Quantile on Quantiles*

Martina Pons†

February 12, 2026

Abstract

I develop an econometric framework to analyze the impact of a treatment on inequality within and between groups. Outcomes are represented by a two-dimensional quantile surface mapping within-group and between-group ranks to outcome levels. This representation captures the distributional structure of the data without imposing normative assumptions, providing a foundation for assessing trade-offs between different dimensions of inequality. Within a broad class of linear welfare functions, the two-dimensional quantile surface is the empirical primitive for welfare evaluation. I propose a two-step quantile regression estimator and establish its weak convergence to a bivariate Gaussian process. An application to business training in Kenya shows that treatment effects are concentrated among high-performing firms in strong markets, highlighting complementarities between individual and group performance.

Keywords: Quantile Regression, Grouped Data, Multidimensional Heterogeneity, Multidimensional Inequality.

1 Introduction

Since Koenker and Bassett (1978), quantile regression has been widely used to study heterogeneous effects and distributional impacts of policies. Quantile treatment effects are particularly appealing when policymakers care about distributional consequences, such as targeting low-income households or addressing inequality rather than maximizing aggregate outcomes. Yet many real-world policy objectives are inherently multidimensional: societies often care about inequality both within and between groups—across regions,

*I am grateful to Blaise Melly and Anna Mikusheva for their support. I also thank Alberto Abadie, Ingvild Almås, Isaiah Andrews, Jean-Michel Benkert, Maya Eden, Ivan Fernández-Val, Joachim Freyberger, Bo Honoré, Frédéric Kluser, Damian Kozbur, Jonas Meier, Sami Stouli, Maximilian von Ehrlich, Michael Wolf, Kaspar Wüthrich, and seminar participants for helpful comments and discussions. I gratefully acknowledge the support from a Doc.Mobility Grant from the University of Bern to visit the Massachusetts Institute of Technology.

†Department of Economics, University of Zurich, Schönberggasse 1, 8001 Zurich, Switzerland. martina.pons@econ.uzh.ch.

industries, or socioeconomic backgrounds—and not all forms of inequality are viewed as equally problematic. These concerns are explicit in policy agendas: the United Nations' Sustainable Development Goals call for reducing inequality "within and among countries," while *equality of opportunity* principles emphasize compensating for differences due to circumstances while respecting differences due to effort. Similarly, the European Union's Cohesion Policy aims to foster convergence across regions; yet also in this case, the within-region component cannot be ignored, as meaningful convergence requires distributions to converge at all percentiles rather than only in their averages. These goals raise a practical evaluation question: how should economists assess whether a policy intervention advances such objectives? A coherent assessment requires both (i) a description of how it reshapes outcomes across the relevant dimensions and (ii) an aggregation of those changes into a welfare assessment.

This paper develops an econometric framework that represents outcomes through a two-dimensional quantile surface, capturing heterogeneity both within and between groups simultaneously. The surface provides a parsimonious yet flexible representation of multidimensional heterogeneity, without imposing restrictions on how groups differ or on how individual and group characteristics interact. The framework describes how inequality evolves across these dimensions and links this structure to welfare analysis by modeling how society trades off differences within and between groups. I show that this representation delivers a natural foundation for multidimensional welfare comparisons and develop an estimator for the surface with uniform asymptotic theory, enabling inference on multidimensional distributional effects and welfare implications.

Existing empirical work suggests that within- and between-group inequality are both relevant and need not move in the same direction, yet these dimensions are often studied separately. For example, place-based policies have been shown to stimulate local growth and employment in lagging regions (Becker et al., 2010; Busso et al., 2013; Ehrlich and Seidel, 2018), while at the same time increasing within-region inequality (Lang et al., 2023; Albanese et al., 2023). These findings underscore that the two dimensions are interdependent: policies may improve outcomes along one dimension while worsening them along the other. A coherent assessment, therefore, requires modeling both dimensions together.[1]

Jointly modeling heterogeneity within and between groups requires going beyond standard conditional distributional tools. Quantile and distribution regression characterize heterogeneity along a single index, potentially conditional on covariates and group indicators. Even with rich interactions and group-specific coefficients, these approaches deliver a collection of group-specific distributions rather than a unified object that simultaneously organizes heterogeneity along both margins. As a result, while existing methods can estimate group-specific quantile functions, they do not, by themselves, provide a canonical way to jointly summarize between-group and within-group heterogeneity in a flexible manner that is amenable to covariates and directly usable for welfare analysis.

Analyzing inequality along multiple dimensions raises fundamental conceptual challenges. First, the extent to which distributions or groups can be ranked depends on the strength of the normative assumptions. For instance, the same region can exhibit high mobility for some parts of the parental income distribution but low mobility for others (Chetty and Hendren, 2018a,b). Therefore, without strong assumptions, only partial orderings of groups can be obtained.

Second, in the one-dimensional case, standard dominance criteria often allow for an (almost) complete ranking of distributions: society can generally determine whether one distribution is more unequal than another for a broad class of welfare functions. By contrast, when inequality has several dimensions, comparisons are generally incomplete without specifying how society values improvements in one dimension relative to deteriorations in another (see, e.g., Atkinson and Bourguignon, 1987).

Section 2 shows that outcomes can be represented parsimoniously by a two-dimensional quantile surface capturing heterogeneity both within and between groups. Each individual outcome depends on a scalar within-group rank, a vector of group characteristics, and

[1] A related point appears in the misallocation literature: Kehrig and Vincent (2025) shows that dispersion within firms (across plants) can arise from efficient internal reallocation ("good" dispersion), whereas dispersion between firms can be interpreted as reflecting misallocation. This underscores that different dimensions of dispersion need not have the same welfare meaning. Empirical evidence also highlights the importance of both dimensions on inequality: within- and between-group disparities have been linked to adverse social and economic outcomes (e.g., Breza et al., 2018; Fehr et al., 2020; Glaeser et al., 2009; Chetty et al., 2016; Chetty and Hendren, 2018b,a), and regional disparities in economic performance have been associated with negative effects (e.g., Pierce and Schott, 2016; Autor et al., 2021). At the same time, interpersonal comparisons further underscore the relevance of the within-group dimension (Galí, 1994; Luttmer, 2005; Card et al., 2012).

possibly additional covariates. Rather than imposing a single, global ranking of groups, the two-dimensional quantile surface ranks groups at each point of the within-group distribution by an ordinal index of group advantage. Importantly, this representation imposes no restrictions on how groups differ and does not assume that groups can be unambiguously ordered: a group may offer favorable conditions for some parts of the distribution while being disadvantaged at others. Viewed along one dimension, the surface shows how group-specific percentiles shift as one moves to higher-ranked groups; viewed along the other, it traces the quantile envelope formed by group-specific percentile curves.

I then propose a general class of welfare criteria that captures trade-offs between within- and between-group inequality through generalized social marginal welfare weights (Saez and Stantcheva, 2016). Throughout, these criteria are interpreted as the objective of a benevolent social planner. Welfare is defined as a weighted aggregation of the two-dimensional outcome surface, with weights that may vary with an individual's position within the group and the group's relative standing. This formulation embeds normative choices directly in the weights rather than in the curvature of a social utility function, yielding a tractable and flexible criterion. Importantly, the outcome surface is the empirical primitive for this class of rank-based welfare criteria: any welfare index in the class can be written as a weighted integral of the surface. The class encompasses well-known cases such as equality of opportunity, rank-dependent criteria, and certain standard welfarist specifications, and it permits different trade-offs across dimensions of inequality, for example, through a two-dimensional extension of the Gini welfare function. This flexibility makes the framework applicable to a wide range of contexts, from evaluating regional convergence policies to assessing opportunity gaps across socioeconomic backgrounds and studying multidimensional inequality more generally.

Section 3 develops the econometric representation of the two-dimensional surface and introduces covariates into the model. Within-group heterogeneity is captured by group-specific conditional quantile functions, while the between-group component is obtained by modeling the quantile function of these group-level quantiles.

In Section 4, I introduce the estimation approach, which proceeds in two stages. The first stage consists of group-by-group quantile regressions of the outcome on covariates that vary within groups, yielding estimates of the group-specific conditional quantile functions. Similar first-stage estimators have been proposed by Galvao and Wang (2015), Chetverikov et al. (2016), and Melly and Pons (2025). In the second stage, the first-stage fitted values are regressed on all covariates using quantile regression, separately at each within-group quantile. Intuitively, the first stage captures heterogeneity within groups, while the second aggregates these results to recover heterogeneity across groups at each point of the within-group distribution. The estimator is flexible, allowing the coefficients to vary freely along both dimensions.[2]

Section 5 presents the asymptotic results. Establishing the asymptotic properties of the estimator involves three challenges: the non-smoothness of the objective function, the generated dependent variable in the second stage, and the different convergence rates of the two steps. Each first-stage regression, which uses only observations within a group, converges at rate $\sqrt{n}$, where n is the number of observations per group. By contrast, the second stage, which captures heterogeneity across groups, converges at rate $\sqrt{m}$, where m is the number of groups. The setting is related to Chen et al. (2003), who analyze estimators with non-smooth objective functions and nonparametric first steps, and to Volgushev et al. (2019) and Galvao et al. (2020), who provide detailed Bahadur representations for the first step quantile regression.[3] Building on this work and the process results in Angrist et al. (2006), I establish weak convergence of the full two-dimensional process in a framework where both n and m diverge, subject to the mild condition $\frac{\sqrt{m}\log n}{n} \to 0$. I also propose an inference procedure for uniform hypotheses.

In Section 6, as an empirical application, I extend the analysis of McKenzie and Puerto (2021) by assessing the impact of business training on income from work in Kenya, considering distributional effects both within and between markets. A standard quantile regression of income on the treatment dummy captures heterogeneity across firms, but it does not account for whether a median firm (in the unconditional distribution) operates in a high- or low-performing market, even though the treatment may affect these firms differently. Poorly performing markets may face disadvantages such as location or low consumer traffic, and interactions between individual ability and market quality may also

[2] I provide an R package that implements the proposed estimator.

[3] See also Ma and Koenker (2006) and Chen et al. (2021), on quantile regression with generated regressors or dependent variables. In contrast, my setting allows the dimension of the first stage to grow with the number of groups, features two stages converging at different rates, and establishes uniform asymptotic results.

be important, factors that a traditional quantile regression cannot capture. The results show larger treatment effects for firms that perform well within successful markets. Specifically, effects increase with both a firm's rank within its market and the market's overall rank, providing evidence of complementarities between individual and group ranks. This finding is useful for policymakers in deciding which firms and markets to target, depending on their objective. For example, if the goal is to maximize total income growth, reduce inequality, or balance the two, the treatment can be directed toward different parts of the distribution or groups.

Distributional effects and inequality, particularly within groups, have been widely studied in theoretical and applied work. On the methodological side, Galvao and Wang (2015), Chetverikov et al. (2016), and Melly and Pons (2025) develop quantile estimators to model heterogeneity in treatment effects along the within-group distribution. On the applied side, Autor et al. (2021) and Friedrich (2022) analyze how import competition and trade shocks affect wage distributions within local labor markets and firms, respectively. Further work, such as Autor et al. (2016) and Engbom and Moser (2022), examines how minimum wages affect within-state inequality in the United States and Brazil. These papers, however, speak to within-group distributional effects, leaving cross-group heterogeneity outside their scope.

Some papers consider both within- and between-group inequality, and these mostly take a descriptive approach. For instance, Bourguignon and Morrisson (2002) and Milanovic (2002) analyze the evolution of income inequality both within and across countries. Akerman et al. (2013) and Helpman et al. (2017) study wage inequality between and within firms, sectors, and occupations, and Haltiwanger et al. (2024) shows that the differences across industries account for a substantial share of the recent rise in earnings inequality. Typically, these papers rely on decompositions that separate overall inequality into within- and between-group components rather than examining both margins simultaneously. While informative, standard within–between decompositions (e.g., variance- or Theil-type) reduce inequality to additively separable components and therefore cannot capture interactions across the two dimensions nor distributional changes beyond low-order moments.[4]

This paper contributes to the theoretical literature on multidimensional unobserved heterogeneity, where marginal effects can vary across multiple dimensions. Existing approaches that consider both dimensions typically achieve tractability by imposing restrictions such as requiring within-group variation for identification of between-group components, modeling between-group heterogeneity through low-dimensional latent factors, or adopting additive/separable structures that limit cross-dimensional interactions—rather than by providing a direct distributional representation across dimensions. For instance, in a distribution regression framework, Fernández-Val et al. (2022) model between-group heterogeneity through group-specific coefficients. As a result, identifying these between-group components requires within-group variation in the covariate of interest. Arellano and Bonhomme (2016) model group effects as latent variables in a correlated random-effects framework, generating heterogeneity through an individual rank variable and low-dimensional latent group effects. Other contributions impose additional structure on cross-dimensional interactions: Frumento et al. (2021) let individual-level covariates govern within-group dimension while group-level covariates shift the between-group dimension, and Liu (2024) allows individual-level effects to vary with a group-level rank variable while keeping the individual error additive. In contrast, the framework developed here targets the entire two-dimensional quantile surface without such restrictions, allowing heterogeneity to vary freely along both margins.

Finally, the paper also connects to the normative tradition on welfare measurement and inequality evaluation. Standard welfarism, which evaluates allocations solely through individual utilities, struggles to capture multidimensional concerns (see, e.g., Carroll, 2025). The generalized approach of Saez and Stantcheva (2016) expresses social preferences directly through welfare weights that may depend on ethically relevant characteristics, linking naturally to the equality of opportunity framework of Roemer (1998) and to the multidimensional-inequality literature of Atkinson and Bourguignon (1987). By expressing welfare weights as functions of within- and between-group ranks, the frame-

[4]In particular, for widely used measures such as generalized entropy indices (including Theil's indices), the between-group component is driven by differences in subgroup means and is therefore largely silent about cross-group differences in distributional shape. An exception is Heikkuri and Schief (2025), who develops an axiomatic Gini decomposition in which the between-group component explicitly depends on cross-group distributional differences.

work developed here provides a tractable and empirically implementable way to analyze multidimensional heterogeneity.

2 Model

2.1 Outcome Model

Consider a population divided into m groups indexed by $j = 1, \ldots, m$. Within each group j, individuals are indexed by $i = 1, \ldots, n$. Let each individual's outcome be

$$y_{ij} = q(u_{ij}, v_j), \tag{1}$$

where $q(\cdot, \cdot)$ is a function of some individual level rank variable u_{ij} normalized to be $U(0, 1)$ within each group and a vector (of possibly unknown dimension) containing group characteristics or circumstances v_j.

A central challenge in analyzing multidimensional inequality is to construct a flexible representation that captures all relevant distributional features. Given that we consider inequality over two dimensions, optimally, we could construct a bivariate function.

Modeling heterogeneity within groups is straightforward. Conditional on each group j, the within-group rank satisfies $u_{ij}|v_j \sim U(0, 1)$, and we impose strict monotonicity of q with respect to this scalar rank. This delivers a function $q(u, v_j)$ that maps within-group ranks u to outcome levels, so $q(u, v_j)$ is the quantile function of group j.

This approach would not work for the between dimension as the assumptions required to order distributions are typically too strong to be broadly acceptable (see, e.g., Atkinson and Bourguignon, 1987; Roemer and Trannoy, 2016). Often, researchers tackle this difficulty by focusing on the group mean or median outcome, which provides a tractable way to summarize group differences. However, this approach may mask substantial differences across the distributions, ignore potentially important variation within the group, and poorly reflect the conditions experienced by most individuals—features that are particularly relevant when inequality itself is central to the analysis. In the setting of model (1), this challenge implies that the group component cannot, in general, be represented by a single scalar rank variable.

To see this, consider a naive attempt to construct such a model in which $y_{ij} = q(u_{ij}, v_j)$ is strictly increasing in a scalar group rank v_j. Under this assumption, groups can be ordered unambiguously: for any two groups h and l, with $v_h > v_l, q(u, v_h) > q(u, v_l)$ for all $u \in (0, 1)$. This means that if one group has a higher first decile, it must also have a higher value for all other percentiles. Such a restriction substantially limits the degree of heterogeneity between groups and would hold, for example, under additive separability: $y = h(u) + f(v)$, where h and f are strictly increasing. In this case, groups have identical distributions up to a location shift.[5] To see that the data depart from this pattern, consider the following example.

Example 1. Figure 1 illustrates this departure using income data from the 2015–2019 ACS 5-year sample for a selected set of U.S. labor markets (Ruggles et al., 2025). As in Autor et al. (2019), I approximate labor markets using commuting zones (CZs).[6] The sample is restricted to individuals aged 24 to 64 and includes 13,170,346 observations across 665 CZs, and I adjust incomes for cross-state price differences using the BEA's 2020 regional price parities. For expositional clarity, I abstract from estimation error in this illustration.

Given the large number of CZs, the left panel displays the full quantile functions for four selected CZs, while for all others it reports only the 20th, 50th, and 80th percentiles (in gray). The figure shows that CZ-level quantile functions often cross, implying the absence of a clear group ordering. Income distributions differ markedly across CZs: the Houston and Seattle CZs exhibit substantially more dispersed wage distributions, whereas the Minneapolis CZ displays a more compressed distribution with higher wages at the bottom but lower wages at the top.

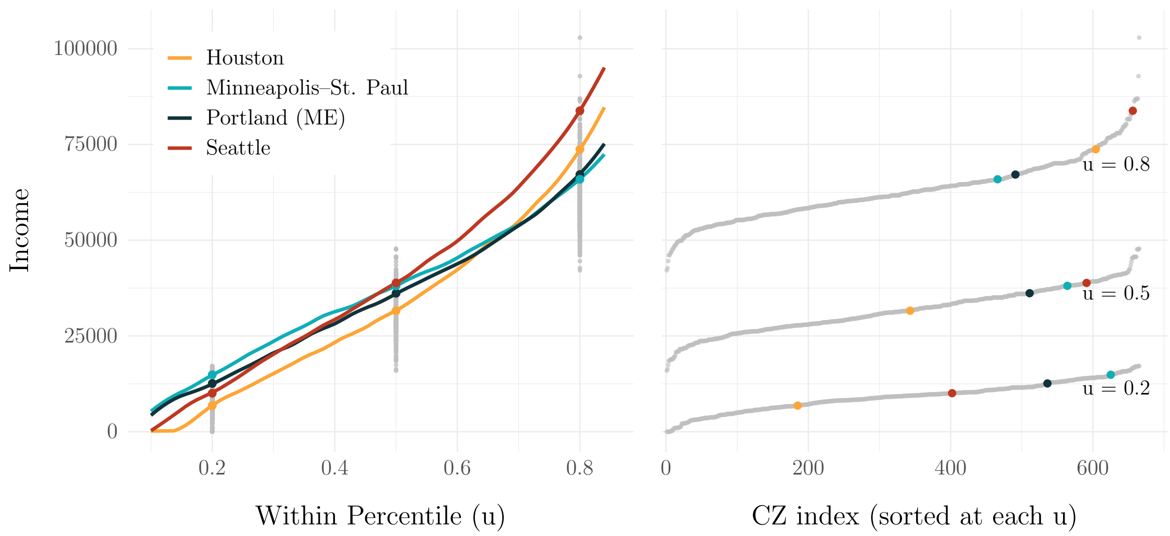
Since requiring a single scalar ranking of groups is both highly restrictive and inconsistent with many empirical patterns, I let v_j be multidimensional, allowing for richer forms of between-group heterogeneity. In practice, a group need not be uniformly better or worse for all individuals: it may offer favorable conditions for some parts of the distribution while being relatively disadvantaged for others.

Still, modeling the between dimension requires specifying how the multidimensional vector v_j enters q in a way that does not constrain heterogeneity across groups. Since a

[5] This restriction would also hold if there were no overlap between groups, but such cases are rare in economic applications. Formally, a sufficient condition ensuring constant ranks is strict first-order stochastic dominance between groups.

[6] Individuals are assigned to CZs using the crosswalk provided by Autor et al. (2019).

Figure 1: Income Quantiles across Commuting Zones



Notes: The left panel plots CZ-level quantile functions of yearly income for four selected CZs, together with the 20th, 50th, and 80th percentiles for all remaining CZs (in gray). The right panel displays the sorted CZ-level percentiles: for a fixed within-group percentile u, each CZ contributes a single value $q(v_j, u)$ which can be unambiguously sorted.

total ordering of groups is generally unfeasible, I instead order them separately at each percentile of the within-group distribution. That is, for every percentile u, groups can be unambiguously ranked based on the corresponding quantile $q(u, v_j)$. Formally, even if v_j is multidimensional, its contribution at a fixed u can be represented by a scalar-valued function $v_j(u)$, so there exists a mapping $\tilde{q}$, increasing in both arguments, such that

$$q(u, v_j) = \tilde{q}(u, v_j(u)).$$

Without loss of generality, I normalize $v_j(u)$ to be uniformly distributed on $(0, 1)$ for each u and, for simplicity, use v to denote this scalar index, with the understanding that, within each group, it may vary with u. Importantly, v does not label a fixed group; it represents an ordinal index of group advantage that is defined separately at each within-group quantile u. This yields a two-dimensional surface $\tilde{q}(u, v)$ that summarizes joint heterogeneity by ranking individuals within groups (via u) and ranking groups at each u (via v), capturing the notion that a better within-group rank or a higher group rank for a given u implies a higher outcome level. This formulation is entirely flexible: it imposes no restrictions on the joint distribution of outcomes or ranks; it simply maps multidimensional group characteristics into a single index that describes group differences at specific points of the within distribution.

For a fixed u, $v \mapsto \tilde{q}(u,v)$, is the quantile function of the uth group-level percentiles, recording how the corresponding group-specific quantiles vary across groups as v increases. Conversely, fixing v yields $u \mapsto \tilde{q}(u,v)$, which at each within-group percentile u, selects the cross-group v-quantile of the group-level quantiles $q(u,v_j)$. As v varies, these curves form the family of quantile envelopes of the group-level quantile functions, with $v \downarrow 0$ and $v \uparrow 1$ corresponding to the lower and upper envelopes, respectively. The two-dimensional surface $\tilde{q}(u,v)$ can be written using nested quantile operators $Q(v, Q(u, y_{ij}|v_j))$, where $Q(u, y_{ij}|v_j)$ is the within-group u-quantile of y_{ij}, and $Q(v, Q(u, y_{ij}|v_j))$ is the v-quantile of that object across groups. Throughout, I use uppercase Q for (conditional) quantile operators and lowercase q (and $\tilde{q}$) for latent-rank mappings from ranks to outcome levels.

Example 2 (Continuation of Example 1)**.** Returning to Figure 1, we can now visualize how the second dimension of the surface is constructed. Consider the CZ-level median ($u = 0.5$). Each CZ has a well-defined median; across all CZs, this yields a distribution of group medians, which can be unambiguously ordered, as shown in the right Panel, where the same medians displayed in the left Panel are sorted and plotted against a group-level index. This produces a monotone function in a scalar index $v_j(u)$, which we may subsequently normalize without loss of generality. Repeating this construction for every percentile $u \in (0,1)$ yields a family of such orderings, thereby producing the full between-group dimension of the surface.

One might wonder why not bypass monotonicity in a scalar group index and work directly with group-level quantile functions $q(u,v_j)$, treating v_j as a group-specific scalar rank variable. However, a single scalar v_j still restricts heterogeneity across group-level quantile functions. Working instead with the unrestricted collection $\{q(u,v_j)\}_{j=1}^{m}$ is also not tractable: without a scalar index that places groups on a common scale, the quantile curves cannot be easily compared, no coherent between-group dimension exists, and differences across groups cannot be related to group-level characteristics. This makes it difficult to construct a unified representation of the joint distribution or to apply welfare criteria. The two-dimensional formulation $\tilde{q}(u,v)$ resolves these issues, making the model tractable even with covariates and many groups.

2.2 Social Welfare

Having constructed a representation of joint within- and between-group heterogeneity, this section describes how a policymaker can aggregate the two-dimensional surface for welfare evaluation. I adopt the *marginal social welfare weights* approach of Saez and Stantcheva (2016), allowing the weights to depend on the rank variables:

$$W_w(\tilde{q}) = \int_0^1 \int_0^1 w(u, v)\, \tilde{q}(u, v)\, du\, dv, \tag{2}$$

where $w(u, v) \geq 0$ denotes the social marginal welfare weight assigned to the individual at within-group rank u and group rank v.[7] Hence, welfare is a weighted average of outcomes, with weights depending on both rank variables. Section 3 introduces covariates, and the framework naturally extends to accommodate observable characteristics that may be normatively relevant, allowing inequality to be analyzed conditional on them.

Weights may decrease in both u and v, reflecting concern for inequality within and between groups. More precisely, one could set $w(u, v)$ to decline with u, giving greater weight to individuals who are worse off given their v; and for each u, the weights decline with v, assigning a higher value to improving outcomes of individuals in less advantaged groups. More generally, the framework captures trade-offs across different dimensions of inequality and accommodates fairness views such as equality of opportunity. It recognizes that different sources of inequality need not be viewed as equally objectionable, and that society may place greater weight on disparities arising from factors beyond individual control. The framework also allows the strength of these trade-offs to vary with the level of inequality, reflecting the possibility of increasing marginal social costs along either dimension. This perspective clarifies the tensions policymakers face: as Glaeser et al. (2009) notes, "*local inequality is actually the inverse of area-level income segregation*," so whether reducing inequality in one dimension at the cost of increasing it in another is desirable ultimately depends on the underlying welfare weights.

Equation (2) defines a class of welfare functionals parameterized by the social marginal weights $w(u, v)$. In this framework, the weight function $w(u, v)$ is the primitive normative

[7] Although $v_j(u)$ is defined only conditional on u, the double integral in Equation (2) should be interpreted as integrating over the two-dimensional rank surface $(u, v) \in (0, 1)^2$: for each u, v indexes the relative position of groups across the between-group dimension. This representation is equivalent to averaging across groups at each u and then integrating over u.

object; alternative social objectives correspond to alternative choices of weights. The next definition and lemma give a useful characterization: welfare comparisons within this class depend on the outcome distribution only through the surface $\tilde{q}(u, v)$, and any welfare index in the class is a weighted integral of that surface. The proof is in the Appendix.

Definition 1 (Class of welfare functionals)**.** Let $\mathcal{W}$ be the set of nonnegative integrable weight functions on $(0, 1)^2$. Each $w \in \mathcal{W}$ defines a welfare functional W_w via Equation (2). Two outcome surfaces $\tilde{q}_1$ and $\tilde{q}_2$ are *$\mathcal{W}$-equivalent* if $W_w(\tilde{q}_1) = W_w(\tilde{q}_2)$ for all $w \in \mathcal{W}$.

Lemma 1 (Characterization of $\mathcal{W}$-equivalence)**.** *Let $\mathcal{W}$ be as in Definition 1. Then, for any two outcome surfaces $\tilde{q}_1$ and $\tilde{q}_2$:*

(i) ***Sufficiency.*** *If $\tilde{q}_1 = \tilde{q}_2$ a.e., then $W_w(\tilde{q}_1) = W_w(\tilde{q}_2)$ for all $w \in \mathcal{W}$.*

(ii) ***Completeness.*** *If $W_w(\tilde{q}_1) = W_w(\tilde{q}_2)$ for all $w \in \mathcal{W}$, then $\tilde{q}_1 = \tilde{q}_2$ a.e.*

Because welfare aggregates information along both dimensions, it provides a natural basis to assess distributional consequences of policy interventions. Let $D \in \{0, 1\}$ denote the policy, and suppose the potential outcome surfaces $\tilde{q}_d(u, v)$ are identified. The welfare effect of adopting the policy is then

$$\Delta W_w = \int_0^1 \int_0^1 w(u, v) \, [\tilde{q}_1(u, v) - \tilde{q}_0(u, v)] \, du \, dv.$$

The class of welfare functionals satisfies natural properties: welfare weakly increases when the outcome surface increases at any rank position (monotonicity), varies continuously with the outcome surface (continuity), and is invariant to relabeling within groups and across groups (conditional anonymity). This class is highly general and can accommodate a wide range of normative perspectives; below, I show how well-known criteria such as equality of opportunity, utilitarian and rank-dependent welfare, and a two-dimensional extension of the Gini criterion arise naturally within this framework.

Importantly, Equation (2) allows marginal social welfare weights to depend on rank-based context (u, v), rather than only on utility levels. Under classical welfarism, marginal weights are functions of utility alone, so individuals with identical utilities must receive identical weights. Here, two individuals with the same utility may receive different weights depending on their position within the group and their group's relative standing. This follows Saez and Stantcheva (2016) in treating marginal welfare weights as the primitive normative objects and allowing them to vary with ethically relevant characteristics. In contrast, Fleurbaey and Maniquet (2018) incorporates fairness through the definition of individual advantage rather than through aggregation weights.

Unlike a standard rank-dependent welfare function, Equation (2) distinguishes between allocations that share the same marginal outcome distribution but differ in their within- and between-group structure. I now illustrate a few examples of welfare weights, as well as special cases in which Equation (2) reduces to familiar welfare criteria.

Two-dimensional Gini Social Welfare Function. A natural starting point is a two-dimensional extension of the Gini social welfare function, which captures inequality aversion through bilinear rank weights:[8]

$$w(u, v) = 4(1-u)(1-v),$$

assigning the highest weight to individuals who are disadvantaged both within and across groups. This weighting scheme induces a two-dimensional Lorenz surface

$$L(s, t) = \frac{1}{E[Y]} \int_0^s \int_0^t \tilde{q}(u, v) du dv,$$

and yields the corresponding two-dimensional Gini index

$$I_{\text{Gini}}^{(2)} = 1 - 4 \int_0^1 \int_0^1 L(s, t)\, ds\, dt.$$

Geometrically, $I_{\text{Gini}}^{(2)}$ measures four times the volume between the Lorenz surface $L(s, t)$ and the 45° plane, providing a natural generalization of the one-dimensional Gini coefficient.[9]

Additively Separable Weights. Consider the additively separable weight function

$$w(u, v) = 2\big[\omega(1-u) + (1-\omega)(1-v)\big], \qquad \omega \in [0, 1],$$

[8]For the one-dimensional Gini welfare function, see Blackorby and Donaldson (1978) and Weymark (1981).

[9]The two-dimensional Gini index is always weakly larger than the standard (one-dimensional) Gini, since the bilinear weighting scheme jointly penalizes high ranks in both dimensions. Equality holds when inequality arises along only one dimension.

where larger ω places greater emphasis on within-group inequality. Because the weights are additively separable in (u,v), welfare decomposes additively across the two rank dimensions:

$$W_w(\tilde{q}) = \int_0^1 2\omega(1-u)\Big(\int_0^1 \tilde{q}(u,v)\,dv\Big)\,du + \int_0^1 2(1-\omega)(1-v)\Big(\int_0^1 \tilde{q}(u,v)\,du\Big)\,dv.$$

The first term downweights higher within-group ranks u (after averaging outcomes across groups), while the second downweights higher group ranks v (after averaging outcomes across within-ranks), so ω governs the trade-off between these two margins.[10]

Equality of Opportunity. If groups are interpreted as exogenous *circumstances*, the class of welfare functional naturally aligns with the equality of opportunity perspective (Roemer, 1998). In this view, outcomes reflect both *effort* and *circumstances*, and society may wish to compensate for differences due to circumstances but not for those due to effort. Following the standard approach, *accountable effort* can be measured by an individual's rank within the distribution of those sharing the same circumstances, so that u captures effort. Fixing u and comparing across v thus isolates inequality due purely to circumstances. A convenient way to implement this principle is to use weights that depend only on v and place greater emphasis on worse circumstances:

$$W_w(\tilde{q}) \;=\; \int_0^1\!\!\int_0^1 w(v)\,\tilde{q}(u,v)\,dv\,du, \qquad w'(v) \le 0.$$

Roemer's criterion is obtained by concentrating all weight on the lowest circumstance rank v at each effort tranche u:

$$w_\varepsilon(v) = \frac{\mathbf{1}\{0\le v\le \varepsilon\}}{\varepsilon}, \qquad W^{\mathrm{R}}(\tilde{q}) = \lim_{\varepsilon\downarrow 0}\int_0^1\!\!\int_0^1 w_\varepsilon(v)\tilde{q}(u,v)dvdu = \int_0^1 \lim_{v\downarrow 0}\tilde{q}(u,v)du.$$

[10] Equivalently, welfare under this specification can be written in terms of two one-dimensional Gini indices computed along the two rank dimensions:

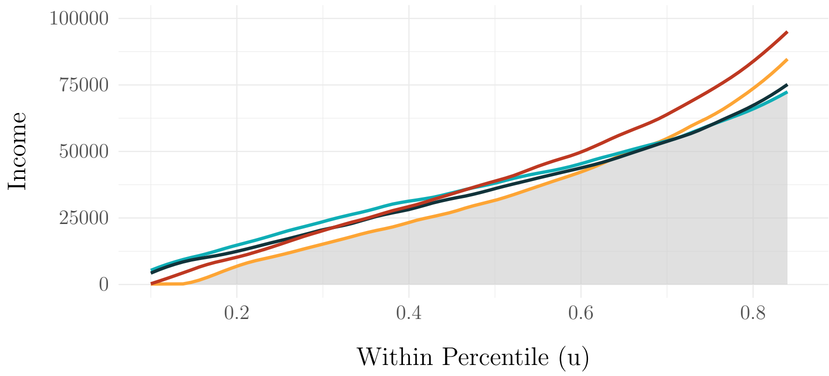
$$W_w(\tilde{q}) = E[Y]\Big(1 - \omega I_u - (1-\omega)I_v\Big),$$

where

$$I_u = 1 - \frac{\int_0^1 2(1-u)\left(\int_0^1 \tilde{q}(u,v)\,dv\right)du}{E[Y]}, \qquad I_v = 1 - \frac{\int_0^1 2(1-v)\left(\int_0^1 \tilde{q}(u,v)\,du\right)dv}{E[Y]}.$$

Here, I_u is the Gini index of the rank profile obtained by averaging outcomes across groups at each within-rank u, while I_v is the corresponding Gini index obtained by averaging outcomes across within-ranks at each group-rank v.

Figure 2: Roemer's (1998) Criterion



Notes: The figure plots the yearly income quantile functions for four CZs. The shaded region shows the area under the lowest envelope of the four CZ quantile functions. Under Roemer's equality of opportunity criterion, an optimal policy is the one that maximizes this area.

Graphically, when focusing on the four CZs from Example 1, the criterion corresponds to maximizing the area under the lowest envelope of the group-level quantile functions. Figure 2 shows this area in grey.

Allowing $w(v)$ to vary smoothly generalizes the strict maximin rule to intermediate degrees of concern for circumstance-based inequality. This yields a flexible yet empirically tractable framework for measuring inequality of opportunity and evaluating development while explicitly accounting for systematic opportunity differences (Roemer and Trannoy, 2016).

Special Cases. Equation (2) nests other familiar welfare criteria. If $w(u, v) = 1$, welfare reduces to the utilitarian mean $W_w(\tilde{q}) = E[Y]$, implying indifference toward inequality. If $w(u, v) = w(u)$, only within-group inequality matters, and welfare can be written as a function of the average of group-level quantiles. Finally, when $w(u, v) = \tilde{w}(F_Y(\tilde{q}(u, v)))$ for some $\tilde{w}(\cdot)$, the welfare depends solely on unconditional ranks, corresponding to the standard rank-dependent welfare function.

Welfare-Based Group Ordering. The framework can also be used when a policy-maker wishes to derive a single measure of group deprivation or priority—for example, to guide eligibility criteria. While a total ordering of groups is generally infeasible without strong assumptions, aggregating welfare weights within groups yields coherent normative rankings even in the absence of stochastic dominance. Constructing these rankings requires mapping each observation to its implied within-group rank and group rank using $\tilde{q}(u, v)$.

Taken together, these uses illustrate the versatility of the two-dimensional framework. The surface $\tilde{q}(u, v)$ serves as the empirical primitive: once it is known, any welfare evaluation reduces to a simple weighted integral.

3 Econometric Representation

This section generalizes the baseline model in Equation (1) to include covariates and shows that the two-dimensional structure is identified by a two-level conditional quantile function. Within each group, conditional quantile functions characterize the distribution of individual outcomes given the covariates, while conditional quantiles of these group-level quantities capture the between-group component.

This setup applies to settings in which researchers seek to estimate the outcome surface or to characterize the distributional effects of covariates. Assuming a linear specification, the resulting two-dimensional conditional quantile model is

$$Q(v, Q(u, y_{ij} \mid x_{1ij}, x_{2j}, v_j) \mid x_{1ij}, x_{2j}) = x'_{1ij}\beta(u, v) + x'_{2j}\gamma(u, v) + \alpha(u, v), \qquad (3)$$

where x_{1ij} contains individual-level covariates, x_{2j} group-level covariates, and $\alpha(u, v)$ is an intercept.

I refer to equation (3) as the two-dimensional conditional quantile function: the conditional quantile (indexed by v) of the group-specific conditional quantiles (indexed by u). This is the main econometric object studied in the paper. For group j, $Q(u, y_{ij} \mid x_{1ij}, x_{2j}, v_j)$ denotes the u-th conditional quantile of the individual outcome given covariates and group membership, while the outer operator $Q(v, \cdot \mid x_{1ij}, x_{2j})$ takes the v-th conditional quantile across groups, conditional on the covariates.[11]

[11] The unconditional two-dimensional quantile function is obtained by integrating out the covariates. Moreover, the two-dimensional conditional quantile function fully determines the conditional distribution of y:

$$F_{Y|X}(y \mid x) = \int_0^1 \int_0^1 \mathbf{1}\{q(x, u, v) \leq y\} \, dv \, du. \qquad (4)$$

Inverting this expression yields the standard one-dimensional conditional quantile function, so no information is lost by modeling both dimensions.

All coefficients are allowed to vary freely along both dimensions, capturing heterogeneity within groups (u) and across groups (v), and admit a distributional interpretation. $\beta(u, v)$ describe how a one-unit change in x_{1ij} shifts the two-dimensional conditional quantile function. Fixing u, the function $v \mapsto \beta(u, v)$ describes how x_{1ij} affects the distribution of the group-level u-quantiles (e.g., group medians when $u = 0.5$). Fixing v, the function $u \mapsto \beta(u, v)$ describes how x_{1ij} shifts the v-quantile envelope of the group-level quantile functions. The interpretation of $\gamma(u, v)$ is analogous.[12] Taken together, the model generates rich patterns of distributional heterogeneity, which can be aggregated under explicit welfare criteria, as discussed in Section 2.

Remark 1 (Distributional vs. structural interpretation)**.** Equation (3) is defined entirely in terms of conditional quantiles and has a *distributional* interpretation: it characterizes how the conditional outcome distribution varies along the within-group index u and the across-group index v. Accordingly, the conditional quantile function admits the usual interpretation in terms of the subpopulation at the corresponding conditional quantiles (see, e.g., Hoderlein and Mammen, 2007; Sasaki, 2015). No assumption that latent heterogeneity is scalar, nor any monotonicity or rank-invariance restriction, is needed for this interpretation or for the welfare analysis in Section 2.

If one additionally wishes to interpret the coefficients in (3) as effects for units at fixed ranks (u, v) across counterfactual changes in regressors, one may impose additional structure that delivers a structural representation (see Matzkin, 2003; Torgovitsky, 2015).

In particular, suppose the outcome of individual i in group j admits the representation $y_{ij} = q(x_{1ij}, x_{2j}, u_{ij}, v_j)$ with a scalar $u_{ij} \in (0, 1)$ and suppose there exists a scalar index $v_j(u)$ such that $q(x_{1ij}, x_{2j}, u, v_j) = \tilde{q}(x_{1ij}, x_{2j}, u, v_j(u))$, with $\tilde{q}(\cdot)$ strictly increasing in u and in $v_j(u)$ conditional on (x_{1ij}, x_{2j}).[13] This representation implies rank invariance along both dimensions, so that the coefficients in Model (3) can be interpreted as individual-level

[12]Because group heterogeneity is multidimensional, the scalar index v orders groups by their u-conditional quantiles, and this ordering may vary with u. In addition, as in standard quantile models, individual-level covariates allow ranks to differ even within groups. For example, a region may occupy a different position in the distribution for high- versus low-educated individuals.

[13]After fixing u and (x_{1ij}, x_{2j}), it is always possible to find a scalar-valued function $v_j(u)$ that ranks groups; hence, one can always define the cross-group distribution of group-level u-quantiles. However, without imposing the additional single-index and monotonicity structure in this remark, this $v_j(u)$ need not satisfy $q(x_{1ij}, x_{2j}, u, v_j) = \tilde{q}(x_{1ij}, x_{2j}, u, v_j(u))$.

effects for individuals at ranks (u, v).[14] Such individual effects can then be summarized under alternative group orderings, such as the median rank or the welfare-based ordering in Section 2, and used to design optimal treatment rules. The Supplemental Material discusses this last extension and shows how the two-dimensional structure allows decision makers to exploit treatment-effect heterogeneity both within and between groups.

4 Estimation

Estimation proceeds via a two–step quantile regression that mirrors the within– and between–group structure of the model. For simplicity, I consider the same $\#\tau$-dimensional vector of quantiles to model both dimensions, although this is not a requirement.

The first stage consists of group-by-group quantile regressions: for each group j and quantile u, estimate the quantile regression of the outcome on the individual-level covariates x_{1ij} and save the fitted values. Formally, for each (j, u):

$$\hat{\beta}_j(u) = \left(\hat{\beta}_{1,j}(u), \hat{\beta}_{2,j}(u)'\right)' = \underset{(b_1,b_2)}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \rho_u(y_{ij} - b_1 - x'_{1ij}b_2), \tag{5}$$

where $\rho_u(x) = (u - 1\{x < 0\})x$ for $x \in \mathbb{R}$ is the check function. For group j, the fitted values $\hat{y}_{ij}(u) = \hat{\beta}_{1,j}(u) + x'_{1ij}\hat{\beta}_{2,j}(u)$ are estimates of the u-conditional quantile function $Q(u, y_{ij} \mid x_{ij}, v_j)$, where the effect of the group-level regressors is absorbed by the intercept.

In the second stage, for each u, regress the first-stage fitted values on all covariates using the full sample, again via quantile regression for each v. Specifically, for all (u, v):

$$\hat{\delta}(\hat{\beta}(u), v) = \underset{(a,b,g)}{\arg\min} \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} \rho_v(\hat{y}_{ij}(u) - x'_{2j}g - x'_{1ij}b - a), \tag{6}$$

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

Estimation requires $\#\tau \times m$ first-stage quantile regressions and $\#\tau \times \#\tau$ second-stage quantile regressions. The procedure relies only on standard quantile regression routines and is implemented in an R package. While the absence of a closed-form solution can increase computing time, recent algorithms allow simultaneous estimation of multiple

[14] In this setting, rank invariance means that the ranking of individuals within groups and the ranking of groups at each value of u are unaffected by changes in the regressors. To identify the effect of a specific regressor, the corresponding rank variables must remain invariant to changes in that regressor so that the same individuals and groups are indexed across counterfactuals. When rank invariance fails, individual treatment effects are not identified; see Chernozhukov et al. (2023) for alternative approaches based on conditional prediction intervals.

quantiles, greatly improving speed. The procedure is also highly parallelizable: all first-stage regressions run independently across groups, and the second stage can be parallelized over u.[15]

Remark 2 (**Endogenous Regressors: Control Function Extensions**)**.** Model (3) assumes that both x_{1ij} and x_{2j} are exogenous, so that quantile-regression in both stages identifies the parameters. If this assumption fails, the framework can be extended to allow for endogeneity. For example, when the endogenous regressor exhibits a triangular structure, the model can be embedded in a control function approach instead (see e.g., Chernozhukov et al., 2020 and Newey and Stouli, 2025). A full development of this extension is beyond the scope of this paper and is left for future research.

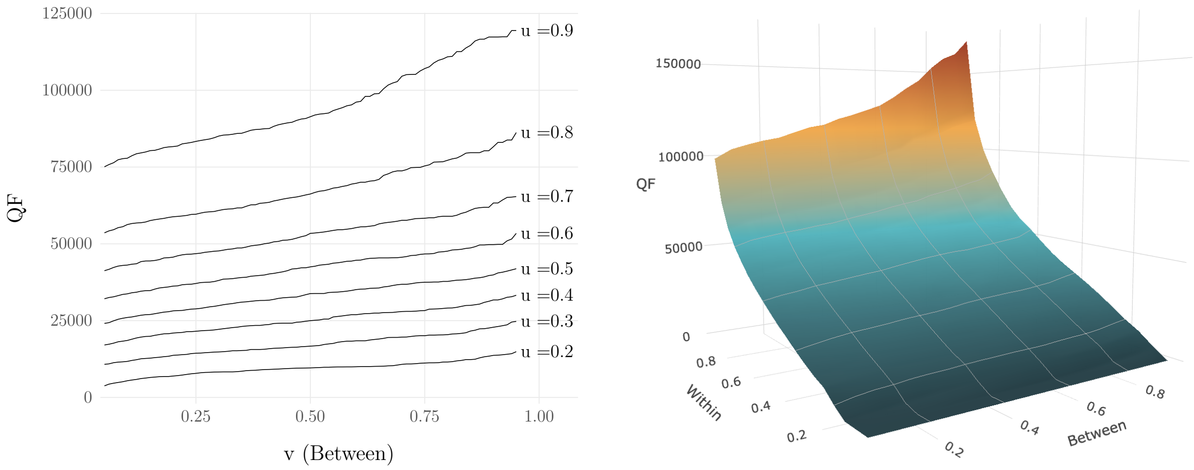
Example 3 (Continuation of Example 1)**.** Returning to Example 1, we can look at the estimation in the setting with no regressors. The goal is to analyze income heterogeneity within and between US labor markets. With no covariates, the first stage reduces to computing sample percentiles within each group. In the second stage, I then take the percentiles across groups of these group-level percentiles, using the grid $\{0.05, 0.06, \ldots, 0.99\}$ for both stages. Since the first-stage fitted values are constant within groups, the data can be collapsed to the group level to speed up computation. In this case, weights should be included in the second step.[16]

Figure 3 provides two representations of the two-dimensional quantile function. Panel (a) plots selected slices $v \mapsto \tilde{q}(u, v)$ for fixed within-group percentiles u, while Panel (b) displays the corresponding 3D surface. Both representations reveal substantial heterogeneity along both dimensions. While differences between labor markets appear especially pronounced at the top, sizeable disparities already emerge in the lower tail—for instance, some regions exhibit third-decile incomes nearly twice as large as those of others. This pattern highlights substantial differences between labor markets, particularly in the tails of the within distribution, with far less divergence in its center.

[15] Ensuring monotonicity of the estimated two-level quantile functions may require a rearrangement procedure (Chernozhukov et al., 2009, 2010). Given the nested structure, rearrangement along u should follow the first stage. If monotonicity holds in the first stage for all groups, the second-stage regressions remain monotonic in u. Rearrangement along v can be applied after the second stage.

[16] All estimates use individual person weights.

Figure 3: Two-Dimensional Quantile Function



(a) Slices (b) 3D Surface

Notes: Figure 3 displays the two-dimensional quantile function of yearly income across U.S. commuting zones. Panel (a) plots selected slices $u \mapsto \tilde{q}(u, v)$ at fixed within-group percentiles, while Panel (b) shows the corresponding 3D surface representation.

5 Asymptotic Theory

This section develops the asymptotic properties of the two-step estimator. I establish uniform consistency and weak convergence of the two-dimensional quantile process, and I show that a clustered bootstrap procedure provides valid inference. The proofs of the main results are in Appendix B.

Notation - Let $\tau = (u, v)$. For each j, let $\beta_{j,0}(u)$ denote the true first-stage coefficient vector, and define the stacked vector $\beta_0(u) = \left(\beta_{1,0}(u)', \ldots, \beta_{m,0}(u)'\right)'$. Let $\delta_0(\beta_0, \tau)$ denote the true second-stage parameter vector. To simplify notation, I suppress the dependency of δ and β_j on u and v, unless necessary. For a random variable h_{ij}, $E_{i|j}[h_{ij}]$ denotes the expectation of h_{ij} 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, define $K = K_1 + K_2 + 1$ to be the total number of regressors. Let the $(K_1 + 1)$-dimensional vector of first stage regressors be $\tilde{x}_{ij} = (1, x'_{1ij})'$. I prove weak convergence of the whole quantile regression process for $\tau \in \mathcal{T} \times \mathcal{T}$, where $\mathcal{T}$ is a compact subset of $(0, 1)$. The symbol $\ell^\infty(\mathcal{T} \times \mathcal{T})$ denotes the set of component-wise bounded vector-valued functions on $\mathcal{T} \times \mathcal{T}$ and $\rightsquigarrow$ denotes weak convergence.

5.1 Consistency and Asymptotic Normality

The derivation of asymptotic results faces two primary challenges: the non-smoothness of the quantile regression objective function, the generated dependent variable in the second stage, and the increasing dimension of the first stage as the number of groups diverges to infinity. Several studies have addressed the asymptotic properties of estimators with non-smooth objective functions, leveraging the smoothness of the limiting objective function (see, for example, Newey and McFadden, 1994). Notably, Pakes and Pollard (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 in Chen et al. (2003). Similarly to their paper, my second-stage parameter vector depends on a preliminary first-stage whose dimension increases with the sample size. To this end, I start by making the necessary assumptions 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**). *(i) The processes $\{(y_{ij}, x_{ij}) : i = 1, \dots, n\}$ are i.i.d. across j. (ii) For each j, the observations (y_{ij}, x_{ij}) are i.i.d. across i.*

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 $E_{i|j}[\tilde{x}_{ij}\tilde{x}'_{ij}]$ and $E[x_{ij}x'_{ij}]$ are bounded away from zero and infinity uniformly across j.*

Assumption 3 (**Individual-level heterogeneity**). *The distribution $F_{y_{ij}|x_{ij},v_j}(y|x, v)$ is twice differentiable with respect to y, with the corresponding derivatives $f_{y_{ij}|x_{ij},v_j}(y|x, v)$ and $f'_{y_{ij}|x_{ij},v_j}(y|x, v)$. Further, assume that*

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

and

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

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

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

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

These are standard assumptions in the quantile regression literature. Assumption 1, assumes that the observations are i.i.d. within and between groups. Assumption 2 requires that the regressors are bounded and that both matrices $E_{i|j}[\tilde{x}_{ij}\tilde{x}'_{ij}]$ and $E[x_{ij}x'_{ij}]$ 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 allows us to apply Lemma 3 in Galvao et al. (2020), which provides the Bahadur representation of the first-stage estimator.

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

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

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

and

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

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

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

Assumption 7 (**Compact parameter space**)**.** *For all* τ*,* $\beta_{j,0}(u) \in int(\mathcal{B}_j)$*, where* $\mathcal{B}_j$ *is a compact subset of* $\mathbb{R}^{K_1+1}$*, and* $\delta_0(\beta_0, \tau) \in \mathcal{D}$*, where* $\mathcal{D}$ *is a compact subset of* $\mathbb{R}^K$*. Moreover, there exists* $\eta > 0$ *such that* $\inf_{j=1,\ldots,m,\ u \in \mathcal{T}} \text{dist}\big(\beta_{j,0}(u), \partial\mathcal{B}_j\big) \geq \eta$*.*

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

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

Since quantile regression is consistent but not unbiased, we need the number of observations per group to diverge to infinity. At the same time, the second-stage quantile regression exploits the heterogeneity between groups, which is determined by the heterogeneity of the group-level quantiles, a group-specific term. Thus, the number of groups must also diverge.

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

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

I show that the relative growth rate in Assumption 9(a) is sufficient for consistency of the estimator. This condition is exceptionally weak, as the number of observations per group can increase at an almost arbitrarily slow rate. Differently, weak convergence requires assumption 9(b). This second growth rate is relatively mild as the number of observations per group must grow faster than the square root of the number of groups.[17] It is worth noting that in many empirical applications, groups are defined by industries or geographical units, such as counties, which typically contain a large number of individuals, so that the results in this section provide a useful approximation. Similarly, one

[17] In particular, this condition is weaker than the growth rate required in the two-step quantile model with a linear second stage studied by Chetverikov et al. (2016). Until recently, there has been a substantial gap in the required rate of growth of n relative to m between nonlinear estimators with smooth objective functions and those with non-smooth ones, such as quantile regression. For unbiased asymptotic normality, the bias must decrease more quickly than $1/\sqrt{m}$. Using new results in Volgushev et al. (2019) and Galvao et al. (2020), I show that $m(\log n)/\sqrt{n} \to 0$ is sufficient to ensure unbiased asymptotic normality of the estimator. Galvao et al. (2020) require the stronger condition $m(\log n)^2/n \to 0$, because in their setting the bias must converge more quickly than $1/\sqrt{mn}$ to establish unbiased asymptotic normality.

may ask whether the required growth rate could be relaxed by using smoothed quantile regression and/or bias correction after the first stage. Existing results in the literature suggest that smoothing does not relax the underlying assumptions needed for the asymptotic theory. While smoothing may improve finite-sample performance, it does so at the cost of introducing a tuning parameter. Moreover, the bias-corrected smoothed quantile regression estimator for panel data proposed in Galvao and Kato (2016) is not applicable in the present setting, as it relies on coefficient homogeneity across groups. In recent work, Franguridi et al. (2025) derive an explicit expression for the bias of the leading term in the expansion. However, implementing such a bias correction remains challenging in practice, as it requires estimating objects involving higher-order derivatives, and the selection of tuning parameters becomes even more delicate. A systematic exploration of these approaches is therefore left for future research.

The first result of this paper states weak uniform consistency of the two-step estimator.

Theorem 1 (Uniform Consistency)**.** *Let assumptions 1-8 and 9(a) be satisfied. Then as* $m \to \infty$

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

If the first-stage parameter vector $\beta_0(u)$ were known, the true second-stage parameter vector $\delta_0(\beta_0, \tau)$ would uniquely satisfy:[18]

$$M(\delta_0, \beta_0, \tau) = E[m(\delta_0, \beta_0, \tau)] = 0$$

with $m(\delta, \beta, \tau) = x_{ij}[v - 1(\tilde{x}'_{ij}\beta_j(u) \leq x'_{ij}\delta(\beta, \tau))]$.

While $M(\delta, \beta, \tau)$ is a smooth function, this property does not extend to its sample counterpart $M_{mn}(\delta, \beta, \tau) = \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} m(\delta, \beta, \tau)$. To make the problem tractable, I begin by showing that $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)$ can be approximated by the linear function

$$\Gamma_1(\delta_0, \beta_0, \tau)^{-1} \left(\frac{1}{m} \sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau) \left[\hat{\beta}_j(u) - \beta_{j,0}(u) \right] + M_{mn}(\delta_0, \beta_0, \tau) \right) \tag{7}$$

where

$$\Gamma_1(\delta, \beta_0, \tau) = E\left[f_{Q(u, y_{ij}|x_{ij}, v_j)|x_{ij}} \left(x'_{ij}\delta(\beta_0, \tau) \mid x_{ij} \right) x_{ij}x'_{ij} \right]$$

[18] Under weak regularity conditions.

denotes the derivative of $M(\delta, \beta_0, \tau)$ with respect to δ, while the term with Γ_{2j} is the pathwise derivative of $M(\delta, \beta_0, \tau)$ in the direction of $(\hat{\beta}_j - \beta_{j,0})$:

$$\Gamma_{2j}(\delta, \beta_0, \tau) = E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}\left(x'_{ij}\delta(\beta_0, \tau) \mid x_{ij}\right)x_{ij}\tilde{x}'_{ij}\right].$$

So this component reflects how the first-stage estimation error propagates into the second stage, and $M_{mn}(\delta_0, \beta_0, \tau)$ represents the sampling variation that would arise even if the first stage were known.

Lemma 6 and its corollary in the Supplemental Material show that $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)$ is asymptotically equivalent to the linear approximation in Equation (7), up to a remainder term that converges to zero sufficiently fast, uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$. Consequently, establishing weak convergence of $\hat{\delta}(\hat{\beta}, \tau)$ reduces to studying the asymptotic behavior of the linear function. The following Lemma characterizes the asymptotic properties of the terms in equation (7).

Lemma 2. *Let the model in equation (3) and assumptions 1-7 hold. Then*

(i) Under assumption 9(b), as $m \to \infty$:

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

(ii) As $m \to \infty$:

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

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

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

(iii) Under assumption 9(b), as $m \to \infty$:

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

Lemma 2 shows that the first-stage error converges to zero at a faster rate than the standard deviation of the second stage. Given that the asymptotic distribution of the estimator can be approximated by a linear function of the sum of two terms converging at different rates, the asymptotic behavior will be determined by the term converging at a slower rate, and in the first-order asymptotic distribution, only the second stage matters. Similar results are documented in Chetverikov et al. (2016) and in Melly and Pons (2025).

Given these results, weak convergence follows by asymptotic negligibility of the first-stage error, implying that the limiting distribution of $\hat{\delta}(\hat{\beta}, \tau)$ is identical to that of the infeasible estimator $\hat{\delta}(\beta_0, \tau)$.

Theorem 2 (**Weak Convergence**). *Let assumptions 1-8 and 9(b) be satisfied. Then*

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

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

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

Theorem 2 shows that the entire coefficient vector converges at rate $1/\sqrt{m}$, even though estimation uses mn observations. This slower rate is not specific to the proposed quantile framework but reflects the fact that heterogeneity is identified from cross-group rather than within-group variation. If coefficients on the individual-level regressors were restricted to be common across groups, they could be estimated at the faster $1/\sqrt{mn}$ rate. Allowing $\beta(u, v)$ to vary with v shifts identification to cross-group differences, which necessarily determine the $1/\sqrt{m}$ convergence rate. The same principle arises in least squares: fixed-effects estimators achieve faster rates by exploiting within-group variation but cannot recover between-group heterogeneity.

5.2 Inference

Inference is conducted using a clustered bootstrap procedure that resamples entire groups with replacement. Similar procedures have been suggested in Liao and Yang (2018), Lu and Su (2023), and Fernández-Val et al. (2022).

Because resampling occurs at the group level, the first stage remains unchanged and need not be recomputed. The procedure is thus equivalent to resampling the first-stage fitted values. The following theorem establishes that the bootstrap consistently replicates the distribution of $\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)$. Consequently, inference can be conducted using the bootstrap distribution of $\hat{\delta}^*(\hat{\beta}, \tau) - \hat{\delta}(\hat{\beta}, \tau)$, where $\hat{\delta}^*(\hat{\beta}, \tau)$ denotes the bootstrap estimator. The covariance matrix can then be estimated, for instance, from the variance of the bootstrap replicates or from the interquartile range rescaled by the normal distribution.

Theorem 3 (**Validity of the Bootstrap**). *Assume that the conditions for Theorem 2 are satisfied. Then,*

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

where $\rightsquigarrow^*$ denotes conditional weak convergence of the bootstrap process in $\ell^\infty(\mathcal{T} \times \mathcal{T})$, in probability.

Remark 3 (Heterogeneity, convergence rates, and bootstrap validity)**.** Assumption 5 ensures sufficient heterogeneity across groups so that the second-stage quantile regression is non-degenerate. When this condition holds, cross-group variation generates sampling uncertainty of order $m^{-1/2}$ and determines the estimator's first-order asymptotic distribution.

It is useful to note how the proposed group bootstrap treats the first stage. In the sample linear representation, first-stage estimation error enters through

$$m^{-1} \sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau)\left(\hat{\beta}_j(u) - \beta_{j,0}(u)\right),$$

and the corresponding bootstrap linear representation contains the same term with bootstrap weights attached to each group contribution. Hence, the bootstrap reweights the realized first-stage contribution at the group level in the same way as the leading empirical process term.

In the absence of between-group heterogeneity, the second-stage quantile regression—conditional on the first stage—would exhibit no additional sampling variation: all randomness would originate from first-stage estimation error, and the estimator would be expected to converge at the faster $1/\sqrt{mn}$ rate. This is the case in Melly and Pons (2025), who show that with a linear second step, cluster-robust inference adapts to the convergence rate and accounts for first-stage estimation error. The present paper establishes bootstrap validity for the heterogeneity-driven asymptotic regime. In this regime, the first-stage contribution is of smaller order than the $m^{-1/2}$ fluctuations induced by cross-group heterogeneity and therefore does not affect first-order inference.

Given the large number of estimated coefficients, researchers may wish to test hypotheses involving multiple parameters—for example, whether the effect of a covariate is constant across quantiles. Kolmogorov–Smirnov and Cramér–von Mises type tests are suitable for this purpose. In addition, functional confidence bands providing simultaneous coverage for the entire process can be constructed by inverting the acceptance region of the Kolmogorov–Smirnov statistic. The Supplemental Material provides details on all the procedures.

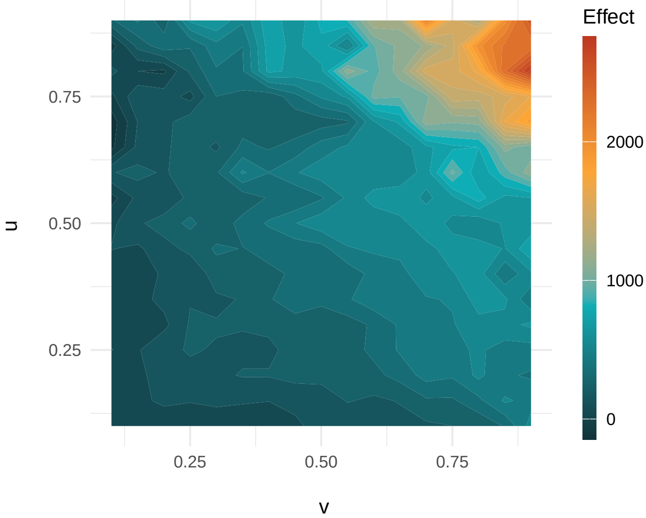
Monte Carlo simulations in the Supplemental Material show strong finite-sample performance of both the estimation and inference procedures. Although $\hat{\beta}(u, v)$ and $\hat{\gamma}(u, v)$ share the same asymptotic distribution, they exhibit distinct finite-sample behavior. The bias of $\hat{\beta}$ declines with both the number of groups m and the within-group size n, whereas the bias of $\hat{\gamma}$ falls mainly as m increases. For both vectors, variance decreases with m; in addition, the variance of $\hat{\beta}$ also declines with n. Overall, gains from increasing n taper quickly, for both bias and variance. Finally, the 95% bootstrap confidence intervals achieve coverage close to the nominal level.

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

This section complements the findings of McKenzie and Puerto (2021) by examining how business training shifts the distribution of income from work among female-owned businesses, both within and across markets. While their study examines the effects of business training on female-owned firms and potential spillovers to others, I use the two-level quantile framework to capture distributional effects that vary both with a firm's position within its local market and with the relative rank of the market itself.[19] The sample consists of 3,537 female-owned businesses across 157 rural markets in Kenya. Training was assigned through a two-stage randomization. First, 93 markets were ran-

[19]The training program lasted five days and covered bookkeeping, recordkeeping, marketing, financial concepts, costing and pricing, and the development of new business ideas.

Figure 4: Effect of Training Assignment on Income from Work



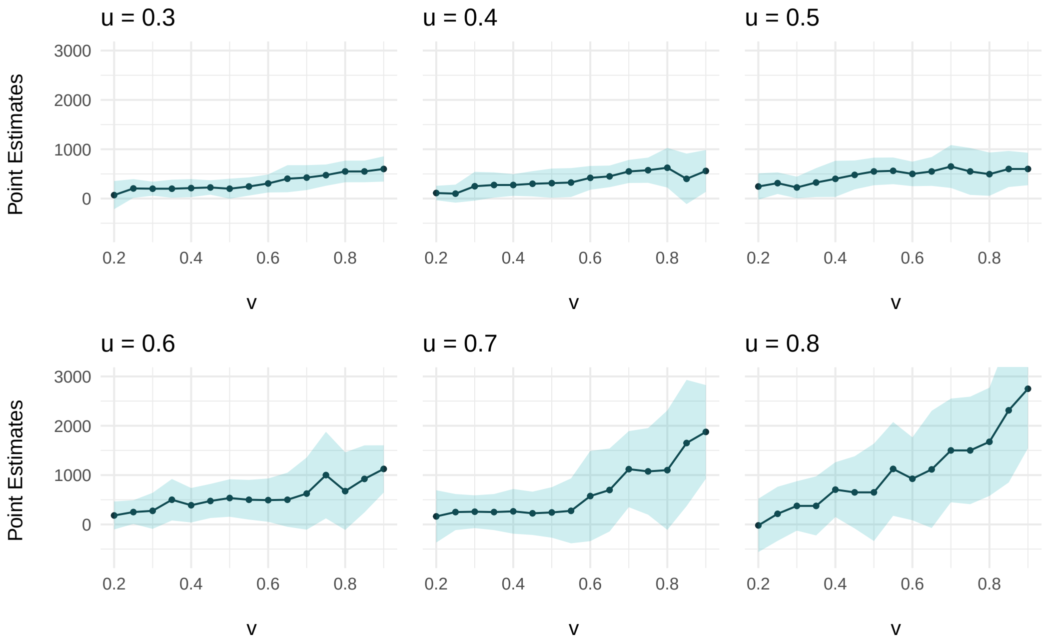
Notes: The figure shows the treatment effect on the within-market u and the between-market distribution v of income from work.

domly selected for treatment, and 64 served as controls. Within treated markets, firms were further randomized into training or control businesses, yielding 1,172 treated firms and 2,365 controls. Randomization was stratified by region, market size, and weekly profit quartiles. Most firms operate in retail (e.g., fruits, vegetables, grains), while a substantial share run restaurants or tailoring businesses. They are generally small-scale operations: only about 20% report having any employees.

McKenzie and Puerto (2021) find that training improves business survival after three years and increases average weekly sales and profits by 18 and 15 percent, respectively. Firm owners also report better mental health and a higher subjective standard of living. By contrast, spillover effects in non-participating businesses in treatment markets remain unclear, with point estimates being small and not statistically significant. In addition, they estimate distributional effects of training on profits and sales using data from two follow-up surveys conducted three years after the program. They find that the gains are concentrated in the upper tail of the outcome distribution.

For my analysis, I use data from the same two follow-up waves. When a business is observed in both waves, I average its outcomes across them. Groups are defined at the market level, and to ensure sufficient observations for estimation, I exclude markets with fewer than 15 businesses (results are similar with alternative cutoffs). The final dataset comprises 2,922 firms across 116 markets, with an average of 27 firms per market.

Figure 5: Effect of Training Assignment on Income From Work



Notes: The figure shows the effect of the treatment on the between-market distribution v of income from work for selected quantiles of the within-market distribution u. The shaded areas show the 95% confidence intervals estimated using clustered bootstrap standard errors computed with 1,000 replications.

I use income from work of the business owner as the main outcome of interest.[20] In the sample, the average income from work amounts to 2,300 Kenyan shillings.[21] The dependent variable exhibits a mass point at zero, partly because firms that exit the market are coded with zero outcomes. In the final dataset, about 11% of firms did not survive after three years, and 10% of owners report no income. To avoid distortions from this censoring, I refrain from estimating effects too far into the lower tail. Since these mass points can undermine inference, I do not report confidence intervals for quantiles affected by the issue.[22] Beyond the lower tail, results and inference remain unaffected by this mass point. I estimate the following model:

$$Q(v, Q(u, y_{ij}|d_{ij}, s_{ij}, v_j)|d_{ij}, s_{ij}) = \beta_1(u, v) \cdot d_{ij} + \beta_2(u, v) \cdot s_{ij} + \alpha(u, v) \tag{8}$$

where y_{ij} denotes the income of the owner of firm i in market j, d_{ij} is a treatment indicator, and s_{ij} is a spillover dummy equal to 1 for firms in treatment markets that

[20] Results are similar when using weekly sales or weekly profits.

[21] In December 2025, 1,000 Kenyan shillings is equivalent to about 7.75 USD.

[22] If the second-stage fitted values equal zero for at least one observation, I classify the cell as affected by the mass point. Fitted values of zero suggest a perfect fit for some observations.

were not assigned to training.

Figure 4 shows the treatment effects estimates over the two dimensions for the quantiles indices $\{0.1, 0.15, \ldots, 0.9\}$. The vertical axis represents the within-market dimension (u), while the horizontal axis captures the between-market dimension (v). The heatmap reveals that treatment effects are close to zero at the lower end of both distributions. As v increases, effects become larger, indicating that firms in better-performing markets benefit more from training. The gradient is especially pronounced at higher values of u, where firms already in the upper tail within their markets experience the strongest gains. This pattern highlights that business training not only amplifies differences across markets but also widens gaps within markets, with the largest benefits accruing in the upper tail of both distributions. Figure 5 presents the same results in an alternative format, plotting point estimates together with 95% confidence intervals across the distribution of markets (v) for fixed values of the within-market quantile u. The estimates confirm that heterogeneity along both dimensions is central: the effects are consistently larger in the upper tails. For example, at $u = 0.7$, the effect rises from about 200 Kenyan shillings in weaker markets to over 1,000 in stronger ones. Even within prosperous markets, the within-market dimension matters: the effect grows from roughly 600 shillings at $u = 0.3$ to more than 2,000 at $u = 0.8$. These patterns point to complementarities between individual ability and market quality, suggesting that both dimensions jointly determine who benefits most from training.

Consistent with these patterns, Kolmogorov–Smirnov and Cramér–von Mises tests reject the null hypothesis of treatment effect homogeneity over (u, v) at conventional levels (p-values 0.006 and 0.024), confirming the presence of substantial heterogeneity. Moreover, Table 4 in the Supplementary Material shows that group ranks are not preserved across the within distribution, showing that a single univariate rank would miss important features of the data.

Finally, to assess the impact of the training program, I compute welfare under both the realized outcome and a counterfactual scenario without the intervention, and calculate the percentage change in welfare attributable to the treatment. I use the welfare function in equation (2) and consider four alternative weighting functions capturing different trade-offs between within- and between-group inequality. The results are in Table 1. Abstracting from training costs, the treatment leads to 11.5-15.3% higher welfare, as the treatment effect is positive, and there is no evidence of negative spillovers. Moreover, the welfare gain is larger when the policymaker assigns more weight to within-market inequality, likely because the training program increases the disparity between markets.

Table 1: Welfare Gains

$w(u,v)$	Welfare Gain (%)	
1	15.33	(13.98, 16.98)
$2\big(0.2(1-u)+0.8(1-v)\big)$	11.53	(10.91, 13.76)
$2\big(0.5(1-u)+0.5(1-v)\big)$	13.14	(12.46, 15.40)
$2\big(0.8(1-u)+0.2(1-v)\big)$	15.16	(14.46, 17.40)
$4(1-u)(1-v)$	11.66	(11.57, 14.60)

Notes: The table reports the percentage change in welfare under the treatment scenario relative to the counterfactual without treatment. Welfare is computed using equation (2). 95% Confidence intervals in parentheses are obtained using the percentile bootstrap (1,000 replications).

7 Conclusion

Distributional effects are central to understanding treatment effect heterogeneity. In economics, such heterogeneity arises both within and between groups—such as geographical regions, industries, or firms—and both dimensions are welfare-relevant and often interdependent. Yet most existing work either focuses on one dimension at a time or relies on within–between decompositions that, by construction, cannot capture interactions across the two dimensions.

This paper develops an econometric framework to analyze within- and between-group heterogeneity simultaneously, while remaining agnostic about the policymaker's objective. It proposes a parsimonious representation of these joint distributional features through a two-dimensional quantile surface that ranks individuals within groups and ranks groups at each within-group quantile, thereby addressing the challenge of ranking heterogeneous groups. Specifically, constructing the model involves a trade-off between a simple specification with a unique group rank and a more flexible model that allows for unrestricted heterogeneity and group ranks that vary across the within-group distribution. I adopt the latter, as it provides a more realistic representation of group differences.

By modeling how inequality varies across dimensions, the framework offers a transparent foundation for welfare analysis and policy evaluation. More broadly, it is relevant in a wide range of settings where within- and between-group inequality matters, including the evaluation of policies targeting convergence across regions, opportunity gaps across circumstances, and wage dispersion within and between industries or occupations.

References

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

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

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

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

Atkinson, A. B. and F. Bourguignon (1987): "Income Distribution and Differences in Needs," in *Arrow and the Foundations of the Theory of Economic Policy*, ed. by G. R. Feiwel, London: Palgrave Macmillan UK, 350–370.

Autor, D., D. Dorn, and G. Hanson (2019): "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men," *American Economic Review: Insights*, 1, 161–178.

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

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

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

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

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

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

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

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

Carroll, G. (2025): "Is Equal Opportunity Different from Welfarism?" .

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

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

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

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

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

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

Chernozhukov, V., I. Fernández-Val, W. K. Newey, S. Stouli, and F. Vella (2020): "Semiparametric estimation of structural functions in nonseparable triangular models," *Quantitative Economics*, 11, 503–533.

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

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

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

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

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

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

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

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

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

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

Fleurbaey, M. and F. Maniquet (2018): "Optimal income taxation theory and principles of fairness," .

Franguridi, G., B. Gafarov, and K. Wüthrich (2025): "Bias correction for quantile regression estimators," *Journal of Econometrics*, 251.

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

Frumento, P., M. Bottai, and I. Fernández-Val (2021): "Parametric Modeling of Quantile Regression Coefficient Functions With Longitudinal Data," *Journal of the American Statistical Association*, 116, 783–797.

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

Galvao, A. F., J. Gu, and S. Volgushev (2020): "On the unbiased asymptotic normality of quantile regression with fixed effects," *Journal of Econometrics*, 218, 178–215.

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

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

Glaeser, E. L., M. Resseger, and K. Tobio (2009): "Inequality in cities," *Journal of Regional Science*, 49, 617–646.

Haltiwanger, B. J., H. R. Hyatt, and J. R. Spletzer (2024): "Rising Top, Falling Bottom : Industries and Rising Wage Inequality," *American Economic Review*, 114, 3250–3283.

Haushofer, J., P. Niehaus, C. Paramo, E. Miguel, and M. W. Walker (2025): "Targeting Impact Versus Deprivation," *American Economic Review*, 1936–1974.

Heikkuri, V.-M. and M. Schief (2025): "Subgroup Decomposition of the Gini Coefficient: A New Solution to an Old Problem," .

Helpman, E., O. Itskhoki, M. A. Muendler, and S. J. Redding (2017): "Trade and inequality: From theory to estimation," *Review of Economic Studies*, 84, 357–405.

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

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

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

Kehrig, M. and N. Vincent (2025): "Good Dispersion, Bad Dispersion," *Review of Economic Studies*.

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

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

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

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

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

Liu, X. (2024): "A quantile-based nonadditive fixed effects model," *Working Paper*, 1–34.

Lu, X. and L. Su (2023): "Uniform inference in linear panel data models with two-dimensional heterogeneity," *Journal of Econometrics*, 235, 694–719.

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

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

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

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

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 (2025): "Minimum Distance Estimation of Quantile Panel Data Models," *Working Paper*.

Milanovic, B. (2002): "True world income distribution, 1988 and 1993: First calculation based on household surveys alone," *Economic Journal*, 51–92.

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

Newey, W. K. and S. Stouli (2025): "Identification of Treatment Effects under Limited Exogenous Variation," .

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

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

Roemer, J. E. (1998): *Equality of Opportunity.*

Roemer, J. E. and A. Trannoy (2016): "Equality of opportunity: Theory and measurement," *Journal of Economic Literature*, 54, 1288–1332.

Ruggles, S., S. Flood, M. Sobek, D. Backman, G. Cooper, J. A. Rivera Drew, S. Richards, R. Rodgers, J. Schroeder, and K. C. W. Williams (2025): "IPUMS USA: Version 16.0," .

Saez, E. and S. Stantcheva (2016): "Generalized social marginal welfare weights for optimal tax theory," *American Economic Review*, 106, 24–45.

Sasaki, Y. (2015): "What Do Quantile Regressions Identify For General Structural Functions?" *Econometric Theory*, 31, 1102–1116.

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

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

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

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

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

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

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

A Proof of Characterization of $\mathcal{W}$-equivalence

Proof of Lemma 1. (i) Immediate from $q_1 = q_2$ a.e. and linearity of the integral. (ii) Let $d = q_1 - q_2$. If $d \neq 0$ on a set of positive measure, then either $A_+ = \{d > 0\}$ or $A_- = \{d < 0\}$ has positive measure. Take $w = \mathbf{1}_{A_+}$ (or $\mathbf{1}_{A_-}$); then $w \in \mathcal{W}$ and $\int w\, d = \int_{A_+} d > 0$, so $W_w(q_1) \neq W_w(q_2)$. ∎

B Proofs of Main Asymptotic Results

B.1 Asymptotic Moments

Proof of Lemma 2. **Part (i)** Using equation (4) from Lemma 4 we can write:

$$\frac{1}{m}\sum_{j=1}^{m}\Gamma_{2j}(\delta_0,\beta_0,\tau)(\hat{\beta}_j-\beta_{j,0}) \tag{9}$$

$$=\frac{1}{m}\sum_{j=1}^{m}E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]\left(\frac{1}{n}\sum_{i=1}^{n}\phi_{j,u}(\tilde{x}_{ij},y_{ij})\right) \tag{10}$$

$$+\frac{1}{m}\sum_{j=1}^{m}E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(1)}(u) \tag{11}$$

$$+\frac{1}{m}\sum_{j=1}^{m}E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(2)}(u) \tag{12}$$

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, $f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})$ is also bounded uniformly over τ. It follows directly that the conditional expectation $E_{i|j}\big[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})$ $x_{ij}\tilde{x}_{ij}'\big]$ is bounded uniformly over τ.

Next, consider the third term (12). Together with equation (6), it implies that

$$\sup_{v\in\mathcal{T}}\sup_{u\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m}E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]R_{nj}^{(2)}(u)=O_p\left(\frac{\log n}{n}\right). \tag{13}$$

For the second term (11), Since $\mathrm{Var}\left(R_{nj}^{(1)}(\tau)\right)=o\left(\frac{1}{n}\right)$ by (8), the conditional expectation is bounded and since the observations are independent across groups, we have that

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

In addition, by (7), $\sup_j \sup_{u\in\mathcal{T}} E_{i|j}\left[R^{(1)}_{nj}(u)\right] = O\left(\frac{\log n}{n}\right)$ such that

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

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

$$\sup_{v\in\mathcal{T}}\sup_{u\in\mathcal{T}}\frac{1}{m}\sum_{j=1}^{m} E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x'_{ij}\delta_0|x_{ij})x_{ij}\tilde{x}'_{ij}\right] R^{(1)}_{nj}(u) = o_p\left(\frac{1}{\sqrt{m}}\right). \tag{14}$$

It follows that both (11) and (12) are $o_p\left(\frac{1}{\sqrt{m}}\right)$ uniformity over τ.

Consider now the first term (10):

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

This is a sample mean over mn independent observations denoted by $s_{ij}(\tau)$. The model for the group-level quantile functions implies that $E_{i|j}\left[1(y_{ij}\le\tilde{x}'_{ij}\beta_j(u))|\tilde{x}_{ij}\right] = u$, which gives $E[s_{ij}(\tau)] = 0$. In addition,

$$\begin{aligned}
\operatorname{Var}(s_{ij}(\tau)) =& E_j\left[\Gamma_{2j}(\delta_0,\beta_0,\tau)\operatorname{Var}(\phi_{j,u})\Gamma_{2j}(\delta_0,\beta_0,\tau)'\right]\\
=& E_j[\Gamma_{2j}(\delta_0,\beta_0,\tau)B^{-1}_{j,u}u(1-u)E_{i|j}[\tilde{x}_{ij}\tilde{x}'_{ij}]B^{-1}_{j,\tau}\Gamma_{2j}(\tau,\delta_0,\beta_0)'],
\end{aligned}\tag{15}$$

where $\Gamma_{2j}(\delta_0,\beta_0,\tau) = E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x'_{ij}\delta_0|x_{ij})x_{ij}\tilde{x}'_{ij}\right]$.

It follows from Chebyshev that

$$\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} s_{ij}(\tau) = O_p\left(\frac{1}{\sqrt{mn}}\right). \tag{16}$$

This last results implies that the first term (9) is $o_p\left(\frac{1}{\sqrt{m}}\right)$ pointwise.

To get uniform results, note that

$$\left\{E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x'_{ij}\delta_0|x_{ij})x_{ij}\tilde{x}'_{ij}\right]\left(\frac{-B^{-1}_{j,u}}{n}\sum_{i=1}^{n}\tilde{x}_{ij}(1(y_{ij}\le\tilde{x}'_{ij}\beta_j(u))-u)\right), \tau\in\mathcal{T}\times\mathcal{T}, \beta\in\mathcal{B}\right\}$$

is a Donsker class for any compact set $\mathcal{B}$. This follows by noting that $\{1(y_{ij} \leq \tilde{x}_{ij}'\beta_j(\tau)),$ $\tau \in \mathcal{T} \times \mathcal{T}, \beta \in \mathcal{B}\}$ is a VC subgraph class and hence a bounded Donsker class. Hence,

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{ij}(1(y_{ij} \leq \tilde{x}_{ij}'\beta_j(u) - u), \tau \in \mathcal{T} \times \mathcal{T}, \beta \in \mathcal{B} \right\}$$

is also bounded Donsker with a square-integrable envelope $2 \cdot \max_{i \in 1,\dots,n} |\tilde{x}_{ij}| \leq 2 \cdot C$. The whole function is then Donsker by the boundedness of $E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j))|x_{ij}}(x_{ij}'\delta_0|x_{ij})x_{ij}\tilde{x}_{ij}'\right]$ and $B_{j,u}^{-1}$. It follows that the equation (9) is $o_p\left(\frac{1}{\sqrt{m}}\right)$ uniformly in u and v.

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

Part (iii) By parts (i)–(ii), $\frac{1}{m}\sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau)(\hat{\beta}_j - \beta_{j,0}) = o_p(m^{-1/2})$ and $M_{mn}(\delta_0, \beta_0,$ $\tau) = O_p(m^{-1/2})$ uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$. Hence,

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

■

B.2 Uniform Consistency

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

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

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

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

and therefore that

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

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

Note that by the triangle inequality

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

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

B.3 Weak Convergence

Proof of Theorem 2. The proof has two steps: (i) show that replacing β_0 with $\hat{\beta}$ only affects $\hat{\delta}$ at order $o_p(m^{-1/2})$, and hence the feasible and infeasible estimators coincide asymptotically; (ii) show that $\Gamma_1(\cdot)\sqrt{m}(\hat{\delta}(\beta_0, \cdot) - \delta_0(\beta_0, \cdot)) \rightsquigarrow \mathbb{G}(\cdot)$ in $\ell^\infty(\mathcal{T} \times \mathcal{T})$.

Part 1 — Asymptotic equivalence of feasible and infeasible estimators. By Lemma 2(i),

$$\sup_{\tau \in \mathcal{T} \times \mathcal{T}} \left\| \frac{1}{m} \sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau)\big(\hat{\beta}_j(u) - \beta_{j,0}(u)\big) \right\| = o_p(m^{-1/2}).$$

Moreover, by the linearization result (Lemma 6), uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$,

$$\begin{aligned} M_{mn}(\hat{\delta}, \hat{\beta}, \tau) =& M_{mn}(\delta_0, \beta_0, \tau) + \Gamma_1(\delta_0, \beta_0, \tau)\big(\hat{\delta}(\tau) - \delta_0(\tau)\big) \\ &+ \frac{1}{m} \sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau)\,(\hat{\beta}_j - \beta_{j,0}) + o_p(m^{-1/2}). \end{aligned}$$

It follows that replacing β_0 by $\hat{\beta}$ only affects the second stage at order $o_p(m^{-1/2})$. Hence,

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

so the feasible and infeasible estimators share the same limiting distribution. Moreover, Theorem 1 continues to hold for the infeasible estimator $\hat{\delta}(\beta_0, \tau)$.

Part 2 – Weak Convergence This part of the proof closely follows the work of Angrist et al. (2006). Note that by the subgradient condition for quantile regression (see Koenker and Bassett (1978))

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

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

$$\|\sqrt{m}M_{mn}(\hat{\delta}, \beta_0, u, v)\| = o_p(1), \tag{17}$$

Observe that $\{1(\tilde{x}_{ij}'\beta_j(u) \leq x_{ij}'\delta(\beta,\tau)) : \tau \in \mathcal{T} \times \mathcal{T},\ \beta \in \mathcal{B},\ \delta \in \mathcal{D}\}$ is a VC–subgraph class and hence a bounded Donsker class for any compact $\mathcal{B}$ and $\mathcal{D}$ (compactness by assumption). Since β_0 is fixed, all subsequent arguments hold with $\beta = \beta_0$, so I henceforth omit the index set $\mathcal{B}$ and work over $\mathcal{D} \times \mathcal{T} \times \mathcal{T}$. It then follows that the class

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

is also Donsker with a square–integrable envelope $2\max_{i,j}\|x_{ij}\|$. Hence, the process $\sqrt{m}M_{mn}(\delta, \beta_0, \tau) - M(\delta, \beta_0, \tau)$ is stochastically equicontinuous on $\mathcal{T} \times \mathcal{T} \times \mathcal{D}$ under the usual $L_2(P)$ pseudometric.

Because $f_{Q(u,y|x,v)|x}$ is uniformly continuous and bounded by Assumption 5, the map $\delta \mapsto m(\delta, \beta_0, \tau)$ is Lipschitz in $L_2(P)$, i.e.,

$$\|\, m(\delta', \beta_0, \tau) - m(\delta'', \beta_0, \tau)\, \|_{L_2(P)} \leq C\, \|\delta' - \delta''\|.$$

Together with Theorem 1, which gives $\sup_{\tau\in\mathcal{T}\times\mathcal{T}}\|\hat{\delta}(\beta_0,\tau) - \delta_0(\beta_0,\tau)\| = o_p(1)$, this implies

$$\sup_{\tau\in\mathcal{T}\times\mathcal{T}} \|\, m(\hat{\delta}(\beta_0,\tau), \beta_0, \tau) - m(\delta_0(\beta_0,\tau), \beta_0, \tau)\, \|_{L_2(P)} = o_p(1).$$

By stochastic equicontinuity of the empirical process, it follows that

$$\begin{aligned}
&\sqrt{m}\,[M_{mn}(\hat{\delta}, \beta_0, \cdot) - M(\hat{\delta}, \beta_0, \cdot)] \\
&\quad = \sqrt{m}\,[M_{mn}(\delta_0, \beta_0, \cdot) - M(\delta_0, \beta_0, \cdot)] + o_p(1), \quad \text{in } \ell^\infty(\mathcal{T} \times \mathcal{T}).
\end{aligned} \tag{18}$$

Combining (17), (18), and (12) gives, uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$,

$$\begin{aligned}
o_p(1) =& \left[\Gamma_1(\delta_0, \beta_0, \tau) + o_p(1)\right] \sqrt{m}\left(\hat{\delta}(\tau) - \delta_0(\tau)\right) \\
&+ \sqrt{m}\left\{M_{mn}(\delta_0, \beta_0, \tau) - M(\delta_0, \beta_0, \tau)\right\}.
\end{aligned} \tag{19}$$

Let $\lambda = \inf_{\tau\in\mathcal{T}\times\mathcal{T}} \lambda_{\min}\big(\Gamma_1(\delta_0, \beta_0, \tau)\big) > 0$ (by Assumptions 2 and 6). Taking norms in (19) and using $\|Av\| \geq \lambda_{\min}(A)\,\|v\|$ for symmetric positive definite A, together with the triangle inequality, yields

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}\times\mathcal{T}} \left\|\sqrt{m}\left\{M_{mn}(\delta_0, \beta_0, \tau) - M(\delta_0, \beta_0, \tau)\right\}\right\| \\
&\qquad \geq \big(\lambda + o_p(1)\big) \sup_{\tau\in\mathcal{T}\times\mathcal{T}} \left\|\sqrt{m}\left(\hat{\delta}(\tau) - \delta_0(\tau)\right)\right\| + o_p(1).
\end{aligned} \tag{20}$$

Assumption 8 implies that $\delta_0(\beta_0, \tau)$ is continuous in τ. Because the function class

$$\left\{ m(\delta_0, \beta_0, \tau) = x_{ij}[v - 1(\tilde{x}_{ij}'\beta_{j,0}(u) \leq x_{ij}'\delta_0(\beta_0, \tau))] : \tau \in \mathcal{T} \times \mathcal{T} \right\}$$

is Donsker, the process $\tau \mapsto \sqrt{m}\{M_{mn}(\delta_0, \beta_0, \tau) - M(\delta_0, \beta_0, \tau)\}$ is stochastically equicontinuous on $\mathcal{T} \times \mathcal{T}$ with respect to the usual $L_2(P)$ pseudometric. Hence, the functional central limit theorem implies

$$\sqrt{m}\big(M_{mn}(\delta_0, \beta_0, \cdot) - M(\delta_0, \beta_0, \cdot)\big) \rightsquigarrow \mathbb{G}(\cdot) \quad \text{in } \ell^\infty(\mathcal{T} \times \mathcal{T}), \tag{21}$$

where $\mathbb{G}(\cdot)$ is a mean-zero Gaussian process with covariance function $\Omega_2(\cdot, \cdot)$ as defined in Lemma 2. It follows that the left-hand side of (20) is $O_p(1)$, and therefore

$$\sup_{\tau \in \mathcal{T} \times \mathcal{T}} \|\sqrt{m}(\hat{\delta}(\beta_0, \tau) - \delta_0(\beta_0, \tau))\| = O_p(1).$$

Combining (19)–(21) yields

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

■

SUPPLEMENTAL MATERIAL

A Simulations

Table 1: Bias and Standard Deviation

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

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

To study the finite-sample performance of the estimator, I conduct a Monte Carlo simulation based on the following data-generating process:

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

with $x_{1ij} = 1 + h_j + w_{ij}$, $h_j \sim U[0,1]$, and $w_{ij}, x_{2j}, \eta_j, \nu_{ij} \sim N(0,1)$. This design generates a location–scale–shift model across both quantile indices. I set $\beta = \gamma = 1$. The true coefficients are given by $\beta(u,v) = \gamma(u,v) = 1 + 0.1F^{-1}(u) - 0.1F^{-1}(v)$, where F denotes the standard normal CDF. I consider sample sizes $(m,n) \in \{(25,25),(200,25),(25,200),(200,200),(200,400)\}$ and quantiles $\{0.25, 0.5, 0.75\}$, based on 2,000 Monte Carlo replications.

Table 1 reports biases and standard deviations. Table 2 reports coverage probabilities for 95% confidence intervals. Bootstrap standard errors are computed using 200

Table 2: Coverage Probability

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

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

repetitions. While $\hat{\beta}$ and $\hat{\gamma}$ share the same asymptotic properties, their finite-sample performance differs. The bias of $\hat{\beta}$ declines with both larger n and m, whereas the bias of $\hat{\gamma}$ falls mainly as m increases. Likewise, the variance of $\hat{\gamma}$ is only marginally affected by larger n, while the variance of $\hat{\beta}$ declines more noticeably with additional within-group observations. Nonetheless, the most substantial reduction in variance comes from increasing the number of groups. As n grows, further increases in within-group observations have little impact on the variance of $\hat{\beta}$. Specifically, the initial decline in the bias and variance of $\hat{\beta}$ – primarily driven by a reduction in first-stage estimation error – quickly levels off, and the difference between $n = 200$ and $n = 400$ is minimal.

The coverage probabilities in Table 2 are generally close to the nominal level, with small deviations that shrink as m and n grow. For $(m, n) = (200, 25)$, the intervals tend to undercover, which is consistent with the relatively larger bias that arises with this sample size configuration (see Table 1).

Table 3 shows the rejection probabilities of 5% Kolmogorov-Smirnov and Cramér-

Table 3: Rejection Probability of the KS and CvM Tests

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

Notes: The table shows the rejection probabilities of a 5% test of the null hypothesis of homogeneous effects along both dimensions. The test uses the Kolmogorov–Smirnov and Cramér–von Mises test statistics. The results are based on 1,000 Monte Carlo simulations with 100 bootstrap replications.

von Mises tests for the null hypotheses that $\beta(\tau) = \bar{\beta}$ and that $\gamma(\tau) = \bar{\gamma}$.[23] For these simulations, I use a variation of the baseline data-generating process:

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

with the same distributions for all variables as in Section A. The parameter ϕ governs heterogeneity over u and ψ governs heterogeneity over v. I consider five (ϕ, ψ) specifications and test the null of homogeneous effects for both x_{1ij} and x_{2j}. The rejection probabilities are based on 1,000 Monte Carlo simulations, each with 100 bootstrap replications, using quantiles $\{0.1, 0.2, \ldots, 0.9\}$.

In the first column, where the null hypothesis is true, the rejection probability should equal 5% (empirical size). In all other cases, the rejection probability indicates the power of the test. In smaller samples, the tests are somewhat conservative, and with little

[23]The Cramér-von Mises test statistic is $t^{CvM} = \int_{\mathcal{T}} \int_{\mathcal{T}} \left(\hat{\delta}_k(\tau) - \bar{\delta}_k\right)' \hat{V}_k(\tau) \left(\hat{\delta}_k(\tau) - \bar{\delta}_k\right) dudv$, and its critical values are obtained by bootstrapping the statistic, as for the Kolmogorov–Smirnov test.

heterogeneity, their ability to detect differences is reduced. However, the empirical size approaches its nominal level as the sample grows, and columns (2)–(5) show that power rises quickly as either heterogeneity or sample size increases.

The performance also differs between β and γ. For γ, power is much lower when the number of groups is small, but this gap narrows as m increases. In contrast, for β, it is easier to detect heterogeneity in small samples, and power improves notably with larger n, whereas for γ, increases in n yield only modest gains.

Across designs, the Cramér–von Mises test is generally more powerful than the Kolmogorov–Smirnov test. Still, different data-generating processes may favor one over the other, as neither test is uniformly more powerful, and each test is better at detecting certain types of deviations.

B Inference

B.1 Bootstrap Algorithm

To conduct inference, I implement a clustered (group-level) bootstrap that resamples entire groups with replacement. The bootstrap proceeds as follows:

(i) Draw m groups with replacement from the original sample to form a bootstrap sample. Because resampling occurs at the group level, the first-stage estimates $\hat{\beta}_j(u)$ remain fixed.

(ii) Using the resampled data, re-estimate the second-stage quantile regression to obtain $\hat{\delta}^*(\hat{\beta}, \tau)$.

(iii) Repeat steps (i)–(ii) for $b = 1, \dots, B$ to generate a set of bootstrap estimates $\{\hat{\delta}^*_b(\hat{\beta}, \tau)\}_{b=1}^B$.

(iv) Estimate the asymptotic covariance matrix $\Gamma_1^{-1}\Omega_2(\tau)\Gamma_1^{-1}$ using, for instance, (a) the empirical variance of the bootstrap estimates, or (b) a robust interquartile-range estimator scaled by the standard normal quantile difference:

$$\hat{\Gamma}_1^{*-1}\hat{\Omega}_2^*(\tau)\hat{\Gamma}_1^{*-1} = \frac{q_{0.75}(\tau) - q_{0.25}(\tau)}{z_{0.75} - z_{0.25}},$$

where $q_p(\tau)$ is the pth percentile of the bootstrap replicates.

B.2 Kolmogorov–Smirnov and Cramér–von Mises Tests

This appendix provides details on the implementation of Kolmogorov–Smirnov (KS) and Cramér–von Mises (CvM) type tests for hypotheses involving multiple coefficients in the two-step quantile regression framework. For instance, one might want to test whether a subvector of $\delta(u, v)$, denoted by $\delta_k(u, v)$ for a given index set k, is constant across quantiles:

$$H_0 : \delta_k(\tau) = \bar{\delta}_k, \ \forall \ u, v \in \mathcal{T} \times \mathcal{T},$$

with $\bar{\delta}_k = \int_v \int_u \hat{\delta}(u, v) dudv$.

KS and CvM type tests are suitable in these settings. The KS test statistic is:

$$t^{KS} = \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \sqrt{\left(\hat{\delta}_k(\tau) - \bar{\delta}_k\right)' \hat{V}_k(\tau)^{-1} \left(\hat{\delta}_k(\tau) - \bar{\delta}_k\right)}, \tag{1}$$

where $\hat{V}_k(\tau)$ is a bootstrap estimate of the asymptotic variance of $\hat{\delta}_k(\tau)$. To obtain the critical values, I follow Chernozhukov and Fernández-Val (2005) and use the bootstrap to mimic the test statistic. However, instead of recentering, I impose the null hypothesis using the parametric bootstrap based on the estimated quantile regression process.[24] For each bootstrap iteration, the corresponding test statistic is computed by:

$$t_b^{KS} = \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \sqrt{\left(\hat{\delta}_k^{*b}(\tau) - \hat{\delta}_k^{*b}\right)' \hat{V}_k(\tau)^{-1} \left(\hat{\delta}_k^{*b}(\tau) - \hat{\delta}_k^{*b}\right)}, \tag{2}$$

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

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

$$t_b^{KS} = \sup_{\tau \in \mathcal{T} \times \mathcal{T}} \sqrt{\left(\hat{\delta}_k^{*b}(\tau) - \hat{\delta}_k(\tau)\right)' \hat{V}_k(\tau)^{-1} \left(\hat{\delta}_k^{*b}(\tau) - \hat{\delta}_k(\tau)\right)}. \tag{3}$$

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

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

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

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

C Additional Asymptotic Results

C.1 Preliminary Lemmas and Linear Representation

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

Lemma 3 (Uniform consistency of $\hat{\beta}_j(u)$)**.** *Under Assumptions 1-4 and 9(a), we have*

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

Proof. The proof follows directly from the proof of Lemma 3 in Melly and Pons (2025) after noting that $\hat{\beta}(u)$ does not depend on v. ∎

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

$$\hat{\beta}_j(u) - \beta_{j,0}(u) = \frac{1}{n}\sum_{i=1}^{n} \phi_{j,u}(\tilde{x}_{ij}, y_{ij}) + R_{nj}^{(1)}(u) + R_{nj}^{(2)}(u), \tag{4}$$

where

$$\phi_{j,u}(\tilde{x}_{ij}, y_{ij}) = -B_{j,u}^{-1}\tilde{x}_{ij}(1(y_{ij}\le \tilde{x}_{ij}'\beta_j(u)) - u), \tag{5}$$

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

$$\sup_j \sup_{u\in\mathcal{T}} \left\|R_{nj}^{(2)}(u)\right\| = O_p\left(\frac{\log n}{n}\right) \tag{6}$$

$$\sup_j \sup_{u\in\mathcal{T}} \left\|E_{i|j}\left[R_{nj}^{(1)}(u)\right]\right\| = O\left(\frac{\log n}{n}\right) \tag{7}$$

$$\sup_j \sup_{\tau\in\mathcal{T}} \left\|E\left[\left(R_{nj}^{(1)}(u) - E_{i|j}[R_{nj}^{(1)}(u)]\right)\left(R_{nj}^{(1)}(\tau) - E_{i|j}[R_{nj}^{(1)}(u)]\right)'\right]\right\|$$
$$= O\left(\left(\frac{\log n}{n}\right)^{3/2}\right). \tag{8}$$

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

Lemma 5. *Under Assumptions 1–2 and 7,*

$$\sup_{\|\beta-\beta_0\|_{\mathcal{B}}\le\zeta_m,\ \|\delta-\delta_0\|\le\zeta_m} \left\|M_{mn}(\delta,\beta,\tau) - M(\delta,\beta,\tau) - M_{mn}(\delta_0,\beta_0,\tau)\right\| = o_p(m^{-1/2}), \tag{9}$$

for all positive sequences $\zeta_m = o(1)$*. That is, Condition* $(2.5')$ *of Chen et al. (2003) holds.*

Proof. This is an application of Theorem 3 in Chen et al. (2003). I verify that its conditions are satisfied.

Condition (3.1). Recall $m(\delta,\beta,\tau) = \rho_v(\tilde{x}_{ij}'\beta_j - x_{ij}'\delta)$. By the Lipschitz property of the check loss,

$$\begin{aligned} |m(\delta',\beta',\tau) - m(\delta'',\beta'',\tau)| &\le |\rho_v(\tilde{x}_{ij}'\beta_j' - x_{ij}'\delta') - \rho_v(\tilde{x}_{ij}'\beta_j'' - x_{ij}'\delta')| \\ &\quad + |\rho_v(\tilde{x}_{ij}'\beta_j'' - x_{ij}'\delta') - \rho_v(\tilde{x}_{ij}'\beta_j'' - x_{ij}'\delta'')| \\ &\le 2\|\tilde{x}_{ij}\|\,\|\beta_j' - \beta_j''\| + 2\|x_{ij}\|\,\|\delta' - \delta''\|. \end{aligned} \tag{10}$$

Thus $m(\delta,\beta,\tau)$ is Hölder continuous in (δ,β) since the regressors are bounded by Assumption 2.

Condition (3.2). Using the bound (10) and squaring, we obtain

$$E\left[\sup_{\|\beta-\beta'\|_{\mathcal{B}}\le\zeta,\ \|\delta-\delta'\|\le\zeta} |m(\delta,\beta,\tau) - m(\delta',\beta',\tau)|^2\right] \ \le\ K\zeta^2,$$

for some finite K, since the regressors are bounded. Hence $(\delta,\beta)\mapsto m(\delta,\beta,\tau)$ is locally uniformly L_2–continuous, as required.

Condition (3.3). By Assumption 7, δ lies in a compact set $\mathcal{D}\subset\mathbb{R}^K$ and each β_j in a compact set $\mathcal{B}_j\subset\mathbb{R}^{K_1}$. Therefore $\beta = (\beta_1,\ldots,\beta_m)$ lies in $\mathcal{B} = \prod_{j=1}^m \mathcal{B}_j$, which is compact by Tychonoff's theorem. Compactness ensures finite covering numbers, so (3.3) is satisfied.

Since conditions (3.1)–(3.3) hold, Theorem 3 in Chen et al. (2003) applies, implying Condition (2.5′). Finally, by Remark 2(i) in Chen et al. (2003), (2.5′) implies Condition (2.5). ∎

Lemma 6 (Linearization)**.** *Let Assumptions 1–7 and 9 hold. Define, for $\tau\in\mathcal{T}\times\mathcal{T}$,*

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

Then, uniformly in $\tau\in\mathcal{T}\times\mathcal{T}$,

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

Proof of Lemma 6. The proof consists of two parts. First, I show that the linearization holds pointwise. Then I show that it holds uniformly in $\tau\in\mathcal{T}\times\mathcal{T}$.

Part 1. (Pointwise) For this first part, I rely on a result from the proof of Theorem 2 in Chen et al. (2003) and begin by verifying that the conditions for applying the theorem are satisfied.

First, Assumption 9(b) implies 9(a), so that by Theorem 1 and Lemma 3,

$$\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau) \xrightarrow{p} 0 \quad \text{and} \quad \hat{\beta}_j(u) - \beta_{j,0}(u) \xrightarrow{p} 0$$

uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$ and $1 \leq j \leq m$. Hence, we can replace the parameter space with a small neighborhood of the true parameter. By uniform consistency and compactness, $(\hat{\delta}, \hat{\beta})$ remains in such a neighborhood with probability tending to one, so an explicit construction of shrinking balls as in Chen et al. (2003) is not needed here.

Condition (2.1). 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.

Condition (2.2). By Assumption 5 the derivative of $M(\delta, \beta_0, \tau)$ with respect to δ exists and is given by

$$\Gamma_1(\delta, \beta_0, \tau) = E\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta|x_{ij})\, x_{ij}x_{ij}'\right].$$

Assumptions 2 and 6 ensure that $\Gamma_1(\delta_0, \beta_0, \tau)$ is full rank. Hence, Condition (2.2) is satisfied.

Condition (2.3). Denote the pathwise derivative of $M(\delta, \beta_0, \tau)$

$$\tilde{\Gamma}_2(\delta, \beta_0, \tau)[\beta(u) - \beta_0(u)] = \sum_{j=1}^{m} \tilde{\Gamma}_{2j}(\delta, \beta_0, \tau)(\beta_j(u) - \beta_{j,0}(u)),$$

where

$$-\tilde{\Gamma}_{2j}(\delta, \beta_0, \tau) = -E_{i|j}\left[f_{Q(u,y_{ij}|x_{ij},v_j)|x_{ij}}(x_{ij}'\delta(\beta_0, \tau) \mid x_{ij})\, x_{ij}\tilde{x}_{ij}'\right].$$

For notational convenience, in the paper, I define $\tilde{\Gamma}_{2j}(\delta, \beta_0, \tau) = \frac{1}{m}\Gamma_{2j}(\delta, \beta_0, \tau)$.

By Assumption 5, the derivative exists in all directions $(\beta_j - \beta_{j,0}) \in \mathcal{B}_j$. Condition (2.3) requires that, for parameters (δ, β) in a sufficiently small neighborhood of (δ_0, β_0):

(i)

$$\|M(\delta, \beta, \tau) - M(\delta, \beta_0, \tau) - \sum_{j=1}^{m} \tilde{\Gamma}_{2j}(\delta, \beta_0, \tau)(\beta_j - \beta_{j,0})\| \leq c \cdot \|\beta - \beta_0\|_{\mathcal{B}}^2,$$

for some constant $c \geq 0$, which follows from twice continuous differentiability and Lemma 3; and (ii)

$$\Big\| \sum_{j=1}^{m} \tilde{\Gamma}_{2j}(\delta, \beta_0, \tau)(\beta_j - \beta_{j,0}) - \sum_{j=1}^{m} \tilde{\Gamma}_{2j}(\delta_0, \beta_0, \tau)(\beta_j - \beta_{j,0}) \Big\| \leq o(1) \cdot \|\beta - \beta_0\|_{\mathcal{B}}.$$

Assumption 5 implies that $\Gamma_{2j}(\delta, \beta_0, \tau)$ is continuous in δ. Hence the difference between evaluating at δ and δ_0 is of order $O(\|\delta - \delta_0\| \cdot \|\beta - \beta_0\|_{\mathcal{B}})$, which is $o(1)\|\beta - \beta_0\|_{\mathcal{B}}$. Thus, condition (ii) holds.

Condition (2.4). Condition (2.4) requires that $||\hat{\beta} - \beta_0||_{\mathcal{B}} = o_p(m^{-1/4})$. The proof of Lemma 1 in Galvao and Wang (2015) implies that

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

If $\frac{\sqrt{m}\log n}{n} \to 0$ (Assumption 9(b)), $\sup ||\hat{\beta}_j - \beta_0||_{\mathcal{B}} = o_p(m^{-1/4})$, so that condition (2.4) in Chen et al. (2003) is satisfied.

Condition (2.5). Directly implied by Lemma 5.

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

$$||M_{mn}(\hat{\delta}, \hat{\beta}, \tau) - \mathcal{L}_{mn}(\hat{\delta}, \tau)|| = o_p(m^{-1/2}) \tag{11}$$

Part 2. (Uniform in τ) I now show that the linearization holds uniformly over $\tau \in \mathcal{T} \times \mathcal{T}$.

For any $(\delta', \beta', \tau'), (\delta'', \beta'', \tau'')$,

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

From the proof of Lemma 5 we know that

$$|m(\delta', \beta', \tau') - m(\delta'', \beta'', \tau')| \;\leq\; 2\|\tilde{x}_{ij}\| \, \|\beta' - \beta''\|_{\mathcal{B}} + 2\|x_{ij}\| \, \|\delta' - \delta''\|.$$

By the Lipschitz property of the check function, we also have

$$|m(\delta'', \beta'', \tau') - m(\delta'', \beta'', \tau'')| \;\leq\; |v' - v''| \cdot \|x_{ij}\| \big(\|\beta''\|_{\mathcal{B}} + \|\delta''\|\big).$$

From the two equations above and boundedness of the regressors (Assumption 2) together with compactness (Assumption 7), it follows that there exists $C < \infty$ such that

$$|m(\delta', \beta', \tau') - m(\delta'', \beta'', \tau'')| \leq C\Big(\|\beta' - \beta''\|_{\mathcal{B}} + \|\delta' - \delta''\| + |v' - v''|\Big),$$

uniformly in (δ, β, τ). Averaging over i, j preserves the bound, so $(\delta, \beta, \tau) \mapsto M_{mn}(\delta, \beta, \tau)$ is stochastically equicontinuous.

By a mean–value expansion in δ, for each $\tau \in \mathcal{T} \times \mathcal{T}$ there exists $\delta^*(\tau)$ on the segment joining $\delta_0(\beta_0, \tau)$ and $\hat{\delta}(\hat{\beta}, \tau)$ such that

$$M(\delta, \beta_0, \tau)\big|_{\delta=\hat{\delta}(\hat{\beta},\tau)} = \Gamma_1\big(\delta^*(\tau), \beta_0, \tau\big)\ \big(\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)\big).$$

By Theorem 1, $\sup_{\tau\in\mathcal{T}\times\mathcal{T}} \|\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)\| = o_p(1)$, and by Assumption 5 (continuous and bounded density) together with Assumption 2, $\Gamma_1(\delta, \beta_0, \tau)$ is continuous in δ uniformly over $\tau \in \mathcal{T} \times \mathcal{T}$. Hence

$$\sup_{\tau\in\mathcal{T}\times\mathcal{T}} \big\|\Gamma_1(\delta^*(\tau), \beta_0, \tau) - \Gamma_1(\delta_0, \beta_0, \tau)\big\| = o_p(1),$$

and therefore

$$M(\delta, \beta_0, \tau)\big|_{\delta=\hat{\delta}(\hat{\beta},\tau)} = \big[\Gamma_1(\delta_0, \beta_0, \tau) + o_p(1)\big]\ \big(\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)\big), \tag{12}$$

uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$. Hence, $\tau \mapsto \Gamma_1(\delta_0, \beta_0, \tau)\big(\hat{\delta}(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)\big)$ is stochastically equicontinuous.

Similarly, a mean–value expansion in β around β_0 gives

$$M(\delta_0, \beta, \tau)\big|_{\beta=\hat{\beta}} = \frac{1}{m}\sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta^*, \tau)\ (\hat{\beta}_j(u) - \beta_{j,0}(u)),$$

for some intermediate points $\beta_j^*(u)$ on the segment joining $\beta_{j,0}(u)$ and $\hat{\beta}_j(u)$.

Using the same argument as above, and by uniform consistency of $\hat{\beta}_j$ over j and τ (Theorem 3), it follows that

$$M(\delta_0, \beta, \tau)\big|_{\beta=\hat{\beta}} = \left[\frac{1}{m}\sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau) + o_p(1)\right] (\hat{\beta}_j(u) - \beta_{j,0}(u)),$$

By Assumptions 5 and 2, $\tau \mapsto \Gamma_{2j}(\delta_0, \beta_0, \tau)$ is continuous and uniformly bounded. It then follows that

$$\tau \ \mapsto \ \frac{1}{m}\sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau)\ (\hat{\beta}_j(u) - \beta_{j,0}(u))$$

is asymptotically equicontinuous.

Since each component of $\mathcal{L}(\delta)$ and $M_{mn}(\delta, \beta, \tau)$ is asymptotically equicontinuous, it follows that the linearization holds uniformly over $\tau \in \mathcal{T} \times \mathcal{T}$. ∎

The next corollary restates Lemma 6 in terms of $\hat{\delta}$.

Corollary 1 (Corollary to Lemma 6)**.** *Under the conditions of Lemma 6, uniformly over $\tau \in \mathcal{T} \times \mathcal{T}$,*

$$\begin{aligned}\sqrt{m}\big(\hat{\delta}(\hat{\beta},\tau) - \delta_0(\beta_0,\tau)\big) = &- \Gamma_1(\delta_0,\beta_0,\tau)^{-1} \\ &\sqrt{m}\left(M_{mn}(\delta_0,\beta_0,\tau) + \frac{1}{m}\sum_{j=1}^{m}\Gamma_{2j}(\delta_0,\beta_0,\tau)\left(\hat{\beta}_j(u) - \beta_{j,0}(u)\right)\right) + o_p(1).\end{aligned}$$

C.2 Asymptotic Normality

Theorem 4 (**Asymptotic Normality**)**.** *Let assumptions 1-7 and 9(b) be satisfied. Then*

$$\sqrt{m}\left(\hat{\delta}(\hat{\beta},\tau) - \delta_0(\beta_0,\tau)\right) \xrightarrow{d} N(0, \Gamma_1^{-1}\Omega_2(\tau)\Gamma_1'^{-1}) \tag{13}$$

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

Proof of Theorem 4. By Lemma 2 and 6 it follows directly that

$$\sqrt{m}\left(M_{mn}(\delta_0,\beta_0,\tau) + \frac{1}{m}\sum_{j=1}^{m}\Gamma_{2,j}(\beta_0,\tau)\left(\hat{\beta}_j(u) - \beta_{j,0}(u)\right)\right) \xrightarrow{d} N(0,\Omega_2(\tau)), \tag{14}$$

which implies the result. ■

C.3 Validity of the Bootstrap

Part 1 - Linearization. Since the bootstrap algorithm samples entire groups, the first stage is the same in all bootstrap replications. The only source of randomness in the bootstrap world is therefore which groups are resampled. Let $w_1^*, \ldots, w_m^*$ denote the bootstrap weights, where w_j^* is the number of times group j appears in the bootstrap resample (so that $\sum_{j=1}^{m} w_j^* = m$). For any group-level quantity a_j, its bootstrap analogue enters through the weighted average $\frac{1}{m}\sum_{j=1}^{m} w_j^* a_j$. In particular, define

$$M_{mn}(\delta,\beta,\tau) = \frac{1}{m}\sum_{j=1}^{m}\bar{m}_j(\delta,\beta,\tau), \qquad M_{mn}^*(\delta,\beta,\tau) = \frac{1}{m}\sum_{j=1}^{m} w_j^*\,\bar{m}_j(\delta,\beta,\tau).$$

Moreover, recall that the first stage is not re-estimated in the bootstrap world. It can be shown that

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

Together with Theorem 4, this implies

$$\hat{\delta}^*(\hat{\beta},\tau) - \delta_0(\beta_0,\tau) = O_p^*\big(m^{-1/2}\big)\,. \tag{15}$$

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

$$\begin{aligned}\sqrt{m}\Big(\hat{\delta}^*(\hat{\beta},\tau) - \delta_0(\beta_0,\tau)\Big) &= \Gamma_1(\delta_0,\beta_0,\tau)^{-1} \\ &\times \sqrt{m}\left(\frac{1}{m}\sum_{j=1}^{m} w_j^*\,\Gamma_{2j}(\delta_0,\beta_0,\tau)\big[\hat{\beta}_j(u) - \beta_{j,0}(u)\big] + M_{mn}^*(\delta_0,\beta_0,\tau)\right) + o_p^*(1),\end{aligned} \tag{16}$$

where $M_{mn}^*(\delta_0,\beta_0,\tau) = \frac{1}{m}\sum_{j=1}^{m} w_j^*\,\bar{m}_j(\delta_0,\beta_0,\tau)$. For this part of the proof, I rely on the results from Chen et al. (2003).

Define the linear approximation (suppressing the dependence on τ for ease of notation):

$$\mathcal{L}_{mn}^*(\delta) = M_{mn}^*(\delta_0,\beta_0) + \Gamma_1(\delta_0,\beta_0)(\delta-\delta_0) + \frac{1}{m}\sum_{j=1}^{m} w_j^*\,\Gamma_{2j}(\delta_0,\beta_0)\big(\hat{\beta}_j - \beta_{j,0}\big). \tag{17}$$

The first step is to show that,

$$\big\|M_{mn}^*(\hat{\delta}^*,\hat{\beta}) - \mathcal{L}_{mn}^*(\hat{\delta}^*)\big\| = o_p^*(m^{-1/2}). \tag{18}$$

By the triangle inequality,

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

We bound the three terms on the right-hand side in turn.

(i) First term. By Condition 2.5 in Chen et al. (2003)

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

(ii) Second term. Using the same local expansion arguments as in the proof of Theorem 4, and noting that the bootstrap enters only through the weights w_j^* (which affect the

group average but do not change the first-stage estimates), we have

$$
\begin{aligned}
&\left\| M(\hat{\delta}^*, \hat{\beta}) - M(\hat{\delta}^*, \beta_0) + \frac{1}{m}\sum_{j=1}^{m} w_j^* \, \Gamma_{2j}(\delta_0, \beta_0)\big(\hat{\beta}_j - \beta_{j,0}\big) \right\| \\
&\le \left\| M(\hat{\delta}^*, \hat{\beta}) - M(\hat{\delta}^*, \beta_0) + \frac{1}{m}\sum_{j=1}^{m} w_j^* \, \Gamma_{2j}(\hat{\delta}^*, \beta_0)\big(\hat{\beta}_j - \beta_{j,0}\big) \right\| \\
&\quad + \left\| \frac{1}{m}\sum_{j=1}^{m} w_j^* \Big(\Gamma_{2j}(\hat{\delta}^*, \beta_0) - \Gamma_{2j}(\delta_0, \beta_0)\Big)\big(\hat{\beta}_j - \beta_{j,0}\big) \right\| \\
&= O_p^*\big(\|\hat{\beta} - \beta_0\|_{\mathcal{B}}^2\big) + o_p^*(1) \cdot \|\hat{\delta}^* - \delta_0\| = o_p^*(m^{-1/2}),
\end{aligned}
$$

since $\sup_j \|\hat{\beta}_j - \beta_{j,0}\| = o_p(m^{-1/4})$ as shown in the proof of Theorem 4 and $\hat{\delta}^* - \delta_0 = O_p^*(m^{-1/2})$.

(iii) Third term. By differentiability of $M(\delta, \beta_0)$ in δ at δ_0 and (15),

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

Combining (i)–(iii) yields (18).

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 $\mathcal{L}_{mn}^*(\delta)$. Since $\mathcal{L}_{mn}^*(\delta)$ is affine in δ, $\bar{\delta}^*$ satisfies that $\mathcal{L}_{mn}^*(\bar{\delta}^*)$ is orthogonal to the column space of $\Gamma_1(\delta_0, \beta_0)$ and admits the explicit representation

$$
\begin{aligned}
\bar{\delta}^* - \delta_0 &= \Gamma_1(\delta_0, \beta_0)^{-1} \left(\frac{1}{m}\sum_{j=1}^{m} w_j^* \, \Gamma_{2j}(\delta_0, \beta_0)\big(\hat{\beta}_j - \beta_{j,0}\big) + M_{mn}^*(\delta_0, \beta_0) \right) \qquad (19) \\
&= \Gamma_1(\delta_0, \beta_0)^{-1} \left(\frac{1}{m}\sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0)\big(\hat{\beta}_j - \beta_{j,0}\big) + M_{mn}(\delta_0, \beta_0) \right) + O_p^*(m^{-1/2}) \\
&= \hat{\delta} - \delta_0 + O_p^*(m^{-1/2}) \\
&= O_p^*(m^{-1/2}), \qquad (20)
\end{aligned}
$$

where the third line follows from Theorem 4, and the second line uses the standard bootstrap fact that

$$
\frac{1}{m}\sum_{j=1}^{m}(w_j^* - 1)a_j = O_p^*(m^{-1/2}) \quad \text{whenever} \quad \frac{1}{m}\sum_{j=1}^{m}\|a_j\|^2 = O_p(1),
$$

applied to $a_j = \Gamma_{2j}(\delta_0, \beta_0)(\hat{\beta}_j - \beta_{j,0})$ and to $a_j = \bar{m}_j(\delta_0, \beta_0)$.

By the triangle inequality,

$$\big\|M^*_{mn}(\bar\delta^*,\hat\beta)-\mathcal{L}^*_{mn}(\bar\delta^*)\big\| \le \Big\|M(\bar\delta^*,\hat\beta)-M(\bar\delta^*,\beta_0)-\frac{1}{m}\sum_{j=1}^{m} w_j^*\,\Gamma_{2j}(\delta_0,\beta_0)\big(\hat\beta_j-\beta_{j,0}\big)\Big\|$$
$$+\Big\|M(\bar\delta^*,\beta_0)-\Gamma_1(\delta_0,\beta_0)\big(\bar\delta^*-\delta_0\big)\Big\|$$
$$+\Big\|M^*_{mn}(\bar\delta^*,\hat\beta)-M(\bar\delta^*,\hat\beta)-M^*_{mn}(\delta_0,\beta_0)\Big\|.$$

For the first term, add and subtract the same linear term evaluated at $\bar\delta^*$ and use the usual second-order remainder in β:

$$\begin{aligned}
&\Big\|M(\bar\delta^*,\hat\beta)-M(\bar\delta^*,\beta_0)-\frac{1}{m}\sum_{j=1}^{m} w_j^*\,\Gamma_{2j}(\delta_0,\beta_0)\big(\hat\beta_j-\beta_{j,0}\big)\Big\|\\
&\le \Big\|M(\bar\delta^*,\hat\beta)-M(\bar\delta^*,\beta_0)-\frac{1}{m}\sum_{j=1}^{m} w_j^*\,\Gamma_{2j}(\bar\delta^*,\beta_0)\big(\hat\beta_j-\beta_{j,0}\big)\Big\|\\
&\quad+\Big\|\frac{1}{m}\sum_{j=1}^{m} w_j^*\Big(\Gamma_{2j}(\bar\delta^*,\beta_0)-\Gamma_{2j}(\delta_0,\beta_0)\Big)\big(\hat\beta_j-\beta_{j,0}\big)\Big\|\\
&= O_p^*\big(\|\hat\beta-\beta_0\|^2_{\mathcal{B}}\big)\;+\;o_p^*(1)\cdot\|\bar\delta^*-\delta_0\|\;=\;o_p^*(m^{-1/2}),
\end{aligned}$$

where the last equality uses $\sup_j\|\hat\beta_j-\beta_{j,0}\|=o_p(m^{-1/4})$ and $\bar\delta^*-\delta_0=O_p^*(m^{-1/2})$ (shown below).

For the second term, differentiability of $M(\delta,\beta_0)$ in δ yields the expansion

$$M(\bar\delta^*,\beta_0)=M(\delta_0,\beta_0)+\Gamma_1(\delta_0,\beta_0)(\bar\delta^*-\delta_0)+o\big(\|\bar\delta^*-\delta_0\|\big),$$

and since $M(\delta_0,\beta_0)=0$ and $\bar\delta^*-\delta_0=O_p^*(m^{-1/2})$,

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

For the third term, Condition 2.5 in Chen et al. (2003) implies

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

Combining the three bounds yields

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

To conclude the linearization, it remains to show that $\|\hat\delta^*-\bar\delta^*\|=o_p^*(m^{-1/2})$. Following Pakes and Pollard (1989), we know that $M^*_{mn}(\delta,\beta_0)$ and $\mathcal{L}^*(\delta)$ are close at both $\hat\delta^*$ which almost minimizes $||M^*_{mn}(\delta, \hat{\beta})||$ and at $\bar{\delta}^*$ which minimizes $\mathcal{L}^*(\delta)$. This means that $\hat{\delta}^*$ has to be close to minimizing $\mathcal{L}^*(\delta)$:

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

Hence,

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

and squaring both sides yields

$$\|\mathcal{L}^*_{mn}(\hat{\delta}^*)\|^2 = \|\mathcal{L}^*_{mn}(\bar{\delta}^*)\|^2 + o^*_p(m^{-1}), \tag{21}$$

where the cross product is also $o_p(m^{-1})$ because $||\mathcal{L}(\bar{\delta}^*)||$ is of order $O^*_p(m^{-1/2})$. The map $\delta \mapsto \|\mathcal{L}^*_{mn}(\delta)\|^2$ admits the quadratic expansion

$$\|\mathcal{L}^*_{mn}(\delta)\|^2 = \|\mathcal{L}^*_{mn}(\bar{\delta}^*)\|^2 + \|\Gamma_1(\delta - \bar{\delta}^*)\|^2, \tag{22}$$

around its global minimum $\bar{\delta}^*$. The cross-product term vanishes because the residual vector $\mathcal{L}^*_{mn}(\bar{\delta}^*)$ is orthogonal to the column space of Γ_1. Setting $\delta = \hat{\delta}*$ and combining (21) and (22) gives

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

and, since Γ_1 is full rank,

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

Substituting $\hat{\delta}^* = \bar{\delta}^* + o^*_p(m^{-1/2})$ into (19) yields (16).

Similarly, to Part 1 of the proof of Theorem 2, one can strengthen the preceding argument to obtain the uniform version of (16) over $\tau \in \mathcal{T} \times \mathcal{T}$:

$$\begin{aligned}&\sqrt{m}\Big(\hat{\delta}^*(\hat{\beta}, \tau) - \delta_0(\beta_0, \tau)\Big) = \Gamma_1(\delta_0, \beta_0, \tau)^{-1} \\ &\quad \sqrt{m}\left(\frac{1}{m}\sum_{j=1}^{m} w^*_j\, \Gamma_{2j}(\delta_0, \beta_0, \tau)\big[\hat{\beta}_j(u) - \beta_{j,0}(u)\big] + M^*_{mn}(\delta_0, \beta_0, \tau)\right) + o^*_p(1).\end{aligned} \tag{23}$$

Part 2 – Asymptotic distribution of $\hat{\delta}^* - \hat{\delta}$. For this part of the proof, I borrow from the proof of Proposition H.1 in Fernández-Val et al. (2022).

Define, for $\tau \in \mathcal{T} \times \mathcal{T}$,

$$\theta^*_{mn}(\tau) = \Gamma_1(\delta_0, \beta_0, \tau)^{-1} \left[\frac{1}{m} \sum_{j=1}^{m} w^*_j \Gamma_{2j}(\delta_0, \beta_0, \tau) \big(\hat{\beta}_j(u) - \beta_{j,0}(u)\big) + M^*_{mn}(\delta_0, \beta_0, \tau) \right], \tag{24}$$

where $M^*_{mn}(\delta_0, \beta_0, \tau) = \frac{1}{m} \sum_{j=1}^{m} w^*_j \, \bar{m}_j(\delta_0, \beta_0, \tau)$.
Since $E^*[w^*_j] = 1$ and $E^*[M^*_{mn}(\delta_0, \beta_0, \tau)] = M_{mn}(\delta_0, \beta_0, \tau)$, we have

$$E^* \left[\theta^*_{mn}(\tau)\right] = \Gamma_1(\delta_0, \beta_0, \tau)^{-1} \left[\frac{1}{m} \sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau) \big(\hat{\beta}_j(u) - \beta_{j,0}(u)\big) + M_{mn}(\delta_0, \beta_0, \tau) \right]. \tag{25}$$

Combining the linear representations for $\hat{\delta}^*(\hat{\beta}, \tau)$ (Equation (23)) and $\hat{\delta}(\hat{\beta}, \tau)$ yields, uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$,

$$\hat{\delta}^*(\hat{\beta}, \tau) - \hat{\delta}(\hat{\beta}, \tau) = \theta^*_{mn}(\tau) - E^* \left[\theta^*_{mn}(\tau)\right] + o^*_p(m^{-1/2}). \tag{26}$$

Fix any $\tau^{(1)}, \ldots, \tau^{(T)} \in \mathcal{T} \times \mathcal{T}$ and define the T-vector

$$\Theta^*_{mn} = \big(\theta^*_{mn}(\tau^{(1)}) - E^*[\theta^*_{mn}(\tau^{(1)})], \ldots, \theta^*_{mn}(\tau^{(T)}) - E^*[\theta^*_{mn}(\tau^{(T)})]\big)'.$$

Let

$$\Sigma(\tau, \tau') = \Gamma_1(\tau)^{-1} \, \Omega_2(\tau, \tau') \left(\Gamma_1(\tau')'\right)^{-1}, \qquad \Sigma = \left(\Sigma(\tau^{(t)}, \tau^{(t')})\right)_{t,t'=1}^{T}.$$

Under the maintained conditions ensuring $\frac{1}{m} \sum_{j=1}^{m} \|a_j(\tau)\|^2 = O_p(1)$ for the relevant group-level vectors $a_j(\tau)$, the bootstrap CLT implies that

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

In particular, by (26),

$$\sqrt{m}\Big(\hat{\delta}^*(\hat{\beta}, \tau^{(t)}) - \hat{\delta}(\hat{\beta}, \tau^{(t)})\Big)_{t=1}^{T}$$

converges conditionally to the same Gaussian limit as

$$\sqrt{m}\Big(\hat{\delta}(\hat{\beta}, \tau^{(t)}) - \delta_0(\beta_0, \tau^{(t)})\Big)_{t=1}^{T}.$$

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

Part 3 – Weak convergence of the bootstrap. By Lemma 2, uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$,

$$\frac{1}{m}\sum_{j=1}^{m} \Gamma_{2j}(\delta_0, \beta_0, \tau)\big(\hat{\beta}_j(u) - \beta_{j,0}(u)\big) = o_p\bigg(\frac{1}{\sqrt{m}}\bigg). \tag{27}$$

Since the bootstrap only reweights groups, the same argument yields the bootstrap analogue:

$$\frac{1}{m}\sum_{j=1}^{m} w_j^* \, \Gamma_{2j}(\delta_0, \beta_0, \tau)\big(\hat{\beta}_j(u) - \beta_{j,0}(u)\big) = o_p^*\bigg(\frac{1}{\sqrt{m}}\bigg), \tag{28}$$

uniformly in $\tau \in \mathcal{T} \times \mathcal{T}$.

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

$$\hat{\delta}^*(\hat{\beta}, \tau) - \hat{\delta}(\hat{\beta}, \tau) = \Gamma_1(\delta_0, \beta_0, \tau)^{-1}\left(M^*_{mn}(\delta_0, \beta_0, \tau) - M_{mn}(\delta_0, \beta_0, \tau)\right) + o_p^*\bigg(\frac{1}{\sqrt{m}}\bigg). \tag{29}$$

From Part (i) of the proof of Theorem 2, the class

$$\mathcal{F} = \{\bar{m}_j(\delta_0, \beta_0, \tau) : \tau \in \mathcal{T} \times \mathcal{T}\}$$

satisfies the conditions ensuring that the empirical process

$$\mathbb{G}_m(\tau) = \sqrt{m}\Big(M_{mn}(\delta_0, \beta_0, \tau) - M(\delta_0, \beta_0, \tau)\Big)$$

is asymptotically tight in $\ell^\infty(\mathcal{T} \times \mathcal{T})$ and $\mathbb{G}_m \rightsquigarrow \mathbb{G}$ for a tight mean-zero Gaussian process $\mathbb{G}$ with covariance kernel $\Omega_2(\tau, \tau')$.

Consider the bootstrap empirical process

$$\mathbb{G}^*_m(\tau) = \sqrt{m}\Big(M^*_{mn}(\delta_0, \beta_0, \tau) - M_{mn}(\delta_0, \beta_0, \tau)\Big) = \frac{1}{\sqrt{m}}\sum_{j=1}^{m}(w_j^* - 1)\,\bar{m}_j(\delta_0, \beta_0, \tau).$$

Under the same entropy/moment conditions (and since the bootstrap only reweights the i.i.d. group-level contributions), the multinomial bootstrap empirical-process theorem implies

$$\mathbb{G}^*_m \rightsquigarrow^* \mathbb{G} \quad \text{in } \ell^\infty(\mathcal{T} \times \mathcal{T}),$$

conditionally on the data.

Combining this with the linear representation (29) and the continuous mapping theorem yields

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

D 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 (2025) consider a similar model where the heterogeneity arises from the individual rank variable u_{ij} and the focus is on the within distribution. Starting from equation (3) and assuming that $(x_{1ij}, x_{2j}) \perp\!\!\!\perp v_j$, it is possible to obtain their model by integrating over v:

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

Hence, when model (3) holds, they identify the average effects over groups at the u quantile of the within distribution.

If only the heterogeneity of group averages is of interest, one could consider the conditional quantile function of the conditional expectation function in each group. Specifically, by assuming that $(x_{1ij}, x_{2j}) \perp\!\!\!\perp u_{ij}$, one can attain

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

with

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

This setting is common in empirical research where only aggregated data is available.

If the primary focus is heterogeneity between groups, one may prefer to study heterogeneities in group medians rather than averages. This choice aligns with the framework suggested in this paper, where the specific quantile of $u = 0.5$ is considered.

E Application to Optimal Treatment Assignment

Under the assumption of rank invariance across treatment states, the framework identifies individual treatment effects (see Remark 1 in the paper). This feature makes it particularly useful for optimal treatment assignment when policymakers maximize some social welfare function and baseline outcomes are not observed (see, e.g., Manski, 2004; Kitagawa and Tetenov, 2018, 2021). In such cases, one might want to construct an assignment rule that exploits effect heterogeneity over the outcome distribution, both within and between groups. The model is relevant when treatment is assigned either at the group or individual level. Group-level interventions are common in economics—for instance, place-based policies and infrastructure projects such as highways, railways, and sanitation systems affect all individuals in a locality, and educational policies are often implemented at the jurisdiction, school, or classroom level.

The contribution of the framework is that it exploits treatment effect heterogeneity along both the within- and between-group dimensions, enabling more efficient treatment allocation. This remains true even if the welfare function is utilitarian. For generality, I consider a welfare function as in equation (2) and to keep the exposition focused on the main ideas, I suppress covariates throughout this section. The framework easily extends to include covariates, for example, when they are required for identification. The unconditional quantile function can then be obtained by integrating over the covariates. I use the conventional potential outcome framework with a treatment variable and assume that the two-level quantile function of potential outcomes is identified. Let $y_{ij}(d)$ denote the potential outcome under treatment state $d \in \{0, 1\}$. The quantile treatment effects over both dimensions are given by:[25]

$$\tilde{q}_1(u, v) - \tilde{q}_0(u, v),$$

which under rank invariance equals the treatment effect for an individual at quantiles (u, v).

I consider a policymaker who designs a treatment rule for a *target* population under a capacity or budget constraint, using information from a *sample* population. I assume a static setting where the policymaker chooses whom to treat out of a pool of individuals or groups based on their *unobserved* ranks. This is in contrast to Kitagawa and Tetenov (2021), where the goal is to optimally assign individuals to treatment based on *observable*

[25]Integrating the conditional quantile treatment effects over both u and v yields average treatment effects:

$$ATE = \int_0^1 \int_0^1 [\tilde{q}_1(u, v) - \tilde{q}_0(u, v)]\, dvdu.$$

covariates. Baseline outcomes may be among the covariates but are not always available (e.g., Tarozzi et al., 2015). Formally, the policymaker maximizes welfare over a class of feasible treatment rules $\mathcal{G}$, where each $G \in \mathcal{G}$ assigns individuals to treatment as a function of their ranks $(u_{ij}, v_j(u_{ij})) \in (0,1)^2$.[26] If these ranks were observed, we could include them in the set of covariates, and the problem would coincide with the one in Kitagawa and Tetenov (2021). Since they are unobserved, we instead rely on distributional methods to identify them. This setting also differs from Kaji and Cao (2023), which allows heterogeneity only in one dimension. Instead, with grouped data, one might want to exploit treatment effect heterogeneity within and between groups to more efficiently allocate the treatment.[27]

When a treatment rule G is applied to the target population, the social welfare is proportional to

$$W_w(G) = \int_0^1 \int_0^1 \tilde{q}_G(u,v) w(u,v) dv du, \tag{30}$$

where q_G is the induced quantile function of the outcomes under G:

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

and the optimal treatment rule solves[28]

$$G^* \in \arg\max_{G \in \mathcal{G}} W_w(G). \tag{31}$$

Given a sample population, Kitagawa and Tetenov (2018) derive the optimal treatment assignment rule by assuming that the joint distribution of potential outcomes and covariates is the same in the sample and target populations. In my setting, this requires equality of the joint distribution of $(y_{ij}(1), y_{ij}(0), u_{ij}, v_j)$. Since u_{ij} is normalized, this reduces to assuming that the conditional distribution of $(y_{ij}(1), y_{ij}(0), v_j(u_{ij}))$ given u_{ij} is the same in both populations.

[26] In the following, I write v_j for brevity, though it should be understood as the group's rank at a given within-group percentile.

[27] As in the one-dimensional case, targeting the most deprived units is not necessarily optimal (Haushofer et al., 2025); there may be a trade-off between reaching the most deprived and those who benefit most, especially when group membership strongly shapes treatment impacts.

[28] Solving problem (31) is nontrivial, even with known potential outcomes (Kitagawa and Tetenov, 2021), since an individual's welfare weight may depend on the treatment status of others. Intuitively, an individual's welfare weight increases with the outcomes of others.

Summarizing, identifying $\tilde{q}_1(u,v)$ and $\tilde{q}_0(u,v)$ provide the key inputs for optimal treatment assignment: under rank invariance, they identify individual treatment effects at each (u,v). This pins down the induced outcome distribution under any feasible rule G and the resulting welfare, so the welfare-maximizing rule in (31) is identified.

This Appendix illustrates how the two-dimensional framework can be used for optimal treatment assignment under the assumption of rank invariance. Yet, this assumption may not be realistic in certain settings. Recent work by Kaji and Cao (2023) offers a partial identification alternative based on quantile functions of potential outcomes, showing that meaningful insights about treatment effect heterogeneity are attainable without rank invariance. Exploring such approaches in the multidimensional case is left for future research.[29]

F Additional Empirical Results

F.1 Supplement to Empirical Application - Rank Correlations

Table 4: Correlation of Ranks over u for Income from Work

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

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

[29] See also Chernozhukov et al. (2023), who develop robust conformal inference methods for individual treatment effects without assuming identification of the joint distribution of potential outcomes.