "Distributed inference for quantile regression processes," *The Annals of Statistics*, 47, 1634–1662.

Weymark, J. A. (1981): "Generalized gini inequality indices," *Mathematical Social Sciences*, 1, 409–430.

Zhang, Y., H. J. Wang, and Z. Zhu (2019): "Quantile-Regression-Based Clustering for Panel Data," *Journal of Econometrics*, 213, 54–67.

A Preliminary Lemmas

Let $\mathcal{B}$ is a vector space endowed with a pseudo-metric $||\cdot||_{\mathcal{B}}$, which is a sup-norm metric in the sense that $||\beta - \beta'||_{\mathcal{B}} = \sup_j ||\beta_j - \beta_j'||$.

Lemma 2 (Uniform consistency of $\hat{\beta}_j(\tau_1)$)**.** *Under Assumptions 1-4 and 9(a), we have*

$$\sup_{\tau \in \mathcal{T}} \max_{1 \leq j \leq m} \|\hat{\beta}_j(\tau_1) - \beta_{j,0}(\tau_1)\| = o_p(1).$$

Proof. The proof follows directly by the proof of Lemma 3 in Melly and Pons (2024) after noting that $\hat{\beta}(\tau_1)$ does not depend on τ_2. ∎

Lemma 3 (Bahadur representation of the first stage estimator)**.** *Let assumption 1-4 be satisfied. Then,*

$$\hat{\beta}_j(\tau_1) - \beta_j(\tau_1) = \frac{1}{n}\sum_{i=1}^{n} \phi_{j,\tau_1}(\tilde{x}_{ij}, y_{ij}) + R_{nj}^{(1)}(\tau_1) + R_{nj}^{(2)}(\tau_1), \tag{29}$$

where

$$\phi_{j,\tau_1}(\tilde{x}_{ij}, y_{ij}) = -B_{j,\tau_1}^{-1}\tilde{x}_{ij}(1(y_{ij} \leq \tilde{x}_{ij}\beta_j(\tau_1)) - \tau_1), \tag{30}$$

with $B_{j,\tau_1} = \mathbb{E}_{i|j}[f_{y|x}(Q_{y|x}(\tau_1|\tilde{x}_{ij}'\beta_j)|\tilde{x}_{ij})\tilde{x}_{ij}\tilde{x}_{ij}']$ *and*

$$\sup_j \sup_{\tau_1 \in \mathcal{T}} \left\| R_{nj}^{(2)}(\tau_1) \right\| = O_p\left(\frac{\log n}{n}\right) \tag{31}$$

$$\sup_j \sup_{\tau_1 \in \mathcal{T}} \left\| \mathbb{E}_{i|j}\left[R_{nj}^{(1)}(\tau_1)\right] \right\| = O\left(\frac{\log n}{n}\right) \tag{32}$$

$$\sup_j \sup_{\tau \in \mathcal{T}} \left\| \mathbb{E}\left[\left(R_{nj}^{(1)}(\tau_1) - \mathbb{E}_{i|j}[R_{nj}^{(1)}(\tau_1)]\right)\left(R_{nj}^{(1)}(\tau) - \mathbb{E}_{i|j}[R_{nj}^{(1)}(\tau_1)]\right)'\right] \right\| = O\left(\left(\frac{\log n}{n}\right)^{3/2}\right). \tag{33}$$

Proof. See Lemma 3 in Galvao et al. (2020). ∎

Lemma 4. *Under assumptions 1-2 and 7 ,*

$$\sup_{||\beta-\beta_0||_{\mathcal{B}} \leq \zeta_m,\ ||\delta-\delta_0|| \leq \zeta_m} ||M_{mn}(\delta, \beta, \tau) - M(\delta, \beta, \tau) - M_{mn}(\delta_0, \beta_0, \tau)|| = o_p(m^{-1/2}), \tag{34}$$

for all positive sequences $\zeta_m = o(1)$.

Proof. This result is implied by Theorem 3 in Chen et al. (2003). Hence, I show now that the conditions to apply the theorem are satisfied. First, recall that $\mathbb{E}[m(\delta_0, \beta_0, \tau)] = 0$ and that by Assumption 1 the data is i.i.d. To check condition (3.1), note that $m(\delta, \beta, \tau) = \rho_{\tau_2}(\tilde{x}_{ij}'\beta_j - x_{ij}'\delta)$. By the properties of the check function $\rho_{\tau_2}(y+z) - \rho_{\tau_2}(y) \leq 2 \cdot ||z||$. Hence,

$$\begin{aligned}
m(\delta',&\beta', \tau) - m(\delta'', \beta'', \tau) \\
&= \rho_{\tau_2}(\tilde{x}_{ij}'\beta_j' - x_{ij}'\delta') - \rho_{\tau_2}(\tilde{x}_{ij}'\beta_j'' - x_{ij}'\delta') + \rho_{\tau_2}(\tilde{x}_{ij}'\beta_j'' - x_{ij}'\delta') - \rho_{\tau_2}(\tilde{x}_{ij}'\beta_j'' - x_{ij}'\delta'') \\
&\leq 2||\tilde{x}_{ij}'(\beta_j' - \beta_j'')||_{\mathcal{B}} + 2||x_{ij}'(\delta' - \delta'')||.
\end{aligned} \tag{35}$$

It follows that $m(\delta, \beta, \tau)$ is Hölder continuous because

$$|m(\delta', \beta', \tau) - m(\delta'', \beta'', \tau)| \leq C_1 ||\beta_j' - \beta_j''||_{\mathcal{B}} + C_2||\delta' - \delta''||$$

with $C_1 = 2 \cdot ||\tilde{x}_{ij}||$ and $C_2 = 2 \cdot ||x_{ij}||$. which are bounded by assumption 2. This implies that condition (3.1) in Chen et al. (2003) is satisfied.

Condition (3.2) is satisfied as

$$\begin{aligned}
&\mathbb{E}\Bigg[\sup_{||\beta_j' - \beta_j|| \leq \zeta,\ ||\delta' - \delta|| \leq \zeta} |m(\delta, \beta, \tau) - m(\delta', \beta', \tau)|^2 \Bigg] \\
&\qquad \leq \mathbb{E}\big[||x_{ij}||^2 |1(\tilde{x}_{ij}'\beta_j(\tau_1) \leq x_{ij}'\delta) - 1(\tilde{x}_{ij}'\beta_j(\tau_1) \leq x_{ij}'(\delta + \zeta))| \\
&\qquad\quad + ||x_{ij}||^2 |1(\tilde{x}_{ij}'\beta_j(\tau_1) \leq x_{ij}'(\delta + \zeta)) - 1(\tilde{x}_{ij}'(\beta_j(\tau_1) + \zeta) \leq x_{ij}'(\delta + \zeta))|\big] \\
&\qquad \leq \mathbb{E}\big[||x_{ij}||^2 |F_{Q(\tau_1, y|x,\nu)|x}(x_{ij}'\delta) - F_{Q(\tau_1, y|x,\nu)|x}(x_{ij}'(\delta + \zeta))| \\
&\qquad\quad + ||x_{ij}||^2 |F_{Q(\tau_1, y|x,\nu)|x}(x_{ij}'\delta) - F_{Q(\tau_1, y|x,\nu)|x}(x_{ij}'(\delta + \zeta) - \tilde{x}_{ij}'\zeta))|\big] \\
&\qquad \leq K \cdot \zeta
\end{aligned}$$

for some $K < \infty$, since x_{ij} is bounded by assumption 2.
To check condition (3.3), I start by noting that by assumption 7, $\mathcal{D}$ is a compact subset of $\mathbb{R}^K$. Further $\beta_j \in \mathcal{B}_j$ for all j, where $\mathcal{B}_j$ is a compact set of $\mathbb{R}^{K_1}$. It follows by Tychonoff's Theorem that $\beta \in \mathcal{B}$, where $\mathcal{B} = \prod_{j=1}^m \mathcal{B}_j$ is also compact. Since both sets are compact, the covering numbers of $\mathcal{B}$ and $\mathcal{D}$ are known, and the condition is satisfied.

■

B Proofs of Asymptotic Results

B.1 Pointwise Results

B.1.1 Consistency

Theorem 4 **(Consistency).** *Let assumptions 1-7 and 9(a) be satisfied. Then,* $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau) \xrightarrow{p} 0$.

Proof of Theorem 4. To prove the results, I apply Theorem 1 in Chen et al. (2003) and start by showing that the conditions to apply the theorem are satisfied. First, by definition $M(\delta_0, \beta_0, \tau) = 0$ and $||M_{mn}(\hat{\delta}, \hat{\beta}, \tau)|| \leq \inf_{\delta \in \mathcal{D}_\zeta} ||M_{mn}(\delta, \hat{\beta}, \tau)|| + o_p(1)$ so that condition (1.1) is satisfied. Condition (1.4) is implied by Lemma 2 and (1.5) is implied by Lemma 4. Condition (1.3) is satisfied since $M(\delta, \beta, \tau)$ is Lipschitz-continuous over β_j at $\beta_j = \beta_{j,0}$ with respect to the metric $|| \cdot ||_{\mathcal{B}}$. Condition (1.2) is satisfied as $M(\delta, \beta_0)$ is uniquely minimized at $M(\delta_0, \beta_0)$, since $\mathbb{E}[x_{ij}x_{ij}']$ is full rank (Assumption 2) and by Assumption 6. Since all the conditions are satisfied, the result follows by Theorem 1 in Chen et al. (2003).

■

Proof of Lemma 1. **B.1.2 Asymptotic Normality**

Part (i) Inserting equation (29) from Lemma 3 in equation (19) gives:

$$\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)(\hat{\beta}_j-\beta_{j,0}) \tag{36}$$

$$=\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau_1}(\tilde{x}_{ij},y_{ij})\right) \tag{37}$$

$$+\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(1)}(\tau_1) \tag{38}$$

$$+\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(2)}(\tau_1) \tag{39}$$

First, note that by Assumption 2, x_{ij} is bounded by a constant C such that $x_{ij}\tilde{x}_{ij}'$ is also bounded. Further, by Assumption 5, $f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})$ is also bounded uniformly over τ. It follows directly that the conditional expectation $\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]$ is bounded uniformly over τ.

Next, consider the third term (39). Together with equation (31), it implies that

$$\sup_{\tau_2\in\mathcal{T}}\sup_{\tau_1\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(2)}(\tau_1)=O_p\left(\frac{\log n}{n}\right). \tag{40}$$

For the second term (38), Since $\mathrm{Var}\left(R_{nj}^{(1)}(\tau)\right)=o\left(\frac{1}{n}\right)$ by (33), the conditional expectation is bounded and since the observations are independent across groups, we have that

$$\mathrm{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(1)}(\tau_1)\right)=o_p\left(\frac{1}{mn}\right).$$

In addition, by (32), $\sup_j\sup_{\tau_1\in\mathcal{T}}\mathbb{E}_{i|j}\left[R_{nj}^{(1)}(\tau_1)\right]=O\left(\frac{\log n}{n}\right)$ such that

$$\sup_{\tau_2\in\mathcal{T}}\sup_{\tau_1\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(1)}(\tau_1)=O\left(\frac{\log n}{n}\right)$$

where the uniformity over τ_2 follows since $R_{nj}^{(1)}(\tau_1)$ does not depend on τ_2. Putting this together, by the Chebyshev inequality and under Assumption 9(b), we have that

$$\sup_{\tau_2\in\mathcal{T}}\sup_{\tau_1\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(1)}(\tau_1)=o_p\left(\frac{1}{\sqrt{m}}\right). \tag{41}$$

It follows that both (38) and (39) are $o_p\left(\frac{1}{\sqrt{m}}\right)$ uniformity over τ.

Consider now the first term (37):

$$
\begin{aligned}
&\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau_1}(\tilde{x}_{ij},y_{ij})\right)\\
&\qquad=\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]\\
&\qquad\qquad\cdot\left(-\frac{B_{j,\tau_1}^{-1}}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\le\tilde{x}_{ij}\beta_j(\tau_1))-\tau_1)\right)\\
&\qquad=\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}s_{ij}(\tau).
\end{aligned}
$$

This is a sample mean over mn i.i.d. observations denoted by $s_{ij}(\tau)$. The model in equation (2) implies that $\mathbb{E}_{i|j}\left[1(y_{ij}\le\tilde{x}_{ij}\beta_j(\tau))|\tilde{x}_{ij}\right]=\tau_1$, which together with Assumption 2(iii) gives $\mathbb{E}[s_{ij}(\tau)]=0$. In addition,

$$
\begin{aligned}
\mathrm{Var}(s_{ij}(\tau_1)) =&\mathbb{E}_j\left[\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)\,\mathrm{Var}(\phi_{j,\tau_1})\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)'\right]\\
=&\mathbb{E}_j[\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)B_{j,\tau}^{-1}\tau(1-\tau)\mathbb{E}_{i|j}[\tilde{x}_{ij}\tilde{x}_{ij}']B_{j,\tau}^{-1}\bar{\Gamma}_{2j}(\tau,\delta_0,\beta_0)'],
\end{aligned}\tag{42}
$$

where $\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)=\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]$.

It follows by the Lindeberg-Lévy Central Limit Theorem that

$$
\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}s_{ij}(\tau)=O_p\left(\frac{1}{\sqrt{mn}}\right).\tag{43}
$$

This last results implies that the first term (36) is $o_p\left(\frac{1}{\sqrt{m}}\right)$ pointwise. To get uniform results, note that

$$
\left\{\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]\left(\frac{-B_{j,\tau_1}^{-1}}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\le\tilde{x}_{ij}\beta(\tau_1))-\tau_1)\right),\tau_1\in\mathcal{T},\tau_2\in\mathcal{T},\beta\in\mathcal{B}\right\}
$$

is a Donsker class for any compact set $\mathcal{B}$. This follows by noting that $\{1(y_{ij}\le\tilde{x}_{ij}\beta_j(\tau)),\tau_1\in\mathcal{T},\tau_2\in\mathcal{T},\beta\in\mathcal{B}\}$ is a VC subgraph class and hence a bounded Donsker class. Hence,

$$
\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\le\tilde{x}_{ij}\beta)-\tau),\tau_1\in\mathcal{T},\tau_2\in\mathcal{T},\beta\in\mathcal{B}\right\}
$$

is also bounded Donsker with a square-integrable envelope $2\cdot\max_{i\in 1,\dots,n}|\tilde{x}_{ij}|\le 2\cdot C$. The whole function is then Donsker by the boundedness of $\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]$ and B_{j,τ_1}^{-1}. Hence, it follows that the equation (36) is $o_p\left(\frac{1}{\sqrt{m}}\right)$ uniformly in τ_1 and τ_2.

Part (ii) This part of the proof is implied by the proof of Theorem 2.

Part (iii) Note that $\sum_{j=1}^{m}\frac{1}{m}\bar{\Gamma}_{2,j}(\beta_0,\tau)\left(\hat{\beta}_j-\beta_{j,0}\right)$ is asymptotically equivelent to (up to a $o_p\left(\frac{1}{\sqrt{m}}\right)$ term)

$$
\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]\left(-\frac{B_{j,\tau_1}^{-1}}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\le\tilde{x}_{ij}\beta_j(\tau_1))-\tau_1)\right)
$$

Since the observations are independent over j and i (Assumption 1) we only need to analyze the covariance for a given i and j:

$$\begin{aligned}
&\text{Cov}\left(\tilde{x}_{ij}(1(y_{ij} \leq \tilde{x}_{ij}\beta_j(\tau_1)) - \tau_1), x'_{ij}1(\tilde{x}'_{ij}\beta_{j,0}(\tau_1) \leq x'_{ij}\delta_0(\beta_0, \tau) - \tau_2)\right) \\
&\quad = \text{Cov}\left(\tilde{x}_{ij}(1(y_{ij} \leq \tilde{x}_{ij}\beta_{j,0}(\tau_1)) - \tau_1), x'_{ij}1(x'_{ij}[\delta(\tau_1, v_j) - \delta_0(\beta_0, \tau)] + \alpha(\tau_1, v_j) \leq 0) - \tau_2)\right) \\
&\quad = \mathbb{E}\left[\tilde{x}_{ij}(1(y_{ij} \leq \tilde{x}_{ij}\beta_{j,0}(\tau_1)) - \tau_1)x'_{ij}1(x'_{ij}[\delta(\tau_1, v_j) - \delta_0(\beta_0, \tau)] + \alpha(\tau_1, v_j) \leq 0) - \tau_2]\right] \\
&\quad = \mathbb{E}_j\left[\mathbb{E}_{i|j}\left[\tilde{x}_{ij}(1(y_{ij} \leq \tilde{x}_{ij}\beta_{j,0}(\tau_1)) - \tau_1)|x_{ij}\right] x'_{ij}1(x'_{ij}[\delta(\tau_1, v_j) - \delta_0(\beta_0, \tau)] + \alpha(\tau_1, v_j) \leq 0) - \tau_2]\right] = 0.
\end{aligned}$$

Where the second line follows since both terms have mean zero.

This implies that uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$,

$$\text{Cov}\left(M_{mn}(\delta_0, \beta_0, \tau), \frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2,j}(\beta_0, \tau)\left(\hat{\beta}_j - \beta_{j,0}\right)\right) = o_p\left(\frac{1}{\sqrt{m}}\right)$$

■

Theorem 5 (**Asymptotic Normality**). *Let assumptions 1-7 and 9(b) be satisfied. Then*

$$\sqrt{m}\left(\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)\right) \xrightarrow{d} N(0, \Gamma_1^{-1}\Omega_2(\tau)\Gamma_1'^{-1}) \tag{44}$$

with $\Gamma_1 = \Gamma_1(\delta_0, \beta_0, \tau)$.

Proof of Theorem 5. To prove the results, I apply Theorem 2 in Chen et al. (2003) and start by showing that the conditions to apply the theorem are satisfied. First, assumption 9(b) implies 9(a) so that by Theorem 4, $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau) \xrightarrow{p} 0$. Therefore, following Chen et al. (2003), I can replace the parameter space with a small or shrinking ball around the true parameter. Let $\mathcal{D}_{\zeta,\tau} = \{\delta \in \mathcal{D} : ||\delta - \delta_0(\tau)|| \leq \zeta_m\}$ and $\mathcal{B}_{\zeta,\tau_1} = \{\beta \in \mathcal{B} : ||\beta - \beta_0(\tau_1)|| \leq \zeta_m\}$.

Next, by definition

$$||M_{mn}(\hat{\delta}, \hat{\beta}, \tau)|| \leq \inf_{\delta \in \mathcal{D}_\zeta} ||M_{mn}(\delta, \hat{\beta}, \tau)|| + o_p(m^{-1/2})$$

so that condition (2.1) is trivially satisfied.

Recall the matrix

$$\Gamma_1(\delta, \beta_0, \tau_2) = \frac{\partial M(\delta, \beta_0, \tau_2)}{\partial \delta} = \mathbb{E}[f_{Q(\tau_1, y_{ij}|x_{ij}, v_j)|x_{ij}}(x'_{ij}\delta|x_{ij})x_{ij}x'_{ij}]. \tag{45}$$

By assumption 5, $\Gamma_1(\delta, \beta_0, \tau_2)$ exist, is continuous at $\delta = \delta_0$. Further, $\Gamma_1(\delta_0, \beta_0, \tau_2)$ is full rank by assumptions 2 and 6. Hence, condition (2.2) is satisfied.

Denote $\Gamma_2(\delta,\beta_0,\tau)[\beta(\tau_1)-\beta_0(\tau_1)]=\sum_{j=1}^m \Gamma_{2j}(\delta,\beta_0,\tau)[\beta_j(\tau_1)-\beta_{j,0}(\tau_1)]$ the pathwise derivative of $M(\delta,\beta_0,\tau_1)$ in the direction $(\beta-\beta_0,\tau_1)$, where

$$\begin{aligned}\Gamma_{2j}(\delta,\beta_0,\tau)=&\frac{\partial}{\partial\beta_j}\left[\mathbb{E}[\tau_2-F_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta(\beta_0,\tau)|x_{ij})x_{ij}]\right]\\=&\frac{\partial}{\partial\beta_j}\left[\mathbb{E}_j\left[\mathbb{E}_{i|j}\left[\tau_2-F_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta(\beta_0,\tau)|x_{ij})x_{ij}\right]\right]\right]\\=&-\frac{1}{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta(\beta_0,\tau)|x_i)x_{ij}\tilde{x}_{ij}'\right].\end{aligned}$$

By assumption 5 the pathwise derivative will exist in all directions $(\beta_j-\beta_{j,0})\in\mathcal{B}_j$.

Condition (2.3) requires that for all $(\beta_j,\delta)\in\mathcal{B}_{\zeta_m}\times\mathcal{D}_{\zeta_m}$ with a positive sequence $\zeta_m=o_p(1)$, (i) $||M(\delta,\beta,\tau)-M(\delta,\beta_0,\tau)-\sum_{j=1}^m\Gamma_{2j}(\delta,\beta_0,\tau)[\beta_j-\beta_{j,0}]||\leq c\cdot\sup_j||\beta_j-\beta_{j,0}||^2$ for some constant $c\geq 0$; and (ii) $||\sum_{j=1}^m\Gamma_{2j}(\delta,\beta_0,\tau)[\beta_j-\beta_{j,0}]-\sum_{j=1}^m\Gamma_{2j}(\delta_0,\beta_0,\tau)[\beta_j-\beta_{j,0}]||\leq o(1)\zeta_m$. Note that sequence ζ_m is also defined in terms of the radius of the ball around β_0 and δ_0. Hence, we need this sequence converge to zero at a rate weakly slower than $(\beta-\beta_0)$ and $(\delta-\delta_0)$. Using a Taylor approximation and since Assumption 5 implies that $M(\delta,\beta,\tau)$ is twice continuously differentiable we have that

$$M(\delta,\beta,\tau)-M(\delta,\beta_0,\tau)=\sum_{j=1}^m\Gamma_{2j}(\delta,\beta_0,\tau)[\beta_j-\beta_{j,0}]+O_p(||\beta-\beta_0||_{\mathcal{B}}^2)\tag{46}$$

which implies (i). Condition (2.3ii) is trivially satisfied by Assumption 5.

Condition (2.4), is satisfied if $||\hat{\beta}_j-\beta_{j,0}||_{\mathcal{B}}=o_p(m^{-1/4})$. The proof of Lemma 1 in Galvao and Wang (2015) implies that

$$\mathrm{P}\left\{\max_{1\leq j\leq m}||\hat{\beta}_j(\tau)-\beta_{j,0}(\tau)||>\zeta\right\}\leq O(m\exp{(-n)}).$$

If $\frac{\sqrt{m}\log n}{n}\to 0$ (Assumption 9(b)), $\sup_j||\hat{\beta}_j-\beta_{j,0}||=o_p(m^{-1/4})$, so that condition (2.4) in Chen et al. (2003) is satisfied.

Condition (2.5) is implied by Lemma 4.

Under these conditions, it follows by the proof of Theorem 2 in Chen et al. (2003) that

$$\begin{aligned}&\sqrt{m}\left(\hat{\delta}(\hat{\beta},\tau)-\delta_0(\beta_0,\tau)\right)\\&=-\Gamma_1(\delta_0,\beta_0,\tau)^{-1}\sqrt{m}\left(\frac{1}{m}\sum_{j=1}^m\bar{\Gamma}_{2j}(\tau,\delta_0,\beta_0)[\hat{\beta}_j(\tau_1)-\beta_{j,0}(\tau_1)]+M_{mn}(\delta_0,\beta_0,\tau)\right)\\&\quad+o_p(1)\end{aligned}\tag{47}$$

Then, adding condition (2.6), which follows directly by Lemma 1 with

$$\sqrt{m}\left(M_{mn}(\delta_0,\beta_0,\tau)+\frac{1}{m}\sum_{j=1}^m\bar{\Gamma}_{2,j}(\beta_0,\tau)\left(\hat{\beta}_j-\beta_{j,0}\right)\right)\xrightarrow{d}N(0,\Omega_2(\tau)).\tag{48}$$

The final result follows by Theorem 2 in Chen et al. (2003). ∎

B.2 Uniform Results

B.2.1 Uniform Consistency

Proof of Theorem 1. Note that for all $\zeta > 0$ there exist $\epsilon(\zeta)$ such that

$$\inf_{\tau \in \mathcal{T} \times \mathcal{T}} \inf_{||\delta - \delta_0(\tau)|| > \zeta} \|M(\delta, \beta_0, \tau)\| > \epsilon(\zeta).$$

This implies that uniformly over $\tau \in \mathcal{T} \times \mathcal{T}$, if δ is more than ζ away from $\delta_0(\tau)$, then $M(\delta, \beta, \tau)$ is at least $\epsilon(\zeta)$. Hence, for any $\tau \in \mathcal{T}$, $||\delta - \delta_0(\tau)|| > \zeta$ implies $\|M(\delta, \beta_0, \tau)\| > \epsilon(\zeta)$. Since, it must also hold for the supremum over $\tau \in \mathcal{T} \times \mathcal{T}$, it follows that

$$\left\{ \sup_{\tau \in \mathcal{T} \times \mathcal{T}} ||\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\tau)|| > \zeta \right\} \subseteq \left\{ \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \|M(\hat{\delta}(\hat{\beta}, \tau), \beta_0, \tau)\| > \epsilon(\zeta) \right\}$$

and therefore that

$$P\left(\sup_{\tau \in \mathcal{T} \times \mathcal{T}} ||\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\tau)|| > \zeta \right) \leq P\left(\sup_{\tau \in \mathcal{T} \times \mathcal{T}} \|M(\hat{\delta}(\hat{\beta}, \tau), \beta_0, \tau)\| > \epsilon(\zeta) \right)$$

.

I need to show that $\sup_{\tau \in \mathcal{T} \times \mathcal{T}} \|M(\hat{\delta}, \beta_0, \tau)\| = o_p(1)$.

Note that by triangle inequality

$$\begin{aligned}
\sup_{\tau \in \mathcal{T} \times \mathcal{T}} \|M(\hat{\delta}, \beta_0, \tau)\| \leq & \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \|M(\hat{\delta}, \beta_0, \tau) - M(\hat{\delta}, \hat{\beta}, \tau)\| \\
& + \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \|M(\hat{\delta}, \hat{\beta}, \tau) - M_{mn}(\hat{\delta}, \hat{\beta}, \tau)\| \\
& + \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \|M_{mn}(\hat{\delta}, \hat{\beta}, \tau)\|
\end{aligned}$$

The first term is $o_p(1)$ by continuity and uniform consistency of $\hat{\beta}(\tau)$. Next, note that $\mathcal{M} = \{m(\delta, \beta, \tau) : \delta \in \mathcal{D}, \beta \in \mathcal{B}, \tau \in \mathcal{T} \times \mathcal{T}\}$ is Lipschitz continuous, hence by Theorem 2.7.11 in van der Vaart and Wellner (1996) we can directly bound $N_{[\,]}(\varepsilon, \mathcal{M}, \|\cdot\|_{L_2(P)})$ from above by the covering number of the class $\{\delta, \in \mathcal{D}, \beta \in \mathcal{B}, \tau \in \mathcal{T} \times \mathcal{T}\}$ which is finite for any $\varepsilon > 0$ by assumption 7. It follows directly by Theorem 19.4 in van der Vaart (1998), that the class is Glivenko-Cantelli. Hence, the second term is also $o_p(1)$. The third term is also $o_p(1)$ by construction. ∎

B.2.2 Weak Convergence

Proof of Theorem 2. The proof consists of three parts. First, I show that the lineralization in equation (47) holds uniformly over τ_1, τ_2. Second, I show that uniformly over τ_1, τ_2, $\hat{\delta}(\hat{\beta}, \cdot) - \hat{\delta}(\beta_0, \cdot) = o_p\left(m^{-1/2}\right)$. With this result, weak convergence of $\sqrt{m}\left(\hat{\delta}(\beta_0, \cdot) - \delta_0(\beta_0, \cdot)\right)$ directly implies weak convergence of $\sqrt{m}\left(\hat{\delta}(\hat{\beta}, \cdot) - \delta_0(\beta_0, \cdot)\right)$. Third, I show that $\Gamma_1(\cdot)\sqrt{m}\left(\hat{\delta}(\beta_0, \cdot) - \delta_0(\beta_0, \cdot)\right) \rightsquigarrow \mathbb{G}(\cdot)$ in $\ell^\infty(\mathcal{T} \times \mathcal{T})$.

Part 1 – Linearization

By the proof of Theorem 2 in Chen et al. (2003) we have

$$||M_{mn}(\hat{\delta},\hat{\beta})-\mathcal{L}_{mn}(\hat{\delta})||=o_p(m^{-1/2})$$

with

$$\mathcal{L}_{mn}(\hat{\delta})=M_{mn}(\delta_0,\beta_0)+\Gamma_1(\delta_0,\beta_0)(\hat{\delta}-\delta_0)+\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2,j}(\delta_0,\beta_0)(\hat{\beta}_j-\beta_{j,0})$$

Now, I show that this linearization holds uniformly in $\tau\in\mathcal{T}\times\mathcal{T}$. If I can show that all pieces are asymptotically equicontinuous, then the entire thing is asymptotically equicontinuous, and the approximation is uniform by Newey (1991).

I start by noting that

$$m(\delta',\beta',\tau')-m(\delta'',\beta'',\tau'')=m(\delta',\beta',\tau')-m(\delta'',\beta'',\tau')+m(\delta'',\beta'',\tau')-m(\delta'',\beta'',\tau'').$$

From the proof of Lemma 4, we know that $m(\delta',\beta',\tau')-m(\delta'',\beta'',\tau')\leq 2||\tilde{x}_{ij}||\cdot||(\beta'_j-\beta''_j)||_{\mathcal{B}}+2||x_{ij}||\cdot||(\delta'-\delta'')||$. Next, by the properties of the check function

$$\begin{aligned}m(\delta'',\beta'',\tau')-m(\delta'',\beta'',\tau'')\leq&|\tau''_2-\tau'_2|\cdot||\tilde{x}'_{ij}\beta''_j-x'_{ij}\delta''||\\ \leq&|\tau''_2-\tau'_2|\cdot||x_{ij}||\left(||\beta''_j||_{\mathcal{B}}+||\delta''||\right),\end{aligned}$$

where x_{ij} is bounded by assumption 2. Hence, $(\delta,\beta,\tau)\mapsto m(\delta,\beta,\tau)$ is asymptotically equicontinuous.

By Taylor expansion

$$M(\delta,\beta,\tau)|_{\delta=\hat{\delta}(\tau)}=\mathbb{E}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x'_{ij}d(\tau)|x)x_{ij}x'_{ij}\right]|_{d(\tau)=\delta^*(\tau)}\left(\hat{\delta}(\tau)-\delta_0(\tau)\right)$$

where $\delta^*(\tau)$ is on the line connecting $\hat{\delta}(\tau)$ with $\delta(\tau)$ for each $\tau\in\mathcal{T}\times\mathcal{T}$ and is allowed to vary for each row of the Jacobian matrix. Then, by uniform consistency of $\hat{\delta}(\tau_1,\tau_2)$ (Theorem 1) and since by Assumption 5, $f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x'_{ij}d(\tau)|x)$ is uniformly continuous and bounded it follows that

$$M(\delta,\beta,\tau)|_{\delta=\hat{\delta}(\tau)}=[\Gamma_1(\delta_0,\beta_0,\tau)+o_p(1)]\left(\hat{\delta}(\tau)-\delta_0(\tau)\right)\tag{49}$$

uniformly over τ_1,τ_2. Hence, $\tau\mapsto\Gamma_1(\delta_0,\beta_0)(\hat{\delta}-\delta_0)$ is stochastically equicontinuous.

Similarly, by Taylor expansion

$$M(\delta_0,\beta,\tau)|_{\beta=\hat{\beta}}=\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_2(\delta_0,\beta,\tau)|_{\beta=\hat{\beta}^*}(\hat{\beta}_j-\beta_{j,0}),$$

where $\beta^*(\tau_1)$ is on the line connecting $\hat{\beta}(\tau_1)$ with $\beta_0(\tau_1)$ for each $\tau\in\mathcal{T}\times\mathcal{T}$. Using the same argument as above, and by uniform consistency of $\hat{\beta}_j$ over j and τ (Theorem 2), it follows that

$$M(\delta_0,\beta,\tau)|_{\beta=\hat{\beta}}=\frac{1}{m}\sum_{j=1}^{m}\left[\bar{\Gamma}_2(\delta_0,\beta,\tau)+o_p(1)\right](\hat{\beta}_j-\beta_{j,0})$$

uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$. Hence, $\tau \mapsto \frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_2(\delta_0,\beta,\tau)(\hat{\beta}_j - \beta_{j,0})$ is asymptotically equicontinuous.

Since all three terms of $\mathcal{L}(\delta)$ as well as $M_{mn}(\delta,\beta,\tau)$ are asymptotically equicontinuous, it follows that the linearization holds uniformly over τ_1,τ_2. Hence, uniformly over τ_1,τ_2, the asymptotic behavior of $\sqrt{m}\left(\hat{\delta}(\hat{\beta},\tau) - \delta_0(\beta_0,\tau)\right)$ is determined by

$$-\Gamma_1(\delta_0,\beta_0,\tau)^{-1}\sqrt{m}\left(\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2j}(\tau,\delta_0,\beta_0)[\hat{\beta}_j(\tau_1) - \beta_{j,0}(\tau_1)] + M_{mn}(\delta_0,\beta_0,\tau)\right) \tag{50}$$

Part 2

From the proof of Lemma 1, we know that the first stage error is $o_p\left(\frac{1}{\sqrt{m}}\right)$ uniformly over $\tau_1,\tau_2 \in \mathcal{T}\times\mathcal{T}$. Hence, the limiting distribution of $\hat{\delta}(\hat{\beta},\tau)$ is the same as the one of the infeasible estimator $\hat{\delta}(\beta_0,\tau)$. Formally,

$$\sup_{\tau_1,\tau_2\in\mathcal{T}\times\mathcal{T}}\|\hat{\delta}(\hat{\beta},\tau) - \hat{\delta}(\beta_0,\tau)\| = o_p\left(m^{-1/2}\right).$$

Part 3 – Weak Convergence of $\sqrt{m}\left(\hat{\delta}(\beta_0,\cdot) - \delta_0(\hat{\beta},\cdot)\right)$

This part of the proof closely follows the work of Angrist et al. (2006). Let $\mathbb{Q}_{mn}(\delta,\beta,\tau_1,\tau_2) = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n}\rho_{\tau_2}(\tilde{x}_{ij}'\beta_{j,0}-x_{ij}\delta)-\rho_{\tau}(\tilde{x}_{ij}'\beta_{j,0}(\tau_1)-x_{ij}\delta_0(\tau))$ and $Q(\tau_1,\tau_2,\delta,\beta) = \mathbb{E}[\rho_{\tau_2}(\tilde{x}_{ij}'\beta_{j,0}-x_{ij}\delta) - \rho_{\tau}(\tilde{x}_{ij}'\beta_{j,0}(\tau_1) - x_{ij}\delta_0(\tau))]$. Similarly to Angrist et al. (2006), the empirical processes $(\beta,\delta,\tau_1,\tau_2) \mapsto \mathbb{Q}_{mn}(\delta,\beta,\tau_1,\tau_2)$ are stochastically equicontinuous because

$$\left|\mathbb{Q}_{mn}(\delta',\beta',\tau_1',\tau_2') - \mathbb{Q}_{mn}(\delta'',\beta'',\tau_1'',\tau_2'')\right| \leq C_1\cdot|\tau_1'-\tau_1''|+C_2\cdot|\tau_2'-\tau_2''|+C_3\cdot\|\beta'-\beta''\|_{\mathcal{B}}+C_3\cdot\|\delta'-\delta''\|,$$

where $C_1 = 2\cdot C\cdot\sup_{\beta\in\mathcal{B}}\|\beta\|_{\mathcal{B}}$, $C_2 = 2\cdot C\cdot\sup_{\delta\in\mathcal{D}}\|\delta\|$ and $C_3 = 2\cdot C$. The constant C is defined in Assumption 2. Note that C_1, C_2 and C_3 are neither functions of τ_1 nor τ_2. It follows that, $\mathbb{Q}_{mn}(\delta',\beta',\tau_1',\tau_2')$ is Glivenko-Cantelli so that uniformly in $(\tau,\beta,\delta) \in \{\mathcal{T}\times\mathcal{T},\mathcal{B},\mathcal{D}\}$,

$$\mathbb{Q}_{mn}(\delta,\beta,\tau_1,\tau_2) = Q(\delta,\beta,\tau_1,\tau_2) + o_p(1). \tag{51}$$

Next consider a collection of closed balls $B_M(\delta(\tau))$ with radius M centered around $\delta(\tau)$, and let $\delta_M(\tau) = \delta(\tau) + \xi_M(\tau)\cdot\nu(\tau)$, where $\nu(\tau)$ is a direction vector with $||\nu(\tau)|| = 1$ and $\xi(\tau)$ is a positive scalar such that $\xi_M(\tau) \geq M$. Then uniformly in $\tau \in \mathcal{T}\times\mathcal{T}$:

$$\begin{aligned}\left(\frac{M}{\xi_M(\tau)}\right)&\cdot(\mathbb{Q}_{mn}(\delta_M(\tau),\beta,\tau_1,\tau_2) - \mathbb{Q}_{mn}(\delta(\tau),\beta,\tau_1,\tau_2))\\ &\geq \mathbb{Q}_{mn}(\delta_M^*(\tau),\beta,\tau_1,\tau_2) - \mathbb{Q}_{mn}(\delta(\tau),\beta,\tau_1,\tau_2)\\ &\geq Q(\delta_M^*(\tau),\beta,\tau_1,\tau_2) - Q(\delta(\tau),\beta,\tau_1,\tau_2) + o_p(1)\\ &> \epsilon_M + o_p(1),\end{aligned}$$

for some $\epsilon_M > 0$, where the first inequality follows by convexity in δ for $\delta_M^*(\tau)$ (i.e. the point on the boundary of $B_M(\delta(\tau))$, that lies on the line connecting $\delta_M(\tau)$ and $\delta(\tau)$). The second inequality is implied by equation (51); and the third follows because by convexity and by Assumption 2 and Assumption 6, $\delta(\tau)$ is the unique minimizer of $Q(\delta, \tau_1, \tau_2)$ uniformly in $\tau \in \mathcal{T}$. It then follows that,

$$\sup_{\tau_1, \tau_2 \in \mathcal{T} \times \mathcal{T}} \left\| \hat{\delta}(\beta_0, \tau_1, \tau_2) - \delta_0(\beta_0, \tau_1, \tau_2) \right\| \xrightarrow{p} 0. \tag{52}$$

Next, note that

$$\| \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} [\rho_{\tau_2}(\tilde{x}_{ij}' \beta_{j,0}(\tau_1) - x_{ij}' \hat{\delta}(\beta_0, \tau))] \| \leq C \cdot \sup_i \sup_j \|x_{ij}\| / m.$$

By assumption 2, $\sup_i \sup_j \|x_{ij}\| = O_p(1)$. Hence, uniformly in $\tau \in \mathcal{T}$

$$\sqrt{m} \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} \left[\rho_{\tau_2}(\tilde{x}_{ij}' \beta_{j,0}(\tau_1) - x_{ij}' \hat{\delta}(\beta_0, \tau)) \right] = \sqrt{m} M_{mn}(\hat{\delta}, \beta_0, \tau_1, \tau_2) = o_p(1), \tag{53}$$

Note that,

$$(\tau_1, \tau_2, \delta, \beta) \mapsto \sqrt{m} \left[M_{mn}(\hat{\delta}, \beta_0, \tau_1, \tau_2) - M(\hat{\delta}, \beta_0, \tau_1, \tau_2) \right]$$

is stochastic equicontinuous over $\mathcal{T} \times \mathcal{T} \times \mathcal{D} \times \mathcal{B}$, with respect to the $L_2(P)$ pseudometric

$$\xi((\delta', \beta', \tau_1', \tau_2'), (\delta'', \beta'', \tau_1'', \tau_2''))^2 = \max_{k=1,\dots,K} \mathbb{E} \left[\left(m_k(\delta', \beta', \tau_1', \tau_2') - m_k(\delta'', \beta'', \tau_1'', \tau_2'') \right)^2 \right]$$

where $k = 1, \dots, K$ indexes the elements of $m(\cdot)$.

Observe that $\{1(\tilde{x}_{ij}' \beta_j(\tau_1) \leq x_{ij}' \delta(\beta, \tau)), \tau_1 \in \mathcal{T}, \tau_2 \in \mathcal{T}, \beta \in \mathcal{B}, \delta \in \mathcal{D}\}$ is a VC subgraph class and hence a bounded Donsker class for any compact sets $\mathcal{B}$ and $\mathcal{D}$, where the sets are compact by assumption. Hence, it follows that the function class

$$\left\{ x_{ij}' [\tau_2 - 1(\tilde{x}_{ij}' \beta_j(\tau_1) \leq x_{ij}' \delta(\beta, \tau))], \tau_1 \in \mathcal{T}, \tau_2 \in \mathcal{T}, \beta \in \mathcal{B}, \delta \in \mathcal{D} \right\}$$

is also Donsker with a square-integrable envelope $2 \cdot \max_{i,j} \|x_{ij}\|$, and stochastic equicontinuity follows.

Next, stochastic equicontinuity of $(\tau_1, \tau_2, \delta) \mapsto \sqrt{m} \left[M_{mn}(\hat{\delta}, \beta_0, \tau_1, \tau_2) - M(\hat{\delta}, \beta_0, \tau_1, \tau_2) \right]$ together with uniform consistency of $\hat{\delta}(\beta_0, \tau_1, \tau_2)$ and the resulting convergence with respect to the pseudometric $\sup_{\tau_1, \tau_2 \in \mathcal{T} \times \mathcal{T}} \xi \left((\hat{\delta}(\tau), \beta_0(\tau_1), \tau_1, \tau_2)), (\delta(\tau), \beta_0(\tau_1), \tau_1, \tau_2)) \right) = o_p(1)$ imply

$$\begin{aligned} \sqrt{m} \Big[M_{mn}(\hat{\delta}, &\beta_0, \tau_1, \tau_2) - M(\hat{\delta}, \beta_0, \tau_1, \tau_2) \Big] \\ &= \sqrt{m} \left[M_{mn}(\delta_0, \beta_0, \tau_1, \tau_2) - M(\delta_0, \beta_0, \tau_1, \tau_2) \right] + o_p(1), \quad \text{in } \ell^\infty(\mathcal{T} \times \mathcal{T}). \end{aligned} \tag{54}$$

Note that convergence with respect to the pseudometric follows from

$$\begin{aligned} &\sup_{\tau \in \mathcal{T} \times \mathcal{T}} \max_{k=1,\dots,K} \xi \left(d(\tau), \beta_0(\tau_1), \tau_1, \tau_2)), (\delta(\tau), \beta_0(\tau_1), \tau_1, \tau_2)) \right)^2 \\ &\quad = \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \max_{k=1,\dots,K} \mathbb{E} \Big[\Big(x_{ijk} \left(1(\tilde{x}_{ij}' \beta_{j,0}(\tau_1) \leq x_{ij}' d(\tau)) - 1(\tilde{x}_{ij}' \beta_{j,0}(\tau_1) \leq x_{ijk}' (\delta(\tau))) \right) \Big)^2 \Big] \\ &\quad \leq \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \max_{k=1,\dots,K} \mathbb{E} \Big[\Big(x_{ijk} \left(F_{Q(\tau_1, y|x, \nu)|x}(x_{ij}' d(\tau)) - F_{Q(\tau_1, y|x, \nu)|x}(x_{ij}'(\delta)) \right) \Big)^2 \Big] \\ &\quad \leq \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \max_{k=1,\dots,K} \mathbb{E} \big[x_{ijk}^2 \big] \mathbb{E} \Big[\big(F_{Q(\tau_1, y|x, \nu)|x}(x_{ij}' d(\tau)) - F_{Q(\tau_1, y|x, \nu)|x}(x_{ij}'(\delta)) \big)^2 \Big] \\ &\quad \leq C^2 \mathbb{E} \left[f_Q^{max} \cdot ||x_{ij}||^2 \right] \sup_{\tau \in \mathcal{T} \times \mathcal{T}} ||d(\tau) - \hat{\delta}(\tau)||^2, \end{aligned}$$

where f_Q^{max} is an uniform upper bound on the density of $Q(\tau_1, y|x, \eta)$ given x defined in Assumption 5. The first inequality follows by the law of iterated expectation, the second by Cauchy-Schwarz, and the third by a Taylor expansion.
Combining equations (53), (54) and (49) yields

$$o_p(1) = \left[\Gamma_1(\delta_0, \beta_0, \cdot) + o_p(1)\right]\sqrt{m}\left(\hat{\delta}(\cdot) - \delta_0(\cdot)\right) + \sqrt{m}\left[M_{mn}(\hat{\delta}, \beta_0, \cdot) - M(\hat{\delta}_0, \beta_0, \cdot)\right]\Big]. \tag{55}$$

Since the Eigenvalues of Γ_1 are bounded away from zero by assumption, we obtain that uniformly in $\tau_1, \tau_2 \in \mathcal{T} \times \mathcal{T}$

$$\begin{aligned}\sup_{\tau_1,\tau_2 \in \mathcal{T}\times\mathcal{T}} &\|\sqrt{m}\left[M_{mn}(\delta_0, \beta_0, \tau_1, \tau_2) - M(\delta_0, \beta_0, \tau_1, \tau_2)\right] + o_p(1)\| \\ &\geq \left(\sqrt{\lambda} + o_p(1)\right) \cdot \sup_{\tau_1,\tau_2 \in \mathcal{T}\times\mathcal{T}} \sqrt{m}\|\hat{\delta}(\tau) - \delta_0(\tau)\|. \end{aligned} \tag{56}$$

where λ is the minimum eigenvalue of $\Gamma_1(\delta_0, \beta_0, \tau)$. Observe that $\lambda > 0$ by Assumptions 2 and 6.
Assumption 8 implies that $\delta(\beta_0, \tau)$ is continuous with respect to τ. It follows that $\tau \mapsto \sqrt{m}M_{mn}(\delta_0, \beta_0, \tau_1, \tau_2) - M(\delta_0, \beta_0, \tau_1, \tau_2)$ is stochastic equicontinuous over $\mathcal{T} \times \mathcal{T}$ with respect to the pseudometric, so that the functional central limit theorem implies

$$\sqrt{m}\left(M_{mn}(\delta_0, \beta_0, \cdot) - M(\delta_0, \beta_0, \cdot)\right) \rightsquigarrow \mathbb{G}(\cdot) \quad \text{in } \ell^\infty(\mathcal{T} \times \mathcal{T}) \tag{57}$$

where $\mathbb{G}(\cdot)$ is a Gaussian process with covariance function $\Omega_2(\cdot, \cdot)$ defined in Lemma 1. Hence, the left-hand side of (56) is $O_p(1)$ such that $\sup_{\tau_1,\tau_2 \in \mathcal{T}\times\mathcal{T}}\|\sqrt{m}(\hat{\delta}(\beta_0, \tau) - \delta_0(\beta_0, \tau))\| = O_p(1)$. Together with equations (55)-(57) this implies

$$\begin{aligned}&\Gamma_1(\cdot)\sqrt{m}\left(\hat{\delta}(\beta_0, \cdot) - \delta_0(\beta_0, \cdot)\right) \\ &\quad = -\sqrt{m}\left(M_{mn}(\delta_0, \beta_0, \cdot) - M(\delta_0, \beta_0, \cdot)\right) + o_p(1) \rightsquigarrow \mathbb{G}(\cdot) \quad \text{in } \ell^\infty(\mathcal{T} \times \mathcal{T})\end{aligned}$$

■

B.3 Inference

Proof of Theorem 3. This proof uses a similar strategy to the proof of Theorem 5.4 in Fernández-Val et al. (2022) The idea is to prove the result in two steps. First, show that $\hat{\delta}^* - \delta_0$ can be approximated by a linear function with an error of order $o_p^*(m^{-1/2})$. Then, show that the $\hat{\delta}^* - \hat{\delta}$ follow the same distribution as $\hat{\delta} - \delta_0$.

Part 1 - Linearization Since the bootstrap algorithm that I consider samples entire groups, the first stage is the same in all bootstrap replications. Instead, the source of randomness is which groups are sampled. In this section, I make the dependency of β on the data z explicit and denote the vector containing first-stage estimates in the bootstrap world $\hat{\beta}(z^*)$.
It can be shown that

$$\hat{\delta}^*(\hat{\beta}(z^*),\tau) - \hat{\delta}(\hat{\beta},\tau) = O_p^*\left(m^{-1/2}\right)$$

which together with Theorem 5 implies

$$\hat{\delta}^*(\hat{\beta}(z^*),\tau) - \delta_0(\beta_0,\tau) = O_p^*\left(m^{-1/2}\right). \tag{58}$$

Next, the idea is to approximate $\sqrt{m}(\hat{\delta}^*(\hat{\beta}(z^*),\tau) - \delta_0(\beta_0,\tau))$ with a linear function. Hence, the goal is to show that

$$\begin{aligned}\sqrt{m}(\hat{\delta}^*(\hat{\beta}(z^*),\tau) - \delta_0(\beta_0,\tau)) &= \Gamma_1(\delta_0,\beta_0,\tau)^{-1} \\ \times\sqrt{m}&\left(\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2j}^*(\delta_0,\beta_0,\tau)[\hat{\beta}_j(\tau_1,z^*) - \beta_{j,0}(\tau_1,z^*)] + M_{mn}^*(\delta_0,\beta_0,\tau)\right) + o_p^*(1),\end{aligned} \tag{59}$$

where $M_{mn}^*(\delta_0,\beta_0) = \frac{1}{m}\sum_{j=1}^m \bar{m}_j^*(\delta_0,\beta_0)$ and $\bar{m}_j(\delta,\beta,\tau) = \frac{1}{n}\sum_{i=1}^n x_{ij}'[\tau_2 - 1(\tilde{x}_{ij}'\beta_j(\tau_1) \leq x_{ij}'\delta(\beta),\tau)]$. For this part of the proof, I rely on the results from Chen et al. (2003). Define the linearization where the dependencies on τ are suppressed for ease of notation:

$$\mathcal{L}_{mn}^*(\delta) = M_{mn}^*(\delta_0,\beta_0) + \Gamma_1(\delta_0,\beta_0)(\delta-\delta_0) + \frac{1}{m}\sum_j \bar{\Gamma}_{2,j}^*(\delta_0,\beta_0)(\hat{\beta}_j(z^*) - \beta_{j,0}(z^*)). \tag{60}$$

The first step is to show that we can approximate $M_{mn}^*(\hat{\delta}^*,\hat{\beta}(z^*))$ by $\mathcal{L}_{mn}^*(\hat{\delta}^*)$ with an error of order $o_p^*(m^{-1/2})$ within a $O_p(m^{-1/2})$ neighborhood of δ_0. Hence, I want to show that

$$||M_{mn}^*(\hat{\delta}^*,\hat{\beta}(z^*)) - \mathcal{L}_{mn}^*(\hat{\delta}^*)|| = o_p^*(m^{-1/2}).$$

By the triangle inequality:

$$\begin{aligned}||M_{mn}^*(\hat{\delta}^*,\hat{\beta}(z^*)) - \mathcal{L}_{mn}(\hat{\delta}^*)|| \leq &||M(\hat{\delta}^*,\hat{\beta}(z^*)) - M(\hat{\delta}^*,\beta_0(z^*)) - \frac{1}{m}\sum_{j=1}^m \bar{\Gamma}_{2,j}^*(\delta_0,\beta_0)(\hat{\beta}_j(z^*) - \beta_{j,0}(z^*))|| \\ &+ ||M(\hat{\delta}^*,\beta_0(z^*)) - \Gamma_1(\delta_0,\beta_0)(\hat{\delta}^*-\delta_0)|| \\ &+ ||M_{mn}^*(\hat{\delta}^*,\hat{\beta}(z^*)) - M(\hat{\delta}^*,\hat{\beta}(z^*)) - M_{mn}^*(\delta_0,\beta_0)||.\end{aligned}$$

Where for the first term we have:

$$\begin{aligned}||M(\hat{\delta}^*,\hat{\beta}(z^*)) - &M(\hat{\delta}^*,\beta_0(z^*)) - \frac{1}{m}\sum_{j=1}^m \bar{\Gamma}_{2,j}^*(\delta_0,\beta_0)(\hat{\beta}_j(z^*) - \beta_{j,0}(z^*))|| \\ \leq &||M(\hat{\delta}^*,\hat{\beta}(z^*)) - M(\hat{\delta}^*,\beta_0(z^*)) - \frac{1}{m}\sum_{j=1}^m \bar{\Gamma}_{2,j}^*(\hat{\delta}^*,\beta_0)(\hat{\beta}_j(z^*) - \beta_{j,0}(z^*))|| \\ &+ ||\frac{1}{m}\sum_{j=1}^m \bar{\Gamma}_{2,j}^*(\hat{\delta}^*,\beta_0)(\hat{\beta}_j(z^*) - \beta_{j,0}(z^*)) - \frac{1}{m}\sum_{j=1}^m \bar{\Gamma}_{2,j}^*(\delta_0,\beta_0)(\hat{\beta}_j(z^*) - \beta_{j,0}(z^*))|| \\ = &O_p^*(||\hat{\beta}(z^*) - \beta_0||_{\mathcal{B}}^2) + o_p^*(1) \times ||\hat{\delta}^* - \delta_0|| = o_p^*(m^{-1/2}).\end{aligned}$$

since $\sup_j ||\hat{\beta}_j - \beta_{j,0}|| = o_p(m^{-1/4})$ as shown in the proof of Theorem 5.

For the second term, a Taylor approximation combined with $(\hat{\delta}^* - \delta_0) = O_p^*(m^{-1/2})$ implies:

$$||M(\hat{\delta}^*, \beta_0(z^*)) - \Gamma_1(\delta_0, \beta_0)(\hat{\delta}^* - \delta_0)|| = o_p^*(m^{-1/2}).$$

For the third term, we have:

$$||M_{mn}^*(\hat{\delta}^*, \hat{\beta}(z^*)) - M(\hat{\delta}^*, \hat{\beta}(z^*)) - M_{mn}^*(\delta_0, \beta_0)|| = o_p^*(m^{-1/2})$$

by condition 2.5 in Chen et al. (2003). Hence,

$$||M_{mn}^*(\hat{\delta}^*, \hat{\beta}(z^*)) - \mathcal{L}_{mn}^*(\hat{\delta}^*)|| = o_p^*(m^{-1/2}).$$

Similarly, I now shown that $||M_{mn}^*(\bar{\delta}^*, \hat{\beta}(z^*)) - \mathcal{L}_{mn}^*(\bar{\delta}^*)|| = o_p^*(m^{-1/2})$, where $\bar{\delta}^*$ is the value of δ that minimizes $\mathcal{L}_{mn}^*(\delta)$.

First, for $\bar{\delta}^*$ to be the value of δ that minimizes $\mathcal{L}^*(\delta)$ it must be equal to:

$$\begin{aligned}
\bar{\delta}^* - \delta_0 =& \Gamma_1(\delta_0, \beta_0)^{-1} \left(\frac{1}{m} \sum_{j=1}^{m} \bar{\Gamma}_{2j}^*(\delta_0, \beta_0)[\hat{\beta}_j(z^*) - \beta_{j,0}(z^*)] + \frac{1}{m} \sum_{j=1}^{m} \bar{m}_j^*(\delta_0, \beta_0) \right) && (61)\\
=& \Gamma_1(\delta_0, \beta_0)^{-1} \left(\frac{1}{m} \sum_{j=1}^{m} \bar{\Gamma}_{2j}(\delta_0, \beta_0)[\hat{\beta}_j - \beta_{j,0}] + \frac{1}{m} \sum_{j=1}^{m} \bar{m}_j(\delta_0, \beta_0) \right) + O_p^*(m^{-1/2})\\
=& \hat{\delta} - \delta_0 + O_p^*(m^{-1/2})\\
=& O_p^*(m^{-1/2}). && (62)
\end{aligned}$$

where the third line is implied by Theorem 5.

By the triangle inequality

$$\begin{aligned}
||M_{mn}^*(\bar{\delta}^*, \hat{\beta}(z^*)) - \mathcal{L}_{mn}^*(\bar{\delta}^*)|| \leq & ||M(\bar{\delta}^*, \hat{\beta}(z^*)) - M(\bar{\delta}^*, \beta_0(z^*)) - \sum_{j=1}^{m} \Gamma_{2,j}^*(\delta_0, \beta_0)(\hat{\beta}_j(z^*) - \beta_{j,0}(z^*))||\\
&+ ||M(\bar{\delta}^*, \beta_0^*) - \Gamma_1(\delta_0, \beta_0)(\bar{\delta}^* - \delta_0)||\\
&+ ||M_{mn}^*(\bar{\delta}^*, \hat{\beta}(z^*)) - M(\bar{\delta}^*, \hat{\beta}(z^*)) - M_{mn}^*(\delta_0, \beta_0)||.
\end{aligned}$$

For the first term, we have:

$$\begin{aligned}
||M(\bar{\delta}^*, \hat{\beta}(z^*)) - & M(\bar{\delta}^*, \beta_0(z^*)) - \sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0)[\hat{\beta}_j(z^*) - \hat{\beta}_{j,0}(z^*)]||\\
\leq & ||M(\bar{\delta}^*, \hat{\beta}(z^*)) - M(\bar{\delta}^*, \beta_0^*) - \sum_{j=1}^{m} \Gamma_{2j}(\bar{\delta}, \beta_0)[\hat{\beta}_j(z^*) - \beta_{j,0}(z^*)]||\\
&+ ||\sum_{j=1}^{m} \Gamma_{2j}(\bar{\delta}^*, \beta_0)[\hat{\beta}_j(z^*) - \beta_{j,0}(z^*)] - \sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0)[\hat{\beta}_j(z^*) - \beta_{j,0}(z^*)]||\\
&+ O_p(||\hat{\beta}_j - \beta_{j,0}||_{\mathcal{B}}^2) + o_p^*(1) \times ||\bar{\delta}^* - \delta_0|| = o_p^*(m^{-1/2})
\end{aligned}$$

For the second term, by differentiability of $M(\delta, \beta_0, \tau)$ and using equation (62) yields

$$||M(\bar{\delta}^*, \beta_0) - \Gamma_1(\bar{\delta}^* - \delta_0)|| = ||o(\bar{\delta}^* - \delta_0)|| = o_p^*(m^{-1/2}).$$

For the third term, by condition 2.5 in Chen et al. (2003), we have

$$||M^*_{mn}(\bar{\delta}^*, \hat{\beta}) - M(\bar{\delta}^*, \hat{\beta}) - M^*_{mn}(\delta_0, \beta_0)|| = o_p(m^{-1/2}).$$

Hence, it follows that

$$||M^*_{mn}(\bar{\delta}^*, \hat{\beta}(z^*)) - \mathcal{L}^*_{mn}(\bar{\delta}^*)|| = o^*_p(m^{-1/2}).$$

so that if I can show that $\bar{\delta}^* - \hat{\delta}^* = o^*_p(m^{-1/2})$, it follows from equation (61) that equation (59) holds. Following Pakes and Pollar (1989), we know that $M^*_{mn}(\delta, \beta_0)$ and $\mathcal{L}^*(\delta)$ are close at both $\hat{\delta}^*$ which almost minimizes $||M^*_{mn}(\delta, \hat{\beta}(z^*))||$ and at $\bar{\delta}^*$ which minimizes $\mathcal{L}^*(\delta)$. This means that $\hat{\delta}^*$ has to be close to minimizing $\mathcal{L}^*(\delta)$:

$$\begin{aligned}
||\mathcal{L}(\hat{\delta}^*)|| - o^*_p(m^{-1/2}) \leq& ||M_{mn}(\hat{\delta}^*, \hat{\beta}(z^*))|| \\
\leq& ||M_{mn}(\bar{\delta}^*, \hat{\beta}(z^*))|| + o^*_p(m^{-1/2}) \\
\leq& ||\mathcal{L}(\bar{\delta}^*)|| + o^*_p(m^{-1/2}).
\end{aligned}$$

This implies

$$||\mathcal{L}(\hat{\delta}^*)|| = ||\mathcal{L}(\bar{\delta}^*)|| + o^*_p(m^{-1/2}),$$

and squaring both sides

$$||\mathcal{L}(\hat{\delta}^*)||^2 = ||\mathcal{L}(\bar{\delta}^*)||^2 + o^*_p(m^{-1}), \tag{63}$$

where the cross product is also $o_p(m^{-1})$ because $||\mathcal{L}(\bar{\delta}^*)||$ is of order $O^*_p(m^{-1/2})$. The term $||\mathcal{L}(\delta)||^2$ has the simple expansion

$$||\mathcal{L}(\delta)||^2 = ||\mathcal{L}(\bar{\delta}^*)||^2 + ||\Gamma_1(\delta - \bar{\delta}^*)||^2 \tag{64}$$

around its global minimum. The cross-product term vanished because the residual vector $\mathcal{L}(\bar{\delta}^*)$, must be orthogonal to the columns of Γ_1. Let $\delta = \bar{\delta}^*$, and equations (63) and (64) give that

$$||\Gamma_1(\hat{\delta}^* - \bar{\delta}^*)|| = o^*_p(m^{-1/2}).$$

which implies

$$||(\hat{\delta}^* - \bar{\delta}^*)|| = o^*_p(m^{-1/2}).$$

as Γ_1 is full rank. Hence, equation (59) holds.

Similarly, to part 1 of the proof of Theorem 2, we can show that uniformly over $\tau_1, \tau_2 \in \mathcal{T} \times \mathcal{T}$,

$$\begin{aligned}
&\sqrt{m}(\hat{\delta}^*(\hat{\beta}(z^*), \tau) - \delta_0(\beta_0, \tau)) \\
=&\Gamma_1(\delta_0, \beta_0, \tau)^{-1}\sqrt{m}\left(\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}^*_{2j}(\delta_0, \beta_0, \tau)[\hat{\beta}_j(\tau_1, z^*) - \beta_{j,0}(\tau_1, z^*)] + M^*_{mn}(\delta_0, \beta_0, \tau)\right) + o^*_p(m^{-1/2}),
\end{aligned} \tag{65}$$

Part 2 - Asymptotic distribution of $\hat{\delta}^* - \hat{\delta}$

For this last part of the proof, I borrow from the proof of Proposition H.1. in Fernández-Val et al. (2022). First, denote

$$\theta^*_{mn}(\tau) = \Gamma_1(\delta_0, \beta_0, \tau)^{-1} \frac{1}{m} \sum_{j=1}^{m} \left(\bar{\Gamma}^*_{2j}(\delta, \beta_0, \tau)[\hat{\beta}_j(z^*, \tau) - \beta_{j,0}(z^*, \tau)] + \bar{m}^*_j(\delta_0, \beta_0, \tau) \right).$$

Since the bootstrap algorithm samples entire groups, we can write:

$$\mathbb{E}^* \left[\theta^*_{mn}(\tau)\right] = \Gamma_1(\delta_0, \beta_0, \tau)^{-1} \frac{1}{m} \sum_{j=1}^{m} \bar{\Gamma}_{2j}(\delta, \beta_0)[\hat{\beta}_j(\tau) - \beta_{j,0}(\tau)] + \Gamma_1(\delta_0, \beta_0, \tau)^{-1} M_{mn}(\delta_0, \beta_0, \tau).$$

Combining the expressions for $\hat{\delta}^*(\hat{\beta}(z^*), \tau)$ (Equation 65) and $\hat{\delta}(\hat{\beta}, \tau)$ (Equation 50) yields that uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$:

$$\hat{\delta}^*(\hat{\beta}(z^*), \tau) - \hat{\delta}(\hat{\beta}, \tau) = (\theta^*_{mn}(\tau) - \mathbb{E}^*[\theta^*_{mn}(\tau)]) + o^*_p(m^{-1/2})$$

For any $\tau^{(1)}, \ldots, ^{(T)}$, let $\Theta^*_{mn} = \left(\theta^*_{mn}(\tau^{(1)}) - \mathbb{E}^*[\theta^*_{mn}(\tau^{(1)})], \ldots, \theta^*_{mn}(\tau^{(T)}) - \mathbb{E}^*[\theta^*_{mn}(\tau^{(T)})]\right)$.

Let $\Sigma(\tau, \tau') = \Gamma_1(\tau)^{-1}\Omega_2(\tau, \tau')\Gamma_1(\tau')'^{-1}$. and $\Sigma = (\Sigma(\tau, \tau'))_{T \times T}$ and note that, $\mathrm{Var}^*(\sqrt{m}\Theta^*_{mn}) = \Sigma$. Then by the central limit theorem for i.i.d. data:

$$\sqrt{m}\Theta^* \xrightarrow{d^*} N(0, \Sigma).$$

Hence, $\hat{\delta}^* - \hat{\delta}$ has the same asymptotic distribution as $\hat{\delta} - \delta_0$.

Part 3 - Weak Convergence of the Bootstrap

By Lemma 1 we know that uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$,

$$\frac{1}{m} \sum_{j=1}^{m} \bar{\Gamma}_{2,j}(\delta_0, \beta_0, \tau) \left(\hat{\beta}_j(\tau_1) - \beta_{j,0}(\tau_1)\right) = o_p\left(\frac{1}{\sqrt{m}}\right),$$

and it directly follows that

$$\frac{1}{m} \sum_{j=1}^{m} \bar{\Gamma}^*_{2,j}(\delta_0, \beta_0, \tau) \left(\hat{\beta}_j(z^*, \tau_1) - \beta_{j,0}(z^*, \tau_1)\right) = o^*_p\left(\frac{1}{\sqrt{m}}\right).$$

Hence, uniformly over $\tau \in \mathcal{T} \times \mathcal{T}$

$$\hat{\delta}^*(\hat{\beta}(z^*), \tau) - \hat{\delta}(\hat{\beta}, \tau) = \Gamma_1(\delta_0, \beta_0, \tau)^{-1} \left(\frac{1}{m} \sum_{j=1}^{m} \bar{m}^*_j(\delta_0, \beta_0, \tau) - \bar{m}_j(\delta_0, \beta_0, \tau) \right) + o^*_p\left(\frac{1}{\sqrt{m}}\right)$$

From part (i) of the proof of Theorem 2, we know that $\tau \mapsto M_{mn}(\delta_0, \beta_0, \tau)$ is asymptotically equicontinuous. Given that $\frac{1}{m}\sum_{j=1}^{m} \bar{m}^*_j(\delta_0, \beta_0, \tau)$ is constructed by resampling with replacement elements from $\bar{m}_j(\delta_0, \beta_0, \tau)$ it follow that $\tau \mapsto \frac{1}{m}\sum_{j=1}^{m} \bar{m}^*_j(\delta_0, \beta_0, \tau)$ is also asymptotically equicontinuous. Hence, by Theorem 18.14 in van der Vaart and Wellner (1996) weak convergence follow:

$$\sqrt{m}\left(\hat{\delta}^*(\hat{\beta}(z^*), \cdot) - \hat{\delta}(\hat{\beta}, \cdot)\right) \rightsquigarrow^* \Gamma_1^{-1}(\cdot)\mathbb{G}(\cdot) \quad \textit{in } \ell^\infty(\mathcal{T} \times \mathcal{T}).$$

■

C Additional Simulation Results

Table 8: Standard Errors relative to Standard Deviation

	β			γ		
$\tau_1 \setminus \tau_2$	0.25	0.5	0.75	0.25	0.5	0.75
			(m, n) = (25,25)			
0.25	1.180	1.146	1.200	1.148	1.098	1.261
0.5	1.191	1.133	1.242	1.190	1.115	1.327
0.75	1.213	1.119	1.230	1.167	1.107	1.357
			(m, n) = (25,200)			
0.25	1.321	1.231	1.401	1.275	1.138	1.649
0.5	1.381	1.229	1.457	1.332	1.138	1.720
0.75	1.358	1.199	1.443	1.352	1.126	1.724
			(m, n) = (200,25)			
0.25	1.028	1.031	1.048	1.002	1.056	1.043
0.5	1.017	1.052	1.069	1.028	1.052	1.063
0.75	1.025	1.063	1.050	1.027	1.053	1.052
			(m, n) = (200,200)			
0.25	1.089	1.081	1.080	1.056	0.995	1.021
0.5	1.064	1.081	1.081	1.052	1.000	1.014
0.75	1.075	1.082	1.095	1.036	1.004	1.018
			(m, n) = (200,400)			
0.25	1.081	1.111	1.078	1.044	1.003	1.011
0.5	1.089	1.092	1.088	1.039	1.004	1.009
0.75	1.092	1.092	1.082	1.037	1.005	1.008

Notes: Results based on 2,000 Monte Carlo simulations. The table provides clustered standard errors relative to standard deviation.

Table 9: Coverage Probability

	β			γ		
$\tau_1 \setminus \tau_2$	0.25	0.5	0.75	0.25	0.5	0.75
(m, n) = (25,25)						
0.25	0.970	0.973	0.969	0.948	0.954	0.953
0.5	0.972	0.973	0.970	0.949	0.951	0.948
0.75	0.971	0.968	0.972	0.949	0.958	0.946
(m, n) = (25,200)						
0.25	0.985	0.987	0.985	0.957	0.959	0.965
0.5	0.986	0.985	0.981	0.956	0.956	0.964
0.75	0.988	0.988	0.987	0.955	0.953	0.954
(m, n) = (200,25)						
0.25	0.916	0.948	0.899	0.929	0.943	0.928
0.5	0.905	0.955	0.925	0.936	0.954	0.932
0.75	0.878	0.952	0.931	0.941	0.959	0.943
(m, n) = (200,200)						
0.25	0.964	0.965	0.954	0.948	0.938	0.940
0.5	0.955	0.961	0.956	0.945	0.940	0.944
0.75	0.961	0.963	0.961	0.947	0.942	0.947
(m, n) = (200,400)						
0.25	0.957	0.958	0.961	0.948	0.936	0.939
0.5	0.963	0.961	0.961	0.946	0.938	0.934
0.75	0.959	0.963	0.959	0.944	0.940	0.931

Notes: Results based on 2,000 Monte Carlo simulations. The table provides the coverage probability of the 95% confidence intervals computed using clustered standard errors.