IV QUANTILE REGRESSION FOR GROUP-LEVEL TREATMENTS, WITH AN APPLICATION TO THE DISTRIBUTIONAL EFFECTS OF TRADE

DENIS CHETVERIKOV, BRADLEY LARSEN, AND CHRISTOPHER PALMER

Abstract. We present a methodology for estimating the distributional effects of an endogenous treatment that varies at the group level when there are group-level unobservables, a quantile extension of Hausman and Taylor (1981). Because of the presence of group-level unobservables, standard quantile regression techniques are inconsistent in our setting even if the treatment is independent of unobservables. In contrast, our estimation technique is consistent as well as computationally simple, consisting of group-by-group quantile regression followed by two-stage least squares. Using the Bahadur representation of quantile estimators, we derive weak conditions on the growth of the number of observations per group that are sufficient for consistency and asymptotic zero-mean normality of our estimator. As in Hausman and Taylor (1981), micro-level covariates can be used as internal instruments for the endogenous group-level treatment if they satisfy relevance and exogeneity conditions. An empirical application indicates that low-wage earners in the US from 1990–2007 were significantly more affected by increased Chinese import competition than high-wage earners. Our approach applies to a broad range of settings in labor, industrial organization, trade, public finance, and other applied fields.

Keywords: quantile regression, instrumental variables, panel data, income inequality, import competition

JEL Classification: C21, C31, C33, C36, F16, J30

Date: First version: June 2012. This version: March 2015.

We thank Moshe Buchinsky, Ivan Canay, Brigham Frandsen, Antonio Galvao, Wenshu Guo, Jerry Hausman, Rosa Matzkin, Whitney Newey, and Christopher Taber for helpful comments; and Yuqi Song and Caio Waisman for meticulous research assistance. We are especially grateful to Jin Hahn for many useful discussions.

Chetverikov: UCLA Department of Economics; chetverikov@econ.ucla.edu.

Larsen: Stanford University Department of Economics and NBER; bjlarsen@stanford.edu.

Palmer: University of California at Berkeley Haas School of Business; cjpalmer@berkeley.edu.

1. Introduction

In classical panel-data models for mean regression, fixed effects are commonly used to obtain identification when time-invariant unobservables are correlated with included variables. While this approach yields consistent estimates of the coefficients on time-varying variables, it precludes identification of the coefficients of any time-invariant variables, as these variables are eliminated by the within-group transformation. In an influential paper, Hausman and Taylor (1981) demonstrated that exogenous between variation of time-varying variables can help to identify the coefficients of time-invariant variables after their within variation has been used to identify the coefficients on time-varying variables, thus yielding identification of the whole model without external instruments. Our paper provides a quantile extension of the Hausman and Taylor (1981) classical linear panel estimator.

We present our model in Section 2. To clarify the range of potential applications of our estimator, we depart in the model from the usual panel-data terminology and refer to panel units as groups (instead of as individuals; groups might be states, cities, schools, etc.) and to within-group observations as individuals or micro-level observations (instead of as time observations; individuals might be students, families, firms, etc.).[1] The model is of practical significance when the researcher has data on a group-level endogenous treatment and has microdata on the outcome of interest within each group. For example, a researcher may be interested in the effect of a policy which varies across states and years (a "group") on the within-group distribution of micro-level outcomes. In Section 2, we also explain how the problem we solve differs from others in the quantile regression literature, and we demonstrate that, as in Hausman and Taylor (1981), micro-level covariates can be used as internal instruments for the endogenous group-level treatment if they satisfy relevance and exogeneity conditions. This last feature of the model is especially appealing because in practice it may be difficult to find external instruments.

We introduce our estimator in Section 3. The estimator is computationally simple to implement and consists of two steps: 1) perform quantile regression within each group to estimate effects of micro-level covariates, or, if no micro-level covariates are included, calculate the desired quantile for the outcome within each group; and 2) regress the estimated group-specific effects on group-level covariates using either 2SLS, if the group-level covariates are endogenous, or OLS, if the group-level covariates are exogenous, either of which cases would render standard quantile regression (e.g. Koenker and Bassett 1978) inconsistent.[2] Section 3 also highlights a variety of applied micro settings in which our estimator is useful (with detailed example applications discussed in

[1] Similar terminology is used, for example, by Altonji and Matzkin (2005).

[2] Even in the absence of endogeneity, the Koenker and Bassett (1978) estimator will be inconsistent in our setting because of group-level unobservables, akin to left-hand side measurement error; see Section 2 for details on our setting. While posing no problems for linear models, left-hand side errors-in-variables can bias quantile estimation (see Hausman (2001) and Hausman, Luo, and Palmer (2014)).

Appendix A) and discusses Monte Carlo simulations (found in Appendix B) that demonstrate that our estimator has much lower bias than that of the standard quantile regression estimator when the group-level treatment is endogenous, even in small samples, and at larger sample sizes our estimator outperforms quantile regression even when the treatment is exogenous. Section 3 also highlights additional computational benefits of our estimator.

We derive theoretical properties of the estimator in Section 4. The results are based on asymptotics where both the number of groups and the number of observations per group grow to infinity. While linear panel models, including Hausman and Taylor (1981), admit a simple *unbiased* fixed effects estimator and hence do not require asymptotics in the number of observations per group, quantile estimators are *biased in finite samples* leading to inconsistency of our estimator if the number of observations per group remains small as the number of groups increases, and making the estimator inappropriate in the settings with a small number of observations per group and a large number of groups. However, since quantile estimators are *asymptotically unbiased*, we are able to employ Bahadur's representation of quantile estimators to derive weak conditions on the growth of the number of observations per group that are sufficient for the consistency and asymptotic zero-mean normality of our estimator. Importantly, the attractive theoretical properties of the estimator remain valid even if the number of observations per group is relatively small in comparison with the number of groups. We demonstrate that standard errors for the proposed estimator can be obtained using traditional robust variance estimators for 2SLS (heteroskedasticity-robust and clustered), making inference particularly simple. Finally, we show how to construct confidence bands for the coefficient of interest which hold uniformly over a set of quantiles of interest via multiplier bootstrap procedure.

Section 5 presents an empirical application which studies the effect of trade on the distribution of wages within local labor markets. We build on the work of Autor, Dorn, and Hanson (2013), who studied the effect of Chinese import competition on average wages in local labor markets. Using the grouped IV quantile regression approach developed here, we find that Chinese import competition reduced the wages of low-wage earners (individuals at the bottom quartile of the conditional wage distribution) more than high-wage earners, particularly for females, heterogeneity which is missed by focusing on traditional 2SLS estimates.

To the best of our knowledge, our paper is the first to present a framework for estimating distributional effects as a function of group-level covariates. There is, however, a large literature studying quantile models for panel data when the researcher wishes to estimate distributional effects of micro-level covariates. See, for example, Koenker (2004), Abrevaya and Dahl (2008), Lamarche (2010), Canay (2011), Galvao (2011), Kato and Galvao (2011), Ponomareva (2011), Kato, Galvao, and Montes-Rojas (2012), Rosen (2012), Arellano and Bonhomme (2013), and Galvao and Wang (2013). Our paper also contributes to the growing literature on IV treatment effects in quantile models, such as Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005, 2006, 2008), Lee (2007), Chesher (2003), and Imbens and Newey (2009). Our paper differs, however, in that this literature focuses on the case where individual-level unobserved heterogeneity is correlated with an individual-level treatment, whereas we focus on the case where a group-level, additively separable unobservable is correlated with a group-level treatment.

Throughout the paper, we use the following notation. The symbol $\|\cdot\|$ denotes the Euclidean norm. The symbol $\Rightarrow$ signifies weak convergence, and $l^\infty(\mathcal{U})$ represents the set of bounded functions on $\mathcal{U}$. With some abuse of notation, $\ell^\infty(\mathcal{U})$ also denotes the set of component-wise bounded vector-valued functions on $\mathcal{U}$. All equalities and inequalities concerning random variables are implicitly assumed to hold almost surely. All proofs and some extensions of our results are contained in the Appendix.

2. Model

We study a panel data quantile regression model for a response variable y_{ig} of individual i in group g. We first present the model in its most general form in equations (1) and (2) below and then discuss a particular operationalization of the model in equation (3) that will be particularly appealing for applied work. In the general model, we assume that the uth quantile of the conditional distribution of y_{ig} is given by

$$Q_{y_{ig}|z_{ig},x_g,\alpha_g}(u) = z_{ig}'\alpha_g(u), \quad u \in \mathcal{U} \tag{1}$$

$$\alpha_{g,1}(u) = x_g'\beta(u) + \varepsilon_g(u), \quad u \in \mathcal{U}, \tag{2}$$

where $Q_{y_{ig}|z_{ig},x_g,\alpha_g}(u)$ is the uth conditional quantile of y_{ig} given (z_{ig}, x_g, α_g), z_{ig} is a d_z-vector of observable individual-level covariates (which we sometimes refer to as micro-level covariates), $\alpha_g = \{\alpha_g(u), u \in \mathcal{U}\}$ is a set of group-specific effects with $\alpha_{g,1}(u)$ being the first component of the vector $\alpha_g(u) = (\alpha_{g,1}(u), \dots, \alpha_{g,d_z}(u))'$, x_g is a d_x-vector of observable group-level covariates (x_g contains a constant), $\beta(u)$ is a d_x-vector of coefficients, $\varepsilon_g = \{\varepsilon_g(u), u \in \mathcal{U}\}$ is a set of unobservable group-level random scalar shifters,[3] and $\mathcal{U}$ is a set of quantile indices of interest. Thus, we assume that the response variable y_{ig} satisfies the quantile regression model in (1) with group-specific effects $\alpha_g(u)$. We are primarily interested in studying how these effects depend on the group-level covariates x_g, and, without loss of generality, we focus on $\alpha_{g,1}(u)$, the first component of the vector $\alpha_g(u)$. To make the problem operational, we assume that $\alpha_{g,1}(u)$ satisfies the linear regression model (2), in which we are interested in estimating the vector of coefficients $\beta(u)$.

In empirical work, we envision that the most useful variant of the model (1)-(2) would be case where the first element of z_{ig} corresponds to a constant and where coefficients on micro-level

[3] One interpretation of the term $\varepsilon_g(u)$ in (2) is that it accounts for all unobservable group-level covariates η_g that affect $\alpha_{g,1}(u)$ but are not included in x_g. In this case, $\varepsilon_g(u) = \varepsilon(u, \eta_g)$. Note that we do not impose any parametric restrictions on $\varepsilon(u, \eta_g)$, and so we allow for arbitrary nonlinear effects of the group-level unobservable covariates that can affect different quantiles in different ways.

covariates do not vary by group, given by the model

$$Q_{y_{ig}|\widetilde{z}_{ig},x_g,\varepsilon_g}(u) = \widetilde{z}_{ig}'\gamma(u) + x_g'\beta(u) + \varepsilon_g(u), \quad u \in \mathcal{U}, \tag{3}$$

which is obtained from (1)-(2) by assuming that $(\alpha_{g,2}(u), \ldots, \alpha_{g,d_z}(u))' = \gamma(u)$ for some non-stochastic $(d_z - 1)$-vector $\gamma(u)$ and all $g = 1, \ldots, G$, setting $z_{ig} = (1, \widetilde{z}_{ig})'$, and substituting (2) into (1). This model allows for the analysis of location-shift effects of the group-level covariates x_g on the conditional distribution of y_{ig} in the group g.

As an example of where the above modeling framework is useful, consider a case in which a researcher wishes to model the effects of a policy, contained in x_g, which varies at the state-by-year level (a "group" in this setting) on the distribution of micro-level outcomes (such as individuals' wages within each state-by-year combination), denoted y_{ig}, conditional on micro-level covariates, such as education level, denoted z_{ig}. The framework in (3) would model the location-shift effect of the policy on conditional quantiles of wages within a group, given by $\beta(u)$. The additional flexibility of (1)-(2) would also allow for interaction effects. For example, a policy x_g may have differential effects on lower wage quantiles for the less-educated than for the higher-educated; model (1) would capture this idea by allowing the researcher to specify a linear regression model of the form of (2) for the component of α_g that is the coefficient on education level, allowing the researcher to study how the effect of education level on the wage distribution varies as a function of x_g, the policy.[4]

In many applications, it is likely that the group-level covariates x_g may be endogenous in the sense that $E[x_g\varepsilon_g(u)] \neq 0$, at least for some values of the quantile index $u \in \mathcal{U}$. Therefore, to increase applicability of our results, we assume that there exists a d_w-vector of observable instruments w_g such that $E[w_g\varepsilon_g(u)] = 0$ for all $u \in \mathcal{U}$, $E[w_g x_g']$ is nonsingular, and y_{ig} is independent of w_g conditional on (z_{ig}, x_g, α_g).[5] The first two conditions are familiar from the classical linear instrumental variable regression analysis, and the third condition requires the distribution of y_{ig} to be independent of w_g once we control for z_{ig}, x_g, and α_g. It implies, in particular, that $Q_{y_{ig}|z_{ig},x_g,\alpha_g,w_g}(u) = z_{ig}'\alpha_g(u)$ for all $u \in \mathcal{U}$.[6]

[4] If the researcher is interested in modeling several effects, for example location-shift and some interaction effects, she can specify a linear regression model of the form (2) for each effect.

[5] To understand the assumption that $E[w_g\varepsilon_g(u)] = 0$ holds jointly for all $u \in \mathcal{U}$, assume, for example, that $\varepsilon_g(u) = \varepsilon(u, \eta_g)$ where η_g is a vector of group-level omitted variables in regression (2). Then a sufficient condition for the assumption $E[w_g\varepsilon_g(u)] = E[w_g\varepsilon(u, \eta_g)] = 0$ is that $E[\varepsilon(u, \eta_g)|w_g] = 0$. In turn, the restriction of the condition $E[\varepsilon(u, \eta_g)|w_g] = 0$ is that $E[\varepsilon(u, \eta_g)|w_g]$ does not depend on w_g, which occurs (for example) if η_g is independent of w_g. Once we assume that $E[\varepsilon(u, \eta_g)|w_g]$ does not depend on w_g, the further restriction that $E[\varepsilon(u, \eta_g)|w_g] = 0$ is a normalization of the component of the vector $\beta(u)$ corresponding to the constant in the vector x_g.

[6] The setting we model differs from other IV quantile settings, such as Chernozhukov and Hansen (2005, 2006, 2008). Consider, for simplicity, our model (3) and assume that $\mathcal{U} = [0, 1]$. Then the Skorohod representation implies that $y_{ig} = \widetilde{z}_{ig}'\gamma(u_{ig}) + x_g'\beta(u_{ig}) + \varepsilon_g(u_{ig})$ where u_{ig} is a random variable that is distributed uniformly on $[0, 1]$ and is independent of $(\widetilde{z}_{ig}, x_g, \varepsilon_g)$. Here, one can think of u_{ig} as unobserved individual-level heterogeneity. In this model, the unobserved group-level component $\varepsilon_g(\cdot)$ is modeled as an additively separable term. In contrast, the model in

We assume that a researcher has data on G groups and N_g individuals within group $g = 1, ..., G$. Thus, the data consist of observations on $\{(z_{ig}, y_{ig}), i = 1, \dots, N_g\}$, x_g, and w_g for $g = 1, \dots, G$. Throughout the paper, we denote $N_G = \min_{1 \le g \le G} N_g$. For our asymptotic theory in Section 4, we will assume that N_G gets large as $G \to \infty$. Specifically, for the asymptotic zero-mean normality of our estimator $\hat{\beta}(u)$ of $\beta(u)$, we will assume that $G^{2/3}(\log N_G)/N_G \to 0$ as $G \to \infty$; see Assumption 3 below. Thus, our results are useful when both G and N_G are large, which occurs in many empirical applications, but we also note that our results apply even if the number of observations per group is relatively small in comparison with the number of groups.

We also emphasize that, like in the original panel data *mean* regression model of Hausman and Taylor (1981), an important feature of our panel data *quantile* regression model is that it allows for *internal* instruments. Specifically, if some component of the vector z_{ig}, say $z_{ig,k}$, is exogenous in the sense that $E[z_{ig,k}\varepsilon_g(u)] = 0$ for all $u \in \mathcal{U}$, we can use, for example, $N_g^{-1/2} \sum_{i=1}^{N_g} z_{ig,k}$ as an additional instrument provided it is correlated with x_g, including it into the vector w_g. Since in practice it is often difficult to find an appropriate external instrument, allowing for internal instruments greatly increases applicability of our results.

Our problem in this paper is different from that studied in Koenker (2004), Kato, Galvao, and Montes-Rojas (2012), and Kato and Galvao (2011).[7] Specifically, they considered the panel data quantile regression model

$$Q_{y_{ig}|z_{ig},\alpha_g}(u) = z_{ig}'\gamma(u) + \alpha_g(u), \quad u \in \mathcal{U}, \tag{4}$$

and developed estimators of $\gamma(u)$. Building on Koenker (2004), Kato, Galvao, and Montes-Rojas (2012) suggested estimating $\gamma(u)$ in this model by running a quantile regression estimator of Koenker and Bassett (1978) on the pooled data, treating $\{\alpha_g(u), g = 1, \dots, G\}$ as a set of parameters to be estimated jointly with the vector of parameters $\gamma(u)$ (the same technique can be used to estimate $\gamma(u)$ in our model (3) by setting $\alpha_g(u) = x_g'\beta(u) + \varepsilon_g(u)$). They showed that their estimator is asymptotically zero-mean normal if $G^2(\log G)^3/N_G \to 0$ as $G \to \infty$. Making further progress, Kato and Galvao (2011) suggested an interesting smoothed quantile regression estimator

Chernozhukov and Hansen (2005, 2006, 2008) assumes that $\varepsilon_g(u) = 0$ for all $u \in [0, 1]$ and instead assumes that u_{ig} is not independent of $(\tilde{z}_{ig}, x_g)$. Thus, these two models are different and require different analysis.

[7]Our paper is also related to but different from Graham and Powell (2012) who studied the model that in our notation would take the form $y_{ig} = z_{ig}'\alpha_g(u_{ig})$ where u_{ig} represents (potentially multi-dimensional) random unobserved heterogeneity, and developed an interesting identification and estimation strategy for the parameter $E[\alpha_g(u_{ig})]$, achieving identification when the number of observations per group remains small as the number of groups gets large and, under certain conditions, allowing $\alpha_g(\cdot) = \alpha_{ig}(\cdot)$ to depend on i.

of $\gamma(u)$ that is asymptotically zero-mean normal if $G/N_G \to 0$.[8] These papers do not provide a model for our estimator of $\beta(u)$, our primary object of interest, but instead focus solely on $\gamma(u)$.

Our model is also different from that studied in Hahn and Meinecke (2005), who considered an extension of Hausman and Taylor (1981) to cover non-linear panel data models. Formally, they considered a non-linear panel data model defined by the following equation:

$$E\left[\varphi(y_{ig}, z_{ig}'\gamma + x_g'\beta + \varepsilon_g)\right] = 0$$

where $\varphi(\cdot,\cdot)$ is a vector of moment functions and $x_g'\beta + \varepsilon_g$ is the group-specific effect. As in this paper, the authors were interested in estimating the effect of group-level covariates (coefficient β) without assuming that ε_g is independent (or mean-independent) of x_g but assuming instead that there exists an instrument w_g satisfying $E[w_g\varepsilon_g] = 0$. Importantly, however, they assumed that $\varphi(\cdot,\cdot)$ is a vector of *smooth* functions, so that their results do not apply immediately to our model. In addition, Hahn and Meinecke (2005) required that $N_G/G > c$ for some $c > 0$ uniformly over all G to prove that their estimator is asymptotically zero-mean normal. In contrast, as emphasized above, we only require that $G^{2/3}(\log N_G)/N_G \to 0$ as $G \to \infty$, with the improvement coming from a better control of the residuals in the Bahadur representation.[9]

3. Estimator

In this section we develop our estimator. Our main emphasis is to derive a computationally simple, yet consistent, estimator. The estimator consists of the following two stages.

Stage 1: For each group g and each quantile index u from the set $\mathcal{U}$ of indices of interest, estimate uth quantile regression of y_{ig} on z_{ig} using the data $\{(y_{ig}, z_{ig}) : i = 1, ..., N_g\}$ by the classical quantile regression estimator of Koenker and Bassett (1978):

$$\hat{\alpha}_g(u) = \arg\min_{a\in\mathbb{R}^{d_z}} \sum_{i=1}^{N_g} \rho_u(y_{ig} - z_{ig}'a),$$

[8] To clarify the difference between the growth condition in our paper, which is $G^{2/3}(\log N_G)/N_G \to 0$, and the growth condition, for example, in Kato, Galvao, and Montes-Rojas (2012), which is $G^2(\log G)^3/N_G \to 0$, assume, for simplicity, that $d_x = 1$, $d_z = 2$, and x_g and the second component of z_{ig} are constants, that is, $x_g = 1$ and $z_{ig} = (\widetilde{z}_{ig}, 1)'$. Then our model (1)-(2) reduces to $Q_{y_{ig}|\widetilde{z}_{ig},\varepsilon_g,\alpha_g}(u) = \widetilde{z}_{ig}(\beta(u)+\varepsilon_g(u)) + \alpha_g(u)$, which is similar to the model (4) studied in Kato, Galvao, and Montes-Rojas (2012) with the exception that we allow for additional group-specific random shifter $\varepsilon_g(u)$. When $\varepsilon_g(u)$ is present, our estimator $\hat{\beta}(u)$ of $\beta(u)$ satisfies $G^{1/2}(\hat{\beta}(u)-\beta(u)) \Rightarrow N(0, V_1)$ for some non-vanishing variance V_1; see Section 4. When $\varepsilon_g(u)$ is set to zero, however, V_1 vanishes, making the limiting distribution degenerate and leading to faster convergence rate of the estimator $\hat{\beta}(u)$. In fact, when V_1 vanishes, one obtains $(GN_G)^{1/2}(\hat{\beta}(u) - \beta(u)) \Rightarrow N(0, V_2)$ for some non-vanishing variance V_2. An additional $N_G^{1/2}$ factor in turn appears in the residual terms of the Bahadur representation of the estimator $\hat{\beta}(u)$, which eventually lead to stronger requirements on the growth of the number of observations per group N_G relative to the number of groups, explaining the difference between the growth condition in Kato, Galvao, and Montes-Rojas (2012) and our growth condition.

[9] Appendix F contains additional discussion of the model, including an extension to a random coefficients setting.

where $\rho_u(x) = (u - 1\{x < 0\})x$ for $x \in \mathbb{R}$. Denote $\hat{\alpha}_g(u) = (\hat{\alpha}_{g,1}(u), \ldots, \hat{\alpha}_{g,d_z})'$.

Stage 2: Estimate a 2SLS regression of $\hat{\alpha}_{g,1}(u)$ on x_g using w_g as an instrument to get an estimator $\hat{\beta}(u)$ of $\beta(u)$, that is,

$$\hat{\beta}(u) = \left(X' P_W X\right)^{-1} \left(X' P_W \hat{A}(u)\right)$$

where $X = (x_1, ..., x_G)'$, $W = (w_1, ..., w_G)'$, $\hat{A}(u) = (\hat{\alpha}_{1,1}(u), \ldots, \hat{\alpha}_{G,1}(u))'$, and $P_W = W(W'W)^{-1}W'$.

Intuitively, as the number of observations per group increases, $\hat{\alpha}_{g,1} - \alpha_{g,1}$ shrinks to zero uniformly over $g = 1, \ldots, G$, and we obtain a classical instrumental variables problem. The theory presented below provides a mild condition on the growth of the number of observations per group that is sufficient to achieve consistency and asymptotic zero-mean normality of $\hat{\beta}(u)$.

Several special cases of our estimator are worth noting. First, when the model is given by equation (3), the steps of our estimator consist of 1) group-by-group quantile regression of y_{ig} on $\tilde{z}_{ig}$ and on a constant, saving the estimated coefficient $\hat{\alpha}_{g,1}(u)$ corresponding to the constant, $\alpha_{g,1}(u) = x_g'\beta(u) + \varepsilon_g(u)$, in each group; and 2) regressing those saved coefficients $\hat{\alpha}_{g,1}(u)$ on x_g via 2SLS using w_g as instruments. Second, if z_{ig} contains *only* a constant, the first stage simplifies to selecting the uth quantile of the outcome variable y_{ig} within each group. Third, if x_g is exogenous, that is, $E[x_g \varepsilon_g(u)] = 0$, OLS of $\hat{\alpha}_{g,1}(u)$ on x_g may be used rather than 2SLS in the second stage. In this latter case, the grouped quantile estimation approach provides the advantage of handling group-level unobservables (or, alternatively, left-hand-side measurement error), which would bias the traditional Koenker and Bassett (1978) estimator. When z_{ig} only includes a constant and x_g is exogenous, the grouped IV quantile regression estimator $\hat{\beta}(u)$ simplifies to the minimum distance estimator described in Chamberlain (1994) (see also Angrist, Chernozhukov, and Fernandez-Val 2006).

This estimator has several computational benefits relative to alternative methods. First, note that when the model is given by equation (3), another approach to perform the first stage of our estimator would be to denote $\alpha_{g,1}(u) = x_g'\beta(u) + \varepsilon_g(u)$ and estimate parameters $\gamma(u)$ and $\{\alpha_{g,1}(u) : g = 1, \ldots, G\}$ jointly from the pooled dataset as in Kato, Galvao, and Montes-Rojas (2012). This would provide an efficiency gain given that in this case, individual-level effects $\gamma(u)$ are group-independent. Although the method we use is less efficient, it is computationally much less demanding since only few parameters are estimated in each regression, which can greatly reduce computation times in large datasets with many fixed effects.[10] Second, even if no group-level unobservables exist (consider model (3) with $\varepsilon_g(u) = 0$ for all $g = 1, \ldots, G$), the grouped estimation approach can be considerably faster than the traditional Koenker and Bassett (1978)

[10]In Monte Carlo experiments in Appendix B, we find that jointly estimating group-level effects can take over 150 times as long as the grouped quantile approach when $G = 200$. With $G > 200$, the computation time ratio drastically increases further, with standard optimization packages often failing to converge appropriately.

estimator (though both estimators will be consistent). This computational advantage occurs when the dimension of x_g is large: standard quantile regression estimates $\beta(u)$ in a single, nonlinear step, whereas the grouped quantile approach estimates $\beta(u)$ in a linear second stage.[11]

Monte Carlo simulations in Appendix B highlight the performance of our estimator for $\beta(u)$ in (3) relative to the traditional Koenker and Bassett (1978) estimator (which ignores endogeneity of x_g as well as the existence of $\varepsilon_g(u)$). Even when N_G and G are both small, the grouped IV quantile approach has lower bias than traditional quantile regression when x_g is endogenous. When x_g is exogenous but group-level unobservables $\varepsilon_g(u)$ are still present, the bias of the grouped quantile approach shrinks quickly to zero as N_G grows but the bias of traditional quantile estimator does not. When no group-level unobservables are present, and hence both the grouped estimation approach and traditional quantile regression should be consistent, our estimator still has small bias, although traditional quantile regression outperforms our method in this case.

As we demonstrate in Section 4 below, standard errors for our estimator $\hat{\beta}(u)$ may be obtained using standard heteroskedasticity-robust or clustering approaches for 2SLS or OLS as if there were no first stage.[12] Section 4 also describes a multiplier bootstrap procedure that is suitable for constructing uniform confidence bands for the case when the researcher is interested in the set $\mathcal{U}$ of quantile indices u.

To conclude this section, we note that our estimator applies to a wide variety of settings in labor, industrial organization, trade, public finance, development, and other applied fields. Appendix A illustrates examples from Angrist and Lang (2004), Larsen (2014), Palmer (2011), and Backus (2014).

4. Asymptotic Theory

In this section, we formulate our assumptions and present the main theoretical results of the paper.

4.1. Assumptions.

Let c_M, c_f, C_M, C_f, C_L be strictly positive constants whose values are fixed throughout the paper. Recall that $N_G = \min_{g=1,\dots,G} N_g$. We start with specifying our main assumptions.

A1 (Design). *(i) Observations are independent across groups. (ii) For all $g = 1, \dots, G$, the pairs (z_{ig}, y_{ig}) are i.i.d. across $i = 1, \dots, N_g$ conditional on (x_g, α_g).*

A 2 (Instruments). *(i) For all $u \in \mathcal{U}$ and $g = 1, \dots, G$, $E[w_g \varepsilon_g(u)] = 0$. (ii) As $G \to \infty$, $G^{-1} \sum_{g=1}^{G} E[x_g w_g'] \to Q_{xw}$ and $G^{-1} \sum_{g=1}^{G} E[w_g w_g'] \to Q_{ww}$ where Q_{xw} and Q_{ww} are matrices with*

[11] One such example would be a case where a group is a state-by-year combination, and x_g contains many state and year fixed effects, in addition to the treatment of interest, as in Example 2 of Appendix A.

[12] Note that clustering in the second stage refers to dependence *across* groups, not within groups. For example, if a group is a state-by-year combination, the researcher may wish to use standard errors which are clustered at the state level.

singular values bounded in absolute value from below by c_M and from above by C_M. (iii) For all $g = 1, \ldots, G$ and $i = 1, \ldots, N_g$, y_{ig} is independent of w_g conditional on (z_{ig}, x_g, α_g). (iv) For all $g = 1, \ldots, G$, $E[\|w_g\|^{4+c_M}] \leq C_M$.

A3 (Growth Condition)**.** *As $G \to \infty$, we have $G^{2/3}(\log N_G)/N_G \to 0$.*

Assumption 1(i) holds, for example, if groups are sampled randomly from some population of groups. This assumption precludes the possibility of clustering across groups (for example, if a group is a state-by-year combination, there may be clustering on the state level). Since clustered standard errors are important in practice, however, we derive an extension of our results relaxing the independence across groups condition and allowing for clustering in Appendix E. Assumption 1(ii) allows for inter-dependence (clustering) within groups but imposes the restriction that the inter-dependence between observations within the group g is fully controlled for by the group-level covariates x_g and the group-specific effect α_g. Assumption 2 is our main identification condition. Note that Assumption 2 allows for internal instruments. In particular, if $w_g = N_g^{-1/2} \sum_{i=1}^{N_g} z_{ig,k}$ for some k, then Assumption 2(iii) automatically follows from Assumption 1(ii). Assumption 3 implies that the number of observations per group grows sufficiently fast as G gets large, and gives a particular growth rate that suffices for our results. Note that our growth condition is rather weak and, most importantly, allows for the case when the number of observations per group is small relative to the number of groups.[13]

Next, we specify technical conditions that are required for our analysis. Let $E_g[\cdot] = E[\cdot|x_g, \alpha_g]$, and let $f_g(\cdot)$ denote the conditional density function of y_{1g} given (z_{1g}, x_g, α_g) (dependence of $f_g(\cdot)$ on z_{1g} is not shown explicitly for brevity of notation). Also denote $B_g(u, c) = (z_{1g}'\alpha_g(u) - c, z_{1g}'\alpha_g(u) + c)$ for $c > 0$. We will assume the following regularity conditions:

A4 (Covariates)**.** *(i) For all $g = 1, \ldots, G$ and $i = 1, \ldots, N_g$, random vectors z_{ig} and x_g satisfy $\|z_{ig}\| \leq C_M$ and $\|x_g\| \leq C_M$. (ii) For all $g = 1, \ldots, G$, all eigenvalues of $E_g[z_{1g}z_{1g}']$ are bounded from below by c_M.*

A5 (Coefficients)**.** *For all $u_1, u_2 \in \mathcal{U}$ and $g = 1, \ldots, G$, $\|\alpha_g(u_2) - \alpha_g(u_1)\| \leq C_L|u_2 - u_1|$.*

A6 (Noise)**.** *(i) For all $g = 1, \ldots, G$, $E[\sup_{u \in \mathcal{U}} |\varepsilon_g(u)|^{4+c_M}] \leq C_M$. (ii) For some (matrix-valued) function $J : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^{d_w \times d_w}$, $G^{-1} \sum_{g=1}^{G} E[\varepsilon_g(u_1)\varepsilon_g(u_2)w_g w_g'] \to J(u_1, u_2)$ uniformly over $u_1, u_2 \in \mathcal{U}$. (iii) For all $u_1, u_2 \in \mathcal{U}$, $|\varepsilon_g(u_2) - \varepsilon_g(u_1)| \leq C_L|u_2 - u_1|$.*

A7 (Density)**.** *(i) For all $u \in \mathcal{U}$ and $g = 1, \ldots, G$, the conditional density function $f_g(\cdot)$ is continuously differentiable on $B_g(u, c_f)$ with the derivative $f_g'(\cdot)$ satisfying $|f_g'(y)| \leq C_f$ for all $y \in B_g(u, c_f)$*

[13] Using the more common notation of panel data models, where N is the number of individuals (groups) and T is the number of time periods (individuals within the group), Assumption 3 would take the form: $N^{2/3}(\log T)/T \to 0$ as $N \to \infty$.

and $|f'_g(z'_{1g}\alpha_g(u))| \geq c_f$*. (ii) For all* $u \in \mathcal{U}$ *and* $g = 1, \dots, G$*,* $f_g(y) \leq C_f$ *for all* $y \in B_g(u, c_f)$ *and* $f_g(z'_{1g}\alpha_g(u)) \geq c_f$*.*

A8 (Quantile indices)**.** *The set of quantile indices* $\mathcal{U}$ *is a compact set included in* $(0,1)$*.*

Assumption 4(i) requires that both individual and group-level observable covariates z_{ig} and x_g are bounded. Assumption 4(ii) is a familiar identification condition in regression analysis. Assumption 5 is a mild continuity condition. Assumption 6(i) requires sufficient integrability of the noise $\varepsilon_g(u)$, which is a mild regularity condition. In fact, under Assumption 6(iii), which is also a mild continuity condition, Assumption 6(i) is satisfied as long as $E[|\varepsilon_g(u)|^{4+c_M}] \leq C_M$ for some $u \in \mathcal{U}$ (with a possibly different constant C_M). Assumption 6(ii) is trivially satisfied if the pairs (w_g, ε_g) are i.i.d. across g. Assumption 7 is a mild regularity condition that is typically imposed in the quantile regression analysis. Finally, Assumption 8 excludes quantile indices that are too close to either 0 or 1 (when the quantile index u is close to either 0 or 1, one obtains a so called extremal quantile model, which requires a rather different analysis; see, for example, Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011)).

4.2. **Results.** We now present our main results. We start by deriving the asymptotic distribution of our estimator in Theorem 1. Further, we show how to estimate the asymptotic covariance of our estimator in Theorem 2. Finally, we demonstrate how to obtain uniform over $u \in \mathcal{U}$ confidence bands for the parameter of interest $\{\beta(u), u \in \mathcal{U}\}$ via a multiplier bootstrap method in Theorem 3. The first theorem derives the asymptotic distribution of our estimator.

Theorem 1 (Asymptotic Distribution)**.** *Let Assumptions 1-8 hold. Then*

$$\sqrt{G}(\hat{\beta}(\cdot) - \beta(\cdot)) \Rightarrow \mathbb{G}(\cdot), \text{ in } \ell^\infty(\mathcal{U})$$

where $\mathbb{G}(\cdot)$ *is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function* $\mathcal{C}(u_1, u_2) = SJ(u_1, u_2)S'$ *where* $S = \left(Q_{xw}Q_{ww}^{-1}Q'_{xw}\right)^{-1} Q_{xw}Q_{ww}^{-1}$*,* Q_{xw} *and* Q_{ww} *appear in Assumption 2, and* $J(u_1, u_2)$ *in Assumption 6.*

Remark 1. (i) This is our main convergence result that establishes the asymptotic behavior of our estimator. Note that we provide the *joint* asymptotic distribution of our estimator for all $u \in \mathcal{U}$. In addition, Theorem 1 implies that for any $u \in \mathcal{U}$,

$$\sqrt{G}(\hat{\beta}(u) - \beta(u)) \Rightarrow N(0, V)$$

where $V = SJ(u,u)S'$, which is the asymptotic distribution of the classical 2SLS estimator.

(ii) In order to establish the joint asymptotic distribution of our estimator for all $u \in \mathcal{U}$, we have to deal with G independent quantile processes $\{\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u), u \in \mathcal{U}\}$. Since $G \to \infty$, classical functional central limit theorems do not apply. Therefore, we employ a non-standard but powerful Bracketing by Gaussian Hypotheses Theorem, which is also related to majorizing measures for Gaussian processes; see Theorem 2.11.11 in Van der Vaart and Wellner (1996).

(iii) Since quantile regression estimators are biased in finite samples, our estimator $\hat{\alpha}_{g,1}(u)$ of $\alpha_{g,1}(u)$ does not necessarily satisfy $E[(\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u))w_g] = 0$. For this reason, our estimator $\hat{\beta}(u)$ of $\beta(u)$ is not consistent if N_g is bounded from above uniformly over $g = 1, \dots, G$ and $G \geq 2$. We note, however, that quantile estimators are *asymptotically* unbiased, and so we use the Bahadur representation of quantile estimators to derive weak condition on the growth of $N_G = \min_{1 \leq g \leq G} N_g$ relative to G, so that consistent estimation of $\beta(u)$ is indeed possible. Specifically, we prove consistency and asymptotic zero-mean normality under Assumption 3 that states that $G^{2/3}(\log N_G)/N_G \to 0$ as $G \to \infty$, which is a mild growth condition. In principle, it is also possible to consider bias correction of the quantile regression estimators. This would further relax the growth condition on N_G relative to G at the expense of stronger side assumptions and more complicated estimation procedures.

(iv) The requirement that $N_G \to \infty$ as $G \to \infty$ is in contrast with the classical results of Hausman and Taylor (1981) on estimation of panel data mean regression model. The main difference is that the fixed effect estimator in the panel data mean regression model is unbiased even in finite samples leading to consistent estimators of the effects of group-level covariates with the number of observations per group being fixed. □

The result in Theorem 1 derives asymptotic behavior of our estimator. In order to perform inference, we also need an estimator of the asymptotic covariance function. We suggest using an estimator $\hat{\mathcal{C}}(\cdot, \cdot)$ that is defined for all $u_1, u_2 \in \mathcal{U}$ as

$$\hat{\mathcal{C}}(u_1, u_2) = \hat{S}\hat{J}(u_1, u_2)\hat{S}'$$

where

$$\hat{J}(u_1, u_2) = \frac{1}{G}\sum_{g=1}^{G}\Big((\hat{\alpha}_{g,1}(u_1) - x_g'\hat{\beta}(u_1))(\hat{\alpha}_{g,2}(u_2) - x_g'\hat{\beta}(u_2))w_g w_g'\Big),$$

$\hat{S} = (\hat{Q}_{xw}\hat{Q}_{ww}^{-1}\hat{Q}_{xw}')^{-1}\hat{Q}_{xw}\hat{Q}_{ww}^{-1}$, $\hat{Q}_{xw} = X'W/G$, and $\hat{Q}_{ww} = W'W/G$. In the theorem below, we show that $\hat{\mathcal{C}}(u_1, u_2)$ is consistent for $\mathcal{C}(u_1, u_2)$ uniformly over $u_1, u_2 \in \mathcal{U}$.

Theorem 2 (Estimating $\mathcal{C}$)**.** *Let Assumptions 1-8 hold. Then* $\|\hat{\mathcal{C}}(u_1, u_2) - \mathcal{C}(u_1, u_2)\| = o_p(1)$ *uniformly over* $u_1, u_2 \in \mathcal{U}$.

Remark 2. Theorems 1 and 2 can be used for hypothesis testing concerning $\beta(u)$ for a given quantile index $u \in \mathcal{U}$. In particular, we have that

$$\sqrt{G}\hat{\mathcal{C}}(u, u)^{-1/2}(\hat{\beta}(u) - \beta(u)) \Rightarrow N(0, 1). \tag{5}$$

Importantly for applied researchers, Theorems 1 and 2 demonstrate that heteroskedasticity-robust standard errors for our estimator can be obtained by the traditional White (1980) standard errors where we proceed as if $\hat{\alpha}_{g,1}(u)$ were equal to $\alpha_{g,1}(u)$, that is, as if there were no first-stage estimation error. Appendix E extends this result for clustered standard errors. □

Finally, we show how to obtain confidence bands for $\beta(u)$ that hold uniformly over $\mathcal{U}$.[14] Observe that $\beta(u)$ is a d_x-vector, that is, $\beta(u) = (\beta_1(u), \dots, \beta_{d_x}(u))'$. Without loss of generality, we focus on $\beta_1(u)$, the first component of $\beta(u)$. Let $\hat{\beta}_1(u)$, $V(u)$, and $\hat{V}(u)$ denote the first component of $\hat{\beta}(u)$, the $(1,1)$ component of $\mathcal{C}(u,u)$, and the $(1,1)$ component of $\hat{\mathcal{C}}(u,u)$, respectively. Define

$$T = \sup_{u \in \mathcal{U}} \sqrt{G} |\hat{V}(u)^{-1/2}(\hat{\beta}_1(u) - \beta_1(u))|, \tag{6}$$

and let $c_{1-\alpha}$ denote the $(1-\alpha)$ quantile of T. Then uniform confidence bands of level α for $\beta_1(u)$ could be constructed as

$$\left[\hat{\beta}_1(u) - c_{1-\alpha}\sqrt{\frac{\hat{V}(u)}{G}}, \hat{\beta}_1(u) + c_{1-\alpha}\sqrt{\frac{\hat{V}(u)}{G}}\right]. \tag{7}$$

These confidence bands are infeasible, however, because $c_{1-\alpha}$ is unknown. We suggest estimating $c_{1-\alpha}$ by the multiplier bootstrap method. To describe the method, let $\epsilon_1, ..., \epsilon_G$ be an i.i.d. sequence of $N(0,1)$ random variables that are independent of the data. Also, let $\hat{w}^S_{g,1}$ denote the 1st component of the vector $\hat{S}w_g$. Then the multiplier bootstrap statistic is

$$T^{MB} = \sup_{u \in \mathcal{U}} \frac{1}{\sqrt{G\hat{V}(u)}} \sum_{g=1}^{G} \left(\epsilon_g(\hat{\alpha}_{g,1}(u) - x_g'\hat{\beta}(u))\hat{w}^S_{g,1}\right)$$

The multiplier bootstrap critical value $\hat{c}_{1-\alpha}$ is the conditional $(1-\alpha)$ quantile of T^{MB} given the data. Then a feasible version of uniform confidence bands is given by equation (7) with $\hat{c}_{1-\alpha}$ replacing $c_{1-\alpha}$. The validity of the method is established in the following theorem using the results of Chernozhukov, Chetverikov, and Kato (2013).

Theorem 3 (Uniform Confidence Bands via Multiplier Bootstrap)**.** *Let Assumptions 1-8 hold. In addition, suppose that all eigenvalues of $J(u,u)$ are bounded away from zero uniformly over $u \in \mathcal{U}$. Then*

$$P\left(\beta_1(u) \in \left[\hat{\beta}_1(u) - \hat{c}_{1-\alpha}\sqrt{\frac{\hat{V}(u)}{G}}, \hat{\beta}_1(u) + \hat{c}_{1-\alpha}\sqrt{\frac{\hat{V}(u)}{G}}\right] \text{ for all } u \in \mathcal{U}\right) \to 1 - \alpha$$

as $G \to \infty$.

Remark 3. *Uniform* confidence bands are typically larger than the *point-wise* confidence bands based on the result (5). The reason is that uniform confidence bands are constructed so that the *whole* function $\{\beta(u), u \in \mathcal{U}\}$ is contained in the bands with approximately $1-\alpha$ probability whereas point-wise bands are constructed so that for any given $u \in \mathcal{U}$, $\beta(u)$ is contained in the bands with approximately $1-\alpha$ probability. Which confidence bands to use depends on the specific purposes of the researcher. □

[14]In addition, Appendix C presents an approach for uniform inference on the $\{\alpha_{g,1}(u)\}$ in the model (1)–(2). In particular, we construct the confidence bands $[\hat{\alpha}^l_{g,1}(u), \hat{\alpha}^r_{g,1}(u)]$ that cover the true group-specific effects $\alpha_{g,1}$ for all $g = 1, \dots, G$ simultaneously with probability approximately $1-\alpha$.

5. The effect of Chinese import competition on the local wage distribution

5.1. **Background on wage inequality.** Over the past 40 years, wage inequality within the United States has increased drastically.[15] Economists have engaged in heated debates about the primary causes of the rising wage inequality—such as globalization, skill-biased technological change, or the declining real minimum wage—and how the importance of these factors has changed over the years.[16] Recent work in Autor, Dorn, and Hanson (2013) (hereafter ADH) focused on import competition and its effects on wages and employment in US local labor markets. ADH studied the period 1990–2007, when the share of US spending on Chinese imports increased dramatically from 0.6% to 4.6%. For identification, the authors used spatial variation in manufacturing concentration, showing that localized US labor markets which specialize in manufacturing were more affected by increased import competition from China. The authors found that those markets which were more exposed to increased import competition in turn had lower employment and lower wages.

We contribute to this debate by studying the effect of increased trade, in the form of increased import competition, on the distribution of local wages (rather than on the average local wages as in ADH). Given that we exploit the same variation in import competition as in ADH, we first describe the ADH framework below and then present our results.

5.2. **Framework of Autor, Dorn, and Hanson (2013).** To study the effect of Chinese import competition on average domestic wages, ADH used Census microdata to calculate the mean wage within each Commuting Zone (CZ) in the United States.[17] The authors then estimated the following regression:

$$\Delta\overline{\ln w}_g = \beta_1 \Delta IPW_g^U + X_g' \beta_2 + \varepsilon_g \tag{8}$$

where $\Delta\overline{\ln w}_g$ is the change in average individual log weekly wage in a given CZ in a given decade, X_g are characteristics of the CZ and decade, including indicator variables for each decade. Note that we have changed the notation slightly from that in ADH in order to improve clarity for our application—a "group" g in this setting is a given CZ in a given decade. The variable of interest is ΔIPW_g^U, which represents the decadal change in Chinese imports per US worker for the CZ and decade corresponding to group g.[18]

[15] Autor, Katz, and Kearney (2008) documented that, from 1963 to 2005, the change in wages for the 90^{th} percentile earner was 55% higher than for the 10^{th} percentile earner.

[16] See, for example, Leamer (1994), Krugman (2000), Feenstra and Hanson (1999), Katz and Autor (1999), as well as many other papers cited in Feenstra (2010) or in Haskel, Lawrence, Leamer, and Slaughter (2012).

[17] The United States is covered exhaustively by 722 Commuting Zones (Tolbert and Sizer 1996), each roughly corresponding to a local labor market.

[18] Due to data limitations, ADH proxy for the change in actual local imports per worker with the weighted average of industry-level changes in the value of Chinese imports to the US with the weights corresponding to the beginning of decade employment share of each industry in each CZ.

To address endogeneity concerns (i.e. that imports from China may be correlated with unobserved labor demand shocks), the authors instrumented for imports per last-period worker using ΔIPW_g^O, a measure of import exposure that replaces the change in Chinese imports to the US in a given industry with the change in Chinese imports to other similarly developed nations for the same industry and uses one decade lagged employment shares in calculating the weighted average. Using this 2SLS approach, the authors found that a \$1,000 increase in Chinese imports per worker in a CZ decreases average log weekly wage by -0.76 log points, corresponding to decrease in wages for the average CZ of 0.9% from 1990–2000 and 1.4% from 2000–2007. When estimated separately by gender, the effect was more negative for males (-0.89 log points) and less so for females (-0.61 log points).[19]

5.3. **Distributional effects of increased import competition.** We build on the ADH framework to analyze whether low-wage earners were more adversely affected than high-wage earners by Chinese import competition. To apply the grouped IV quantile regression estimator to this setting, we replace $\Delta\overline{\ln w}_g$, the change in the average log weekly wage in equation (8) with $\Delta \ln w_g^u$, the change in the u-quantile of log wages in the CZ and decade corresponding to group g. We calculate these quantiles using micro-level observations from the Census Integrated Public Use Micro Samples for 1990 and 2000 and the American Community Survey for 2006-2008, matching these observations to CZs following the strategy described in ADH.[20] We instrument for ΔIPW_g^U using ΔIPW_g^O as described above. Recall that existing methods for handling endogeneity in quantile models are suited for the case where the individual-level unobserved conditional quantile itself is correlated with the treatment and would be inconsistent in this setting because the endogeneity consists of a group-level treatment being correlated with the group-level unobservable additive term.

Figures 1, 2, and 3 display the results of the grouped IV quantile regression estimator for the full sample, for males only, and for females only. Each figure displays u-quantile estimates for $u \in \{0.05, 0.1, ..., 0.95\}$, along with pointwise 95% confidence bands about each estimate. The figures also display the 2SLS effect found in ADH and 95% confidence intervals corresponding to their IV estimate of Chinese import penetration on the change in CZ-level average wages.

[19] As discussed by ADH, the existence of an extensive-margin labor supply response—imports affecting whether individuals are employed—makes these results likely a lower-bound for the effect on all workers because we don't observe wages for the unemployed population.

[20] The thought experiment behind the asymptotics in this application is that the estimator is consistent as the number of groups ($G = 722$ CZs $\times$ two decades) and the number of individuals within each group ($N_G = 543$, the size of the smallest group) both grow large. We follow ADH by clustering at the state level and weighting by start-of-decade CZ population in the second stage of our estimator. To cluster, we are relying on Appendix E, which relaxes Assumption 1 to allow for observations to be dependent across groups. We also follow the ADH individual weighting procedure in the first stage given that not all individuals can be mapped to a unique CZ.

Each figure provides evidence that Chinese import competition affected the wages of low-wage earners more than high-wage earners, demonstrating how increases in trade can causally exacerbate local income inequality. For all three samples, the magnitude of the estimated causal effect of Chinese import penetration is much larger for lower quantiles of the conditional wage distribution. The point estimates suggest that the average negative effect of Chinese import penetration estimated by ADH is primarily driven by large negative effects for those in the bottom tercile, where the effect is twice as large as the average effect.[21] Wages not in the bottom tercile were less affected than the average—Figure 1 shows that for most wage-earners (from the 0.35 quantile and above) the effect of Chinese import competition was one-third smaller in magnitude than the effect on the average estimated by ADH. Comparing the pattern of the coefficients across two gender subsamples in Figures 2 and 3, there is more distributional heterogeneity for females than males, a finding that additional testing shows is even more pronounced for non-college educated females. For each sample, we can reject an effect size of zero for almost all quantiles below the median but cannot for all quantiles above the median.

[21] A coefficient of -1.4 log points, e.g. for the lower quantiles of Figure 1, corresponds to a 2.6% decrease in wages from 2000–2007 for the average commuting zone's change in Chinese import exposure.

References

Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, **70**, 91–117.

Abrevaya, J. and Dahl, C. (2008). The effects of birth inputs on birthweight. *Journal of Business and Economic Statistics*, **26**, 379–397.

Altonji, J. and Matzkin, R. (2005). Cross section and panel data estimators for non separable models with endogenous regressors. *Econometrica*, **73**, 1053–1102.

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

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

Arellano, M. and Bonhomme, S. (2013). Random Effects Quantile Regression. Working paper.

Autor, D., Dorn, D., and Hanson, G. (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, **103(6)**, 2121–2168.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in U.S. wage inequality: Revising the revisionists. *Review of Economics and Statistics*, **90**, 300–323.

Backus, M. (2014). Why is productivity correlated with competition? Working paper.

Baum-Snow, N. (2007). Did highways cause suburbanization? *Quarterly Journal of Economics*, **122**, 775-805.

Belloni, A., Chernozhukov, V., and Hansen, C. (2006). Conditional quantile processes based on series or many regressors. Working paper.

Canay, I. (2011). A simple approach to quantile regression for panel data. *Econometrics Journal*, **14**, 368–386.

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. *Advances in Econometrics, Sixth World Congress*, 171–209.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics*, **33**, 806–839.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, **41**, 2786–2819.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014a). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, **42**, 1564-1597.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014b). Anti-concentration and honest adaptive confidence bands. *The Annals of Statistics, forthcoming.*

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014c). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields, forthcoming.* Available at arXiv:1301.4807v3.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014d). Central limit theorems and bootstrap in high dimensions. Available at arXiv:1412.3661v2.

Chernozhukov, V. and Fernández-Val, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *The Review of Economic Studies*.

Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica*, **73**, 245–261.

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

Chernozhukov, V. and Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, **142**, 379–398.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, **71**, 1405–1441.

Feenstra, R. C. (2010). Offshoring in the global economy: Microeconomic structure and macroeconomic implications. Cambridge, MA: MIT Press.

Feenstra, R. C. and Hanson, G. H. (1999). The Impact of outsourcing and high-technology capital on wages: Estimates for the U.S., 1979–1990. *Quarterly Journal of Economics*, **114(3)**, 907–940.

Florens, J. P., Heckman, J., Meghir, C., and Vytlacil, E. (2008). Identification of treatment effects using control functions in models with continuous endogenous treatment and heterogeneous effects. *Econometrica*, **76**, 1191–1206.

Galvao, A. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, **164**, 142–157.

Galvao, A. and Wang, L. (2013). Efficient minimum distance estimator for quantile regression fixed effects panel data. Working paper.

Graham, B. and Powell, J. (2012). Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica*, **80**, 2105–2152.

Hahn, J. and Meinecke, J. (2005). Time-invariant regressor in nonlinear panel model with fixed effects. *Econometric Theory*, **21**, 455–469.

Haskel, J., Lawrence, R., Leamer, E. E., and Slaughter, M. J. (2012). Globalization and U.S. wages: Modifying classic theory to explain recent facts. *Journal of Economic Perspectives*, **26(2)**: 119–40.

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspectives*, **15**, 57–67.

Hausman, J. and Taylor, W. (1981). Panel data and unobservable individual effects. *Econometrica*, **49**, 1377–1398.

Hausman, J., Luo, Y., and Palmer, C. (2014). Errors in the dependent variable of quantile regression models. Working paper.

Heckman, J. and Vytlacil, E. (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *The Journal of Human Resources*, **33**, 974-987.

Imbens, G. and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62**, 467-475.

Imbens, G. and Newey, W. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, **77**, 1481–1512.

Kato, K. and Galvao, A. (2011). Smoothed quantile regression for panel data. Working paper.

Kato, K., Galvao, A., and Montes-Rojas, G. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, **170**, 76–91.

Katz, L. F. and Autor, D. (1999). Changes in the wage structure and earnings inequality, in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam: Elsevier Science, 1463–1555.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74–89.

Koenker, R. (2005). *Quantile regression.* Econometric Society Monographs.

Koenker, R. and Bassett, J. (1978). Regression quantiles. *Econometrica*, **46**, 33–50.

Krugman, P. (2000). Technology, trade and factor prices. *Journal of International Economics* , **50(1)**, 51–71.

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, **157**, 396–408.

Larsen, B. (2014). Occupational licensing and quality: Distributional and heterogeneous effects in the teaching profession. Working paper.

Leamer, E. E. (1994). Trade, wages and revolving door ideas. NBER working paper no. 4716.

Lee, S. (2007). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics*, **141**, 1131–1158.

Lehmann, E. and Romano, J. (2005). *Testing statistical hypotheses.* Springer Texts in Statistics.

Masten, M. and Torgovitsky, A. (2014). Instrumental variables estimation of a generalized correlated random coefficients model. Working paper.

Palmer, C. (2011). Suburbanization and urban decline. Working paper.

Ponomareva, M. (2011). Identification in quantile regression panel data models with fixed effects and small T. Working paper.

Powell, J. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, **25**, 303-325.

Rosen, A. (2012). Set identification via quantile restrictions in short panels. *Journal of Econometrics*, **166**, 127–137.

Tao, T. (2012). *Topics in random matrix theory.* American Mathematica Society.

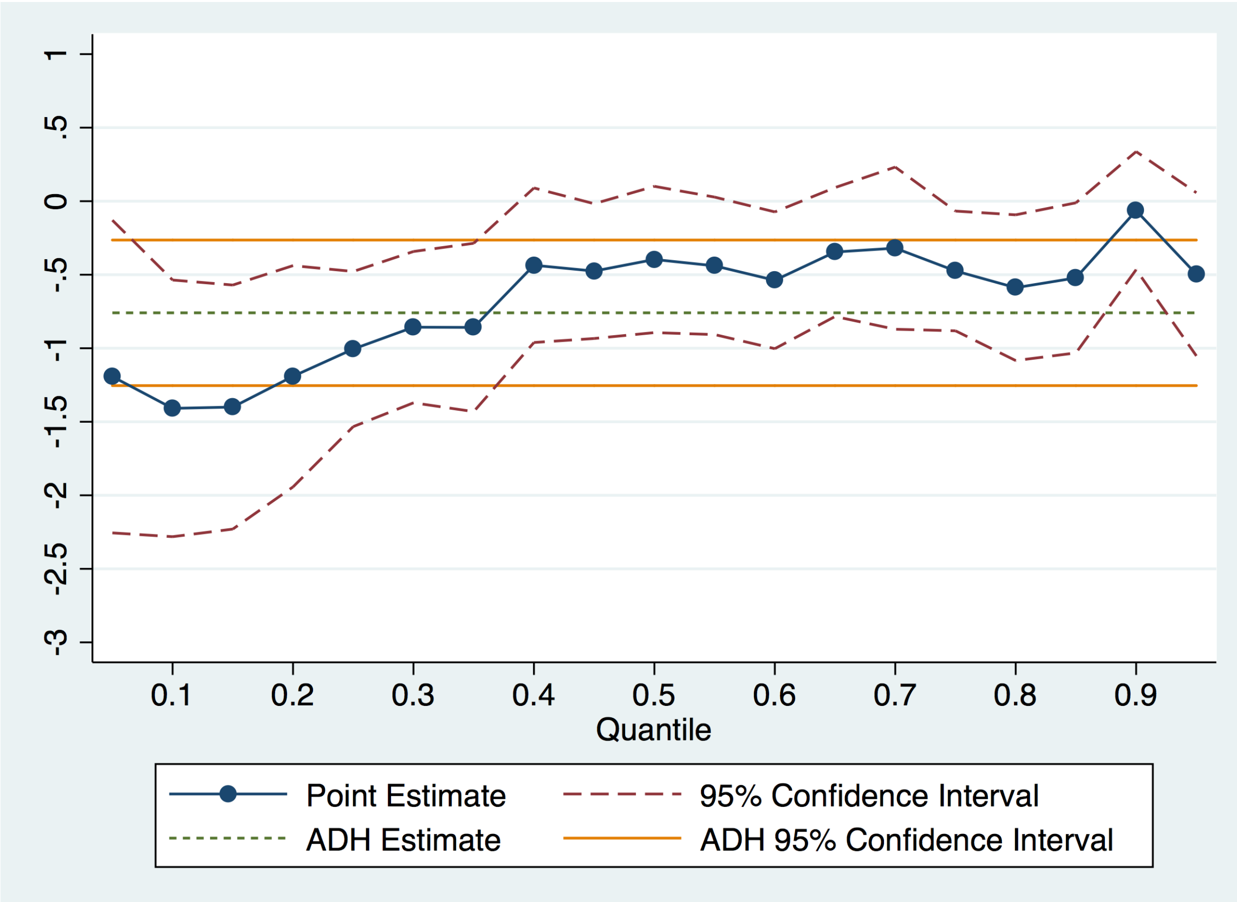
Tolbert, C. M. and Sizer, M. (1996). U.S. commuting zones and labor market areas. A 1990 update. Economic Research Service Staff Paper No. 9614.

Van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press.

Van der Vaart, A. and Wellner, J. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics.

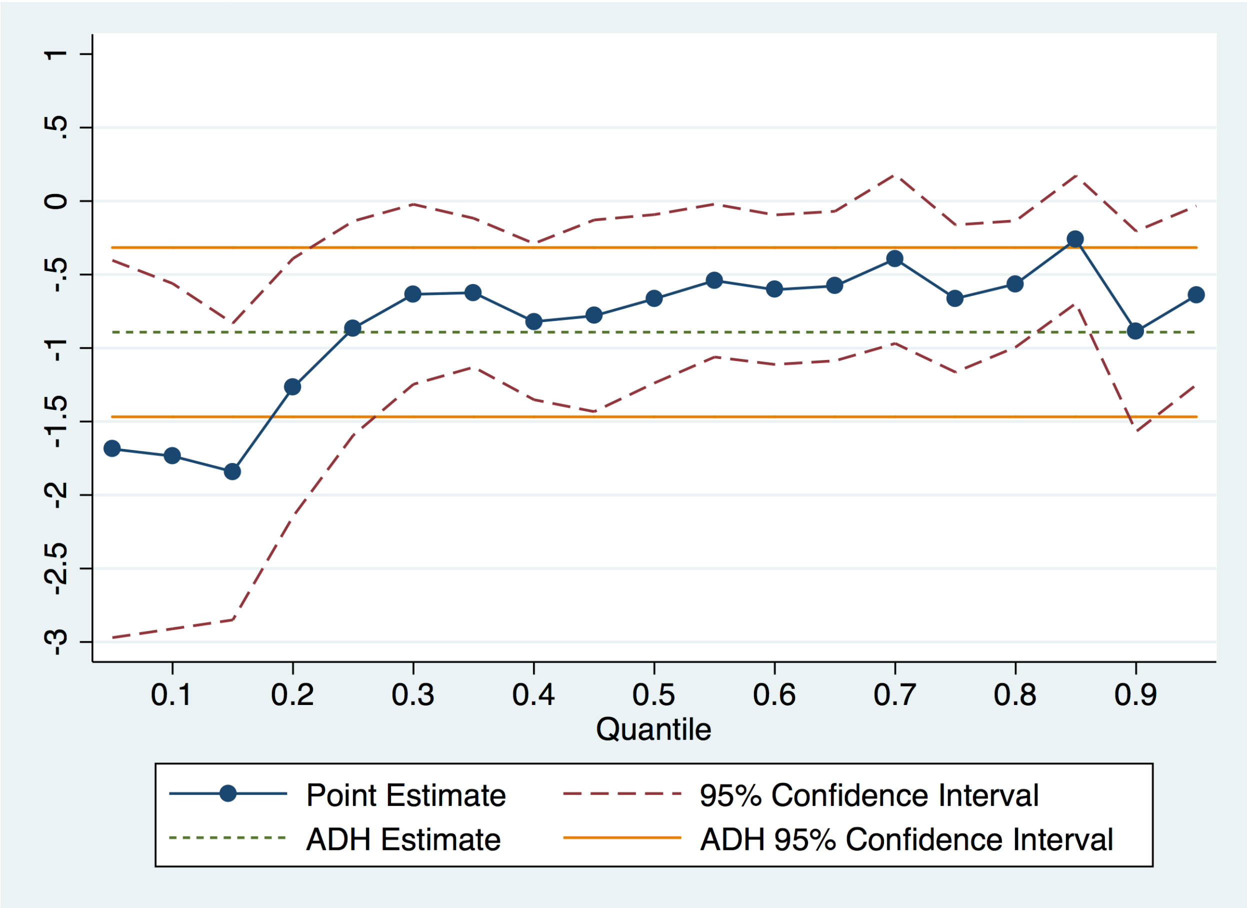
White, H. (1980). A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

Figure 1. Effect of Chinese Import Competition on Conditional Wage Distribution: Full Sample



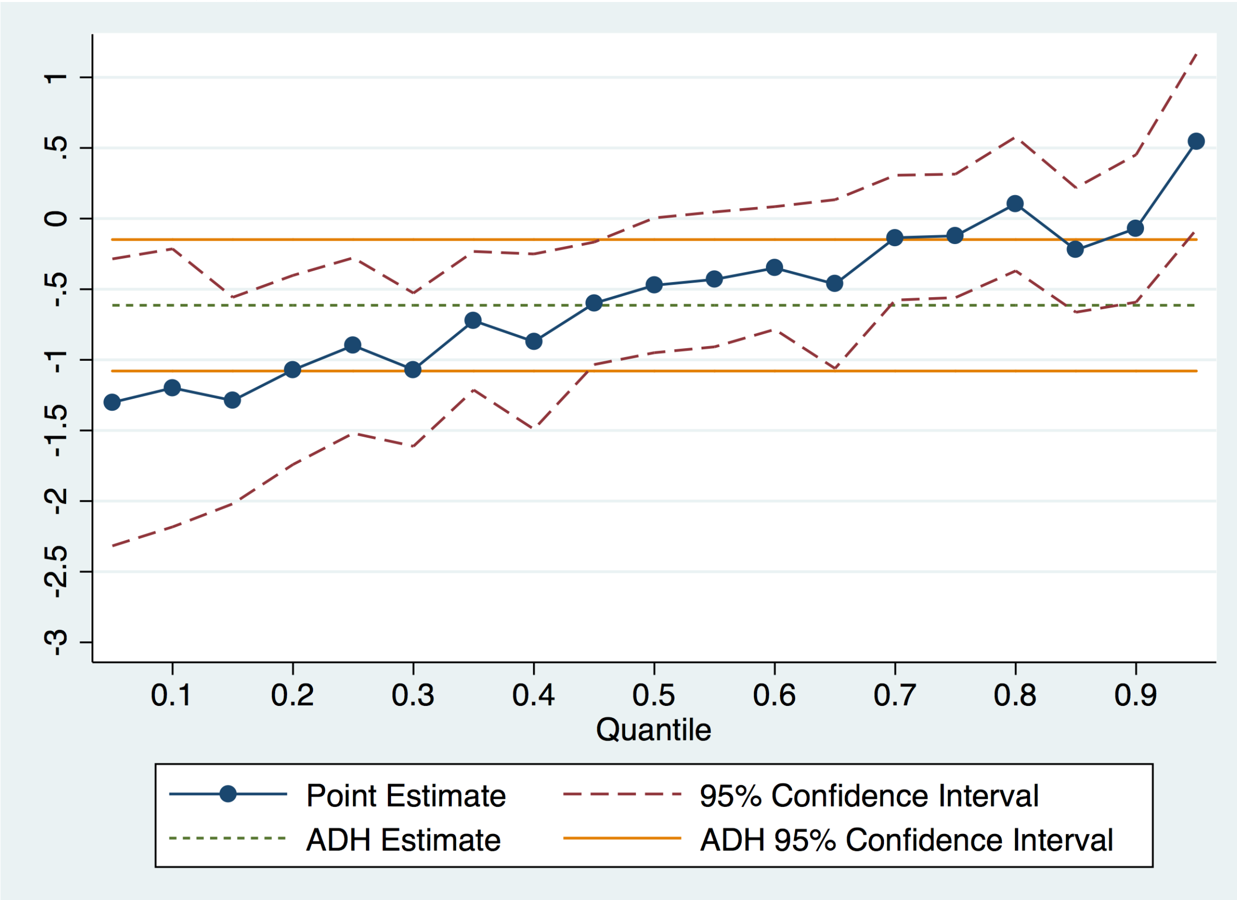
Notes: Figure plots grouped IV quantile regression estimates of the effect of a \$1,000 increase in Chinese imports per worker on the conditional wage distribution (β_1 in equation (8) in the text when the change in average log wages for the commuting zone and decade corresponding to group g, $\Delta \overline{\ln w}_g$, is replaced with the change in the u-quantile of log wages $\Delta \ln w_g^u$). The dashed horizontal line is the ADH estimate of β_1 in equation (8). 95% pointwise confidence intervals are constructed from robust standard errors clustered by state and observations are weighted by CZ population, as in ADH. Units on the vertical axis are log points.

FIGURE 2. Effect of Chinese Import Competition on Conditional Wage Distribution: Males Only



Notes: Figure plots grouped IV quantile regression estimates for the male-only sample of the effect of a \$1,000 increase in Chinese imports per worker on the male conditional wage distribution (β_1 in equation (8) in the text when the change in average log wages for the commuting zone and decade corresponding to group g, $\Delta\overline{\ln w}_g$, is replaced with the change in the u-quantile of log wages $\Delta \ln w_g^u$). The dashed horizontal line is the ADH estimate of β_1 in equation (8). 95% pointwise confidence intervals are constructed from robust standard errors clustered by state and observations are weighted by CZ population, as in ADH. Units on the vertical axis are log points.

Figure 3. Effect of Chinese Import Competition on Conditional Wage Distribution: Females Only



Notes: Figure plots grouped IV quantile regression estimates for the female-only sample of the effect of a \$1,000 increase in Chinese imports per worker on the female conditional wage distribution (β_1 in equation (8) in the text when the change in average log wages for the commuting zone and decade corresponding to group g, $\Delta\overline{\ln w_g}$, is replaced with the change in the u-quantile of log wages $\Delta \ln w_g^u$). The dashed horizontal line is the ADH estimate of β_1 in equation (8). 95% pointwise confidence intervals are constructed from robust standard errors clustered by state and observations are weighted by CZ population, as in ADH. Units on the vertical axis are log points.

Appendix A. Examples of Grouped IV Quantile Regression

To help the reader envision applications of our estimator, in this section, we provide several motivating examples of settings for which our estimator may be useful. Example 2 also provides additional discussion of computational advantages of our estimator. Note that each of the following examples involves estimation of a treatment effect that varies at the group level with all endogeneity concerns also existing only at the group level.[22]

Example 1: Peer Effects of School Integration. Angrist and Lang (2004) studied how suburban student test scores were affected by the reassignment of participating urban students to suburban schools through Boston's Metco program. Before estimating their main instrumental variables model, the authors tested for a relationship between the presence of urban students in the classroom and the second decile of student test scores by estimating

$$Q_{y_{igjt}|m_{gjt},s_{gjt},\xi_{gjt},\alpha_g,\beta_j,\gamma_t}(0.2) = \alpha_g + \beta_j + \gamma_t + \delta m_{gjt} + \lambda s_{gjt} + \xi_{gjt} \tag{9}$$

where the left-hand-side represents the second decile of student test scores within a group, where each group is a grade g × school j × year t cell. The variables s_{gjt} and m_{gjt} denote the class size and the fraction of Metco students within each $g \times j \times t$ cell, and α_g, β_j, and γ_t represent grade, school, and year effects. The unobserved component, ξ_{gjt}, is analogous to the $\varepsilon_g(0.2)$ of the special form (3) of our model (1)–(2).

Angrist and Lang (2004) estimated equation (9) by OLS, which is equivalent to the non-IV application of our estimator with no micro-level covariates. Similar to their OLS results on average test scores, they found that classrooms with higher proportions of urban students have lower second decile test scores. Once they instrumented for a classroom's level of Metco exposure, the authors found no effect on *average* test scores. However, by not estimating model (9) by 2SLS, they were unable to address the causal *distributional* effects of Metco exposure.

In estimating (9), Angrist and Lang (2004) used heteroskedasticity-robust standard errors, which we demonstrate in Section 4 is valid. The extension in Appendix E implies that the authors could have instead allowed for clustering across groups in computing standard errors (for example, clustering at the school level).

Example 2: Occupational Licensing and Quality. Larsen (2014) applied the estimator developed in this paper to study the effects of occupational licensing laws on the distribution of quality within the teaching profession. This application uses a difference-in-differences approach. Similar to Example 1, the explanatory variable of interest is treated as exogenous and the researcher is concerned that there may be unobserved group-level disturbances. In this application, a group is

[22]This is in contrast to settings where the endogeneity exists at the individual level, i.e. when the individual unobserved heterogeneity is correlated with treatment. Such situations require a different approach than the one presented here, e.g. Chernozhukov and Hansen (2005), Abadie, Angrist, and Imbens (2002), or other approaches referenced in Section 1.

a state-year combination (s,t), and micro-level data consists of teachers within a particular state in a given year.

Let q_{ist} represent the quality of teacher i in state s who began teaching in survey year t, where quality is proxied for by a continuous measure of the selectivity of the teacher's undergraduate institution. The conditional uth quantile of quality is modeled as

$$Q_{q_{ist}|Law_{st},\varepsilon_{st}}(u) = \gamma_s(u) + \lambda_t(u) + Law'_{st}\delta(u) + \varepsilon_{st}(u) \tag{10}$$

where γ_s is a state effect; λ_t is a year effect; Law_{st} is a three-element vector containing dummies equal to one if a subject test, basic skills test, or professional knowledge test was required in state s in year t; and $\varepsilon_{st}(u)$ represents group-level unobservables.

Because no micro-level covariates are included, the first stage of the grouped quantile estimator is obtained by simply selecting the uth quantile of quality in a given state-year cell. The second stage is obtained via OLS, and the author used heteroskedasticity-robust standard errors, which are valid by the results in Section 4. Using the grouped quantile estimator, Larsen (2014) found that, for first-year teachers, occupational licensing laws requiring teachers to pass a subject test lead to a small but significant decrease in the upper tail of quality, suggestive that these laws may drive some highly qualified candidates from the occupation.

In this setting, if micro-level covariates, z_{ist}, were included in the first stage of estimation, the researcher could also estimate interaction effects of the group-level treatment and a micro-level covariate, such as the percent of minority students at the teacher's school. This would be done by 1) estimating quantile regression of q_{ist} on z_{ist} (which includes the percent of minority students measure) separately for each (s,t) group and 2) saving each group-level estimate for the coefficient corresponding to the percent minority variable, and then estimating a linear regression of these coefficients on Law_{st} and on the state and year fixed effects.

This example highlights another useful feature of grouped IV quantile regression. Including many variables in a standard quantile regression drastically increases the computational time (see Koenker (2004), Lamarche (2010), Galvao and Wang (2013), and Galvao (2011) for further discussion) and, in our experience, can often lead standard optimization packages to fail to converge. The grouped quantile approach, on the other hand, can handle large numbers of variables easily when these variables happen to be constant within group, as in the case of state and year fixed effects in this example, because the coefficients corresponding to these variables can be estimated in the second-stage linear model, greatly reducing the number of parameters to be estimated in the nonlinear first stage and hence reducing the computational burden significantly. Furthermore, this specific computational advantage of the grouped quantile regression estimator exists even in cases where both standard quantile regression and the grouped approach are valid (i.e. when no group-level unobservables are present). Larsen (2014) found that the grouped approach was significantly faster than estimating parameters in a single quantile regression.

Example 3: Distributional Effects of Suburbanization. Palmer (2011) applied the grouped quantile estimator to study the effects of suburbanization on resident outcomes. This application illustrates the use of our estimator in an IV setting. In this application, a group is a metropolitan statistical area (MSA), and individuals are MSA residents. As an identification strategy, Palmer (2011) used the results of Baum-Snow (2007) in instrumenting suburbanization with planned highways.[23]

The model is

$$\Delta Q_{y_{igt}|x_g,s_g,\varepsilon_g}(u) = \beta(u) \cdot \text{suburbanization}_g + x_g'\gamma_1(u) + \varepsilon_g(u)$$
$$\text{suburbanization}_g = \pi(u) \cdot \text{planned highway rays}_g + x_g'\gamma_2(u) + v_g(u)$$

where $\Delta Q_{y|x_g,s_g,\varepsilon_g}(u)$ is the change in the uth quantile of log wages y_{igt} within an MSA between 1950 and 1990 and x_g is a vector of controls (including a constant) conditional upon which planned highway rays$_g$ is uncorrelated with $\varepsilon_g(u)$ and $v_g(u)$. The variable suburbanization$_g$ is a proxy measure of population decentralization, such as the amount of decline of central city population density. $\beta(u)$ is the coefficient of interest, capturing the effect of suburbanization on the within-MSA conditional wage distribution. For example, if the process of suburbanization had particularly acute effects on the prospects of low-wage workers, we may expect $\beta(u)$ to be negative for $u = 0.1$. For a given u, the grouped IV quantile approach estimates $\beta(u)$ through a 2SLS regression.

Example 4: The Relationship Between Productivity and Competition. Backus (2014) studied the relationship between competition and productivity in the ready-mix concrete industry. The author discussed the fact that competition and productivity are positively correlated, and studied whether this relationship is similar for firms of all productivity levels (e.g. through encouraging better monitoring of firm managers or better investments), or whether increased competition primarily affects the lower tail of the productivity distribution (driving out less productive firms).

Let ρ_{imt} represent a measure of productivity of firm i in market m and time period t. Using our notation, define a group as a pair $m \times t$. The author assumes that ρ_{imt} satisfies the following quantile regression model:

$$Q_{\rho_{mt}|c_{mt},n_{mt},\varepsilon_{mt}}(u) = \beta_t(u) + c_{mt}\beta_c(u) + g(n_{mt}, u) + \varepsilon_{mt}(u) \tag{11}$$

where c_{mt} is a group-level measure of competition, n_{mt} is the number of firms in the group, $g(n_{mt}, u)$ is the third order polynomial of n_{mt}), and ε_{mt} is an unobserved group-level disturbance, which is possibly correlated with c_{mt}.

Backus (2014) instrumented for c_{mt} using group-level measures which shift the demand for concrete. Thus, the IV regression in (11) represents an application of our estimator when group-level

[23] Baum-Snow (2007) instrumented for actual constructed highways with planned highways and estimated that each highway ray emanating out of a city caused an 18% decline in central-city population.

shocks are endogenous and no micro-level covariates are present. The author found some evidence that the effect of competition on the left tail of the productivity distribution may be more positive than at some quantiles in the middle of the distribution (consistent with selection of low-productivity firms out of the industry), but was unable to reject the hypothesis of a constant effect.

Appendix B. Simulations

In order to investigate the properties of our estimator and compare to traditional quantile regression, we generate data according to the following model:

$$y_{ig} = z_{ig}\gamma(u_{ig}) + \delta(u) + x_g\beta(u_{ig}) + \varepsilon_g(u_{ig}) \tag{12}$$

$$x_g = \pi w_g + \eta_g + \nu_g \tag{13}$$

$$\varepsilon_g(u) = u\eta_g - \frac{u}{2} \tag{14}$$

where w_g, ν_g, and z_{ig} are each distributed $\exp(0.25^* N[0,1])$; u_{ig} and η_g are both distributed $U[0,1]$; and random variables w_g, ν_g, z_{ig}, u_{ig}, and η_g are mutually independent. Note that the form $\varepsilon_g(u) = u\eta_g - \frac{u}{2}$ implies $E[\varepsilon_g(u)|w_g] = E[u\eta_g - u/2|w_g] = E[u\eta_g - u/2] = u/2 - u/2 = 0$. The quantile coefficient functions are $\gamma(u) = \beta(u) = u^{1/2}$ and $\delta(u) = u/2$. The parameter $\pi = 1$.

We employ three variants of the data generating process described in (12)–(14). The first case is exactly as in (12)–(14), with the group-level treatment of interest, x_g, being endogenous (correlated with ε_g through η_g). We estimate $\beta(u)$ in this case using the grouped IV quantile estimator as well as standard quantile regression (which ignores the endogeneity as well as the existence of ε_g). In the second case x_g is exogenous, where we set $x_g = w_g$ in (13). We estimate $\beta(u)$ again in this case using the grouped quantile approach as well as standard quantile regression, where the latter ignores the existence of ε_g. In the third case x_g is exogenous *and* no group-level unobservables are included, where we set $x_g = w_g$ and $\varepsilon_g = 0$. In this latter case, both grouped quantile regression and standard quantile regression should be consistent.

We perform these exercises with the number of groups (G) and the number of observations per group (N) given by (N, G) =(25,25), (200,25), (25,200), (200,200). 1,000 Monte Carlo replications were used. The results are displayed in Table I. Each panel displays the bias from the procedure for each decile ($u = 0.1, ..., 0.9$) as well as the average absolute value of that bias, averaged over the nine deciles.

The top panel of Table I demonstrates that in the endogenous group-level treatment case the magnitude of the bias is much smaller in our estimator than in standard quantile regression, and the bias of our estimator disappears as N and G increase, while the bias of quantile regression remains constant (0.196 on average). The middle panel considers the case where x_g is exogenous but group-level unobservables are present (or, equivalently, left-hand-side measurement error exists in the quantile regression). At some quantiles, standard quantile regression has a bias which is smaller in magnitude than the grouped approach, in particular in the cases where $N = 25$. However, as N increases, the magnitude of the bias of the grouped estimator falls close to zero on average while that of standard quantile regression remains about three times as high at 0.01. Finally, the bottom panel focuses on the case in which no group-level unobservables exist and hence standard quantile regression is unbiased. In this case, we find that the bias of standard quantile regression is indeed lower than that of the grouped quantile approach, but the bias of the grouped quantile method also diminishes rapidly as N and G grow.

To illustrate the computational burden which our estimator overcomes, we redid the first stage estimation with $\gamma(\cdot)$ and group-level fixed effects—α_g from Section 2—estimated jointly in one large quantile regression rather than estimating group-by-group quantile regression. We performed 100 replications due to the computational burden of the joint estimation. We found that in the $(N, G) = (25, 25)$ case the joint estimation took only slightly longer than than the group-by-group approach; with $(N, G) = (200, 25)$ the group-by-group approach was ten times faster; with $(N, G) = (25, 200)$ the group-by-group approach was over forty times as fast; and in the $(N, G) = (200, 200)$ the group-by-group approach was over 150 times as fast, with estimation on a single replication sample for the nine deciles taking over three minutes, while the the grouped quantile approach performed the same exercise in 1.22 seconds.[24] This exercise illustrates the benefit of the group-by-group approach to estimating α_g and also illustrates that, in general, standard quantile regression can be very slow when a large number of explanatory variable is included. The grouped quantile approach can greatly reduce this computational burden by handling all group-level explanatory variables linearly in the second stage (implying that the grouped quantile approach can be especially beneficial if the dimension of x_g is large).

Appendix C. Joint Inference on Group-Specific Effects

In this section, we are concerned with inference on group-specific effects $\alpha_{g,1}(u)$, $g = 1, \dots, G$, in the model (1)-(2) defined in Section 2. In particular, we are interested in constructing the confidence bands $[\hat{\alpha}^l_{g,1}, \hat{\alpha}^r_{g,1}]$ for $\alpha_{g,1}(u)$ that are adjusted for multiplicity of the effects, that is, we would like to have the bands satisfying

$$P(\alpha_{g,1}(u) \in [\hat{\alpha}^l_{g,1}, \hat{\alpha}^r_{g,1}] \text{ for all } g = 1, \dots, G) \to 1 - \alpha. \tag{15}$$

Thus, the confidence bands $[\hat{\alpha}^l_{g,1}, \hat{\alpha}^r_{g,1}]$ cover the true group-specific effects $\alpha_{g,1}$ for all $g = 1, \dots, G$ simultaneously with probability approximately $1 - \alpha$.

The main challenge here is that we have G parameters $\alpha_{g,1}(u)$, $g = 1, \dots, G$, and only N_g observations to estimate $\alpha_{g,1}$ where N_g is potentially smaller than G (recall that we impose Assumption

[24]With $G > 200$, the computation time ratio drastically increases further, with standard optimization packages often failing to converge appropriately.

3, according to which $G^{2/3}(\log N_G)/N_G \to 0$ as $G \to \infty$ where $N_G = \min_{g=1,\dots,G} N_g$). To decrease technicalities, in this section we assume that $\mathcal{U} = \{u\}$, that is, $\mathcal{U}$ is a singleton.

It is well-known that as $N_g \to \infty$, $N_g^{1/2}(\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)) \Rightarrow N(0, I_g)$ where I_g is the $(1,1)$th element of the matrix $u(1-u)J_g(u)^{-1}E_g[z_{ig}z_{ig}']J_g(u)^{-1}$; see, for example, Koenker (2005). Therefore, letting $c_{1-\alpha}$ be the $(1-\alpha)$ quantile of $|Y|$ where $Y \sim N(0,1)$, we obtain

$$P\left(\alpha_{g,1}(u) \in \left[\hat{\alpha}_{g,1}(u) - c_{1-\alpha}\sqrt{\frac{I_g}{N_g}}, \hat{\alpha}_{g,1}(u) + c_{1-\alpha}\sqrt{\frac{I_g}{N_g}}\right]\right) \to 1-\alpha \text{ as } N_g \to \infty. \tag{16}$$

In practice, I_g is typically unknown, however, and has to be estimated from the data. For example, one can use a method developed in Powell (1984). Letting $\hat{I}_g$ denote a suitable estimator of I_g, it is standard to show that (16) continues to hold if we replace I_g with $\hat{I}_g$ as long as $\hat{I}_g \to_p I_g$.

The drawback of the confidence bands in (16), however, is that they do not take into account multiplicity of the effects $\alpha_{g,1}(u)$, $g = 1,\dots,G$. This is especially important given that G is large. To fix this problem, we would like to adjust the constant $c_{1-\alpha}$ in (16) so that the events under the probability sign in (16) hold simultaneously for all $g = 1,\dots,G$ with probability asymptotically equal to $1-\alpha$. The theorem below shows that this can be achieved by replacing $c_{1-\alpha}$ with $c_{1-\alpha}^M$, the $(1-\alpha)$ quantile of $\max_{1\le g\le G}|Y_g|$ where $Y_1,\dots,Y_G$ are i.i.d. $N(0,1)$ random variables. To decrease technicalities, we assume in the theorem that all I_g's are known.

Theorem 4 (Joint Inference on Group-Specific Effects). *Let Assumptions 1-8 hold. In addition, suppose that $I_g \ge c_M$ for all $g = 1,\dots,G$ and $\bar{N}_G/N_G \le C_M$ where $N_G = \min_{1\le g\le G} N_g$ and $\bar{N}_G = \max_{1\le g\le G} N_g$. Let $c_{1-\alpha}^M$ be the $(1-\alpha)$ quantile of $\max_{1\le g\le G}|Y_g|$ where $Y_1,\dots,Y_G$ are i.i.d. $N(0,1)$ random variables. Then*

$$P\left(\alpha_{g,1}(u) \in \left[\hat{\alpha}_{g,1}(u) - c_{1-\alpha}^M\sqrt{\frac{I_g}{N_g}}, \hat{\alpha}_{g,1}(u) + c_{1-\alpha}^M\sqrt{\frac{I_g}{N_g}}\right] \text{ for all } g = 1,\dots,G\right) \to 1-\alpha$$

as $G \to \infty$.

Appendix D. Sub-gaussian Tail Bound

In this section, we derive the sub-gaussian tail bound for the quantile regression estimator. This bound plays an important role in deriving the asymptotic distribution of our estimator, which is given in Theorem 1.

Theorem 5 (Sub-Gaussian Tail Bound for Quantile Estimator). *Let Assumptions 1-8 hold. Then there exist constants $\bar{c}, c, C > 0$ that depend only on c_M, c_f, C_M, C_f, C_L such that for all $g = 1,...,G$ and $x \in (0,\bar{c})$,*

$$P\left(\sup_{u\in\mathcal{U}}\|\hat{\alpha}_g(u) - \alpha_g(u)\| > x\right) \le Ce^{-cx^2N_g}. \tag{17}$$

Remark 4. The bound provided in Theorem 5 is *non-asymptotic*. In principle, it is also possible to calculate the exact constants in the inequality (17). We do not calculate these constants because they are not needed for our results. Since $\hat{\alpha}_{g,1}(u)$ is the classical Koenker and Bassett's (1978) quantile regression estimator of $\alpha_g(u)$, Theorem 5 may also be of independent interest. The theorem implies that large deviations of the quantile estimator from the true value are extremely unlikely under our conditions. $\square$

Appendix E. Clustered Standard Errors

In this section, we consider the model from the main text, which is defined in equations (1)–(2), but we seek to relax the independence across groups condition appearing in Assumption 1(i). In particular, in this section we allow for cluster sampling and derive the results that are analogous to Theorems 1 - 3 in the main text.

We assume that the data consist of $M = M_G$ clusters of groups, and that there exists a correspondence $\mathbb{C}_G : \{1, \ldots, M\} \rightrightarrows \{1, \ldots, G\}$ such that (i) for each $m = 1, \ldots, M$, $\mathbb{C}_G(m)$ denotes the set of groups corresponding to cluster m, (ii) for $m, m' = 1, \ldots, M$ with $m \neq m'$, the set $\mathbb{C}_G(m) \cap \mathbb{C}_G(m')$ is empty, and (iii) for any $g = 1, \ldots, G$, there exists $m = 1, \ldots, M$ such that $g \in \mathbb{C}_G(m)$. Thus, the correspondence $\mathbb{C}_G(\cdot)$ partitions groups into M clusters. Using this notation, we replace Assumption 1 with the following condition:

A1$'$ (Design). *(i) Observations are independent across clusters $m = 1, \ldots, M$. (ii) For all $g = 1, \ldots, G$, the pairs (z_{ig}, y_{ig}) are i.i.d. across $i = 1, \ldots, N_g$ conditional on (x_g, α_g). (iii) For each $m = 1, \ldots, M$, the number of elements in the set $\mathbb{C}_G(m)$ is bounded from above by some constant $\bar{C}$, which is independent of G.*

Assumption 1$'$(i) relaxes Assumption 1(i) from the main text by requiring independence across clusters instead of independence across groups. Assumption 1$'$(ii) is the same as Assumption 1(ii). Assumption 1$'$(iii) imposes the condition that the number of groups within each cluster remains small as the number of groups gets large.

In addition, we replace Assumption 6 with the following condition:

A6$'$ (Noise). *(i) For all $g = 1, \ldots, G$, $E[\sup_{u \in \mathcal{U}} |\varepsilon_g(u)|^{4+c_M}] \leq C_M$. (ii) For some (matrix-valued) function $J^{CS} : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^{d_w \times d_w}$,*

$$\frac{1}{G} \sum_{m=1}^{M} E\left[\left(\sum_{g \in \mathbb{C}_G(m)} \varepsilon_g(u_1) w_g \right) \left(\sum_{g \in \mathbb{C}_G(m)} \varepsilon_g(u_1) w_g' \right) \right] \to J^{CS}(u_1, u_2)$$

uniformly over $u_1, u_2 \in \mathcal{U}$. (iii) For all $u_1, u_2 \in \mathcal{U}$, $|\varepsilon_g(u_2) - \varepsilon_g(u_1)| \leq C_L |u_2 - u_1|$.

Assumptions 6$'$(i) and 6$'$(iii) are the same as Assumptions 6(i) and 6(iii). Assumption 6$'$(ii) is a modification of Assumption 6(ii) adjusting the asymptotic covariance function of $G^{-1/2} \sum_{g=1}^{G} \varepsilon_g(\cdot) w_g$

to allow for clustering. When $\mathbb{C}_G(m)$ contains only one group for each $m = 1, \ldots, M$, Assumption $6'$(ii) reduces to Assumption 6(ii).

Like in the classical cross-section cluster sampling setup, allowing for clustering in our model does not require adjusting the estimator. Therefore, we study the properties of the estimator $\hat{\beta}(u)$ of parameter $\beta(u)$, $u \in \mathcal{U}$, defined in Section 3. Our first theorem in this section describes the asymptotic distribution of $\hat{\beta}(u)$.

Theorem 6 (Asymptotic Distribution under Cluster Sampling)**.** *Let Assumptions 1′, 2-5, 6′, 7, and 8 hold. Then*

$$\sqrt{G}(\hat{\beta}(\cdot) - \beta(\cdot)) \Rightarrow \mathbb{G}^{CS}(\cdot), \text{ in } \ell^\infty(\mathcal{U})$$

where $\mathbb{G}^{CS}(\cdot)$ is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function $\mathcal{C}^{CS}(u_1, u_2) = SJ^{CS}(u_1, u_2)S'$ where $S = (Q_{xw}Q_{ww}^{-1}Q_{xw}')^{-1}Q_{xw}Q_{ww}^{-1}$, Q_{xw} and Q_{ww} appear in Assumption 2, and $J^{CS}(u_1, u_2)$ in Assumption 6′.

Next, we discuss how to estimate the covariance function $\mathcal{C}^{CS}(\cdot,\cdot)$ of the limiting Gaussian process $\mathbb{G}^{CS}(\cdot)$. We suggest estimating $\mathcal{C}^{CS}(\cdot,\cdot)$ by $\hat{\mathcal{C}}^{CS}(\cdot,\cdot)$ defined for all $u_1, u_2 \in \mathcal{U}$ as

$$\hat{\mathcal{C}}^{CS}(u_1, u_2) = \hat{S}\hat{J}^{CS}(u_1, u_2)\hat{S}'$$

where

$$\hat{J}^{CS}(u_1, u_2) = \frac{1}{G}\sum_{m=1}^{M}\left(\sum_{g\in\mathbb{C}_G(m)}(\hat{\alpha}_{g,1}(u_1) - x_g'\hat{\beta}(u_1))w_g\right)\left(\sum_{g\in\mathbb{C}_G(m)}(\hat{\alpha}_{g,2}(u_2) - x_g'\hat{\beta}(u_2))w_g'\right),$$

$\hat{S} = (\hat{Q}_{xw}\hat{Q}_{ww}^{-1}\hat{Q}_{xw}')^{-1}\hat{Q}_{xw}\hat{Q}_{ww}^{-1}$, $\hat{Q}_{xw} = X'W/G$, $\hat{Q}_{ww} = W'W/G$. In the theorem below, we show that $\hat{\mathcal{C}}^{CS}(u_1, u_2)$ is consistent for $\mathcal{C}^{CS}(u_1, u_2)$ uniformly over $u_1, u_2 \in \mathcal{U}$.

Theorem 7 (Estimating $\mathcal{C}^{CS}$ under Cluster Sampling)**.** *Let Assumptions 1′, 2-5, 6′, 7, and 8 hold. Then $\|\hat{\mathcal{C}}^{CS}(u_1, u_2) - \mathcal{C}^{CS}(u_1, u_2)\| = o_p(1)$ uniformly over $u_1, u_2 \in \mathcal{U}$.*

Finally, we show how to obtain confidence bands for $\beta(u)$ that hold uniformly over $\mathcal{U}$. Observe that $\beta(u)$ is a d_x-vector, that is, $\beta(u) = (\beta_1(u), \ldots, \beta_{d_x}(u))'$. As in the main text, we focus on $\beta_1(u)$, the first component of $\beta(u)$, and we suggest constructing uniform confidence bands via multiplier bootstrap method. An important difference from the results in the main text is that now we should bootstrap on the cluster level.

Specifically, let $\hat{\beta}_1(u)$, $V^{CS}(u)$, and $\hat{V}^{CS}(u)$ denote the 1st component of $\hat{\beta}(u)$, the $(1,1)$st component of $\mathcal{C}^{CS}(u, u)$, and the $(1,1)$st component of $\hat{\mathcal{C}}^{CS}(u, u)$, respectively. Define

$$T = \sup_{u\in\mathcal{U}} \sqrt{G}|\hat{V}(u)^{-1/2}(\hat{\beta}_1(u) - \beta_1(u))|, \tag{18}$$

and let $c_{1-\alpha}$ denote the $(1-\alpha)$ quantile of T. As in the main text, we estimate $c_{1-\alpha}$ by the multiplier bootstrap method. Let $\epsilon_1, ..., \epsilon_M$ be an i.i.d. sequence of $N(0,1)$ random variables that are independent of the data. Also, let $\hat{w}^S_{g,1}$ denote the 1st component of the vector $\hat{S}w_g$. Then the multiplier bootstrap statistic is

$$T^{MB} = \sup_{u \in \mathcal{U}} \frac{1}{\sqrt{G\hat{V}(u)}} \sum_{m=1}^{M} \epsilon_m \left(\sum_{g \in \mathbb{C}_G(m)} (\hat{\alpha}_{g,1}(u) - x_g' \hat{\beta}(u)) \hat{w}^S_{g,1} \right)$$

The multiplier bootstrap critical value $\hat{c}_{1-\alpha}$ is the conditional $(1-\alpha)$ quantile of T^{MB} given the data. Our final theorem in this section explains how to construct uniform confidence bands using $\hat{c}_{1-\alpha}$.

Theorem 8 (Uniform Confidence Bands via Multiplier Bootstrap under Cluster Sampling)**.** *Let Assumptions 1′, 2-5, 6′, 7, and 8 hold. In addition, suppose that all eigenvalues of $J^{CS}(u,u)$ are bounded away from zero uniformly over $u \in \mathcal{U}$. Then*

$$P\left(\beta_1(u) \in \left[\hat{\beta}_1(u) - \hat{c}_{1-\alpha}\sqrt{\frac{\hat{V}(u)}{G}}, \hat{\beta}_1(u) + \hat{c}_{1-\alpha}\sqrt{\frac{\hat{V}(u)}{G}} \right] \text{ for all } u \in \mathcal{U} \right) \to 1 - \alpha$$

as $G \to \infty$.

Appendix F. Further Discussion of the Model in Section 2

In this section, we provide further discussion of our model in Section 2, give a structural interpretation, and outline possible extensions.

F.1. **Structural Model Justifying the Model in Section 2.** Consider the following structural model:

$$y_{ig} = z_{ig}' \widetilde{\alpha}_g(\widetilde{u}_{ig}) \tag{19}$$

where y_{ig} is the response variable of individual i in group g, z_{ig} is a vector of observable individual-level covariates, $\widetilde{u}_{ig}$ is unobserved scalar heterogeneity with values in $[0,1]$, and $\widetilde{\alpha}_g = \{\widetilde{\alpha}_g(u), u \in [0,1]\}$ is a group-specific effect. We assume that the group-specific effect $\widetilde{\alpha}_g$ is determined by vectors of observable and unobservable group-level covariates x_g and ψ_g, respectively, that is, $\widetilde{\alpha}_g(u) = \widetilde{\alpha}(u, x_g, \psi_g)$ for some function $\widetilde{\alpha}$.

In many empirical settings, it is natural to expect that the distribution of $\widetilde{u}_{ig}$ varies across groups, so that the distribution function $F_g : [0,1] \to [0,1]$ of $\widetilde{u}_{ig}$ in group g is indexed by g. We assume that F_g is determined by a vector of unobservable group-level covariates ν_g, that is, $F_g(u) = F(u, \nu_g)$ for some function F. Let $F^{-1}(u, \nu_g)$ denote the (generalized) inverse of the function $u \mapsto F(u, \nu_g)$.

Further, we assume that $\widetilde{u}_{ig}$ is independent of z_{ig} conditional on (x_g, ψ_g, ν_g), which can be considered analogous to the usual independence condition of quantile regression analysis for cross-sectional data adapted to group/panel data as considered here. Under this condition,

$$\widetilde{u}_{ig} = F^{-1}(u_{ig}, \nu_g)$$

for a random variable u_{ig} that is distributed uniformly on $[0,1]$ and that is independent of $(z_{ig}, x_g, \psi_g, \nu_g)$. Therefore, denoting

$$\alpha_g(u) = \alpha(u, x_g, \psi_g, \nu_g) = \widetilde{\alpha}(F^{-1}(u, \nu_g), x_g, \psi_g),$$

rewriting the model (19) as

$$y_{ig} = z_{ig}'\alpha(u_{ig}, x_g, \psi_g, \nu_g),$$

and assuming that the function $u \mapsto z_{ig}'\alpha(u_{ig}, x_g, \psi_g, \nu_g)$ is strictly increasing with probability one, we obtain the following quantile regression model:

$$Q_{y_{ig}|z_{ig},x_g,\psi_g,\nu_g}(u) = z_{ig}'\alpha_g(u), \ \ u \in [0,1],$$

which in turn implies that

$$Q_{y_{ig}|z_{ig},x_g,\alpha_g}(u) = z_{ig}'\alpha_g(u), \ \ u \in [0,1], \tag{20}$$

where $Q_{y_{ig}|z_{ig},x_g,\psi_g,\nu_g}(u)$ denotes the uth quantile of the conditional distributional of y_{ig} given $(z_{ig}, x_g, \psi_g, \nu_g)$ and $Q_{y_{ig}|z_{ig},x_g,\alpha_g}(u)$ denotes the uth quantile of the conditional distribution of y_{ig} given (z_{ig}, x_g, α_g) with $\alpha_g = \{\alpha_g(u), u \in [0,1]\}$.

Equation (1) with any $\mathcal{U} \subset [0,1]$ in Section 2 follows from (20). In turn, equation (2) in Section 2 arises if $\alpha_1(u, x_g, \psi_g, \nu_g)$, the first component of the vector $\alpha(u, x_g, \psi_g, \nu_g)$, can be reasonably well approximated by $x_g'\beta(u) + \varepsilon_g(u)$ where $\varepsilon_g(u) = \varepsilon(u, \psi_g, \nu_g)$, which is a typical assumption in applied regression analysis. This provides a structural interpretation of the model in Section 2.

An advantage of this interpretation is that it yields additional intuition behind the condition that $E[w_g\varepsilon_g(u)] = 0$ for all $u \in \mathcal{U}$ imposed on the instrument w_g in Section 2. In particular, as explained in footnote 5, as long as the vector x_g contains the constant, this condition follows if w_g is independent of η_g where η_g is such that $\varepsilon_g(u) = \varepsilon(u, \eta_g)$. In this section, we have $\varepsilon_g(u) = \varepsilon(u, \psi_g, \nu_g)$, so that $\eta_g = (\psi_g, \nu_g)$. Thus, the instrument w_g should be independent both of ψ_g, a vector of unobserved group-level covariates governing group-specific effects $\widetilde{\alpha}_g(u) = \widetilde{\alpha}(u, x_g, \psi_g)$, and of ν_g, a vector of unobserved group-level covariates governing the distribution of unobserved heterogeneity $F_g(u) = F(u, \nu_g)$. Both of these conditions are reasonable in our empirical application in Section 5.

F.2. **Extension based on a Random Coefficient Model.** One of the conditions we used in the discussion above is a functional form assumption that $\alpha_1(u, x_g, \psi_g, \nu_g)$ can be reasonably well approximated by a linear form $x_g'\beta(u) + \varepsilon_g(u)$ where $\varepsilon_g(u) = \varepsilon(u, \psi_g, \nu_g)$. Here linearity in x_g is a rather flexible assumption because we can always replace x_g by a set of different transformations of x_g whose linear combinations can approximate the function $x_g \mapsto \alpha_1(u, x_g, \psi_g, \nu_g)$ sufficiently well. On the other hand, additive separability of $x_g'\beta(u)$ and $\varepsilon_g(u)$ may be difficult to justify on theoretical grounds. If this is the case, a better approximation of $\alpha_1(u, x_g, \psi_g, \nu_g)$ can be given by

$x_g'\beta_g(u)$ where $\beta_g(u) = \beta(u, \psi_g, \nu_g)$. Therefore, in this section, we briefly comment on how one can estimate the model given by

$$
\begin{aligned}
Q_{y_{ig}|z_{ig},x_g,\alpha_g}(u) &= z_{ig}'\alpha_g(u), \;\; u \in \mathcal{U}, \\
\alpha_{g,1}(u) &= x_g'\beta_g(u), \;\; u \in \mathcal{U},
\end{aligned} \tag{21}
$$

where we use the same notation as above and where (21) can be thought of as a random coefficient model since $\beta_g(u) = \beta(u, \psi_g, \nu_g)$. Throughout this section, we assume that the instrument w_g is independent of the pair (ψ_g, ν_g), which, as explained above, strengthens the condition $E[w_g\varepsilon_g(u)] = 0$ for all $u \in \mathcal{U}$ used in Section 2. Observe that this assumption implies that $\beta_g(u)$ is independent of w_g.

In this model, one can use the same first stage procedure to estimate group-specific effects $\alpha_g(u)$, that is, one can run a quantile regression on the data $\{(z_{ig}, y_{ig}), i = 1, \ldots, N_g\}$ separately in each group g to find the estimators $\hat{\alpha}_g(u)$ of $\alpha_g(u)$. If the number of observations per group grows sufficiently fast as the number of groups gets large, $\hat{\alpha}_g(u)$ will consistently estimate $\alpha_g(u)$ uniformly over $g = 1, \ldots, G$. In the second stage, we will have to replace the 2SLS estimator suitable for (2) by an estimator suitable for (21). Given that $\beta_g(u)$ is independent of w_g, several approaches developed in the literature can be applied to learn some features of the distribution of $\beta_g(u) = \beta(u, \psi_g, \nu_g)$ depending on what side conditions we impose on the model; see, for example, Imbens and Angrist (1994), Heckman and Vytlacil (1998), Florens, Heckman, Meghir, and Vytlacil (2008), and Masten and Torgovitsky (2014). For concreteness, we describe here the approach developed by Masten and Torgovitsky (2014), which, under certain control variable assumptions and some other technical assumptions, yields consistent estimates of $\bar{\beta}(u) = E[\beta_g(u)]$.

To explain their procedure, assume, for simplicity, that there is only one endogenous covariate among the vector of covariates x_g, that is, $x_g = (1, \widetilde{x}_g, x_{g,d_x})'$ where $\widetilde{x}_g = (x_{g,2}, \ldots, x_{g,d_x-1})'$ is independent of (ψ_g, ν_g). Assume that x_g is continuously distributed, and $x_{g,d_x} = h(w_g, v_g)$ where the function $v \mapsto h(w_g, v)$ is increasing with probability one, and the (scalar) random variable v_g is such that w_g is independent of (ψ_g, ν_g, v_g) (control variable assumption). Then Masten and Torgovitsky (2014) show that under some further technical conditions, $\bar{\beta}(u) = E[\beta_g(u)]$ can be consistently estimated by

$$
\hat{\beta}(u) = \int_0^1 \hat{\beta}(u, r)dr
$$

where

$$
\hat{\beta}(u, r) = \left(\frac{1}{G} \sum_{g=1}^{G} \hat{k}_g(r) x_g x_g' \right)^{-1} \left(\frac{1}{G} \sum_{g=1}^{G} \hat{k}_g(u) x_g \alpha_g(u) \right), \tag{22}
$$

$\hat{k}_g(r) = h^{-1}K(h^{-1}(\hat{R}_g - r))$, h is the bandwidth value satisfying $h = h_G \to 0$ as $G \to \infty$, K is the kernel function, $\hat{R}_g = \hat{F}(x_{g,d_x}|w_g)$, and $\hat{F}(x|w)$ is an estimator of $F(x|w)$, the conditional probability that $x_{g,d_x} \leq x$ given $w_g = w$.

Note, however, that $\alpha_g(u)$ is unknown in our setting, and so this estimator is infeasible. To obtain a feasible estimator, one can substitute $\hat{\alpha}_g(u)$ calculated in the first stage instead of $\alpha_g(u)$ in (22). Using the same techniques as those developed in this paper, it is then possible to show that the feasible estimator is asymptotically equivalent to the infeasible estimator under weak condition on the growth of the number of observations per group as the number of groups gets large, and so the feasible estimator has the same asymptotic properties as those of $\hat{\beta}(u)$, which are in turn developed in Masten and Torgovitsky (2014).

Appendix G. Proofs

In this Appendix, we first prove some preliminary lemmas. Then we present the proofs of the theorems stated in the main text as well as the proof of Theorems 4 and 5 stated in Appendices C and D. In all proofs, c and C denote strictly positive generic constants that depend only on c_M, c_f, C_M, C_f, C_L whose values can change at each appearance.

We will use the following notation in addition to that appearing in the main text. Let

$$\begin{aligned} A(u) &= (\alpha_{1,1}(u), ..., \alpha_{G,1}(u))', \\ \widetilde{\beta}(u) &= (X'P_W X)^{-1}(X'P_W A(u)), \\ J_g(u) &= E_g[z_{1g}z_{1g}' f_g(z_{1g}'\alpha_g(u))]. \end{aligned} \tag{23}$$

For $\eta, \alpha \in \mathbb{R}^{d_z}$, and $u \in \mathcal{U}$, consider the function $f_{\eta,\alpha,u} : \mathbb{R}^{d_z} \times \mathbb{R} \to \mathbb{R}$ defined by

$$f_{\eta,\alpha,u}(z, y) = (z'\eta) \cdot (1\{y \le z'\alpha\} - u). \tag{24}$$

Let $\mathcal{F} = \{f_{\eta,\alpha,u} : \eta, \alpha \in \mathbb{R}^{d_z}; u \in \mathcal{U}\}$; that is, $\mathcal{F}$ is the class of functions $f_{\eta,\alpha,u}$ as η, α vary over $\mathbb{R}^{d_z}$ and u varies over $\mathcal{U}$. For $\alpha \in \mathbb{R}^{d_z}$ and $u \in \mathcal{U}$, let the function $h_{\alpha,u} : \mathbb{R}^{d_z} \times \mathbb{R} \to \mathbb{R}^{d_z}$ be defined by

$$h_{\alpha,u}(z, y) = z(1\{y \le z'\alpha\} - u),$$

and let $h_{k,\alpha,u}$ denote kth component of $h_{\alpha,u}$. Let $\mathcal{H}_k = \{h_{k,\alpha,u} : \alpha \in \mathbb{R}^{d_z}; u \in \mathcal{U}\}$. Note that $\mathcal{H}_k \subset \mathcal{F}$ for all $k = 1, ..., d_z$.

We will also use the following notation from the empirical process literature,

$$\mathbb{G}^g(f) = \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N_g} (f(z_{ig}, y_{ig}) - E_g[f(z_{ig}, y_{ig})])$$

for $f \in \mathcal{F}, \mathcal{H}$, or $\mathcal{H}_k$, $k = 1, \dots, d_z$.

Preliminary Lemmas. In all lemmas, we implicitly impose Assumptions 1-8.

Lemma 1. *As $G \to \infty$,*

$$\hat{Q}_{xw} = \frac{1}{G} \sum_{g=1}^{G} x_g w_g' \to_p Q_{xw}, \tag{25}$$

$$\hat{Q}_{ww} = \frac{1}{G}\sum_{g=1}^{G} w_g w_g' \to_p Q_{ww} \tag{26}$$

where Q_{xw} and Q_{ww} appear in Assumption 2.

Proof. We only prove (25). The proof of (26) is similar. To prove (25), observe that $G^{-1}\sum_{g=1}^{G} E[x_g w_g'] \to Q_{xw}$ by Assumption 2. Therefore, it suffices to prove that

$$\frac{1}{G}\sum_{g=1}^{G}\left(x_g w_g' - E[x_g w_g']\right) \to_p 0. \tag{27}$$

In turn, (27) follows from Assumptions 2(iv) and 4(i) and Chebyshev's inequality. Hence, (25) follows. This completes the proof of the lemma. □

Lemma 2. *As $G \to \infty$,*

$$\frac{1}{G}\sum_{g=1}^{G} \varepsilon_g(u_1)\varepsilon_g(u_2) w_g w_g' \to_p J(u_1, u_2)$$

uniformly over $u_1, u_2 \in \mathcal{U}$.

Proof. Observe that we cannot apply a uniform law of large numbers with bracketing directly because the data are not necessarily i.i.d. across g. Therefore, we provide a complete proof.

Since

$$\frac{1}{G}\sum_{g=1}^{G} E\left[\varepsilon_g(u_1)\varepsilon_g(u_2) w_g w_g'\right] \to J(u_1, u_2)$$

uniformly over $u_1, u_2 \in \mathcal{U}$ by Assumption 6(ii), it suffices to prove that

$$\frac{1}{G}\sum_{g=1}^{G}\left(\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l} - E\left[\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l}\right]\right) \to_p 0 \tag{28}$$

uniformly over $u_1, u_2 \in \mathcal{U}$ for all $k, l = 1, \dots, d_w$.

To this end, fix $u_1, u_2 \in \mathcal{U}$ and $k, l = 1 \dots, d_w$. We first show (28) for these values of u_1, u_2, k, and l. Note that we cannot use Chebyshev's inequality here because $E[(\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l})^2]$ is not necessarily finite. Instead, we use a more delicate method as presented in Theorem 2.1.7 of Tao (2012). Let $\delta = c_M/4$. Then by Hölder's inequality,

$$E[|\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l}|^{1+\delta}] \le \left(E[|\varepsilon_g(u_1)\varepsilon_g(u_2)|^{2+2\delta}] \cdot E[|w_{g,k} w_{g,l}|^{2+2\delta}]\right)^{1/2}.$$

In turn,

$$E[|\varepsilon_g(u_1)\varepsilon_g(u_2)|^{2+2\delta}] \le E\left[\sup_{u\in\mathcal{U}} |\varepsilon_g(u)|^{4+4\delta}\right] \le C_M,$$

$$E[|w_{g,k} w_{g,l}|^{2+2\delta}] \le E\left[\|w_g\|^{4+4\delta}\right] \le C_M$$

by Assumptions 6(i) and 2(iv). Hence,

$$E[|\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l}|^{1+\delta}] \le C_M,$$

and so denoting $X_g = \varepsilon_g(u_1)\varepsilon_g(u_2)w_{g,k}w_{g,l} - E[\varepsilon_g(u_1)\varepsilon_g(u_2)w_{g,k}w_{g,l}]$, we obtain

$$E[|X_g|^{1+\delta}] \leq C. \tag{29}$$

With this notation, (28) is equivalent to $G^{-1}\sum_{g=1}^G X_g \to_p 0$.

Now for $N > 0$ to be chosen later, denote $X_{g,\leq N} = X_g \cdot 1\{|X_g| \leq N\}$ and $X_{g,>N} = X_g \cdot 1\{|X_g| > N\}$. Then by Fubini's theorem and Markov's inequality,

$$\begin{aligned}
|E[X_{g,>N}]| \leq E[|X_{g,>N}|] &= \int_0^\infty P(|X_g| \cdot 1\{|X_g| > N\} > t)dt \\
&= \int_0^N P(|X_g| > N)dt + \int_N^\infty P(|X_g| > t)dt \\
&\leq N \cdot \frac{E[|X_g|^{1+\delta}]}{N^{1+\delta}} + \int_N^\infty \frac{E[|X_g|^{1+\delta}]}{t^{1+\delta}}dt \\
&= \frac{E[|X_g|^{1+\delta}]}{N^\delta} + \frac{E[|X_g|^{1+\delta}]}{\delta N^\delta} \leq CN^{-\delta}
\end{aligned}$$

where in the last inequality we used (29). Hence, by Markov's inequality, for any $\varepsilon > 0$,

$$P\Big(\Big|\frac{1}{G}\sum_{g=1}^G X_{g,>N}\Big| > \varepsilon\Big) \leq \frac{1}{\varepsilon G}\sum_{g=1}^G E[|X_{g,>N}|] \leq \frac{C}{\varepsilon N^\delta},$$

and since $|E[X_{g,\leq N}]| = |E[X_{g,>N}]| \leq CN^{-\delta}$,

$$\begin{aligned}
P\Big(\Big|\frac{1}{G}\sum_{g=1}^G X_{g,\leq N}\Big| > \varepsilon + CN^{-\delta}\Big) &\leq P\Big(\Big|\frac{1}{G}\sum_{g=1}^G (X_{g,\leq N} - E[X_{g,\leq N}])\Big| > \varepsilon\Big) \\
&\leq \frac{1}{\varepsilon G^2}\sum_{g=1}^G E[X_{g,\leq N}^2] \leq \frac{N^2}{\varepsilon G}.
\end{aligned}$$

Thus, setting $N = G^{1/3}$, we obtain $G^{-1}\sum_{g=1}^G X_g \to_p 0$, which is equivalent to (28) for given u_1, u_2, k, and l.

Next, to show that (28) holds uniformly over $u_1, u_2 \in \mathcal{U}$, for $\delta > 0$, let $\mathcal{U}_\delta$ be a finite subset of $\mathcal{U}$ such that for any $u \in \mathcal{U}$, there exists $u' \in \mathcal{U}_\delta$ satisfying $|\varepsilon_g(u) - \varepsilon_g(u')| \leq \delta$. Existence of such a set $\mathcal{U}_\delta$ follows from Assumption 6(iii). Then

$$\begin{aligned}
&\sup_{u_1,u_2\in\mathcal{U}}\Big|\frac{1}{G}\sum_{g=1}^G (\varepsilon_g(u_1)\varepsilon_g(u_2)w_{g,k}w_{g,l} - E[\varepsilon_g(u_1)\varepsilon_g(u_2)w_{g,k}w_{g,l}])\Big| \\
&\leq \max_{u_1,u_2\in\mathcal{U}_\delta}\Big|\frac{1}{G}\sum_{g=1}^G (\varepsilon_g(u_1)\varepsilon_g(u_2)w_{g,k}w_{g,l} - E[\varepsilon_g(u_1)\varepsilon_g(u_2)w_{g,k}w_{g,l}])\Big| \\
&\quad + \frac{2\delta}{G}\sum_{g=1}^G \left(\sup_{u\in\mathcal{U}}|\varepsilon_g(u)| \cdot |w_{g,k}w_{g,l}| + E\left[\sup_{u\in\mathcal{U}}|\varepsilon_g(u)| \cdot |w_{g,k}w_{g,l}|\right]\right) = o_p(1) + \delta \cdot O_p(1)
\end{aligned}$$

by the result above and Chebyshev's inequality. Since δ is arbitrary, this completes the proof. □

Lemma 3. *As $G \to \infty$,*

$$\frac{1}{\sqrt{G}} \sum_{g=1}^{G} w_g \varepsilon_g(\cdot) \Rightarrow \mathbb{G}^0(\cdot), \text{ in } \ell^\infty(\mathcal{U})$$

where $\mathbb{G}^0$ is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function $J(u_1, u_2)$ for all u_1, u_2 appearing in Assumption 6.

Proof. For any finite set $\mathcal{U}' \subset \mathcal{U}$, it follows from Assumption 6(ii), Lindeberg's Central Limit Theorem, and the Cramér-Wold device (see, for example, Theorems 11.2.5 and 11.2.3 in Lehmann and Romano (2005)) that

$$\Big(\frac{1}{\sqrt{G}} \sum_{g=1}^{G} w_g \varepsilon_g(u)\Big)_{u \in \mathcal{U}'} \Rightarrow (N(u))_{u \in \mathcal{U}'}$$

where $(N(u))_{u\in\mathcal{U}'}$ is a zero-mean Gaussian vector with covariance function $J(u_1, u_2)$ for all $u_1, u_2 \in \mathcal{U}'$. Therefore, it follows from the second part of Theorem 14 that the asserted claim of the lemma holds if for any $k = 1, \dots, d_w$ and $Z_g(u) = G^{-1/2} w_{g,k} \varepsilon_g(u)$, $g = 1, \dots, G$ and $u \in \mathcal{U}$, the sequence $\sum_{g=1}^G Z_g(\cdot)$ is asymptotically tight in $\ell^\infty(\mathcal{U})$. Fix $k = 1, \dots, d_w$. To prove that $\sum_{g=1}^G Z_g(\cdot)$ is asymptotically tight in $\ell^\infty(\mathcal{U})$, we apply the first part of Theorem 14 with Gaussian-dominated semi-metric $\rho : \mathcal{U} \times \mathcal{U} \to \mathbb{R}_+$ defined by $\rho(u_1, u_2) = C|u_2 - u_1|$ for sufficiently large constant $C > 0$; see discussion in front of Theorem 14 for the definition of Gaussian-dominated semi-metrics.

Condition (i) of Theorem 14 holds because for any $\eta > 0$ and $\delta = 1 + c_M/2$,

$$\begin{aligned}
\sum_{g=1}^{G} E\left[\sup_{u\in\mathcal{U}} |Z_g(u)| \cdot 1\left\{\sup_{u\in\mathcal{U}} |Z_g(u) > \eta\right\}\right] &\le \frac{1}{\eta^\delta G^{1/2+\delta/2}} \sum_{g=1}^{G} E\left[\sup_{u\in\mathcal{U}} |\varepsilon_g(u)|^{1+\delta} |w_{g,k}|^{1+\delta}\right] \\
&\le \frac{1}{\eta^\delta G^{1/2+\delta/2}} \sum_{g=1}^{G} \left(E\left[\sup_{u\in\mathcal{U}} |\varepsilon_g(u)|^{2+2\delta}\right] \cdot E\left[|w_{j,k}|^{2+2\delta}\right]\right)^{1/2} \to 0
\end{aligned}$$

by Hölder's inequality and Assumptions 2(iv) and 6(i).

Condition (ii) of Theorem 14 holds because for any $u_1, u_2 \in \mathcal{U}$,

$$\begin{aligned}
\sum_{g=1}^{G} E[(Z(u_2) - Z(u_1))^2] &= \frac{1}{G} \sum_{g=1}^{G} E[(w_{g,k}\varepsilon_g(u_2) - w_{g,k}\varepsilon_g(u_1))^2] \\
&\le \frac{C}{G} \sum_{g=1}^{G} E[w_{g,k}^2 |u_2 - u_1|^2] \le C|u_2 - u_1|^2 \le \rho^2(u_1, u_2)
\end{aligned}$$

by Assumptions 2(iv) and 6(iii) since the constant C in the definition of $\rho(u_1, u_2)$ is large enough.

Finally, condition (iii) of Theorem 14 holds because by Markov's inequality for any $\epsilon > 0$,

$$\sup_{t>0} \sum_{g=1}^{G} t^2 P\left(\sup_{\rho(u_1,u_2)\le 2\epsilon} |Z_g(u_2) - Z_g(u_1)| > t\right)$$

$$\leq \frac{1}{G}\sum_{g=1}^{G} E\left[\sup_{\rho(u_1,u_2)\leq 2\epsilon} |w_{g,k}\varepsilon_g(u_2) - w_{g,k}\varepsilon(u_1)|^2\right] \leq C \sup_{\rho(u_1,u_2)\leq 2\epsilon} |u_2 - u_1|^2 \leq \epsilon^2$$

by Assumptions 2(iv) and 6(iii) since the constant C in the definition of $\rho(u_1, u_2)$ is large enough.

Therefore, Theorem 14 implies that the sequence $\sum_{g=1}^{G} Z_g(\cdot)$ is asymptotically tight in $\ell^\infty(\mathcal{U})$. The asserted claim follows. □

Lemma 4. *There exist constants $c, C > 0$ such that (i) for all $u \in \mathcal{U}$ and $g = 1, \ldots, G$, all eigenvalues of $J_g(u)$ are bounded from below by c, and (ii) for all $u_1, u_2 \in \mathcal{U}$ and $g = 1 \ldots, G$, $\|J_g^{-1}(u_2) - J_g^{-1}(u_1)\| \leq C|u_2 - u_1|$.*

Proof. For any $u \in \mathcal{U}$ and $\alpha \in \mathbb{R}^{d_z}$ with $\|\alpha\| = 1$,

$$\alpha' J_g(u)\alpha \geq c_f \alpha' E_g[z_{1g}z_{1g}']\alpha \geq c_f c_M \tag{30}$$

where the first inequality follows from Assumption 7(ii) and the second from Assumption 4(ii). This gives the first asserted claim.

To prove the second claim, observe that

$$\|J_g^{-1}(u_2) - J_g^{-1}(u_1)\| \leq \|J_g^{-1}(u_1)\|\|J_g^{-1}(u_2)\|\|J_g(u_2) - J_g(u_1)\| \leq \frac{\|J_g(u_2) - J_g(u_1)\|}{(c_f c_M)^2}$$

where the second inequality follows from (30). Hence, it suffices to show that $\|J_g(u_2) - J_g(u_1)\| \leq C|u_2 - u_1|$ for some $C > 0$. To this end, note that

$$|z_{1g}'\alpha_g(u_2) - z_{1g}'\alpha_g(u_1)| \leq \|z_{1g}\|\|\alpha_g(u_2) - \alpha_g(u_1)\| \leq C_M C_L |u_2 - u_1|$$

where the second inequality follows from Assumptions 4(i) and 5. Thus, if $|u_2 - u_1| < c_f/(C_M C_L)$, then $z_{1g}'\alpha_g(u_2) \in B_g(u_1, c_f)$, and so

$$\begin{aligned}\|J_g(u_2) - J_g(u_1)\| &\leq \left\|E_g[z_{1g}z_{1g}' \cdot \left|f_g(z_{1g}'\alpha_g(u_2)) - f_g(z_{1g}'\alpha_g(u_1))\right|]\right\| \\ &\leq C_f C_M C_L |u_2 - u_1| \cdot \|E_g[z_{1g}z_{1g}']\| \leq C_f C_M^3 C_L |u_2 - u_1|\end{aligned}$$

where the second inequality follows from Assumption 7(i) and the derivation above, and the third from Assumption 4(i). On the other hand, if $|u_2 - u_1| \geq c_f/(C_M C_L)$, then

$$\begin{aligned}\|J_g(u_2) - J_g(u_1)\| &\leq \|J_g(u_1)\| + \|J_g(u_2)\| \leq 2C_f\|E_g[z_{1g}z_{1g}']\| \\ &\leq 2C_f C_M^2 \leq c_f^{-1} C_f C_M^3 C_L |u_2 - u_1|\end{aligned}$$

where the first inequality follows from the triangle inequality, the second from Assumption 7(ii), and the third from Assumption 4(i). This gives the second asserted claim and completes the proof of the lemma. □

Lemma 5. *There exist constants $c, C > 0$ such that for all $g = 1, \ldots, G$,*

$$\|E_g[h_{\alpha,u}(z_{1g}, y_{1g})] - J_g(u)(\alpha - \alpha_g(u))\| \leq C\|\alpha - \alpha_g(u)\|^2, \tag{31}$$

$$E_g[(\alpha - \alpha_g(u))' h_{\alpha,u}(z_{1g}, y_{1g})] \geq c\|\alpha - \alpha_g(u)\|^2. \tag{32}$$

for all $u \in \mathcal{U}$ and $\alpha \in \mathbb{R}^{d_z}$ satisfying $\|\alpha - \alpha_g(u)\| \leq c$.

Proof. Second-order Taylor expansion around $\alpha_g(u)$ and the law of iterated expectation give

$$\begin{aligned} E_g[h_{\alpha,u}(z_{1g}, y_{1g})] &= E_g[z_{1g}(1\{y_{1g} \leq z_{1g}'\alpha\} - u)] = E_g[z_{1g}(F_g(z_{1g}'\alpha) - u)] \\ &= E_g[z_{1g}(F_g(z_{1g}'\alpha_g(u)) - u)] + J_g(u)(\alpha - \alpha_g(u)) + r_n(u), \end{aligned}$$

where $r_n(u)$ is the remainder and $F_g(\cdot)$ is the conditional distribution function of y_{1g} given (z_{1g}, α_g). The first claim of the lemma follows from $E_g[z_{1g}(F_g(z_{1g}'\alpha_g(u)) - u)] = 0$, which holds because $z_{1g}'\alpha_g(u)$ is the uth quantile of the conditional distribution of y_{1g}, and from $\|r_n(u)\| \leq C\|\alpha - \alpha_g(u)\|^2$ for some $C > 0$, which holds by Assumptions 4(i) and 7(i).

To prove the second claim, note that if $\|\alpha - \alpha_g(u)\|$ is sufficiently small, then $\|(\alpha - \alpha_g(u))' r_n(u)\| \leq c\|\alpha - \alpha_g(u)\|^2$ for an arbitrarily small constant $c > 0$. On the other hand,

$$(\alpha - \alpha_g(u))' J_g(u)(\alpha - \alpha_g(u)) \geq c\|\alpha - \alpha_g(u)\|^2$$

by Lemma 4. Combining these inequalities gives the second claim. □

Lemma 6. *The function class $\mathcal{F}$, defined in the beginning of this section, is a VC subgraph class of functions. Moreover, for all $k = 1, ..., d_z$, $\mathcal{H}_k$ is a VC subgraph class of functions as well.*

Proof. A similar proof can be found in Belloni, Chernozhukov, and Hansen (2006). We present the proof here for the sake of completeness. Consider the class of sets $\{x \in \mathbb{R}^{d_z+1} : a'x \leq 0\}$ with a varying over $\mathbb{R}^{d_z+1}$. It is well known that this is a VC subgraph class of sets; see, for example, exercise 14 of chapter 2.6 in Van der Vaart and Wellner (1996). Further, note that

$$\begin{aligned} \{(z, y, t) : f_{\eta,\alpha,u}(z, y) > t\} = &\left(\{y \leq z'\alpha\} \cap \{z'\eta > t/(1-u)\}\right) \\ &\cup \left(\{y > z'\alpha\} \cap \{z'\eta < -t/u\}\right). \end{aligned}$$

Therefore, the first result follows from Lemma 2.6.17(ii,iii) in Van der Vaart and Wellner (1996). The second result follows from the fact that $\mathcal{H}_k \subset \mathcal{F}$. □

Lemma 7. *For any $\varphi \geq 1$, there exists a constant $C > 0$ such that for all $g = 1, \ldots, G$*

$$E_g\left[\sup_{u \in \mathcal{U}} \|\mathbb{G}^g(h_{\alpha_g(u),u})\|^\varphi\right] \leq C.$$

Proof. Observe that

$$E_g\left[\sup_{u \in \mathcal{U}} \|\mathbb{G}^g(h_{\alpha_g(u),u})\|^\varphi\right] \leq C \sum_{k=1}^{d_z} E_g\left[\sup_{u \in \mathcal{U}} |\mathbb{G}^g(h_{k,\alpha_g(u),u})|^\varphi\right] \leq C \sum_{k=1}^{d_z} E_g\left[\sup_{f \in \mathcal{H}_k} |\mathbb{G}^g(f)|^\varphi\right].$$

Further, all functions in $\mathcal{H}_k$ are bounded by some constant $C > 0$ by Assumption 4(i) and the set of functions $\mathcal{H}_k$ is a VC subgraph class by Lemma 6. Therefore, combining Theorems 9 and 11 gives $E_g[\sup_{f\in\mathcal{H}_k}|\mathbb{G}^g(f)|] \leq C$, and so Theorem 13 shows that

$$E_g\left[\sup_{f\in\mathcal{H}_k}|\mathbb{G}^g(f)|^{\varphi}\right] \leq C.$$

The asserted claim follows. □

Lemma 8. *There exist constants $c, C > 0$ such that for all $g = 1,\dots,G$,*

$$E_g\left[\sup_{u_2\in\mathcal{U}:|u_2-u_1|\leq\epsilon}\|\mathbb{G}^g(h_{\alpha_g(u_2),u_2}) - \mathbb{G}^g(h_{\alpha_g(u_1),u_1})\|^4\right] \leq C\epsilon$$

for all $\epsilon \in (0,c)$ and $u_1 \in \mathcal{U}$.

Proof. Fix some $u_1 \in \mathcal{U}$. Observe that

$$\begin{aligned}
&E_g\left[\sup_{u_2\in\mathcal{U}:|u_2-u_1|\leq\epsilon}\|\mathbb{G}^g(h_{\alpha_g(u_2),u_2}) - \mathbb{G}^g(h_{\alpha_g(u_1),u_1})\|^4\right]\\
&\qquad \leq C\sum_{k=1}^{d_z}E_g\left[\sup_{u_2\in\mathcal{U}:|u_2-u_1|\leq\epsilon}|\mathbb{G}^g(h_{k,\alpha_g(u_2),u_2}) - \mathbb{G}^g(h_{k,\alpha_g(u_1),u_1})|^4\right].
\end{aligned}$$

Consider the function $F : \mathbb{R}^{d_z}\times\mathbb{R} \to \mathbb{R}$ given by

$$F(z,y) = C\left(1\{|y - z'\alpha_g(u_1)| \leq C\epsilon\} + \epsilon\right)$$

for some sufficiently large $C > 0$. By Assumptions 4(i) and 5, $|z_{ig}'(\alpha_g(u_2) - \alpha_g(u_1))| \leq C|u_2 - u_1|$ for some $C > 0$. Therefore, for all $u_2 \in \mathcal{U}$ satisfying $|u_2 - u_1| \leq \epsilon$,

$$\left|h_{k,\alpha_g(u_2),u_2}(z_{ig},y_{ig}) - h_{k,\alpha_g(u_1),u_1}(z_{ig},y_{ig})\right| \leq F(z_{ig},y_{ig})$$

by Assumption 4(i). Note that $E_g[F^2(z_{ig},y_{ig})] \leq C\epsilon$ for some $C > 0$ by Assumption 7(ii) if $\epsilon \leq 1$. Also, for $M = \max_{1\leq i\leq N_g}F(z_{ig},y_{ig})$, we have $E[M^2] \leq Cn\epsilon$. Further, by Lemma 6, $\mathcal{H}_k$ is a VC subgraph class of functions, so that the function class $\tilde{\mathcal{H}}_k = \{h_{k,\alpha_g(u_2),u_2} - h_{k,\alpha_g(u_1),u_1} : u_2 \in [u_1-\epsilon, u_1+\epsilon]\}$ is a VC type class by Theorem 9. So, applying Theorem 11 with F as an envelope yields

$$E_g\left[\sup_{u_2\in\mathcal{U}:|u_2-u_1|\leq\epsilon}|\mathbb{G}^g(h_{k,\alpha_g(u_2),u_2}) - \mathbb{G}^g(h_{k,\alpha_g(u_1),u_1})|\right] \leq C\sqrt{\epsilon},$$

and so Theorem 13 shows that

$$E_g\left[\sup_{u_2\in\mathcal{U}:|u_2-u_1|\leq\epsilon}|\mathbb{G}^g(h_{k,\alpha_g(u_2),u_2}) - \mathbb{G}^g(h_{k,\alpha_g(u_1),u_1})|^4\right] \leq C\epsilon.$$

The asserted claim follows. □

Lemma 9. *There exist constants $c, C > 0$ such that for all $g = 1, \dots, G$,*

$$E_g\left[\sup_{u\in\mathcal{U}}\sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\le\epsilon}\|\mathbb{G}^g(h_{\alpha,u})-\mathbb{G}^g(h_{\alpha_g(u),u})\|^2\right]\le C\Big(\epsilon\log(1/\epsilon)+N_g^{-1}\log^2(1/\epsilon)\Big)$$

for all $\epsilon \in (0, c)$.

Proof. Observe that

$$\begin{aligned}&E_g\left[\sup_{u\in\mathcal{U}}\sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\le\epsilon}\|\mathbb{G}^g(h_{\alpha,u})-\mathbb{G}^g(h_{\alpha_g(u),u})\|^2\right] &(33)\\ &\quad\le C\sum_{k=1}^{d_z}E_g\left[\sup_{u\in\mathcal{U}}\sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\le\epsilon}|\mathbb{G}^g(h_{k,\alpha,u})-\mathbb{G}^g(h_{k,\alpha_g(u),u})|^2\right]. &(34)\end{aligned}$$

Consider the function class

$$\tilde{\mathcal{H}}_k=\{h_{k,\alpha,u}-h_{k,\alpha_g(u),u}:u\in\mathcal{U};\alpha\in\mathbb{R}^{d_z};\|\alpha-\alpha_g(u)\|\le\epsilon\}.$$

By Lemma 6 and Theorem 9, $\mathcal{F}$ is a VC type class, and so Theorem 10 implies that $\tilde{\mathcal{H}}_k \subset \mathcal{F} - \mathcal{F}$ is also a VC type class. In addition, all functions from $\tilde{\mathcal{H}}_k$ are bounded in absolute value by some constant $C > 0$ by Assumption 4(i). Moreover, for any $f \in \tilde{\mathcal{H}}_k$, $E_g[f(z_{ig}, y_{ig})^2] \le C\epsilon$ if $\epsilon \le 1$. Thus, applying Theorem 11 with the function class $\tilde{\mathcal{H}}_k$ yields

$$E_g\left[\sup_{u\in\mathcal{U}}\sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\le\epsilon}|\mathbb{G}^g(h_{k,\alpha,u})-\mathbb{G}^g(h_{k,\alpha_g(u),u})|\right]\le C\left(\sqrt{\epsilon\log(1/\epsilon)}+N_g^{-1/2}\log(1/\epsilon)\right),$$

and so Theorem 13 gives

$$E_g\left[\sup_{u\in\mathcal{U}}\sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\le\epsilon}|\mathbb{G}^g(h_{k,\alpha,u})-\mathbb{G}^g(h_{k,\alpha_g(u),u})|^2\right]\le C\Big(\epsilon\log(1/\epsilon)+N_g^{-1}\log^2(1/\epsilon)\Big).$$

The asserted claim follows. ☐

Lemma 10. *Uniformly over $u \in \mathcal{U}$,*

$$\frac{1}{\sqrt{G}}\sum_{g=1}^{G}J_g^{-1}(u)\mathbb{G}^g(h_{\alpha_g(u),u})w_g'=O_p(1).$$

Proof. To prove this lemma, we use Theorem 14 with the semi-metric $\rho(u_1, u_2) = C|u_2 - u_1|^{1/4}$ defined for all $u_1, u_2 \in \mathcal{U}$ and some sufficiently large constant $C > 0$. Clearly, ρ is Gaussian-dominated; see discussion before Theorem 14 for the definition. Define $v_g(u) = J_g^{-1}(u)\mathbb{G}^g(h_{\alpha_g(u),u})$ and

$$Z_{g,k,m}(u)=v_{g,k}(u)w_{g,m}/\sqrt{G}$$

where $v_{g,k}(u)$ and $w_{g,m}$ denote kth and mth components of $v_g(u)$ and w_g, respectively. Then the asserted claim is equivalent to the statement that

$$\sum_{g=1}^{G}Z_{g,k,m}(u)=O_p(1)\text{ uniformly over }u\in\mathcal{U}\tag{35}$$

for all k and m. To prove (35), observe first that by Assumptions 1(i) and 2(iii), zero-mean processes $Z_{g,k,m}(\cdot)$ are independent across g. Also, for any $a > 0$,

$$\sum_{g=1}^{G} E\left[\sup_{u\in\mathcal{U}} |Z_{g,k,m}(u)| \cdot 1\left\{\sup_{u\in\mathcal{U}} |Z_{g,k,m}(u)| > a\right\}\right]$$
$$\leq a^{-1}\sum_{g=1}^{G} E\left[\sup_{u\in\mathcal{U}} Z^2_{g,k,m}(u) \cdot 1\left\{\sup_{u\in\mathcal{U}} |Z_{g,k,m}(u)| > a\right\}\right]$$
$$\leq \frac{1}{aG}\sum_{g=1}^{G} E\left[\sup_{u\in\mathcal{U}} (v_{g,k}(u)w_{g,m})^2 \cdot 1\left\{\sup_{u\in\mathcal{U}} |v_{g,k}(u)w_{g,m}| > \sqrt{G}a\right\}\right]. \tag{36}$$

Further, pick some $0 < \varphi < 2$. The expression under the sum in (36) is bounded from above by Lemma 4 by

$$\frac{C}{a^{\varphi}G^{\varphi/2}} E\left[\sup_{u\in\mathcal{U}} \|\mathbb{G}^g(h_{\alpha_g(u),u})\|^{2+\varphi}\|w_g\|^{2+\varphi}\right]$$
$$\leq \frac{C}{a^{\varphi}G^{\varphi/2}}\left(E\left[\sup_{u\in\mathcal{U}} \|\mathbb{G}^g(h_{\alpha_g(u),u})\|^{\frac{4(2+\varphi)}{2-\varphi}}\right]\right)^{\frac{2-\varphi}{4}} \left(E\left[\|w_g\|^4\right]\right)^{\frac{2+\varphi}{4}} \leq \frac{C}{a^{\varphi}G^{\varphi/2}} \to 0$$

uniformly over $g = 1, \ldots, G$ where the second line follows from Hölder's inequality, Assumption 2(iv), and Lemma 7. This gives condition (i) of Theorem 14.

Next, we verify condition (ii) of Theorem 14. For any $u_1, u_2 \in \mathcal{U}$,

$$\sum_{g=1}^{G} E\left[(Z_{g,k,m}(u_2) - Z_{g,k,m}(u_1))^2\right] = \frac{1}{G}\sum_{g=1}^{G}\left(E[w^4_{g,m}]\right)^{1/2} \cdot \left(E[(v_{g,k}(u_2) - v_{g,k}(u_1))^4]\right)^{1/2}.$$

Further, using an elementary inequality $(a+b)^4 \leq C(a^4+b^4)$ for all $a, b \in \mathbb{R}^p$ gives

$$\begin{aligned}E_g[(v_{g,k}(u_2) - v_{g,k}(u_1))^4] &\leq CE_g[\|J_g^{-1}(u_2)\|^4 \cdot \|\mathbb{G}^g(h_{\alpha_g(u_2),u_2} - h_{\alpha_g(u_1),u_1})\|^4] \\ &\quad + CE_g[\|J_g^{-1}(u_2) - J_g^{-1}(u_1)\|^4 \cdot \|\mathbb{G}^g(h_{\alpha_g(u_1),u_1})\|^4] \\ &\leq CE_g[\|\mathbb{G}^g(h_{\alpha_g(u_2),u_2} - h_{\alpha_g(u_1),u_1})\|^4] \\ &\quad + CE_g[\|\mathbb{G}^g(h_{\alpha_g(u_1),u_1})\|^4] \cdot |u_2 - u_1|^4\end{aligned}$$

where the second inequality follows from Lemma 4. In addition,

$$E_g[\|\mathbb{G}^g(h_{\alpha_g(u_2),u_2} - h_{\alpha_g(u_1),u_1})\|^4] \leq C|u_2 - u_1| \text{ and } E_g[\|\mathbb{G}^g(h_{\alpha_g(u_1),u_1})\|^4] \leq C \tag{37}$$

where the first inequality follows from Lemma 8 and the second is easy to check directly. Therefore,

$$E_g[(v_{g,k}(u_2) - v_{g,k}(u_1))^4] \leq C|u_2 - u_1|,$$

and so

$$\sum_{g=1}^{G} E\left[(Z_{g,k,m}(u_2) - Z_{g,k,m}(u_1))^2\right] \leq C|u_2 - u_1|^{1/2} \leq \rho^2(u_1, u_2)$$

by Assumption 2(iv) since the constant C in the definition of $\rho(u_1,u_2)$ is sufficiently large. This gives condition (ii) of Theorem 14.

Finally, condition (iii) of Theorem 14 holds because for any $\epsilon>0$ and $u_1\in\mathcal{U}$,

$$\begin{aligned}\sup_{t>0}\sum_{g=1}^{G}t^2P&\left(\sup_{u_2\in\mathcal{U}:\rho(u_1,u_2)\le\epsilon}|Z_{g,k,m}(u_2)-Z_{g,k,m}(u_1)|>t\right)\\&\le\sum_{g=1}^{G}E\left[\sup_{u_2\in\mathcal{U}:\rho(u_1,u_2)\le\epsilon}|Z_{g,k,m}(u_2)-Z_{g,k,m}(u_1)|^2\right]\\&=\frac{1}{G}\sum_{g=1}^{G}E\left[\sup_{u_2\in\mathcal{U}:\rho(u_1,u_2)\le\epsilon}|v_{g,k}(u_2)-v_{g,k}(u_1)|^2w_{g,m}^2\right]\le\epsilon^2\end{aligned}$$

where the second line follows from Markov's inequality, and the last inequality follows by selecting sufficiently large constant C in the definition of ρ and using the same argument as that in verification of condition (ii) since the first inequality in (37) used in the verification of condition (ii) can be replaced by

$$E_g\left[\sup_{u_2\in\mathcal{U}:\rho(u_1,u_2)\le\epsilon}\|\mathbb{G}^g(h_{\alpha_g(u_2),u_2}-h_{\delta_g(u_1),u_1})\|^4\right]\le c\epsilon^4$$

for arbitrarily small $c>0$ by selecting the constant C in the definition of $\rho(u_1,u_2)$ large enough and using Lemma 8. The claim of the lemma now follows by applying Theorem 14. □

Proofs of Theorems.

Proof of Theorem 1. The proof consists of two steps. First, we show that $\sqrt{G}(\hat\beta(u)-\widetilde\beta(u))=o_p(1)$ uniformly over $u\in\mathcal{U}$ where $\widetilde\beta(u)$ is defined in (23). Second, we show that $\sqrt{G}(\widetilde\beta(\cdot)-\beta(\cdot))\Rightarrow\mathbb{G}(\cdot)$ in $\ell^\infty(\mathcal{U})$. Combining these steps gives the result.

Step 1. Denote $\hat Q_{xw}=X'W/G$ and $\hat Q_{ww}=W'W/G$. Then

$$\sqrt{G}(\hat\beta(u)-\widetilde\beta(u))=\left(\hat Q_{xw}\hat Q_{ww}^{-1}\hat Q_{xw}'\right)^{-1}\hat Q_{xw}\hat Q_{ww}^{-1}\left(W'(\hat A(u)-A(u))/\sqrt{G}\right).$$

By Lemma 1, $X'W/G\to_p Q_{xw}$ and $W'W/G\to_p Q_{ww}$ where matrices Q_{xw} and Q_{ww} have singular values bounded in absolute values from above and away from zero by Assumption 2(ii), and so

$$\hat S=\left(\hat Q_{xw}\hat Q_{ww}^{-1}\hat Q_{xw}'\right)^{-1}\hat Q_{xw}\hat Q_{ww}^{-1}\to_p\left(Q_{xw}Q_{ww}Q_{xw}'\right)^{-1}Q_{xw}Q_{ww}^{-1}=S. \tag{38}$$

Therefore, to prove the first step, it suffices to show that

$$S(u)=\frac{1}{\sqrt{G}}\sum_{g=1}^{G}(\hat\alpha_g(u)-\alpha_g(u))w_g'=o_p(1)$$

uniformly over $u\in\mathcal{U}$. To this end, write $S(u)=S_1(u)+S_2(u)$ where

$$S_1(u)=-\frac{1}{\sqrt{G}}\sum_{g=1}^{G}J_g^{-1}(u)\mathbb{G}^g(h_{\alpha_g(u),u})w_g'/\sqrt{N_g},$$

$$S_2(u) = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \left(J_g^{-1}(u)\mathbb{G}^g(h_{\alpha_g(u),u}) + \sqrt{N_g}(\hat{\alpha}_g(u) - \alpha_g(u)) \right) w_g' / \sqrt{N_g}.$$

Since $N_G = \min_{g=1,\dots,G} N_g \to \infty$ by Assumption 3, Lemma 10 implies that $S_1(u) = o_p(1)$ uniformly over $u \in \mathcal{U}$.

Consider $S_2(u)$. Let

$$K_g = C\sqrt{N_g^{-1} \log N_g} \tag{39}$$

for sufficiently large constant $C > 0$ so that Theorem 5 implies that

$$P\left(\sup_{u \in \mathcal{U}} \|\hat{\alpha}_g(u) - \alpha_g(u)\| > K_g \right) \le C N_g^{-3}.$$

Let $\mathcal{D}_G$ be the event that

$$\max_{g=1,\dots,G} \sup_{u \in \mathcal{U}} \|\hat{\alpha}_g(u) - \alpha_g(u)\| \le K_g,$$

and let $\mathcal{D}_G^c$ be the event that $\mathcal{D}_G$ does not hold. By the union bound, $P(\mathcal{D}_G^c) \le CGN_g^{-3}$. By Assumption 3, $CGN_g^{-3} \to 0$. Therefore,

$$S_2(u) = S_2(u)1\{\mathcal{D}_G\} + S_2(u)1\{\mathcal{D}_G^c\} = S_2(u)1\{\mathcal{D}_G\} + o_p(1)$$

uniformly over $u \in \mathcal{U}$. Further, $\|S_2(u)\|1\{\mathcal{D}_G\} \le C\sum_{g=1}^{G}(r_{1,g} + r_{2,g} + r_{3,g})/\sqrt{GN_g}$ where

$$r_{1,g} = \sup_{u \in \mathcal{U}} \sup_{\alpha \in \mathbb{R}^{d_z} : \|\alpha - \alpha_g(u)\| \le K_g} \left\| J_g^{-1}(u)(\mathbb{G}^g(h_{\alpha,u}) - \mathbb{G}^g(h_{\alpha_g(u),u})) \right\| \|w_g\|,$$

$$r_{2,g} = \sup_{u \in \mathcal{U}} \left\| J_g^{-1}(u) \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N_g} h_{\hat{\alpha}_g(u),u}(z_{ig}, y_{ig}) \right\| \|w_g\|,$$

$$r_{3,g} = \sup_{u \in \mathcal{U}} \sup_{\alpha \in \mathbb{R}^{d_z} : \|\alpha - \alpha_g(u)\| \le K_g} \left\| E_g\left[\sqrt{N_g}(J_g^{-1}(u)h_{\alpha,u}(z_{ig}, y_{ig}) - (\alpha - \alpha_g(u))) \right] \right\| \|w_g\|.$$

We bound the three terms $r_{1,g}$, $r_{2,g}$, and $r_{3,g}$ in turn. By Lemma 4 and Hölder's inequality,

$$\begin{aligned} E[r_{1,g}] &\le \left(E[\|w_g\|^2]\right)^{1/2} \left(E\left[\sup_{u \in \mathcal{U}} \sup_{\alpha \in \mathbb{R}^{d_z} : \|\alpha - \alpha_g(u)\| \le K_g} \left\| \mathbb{G}^g(h_{\alpha,u}) - \mathbb{G}^g(h_{\alpha_g(u),u}) \right\|^2 \right] \right)^{1/2} \\ &\le C\left(\sqrt{\frac{\log N_g}{N_g}} \log N_g \right)^{1/2} = \frac{(\log N_g)^{3/4}}{N_g^{1/4}} \end{aligned}$$

where the second line follows from the definition of K_g, Assumption 2(iv), and Lemma 9. Further, using Lemma 4 again gives

$$\sup_{u \in \mathcal{U}} \left\| J_g^{-1}(u) \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N_g} h_{\hat{\alpha}_g(u),u}(z_{ig}, y_{ig}) \right\| \le C \sup_{u \in \mathcal{U}} \left\| \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N_g} h_{\hat{\alpha}_g(u),u}(z_{ig}, y_{ig}) \right\| \le \frac{C}{\sqrt{N_g}}$$

by the optimality of $\hat{\alpha}_g(u)$ and since y_{ig} has a continuous conditional distribution. Hence, $E[r_{2,g}] \leq C/\sqrt{N_g}$. Finally, by Lemmas 4 and 5,

$$E[r_{3,g}] \leq C\sqrt{N_g}K_g^2 \leq \frac{C\log N_g}{\sqrt{N_g}}.$$

Hence, by Assumption 3,

$$E\left[\sup_{u\in\mathcal{U}}\|S_2(u)\|1\{\mathcal{D}_G\}\right] \leq \frac{C\sqrt{G}(\log N_G)^{3/4}}{N_G^{3/4}} = o(1),$$

implying that $\sqrt{G}(\hat{\beta}(u) - \widetilde{\beta}(u)) = o_p(1)$ uniformly over $u \in \mathcal{U}$ and completing the first step.

Step 2. To prove that $\sqrt{G}(\widetilde{\beta}(\cdot) - \beta(\cdot)) \Rightarrow \mathbb{G}(\cdot)$ in $\ell^\infty(\mathcal{U})$, observe that

$$\sqrt{G}(\widetilde{\beta}(\cdot) - \beta(\cdot)) = \hat{S} \cdot \frac{1}{\sqrt{G}}\sum_{g=1}^{G} w_g\varepsilon_g(\cdot).$$

As explained in Step 1, $\hat{S} \to_p S$. Also, by Lemma 3,

$$\frac{1}{\sqrt{G}}\sum_{g=1}^{G} w_g\varepsilon_g(\cdot) \Rightarrow \mathbb{G}^0(\cdot),\ in\, \ell^\infty(\mathcal{U})$$

where $\mathbb{G}^0$ is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function $J(u_1, u_2)$. Therefore, by Slutsky's theorem,

$$\sqrt{G}(\widetilde{\beta}(\cdot) - \beta(\cdot)) \Rightarrow \mathbb{G}(\cdot),\ in\, \ell^\infty(\mathcal{U}) \tag{40}$$

where $\mathbb{G}$ is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function $\mathcal{C}(u_1, u_2) = SJ(u_1, u_2)S'$. Combining (40) with Step 1 gives the asserted claim and completes the proof of the theorem. □

Proof of Theorem 2. Equation (38) in the proof of Theorem 1 gives $\hat{S} \to_p S$. Therefore, it suffices to prove that $\|\hat{J}(u_1, u_2) - J(u_1, u_2)\| = o_p(1)$ uniformly over $u_1, u_2 \in \mathcal{U}$. Note that $\alpha_{g,1}(u) - x_g'\beta(u) = \varepsilon_g(u)$. Hence,

$$\begin{aligned}\hat{\alpha}_{g,1}(u) - x_g'\hat{\beta}(u) &= (\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)) - x_g'(\hat{\beta}(u) - \beta(u)) + \varepsilon_g(u)\\ &= I_{1,g}(u) - I_2(u) + \varepsilon_g(u)\end{aligned}$$

where $I_{1,g}(u) = \hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)$ and $I_2(u) = x_g'(\hat{\beta}(u) - \beta(u))$. Further, we have

$$\frac{1}{G}\sum_{g=1}^{G}\varepsilon_g(u_1)\varepsilon_g(u_2)w_g w_g' \to_p J(u_1, u_2)$$

uniformly over $u_1, u_2 \in \mathcal{U}$ by Lemma 2. In addition, it was demonstrated in the proof of Theorem 1 that

$$P\left(\max_{g=1,\ldots,G}\sup_{u\in\mathcal{U}}\|\hat{\alpha}_g(u) - \alpha_g(u)\| > K_g\right) \leq CGN_g^{-3} = o(1)$$

by Assumption 3 where $K_g = C(N_g^{-1}\log N_g)^{1/2}$ for sufficiently large constant C. Thus, setting $K_G = \max_{g=1,\dots,G} K_g$, we obtain

$$\begin{aligned}\left\|\frac{1}{G}\sum_{g=1}^{G} I_{1,g}(u_1)I_{1,g}(u_2)w_g w_g'\right\| &\leq \frac{K_G^2}{G}\sum_{g=1}^{G}\|w_g\|^2 + o_p(1)\\ &\leq O_p(K_G^2) + o_p(1) = o_p(1)\end{aligned}$$

uniformly over $u_1, u_2 \in \mathcal{U}$ by Assumption 2(iv) and Chebyshev's inequality. Further,

$$\begin{aligned}\left\|\frac{1}{G}\sum_{g=1}^{G} I_{1,g}(u_1)\varepsilon_g(u_2)w_g w_g'\right\| &\leq \frac{K_G}{G}\sum_{g=1}^{G}|\varepsilon_g(u_2)|\|w_g\|^2 + o_p(1)\\ &\leq \frac{K_G}{G}\sum_{g=1}^{G}\sup_{u\in\mathcal{U}}|\varepsilon_g(u)|\|w_g\|^2 + o_p(1) = o_p(1)\end{aligned}$$

uniformly over $u_1, u_2 \in \mathcal{U}$ by same argument as that used in the proof of Lemma 2 since Hölder's inequality implies that

$$E\left[\sup_{u\in\mathcal{U}}|\varepsilon_g(u)|\|w_g\|^2\right] \leq \left(E\left[\sup_{u\in\mathcal{U}}|\varepsilon_g(u)|^2\right]\right)^{1/2}\left(E[\|w_g\|^4]\right)^{1/2} \leq C$$

by Assumptions 2(iv) and 6(i). Similarly,

$$\begin{aligned}\left\|\frac{1}{G}\sum_{g=1}^{G} I_2(u_1)I_2(u_2)w_g w_g'\right\| &\leq \frac{C}{G}\sum_{g=1}^{G}\|w_g\|^2\sup_{u\in\mathcal{U}}\|\hat\beta(u)-\beta(u)\|^2 = o_p(1),\\ \left\|\frac{1}{G}\sum_{g=1}^{G} I_2(u_1)\varepsilon_g(u_2)w_g w_g'\right\| &\leq \frac{C}{G}\sum_{g=1}^{G}|\varepsilon_g(u_2)|\|w_g\|^2\sup_{u\in\mathcal{U}}\|\hat\beta(u)-\beta(u)\| = o_p(1)\end{aligned}$$

uniformly over $u_1, u_2 \in \mathcal{U}$ by Assumption 4(i). Finally,

$$\left\|\frac{1}{G}\sum_{g=1}^{G} I_{1,g}(u_1)I_{2,g}(u_2)w_g w_g'\right\| \leq \frac{CK_G}{G}\sum_{g=1}^{G}\|w_g\|^2\|\sup_{u\in\mathcal{U}}\|\hat\beta(u)-\beta(u)\| + o_p(1) = o_p(1)$$

uniformly over $u_1, u_2 \in \mathcal{U}$. Combining these inequalities gives the asserted claim. □

Proof of Theorem 3. Observe that the statement

$$\beta_1(u) \notin \left[\hat\beta_1(u) - \hat c_{1-\alpha}\sqrt{\frac{\hat V(u)}{G}}, \hat\beta_1(u) + \hat c_{1-\alpha}\sqrt{\frac{\hat V(u)}{G}}\right] \text{ for some } u \in \mathcal{U}$$

is equivalent to the statement that $T > \hat c_{1-\alpha}$. Therefore, it suffices to prove that

$$P(T > \hat c_{1-\alpha}) \to \alpha. \tag{41}$$

To prove (41), recall the process $\mathbb{G}(\cdot) = (\mathbb{G}_1(u), \dots, \mathbb{G}_{d_x}(u))'$ appearing in Theorem 1. Define a Gaussian process $\widetilde{\mathbb{G}}(\cdot)$ on $\mathcal{U}$ with values in $\mathbb{R}$ by

$$\widetilde{\mathbb{G}}(u) = V(u)^{-1/2}\mathbb{G}_1(u), \ \ u \in \mathcal{U}$$

where $V(u) = \mathcal{C}_{1,1}(u,u)$, the $(1,1)$st component of $\mathcal{C}(u,u) = SJ(u,u)S'$. It follows from conditions of the theorem that $V(u)$ is bounded away from zero uniformly over $u \in \mathcal{U}$. Therefore, since $\mathbb{G}(\cdot)$ has uniformly continuous sample paths, the process $\widetilde{\mathbb{G}}(\cdot)$ also has uniformly continuous sample paths. The covariance function of the process $\widetilde{\mathbb{G}}(\cdot)$ is

$$\widetilde{\mathcal{C}}(u_1,u_2) = V(u_1)^{-1/2}\mathcal{C}_{1,1}(u_1,u_2)V(u_2)^{-1/2}.$$

Further, for $G \geq 1$, define processes $\widehat{\mathbb{G}}_G(\cdot)$ and $\widetilde{\mathbb{G}}_G(\cdot)$ on $\mathcal{U}$ with values in $\mathbb{R}$ by

$$\widehat{\mathbb{G}}_G(u) = \frac{1}{\sqrt{G\hat{V}(u)}}\sum_{g=1}^{G}\left(\epsilon_g(\hat{\alpha}_{g,1}(u) - x_g'\hat{\beta}(u))\hat{w}_{g,1}^S\right), \ \ u \in \mathcal{U}$$

$$\widetilde{\mathbb{G}}_G(u) = \frac{1}{\sqrt{GV(u)}}\sum_{g=1}^{G}\epsilon_g\varepsilon_g(u)w_{g,1}^S, \ \ u \in \mathcal{U}$$

where $w_{g,1}^S$ and $\hat{w}_{g,1}^S$ are the 1st component of the vectors Sw_g and $\hat{S}w_g$, respectively, and $\hat{V}(u) = \hat{\mathcal{C}}_{1,1}(u,u)$.

Observe that $\hat{c}_{1-\alpha}$ is the $(1-\alpha)$ conditional quantile of $\sup_{u\in\mathcal{U}}|\hat{\mathbb{G}}_G(u)|$ given the data. Also, for $\beta \in (0,1)$ and $\mathcal{V} \subset \mathcal{U}$, let $c^0_{\beta,\mathcal{V}}$ be the βth quantile of $\sup_{u\in\mathcal{V}}|\widetilde{\mathbb{G}}(u)|$, and let $c_{\beta,\mathcal{V},G}$ be the βth quantile of $\sup_{u\in\mathcal{V}}|\widetilde{\mathbb{G}}_G(u)|$ given the data.

Now, since the process $\widetilde{\mathbb{G}}(\cdot)$ has uniformly continuous sample paths, it follows that $\sup_{u\in\mathcal{U}}|\widetilde{\mathbb{G}}(u)| < \infty$, and so Theorem 2.1 of Chernozhukov, Chetverikov, and Kato (2014b) implies that $\sup_{u\in\mathcal{U}}|\widetilde{\mathbb{G}}(u)|$ has continuous distribution. Therefore, for any $\delta > 0$, there exists $\eta > 0$ such that

$$P\left(\sup_{u\in\mathcal{U}}|\widetilde{\mathbb{G}}(u)| > c^0_{1-\alpha-\eta,\mathcal{U}} - \eta\right) \leq \alpha + \delta,$$

$$P\left(\sup_{u\in\mathcal{U}}|\widetilde{\mathbb{G}}(u)| > c^0_{1-\alpha+\eta,\mathcal{U}} + \eta\right) \geq \alpha - \delta.$$

In addition, Theorem 1 combined with continuous mapping theorem implies that $T \Rightarrow \sup_{u\in\mathcal{U}}|\widetilde{\mathbb{G}}(u)|$, and so

$$P(T > c^0_{1-\alpha-\eta,\mathcal{U}} - \eta) \leq \alpha + \delta + o(1),$$

$$P(T > c^0_{1-\alpha+\eta,\mathcal{U}} + \eta) \geq \alpha - \delta + o(1).$$

Hence, to prove (41), it suffices to show that for any $\eta > 0$,

$$P(c^0_{1-\alpha-\eta,\mathcal{U}} - \eta \leq \hat{c}_{1-\alpha} \leq c^0_{1-\alpha+\eta,\mathcal{U}} + \eta) \to 1. \tag{42}$$

To prove (42), fix some $\eta > 0$. Since $\widetilde{\mathbb{G}}(\cdot)$ has uniformly continuous sample paths, there exists a finite $\mathcal{U}(\eta,1) \subset \mathcal{U}$ such that

$$c^0_{1-\alpha-\eta,\mathcal{U}} - \eta \le c^0_{1-\alpha-\eta/2,\mathcal{U}(\eta,1)} - \eta/2, \tag{43}$$

$$c^0_{1-\alpha+\eta,\mathcal{U}} + \eta \ge c^0_{1-\alpha+\eta/2,\mathcal{U}(\eta,1)} + \eta/2. \tag{44}$$

Further, let $\mathcal{A}_G$ be the event that $G^{-1}\sum_{g=1}^G (w^S_{g,1})^2 \le C$ for some sufficiently large $C > 0$. Note that $P(\mathcal{A}_G) \to 1$ as $G \to \infty$. Also, on $\mathcal{A}_G$, for any $u_1, u_2 \in \mathcal{U}$,

$$E_\epsilon\Big[\Big(\frac{1}{\sqrt{G}}\sum_{g=1}^G \epsilon_g(\varepsilon_g(u_2) - \varepsilon_g(u_1))w^S_{g,1}\Big)^2\Big] = \frac{1}{G}\sum_{g=1}^G (\varepsilon_g(u_2) - \varepsilon_g(u_1))^2 (w^S_{g,1})^2 \le C|u_2 - u_1|^2$$

by Assumption 6(iii) where $E_\epsilon[\cdot]$ denotes expectation with respect to the distribution of $\epsilon_1, \dots, \epsilon_G$ (and keeping everything else fixed). Therefore, combining Borell's inequality (see Proposition of Van der Vaart and Wellner (1996)) and Corollary 2.2.8 of Van der Vaart and Wellner (1996) show that one can find finite $\mathcal{U}(\eta,2) \subset \mathcal{U}$ such that on $\mathcal{A}_G$,

$$c_{1-\alpha+\eta/2,\mathcal{U}(\eta,2),G} + \eta/3 \ge c_{1-\alpha+\eta/3,\mathcal{U},G} + \eta/4, \tag{45}$$

$$c_{1-\alpha-\eta/2,\mathcal{U}(\eta,2),G} - \eta/3 \le c_{1-\alpha-\eta/3,\mathcal{U},G} - \eta/4. \tag{46}$$

Now, observe that whenever the inequalities (43) - (46) are satisfied, the same inequalities are also satisfied with $\mathcal{U}(\eta,1)$ and $\mathcal{U}(\eta,2)$ replaced by $\mathcal{U}(\eta) = \mathcal{U}(\eta,1) \cup \mathcal{U}(\eta,2)$.

Next, conditional on the data, $(\widetilde{\mathbb{G}}_G(u))_{u\in\mathcal{U}(\eta)}$ is a zero-mean Gaussian vector with covariance function

$$\widetilde{\mathcal{C}}_G(u_1,u_2) = V(u_1)^{-1/2}\Big(\frac{1}{G}\sum_{g=1}^G \varepsilon_g(u_1)\varepsilon_g(u_2)(w^S_{g,1})^2\Big).$$

By Lemma 2, $\widetilde{\mathcal{C}}_G(u_1,u_2) \to_P \widetilde{C}(u_1,u_2)$ uniformly over $u_1,u_2 \in \mathcal{U}(\eta)$ where $\widetilde{C}(u_1,u_2)$ is the covariance function of a zero-mean Gaussian vector $(\widetilde{\mathbb{G}}(u))_{u\in\mathcal{U}(\eta)}$. Hence, by Lemma 3.1 of Chernozhukov, Chetverikov, and Kato (2013),

$$P(c^0_{1-\alpha+\eta/2,\mathcal{U}(\eta)} + \eta/2 > c_{1-\alpha+\eta/2,\mathcal{U}(\eta),G} + \eta/3) \to 1,$$
$$P(c^0_{1-\alpha-\eta/2,\mathcal{U}(\eta)} - \eta/2 < c_{1-\alpha-\eta/2,\mathcal{U}(\eta),G} - \eta/3) \to 1.$$

Combining this with inequalities (43) - (46) where we replace $\mathcal{U}(\eta,1)$ and $\mathcal{U}(\eta,2)$ by $\mathcal{U}(\eta)$ gives

$$P(c^0_{1-\alpha+\eta,\mathcal{U}} + \eta > c_{1-\alpha+\eta/3,\mathcal{U},G} + \eta/4) \to 1,$$
$$P(c^0_{1-\alpha-\eta,\mathcal{U}} - \eta < c_{1-\alpha-\eta/3,\mathcal{U},G} - \eta/4) \to 1.$$

To complete the proof, it suffices to show that

$$P(c_{1-\alpha-\eta/3,\mathcal{U},G} - \eta/4 \le \hat{c}_{1-\alpha} \le c_{1-\alpha+\eta/3,\mathcal{U}(\eta)} + \eta/4) \to 1. \tag{47}$$

To prove (47), observe that

$$\sup_{u\in\mathcal{U}}\left|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\epsilon_g x_g'(\hat\beta(u)-\beta(u))w_{g,1}^S\right|\leq\sup_{u\in\mathcal{U}}\|\hat\beta(u)-\beta(u)\|\cdot\left\|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\epsilon_g w_{g,1}^S x_g\right\|\to_P 0$$

since $\sup_{u\in\mathcal{U}}\|\hat\beta(u)-\beta(u)\|\to_P 0$ by Theorem 1 and $\|G^{-1/2}\sum_{g=1}^{G}\epsilon_g w_{g,1}^S x_g\|=O_P(1)$ by Assumptions 2(iv) and 4(i). Also,

$$\sup_{u\in\mathcal{U}}\left|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\epsilon_g(\hat\alpha_{g,1}(u)-\alpha_{g,1}(u))w_{g,1}^S\right|\to_P 0$$

by the same argument as that used in Step 1 of the proof of Theorem 1. Therefore, since $\varepsilon_g(u)=\alpha_{g,1}(u)-x_g'\beta(u)$, $\sup_{u\in\mathcal{U}}|\hat V(u)-V(u)|\to_P 0$ by Theorem 2, $V(u)$ is bounded away from zero uniformly over $u\in\mathcal{U}$, and $\hat S\to_P S$ as in the proof of Theorem 1, we obtain

$$\sup_{u\in\mathcal{U}}\|\widetilde{\mathbb{G}}_G(u)-\widehat{\mathbb{G}}_G(u)\|\to_p 0.$$

Since $\hat c_{1-\alpha}$ is the $(1-\alpha)$ conditional quantile of $\sup_{u\in\mathcal{U}}|\widehat{\mathbb{G}}(u)|$ given the data and $c_{\beta,\mathcal{U},G}$ is the βth conditional quantile of $\sup_{u\in\mathcal{U}}|\widetilde{\mathbb{G}}(u)|$ given the data, (47) follows. This completes the proof of the theorem. $\square$

Proof of Theorem 4. We split the proof into two steps.

Step 1. Here we wish to show that for sufficiently large $C>0$,

$$P\left(\max_{1\leq g\leq G}\left\|J_g^{-1}(u)\mathbb{G}^g(h_{\alpha_g(u),u})+\sqrt{N_g}(\hat\alpha_g-\alpha_g)\right\|>\frac{C(\log N_G)^{3/4}}{N_G^{1/4}}\right)\to 0 \tag{48}$$

Set $K_g=C(N_g^{-1}\log N_g)^{1/2}$ for sufficiently large $C>0$ so that Theorem 5 implies that

$$P\left(\|\hat\alpha_g(u)-\alpha_g(u)\|>K_g\right)\leq CN_g^{-3}.$$

Let $\mathcal{D}_G$ be the event that

$$\max_{1\leq g\leq G}\|\hat\alpha_g(u)-\alpha_g(u)\|\leq K_g$$

and let $\mathcal{D}_G^c$ be the event that $\mathcal{D}_G$ does not hold. By the union bound, $P(\mathcal{D}_G^c)\leq CGN_g^{-3}\to 0$.

Now, on the event $\mathcal{D}_G$,

$$\left\|J_g^{-1}(u)\mathbb{G}^g(h_{\alpha_g(u),u})+\sqrt{N_g}(\hat\alpha_g-\alpha_g)\right\|\leq r_{1,g}+r_{2,g}+r_{3,g}$$

where

$$r_{1,g}=\sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\leq K_g}\|J_g^{-1}(u)(\mathbb{G}^g(h_{\alpha,u})-\mathbb{G}^g(h_{\alpha_g(u),u}))\|,$$

$$r_{2,g}=\left\|J_g^{-1}(u)\frac{1}{\sqrt{N_g}}\sum_{i=1}^{N_g}h_{\hat\alpha_g(u),u}(z_{ig},y_{ig})\right\|,$$

$$r_{3,g} = \sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\leq K_g} \|E_g[\sqrt{N_g}(J_g^{-1}(u)h_{\alpha,u}(z_{ig},y_{ig})-(\alpha-\alpha_g(u)))]\|.$$

By Lemma 4 and optimality of $\hat{\alpha}_g(u)$,

$$r_{2,g} \leq \left\| \frac{C}{\sqrt{N_g}} \sum_{i=1}^{N_g} h_{\hat{\alpha}_g(u),u}(z_{ig},y_{ig}) \right\| \leq \frac{C}{\sqrt{N_g}}.$$

Also, by Lemmas 4 and 5,

$$r_{3,g} \leq C\sqrt{N_g}K_g^2 \leq \frac{C\log N_g}{\sqrt{N_g}}.$$

Finally, by Lemma 4 and Talagrand's inequality (see, for example, Theorem B.1 in Chernozhukov, Chetverikov, and Kato (2014b)),

$$r_{1,g} \leq C\sup \sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\leq K_g} \|\mathbb{G}^g(h_{\alpha,u}) - \mathbb{G}^g(h_{\alpha_g(u),u})\| \leq C\sqrt{K_g\log G} = \frac{C\log^{3/4} N_g}{N_g^{1/4}}$$

with probability at least $1-G^{-2}$. Combining these bounds gives (48) and completes this step.

Step 2. Here we complete the proof. For $g=1,\dots,G$ and $i=1,\dots,\bar{N}_G$, define q_{ig} as follows. If $i>N_g$, set $q_{ig}=0$. If $i\leq N_g$, set

$$q_{ig} = (\bar{N}_G/N_g)^{1/2} I_g^{-1/2}\bar{z}_{ig}(1\{y_{ig}\leq z_{ig}'\alpha_g(u)\}-u)$$

where $\bar{z}_{ig}$ denotes the first component of the vector $J_g^{-1}(u)z_{ig}$. By Step 1 and assumptions that $I_g\geq c_M$ and $\bar{N}_G/N_G\leq C_M$, it follows that

$$\begin{aligned} &P\left(\max_{1\leq g\leq G}\sqrt{N_g/I_g}|\hat{\alpha}_{g,1}(u)-\alpha_{g,1}(u)|\leq c_{1-\alpha}^M\right) \\ &\quad\leq P\left(\max_{1\leq g\leq G}\left|\frac{1}{\sqrt{\bar{N}_G}}\sum_{g=1}^{\bar{N}_G}(q_{ig}-E_g[q_{ig}])\right|\leq c_{1-\alpha}^M + \frac{C\log^{3/4}N_g}{N_g^{1/4}}\right)+o(1) \end{aligned} \tag{49}$$

In turn, since under our assumptions $|q_{ig}|\leq C$, by Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2014d), the probability in (49) is bounded from above by

$$\begin{aligned} &P\left(\max_{1\leq g\leq G}|Y_g|\leq c_{1-\alpha}^M+\frac{C\log^{3/4}N_G}{N_G^{1/4}}\right)+o(1) \\ &\quad\leq P\left(\max_{1\leq g\leq G}|Y_g|\leq c_{1-\alpha}^M\right)+\frac{C(\log^{3/4}N_G)\cdot(\log^{1/2}G)}{N_G^{1/4}}+o(1)=1-\alpha+o(1) \end{aligned}$$

where in the second line we used Theorem 3 in Chernozhukov, Chetverikov, and Kato (2014c). Thus,

$$P\left(\max_{1\leq g\leq G}\sqrt{N_g/I_g}|\hat{\alpha}_{g,1}(u)-\alpha_{g,1}(u)|\leq c_{1-\alpha}^M\right)\leq 1-\alpha+o(1). \tag{50}$$

Similar arguments also give

$$P\left(\max_{1\leq g\leq G}\sqrt{N_g/I_g}|\hat{\alpha}_{g,1}(u)-\alpha_{g,1}(u)|\leq c_{1-\alpha}^M\right)\geq 1-\alpha-o(1). \tag{51}$$

Rearranging the terms under the probability signs in (50) and (51) completes the proof of the theorem. □

Proof of Theorem 5. Recall the definition of the function $f_{\eta,\alpha,u}$ in (24). Since $x \mapsto \rho_u(x) = (u - I\{x < 0\})x$ is convex, for $x > 0$, $\|\hat{\alpha}_g(u) - \alpha_g(u)\| \leq x$ for all $u \in \mathcal{U}$ if

$$\inf_{u\in\mathcal{U}} \inf_{\eta\in\mathbb{R}^{d_z};\|\eta\|=1} \sum_{i=1}^{N_g} f_{\eta,\alpha_g(u)+x\eta,u}(z_{ig}, y_{ig})/N_g > 0. \tag{52}$$

Now, since $f_{\eta,\alpha,u} = \eta' h_{\alpha,u}$, Lemma 5 implies that

$$\inf_{u\in\mathcal{U}} \inf_{\eta\in\mathbb{R}^{d_z};\|\eta\|=1} E_g[f_{\eta,\alpha_g(u)+x\eta,u}(z_{ig}, y_{ig})] > cx$$

if the constant $\bar{c}$ in the statement of the theorem is sufficiently small. Therefore, it follows that (52) holds if

$$\inf_{u\in\mathcal{U}} \inf_{\eta\in\mathbb{R}^{d_z};\|\eta\|=1} \sum_{i=1}^{N_g} \left(f_{\eta,\alpha_g(u)+x\eta,u}(z_{ig}, y_{ig}) - E_g[f_{\eta,\alpha_g(u)+x\eta,u}(z_{ig}, y_{ig})]\right)/N_g \geq -cx,$$

which in turn follows if

$$\inf_{u\in\mathcal{U}} \inf_{\eta,\alpha\in\mathbb{R}^{d_z};\|\eta\|=1} \mathbb{G}^g(f_{\eta,\alpha,u}) \geq -cx\sqrt{N_g}. \tag{53}$$

Note that for any $\eta \in \mathbb{R}^{d_z}$ satisfying $\|\eta\| = 1$, $|f_{\eta,\alpha,u}| \leq 2\|z_{ig}\| \leq C$ for some $C > 0$ by Assumption 4(i). In addition, it follows from Lemma 6 and Theorem 9 that the conditions of Theorem 12 hold for the function class $\{f_{\eta,\alpha,u} \in \mathcal{F} : u \in \mathcal{U}; \eta, \alpha \in \mathbb{R}^{d_z}; \|\eta\| = 1\}$. Therefore, Theorem 12 shows that (53) holds with probability not smaller than

$$1 - C\exp(-cx^2 N_g)$$

for some $c, C > 0$. The asserted claim follows. □

Appendix H. Proofs of Theorems 6-8

The proofs are analogous to those of Theorems 1-3. Therefore, we only discuss important differences. First, the constants $c, C > 0$ in the proofs now depend on c_M, c_f, C_M, C_f, C_L, and $\bar{C}$. Second, among Lemmas 1 - 10, Lemmas 4 - 9 deal with within group variation, and so apply under our conditions without changes. The statement of Lemma 1 holds without changes but in the proof, Chebyshev's inequality applies on cluster level, that is, for $k = 1, \dots, d_x$ and $l = 1, \dots, d_w$,

$$\begin{aligned} E\Big[\Big(\frac{1}{G}\sum_{g=1}^{G}(x_{g,k}w_{g,l} - E[x_{g,k}w_{g,l}])\Big)^2\Big] &= \frac{1}{G^2}\sum_{m=1}^{M} E\Big[\Big(\sum_{g\in\mathbb{C}_G(m)}(x_{g,k}w_{g,l} - E[x_{g,k}w_{g,l}])\Big)^2\Big] \\ &\leq \frac{C}{G^2}\sum_{m=1}^{M} E\Big[\sum_{g\in\mathbb{C}_G(m)}(x_{g,k}w_{g,l} - E[x_{g,k}w_{g,l}])^2\Big] \end{aligned}$$

$$= \frac{C}{G^2} \sum_{g=1}^{G} E[(x_{g,k} w_{g,l} - E[x_{g,k} w_{g,l}])^2] \to 0$$

where in the second line we used Assumption 1′(iii) that the number of groups in each cluster is bounded from above by $\bar{C}$.

Lemma 2 should be replaced with the statement that $G \to \infty$,

$$\frac{1}{G} \sum_{m=1}^{M} \Big(\sum_{g \in \mathbb{C}_G(m)} \varepsilon_g(u_1) w_g \Big) \Big(\sum_{g \in \mathbb{C}_G(m)} \varepsilon_g(u_1) w_g' \Big) \to_P J^{CS}(u_1, u_2) \tag{54}$$

uniformly over $u_1, u_2 \in \mathcal{U}$. To prove this statement, observe that by Assumption 6′(ii),

$$\frac{1}{G} \sum_{m=1}^{M} E\Big[\Big(\sum_{g \in \mathbb{C}_G(m)} \varepsilon_g(u_1) w_g \Big) \Big(\sum_{g \in \mathbb{C}_G(m)} \varepsilon_g(u_1) w_g' \Big) \Big] \to J^{CS}(u_1, u_2)$$

uniformly over $u_1, u_2 \in \mathcal{U}$. Further, for $\delta = c_M/4$ and $k, l = 1, \dots, d_w$,

$$\begin{aligned} E\Big[\Big| \Big(\sum_{g \in \mathbb{C}_G(m)} \varepsilon_g(u_1) w_{g,k} \Big) \Big(\sum_{g \in \mathbb{C}_G(m)} \varepsilon_g(u_2) w_{g,l} \Big) \Big|^{1+\delta} \Big] \\ \le CE\Big[\sum_{g,g' \in \mathbb{C}_G(m)} |\varepsilon_g(u_1) w_{g,k} \varepsilon_{g'}(u_2) w_{g',l}|^{1+\delta} \Big] \\ \le CE\Big[\sum_{g,g' \in \mathbb{C}_G(m)} \Big(|\varepsilon_g(u_1) w_{g,k}|^{2+2\delta} + |\varepsilon_{g'}(u_2) w_{g',l}|^{2+2\delta} \Big) \Big] \le C, \end{aligned}$$

where the last inequality can be proven by the same argument as that used in the proof of Lemma 2. From this point, the proof of 54 is analogous to the proof used in Lemma 2.

The statement of Lemma 3 holds with $J(u_1, u_2)$ replaced by $J^{CS}(u_1, u_2)$. To prove the new statement, first observe that for any finite $\mathcal{U}' \subset \mathcal{U}$,

$$\Big(\frac{1}{\sqrt{G}} \sum_{g=1}^{G} w_g \varepsilon_g(u) \Big)_{u \in \mathcal{U}'} \Rightarrow (N(u))_{u \in \mathcal{U}'}$$

where $(N(u))_{u \in \mathcal{U}'}$ is a zero-mean Gaussian vector with covariance function $J^{CS}(u_1, u_2)$ for all $u_1, u_2 \in \mathcal{U}'$. The rest of the proof follows from Theorem 14 by the same arguments as those used in Lemma 3 and those explained above where we replace $Z_g(u) = G^{-1/2} w_{g,k} \varepsilon_g(u)$ by $Z_m(u) = G^{-1/2} \sum_{g \in \mathbb{C}_G(m)} w_{g,k} \varepsilon_g(u)$, and we replace sums over $g = 1, \dots, G$ by sums over $m = 1, \dots, M$ where appropriate.

The statement of Lemma 10 holds without changes but in the proof, we replace $Z_{g,k,l}(u) = v_{g,k}(u) w_{g,l}/\sqrt{G}$ by $Z_{m,k,l}(u) = \sum_{g \in \mathbb{C}_G(m)} v_{g,k}(u) w_{g,l}/\sqrt{G}$ and we replace sums over $g = 1, \dots, G$ by sums over $m = 1, \dots, M$ where appropriate, and employ the arguments explained above.

With the new versions of Lemmas 1 - 10, the proof of Theorem 6 is the same as the proof of Theorem 1. The proof of Theorem 7 is analogous to that of Theorem 2 where, using the same notation as that in the proof of Theorem 2, we employ the bound

$$\begin{aligned}&\Big\|\frac{1}{G}\sum_{m=1}^{M}\Big(\sum_{g\in\mathbb{C}_G(m)} I_{1,g}(u_1)w_g\Big)\Big(\sum_{g\in\mathbb{C}_G(m)} I_{1,g}(u_2)w_g'\Big)\Big\|\\ &\quad\le \frac{1}{G}\sum_{m=1}^{M}\sum_{g,g'\in\mathbb{C}_G(m)}\|I_{1,g}(u_1)I_{1,g'}(u_2)w_g w_g'\|\le \frac{K_g^2}{G}\sum_{g=1}^{G}\|w_g\|^2+o_P(1)=o_P(1),\end{aligned}$$

and we bound all other terms in the proof similarly. The proof of Theorem 8 is analogous to that of Theorem 3.

Appendix I. Tools

In Appendix G, we used several results from the empirical process theory. For ease of reference, we describe these results in this section.

Let (T,ρ) be a semi-metric space. For $\varepsilon > 0$, an ε-net of (T,ρ) is a subset T_ε of T such that for every $t\in T$, there exists a point $t_\varepsilon\in T_\varepsilon$ with $\rho(t,t_\varepsilon)<\varepsilon$. The ε-covering number $N(\varepsilon,T,\rho)$ of T is the infimum of the cardinality of ε-nets of T, that is, $N(\varepsilon,T,\rho)=\inf\{\mathrm{Card}(T_\varepsilon): T_\varepsilon$ is an ε net of $T\}$.

Let $\mathcal{F}$ be a class of measurable functions defined on some measurable space $(S,\mathcal{S})$. For any probability measure Q on $(S,\mathcal{S})$ and $p\ge 1$, let $L_p(Q)$ denote the space of functions f on S with the norm $\|f\|_{Q,p}=(\int|f(s)|^p dQ(s))^{1/p}<\infty$. The function class $\mathcal{F}$ is called VC-subgraph class if the collection of all subgraphs of the functions in $\mathcal{F}$ forms a VC-class of sets; see Section 2.6.2 of Van der Vaart and Wellner (1996) for the definitions. In addition, we say that the function class $\mathcal{F}$ is VC type class of functions with an envelope $F:S\to\mathbb{R}_+$ and constants $A\ge e$, and $v\ge 1$ if all functions in $\mathcal{F}$ are bounded in absolute value by F and the following condition holds:

$$\sup_Q N(\varepsilon\|F\|_{Q,2},\mathcal{F},L_2(Q))\le (A/\varepsilon)^v$$

for all $\varepsilon\in(0,1)$ where the supremum is taken over all finitely discrete probability measures Q on $(S,\mathcal{S})$.

Finally, let $X_1,\dots,X_n$ be an i.i.d. sequence of random variables taking values in $(S,\mathcal{S})$ with a common distribution P. Define the empirical process:

$$\mathbb{G}_n(f)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(f(X_i)-E[f(X_i)]\Big),\ \ f\in\mathcal{F}.$$

The following theorems are used in Appendix G:

Theorem 9. *There exists a universal constant K such that for any VC subgraph class $\mathcal{F}$ of functions with an envelope F, any $p\ge 1$, and $0<\varepsilon<1$,*

$$\sup_Q N(\varepsilon\|F\|_{Q,p},\mathcal{F},L_p(Q))\le KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\varepsilon}\right)^{r(V(\mathcal{F})-1)}$$

where $V(\mathcal{F})$ is a finite constant that depends only on the function class $\mathcal{F}$ (and called VC dimension of the class $\mathcal{F}$). Thus, any VC-subgraph class of functions $\mathcal{F}$ is also a VC type class of functions with some constants $A \geq e$ and $v \geq 1$ depending only on $\mathcal{F}$.

Proof. See Lemma 19.15 in Van der Vaart (1998). □

Theorem 10. *Let $\mathcal{F}_1, \ldots, \mathcal{F}_k$ be classes of measurable functions $S \to \mathbb{R}$ to which measurable envelopes $F_1, \ldots, F_k$ are attached, respectively, and let $\phi : \mathbb{R}^k \to \mathbb{R}$ be a map that is Lipschitz in the sense that*

$$|\phi \circ f(s) - \phi \circ g(s)|^2 \leq \sum_{j=1}^{k} L_j^2(s)|f_j(s) - g_j(s)|^2,$$

for every $f = (f_1, \ldots, f_k)$, $g = (g_1, \ldots, g_k) \in \mathcal{F}_1 \times \ldots \mathcal{F}_k = \mathcal{F}$ and every $s \in S$, where $L_1, \ldots, L_k$ are non-negative measurable functions on S. Consider the class of functions $\phi(\mathcal{F}) = \{\phi \circ f : f \in \mathcal{F}\}$. Denote $(\sum_{j=1}^{k} L_j^2 F_j^2)^{1/2}$ by $L \cdot F$. Then we have

$$\sup_Q N(\varepsilon \|L \cdot F\|_{Q,2}, \phi(\mathcal{F}), L_2(Q)) \leq \prod_{j=1}^{k} \sup_{Q_j} N(\varepsilon \|F_j\|_{Q_j,2}, \mathcal{F}_j, L_2(Q_j))$$

for every $0 < \varepsilon < 1$.

Proof. See Lemma A.6 in Chernozhukov, Chetverikov, and Kato (2014a). □

Theorem 11. *Let $\mathcal{F}$ be a VC type class of functions with an envelope F and constants $A \geq e$ and $v \geq 1$. Denote $\sigma^2 = \sup_{f \in \mathcal{F}} E[f(X_1)^2]$ and $M = \max_{1 \leq i \leq n} F(X_i)$. Then*

$$E\left[\sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)|\right] \leq K\left(\sqrt{v\sigma^2 \log\left(\frac{A\|F\|_{P,2}}{\sigma}\right)} + \frac{v\|M\|_2}{\sqrt{n}} \log\left(\frac{A\|F\|_{P,2}}{\sigma}\right)\right)$$

for some absolute constant K where $\|M\|_2 = (E[M^2])^{1/2}$.

Proof. See Corollary 5.1 of Chernozhukov, Chetverikov, and Kato (2014a). □

Theorem 12. *Let $\mathcal{F}$ be a class of functions $f : \mathcal{X} \to [0, 1]$ that satisfies*

$$\sup_Q N(\varepsilon, \mathcal{C}, L_2(Q)) \leq \left(\frac{K}{\varepsilon}\right)^V, \text{ for every } 0 < \varepsilon < K$$

where supremum is taken over all probability measures Q. Then for every $t > 0$,

$$P\left(\sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)| > t\right) \leq \left(\frac{Dt}{\sqrt{V}}\right)^V e^{-2t^2}$$

for a constant D that depends on K only.

Proof. See Theorem 2.14.9 in Van der Vaart and Wellner (1996). □

Theorem 13. *Let $X_1, \ldots, X_n$ be independent, zero-mean stochastic processes indexed by an arbitrary index set T with joint probability measure P. Then*

$$\Big\| \|S_n\| \Big\|_{P,p} \leq K \frac{p}{\log p} \left(\Big\| \|S_n\| \Big\|_{P,1} + \Big\| \max_{1 \leq i \leq n} \|X_i\| \Big\|_{P,p} \right)$$

for any $p > 1$ where $S_n = X_1 + \cdots + X_n$, $\|S_n\| = \sup_{t \in T} |S_n(t)|$, $\|X_i\| = \sup_{t \in T} |X_i(t)|$, and K is a universal constant.

Proof. See Proposition A.1.6 in Van der Vaart and Wellner (1996). □

Finally, we provide a reference for Central Limit Theorem with bracketing by Gaussian hypotheses, which we use several times in Section G. A semi-metric $\rho : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_+$ is called Gaussian if it can be defined as

$$\rho(f, g) = \big(E[(G(f) - G(g))^2] \big)^{1/2}$$

where G is a tight, zero-mean, Gaussian random element in $l^\infty(\mathcal{F})$. A semi-metric ρ is called Gaussian-dominated if it is bounded from above by Gaussian metric. In particular, it is known that any semi-metric ρ satisfying

$$\int_0^\infty \sqrt{\log N(\varepsilon, \mathcal{F}, \rho)} d\varepsilon < \infty$$

is Gaussian-dominated; see discussion on page 212 in Van der Vaart and Wellner (1996).

Theorem 14 (Bracketing by Gaussian Hypotheses)**.** *For each n, let $Z_{n1}, ..., Z_{nm_n}$ be independent stochastic processes indexed by an arbitrary index set $\mathcal{F}$. Suppose that there exists a Gaussian-dominated semi-metric ρ on $\mathcal{F}$ such that*

$$(i) \qquad \sum_{i=1}^{m_n} E\left[\|Z_{ni}\|_{\mathcal{F}} \cdot 1\{\|Z_{ni}\|_{\mathcal{F}} > \eta\}\right] \to 0, \text{ for every } \eta > 0,$$

$$(ii) \qquad \sum_{i=1}^{m_n} E\left[(Z_{ni}(f) - Z_{ni}(g))^2\right] \leq \rho^2(f, g), \text{ for every } f, g,$$

$$(iii) \qquad \sup_{t>0} \sum_{i=1}^{m_n} t^2 P\left(\sup_{f,g \in B(\varepsilon)} |Z_{ni}(f) - Z_{ni}(g)| > t \right) \leq \varepsilon^2,$$

for every ρ-ball $B(\varepsilon) \subset \mathcal{F}$ of radius less than ε and for every n. Then the sequence $\sum_{i=1}^{m_n}(Z_{ni} - E[Z_{ni}])$ is asymptotically tight in $l^\infty(\mathcal{F})$. It converges in distribution provided it converges marginally.

Proof. See Theorem 2.11.11 in Van der Vaart and Wellner (1996). □

Table A1. Bias of Grouped IV Quantile Regression vs. Standard Quantile Regression

		(N,G) = (25, 25)		(N,G) = (200, 25)		(N,G) = (25, 200)		(N,G) = (200, 200)	
I. Mean Bias for Endogenous Group-level Treatment									
Quantile (u)	True Coeff.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.
0.1	0.316	0.042	-0.055	0.040	-0.007	0.038	0.018	0.039	-0.005
0.2	0.447	0.076	0.015	0.078	-0.003	0.077	0.008	0.077	0.000
0.3	0.548	0.116	-0.024	0.116	-0.044	0.117	0.005	0.116	-0.003
0.4	0.632	0.155	-0.128	0.154	-0.031	0.154	0.007	0.155	-0.002
0.5	0.707	0.194	-0.182	0.193	-0.023	0.192	0.010	0.194	-0.006
0.6	0.775	0.236	-0.192	0.233	-0.039	0.228	0.003	0.232	-0.006
0.7	0.837	0.273	-0.161	0.270	-0.067	0.267	-0.002	0.270	-0.004
0.8	0.894	0.312	-0.106	0.311	-0.056	0.306	-0.010	0.309	-0.003
0.9	0.949	0.365	-0.106	0.361	-0.060	0.360	-0.013	0.362	-0.001
Avg. abs. bias		**0.197**	**0.108**	**0.195**	**0.037**	**0.193**	**0.008**	**0.195**	**0.003**
II. Mean Bias for Exogenous Group-level Treatment									
Quantile (u)	True Coeff.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.
0.1	0.316	0.005	0.010	-0.004	-0.016	0.002	-0.011	0.001	-0.006
0.2	0.447	0.005	0.027	0.001	-0.010	0.002	-0.018	0.003	-0.008
0.3	0.548	0.006	-0.006	0.006	-0.012	0.003	-0.017	0.005	-0.005
0.4	0.632	0.011	-0.021	0.007	-0.010	0.005	-0.017	0.007	0.002
0.5	0.707	0.008	-0.039	0.008	-0.002	0.007	-0.020	0.009	0.003
0.6	0.775	0.004	-0.021	0.009	-0.004	0.009	-0.015	0.011	0.002
0.7	0.837	0.006	-0.011	0.007	-0.003	0.009	-0.014	0.011	0.000
0.8	0.894	-0.010	-0.007	-0.011	-0.001	-0.011	-0.008	-0.011	0.000
0.9	0.949	-0.031	0.008	-0.038	0.003	-0.028	-0.009	-0.031	-0.001
Avg. abs. bias		**0.010**	**0.017**	**0.010**	**0.007**	**0.009**	**0.014**	**0.010**	**0.003**
III. Mean Bias for Exogenous Group-level Treatment and No Group-level Unobservables									
Quantile (u)	True Coeff.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.	Q. Reg.	Grouped IV Q. Reg.
0.1	0.316	0.002	0.019	0.001	-0.006	0.000	-0.009	0.000	-0.004
0.2	0.447	0.008	0.009	0.003	-0.002	0.000	-0.008	-0.001	-0.007
0.3	0.548	0.005	-0.023	0.004	0.000	0.001	-0.010	-0.001	-0.007
0.4	0.632	0.007	-0.015	0.004	-0.003	0.002	-0.001	0.000	-0.005
0.5	0.707	0.005	-0.027	0.000	-0.003	0.001	-0.002	0.000	-0.004
0.6	0.775	0.004	-0.037	0.001	-0.011	0.000	-0.002	0.000	-0.002
0.7	0.837	0.003	-0.027	0.000	-0.005	0.000	-0.002	0.000	0.000
0.8	0.894	0.000	-0.022	0.000	-0.003	0.001	0.000	0.000	0.002
0.9	0.949	-0.003	-0.023	0.000	-0.003	-0.001	-0.005	0.000	0.001
Avg. abs. bias		**0.004**	**0.023**	**0.002**	**0.004**	**0.001**	**0.004**	**0.000**	**0.004**

Notes: Table shows mean bias for estimation of $\beta(u)$ from 1,000 Monte Carlo simulations using standard quantile regression (Q. Reg.) and our estimator (Grouped IV Q. Reg.) for cases where $(N, G) =$ (25,25), (200,25), (25,200), (200,200). Panel I displays results when the group-level treatment is endogenous, panel II displays results when the group-level treatment is independent of group-level unobservables, and panel III displays results when there are no group-level unobservables. Each panel displays results for quantiles $u \in \{0.1, ..., 0.9\}$ as well as the average absolute value of the bias, averaged over the nine deciles.