Quantile on Quantiles*

Martina Pons[†]

February 8, 2024

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 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

Consider an educational policy designed to improve grades, which, for simplicity, is assumed to be randomized. Distributional effects are particularly interesting in this setting as they offer insights into how the treatment impacts the grade distribution. A policy increasing average grades could have opposite effects on inequality depending on whether the effect is more pronounced in the lower or upper tail. Traditional quantile regression of grades on a treatment dummy, as in Koenker and Bassett (1978), identifies treatment effects at different points of the unconditional grade distribution. More precisely, it provides the effect for high-achieving students and low-achieving students in the sample distribution of grades.[1] Yet, a median student in the unconditional grade distribution could be in the right tail of the distribution in a

*I am grateful to Blaise Melly and Anna Mikusheva. I also thank Isaiah Andrews, Alberto Abadie, Ivan Fernandez-Val, Jonas Meier and seminar participants at the MIT econometrics lunch 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.

[1] In contrast to least squares regression, the definition of quantile treatment effects depends on the included covariates. For example, a quantile regression of grades on the treatment and a gender dummy identifies the effects of the treatment on the distribution of grades within each gender category.

poor-performing school or a weak student in a highly competitive school, and the treatment likely has different effects on these two hypothetical individuals. First, despite being in the same percentile of the unconditional distribution, these two students experience substantially different environments. Second, the treatment effect could be lower in poor-performing schools, which may have to deal with additional challenges, or in high-performing schools, where superior teaching practices and infrastructure may already be in place. Another perspective is that there might be complementarities between individual abilities and school quality, which a traditional quantile regression model fails to capture. Hence, while allowing for heterogeneous effects based on an individual-level rank in the sample distribution provides interesting insights, it is important to recognize that inequality (or heterogeneity) has multiple layers.

This paper aims to study these complementarities by simultaneously analyzing treatment effect heterogeneities within and between groups. Geographical regions, firms, or industries could define such groups. In the example above, schools define the groups, and the model allows the treatment effect to flexibly vary over the distribution of grades within schools and over the distribution of schools. 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. This yields a quantile function of group-level quantile functions, offering insights, for instance, into the quantile function of median grades across schools.

The estimation is performed in two stages. The first stage consists of group-by-group quantile regression 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 saved and 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 within the 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 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 quantile regression.[2] Building on these results, I show that the estimator is asymptotically normally distributed in a framework where the number of observations per group n and the number of groups m diverge to infinity satisfying $\frac{\sqrt{m}\log n}{n} \to 0$.

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 decomposition 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 gives insights into which parts of the within distribution drive inequality between groups. Further, I show that the framework considered in the paper is useful for optimal policy targeting 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 can be useful 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 (highway, railways, sanitation) affect all people nearby, 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 results indicate larger effects for firms that perform well within their successful markets. More precisely, the effects increase both in the firm rank within a market and in the market rank. This suggests the existence of complementarities between individual and group ranks.

Distributional effects and inequalities within groups are studied both in the applied and theoretical literature. For example, Chetverikov et al. (2016), Galvao and Wang (2015), and Melly and Pons (2022) 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. 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. In contrast, papers studying the effectiveness of place-based policies in supporting laggard or underdeveloped regions provide examples focusing on disparities between groups

[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 among countries.

[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.

(see, e.g., Busso et al., 2013; Ehrlich and Seidel, 2018; Ehrlich and Overman, 2020).[5] Only a few papers focus on both within- and between-group inequality, mostly using a descriptive approach. For example, Bourguignon and Morrisson (2002) analyze the historical evolution of within and between countries income inequality, and Akerman et al. (2013) study wage inequality between and within different groups, including firms, sectors, and occupations. To this end, they decompose the variance into a within and a between components 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, Fernández-Val, Gao, Liao, and Vella (2022) introduce a model that allows for within and between group heterogeneity. 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, this model requires within-group variation in the variable of interest to identify heterogeneities in both dimensions.[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. Instead, 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 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. I start by considering a simplified version of the model imposing strong assumptions. Later, I relax these assumptions and present the more general model considered in this paper. I specify the following structural function for the outcome variable y_{ij}

[5] See Neumark and Simpson (2015) for an overview of the literature on place-based policies.

[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.

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 third and fourth arguments. Further, I assume that

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

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:

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

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 groups.

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 obtain 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 conditional quantile function of the 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. Assume for a moment that there are no covariates. Take groups h and l with v_h and v_l such that $v_h > v_l$. Strict monotonicity of $q(\tau_1, v_j)$ with respect to v_j implies

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

for all τ_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 would be generated by $y_{ij} = h(x_{1ij}, x_{2j}, u_{ij}) + f(x_{1ij}, x_{2j}, v_j)$.[7] That is, if conditional on the covariates, all groups share the same distribution

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

of the outcome up to a location parameter. This requires that, conditional on the covariates (x_{1ij}, x_{2j}), v_j enter as a pure location shifter.

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 allow 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 be to consider a case with a bivariate v_j where one element determines the mean and the other the variance. This corresponds to $y_{ij} = h(u_{ij}, v_j^{(1)}, x_{1ij}, x_{2j}) + f(v_j^{(2)}, x_{1ij}, x_{2j})$. Hence, conditional on the covariates x_{1ij} and x_{2ij}, the outcome has the same distribution 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, v_j is two-dimensional, and as it is not feasible to completely separate u_{ij} and v_j, the group rank varies 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, so that $q(v_j, \tau_1) = \tau_1(v_j^{(1)} + \epsilon) + v_j^{(2)} = \tau_1\epsilon + \tau_1 v_j^{(1)} + v_j^{(2)}$ 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 group 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.

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 as well as for different values of the covariates. For instance, the groups in the upper tail for $\tau_1 = 0.1$ might differ from groups in the upper tail of the distribution for $\tau_1 = 0.9$. At the same time, I maintain the assumptions on the scalar u_{ij}. Thus, the τ_1-conditional quantile function 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. Equation (2) 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.

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, I assume that for each $\tau_1 \in (0,1)$, there exists a scalar-valued function $v_j(\tau_1)$ such that

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

If I impose strict monotonicity of $q(x_{1ij}, x_{2j}, v_j(\tau_1), \tau_1)$ with respect to $v_j(\tau_1)$, we obtain 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 ranks over groups are allowed to be different at different points of the within distribution.

The main advantage of this way of ordering group is that it remains agnostic with respect to the social welfare function of the policymaker. When groups contain non-homogenous 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 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 decomposition. 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 locations in their conditional distribution, 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 requires 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 rank 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.[8] 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.[9] 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 rank invariance in the population 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, she will prefer groups with relatively high low wages and compressed wage distribution. Differently, individuals at the top of the within-group wage distribution will favor groups with high top wages.

Since no information is lost when modeling the two-level conditional quantile function, 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

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

[9] A rank invariance (or rank preservation) assumption is used, for example, in Chernozhukov and Hansen (2005); Firpo (2007).

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.

This model sheds light on the variation of the within percentiles of the outcome over groups. 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. 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 lower end of the wage distribution varies across groups. It thus offers a nuanced perspective on the dynamics of inequality within and between geographical regions.

To identify these heterogeneous coefficients, I suggest a two-step quantile regression estimator. (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 (2022) consider a similar model where the heterogeneity arises from the individual rank variable u_{ij} and the focus is on the within distribution.[10] 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:

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

[10]Note that Melly and Pons (2022) include the intercept in x_{2j}.

Hence, when model (6) holds, they identify the average effects over groups at the τ_1 quantile of the within distribution. In the social case that there is no group-level heterogeneity at a given τ_1, they identify the coefficients $\alpha(\tau_1, \tau_2)$, $\beta(\tau_1, \tau_2)$ and $\gamma(\tau_1, \tau_2)$. 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. 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) 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 Potential Application: Empirical Welfare Maximization

Two-dimensional quantile treatment effects can be used to optimally assign groups or individuals to treatment. Consider a policymaker who observes data from a *sample* population with a given group structure and has to decide whom to treat in a given *target* population (subject to some capacity/budget constraint) by maximizing a rank-dependent social welfare function:

$$W \equiv \int Y_{ij} \cdot w(\text{Rank}(Y_{ij}))di\ dj, \tag{7}$$

where $\text{Rank}(Y_{ij})$ is the rank of Y_{ij} in the population and $w(\cdot) \geq 0$ is a weight associated to a given rank. I consider a 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 between groups to more efficiently allocate the treatment. At the same time, the rank-dependent social welfare function implies that we are also interested in the heterogeneities within the groups.

In this section, I use the conventional potential outcome framework with a treatment variable D. Let y_d for $d \in \{0,1\}$ denote the potential outcome under treatment state d. The object of interest in this paper is the two-level quantile function of the potential outcomes under treatment d, conditional on observed characteristics $x = (x_1', x_2')'$:

$$q(d, x, \tau_1, \tau_2)$$

as well as the conditional quantile treatment effects over both dimensions:[11]

$$q(1, x, \tau_1, \tau_2) - q(0, x, \tau_1, \tau_2).$$

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

$$F_{Y|X}(y|x) = \int_0^1 \int_0^1 1\{q(y, x, \tau_1, \tau_2) \leq y\} d\tau_2 d\tau_1, \tag{8}$$

where $F_{Q(\tau_1, y|x)|X} = \int_0^1 1\{q(y, x, \tau_1, \tau_2) \leq y\} d\tau_2$, and $F_{Y|X}(y|x) = \int_0^1 F_{Q(\tau_1, y|x)|X} d\tau_1$. Inverting the cdf yields the one-dimensional 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.

For simplicity, I consider the case where there are no covariates; however, the framework can be easily extended to include other variables.[12] The goal is to select a treatment rule that assigns individuals to treatment depending on their ranks (u_{ij}, v_j).[13] If these ranks were observed, the problem would coincide with the setting considered in Kitagawa and Tetenov (2021).
When a treatment rule G is applied to the target population, the social welfare is proportional to:[14]

$$W_\Lambda(F_G) = \int_0^\infty \Lambda(F_G(y)) dy \tag{9}$$

where $\Lambda(\cdot) : [0,1] \to [0,1]$ is a nonincreasing, nonnegative convex function with $\Lambda(0) = 1$ and $\Lambda(1) = 0$, while $F_G(Y)$ is the distribution of the outcome under treatment rule G:

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

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

$$ATE = \int_0^1 \int_0^1 [q(1, x, \tau_1, \tau_2) - q(0, x, \tau_1, \tau_2)] \, d\tau_2 d\tau_1.$$

[12] 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.

[13] 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.

[14] This social welfare function comprises, for example, the extended Gini family.

Given that $\Lambda(\cdot)$ is convex, we can equivalently write the social welfare function in equation (9) as a weighted average of the outcomes:

$$W_{\Lambda}(F_G) = \int_0^1 F_G^{-1}(\theta) w(\theta) d\theta, \tag{10}$$

where the weights $w(\theta) \equiv \frac{d(1-\Lambda(\theta))}{d\theta}$ depend on the population quantiles.

We want to maximize the social welfare over a class of feasible policies $\mathcal{G} \in \{u_{ij}, v_j(u_{ij}) \in (0,1) \times (0,1)\}$. Hence, the optimal treatment rule solves[15]

$$G^* = \arg\max_{G \in \mathcal{G}} W(G). \tag{11}$$

To make the problem operational, we need to identify individual treatment effects and assign welfare weights to each observation under each policy rule. With rank invariance, $q(1, \tau_1, \tau_2) - q(0, \tau_1, \tau_2)$ is the treatment effect for an individual at quantiles (τ_1, τ_2).[16] Further, using equation (8), we can identify the conditional quantile function of the potential outcomes in the population. These objects can then be used to assign a rank 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_0 share identical welfare weights. Summing the welfare weights within groups provides the weights assigned to group j:[17]

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

While the rank of a group changes over τ_1, these groups' weights are constant over the entire within distribution. Consequently, these weights can offer a welfare-based measure of a group's rank or priority.

To find the optimal treatment assignment rules that maximize 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.[18] I will assume 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}, v_j follow the same distribution in both populations by construction. Therefore, one can equivalently assume that for all u_{ij}, v_j, the joint distribution of $Y_{ij}(1), Y_{ij}(0)|u_{ij}, v_j$ is the same in the sample and target populations. Hence, if the quantile function of the potential outcomes is identified, we can identify individual treatment effects that depend on two different rank variables. These

[15]Solving problem (11) is nontrivial as it lacks a closed-form solution even if we knew the distribution of the 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.

[16]Without a rank stability assumption, individual treatment effects are not identified. Chernozhukov et al. (2023) suggests conditional prediction intervals that can be obtained with a relaxation of this assumption.

[17]In the case of unbalanced groups, larger groups are more likely to have a higher welfare weight. This feature is 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.

[18]Kitagawa and Tetenov (2018) assume that the joint distribution of the potential outcome and covariates is the same in both populations.

can be viewed as counterparts to conditional average treatment effects, so the setting considered in this paper fits the equality-minded treatment choice framework in Kitagawa and Tetenov (2021).

In summary, the two-level conditional quantile function of potential outcomes in the sample population enables the estimation of treatment effects for an individual at a given rank as well as individual ranks. Utilizing this information, we can estimate $Y_{ij}(1)$ for all i and j. Subsequently, for each $G \in \mathcal{G}$, we compute 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 welfare.

4 Estimator

Let $\mathcal{T}$ be the set of quantiles of interest. For simplicity of notation, I consider the same set of quantiles to model both dimensions, although this is not a requirement. I propose a two-step quantile regression estimator to estimate model (6). The first stage consists of group-by-group quantile regressions. For each group j and quantile $\tau_1 \in \mathcal{T}$, the outcome is regressed on the individual level variables x_{1ij} using quantile regression. Then, for each group and each $\tau_1 \in \mathcal{T}$, the fitted values are saved. In the second stage, for each $\tau_1 \in \mathcal{T}$, 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 $\tau_2 \in \mathcal{T}$. Thus, if $\mathcal{T} = \{0.1, 0.2, \ldots, 0.9\}$ there are $9 \times m$ first stage regression and $9 \times 9 = 81$ second stages. Formally, the first-stage quantile regression solves the following minimization problem for each group j and quantile $\tau_1 \in \mathcal{T}$ separately:

$$\hat{\beta}_j(\tau) \equiv \left(\hat{\beta}_{1,j}(\tau_1), \hat{\beta}_{2,j}(\tau_1)'\right)' = \underset{(b_1,b_2)\in\mathbb{R}^{dim(x_1)+1}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \rho_{\tau_1}(y_{ij} - b_1 - x_{1ij}'b_2), \tag{12}$$

where $\rho_\tau(x) = (\tau - 1\{x < 0\})x$ for $x \in \mathbb{R}$ is the check function. For group, 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 $(\tau_1, \tau_2) \in \mathcal{T} \times \mathcal{T}$:

$$\hat{\delta}(\hat{\beta}(\tau_1), \tau_2) = \underset{(a,b,g)\in\mathbb{R}^{dim(x)+1}}{\arg\min} \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} \rho_{\tau_2}(\hat{y}_{ij}(\tau_1) - x_{2j}'g - x_{1ij}'b - a), \tag{13}$$

where the notation makes the dependency on the first step explicit 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 embarrassingly parallelizable, as all first-stage quantile regressions run independently across the groups.

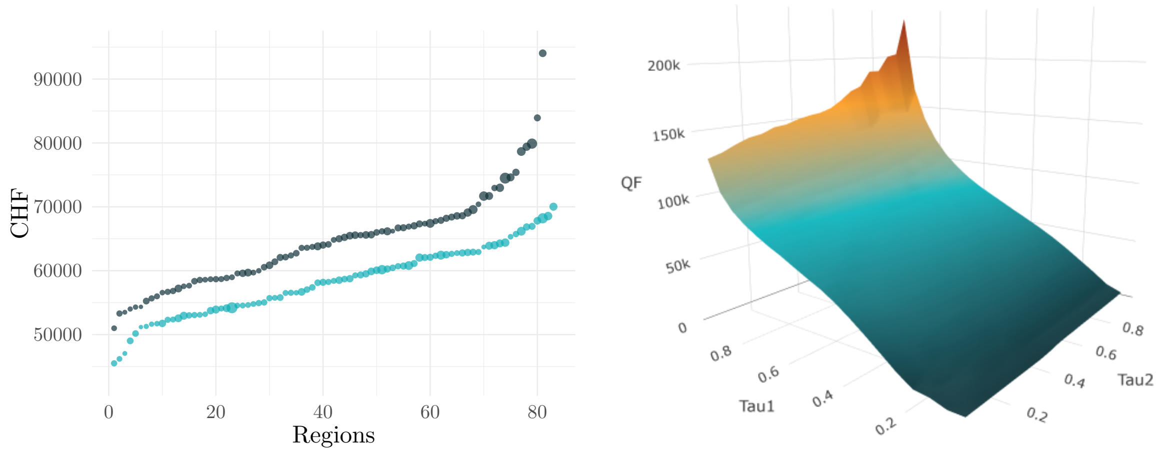
Ensuring the monotonicity of the estimated two-level quantile functions across both dimensions might require a rearrangement operation, as suggested in Chernozhukov et al. (2009, 2010).

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} is exogenous so that quantile regression in both stages yields consistent estimates. It 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).

Remark 3 (**Alternative estimators - distribution regression**)**.** As an alternative to quantile regression, one might specify a model for the distribution function and perform estimation using distribution regression. This, however, has some complications. First, the first stage is well defined only for outcome values that are on the support of the outcome of every j (see Fernández-Val et al., 2022). Second, a similar procedure applied using distribution regression would yield a cdf of a cdf, making the interpretation of the results more complicated.

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 to analyze 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 smaller grid 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, estimation consists in regressing the outcome on a constant separately for each group and quantile τ_1 in the first stage. Subsequently, in the second stage, first-stage fitted values are also regressed on a constant. I consider the set of quantile $\mathcal{T} = \{0.01, 0.02, \ldots, 0.99\}$ in both stages.
Figure 1 shows the regional averages and medians of yearly income on the left and the two-dimensional quantile function of the same variable on the right. Both regional averages and medians are arranged from low to high. The darker dots in Figure 1a reveal substantial differences in average income across regions. However, figure 1b 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 distribution τ_1, the differences across regions are substantially smaller compared to the right tail of the within distribution. Thus, the differences in average wages, as shown in Figure 1a, not only mask substantial within-region income heterogeneity but are predominantly driven by differences in top incomes.



(a) Average Income by Region (b) Two-dimensional quantile function of Income

Notes: Figure 1a show the heterogeneity in average (dark blue) and median (light blue) yearly income across regions defined by 2-digit ZIP codes. Figure 1b shows the two-dimensional quantile function of yearly income within and between regions.

Figure 1: Income Heterogeneity within and between regions

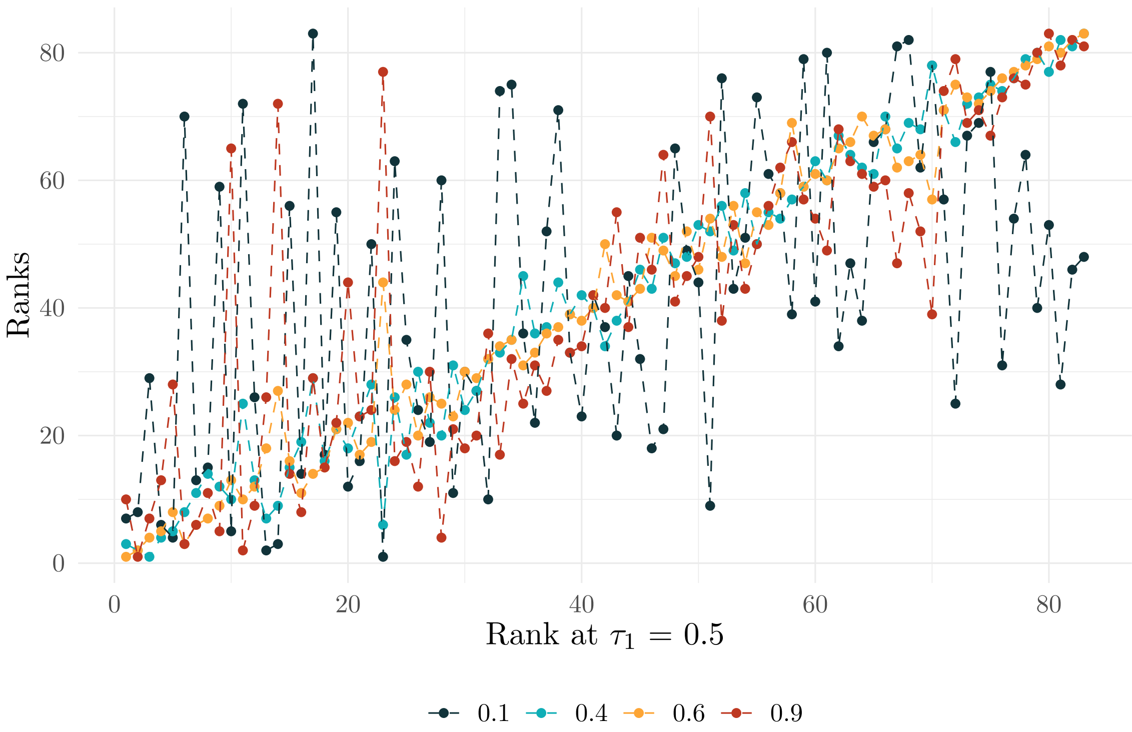
The lighter dots in Figure 1a show that the heterogeneity in median wages across regions is substantially smaller than the heterogeneity in average income. However, this measure solely reflects the heterogeneity at one point of the within distribution, potentially overlooking the labor market situation 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 need to understand how group ranks evolve over the distribution of τ_1. Figure 2 plots the ranks at $\tau_1 = 0.5$ against the ranks at other values of τ_1. As expected, the group ranks at the median are highly correlated with the ranks at $\tau_1 = \{0.4, 0, 6\}$. However, this correlation substantially declines as we move towards the tails, particularly in the lower tail.

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 notational dependency of δ and β_j on τ_1 and τ_2, unless necessary. For a random variable h_{ij}, $\mathbb{E}_{i|j}[h_{ij}]$ is 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 $\tilde{x}_{ij} = (1, x'_{1ij})'$.

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



Notes: The Figure shows the group ranks at different values of τ_1 as a function of their group rank at $\tau_1 = 0.5$.

Figure 2: Evolution of Group Ranks over τ_1

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, 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. 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. (iii) As $m \to \infty$,*

$$\lim_{m\to\infty} \frac{1}{m} \sum_{j=1}^{m} \mathbb{E}_{i|j} \left[f_{Q(\tau_1, y_{ij}|\nu_j)|x_{ij}}(x'_{ij}\delta_0|x) x_{ij} \tilde{x}'_{ij} \right] = \mathbb{E}[f_{Q(\tau_1, y_{ij}|\nu_j)|x_{ij}}(x'_{ij}\delta_0|x) x_{ij} \tilde{x}'_{ij}]$$

where the eigenvalues of $\mathbb{E}[f_{Q(\tau_1, y_{ij}|\nu_j)|x_{ij}}(x'_{ij}\delta_0|x) x_{ij} x'_{ij}]$ are bounded from below and above.

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

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

and

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

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*

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

These are quite 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 (**Conditional distribution II**). *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*

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

and

$$\bar{f}'_Q := \sup_{q\in\mathbb{R}, x\in\mathcal{X}} |f'_{Q(\tau_1, y_{ij}|x_{ij}, v_j)|x_{ij}}(q|x)| < \infty.$$

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*

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

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* $\mathbb{R}^{K_1}$ *and* $\mathbb{R}^K$, *respectively.*

Assumptions 5-6 are the second stage counterpart of of assumptions 3-4, 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 *continously* differentiable. This additional assumption on the distribution of the second stage dependent variable is sufficient to ensure that its second derivative is Lipschitz continous. 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); Zhang et al. (2019). Compactness of $\mathcal{D}$ is necessary to use the results in Chen et al. (2003). Whereas compactness of $\mathcal{B}_j$ is useful as it directly implies that the covering integral is finite but could easily be relaxed.

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 8 (**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 8(a) is sufficient for consistency of the estimator. While asymptotic normality requires the stronger assumption 8(b). The first result of this paper states consistency of the two-step estimator.

Theorem 1 (Consistency). *Let assumptions 1-7 and 8(a) be satisfied. Then,* $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau) \xrightarrow{p} 0$.

To establish asymptotic normality, I start by showing that $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)$ can be approximated by the sum of two terms that account for estimation error arising at different steps of the estimation. If the first stage parameter vector $\beta_0(\tau_1) = (\beta_{0,1}(\tau_1)', \ldots, \beta_{0,m}(\tau_1)')'$ were known, the true second-stage parameter vector $\delta_0(\beta_0, \tau)$ uniquely[19] satisfies:

$$\mathbb{E}[m(\delta_0, \beta_0, \tau)] = 0 \quad (14)$$

[19]Under weak regularity conditions.

with

$$m(\delta,\beta,\tau) = x'_{ij}[\tau_2 - 1(\tilde{x}'_{ij}\beta_j(\tau_1) \leq x'_{ij}\delta(\beta,\tau))]. \tag{15}$$

Let $M(\delta,\beta,\tau) = \mathbb{E}[m(\delta,\beta,\tau)]$ and denote the sample counterpart $M_{mn}(\delta,\beta,\tau) = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} m(\delta,\beta,\tau)$. 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 establish asymptotic normality. (i) The sample moment evaluated at the true parameters:

$$M_{mn}(\delta_0,\beta_0,\tau) = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} m(\delta_0,\beta_0,\tau), \tag{16}$$

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

$$\Gamma_2(\delta,\beta_0,\tau)[\beta-\beta_0] = \frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2j}(\delta,\beta_0,\tau)[\beta_j-\beta_{j,0}], \tag{17}$$

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 $\frac{1}{m}\bar{\Gamma}_{2j}(\tau,\delta,\beta_0)\equiv\Gamma_{2j}(\tau,\delta,\beta_0)$ with

$$\Gamma_{2j}(\delta,\beta_0,\tau) = \frac{\partial}{\partial\beta_j}M(\delta,\beta_0,\tau) = -\frac{1}{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x'_{ij}\delta(\beta_0,\tau)|x_i)x_{ij}\tilde{x}'_{ij}\right]. \tag{18}$$

The expression in equation (16) is directly related to the leading term of a Bahadur expansion of the unfeasible estimator $\hat{\delta}(\beta_0,\tau)$:

$$\hat{\delta}(\beta_0,\tau)-\delta(\beta_0,\tau) = \Gamma_1(\delta_0,\beta_0,\tau)^{-1}\cdot\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} m(\delta_0,\beta_0,\tau)+o_p(1) \tag{19}$$

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 (16) captures the estimation error that would arise due to random variation in the second stage if we knew the true first stage and equation (17) 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 $\Gamma_1(\delta_0,\beta_0,\tau)^{-1}\sqrt{m}\big(\sum_{j=1}^{m}\Gamma_{2j}(\delta,\beta_0,\tau)\ [\hat{\beta}_j(\tau_1)-\beta_{j,0}(\tau_1)] + M_{mn}(\delta_0,\beta_0,\tau)\big)$. 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

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

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

The following Lemma establishes the asymptotic properties of equations (17) and (16).

Lemma 1. *Let the model in equation (6) and assumptions 1-7 hold. Then* *(i) Under assumptions 8(b):*

$$\sqrt{m}\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2,j}(\delta_0,\beta_0,\tau)\left(\hat{\beta}_j(\tau_1)-\beta_{j,0}(\tau_1)\right)\xrightarrow{d}N\left(0,\Omega_1(\tau)/n\right),$$

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

(ii) Assume that $Pr(\tilde{x}_{ij}'\beta_{j,0}(\tau_1)\leq x_{ij}'\delta_0(\beta_0,\tau))-\tau_2>\epsilon>0$ *for all* (τ_1,τ_2)*, then*

$$\sqrt{m}\left(M_{mn}(\delta_0,\beta_0,\tau)\right)\xrightarrow{d}N\left(0,\Omega_2(\tau)\right), \tag{20}$$

where $\Omega_2(\tau)=\mathbb{E}\left[[\tau_2-1(\tilde{x}_{ij}'\beta_{j,0}(\tau_1)\leq x_{ij}'\delta_0(\beta_0,\tau)]^2x_{ij}x_{ij}'\right]=\tau_2(1-\tau_2)\mathbb{E}[x_{ij}x_{ij}']$.

(iii) Under assumption 8(b):

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

Hence, the asymptotic distribution of the estimator can be approximated by a linear function of the sum of two terms converging at a different rate, and the asymptotic behavior will be determined by the term converging at the slower rate.

Theorem 2. *Let assumptions 1-7 and 8(b) be satisfied. Assume that for all* τ_1,τ_2*,* $Pr(\tilde{x}_{ij}'\beta_{j,0}(\tau_1)\leq x_{ij}'\delta_0(\tau))-\tau_2>\epsilon>0$*. Then*

$$\sqrt{m}\left(\hat{\delta}(\hat{\beta},\tau)-\delta_0(\beta_0,\tau)\right)\xrightarrow{d}N(0,\Gamma_1^{-1}\Omega(\tau)\Gamma_1'^{-1}) \tag{21}$$

with $\Gamma_1=\Gamma_1(\delta_0,\beta_0,\tau)$ *and* $\Omega=\frac{\Omega_1}{n}+\Omega_2$*, where* Ω_1 *and* Ω_2 *are defined in Lemma 1.*

The variance of the component capturing the first-stage error shrinks to zero at a rate of $1/n$ and, therefore, does not show up in the first-order asymptotic distribution. To improve finite sample inference, I suggest below a covariance matrix estimator that takes the first stage error into account.

The requirement that $Pr(\tilde{x}_{ij}'\beta_{j,0}(\tau_1)\leq x_{ij}'\delta_0(\tau))-\tau_2>\epsilon>0$ for all (τ_1,τ_2) ensures that there is heterogeneity between groups. In the follwoing Remark 4, I discuss a special case where $Pr(\tilde{x}_{ij}'\beta_{j,0}(\tau_1)\leq x_{ij}'\delta_0(\tau))=\tau_2$ for all j.

Remark 4 (Degree of heterogeneity and growth condition)**.** If there is no heterogeneity between group – that is, if $Pr(\tilde{x}_{ij}'\beta_{j,0}(\tau_1)\leq x_{ij}'\delta_0(\tau))=\tau_2$ for a given (τ_1,τ_2)– it should be possible to show that $\sqrt{nm}\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2,j}(\delta_0,\beta_0,\tau)\left(\hat{\beta}_j(\tau_1)-\beta_{j,0}(\tau_1)\right)\xrightarrow{d}N\left(0,\Omega_1(\tau)\right)$ under the stronger requirement that $\frac{\log(n)^2m}{n}\to 0$. Since in this case $M_{mn}(\delta_0,\beta_0,\tau)=0$, the first stage error dominates the asymptotic distribution and the estimator converges at the $\sqrt{mn}$-rate. This stronger growth condition is used, for example, in Galvao et al. (2020) quantile regression fixed effect estimators, and is required in Melly and Pons (2022) for the adaptive results in the degree of heterogeneity and convergence rates. In this paper, given the central interest in analyzing the heterogeneity between groups, I do not explore this possibility.

5.2 Inference

To perform inference, I suggest a clustered bootstrap procedure where entire groups are resampled with replacement. 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. More precisely, for each bootstrap replication $b = 1, \ldots, B$, draw a random sample with replacement $\{(\hat{y}^*_{1j}, \ldots, y^*_{nj}), (x^*_{1j}, \ldots, x^*_{nj}) : j = 1, \ldots, m\}$ from $\{(\hat{y}_{1j}, \ldots, \hat{y}_{nj}), (x_{1j}, \ldots, x_{nj}) : j = 1, \ldots, m\}$, and run the second step estimator using the resampled data. I show that the asymptotic distribution of $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)$ can be approximated with the distribution of $\hat{\delta}^*(\hat{\beta}, \tau) - \hat{\delta}(\hat{\beta}, \tau)$ and that this procedure also takes into account the first-stage error.

Theorem 3. *Assume that the condition for Theorem 2 are satisfied. Then,*

$$\sqrt{m}\left(\hat{\delta}^*(\hat{\beta}, \tau) - \hat{\delta}(\hat{\beta}, \tau)\right) \xrightarrow{d^*} N\left(0, \Gamma_1^{-1}\Omega(\tau)\Gamma_1^{-1}\right).$$

Remark 4 discusses the impact of heterogeneity on the rate of convergence of the estimator. In similar situations, Liao and Yang (2018); Lu and Su (2022); Fernández-Val et al. (2022) show that the clustered bootstrap is uniformly valid in the rate of convergence of the estimator.[20] Simulations show similar results for the estimator in this paper, hence providing some anecdotal evidence that the bootstrap standard errors might also be adaptive in this case. In a similar setting, Melly and Pons (2022), use a clustered covariance matrix estimator in the second stage. However, based on simulations, the bootstrap method appears to outperform the clustered covariance matrix estimator with this estimator. Consequently, I recommend using the bootstrap in this paper.

6 Simulations

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

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

with $x_{1ij} = 1 + h_j + w_{ij}$, where $h_j \sim U[0, 1]$, $w_{ij}, x_{2j}, \eta_j, \nu_{ij}$ are $N(0, 1)$. This is a location-scale-shift model over both quantile indices. Let $\beta = \gamma = 1$. The true coefficients on the individual-level variable take the form $\beta(\tau_1, \tau_2) = 1 + 0.1 \cdot F^{-1}(\tau_1) + 0.1 \cdot F^{-1}(\tau_2)$ and the true coefficients on the group-level regressor equals $\gamma(\tau_1, \tau_2) = 1 - 0.1 \cdot F^{-1}(\tau_1) - 0.1 \cdot F^{-1}(\tau_2)$ where F is the standard normal cdf. I consider the sample sizes $(m, n) = \{(25, 25), (200, 25), (25, 200), (200, 200)\}$ and focus on the set of quantiles $\mathcal{T} \in \{0.25, 0.5, 0.75\}$ using 2'000 Monte Carlo simulations.

[20] These papers refer to this bootstrap procedure as cross-sectional bootstrap since the focus is on panel data models. See Melly and Pons (2022) for a similar result using clustered standard errors.

$\tau_1 \setminus \tau_2$	β 0.25	β 0.5	β 0.75	γ 0.25	γ 0.5	γ 0.75
			(m, n) = (25,25)			
0.25	-0.026	0.001	0.031	-0.019	0.004	0.028
	(0.114)	(0.109)	(0.118)	(0.237)	(0.223)	(0.243)
0.5	-0.027	-0.004	0.023	-0.020	0.000	0.023
	(0.112)	(0.104)	(0.109)	(0.241)	(0.219)	(0.240)
0.75	-0.034	-0.006	0.022	-0.020	-0.002	0.026
	(0.116)	(0.109)	(0.114)	(0.239)	(0.220)	(0.241)
			(m, n) = (25,200)			
0.25	-0.010	-0.001	0.007	-0.008	0.000	0.007
	(0.074)	(0.066)	(0.072)	(0.234)	(0.219)	(0.230)
0.5	-0.008	-0.002	0.004	-0.008	-0.001	0.004
	(0.072)	(0.065)	(0.069)	(0.234)	(0.221)	(0.231)
0.75	-0.012	-0.003	0.005	-0.010	-0.002	0.004
	(0.074)	(0.067)	(0.070)	(0.235)	(0.219)	(0.231)
			(m, n) = (200,25)			
0.25	-0.022	0.006	0.031	-0.022	-0.001	0.021
	(0.042)	(0.038)	(0.042)	(0.079)	(0.073)	(0.079)
0.5	-0.025	-0.001	0.023	-0.020	-0.002	0.017
	(0.041)	(0.037)	(0.039)	(0.078)	(0.073)	(0.078)
0.75	-0.033	-0.007	0.020	-0.023	-0.003	0.018
	(0.042)	(0.038)	(0.041)	(0.079)	(0.074)	(0.081)
			(m, n) = (200,200)			
0.25	-0.005	0.002	0.007	-0.004	0.000	0.006
	(0.028)	(0.026)	(0.028)	(0.076)	(0.070)	(0.078)
0.5	-0.004	0.001	0.005	-0.003	0.000	0.006
	(0.027)	(0.025)	(0.028)	(0.075)	(0.070)	(0.079)
0.75	-0.006	0.000	0.006	-0.004	0.000	0.006
	(0.027)	(0.026)	(0.028)	(0.076)	(0.070)	(0.078)

Note:
Results based on 2000 Monte Carlo simulations. The table provides standard errors relative to standard deviation.

Table 1: Bias and Standard Deviation

Table 1 shows the bias and standard deviation. Table 2 shows the bootstrap standard errors relative to the standard deviation, and Table 3 shows the coverage probability of the 95% confidence interval. Bootstrap standard errors are computed using 200 repetitions.

While β and γ have the same asymptotic behavior, we see differences in their finite sample properties. The simulations show that the bias of β decreases both as n or m increases, while the bias of γ decreases only when m increases. Similarly, the variance of γ is only minimally affected by an increase in the number of observations per group. On the other hand, the variance of β shows a larger improvement when the n increases. Still, an increase in the number of groups yields the largest improvement in the variance.[21]

Table 2 shows the bootstrap standard errors relative to the standard deviation. The simulation shows that the bootstrap standard errors are conservative. The standard errors are

[21] As n becomes larger, further increases n will not improve the variance of either β nor γ. More precisely, the relatively large decrease in the variance β here quickly converges to zero as n increases.

$\tau_1 \setminus \tau_2$	β 0.25	0.5	0.75	γ 0.25	0.5	0.75
			(m, n) = (25,25)			
0.25	1.203	1.114	1.204	1.114	1.088	1.264
0.5	1.207	1.140	1.202	1.138	1.085	1.295
0.75	1.184	1.115	1.229	1.127	1.077	1.267
			(m, n) = (25,200)			
0.25	1.249	1.206	1.350	1.251	1.122	1.553
0.5	1.314	1.216	1.439	1.292	1.126	1.651
0.75	1.330	1.172	1.386	1.324	1.119	1.593
			(m, n) = (200,25)			
0.25	1.054	1.025	1.019	1.035	1.029	1.019
0.5	1.036	1.022	1.015	1.003	1.017	1.025
0.75	1.018	1.012	1.033	1.005	0.998	1.021
			(m, n) = (200,200)			
0.25	1.075	1.033	1.059	1.033	1.078	1.053
0.5	1.069	1.065	1.062	1.022	1.078	1.052
0.75	1.067	1.070	1.046	1.030	1.068	1.059

Note:
Results based on 2000 Monte Carlo simulations. The table provides standard errors relative to standard deviation.

Table 2: Bootstrap Standard Errors relative to Standard Deviation

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 3 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 larger bias. In the appendix C, I include simulation results for the standard errors and coverage probability computed using clustered standard errors in the second stage. The clustered standard errors are smaller than the bootstrap standard errors, and consequently, the confidence intervals have lower coverage.

7 Empirical Application

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 designated as treatment markets, and the remaining 64 serve as control markets. In the second stage, individual-level randomization assigns firms in the treatment markets to training randomly. Randomization is stratified by geographical region, market size, and quartiles of

$\tau_1 \setminus \tau_2$	β			γ		
	0.25	0.5	0.75	0.25	0.5	0.75
(m, n) = (25,25)						
0.25	0.972	0.968	0.963	0.933	0.950	0.938
0.5	0.976	0.973	0.966	0.931	0.947	0.947
0.75	0.969	0.967	0.969	0.939	0.943	0.935
(m, n) = (25,200)						
0.25	0.987	0.986	0.985	0.946	0.954	0.960
0.5	0.984	0.982	0.986	0.951	0.953	0.959
0.75	0.986	0.986	0.982	0.949	0.952	0.954
(m, n) = (200,25)						
0.25	0.925	0.950	0.888	0.940	0.935	0.926
0.5	0.912	0.949	0.904	0.929	0.941	0.936
0.75	0.881	0.944	0.921	0.924	0.929	0.925
(m, n) = (200,200)						
0.25	0.956	0.953	0.947	0.939	0.949	0.943
0.5	0.952	0.962	0.953	0.944	0.945	0.942
0.75	0.946	0.960	0.956	0.945	0.952	0.950

Note:
Results based on 2000 Monte Carlo simulations. The table provides the coverage probability of the 95 confidence invervals.

Table 3: Coverage Probability of Bootstrap Confidence Interval

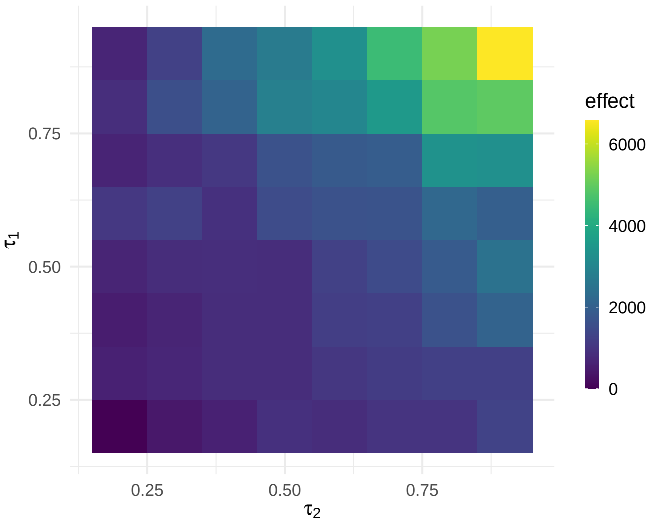


Figure 3: Effect of training Assignment on Weekly Sales

weekly profits to ensure a balanced sample. This results in 1,172 individual firms assigned to training and 988 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

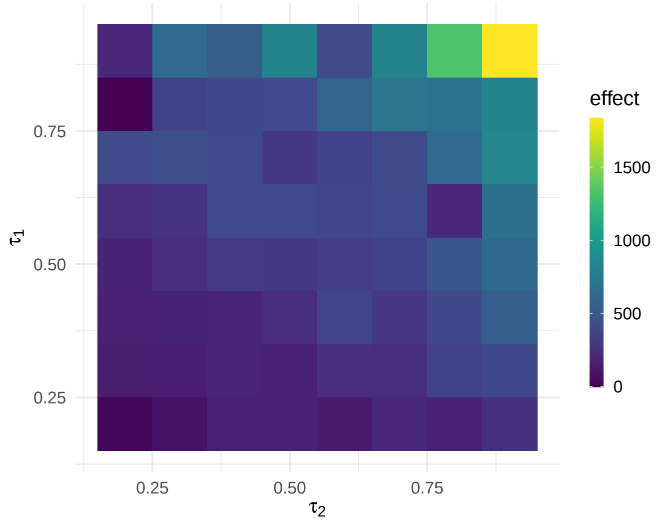


Figure 4: Effect of Training Assignment on Weekly Profits

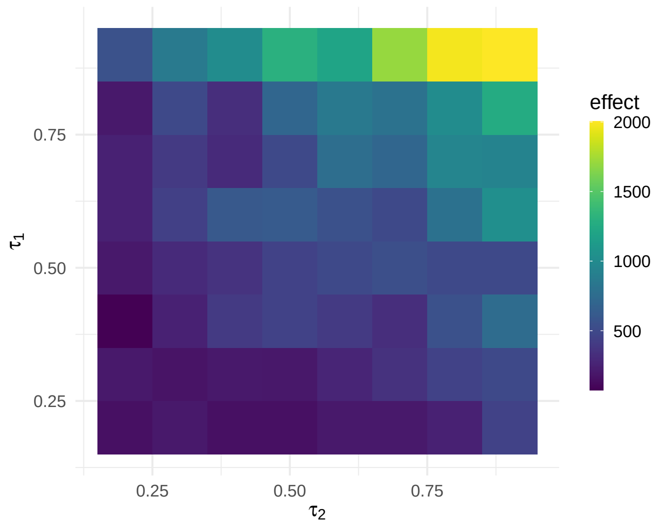


Figure 5: Effect of Training Assignment on Income from Work

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 weekly average 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 markets not assigned to the program remain unclear, with point estimates being small and not statistically significant. 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. Their findings indicate larger effects in the upper tail of the outcome distribution.

In my analysis, I use data from the same two waves and define groups based on markets. To ensure that I have enough observations for the estimation, I drop markets with fewer than

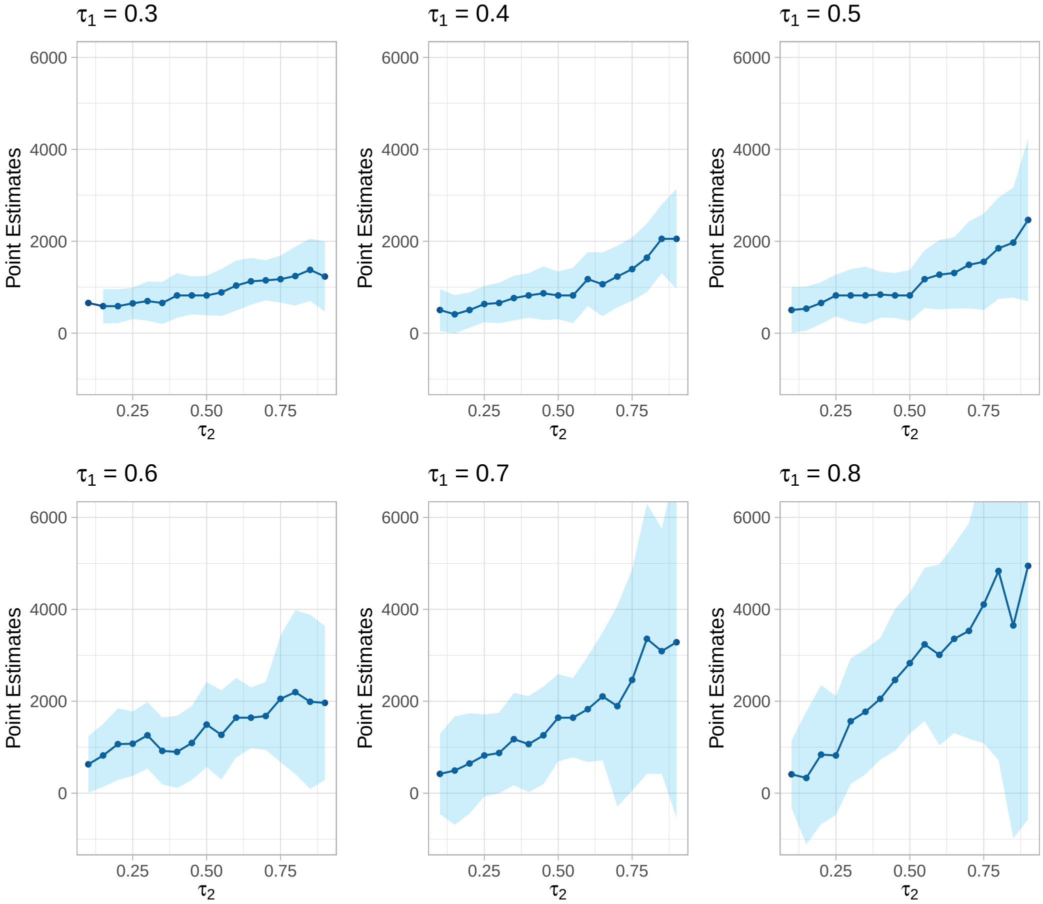


Figure 6: Effect of Training Assignment on Sales

25 businesses.[22] The final dataset includes 5,773 observations, corresponding to 3,032 firms operating in 124 markets. On average, there are 52 observations per market. I consider three different outcome variables: weekly sales, weekly profits, and income from work with averages of 6,100, 1,500, and 2,300 Kenyan shillings.[23]

All three variables have mass points at zero. One reason is that these variables are recorded 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 18% of the observations. Further, 14% of people report having no income. Due to this censoring issue, I refrain from computing quantile regression 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.[24] Further away from the lower tail, this problem does not affect the results.

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

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

[24] 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.

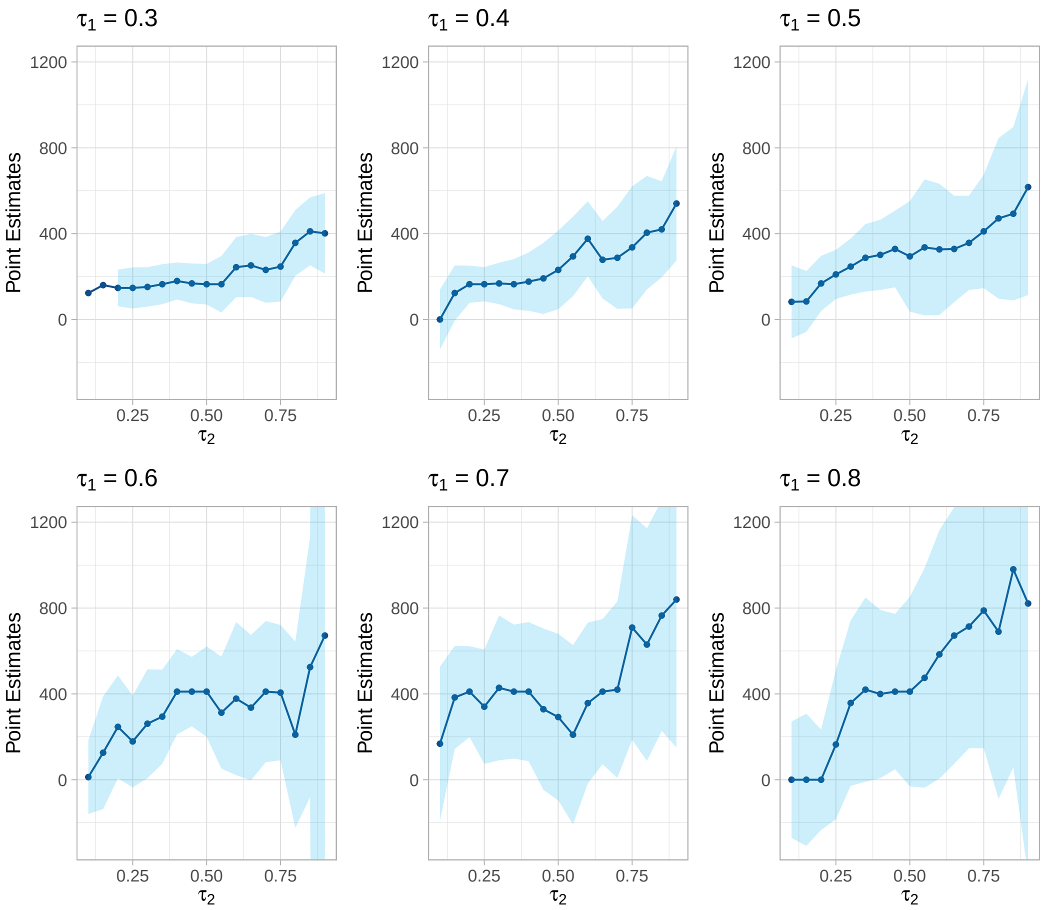


Figure 7: Effect of Training Assignment on Profits

I estimate the following model:

$$Q(\tau_2, Q(\tau_1, y_{ijt}|d_{ij}, s_{ij}, w_{5,ij}, v_j)|d_{ij}, s_{ij}, w_{5,ijt}) = \beta_1(\tau)\cdot d_{ij} + \beta_2(\tau)\cdot s_{ij} + \beta_3(\tau)\cdot w_{5,ijt} + \alpha(\tau), \quad (22)$$

where y_{ijt} is the outcome of firm i operating in market j in wave t, d_{ij} is the treatment dummy, 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, and $w_{5,ijt}$ is an indicator variable for the last wave.

Figures 3-5 show the treatment effects estimates for sales, profits, and income from work over the two dimensions for the quantiles indices $\{0.2, 0.3, \ldots, 0.9\}$. Figure Figures 6-8 plot the point estimates and confidence intervals over the distribution of markets τ_2 when fixing the within-market rank τ_1. The results for all three variables show a similar pattern, where both within-group and between-group inequalities play essential roles, resulting in larger positive treatment effects in the upper tail of their respective distributions. For instance, at $\tau_1 = 0.5$, the effect on profits increases from 100 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 at $\tau_1 = 0.3$ to 800 at

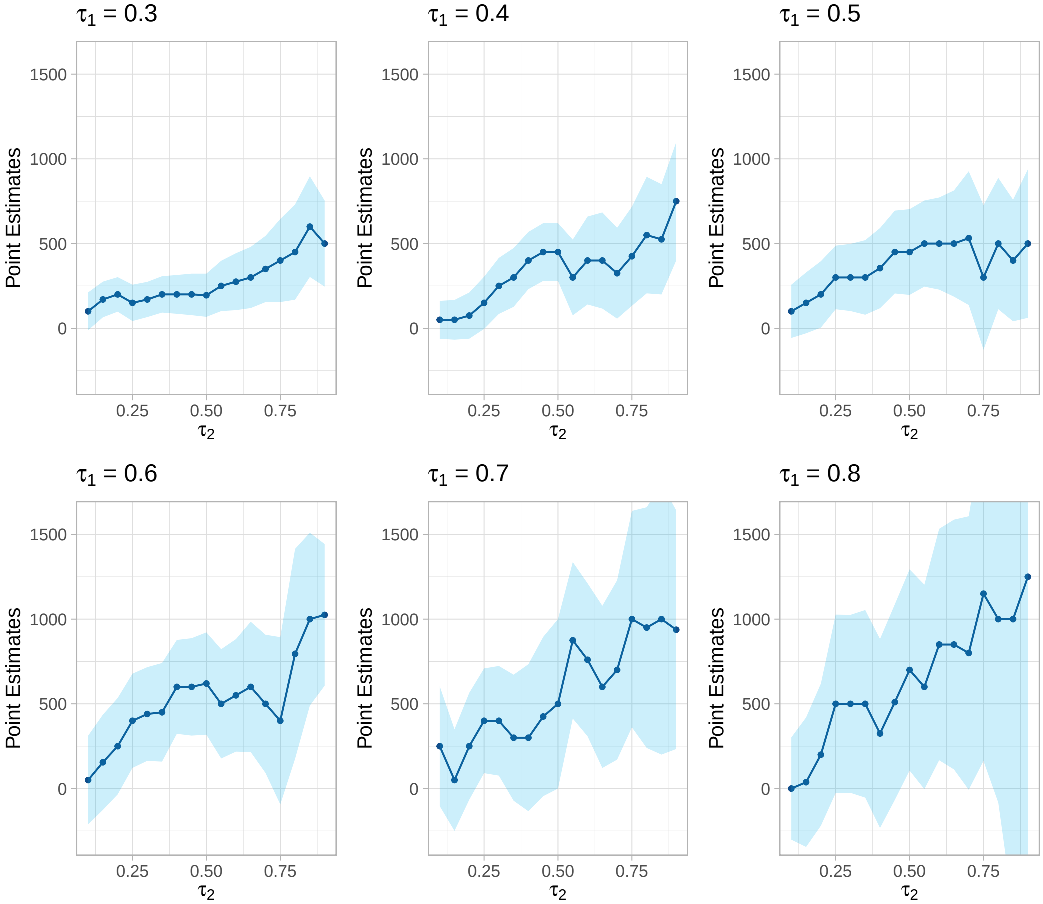


Figure 8: Effect of Training Assignment on Income From Work

$\tau_1 = 0.7$.

	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.2	1							
0.3	0.81	1						
0.4	0.67	0.81	1					
0.5	0.61	0.76	0.87	1				
0.6	0.53	0.64	0.78	0.86	1			
0.7	0.45	0.54	0.69	0.72	0.82	1		
0.8	0.35	0.44	0.58	0.62	0.69	0.8	1	
0.9	0.22	0.32	0.48	0.5	0.56	0.6	0.73	1

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

Table 4: Correlation of Ranks over τ_1

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 4 presents the rank correlation across the τ_1 dimension. The table provides the correlation of the 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. A lower correlation suggests that we would rank groups differently at different points of the within distribution, indicating a variation in the notion of poor-performing market vs. good-performing market across τ_1. The correlation matrix shows that the ranks vary across the τ_1, however, the rank at $\tau_1 = 0.5$ is still highly predictive of the ranks at different points in the distribution.

8 Conclusion

Distributional effects are particularly interesting in analyzing treatment effect heterogeneity. In economics, heterogeneity manifests itself across various dimensions, encompassing within and between groups. This paper aims to provide a method to analyze heterogeneity and distributional effects within and between groups simultaneously. 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 tradeoff between a simple model with a unique group rank and a more flexible model that allows for unrestricted heterogeneity between groups but requires that group ranks can change over the within distribution. This paper follows the second approach as it offers a more realistic model.

I suggest a two-step quantile regression estimator with within-group by-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 grows to infinity. I show that the estimator can provide new insights about two-dimensional heterogeneity in grouped data. In a descriptive illustration, I study income heterogeneity within and between the labor market 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. Furthermore, the data show that group ranks change substantially over the within distribution, suggesting that comparing differences in median wage between regions does 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. I find stronger effects for good-performing firms (in their markets) operating in thriving markets, indicating that there might be complementarities between individual and group ranks.

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.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework," *Econometrica*, 85, 693–734.

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

Athey, S. and G. W. Imbens (2006): "Identification and inference in nonlinear difference-in-differences models," *Econometrica*, 74, 431–497.

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.

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

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

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.

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., 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. 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.

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.

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.

Firpo, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75, 259–276.

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.

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

Honoré, B., S. Khan, and J. L. Powell (2002): "Quantile regression under random censoring," *Journal of Econometrics*, 109, 67–105.

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.

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*.

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.

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 (2022): "Minimum Distance Estimation of Quantile Panel Data Models," *Working Paper*.

Neumark, D. and H. Simpson (2015): "Place-Based Policies," in *Handbook of Regional and Urban Economics*, Elsevier B.V., vol. 5, chap. 18, 1197–1287, 1 ed.

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.

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

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

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 $\hat{\beta}_j(\tau)$)**.** *Under Assumptions 1-4, 7 and 8(a), we have*

$$\max_{1\le j\le m} \|\hat{\beta}_j(\tau) - \beta_j\| = o_p(1).$$

Proof. The proof follows directly by the proof of Lemma 1 in Galvao and Wang (2015). ∎

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

$$\hat{\beta}_j(\tau_1) - \beta_j(\tau_1) = \frac{1}{n}\sum_{i=1}^{n} \phi_{j,\tau_1}(\tilde{x}_{ij}, y_{ij}) + R_{nj}^{(1)}(\tau_1) + R_{nj}^{(2)}(\tau_1), \tag{23}$$

where

$$\phi_{j,\tau_1}(\tilde{x}_{ij}, y_{ij}) = -B_{j,\tau_1}^{-1}\tilde{x}_{ij}(1(y_{ij} \le \tilde{x}_{ij}\beta_j(\tau_1)) - \tau_1), \tag{24}$$

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*

$$\sup_j \sup_{\tau_1\in\mathcal{T}} \left\| R_{nj}^{(2)}(\tau_1)\right\| = O_p\left(\frac{\log n}{n}\right) \tag{25}$$

$$\sup_j \sup_{\tau_1\in\mathcal{T}} \left\| \mathbb{E}_{i|j}\left[R_{nj}^{(1)}(\tau_1)\right]\right\| = O\left(\frac{\log n}{n}\right) \tag{26}$$

$$\sup_j \sup_{\tau\in\mathcal{T}} \left\| \mathbb{E}\left[\left(R_{nj}^{(1)}(\tau_1) - \mathbb{E}_{i|j}[R_{nj}^{(1)}(\tau_1)]\right)\left(R_{nj}^{(1)}(\tau) - \mathbb{E}_{i|j}[R_{nj}^{(1)}(\tau_1)]\right)'\right]\right\| = O\left(\left(\frac{\log n}{n}\right)^{3/2}\right). \tag{27}$$

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

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

$$\sup_{||\beta-\beta_0||_{\mathcal{B}}\le\zeta_m,\ ||\delta-\delta_0||\le\zeta_m} ||M_{mn}(\delta,\beta,\tau) - M(\delta,\beta,\tau) - M_{mn}(\delta_0,\beta_0,\tau)|| = o_p(m^{-1/2}), \tag{28}$$

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 $m(\delta,\beta,\tau_2) = \rho_{\tau_2}(\tilde{x}_{ij}\beta_j - x_{ij}\delta)$. By the properties of the check function $\rho_{\tau_2}(y+z) - \rho_{\tau_2}(y) \le 2\cdot||z||$. Hence,

$$\begin{aligned}
m(\delta',\beta',\tau) &- m(\delta'',\beta'',\tau)\\
=&\rho_{\tau_2}(\tilde{x}_{ij}\beta'_j - x_{ij}\delta') - \rho_{\tau_2}(\tilde{x}_{ij}\beta''_j - x_{ij}\delta') + \rho_{\tau_2}(\tilde{x}_{ij}\beta''_j - x_{ij}\delta') - \rho_{\tau_2}(\tilde{x}_{ij}\beta''_j - x_{ij}\delta'')\\
\le& 2||\tilde{x}_{ij}(\beta'_j - \beta''_j)||_{\mathcal{B}} + 2||x_{ij}(\delta' - \delta'')||.
\end{aligned} \tag{29}$$

It follows that $m_{mn}(\delta, \beta, \tau_2)$ is stochastic equicontinous because

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

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

Condition (3.2) is satisfied as, for $k = 1, 2, \ldots, K$,

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

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 Resutls

B.1 Consistency and Asymptotic Normality

Proof of Theorem 1. 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-continous 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. **Part (i)** Inserting equation (23) from Lemma 3 in equation (17) gives:

$$\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2j}(\delta,\beta_0,\tau)(\hat{\beta}_j-\beta_{j,0}) \tag{30}$$

$$=\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,\tau_1}(\tilde{x}_{ij},y_{ij})\right) \tag{31}$$

$$+\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(1)}(\tau_1) \tag{32}$$

$$+\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(2)}(\tau_1) \tag{33}$$

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. It follows directly that the conditional expectation $\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta|x_{ij})x_{ij}\tilde{x}_{ij}'\right]$ is bounded. Next, consider the third term (33). Together with equation (25), it implies that

$$\sup_{\tau_1\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(2)}(\tau_1)=O_p\left(\frac{\log n}{n}\right). \tag{34}$$

For the second term (32), Since $\text{Var}\left(R_{nj}^{(1)}(\tau)\right)=o\left(\frac{1}{n}\right)$ by (27), the conditional expectation is bounded and since the observations are independent across groups, we have that

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

In addition, by (26), $\sup_j\sup_{\tau_1\in\mathcal{T}}\mathbb{E}_{i|j}\left[R_{nj}^{(1)}(\tau_1)\right]=O\left(\frac{\log n}{n}\right)$ such that

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

Putting this together, by the Chebyshev inequality and under Assumption 8(b), we have that

$$\sup_{\tau_1\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m}\mathbb{E}_{i|j}\left[f_{Q(\tau_1,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(1)}(\tau_1)=o_p\left(\frac{1}{\sqrt{m}}\right). \tag{35}$$

It follows that both (32) and (33) are $o_p\left(\frac{1}{\sqrt{m}}\right)$.

Consider now the first term (31):

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

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

$$:=\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}s_{ij}(\tau).$$

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

$$\begin{aligned}\text{Var}(s_{ij}(\tau_1)) =&\mathbb{E}_j\left[\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)\,\text{Var}(\phi_{j,\tau_1})\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)'\right]\\ =&\mathbb{E}_j[\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)B_{j,\tau}^{-1}\tau(1-\tau)\mathbb{E}_{i|j}[\tilde{x}_{ij}\tilde{x}_{ij}']B_{j,\tau}^{-1}\bar{\Gamma}_{2j}(\tau,\delta_0,\beta_0)'],\end{aligned} \tag{36}$$

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

$$\sqrt{m}\frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)(\hat{\beta}_j-\beta_{j,0}) \xrightarrow{d} N\left(0,\frac{\Omega_1}{n}\right), \tag{37}$$

where $\Omega_1 = \mathbb{E}_j\left[\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)\text{Var}(\phi_{j,\tau})\bar{\Gamma}_{2j}(\delta_0,\beta_0,\tau)'\right]$

Part (ii)

$$M_{mn}(\delta_0,\beta_0,\tau) = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}x_{ij}'[\tau_2 - 1(\tilde{x}_{ij}'\beta_{j,0}(\tau_1) \leq x_{ij}'\delta_0(\beta_0,\tau)]$$

It follows directly that

$$\sqrt{m}\left(M_{mn}(\delta_0,\beta_0,\tau)\right) \xrightarrow{d} N\left(0,\Omega_2(\tau)\right) \tag{38}$$

where $\Omega_2(\tau) = \mathbb{E}\left[[\tau_2 - 1(\tilde{x}_{ij}'\beta_{j,0}(\tau_1) \leq x_{ij}'\delta_0(\beta_0,\tau)]^2 x_{ij}x_{ij}'\right] = \tau_2(1-\tau_2)\mathbb{E}[x_{ij}x_{ij}']$ where the last equality follows by correct specification.

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 $o_p\left(\frac{1}{\sqrt{mn}}\right)$ term)

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

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

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

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

This implies that

$$\text{Cov}\left(M_{mn}(z,\delta_0,\beta_0,\tau_2), \frac{1}{m}\sum_{j=1}^{m}\bar{\Gamma}_{2,j}(\beta_0,\tau)\left(\hat{\beta}_j-\beta_{j,0}\right)\right) = o_p\left(\frac{1}{\sqrt{mn}}\right)$$

■

Proof of Theorem 2. 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 8(b) implies 8(a) so that by Theorem 1, $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau) \xrightarrow{p} 0$. Therefore, following Chen et al. (2003), I can replace the parameter space with a small or shrinking ball around the true parameter. Let $\mathcal{D}_{\zeta,\tau} := \{\delta \in \mathcal{D} : ||\delta - \delta_0(\tau)|| \leq \zeta\}$ and $\mathcal{B}_{\zeta,\tau_1} := \{\beta \in \mathcal{B} : ||\beta - \beta_0(\tau_1)|| \leq \zeta\}$.

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

$$\Gamma_1(\delta, \beta_0, \tau_2) := \frac{\partial M(\delta, \beta_0, \tau_2)}{\partial \delta} = \mathbb{E}[f_{Q(\tau_1, y_{ij}|x_{ij}, v_j)|x_{ij}}(x'_{ij}\delta | x_{ij}, \beta) x_{ij} x'_{ij}]. \tag{39}$$

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

Denote $\Gamma_2(\delta, \beta_0)[\beta - \beta_0] = \sum_{j=1}^m \Gamma_{2j}(\delta, \beta_0)[\beta_j - \beta_{j,0}]$ the pathwise derivative of $M(\delta, \beta_0)$ in the direction $(\beta - \beta_0)$, 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) $||M(\delta, \beta, \tau) - M(\delta, \beta_0) - \sum_{j=1}^m \Gamma_{2j}(\delta, \beta_0, \tau)[\beta_j - \beta_{j,0}]|| \leq c \cdot \sup_j ||\beta_j - \beta_{j,0}||^2_{\mathcal{B}}$ for some constant $c \geq 0$; and (ii) $|| \sum_{j=1}^m \Gamma_{2j}(\delta, \beta_0, \tau)[\beta_j - \beta_{j,0}] - \sum_{j=1}^m \Gamma_{2j}(\delta_0, \beta_0, \tau)[\beta_j - \beta_{j,0}]|| \leq o(1)\zeta_m$.
The 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

$$M(\delta, \beta, \tau) - M(\delta, \beta_0, \tau) = \sum_{j=1}^m \Gamma_{2j}(\delta, \beta_0, \tau)[\beta_j - \beta_{j,0}] + O_p(||\beta - \beta_0||^2_{\mathcal{B}}) \tag{40}$$

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 8(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 and Condition (2.6) follows directly by Lemma 1, with

$$\sqrt{m}\left(M_{mn}(\delta_0, \beta_0, \tau_2) + \frac{1}{m}\sum_{j=1}^{m} \bar{\Gamma}_{2,j}(\beta_0, \tau)\left(\hat{\beta}_j - \beta_{j,0}\right)\right) \xrightarrow{d} N(0, \Omega(\tau)) \tag{41}$$

where $\Omega(\tau) = \frac{\Omega_1(\tau)}{n} + \Omega_2(\tau)$.

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

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 In this part of the proof, I start by showing that $\hat{\delta}^*(\hat{\beta}, \tau) - \hat{\delta}(\hat{\beta}, \tau) = O_p^*(m^{-1/2})$. 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.

It can be shown that

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

which together with Theorem 2 implies

$$\hat{\delta}^*(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau) = O_p^*\left(m^{-1/2}\right). \tag{42}$$

Second, 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 a linear function.

Hence, I want to show that

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

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 second part, I rely on the results on Chen et al. (2003).

Define the linearization where the dependencies on τ are suppressed for ease of notation:

$$\mathcal{L}^*_{mn}(\delta) = M^*_{mn}(\delta_0, \beta_0) + \Gamma_1(\delta_0, \beta_0)(\delta - \delta_0) + \frac{1}{m}\sum_{j}\bar{\Gamma}^*_{2,j}(\hat{\delta}^*, \beta_0)(\hat{\beta}^*_j - \beta_{j,0}) \tag{43}$$

Similarly to the proof theorem 2 in Chen et al. (2003), the goal is to show that we can approximate $M^*_{mn}(\hat{\delta}^*, \hat{\beta}^*)$ 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, we want to show that

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

By the triangle inequality:

$$\begin{aligned}||M^*_{mn}(\hat{\delta}^*, \hat{\beta}^*, \tau) - \mathcal{L}_{mn}(\hat{\delta}^*)|| \leq & ||M(\hat{\delta}^*, \hat{\beta}^*, \tau) - M(\hat{\delta}^*, \beta^*_0, \tau) - \frac{1}{m}\sum_{j=1}^{m} \bar{\Gamma}^*_{2,j}(\hat{\delta}^*, \beta^*)(\hat{\beta}^*_j - \beta^*_{j,0})|| \\ & + ||M(\hat{\delta}^*, \beta^*_0, \tau) - \Gamma_1(\delta_0, \beta_0)(\hat{\delta}^* - \delta_0)|| \\ & + ||M^*_{mn}(\hat{\delta}^*, \hat{\beta}^*, \tau) - M(\hat{\delta}^*, \hat{\beta}^*, \tau) - M^*_{mn}(\delta_0, \beta_0)||\end{aligned}$$

Where for the first term we have:

$$||M(\hat{\delta}^*, \hat{\beta}^*, \tau) - M(\hat{\delta}^*, \beta^*_0, \tau) - \frac{1}{m}\sum_{j=1}^{m} \bar{\Gamma}^*_{2,j}(\hat{\delta}^*, \beta^*)(\hat{\beta}^*_j - \beta^*_{j,0})|| = O^*_p(||\hat{\beta}^* - \beta_0||^2_{\mathcal{B}}) = o^*_p(m^{-1/2}).$$

since $\sup_j ||\hat{\beta}_j - \beta_{j,0}|| = o_p(m^{-1/4})$ as shown in the proof of Theorem 2.

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

$$||M(\hat{\delta}^*, \beta^*_0, \tau) - \Gamma_1(\delta_0, \beta_0)(\hat{\delta}^* - \delta_0)|| = o^*_p(m^{-1/2}).$$

For the third term, we have:

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

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

Hence,

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

Similarly, I now shown that $||M^*_{mn}(\bar{\delta}^*, \hat{\beta}) - \mathcal{L}^*_{mn}(\bar{\delta}^*)|| = o^*_p(m^{-1/2})$, 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:

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

Note that by the proof of Theorem 2

$$\begin{aligned}\bar{\delta}^* - \delta_0 = & \Gamma_1(\delta_0, \beta_0)^{-1} \left(\frac{1}{m}\sum_{j=1}^{m} \bar{\Gamma}^*_{2j}(\delta, \beta_0)[\hat{\beta}^*_j - \beta_{j,0}] + \frac{1}{m}\sum_{j=1}^{m} \bar{m}^*_j(\delta_0, \beta_0) \right) \\ = & \Gamma_1(\delta_0, \beta_0)^{-1} \left(\frac{1}{m}\sum_{j=1}^{m} \bar{\Gamma}_{2j}(\delta, \beta_0)[\hat{\beta}_j - \beta_{j,0}] + \frac{1}{m}\sum_{j=1}^{m} \bar{m}_j(\delta_0, \beta_0) \right) + O^*_p(m^{-1/2}) \\ = & \hat{\delta} - \delta_0 + O^*_p(m^{-1/2}) \\ = & O^*_p(m^{-1/2}) \end{aligned} \tag{44}$$

By the triangle inequality

$$||M^*_{mn}(\bar{\delta}^*, \hat{\beta}^*) - \mathcal{L}^*_{mn}(\bar{\delta}^*)|| \leq ||M(\bar{\delta}^*, \hat{\beta}^*) - M(\bar{\delta}^*, \beta^*_0) - \sum_{j=1}^{m} \Gamma^*_{2,j}(\bar{\delta}^*, \beta^*)(\hat{\beta}^* - \beta^*)|| \\ + ||M(\bar{\delta}^*, \beta_0) - \Gamma_1(\delta_0, \beta_0)(\bar{\delta}^* - \delta_0)|| \\ + ||M^*_{mn}(\bar{\delta}^*, \hat{\beta}^*) - M(\bar{\delta}^*, \hat{\beta}^*) - M^*_{mn}(\delta_0, \beta_0)||.$$

For the first term, we have:

$$||M(\bar{\delta}^*, \hat{\beta},) - M(\bar{\delta}^*, \beta_0)|| \leq ||\sum_{j=1}^{m} \Gamma_{2j}(\bar{\delta}^*, \beta_0)[\beta_j - \beta_{j,0}]|| + O_p(||\beta_j - \beta_{j,0}||^2_{\mathcal{B}}) = o^*_p(m^{-1/2})$$

For the second term, by differentiability of $M(\delta, \beta_0, \tau)$ and using equation (44) yields

$$||M(\bar{\delta}^*, \beta_0, \tau)|| \leq ||\Gamma_1(\bar{\delta}^*(\beta_0) - \delta_0)|| + ||o(\bar{\delta}^*(\beta^*_0) - \delta_0)|| = o^*_p(m^{-1/2}).$$

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

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

Hence, it follows that

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

Following Chen et al. (2003), if I can show that $\bar{\delta}^* - \hat{\delta}^* = o^*_p(m^{-1/2})$, it implies that:

$$\hat{\delta}^*(\hat{\beta}^*, \tau) - \delta_0(\beta_0, \tau) \\ = \Gamma_1(\delta_0, \beta_0, \tau)^{-1} \left(\sum_{j=1}^{m} \Gamma^*_{2j}(\delta, \beta_0, \tau)[\hat{\beta}^*_j(\tau_1) - \beta_{j,0}(\tau_1)] + M^*_{mn}(\delta_0, \beta^*_0, \tau) \right) + o^*_p \left(m^{-1/2} \right) \quad (45)$$

Now I show $\bar{\delta}^* - \hat{\delta}^* = o_p(m^{-1/2})$. Following Pakes and Pollar (1989), we know that $M^*_{mn}(\delta, \beta^*_0)$ and $\mathcal{L}^*(\delta)$ are close at both $\hat{\delta}^*(\hat{\beta}^*)$ which almost minimizes $||M^*_{mn}(\delta, \hat{\beta}^*)||$ and at $\bar{\delta}^*$ which minimizes $\mathcal{L}^*(\delta)$. This means that $\hat{\delta}^*$ has to be close to minimizing $\mathcal{L}^*(\delta)$:

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

This implies

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

and squaring both sides

$$||\mathcal{L}(\hat{\delta}^*)||^2 = ||\mathcal{L}(\bar{\delta}^*)||^2 + o_p^*(m^{-1}) \tag{46}$$

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

$$||\mathcal{L}(\delta)||^2 = ||\mathcal{L}(\bar{\delta}^*(\hat{\beta}^*)||^2 + ||\Gamma_1(\delta - \delta(\hat{\beta}^*))||^2 \tag{47}$$

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 (46) and (47) 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 (45) holds.

Part 2 - Distribution of $\hat{\delta}^* - \hat{\delta}_0$

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

$$\theta_j^*(\tau) = \Gamma_1(\delta_0, \beta_0)^{-1}\frac{1}{m}\sum_{j=1}^{m}\left(\bar{\Gamma}_{2j}^*(\delta, \beta_0, \tau)[\hat{\beta}_j^*(\tau_1) - \beta_{j,0}(\tau_1)] + \bar{m}_j^*(\delta_0, \beta_0^*, \tau)\right).$$

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

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

The goal is to show that for a nonrandom vector $\eta > 0$,

$$\frac{\eta'(\theta_j^* - \mathbb{E}^*[\theta_j^*])}{\sqrt{\eta'\Gamma_1^{-1}\Omega\Gamma_1'^{-1}\eta}} \xrightarrow{d} N(0,1).$$

Combining the expressions for $\hat{\delta}^*(\hat{\beta}^*, \tau)$ and $\hat{\delta}(\hat{\beta}^*, \tau)$ yields:

$$\hat{\delta}^*(\hat{\beta}^*, \tau) - \hat{\delta}(\hat{\beta}^*, \tau) = \left(\theta_j^*(\tau) - \mathbb{E}^*[\theta_j^*(\tau)]\right) + o_p^*(m^{-1/2})$$

Then note that, $\text{Var}^*(\eta'\theta_j^*) \xrightarrow{p} \eta'\Gamma_1^{-1}\Omega\Gamma_1'^{-1}\eta$. Which implies that $\text{Var}^*(\eta'\theta_j^*) = \eta'\Gamma_1^{-1}\Omega\Gamma_1^{-1\prime}\eta + o_p(\eta'\Gamma_1\Omega\Gamma_1\eta)$ since $\Lambda_{min}(\Omega) > c > 0$. Then by the central limit theorem for i.i.d. data:

$$\frac{\eta'\theta_j^*}{\sqrt{\eta'\Gamma_1\Omega\Gamma_1'\eta}} = \frac{\eta'\theta_j^*}{\sqrt{\text{Var}^*(\eta'\theta_j^*)}} + o_p^*(1) \xrightarrow{d^*} N(0,1).$$

Hence, $\hat{\delta}^*(\hat{\beta}) - \hat{\delta}(\hat{\beta})$ has the same asymptotic distribution as $\hat{\delta}(\hat{\beta}) - \delta_0(\beta_0)$.

∎

C Additional Simulation Results

$\tau_1 \setminus \tau_2$	β 0.25	β 0.5	β 0.75	γ 0.25	γ 0.5	γ 0.75
	(m, n) = (25,25)					
0.25	1.083	1.117	1.046	0.943	1.015	0.920
0.5	1.064	1.124	1.091	0.910	1.026	0.931
0.75	1.057	1.121	1.104	0.924	1.030	0.932
	(m, n) = (25,200)					
0.25	1.074	1.126	1.093	0.860	0.873	0.845
0.5	1.088	1.137	1.116	0.849	0.855	0.843
0.75	1.079	1.127	1.124	0.849	0.871	0.847
	(m, n) = (200,25)					
0.25	1.036	1.067	1.025	1.019	1.039	1.006
0.5	1.022	1.057	1.070	1.019	1.031	1.007
0.75	1.027	1.074	1.056	1.007	1.025	0.986
	(m, n) = (200,200)					
0.25	1.035	1.026	1.039	0.981	1.004	0.965
0.5	1.045	1.035	1.039	0.994	1.002	0.960
0.75	1.056	1.034	1.025	0.996	1.005	0.967

Note:
Results based on 2000 Monte Carlo simulations. The table provides standard errors relative to standard deviation.

Table 5: Standard Errors relative to 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	0.960	0.967	0.947	0.909	0.935	0.893
0.5	0.947	0.965	0.958	0.897	0.934	0.898
0.75	0.944	0.965	0.967	0.899	0.942	0.900
	(m, n) = (25,200)					
0.25	0.946	0.967	0.953	0.831	0.871	0.828
0.5	0.944	0.965	0.955	0.826	0.861	0.829
0.75	0.945	0.962	0.962	0.828	0.865	0.830
	(m, n) = (200,25)					
0.25	0.918	0.964	0.889	0.933	0.949	0.925
0.5	0.908	0.955	0.926	0.929	0.954	0.934
0.75	0.875	0.960	0.926	0.932	0.945	0.923
	(m, n) = (200,200)					
0.25	0.943	0.947	0.941	0.924	0.932	0.920
0.5	0.946	0.949	0.950	0.922	0.929	0.921
0.75	0.944	0.946	0.936	0.921	0.932	0.920

Note:
Results based on 2000 Monte Carlo simulations. The table provides the coverage probability of the 95 confidence invervals.

Table 6: Coverage Probability